Amazon SageMaker

Amazon Sagemaker API Reference
Amazon SageMaker: Amazon Sagemaker API Reference
Copyright © 2023 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Amazon's trademarks and trade dress may not be used in connection with any product or service that is not Amazon's, in any manner that is likely to cause confusion among customers, or in any manner that disparages or discredits Amazon. All other trademarks not owned by Amazon are the property of their respective owners, who may or may not be affiliated with, connected to, or sponsored by Amazon.
# Table of Contents

Welcome .................................................................................................................. 1
Amazon SageMaker Service .................................................................................... 1
Amazon SageMaker Runtime .................................................................................. 1
Amazon Sagemaker Edge Manager .................................................................... 1
Amazon SageMaker Feature Store Runtime .......................................................... 1
Amazon SageMaker geospatial capabilities ......................................................... 1
Amazon SageMaker Metrics Service .................................................................. 1

Actions .................................................................................................................. 2
Amazon SageMaker Service .................................................................................. 10
   AddAssociation ................................................................................................. 18
   AddTags .............................................................................................................. 21
   AssociateTrialComponent ............................................................................... 23
   BatchDescribeModelPackage ......................................................................... 25
   CreateAction ..................................................................................................... 28
   CreateAlgorithm ............................................................................................... 31
   CreateApp .......................................................................................................... 37
   CreateAppImageConfig .................................................................................... 40
   CreateArtifact ................................................................................................... 43
   CreateAutoMLJob .............................................................................................. 46
   CreateAutoMLJobV2 ......................................................................................... 51
   CreateCluster .................................................................................................... 56
   CreateCodeRepository ....................................................................................... 59
   CreateCompilationJob ...................................................................................... 61
   CreateContext ................................................................................................... 65
   CreateDataQualityJobDefinition ...................................................................... 68
   CreateDeviceFleet .............................................................................................. 73
   CreateDomain .................................................................................................... 76
   CreateEdgeDeploymentPlan ............................................................................. 84
   CreateEdgeDeploymentStage ......................................................................... 87
   CreateEdgePackagingJob ............................................................................... 89
   CreateEndpoint .................................................................................................. 92
   CreateEndpointConfig ..................................................................................... 96
   CreateExperiment .............................................................................................. 102
   CreateFeatureGroup ....................................................................................... 105
   CreateFlowDefinition ..................................................................................... 110
   CreateHub ......................................................................................................... 113
   CreateHumanTaskUi .......................................................................................... 116
   CreateHyperParameterTuningJob ................................................................... 118
   CreateImage ...................................................................................................... 127
   CreateImageVersion ......................................................................................... 130
   CreateInferenceComponent .......................................................................... 134
   CreateInferenceExperiment .......................................................................... 137
   CreateInferenceRecommendationsJob ............................................................ 142
   CreateLabelingJob ........................................................................................... 147
   CreateModel ..................................................................................................... 153
   CreateModelBiasJobDefinition ...................................................................... 158
   CreateModelCard ............................................................................................. 163
   CreateModelCardExportJob .......................................................................... 166
   CreateModelExplainabilityJobDefinition ....................................................... 169
   CreateModelPackage ....................................................................................... 174
   CreateModelPackageGroup ............................................................................. 183
   CreateModelQualityJobDefinition .................................................................. 185
   CreateMonitoringSchedule ............................................................................ 190
   CreateNotebookInstance ................................................................................. 194
<table>
<thead>
<tr>
<th>Function Name</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>CreateNotebookInstanceLifecycleConfig</td>
<td>200</td>
</tr>
<tr>
<td>CreatePipeline</td>
<td>203</td>
</tr>
<tr>
<td>CreatePresignedDomainUrl</td>
<td>207</td>
</tr>
<tr>
<td>CreatePresignedNotebookInstanceUrl</td>
<td>210</td>
</tr>
<tr>
<td>CreateProcessingJob</td>
<td>212</td>
</tr>
<tr>
<td>CreateProject</td>
<td>217</td>
</tr>
<tr>
<td>CreateSpace</td>
<td>220</td>
</tr>
<tr>
<td>CreateStudioLifecycleConfig</td>
<td>224</td>
</tr>
<tr>
<td>CreateTrainingJob</td>
<td>227</td>
</tr>
<tr>
<td>CreateTransformJob</td>
<td>237</td>
</tr>
<tr>
<td>CreateTrial</td>
<td>243</td>
</tr>
<tr>
<td>CreateTrialComponent</td>
<td>246</td>
</tr>
<tr>
<td>CreateUserProfile</td>
<td>250</td>
</tr>
<tr>
<td>CreateWorkforce</td>
<td>255</td>
</tr>
<tr>
<td>CreateWorkteam</td>
<td>258</td>
</tr>
<tr>
<td>DeleteAction</td>
<td>261</td>
</tr>
<tr>
<td>DeleteAlgorithm</td>
<td>263</td>
</tr>
<tr>
<td>DeleteApp</td>
<td>264</td>
</tr>
<tr>
<td>DeleteAppImageConfig</td>
<td>266</td>
</tr>
<tr>
<td>DeleteArtifact</td>
<td>268</td>
</tr>
<tr>
<td>DeleteAssociation</td>
<td>270</td>
</tr>
<tr>
<td>DeleteCluster</td>
<td>272</td>
</tr>
<tr>
<td>DeleteCodeRepository</td>
<td>274</td>
</tr>
<tr>
<td>DeleteContext</td>
<td>275</td>
</tr>
<tr>
<td>DeleteDataQualityJobDefinition</td>
<td>277</td>
</tr>
<tr>
<td>DeleteDeviceFleet</td>
<td>279</td>
</tr>
<tr>
<td>DeleteDomain</td>
<td>281</td>
</tr>
<tr>
<td>DeleteEdgeDeploymentPlan</td>
<td>283</td>
</tr>
<tr>
<td>DeleteEdgeDeploymentStage</td>
<td>285</td>
</tr>
<tr>
<td>DeleteEndpoint</td>
<td>287</td>
</tr>
<tr>
<td>DeleteEndpointConfig</td>
<td>289</td>
</tr>
<tr>
<td>DeleteExperiment</td>
<td>291</td>
</tr>
<tr>
<td>DeleteFeatureGroup</td>
<td>293</td>
</tr>
<tr>
<td>DeleteFlowDefinition</td>
<td>295</td>
</tr>
<tr>
<td>DeleteHub</td>
<td>297</td>
</tr>
<tr>
<td>DeleteHubContent</td>
<td>299</td>
</tr>
<tr>
<td>DeleteHumanTaskUI</td>
<td>301</td>
</tr>
<tr>
<td>DeleteImage</td>
<td>303</td>
</tr>
<tr>
<td>DeleteImageVersion</td>
<td>305</td>
</tr>
<tr>
<td>DeleteInferenceComponent</td>
<td>307</td>
</tr>
<tr>
<td>DeleteInferenceExperiment</td>
<td>308</td>
</tr>
<tr>
<td>DeleteModel</td>
<td>310</td>
</tr>
<tr>
<td>DeleteModelBiasJobDefinition</td>
<td>312</td>
</tr>
<tr>
<td>DeleteModelCard</td>
<td>314</td>
</tr>
<tr>
<td>DeleteModelExplainabilityJobDefinition</td>
<td>316</td>
</tr>
<tr>
<td>DeleteModelPackage</td>
<td>318</td>
</tr>
<tr>
<td>DeleteModelPackageGroup</td>
<td>320</td>
</tr>
<tr>
<td>DeleteModelPackageGroupPolicy</td>
<td>322</td>
</tr>
<tr>
<td>DeleteModelQualityJobDefinition</td>
<td>323</td>
</tr>
<tr>
<td>DeleteMonitoringSchedule</td>
<td>325</td>
</tr>
<tr>
<td>DeleteNotebookInstance</td>
<td>327</td>
</tr>
<tr>
<td>DeleteNotebookInstanceLifecycleConfig</td>
<td>329</td>
</tr>
<tr>
<td>DeletePipeline</td>
<td>330</td>
</tr>
<tr>
<td>DeleteProject</td>
<td>332</td>
</tr>
<tr>
<td>DeleteSpace</td>
<td>334</td>
</tr>
<tr>
<td>DeleteStudioLifecycleConfig</td>
<td>336</td>
</tr>
<tr>
<td>DeleteTags</td>
<td>338</td>
</tr>
</tbody>
</table>
DescribeSpace ................................................................. 573
DescribeStudioLifecycleConfig ........................................... 578
DescribeSubscribedWorkteam ................................................ 581
DescribeTrainingJob ........................................................... 583
DescribeTransformJob .......................................................... 594
DescribeTrial ................................................................. 599
DescribeTrialComponent ...................................................... 602
DescribeUserProfile .......................................................... 607
DescribeWorkforce ............................................................ 613
DescribeWorkteam ............................................................. 615
DisableSageMakerServicecatalogPortfolio ................................ 617
DisassociateTrialComponent .................................................. 618
EnableSageMakerServicecatalogPortfolio .................................. 620
GetDeviceFleetReport .......................................................... 621
GetLineageGroupPolicy ......................................................... 624
GetModelPackageGroupPolicy ............................................... 626
GetSageMakerServicecatalogPortfolioStatus ................................ 628
GetScalingConfigurationRecommendation ................................. 629
GetSearchSuggestions .......................................................... 633
GetImportHubContent .......................................................... 635
ListActions ........................................................................ 639
ListAlgorithms .................................................................... 642
ListAliases ........................................................................ 645
ListAppImageConfigs .............................................................. 648
ListApps ........................................................................... 652
ListArtifacts ...................................................................... 655
ListAssociations .................................................................. 658
ListAutoMLJobs .................................................................. 662
ListCandidatesForAutoMLJob ................................................ 666
ListClusterNodes .................................................................. 670
ListClusters ....................................................................... 674
ListCodeRepositories ............................................................. 677
ListCompilationJobs ............................................................... 681
ListContexts ....................................................................... 685
ListDataQualityJobDefinitions ................................................. 688
ListDeviceFleets ................................................................. 691
ListDevices ....................................................................... 694
ListDomains ....................................................................... 697
ListEdgeDeploymentPlans ...................................................... 699
ListEdgePackagingJobs .......................................................... 703
ListEndpointConfigs .............................................................. 707
ListEndpoints .................................................................... 710
ListExperiments .................................................................. 713
ListFeatureGroups ............................................................... 716
ListFlowDefinitions ............................................................... 719
ListHubContents ................................................................. 722
ListHubContentVersions .......................................................... 726
ListHubs ........................................................................... 730
ListHumanTaskUis ................................................................. 733
ListHyperParameterTuningJobs ................................................ 736
ListImages ........................................................................ 740
ListImageVersions ................................................................. 743
ListInferenceComponents ...................................................... 746
ListInferenceExperiments ...................................................... 750
ListInferenceRecommendationsJobs ......................................... 754
ListInferenceRecommendationsJobSteps .................................... 758
ListLabelingJobs ................................................................. 762
StopInferenceRecommendationsJob ................................................................. 967
StopLabelingJob ......................................................................................... 969
StopMonitoringSchedule ................................................................. 971
StopNotebookInstance ................................................................. 973
StopPipelineExecution ........................................................................... 975
StopProcessingJob .................................................................................. 977
StopTrainingJob ....................................................................................... 979
StopTransformJob ..................................................................................... 981
UpdateAction ......................................................................................... 983
UpdateApplImageConfig ....................................................................... 986
UpdateArtifact ......................................................................................... 988
UpdateCluster ......................................................................................... 991
UpdateCodeRepository .......................................................................... 994
UpdateContext ......................................................................................... 996
UpdateDeviceFleet .................................................................................. 999
UpdateDevices .........................................................................................1001
UpdateDomain ......................................................................................... 1003
UpdateEndpoint ....................................................................................... 1009
UpdateEndpointWeightsAndCapacities ......................................................... 1013
UpdateExperiment .................................................................................... 1015
UpdateFeatureGroup ............................................................................. 1017
UpdateFeatureMetadata ......................................................................... 1020
UpdateHub ................................................................................................. 1022
UpdateImage ............................................................................................. 1025
UpdateImageVersion .............................................................................. 1028
UpdateInferenceComponent .................................................................... 1032
UpdateInferenceComponentRuntimeConfig ................................................. 1035
UpdateInferenceExperiment ..................................................................... 1037
UpdateModelCard .................................................................................... 1040
UpdateModelPackage ............................................................................. 1043
UpdateMonitoringAlert ........................................................................... 1046
UpdateMonitoringSchedule .................................................................... 1049
UpdateNotebookInstance ....................................................................... 1053
UpdateNotebookInstanceLifecycleConfig .................................................... 1058
UpdatePipeline .......................................................................................... 1060
UpdatePipelineExecution ........................................................................... 1063
UpdateProject ............................................................................................ 1066
UpdateSpace .............................................................................................. 1069
UpdateTrainingJob .................................................................................... 1073
UpdateTrial ................................................................................................. 1076
UpdateTrialComponent ............................................................................ 1078
UpdateUserProfile ................................................................................... 1082
UpdateWorkforce ........................................................................................ 1087
UpdateWorkteam ....................................................................................... 1090

Amazon SageMaker Runtime ...................................................................... 1092
  InvokeEndpoint ......................................................................................... 1093
  InvokeEndpointAsync ............................................................................. 1099
  InvokeEndpointWithResponseStream ......................................................... 1103

Amazon SageMaker Edge Manager .............................................................. 1107
  GetDeployments ..................................................................................... 1109
  GetDeviceRegistration .......................................................................... 1111
  SendHeartbeat ........................................................................................ 1113

Amazon SageMaker Feature Store Runtime .................................................. 1115
  BatchGetRecord ...................................................................................... 1115
  DeleteRecord ........................................................................................... 1119
  GetRecord ................................................................................................. 1122
  PutRecord ................................................................................................. 1125
Amazon SageMaker geospatial capabilities ................................................................. 1127
DeleteEarthObservationJob ......................................................................................... 1129
DeleteVectorEnrichmentJob ......................................................................................... 1131
ExportEarthObservationJob ......................................................................................... 1133
ExportVectorEnrichmentJob ......................................................................................... 1137
GetEarthObservationJob .............................................................................................. 1141
GetRasterDataCollection ............................................................................................. 1145
GetTile ........................................................................................................................ 1148
GetVectorEnrichmentJob .............................................................................................. 1151
ListEarthObservationJobs ........................................................................................... 1155
ListRasterDataCollections .......................................................................................... 1158
ListTagsForResource ................................................................................................... 1161
ListVectorEnrichmentJobs .......................................................................................... 1163
SearchRasterDataCollection ....................................................................................... 1166
StartEarthObservationJob ............................................................................................ 1169
StartVectorEnrichmentJob ........................................................................................... 1174
StopEarthObservationJob ............................................................................................ 1179
StopVectorEnrichmentJob ........................................................................................... 1181
TagResource ................................................................................................................ 1183
UntagResource ............................................................................................................ 1185
Amazon SageMaker Metrics Service ........................................................................... 1186
BatchPutMetrics ........................................................................................................... 1187
Data Types .................................................................................................................. 1189
Amazon SageMaker Service ......................................................................................... 1203
ActionSource ................................................................................................................ 1216
ActionSummary ............................................................................................................ 1217
AdditionalInferenceSpecificationDefinition ............................................................... 1219
AdditionalS3DataSource ............................................................................................... 1222
AgentVersion ................................................................................................................ 1223
Alarm .......................................................................................................................... 1224
AlgorithmSpecification ............................................................................................... 1225
AlgorithmStatusDetails ............................................................................................... 1228
AlgorithmStatusItem .................................................................................................... 1229
AlgorithmSummary ..................................................................................................... 1230
AlgorithmValidationProfile .......................................................................................... 1232
AlgorithmValidationSpecification ............................................................................... 1233
AnnotationConsolidationConfig .................................................................................. 1234
AppDetails ................................................................................................................... 1246
AppImageConfigDetails .............................................................................................. 1248
AppSpecification ......................................................................................................... 1250
ArtifactSource .............................................................................................................. 1251
ArtifactSourceType ..................................................................................................... 1252
ArtifactSummary .......................................................................................................... 1253
AssociationSummary .................................................................................................... 1255
AsyncInferenceClientConfig ......................................................................................... 1257
AsyncInferenceConfig .................................................................................................. 1258
AsyncInferenceNotificationConfig .............................................................................. 1259
AsyncInferenceOutputConfig ....................................................................................... 1260
AthenaDatasetDefinition .............................................................................................. 1262
AutoMLAlgorithmConfig ............................................................................................. 1264
AutoMLCandidate ......................................................................................................... 1265
AutoMLCandidateGenerationConfig .......................................................................... 1268
AutoMLCandidateStep .................................................................................................. 1270
AutoMLChannel ........................................................................................................... 1271
AutoMLContainerDefinition ....................................................................................... 1273
AutoMLDataSource ...................................................................................................... 1275
AutoMLDataSplitConfig ............................................................................................... 1276
EdgeDeploymentModelConfig ................................................................. 1435
EdgeDeploymentPlanSummary ................................................................. 1436
EdgeDeploymentStatus ........................................................................ 1438
EdgeModel ....................................................................................... 1440
EdgeModelStat ................................................................................ 1441
EdgeModelSummary .......................................................................... 1443
EdgeOutputConfig .................................................................................. 1444
EdgePackagingJobSummary ................................................................. 1446
EdgePresetDeploymentOutput .............................................................. 1448
EFSFileSystem ...................................................................................... 1449
EFSFileSystemConfig ........................................................................... 1450
EMRStepMetadata .................................................................................. 1451
Endpoint .................................................................................................. 1452
EndpointConfigSummary .......................................................................... 1455
EndpointInfo .......................................................................................... 1456
EndpointInput .......................................................................................... 1457
EndpointInputConfiguration ................................................................. 1458
EndpointMetadata .................................................................................. 1460
EndpointOutputConfiguration ............................................................... 1463
EndpointPerformance ............................................................................. 1466
EndpointSummary ................................................................................... 1467
EnvironmentParameter ............................................................................ 1469
EnvironmentParameterRanges ................................................................. 1470
Experiment ............................................................................................. 1471
ExperimentConfig .................................................................................... 1473
ExperimentSource .................................................................................... 1475
ExperimentSummary ................................................................................ 1476
Explainability ........................................................................................... 1478
ExplainerConfig ...................................................................................... 1479
FailStepMetadata .................................................................................... 1480
FeatureDefinition .................................................................................... 1481
FeatureGroup ........................................................................................... 1483
FeatureGroupSummary .............................................................................. 1487
FeatureMetadata ...................................................................................... 1489
FeatureParameter ..................................................................................... 1491
FileSource ................................................................................................. 1492
FileSystemConfig .................................................................................... 1493
FileSystemDataSource ............................................................................. 1494
Filter .......................................................................................................... 1496
FinalAutoMLJobObjectiveMetric ............................................................ 1499
FinalHyperParameterTuningJobObjectiveMetric ..................................... 1501
FlowDefinitionOutputConfig .................................................................... 1502
FlowDefinitionSummary ........................................................................... 1503
GitConfig ................................................................................................. 1505
GitConfigForUpdate ................................................................................ 1506
HolidayConfigAttributes .......................................................................... 1507
HubContentDependency .......................................................................... 1508
HubContentInfo ....................................................................................... 1509
HubInfo ..................................................................................................... 1512
HubS3StorageConfig ................................................................................ 1514
HumanLoopActivationConditionsConfig .................................................. 1515
HumanLoopActivationConfig ................................................................... 1516
HumanLoopConfig ................................................................................... 1517
HumanLoopRequestSource ...................................................................... 1522
HumanTaskConfig ..................................................................................... 1523
HumanTaskUISummary .............................................................................. 1537
HyperbandStrategyConfig ....................................................................... 1538
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LabelingJobInputConfig</td>
<td>1632</td>
</tr>
<tr>
<td>LabelingJobOutput</td>
<td>1633</td>
</tr>
<tr>
<td>LabelingJobOutputConfig</td>
<td>1634</td>
</tr>
<tr>
<td>LabelingJobResourceConfig</td>
<td>1636</td>
</tr>
<tr>
<td>LabelingJobS3DataSource</td>
<td>1637</td>
</tr>
<tr>
<td>LabelingJobSnsDataSource</td>
<td>1638</td>
</tr>
<tr>
<td>LabelingJobStoppingConditions</td>
<td>1639</td>
</tr>
<tr>
<td>LabelingJobSummary</td>
<td>1640</td>
</tr>
<tr>
<td>LambdaStepMetadata</td>
<td>1643</td>
</tr>
<tr>
<td>LastUpdateStatus</td>
<td>1644</td>
</tr>
<tr>
<td>LineageGroupSummary</td>
<td>1645</td>
</tr>
<tr>
<td>MemberDefinition</td>
<td>1647</td>
</tr>
<tr>
<td>MetadataProperties</td>
<td>1648</td>
</tr>
<tr>
<td>MetricData</td>
<td>1650</td>
</tr>
<tr>
<td>MetricDatum</td>
<td>1651</td>
</tr>
<tr>
<td>MetricDefinition</td>
<td>1653</td>
</tr>
<tr>
<td>MetricSpecification</td>
<td>1654</td>
</tr>
<tr>
<td>MetricsSource</td>
<td>1655</td>
</tr>
<tr>
<td>Model</td>
<td>1656</td>
</tr>
<tr>
<td>ModelAccessConfig</td>
<td>1658</td>
</tr>
<tr>
<td>ModelArtifacts</td>
<td>1659</td>
</tr>
<tr>
<td>ModelBiasAppSpecification</td>
<td>1660</td>
</tr>
<tr>
<td>ModelBiasBaselineConfig</td>
<td>1661</td>
</tr>
<tr>
<td>ModelBiasJobInput</td>
<td>1662</td>
</tr>
<tr>
<td>ModelCard</td>
<td>1663</td>
</tr>
<tr>
<td>ModelCardExportArtifacts</td>
<td>1666</td>
</tr>
<tr>
<td>ModelCardExportJobSummary</td>
<td>1667</td>
</tr>
<tr>
<td>ModelCardExportOutputConfig</td>
<td>1669</td>
</tr>
<tr>
<td>ModelCardSecurityConfig</td>
<td>1670</td>
</tr>
<tr>
<td>ModelCardSummary</td>
<td>1671</td>
</tr>
<tr>
<td>ModelCardVersionSummary</td>
<td>1673</td>
</tr>
<tr>
<td>ModelClientConfig</td>
<td>1675</td>
</tr>
<tr>
<td>ModelConfiguration</td>
<td>1676</td>
</tr>
<tr>
<td>ModelDashboardEndpoint</td>
<td>1677</td>
</tr>
<tr>
<td>ModelDashboardIndicatorAction</td>
<td>1679</td>
</tr>
<tr>
<td>ModelDashboardModel</td>
<td>1680</td>
</tr>
<tr>
<td>ModelDashboardModelCard</td>
<td>1681</td>
</tr>
<tr>
<td>ModelDashboardMonitoringSchedule</td>
<td>1684</td>
</tr>
<tr>
<td>ModelDataQuality</td>
<td>1687</td>
</tr>
<tr>
<td>ModelDataSource</td>
<td>1688</td>
</tr>
<tr>
<td>ModelDeployConfig</td>
<td>1689</td>
</tr>
<tr>
<td>ModelDeployResult</td>
<td>1690</td>
</tr>
<tr>
<td>ModelDigests</td>
<td>1691</td>
</tr>
<tr>
<td>Model ExplainabilityAppSpecification</td>
<td>1692</td>
</tr>
<tr>
<td>Model ExplainabilityBaselineConfig</td>
<td>1693</td>
</tr>
<tr>
<td>Model ExplainabilityJobInput</td>
<td>1694</td>
</tr>
<tr>
<td>Model InfrastructureConfig</td>
<td>1695</td>
</tr>
<tr>
<td>ModelInput</td>
<td>1696</td>
</tr>
<tr>
<td>Model Latency Threshold</td>
<td>1697</td>
</tr>
<tr>
<td>Model Metadata Filter</td>
<td>1698</td>
</tr>
<tr>
<td>Model Metadata Search Expression</td>
<td>1699</td>
</tr>
<tr>
<td>Model Metadata Summary</td>
<td>1700</td>
</tr>
<tr>
<td>Model Metrics</td>
<td>1701</td>
</tr>
<tr>
<td>ModelPackage</td>
<td>1702</td>
</tr>
<tr>
<td>ModelPackageContainerDefinition</td>
<td>1707</td>
</tr>
<tr>
<td>ModelPackageGroup</td>
<td>1710</td>
</tr>
<tr>
<td>ModelPackageGroupSummary</td>
<td>1712</td>
</tr>
<tr>
<td>ModelPackageStatusDetails</td>
<td>1714</td>
</tr>
<tr>
<td>--------------------------</td>
<td>------</td>
</tr>
<tr>
<td>ModelPackageStatusItem</td>
<td>1715</td>
</tr>
<tr>
<td>ModelPackageSummary</td>
<td>1716</td>
</tr>
<tr>
<td>ModelPackageValidationProfile</td>
<td>1718</td>
</tr>
<tr>
<td>ModelPackageValidationSpecification</td>
<td>1719</td>
</tr>
<tr>
<td>ModelQuality</td>
<td>1720</td>
</tr>
<tr>
<td>ModelQualityAppSpecification</td>
<td>1721</td>
</tr>
<tr>
<td>ModelQualityBaselineConfig</td>
<td>1723</td>
</tr>
<tr>
<td>ModelQualityJobInput</td>
<td>1724</td>
</tr>
<tr>
<td>ModelRegisterSettings</td>
<td>1725</td>
</tr>
<tr>
<td>ModelStepMetadata</td>
<td>1726</td>
</tr>
<tr>
<td>ModelSummary</td>
<td>1727</td>
</tr>
<tr>
<td>ModelVariantConfig</td>
<td>1728</td>
</tr>
<tr>
<td>ModelVariantConfigSummary</td>
<td>1729</td>
</tr>
<tr>
<td>MonitoringAlertActions</td>
<td>1731</td>
</tr>
<tr>
<td>MonitoringAlertHistorySummary</td>
<td>1732</td>
</tr>
<tr>
<td>MonitoringAlertSummary</td>
<td>1733</td>
</tr>
<tr>
<td>MonitoringAppSpecification</td>
<td>1735</td>
</tr>
<tr>
<td>MonitoringBaselineConfig</td>
<td>1737</td>
</tr>
<tr>
<td>MonitoringClusterConfig</td>
<td>1738</td>
</tr>
<tr>
<td>MonitoringConstraintsResource</td>
<td>1740</td>
</tr>
<tr>
<td>MonitoringCsvDatasetFormat</td>
<td>1741</td>
</tr>
<tr>
<td>MonitoringDatasetFormat</td>
<td>1742</td>
</tr>
<tr>
<td>MonitoringExecutionSummary</td>
<td>1743</td>
</tr>
<tr>
<td>MonitoringGroundTruthS3Input</td>
<td>1745</td>
</tr>
<tr>
<td>MonitoringInput</td>
<td>1746</td>
</tr>
<tr>
<td>MonitoringJobDefinition</td>
<td>1747</td>
</tr>
<tr>
<td>MonitoringJobDefinitionSummary</td>
<td>1749</td>
</tr>
<tr>
<td>MonitoringJsonDatasetFormat</td>
<td>1751</td>
</tr>
<tr>
<td>MonitoringNetworkConfig</td>
<td>1752</td>
</tr>
<tr>
<td>MonitoringOutput</td>
<td>1753</td>
</tr>
<tr>
<td>MonitoringOutputConfig</td>
<td>1754</td>
</tr>
<tr>
<td>MonitoringParquetDatasetFormat</td>
<td>1755</td>
</tr>
<tr>
<td>MonitoringResources</td>
<td>1756</td>
</tr>
<tr>
<td>MonitoringS3Output</td>
<td>1757</td>
</tr>
<tr>
<td>MonitoringSchedule</td>
<td>1758</td>
</tr>
<tr>
<td>MonitoringScheduleConfig</td>
<td>1761</td>
</tr>
<tr>
<td>MonitoringScheduleSummary</td>
<td>1762</td>
</tr>
<tr>
<td>MonitoringStatisticsResource</td>
<td>1764</td>
</tr>
<tr>
<td>MonitoringStoppingCondition</td>
<td>1765</td>
</tr>
<tr>
<td>MultiModelConfig</td>
<td>1766</td>
</tr>
<tr>
<td>NeoVpcConfig</td>
<td>1767</td>
</tr>
<tr>
<td>NestedFilters</td>
<td>1768</td>
</tr>
<tr>
<td>NetworkConfig</td>
<td>1769</td>
</tr>
<tr>
<td>NotebookInstanceLifecycleConfigSummary</td>
<td>1770</td>
</tr>
<tr>
<td>NotebookInstanceLifecycleHook</td>
<td>1771</td>
</tr>
<tr>
<td>NotebookInstanceSummary</td>
<td>1772</td>
</tr>
<tr>
<td>NotificationConfiguration</td>
<td>1775</td>
</tr>
<tr>
<td>ObjectiveStatusCounters</td>
<td>1776</td>
</tr>
<tr>
<td>OfflineStoreConfig</td>
<td>1777</td>
</tr>
<tr>
<td>OfflineStoreStatus</td>
<td>1778</td>
</tr>
<tr>
<td>OidcConfig</td>
<td>1779</td>
</tr>
<tr>
<td>OidcConfigForResponse</td>
<td>1781</td>
</tr>
<tr>
<td>OidcMemberDefinition</td>
<td>1783</td>
</tr>
<tr>
<td>OnlineStoreConfig</td>
<td>1784</td>
</tr>
<tr>
<td>OnlineStoreConfigUpdate</td>
<td>1786</td>
</tr>
<tr>
<td>OnlineStoreSecurityConfig</td>
<td>1787</td>
</tr>
<tr>
<td>RecommendationJobCompiledOutputConfig</td>
<td>1886</td>
</tr>
<tr>
<td>---------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>RecommendationJobContainerConfig</td>
<td>1887</td>
</tr>
<tr>
<td>RecommendationJobInferenceBenchmark</td>
<td>1890</td>
</tr>
<tr>
<td>RecommendationJobInputConfig</td>
<td>1892</td>
</tr>
<tr>
<td>RecommendationJobOutputConfig</td>
<td>1895</td>
</tr>
<tr>
<td>RecommendationJobPayloadConfig</td>
<td>1896</td>
</tr>
<tr>
<td>RecommendationJobResourceLimit</td>
<td>1897</td>
</tr>
<tr>
<td>RecommendationJobStoppingConditions</td>
<td>1898</td>
</tr>
<tr>
<td>RecommendationJobVpcConfig</td>
<td>1899</td>
</tr>
<tr>
<td>RecommendationMetrics</td>
<td>1900</td>
</tr>
<tr>
<td>RedshiftDatasetDefinition</td>
<td>1902</td>
</tr>
<tr>
<td>RegisterModelStepMetadata</td>
<td>1904</td>
</tr>
<tr>
<td>RenderableTask</td>
<td>1905</td>
</tr>
<tr>
<td>RenderingError</td>
<td>1906</td>
</tr>
<tr>
<td>RepositoryAuthConfig</td>
<td>1907</td>
</tr>
<tr>
<td>ResolvedAttributes</td>
<td>1908</td>
</tr>
<tr>
<td>ResourceCatalog</td>
<td>1909</td>
</tr>
<tr>
<td>ResourceConfig</td>
<td>1910</td>
</tr>
<tr>
<td>ResourceConfigForUpdate</td>
<td>1913</td>
</tr>
<tr>
<td>ResourceLimits</td>
<td>1914</td>
</tr>
<tr>
<td>ResourceSpec</td>
<td>1915</td>
</tr>
<tr>
<td>RetentionPolicy</td>
<td>1917</td>
</tr>
<tr>
<td>RetryStrategy</td>
<td>1918</td>
</tr>
<tr>
<td>RollingUpdatePolicy</td>
<td>1919</td>
</tr>
<tr>
<td>RSessionAppSettings</td>
<td>1920</td>
</tr>
<tr>
<td>RStudioServerProAppSettings</td>
<td>1921</td>
</tr>
<tr>
<td>RStudioServerProDomainSettings</td>
<td>1922</td>
</tr>
<tr>
<td>RStudioServerProDomainSettingsForUpdate</td>
<td>1923</td>
</tr>
<tr>
<td>S3DataSource</td>
<td>1924</td>
</tr>
<tr>
<td>S3ModelDataSource</td>
<td>1927</td>
</tr>
<tr>
<td>S3StorageConfig</td>
<td>1929</td>
</tr>
<tr>
<td>ScalingPolicy</td>
<td>1931</td>
</tr>
<tr>
<td>ScalingPolicyMetric</td>
<td>1932</td>
</tr>
<tr>
<td>ScalingPolicyObjective</td>
<td>1933</td>
</tr>
<tr>
<td>ScheduleConfig</td>
<td>1934</td>
</tr>
<tr>
<td>SearchExpression</td>
<td>1936</td>
</tr>
<tr>
<td>SearchRecord</td>
<td>1938</td>
</tr>
<tr>
<td>SecondaryStatusTransition</td>
<td>1941</td>
</tr>
<tr>
<td>SelectedStep</td>
<td>1943</td>
</tr>
<tr>
<td>SelectiveExecutionConfig</td>
<td>1944</td>
</tr>
<tr>
<td>SelectiveExecutionResult</td>
<td>1945</td>
</tr>
<tr>
<td>ServiceCatalogProvisionedProductDetails</td>
<td>1946</td>
</tr>
<tr>
<td>ServiceCatalogProvisioningDetails</td>
<td>1947</td>
</tr>
<tr>
<td>ServiceCatalogProvisioningUpdateDetails</td>
<td>1949</td>
</tr>
<tr>
<td>ShadowModeConfig</td>
<td>1950</td>
</tr>
<tr>
<td>ShadowModelVariantConfig</td>
<td>1951</td>
</tr>
<tr>
<td>SharingSettings</td>
<td>1952</td>
</tr>
<tr>
<td>ShuffleConfig</td>
<td>1953</td>
</tr>
<tr>
<td>SourceAlgorithm</td>
<td>1954</td>
</tr>
<tr>
<td>SourceAlgorithmSpecification</td>
<td>1955</td>
</tr>
<tr>
<td>SourceAppConfig</td>
<td>1956</td>
</tr>
<tr>
<td>SpaceCodeEditorAppSettings</td>
<td>1957</td>
</tr>
<tr>
<td>SpaceDetails</td>
<td>1958</td>
</tr>
<tr>
<td>SpaceJupyterLabAppSettings</td>
<td>1960</td>
</tr>
<tr>
<td>SpaceSettings</td>
<td>1961</td>
</tr>
<tr>
<td>SpaceSettingsSummary</td>
<td>1963</td>
</tr>
<tr>
<td>SpaceSharingSettings</td>
<td>1964</td>
</tr>
<tr>
<td>Topic</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>SpaceSharingSettingsSummary</td>
<td>1965</td>
</tr>
<tr>
<td>SpaceStorageSettings</td>
<td>1966</td>
</tr>
<tr>
<td>Stairs</td>
<td>1967</td>
</tr>
<tr>
<td>StoppingCondition</td>
<td>1968</td>
</tr>
<tr>
<td>StudioLifecycleConfigDetails</td>
<td>1970</td>
</tr>
<tr>
<td>SubscribedWorkteam</td>
<td>1972</td>
</tr>
<tr>
<td>SuggestionQuery</td>
<td>1974</td>
</tr>
<tr>
<td>TabularJobConfig</td>
<td>1975</td>
</tr>
<tr>
<td>TabularResolvedAttributes</td>
<td>1978</td>
</tr>
<tr>
<td>Tag</td>
<td>1979</td>
</tr>
<tr>
<td>TargetPlatform</td>
<td>1980</td>
</tr>
<tr>
<td>TargetTrackingScalingPolicyConfiguration</td>
<td>1982</td>
</tr>
<tr>
<td>TensorBoardAppSettings</td>
<td>1983</td>
</tr>
<tr>
<td>TensorBoardOutputConfig</td>
<td>1984</td>
</tr>
<tr>
<td>TextClassificationJobConfig</td>
<td>1985</td>
</tr>
<tr>
<td>TextGenerationJobConfig</td>
<td>1986</td>
</tr>
<tr>
<td>TextGenerationResolvedAttributes</td>
<td>1988</td>
</tr>
<tr>
<td>TimeSeriesConfig</td>
<td>1989</td>
</tr>
<tr>
<td>TimeSeriesForecastingJobConfig</td>
<td>1991</td>
</tr>
<tr>
<td>TimeSeriesForecastingSettings</td>
<td>1994</td>
</tr>
<tr>
<td>TimeSeriesTransformations</td>
<td>1995</td>
</tr>
<tr>
<td>TrafficPattern</td>
<td>1997</td>
</tr>
<tr>
<td>TrafficRoutingConfig</td>
<td>1998</td>
</tr>
<tr>
<td>TrainingImageConfig</td>
<td>2000</td>
</tr>
<tr>
<td>TrainingJob</td>
<td>2001</td>
</tr>
<tr>
<td>TrainingJobDefinition</td>
<td>2009</td>
</tr>
<tr>
<td>TrainingJobStatusCounters</td>
<td>2011</td>
</tr>
<tr>
<td>TrainingJobStepMetadata</td>
<td>2013</td>
</tr>
<tr>
<td>TrainingJobSummary</td>
<td>2014</td>
</tr>
<tr>
<td>TrainingRepositoryAuthConfig</td>
<td>2016</td>
</tr>
<tr>
<td>TrainingSpecification</td>
<td>2017</td>
</tr>
<tr>
<td>TransformDataSource</td>
<td>2019</td>
</tr>
<tr>
<td>TransformInput</td>
<td>2020</td>
</tr>
<tr>
<td>TransformJob</td>
<td>2022</td>
</tr>
<tr>
<td>TransformJobDefinition</td>
<td>2027</td>
</tr>
<tr>
<td>TransformJobStepMetadata</td>
<td>2029</td>
</tr>
<tr>
<td>TransformJobSummary</td>
<td>2030</td>
</tr>
<tr>
<td>TransformOutput</td>
<td>2032</td>
</tr>
<tr>
<td>TransformResources</td>
<td>2034</td>
</tr>
<tr>
<td>TransformS3DataSource</td>
<td>2036</td>
</tr>
<tr>
<td>Trial</td>
<td>2038</td>
</tr>
<tr>
<td>TrialComponent</td>
<td>2041</td>
</tr>
<tr>
<td>TrialComponentArtifact</td>
<td>2045</td>
</tr>
<tr>
<td>TrialComponentMetricSummary</td>
<td>2046</td>
</tr>
<tr>
<td>TrialComponentParameterValue</td>
<td>2048</td>
</tr>
<tr>
<td>TrialComponentSimpleSummary</td>
<td>2049</td>
</tr>
<tr>
<td>TrialComponentSource</td>
<td>2051</td>
</tr>
<tr>
<td>TrialComponentSourceDetail</td>
<td>2052</td>
</tr>
<tr>
<td>TrialComponentStatus</td>
<td>2053</td>
</tr>
<tr>
<td>TrialComponentSummary</td>
<td>2054</td>
</tr>
<tr>
<td>TrialSource</td>
<td>2056</td>
</tr>
<tr>
<td>TrialSummary</td>
<td>2057</td>
</tr>
<tr>
<td>TtlDuration</td>
<td>2059</td>
</tr>
<tr>
<td>TuningJobCompletionCriteria</td>
<td>2060</td>
</tr>
<tr>
<td>TuningJobStepMetaData</td>
<td>2061</td>
</tr>
<tr>
<td>UICfg</td>
<td>2062</td>
</tr>
<tr>
<td>UiTemplate</td>
<td>2064</td>
</tr>
<tr>
<td>Class/Function</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>UiTemplateInfo</td>
<td>2065</td>
</tr>
<tr>
<td>USD</td>
<td>2066</td>
</tr>
<tr>
<td>UserContext</td>
<td>2067</td>
</tr>
<tr>
<td>UserProfileDetails</td>
<td>2068</td>
</tr>
<tr>
<td>UserSettings</td>
<td>2070</td>
</tr>
<tr>
<td>VariantProperty</td>
<td>2073</td>
</tr>
<tr>
<td>VectorConfig</td>
<td>2074</td>
</tr>
<tr>
<td>Vertex</td>
<td>2075</td>
</tr>
<tr>
<td>VpcConfig</td>
<td>2076</td>
</tr>
<tr>
<td>WarmPoolStatus</td>
<td>2077</td>
</tr>
<tr>
<td>Workforce</td>
<td>2079</td>
</tr>
<tr>
<td>WorkforceVpcConfigRequest</td>
<td>2081</td>
</tr>
<tr>
<td>WorkforceVpcConfigResponse</td>
<td>2082</td>
</tr>
<tr>
<td>WorkspaceSettings</td>
<td>2084</td>
</tr>
<tr>
<td>Workteam</td>
<td>2085</td>
</tr>
<tr>
<td>Amazon SageMaker Runtime</td>
<td>2086</td>
</tr>
<tr>
<td>PayloadPart</td>
<td>2088</td>
</tr>
<tr>
<td>ResponseStream</td>
<td>2089</td>
</tr>
<tr>
<td>Amazon Sagemaker Edge Manager</td>
<td>2089</td>
</tr>
<tr>
<td>Checksum</td>
<td>2091</td>
</tr>
<tr>
<td>Definition</td>
<td>2092</td>
</tr>
<tr>
<td>DeploymentModel</td>
<td>2093</td>
</tr>
<tr>
<td>DeploymentResult</td>
<td>2095</td>
</tr>
<tr>
<td>EdgeDeployment</td>
<td>2097</td>
</tr>
<tr>
<td>EdgeMetric</td>
<td>2098</td>
</tr>
<tr>
<td>Model</td>
<td>2099</td>
</tr>
<tr>
<td>Amazon SageMaker Feature Store Runtime</td>
<td>2100</td>
</tr>
<tr>
<td>BatchGetRecordError</td>
<td>2101</td>
</tr>
<tr>
<td>BatchGetRecordIdentifier</td>
<td>2103</td>
</tr>
<tr>
<td>BatchGetRecordResultDetail</td>
<td>2105</td>
</tr>
<tr>
<td>FeatureValue</td>
<td>2106</td>
</tr>
<tr>
<td>TtlDuration</td>
<td>2107</td>
</tr>
<tr>
<td>Amazon SageMaker geospatial capabilities</td>
<td>2107</td>
</tr>
<tr>
<td>AreaOfInterest</td>
<td>2110</td>
</tr>
<tr>
<td>AreaOfInterestGeometry</td>
<td>2111</td>
</tr>
<tr>
<td>AssetValue</td>
<td>2112</td>
</tr>
<tr>
<td>BandMathConfigInput</td>
<td>2113</td>
</tr>
<tr>
<td>CloudMaskingConfigInput</td>
<td>2114</td>
</tr>
<tr>
<td>CloudRemovalConfigInput</td>
<td>2115</td>
</tr>
<tr>
<td>CustomIndicesInput</td>
<td>2116</td>
</tr>
<tr>
<td>EarthObservationJobErrorDetails</td>
<td>2117</td>
</tr>
<tr>
<td>EoCloudCoverInput</td>
<td>2118</td>
</tr>
<tr>
<td>ExportErrorDetails</td>
<td>2119</td>
</tr>
<tr>
<td>ExportErrorDetailsOutput</td>
<td>2120</td>
</tr>
<tr>
<td>ExportS3DataInput</td>
<td>2121</td>
</tr>
<tr>
<td>ExportVectorEnrichmentJobOutputConfig</td>
<td>2122</td>
</tr>
<tr>
<td>Filter</td>
<td>2123</td>
</tr>
<tr>
<td>Geometry</td>
<td>2124</td>
</tr>
<tr>
<td>GeoMosaicConfigInput</td>
<td>2125</td>
</tr>
<tr>
<td>InputConfigInput</td>
<td>2126</td>
</tr>
<tr>
<td>InputConfigOutput</td>
<td>2127</td>
</tr>
<tr>
<td>ItemSource</td>
<td>2128</td>
</tr>
<tr>
<td>JobConfigInput</td>
<td>2129</td>
</tr>
<tr>
<td>LandCoverSegmentationConfigInput</td>
<td>2131</td>
</tr>
<tr>
<td>LandsatCloudCoverLandInput</td>
<td>2132</td>
</tr>
<tr>
<td>ListEarthObservationJobOutputConfig</td>
<td>2133</td>
</tr>
<tr>
<td>ListVectorEnrichmentJobOutputConfig</td>
<td>2135</td>
</tr>
</tbody>
</table>
Welcome

Amazon SageMaker Service

Provides APIs for creating and managing SageMaker resources.

Other Resources:

- SageMaker Developer Guide
- Amazon Augmented AI Runtime API Reference

Amazon SageMaker Runtime

The Amazon SageMaker runtime API.

Amazon Sagemaker Edge Manager

SageMaker Edge Manager dataplane service for communicating with active agents.

Amazon SageMaker Feature Store Runtime

Contains all data plane API operations and data types for the Amazon SageMaker Feature Store. Use this API to put, delete, and retrieve (get) features from a feature store.

Use the following operations to configure your OnlineStore and OfflineStore features, and to create and manage feature groups:

- CreateFeatureGroup
- DeleteFeatureGroup
- DescribeFeatureGroup
- ListFeatureGroups

Amazon SageMaker geospatial capabilities

Provides APIs for creating and managing SageMaker geospatial resources.

Amazon SageMaker Metrics Service

Contains all data plane API operations and data types for Amazon SageMaker Metrics. Use these APIs to put and retrieve (get) features related to your training run.

- BatchPutMetrics
Actions

The following actions are supported by Amazon SageMaker Service:

- AddAssociation (p. 18)
- AddTags (p. 21)
- AssociateTrialComponent (p. 23)
- BatchDescribeModelPackage (p. 25)
- CreateAction (p. 28)
- CreateAlgorithm (p. 31)
- CreateApp (p. 37)
- CreateAppImageConfig (p. 40)
- CreateArtifact (p. 43)
- CreateAutoMLJob (p. 46)
- CreateAutoMLJobV2 (p. 51)
- CreateCluster (p. 56)
- CreateCodeRepository (p. 59)
- CreateCompilationJob (p. 61)
- CreateContext (p. 65)
- CreateDataQualityJobDefinition (p. 68)
- CreateDeviceFleet (p. 73)
- CreateDomain (p. 76)
- CreateEdgeDeploymentPlan (p. 84)
- CreateEdgeDeploymentStage (p. 87)
- CreateEdgePackagingJob (p. 89)
- CreateEndpoint (p. 92)
- CreateEndpointConfig (p. 96)
- CreateExperiment (p. 102)
- CreateFeatureGroup (p. 105)
- CreateFlowDefinition (p. 110)
- CreateHub (p. 113)
- CreateHumanTaskUi (p. 116)
- CreateHyperParameterTuningJob (p. 118)
- CreateImage (p. 127)
- CreateImageVersion (p. 130)
- CreateInferenceComponent (p. 134)
- CreateInferenceExperiment (p. 137)
- CreateInferenceRecommendationsJob (p. 142)
- CreateLabelingJob (p. 147)
- CreateModel (p. 153)
- CreateModelBiasJobDefinition (p. 158)
- CreateModelCard (p. 163)
- CreateModelCardExportJob (p. 166)
• CreateModelExplainabilityJobDefinition (p. 169)
• CreateModelPackage (p. 174)
• CreateModelPackageGroup (p. 183)
• CreateModelQualityJobDefinition (p. 185)
• CreateMonitoringSchedule (p. 190)
• CreateNotebookInstance (p. 194)
• CreateNotebookInstanceLifecycleConfig (p. 200)
• CreatePipeline (p. 203)
• CreatePresignedDomainUrl (p. 207)
• CreatePresignedNotebookInstanceUrl (p. 210)
• CreateProcessingJob (p. 212)
• CreateProject (p. 217)
• CreateSpace (p. 220)
• CreateStudioLifecycleConfig (p. 224)
• CreateTrainingJob (p. 227)
• CreateTransformJob (p. 237)
• CreateTrial (p. 243)
• CreateTrialComponent (p. 246)
• CreateUserProfile (p. 250)
• CreateWorkforce (p. 255)
• CreateWorkteam (p. 258)
• DeleteAction (p. 261)
• DeleteAlgorithm (p. 263)
• DeleteApp (p. 264)
• DeleteAppImageConfig (p. 266)
• DeleteArtifact (p. 268)
• DeleteAssociation (p. 270)
• DeleteCluster (p. 272)
• DeleteCodeRepository (p. 274)
• DeleteContext (p. 275)
• DeleteDataQualityJobDefinition (p. 277)
• DeleteDeviceFleet (p. 279)
• DeleteDomain (p. 281)
• DeleteEdgeDeploymentPlan (p. 283)
• DeleteEdgeDeploymentStage (p. 285)
• DeleteEndpoint (p. 287)
• DeleteEndpointConfig (p. 289)
• DeleteExperiment (p. 291)
• DeleteFeatureGroup (p. 293)
• DeleteFlowDefinition (p. 295)
• DeleteHub (p. 297)
• DeleteHubContent (p. 299)
• DeleteHumanTaskUi (p. 301)
• DeleteImage (p. 303)
• DeleteImageVersion (p. 305)
• DeleteInferenceComponent (p. 307)
- **DeleteInferenceExperiment** (p. 308)
- **DeleteModel** (p. 310)
- **DeleteModelBiasJobDefinition** (p. 312)
- **DeleteModelCard** (p. 314)
- **DeleteModelExplainabilityJobDefinition** (p. 316)
- **DeleteModelPackage** (p. 318)
- **DeleteModelPackageGroup** (p. 320)
- **DeleteModelPackageGroupPolicy** (p. 322)
- **DeleteModelQualityJobDefinition** (p. 323)
- **DeleteMonitoringSchedule** (p. 325)
- **DeleteNotebookInstance** (p. 327)
- **DeleteNotebookInstanceLifecycleConfig** (p. 329)
- **DeletePipeline** (p. 330)
- **DeleteProject** (p. 332)
- **DeleteSpace** (p. 334)
- **DeleteStudioLifecycleConfig** (p. 336)
- **DeleteTags** (p. 338)
- **DeleteTrial** (p. 340)
- **DeleteTrialComponent** (p. 342)
- **DeleteUserProfile** (p. 344)
- **DeleteWorkforce** (p. 346)
- **DeleteWorkteam** (p. 348)
- **DeregisterDevices** (p. 350)
- **DescribeAction** (p. 352)
- **DescribeAlgorithm** (p. 356)
- **DescribeApp** (p. 362)
- **DescribeAppImageConfig** (p. 366)
- **DescribeArtifact** (p. 369)
- **DescribeAutoMLJob** (p. 373)
- **DescribeAutoMLJobV2** (p. 379)
- **DescribeCluster** (p. 385)
- **DescribeClusterNode** (p. 388)
- **DescribeCodeRepository** (p. 390)
- **DescribeCompilationJob** (p. 392)
- **DescribeContext** (p. 397)
- **DescribeDataQualityJobDefinition** (p. 401)
- **DescribeDevice** (p. 405)
- **DescribeDeviceFleet** (p. 409)
- **DescribeDomain** (p. 412)
- **DescribeEdgeDeploymentPlan** (p. 420)
- **DescribeEdgePackagingJob** (p. 424)
- **DescribeEndpoint** (p. 428)
- **DescribeEndpointConfig** (p. 436)
- **DescribeExperiment** (p. 441)
- **DescribeFeatureGroup** (p. 444)
- **DescribeFeatureMetadata** (p. 449)
• DescribeFlowDefinition (p. 452)
• DescribeHub (p. 455)
• DescribeHubContent (p. 458)
• DescribeHumanTaskUi (p. 463)
• DescribeHyperParameterTuningJob (p. 465)
• DescribeImage (p. 475)
• DescribeImageVersion (p. 478)
• DescribeInferenceComponent (p. 483)
• DescribeInferenceExperiment (p. 486)
• DescribeInferenceRecommendationsJob (p. 491)
• DescribeLabelingJob (p. 496)
• DescribeLineageGroup (p. 502)
• DescribeModel (p. 505)
• DescribeModelBiasJobDefinition (p. 509)
• DescribeModelCard (p. 513)
• DescribeModelCardExportJob (p. 517)
• DescribeModelExplainabilityJobDefinition (p. 520)
• DescribeModelPackage (p. 524)
• DescribeModelPackageGroup (p. 533)
• DescribeModelQualityJobDefinition (p. 536)
• DescribeMonitoringSchedule (p. 540)
• DescribeNotebookInstance (p. 545)
• DescribeNotebookInstanceLifecycleConfig (p. 550)
• DescribePipeline (p. 553)
• DescribePipelineDefinitionForExecution (p. 557)
• DescribePipelineExecution (p. 559)
• DescribeProcessingJob (p. 563)
• DescribeProject (p. 569)
• DescribeSpace (p. 573)
• DescribeStudioLifecycleConfig (p. 578)
• DescribeSubscribedWorkteam (p. 581)
• DescribeTrainingJob (p. 583)
• DescribeTransformJob (p. 594)
• DescribeTrial (p. 599)
• DescribeTrialComponent (p. 602)
• DescribeUserProfile (p. 607)
• DescribeWorkforce (p. 613)
• DescribeWorkteam (p. 615)
• DisableSagemakerServicecatalogPortfolio (p. 617)
• DisassociateTrialComponent (p. 618)
• EnableSagemakerServicecatalogPortfolio (p. 620)
• GetDeviceFleetReport (p. 621)
• GetLineageGroupPolicy (p. 624)
• GetModelPackageGroupPolicy (p. 626)
• GetSagemakerServicecatalogPortfolioStatus (p. 628)
- GetScalingConfigurationRecommendation (p. 629)
- GetSearchSuggestions (p. 633)
- ImportHubContent (p. 635)
- ListActions (p. 639)
- ListAlgorithms (p. 642)
- ListAliases (p. 645)
- ListAppImageConfigs (p. 648)
- ListApps (p. 652)
- ListArtifacts (p. 655)
- ListAssociations (p. 658)
- ListAutoMLJobs (p. 662)
- ListCandidatesForAutoMLJob (p. 666)
- ListClusterNodes (p. 670)
- ListClusters (p. 674)
- ListCodeRepositories (p. 677)
- ListCompilationJobs (p. 681)
- ListContexts (p. 685)
- ListDataQualityJobDefinitions (p. 688)
- ListDeviceFleets (p. 691)
- ListDevices (p. 694)
- ListDomains (p. 697)
- ListEdgeDeploymentPlans (p. 699)
- ListEdgePackagingJobs (p. 703)
- ListEndpointConfigs (p. 707)
- ListEndpoints (p. 710)
- ListExperiments (p. 713)
- ListFeatureGroups (p. 716)
- ListFlowDefinitions (p. 719)
- ListHubContents (p. 722)
- ListHubContentVersions (p. 726)
- ListHubs (p. 730)
- ListHumanTaskUis (p. 733)
- ListHyperParameterTuningJobs (p. 736)
- ListImages (p. 740)
- ListImageVersions (p. 743)
- ListInferenceComponents (p. 746)
- ListInferenceExperiments (p. 750)
- ListInferenceRecommendationsJobs (p. 754)
- ListInferenceRecommendationsJobSteps (p. 758)
- ListLabelingJobs (p. 762)
- ListLabelingJobsForWorkteam (p. 766)
- ListLineageGroups (p. 769)
- ListModelBiasJobDefinitions (p. 772)
- ListModelCardExportJobs (p. 775)
- ListModelCards (p. 779)
- ListModelCardVersions (p. 782)
• ListModelExplainabilityJobDefinitions (p. 785)
• ListModelMetadata (p. 788)
• ListModelPackageGroups (p. 790)
• ListModelPackages (p. 793)
• ListModelQualityJobDefinitions (p. 797)
• ListModels (p. 800)
• ListMonitoringAlertHistory (p. 803)
• ListMonitoringAlerts (p. 807)
• ListMonitoringExecutions (p. 810)
• ListMonitoringSchedules (p. 814)
• ListNotebookInstanceLifecycleConfigs (p. 818)
• ListNotebookInstances (p. 821)
• ListPipelineExecutions (p. 825)
• ListPipelineExecutionSteps (p. 828)
• ListPipelineParametersForExecution (p. 832)
• ListPipelines (p. 835)
• ListProcessingJobs (p. 838)
• ListProjects (p. 841)
• ListResourceCatalogs (p. 844)
• ListSpaces (p. 847)
• ListStageDevices (p. 850)
• ListStudioLifecycleConfigs (p. 853)
• ListSubscribedWorkteams (p. 857)
• ListTags (p. 859)
• ListTrainingJobs (p. 861)
• ListTrainingJobsForHyperParameterTuningJob (p. 865)
• ListTransformJobs (p. 868)
• ListTrialComponents (p. 871)
• ListTrials (p. 875)
• ListUserProfiles (p. 878)
• ListWorkforces (p. 881)
• ListWorkteams (p. 884)
• PutModelPackageGroupPolicy (p. 887)
• QueryLineage (p. 889)
• RegisterDevices (p. 893)
• RenderUiTemplate (p. 895)
• RetryPipelineExecution (p. 898)
• Search (p. 901)
• SendPipelineExecutionStepFailure (p. 940)
• SendPipelineExecutionStepSuccess (p. 942)
• StartEdgeDeploymentStage (p. 944)
• StartInferenceExperiment (p. 946)
• StartMonitoringSchedule (p. 948)
• StartNotebookInstance (p. 950)
• StartPipelineExecution (p. 952)
• StopAutoMLJob (p. 955)
• StopCompilationJob (p. 957)
• StopEdgeDeploymentStage (p. 959)
• StopEdgePackagingJob (p. 961)
• StopHyperParameterTuningJob (p. 962)
• StopInferenceExperiment (p. 964)
• StopInferenceRecommendationsJob (p. 967)
• StopLabelingJob (p. 969)
• StopMonitoringSchedule (p. 971)
• StopNotebookInstance (p. 973)
• StopPipelineExecution (p. 975)
• StopProcessingJob (p. 977)
• StopTrainingJob (p. 979)
• StopTransformJob (p. 981)
• UpdateAction (p. 983)
• UpdateAppImageConfig (p. 986)
• UpdateArtifact (p. 988)
• UpdateCluster (p. 991)
• UpdateCodeRepository (p. 994)
• UpdateContext (p. 996)
• UpdateDeviceFleet (p. 999)
• UpdateDevices (p. 1001)
• UpdateDomain (p. 1003)
• UpdateEndpoint (p. 1009)
• UpdateEndpointWeightsAndCapacities (p. 1013)
• UpdateExperiment (p. 1015)
• UpdateFeatureGroup (p. 1017)
• UpdateFeatureMetadata (p. 1020)
• UpdateHub (p. 1022)
• UpdateImage (p. 1025)
• UpdateImageVersion (p. 1028)
• UpdateInferenceComponent (p. 1032)
• UpdateInferenceComponentRuntimeConfig (p. 1035)
• UpdateInferenceExperiment (p. 1037)
• UpdateModelCard (p. 1040)
• UpdateModelPackage (p. 1043)
• UpdateMonitoringAlert (p. 1046)
• UpdateMonitoringSchedule (p. 1049)
• UpdateNotebookInstance (p. 1053)
• UpdateNotebookInstanceLifecycleConfig (p. 1058)
• UpdatePipeline (p. 1060)
• UpdatePipelineExecution (p. 1063)
• UpdateProject (p. 1066)
• UpdateSpace (p. 1069)
• UpdateTrainingJob (p. 1073)
• UpdateTrial (p. 1076)
• UpdateTrialComponent (p. 1078)
The following actions are supported by Amazon SageMaker Runtime:

- **InvokeEndpoint** (p. 1093)
- **InvokeEndpointAsync** (p. 1099)
- **InvokeEndpointWithResponseStream** (p. 1103)

The following actions are supported by Amazon Sagemaker Edge Manager:

- **GetDeployments** (p. 1109)
- **GetDeviceRegistration** (p. 1111)
- **SendHeartbeat** (p. 1113)

The following actions are supported by Amazon SageMaker Feature Store Runtime:

- **BatchGetRecord** (p. 1116)
- **DeleteRecord** (p. 1119)
- **GetRecord** (p. 1122)
- **PutRecord** (p. 1125)

The following actions are supported by Amazon SageMaker geospatial capabilities:

- **DeleteEarthObservationJob** (p. 1129)
- **DeleteVectorEnrichmentJob** (p. 1131)
- **ExportEarthObservationJob** (p. 1133)
- **ExportVectorEnrichmentJob** (p. 1137)
- **GetEarthObservationJob** (p. 1141)
- **GetRasterDataCollection** (p. 1145)
- **GetTile** (p. 1148)
- **GetVectorEnrichmentJob** (p. 1151)
- **ListEarthObservationJobs** (p. 1155)
- **ListRasterDataCollections** (p. 1158)
- **ListTagsForResource** (p. 1161)
- **ListVectorEnrichmentJobs** (p. 1163)
- **SearchRasterDataCollection** (p. 1166)
- **StartEarthObservationJob** (p. 1169)
- **StartVectorEnrichmentJob** (p. 1174)
- **StopEarthObservationJob** (p. 1179)
- **StopVectorEnrichmentJob** (p. 1181)
- **TagResource** (p. 1183)
- **UntagResource** (p. 1185)

The following actions are supported by Amazon SageMaker Metrics Service:

- **BatchPutMetrics** (p. 1187)
Amazon SageMaker Service

The following actions are supported by Amazon SageMaker Service:

- AddAssociation (p. 18)
- AddTags (p. 21)
- AssociateTrialComponent (p. 23)
- BatchDescribeModelPackage (p. 25)
- CreateAction (p. 28)
- CreateAlgorithm (p. 31)
- CreateApp (p. 37)
- CreateAppImageConfig (p. 40)
- CreateArtifact (p. 43)
- CreateAutoMLJob (p. 46)
- CreateAutoMLJobV2 (p. 51)
- CreateCluster (p. 56)
- CreateCodeRepository (p. 59)
- CreateCompilationJob (p. 61)
- CreateContext (p. 65)
- CreateDataQualityJobDefinition (p. 68)
- CreateDeviceFleet (p. 73)
- CreateDomain (p. 76)
- CreateEdgeDeploymentPlan (p. 84)
- CreateEdgeDeploymentStage (p. 87)
- CreateEdgePackagingJob (p. 89)
- CreateEndpoint (p. 92)
- CreateEndpointConfig (p. 96)
- CreateExperiment (p. 102)
- CreateFeatureGroup (p. 105)
- CreateFlowDefinition (p. 110)
- CreateHub (p. 113)
- CreateHumanTaskUi (p. 116)
- CreateHyperParameterTuningJob (p. 118)
- CreateImage (p. 127)
- CreateImageVersion (p. 130)
- CreateInferenceComponent (p. 134)
- CreateInferenceExperiment (p. 137)
- CreateInferenceRecommendationsJob (p. 142)
- CreateLabelingJob (p. 147)
- CreateModel (p. 153)
- CreateModelBiasJobDefinition (p. 158)
- CreateModelCard (p. 163)
- CreateModelCardExportJob (p. 166)
- CreateModelExplainabilityJobDefinition (p. 169)
- CreateModelPackage (p. 174)
- CreateModelPackageGroup (p. 183)
• CreateModelQualityJobDefinition (p. 185)
• CreateMonitoringSchedule (p. 190)
• CreateNotebookInstance (p. 194)
• CreateNotebookInstanceLifecycleConfig (p. 200)
• CreatePipeline (p. 203)
• CreatePresignedDomainUrl (p. 207)
• CreatePresignedNotebookInstanceUrl (p. 210)
• CreateProcessingJob (p. 212)
• CreateProject (p. 217)
• CreateSpace (p. 220)
• CreateStudioLifecycleConfig (p. 224)
• CreateTrainingJob (p. 227)
• CreateTransformJob (p. 237)
• CreateTrial (p. 243)
• CreateTrialComponent (p. 246)
• CreateUserProfile (p. 250)
• CreateWorkforce (p. 255)
• CreateWorkteam (p. 258)
• DeleteAction (p. 261)
• DeleteAlgorithm (p. 263)
• DeleteApp (p. 264)
• DeleteAppImageConfig (p. 266)
• DeleteArtifact (p. 268)
• DeleteAssociation (p. 270)
• DeleteCluster (p. 272)
• DeleteCodeRepository (p. 274)
• DeleteContext (p. 275)
• DeleteDataQualityJobDefinition (p. 277)
• DeleteDeviceFleet (p. 279)
• DeleteDomain (p. 281)
• DeleteEdgeDeploymentPlan (p. 283)
• DeleteEdgeDeploymentStage (p. 285)
• DeleteEndpoint (p. 287)
• DeleteEndpointConfig (p. 289)
• DeleteExperiment (p. 291)
• DeleteFeatureGroup (p. 293)
• DeleteFlowDefinition (p. 295)
• DeleteHub (p. 297)
• DeleteHubContent (p. 299)
• DeleteHumanTaskUi (p. 301)
• DeleteImage (p. 303)
• DeleteImageVersion (p. 305)
• DeleteInferenceComponent (p. 307)
• DeleteInferenceExperiment (p. 308)
• DeleteModel (p. 310)
• **DeleteModelBiasJobDefinition** (p. 312)
  • **DeleteModelCard** (p. 314)
  • **DeleteModelExplainabilityJobDefinition** (p. 316)
  • **DeleteModelPackage** (p. 318)
  • **DeleteModelPackageGroup** (p. 320)
  • **DeleteModelPackageGroupPolicy** (p. 322)
  • **DeleteModelQualityJobDefinition** (p. 323)
  • **DeleteMonitoringSchedule** (p. 325)
  • **DeleteNotebookInstance** (p. 327)
  • **DeleteNotebookInstanceLifecycleConfig** (p. 329)
  • **DeletePipeline** (p. 330)
  • **DeleteProject** (p. 332)
  • **DeleteSpace** (p. 334)
  • **DeleteStudioLifecycleConfig** (p. 336)
  • **DeleteTags** (p. 338)
  • **DeleteTrial** (p. 340)
  • **DeleteTrialComponent** (p. 342)
  • **DeleteUserProfile** (p. 344)
  • **DeleteWorkforce** (p. 346)
  • **DeleteWorkteam** (p. 348)
  • **DeregisterDevices** (p. 350)
  • **DescribeAction** (p. 352)
  • **DescribeAlgorithm** (p. 356)
  • **DescribeApp** (p. 362)
  • **DescribeAppImageConfig** (p. 366)
  • **DescribeArtifact** (p. 369)
  • **DescribeAutoMLJob** (p. 373)
  • **DescribeAutoMLJobV2** (p. 379)
  • **DescribeCluster** (p. 385)
  • **DescribeClusterNode** (p. 388)
  • **DescribeCodeRepository** (p. 390)
  • **DescribeCompilationJob** (p. 392)
  • **DescribeContext** (p. 397)
  • **DescribeDataQualityJobDefinition** (p. 401)
  • **DescribeDevice** (p. 405)
  • **DescribeDeviceFleet** (p. 409)
  • **DescribeDomain** (p. 412)
  • **DescribeEdgeDeploymentPlan** (p. 420)
  • **DescribeEdgePackagingJob** (p. 424)
  • **DescribeEndpoint** (p. 428)
  • **DescribeEndpointConfig** (p. 436)
  • **DescribeExperiment** (p. 441)
  • **DescribeFeatureGroup** (p. 444)
  • **DescribeFeatureMetadata** (p. 449)
  • **DescribeFlowDefinition** (p. 452)
  • **DescribeHub** (p. 455)
• DescribeHubContent (p. 458)
• DescribeHumanTaskUi (p. 463)
• DescribeHyperParameterTuningJob (p. 465)
• DescribeImage (p. 475)
• DescribeImageVersion (p. 478)
• DescribeInferenceComponent (p. 483)
• DescribeInferenceExperiment (p. 486)
• DescribeInferenceRecommendationsJob (p. 491)
• DescribeLabelingJob (p. 496)
• DescribeLineageGroup (p. 502)
• DescribeModel (p. 505)
• DescribeModelBiasJobDefinition (p. 509)
• DescribeModelCard (p. 513)
• DescribeModelCardExportJob (p. 517)
• DescribeModelExplainabilityJobDefinition (p. 520)
• DescribeModelPackage (p. 524)
• DescribeModelPackageGroup (p. 533)
• DescribeModelQualityJobDefinition (p. 536)
• DescribeMonitoringSchedule (p. 540)
• DescribeNotebookInstance (p. 545)
• DescribeNotebookInstanceLifecycleConfig (p. 550)
• DescribePipeline (p. 553)
• DescribePipelineDefinitionForExecution (p. 557)
• DescribePipelineExecution (p. 559)
• DescribeProcessingJob (p. 563)
• DescribeProject (p. 569)
• DescribeSpace (p. 573)
• DescribeStudioLifecycleConfig (p. 578)
• DescribeSubscribedWorkteam (p. 581)
• DescribeTrainingJob (p. 583)
• DescribeTransformJob (p. 594)
• DescribeTrial (p. 599)
• DescribeTrialComponent (p. 602)
• DescribeUserProfile (p. 607)
• DescribeWorkforce (p. 613)
• DescribeWorkteam (p. 615)
• DisableSagemakerServicecatalogPortfolio (p. 617)
• DisassociateTrialComponent (p. 618)
• EnableSagemakerServicecatalogPortfolio (p. 620)
• GetDeviceFleetReport (p. 621)
• GetLineageGroupPolicy (p. 624)
• GetModelPackageGroupPolicy (p. 626)
• GetSagemakerServicecatalogPortfolioStatus (p. 628)
• GetScalingConfigurationRecommendation (p. 629)
• GetSearchSuggestions (p. 633)
• ImportHubContent (p. 635)
• ListActions (p. 639)
• ListAlgorithms (p. 642)
• ListAliases (p. 645)
• ListAppImageConfigs (p. 648)
• ListApps (p. 652)
• ListArtifacts (p. 655)
• ListAssociations (p. 658)
• ListAutoMLJobs (p. 662)
• ListCandidatesForAutoMLJob (p. 666)
• ListClusters (p. 669)
• ListCodeRepositories (p. 677)
• ListCompilationJobs (p. 681)
• ListContexts (p. 685)
• ListDataQualityJobDefinitions (p. 688)
• ListDeviceFleets (p. 691)
• ListDevices (p. 694)
• ListDomains (p. 697)
• ListEdgeDeploymentPlans (p. 699)
• ListEdgePackagingJobs (p. 703)
• ListEndpointConfigs (p. 707)
• ListEndpoints (p. 710)
• ListExperiments (p. 713)
• ListFeatureGroups (p. 716)
• ListFlowDefinitions (p. 719)
• ListHubContents (p. 722)
• ListHubContentVersions (p. 726)
• ListHubs (p. 730)
• ListHumanTaskUis (p. 733)
• ListHyperParameterTuningJobs (p. 736)
• ListImages (p. 740)
• ListImageVersions (p. 743)
• ListInferenceComponents (p. 746)
• ListInferenceExperiments (p. 750)
• ListInferenceRecommendationsJobs (p. 754)
• ListInferenceRecommendationsJobSteps (p. 758)
• ListLabelingJobs (p. 762)
• ListLabelingJobsForWorkteam (p. 766)
• ListLineageGroups (p. 769)
• ListModelBiasJobDefinitions (p. 772)
• ListModelCardExportJobs (p. 775)
• ListModelCards (p. 779)
• ListModelCardVersions (p. 782)
• ListModelExplainabilityJobDefinitions (p. 785)
• ListModelMetadata (p. 788)
• ListModelPackageGroups (p. 790)
*ListModelPackages (p. 793)*
*ListModelQualityJobDefinitions (p. 797)*
*ListModel (p. 800)*
*ListMonitoringAlertHistory (p. 803)*
*ListMonitoringAlerts (p. 807)*
*ListMonitoringExecutions (p. 810)*
*ListMonitoringSchedules (p. 814)*
*ListNotebookInstanceLifecycleConfigs (p. 818)*
*ListNotebookInstances (p. 821)*
*ListPipelineExecutions (p. 825)*
*ListPipelineExecutionSteps (p. 828)*
*ListPipelineParametersForExecution (p. 832)*
*ListPipelines (p. 835)*
*ListProcessingJobs (p. 838)*
*ListProjects (p. 841)*
*ListResourceCatalogs (p. 844)*
*ListSpaces (p. 847)*
*ListStageDevices (p. 850)*
*ListStudioLifecycleConfigs (p. 853)*
*ListSubscribedWorkteams (p. 857)*
*ListTags (p. 859)*
*ListTrainingJobs (p. 861)*
*ListTrainingJobsForHyperParameterTuningJob (p. 865)*
*ListTransformJobs (p. 868)*
*ListTrialComponents (p. 871)*
*ListTrials (p. 875)*
*ListUserProfiles (p. 878)*
*ListWorkforces (p. 881)*
*ListWorkteams (p. 884)*
*PutModelPackageGroupPolicy (p. 887)*
*QueryLineage (p. 889)*
*RegisterDevices (p. 893)*
*RenderUiTemplate (p. 895)*
*RetryPipelineExecution (p. 898)*
*Search (p. 901)*
*SendPipelineExecutionStepFailure (p. 940)*
*SendPipelineExecutionStepSuccess (p. 942)*
*StartEdgeDeploymentStage (p. 944)*
*StartInferenceExperiment (p. 946)*
*StartMonitoringSchedule (p. 948)*
*StartNotebookInstance (p. 950)*
*StartPipelineExecution (p. 952)*
*StopAutoMLJob (p. 955)*
*StopCompilationJob (p. 957)*
*StopEdgeDeploymentStage (p. 959)*
*StopEdgePackagingJob (p. 961)
• StopHyperParameterTuningJob (p. 962)
• StopInferenceExperiment (p. 964)
• StopInferenceRecommendationsJob (p. 967)
• StopLabelingJob (p. 969)
• StopMonitoringSchedule (p. 971)
• StopNotebookInstance (p. 973)
• StopPipelineExecution (p. 975)
• StopProcessingJob (p. 977)
• StopTrainingJob (p. 979)
• StopTransformJob (p. 981)
• UpdateAction (p. 983)
• UpdateAppImageConfig (p. 986)
• UpdateArtifact (p. 988)
• UpdateCluster (p. 991)
• UpdateCodeRepository (p. 994)
• UpdateContext (p. 996)
• UpdateDeviceFleet (p. 999)
• UpdateDevices (p. 1001)
• UpdateDomain (p. 1003)
• UpdateEndpoint (p. 1009)
• UpdateEndpointWeightsAndCapacities (p. 1013)
• UpdateExperiment (p. 1015)
• UpdateFeatureGroup (p. 1017)
• UpdateFeatureMetadata (p. 1020)
• UpdateHub (p. 1022)
• UpdateImage (p. 1025)
• UpdateImageVersion (p. 1028)
• UpdateInferenceComponent (p. 1032)
• UpdateInferenceComponentRuntimeConfig (p. 1035)
• UpdateInferenceExperiment (p. 1037)
• UpdateModelCard (p. 1040)
• UpdateModelPackage (p. 1043)
• UpdateMonitoringAlert (p. 1046)
• UpdateMonitoringSchedule (p. 1049)
• UpdateNotebookInstance (p. 1053)
• UpdateNotebookInstanceLifecycleConfig (p. 1058)
• UpdatePipeline (p. 1060)
• UpdatePipelineExecution (p. 1063)
• UpdateProject (p. 1066)
• UpdateSpace (p. 1069)
• UpdateTrainingJob (p. 1073)
• UpdateTrial (p. 1076)
• UpdateTrialComponent (p. 1078)
• UpdateUserProfile (p. 1082)
• UpdateWorkforce (p. 1087)
• UpdateWorkteam (p. 1090)
AddAssociation
Service: Amazon SageMaker Service

Creates an association between the source and the destination. A source can be associated with multiple destinations, and a destination can be associated with multiple sources. An association is a lineage tracking entity. For more information, see Amazon SageMaker ML Lineage Tracking.

Request Syntax

```
{
  "AssociationType": "string",
  "DestinationArn": "string",
  "SourceArn": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**AssociationType (p. 18)**

The type of association. The following are suggested uses for each type. Amazon SageMaker places no restrictions on their use.

- **ContributedTo** - The source contributed to the destination or had a part in enabling the destination. For example, the training data contributed to the training job.
- **AssociatedWith** - The source is connected to the destination. For example, an approval workflow is associated with a model deployment.
- **DerivedFrom** - The destination is a modification of the source. For example, a digest output of a channel input for a processing job is derived from the original inputs.
- **Produced** - The source generated the destination. For example, a training job produced a model artifact.

Type: String

Valid Values: ContributedTo | AssociatedWith | DerivedFrom | Produced

Required: No

**DestinationArn (p. 18)**

The Amazon Resource Name (ARN) of the destination.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:(experiment|experiment-trial-component|artifact|action|context)/.*

Required: Yes

**SourceArn (p. 18)**

The ARN of the source.

Type: String
AddAssociation

Length Constraints: Maximum length of 256.

Pattern: `arn:aws[a-z-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:(experiment|experiment-trial-component|artifact|action|context)/.*`

Required: Yes

Response Syntax

```
{
    "DestinationArn": "string",
    "SourceArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**DestinationArn (p. 19)**

The Amazon Resource Name (ARN) of the destination.

Type: String

Length Constraints: Maximum length of 256.

Pattern: `arn:aws[a-z-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:(experiment|experiment-trial-component|artifact|action|context)/.*`

**SourceArn (p. 19)**

The ARN of the source.

Type: String

Length Constraints: Maximum length of 256.

Pattern: `arn:aws[a-z-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:(experiment|experiment-trial-component|artifact|action|context)/.*`

Errors

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](#).

**ResourceLimitExceeded**

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400
See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
AddTags
Service: Amazon SageMaker Service

Adds or overwrites one or more tags for the specified SageMaker resource. You can add tags to notebook instances, training jobs, hyperparameter tuning jobs, batch transform jobs, models, labeling jobs, work teams, endpoint configurations, and endpoints.

Each tag consists of a key and an optional value. Tag keys must be unique per resource. For more information about tags, see AWS Tagging Strategies.

Note
Tags that you add to a hyperparameter tuning job by calling this API are also added to any training jobs that the hyperparameter tuning job launches after you call this API, but not to training jobs that the hyperparameter tuning job launched before you called this API. To make sure that the tags associated with a hyperparameter tuning job are also added to all training jobs that the hyperparameter tuning job launches, add the tags when you first create the tuning job by specifying them in the Tags parameter of CreateHyperParameterTuningJob.

Note
Tags that you add to a SageMaker Domain or User Profile by calling this API are also added to any Apps that the Domain or User Profile launches after you call this API, but not to Apps that the Domain or User Profile launched before you called this API. To make sure that the tags associated with a Domain or User Profile are also added to all Apps that the Domain or User Profile launches, add the tags when you first create the Domain or User Profile by specifying them in the Tags parameter of CreateDomain or CreateUserProfile.

Request Syntax

```json
{
  "ResourceArn": "string",
  "Tags": [
    {
      "Key": "string",
      "Value": "string"
    }
  ]
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

ResourceArn (p. 21)

The Amazon Resource Name (ARN) of the resource that you want to tag.

Type: String

Length Constraints: Maximum length of 256.

Pattern: `arn:aws[a-z-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:.` +

Required: Yes

Tags (p. 21)

An array of key-value pairs. You can use tags to categorize your AWS resources in different ways, for example, by purpose, owner, or environment. For more information, see Tagging AWS Resources.
AddTags

Type: Array of Tag objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: Yes

Response Syntax

```json
{
  "Tags": [
    {
      "Key": "string",
      "Value": "string"
    }
  ]
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

Tags (p. 22)

A list of tags associated with the SageMaker resource.

Type: Array of Tag objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3

22
**AssociateTrialComponent**

Service: Amazon SageMaker Service

Associates a trial component with a trial. A trial component can be associated with multiple trials. To disassociate a trial component from a trial, call the [DisassociateTrialComponent](#) API.

**Request Syntax**

```json
{
   "TrialComponentName": "string",
   "TrialName": "string"
}
```

**Request Parameters**

For information about the parameters that are common to all actions, see [Common Parameters](#).

The request accepts the following data in JSON format.

**TrialComponentName** *(p. 23)*

The name of the component to associated with the trial.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: `^[a-zA-Z0-9]*(-[a-zA-Z0-9]*){0,119}`

Required: Yes

**TrialName** *(p. 23)*

The name of the trial to associate with.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: `^[a-zA-Z0-9]*(-[a-zA-Z0-9]*){0,119}`

Required: Yes

**Response Syntax**

```json
{
   "TrialArn": "string",
   "TrialComponentArn": "string"
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.
**TrialArn (p. 23)**

The Amazon Resource Name (ARN) of the trial.

Type: String

Length Constraints: Maximum length of 256.

Pattern: `arn:aws[a-z-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:experiment-trial/.*`

**TrialComponentArn (p. 23)**

The Amazon Resource Name (ARN) of the trial component.

Type: String

Length Constraints: Maximum length of 256.

Pattern: `arn:aws[a-z-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:experiment-trial-component/.*`

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](#).

**ResourceLimitExceeded**

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
BatchDescribeModelPackage

Service: Amazon SageMaker Service

This action batch describes a list of versioned model packages

Request Syntax

```
{
  "ModelPackageArnList": [ "string" ]
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**ModelPackageArnList (p. 25)**

The list of Amazon Resource Name (ARN) of the model package groups.

Type: Array of strings

Array Members: Minimum number of 1 item. Maximum number of 100 items.

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(-cn|-us-gov)?:sagemaker:[a-z0-9\-]{9,16}:[0-9]{12}:model-package/[^\$]{1,2048}$`

Required: Yes

Response Syntax

```
{
  "BatchDescribeModelPackageErrorMap": {
    "string": {
      "ErrorCode": "string",
      "ErrorResponse": "string"
    }
  },
  "ModelPackageSummaries": {
    "string": {
      "CreationTime": number,
      "InferenceSpecification": {
        "Containers": [{
          "AdditionalS3DataSource": {
            "CompressionType": "string",
            "S3DataType": "string",
            "S3Uri": "string"
          },
          "ContainerHostname": "string",
          "Environment": {
            "string": "string"
          },
          "Framework": "string",
          "FrameworkVersion": "string"
        }
      }
    }
  }
}
```

25
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**BatchDescribeModelPackageErrorMap (p. 25)**

A map of the resource and BatchDescribeModelPackageError objects reporting the error associated with describing the model package.

Type: String to BatchDescribeModelPackageError (p. 1302) object map

Key Length Constraints: Minimum length of 1. Maximum length of 2048.

Key Pattern: ^arn:aws(-cn|-us-gov)?:sagemaker:[a-z0-9\-]{9,16}:[0-9]{12}:model-package/[^\s]{1,2048}$

**ModelPackageSummaries (p. 25)**

The summaries for the model package versions

Type: String to BatchDescribeModelPackageSummary (p. 1303) object map

Key Length Constraints: Minimum length of 1. Maximum length of 2048.

Key Pattern: ^arn:aws(-cn|-us-gov)?:sagemaker:[a-z0-9\-]{9,16}:[0-9]{12}:model-package/[^\s]{1,2048}$

**Errors**

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:
• AWS Command Line Interface
• AWS SDK for .NET
• AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
CreateAction
Service: Amazon SageMaker Service

Creates an action. An action is a lineage tracking entity that represents an action or activity. For example, a model deployment or an HPO job. Generally, an action involves at least one input or output artifact. For more information, see Amazon SageMaker ML Lineage Tracking.

Request Syntax

```
{
    "ActionName": "string",
    "ActionType": "string",
    "Description": "string",
    "MetadataProperties": {
        "CommitId": "string",
        "GeneratedBy": "string",
        "ProjectId": "string",
        "Repository": "string"
    },
    "Properties": {
        "string": "string"
    },
    "Source": {
        "SourceId": "string",
        "SourceType": "string",
        "SourceUri": "string"
    },
    "Status": "string",
    "Tags": [
        {
            "Key": "string",
            "Value": "string"
        }
    ]
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**ActionName (p. 28)**

The name of the action. Must be unique to your account in an AWS Region.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: ^[a-zA-Z0-9][-]*[a-zA-Z0-9]{0,119}

Required: Yes

**ActionType (p. 28)**

The action type.

Type: String
Length Constraints: Maximum length of 256.

Required: Yes

**Description (p. 28)**

The description of the action.

Type: String

Length Constraints: Maximum length of 3072.

Pattern: .*

Required: No

**MetadataProperties (p. 28)**

Metadata properties of the tracking entity, trial, or trial component.

Type: MetadataProperties (p. 1648) object

Required: No

**Properties (p. 28)**

A list of properties to add to the action.

Type: String to string map

Map Entries: Maximum number of 30 items.

Key Length Constraints: Maximum length of 256.

Key Pattern: .*

Value Length Constraints: Maximum length of 256.

Value Pattern: .*

Required: No

**Source (p. 28)**

The source type, ID, and URI.

Type: ActionSource (p. 1216) object

Required: Yes

**Status (p. 28)**

The status of the action.

Type: String

Valid Values: Unknown | InProgress | Completed | Failed | Stopping | Stopped

Required: No

**Tags (p. 28)**

A list of tags to apply to the action.

Type: Array of Tag (p. 1979) objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.
Required: No

Response Syntax

```json
{
    "ActionArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.
The following data is returned in JSON format by the service.

**ActionArn (p. 30)**

The Amazon Resource Name (ARN) of the action.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]\{12\}:action/.*

Errors

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](#).

**ResourceLimitExceeded**

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
CreateAlgorithm
Service: Amazon SageMaker Service

Create a machine learning algorithm that you can use in SageMaker and list in the AWS Marketplace.

Request Syntax

```
{
   "AlgorithmDescription": "string",
   "AlgorithmName": "string",
   "CertifyForMarketplace": boolean,
   "InferenceSpecification": {
      "Containers": [
         {
            "AdditionalS3DataSource": {
               "CompressionType": "string",
               "S3DataType": "string",
               "S3Uri": "string"
            },
            "ContainerHostname": "string",
            "Environment": {
               "string": "string"
            },
            "Framework": "string",
            "FrameworkVersion": "string",
            "Image": "string",
            "ImageDigest": "string",
            "ModelDataUrl": "string",
            "ModelInput": {
               "DataInputConfig": "string"
            },
            "NearestModelName": "string",
            "ProductID": "string"
         }
      ],
      "SupportedContentTypes": [ "string" ],
      "SupportedRealtimeInferenceInstanceTypes": [ "string" ],
      "SupportedResponseMIMETypes": [ "string" ],
      "SupportedTransformInstanceTypes": [ "string" ]
   },
   "Tags": [
      {
         "Key": "string",
         "Value": "string"
      }
   ],
   "TrainingSpecification": {
      "AdditionalS3DataSource": {
         "CompressionType": "string",
         "S3DataType": "string",
         "S3Uri": "string"
      },
      "MetricDefinitions": [
         {
            "Name": "string",
            "Regex": "string"
         }
      ],
      "SupportedHyperParameters": [
         {
            "DefaultValue": "string",
            "Description": "string",
            "IsRequired": boolean
         }
      ]
   }
}
```
"IsTunable": boolean,
"Name": "string",
"Range": {
  "CategoricalParameterRangeSpecification": {
    "Values": [ "string" ]
  },
  "ContinuousParameterRangeSpecification": {
    "MaxValue": "string",
    "MinValue": "string"
  },
  "IntegerParameterRangeSpecification": {
    "MaxValue": "string",
    "MinValue": "string"
  }
},
"Type": "string"
},
"SupportedTrainingInstanceTypes": [ "string" ],
"SupportedTuningJobObjectiveMetrics": [ {
  "MetricName": "string",
  "Type": "string"
} ],
"SupportsDistributedTraining": boolean,
"TrainingChannels": [ {
  "Description": "string",
  "IsRequired": boolean,
  "Name": "string",
  "SupportedCompressionTypes": [ "string" ],
  "SupportedContentTypes": [ "string" ],
  "SupportedInputModes": [ "string" ]
} ],
"TrainingImage": "string",
"TrainingImageDigest": "string"
},
"ValidationSpecification": { "ValidationProfiles": [ {
  "ProfileName": "string",
  "TrainingJobDefinition": {
    "HyperParameters": {
      "string": "string"
    },
    "InputDataConfig": [ {
      "ChannelName": "string",
      "CompressionType": "string",
      "ContentType": "string",
      "DataSource": {
        "FileSystemDataSource": {
          "DirectoryPath": "string",
          "FileSystemAccessMode": "string",
          "FileSystemId": "string",
          "FileSystemType": "string"
        },
        "S3DataSource": {
          "AttributeNames": [ "string" ],
          "InstanceGroupNames": [ "string" ],
          "S3DataDistributionType": "string",
          "S3DataType": "string",
          "S3Uri": "string"
        }
      }
    }
  }
} ]}


```
],
"InputMode": "string",
"RecordWrapperType": "string",
"ShuffleConfig": {
  "Seed": number
}
},
"OutputDataConfig": {
  "CompressionType": "string",
  "KmsKeyId": "string",
  "S3OutputPath": "string"
},
"ResourceConfig": {
  "InstanceCount": number,
  "InstanceGroups": [
    {
      "InstanceCount": number,
      "InstanceGroupName": "string",
      "InstanceType": "string"
    }
  ],
  "InstanceType": "string",
  "KeepAlivePeriodInSeconds": number,
  "VolumeKmsKeyId": "string",
  "VolumeSizeInGB": number
},
"StoppingCondition": {
  "MaxPendingTimeInSeconds": number,
  "MaxRuntimeInSeconds": number,
  "MaxWaitTimeInSeconds": number
},
"TrainingInputMode": "string"
},
"TransformJobDefinition": {
  "BatchStrategy": "string",
  "Environment": {
    "string": "string"
  },
  "MaxConcurrentTransforms": number,
  "MaxPayloadInMB": number,
  "TransformInput": {
    "CompressionType": "string",
    "ContentType": "string",
    "DataSource": {
      "S3DataSource": {
        "S3DataType": "string",
        "S3Uri": "string"
      }
    },
    "SplitType": "string"
  },
  "TransformOutput": {
    "Accept": "string",
    "AssembleWith": "string",
    "KmsKeyId": "string",
    "S3OutputPath": "string"
  },
  "TransformResources": {
    "InstanceCount": number,
    "InstanceType": "string",
    "VolumeKmsKeyId": "string"
  }
}
```
Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**AlgorithmDescription (p. 31)**

A description of the algorithm.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: \[\p{L}\p{M}\p{Z}\p{S}\p{N}\p{P}]\* 

Required: No

**AlgorithmName (p. 31)**

The name of the algorithm.

Type: String


Pattern: ^\[a-zA-Z0-9\](-*[a-zA-Z0-9]\{0,62}\}$

Required: Yes

**CertifyForMarketplace (p. 31)**

Whether to certify the algorithm so that it can be listed in AWS Marketplace.

Type: Boolean

Required: No

**InferenceSpecification (p. 31)**

Specifies details about inference jobs that the algorithm runs, including the following:
- The Amazon ECR paths of containers that contain the inference code and model artifacts.
- The instance types that the algorithm supports for transform jobs and real-time endpoints used for inference.
- The input and output content formats that the algorithm supports for inference.

Type: InferenceSpecification (p. 1601) object

Required: No

**Tags (p. 31)**

An array of key-value pairs. You can use tags to categorize your AWS resources in different ways, for example, by purpose, owner, or environment. For more information, see Tagging AWS Resources.

Type: Array of Tag (p. 1979) objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.
Required: No

**TrainingSpecification (p. 31)**

Specifies details about training jobs run by this algorithm, including the following:

- The Amazon ECR path of the container and the version digest of the algorithm.
- The hyperparameters that the algorithm supports.
- The instance types that the algorithm supports for training.
- Whether the algorithm supports distributed training.
- The metrics that the algorithm emits to Amazon CloudWatch.
- Which metrics that the algorithm emits can be used as the objective metric for hyperparameter tuning jobs.
- The input channels that the algorithm supports for training data. For example, an algorithm might support `train`, `validation`, and `test` channels.

Type: *TrainingSpecification (p. 2017)* object

Required: Yes

**ValidationSpecification (p. 31)**

Specifies configurations for one or more training jobs and that SageMaker runs to test the algorithm's training code and, optionally, one or more batch transform jobs that SageMaker runs to test the algorithm's inference code.

Type: *AlgorithmValidationSpecification (p. 1233)* object

Required: No

**Response Syntax**

```
{
    "AlgorithmArn": "string"
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**AlgorithmArn (p. 35)**

The Amazon Resource Name (ARN) of the new algorithm.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: ^arn:aws(-cn|-us-gov)?:sagemaker:[a-z0-9\-]{9,16}:[0-9]\{12\}:algorithm/\[\S\]{1,2048}$

**Errors**

For information about the errors that are common to all actions, see *Common Errors (p. 2180).*
See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
CreateApp
Service: Amazon SageMaker Service

Creates a running app for the specified UserProfile. This operation is automatically invoked by Amazon SageMaker upon access to the associated Domain, and when new kernel configurations are selected by the user. A user may have multiple Apps active simultaneously.

Request Syntax

```
{
    "AppName": "string",
    "AppType": "string",
    "DomainId": "string",
    "ResourceSpec": {
        "InstanceType": "string",
        "LifecycleConfigArn": "string",
        "SageMakerImageArn": "string",
        "SageMakerImageVersionAlias": "string",
        "SageMakerImageVersionArn": "string"
    },
    "SpaceName": "string",
    "Tags": [
        {
            "Key": "string",
            "Value": "string"
        }
    ],
    "UserProfileName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

AppName (p. 37)

The name of the app.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](\-*[a-zA-Z0-9])\{0,62}$

Required: Yes

AppType (p. 37)

The type of app.

Type: String

Valid Values: JupyterServer | KernelGateway | TensorBoard | RStudioServerPro | RSessionGateway | JupyterLab | CodeEditor

Required: Yes
DomainId (p. 37)

The domain ID.

Type: String

Length Constraints: Maximum length of 63.

Required: Yes

ResourceSpec (p. 37)

The instance type and the Amazon Resource Name (ARN) of the SageMaker image created on the instance.

Note

The value of InstanceType passed as part of the ResourceSpec in the CreateApp call overrides the value passed as part of the ResourceSpec configured for the user profile or the domain. If InstanceType is not specified in any of those three ResourceSpec values for a KernelGateway app, the CreateApp call fails with a request validation error.

Type: ResourceSpec (p. 1915) object

Required: No

SpaceName (p. 37)

The name of the space. If this value is not set, then UserProfileName must be set.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}

Required: No

Tags (p. 37)

Each tag consists of a key and an optional value. Tag keys must be unique per resource.

Type: Array of Tag (p. 1979) objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

UserProfileName (p. 37)

The user profile name. If this value is not set, then SpaceName must be set.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}

Required: No

Response Syntax

{  
  "AppArn": "string"  
}
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**AppArn (p. 38)**

The Amazon Resource Name (ARN) of the app.

Type: String

Length Constraints: Maximum length of 256.

Pattern: `arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:app/.*`

Errors

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](https://docs.aws.amazon.com/en_us/sagemaker/latest/dg/creating-app.html).

**ResourceInUse**

Resource being accessed is in use.

HTTP Status Code: 400

**ResourceLimitExceeded**

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for .NET](https://docs.aws.amazon.com/en_us/sagemaker/latest/dg/creating-app.html)
- [AWS SDK for C++](https://docs.aws.amazon.com/en_us/sagemaker/latest/dg/creating-app.html)
- [AWS SDK for Go](https://docs.aws.amazon.com/en_us/sagemaker/latest/dg/creating-app.html)
- [AWS SDK for Java V2](https://docs.aws.amazon.com/en_us/sagemaker/latest/dg/creating-app.html)
- [AWS SDK for JavaScript V3](https://docs.aws.amazon.com/en_us/sagemaker/latest/dg/creating-app.html)
- [AWS SDK for PHP V3](https://docs.aws.amazon.com/en_us/sagemaker/latest/dg/creating-app.html)
- [AWS SDK for Python](https://docs.aws.amazon.com/en_us/sagemaker/latest/dg/creating-app.html)
CreateAppImageConfig
Service: Amazon SageMaker Service

Creates a configuration for running a SageMaker image as a KernelGateway app. The configuration specifies the Amazon Elastic File System (EFS) storage volume on the image, and a list of the kernels in the image.

Request Syntax

```json
{
    "AppImageConfigName": "string",
    "JupyterLabAppImageConfig": {
        "ContainerConfig": {
            "ContainerArguments": [ "string" ],
            "ContainerEntrypoint": [ "string" ],
            "ContainerEnvironmentVariables": {
                "string": "string"
            }
        }
    },
    "KernelGatewayImageConfig": {
        "FileSystemConfig": {
            "DefaultGid": number,
            "DefaultUid": number,
            "MountPath": "string"
        },
        "KernelSpecs": [
            { "DisplayName": "string", "Name": "string" }
        ],
        "Tags": [
            { "Key": "string", "Value": "string" }
        ]
    }
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 217).

The request accepts the following data in JSON format.

**AppImageConfigName (p. 40)**

The name of the AppImageConfig. Must be unique to your account.

Type: String

Length Constraints: Maximum length of 63.

Pattern: `^\([a-zA-Z0-9-]*[a-zA-Z0-9]\)\([^a-zA-Z0-9]\)\{0,62\}`

Required: Yes
**JupyterLabAppImageConfig (p. 40)**

The JupyterLabAppImageConfig. You can only specify one image kernel in the AppImageConfig API. This kernel is shown to users before the image starts. After the image runs, all kernels are visible in JupyterLab.

Type: [JupyterLabAppImageConfig (p. 1615)] object

Required: No

**KernelGatewayImageConfig (p. 40)**

The KernelGatewayImageConfig. You can only specify one image kernel in the AppImageConfig API. This kernel will be shown to users before the image starts. Once the image runs, all kernels are visible in JupyterLab.

Type: [KernelGatewayImageConfig (p. 1621)] object

Required: No

**Tags (p. 40)**

A list of tags to apply to the AppImageConfig.

Type: Array of [Tag (p. 1979)] objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

**Response Syntax**

```json
{
   "AppImageConfigArn": "string"
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**AppImageConfigArn (p. 41)**

The Amazon Resource Name (ARN) of the AppImageConfig.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-\*:sagemaker:[a-z0-9\-\*:\[0-9][12]:app-image-config/.*

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)].

**ResourceInUse**

Resource being accessed is in use.
HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
CreateArtifact
Service: Amazon SageMaker Service

Creates an artifact. An artifact is a lineage tracking entity that represents a URI addressable object or data. Some examples are the S3 URI of a dataset and the ECR registry path of an image. For more information, see Amazon SageMaker ML Lineage Tracking.

Request Syntax

```json
{
    "ArtifactName": "string",
    "ArtifactType": "string",
    "MetadataProperties": {
        "CommitId": "string",
        "GeneratedBy": "string",
        "ProjectId": "string",
        "Repository": "string"
    },
    "Properties": {
        "string": "string"
    },
    "Source": {
        "SourceTypes": [
            {
                "SourceIdType": "string",
                "Value": "string"
            }
        ],
        "SourceUri": "string"
    },
    "Tags": [
        {
            "Key": "string",
            "Value": "string"
        }
    ]
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**ArtifactName (p. 43)**

The name of the artifact. Must be unique to your account in an AWS Region.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: ^[a-zA-Z0-9](\-*[a-zA-Z0-9])\{0,119}$

Required: No

**ArtifactType (p. 43)**

The artifact type.
CreateArtifact

Type: String
Length Constraints: Maximum length of 256.
Required: Yes

**MetadataProperties (p. 43)**
Metadata properties of the tracking entity, trial, or trial component.
Type: MetadataProperties (p. 1648) object
Required: No

**Properties (p. 43)**
A list of properties to add to the artifact.
Type: String to string map
Map Entries: Maximum number of 30 items.
Key Length Constraints: Maximum length of 256.
Key Pattern: .*
Value Length Constraints: Maximum length of 256.
Value Pattern: .*
Required: No

**Source (p. 43)**
The ID, ID type, and URI of the source.
Type: ArtifactSource (p. 1251) object
Required: Yes

**Tags (p. 43)**
A list of tags to apply to the artifact.
Type: Array of Tag (p. 1979) objects
Array Members: Minimum number of 0 items. Maximum number of 50 items.
Required: No

**Response Syntax**

```json
{
   "ArtifactArn": "string"
}
```

**Response Elements**
If the action is successful, the service sends back an HTTP 200 response.
The following data is returned in JSON format by the service.
**ArtifactArn (p. 44)**

The Amazon Resource Name (ARN) of the artifact.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:artifact/.*

**Errors**

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceLimitExceeded**

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
CreateAutoMLJob
Service: Amazon SageMaker Service

Creates an Autopilot job also referred to as Autopilot experiment or AutoML job.

**Note**
We recommend using the new versions `CreateAutoMLJobV2` and `DescribeAutoMLJobV2`, which offer backward compatibility.

CreateAutoMLJobV2 can manage tabular problem types identical to those of its previous version CreateAutoMLJob, as well as time-series forecasting, non-tabular problem types such as image or text classification, and text generation (LLMs fine-tuning).

Find guidelines about how to migrate a CreateAutoMLJob to CreateAutoMLJobV2 in [Migrate a CreateAutoMLJob to CreateAutoMLJobV2](#).

You can find the best-performing model after you run an AutoML job by calling `DescribeAutoMLJobV2` (recommended) or `DescribeAutoMLJob`.

**Request Syntax**

```json
{
    "AutoMLJobConfig": {
        "CandidateGenerationConfig": {
            "AlgorithmsConfig": [
                "AutoMLAlgorithms": [ "string" ]
            ],
            "FeatureSpecificationS3Uri": "string"
        },
        "CompletionCriteria": {
            "MaxAutoMLJobRuntimeInSeconds": number,
            "MaxCandidates": number,
            "MaxRuntimePerTrainingJobInSeconds": number
        },
        "DataSplitConfig": {
            "ValidationFraction": number
        },
        "Mode": "string",
        "SecurityConfig": {
            "EnableInterContainerTrafficEncryption": boolean,
            "VolumeKmsKeyId": "string",
            "VpcConfig": {
                "SecurityGroupId": [ "string" ],
                "Subnets": [ "string" ]
            }
        }
    },
    "AutoMLJobName": "string",
    "AutoMLJobObjective": {
        "MetricName": "string"
    },
    "GenerateCandidateDefinitionsOnly": boolean,
    "InputDataConfig": [
        { "SourceType": "string",
        "CompressionType": "string",
        "ContentType": "string",
        "DataSource": { "S3DataSource": { "S3DataType": "string", "S3Uri": "string" } }
    }]
}
```
Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**AutoMLJobConfig (p. 46)**

A collection of settings used to configure an AutoML job.

Type: AutoMLJobConfig (p. 1282) object

Required: No

**AutoMLJobName (p. 46)**

Identifies an Autopilot job. The name must be unique to your account and is case insensitive.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,31}$

Required: Yes

**AutoMLJobObjective (p. 46)**

Specifies a metric to minimize or maximize as the objective of a job. If not specified, the default objective metric depends on the problem type. See AutoMLJobObjective for the default values.

Type: AutoMLJobObjective (p. 1284) object

Required: No

**GenerateCandidateDefinitionsOnly (p. 46)**

Generates possible candidates without training the models. A candidate is a combination of data preprocessors, algorithms, and algorithm parameter settings.
CreateAutoMLJob

Type: Boolean

Required: No

**InputDataConfig (p. 46)**

An array of channel objects that describes the input data and its location. Each channel is a named input source. Similar to InputDataConfig supported by HyperParameterTrainingJobDefinition. Format(s) supported: CSV, Parquet. A minimum of 500 rows is required for the training dataset. There is not a minimum number of rows required for the validation dataset.

Type: Array of AutoMLChannel (p. 1271) objects

Array Members: Minimum number of 1 item. Maximum number of 2 items.

Required: Yes

**ModelDeployConfig (p. 46)**

Specifies how to generate the endpoint name for an automatic one-click Autopilot model deployment.

Type: ModelDeployConfig (p. 1689) object

Required: No

**OutputDataConfig (p. 46)**

Provides information about encryption and the Amazon S3 output path needed to store artifacts from an AutoML job. Format(s) supported: CSV.

Type: AutoMLOutputDataConfig (p. 1289) object

Required: Yes

**ProblemType (p. 46)**

Defines the type of supervised learning problem available for the candidates. For more information, see Amazon SageMaker Autopilot problem types.

Type: String

Valid Values: BinaryClassification | MulticlassClassification | Regression

Required: No

**RoleArn (p. 46)**

The ARN of the role that is used to access the data.

Type: String


Pattern: ^arn:aws[a-z\-]*:iam::\d{12}:role/?[a-zA-Z_0-9+=,.@\-_/>]+$?

Required: Yes

**Tags (p. 46)**

An array of key-value pairs. You can use tags to categorize your AWS resources in different ways, for example, by purpose, owner, or environment. For more information, see Tagging AWSResources. Tag keys must be unique per resource.

Type: Array of Tag (p. 1979) objects
Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

Response Syntax

```json
{
   "AutoMLJobArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**AutoMLJobArn (p. 49)**

The unique ARN assigned to the AutoML job when it is created.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 256.

Pattern: `arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:automl-job/.*`

Errors

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](#).

**ResourceInUse**

Resource being accessed is in use.

HTTP Status Code: 400

**ResourceLimitExceeded**

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
• AWS SDK for Python
• AWS SDK for Ruby V3
CreateAutoMLJobV2

Create an Autopilot job also referred to as Autopilot experiment or AutoML job V2.

**Note**

CreateAutoMLJobV2 and DescribeAutoMLJobV2 are new versions of CreateAutoMLJob and DescribeAutoMLJob which offer backward compatibility. CreateAutoMLJobV2 can manage tabular problem types identical to those of its previous version CreateAutoMLJob, as well as time-series forecasting, non-tabular problem types such as image or text classification, and text generation (LLMs fine-tuning).

Find guidelines about how to migrate a CreateAutoMLJob to CreateAutoMLJobV2 in [Migrate a CreateAutoMLJob to CreateAutoMLJobV2](#).

For the list of available problem types supported by CreateAutoMLJobV2, see [AutoMLProblemTypeConfig](#).

You can find the best-performing model after you run an AutoML job V2 by calling [DescribeAutoMLJobV2](#).

### Request Syntax

```json
{
  "AutoMLJobInputDataConfig": [
    {
      "ChannelType": "string",
      "CompressionType": "string",
      "ContentType": "string",
      "DataSource": {
        "S3DataSource": {
          "S3DataType": "string",
          "S3Uri": "string"
        }
      }
    },
    ...
  ],
  "AutoMLJobName": "string",
  "AutoMLJobObjective": {
    "MetricName": "string"
  },
  "AutoMLProblemTypeConfig": {...}
  "DataSplitConfig": {
    "ValidationFraction": number
  },
  "ModelDeployConfig": {
    "AutoGenerateEndpointName": boolean,
    "EndpointName": "string"
  },
  "OutputDataConfig": {
    "KmsKeyId": "string",
    "S3OutputPath": "string"
  },
  "RoleArn": "string",
  "SecurityConfig": {
    "EnableInterContainerTrafficEncryption": boolean,
    "VolumeKmsKeyId": "string",
    "VpcConfig": {
      "SecurityGroupIds": [ "string" ],
      "Subnets": [ "string" ]
    }
  },
  "Tags": [ ]
}
```
Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**AutoMLJobInputDataConfig (p. 51)**

An array of channel objects describing the input data and their location. Each channel is a named input source. Similar to the `InputDataConfig` attribute in the `CreateAutoMLJob` input parameters. The supported formats depend on the problem type:

- For tabular problem types: `S3Prefix, ManifestFile`.
- For image classification: `S3Prefix, ManifestFile, AugmentedManifestFile`.
- For text classification: `S3Prefix`.
- For time-series forecasting: `S3Prefix`.
- For text generation (LLMs fine-tuning): `S3Prefix`.

Type: Array of `AutoMLJobChannel` (p. 1278) objects

Array Members: Minimum number of 1 item. Maximum number of 2 items.

Required: Yes

**AutoMLJobName (p. 51)**

Identifies an Autopilot job. The name must be unique to your account and is case insensitive.

Type: String


Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,31}$`

Required: Yes

**AutoMLJobObjective (p. 51)**

Specifies a metric to minimize or maximize as the objective of a job. If not specified, the default objective metric depends on the problem type. For the list of default values per problem type, see `AutoMLJobObjective`.

Note

- For tabular problem types: You must either provide both the `AutoMLJobObjective` and indicate the type of supervised learning problem in `AutoMLProblemTypeConfig` (TabularJobConfig.ProblemType), or none at all.
- For text generation problem types (LLMs fine-tuning): Fine-tuning language models in Autopilot does not require setting the `AutoMLJobObjective` field. Autopilot fine-tunes LLMs without requiring multiple candidates to be trained and evaluated. Instead, using your dataset, Autopilot directly fine-tunes your target model to enhance a default objective metric, the cross-entropy loss. After fine-tuning a language model, you can evaluate the quality of its generated text using different metrics. For a list of the available metrics, see Metrics for fine-tuning LLMs in Autopilot.
Type: **AutoMLJobObjective** *(p. 1284)* object

Required: No

**AutoMLProblemTypeConfig** *(p. 51)*

Defines the configuration settings of one of the supported problem types.

Type: **AutoMLProblemTypeConfig** *(p. 1291)* object

**Note**: This object is a Union. Only one member of this object can be specified or returned.

Required: Yes

**DataSplitConfig** *(p. 51)*

This structure specifies how to split the data into train and validation datasets.

The validation and training datasets must contain the same headers. For jobs created by calling `CreateAutoMLJob`, the validation dataset must be less than 2 GB in size.

**Note**

This attribute must not be set for the time-series forecasting problem type, as Autopilot automatically splits the input dataset into training and validation sets.

Type: **AutoMLDataSplitConfig** *(p. 1276)* object

Required: No

**ModelDeployConfig** *(p. 51)*

Specifies how to generate the endpoint name for an automatic one-click Autopilot model deployment.

Type: **ModelDeployConfig** *(p. 1689)* object

Required: No

**OutputDataConfig** *(p. 51)*

Provides information about encryption and the Amazon S3 output path needed to store artifacts from an AutoML job.

Type: **AutoMLOutputDataConfig** *(p. 1289)* object

Required: Yes

**RoleArn** *(p. 51)*

The ARN of the role that is used to access the data.

Type: String


Pattern: `^arn:aws[a-z\-]*:iam::\d{12}:role/?[a-zA-Z0-9-_]*$`

Required: Yes

**SecurityConfig** *(p. 51)*

The security configuration for traffic encryption or Amazon VPC settings.

Type: **AutoMLSecurityConfig** *(p. 1297)* object

Required: No
Tags (p. 51)

An array of key-value pairs. You can use tags to categorize your AWS resources in different ways, such as by purpose, owner, or environment. For more information, see Tagging AWSResources. Tag keys must be unique per resource.

Type: Array of Tag (p. 1979) objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

Response Syntax

```
{
  "AutoMLJobArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

AutoMLJobArn (p. 54)

The unique ARN assigned to the AutoMLJob when it is created.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 256.

Pattern: arn:aws[a-zA-Z-]*:sagemaker:[a-zA-Z0-9-]*:[0-9]{12}:automl-job/.*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceInUse

Resource being accessed is in use.

HTTP Status Code: 400

ResourceLimitExceeded

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
• AWS SDK for .NET
• AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
CreateCluster

Service: Amazon SageMaker Service

Creates a SageMaker HyperPod cluster. SageMaker HyperPod is a capability of SageMaker for creating and managing persistent clusters for developing large machine learning models, such as large language models (LLMs) and diffusion models. To learn more, see Amazon SageMaker HyperPod in the Amazon SageMaker Developer Guide.

Request Syntax

```
{
    "ClusterName": "string",
    "InstanceGroups": [
        {
            "ExecutionRole": "string",
            "InstanceCount": number,
            "InstanceGroupName": "string",
            "InstanceType": "string",
            "LifeCycleConfig": {
                "OnCreate": "string",
                "SourceS3Uri": "string"
            },
            "ThreadsPerCore": number,
        }
    ],
    "Tags": [
        {
            "Key": "string",
            "Value": "string"
        }
    ],
    "VpcConfig": {
        "SecurityGroupIds": [ "string" ],
        "Subnets": [ "string" ]
    }
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**ClusterName (p. 56)**

The name for the new SageMaker HyperPod cluster.

Type: String


Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9])*$`

Required: Yes

**InstanceGroups (p. 56)**

The instance groups to be created in the SageMaker HyperPod cluster.

Type: Array of ClusterInstanceGroupSpecification (p. 1343) objects
Array Members: Minimum number of 1 item. Maximum number of 5 items.

Required: Yes

**Tags (p. 56)**

Custom tags for managing the SageMaker HyperPod cluster as an AWS resource. You can add tags to your cluster in the same way you add them in other AWS services that support tagging. To learn more about tagging AWS resources in general, see [Tagging AWS Resources User Guide](#).

Type: Array of Tag (p. 1979) objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

**VpcConfig (p. 56)**

Specifies an Amazon Virtual Private Cloud (VPC) that your SageMaker jobs, hosted models, and compute resources have access to. You can control access to and from your resources by configuring a VPC. For more information, see [Give SageMaker Access to Resources in your Amazon VPC](#).

Type: VpcConfig (p. 2076) object

Required: No

**Response Syntax**

```
{
  "ClusterArn": "string"
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**ClusterArn (p. 57)**

The Amazon Resource Name (ARN) of the cluster.

Type: String

Length Constraints: Maximum length of 256.

Pattern: ^arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:cluster/[a-z0-9]{12}$

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](#).

**ResourceInUse**

Resource being accessed is in use.

HTTP Status Code: 400
ResourceLimitExceeded

You have exceeded a SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
CreateCodeRepository

Service: Amazon SageMaker Service

Creates a Git repository as a resource in your SageMaker account. You can associate the repository with notebook instances so that you can use Git source control for the notebooks you create. The Git repository is a resource in your SageMaker account, so it can be associated with more than one notebook instance, and it persists independently from the lifecycle of any notebook instances it is associated with.

The repository can be hosted either in AWS CodeCommit or in any other Git repository.

Request Syntax

```json
{
   "CodeRepositoryName": "string",
   "GitConfig": {
      "Branch": "string",
      "RepositoryUrl": "string",
      "SecretArn": "string"
   },
   "Tags": [
      {
         "Key": "string",
         "Value": "string"
      }
   ]
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CodeRepositoryName (p. 59)**

The name of the Git repository. The name must have 1 to 63 characters. Valid characters are a-z, A-Z, 0-9, and - (hyphen).

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}$

Required: Yes

**GitConfig (p. 59)**

Specifies details about the repository, including the URL where the repository is located, the default branch, and credentials to use to access the repository.

Type: GitConfig (p. 1505) object

Required: Yes

**Tags (p. 59)**

An array of key-value pairs. You can use tags to categorize your AWS resources in different ways, for example, by purpose, owner, or environment. For more information, see Tagging AWS Resources.
CreateCodeRepository

Type: Array of Tag objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

Response Syntax

```json
{
   "CodeRepositoryArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**CodeRepositoryArn (p. 60)**

The Amazon Resource Name (ARN) of the new repository.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(-cn|-us-gov)?:sagemaker:[a-z0-9-]{9,16}:[0-9]{12}:code-repository/[\S]{1,2048}$`

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
CreateCompilationJob

Service: Amazon SageMaker Service

Starts a model compilation job. After the model has been compiled, Amazon SageMaker saves the resulting model artifacts to an Amazon Simple Storage Service (Amazon S3) bucket that you specify.

If you choose to host your model using Amazon SageMaker hosting services, you can use the resulting model artifacts as part of the model. You can also use the artifacts with AWS IoT Greengrass. In that case, deploy them as an ML resource.

In the request body, you provide the following:

- A name for the compilation job
- Information about the input model artifacts
- The output location for the compiled model and the device (target) that the model runs on
- The Amazon Resource Name (ARN) of the IAM role that Amazon SageMaker assumes to perform the model compilation job.

You can also provide a Tag to track the model compilation job’s resource use and costs. The response body contains the CompilationJobArn for the compiled job.

To stop a model compilation job, use StopCompilationJob. To get information about a particular model compilation job, use DescribeCompilationJob. To get information about multiple model compilation jobs, use ListCompilationJobs.

Request Syntax

```json
{
    "CompilationJobName": "string",
    "InputConfig": {
        "DataInputConfig": "string",
        "Framework": "string",
        "FrameworkVersion": "string",
        "S3Uri": "string"
    },
    "ModelPackageVersionArn": "string",
    "OutputConfig": {
        "CompilerOptions": "string",
        "KmsKeyId": "string",
        "S3OutputLocation": "string",
        "TargetDevice": "string",
        "TargetPlatform": {
            "Accelerator": "string",
            "Arch": "string",
            "Os": "string"
        }
    },
    "RoleArn": "string",
    "StoppingCondition": {
        "MaxPendingTimeInSeconds": number,
        "MaxRuntimeInSeconds": number,
        "MaxWaitTimeInSeconds": number
    },
    "Tags": [
        {
            "Key": "string",
            "Value": "string"
        }
    ]
}
```
"VpcConfig": {
  "SecurityGroupIds": [ "string" ],
  "Subnets": [ "string" ]
}

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CompilationJobName (p. 61)**

A name for the model compilation job. The name must be unique within the AWS Region and within your AWS account.

Type: String


Pattern: ^[a-zA-Z0-9][-\*[a-zA-Z0-9]]{0,62}$

Required: Yes

**InputConfig (p. 61)**

Provides information about the location of input model artifacts, the name and shape of the expected data inputs, and the framework in which the model was trained.

Type: InputConfig (p. 1605) object

Required: No

**ModelPackageVersionArn (p. 61)**

The Amazon Resource Name (ARN) of a versioned model package. Provide either a ModelPackageVersionArn or an InputConfig object in the request syntax. The presence of both objects in the CreateCompilationJob request will return an exception.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: ^arn:aws(-cn|-us-gov):sagemaker:([a-zA-Z0-9]([-*][a-zA-Z0-9])){0,62}$

Required: No

**OutputConfig (p. 61)**

Provides information about the output location for the compiled model and the target device the model runs on.

Type: OutputConfig (p. 1788) object

Required: Yes

**RoleArn (p. 61)**

The Amazon Resource Name (ARN) of an IAM role that enables Amazon SageMaker to perform tasks on your behalf.
During model compilation, Amazon SageMaker needs your permission to:

- Read input data from an S3 bucket
- Write model artifacts to an S3 bucket
- Write logs to Amazon CloudWatch Logs
- Publish metrics to Amazon CloudWatch

You grant permissions for all of these tasks to an IAM role. To pass this role to Amazon SageMaker, the caller of this API must have the `iam:PassRole` permission. For more information, see Amazon SageMaker Roles.

**Type:** String

**Length Constraints:** Minimum length of 20. Maximum length of 2048.

**Pattern:** ^arn:aws[a-z-]*:iam::\d{12}:role/[a-zA-Z_0-9+=,.@-_/]+$

**Required:** Yes

**StoppingCondition (p. 61)**

Specifies a limit to how long a model compilation job can run. When the job reaches the time limit, Amazon SageMaker ends the compilation job. Use this API to cap model training costs.

**Type:** StoppingCondition (p. 1968) object

**Required:** Yes

**Tags (p. 61)**

An array of key-value pairs. You can use tags to categorize your AWS resources in different ways, for example, by purpose, owner, or environment. For more information, see Tagging AWS Resources.

**Type:** Array of Tag (p. 1979) objects

**Array Members:** Minimum number of 0 items. Maximum number of 50 items.

**Required:** No

**VpcConfig (p. 61)**

A VpcConfig object that specifies the VPC that you want your compilation job to connect to. Control access to your models by configuring the VPC. For more information, see Protect Compilation Jobs by Using an Amazon Virtual Private Cloud.

**Type:** NeoVpcConfig (p. 1767) object

**Required:** No

**Response Syntax**

```
{
  "CompilationJobArn": "string"
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.
CompilationJobArn (p. 63)

If the action is successful, the service sends back an HTTP 200 response. Amazon SageMaker returns the following data in JSON format:

- CompilationJobArn: The Amazon Resource Name (ARN) of the compiled job.

  Type: String

  Length Constraints: Maximum length of 256.

  Pattern: arn:aws[a-z\-\*:sagemaker:[a-z0-9\-\*:\[0-9\]{12}:compilation-job/.*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceInUse

- Resource being accessed is in use.
- HTTP Status Code: 400

ResourceLimitExceeded

- You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.
- HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
CreateContext

Service: Amazon SageMaker Service

Creates a context. A context is a lineage tracking entity that represents a logical grouping of other tracking or experiment entities. Some examples are an endpoint and a model package. For more information, see Amazon SageMaker ML Lineage Tracking.

Request Syntax

```
{
  "ContextName": "string",
  "ContextType": "string",
  "Description": "string",
  "Properties": {
    "string": "string"
  },
  "Source": {
    "SourceId": "string",
    "SourceType": "string",
    "SourceUri": "string"
  },
  "Tags": [
    {
      "Key": "string",
      "Value": "string"
    }
  ]
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**ContextName** (p. 65)

The name of the context. Must be unique to your account in an AWS Region.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,119}

Required: Yes

**ContextType** (p. 65)

The context type.

Type: String

Length Constraints: Maximum length of 256.

Required: Yes

**Description** (p. 65)

The description of the context.
Type: String
Length Constraints: Maximum length of 3072.
Pattern: . *
Required: No

Properties (p. 65)
A list of properties to add to the context.
Type: String to string map
Map Entries: Maximum number of 30 items.
Key Length Constraints: Maximum length of 256.
Key Pattern: . *
Value Length Constraints: Maximum length of 256.
Value Pattern: . *
Required: No

Source (p. 65)
The source type, ID, and URI.
Type: ContextSource (p. 1369) object
Required: Yes

Tags (p. 65)
A list of tags to apply to the context.
Type: Array of Tag (p. 1979) objects
Array Members: Minimum number of 0 items. Maximum number of 50 items.
Required: No

Response Syntax

```json
{
   "ContextArn": "string"
}
```

Response Elements
If the action is successful, the service sends back an HTTP 200 response.
The following data is returned in JSON format by the service.

ContextArn (p. 66)
The Amazon Resource Name (ARN) of the context.
Type: String
Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:context/.*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceLimitExceeded

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
CreateDataQualityJobDefinition

Service: Amazon SageMaker Service

Creates a definition for a job that monitors data quality and drift. For information about model monitor, see Amazon SageMaker Model Monitor.

Request Syntax

```json
{
    "DataQualityAppSpecification": {
        "ContainerArguments": [ "string" ],
        "ContainerEntrypoint": [ "string" ],
        "Environment": {
            "string": "string"
        },
        "ImageUri": "string",
        "PostAnalyticsProcessorSourceUri": "string",
        "RecordPreprocessorSourceUri": "string"
    },
    "DataQualityBaselineConfig": {
        "BaseliningJobName": "string",
        "ConstraintsResource": {
            "S3Uri": "string"
        },
        "StatisticsResource": {
            "S3Uri": "string"
        }
    },
    "DataQualityJobInput": {
        "BatchTransformInput": {
            "DataCapturedDestinationS3Uri": "string",
            "DatasetFormat": {
                "Csv": {
                    "Header": boolean
                },
                "Json": {
                    "Line": boolean
                },
                "Parquet": {
                }
            },
            "EndTimeOffset": "string",
            "ExcludeFeaturesAttribute": "string",
            "FeaturesAttribute": "string",
            "InferenceAttribute": "string",
            "LocalPath": "string",
            "ProbabilityAttribute": "string",
            "ProbabilityThresholdAttribute": number,
            "S3DataDistributionType": "string",
            "S3InputMode": "string",
            "StartTimeOffset": "string"
        },
        "EndpointInput": {
            "EndpointName": "string",
            "EndTimeOffset": "string",
            "ExcludeFeaturesAttribute": "string",
            "FeaturesAttribute": "string",
            "InferenceAttribute": "string",
            "LocalPath": "string",
            "ProbabilityAttribute": "string",
            "ProbabilityThresholdAttribute": number,
            "S3DataDistributionType": "string",
            "S3InputMode": "string"
        }
    }
}
```
"StartTimeOffset": "string"

"DataQualityJobOutputConfig": {
"KmsKeyId": "string",
"MonitoringOutputs": [
{
"S3Output": {
"LocalPath": "string",
"S3UploadMode": "string",
"S3Uri": "string"
}
}
]

"JobDefinitionName": "string",
"JobResources": {
"ClusterConfig": {
"InstanceCount": number,
"InstanceType": "string",
"VolumeKmsKeyId": "string",
"VolumeSizeInGB": number
}

"NetworkConfig": {
"EnableInterContainerTrafficEncryption": boolean,
"EnableNetworkIsolation": boolean,
"VpcConfig": {
"SecurityGroupIds": [ "string" ],
"Subnets": [ "string" ]
}

"RoleArn": "string",
"StoppingCondition": {
"MaxRuntimeInSeconds": number
}

"Tags": [
{
"Key": "string",
"Value": "string"
}
]

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**DataQualityAppSpecification (p. 68)**

Specifies the container that runs the monitoring job.

Type: DataQualityAppSpecification (p. 1388) object

Required: Yes

**DataQualityBaselineConfig (p. 68)**

Configures the constraints and baselines for the monitoring job.

Type: DataQualityBaselineConfig (p. 1390) object
CreateDataQualityJobDefinition

Required: No

**DataQualityJobInput (p. 68)**

A list of inputs for the monitoring job. Currently endpoints are supported as monitoring inputs.

Type: [DataQualityJobInput (p. 1391)] object

Required: Yes

**DataQualityJobOutputConfig (p. 68)**

The output configuration for monitoring jobs.

Type: [MonitoringOutputConfig (p. 1754)] object

Required: Yes

**JobDefinitionName (p. 68)**

The name for the monitoring job definition.

Type: String


Pattern: `^[a-zA-Z0-9-]*[a-zA-Z0-9]{0,62}$`

Required: Yes

**JobResources (p. 68)**

Identifies the resources to deploy for a monitoring job.

Type: [MonitoringResources (p. 1756)] object

Required: Yes

**NetworkConfig (p. 68)**

Specifies networking configuration for the monitoring job.

Type: [MonitoringNetworkConfig (p. 1752)] object

Required: No

**RoleArn (p. 68)**

The Amazon Resource Name (ARN) of an IAM role that Amazon SageMaker can assume to perform tasks on your behalf.

Type: String


Pattern: `^[arn:aws[a-zA-Z\-]*:iam::\d{12}:role/\?a-zA-Z_0-9+=,.@\-_]+$`

Required: Yes

**StoppingCondition (p. 68)**

A time limit for how long the monitoring job is allowed to run before stopping.

Type: [MonitoringStoppingCondition (p. 1765)] object

Required: No
Tags (p. 68)

(Optional) An array of key-value pairs. For more information, see Using Cost Allocation Tags in the AWS Billing and Cost Management User Guide.

Type: Array of Tag (p. 1979) objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

Response Syntax

```json
{
  "JobDefinitionArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

JobDefinitionArn (p. 71)

The Amazon Resource Name (ARN) of the job definition.

Type: String

Length Constraints: Maximum length of 256.

Pattern: .*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceInUse

Resource being accessed is in use.

HTTP Status Code: 400

ResourceLimitExceeded

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
• AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
CreateDeviceFleet

Service: Amazon SageMaker Service

Creates a device fleet.

Request Syntax

```
{
    "Description": "string",
    "DeviceFleetName": "string",
    "EnableIotRoleAlias": boolean,
    "OutputConfig": {
        "KmsKeyId": "string",
        "PresetDeploymentConfig": "string",
        "PresetDeploymentType": "string",
        "S3OutputLocation": "string"
    },
    "RoleArn": "string",
    "Tags": [
        {
            "Key": "string",
            "Value": "string"
        }
    ]
}
```

Request Parameters

For information about the parameters that are common to all actions, see [Common Parameters](p. 2178).

The request accepts the following data in JSON format.

**Description (p. 73)**

A description of the fleet.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 800.

Pattern: `^[a-zA-Z0-9,;:]*$`

Required: No

**DeviceFleetName (p. 73)**

The name of the fleet that the device belongs to.

Type: String


Pattern: `^[a-zA-Z0-9,;:]*[a-zA-Z0-9]$`

Required: Yes

**EnableIotRoleAlias (p. 73)**

Whether to create an AWS IoT Role Alias during device fleet creation. The name of the role alias generated will match this pattern: "SageMakerEdge-{DeviceFleetName}".
For example, if your device fleet is called "demo-fleet", the name of the role alias will be "SageMakerEdge-demo-fleet".

Type: Boolean

Required: No

OutputConfig (p. 73)

The output configuration for storing sample data collected by the fleet.

Type: EdgeOutputConfig (p. 1444) object

Required: Yes

RoleArn (p. 73)

The Amazon Resource Name (ARN) that has access to AWS Internet of Things (IoT).

Type: String


Pattern: ^arn:aws[a-z\-]*:iam::\d{12}:role/?[a-zA-Z_0-9+=,.@\-_]+$ 

Required: No

Tags (p. 73)

Creates tags for the specified fleet.

Type: Array of Tag (p. 1979) objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceInUse

Resource being accessed is in use.

HTTP Status Code: 400

ResourceLimitExceeded

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:
- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
CreateDomain
Service: Amazon SageMaker Service

Creates a Domain. A domain consists of an associated Amazon Elastic File System (EFS) volume, a list of authorized users, and a variety of security, application, policy, and Amazon Virtual Private Cloud (VPC) configurations. Users within a domain can share notebook files and other artifacts with each other.

EFS storage
When a domain is created, an EFS volume is created for use by all of the users within the domain. Each user receives a private home directory within the EFS volume for notebooks, Git repositories, and data files.

SageMaker uses the AWS Key Management Service (AWS KMS) to encrypt the EFS volume attached to the domain with an AWS managed key by default. For more control, you can specify a customer managed key. For more information, see Protect Data at Rest Using Encryption.

VPC configuration
All traffic between the domain and the EFS volume is through the specified VPC and subnets. For other traffic, you can specify the AppNetworkAccessType parameter. AppNetworkAccessType corresponds to the network access type that you choose when you onboard to the domain. The following options are available:

- PublicInternetOnly - Non-EFS traffic goes through a VPC managed by Amazon SageMaker, which allows internet access. This is the default value.
- VpcOnly - All traffic is through the specified VPC and subnets. Internet access is disabled by default. To allow internet access, you must specify a NAT gateway.

When internet access is disabled, you won't be able to run a Amazon SageMaker Studio notebook or to train or host models unless your VPC has an interface endpoint to the SageMaker API and runtime or a NAT gateway and your security groups allow outbound connections.

Important
NFS traffic over TCP on port 2049 needs to be allowed in both inbound and outbound rules in order to launch a Amazon SageMaker Studio app successfully.

For more information, see Connect Amazon SageMaker Studio Notebooks to Resources in a VPC.

Request Syntax

```json
{
  "AppNetworkAccessType": "string",
  "AppSecurityGroupManagement": "string",
  "AuthMode": "string",
  "DefaultSpaceSettings": {
    "ExecutionRole": "string",
    "JupyterServerAppSettings": {
      "CodeRepositories": [{
        "RepositoryUrl": "string"
      }]
    },
    "DefaultResourceSpec": {
      "InstanceType": "string",
      "LifecycleConfigArn": "string",
      "SageMakerImageArn": "string",
      "SageMakerImageVersionAlias": "string",
      "SageMakerImageVersionArn": "string"
    }
  }
}
```
"LifecycleConfigArns": [ "string" ],
"
"KernelGatewayAppSettings": {
"
"CustomImages": [ {
"
"AppNameImageConfigName": "string",
"
"ImageName": "string",
"
"ImageVersionNumber": number
}
],
"
"DefaultResourceSpec": {
"
"InstanceType": "string",
"
"LifecycleConfigArn": "string",
"
"SageMakerImageArn": "string",
"
"SageMakerImageVersionAlias": "string",
"
"SageMakerImageVersionArn": "string"
},
"
"LifecycleConfigArns": [ "string" ]
],
"
"SecurityGroups": [ "string" ]
],
"
"DefaultUserSettings": {
"
"CanvasAppSettings": {
"
"DirectDeploySettings": {
"
"Status": "string"
},
"
"IdentityProviderOAuthSettings": [ {
"
"DataSourceName": "string",
"
"SecretArn": "string",
"
"Status": "string"
} ],
"
"KendraSettings": { "Status": "string" },
"
"ModelRegisterSettings": { "CrossAccountModelRegisterRoleArn": "string", "Status": "string" },
"
"TimeSeriesForecastingSettings": { "AmazonForecastRoleArn": "string", "Status": "string" },
"
"WorkspaceSettings": { "S3ArtifactPath": "string", "S3KmsKeyId": "string" }
],
"
"CodeEditorAppSettings": {
"
"DefaultResourceSpec": { "InstanceType": "string", "LifecycleConfigArn": "string", "SageMakerImageArn": "string", "SageMakerImageVersionAlias": "string", "SageMakerImageVersionArn": "string" },
"
"LifecycleConfigArns": [ "string" ]
],
"
"CustomFileSystemConfigs": [ { } ],
"
"CustomPosixUserConfig": { "Gid": number, "Uid": number
}
"DefaultLandingUri": "string",
"ExecutionRole": "string",
"JupyterLabAppSettings": {
  "CodeRepositories": [
    { "RepositoryUrl": "string"
  }
],
  "CustomImages": [
    { "AppImageConfigName": "string",
      "ImageName": "string",
      "ImageVersionNumber": number
    }
  ],
  "DefaultResourceSpec": {
    "InstanceType": "string",
    "LifecycleConfigArn": "string",
    "SageMakerImageArn": "string",
    "SageMakerImageVersionAlias": "string",
    "SageMakerImageVersionArn": "string"
  },
  "LifecycleConfigArns": [ "string" ]
},
"JupyterServerAppSettings": {
  "CodeRepositories": [
    { "RepositoryUrl": "string"
  }
],
  "DefaultResourceSpec": {
    "InstanceType": "string",
    "LifecycleConfigArn": "string",
    "SageMakerImageArn": "string",
    "SageMakerImageVersionAlias": "string",
    "SageMakerImageVersionArn": "string"
  },
  "LifecycleConfigArns": [ "string" ]
},
"KernelGatewayAppSettings": {
  "CustomImages": [
    { "AppImageConfigName": "string",
      "ImageName": "string",
      "ImageVersionNumber": number
    }
  ],
  "DefaultResourceSpec": {
    "InstanceType": "string",
    "LifecycleConfigArn": "string",
    "SageMakerImageArn": "string",
    "SageMakerImageVersionAlias": "string",
    "SageMakerImageVersionArn": "string"
  },
  "LifecycleConfigArns": [ "string" ]
},
"RSessionAppSettings": {
  "CustomImages": [
    { "AppImageConfigName": "string",
      "ImageName": "string",
      "ImageVersionNumber": number
    }
  ],
  "DefaultResourceSpec": {
    "InstanceType": "string",

"LifecycleConfigArn": "string",
"SageMakerImageArn": "string",
"SageMakerImageVersionAlias": "string",
"SageMakerImageVersionArn": "string"
},
"RStudioServerProAppSettings": {
  "AccessStatus": "string",
  "UserGroup": "string"
},
"SecurityGroups": [ "string" ],
"SharingSettings": {
  "NotebookOutputOption": "string",
  "S3KmsKeyId": "string",
  "S3OutputPath": "string"
},
"SpaceStorageSettings": {
  "DefaultEbsStorageSettings": {
    "DefaultEbsVolumeSizeInGb": number,
    "MaximumEbsVolumeSizeInGb": number
  }
},
"StudioWebPortal": "string",
"TensorBoardAppSettings": {
  "DefaultResourceSpec": {
    "InstanceType": "string",
    "LifecycleConfigArn": "string",
    "SageMakerImageArn": "string",
    "SageMakerImageVersionAlias": "string",
    "SageMakerImageVersionArn": "string"
  }
},
"DomainName": "string",
"DomainSettings": {
  "ExecutionRoleIdentityConfig": "string",
  "RStudioServerProDomainSettings": {
    "DefaultResourceSpec": {
      "InstanceType": "string",
      "LifecycleConfigArn": "string",
      "SageMakerImageArn": "string",
      "SageMakerImageVersionAlias": "string",
      "SageMakerImageVersionArn": "string"
    },
    "DomainExecutionRoleArn": "string",
    "RStudioConnectUrl": "string",
    "RStudioPackageManagerUrl": "string"
  },
  "SecurityGroupIds": [ "string" ]
},
"HomeEfsFileSystemKmsKeyId": "string",
"KmsKeyId": "string",
"SubnetIds": [ "string" ],
"Tags": [
  {
    "Key": "string",
    "Value": "string"
  }
],
"VpcId": "string"}
Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**AppNetworkAccessType (p. 76)**

Specifies the VPC used for non-EFS traffic. The default value is PublicInternetOnly.
- PublicInternetOnly - Non-EFS traffic is through a VPC managed by Amazon SageMaker, which allows direct internet access.
- VpcOnly - All traffic is through the specified VPC and subnets.

Type: String

Valid Values: PublicInternetOnly | VpcOnly

Required: No

**AppSecurityGroupManagement (p. 76)**

The entity that creates and manages the required security groups for inter-app communication in VpcOnly mode. Required when CreateDomain.AppNetworkAccessType is VpcOnly and DomainSettings.RStudioServerProDomainSettings.DomainExecutionRoleArn is provided. If setting up the domain for use with RStudio, this value must be set to Service.

Type: String

Valid Values: Service | Customer

Required: No

**AuthMode (p. 76)**

The mode of authentication that members use to access the domain.

Type: String

Valid Values: SSO | IAM

Required: Yes

**DefaultSpaceSettings (p. 76)**

The default settings used to create a space.

Type: [DefaultSpaceSettings (p. 1402)] object

Required: No

**DefaultUserSettings (p. 76)**

The default settings to use to create a user profile when UserSettings isn't specified in the call to the CreateUserProfile API.

SecurityGroups is aggregated when specified in both calls. For all other settings in UserSettings, the values specified in CreateUserProfile take precedence over those specified in CreateDomain.

Type: [UserSettings (p. 2070)] object

Required: Yes
DomainName (p. 76)

A name for the domain.

Type: String
Length Constraints: Maximum length of 63.
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$
Required: Yes

DomainSettings (p. 76)

A collection of Domain settings.

Type: DomainSettings (p. 1424) object
Required: No

HomeEfsFileSystemKmsKeyId (p. 76)

This parameter has been deprecated.
Use KmsKeyId.

Type: String
Length Constraints: Maximum length of 2048.
Pattern: .*
Required: No

KmsKeyId (p. 76)

SageMaker uses AWS KMS to encrypt the EFS volume attached to the domain with an AWS managed key by default. For more control, specify a customer managed key.

Type: String
Length Constraints: Maximum length of 2048.
Pattern: .*
Required: No

SubnetIds (p. 76)

The VPC subnets that the domain uses for communication.

Type: Array of strings
Array Members: Minimum number of 1 item. Maximum number of 16 items.
Length Constraints: Maximum length of 32.
Pattern: [-0-9a-zA-Z]+
Required: Yes

Tags (p. 76)

Tags to associated with the Domain. Each tag consists of a key and an optional value. Tag keys must be unique per resource. Tags are searchable using the Search API.
Tags that you specify for the Domain are also added to all Apps that the Domain launches.

Type: Array of Tag objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

VpcId (p. 76)

The ID of the Amazon Virtual Private Cloud (VPC) that the domain uses for communication.

Type: String

Length Constraints: Maximum length of 32.

Pattern: [-0-9a-zA-Z]+

Required: Yes

Response Syntax

```
{
    "DomainArn": "string",
    "Url": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

DomainArn (p. 82)

The Amazon Resource Name (ARN) of the created domain.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:domain/.*

Url (p. 82)

The URL to the created domain.

Type: String

Length Constraints: Maximum length of 1024.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceInUse

Resource being accessed is in use.
HTTP Status Code: 400

**ResourceLimitExceeded**

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
CreateEdgeDeploymentPlan

Service: Amazon SageMaker Service

Creates an edge deployment plan, consisting of multiple stages. Each stage may have a different deployment configuration and devices.

Request Syntax

```
{
  "DeviceFleetName": "string",
  "EdgeDeploymentPlanName": "string",
  "ModelConfigs": [
    {
      "EdgePackagingJobName": "string",
      "ModelHandle": "string"
    }
  ],
  "Stages": [
    {
      "DeploymentConfig": {
        "FailureHandlingPolicy": "string"
      },
      "DeviceSelectionConfig": {
        "DeviceNameContains": "string",
        "DeviceNames": [ "string" ],
        "DeviceSubsetType": "string",
        "Percentage": "number"
      },
      "StageName": "string"
    }
  ],
  "Tags": [
    {
      "Key": "string",
      "Value": "string"
    }
  ]
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**DeviceFleetName (p. 84)**

The device fleet used for this edge deployment plan.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}$

Required: Yes

**EdgeDeploymentPlanName (p. 84)**

The name of the edge deployment plan.
CreateEdgeDeploymentPlan

Type: String
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}$
Required: Yes

ModelConfigs (p. 84)
List of models associated with the edge deployment plan.
Type: Array of EdgeDeploymentModelConfig (p. 1435) objects
Required: Yes

Stages (p. 84)
List of stages of the edge deployment plan. The number of stages is limited to 10 per deployment.
Type: Array of DeploymentStage (p. 1407) objects
Required: No

Tags (p. 84)
List of tags with which to tag the edge deployment plan.
Type: Array of Tag (p. 1979) objects
Array Members: Minimum number of 0 items. Maximum number of 50 items.
Required: No

Response Syntax

```json
{
    "EdgeDeploymentPlanArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.
The following data is returned in JSON format by the service.

EdgeDeploymentPlanArn (p. 85)
The ARN of the edge deployment plan.
Type: String
Pattern: ^arn:aws[a-z\-]*:sagemaker:[a-z\-]*:d\{12\}:edge-deployment/?[a-zA-Z_0-9=+.\-_\/]\+$

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).
ResourceLimitExceeded

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
CreateEdgeDeploymentStage

Service: Amazon SageMaker Service

Creates a new stage in an existing edge deployment plan.

Request Syntax

```
{
  "EdgeDeploymentPlanName": "string",
  "Stages": [
    {
      "DeploymentConfig": {
        "FailureHandlingPolicy": "string"
      },
      "DeviceSelectionConfig": {
        "DeviceNameContains": "string",
        "DeviceNames": [ "string" ],
        "DeviceSubsetType": "string",
        "Percentage": number
      }
    },
    "StageName": "string"
  ]
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**EdgeDeploymentPlanName (p. 87)**

The name of the edge deployment plan.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}$

Required: Yes

**Stages (p. 87)**

List of stages to be added to the edge deployment plan.

Type: Array of DeploymentStage (p. 1407) objects

Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).
ResourceLimitExceeded

You have exceeded a SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
CreateEdgePackagingJob
Service: Amazon SageMaker Service

Starts a SageMaker Edge Manager model packaging job. Edge Manager will use the model artifacts from the Amazon Simple Storage Service bucket that you specify. After the model has been packaged, Amazon SageMaker saves the resulting artifacts to an S3 bucket that you specify.

Request Syntax

```json
{
    "CompilationJobName": "string",
    "EdgePackagingJobName": "string",
    "ModelName": "string",
    "ModelVersion": "string",
    "OutputConfig": {
        "KmsKeyId": "string",
        "PresetDeploymentConfig": "string",
        "PresetDeploymentType": "string",
        "S3OutputLocation": "string"
    },
    "ResourceKey": "string",
    "RoleArn": "string",
    "Tags": [
        {
            "Key": "string",
            "Value": "string"
        }
    ]
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

CompilationJobName (p. 89)

The name of the SageMaker Neo compilation job that will be used to locate model artifacts for packaging.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

Required: Yes

EdgePackagingJobName (p. 89)

The name of the edge packaging job.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

Required: Yes
**ModelName (p. 89)**

The name of the model.

Type: String


Pattern: \^[a-zA-Z0-9\-\*\(\)\[\]{{0,62}}$

Required: Yes

**ModelVersion (p. 89)**

The version of the model.

Type: String


Pattern: [a-zA-Z0-9\-\*\(\)\[\]{{1,30}}$

Required: Yes

**OutputConfig (p. 89)**

Provides information about the output location for the packaged model.

Type: EdgeOutputConfig (p. 1444) object

Required: Yes

**ResourceKey (p. 89)**

The AWS KMS key to use when encrypting the EBS volume the edge packaging job runs on.

Type: String

Length Constraints: Maximum length of 2048.

Pattern: .*

Required: No

**RoleArn (p. 89)**

The Amazon Resource Name (ARN) of an IAM role that enables Amazon SageMaker to download and upload the model, and to contact SageMaker Neo.

Type: String


Pattern: ^arn\:aws\[a-zA-Z\-\]*\:iam\:\d\(12\):role/\?\[a-zA-Z0-9\-\*\+\@\-\_]\+\$

Required: Yes

**Tags (p. 89)**

Creates tags for the packaging job.

Type: Array of Tag (p. 1979) objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No
Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceLimitExceeded

You have exceeded a SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
CreateEndpoint
Service: Amazon SageMaker Service

Creates an endpoint using the endpoint configuration specified in the request. SageMaker uses the endpoint to provision resources and deploy models. You create the endpoint configuration with the CreateEndpointConfig API.

Use this API to deploy models using SageMaker hosting services.

**Note**
You must not delete an EndpointConfig that is in use by an endpoint that is live or while the UpdateEndpoint or CreateEndpoint operations are being performed on the endpoint. To update an endpoint, you must create a new EndpointConfig.

The endpoint name must be unique within an AWS Region in your AWS account.

When it receives the request, SageMaker creates the endpoint, launches the resources (ML compute instances), and deploys the model(s) on them.

**Note**
When you call CreateEndpoint, a load call is made to DynamoDB to verify that your endpoint configuration exists. When you read data from a DynamoDB table supporting Eventually Consistent Reads, the response might not reflect the results of a recently completed write operation. The response might include some stale data. If the dependent entities are not yet in DynamoDB, this causes a validation error. If you repeat your read request after a short time, the response should return the latest data. So retry logic is recommended to handle these possible issues. We also recommend that customers call DescribeEndpointConfig before calling CreateEndpoint to minimize the potential impact of a DynamoDB eventually consistent read.

When SageMaker receives the request, it sets the endpoint status to Creating. After it creates the endpoint, it sets the status to InService. SageMaker can then process incoming requests for inferences. To check the status of an endpoint, use the DescribeEndpoint API.

If any of the models hosted at this endpoint get model data from an Amazon S3 location, SageMaker uses AWS Security Token Service to download model artifacts from the S3 path you provided. AWS STS is activated in your AWS account by default. If you previously deactivated AWS STS for a region, you need to reactivate AWS STS for that region. For more information, see Activating and Deactivating AWS STS in an AWS Region in the AWS Identity and Access Management User Guide.

**Note**
To add the IAM role policies for using this API operation, go to the IAM console, and choose Roles in the left navigation pane. Search the IAM role that you want to grant access to use the CreateEndpoint and CreateEndpointConfig API operations, add the following policies to the role.

- Option 1: For a full SageMaker access, search and attach the AmazonSageMakerFullAccess policy.
- Option 2: For granting a limited access to an IAM role, paste the following Action elements manually into the JSON file of the IAM role:

  ```json
  "Action": ["sagemaker:CreateEndpoint",
  "sagemaker:CreateEndpointConfig"]
  "Resource": [
  "arn:aws:sagemaker:region:account-id:endpoint/endpointName"
  "arn:aws:sagemaker:region:account-id:endpoint-config/endpointConfigName"
  ```
Request Syntax

```json
{
  "DeploymentConfig": {
    "AutoRollbackConfiguration": {
      "Alarms": [
        {
          "AlarmName": "string"
        }
      ],
      "BlueGreenUpdatePolicy": {
        "MaximumExecutionTimeoutInSeconds": number,
        "TerminationWaitInSeconds": number,
        "TrafficRoutingConfiguration": {
          "canarySize": {
            "Type": "string",
            "Value": number
          },
          "LinearStepSize": {
            "Type": "string",
            "Value": number
          },
          "Type": "string",
          "WaitIntervalInSeconds": number
        }
      },
      "RollingUpdatePolicy": {
        "MaximumBatchSize": {
          "Type": "string",
          "Value": number
        },
        "MaximumExecutionTimeoutInSeconds": number,
        "RollbackMaximumBatchSize": {
          "Type": "string",
          "Value": number
        },
        "WaitIntervalInSeconds": number
      }
    },
    "EndpointConfigName": "string",
    "EndpointName": "string",
    "Tags": [
      {
        "Key": "string",
        "Value": "string"
      }
    ]
  }
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.
DeploymentConfig (p. 93)

The deployment configuration for an endpoint, which contains the desired deployment strategy and rollback configurations.

Type: DeploymentConfig (p. 1405) object

Required: No

EndpointConfigName (p. 93)

The name of an endpoint configuration. For more information, see CreateEndpointConfig.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

Required: Yes

EndpointName (p. 93)

The name of the endpoint. The name must be unique within an AWS Region in your AWS account. The name is case-insensitive in CreateEndpoint, but the case is preserved and must be matched in InvokeEndpoint.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

Required: Yes

Tags (p. 93)

An array of key-value pairs. You can use tags to categorize your AWS resources in different ways, for example, by purpose, owner, or environment. For more information, see Tagging AWS Resources.

Type: Array of Tag (p. 1979) objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

Response Syntax

```json
{
  "EndpointArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

EndpointArn (p. 94)

The Amazon Resource Name (ARN) of the endpoint.
CreateEndpoint

Type: String


Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:endpoint/.*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceLimitExceeded

You have exceeded a SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
CreateEndpointConfig
Service: Amazon SageMaker Service

Creates an endpoint configuration that SageMaker hosting services uses to deploy models. In the configuration, you identify one or more models, created using the CreateModel API, to deploy and the resources that you want SageMaker to provision. Then you call the CreateEndpoint API.

**Note**
Use this API if you want to use SageMaker hosting services to deploy models into production.

In the request, you define a ProductionVariant, for each model that you want to deploy. Each ProductionVariant parameter also describes the resources that you want SageMaker to provision. This includes the number and type of ML compute instances to deploy.

If you are hosting multiple models, you also assign a VariantWeight to specify how much traffic you want to allocate to each model. For example, suppose that you want to host two models, A and B, and you assign traffic weight 2 for model A and 1 for model B. SageMaker distributes two-thirds of the traffic to Model A, and one-third to model B.

**Note**
When you call CreateEndpoint, a load call is made to DynamoDB to verify that your endpoint configuration exists. When you read data from a DynamoDB table supporting Eventually Consistent Reads, the response might not reflect the results of a recently completed write operation. The response might include some stale data. If the dependent entities are not yet in DynamoDB, this causes a validation error. If you repeat your read request after a short time, the response should return the latest data. So retry logic is recommended to handle these possible issues. We also recommend that customers call DescribeEndpointConfig before calling CreateEndpoint to minimize the potential impact of a DynamoDB eventually consistent read.

**Request Syntax**

```json
{
   "AsyncInferenceConfig": {
      "ClientConfig": {
         "MaxConcurrentInvocationsPerInstance": number
      },
      "OutputConfig": {
         "KmsKeyId": "string",
         "NotificationConfig": {
            "ErrorTopic": "string",
            "IncludeInferenceResponseIn": [ "string" ],
            "SuccessTopic": "string"
         },
         "S3FailurePath": "string",
         "S3OutputPath": "string"
      }
   },
   "DataCaptureConfig": {
      "CaptureContentTypeHeader": {
         "CsvContentTypes": [ "string" ],
         "JsonContentTypes": [ "string" ]
      },
      "CaptureOptions": [ {
         "CaptureMode": "string"
      } ],
      "DestinationS3Uri": "string",
      "EnableCapture": boolean,
      "InitialSamplingPercentage": number,
      "KmsKeyId": "string"
   }
}
```
"EnableNetworkIsolation": boolean,
"EndpointConfigName": "string",
"ExecutionRoleArn": "string",
"ExplainerConfig": {
  "ClarifyExplainerConfig": {
    "EnableExplanations": "string",
    "InferenceConfig": {
      "ContentTemplate": "string",
      "FeatureHeaders": ["string"],
      "FeaturesAttribute": "string",
      "FeatureTypes": ["string"],
      "LabelAttribute": "string",
      "LabelHeaders": ["string"],
      "LabelIndex": number,
      "MaxPayloadInMB": number,
      "MaxRecordCount": number,
      "ProbabilityAttribute": "string",
      "ProbabilityIndex": number
    },
    "ShapConfig": {
      "NumberOfSamples": number,
      "Seed": number,
      "ShapBaselineConfig": {
        "MimeType": "string",
        "ShapBaseline": "string",
        "ShapBaselineUri": "string"
      },
      "TextConfig": {
        "Granularity": "string",
        "Language": "string"
      },
      "UseLogit": boolean
    }
  }
},
"KmsKeyId": "string",
"ProductionVariants": [ {
  "AcceleratorType": "string",
  "ContainerStartupHealthCheckTimeoutInSeconds": number,
  "CoreDumpConfig": {
    "DestinationS3Uri": "string",
    "KmsKeyId": "string"
  },
  "EnableSSMAccess": boolean,
  "InitialInstanceCount": number,
  "InitialVariantWeight": number,
  "InstanceType": "string",
  "ManagedInstanceScaling": {
    "MaxInstanceCount": number,
    "MinInstanceCount": number,
    "Status": "string"
  },
  "ModelDataDownloadTimeoutInSeconds": number,
  "ModelName": "string",
  "RoutingConfig": {
    "RoutingStrategy": "string"
  },
  "ServerlessConfig": {
    "MaxConcurrency": number,
    "MemorySizeInMB": number,
    "ProvisionedConcurrency": number
  },
  "VariantName": "string",
  "VolumeSizeInGB": number
}]
}
Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

AsyncInferenceConfig (p. 96)

Specifies configuration for how an endpoint performs asynchronous inference. This is a required field in order for your Endpoint to be invoked using InvokeEndpointAsync.

Type: AsyncInferenceConfig (p. 1258) object

Required: No

DataCaptureConfig (p. 96)

Configuration to control how SageMaker captures inference data.
**DataCaptureConfig (p. 1381)** object

Required: No

**EnableNetworkIsolation (p. 96)**

Sets whether all model containers deployed to the endpoint are isolated. If they are, no inbound or outbound network calls can be made to or from the model containers.

Type: Boolean

Required: No

**EndpointConfigName (p. 96)**

The name of the endpoint configuration. You specify this name in a [CreateEndpoint](#) request.

Type: String

Length Constraints: Maximum length of 63.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62\}`

Required: Yes

**ExecutionRoleArn (p. 96)**

The Amazon Resource Name (ARN) of an IAM role that Amazon SageMaker can assume to perform actions on your behalf. For more information, see [SageMaker Roles](#).

**Note**

To be able to pass this role to Amazon SageMaker, the caller of this action must have the `iam:PassRole` permission.

Type: String


Pattern: `^arn:aws[a-zA-Z\-]*:iam::\d{12}:role/?[a-zA-Z-A-Z0-9\+=,.@-_\/>]+$`

Required: No

**ExplainerConfig (p. 96)**

A member of CreateEndpointConfig that enables explainers.

Type: [ExplainerConfig (p. 1479)](#) object

Required: No

**KmsKeyId (p. 96)**

The Amazon Resource Name (ARN) of a AWS Key Management Service key that SageMaker uses to encrypt data on the storage volume attached to the ML compute instance that hosts the endpoint.

The KmsKeyId can be any of the following formats:

- **Key ID:** `1234abcd-12ab-34cd-56ef-1234567890ab`
- **Key ARN:** `arn:aws:kms:us-west-2:111122223333:key/1234abcd-12ab-34cd-56ef-1234567890ab`
- **Alias name:** `alias/ExampleAlias`
- **Alias name ARN:** `arn:aws:kms:us-west-2:111122223333:alias/ExampleAlias`

The KMS key policy must grant permission to the IAM role that you specify in your [CreateEndpoint](#), [UpdateEndpoint](#) requests. For more information, refer to the AWS Key Management Service section [Using Key Policies in AWS KMS](#)

---

99
Note
Certain Nitro-based instances include local storage, dependent on the instance type. Local storage volumes are encrypted using a hardware module on the instance. You can't request a KmsKeyId when using an instance type with local storage. If any of the models that you specify in the ProductionVariants parameter use nitro-based instances with local storage, do not specify a value for the KmsKeyId parameter. If you specify a value for KmsKeyId when using any nitro-based instances with local storage, the call to CreateEndpointConfig fails.
For a list of instance types that support local instance storage, see Instance Store Volumes.
For more information about local instance storage encryption, see SSD Instance Store Volumes.

Type: String
Length Constraints: Maximum length of 2048.
Pattern: .*
Required: No

ProductionVariants (p. 96)
An array of ProductionVariant objects, one for each model that you want to host at this endpoint.
Type: Array of ProductionVariant (p. 1847) objects
Array Members: Minimum number of 1 item. Maximum number of 10 items.
Required: Yes

ShadowProductionVariants (p. 96)
An array of ProductionVariant objects, one for each model that you want to host at this endpoint in shadow mode with production traffic replicated from the model specified on ProductionVariants. If you use this field, you can only specify one variant for ProductionVariants and one variant for ShadowProductionVariants.
Type: Array of ProductionVariant (p. 1847) objects
Array Members: Minimum number of 1 item. Maximum number of 10 items.
Required: No

Tags (p. 96)
An array of key-value pairs. You can use tags to categorize your AWS resources in different ways, for example, by purpose, owner, or environment. For more information, see Tagging AWS Resources.
Type: Array of Tag (p. 1979) objects
Array Members: Minimum number of 0 items. Maximum number of 50 items.
Required: No

VpcConfig (p. 96)
Specifies an Amazon Virtual Private Cloud (VPC) that your SageMaker jobs, hosted models, and compute resources have access to. You can control access to and from your resources by configuring a VPC. For more information, see Give SageMaker Access to Resources in your Amazon VPC.
Type: VpcConfig (p. 2076) object
Required: No
Response Syntax

```json
{
   "EndpointConfigArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**EndpointConfigArn (p. 101)**

The Amazon Resource Name (ARN) of the endpoint configuration.

Type: String


Pattern: `arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:endpoint-config/.*`

Errors

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](https://docs.aws.amazon.com/AmazonSageMaker/latest/APIReference/MsgResourceLimitExceeded.html).

**ResourceLimitExceeded**

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](https://docs.aws.amazon.com/cli/index.html)
- [AWS SDK for .NET](https://docs.aws.amazon.com/sdk-for-net/v3/developer-guide/welcome.html)
- [AWS SDK for C++](https://docs.aws.amazon.com/sdk-for-cpp/v1/developer-guide/welcome.html)
- [AWS SDK for Go](https://docs.aws.amazon.com/sdk-for-golang/v2/developer-guide/welcome.html)
- [AWS SDK for Java V2](https://docs.aws.amazon.com/java-sdk/index.html)
- [AWS SDK for JavaScript V3](https://docs.aws.amazon.com/js-sdk/index.html)
- [AWS SDK for PHP V3](https://docs.aws.amazon.com/php-sdk/index.html)
- [AWS SDK for Python](https://docs.aws.amazon.com/sdk-for-python/
- [AWS SDK for Ruby V3](https://docs.aws.amazon.com/sdk-for-ruby/index.html)
CreateExperiment
Service: Amazon SageMaker Service

Creates a SageMaker experiment. An experiment is a collection of trials that are observed, compared and evaluated as a group. A trial is a set of steps, called trial components, that produce a machine learning model.

Note
In the Studio UI, trials are referred to as run groups and trial components are referred to as runs.

The goal of an experiment is to determine the components that produce the best model. Multiple trials are performed, each one isolating and measuring the impact of a change to one or more inputs, while keeping the remaining inputs constant.

When you use SageMaker Studio or the SageMaker Python SDK, all experiments, trials, and trial components are automatically tracked, logged, and indexed. When you use the AWS SDK for Python (Boto), you must use the logging APIs provided by the SDK.

You can add tags to experiments, trials, trial components and then use the Search API to search for the tags.

To add a description to an experiment, specify the optional Description parameter. To add a description later, or to change the description, call the UpdateExperiment API.

To get a list of all your experiments, call the ListExperiments API. To view an experiment's properties, call the DescribeExperiment API. To get a list of all the trials associated with an experiment, call the ListTrials API. To create a trial call the CreateTrial API.

Request Syntax

```
{
   "Description": "string",
   "DisplayName": "string",
   "ExperimentName": "string",
   "Tags": [
   {
   "Key": "string",
   "Value": "string"
   }
   ]
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**Description (p. 102)**

The description of the experiment.

Type: String

Length Constraints: Maximum length of 3072.

Pattern: .*
Required: No

**DisplayName (p. 102)**

The name of the experiment as displayed. The name doesn't need to be unique. If you don't specify `DisplayName`, the value in `ExperimentName` is displayed.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,119}`

Required: No

**ExperimentName (p. 102)**

The name of the experiment. The name must be unique in your AWS account and is not case-sensitive.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,119}`

Required: Yes

**Tags (p. 102)**

A list of tags to associate with the experiment. You can use `Search API` to search on the tags.

Type: Array of `Tag (p. 1979)` objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

**Response Syntax**

```json
{
  "ExperimentArn": "string"
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**ExperimentArn (p. 103)**

The Amazon Resource Name (ARN) of the experiment.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-zA-Z\-]*:sagemaker:[a-zA-Z0-9\-]*:[0-9]{12}:experiment/.*
Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceLimitExceeded

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
CreateFeatureGroup

Service: Amazon SageMaker Service

Create a new FeatureGroup. A FeatureGroup is a group of Features defined in the FeatureStore to describe a Record.

The FeatureGroup defines the schema and features contained in the FeatureGroup. A FeatureGroup definition is composed of a list of Features, a RecordIdentifierFeatureName, an EventTimeFeatureName and configurations for its OnlineStore and OfflineStore. Check AWS service quotas to see the FeatureGroup's quota for your AWS account.

Note that it can take approximately 10-15 minutes to provision an OnlineStore FeatureGroup with the InMemory StorageType.

**Important**
You must include at least one of OnlineStoreConfig and OfflineStoreConfig to create a FeatureGroup.

Request Syntax

```json
{
  "Description": "string",
  "EventTimeFeatureName": "string",
  "FeatureDefinitions": [
    {
      "CollectionConfig": { ... },
      "CollectionType": "string",
      "FeatureName": "string",
      "FeatureType": "string"
    }
  ],
  "FeatureGroupName": "string",
  "OfflineStoreConfig": {
    "DataCatalogConfig": {
      "Catalog": "string",
      "Database": "string",
      "TableName": "string"
    },
    "DisableGlueTableCreation": boolean,
    "S3StorageConfig": {
      "KmsKeyId": "string",
      "ResolvedOutputS3Uri": "string",
      "S3Uri": "string"
    },
    "TableFormat": "string"
  },
  "OnlineStoreConfig": {
    "EnableOnlineStore": boolean,
    "SecurityConfig": {
      "KmsKeyId": "string"
    },
    "StorageType": "string",
    "TtlDuration": {
      "Unit": "string",
      "Value": number
    }
  },
  "RecordIdentifierFeatureName": "string",
  "RoleArn": "string",
  "Tags": [
    {
      "Key": "string",
    }
  ]
}
```
"Value": "string"}
]
}

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**Description (p. 105)**

A free-form description of a FeatureGroup.

Type: String

Length Constraints: Maximum length of 128.

Required: No

**EventTimeFeatureName (p. 105)**

The name of the feature that stores the EventTime of a Record in a FeatureGroup.

An EventTime is a point in time when a new event occurs that corresponds to the creation or update of a Record in a FeatureGroup. All Records in the FeatureGroup must have a corresponding EventTime.

An EventTime can be a String or Fractional.

- Fractional: EventTime feature values must be a Unix timestamp in seconds.
- String: EventTime feature values must be an ISO-8601 string in the format. The following formats are supported yyyy-MM-dd'T'HH:mm:ssZ and yyyy-MM-dd'T'HH:mm:ss.SSSZ where yyyy, MM, and dd represent the year, month, and day respectively and HH, mm, ss, and if applicable, SSS represent the hour, month, second and milliseconds respectively. 'T' and Z are constants.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 64.

Pattern: ^[a-zA-Z0-9]([-_]*)[a-zA-Z0-9]{0,63}

Required: Yes

**FeatureDefinitions (p. 105)**

A list of Feature names and types. Name and Type is compulsory per Feature.

Valid feature FeatureTypes are Integral, Fractional and String.

FeatureNames cannot be any of the following: is_deleted, write_time, api_invocation_time

You can create up to 2,500 FeatureDefinitions per FeatureGroup.

Type: Array of FeatureDefinition (p. 1481) objects

Array Members: Minimum number of 1 item. Maximum number of 2500 items.

Required: Yes
**FeatureGroupName (p. 105)**

The name of the FeatureGroup. The name must be unique within an AWS Region in an AWS account. The name:
- Must start and end with an alphanumeric character.
- Can only contain alphanumeric character and hyphens. Spaces are not allowed.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 64.

Pattern: `^[a-zA-Z0-9](\[-_]*[a-zA-Z0-9])\{0,63}\]

Required: Yes

**OfflineStoreConfig (p. 105)**

Use this to configure an OfflineFeatureStore. This parameter allows you to specify:
- The Amazon Simple Storage Service (Amazon S3) location of an OfflineStore.
- A configuration for an AWS Glue or AWS Hive data catalog.
- An KMS encryption key to encrypt the Amazon S3 location used for OfflineStore. If KMS encryption key is not specified, by default we encrypt all data at rest using AWS KMS key. By defining your bucket-level key for SSE, you can reduce AWS KMS requests costs by up to 99 percent.
- Format for the offline store table. Supported formats are Glue (Default) and Apache Iceberg.

To learn more about this parameter, see **OfflineStoreConfig**.

Type: `OfflineStoreConfig (p. 1777)` object

Required: No

**OnlineStoreConfig (p. 105)**

You can turn the OnlineStore on or off by specifying True for the EnableOnlineStore flag in OnlineStoreConfig.

You can also include an AWS KMS key ID (KMSKeyId) for at-rest encryption of the OnlineStore.

The default value is False.

Type: `OnlineStoreConfig (p. 1784)` object

Required: No

**RecordIdentifierFeatureName (p. 105)**

The name of the Feature whose value uniquely identifies a Record defined in the FeatureStore. Only the latest record per identifier value will be stored in the OnlineStore. RecordIdentifierFeatureName must be one of feature definitions' names.

You use the RecordIdentifierFeatureName to access data in a FeatureStore.

This name:
- Must start and end with an alphanumeric character.
- Can only contains alphanumeric characters, hyphens, underscores. Spaces are not allowed.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 64.
Pattern: ^[a-zA-Z0-9-][\-_]*[a-zA-Z0-9]\{0,63\}$
Required: Yes

RoleArn (p. 105)
The Amazon Resource Name (ARN) of the IAM execution role used to persist data into the OfflineStore if an OfflineStoreConfig is provided.
Type: String
Pattern: ^arn:aws[a-z\-]*:iam::\d{12}:role/?[a-zA-Z\-+,.@\-_/]+$ Required: No
Tags (p. 105)
Tags used to identify Features in each FeatureGroup.
Type: Array of Tag (p. 1979) objects
Array Members: Minimum number of 0 items. Maximum number of 50 items.
Required: No

Response Syntax

```
{
  "FeatureGroupArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.
The following data is returned in JSON format by the service.

FeatureGroupArn (p. 108)
The Amazon Resource Name (ARN) of the FeatureGroup. This is a unique identifier for the feature group.
Type: String
Length Constraints: Maximum length of 256.
Pattern: arn:aws[a-z\-]*:sagemaker:[a-zA-Z0-9\-]*:[0-9]{12}:feature-group/.*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceInUse
Resource being accessed is in use.
HTTP Status Code: 400
ResourceLimitExceeded

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
CreateFlowDefinition
Service: Amazon SageMaker Service

Creates a flow definition.

Request Syntax

```json
{
    "FlowDefinitionName": "string",
    "HumanLoopActivationConfig": {
        "HumanLoopActivationConditionsConfig": {
            "HumanLoopActivationConditions": "string"
        }
    },
    "HumanLoopConfig": {
        "HumanTaskUiArn": "string",
        "PublicWorkforceTaskPrice": {
            "AmountInUsd": {
                "Cents": number,
                "Dollars": number,
                "TenthFractionsOfACent": number
            }
        },
        "TaskAvailabilityLifetimeInSeconds": number,
        "TaskCount": number,
        "TaskDescription": "string",
        "TaskKeywords": [ "string" ],
        "TaskTimeLimitInSeconds": number,
        "TaskTitle": "string",
        "WorkteamArn": "string"
    },
    "HumanLoopRequestSource": {
        "AwsManagedHumanLoopRequestSource": "string"
    },
    "OutputConfig": {
        "KmsKeyId": "string",
        "S3OutputPath": "string"
    },
    "RoleArn": "string",
    "Tags": [
        {
            "Key": "string",
            "Value": "string"
        }
    ]
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**FlowDefinitionName (p. 110)**

The name of your flow definition.

Type: String

Pattern: ^[a-z0-9](-*[a-z0-9])\{0,62\}
Required: Yes

**HumanLoopActivationConfig (p. 110)**

An object containing information about the events that trigger a human workflow.

Type: `HumanLoopActivationConfig (p. 1516)` object

Required: No

**HumanLoopConfig (p. 110)**

An object containing information about the tasks the human reviewers will perform.

Type: `HumanLoopConfig (p. 1517)` object

Required: Yes

**HumanLoopRequestSource (p. 110)**

Container for configuring the source of human task requests. Use to specify if Amazon Rekognition or Amazon Textract is used as an integration source.

Type: `HumanLoopRequestSource (p. 1522)` object

Required: No

**OutputConfig (p. 110)**

An object containing information about where the human review results will be uploaded.

Type: `FlowDefinitionOutputConfig (p. 1502)` object

Required: Yes

**RoleArn (p. 110)**

The Amazon Resource Name (ARN) of the role needed to call other services on your behalf. For example, `arn:aws:iam::1234567890:role/service-role/AmazonSageMaker-ExecutionRole-20180111T151298`.

Type: String


Pattern: ^arn:aws[a-zA-Z0-9\-]*:iam::d{12}:\d{12}:role/\^[a-zA-Z0-9_\-@=+/]+,..@-/.]+$

Required: Yes

**Tags (p. 110)**

An array of key-value pairs that contain metadata to help you categorize and organize a flow definition. Each tag consists of a key and a value, both of which you define.

Type: Array of `Tag (p. 1979)` objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

**Response Syntax**

```json
{
}
```
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**FlowDefinitionArn** *(p. 111)*

The Amazon Resource Name (ARN) of the flow definition you create.

- **Type:** String
- **Length Constraints:** Maximum length of 1024.
- **Pattern:** `arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:flow-definition/.*`

Errors

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](#).

**ResourceInUse**

Resource being accessed is in use.

HTTP Status Code: 400

**ResourceLimitExceeded**

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
CreateHub
Service: Amazon SageMaker Service

Create a hub.

Note
Hub APIs are only callable through SageMaker Studio.

Request Syntax

```json
{
    "HubDescription": "string",
    "HubDisplayName": "string",
    "HubName": "string",
    "HubSearchKeywords": [ "string" ],
    "S3OutputConfig": {
        "S3OutputPath": "string"
    },
    "Tags": [
        { "Key": "string", "Value": "string" }
    ]
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**HubDescription (p. 113)**

A description of the hub.

Type: String

Length Constraints: Maximum length of 1023.

Pattern: .*

Required: Yes

**HubDisplayName (p. 113)**

The display name of the hub.

Type: String

Length Constraints: Maximum length of 255.

Pattern: .*

Required: No

**HubName (p. 113)**

The name of the hub to create.
**CreateHub**

**Type:** String

**Length Constraints:** Maximum length of 63.

**Pattern:** ^[a-zA-Z0-9](\[^a-zA-Z0-9\])*^[a-zA-Z0-9]{0,62}$

**Required:** Yes

**HubSearchKeywords (p. 113)**

The searchable keywords for the hub.

**Type:** Array of strings

**Array Members:** Maximum number of 50 items.

**Length Constraints:** Maximum length of 255.

**Pattern:** ^[^A-Z]*$

**Required:** No

**S3StorageConfig (p. 113)**

The Amazon S3 storage configuration for the hub.

**Type:** HubS3StorageConfig (p. 1514) object

**Required:** No

**Tags (p. 113)**

Any tags to associate with the hub.

**Type:** Array of Tag (p. 1979) objects

**Array Members:** Minimum number of 0 items. Maximum number of 50 items.

**Required:** No

**Response Syntax**

```
{
  "HubArn": "string"
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**HubArn (p. 114)**

The Amazon Resource Name (ARN) of the hub.

**Type:** String

**Length Constraints:** Maximum length of 255.

**Pattern:** .*
Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceInUse

Resource being accessed is in use.

HTTP Status Code: 400

ResourceLimitExceeded

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
CreateHumanTaskUi
Service: Amazon SageMaker Service

Defines the settings you will use for the human review workflow user interface. Reviewers will see a three-panel interface with an instruction area, the item to review, and an input area.

Request Syntax

```json
{
  "HumanTaskUiName": "string",
  "Tags": [ {
    "Key": "string",
    "Value": "string"
  } ],
  "UiTemplate": { 
    "Content": "string"
  }
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

HumanTaskUiName (p. 116)

The name of the user interface you are creating.

Type: String


Pattern: ^[a-z0-9](-[a-z0-9]*)*

Required: Yes

Tags (p. 116)

An array of key-value pairs that contain metadata to help you categorize and organize a human review workflow user interface. Each tag consists of a key and a value, both of which you define.

Type: Array of Tag (p. 1979) objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

UiTemplate (p. 116)

The Liquid template for the worker user interface.

Type: UiTemplate (p. 2064) object

Required: Yes
Response Syntax

```
{
  "HumanTaskUiArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**HumanTaskUiArn (p. 117)**

The Amazon Resource Name (ARN) of the human review workflow user interface you create.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: `arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:human-task-ui/.*`

Errors

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](#).

**ResourceInUse**

Resource being accessed is in use.

HTTP Status Code: 400

**ResourceLimitExceeded**

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
CreateHyperParameterTuningJob

Service: Amazon SageMaker Service

Starts a hyperparameter tuning job. A hyperparameter tuning job finds the best version of a model by running many training jobs on your dataset using the algorithm you choose and values for hyperparameters within ranges that you specify. It then chooses the hyperparameter values that result in a model that performs the best, as measured by an objective metric that you choose.

A hyperparameter tuning job automatically creates Amazon SageMaker experiments, trials, and trial components for each training job that it runs. You can view these entities in Amazon SageMaker Studio. For more information, see View Experiments, Trials, and Trial Components.

Important
Do not include any security-sensitive information including account access IDs, secrets or tokens in any hyperparameter field. If the use of security-sensitive credentials are detected, SageMaker will reject your training job request and return an exception error.

Request Syntax

```
{
    "Autotune": {
        "Mode": "string"
    },
    "HyperParameterTuningJobConfig": {
        "HyperparameterTuningJobObjective": {
            "MetricName": "string",
            "Type": "string"
        },
        "ParameterRanges": {
            "AutoParameters": [
                {
                    "Name": "string",
                    "ValueHint": "string"
                }
            ],
            "CategoricalParameterRanges": [
                {
                    "Name": "string",
                    "Values": [ "string" ]
                }
            ],
            "ContinuousParameterRanges": [
                {
                    "MaxValue": "string",
                    "MinValue": "string",
                    "Name": "string",
                    "ScalingType": "string"
                }
            ],
            "IntegerParameterRanges": [
                {
                    "MaxValue": "string",
                    "MinValue": "string",
                    "Name": "string",
                    "ScalingType": "string"
                }
            ]
        }
    },
    "RandomSeed": number,
    "ResourceLimits": {
        "MaxNumberOfTrainingJobs": number,
        "MaxParallelTrainingJobs": number
    }
}
```
"MaxRuntimeInSeconds": number,
"Strategy": "string",
"StrategyConfig": {
  "HyperbandStrategyConfig": {
    "MaxResource": number,
    "MinResource": number
  }
},
"TrainingJobEarlyStoppingType": "string",
"TuningJobCompletionCriteria": {
  "BestObjectiveNotImproving": {
    "MaxNumberOfTrainingJobsNotImproving": number
  },
  "ConvergenceDetected": {
    "CompleteOnConvergence": "string"
  },
  "TargetObjectiveMetricValue": number
}
},
"HyperParameterTuningJobName": "string",
"Tags": [
  {
    "Key": "string",
    "Value": "string"
  }
],
"TrainingJobDefinition": {
  "AlgorithmSpecification": {
    "AlgorithmName": "string",
    "MetricDefinitions": [
      {
        "Name": "string",
        "Regex": "string"
      }
    ],
    "TrainingImage": "string",
    "TrainingInputMode": "string"
  },
  "CheckpointConfig": {
    "LocalPath": "string",
    "S3Uri": "string"
  },
  "DefinitionName": "string",
  "EnableInterContainerTrafficEncryption": boolean,
  "EnableManagedSpotTraining": boolean,
  "EnableNetworkIsolation": boolean,
  "Environment": {
    "string": "string"
  },
  "HyperParameterRanges": {
    "AutoParameters": [
      {
        "Name": "string",
        "ValueHint": "string"
      }
    ],
    "CategoricalParameterRanges": [
      {
        "Name": "string",
        "Values": [ "string" ]
      }
    ],
    "ContinuousParameterRanges": [
      {
        "MaxValue": "string",
        "MinValue": "string"
      }
    ]
  }
}
"MinValue": "string",
"Name": "string",
"ScalingType": "string"
},
"IntegerParameterRanges": [
{
"MaxValue": "string",
"MinValue": "string",
"Name": "string",
"ScalingType": "string"
}
],
"HyperParameterTuningResourceConfig": {
"AllocationStrategy": "string",
"InstanceConfigs": [
{
"InstanceCount": "number",
"InstanceType": "string",
"VolumeSizeInGB": "number"
}
],
"InstanceCount": "number",
"InstanceType": "string",
"VolumeKmsKeyId": "string",
"VolumeSizeInGB": "number"
},
"InputDataConfig": [
{
"ChannelName": "string",
"CompressType": "string",
"ContentType": "string",
"DataSource": {
"FileSystemDataSource": {
"DirectoryPath": "string",
"FileSystemAccessMode": "string",
"FileSystemId": "string",
"FileSystemType": "string"
},
"S3DataSource": {
"AttributeNames": [ "string" ],
"InstanceGroupNames": [ "string" ],
"S3DataDistributionType": "string",
"S3DataType": "string",
"S3Uri": "string"
}
},
"InputMode": "string",
"RecordWrapperType": "string",
"ShuffleConfig": {
"Seed": "number"
}
}
],
"OutputDataConfig": {
"CompressionType": "string",
"KmsKeyId": "string",
"S3OutputPath": "string"
},
"ResourceConfig": {
"InstanceCount": "number",
"InstanceGroups": [
{
"InstanceCount": "number",
"InstanceGroupName": "string",
"VolumeSizeInGB": "number",
"VolumeType": "string",
"VolumeKmsKeyId": "string"
}
],
"ScalingType": "string"
}
"InstanceType": "string",
"InstanceType": "string",
"KeepAlivePeriodInSeconds": number,
"VolumeKmsKeyId": "string",
"VolumeSizeInGB": number,
"RetryStrategy": {
  "MaximumRetryAttempts": number,
  "RoleArn": "string",
"StaticHyperParameters": {
  "string": "string"
},
"StoppingCondition": {
  "MaxPendingTimeInSeconds": number,
  "MaxRuntimeInSeconds": number,
  "MaxWaitTimeInSeconds": number,
"TuningObjective": {
  "MetricName": "string",
  "Type": "string",
},
"VpcConfig": {
  "SecurityGroupIds": [ "string" ],
  "Subnets": [ "string" ]
} },
"TrainingJobDefinitions": [
  { "AlgorithmSpecification": {
    "AlgorithmName": "string",
    "MetricDefinitions": [
      { "Name": "string",
      "Regex": "string"
    ]},
    "TrainingImage": "string",
    "TrainingInputMode": "string"
  },
  "CheckpointConfig": {
    "LocalPath": "string",
    "S3Uri": "string"
  },
  "DefinitionName": "string",
  "EnableInterContainerTrafficEncryption": boolean,
  "EnableManagedSpotTraining": boolean,
  "EnableNetworkIsolation": boolean,
  "Environment": {
    "string": "string"
  },
  "HyperParameterRanges": {
    "AutoParameters": [
    { "Name": "string",
      "ValueHint": "string"
    ]},
    "CategoricalParameterRanges": [
    { "Name": "string",
      "Values": [ "string" ]
    ]
  },
"ContinuousParameterRanges": [
    {
        "MaxValue": "string",
        "MinValue": "string",
        "Name": "string",
        "ScalingType": "string"
    }
],
"IntegerParameterRanges": [
    {
        "MaxValue": "string",
        "MinValue": "string",
        "Name": "string",
        "ScalingType": "string"
    }
],
"HyperParameterTuningResourceConfig": {
    "AllocationStrategy": "string",
    "InstanceConfigs": [
        {
            "InstanceCount": number,
            "InstanceType": "string",
            "VolumeSizeInGB": number
        }
    ],
    "InstanceCount": number,
    "InstanceType": "string",
    "VolumeKmsKeyId": "string",
    "VolumeSizeInGB": number
},
"InputDataConfig": [
    {
        "ChannelName": "string",
        "CompressionType": "string",
        "ContentType": "string",
        "DataSource": {
            "FileSystemDataSource": {
                "DirectoryPath": "string",
                "FileSystemAccessMode": "string",
                "FileSystemId": "string",
                "FileSystemType": "string"
            },
            "S3DataSource": {
                "AttributeNames": [ "string" ],
                "InstanceGroupNames": [ "string" ],
                "S3DataDistributionType": "string",
                "S3DataFormatType": "string",
                "S3Uri": "string"
            }
        },
        "InputMode": "string",
        "RecordWrapperType": "string",
        "ShuffleConfig": {
            "Seed": number
        }
    }
],
"OutputDataConfig": {
    "CompressionType": "string",
    "KmsKeyId": "string",
    "S3OutputPath": "string"
},
"ResourceConfig": {
    "InstanceCount": number,
    "InstanceGroups": [
        {
            "InstanceCount": number,
            "InstanceType": "string",
            "VolumeKmsKeyId": "string",
            "VolumeSizeInGB": number
        }
    ]
}
Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

Autotune (p. 118)

Configures SageMaker Automatic model tuning (AMT) to automatically find optimal parameters for the following fields:

- **ParameterRanges**: The names and ranges of parameters that a hyperparameter tuning job can optimize.
- **ResourceLimits**: The maximum resources that can be used for a training job. These resources include the maximum number of training jobs, the maximum runtime of a tuning job, and the maximum number of training jobs to run at the same time.
- **TrainingJobEarlyStoppingType**: A flag that specifies whether or not to use early stopping for training jobs launched by a hyperparameter tuning job.
• **RetryStrategy**: The number of times to retry a training job.
• **Strategy**: Specifies how hyperparameter tuning chooses the combinations of hyperparameter values to use for the training jobs that it launches.
• **ConvergenceDetected**: A flag to indicate that Automatic model tuning (AMT) has detected model convergence.

Type: AutoTune object

**HyperParameterTuningJobConfig** (p. 118)

The **HyperParameterTuningJobConfig** object that describes the tuning job, including the search strategy, the objective metric used to evaluate training jobs, ranges of parameters to search, and resource limits for the tuning job. For more information, see How Hyperparameter Tuning Works.

Type: HyperParameterTuningJobConfig object

**HyperParameterTuningJobName** (p. 118)

The name of the tuning job. This name is the prefix for the names of all training jobs that this tuning job launches. The name must be unique within the same AWS account and AWS Region. The name must have 1 to 32 characters. Valid characters are a-z, A-Z, 0-9, and : + = @ _ % - (hyphen). The name is not case sensitive.

Type: String


Pattern: \^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,31\}

Required: Yes

**Tags** (p. 118)

An array of key-value pairs. You can use tags to categorize your AWS resources in different ways, for example, by purpose, owner, or environment. For more information, see Tagging AWS Resources.

Tags that you specify for the tuning job are also added to all training jobs that the tuning job launches.

Type: Array of Tag objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

**TrainingJobDefinition** (p. 118)

The **HyperParameterTrainingJobDefinition** object that describes the training jobs that this tuning job launches, including static hyperparameters, input data configuration, output data configuration, resource configuration, and stopping condition.

Type: HyperParameterTrainingJobDefinition object

Required: No

**TrainingJobDefinitions** (p. 118)

A list of the **HyperParameterTrainingJobDefinition** objects launched for this tuning job.

Type: Array of HyperParameterTrainingJobDefinition objects
Array Members: Minimum number of 1 item. Maximum number of 10 items.

Required: No

**WarmStartConfig (p. 118)**

Specifies the configuration for starting the hyperparameter tuning job using one or more previous tuning jobs as a starting point. The results of previous tuning jobs are used to inform which combinations of hyperparameters to search over in the new tuning job.

All training jobs launched by the new hyperparameter tuning job are evaluated by using the objective metric. If you specify IDENTICAL_DATA_AND_ALGORITHM as the WarmStartType value for the warm start configuration, the training job that performs the best in the new tuning job is compared to the best training jobs from the parent tuning jobs. From these, the training job that performs the best as measured by the objective metric is returned as the overall best training job.

**Note**

All training jobs launched by parent hyperparameter tuning jobs and the new hyperparameter tuning jobs count against the limit of training jobs for the tuning job.

Type: **HyperParameterTuningJobWarmStartConfig (p. 1564)** object

Required: No

**Response Syntax**

```json
{
    "HyperParameterTuningJobArn": "string"
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**HyperParameterTuningJobArn (p. 125)**

The Amazon Resource Name (ARN) of the tuning job. SageMaker assigns an ARN to a hyperparameter tuning job when you create it.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:hyper-parameter-tuning-job/.*

**Errors**

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceInUse**

Resource being accessed is in use.

HTTP Status Code: 400
ResourceLimitExceeded

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
CreateImage
Service: Amazon SageMaker Service

Creates a custom SageMaker image. A SageMaker image is a set of image versions. Each image version represents a container image stored in Amazon Elastic Container Registry (ECR). For more information, see Bring your own SageMaker image.

Request Syntax

```json
{
  "Description": "string",
  "DisplayName": "string",
  "ImageName": "string",
  "RoleArn": "string",
  "Tags": [
    {
      "Key": "string",
      "Value": "string"
    }
  ]
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**Description (p. 127)**

The description of the image.

Type: String


Pattern: .*

Required: No

**DisplayName (p. 127)**

The display name of the image. If not provided, ImageName is displayed.

Type: String


Pattern: ^\S(.*\S)?$\n
Required: No

**ImageName (p. 127)**

The name of the image. Must be unique to your account.

Type: String

Pattern: ^[a-zA-Z0-9]([-.]?[a-zA-Z0-9])\{0,62}\$
Required: Yes

RoleArn (p. 127)
The ARN of an IAM role that enables Amazon SageMaker to perform tasks on your behalf.
Type: String
Pattern: ^arn:aws([-\n-]+)*:iam::\d{12}:role/\?[a-zA-Z0-9=,\@\-_/]+\$
Required: Yes

Tags (p. 127)
A list of tags to apply to the image.
Type: Array of Tag (p. 1979) objects
Array Members: Minimum number of 0 items. Maximum number of 50 items.
Required: No

Response Syntax
```
{
    "ImageArn": "string"
}
```

Response Elements
If the action is successful, the service sends back an HTTP 200 response.
The following data is returned in JSON format by the service.

ImageArn (p. 128)
The ARN of the image.
Type: String
Length Constraints: Maximum length of 256.
Pattern: ^arn:aws([-\w]+)*:sagemaker:.+:\[0-9\]\{12\}:image/[a-zA-Z0-9]([-.]?[a-zA-Z0-9])*$

Errors
For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceInUse
Resource being accessed is in use.
HTTP Status Code: 400
ResourceLimitExceeded

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
CreateImageVersion
Service: Amazon SageMaker Service

Creates a version of the SageMaker image specified by ImageName. The version represents the Amazon Elastic Container Registry (ECR) container image specified by BaseImage.

Request Syntax

```
{
    "Aliases": [ "string" ],
    "BaseImage": "string",
    "ClientToken": "string",
    "Horovod": boolean,
    "ImageName": "string",
    "JobType": "string",
    "MLFramework": "string",
    "Processor": "string",
    "ProgrammingLang": "string",
    "ReleaseNotes": "string",
    "VendorGuidance": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**Aliases (p. 130)**

A list of aliases created with the image version.

Type: Array of strings


Pattern: (?!^[.-])^([a-zA-Z0-9-_.]+)$

Required: No

**BaseImage (p. 130)**

The registry path of the container image to use as the starting point for this version. The path is an Amazon Elastic Container Registry (ECR) URI in the following format:

```
<acct-id>.dkr.ecr.<region>.amazonaws.com/<repo-name[:tag] or [@digest]>
```

Type: String

Length Constraints: Minimum length of 1. Maximum length of 255.

Pattern: .*

Required: Yes

**ClientToken (p. 130)**

A unique ID. If not specified, the AWS CLI and AWS SDKs, such as the SDK for Python (Boto3), add a unique value to the call.
Horovod (p. 130)

Indicates Horovod compatibility.

Type: Boolean

Required: No

ImageName (p. 130)

The ImageName of the Image to create a version of.

Type: String


Pattern: ^[a-zA-Z0-9\-]([\-.]?[a-zA-Z0-9\-])\{0,62}\$

Required: Yes

JobType (p. 130)

Indicates SageMaker job type compatibility.

- TRAINING: The image version is compatible with SageMaker training jobs.
- INFERENCe: The image version is compatible with SageMaker inference jobs.
- NOTEBOOK_KERNEL: The image version is compatible with SageMaker notebook kernels.

Type: String

Valid Values: TRAINING | INFERENCe | NOTEBOOK_KERNEL

Required: No

MLFramework (p. 130)

The machine learning framework vended in the image version.

Type: String


Pattern: ^[a-zA-Z]+ \d+\.\d+(\.\d+)?$}

Required: No

Processor (p. 130)

Indicates CPU or GPU compatibility.

- CPU: The image version is compatible with CPU.
- GPU: The image version is compatible with GPU.

Type: String

Valid Values: CPU | GPU

Required: No
**ProgrammingLang (p. 130)**

The supported programming language and its version.

Type: String


Pattern: `^[a-zA-Z]+ \d+\.\d+(\d+)?$`

Required: No

**ReleaseNotes (p. 130)**

The maintainer description of the image version.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 255.

Pattern: `.+

Required: No

**VendorGuidance (p. 130)**

The stability of the image version, specified by the maintainer.

- **NOT_PROVIDED**: The maintainers did not provide a status for image version stability.
- **STABLE**: The image version is stable.
- **TO_BE_ARCHIVED**: The image version is set to be archived. Custom image versions that are set to be archived are automatically archived after three months.
- **ARCHIVED**: The image version is archived. Archived image versions are not searchable and are no longer actively supported.

Type: String

Valid Values: NOT_PROVIDED | STABLE | TO_BE_ARCHIVED | ARCHIVED

Required: No

**Response Syntax**

```
{
    "ImageVersionArn": "string"
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**ImageVersionArn (p. 132)**

The ARN of the image version.

Type: String

Length Constraints: Maximum length of 256.
Pattern: ^arn:aws(-[\w]+)*:sagemaker:.+:[0-9]{12}:image-version/[a-z0-9](.-)?[a-z0-9]+$/

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceInUse

Resource being accessed is in use.

HTTP Status Code: 400

ResourceLimitExceeded

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

ResourceNotFound

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
CreateInferenceComponent
Service: Amazon SageMaker Service

Creates an inference component, which is a SageMaker hosting object that you can use to deploy a model to an endpoint. In the inference component settings, you specify the model, the endpoint, and how the model utilizes the resources that the endpoint hosts. You can optimize resource utilization by tailoring how the required CPU cores, accelerators, and memory are allocated. You can deploy multiple inference components to an endpoint, where each inference component contains one model and the resource utilization needs for that individual model. After you deploy an inference component, you can directly invoke the associated model when you use the InvokeEndpoint API action.

Request Syntax

```
{
  "EndpointName": "string",
  "InferenceComponentName": "string",
  "RuntimeConfig": {
    "CopyCount": number
  },
  "Specification": {
    "ComputeResourceRequirements": {
      "MaxMemoryRequiredInMb": number,
      "MinMemoryRequiredInMb": number,
      "NumberOfAcceleratorDevicesRequired": number,
      "NumberOfCpuCoresRequired": number
    },
    "Container": {
      "ArtifactUrl": "string",
      "Environment": {
        "string": "string"
      },
      "Image": "string"
    },
    "ModelName": "string",
    "StartupParameters": {
      "ContainerStartupHealthCheckTimeoutInSeconds": number,
      "ModelDataDownloadTimeoutInSeconds": number
    }
  },
  "Tags": [
    {
      "Key": "string",
      "Value": "string"
    }
  ],
  "VariantName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**EndpointName (p. 134)**

The name of an existing endpoint where you host the inference component.

Type: String
InferenceComponentName (p. 134)

A unique name to assign to the inference component.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9)){0,62}$

Required: Yes

RuntimeConfig (p. 134)

Runtime settings for a model that is deployed with an inference component.

Type: InferenceComponentRuntimeConfig (p. 1582) object

Required: Yes

Specification (p. 134)

Details about the resources to deploy with this inference component, including the model, container, and compute resources.

Type: InferenceComponentSpecification (p. 1584) object

Required: Yes

Tags (p. 134)

A list of key-value pairs associated with the model. For more information, see Tagging AWS resources in the AWS General Reference.

Type: Array of Tag (p. 1979) objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

VariantName (p. 134)

The name of an existing production variant where you host the inference component.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

Required: Yes

Response Syntax

```
{
   "InferenceComponentArn": "string"
}
```
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

InferenceComponentArn (p. 135)

The Amazon Resource Name (ARN) of the inference component.

Type: String


Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceLimitExceeded

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
CreateInferenceExperiment

Service: Amazon SageMaker Service

Creates an inference experiment using the configurations specified in the request.

Use this API to setup and schedule an experiment to compare model variants on a Amazon SageMaker inference endpoint. For more information about inference experiments, see Shadow tests.

Amazon SageMaker begins your experiment at the scheduled time and routes traffic to your endpoint's model variants based on your specified configuration.

While the experiment is in progress or after it has concluded, you can view metrics that compare your model variants. For more information, see View, monitor, and edit shadow tests.

Request Syntax

```
{
  "DataStorageConfig": {
    "ContentType": {
      "CsvContentTypes": [ "string" ],
      "JsonContentTypes": [ "string" ]
    },
    "Destination": "string",
    "KmsKey": "string"
  },
  "Description": "string",
  "EndpointName": "string",
  "KmsKey": "string",
  "ModelVariants": [
    {
      "InfrastructureConfig": {
        "InfrastructureType": "string",
        "RealTimeInferenceConfig": {
          "InstanceCount": number,
          "InstanceType": "string"
        }
      },
      "ModelName": "string",
      "VariantName": "string"
    }
  ],
  "Name": "string",
  "RoleArn": "string",
  "Schedule": {
    "EndTime": number,
    "StartTime": number
  },
  "ShadowModeConfig": {
    "ShadowModeConfig": [ {
      "SamplingPercentage": number,
      "ShadowModelVariantName": "string"
    } ],
    "SourceModelVariantName": "string"
  },
  "Tags": [
    {
      "Key": "string",
      "Value": "string"
    }
  ]
}
```
"Type": "string"
}

**Request Parameters**

For information about the parameters that are common to all actions, see [Common Parameters](p. 2178).

The request accepts the following data in JSON format.

**DataStorageConfig (p. 137)**

The Amazon S3 location and configuration for storing inference request and response data.

This is an optional parameter that you can use for data capture. For more information, see [Capture data](p. 138).

Type: InferenceExperimentDataStorageConfig (p. 1590) object

Required: No

**Description (p. 137)**

A description for the inference experiment.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: .*

Required: No

**EndpointName (p. 137)**

The name of the Amazon SageMaker endpoint on which you want to run the inference experiment.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62} 

Required: Yes

**KmsKey (p. 137)**

The AWS Key Management Service (AWS KMS) key that Amazon SageMaker uses to encrypt data on the storage volume attached to the ML compute instance that hosts the endpoint. The KmsKey can be any of the following formats:

- KMS key ID
  "1234abcd-12ab-34cd-56ef-1234567890ab"
- Amazon Resource Name (ARN) of a KMS key
  "arn:aws:kms:us-west-2:111122223333:key/1234abcd-12ab-34cd-56ef-1234567890ab"
- KMS key Alias
  "alias/ExampleAlias"
- Amazon Resource Name (ARN) of a KMS key Alias

If you use a KMS key ID or an alias of your KMS key, the Amazon SageMaker execution role must include permissions to call kms:Encrypt. If you don't provide a KMS key ID, Amazon SageMaker uses the default KMS key for Amazon S3 for your role's account. Amazon SageMaker uses server-side encryption with KMS managed keys for OutputDataConfig. If you use a bucket policy with an s3:PutObject permission that only allows objects with server-side encryption, set the condition key of s3:x-amz-server-side-encryption to "aws:kms". For more information, see KMS managed Encryption Keys in the Amazon Simple Storage Service Developer Guide.

The KMS key policy must grant permission to the IAM role that you specify in your CreateEndpoint and UpdateEndpoint requests. For more information, see Using Key Policies in AWS KMS in the AWS Key Management Service Developer Guide.

Type: String
Length Constraints: Maximum length of 2048.
Pattern: .*
Required: No

**ModelVariants (p. 137)**

An array of ModelVariantConfig objects. There is one for each variant in the inference experiment. Each ModelVariantConfig object in the array describes the infrastructure configuration for the corresponding variant.

Type: Array of ModelVariantConfig (p. 1728) objects
Array Members: Minimum number of 1 item. Maximum number of 2 items.
Required: Yes

**Name (p. 137)**

The name for the inference experiment.

Type: String
Length Constraints: Minimum length of 1. Maximum length of 120.
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,119}$

Required: Yes

**RoleArn (p. 137)**

The ARN of the IAM role that Amazon SageMaker can assume to access model artifacts and container images, and manage Amazon SageMaker Inference endpoints for model deployment.

Type: String
Pattern: ^arn:aws[a-zA-Z\-]*:iam::\d{12}:role/\?a-zA-Z_0-9+=,.@\-_]{0,}\+$

Required: Yes

**Schedule (p. 137)**

The duration for which you want the inference experiment to run. If you don't specify this field, the experiment automatically starts immediately upon creation and concludes after 7 days.
Type: **InferenceExperimentSchedule (p. 1591)** object

Required: No

**ShadowModeConfig (p. 137)**

The configuration of ShadowMode inference experiment type. Use this field to specify a production variant which takes all the inference requests, and a shadow variant to which Amazon SageMaker replicates a percentage of the inference requests. For the shadow variant also specify the percentage of requests that Amazon SageMaker replicates.

Type: **ShadowModeConfig (p. 1950)** object

Required: Yes

**Tags (p. 137)**

Array of key-value pairs. You can use tags to categorize your AWS resources in different ways, for example, by purpose, owner, or environment. For more information, see Tagging your AWS Resources.

Type: Array of **Tag (p. 1979)** objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

**Type (p. 137)**

The type of the inference experiment that you want to run. The following types of experiments are possible:

- **ShadowMode**: You can use this type to validate a shadow variant. For more information, see Shadow tests.

Type: String

Valid Values: ShadowMode

Required: Yes

---

**Response Syntax**

```json
{
   "InferenceExperimentArn": "string"
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**InferenceExperimentArn (p. 140)**

The ARN for your inference experiment.

Type: String

Length Constraints: Maximum length of 256.
Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:inference-experiment/.*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceInUse

Resource being accessed is in use.

HTTP Status Code: 400

ResourceLimitExceeded

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
CreateInferenceRecommendationsJob
Service: Amazon SageMaker Service

Starts a recommendation job. You can create either an instance recommendation or load test job.

Request Syntax

```json
{
  "InputConfig": {
    "ContainerConfig": {
      "DataInputConfig": "string",
      "Domain": "string",
      "Framework": "string",
      "FrameworkVersion": "string",
      "NearestModelName": "string",
      "PayloadConfig": {
        "SamplePayloadUrl": "string",
        "SupportedContentTypes": [ "string" ]
      },
      "SupportedEndpointType": "string",
      "SupportedInstanceTypes": [ "string" ],
      "SupportedResponseMIMETypes": [ "string" ],
      "Task": "string"
    },
    "SupportedEndpointType": "string",
    "SupportedInstanceTypes": [ "string" ],
    "SupportedResponseMIMETypes": [ "string" ],
    "Task": "string"
  },
  "EndpointConfigurations": [
    {
      "EnvironmentParameterRanges": {
        "CategoricalParameterRanges": [
          {
            "Name": "string",
            "Value": [ "string" ]
          }
        ],
        "InferenceSpecificationName": "string",
        "InstanceType": "string",
        "ServerlessConfig": {
          "MaxConcurrency": number,
          "MemorySizeInMB": number,
          "ProvisionedConcurrency": number
        }
      },
      "Endpoints": [
        {
          "EndpointName": "string"
        }
      ],
      "JobDurationInSeconds": number,
      "ModelName": "string",
      "ModelPackageVersionArn": "string",
      "ResourceLimit": {
        "MaxNumberOfTests": number,
        "MaxParallelOfTests": number
      },
      "TrafficPattern": {
        "Phases": [
          {
            "DurationInSeconds": number,
            "InitialNumberOfUsers": number,
            "SpawnRate": number
          }
        ]
      }
    }
  }
}
```
The request accepts the following data in JSON format.

**InputConfig (p. 142)**

Provides information about the versioned model package Amazon Resource Name (ARN), the traffic pattern, and endpoint configurations.

Type: RecommendationJobInputConfig (p. 1892) object

Required: Yes

**JobDescription (p. 142)**

Description of the recommendation job.

Type: String

Length Constraints: Maximum length of 128.
CreateInferenceRecommendationsJob

**JobName (p. 142)**

A name for the recommendation job. The name must be unique within the AWS Region and within your AWS account. The job name is passed down to the resources created by the recommendation job. The names of resources (such as the model, endpoint configuration, endpoint, and compilation) that are prefixed with the job name are truncated at 40 characters.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 64.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,63}$

**JobType (p. 142)**

Defines the type of recommendation job. Specify Default to initiate an instance recommendation and Advanced to initiate a load test. If left unspecified, Amazon SageMaker Inference Recommender will run an instance recommendation (DEFAULT) job.

Type: String

Valid Values: Default | Advanced

**OutputConfig (p. 142)**

Provides information about the output artifacts and the KMS key to use for Amazon S3 server-side encryption.

Type: RecommendationJobOutputConfig (p. 1895) object

**RoleArn (p. 142)**

The Amazon Resource Name (ARN) of an IAM role that enables Amazon SageMaker to perform tasks on your behalf.

Type: String


Pattern: ^arn:aws[a-z\-]*:iam::\d{12}:role/?[a-zA-Z0-9+=,.@\-_/]{0,63}$

**StoppingConditions (p. 142)**

A set of conditions for stopping a recommendation job. If any of the conditions are met, the job is automatically stopped.

Type: RecommendationJobStoppingConditions (p. 1898) object

**Tags (p. 142)**

The metadata that you apply to AWS resources to help you categorize and organize them. Each tag consists of a key and a value, both of which you define. For more information, see Tagging AWS Resources in the AWS General Reference.
CreateInferenceRecommendationsJob

Type: Array of Tag objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

Response Syntax

```
{
    "JobArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**JobArn (p. 145)**

The Amazon Resource Name (ARN) of the recommendation job.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z-]+:sagemaker:[a-z0-9-]*:[0-9]{12}:inference-recommendations-job/.*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceInUse**

Resource being accessed is in use.

HTTP Status Code: 400

**ResourceLimitExceeded**

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
CreateLabelingJob
Service: Amazon SageMaker Service

Creates a job that uses workers to label the data objects in your input dataset. You can use the labeled data to train machine learning models.

You can select your workforce from one of three providers:

- A private workforce that you create. It can include employees, contractors, and outside experts. Use a private workforce when you want the data to stay within your organization or when a specific set of skills is required.
- One or more vendors that you select from the AWS Marketplace. Vendors provide expertise in specific areas.
- The Amazon Mechanical Turk workforce. This is the largest workforce, but it should only be used for public data or data that has been stripped of any personally identifiable information.

You can also use automated data labeling to reduce the number of data objects that need to be labeled by a human. Automated data labeling uses active learning to determine if a data object can be labeled by machine or if it needs to be sent to a human worker. For more information, see Using Automated Data Labeling.

The data objects to be labeled are contained in an Amazon S3 bucket. You create a manifest file that describes the location of each object. For more information, see Using Input and Output Data.

The output can be used as the manifest file for another labeling job or as training data for your machine learning models.

You can use this operation to create a static labeling job or a streaming labeling job. A static labeling job stops if all data objects in the input manifest file identified in ManifestS3Uri have been labeled. A streaming labeling job runs perpetually until it is manually stopped, or remains idle for 10 days. You can send new data objects to an active (InProgress) streaming labeling job in real time. To learn how to create a static labeling job, see Create a Labeling Job (API) in the Amazon SageMaker Developer Guide. To learn how to create a streaming labeling job, see Create a Streaming Labeling Job.

Request Syntax

```json
{
    "HumanTaskConfig": {
        "AnnotationConsolidationConfig": {
            "AnnotationConsolidationLambdaArn": "string"
        },
        "MaxConcurrentTaskCount": number,
        "NumberOfHumanWorkersPerDataObject": number,
        "PreHumanTaskLambdaArn": "string",
        "PublicWorkforceTaskPrice": {
            "AmountInUsd": {
                "Cents": number,
                "Dollars": number,
                "TenthFractionsOfACent": number
            }
        },
        "TaskAvailabilityLifetimeInSeconds": number,
        "TaskDescription": "string",
        "TaskKeywords": [ "string" ],
        "TaskTimeLimitInSeconds": number,
        "TaskTitle": "string",
        "UiConfig": {
            "HumanTaskUiArn": "string",
        }
    }
}
```
Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**HumanTaskConfig (p. 147)**

Configures the labeling task and how it is presented to workers; including, but not limited to price, keywords, and batch size (task count).

Type: HumanTaskConfig (p. 1523) object

Required: Yes
**InputConfig (p. 147)**

Input data for the labeling job, such as the Amazon S3 location of the data objects and the location of the manifest file that describes the data objects.

You must specify at least one of the following: S3DataSource or SnsDataSource.

- Use SnsDataSource to specify an SNS input topic for a streaming labeling job. If you do not specify an SNS input topic ARN, Ground Truth will create a one-time labeling job that stops after all data objects in the input manifest file have been labeled.
- Use S3DataSource to specify an input manifest file for both streaming and one-time labeling jobs. Adding an S3DataSource is optional if you use SnsDataSource to create a streaming labeling job.

If you use the Amazon Mechanical Turk workforce, your input data should not include confidential information, personal information or protected health information. Use ContentClassifiers to specify that your data is free of personally identifiable information and adult content.

Type: **LabelingJobInputConfig (p. 1632)** object

Required: Yes

**LabelAttributeName (p. 147)**

The attribute name to use for the label in the output manifest file. This is the key for the key/value pair formed with the label that a worker assigns to the object. The LabelAttributeName must meet the following requirements.

- The name can't end with "-metadata".
- If you are using one of the following built-in task types, the attribute name must end with "-ref". If the task type you are using is not listed below, the attribute name must not end with "-ref".
  - Image semantic segmentation (SemanticSegmentation), and adjustment (AdjustmentSemanticSegmentation) and verification (VerificationSemanticSegmentation) labeling jobs for this task type.
  - Video frame object detection (VideoObjectDetection), and adjustment and verification (AdjustmentVideoObjectDetection) labeling jobs for this task type.
  - Video frame object tracking (VideoObjectTracking), and adjustment and verification (AdjustmentVideoObjectTracking) labeling jobs for this task type.
  - 3D point cloud semantic segmentation (3DPointCloudSemanticSegmentation), and adjustment and verification (Adjustment3DPointCloudSemanticSegmentation) labeling jobs for this task type.
  - 3D point cloud object tracking (3DPointCloudObjectTracking), and adjustment and verification (Adjustment3DPointCloudObjectTracking) labeling jobs for this task type.

**Important**

If you are creating an adjustment or verification labeling job, you must use a different LabelAttributeName than the one used in the original labeling job. The original labeling job is the Ground Truth labeling job that produced the labels that you want verified or adjusted. To learn more about adjustment and verification labeling jobs, see **Verify and Adjust Labels**.

Type: String


Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,126}`

Required: Yes
LabelCategoryConfigS3Uri (p. 147)

The S3 URI of the file, referred to as a label category configuration file, that defines the categories used to label the data objects.

For 3D point cloud and video frame task types, you can add label category attributes and frame attributes to your label category configuration file. To learn how, see Create a Labeling Category Configuration File for 3D Point Cloud Labeling Jobs.

For named entity recognition jobs, in addition to "labels", you must provide worker instructions in the label category configuration file using the "instructions" parameter: 

```
{"shortInstruction":"<h1>Add header</h1><p>Add Instructions</p>"},
"fullInstruction": "<p>Add additional instructions.</p>"}
```

For details and an example, see Create a Named Entity Recognition Labeling Job (API).

For all other built-in task types and custom tasks, your label category configuration file must be a JSON file in the following format. Identify the labels you want to use by replacing label_1, label_2,...,label_n with your label categories.

```
{
  "document-version": "2018-11-28",
  "labels": [{"label": "label_1"},{"label": "label_2"},...{"label": "label_n"}]
}
```

Note the following about the label category configuration file:

- For image classification and text classification (single and multi-label) you must specify at least two label categories. For all other task types, the minimum number of label categories required is one.

- Each label category must be unique, you cannot specify duplicate label categories.

- If you create a 3D point cloud or video frame adjustment or verification labeling job, you must include auditLabelAttributeName in the label category configuration. Use this parameter to enter the LabelAttributeName of the labeling job you want to adjust or verify annotations of.

  Type: String

  Length Constraints: Maximum length of 1024.

  Pattern: ^(https|s3):/(/([^/]+)/?([^/]+))$

  Required: No

LabelingJobAlgorithmsConfig (p. 147)

Configures the information required to perform automated data labeling.

Type: LabelingJobAlgorithmsConfig (p. 1626) object

Required: No

LabelingJobName (p. 147)

The name of the labeling job. This name is used to identify the job in a list of labeling jobs. Labeling job names must be unique within an AWS account and region. LabelingJobName is not case sensitive. For example, Example-job and example-job are considered the same labeling job name by Ground Truth.

Type: String

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}$

Required: Yes

**OutputConfig (p. 147)**

The location of the output data and the AWS Key Management Service key ID for the key used to encrypt the output data, if any.

Type: LabelingJobOutputConfig (p. 1634) object

Required: Yes

**RoleArn (p. 147)**

The Amazon Resource Number (ARN) that Amazon SageMaker assumes to perform tasks on your behalf during data labeling. You must grant this role the necessary permissions so that Amazon SageMaker can successfully complete data labeling.

Type: String


Pattern: ^arn:aws[a-zA-Z-]*:iam::\d{12}:role/?[a-zA-Z_0-9+=,.@-_\//]+$

Required: Yes

**StoppingConditions (p. 147)**

A set of conditions for stopping the labeling job. If any of the conditions are met, the job is automatically stopped. You can use these conditions to control the cost of data labeling.

Type: LabelingJobStoppingConditions (p. 1639) object

Required: No

**Tags (p. 147)**

An array of key/value pairs. For more information, see Using Cost Allocation Tags in the AWS Billing and Cost Management User Guide.

Type: Array of Tag (p. 1979) objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

**Response Syntax**

```json
{
   "LabelingJobArn": "string"
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.
**LabelingJobArn (p. 151)**

The Amazon Resource Name (ARN) of the labeling job. You use this ARN to identify the labeling job.

Type: String

Length Constraints: Maximum length of 2048.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:labeling-job/.*

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)].

**ResourceInUse**

Resource being accessed is in use.

HTTP Status Code: 400

**ResourceLimitExceeded**

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
CreateModel
Service: Amazon SageMaker Service

Creates a model in SageMaker. In the request, you name the model and describe a primary container. For the primary container, you specify the Docker image that contains inference code, artifacts (from prior training), and a custom environment map that the inference code uses when you deploy the model for predictions.

Use this API to create a model if you want to use SageMaker hosting services or run a batch transform job.

To host your model, you create an endpoint configuration with the CreateEndpointConfig API, and then create an endpoint with the CreateEndpoint API. SageMaker then deploys all of the containers that you defined for the model in the hosting environment.

For an example that calls this method when deploying a model to SageMaker hosting services, see Create a Model (AWS SDK for Python (Boto 3)).

To run a batch transform using your model, you start a job with the CreateTransformJob API. SageMaker uses your model and your dataset to get inferences which are then saved to a specified S3 location.

In the request, you also provide an IAM role that SageMaker can assume to access model artifacts and docker image for deployment on ML compute hosting instances or for batch transform jobs. In addition, you also use the IAM role to manage permissions the inference code needs. For example, if the inference code access any other AWS resources, you grant necessary permissions via this role.

Request Syntax

```json
{
   "Containers": [
      {
         "ContainerHostname": "string",
         "Environment": {
            "string": "string"
         },
         "Image": "string",
         "ImageConfig": {
            "RepositoryAccessMode": "string",
            "RepositoryAuthConfig": {
               "RepositoryCredentialsProviderArn": "string"
            }
         },
         "InferenceSpecificationName": "string",
         "Mode": "string",
         "ModelDataSource": {
            "S3DataSource": {
               "CompressionType": "string",
               "ModelAccessConfig": {
                  "AcceptEula": boolean
               },
               "S3DataType": "string",
               "S3Uri": "string"
            }
         },
         "ModelDataUrl": "string",
         "ModelPackageName": "string",
         "MultiModelConfig": {
            "ModelCacheSetting": "string"
         }
      }
   }
}
```
Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

Containers (p. 153)

Specifies the containers in the inference pipeline.

Type: Array of ContainerDefinition (p. 1366) objects

Array Members: Maximum number of 15 items.
**EnableNetworkIsolation (p. 153)**

Isolates the model container. No inbound or outbound network calls can be made to or from the model container.

Type: Boolean

Required: No

**ExecutionRoleArn (p. 153)**

The Amazon Resource Name (ARN) of the IAM role that SageMaker can assume to access model artifacts and docker image for deployment on ML compute instances or for batch transform jobs. Deploying on ML compute instances is part of model hosting. For more information, see SageMaker Roles.

**Note**

To be able to pass this role to SageMaker, the caller of this API must have the `iam:PassRole` permission.

Type: String


Pattern: `^arn:aws[a-zA-Z\-]*:iam::\d{12}:role/\?[a-zA-Z0-9\-]+,\@\-[/]+$`

Required: No

**InferenceExecutionConfig (p. 153)**

Specifies details of how containers in a multi-container endpoint are called.

Type: `InferenceExecutionConfig (p. 1589)` object

Required: No

**ModelName (p. 153)**

The name of the new model.

Type: String

Length Constraints: Maximum length of 63.

Pattern: `^[a-zA-Z0-9\-]+(-*[a-zA-Z0-9])*`

Required: Yes

**PrimaryContainer (p. 153)**

The location of the primary docker image containing inference code, associated artifacts, and custom environment map that the inference code uses when the model is deployed for predictions.

Type: `ContainerDefinition (p. 1366)` object

Required: No

**Tags (p. 153)**

An array of key-value pairs. You can use tags to categorize your AWS resources in different ways, for example, by purpose, owner, or environment. For more information, see Tagging AWS Resources.

Type: Array of `Tag (p. 1979)` objects
Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

VpcConfig (p. 153)

A VpcConfig object that specifies the VPC that you want your model to connect to. Control access to and from your model container by configuring the VPC. VpcConfig is used in hosting services and in batch transform. For more information, see Protect Endpoints by Using an Amazon Virtual Private Cloud and Protect Data in Batch Transform Jobs by Using an Amazon Virtual Private Cloud.

Type: VpcConfig (p. 2076) object

Required: No

Response Syntax

{
  "ModelArn": "string"
}

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

ModelArn (p. 156)

The ARN of the model created in SageMaker.

Type: String


Pattern: arn:aws[a-z-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:model/.*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceLimitExceededException

You have exceeded a SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
CreateModelBiasJobDefinition
Service: Amazon SageMaker Service

Creates the definition for a model bias job.

Request Syntax

```json
{
  "JobDefinitionName": "string",
  "JobResources": {
    "ClusterConfig": {
      "InstanceCount": number,
      "InstanceType": "string",
      "VolumeKmsKeyId": "string",
      "VolumeSizeInGB": number
    }
  },
  "ModelBiasAppSpecification": {
    "ConfigUri": "string",
    "Environment": {
      "string": "string"
    },
    "ImageUri": "string"
  },
  "ModelBiasBaselineConfig": {
    "BaseliningJobName": "string",
    "ConstraintsResource": {
      "S3Uri": "string"
    }
  },
  "ModelBiasJobInput": {
    "BatchTransformInput": {
      "DataCapturedDestinationS3Uri": "string",
      "DatasetFormat": {
        "Csv": {
          "Header": boolean
        },
        "Json": {
          "Line": boolean
        },
        "Parquet": {
        }
      },
      "EndTimeOffset": "string",
      "ExcludeFeaturesAttribute": "string",
      "FeaturesAttribute": "string",
      "InferenceAttribute": "string",
      "LocalPath": "string",
      "ProbabilityAttribute": "string",
      "ProbabilityThresholdAttribute": number,
      "S3DataDistributionType": "string",
      "S3InputMode": "string",
      "StartTimeOffset": "string"
    },
    "EndpointInput": {
      "EndpointName": "string",
      "EndTimeOffset": "string",
      "ExcludeFeaturesAttribute": "string",
      "FeaturesAttribute": "string",
      "InferenceAttribute": "string",
      "LocalPath": "string",
      "ProbabilityAttribute": "string",
      "ProbabilityThresholdAttribute": number,
      "ProbabilityThresholdAttribute": number
    }
  }
}
```
"S3DataDistributionType": "string",
"S3InputMode": "string",
"StartTimeOffset": "string"
},
"GroundTruthS3Input": {
"S3Uri": "string"
}
],
"ModelBiasJobOutputConfig": {
"KmsKeyId": "string",
"MonitoringOutputs": [
{
"S3Output": {
"LocalPath": "string",
"S3UploadMode": "string",
"S3Uri": "string"
}
]
},
"NetworkConfig": {
"EnableInterContainerTrafficEncryption": boolean,
"EnableNetworkIsolation": boolean,
"VpcConfig": {
"SecurityGroupIds": [ "string" ],
"Subnets": [ "string" ]
}
},
"RoleArn": "string",
"StoppingCondition": {
"MaxRuntimeInSeconds": number
},
"Tags": [
{
"Key": "string",
"Value": "string"
}
]
}

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**JobDefinitionName (p. 158)**

The name of the bias job definition. The name must be unique within an AWS Region in the AWS account.

Type: String


Pattern: ^[a-zA-Z0-9][^-][a-zA-Z0-9]*[^a-zA-Z0-9](-*[a-zA-Z0-9])*[0-9]$

Required: Yes

**JobResources (p. 158)**

Identifies the resources to deploy for a monitoring job.
Type: **MonitoringResources** *(p. 1756)* object

Required: Yes

**ModelBiasAppSpecification (p. 158)**

Configures the model bias job to run a specified Docker container image.

Type: **ModelBiasAppSpecification** *(p. 1660)* object

Required: Yes

**ModelBiasBaselineConfig (p. 158)**

The baseline configuration for a model bias job.

Type: **ModelBiasBaselineConfig** *(p. 1661)* object

Required: No

**ModelBiasJobInput (p. 158)**

Inputs for the model bias job.

Type: **ModelBiasJobInput** *(p. 1662)* object

Required: Yes

**ModelBiasJobOutputConfig (p. 158)**

The output configuration for monitoring jobs.

Type: **MonitoringOutputConfig** *(p. 1754)* object

Required: Yes

**NetworkConfig (p. 158)**

Networking options for a model bias job.

Type: **MonitoringNetworkConfig** *(p. 1752)* object

Required: No

**RoleArn (p. 158)**

The Amazon Resource Name (ARN) of an IAM role that Amazon SageMaker can assume to perform tasks on your behalf.

Type: String


Pattern: ^arn:aws[a-z\-]*:iam::\d{12}:role/\^[a-zA-Z\_\-0-9\=\,\@\-_]+$/

Required: Yes

**StoppingCondition (p. 158)**

A time limit for how long the monitoring job is allowed to run before stopping.

Type: **MonitoringStoppingCondition** *(p. 1765)* object

Required: No

**Tags (p. 158)**

Type: Array of Tag objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

Response Syntax

```
{
  "JobDefinitionArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**JobDefinitionArn (p. 161)**

The Amazon Resource Name (ARN) of the model bias job.

Type: String

Length Constraints: Maximum length of 256.

Pattern: .*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceInUse**

Resource being accessed is in use.

HTTP Status Code: 400

**ResourceLimitExceeded**

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
CreateModelCard
Service: Amazon SageMaker Service

Creates an Amazon SageMaker Model Card.

For information about how to use model cards, see Amazon SageMaker Model Card.

Request Syntax

```json
{
  "Content": "string",
  "ModelCardName": "string",
  "ModelCardStatus": "string",
  "SecurityConfig": {
    "KmsKeyId": "string"
  },
  "Tags": [
    {
      "Key": "string",
      "Value": "string"
    }
  ]
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**Content (p. 163)**

The content of the model card. Content must be in model card JSON schema and provided as a string.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 100000.

Pattern: .*

Required: Yes

**ModelCardName (p. 163)**

The unique name of the model card.

Type: String


Pattern: ^[a-zA-Z0-9][-][a-zA-Z0-9]{0,62}$

Required: Yes

**ModelCardStatus (p. 163)**

The approval status of the model card within your organization. Different organizations might have different criteria for model card review and approval.
• Draft: The model card is a work in progress.
• PendingReview: The model card is pending review.
• Approved: The model card is approved.
• Archived: The model card is archived. No more updates should be made to the model card, but it can still be exported.

Type: String

Valid Values: Draft | PendingReview | Approved | Archived

Required: Yes

**SecurityConfig (p. 163)**

An optional Key Management Service key to encrypt, decrypt, and re-encrypt model card content for regulated workloads with highly sensitive data.

Type: ModelCardSecurityConfig (p. 1670) object

Required: No

**Tags (p. 163)**

Key-value pairs used to manage metadata for model cards.

Type: Array of Tag (p. 1979) objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

**Response Syntax**

```json
{
  "ModelCardArn": "string"
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**ModelCardArn (p. 164)**

The Amazon Resource Name (ARN) of the successfully created model card.

Type: String

Length Constraints: Maximum length of 256.

Pattern: ^arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]{9,16}:[0-9]{12}:model-card/[a-zA-Z0-9](\*[a-zA-Z0-9]){0,62}$

**Errors**

For information about the errors that are common to all actions, see Common Errors (p. 2180).
ConflictException

There was a conflict when you attempted to modify a SageMaker entity such as an Experiment or Artifact.

HTTP Status Code: 400

ResourceLimitExceeded

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
CreateModelCardExportJob

Service: Amazon SageMaker Service

Creates an Amazon SageMaker Model Card export job.

Request Syntax

```json
{
    "ModelCardExportJobName": "string",
    "ModelCardName": "string",
    "ModelCardVersion": number,
    "OutputConfig": {
        "S3OutputPath": "string"
    }
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**ModelCardExportJobName (p. 166)**

The name of the model card export job.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

Required: Yes

**ModelCardName (p. 166)**

The name or Amazon Resource Name (ARN) of the model card to export.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 256.

Pattern: (arn:aws[a-zA-Z-]*:sagemaker:[a-zA-Z0-9-.]*:[0-9]{12}:model-card/.*)?(a-zA-Z0-9)(-[a-zA-Z0-9])\{0,62\}

Required: Yes

**ModelCardVersion (p. 166)**

The version of the model card to export. If a version is not provided, then the latest version of the model card is exported.

Type: Integer

Required: No

**OutputConfig (p. 166)**

The model card output configuration that specifies the Amazon S3 path for exporting.
Type: `ModelCardExportOutputConfig (p. 1669)` object

Required: Yes

**Response Syntax**

```json
{
  "ModelCardExportJobArn": "string"
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**ModelCardExportJobArn (p. 167)**

The Amazon Resource Name (ARN) of the model card export job.

Type: String

Length Constraints: Maximum length of 256.

Pattern: `arn:aws[a-z\-]*:sagemaker:[a-zA-Z0-9\-]{9,16}:[0-9]{12}:model-card/[a-zA-Z0-9\-\(\)*[a-zA-Z0-9\-]]{0,62}/export-job/[a-zA-Z0-9\-\(\)*[a-zA-Z0-9\-]]{0,62}$`

**Errors**

For information about the errors that are common to all actions, see `Common Errors (p. 2180)`.

**ConflictException**

There was a conflict when you attempted to modify a SageMaker entity such as an Experiment or Artifact.

HTTP Status Code: 400

**ResourceLimitExceeded**

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

**ResourceNotFoundException**

Resource being access is not found.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- `AWS Command Line Interface`
- `AWS SDK for .NET`
• AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
CreateModelExplainabilityJobDefinition

Service: Amazon SageMaker Service

Creates the definition for a model explainability job.

Request Syntax

```json
{
  "JobDefinitionName": "string",
  "JobResources": {
    "ClusterConfig": {
      "InstanceCount": number,
      "InstanceType": "string",
      "VolumeKmsKeyId": "string",
      "VolumeSizeInGB": number
    }
  },
  "ModelExplainabilityAppSpecification": {
    "ConfigUri": "string",
    "Environment": {
      "string": "string"
    },
    "ImageUri": "string"
  },
  "ModelExplainabilityBaselineConfig": {
    "BaseliningJobName": "string",
    "ConstraintsResource": {
      "S3Uri": "string"
    }
  },
  "ModelExplainabilityJobInput": {
    "BatchTransformInput": {
      "DataCapturedDestinationS3Uri": "string",
      "DatasetFormat": {
        "Csv": {
          "Header": boolean
        },
        "Json": {
          "Line": boolean
        },
        "Parquet": {
          ...
        }
      },
      "EndTimeOffset": "string",
      "ExcludeFeaturesAttribute": "string",
      "FeaturesAttribute": "string",
      "InferenceAttribute": "string",
      "LocalPath": "string",
      "ProbabilityAttribute": "string",
      "ProbabilityThresholdAttribute": number,
      "S3DataDistributionType": "string",
      "S3InputMode": "string",
      "StartTimeOffset": "string"
    },
    "EndpointInput": {
      "EndpointName": "string",
      "EndTimeOffset": "string",
      "ExcludeFeaturesAttribute": "string",
      "FeaturesAttribute": "string",
      "InferenceAttribute": "string",
      "LocalPath": "string",
      "ProbabilityAttribute": "string",
      "ProbabilityThresholdAttribute": number,
      "S3OutputMode": "string"
    }
  }
}
```
Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**JobDefinitionName (p. 169)**

The name of the model explainability job definition. The name must be unique within an AWS Region in the AWS account.

Type: String


Pattern: ^[a-zA-Z0-9][-]*[a-zA-Z0-9]{0,62}$

Required: Yes

**JobResources (p. 169)**

Identifies the resources to deploy for a monitoring job.

Type: MonitoringResources (p. 1756) object

Required: Yes
**ModelExplainabilityAppSpecification (p. 169)**

Configures the model explainability job to run a specified Docker container image.

Type: ModelExplainabilityAppSpecification (p. 1692) object

Required: Yes

**ModelExplainabilityBaselineConfig (p. 169)**

The baseline configuration for a model explainability job.

Type: ModelExplainabilityBaselineConfig (p. 1693) object

Required: No

**ModelExplainabilityJobInput (p. 169)**

Inputs for the model explainability job.

Type: ModelExplainabilityJobInput (p. 1694) object

Required: Yes

**ModelExplainabilityJobOutputConfig (p. 169)**

The output configuration for monitoring jobs.

Type: MonitoringOutputConfig (p. 1754) object

Required: Yes

**NetworkConfig (p. 169)**

Networking options for a model explainability job.

Type: MonitoringNetworkConfig (p. 1752) object

Required: No

**RoleArn (p. 169)**

The Amazon Resource Name (ARN) of an IAM role that Amazon SageMaker can assume to perform tasks on your behalf.

Type: String


Pattern: ^arn:aws[a-z\-]*:iam::\d{12}:role/?[a-zA-Z_0-9+=,.@/-_\/%]*$

Required: Yes

**StoppingCondition (p. 169)**

A time limit for how long the monitoring job is allowed to run before stopping.

Type: MonitoringStoppingCondition (p. 1765) object

Required: No

**Tags (p. 169)**

(Optional) An array of key-value pairs. For more information, see Using Cost Allocation Tags in the AWS Billing and Cost Management User Guide.

Type: Array of Tag (p. 1979) objects
Array Members: Minimum number of 0 items. Maximum number of 50 items.
Required: No

Response Syntax

```
{
  "JobDefinitionArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.
The following data is returned in JSON format by the service.

**JobDefinitionArn (p. 172)**

The Amazon Resource Name (ARN) of the model explainability job.
Type: String
Length Constraints: Maximum length of 256.
Pattern: .*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceInUse**

Resource being accessed is in use.
HTTP Status Code: 400

**ResourceLimitExceeded**

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.
HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3

172
- AWS SDK for Python
- AWS SDK for Ruby V3
CreateModelPackage
Service: Amazon SageMaker Service

Creates a model package that you can use to create SageMaker models or list on AWS Marketplace, or a versioned model that is part of a model group. Buyers can subscribe to model packages listed on AWS Marketplace to create models in SageMaker.

To create a model package by specifying a Docker container that contains your inference code and the Amazon S3 location of your model artifacts, provide values for `InferenceSpecification`. To create a model from an algorithm resource that you created or subscribed to in AWS Marketplace, provide a value for `SourceAlgorithmSpecification`.

**Note**
There are two types of model packages:
- Versioned - a model that is part of a model group in the model registry.
- Unversioned - a model package that is not part of a model group.

**Request Syntax**

```
{
  "AdditionalInferenceSpecifications": [
    {
      "Containers": [
        {
          "AdditionalS3DataSource": {
            "CompressionType": "string",
            "S3DataType": "string",
            "S3Uri": "string"
          },
          "ContainerHostname": "string",
          "Environment": {
            "string": "string"
          },
          "Framework": "string",
          "FrameworkVersion": "string",
          "Image": "string",
          "ImageDigest": "string",
          "ModelDataUrl": "string",
          "ModelInput": {
            "DataInputConfig": "string"
          },
          "NearestModelName": "string",
          "ProductId": "string"
        }
      ],
      "Description": "string",
      "Name": "string",
      "SupportedContentTypes": [ "string" ],
      "SupportedRealtimeInferenceInstanceTypes": [ "string" ],
      "SupportedResponseMIMETypes": [ "string" ],
      "SupportedTransformInstanceTypes": [ "string" ]
    }
  ],
  "CertifyForMarketplace": boolean,
  "ClientToken": "string",
  "CustomerMetadataProperties": {
    "string": "string"
  },
  "Domain": "string",
  "DriftCheckBaselines": {
  }
}
```
"Bias": {
  "ConfigFile": {
    "ContentDigest": "string",
    "ContentType": "string",
    "S3Uri": "string"
  },
  "PostTrainingConstraints": {
    "ContentDigest": "string",
    "ContentType": "string",
    "S3Uri": "string"
  },
  "PreTrainingConstraints": {
    "ContentDigest": "string",
    "ContentType": "string",
    "S3Uri": "string"
  }
},
"Explainability": {
  "ConfigFile": {
    "ContentDigest": "string",
    "ContentType": "string",
    "S3Uri": "string"
  },
  "Constraints": {
    "ContentDigest": "string",
    "ContentType": "string",
    "S3Uri": "string"
  }
},
"ModelDataQuality": {
  "Constraints": {
    "ContentDigest": "string",
    "ContentType": "string",
    "S3Uri": "string"
  },
  "Statistics": {
    "ContentDigest": "string",
    "ContentType": "string",
    "S3Uri": "string"
  }
},
"ModelQuality": {
  "Constraints": {
    "ContentDigest": "string",
    "ContentType": "string",
    "S3Uri": "string"
  },
  "Statistics": {
    "ContentDigest": "string",
    "ContentType": "string",
    "S3Uri": "string"
  }
},
"InferenceSpecification": {
  "Containers": [
    {
      "AdditionalS3DataSource": {
        "CompressionType": "string",
        "S3DataType": "string",
        "S3Uri": "string"
      },
      "ContainerHostname": "string",
      "Environment": {
        "string": "string"
      }
    }
  ]
}
"Framework": "string",
"FrameworkVersion": "string",
"Image": "string",
"ImageDigest": "string",
"ModelDataUrl": "string",
"ModelInput": {
  "DataInputConfig": "string"
},
"NearestModelName": "string",
"ProductId": "string"
},
"SupportedContentTypes": [ "string" ],
"SupportedRealtimeInferenceInstanceTypes": [ "string" ],
"SupportedResponseMIMETypes": [ "string" ],
"SupportedTransformInstanceTypes": [ "string" ]
},
"MetadataProperties": {
  "CommitId": "string",
  "GeneratedBy": "string",
  "ProjectId": "string",
  "Repository": "string"
},
"ModelApprovalStatus": "string",
"ModelMetrics": {
  "Bias": {
    "PostTrainingReport": {
      "ContentDigest": "string",
      "ContentType": "string",
      "S3Uri": "string"
    },
    "PreTrainingReport": {
      "ContentDigest": "string",
      "ContentType": "string",
      "S3Uri": "string"
    },
    "Report": {
      "ContentDigest": "string",
      "ContentType": "string",
      "S3Uri": "string"
    }
  },
  "Explainability": {
    "Report": {
      "ContentDigest": "string",
      "ContentType": "string",
      "S3Uri": "string"
    }
  },
  "ModelDataQuality": {
    "Constraints": {
      "ContentDigest": "string",
      "ContentType": "string",
      "S3Uri": "string"
    },
    "Statistics": {
      "ContentDigest": "string",
      "ContentType": "string",
      "S3Uri": "string"
    }
  },
  "ModelQuality": {
    "Constraints": {
      "ContentDigest": "string",
      "ContentType": "string",
      "S3Uri": "string"
    }
  }
}
CreateModelPackage

```
{
  "ModelPackageDescription": "string",
  "ModelPackageGroupName": "string",
  "ModelPackageName": "string",
  "SamplePayloadUrl": "string",
  "SkipModelValidation": "string",
  "SourceAlgorithmSpecification": {
    "SourceAlgorithms": [
      {
        "AlgorithmName": "string",
        "ModelDataUrl": "string"
      }
    ],
    "Tags": [ {
      "Key": "string",
      "Value": "string"
    } ],
    "Task": "string",
    "ValidationSpecification": {
      "ValidationProfiles": [
        {
          "ProfileName": "string",
          "TransformJobDefinition": {
            "BatchStrategy": "string",
            "Environment": { "string": "string" },
            "MaxConcurrentTransforms": number,
            "MaxPayloadInMB": number,
            "TransformInput": { "CompressionType": "string",
            "ContentType": "string",
            "DataSource": { "S3DataSource": { "S3DataType": "string",
            "S3Uri": "string"
          } },
            "SplitType": "string"
          },
          "TransformOutput": { "Accept": "string",
            "AssembleWith": "string",
            "KmsKeyId": "string",
            "S3OutputPath": "string"
          },
          "TransformResources": { "InstanceCount": number, "InstanceType": "string", "VolumeKmsKeyId": "string" }
        }
      ]
    },
    "ValidationRole": "string"
  }
}
```
Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**AdditionalInferenceSpecifications (p. 174)**

An array of additional Inference Specification objects. Each additional Inference Specification specifies artifacts based on this model package that can be used on inference endpoints. Generally used with SageMaker Neo to store the compiled artifacts.

Type: Array of AdditionalInferenceSpecificationDefinition (p. 1219) objects

Array Members: Minimum number of 1 item. Maximum number of 15 items.

Required: No

**CertifyForMarketplace (p. 174)**

Whether to certify the model package for listing on AWS Marketplace.

This parameter is optional for unversioned models, and does not apply to versioned models.

Type: Boolean

Required: No

**ClientToken (p. 174)**

A unique token that guarantees that the call to this API is idempotent.

Type: String


Pattern: ^[a-zA-Z0-9-]+$

Required: No

**CustomerMetadataProperties (p. 174)**

The metadata properties associated with the model package versions.

Type: String to string map

Map Entries: Maximum number of 50 items.

Key Length Constraints: Minimum length of 1. Maximum length of 128.

Key Pattern: ^([\p{L}\p{Z}\p{N}\_\.:\/=\+-@]*):\/$\{1,128\}$

Value Length Constraints: Minimum length of 1. Maximum length of 256.

Value Pattern: ^([\p{L}\p{Z}\p{N}\_\.:\/=\+-@]*.bootstrap)\*[\p{L}\p{Z}\p{N}\_\.:\/=\+-@]*):\/$\{1,256\}$

Required: No

**Domain (p. 174)**

The machine learning domain of your model package and its components. Common machine learning domains include computer vision and natural language processing.
DriftCheckBaselines (p. 174)

Represents the drift check baselines that can be used when the model monitor is set using the model package. For more information, see the topic on Drift Detection against Previous Baselines in SageMaker Pipelines in the Amazon SageMaker Developer Guide.

Type: DriftCheckBaselines (p. 1426) object

InferenceSpecification (p. 174)

Specifies details about inference jobs that can be run with models based on this model package, including the following:

- The Amazon ECR paths of containers that contain the inference code and model artifacts.
- The instance types that the model package supports for transform jobs and real-time endpoints used for inference.
- The input and output content formats that the model package supports for inference.

Type: InferenceSpecification (p. 1601) object

MetadataProperties (p. 174)

Metadata properties of the tracking entity, trial, or trial component.

Type: MetadataProperties (p. 1648) object

ModelApprovalStatus (p. 174)

Whether the model is approved for deployment.

This parameter is optional for versioned models, and does not apply to unversioned models.

For versioned models, the value of this parameter must be set to Approved to deploy the model.

Type: String

Valid Values: Approved | Rejected | PendingManualApproval

ModelMetrics (p. 174)

A structure that contains model metrics reports.

Type: ModelMetrics (p. 1701) object

ModelPackageDescription (p. 174)

A description of the model package.

Type: String

Length Constraints: Maximum length of 1024.
Pattern: [\p{L}\p{M}\p{Z}\p{S}\p{N}\p{P}]*

Required: No

ModelPackageGroupName (p. 174)

The name or Amazon Resource Name (ARN) of the model package group that this model version belongs to.

This parameter is required for versioned models, and does not apply to unversioned models.

Type: String


Pattern: (arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:[a-zA-Z\-]*\/)?(a-zA-Z0-9\+[\-]*[a-zA-Z0-9\-])\{0,62\}\{0,1\}$(?<!-)$

Required: No

ModelPackageName (p. 174)

The name of the model package. The name must have 1 to 63 characters. Valid characters are a-z, A-Z, 0-9, and - (hyphen).

This parameter is required for unversioned models. It is not applicable to versioned models.

Type: String


Pattern: ^[a-zA-Z0-9](--[a-zA-Z0-9])\{0,62\}$(?<!-)$

Required: No

SamplePayloadUrl (p. 174)

The Amazon Simple Storage Service (Amazon S3) path where the sample payload is stored. This path must point to a single gzip compressed tar archive (.tar.gz suffix). This archive can hold multiple files that are all equally used in the load test. Each file in the archive must satisfy the size constraints of the InvokeEndpoint call.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: ^((https|s3):/\/)((([^/]+)/?)+\(.*$)

Required: No

SkipModelValidation (p. 174)

Indicates if you want to skip model validation.

Type: String

Valid Values: All | None

Required: No

SourceAlgorithmSpecification (p. 174)

Details about the algorithm that was used to create the model package.

Type: SourceAlgorithmSpecification (p. 1955) object
**Tags (p. 174)**
A list of key value pairs associated with the model. For more information, see Tagging AWS resources in the AWS General Reference Guide.

If you supply ModelPackageGroupName, your model package belongs to the model group you specify and uses the tags associated with the model group. In this case, you cannot supply a tag argument.

Type: Array of Tag (p. 1979) objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

**Task (p. 174)**
The machine learning task your model package accomplishes. Common machine learning tasks include object detection and image classification. The following tasks are supported by Inference Recommender: "IMAGE_CLASSIFICATION" | "OBJECT_DETECTION" | "TEXT_GENERATION" | "IMAGE_SEGMENTATION" | "FILL_MASK" | "CLASSIFICATION" | "REGRESSION" | "OTHER".

Specify "OTHER" if none of the tasks listed fit your use case.

Type: String

**ValidationSpecification (p. 174)**
Specifies configurations for one or more transform jobs that SageMaker runs to test the model package.

Type: ModelPackageValidationSpecification (p. 1719) object

**Response Syntax**
```
{
   "ModelPackageArn": "string"
}
```

**Response Elements**
If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**ModelPackageArn (p. 181)**
The Amazon Resource Name (ARN) of the new model package.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: ^arn:aws(-cn|-us-gov)?:sagemaker:[a-z0-9-]{9,16}:[0-9]{12}:model-package/[^\s]{1,2048}$
Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ConflictException

There was a conflict when you attempted to modify a SageMaker entity such as an Experiment or Artifact.

HTTP Status Code: 400

ResourceLimitExceeded

You have exceeded a SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
CreateModelPackageGroup

Service: Amazon SageMaker Service

Creates a model group. A model group contains a group of model versions.

Request Syntax

```json
{
    "ModelPackageGroupDescription": "string",
    "ModelPackageGroupName": "string",
    "Tags": [
        {
            "Key": "string",
            "Value": "string"
        }
    ]
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**ModelPackageGroupDescription (p. 183)**

A description for the model group.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: `[/\p{L}\p{M}\p{Z}\p{S}\p{N}\p{P}]*`

Required: No

**ModelPackageGroupName (p. 183)**

The name of the model group.

Type: String


Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$`

Required: Yes

**Tags (p. 183)**

A list of key value pairs associated with the model group. For more information, see Tagging AWS resources in the AWS General Reference Guide.

Type: Array of Tag (p. 1979) objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No
Response Syntax

```json
{
  "ModelPackageGroupArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**ModelPackageGroupArn (p. 184)**

The Amazon Resource Name (ARN) of the model group.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `arn:aws(-cn|-us-gov):sagemaker:[a-z0-9-]{9,16}:[0-9]{12}:model-package-group/[\$\]{1,2048}$`

Errors

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](#).

**ResourceLimitExceeded**

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
CreateModelQualityJobDefinition

Service: Amazon SageMaker Service

Creates a definition for a job that monitors model quality and drift. For information about model monitor, see Amazon SageMaker Model Monitor.

Request Syntax

```json
{
  "JobDefinitionName": "string",
  "JobResources": {
    "ClusterConfig": {
      "InstanceCount": number,
      "InstanceType": "string",
      "VolumeKmsKeyId": "string",
      "VolumeSizeInGB": number
    }
  },
  "ModelQualityAppSpecification": {
    "ContainerArguments": [ "string" ],
    "ContainerEntrypoint": [ "string" ],
    "Environment": {
      "string": "string"
    },
    "ImageUri": "string",
    "PostAnalyticsProcessorSourceUri": "string",
    "ProblemType": "string",
    "RecordPreprocessorSourceUri": "string"
  },
  "ModelQualityBaselineConfig": {
    "BaseliningJobName": "string",
    "ConstraintsResource": {
      "S3Uri": "string"
    }
  },
  "ModelQualityJobInput": {
    "BatchTransformInput": {
      "DataCapturedDestinationS3Uri": "string",
      "DatasetFormat": {
        "Csv": {
          "Header": boolean
        }
      },
      "EndTimeOffset": "string",
      "ExcludeFeaturesAttribute": "string",
      "FeaturesAttribute": "string",
      "InferenceAttribute": "string",
      "LocalPath": "string",
      "ProbabilityAttribute": "string",
      "ProbabilityThresholdAttribute": number,
      "S3DataDistributionType": "string",
      "S3InputMode": "string",
      "StartTimeOffset": "string"
    },
    "EndpointInput": {
      "EndpointName": "string",
      "EndTimeOffset": "string",
      "ExcludeFeaturesAttribute": "string",
```
Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

`JobDefinitionName (p. 185)`

The name of the monitoring job definition.

Type: String


Pattern: `^[a-zA-Z0-9-]*[a-zA-Z0-9]([0-9]$`
JobResources (p. 185)

Identifies the resources to deploy for a monitoring job.

Type: MonitoringResources (p. 1756) object

Required: Yes

ModelQualityAppSpecification (p. 185)

The container that runs the monitoring job.

Type: ModelQualityAppSpecification (p. 1721) object

Required: Yes

ModelQualityBaselineConfig (p. 185)

 Specifies the constraints and baselines for the monitoring job.

Type: ModelQualityBaselineConfig (p. 1723) object

Required: No

ModelQualityJobInput (p. 185)

A list of the inputs that are monitored. Currently endpoints are supported.

Type: ModelQualityJobInput (p. 1724) object

Required: Yes

ModelQualityJobOutputConfig (p. 185)

The output configuration for monitoring jobs.

Type: MonitoringOutputConfig (p. 1754) object

Required: Yes

NetworkConfig (p. 185)

Specifies the network configuration for the monitoring job.

Type: MonitoringNetworkConfig (p. 1752) object

Required: No

RoleArn (p. 185)

The Amazon Resource Name (ARN) of an IAM role that Amazon SageMaker can assume to perform tasks on your behalf.

Type: String


Pattern: ^arn:aws[a-z\-]*:iam::\d{12}:role/?[a-zA-Z0-9+=,.@\-_\/]+$

Required: Yes

StoppingCondition (p. 185)

A time limit for how long the monitoring job is allowed to run before stopping.

Type: MonitoringStoppingCondition (p. 1765) object
CreateModelQualityJobDefinition

Required: No

**Tags (p. 185)**

(Optional) An array of key-value pairs. For more information, see Using Cost Allocation Tags in the AWS Billing and Cost Management User Guide.

Type: Array of Tag (p. 1979) objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

Response Syntax

```json
{
   "JobDefinitionArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**JobDefinitionArn (p. 188)**

The Amazon Resource Name (ARN) of the model quality monitoring job.

Type: String

Length Constraints: Maximum length of 256.

Pattern: .*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceInUse**

Resource being accessed is in use.

HTTP Status Code: 400

**ResourceLimitExceeded**

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
• AWS SDK for .NET
• AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
CreateMonitoringSchedule

Service: Amazon SageMaker Service

Creates a schedule that regularly starts Amazon SageMaker Processing Jobs to monitor the data captured for an Amazon SageMaker Endpoint.

Request Syntax

```json
{
  "MonitoringScheduleConfig": {
    "MonitoringJobDefinition": {
      "BaselineConfig": {
        "BaselineJobName": "string",
        "ConstraintsResource": {
          "S3Uri": "string"
        },
        "StatisticsResource": {
          "S3Uri": "string"
        }
      },
      "Environment": {
        "string": "string"
      },
      "MonitoringAppSpecification": {
        "ContainerArguments": [ "string" ],
        "ContainerEntrypoint": [ "string" ],
        "ImageUri": "string",
        "PostAnalyticsProcessorSourceUri": "string",
        "RecordPreprocessorSourceUri": "string"
      },
      "MonitoringInputs": [
        {
          "BatchTransformInput": {
            "DataCapturedDestinationS3Uri": "string",
            "DatasetFormat": {
              "Csv": {
                "Header": boolean
              },
              "Json": {
                "Line": boolean
              }
            },
            "Parquet": {
            }
          },
          "EndTimeOffset": "string",
          "ExcludeFeaturesAttribute": "string",
          "FeaturesAttribute": "string",
          "InferenceAttribute": "string",
          "LocalPath": "string",
          "ProbabilityAttribute": "string",
          "ProbabilityThresholdAttribute": number,
          "S3DataDistributionType": "string",
          "S3InputMode": "string",
          "StartTimeOffset": "string"
        }
      ],
      "EndpointInput": {
        "EndpointName": "string",
        "EndTimeOffset": "string",
        "ExcludeFeaturesAttribute": "string",
        "FeaturesAttribute": "string",
        "InferenceAttribute": "string",
        "LocalPath": "string",
        "ProbabilityAttribute": "string"
      }
    }
  }
}
```
Request Parameters

For information about the parameters that are common to all actions, see [Common Parameters](p. 2178).

The request accepts the following data in JSON format.
MonitoringScheduleConfig (p. 190)

The configuration object that specifies the monitoring schedule and defines the monitoring job.

Type: MonitoringScheduleConfig (p. 1761) object

Required: Yes

MonitoringScheduleName (p. 190)

The name of the monitoring schedule. The name must be unique within an AWS Region within an AWS account.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}\$

Required: Yes

Tags (p. 190)

(Optional) An array of key-value pairs. For more information, see Using Cost Allocation Tags in the AWS Billing and Cost Management User Guide.

Type: Array of Tag (p. 1979) objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

Response Syntax

```
{
    "MonitoringScheduleArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

MonitoringScheduleArn (p. 192)

The Amazon Resource Name (ARN) of the monitoring schedule.

Type: String

Length Constraints: Maximum length of 256.

Pattern: .*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).
ResourceInUse

Resource being accessed is in use.

HTTP Status Code: 400

ResourceLimitExceeded

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
CreateNotebookInstance
Service: Amazon SageMaker Service

Creates an SageMaker notebook instance. A notebook instance is a machine learning (ML) compute instance running on a Jupyter notebook.

In a CreateNotebookInstance request, specify the type of ML compute instance that you want to run. SageMaker launches the instance, installs common libraries that you can use to explore datasets for model training, and attaches an ML storage volume to the notebook instance.

SageMaker also provides a set of example notebooks. Each notebook demonstrates how to use SageMaker with a specific algorithm or with a machine learning framework.

After receiving the request, SageMaker does the following:

1. Creates a network interface in the SageMaker VPC.
2. (Option) If you specified SubnetId, SageMaker creates a network interface in your own VPC, which is inferred from the subnet ID that you provide in the input. When creating this network interface, SageMaker attaches the security group that you specified in the request to the network interface that it creates in your VPC.
3. Launches an EC2 instance of the type specified in the request in the SageMaker VPC. If you specified SubnetId of your VPC, SageMaker specifies both network interfaces when launching this instance. This enables inbound traffic from your own VPC to the notebook instance, assuming that the security groups allow it.

After creating the notebook instance, SageMaker returns its Amazon Resource Name (ARN). You can't change the name of a notebook instance after you create it.

After SageMaker creates the notebook instance, you can connect to the Jupyter server and work in Jupyter notebooks. For example, you can write code to explore a dataset that you can use for model training, train a model, host models by creating SageMaker endpoints, and validate hosted models.

For more information, see How It Works.

Request Syntax

```json
{
   "AcceleratorTypes": [ "string" ],
   "AdditionalCodeRepositories": [ "string" ],
   "DefaultCodeRepository": "string",
   "DirectInternetAccess": "string",
   "InstanceMetadataServiceConfiguration": {
      "MinimumInstanceMetadataServiceVersion": "string"
   },
   "InstanceType": "string",
   "KmsKeyId": "string",
   "LifecycleConfigName": "string",
   "NotebookInstanceName": "string",
   "PlatformIdentifier": "string",
   "RoleArn": "string",
   "RootAccess": "string",
   "SecurityGroupIds": [ "string" ],
   "SubnetId": "string",
   "Tags": [ 
      { "Key": "string", "Value": "string" }
   ]
}
```
Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**AcceleratorTypes (p. 194)**

A list of Elastic Inference (EI) instance types to associate with this notebook instance. Currently, only one instance type can be associated with a notebook instance. For more information, see Using Elastic Inference in Amazon SageMaker.

Type: Array of strings

Valid Values: ml.eia1.medium | ml.eia1.large | ml.eia1.xlarge | ml.eia2.medium | ml.eia2.large | ml.eia2.xlarge

Required: No

**AdditionalCodeRepositories (p. 194)**

An array of up to three Git repositories to associate with the notebook instance. These can be either the names of Git repositories stored as resources in your account, or the URL of Git repositories in AWS CodeCommit or in any other Git repository. These repositories are cloned at the same level as the default repository of your notebook instance. For more information, see Associating Git Repositories with SageMaker Notebook Instances.

Type: Array of strings

Array Members: Maximum number of 3 items.


Pattern: ^https://(\[^/]+)/(.*)$|^[a-zA-Z0-9](-*[a-zA-Z0-9])*$ |^[a-zA-Z0-9](-*[a-zA-Z0-9])*$ |^[a-zA-Z0-9](-*[a-zA-Z0-9])*$ |^[a-zA-Z0-9](-*[a-zA-Z0-9])*$

Required: No

**DefaultCodeRepository (p. 194)**

A Git repository to associate with the notebook instance as its default code repository. This can be either the name of a Git repository stored as a resource in your account, or the URL of a Git repository in AWS CodeCommit or in any other Git repository. When you open a notebook instance, it opens in the directory that contains this repository. For more information, see Associating Git Repositories with SageMaker Notebook Instances.

Type: String


Pattern: ^https://(\[^/]+)/(.*)$|^[a-zA-Z0-9](-*[a-zA-Z0-9])*$ |^[a-zA-Z0-9](-*[a-zA-Z0-9])*$ |^[a-zA-Z0-9](-*[a-zA-Z0-9])*$ |^[a-zA-Z0-9](-*[a-zA-Z0-9])*$

Required: No

**DirectInternetAccess (p. 194)**

Sets whether SageMaker provides internet access to the notebook instance. If you set this to Disabled this notebook instance is able to access resources only in your VPC, and is not be able to


connect to SageMaker training and endpoint services unless you configure a NAT Gateway in your VPC.

For more information, see Notebook Instances Are Internet-Enabled by Default. You can set the value of this parameter to Disabled only if you set a value for the SubnetId parameter.

Type: String

Valid Values: Enabled | Disabled

Required: No

InstanceMetadataServiceConfiguration (p. 194)

Information on the IMDS configuration of the notebook instance

Type: InstanceMetadataServiceConfiguration (p. 1611) object

Required: No

InstanceType (p. 194)

The type of ML compute instance to launch for the notebook instance.

Type: String

Valid Values: ml.t2.medium | ml.t2.large | ml.t2.xlarge | ml.t2.2xlarge | ml.t3.medium | ml.t3.large | ml.t3.xlarge | ml.t3.2xlarge | ml.m4.xlarge | ml.m4.2xlarge | ml.m4.4xlarge | ml.m4.10xlarge | ml.m4.16xlarge | ml.m5.xlarge | ml.m5.2xlarge | ml.m5.4xlarge | ml.m5.12xlarge | ml.m5.24xlarge | ml.m5d.large | ml.m5d.xlarge | ml.m5d.2xlarge | ml.m5d.4xlarge | ml.m5d.8xlarge | ml.m5d.12xlarge | ml.m5d.16xlarge | ml.m5d.24xlarge | ml.c4.xlarge | ml.c4.2xlarge | ml.c4.4xlarge | ml.c4.8xlarge | ml.c5.xlarge | ml.c5.2xlarge | ml.c5.4xlarge | ml.c5.9xlarge | ml.c5.18xlarge | ml.c5d.xlarge | ml.c5d.2xlarge | ml.c5d.4xlarge | ml.c5d.9xlarge | ml.c5d.18xlarge | ml.p2.xlarge | ml.p2.8xlarge | ml.p2.16xlarge | ml.p3.xlarge | ml.p3.8xlarge | ml.p3.16xlarge | ml.p3dn.24xlarge | ml.g4dn.xlarge | ml.g4dn.2xlarge | ml.g4dn.4xlarge | ml.g4dn.8xlarge | ml.g4dn.12xlarge | ml.g4dn.16xlarge | ml.r5.large | ml.r5.xlarge | ml.r5.2xlarge | ml.r5.4xlarge | ml.r5.8xlarge | ml.r5.12xlarge | ml.r5.16xlarge | ml.r5.24xlarge | ml.g5.xlarge | ml.g5.2xlarge | ml.g5.4xlarge | ml.g5.8xlarge | ml.g5.16xlarge | ml.g5.12xlarge | ml.g5.24xlarge | ml.g5.48xlarge | ml.inf1.xlarge | ml.inf1.2xlarge | ml.inf1.6xlarge | ml.inf1.24xlarge | ml.p4d.24xlarge | ml.p4de.24xlarge

Required: Yes

KmsKeyId (p. 194)

The Amazon Resource Name (ARN) of a AWS Key Management Service key that SageMaker uses to encrypt data on the storage volume attached to your notebook instance. The KMS key you provide must be enabled. For information, see Enabling and Disabling Keys in the AWS Key Management Service Developer Guide.

Type: String

Length Constraints: Maximum length of 2048.

Pattern: .*

Required: No
**LifecycleConfigName (p. 194)**

The name of a lifecycle configuration to associate with the notebook instance. For information about lifestyle configurations, see [Step 2.1: (Optional) Customize a Notebook Instance](#).

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])*$

Required: No

**NotebookInstanceName (p. 194)**

The name of the new notebook instance.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])*$

Required: Yes

**PlatformIdentifier (p. 194)**

The platform identifier of the notebook instance runtime environment.

Type: String

Length Constraints: Maximum length of 15.

Pattern: ^(notebook-al1-v1|notebook-al2-v1|notebook-al2-v2)$

Required: No

**RoleArn (p. 194)**

When you send any requests to AWS resources from the notebook instance, SageMaker assumes this role to perform tasks on your behalf. You must grant this role necessary permissions so SageMaker can perform these tasks. The policy must allow the SageMaker service principal (sagemaker.amazonaws.com) permissions to assume this role. For more information, see [SageMaker Roles](#).

Note

To be able to pass this role to SageMaker, the caller of this API must have the `iam:PassRole` permission.

Type: String


Pattern: ^arn:aws[a-zA-Z\-]*:iam::\d{12}:role/\?([a-zA-Z0-9\_\@\-\/]\+)+$^

Required: Yes

**RootAccess (p. 194)**

Whether root access is enabled or disabled for users of the notebook instance. The default value is `Enabled`.

Note

Lifecycle configurations need root access to be able to set up a notebook instance. Because of this, lifecycle configurations associated with a notebook instance always run with root access even if you disable root access for users.
Type: String

Valid Values: Enabled | Disabled

Required: No

**SecurityGroupIds (p. 194)**

The VPC security group IDs, in the form sg-xxxxxxxx. The security groups must be for the same VPC as specified in the subnet.

Type: Array of strings

Array Members: Maximum number of 5 items.

Length Constraints: Maximum length of 32.

Pattern: [-0-9a-zA-Z]+  

Required: No

**SubnetId (p. 194)**

The ID of the subnet in a VPC to which you would like to have a connectivity from your ML compute instance.

Type: String

Length Constraints: Maximum length of 32.

Pattern: [-0-9a-zA-Z]+  

Required: No

**Tags (p. 194)**

An array of key-value pairs. You can use tags to categorize your AWS resources in different ways, for example, by purpose, owner, or environment. For more information, see Tagging AWS Resources.

Type: Array of Tag (p. 1979) objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

**VolumeSizeInGB (p. 194)**

The size, in GB, of the ML storage volume to attach to the notebook instance. The default value is 5 GB.

Type: Integer


Required: No

**Response Syntax**

```json
{
    "NotebookInstanceArn": "string"
}
```
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**NotebookInstanceArn (p. 198)**

The Amazon Resource Name (ARN) of the notebook instance.

Type: String

Length Constraints: Maximum length of 256.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceLimitExceeded**

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
CreateNotebookInstanceLifecycleConfig
Service: Amazon SageMaker Service

Creates a lifecycle configuration that you can associate with a notebook instance. A lifecycle configuration is a collection of shell scripts that run when you create or start a notebook instance.

Each lifecycle configuration script has a limit of 16384 characters.

The value of the $PATH environment variable that is available to both scripts is /sbin:bin:/usr/sbin:/usr/bin.

View CloudWatch Logs for notebook instance lifecycle configurations in log group /aws/sagemaker/NotebookInstances in log stream [notebook-instance-name]/[LifecycleConfigHook].

Lifecycle configuration scripts cannot run for longer than 5 minutes. If a script runs for longer than 5 minutes, it fails and the notebook instance is not created or started.

For information about notebook instance lifestyle configurations, see Step 2.1: (Optional) Customize a Notebook Instance.

Request Syntax

```json
{
   "NotebookInstanceLifecycleConfigName": "string",
   "OnCreate": [
      {
         "Content": "string"
      }
   ],
   "OnStart": [
      {
         "Content": "string"
      }
   ]
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**NotebookInstanceLifecycleConfigName (p. 200)**

The name of the lifecycle configuration.

- Type: String
- Length Constraints: Maximum length of 63.
- Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])*$
- Required: Yes

**OnCreate (p. 200)**

A shell script that runs only once, when you create a notebook instance. The shell script must be a base64-encoded string.
CreateNotebookInstanceLifecycleConfig

Type: Array of NotebookInstanceLifecycleHook objects

Array Members: Maximum number of 1 item.

Required: No

OnStart (p. 200)

A shell script that runs every time you start a notebook instance, including when you create the notebook instance. The shell script must be a base64-encoded string.

Type: Array of NotebookInstanceLifecycleHook objects

Array Members: Maximum number of 1 item.

Required: No

Response Syntax

```json
{
  "NotebookInstanceLifecycleConfigArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

NotebookInstanceLifecycleConfigArn (p. 201)

The Amazon Resource Name (ARN) of the lifecycle configuration.

Type: String

Length Constraints: Maximum length of 256.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceLimitExceeded

You have exceeded a SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
CreatePipeline

Service: Amazon SageMaker Service

Creates a pipeline using a JSON pipeline definition.

Request Syntax

```json
{
    "ClientRequestToken": "string",
    "ParallelismConfiguration": {
        "MaxParallelExecutionSteps": number
    },
    "PipelineDefinition": "string",
    "PipelineDefinitionS3Location": {
        "Bucket": "string",
        "ObjectKey": "string",
        "VersionId": "string"
    },
    "PipelineDescription": "string",
    "PipelineDisplayName": "string",
    "PipelineName": "string",
    "RoleArn": "string",
    "Tags": [
        {
            "Key": "string",
            "Value": "string"
        }
    ]
}
```

Request Parameters

For information about the parameters that are common to all actions, see [Common Parameters](p. 2178).

The request accepts the following data in JSON format.

**ClientRequestToken (p. 203)**

A unique, case-sensitive identifier that you provide to ensure the idempotency of the operation. An idempotent operation completes no more than one time.

Type: String


Required: Yes

**ParallelismConfiguration (p. 203)**

This is the configuration that controls the parallelism of the pipeline. If specified, it applies to all runs of this pipeline by default.

Type: [ParallelismConfiguration (p. 1797)] object

Required: No

**PipelineDefinition (p. 203)**

The [JSON pipeline definition](p. 203) of the pipeline.

Type: String
Pattern: .*(?:\r\n\t).*
Required: No

**PipelineDefinitionS3Location (p. 203)**

The location of the pipeline definition stored in Amazon S3. If specified, SageMaker will retrieve the pipeline definition from this location.

Type: [PipelineDefinitionS3Location (p. 1814)] object

Required: No

**PipelineDescription (p. 203)**

A description of the pipeline.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 3072.

Pattern: .*

Required: No

**PipelineDisplayName (p. 203)**

The display name of the pipeline.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 256.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,255}$

Required: No

**PipelineName (p. 203)**

The name of the pipeline.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 256.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,255}$

Required: Yes

**RoleArn (p. 203)**

The Amazon Resource Name (ARN) of the role used by the pipeline to access and create resources.

Type: String


Pattern: ^arn:aws[a-zA-Z\-]+:iam::\d{12}:role/\?([a-zA-Z0-9][\._=\+\@\-_]+)\$+

Required: Yes

**Tags (p. 203)**

A list of tags to apply to the created pipeline.
Type: Array of [Tag](p. 1979) objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

**Response Syntax**

```
{
   "PipelineArn": "string"
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**PipelineArn (p. 205)**

The Amazon Resource Name (ARN) of the created pipeline.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:pipeline/.*

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)].

**ResourceLimitExceeded**

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
CreatePresignedDomainUrl
Service: Amazon SageMaker Service

Creates a URL for a specified UserProfile in a Domain. When accessed in a web browser, the user will be automatically signed in to the domain, and granted access to all of the Apps and files associated with the Domain's Amazon Elastic File System (EFS) volume. This operation can only be called when the authentication mode equals IAM.

The IAM role or user passed to this API defines the permissions to access the app. Once the presigned URL is created, no additional permission is required to access this URL. IAM authorization policies for this API are also enforced for every HTTP request and WebSocket frame that attempts to connect to the app.

You can restrict access to this API and to the URL that it returns to a list of IP addresses, Amazon VPCs or Amazon VPC Endpoints that you specify. For more information, see Connect to Amazon SageMaker Studio Through an Interface VPC Endpoint.

Note
The URL that you get from a call to CreatePresignedDomainUrl has a default timeout of 5 minutes. You can configure this value using ExpiresInSeconds. If you try to use the URL after the timeout limit expires, you are directed to the AWS console sign-in page.

Request Syntax

```
{
  "DomainId": "string",
  "ExpiresInSeconds": number,
  "LandingUri": "string",
  "SessionExpirationDurationInSeconds": number,
  "SpaceName": "string",
  "UserProfileName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

DomainId (p. 207)

The domain ID.
Type: String
Length Constraints: Maximum length of 63.
Required: Yes

ExpiresInSeconds (p. 207)

The number of seconds until the pre-signed URL expires. This value defaults to 300.
Type: Integer
Required: No
LandingUri (p. 207)

The landing page that the user is directed to when accessing the presigned URL. Using this value, users can access Studio or Studio Classic, even if it is not the default experience for the domain. The supported values are:

- `studio::relative/path`: Directs users to the relative path in Studio.
- `app:JupyterServer:relative/path`: Directs users to the relative path in the Studio Classic application.
- `app:JupyterLab:relative/path`: Directs users to the relative path in the JupyterLab application.
- `app:RStudioServerPro:relative/path`: Directs users to the relative path in the RStudio application.
- `app:Canvas:relative/path`: Directs users to the relative path in the Canvas application.

Type: String

Length Constraints: Maximum length of 1023.

Required: No

SessionExpirationDurationInSeconds (p. 207)

The session expiration duration in seconds. This value defaults to 43200.

Type: Integer


Required: No

SpaceName (p. 207)

The name of the space.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}

Required: No

UserProfileName (p. 207)

The name of the UserProfile to sign-in as.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}

Required: Yes

Response Syntax

```
{
    "AuthorizedUrl": "string"
}
```
Response Elements

If the action is successful, the service sends back an HTTP 200 response. The following data is returned in JSON format by the service.

**AuthorizedUrl (p. 208)**

The presigned URL.

Type: String

Errors

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](#).

**ResourceNotFoundException**

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
CreatePresignedNotebookInstanceUrl

Service: Amazon SageMaker Service

Returns a URL that you can use to connect to the Jupyter server from a notebook instance. In the SageMaker console, when you choose Open next to a notebook instance, SageMaker opens a new tab showing the Jupyter server home page from the notebook instance. The console uses this API to get the URL and show the page.

The IAM role or user used to call this API defines the permissions to access the notebook instance. Once the presigned URL is created, no additional permission is required to access this URL. IAM authorization policies for this API are also enforced for every HTTP request and WebSocket frame that attempts to connect to the notebook instance.

You can restrict access to this API and to the URL that it returns to a list of IP addresses that you specify. Use the NotIpAddress condition operator and the aw:SourceIP condition context key to specify the list of IP addresses that you want to have access to the notebook instance. For more information, see Limit Access to a Notebook Instance by IP Address.

Note
The URL that you get from a call to CreatePresignedNotebookInstanceUrl is valid only for 5 minutes. If you try to use the URL after the 5-minute limit expires, you are directed to the AWS console sign-in page.

Request Syntax

```
{
    "NotebookInstanceName": "string",
    "SessionExpirationDurationInSeconds": number
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**NotebookInstanceName (p. 210)**

The name of the notebook instance.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9-\[\]*a-zA-Z0-9]*

Required: Yes

**SessionExpirationDurationInSeconds (p. 210)**

The duration of the session, in seconds. The default is 12 hours.

Type: Integer


Required: No
Response Syntax

```json
{
  "AuthorizedUrl": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response. The following data is returned in JSON format by the service.

**AuthorizedUrl (p. 211)**

A JSON object that contains the URL string.

Type: String

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
CreateProcessingJob

Service: Amazon SageMaker Service

Creates a processing job.

Request Syntax

```json
{
    "AppSpecification": {
        "ContainerArguments": [ "string" ],
        "ContainerEntrypoint": [ "string" ],
        "ImageUri": "string"
    },
    "Environment": {
        "string": "string"
    },
    "ExperimentConfig": {
        "ExperimentName": "string",
        "RunName": "string",
        "TrialComponentDisplayName": "string",
        "TrialName": "string"
    },
    "NetworkConfig": {
        "EnableInterContainerTrafficEncryption": boolean,
        "EnableNetworkIsolation": boolean,
        "VpcConfig": {
            "SecurityGroupIds": [ "string" ],
            "Subnets": [ "string" ]
        }
    },
    "ProcessingInputs": [
        {
            "AppManaged": boolean,
            "DatasetDefinition": {
                "AthenaDatasetDefinition": {
                    "Catalog": "string",
                    "Database": "string",
                    "KmsKeyId": "string",
                    "OutputCompression": "string",
                    "OutputFormat": "string",
                    "OutputS3Uri": "string",
                    "QueryString": "string",
                    "WorkGroup": "string"
                },
                "DataDistributionType": "string",
                "InputMode": "string",
                "LocalPath": "string",
                "RedshiftDatasetDefinition": {
                    "ClusterId": "string",
                    "ClusterRoleArn": "string",
                    "Database": "string",
                    "DbUser": "string",
                    "KmsKeyId": "string",
                    "OutputCompression": "string",
                    "OutputFormat": "string",
                    "OutputS3Uri": "string",
                    "QueryString": "string"
                }
            },
            "InputName": "string",
            "S3Input": {
                "LocalPath": "string",
                "S3CompressionType": "string"
            }
        }
    ]
}
```
Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**AppSpecification (p. 212)**

Configures the processing job to run a specified Docker container image.

Type: AppSpecification (p. 1250) object

Required: Yes

**Environment (p. 212)**

The environment variables to set in the Docker container. Up to 100 key and values entries in the map are supported.
Type: String to string map

Map Entries: Maximum number of 100 items.

Key Length Constraints: Maximum length of 256.

Key Pattern: `[a-zA-Z_] [a-zA-Z0-9_]`*

Value Length Constraints: Maximum length of 256.

Value Pattern: `\S\s`*

Required: No

**ExperimentConfig (p. 212)**

Associates a SageMaker job as a trial component with an experiment and trial. Specified when you call the following APIs:

- CreateProcessingJob
- CreateTrainingJob
- CreateTransformJob

Type: **ExperimentConfig (p. 1473)** object

Required: No

**NetworkConfig (p. 212)**

Networking options for a processing job, such as whether to allow inbound and outbound network calls to and from processing containers, and the VPC subnets and security groups to use for VPC-enabled processing jobs.

Type: **NetworkConfig (p. 1769)** object

Required: No

**ProcessingInputs (p. 212)**

An array of inputs configuring the data to download into the processing container.

Type: Array of **ProcessingInput (p. 1832)** objects

Array Members: Minimum number of 0 items. Maximum number of 10 items.

Required: No

**ProcessingJobName (p. 212)**

The name of the processing job. The name must be unique within an AWS Region in the AWS account.

Type: String


Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9])*\{0, 62}$`

Required: Yes

**ProcessingOutputConfig (p. 212)**

Output configuration for the processing job.

Type: **ProcessingOutputConfig (p. 1841)** object
CreateProcessingJob

**Required:** No

**ProcessingResources (p. 212)**

Identifies the resources, ML compute instances, and ML storage volumes to deploy for a processing job. In distributed training, you specify more than one instance.

Type: ProcessingResources (p. 1842) object

**RoleArn (p. 212)**

The Amazon Resource Name (ARN) of an IAM role that Amazon SageMaker can assume to perform tasks on your behalf.

Type: String


Pattern: ^arn:aws[a-z\-]*:iam::\d{12}:role/?[a-zA-Z_0-9+=,.@\-_/]+$

**StoppingCondition (p. 212)**

The time limit for how long the processing job is allowed to run.

Type: ProcessingStoppingCondition (p. 1846) object

**Tags (p. 212)**

(Optional) An array of key-value pairs. For more information, see Using Cost Allocation Tags in the AWS Billing and Cost Management User Guide.

Type: Array of Tag (p. 1979) objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

**Response Syntax**

```
{
  "ProcessingJobArn": "string"
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**ProcessingJobArn (p. 215)**

The Amazon Resource Name (ARN) of the processing job.

Type: String

Length Constraints: Maximum length of 256.
Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceInUse

Resource being accessed is in use.

HTTP Status Code: 400

ResourceLimitExceeded

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

ResourceNotFound

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
CreateProject
Service: Amazon SageMaker Service

Creates a machine learning (ML) project that can contain one or more templates that set up an ML pipeline from training to deploying an approved model.

Request Syntax

```json
{
    "ProjectDescription": "string",
    "ProjectName": "string",
    "ServiceCatalogProvisioningDetails": {
        "PathId": "string",
        "ProductId": "string",
        "ProvisioningArtifactId": "string",
        "ProvisioningParameters": [
            {
                "Key": "string",
                "Value": "string"
            }
        ]
    },
    "Tags": [
        {
            "Key": "string",
            "Value": "string"
        }
    ]
}
```

Request Parameters

For information about the parameters that are common to all actions, see [Common Parameters](#).

The request accepts the following data in JSON format.

**ProjectDescription (p. 217)**

A description for the project.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: `[^\p{L}\p{M}\p{Z}\p{S}\p{N}\p{P}]*`

Required: No

**ProjectName (p. 217)**

The name of the project.

Type: String


Pattern: `^[a-zA-Z0-9-]*[a-zA-Z0-9]{$0,31}`

Required: Yes
ServiceCatalogProvisioningDetails (p. 217)

The product ID and provisioning artifact ID to provision a service catalog. The provisioning artifact ID will default to the latest provisioning artifact ID of the product, if you don't provide the provisioning artifact ID. For more information, see What is AWS Service Catalog.

Type: ServiceCatalogProvisioningDetails (p. 1947) object

Required: Yes

Tags (p. 217)

An array of key-value pairs that you want to use to organize and track your AWS resource costs. For more information, see Tagging AWS resources in the AWS General Reference Guide.

Type: Array of Tag (p. 1979) objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

Response Syntax

```json
{
  "ProjectArn": "string",
  "ProjectId": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

ProjectArn (p. 218)

The Amazon Resource Name (ARN) of the project.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: ^arn:aws(-cn|-us-gov)?:sagemaker:[a-z0-9\-]{9,16}:[0-9]{12}:project/\[\S\]{1,2048}$

ProjectId (p. 218)

The ID of the new project.

Type: String


Pattern: ^[a-zA-Z0-9](\-[a-zA-Z0-9]*[a-zA-Z0-9])*$

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).
ResourceLimitExceeded

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
CreateSpace
Service: Amazon SageMaker Service

Creates a space used for real time collaboration in a Domain.

Request Syntax

```json
{
   "DomainId": "string",
   "OwnershipSettings": {
      "OwnerUserProfileName": "string"
   },
   "SpaceDisplayName": "string",
   "SpaceName": "string",
   "SpaceSettings": {
      "AppType": "string",
      "CodeEditorAppSettings": {
         "DefaultResourceSpec": {
            "InstanceType": "string",
            "LifecycleConfigArn": "string",
            "SageMakerImageArn": "string",
            "SageMakerImageVersionAlias": "string",
            "SageMakerImageVersionArn": "string"
         }
      },
      "CustomFileSystems": [
         ...]
      },
      "JupyterLabAppSettings": {
         "CodeRepositories": [
            {
               "RepositoryUrl": "string"
            }
         ],
         "DefaultResourceSpec": {
            "InstanceType": "string",
            "LifecycleConfigArn": "string",
            "SageMakerImageArn": "string",
            "SageMakerImageVersionAlias": "string",
            "SageMakerImageVersionArn": "string"
         }
      },
      "JupyterServerAppSettings": {
         "CodeRepositories": [
            {
               "RepositoryUrl": "string"
            }
         ],
         "DefaultResourceSpec": {
            "InstanceType": "string",
            "LifecycleConfigArn": "string",
            "SageMakerImageArn": "string",
            "SageMakerImageVersionAlias": "string",
            "SageMakerImageVersionArn": "string"
         },
         "LifecycleConfigArns": [ "string" ]
      },
      "KernelGatewayAppSettings": {
         "CustomImages": [
            {
               "AppImageConfigName": "string",
               "ImageName": "string",
               "ImageVersionNumber": number
            }
         ]
      }
   }
}
```
Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**DomainId (p. 220)**

The ID of the associated Domain.

Type: String

Length Constraints: Maximum length of 63.

Required: Yes

**OwnershipSettings (p. 220)**

A collection of ownership settings.

Type: OwnershipSettings (p. 1795) object

Required: No

**SpaceDisplayName (p. 220)**

The name of the space that appears in the SageMaker Studio UI.

Type: String

Length Constraints: Maximum length of 64.

Pattern: ^(?!(?!s*$).+$

Required: No
SpaceName (p. 220)
The name of the space.
Type: String
Length Constraints: Maximum length of 63.
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$
Required: Yes

SpaceSettings (p. 220)
A collection of space settings.
Type: SpaceSettings (p. 1961) object
Required: No

SpaceSharingSettings (p. 220)
A collection of space sharing settings.
Type: SpaceSharingSettings (p. 1964) object
Required: No

Tags (p. 220)
Tags to associated with the space. Each tag consists of a key and an optional value. Tag keys must be unique for each resource. Tags are searchable using the Search API.
Type: Array of Tag (p. 1979) objects
Array Members: Minimum number of 0 items. Maximum number of 50 items.
Required: No

Response Syntax

```
{
    "SpaceArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

SpaceArn (p. 222)
The space's Amazon Resource Name (ARN).
Type: String
Length Constraints: Maximum length of 256.
Pattern: arn:aws[a-z\-]*:sagemaker:[a-zA-Z0-9-]*:[0-9]{12}:space/.*
Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceInUse

Resource being accessed is in use.

HTTP Status Code: 400

ResourceLimitExceeded

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
CreateStudioLifecycleConfig
Service: Amazon SageMaker Service

Creates a new Amazon SageMaker Studio Lifecycle Configuration.

Request Syntax

```
{
    "StudioLifecycleConfigAppType": "string",
    "StudioLifecycleConfigContent": "string",
    "StudioLifecycleConfigName": "string",
    "Tags": [
        {
            "Key": "string",
            "Value": "string"
        }
    ]
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**StudioLifecycleConfigAppType (p. 224)**

The App type that the Lifecycle Configuration is attached to.

- Type: String
- Valid Values: JupyterServer | KernelGateway | JupyterLab | CodeEditor
- Required: Yes

**StudioLifecycleConfigContent (p. 224)**

The content of your Amazon SageMaker Studio Lifecycle Configuration script. This content must be base64 encoded.

- Type: String
- Pattern: [\S\s]+
- Required: Yes

**StudioLifecycleConfigName (p. 224)**

The name of the Amazon SageMaker Studio Lifecycle Configuration to create.

- Type: String
- Length Constraints: Maximum length of 63.
- Pattern: ^[a-zA-Z0-9-0-9](-*[a-zA-Z0-9])$\{0,62$
- Required: Yes
Tags (p. 224)

Tags to be associated with the Lifecycle Configuration. Each tag consists of a key and an optional value. Tag keys must be unique per resource. Tags are searchable using the Search API.

Type: Array of Tag (p. 1979) objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

Response Syntax

```json
{
    "StudioLifecycleConfigArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

StudioLifecycleConfigArn (p. 225)

The ARN of your created Lifecycle Configuration.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:studio-lifecycle-config/.*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceInUse

Resource being accessed is in use.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
CreateStudioLifecycleConfig

- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
CreateTrainingJob

Service: Amazon SageMaker Service

Starts a model training job. After training completes, SageMaker saves the resulting model artifacts to an Amazon S3 location that you specify.

If you choose to host your model using SageMaker hosting services, you can use the resulting model artifacts as part of the model. You can also use the artifacts in a machine learning service other than SageMaker, provided that you know how to use them for inference.

In the request body, you provide the following:

- **AlgorithmSpecification** - Identifies the training algorithm to use.
- **HyperParameters** - Specify these algorithm-specific parameters to enable the estimation of model parameters during training. Hyperparameters can be tuned to optimize this learning process. For a list of hyperparameters for each training algorithm provided by SageMaker, see [Algorithms](#).
  
  **Important**
  Do not include any security-sensitive information including account access IDs, secrets or tokens in any hyperparameter field. If the use of security-sensitive credentials are detected, SageMaker will reject your training job request and return an exception error.

- **InputDataConfig** - Describes the input required by the training job and the Amazon S3, EFS, or FSx location where it is stored.
- **OutputDataConfig** - Identifies the Amazon S3 bucket where you want SageMaker to save the results of model training.
- **ResourceConfig** - Identifies the resources, ML compute instances, and ML storage volumes to deploy for model training. In distributed training, you specify more than one instance.
- **EnableManagedSpotTraining** - Optimize the cost of training machine learning models by up to 80% by using Amazon EC2 Spot instances. For more information, see [Managed Spot Training](#).
- **RoleArn** - The Amazon Resource Name (ARN) that SageMaker assumes to perform tasks on your behalf during model training. You must grant this role the necessary permissions so that SageMaker can successfully complete model training.
- **StoppingCondition** - To help cap training costs, use `MaxRuntimeInSeconds` to set a time limit for training. Use `MaxWaitTimeInSeconds` to specify how long a managed spot training job has to complete.
- **Environment** - The environment variables to set in the Docker container.
- **RetryStrategy** - The number of times to retry the job when the job fails due to an `InternalServerError`.

For more information about SageMaker, see [How It Works](#).

**Request Syntax**

```json
{
    "AlgorithmSpecification": {
        "AlgorithmName": "string",
        "ContainerArguments": [ "string" ],
        "ContainerEntrypoint": [ "string" ],
        "EnableSageMakerMetricsTimeSeries": boolean,
        "MetricDefinitions": [
            {
                "Name": "string",
                "Regex": "string"
            }
        ]
    },
    "InputDataConfig": {
    },
    "OutputDataConfig": {
    },
    "ResourceConfig": {
    },
    "EnableManagedSpotTraining": boolean,
    "RoleArn": "string",
    "StoppingCondition": {
        "MaxRuntimeInSeconds": long,
        "MaxWaitTimeInSeconds": long
    },
    "Environment": {}
}
```
"TrainingImage": "string",
"TrainingImageConfig": {
  "TrainingRepositoryAccessMode": "string",
  "TrainingRepositoryAuthConfig": {
    "TrainingRepositoryCredentialsProviderArn": "string"
  }
},
"TrainingInputMode": "string"
},
"CheckpointConfig": {
  "LocalPath": "string",
  "S3Uri": "string"
},
"DebugHookConfig": {
  "CollectionConfigurations": [
    {
      "CollectionName": "string",
      "CollectionParameters": {
        "string": "string"
      }
    }
  ],
  "HookParameters": {
    "string": "string"
  },
  "LocalPath": "string",
  "S3OutputPath": "string"
},
"DebugRuleConfigurations": [
  {
    "InstanceType": "string",
    "LocalPath": "string",
    "RuleConfigurationName": "string",
    "RuleEvaluatorImage": "string",
    "RuleParameters": {
      "string": "string"
    },
    "S3OutputPath": "string",
    "VolumeSizeInGB": number
  }
},
"EnableInterContainerTrafficEncryption": boolean,
"EnableManagedSpotTraining": boolean,
"EnableNetworkIsolation": boolean,
"Environment": {
  "string": "string"
},
"ExperimentConfig": {
  "ExperimentName": "string",
  "RunName": "string",
  "TrialComponentDisplayName": "string",
  "TrialName": "string"
},
"HyperParameters": {
  "string": "string"
},
"InfraCheckConfig": {
  "EnableInfraCheck": boolean
},
"InputDataConfig": [
  {
    "ChannelName": "string",
    "CompressionType": "string",
    "ContentType": "string",
    "DataSource": {
      "FileSystemDataSource": {
      }
    }
  }
]
CreateTrainingJob

```
  "DirectoryPath": "string",
  "FileSystemAccessMode": "string",
  "FileSystemId": "string",
  "FileSystemType": "string"
},
  "S3DataSource": {
    "AttributeNames": [ "string" ],
    "InstanceGroupNames": [ "string" ],
    "S3DataDistributionType": "string",
    "S3DataType": "string",
    "S3Uri": "string"
  }
},
"InputMode": "string",
"RecordWrapperType": "string",
"ShuffleConfig": {
  "Seed": number
}
],
"OutputDataConfig": {
  "CompressionType": "string",
  "KmsKeyId": "string",
  "S3OutputPath": "string"
},
"ProfilerConfig": {
  "DisableProfiler": boolean,
  "ProfilingIntervalInMilliseconds": number,
  "ProfilingParameters": {
    "string": "string"
  },
  "S3OutputPath": "string"
},
"ProfilerRuleConfigurations": [
  {
    "InstanceType": "string",
    "LocalPath": "string",
    "RuleConfigurationName": "string",
    "RuleEvaluatorImage": "string",
    "RuleParameters": {
      "string": "string"
    },
    "S3OutputPath": "string",
    "VolumeSizeInGB": number
  }
],
"ResourceConfig": {
  "InstanceCount": number,
  "InstanceGroups": [
    {
      "InstanceCount": number,
      "InstanceGroupName": "string",
      "InstanceType": "string"
    }
  ],
  "InstanceType": "string",
  "KeepAlivePeriodInSeconds": number,
  "VolumeKmsKeyId": "string",
  "VolumeSizeInGB": number
},
"RetryStrategy": {
  "MaximumRetryAttempts": number
},
"RoleArn": "string",
"StoppingCondition": {
  "MaxPendingTimeInSeconds": number,
```

229
Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

AlgorithmSpecification (p. 227)

The registry path of the Docker image that contains the training algorithm and algorithm-specific metadata, including the input mode. For more information about algorithms provided by SageMaker, see Algorithms. For information about providing your own algorithms, see Using Your Own Algorithms with Amazon SageMaker.

Type: AlgorithmSpecification (p. 1225) object

Required: Yes

CheckpointConfig (p. 227)

Contains information about the output location for managed spot training checkpoint data.

Type: CheckpointConfig (p. 1328) object

Required: No

DebugHookConfig (p. 227)

Configuration information for the Amazon SageMaker Debugger hook parameters, metric and tensor collections, and storage paths. To learn more about how to configure the DebugHookConfig parameter, see Use the SageMaker and Debugger Configuration API Operations to Create, Update, and Debug Your Training Job.

Type: DebugHookConfig (p. 1395) object

Required: No

DebugRuleConfigurations (p. 227)

Configuration information for Amazon SageMaker Debugger rules for debugging output tensors.

Type: Array of DebugRuleConfiguration (p. 1397) objects

Array Members: Minimum number of 0 items. Maximum number of 20 items.
To encrypt all communications between ML compute instances in distributed training, choose True. Encryption provides greater security for distributed training, but training might take longer. How long it takes depends on the amount of communication between compute instances, especially if you use a deep learning algorithm in distributed training. For more information, see Protect Communications Between ML Compute Instances in a Distributed Training Job.

Type: Boolean

Required: No

EnableManagedSpotTraining (p. 227)

To train models using managed spot training, choose True. Managed spot training provides a fully managed and scalable infrastructure for training machine learning models. This option is useful when training jobs can be interrupted and when there is flexibility when the training job is run.

The complete and intermediate results of jobs are stored in an Amazon S3 bucket, and can be used as a starting point to train models incrementally. Amazon SageMaker provides metrics and logs in CloudWatch. They can be used to see when managed spot training jobs are running, interrupted, resumed, or completed.

Type: Boolean

Required: No

EnableNetworkIsolation (p. 227)

Isolates the training container. No inbound or outbound network calls can be made, except for calls between peers within a training cluster for distributed training. If you enable network isolation for training jobs that are configured to use a VPC, SageMaker downloads and uploads customer data and model artifacts through the specified VPC, but the training container does not have network access.

Type: Boolean

Required: No

Environment (p. 227)

The environment variables to set in the Docker container.

Type: String to string map

Map Entries: Maximum number of 100 items.

Key Length Constraints: Maximum length of 512.

Key Pattern: [a-zA-Z_]

Value Length Constraints: Maximum length of 512.

Value Pattern: [\S\s]*

Required: No

ExperimentConfig (p. 227)

Associates a SageMaker job as a trial component with an experiment and trial. Specified when you call the following APIs:

* CreateProcessingJob
CreateTrainingJob
CreateTransformJob

Type: **ExperimentConfig** *(p. 1473)* object

Required: No

**HyperParameters (p. 227)**

Algorithm-specific parameters that influence the quality of the model. You set hyperparameters before you start the learning process. For a list of hyperparameters for each training algorithm provided by SageMaker, see *Algorithms*.

You can specify a maximum of 100 hyperparameters. Each hyperparameter is a key-value pair. Each key and value is limited to 256 characters, as specified by the Length Constraint.

**Important**

Do not include any security-sensitive information including account access IDs, secrets or tokens in any hyperparameter field. If the use of security-sensitive credentials are detected, SageMaker will reject your training job request and return an exception error.

Type: String to string map

Map Entries: Minimum number of 0 items. Maximum number of 100 items.

Key Length Constraints: Maximum length of 256.

Key Pattern: .*

Value Length Constraints: Maximum length of 2500.

Value Pattern: .*

Required: No

**InfraCheckConfig (p. 227)**

Contains information about the infrastructure health check configuration for the training job.

Type: **InfraCheckConfig** *(p. 1604)* object

Required: No

**InputDataConfig (p. 227)**

An array of **Channel** objects. Each channel is a named input source. InputDataConfig describes the input data and its location.

Algorithms can accept input data from one or more channels. For example, an algorithm might have two channels of input data, *training_data* and *validation_data*. The configuration for each channel provides the S3, EFS, or FSx location where the input data is stored. It also provides information about the stored data: the MIME type, compression method, and whether the data is wrapped in RecordIO format.

Depending on the input mode that the algorithm supports, SageMaker either copies input data files from an S3 bucket to a local directory in the Docker container, or makes it available as input streams. For example, if you specify an EFS location, input data files are available as input streams. They do not need to be downloaded.

Your input must be in the same AWS region as your training job.

Type: Array of **Channel** *(p. 1324)* objects
Array Members: Minimum number of 1 item. Maximum number of 20 items.

Required: No

**OutputDataConfig (p. 227)**

Specifies the path to the S3 location where you want to store model artifacts. SageMaker creates subfolders for the artifacts.

Type: **OutputDataConfig (p. 1792) object**

Required: Yes

**ProfilerConfig (p. 227)**

Configuration information for Amazon SageMaker Debugger system monitoring, framework profiling, and storage paths.

Type: **ProfilerConfig (p. 1860) object**

Required: No

**ProfilerRuleConfigurations (p. 227)**

Configuration information for Amazon SageMaker Debugger rules for profiling system and framework metrics.

Type: Array of **ProfilerRuleConfiguration (p. 1864) objects**

Array Members: Minimum number of 0 items. Maximum number of 20 items.

Required: No

**ResourceConfig (p. 227)**

The resources, including the ML compute instances and ML storage volumes, to use for model training.

ML storage volumes store model artifacts and incremental states. Training algorithms might also use ML storage volumes for scratch space. If you want SageMaker to use the ML storage volume to store the training data, choose **File** as the TrainingInputMode in the algorithm specification. For distributed training algorithms, specify an instance count greater than 1.

Type: **ResourceConfig (p. 1910) object**

Required: Yes

**RetryStrategy (p. 227)**

The number of times to retry the job when the job fails due to an **InternalServerError**.

Type: **RetryStrategy (p. 1918) object**

Required: No

**RoleArn (p. 227)**

The Amazon Resource Name (ARN) of an IAM role that SageMaker can assume to perform tasks on your behalf.

During model training, SageMaker needs your permission to read input data from an S3 bucket, download a Docker image that contains training code, write model artifacts to an S3 bucket, write logs to Amazon CloudWatch Logs, and publish metrics to Amazon CloudWatch. You grant permissions for all of these tasks to an IAM role. For more information, see [SageMaker Roles](#).
Note
To be able to pass this role to SageMaker, the caller of this API must have the `iam:PassRole` permission.

Type: String
Pattern: ^arn:aws[a-z\-]*:iam::\d{12}:role/?[a-zA-Z0-9\-_]+\+$
Required: Yes

StoppingCondition (p. 227)
Specifies a limit to how long a model training job can run. It also specifies how long a managed Spot training job has to complete. When the job reaches the time limit, SageMaker ends the training job. Use this API to cap model training costs.

To stop a job, SageMaker sends the algorithm the SIGTERM signal, which delays job termination for 120 seconds. Algorithms can use this 120-second window to save the model artifacts, so the results of training are not lost.

Type: StoppingCondition (p. 1968) object
Required: Yes

Tags (p. 227)
An array of key-value pairs. You can use tags to categorize your AWS resources in different ways, for example, by purpose, owner, or environment. For more information, see Tagging AWS Resources.

Type: Array of Tag (p. 1979) objects
Array Members: Minimum number of 0 items. Maximum number of 50 items.
Required: No

TensorBoardOutputConfig (p. 227)
Configuration of storage locations for the Amazon SageMaker Debugger TensorBoard output data.

Type: TensorBoardOutputConfig (p. 1984) object
Required: No

TrainingJobName (p. 227)
The name of the training job. The name must be unique within an AWS Region in an AWS account.

Type: String
Pattern: ^[a-zA-Z0-9\-_0-9]*(\*[^a-zA-Z0-9\-_0-9])\{0,62}$
Required: Yes

VpcConfig (p. 227)
A VpcConfig object that specifies the VPC that you want your training job to connect to. Control access to and from your training container by configuring the VPC. For more information, see Protect Training Jobs by Using an Amazon Virtual Private Cloud.

Type: VpcConfig (p. 2076) object
CreateTrainingJob

Required: No

Response Syntax

```
{
  "TrainingJobArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**TrainingJobArn (p. 235)**

The Amazon Resource Name (ARN) of the training job.

- **Type:** String
- **Length Constraints:** Maximum length of 256.
- **Pattern:** arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:training-job/.*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceInUse**

Resource being accessed is in use.

HTTP Status Code: 400

**ResourceLimitExceeded**

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
CreateTransformJob

Service: Amazon SageMaker Service

Starts a transform job. A transform job uses a trained model to get inferences on a dataset and saves these results to an Amazon S3 location that you specify.

To perform batch transformations, you create a transform job and use the data that you have readily available.

In the request body, you provide the following:

- **TransformJobName** - Identifies the transform job. The name must be unique within an AWS Region in an AWS account.
- **ModelName** - Identifies the model to use. ModelName must be the name of an existing Amazon SageMaker model in the same AWS Region and AWS account. For information on creating a model, see [CreateModel](#).
- **TransformInput** - Describes the dataset to be transformed and the Amazon S3 location where it is stored.
- **TransformOutput** - Identifies the Amazon S3 location where you want Amazon SageMaker to save the results from the transform job.
- **TransformResources** - Identifies the ML compute instances for the transform job.

For more information about how batch transformation works, see [Batch Transform](#).

**Request Syntax**

```json
{
    "BatchStrategy": "string",
    "DataCaptureConfig": {
        "DestinationS3Uri": "string",
        "GenerateInferenceId": boolean,
        "KmsKeyId": "string"
    },
    "DataProcessing": {
        "InputFilter": "string",
        "JoinSource": "string",
        "OutputFilter": "string"
    },
    "Environment": {
        "string": "string"
    },
    "ExperimentConfig": {
        "ExperimentName": "string",
        "RunName": "string",
        "TrialComponentDisplayName": "string",
        "TrialName": "string"
    },
    "MaxConcurrentTransforms": number,
    "MaxPayloadInMB": number,
    "ModelClientConfig": {
        "InvocationsMaxRetries": number,
        "InvocationsTimeoutInSeconds": number
    },
    "ModelName": "string",
    "Tags": [
        {
            "Key": "string",
            "Value": "string"
        }
    ]
}
```
Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**BatchStrategy (p. 237)**

Specifies the number of records to include in a mini-batch for an HTTP inference request. A record is a single unit of input data that inference can be made on. For example, a single line in a CSV file is a record.

To enable the batch strategy, you must set the SplitType property to Line, RecordIO, or TFRecord.

To use only one record when making an HTTP invocation request to a container, set BatchStrategy to SingleRecord and SplitType to Line.

To fit as many records in a mini-batch as can fit within the MaxPayloadInMB limit, set BatchStrategy to MultiRecord and SplitType to Line.

Type: String

Valid Values: MultiRecord | SingleRecord

Required: No

**DataCaptureConfig (p. 237)**

Configuration to control how SageMaker captures inference data.

Type: BatchDataCaptureConfig (p. 1301) object

Required: No
DataProcessing (p. 237)
The data structure used to specify the data to be used for inference in a batch transform job and to associate the data that is relevant to the prediction results in the output. The input filter provided allows you to exclude input data that is not needed for inference in a batch transform job. The output filter provided allows you to include input data relevant to interpreting the predictions in the output from the job. For more information, see Associate Prediction Results with their Corresponding Input Records.

Type: DataProcessing (p. 1386) object

Required: No

Environment (p. 237)
The environment variables to set in the Docker container. We support up to 16 key and values entries in the map.

Type: String to string map

Map Entries: Maximum number of 16 items.

Key Length Constraints: Maximum length of 1024.

Key Pattern: [a-zA-Z_][a-zA-Z0-9_]{0,1023}

Value Length Constraints: Maximum length of 10240.

Value Pattern: [\S\s]*

Required: No

ExperimentConfig (p. 237)
Associates a SageMaker job as a trial component with an experiment and trial. Specified when you call the following APIs:

- CreateProcessingJob
- CreateTrainingJob
- CreateTransformJob

Type: ExperimentConfig (p. 1473) object

Required: No

MaxConcurrentTransforms (p. 237)
The maximum number of parallel requests that can be sent to each instance in a transform job. If MaxConcurrentTransforms is set to 0 or left unset, Amazon SageMaker checks the optional execution-parameters to determine the settings for your chosen algorithm. If the execution-parameters endpoint is not enabled, the default value is 1. For more information on execution-parameters, see How Containers Serve Requests. For built-in algorithms, you don't need to set a value for MaxConcurrentTransforms.

Type: Integer

Valid Range: Minimum value of 0.

Required: No

MaxPayloadInMB (p. 237)
The maximum allowed size of the payload, in MB. A payload is the data portion of a record (without metadata). The value in MaxPayloadInMB must be greater than, or equal to, the size of a single
record. To estimate the size of a record in MB, divide the size of your dataset by the number of records. To ensure that the records fit within the maximum payload size, we recommend using a slightly larger value. The default value is 6 MB.

The value of MaxPayloadInMB cannot be greater than 100 MB. If you specify the MaxConcurrentTransforms parameter, the value of (MaxConcurrentTransforms * MaxPayloadInMB) also cannot exceed 100 MB.

For cases where the payload might be arbitrarily large and is transmitted using HTTP chunked encoding, set the value to 0. This feature works only in supported algorithms. Currently, Amazon SageMaker built-in algorithms do not support HTTP chunked encoding.

Type: Integer
Valid Range: Minimum value of 0.
Required: No

ModelClientConfig (p. 237)
Configures the timeout and maximum number of retries for processing a transform job invocation.
Type: ModelClientConfig (p. 1675) object
Required: No

ModelName (p. 237)
The name of the model that you want to use for the transform job. ModelName must be the name of an existing Amazon SageMaker model within an AWS Region in an AWS account.
Type: String
Length Constraints: Maximum length of 63.
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])*  
Required: Yes

Tags (p. 237)
(Optional) An array of key-value pairs. For more information, see Using Cost Allocation Tags in the AWS Billing and Cost Management User Guide.
Type: Array of Tag (p. 1979) objects
Array Members: Minimum number of 0 items. Maximum number of 50 items.
Required: No

TransformInput (p. 237)
Describes the input source and the way the transform job consumes it.
Type: TransformInput (p. 2020) object
Required: Yes

TransformJobName (p. 237)
The name of the transform job. The name must be unique within an AWS Region in an AWS account.
Type: String
Pattern: ^[a-zA-Z0-9][-][a-zA-Z0-9]{0,62}$

Required: Yes

**TransformOutput (p. 237)**

Describes the results of the transform job.

Type: TransformOutput (p. 2032) object

Required: Yes

**TransformResources (p. 237)**

Describes the resources, including ML instance types and ML instance count, to use for the transform job.

Type: TransformResources (p. 2034) object

Required: Yes

**Response Syntax**

```json
{
   "TransformJobArn": "string"
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**TransformJobArn (p. 241)**

The Amazon Resource Name (ARN) of the transform job.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-zA-Z0-9\-]*:[0-9]{12}:transform-job/.*

**Errors**

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceInUse**

Resource being accessed is in use.

HTTP Status Code: 400

**ResourceLimitExceeded**

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400
ResourceNotFound

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
CreateTrial
Service: Amazon SageMaker Service

Creates an SageMaker trial. A trial is a set of steps called trial components that produce a machine learning model. A trial is part of a single SageMaker experiment.

When you use SageMaker Studio or the SageMaker Python SDK, all experiments, trials, and trial components are automatically tracked, logged, and indexed. When you use the AWS SDK for Python (Boto), you must use the logging APIs provided by the SDK.

You can add tags to a trial and then use the Search API to search for the tags.

To get a list of all your trials, call the ListTrials API. To view a trial's properties, call the DescribeTrial API. To create a trial component, call the CreateTrialComponent API.

Request Syntax

```json
{
    "DisplayName": "string",
    "ExperimentName": "string",
    "MetadataProperties": {
        "CommitId": "string",
        "GeneratedBy": "string",
        "ProjectId": "string",
        "Repository": "string"
    },
    "Tags": [
        {
            "Key": "string",
            "Value": "string"
        }
    ],
    "TrialName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

DisplayName (p. 243)
The name of the trial as displayed. The name doesn't need to be unique. If DisplayName isn't specified,TrialName is displayed.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: ^[a-zA-Z0-9](\-*[a-zA-Z0-9])\{0,119}$

Required: No

ExperimentName (p. 243)

The name of the experiment to associate the trial with.

Type: String
Length Constraints: Minimum length of 1. Maximum length of 120.
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,119}$
Required: Yes

**MetadataProperties (p. 243)**
Metadata properties of the tracking entity, trial, or trial component.
Type: MetadataProperties (p. 1648) object
Required: No

**Tags (p. 243)**
A list of tags to associate with the trial. You can use Search API to search on the tags.
Type: Array of Tag (p. 1979) objects
Array Members: Minimum number of 0 items. Maximum number of 50 items.
Required: No

**TrialName (p. 243)**
The name of the trial. The name must be unique in your AWS account and is not case-sensitive.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 120.
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,119}$
Required: Yes

**Response Syntax**

```json
{
   "TrialArn": "string"
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.
The following data is returned in JSON format by the service.

**TrialArn (p. 244)**
The Amazon Resource Name (ARN) of the trial.
Type: String
Length Constraints: Maximum length of 256.
Pattern: arn:aws[a-z\-]*:sagemaker:[a-zA-Z0-9\-]*:[0-9]{12}:experiment-trial/.*

**Errors**

For information about the errors that are common to all actions, see Common Errors (p. 2180).
ResourceLimitExceeded

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

ResourceNotFound

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
CreateTrialComponent

Service: Amazon SageMaker Service

Creates a trial component, which is a stage of a machine learning trial. A trial is composed of one or more trial components. A trial component can be used in multiple trials.

Trial components include pre-processing jobs, training jobs, and batch transform jobs.

When you use SageMaker Studio or the SageMaker Python SDK, all experiments, trials, and trial components are automatically tracked, logged, and indexed. When you use the AWS SDK for Python (Boto), you must use the logging APIs provided by the SDK.

You can add tags to a trial component and then use the Search API to search for the tags.

Request Syntax

```json
{
    "DisplayName": "string",
    "EndTime": number,
    "InputArtifacts": {
        "string": {
            "MediaType": "string",
            "Value": "string"
        }
    },
    "MetadataProperties": {
        "CommitId": "string",
        "GeneratedBy": "string",
        "ProjectId": "string",
        "Repository": "string"
    },
    "OutputArtifacts": {
        "string": {
            "MediaType": "string",
            "Value": "string"
        }
    },
    "Parameters": {
        "string": {
            "NumberValue": number,
            "StringValue": "string"
        }
    },
    "StartTime": number,
    "Status": {
        "Message": "string",
        "PrimaryStatus": "string"
    },
    "Tags": [
        {
            "Key": "string",
            "Value": "string"
        }
    ],
    "TrialComponentName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).
The request accepts the following data in JSON format.

**DisplayName (p. 246)**

The name of the component as displayed. The name doesn't need to be unique. If DisplayName isn't specified, TrialComponentName is displayed.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,119}

Required: No

**EndTime (p. 246)**

When the component ended.

Type: Timestamp

Required: No

**InputArtifacts (p. 246)**

The input artifacts for the component. Examples of input artifacts are datasets, algorithms, hyperparameters, source code, and instance types.

Type: String to TrialComponentArtifact (p. 2045) object map

Map Entries: Maximum number of 30 items.

Key Length Constraints: Maximum length of 64.

Key Pattern: .*

Required: No

**MetadataProperties (p. 246)**

Metadata properties of the tracking entity, trial, or trial component.

Type: MetadataProperties (p. 1648) object

Required: No

**OutputArtifacts (p. 246)**

The output artifacts for the component. Examples of output artifacts are metrics, snapshots, logs, and images.

Type: String to TrialComponentArtifact (p. 2045) object map

Map Entries: Maximum number of 30 items.

Key Length Constraints: Maximum length of 64.

Key Pattern: .*

Required: No

**Parameters (p. 246)**

The hyperparameters for the component.

Type: String to TrialComponentParameterValue (p. 2048) object map
Map Entries: Maximum number of 150 items.
Key Length Constraints: Maximum length of 256.
Key Pattern: .* 
Required: No

StartTime (p. 246)
When the component started.
Type: Timestamp
Required: No

Status (p. 246)
The status of the component. States include:
• InProgress
• Completed
• Failed
Type: TrialComponentStatus (p. 2053) object
Required: No

Tags (p. 246)
A list of tags to associate with the component. You can use Search API to search on the tags.
Type: Array of Tag (p. 1979) objects
Array Members: Minimum number of 0 items. Maximum number of 50 items.
Required: No

TrialComponentName (p. 246)
The name of the component. The name must be unique in your AWS account and is not case-sensitive.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 120.
Pattern: ^[a-zA-Z0-9]-*[a-zA-Z0-9]{0,119}
Required: Yes

Response Syntax

```json
{
   "TrialComponentArn": "string"
}
```

Response Elements
If the action is successful, the service sends back an HTTP 200 response.
The following data is returned in JSON format by the service.
TrialComponentArn (p. 248)

The Amazon Resource Name (ARN) of the trial component.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:experiment-trial-component/.*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceLimitExceeded

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
CreateUserProfile
Service: Amazon SageMaker Service

Creates a user profile. A user profile represents a single user within a domain, and is the main way to reference a "person" for the purposes of sharing, reporting, and other user-oriented features. This entity is created when a user onboards to a domain. If an administrator invites a person by email or imports them from IAM Identity Center, a user profile is automatically created. A user profile is the primary holder of settings for an individual user and has a reference to the user's private Amazon Elastic File System (EFS) home directory.

Request Syntax

```
{
  "DomainId": "string",
  "SingleSignOnUserIdentifier": "string",
  "SingleSignOnUserValue": "string",
  "Tags": [
    {
      "Key": "string",
      "Value": "string"
    }
  ],
  "UserProfileName": "string",
  "UserSettings": {
    "CanvasAppSettings": {
      "DirectDeploySettings": {
        "Status": "string"
      },
      "IdentityProviderOAuthSettings": [
        {
          "DataSourceName": "string",
          "SecretArn": "string",
          "Status": "string"
        }
      ],
      "KendraSettings": {
        "Status": "string"
      },
      "ModelRegisterSettings": {
        "CrossAccountModelRegisterRoleArn": "string",
        "Status": "string"
      },
      "TimeSeriesForecastingSettings": {
        "AmazonForecastRoleArn": "string",
        "Status": "string"
      },
      "WorkspaceSettings": {
        "S3ArtifactPath": "string",
        "S3KmsKeyId": "string"
      }
    },
    "CodeEditorAppSettings": {
      "DefaultResourceSpec": {
        "InstanceType": "string",
        "LifecycleConfigArn": "string",
        "SageMakerImageArn": "string",
        "SageMakerImageVersionAlias": "string",
        "SageMakerImageVersionArn": "string"
      },
      "LifecycleConfigArns": [ "string" ]
    },
    "CustomFileSystemConfigs": [
      
    ]
  }
}
```
{ ... }

"CustomPosixUserConfig": {
  "Gid": number,
  "Uid": number
},

"DefaultLandingUri": "string",
"ExecutionRole": "string",
"JupyterLabAppSettings": {
  "CodeRepositories": [
  
    
  
    
  },

"DefaultImages": [
  
  ]
},

"DefaultResourceSpec": {
  "InstanceType": "string",
  "LifecycleConfigArn": "string",
  "SageMakerImageArn": "string",
  "SageMakerImageVersionAlias": "string",
  "SageMakerImageVersionArn": "string"
},

"LifecycleConfigArns": [ "string" ]
},

"JupyterServerAppSettings": {
  "CodeRepositories": [
  
    
  ],

"DefaultResourceSpec": {
  "InstanceType": "string",
  "LifecycleConfigArn": "string",
  "SageMakerImageArn": "string",
  "SageMakerImageVersionAlias": "string",
  "SageMakerImageVersionArn": "string"
},

"LifecycleConfigArns": [ "string" ]
},

"KernelGatewayAppSettings": {
  "CustomImages": [
  
  ],

"DefaultResourceSpec": {
  "InstanceType": "string",
  "LifecycleConfigArn": "string",
  "SageMakerImageArn": "string",
  "SageMakerImageVersionAlias": "string",
  "SageMakerImageVersionArn": "string"
},

"LifecycleConfigArns": [ "string" ]
},

"RSessionAppSettings": {
  "CustomImages": [
  
  ]
}
Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**DomainId (p. 250)**

The ID of the associated Domain.

Type: String

Length Constraints: Maximum length of 63.

Required: Yes

**SingleSignOnUserIdentifier (p. 250)**

A specifier for the type of value specified in SingleSignOnUserValue. Currently, the only supported value is "UserName". If the Domain's AuthMode is IAM Identity Center, this field is required. If the Domain's AuthMode is not IAM Identity Center, this field cannot be specified.
CreateUserProfile

**Type**: String

**Pattern**: UserName

**Required**: No

**SingleSignOnUserValue (p. 250)**

The username of the associated AWS Single Sign-On User for this UserProfile. If the Domain's AuthMode is IAM Identity Center, this field is required, and must match a valid username of a user in your directory. If the Domain's AuthMode is not IAM Identity Center, this field cannot be specified.

**Type**: String

**Length Constraints**: Maximum length of 256.

**Required**: No

**Tags (p. 250)**

Each tag consists of a key and an optional value. Tag keys must be unique per resource.

Tags that you specify for the User Profile are also added to all Apps that the User Profile launches.

**Type**: Array of Tag (p. 1979) objects

**Array Members**: Minimum number of 0 items. Maximum number of 50 items.

**Required**: No

**UserProfileName (p. 250)**

A name for the UserProfile. This value is not case sensitive.

**Type**: String

**Length Constraints**: Maximum length of 63.

**Pattern**: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}\{0,62}\}

**Required**: Yes

**UserSettings (p. 250)**

A collection of settings.

**Type**: UserSettings (p. 2070) object

**Required**: No

**Response Syntax**

```
{
  "UserProfileArn": "string"
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.
UserProfileArn (p. 253)
The user profile Amazon Resource Name (ARN).
Type: String
Length Constraints: Maximum length of 256.
Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:user-profile/.*

Errors
For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceInUse
Resource being accessed is in use.
HTTP Status Code: 400

ResourceLimitExceeded
You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.
HTTP Status Code: 400

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
CreateWorkforce
Service: Amazon SageMaker Service

Use this operation to create a workforce. This operation will return an error if a workforce already exists in the AWS Region that you specify. You can only create one workforce in each AWS Region per AWS account.

If you want to create a new workforce in an AWS Region where a workforce already exists, use the `DeleteWorkforce` API operation to delete the existing workforce and then use `CreateWorkforce` to create a new workforce.

To create a private workforce using Amazon Cognito, you must specify a Cognito user pool in `CognitoConfig`. You can also create an Amazon Cognito workforce using the Amazon SageMaker console. For more information, see [Create a Private Workforce (Amazon Cognito)](#).

To create a private workforce using your own OIDC Identity Provider (IdP), specify your IdP configuration in `OidcConfig`. Your OIDC IdP must support groups because groups are used by Ground Truth and Amazon A2I to create work teams. For more information, see [Create a Private Workforce (OIDC IdP)](#).

**Request Syntax**

```json
{
    "CognitoConfig": {
        "ClientId": "string",
        "UserPool": "string"
    },
    "OidcConfig": {
        "AuthorizationEndpoint": "string",
        "ClientId": "string",
        "ClientSecret": "string",
        "Issuer": "string",
        "JwksUri": "string",
        "LogoutEndpoint": "string",
        "TokenEndpoint": "string",
        "UserInfoEndpoint": "string"
    },
    "SourceIpConfig": {
        "Cidrs": [ "string" ]
    },
    "Tags": [
        {
            "Key": "string",
            "Value": "string"
        }
    ],
    "WorkforceName": "string",
    "WorkforceVpcConfig": {
        "SecurityGroupIds": [ "string" ],
        "Subnets": [ "string" ],
        "VpcId": "string"
    }
}
```

**Request Parameters**

For information about the parameters that are common to all actions, see [Common Parameters (p. 2178)](#).

The request accepts the following data in JSON format.
**CognitoConfig (p. 255)**

Use this parameter to configure an Amazon Cognito private workforce. A single Cognito workforce is created using and corresponds to a single [Amazon Cognito user pool](https://docs.aws.amazon.com/sagemaker/latest/dg/API_CreateWorkforce.html).

Do not use OidcConfig if you specify values for CognitoConfig.

Type: **CognitoConfig (p. 1357)** object

Required: No

**OidcConfig (p. 255)**

Use this parameter to configure a private workforce using your own OIDC Identity Provider.

Do not use CognitoConfig if you specify values for OidcConfig.

Type: **OidcConfig (p. 1779)** object

Required: No

**SourceIpConfig (p. 255)**

A list of IP address ranges (CIDRs). Used to create an allow list of IP addresses for a private workforce. Workers will only be able to login to their worker portal from an IP address within this range. By default, a workforce isn't restricted to specific IP addresses.

Type: **SourceIpConfig (p. 1956)** object

Required: No

**Tags (p. 255)**

An array of key-value pairs that contain metadata to help you categorize and organize our workforce. Each tag consists of a key and a value, both of which you define.

Type: Array of **Tag (p. 1979)** objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

**WorkforceName (p. 255)**

The name of the private workforce.

Type: String


Pattern: `^[a-zA-Z0-9]([-a-zA-Z0-9\-]{0,62}$`

Required: Yes

**WorkforceVpcConfig (p. 255)**

Use this parameter to configure a workforce using VPC.

Type: **WorkforceVpcConfigRequest (p. 2081)** object

Required: No

---

**Response Syntax**

```json
{
}
```
"WorkforceArn": "string"
}

## Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

### WorkforceArn (p. 256)

The Amazon Resource Name (ARN) of the workforce.

- **Type:** String
- **Length Constraints:** Maximum length of 256.
- **Pattern:** `arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:workforce/*`

## Errors

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](#).

## See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
CreateWorkteam

Service: Amazon SageMaker Service

Creates a new work team for labeling your data. A work team is defined by one or more Amazon Cognito user pools. You must first create the user pools before you can create a work team.

You cannot create more than 25 work teams in an account and region.

Request Syntax

```json
{
    "Description": "string",
    "MemberDefinitions": [
        {
            "CognitoMemberDefinition": {
                "ClientId": "string",
                "UserGroup": "string",
                "UserPool": "string"
            },
            "OidcMemberDefinition": {
                "Groups": [ "string" ]
            }
        }
    ],
    "NotificationConfiguration": {
        "NotificationTopicArn": "string"
    },
    "Tags": [
        {
            "Key": "string",
            "Value": "string"
        }
    ],
    "WorkforceName": "string",
    "WorkteamName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**Description (p. 258)**

A description of the work team.

Type: String


Pattern: .+

Required: Yes

**MemberDefinitions (p. 258)**

A list of MemberDefinition objects that contains objects that identify the workers that make up the work team.
Workforces can be created using Amazon Cognito or your own OIDC Identity Provider (IdP). For private workforces created using Amazon Cognito use CognitoMemberDefinition. For workforces created using your own OIDC identity provider (IdP) use OidcMemberDefinition. Do not provide input for both of these parameters in a single request.

For workforces created using Amazon Cognito, private work teams correspond to Amazon Cognito user groups within the user pool used to create a workforce. All of the CognitoMemberDefinition objects that make up the member definition must have the same ClientId and UserPool values. To add a Amazon Cognito user group to an existing worker pool, see Adding groups to a User Pool. For more information about user pools, see Amazon Cognito User Pools.

For workforces created using your own OIDC IdP, specify the user groups that you want to include in your private work team in OidcMemberDefinition by listing those groups in Groups.

Type: Array of MemberDefinition (p. 1647) objects

Array Members: Minimum number of 1 item. Maximum number of 10 items.

Required: Yes

**NotificationConfiguration (p. 258)**

Configures notification of workers regarding available or expiring work items.

Type: NotificationConfiguration (p. 1775) object

Required: No

**Tags (p. 258)**

An array of key-value pairs.

For more information, see Resource Tag and Using Cost Allocation Tags in the AWS Billing and Cost Management User Guide.

Type: Array of Tag (p. 1979) objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

**WorkforceName (p. 258)**

The name of the workforce.

Type: String


Pattern: ^[a-zA-Z0-9-]{0,63}$

Required: No

**WorkteamName (p. 258)**

The name of the work team. Use this name to identify the work team.

Type: String


Pattern: ^[a-zA-Z0-9-]*[a-zA-Z0-9-]{0,62}$

Required: Yes
Response Syntax

```
{
  "WorkteamArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**WorkteamArn** *(p. 260)*

The Amazon Resource Name (ARN) of the work team. You can use this ARN to identify the work team.

Type: String

Length Constraints: Maximum length of 256.

Pattern: `arn:aws[a-z-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:workteam/*`

Errors

For information about the errors that are common to all actions, see [Common Errors](p. 2180).

**ResourceInUse**

Resource being accessed is in use.

HTTP Status Code: 400

**ResourceLimitExceeded**

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
DeleteAction
Service: Amazon SageMaker Service

Deletes an action.

Request Syntax

```json
{
   "ActionName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**ActionName (p. 261)**

The name of the action to delete.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,119}$

Required: Yes

Response Syntax

```json
{
   "ActionArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**ActionArn (p. 261)**

The Amazon Resource Name (ARN) of the action.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-zA-Z0-9\-]*:[0-9]{12}:action/.*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).
ResourceNotFound

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DeleteAlgorithm
Service: Amazon SageMaker Service
Removes the specified algorithm from your account.

Request Syntax

```json
{
    "AlgorithmName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**AlgorithmName (p. 263)**

The name of the algorithm to delete.

Type: String


Pattern: ^[a-zA-Z0-9][-*[a-zA-Z0-9]{0,62}$

Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DeleteApp

Service: Amazon SageMaker Service

Used to stop and delete an app.

Request Syntax

```json
{
  "AppName": "string",
  "AppType": "string",
  "DomainId": "string",
  "SpaceName": "string",
  "UserProfileName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**AppName (p. 264)**

The name of the app.

Type: String

Length Constraints: Maximum length of 63.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}`

Required: Yes

**AppType (p. 264)**

The type of app.

Type: String

Valid Values: JupyterServer | KernelGateway | TensorBoard | RStudioServerPro | RSessionGateway | JupyterLab | CodeEditor

Required: Yes

**DomainId (p. 264)**

The domain ID.

Type: String

Length Constraints: Maximum length of 63.

Required: Yes

**SpaceName (p. 264)**

The name of the space. If this value is not set, then UserProfileName must be set.

Type: String
Length Constraints: Maximum length of 63.
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$
Required: No

**UserProfileName** *(p. 264)*

The user profile name. If this value is not set, then SpaceName must be set.
Type: String
Length Constraints: Maximum length of 63.
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$
Required: No

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](p. 2180).

**ResourceInUse**

Resource being accessed is in use.

HTTP Status Code: 400

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DeleteAppImageConfig

Service: Amazon SageMaker Service

Deletes an AppImageConfig.

Request Syntax

```
{
    "AppImageConfigName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**AppImageConfigName (p. 266)**

The name of the AppImageConfig to delete.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
DeleteArtifact
Service: Amazon SageMaker Service

Deletes an artifact. Either ArtifactArn or Source must be specified.

Request Syntax

```
{
    "ArtifactArn": "string",
    "Source": {
        "SourceTypes": [
            {
                "SourceIdType": "string",
                "Value": "string"
            }
        ],
        "SourceUri": "string"
    }
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**ArtifactArn (p. 268)**

The Amazon Resource Name (ARN) of the artifact to delete.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:artifact/.*

Required: No

**Source (p. 268)**

The URI of the source.

Type: ArtifactSource (p. 1251) object

Required: No

Response Syntax

```
{
    "ArtifactArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.
The following data is returned in JSON format by the service.

**ArtifactArn (p. 268)**

The Amazon Resource Name (ARN) of the artifact.

- Type: String
- Length Constraints: Maximum length of 256.
- Pattern: `arn:aws[a-z-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:artifact/.*`

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](Link).

**ResourceNotFoundException**

- Resource being access is not found.
- HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](Link)
- [AWS SDK for .NET](Link)
- [AWS SDK for C++](Link)
- [AWS SDK for Go](Link)
- [AWS SDK for Java V2](Link)
- [AWS SDK for JavaScript V3](Link)
- [AWS SDK for PHP V3](Link)
- [AWS SDK for Python](Link)
- [AWS SDK for Ruby V3](Link)
DeleteAssociation
Service: Amazon SageMaker Service
Deletes an association.

Request Syntax

```
{
  "DestinationArn": "string",
  "SourceArn": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**DestinationArn (p. 270)**

The Amazon Resource Name (ARN) of the destination.

- Type: String
- Length Constraints: Maximum length of 256.
- Pattern: `arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:(experiment|experiment-trial-component|artifact|action|context)/.*`
- Required: Yes

**SourceArn (p. 270)**

The ARN of the source.

- Type: String
- Length Constraints: Maximum length of 256.
- Pattern: `arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:(experiment|experiment-trial-component|artifact|action|context)/.*`
- Required: Yes

Response Syntax

```
{
  "DestinationArn": "string",
  "SourceArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.
**DestinationArn (p. 270)**

The Amazon Resource Name (ARN) of the destination.

Type: String

Length Constraints: Maximum length of 256.

Pattern: `arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:(:experiment|experiment-trial-component|artifact|action|context)/.*`

**SourceArn (p. 270)**

The ARN of the source.

Type: String

Length Constraints: Maximum length of 256.

Pattern: `arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:(:experiment|experiment-trial-component|artifact|action|context)/.*`

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](#).

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
DeleteCluster
Service: Amazon SageMaker Service
Delete a SageMaker HyperPod cluster.

Request Syntax

```
{
    "ClusterName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**ClusterName (p. 272)**

The string name or the Amazon Resource Name (ARN) of the SageMaker HyperPod cluster to delete.

Type: String

Length Constraints: Maximum length of 256.

Pattern: ^arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:cluster/[a-z0-9]{12}([a-zA-Z0-9]-*[a-zA-Z0-9][{0,62}]$ Required: Yes

Response Syntax

```
{
    "ClusterArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**ClusterArn (p. 272)**

The Amazon Resource Name (ARN) of the SageMaker HyperPod cluster to delete.

Type: String

Length Constraints: Maximum length of 256.

Pattern: ^arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:cluster/[a-z0-9]{12}$
Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ConflictException

There was a conflict when you attempted to modify a SageMaker entity such as an Experiment or Artifact.

HTTP Status Code: 400

ResourceNotFoundException

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DeleteCodeRepository
Service: Amazon SageMaker Service

Deletes the specified Git repository from your account.

Request Syntax

```
{
   "CodeRepositoryName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CodeRepositoryName (p. 274)**

The name of the Git repository to delete.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}\$

Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DeleteContext
Service: Amazon SageMaker Service

Deletes an context.

Request Syntax

```
{
   "ContextName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**ContextName (p. 275)**

- The name of the context to delete.
- Type: String
- Length Constraints: Minimum length of 1. Maximum length of 120.
- Pattern: ^[a-zA-Z0-9]+(\-[a-zA-Z0-9]+)*\{0,119\}
- Required: Yes

Response Syntax

```
{
   "ContextArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**ContextArn (p. 275)**

- The Amazon Resource Name (ARN) of the context.
- Type: String
- Length Constraints: Maximum length of 256.
- Pattern: arn:aws[\a-zA-Z-]*:sagemaker:[a-zA-Z0-9-]*:[0-9]{12}:context/.*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).
ResourceNotFound

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
DeleteDataQualityJobDefinition
Service: Amazon SageMaker Service
Deletes a data quality monitoring job definition.

Request Syntax

```
{
   "JobDefinitionName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**JobDefinitionName (p. 277)**

The name of the data quality monitoring job definition to delete.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}\$

Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
DeleteDeviceFleet
Service: Amazon SageMaker

Deletes a fleet.

Request Syntax

```json
{
   "DeviceFleetName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**DeviceFleetName (p. 279)**

  The name of the fleet to delete.

  Type: String


  Pattern: `^[a-zA-Z0-9](\-*[a-zA-Z0-9])*\{0,62}$`

  Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceInUse**

  Resource being accessed is in use.

  HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DeleteDomain
Service: Amazon SageMaker Service

Used to delete a domain. If you onboarded with IAM mode, you will need to delete your domain to onboard again using IAM Identity Center. Use with caution. All of the members of the domain will lose access to their EFS volume, including data, notebooks, and other artifacts.

Request Syntax

```
{
    "DomainId": "string",
    "RetentionPolicy": {
        "HomeEfsFileSystem": "string"
    }
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**DomainId (p. 281)**

The domain ID.

Type: String

Length Constraints: Maximum length of 63.

Required: Yes

**RetentionPolicy (p. 281)**

The retention policy for this domain, which specifies whether resources will be retained after the Domain is deleted. By default, all resources are retained (not automatically deleted).

Type: RetentionPolicy (p. 1917) object

Required: No

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceInUse**

Resource being accessed is in use.

HTTP Status Code: 400

**ResourceNotFound**

Resource being access is not found.
HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DeleteEdgeDeploymentPlan
Service: Amazon SageMaker Service

Deletes an edge deployment plan if (and only if) all the stages in the plan are inactive or there are no stages in the plan.

Request Syntax

```json
{
   "EdgeDeploymentPlanName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**EdgeDeploymentPlanName (p. 283)**

The name of the edge deployment plan to delete.

Type: String


Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$`

Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceInUse

Resource being accessed is in use.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
DeleteEdgeDeploymentStage
Service: Amazon SageMaker Service
Delete a stage in an edge deployment plan if (and only if) the stage is inactive.

Request Syntax

```json
{
  "EdgeDeploymentPlanName": "string",
  "StageName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see [Common Parameters](p. 2178).

The request accepts the following data in JSON format.

**EdgeDeploymentPlanName (p. 285)**

The name of the edge deployment plan from which the stage will be deleted.

Type: String


Pattern: `^[a-zA-Z0-9][-]*[a-zA-Z0-9]{0,62}$`

Required: Yes

**StageName (p. 285)**

The name of the stage.

Type: String


Pattern: `^[a-zA-Z0-9][-]*[a-zA-Z0-9]{0,62}$`

Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see [Common Errors (p. 2180)].

**ResourceInUse**

Resource being accessed is in use.

HTTP Status Code: 400
See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
DeleteEndpoint
Service: Amazon SageMaker Service

Deletes an endpoint. SageMaker frees up all of the resources that were deployed when the endpoint was created.

SageMaker retires any custom KMS key grants associated with the endpoint, meaning you don't need to use the `RevokeGrant` API call.

When you delete your endpoint, SageMaker asynchronously deletes associated endpoint resources such as KMS key grants. You might still see these resources in your account for a few minutes after deleting your endpoint. Do not delete or revoke the permissions for your `ExecutionRoleArn`, otherwise SageMaker cannot delete these resources.

Request Syntax

```json
{
   "EndpointName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see [Common Parameters](p. 2178).

The request accepts the following data in JSON format.

**EndpointName (p. 287)**

The name of the endpoint that you want to delete.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-[a-zA-Z0-9]*){0,62}$

Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](p. 287)
- [AWS SDK for .NET](p. 287)
- [AWS SDK for C++](p. 287)
- [AWS SDK for Go](p. 287)
• AWS SDK for Java V2
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
DeleteEndpointConfig
Service: Amazon SageMaker Service

Deletes an endpoint configuration. The DeleteEndpointConfig API deletes only the specified configuration. It does not delete endpoints created using the configuration.

You must not delete an EndpointConfig in use by an endpoint that is live or while the UpdateEndpoint or CreateEndpoint operations are being performed on the endpoint. If you delete the EndpointConfig of an endpoint that is active or being created or updated you may lose visibility into the instance type the endpoint is using. The endpoint must be deleted in order to stop incurring charges.

Request Syntax

```json
{
    "EndpointConfigName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**EndpointConfigName (p. 289)**

The name of the endpoint configuration that you want to delete.

- Type: String
- Length Constraints: Maximum length of 63.
- Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])*\{0,62}$
- Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
DeleteExperiment

Service: Amazon SageMaker Service

Deletes an SageMaker experiment. All trials associated with the experiment must be deleted first. Use the ListTrials API to get a list of the trials associated with the experiment.

Request Syntax

```json
{
    "ExperimentName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

ExperimentName (p. 291)

The name of the experiment to delete.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: ^[a-zA-Z0-9\-]*[a-zA-Z0-9] {0,119}

Required: Yes

Response Syntax

```json
{
    "ExperimentArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

ExperimentArn (p. 291)

The Amazon Resource Name (ARN) of the experiment that is being deleted.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:experiment/.*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).
ResourceNotFound

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DeleteFeatureGroup

Service: Amazon SageMaker Service

Delete the FeatureGroup and any data that was written to the OnlineStore of the FeatureGroup. Data cannot be accessed from the OnlineStore immediately after DeleteFeatureGroup is called.

Data written into the OfflineStore will not be deleted. The AWS Glue database and tables that are automatically created for your OfflineStore are not deleted.

Note that it can take approximately 10-15 minutes to delete an OnlineStore FeatureGroup with the InMemory StorageType.

Request Syntax

```json
{
    "FeatureGroupName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**FeatureGroupName (p. 293)**

The name of the FeatureGroup you want to delete. The name must be unique within an AWS Region in an AWS account.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 64.

Pattern: ^[a-zA-Z0-9]([-]*[a-zA-Z0-9])\{0,63\}

Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:
- **AWS Command Line Interface**
- **AWS SDK for .NET**
- **AWS SDK for C++**
- **AWS SDK for Go**
- **AWS SDK for Java V2**
- **AWS SDK for JavaScript V3**
- **AWS SDK for PHP V3**
- **AWS SDK for Python**
- **AWS SDK for Ruby V3**
DeleteFlowDefinition
Service: Amazon SageMaker Service
Deletes the specified flow definition.

Request Syntax

```json
{
    "FlowDefinitionName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

FlowDefinitionName (p. 295)

The name of the flow definition you are deleting.

Type: String
Pattern: ^[a-z0-9](-*[a-z0-9])\{0,62
Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceInUse

Resource being accessed is in use.
HTTP Status Code: 400

ResourceNotFound

Resource being access is not found.
HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
• **AWS SDK for C++**
• **AWS SDK for Go**
• **AWS SDK for Java V2**
• **AWS SDK for JavaScript V3**
• **AWS SDK for PHP V3**
• **AWS SDK for Python**
• **AWS SDK for Ruby V3**
DeleteHub

Service: Amazon SageMaker Service

Delete a hub.

**Note**
- Hub APIs are only callable through SageMaker Studio.

**Request Syntax**

```json
{
   "HubName": "string"
}
```

**Request Parameters**

For information about the parameters that are common to all actions, see [Common Parameters](p. 2178).

The request accepts the following data in JSON format.

**HubName (p. 297)**

- The name of the hub to delete.
- Type: String
- Length Constraints: Maximum length of 63.
- Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}`
- Required: Yes

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)].

- **ResourceInUse**
  - Resource being accessed is in use.
  - HTTP Status Code: 400

- **ResourceNotFound**
  - Resource being access is not found.
  - HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:
- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DeleteHubContent

Service: Amazon SageMaker Service

Delete the contents of a hub.

**Note**
Hub APIs are only callable through SageMaker Studio.

**Request Syntax**

```
{
    "HubContentName": "string",
    "HubContentType": "string",
    "HubContentVersion": "string",
    "HubName": "string"
}
```

**Request Parameters**

For information about the parameters that are common to all actions, see [Common Parameters](#).

The request accepts the following data in JSON format.

**HubContentName** *(p. 299)*

- The name of the content that you want to delete from a hub.
  - Type: String
  - Length Constraints: Maximum length of 63.
  - Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}`
  - Required: Yes

**HubContentType** *(p. 299)*

- The type of content that you want to delete from a hub.
  - Type: String
  - Valid Values: Model | Notebook
  - Required: Yes

**HubContentVersion** *(p. 299)*

- The version of the content that you want to delete from a hub.
  - Type: String
  - Pattern: `^\d{1,4}\.\d{1,4}\.\d{1,4}$`
  - Required: Yes

**HubName** *(p. 299)*

- The name of the hub that you want to delete content in.
Type: String
Length Constraints: Maximum length of 63.
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}
Required: Yes

Response Elements
If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors
For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceInUse
Resource being accessed is in use.
HTTP Status Code: 400

ResourceNotFound
Resource being access is not found.
HTTP Status Code: 400

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DeleteHumanTaskUi

Service: Amazon SageMaker Service

Use this operation to delete a human task user interface (worker task template).

To see a list of human task user interfaces (work task templates) in your account, use ListHumanTaskUis. When you delete a worker task template, it no longer appears when you call ListHumanTaskUis.

Request Syntax

```json
{
    "HumanTaskUiName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**HumanTaskUiName** *(p. 301)*

The name of the human task user interface (work task template) you want to delete.

Type: String


Pattern: ^[a-z0-9](-*[a-z0-9])*

Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceNotFoundException**

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
• **AWS SDK for Go**
• **AWS SDK for Java V2**
• **AWS SDK for JavaScript V3**
• **AWS SDK for PHP V3**
• **AWS SDK for Python**
• **AWS SDK for Ruby V3**
DeleteImage
Service: Amazon SageMaker Service

Deletes a SageMaker image and all versions of the image. The container images aren't deleted.

Request Syntax

```
{
   "ImageName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**ImageName (p. 303)**

The name of the image to delete.

  Type: String
  Pattern: ^[a-zA-Z0-9][-\.][a-zA-Z0-9]{0,62}$
  Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceInUse**

  Resource being accessed is in use.
  HTTP Status Code: 400

**ResourceNotFound**

  Resource being access is not found.
  HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
• AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
DeleteImageVersion
Service: Amazon SageMaker Service

Deletes a version of a SageMaker image. The container image the version represents isn't deleted.

Request Syntax

```
{
  "Alias": "string",
  "ImageName": "string",
  "Version": number
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**Alias (p. 305)**

The alias of the image to delete.

Type: String


Pattern: (?![.-])^([a-zA-Z0-9-_.]+)$

Required: No

**ImageName (p. 305)**

The name of the image to delete.

Type: String


Pattern: ^[a-zA-Z0-9][-.][a-zA-Z0-9]{0,62}$

Required: Yes

**Version (p. 305)**

The version to delete.

Type: Integer

Valid Range: Minimum value of 0.

Required: No

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.
Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceInUse

Resource being accessed is in use.

HTTP Status Code: 400

ResourceNotFound

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DeleteInferenceComponent
Service: Amazon SageMaker Service
Deletes an inference component.

Request Syntax

```
{
   "InferenceComponentName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**InferenceComponentName (p. 307)**

The name of the inference component to delete.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](\-[a-zA-Z0-9]*[a-zA-Z0-9])?$

Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DeleteInferenceExperiment

Service: Amazon SageMaker Service

Deletes an inference experiment.

**Note**
This operation does not delete your endpoint, variants, or any underlying resources. This operation only deletes the metadata of your experiment.

**Request Syntax**

```json
{
   "Name": "string"
}
```

**Request Parameters**

For information about the parameters that are common to all actions, see [Common Parameters](p. 2178).

The request accepts the following data in JSON format.

**Name (p. 308)**

The name of the inference experiment you want to delete.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,119}$

Required: Yes

**Response Syntax**

```json
{
   "InferenceExperimentArn": "string"
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**InferenceExperimentArn (p. 308)**

The ARN of the deleted inference experiment.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:inference-experiment/.*
Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ConflictException

There was a conflict when you attempted to modify a SageMaker entity such as an Experiment or Artifact.

HTTP Status Code: 400

ResourceNotFoundException

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DeleteModel

Service: Amazon SageMaker Service

Deletes a model. The DeleteModel API deletes only the model entry that was created in SageMaker when you called the CreateModel API. It does not delete model artifacts, inference code, or the IAM role that you specified when creating the model.

Request Syntax

```json
{
   "ModelName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**ModelName (p. 310)**

The name of the model to delete.

Type: String

Length Constraints: Maximum length of 63.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9])*`

Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DeleteModelBiasJobDefinition

Service: Amazon SageMaker Service

Deletes an Amazon SageMaker model bias job definition.

Request Syntax

```json
{
   "JobDefinitionName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**JobDefinitionName (p. 312)**

The name of the model bias job definition to delete.

Type: String


Pattern: ^[a-zA-Z0-9\-]*[a-zA-Z0-9]\{0,62}$

Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
DeleteModelCard

Service: Amazon SageMaker Service

Deletes an Amazon SageMaker Model Card.

Request Syntax

```json
{
   "ModelCardName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see [Common Parameters](p. 2178).

The request accepts the following data in JSON format.

**ModelCardName (p. 314)**

The name of the model card to delete.

Type: String


Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$`

Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see [Common Errors](p. 2180).

**ConflictException**

There was a conflict when you attempted to modify a SageMaker entity such as an Experiment or Artifact.

HTTP Status Code: 400

**ResourceNotFoundException**

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
• AWS SDK for .NET
• AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
DeleteModelExplainabilityJobDefinition

Service: Amazon SageMaker Service

Deletes an Amazon SageMaker model explainability job definition.

Request Syntax

```
{
    "JobDefinitionName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**JobDefinitionName (p. 316)**

- The name of the model explainability job definition to delete.
- Type: String
- Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$`
- Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceNotFound**

- Resource being access is not found.
- HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
DeleteModelPackage
Service: Amazon SageMaker Service

Deletes a model package.

A model package is used to create SageMaker models or list on AWS Marketplace. Buyers can subscribe to model packages listed on AWS Marketplace to create models in SageMaker.

Request Syntax

```
{
   "ModelPackageName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**ModelPackageName (p. 318)**

The name or Amazon Resource Name (ARN) of the model package to delete.

When you specify a name, the name must have 1 to 63 characters. Valid characters are a-z, A-Z, 0-9, and - (hyphen).

Type: String


Pattern: (arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:[a-z\-]*\/)?(a-zA-Z0-9\[a-zA-Z0-9-\]\{0,62\}[^\-])\(\[/0-9\]{1,5}\)?$

Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ConflictException**

There was a conflict when you attempted to modify a SageMaker entity such as an Experiment or Artifact.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:
• AWS Command Line Interface
• AWS SDK for .NET
• AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
DeleteModelPackageGroup

Service: Amazon SageMaker Service

Deletes the specified model group.

Request Syntax

```
{
    "ModelPackageGroupName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**ModelPackageGroupName (p. 320)**

The name of the model group to delete.

Type: String


Pattern: (arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:[a-zA-Z0-9\-\d\-\(\\)]{0,62}(?!-))?

Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ConflictException**

There was a conflict when you attempted to modify a SageMaker entity such as an Experiment or Artifact.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
DeleteModelPackageGroup

- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DeleteModelPackageGroupPolicy

Service: Amazon SageMaker Service

Deletes a model group resource policy.

Request Syntax

```
{
   "ModelPackageGroupName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**ModelPackageGroupName (p. 322)**

The name of the model group for which to delete the policy.

Type: String


Pattern: ^[a-zA-Z0-9]{0,62}$

Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DeleteModelQualityJobDefinition
Service: Amazon SageMaker Service

Deletes the specified model quality monitoring job definition.

Request Syntax

```
{
  "JobDefinitionName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**JobDefinitionName (p. 323)**

The name of the model quality monitoring job definition to delete.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}\$

Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceNotFound**

Resource being accessed is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DeleteMonitoringSchedule

Service: Amazon SageMaker Service

Deletes a monitoring schedule. Also stops the schedule had not already been stopped. This does not delete the job execution history of the monitoring schedule.

Request Syntax

```
{
    "MonitoringScheduleName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**MonitoringScheduleName (p. 325)**

- The name of the monitoring schedule to delete.
  - Type: String
  - Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$
  - Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceNotFound**

- Resource being access is not found.
- HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DeleteNotebookInstance

Service: Amazon SageMaker Service

Deletes an SageMaker notebook instance. Before you can delete a notebook instance, you must call the StopNotebookInstance API.

Important
When you delete a notebook instance, you lose all of your data. SageMaker removes the ML compute instance, and deletes the ML storage volume and the network interface associated with the notebook instance.

Request Syntax

```json
{
    "NotebookInstanceName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**NotebookInstanceName (p. 327)**

The name of the SageMaker notebook instance to delete.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9][a-zA-Z0-9-]*[^a-zA-Z0-9]$*

Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
DeleteNotebookInstanceLifecycleConfig

Service: Amazon SageMaker Service

Deletes a notebook instance lifecycle configuration.

Request Syntax

```json
{
   "NotebookInstanceLifecycleConfigName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**NotebookInstanceLifecycleConfigName (p. 329)**

The name of the lifecycle configuration to delete.

Type: String

Length Constraints: Maximum length of 63.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]*)*`

Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DeletePipeline

Service: Amazon SageMaker Service

Deletes a pipeline if there are no running instances of the pipeline. To delete a pipeline, you must stop all running instances of the pipeline using the StopPipelineExecution API. When you delete a pipeline, all instances of the pipeline are deleted.

Request Syntax

```
{
    "ClientRequestToken": "string",
    "PipelineName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**ClientRequestToken (p. 330)**

A unique, case-sensitive identifier that you provide to ensure the idempotency of the operation. An idempotent operation completes no more than one time.

- Type: String
- Required: Yes

**PipelineName (p. 330)**

The name of the pipeline to delete.

- Type: String
- Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,255}\$
- Required: Yes

Response Syntax

```
{
    "PipelineArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.
**PipelineArn (p. 330)**

The Amazon Resource Name (ARN) of the pipeline to delete.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:pipeline/.*

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)].

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DeleteProject
Service: Amazon SageMaker Service
Delete the specified project.

Request Syntax

```
{
  "ProjectName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**ProjectName (p. 332)**

The name of the project to delete.

Type: String


Pattern: `^[a-zA-Z0-9\-]*[a-zA-Z0-9]{0,31}$`

Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ConflictException**

There was a conflict when you attempted to modify a SageMaker entity such as an Experiment or Artifact.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
DeleteSpace
Service: Amazon SageMaker Service
Used to delete a space.

Request Syntax

```json
{
    "DomainId": "string",
    "SpaceName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**DomainId (p. 334)**

The ID of the associated Domain.

Type: String

Length Constraints: Maximum length of 63.

Required: Yes

**SpaceName (p. 334)**

The name of the space.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}$

Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceInUse**

Resource being accessed is in use.

HTTP Status Code: 400

**ResourceNotFound**

Resource being access is not found.
HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DeleteStudioLifecycleConfig

Service: Amazon SageMaker Service

Deletes the Amazon SageMaker Studio Lifecycle Configuration. In order to delete the Lifecycle Configuration, there must be no running apps using the Lifecycle Configuration. You must also remove the Lifecycle Configuration from UserSettings in all Domains and UserProfiles.

Request Syntax

```
{
  "StudioLifecycleConfigName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**StudioLifecycleConfigName (p. 336)**

The name of the Amazon SageMaker Studio Lifecycle Configuration to delete.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9][-]\{0,62\}$

Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceInUse**

Resource being accessed is in use.

HTTP Status Code: 400

**ResourceNotFoundException**

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:
• AWS Command Line Interface
• AWS SDK for .NET
• AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
DeleteTags

Service: Amazon SageMaker Service

Deletes the specified tags from an SageMaker resource.

To list a resource's tags, use the ListTags API.

Note
When you call this API to delete tags from a hyperparameter tuning job, the deleted tags are not removed from training jobs that the hyperparameter tuning job launched before you called this API.

Note
When you call this API to delete tags from a SageMaker Domain or User Profile, the deleted tags are not removed from Apps that the SageMaker Domain or User Profile launched before you called this API.

Request Syntax

```
{
    "ResourceArn": "string",
    "TagKeys": [ "string" ]
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**ResourceArn (p. 338)**

The Amazon Resource Name (ARN) of the resource whose tags you want to delete.

Type: String

Length Constraints: Maximum length of 256.

Pattern: awsn:aws[a-z-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:+

Required: Yes

**TagKeys (p. 338)**

An array or one or more tag keys to delete.

Type: Array of strings

Array Members: Minimum number of 1 item. Maximum number of 50 items.


Pattern: ^([\p{L}\p{Z}\p{N}_.:/=+-@]*$)

Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.
Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DeleteTrial
Service: Amazon SageMaker Service

Deletes the specified trial. All trial components that make up the trial must be deleted first. Use the DescribeTrialComponent API to get the list of trial components.

Request Syntax

```json
{
   "TrialName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**TrialName (p. 340)**

The name of the trial to delete.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,119}

Required: Yes

Response Syntax

```json
{
   "TrialArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**TrialArn (p. 340)**

The Amazon Resource Name (ARN) of the trial that is being deleted.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:experiment-trial/.*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).
ResourceNotFound

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DeleteTrialComponent
Service: Amazon SageMaker Service

Deletes the specified trial component. A trial component must be disassociated from all trials before the trial component can be deleted. To disassociate a trial component from a trial, call the DisassociateTrialComponent API.

Request Syntax

```
{
   "TrialComponentName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**TrialComponentName (p. 342)**

The name of the component to delete.

- Type: String
- Length Constraints: Minimum length of 1. Maximum length of 120.
- Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,119}$`
- Required: Yes

Response Syntax

```
{
   "TrialComponentArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**TrialComponentArn (p. 342)**

The Amazon Resource Name (ARN) of the component is being deleted.

- Type: String
- Length Constraints: Maximum length of 256.
- Pattern: `arn:aws[a-z\-]*:sagemaker:[a-zA-Z0-9\-]*:[0-9]{12}:experiment-trial-component/.*`
Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceNotFound

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DeleteUserProfile
Service: Amazon SageMaker Service

Deletes a user profile. When a user profile is deleted, the user loses access to their EFS volume, including data, notebooks, and other artifacts.

Request Syntax

```
{
    "DomainId": "string",
    "UserProfileName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**DomainId (p. 344)**

The domain ID.
Type: String
Length Constraints: Maximum length of 63.
Required: Yes

**UserProfileName (p. 344)**

The user profile name.
Type: String
Length Constraints: Maximum length of 63.
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}$
Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceInUse**

Resource being accessed is in use.
HTTP Status Code: 400

**ResourceNotFound**

Resource being access is not found.
HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
DeleteWorkforce

Service: Amazon SageMaker Service

Use this operation to delete a workforce.

If you want to create a new workforce in an AWS Region where a workforce already exists, use this operation to delete the existing workforce and then use CreateWorkforce to create a new workforce.

**Important**

If a private workforce contains one or more work teams, you must use the DeleteWorkteam operation to delete all work teams before you delete the workforce. If you try to delete a workforce that contains one or more work teams, you will receive a ResourceInUse error.

**Request Syntax**

```json
{
  "WorkforceName": "string"
}
```

**Request Parameters**

For information about the parameters that are common to all actions, see [Common Parameters (p. 2178)](#).

The request accepts the following data in JSON format.

**WorkforceName (p. 346)**

The name of the workforce.

- Type: String
- Pattern: ^[a-zA-Z0-9](^[a-zA-Z0-9-\-]){0,62}$
- Required: Yes

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](#).

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
DeleteWorkteam

Service: Amazon SageMaker Service

Deletes an existing work team. This operation can't be undone.

Request Syntax

```
{
   "WorkteamName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**WorkteamName (p. 348)**

The name of the work team to delete.

- Type: String
- Pattern: `^[a-zA-Z0-9][-\[a-zA-Z0-9\]0,62]`
- Required: Yes

Response Syntax

```
{
   "Success": boolean
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**Success (p. 348)**

- Returns `true` if the work team was successfully deleted; otherwise, returns `false`.
- Type: Boolean

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceLimitExceeded**

- You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.
HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DeregisterDevices
Service: Amazon SageMaker Service

Deregisters the specified devices. After you deregister a device, you will need to re-register the devices.

Request Syntax

```
{
   "DeviceFleetName": "string",
   "DeviceNames": [ "string" ]
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**DeviceFleetName (p. 350)**

The name of the fleet the devices belong to.

Type: String


Pattern: ^[a-zA-Z0-9](\-*[a-zA-Z0-9]){0,62}$

Required: Yes

**DeviceNames (p. 350)**

The unique IDs of the devices.

Type: Array of strings


Pattern: ^[a-zA-Z0-9](\-*[a-zA-Z0-9]){0,62}$

Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
• AWS SDK for .NET
• AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
DescribeAction
Service: Amazon SageMaker Service
Describes an action.

Request Syntax
{
   "ActionName": "string"
}

Request Parameters
For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**ActionName (p. 352)**

The name of the action to describe.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,119}$

Required: Yes

Response Syntax
{
   "ActionArn": "string",
   "ActionName": "string",
   "ActionType": "string",
   "CreatedBy": {
      "DomainId": "string",
      "IamIdentity": {
         "Arn": "string",
         "PrincipalId": "string",
         "SourceIdentity": "string"
      },
      "UserProfileArn": "string",
      "UserProfileName": "string"
   },
   "CreationTime": number,
   "Description": "string",
   "LastModifiedBy": {
      "DomainId": "string",
      "IamIdentity": {
         "Arn": "string",
         "PrincipalId": "string",
         "SourceIdentity": "string"
      },
      "UserProfileArn": "string",
      "UserProfileName": "string"
   }
}
"LastModifiedTime": number,
"LineageGroupArn": "string",
"MetadataProperties": {
  "CommitId": "string",
  "GeneratedBy": "string",
  "ProjectId": "string",
  "Repository": "string"
},
"Properties": {
  "string": "string"
},
"Source": {
  "SourceId": "string",
  "SourceType": "string",
  "SourceUri": "string"
},
"Status": "string"

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**ActionArn (p. 352)**

The Amazon Resource Name (ARN) of the action.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:action/.*

**ActionName (p. 352)**

The name of the action.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 256.

Pattern: (arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:(experiment|experiment-trial|experiment-trial-component|artifact|action|context)\/)?(\[a-zA-Z0-9\-]{{0,119}})

**ActionType (p. 352)**

The type of the action.

Type: String

Length Constraints: Maximum length of 256.

**CreatedBy (p. 352)**

Information about the user who created or modified an experiment, trial, trial component, lineage group, project, or model card.

Type: **UserContext (p. 2067)** object

**CreationTime (p. 352)**

When the action was created.
Type: Timestamp

**Description** *(p. 352)*

The description of the action.

Type: String

Length Constraints: Maximum length of 3072.

Pattern: .*

**LastModifiedBy** *(p. 352)*

Information about the user who created or modified an experiment, trial, trial component, lineage group, project, or model card.

Type: **UserContext** *(p. 2067)* object

**LastModifiedTime** *(p. 352)*

When the action was last modified.

Type: Timestamp

**LineageGroupArn** *(p. 352)*

The Amazon Resource Name (ARN) of the lineage group.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:lineage-group/.*

**MetadataProperties** *(p. 352)*

Metadata properties of the tracking entity, trial, or trial component.

Type: **MetadataProperties** *(p. 1648)* object

**Properties** *(p. 352)*

A list of the action's properties.

Type: String to string map

Map Entries: Maximum number of 30 items.

Key Length Constraints: Maximum length of 256.

Key Pattern: .*

Value Length Constraints: Maximum length of 256.

Value Pattern: .*

**Source** *(p. 352)*

The source of the action.

Type: **ActionSource** *(p. 1216)* object

**Status** *(p. 352)*

The status of the action.

Type: String
Valid Values: Unknown | InProgress | Completed | Failed | Stopping | Stopped

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceNotFound

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
**DescribeAlgorithm**

Service: Amazon SageMaker Service

Returns a description of the specified algorithm that is in your account.

**Request Syntax**

```json
{
  "AlgorithmName": "string"
}
```

**Request Parameters**

For information about the parameters that are common to all actions, see [Common Parameters](p. 2178).

The request accepts the following data in JSON format.

**AlgorithmName (p. 356)**

The name of the algorithm to describe.

Type: String


Pattern: `arn:aws[a-z-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:[a-z-]*\/?([a-zA-Z0-9-]{0,62})([a-zA-Z0-9-]12):[0-9]\{0,62\}\(<\!\-\)$

Required: Yes

**Response Syntax**

```json
{
  "AlgorithmArn": "string",
  "AlgorithmDescription": "string",
  "AlgorithmName": "string",
  "AlgorithmStatus": "string",
  "AlgorithmStatusDetails": {
    "ImageScanStatuses": [
      {
        "FailureReason": "string",
        "Name": "string",
        "Status": "string"
      }
    ],
    "ValidationStatuses": [
      {
        "FailureReason": "string",
        "Name": "string",
        "Status": "string"
      }
    ]
  },
  "CertifyForMarketplace": boolean,
  "CreationTime": number,
  "InferenceSpecification": {
    "Containers": 
  }
}
```
"AdditionalS3DataSource": { 
  "CompressionType": "string",
  "S3DataType": "string",
  "S3Uri": "string"
},
"ContainerHostname": "string",
"Environment": {
  "string": "string"
},
"Framework": "string",
"FrameworkVersion": "string",
"Image": "string",
"ImageDigest": "string",
"ModelDataUrl": "string",
"ModelInput": {
  "DataInputConfig": "string"
},
"NearestModelName": "string",
"ProductId": "string"
],
"SupportedContentTypes": [ "string" ],
"SupportedRealtimeInferenceInstanceTypes": [ "string" ],
"SupportedResponseMIMETypes": [ "string" ],
"SupportedTransformInstanceTypes": [ "string" ]
},
"ProductId": "string",
"TrainingSpecification": {
  "AdditionalS3DataSource": {
    "CompressionType": "string",
    "S3DataType": "string",
    "S3Uri": "string"
  },
  "MetricDefinitions": [ 
    { 
      "Name": "string",
      "Regex": "string"
    }
  ],
  "SupportedHyperParameters": [ 
    { 
      "DefaultValue": "string",
      "Description": "string",
      "IsRequired": boolean,
      "IsTunable": boolean,
      "Name": "string",
      "Range": { 
        "CategoricalParameterRangeSpecification": { 
          "Values": [ "string" ]
        },
        "ContinuousParameterRangeSpecification": { 
          "MaxValue": "string",
          "MinValue": "string"
        },
        "IntegerParameterRangeSpecification": { 
          "MaxValue": "string",
          "MinValue": "string"
        }
      },
      "Type": "string"
    }
  ],
  "SupportedTrainingInstanceTypes": [ "string" ],
  "SupportedTuningJobObjectiveMetrics": { 
    "MetricName": "string",
    "MetricType": "string" 
  }
}
"Type": "string"
],
"SupportsDistributedTraining": boolean,
"TrainingChannels": [
{
"Description": "string",
"IsRequired": boolean,
"Name": "string",
"SupportedCompressionTypes": [ "string" ],
"SupportedContentTypes": [ "string" ],
"SupportedInputModes": [ "string" ]
}],
"TrainingImage": "string",
"TrainingImageDigest": "string"
},
"ValidationSpecification": {
"ValidationProfiles": [
{
"ProfileName": "string",
"TrainingJobDefinition": {
"HyperParameters": {
"string": "string"
},
"InputDataConfig": [
{
"ChannelName": "string",
"CompressionType": "string",
"ContentType": "string",
"DataSource": {
"FileSystemDataSource": {
"DirectoryPath": "string",
"FileSystemAccessMode": "string",
"FileSystemId": "string",
"FileSystemType": "string"
},
"S3DataSource": {
"AttributeName": [ "string" ],
"InstanceGroupName": [ "string" ],
"S3DataDistributionType": "string",
"S3DataType": "string",
"S3Uri": "string"
}
},
"InputMode": "string",
"RecordWrapperType": "string",
"ShuffleConfig": {
"Seed": number
}
}
]
},
"OutputDataConfig": {
"CompressionType": "string",
"KmsKeyId": "string",
"S3OutputPath": "string"
},
"ResourceConfig": {
"InstanceCount": number,
"InstanceGroups": [
{
"InstanceCount": number,
"InstanceGroupName": "string",
"InstanceType": "string"
}
]
DescribeAlgorithm

```
"InstanceType": "string",
"KeepAlivePeriodInSeconds": number,
"VolumeKmsKeyId": "string",
"VolumeSizeInGB": number
},
"StoppingCondition": {
  "MaxPendingTimeInSeconds": number,
  "MaxRuntimeInSeconds": number,
  "MaxWaitTimeInSeconds": number
},
"TrainingInputMode": "string"
},
"TransformJobDefinition": {
  "BatchStrategy": "string",
  "Environment": {
    "string": "string"
  },
  "MaxConcurrentTransforms": number,
  "MaxPayloadInMB": number,
  "TransformInput": {
    "CompressionType": "string",
    "ContentType": "string",
    "DataSource": {
      "S3DataSource": {
        "S3DataType": "string",
        "S3Uri": "string"
      }
    },
    "SplitType": "string"
  },
  "TransformOutput": {
    "Accept": "string",
    "AssembleWith": "string",
    "KmsKeyId": "string",
    "S3OutputPath": "string"
  },
  "TransformResources": {
    "InstanceCount": number,
    "InstanceType": "string",
    "VolumeKmsKeyId": "string"
  }
}
},
"ValidationRole": "string"
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**AlgorithmArn (p. 356)**

The Amazon Resource Name (ARN) of the algorithm.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: ^arn:aws(-cn|-us-gov)?:sagemaker:[a-zA-Z0-9\-]{9,16}:[0-9]\{12\}:algorithm/\[\S\]{1,2048}+$
**AlgorithmDescription (p. 356)**

A brief summary about the algorithm.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: \[\p{L}\p{M}\p{Z}\p{S}\p{N}\p{P}]*

**AlgorithmName (p. 356)**

The name of the algorithm being described.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}$

**AlgorithmStatus (p. 356)**

The current status of the algorithm.

Type: String

Valid Values: Pending | InProgress | Completed | Failed | Deleting

**AlgorithmStatusDetails (p. 356)**

Details about the current status of the algorithm.

Type: AlgorithmStatusDetails (p. 1228) object

**CertifyForMarketplace (p. 356)**

Whether the algorithm is certified to be listed in AWS Marketplace.

Type: Boolean

**CreationTime (p. 356)**

A timestamp specifying when the algorithm was created.

Type: Timestamp

**InferenceSpecification (p. 356)**

Details about inference jobs that the algorithm runs.

Type: InferenceSpecification (p. 1601) object

**ProductId (p. 356)**

The product identifier of the algorithm.

Type: String

Length Constraints: Maximum length of 256.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])*$

**TrainingSpecification (p. 356)**

Details about training jobs run by this algorithm.

Type: TrainingSpecification (p. 2017) object
**ValidationSpecification (p. 356)**

Details about configurations for one or more training jobs that SageMaker runs to test the algorithm.

Type: [AlgorithmValidationSpecification (p. 1233)](p. 1233) object

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](p. 2180).

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](awscli)
- [AWS SDK for .NET](awsdotnet)
- [AWS SDK for C++](awssdkcpp)
- [AWS SDK for Go](awssdkgo)
- [AWS SDK for Java V2](awssdkjava)
- [AWS SDK for JavaScript V3](awssdkjavascript)
- [AWS SDK for PHP V3](awssdkphpv3)
- [AWS SDK for Python](awssdkpython)
- [AWS SDK for Ruby V3](awssdkrubyv3)
DescribeApp
Service: Amazon SageMaker Service
Describes the app.

Request Syntax

```
{
  "AppName": "string",
  "AppType": "string",
  "DomainId": "string",
  "SpaceName": "string",
  "UserProfileName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**AppName (p. 362)**

The name of the app.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}

Required: Yes

**AppType (p. 362)**

The type of app.

Type: String

Valid Values: JupyterServer | KernelGateway

Required: Yes

**DomainId (p. 362)**

The domain ID.

Type: String

Length Constraints: Maximum length of 63.

Required: Yes

**SpaceName (p. 362)**

The name of the space.

Type: String

Length Constraints: Maximum length of 63.
DescribeApp

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62\}

Required: No

UserProfileName (p. 362)

The user profile name. If this value is not set, then SpaceName must be set.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62\}

Required: No

Response Syntax

```
{
  "AppArn": "string",
  "AppName": "string",
  "AppType": "string",
  "CreationTime": number,
  "DomainId": "string",
  "FailureReason": "string",
  "LastHealthCheckTimestamp": number,
  "LastUserActivityTimestamp": number,
  "ResourceSpec": {
    "InstanceType": "string",
    "LifecycleConfigArn": "string",
    "SageMakerImageArn": "string",
    "SageMakerImageVersionAlias": "string",
    "SageMakerImageVersionArn": "string"
  },
  "SpaceName": "string",
  "Status": "string",
  "UserProfileName": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

AppArn (p. 363)

The Amazon Resource Name (ARN) of the app.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-zA-Z\-]*:sagemaker:[a-zA-Z0-9\-]*:[0-9]{12}:app/.*

AppName (p. 363)

The name of the app.

Type: String

Length Constraints: Maximum length of 63.
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62$  

**AppType (p. 363)**  
The type of app.  
Type: String  
Valid Values: JupyterServer | KernelGateway

**CreationTime (p. 363)**  
The creation time.  
Type: Timestamp

**DomainId (p. 363)**  
The domain ID.  
Type: String  
Length Constraints: Maximum length of 63.

**FailureReason (p. 363)**  
The failure reason.  
Type: String  
Length Constraints: Maximum length of 1024.

**LastHealthCheckTimestamp (p. 363)**  
The timestamp of the last health check.  
Type: Timestamp

**LastUserActivityTimestamp (p. 363)**  
The timestamp of the last user's activity. LastUserActivityTimestamp is also updated when SageMaker performs health checks without user activity. As a result, this value is set to the same value as LastHealthCheckTimestamp.  
Type: Timestamp

**ResourceSpec (p. 363)**  
The instance type and the Amazon Resource Name (ARN) of the SageMaker image created on the instance.  
Type: ResourceSpec (p. 1915) object

**SpaceName (p. 363)**  
The name of the space. If this value is not set, then UserProfileName must be set.  
Type: String  
Length Constraints: Maximum length of 63.  
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62$  

**Status (p. 363)**  
The status.  
Type: String
Valid Values: Deleted | Deleting | Failed | InService | Pending

**UserProfileName (p. 363)**

The user profile name.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](p. 2180).

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](AWS Command Line Interface)
- [AWS SDK for .NET](AWS SDK for .NET)
- [AWS SDK for C++](AWS SDK for C++)
- [AWS SDK for Go](AWS SDK for Go)
- [AWS SDK for Java V2](AWS SDK for Java V2)
- [AWS SDK for JavaScript V3](AWS SDK for JavaScript V3)
- [AWS SDK for PHP V3](AWS SDK for PHP V3)
- [AWS SDK for Python](AWS SDK for Python)
- [AWS SDK for Ruby V3](AWS SDK for Ruby V3)
DescribeAppImageConfig
Service: Amazon SageMaker Service
Describes an AppImageConfig.

Request Syntax

```json
{
   "AppImageConfigName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**AppImageConfigName (p. 366)**

The name of the AppImageConfig to describe.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

Required: Yes

Response Syntax

```json
{
   "AppImageConfigArn": "string",
   "AppImageConfigName": "string",
   "CreationTime": number,
   "JupyterLabAppImageConfig": {
      "ContainerConfig": {
         "ContainerArguments": [ "string" ],
         "ContainerEntrypoint": [ "string" ],
         "ContainerEnvironmentVariables": {
            "string": "string"
         }
      }
   },
   "KernelGatewayImageConfig": {
      "FileSystemConfig": {
         "DefaultGid": number,
         "DefaultUid": number,
         "MountPath": "string"
      },
      "KernelSpecs": [
         {
            "DisplayName": "string",
            "Name": "string"
         }
      ]
   }
}
```

366
"LastModifiedTime": number

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**AppImageConfigArn (p. 366)**

The Amazon Resource Name (ARN) of the AppImageConfig.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-zA-Z]*:sagemaker:[a-zA-Z0-9-]*:[0-9]{12}:app-image-config/.*

**AppImageConfigName (p. 366)**

The name of the AppImageConfig.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}

**CreationTime (p. 366)**

When the AppImageConfig was created.

Type: Timestamp

**JupyterLabAppImageConfig (p. 366)**

The configuration of the JupyterLab app.

Type: JupyterLabAppImageConfig (p. 1615) object

**KernelGatewayImageConfig (p. 366)**

The configuration of a KernelGateway app.

Type: KernelGatewayImageConfig (p. 1621) object

**LastModifiedTime (p. 366)**

When the AppImageConfig was last modified.

Type: Timestamp

Errors

For information about the errors that are common to all actions, see [Common Errors (p. 2180)].

**ResourceNotFoundException**

Resource being access is not found.

HTTP Status Code: 400
See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DescribeArtifact
Service: Amazon SageMaker Service
Describes an artifact.

Request Syntax

```json
{
    "ArtifactArn": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

ArtifactArn (p. 369)

The Amazon Resource Name (ARN) of the artifact to describe.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:artifact/.*

Required: Yes

Response Syntax

```json
{
    "ArtifactArn": "string",
    "ArtifactName": "string",
    "ArtifactType": "string",
    "CreatedBy": {
        "DomainId": "string",
        "IamIdentity": {
            "Arn": "string",
            "PrincipalId": "string",
            "SourceIdentity": "string"
        },
        "UserProfileArn": "string",
        "UserProfileName": "string"
    },
    "CreationTime": number,
    "LastModifiedBy": {
        "DomainId": "string",
        "IamIdentity": {
            "Arn": "string",
            "PrincipalId": "string",
            "SourceIdentity": "string"
        },
        "UserProfileArn": "string",
        "UserProfileName": "string"
    },
    "LastModifiedTime": number,
}
```
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**ArtifactArn (p. 369)**

The Amazon Resource Name (ARN) of the artifact.

Type: String

Length Constraints: Maximum length of 256.

Pattern: `arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:artifact/.*`

**ArtifactName (p. 369)**

The name of the artifact.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 256.

Pattern: `(arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:(experiment|experiment-trial|experiment-trial-component|artifact|action|context)\/)?(\-[a-zA-Z0-9]{1,119})`

**ArtifactType (p. 369)**

The type of the artifact.

Type: String

Length Constraints: Maximum length of 256.

**CreatedBy (p. 369)**

Information about the user who created or modified an experiment, trial, trial component, lineage group, project, or model card.

Type: `UserContext (p. 2067)` object
**CreationTime (p. 369)**

When the artifact was created.

Type: Timestamp

**LastModifiedBy (p. 369)**

Information about the user who created or modified an experiment, trial, trial component, lineage group, project, or model card.

Type: `UserContext` object

**LastModifiedTime (p. 369)**

When the artifact was last modified.

Type: Timestamp

**LineageGroupArn (p. 369)**

The Amazon Resource Name (ARN) of the lineage group.

Type: String

Length Constraints: Maximum length of 256.

Pattern: `arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:lineage-group/.*`

**MetadataProperties (p. 369)**

Metadata properties of the tracking entity, trial, or trial component.

Type: `MetadataProperties` object

**Properties (p. 369)**

A list of the artifact's properties.

Type: String to string map

Map Entries: Maximum number of 30 items.

Key Length Constraints: Maximum length of 256.

Key Pattern: `.*`

Value Length Constraints: Maximum length of 256.

Value Pattern: `.*`

**Source (p. 369)**

The source of the artifact.

Type: `ArtifactSource` object

**Errors**

For information about the errors that are common to all actions, see `Common Errors (p. 2180)`.

**ResourceNotFound**

Resource being access is not found.
See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DescribeAutoMLJob

Service: Amazon SageMaker Service

Returns information about an AutoML job created by calling CreateAutoMLJob.

**Note**
AutoML jobs created by calling CreateAutoMLJobV2 cannot be described by DescribeAutoMLJob.

**Request Syntax**

```
{
   "AutoMLJobName": "string"
}
```

**Request Parameters**

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**AutoMLJobName (p. 373)**

Requests information about an AutoML job using its unique name.

- **Type:** String
- **Length Constraints:** Minimum length of 1. Maximum length of 32.
- **Pattern:** `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,31}`
- **Required:** Yes

**Response Syntax**

```
{
   "AutoMLJobArn": "string",
   "AutoMLJobArtifacts": {
      "CandidateDefinitionNotebookLocation": "string",
      "DataExplorationNotebookLocation": "string"
   },
   "AutoMLJobConfig": {
      "CandidateGenerationConfig": {
         "AlgorithmsConfig": [
            { "AutoMLAlgorithms": [ "string" ] }
         ],
         "FeatureSpecificationS3Uri": "string"
      },
      "CompletionCriteria": {
         "MaxAutoMLJobRuntimeInSeconds": number,
         "MaxCandidates": number,
         "MaxRuntimePerTrainingJobInSeconds": number
      },
      "DataSplitConfig": {
         "ValidationFraction": number
      }
   }
}
```
DescribeAutoMLJob

```
},
  "AutoMLJobName": "string",
  "AutoMLJobObjective": {
    "MetricName": "string"
  },
  "AutoMLJobSecondaryStatus": "string",
  "AutoMLJobStatus": "string",
  "BestCandidate": {
    "CandidateName": "string",
    "CandidateProperties": {
      "CandidateArtifactLocations": {
        "BacktestResults": "string",
        "Explainability": "string",
        "ModelInsights": "string"
      },
      "CandidateMetrics": [
        {
          "MetricName": "string",
          "Set": "string",
          "StandardMetricName": "string",
          "Value": number
        }
      ],
      "CandidateStatus": "string",
      "CandidateSteps": [
        {
          "CandidateStepArn": "string",
          "CandidateStepName": "string",
          "CandidateStepType": "string"
        }
      ],
      "CreationTime": number,
      "EndTime": number,
      "FailureReason": "string",
      "FinalAutoMLJobObjectiveMetric": {
        "MetricName": "string",
        "StandardMetricName": "string",
        "Type": "string",
        "Value": number
      },
      "InferenceContainerDefinitions": {
        "string": [
          {
            "Environment": {
              "string": "string"
            },
            "Image": "string",
            "ModelDataUrl": "string"
          }
        ],
        "InferenceContainers": [
          {
            "Environment": {
              "string": "string"
            }
          }
        ]
      }
    }
  }
```
DescribeAutoMLJob

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.
AutoMLJobArn (p. 373)

Returns the ARN of the AutoML job.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 256.

Pattern: arn:aws[a-zA-Z-]*:sagemaker:[a-zA-Z0-9-]*:[0-9]{12}:automl-job/.*

AutoMLJobArtifacts (p. 373)

Returns information on the job's artifacts found in AutoMLJobArtifacts.

Type: AutoMLJobArtifacts (p. 1277) object

AutoMLJobConfig (p. 373)

Returns the configuration for the AutoML job.

Type: AutoMLJobConfig (p. 1282) object

AutoMLJobName (p. 373)

Returns the name of the AutoML job.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,31}$

AutoMLJobObjective (p. 373)

Returns the job's objective.

Type: AutoMLJobObjective (p. 1284) object

AutoMLJobSecondaryStatus (p. 373)

Returns the secondary status of the AutoML job.

Type: String


AutoMLJobStatus (p. 373)

Returns the status of the AutoML job.

Type: String

Valid Values: Completed | InProgress | Failed | Stopped | Stopping

BestCandidate (p. 373)

The best model candidate selected by SageMaker Autopilot using both the best objective metric and lowest InferenceLatency for an experiment.

Type: AutoMLCandidate (p. 1265) object
CreationTime (p. 373)

Returns the creation time of the AutoML job.
Type: Timestamp

EndTime (p. 373)

Returns the end time of the AutoML job.
Type: Timestamp

FailureReason (p. 373)

Returns the failure reason for an AutoML job, when applicable.
Type: String
Length Constraints: Maximum length of 1024.

GenerateCandidateDefinitionsOnly (p. 373)

Indicates whether the output for an AutoML job generates candidate definitions only.
Type: Boolean

InputDataConfig (p. 373)

Returns the input data configuration for the AutoML job.
Type: Array of AutoMLChannel (p. 1271) objects
Array Members: Minimum number of 1 item. Maximum number of 2 items.

LastModifiedTime (p. 373)

Returns the job's last modified time.
Type: Timestamp

ModelDeployConfig (p. 373)

Indicates whether the model was deployed automatically to an endpoint and the name of that endpoint if deployed automatically.
Type: ModelDeployConfig (p. 1689) object

ModelDeployResult (p. 373)

Provides information about endpoint for the model deployment.
Type: ModelDeployResult (p. 1690) object

OutputDataConfig (p. 373)

Returns the job's output data config.
Type: AutoMLOutputDataConfig (p. 1289) object

PartialFailureReasons (p. 373)

Returns a list of reasons for partial failures within an AutoML job.
Type: Array of AutoMLPartialFailureReason (p. 1290) objects
Array Members: Minimum number of 1 item. Maximum number of 5 items.

ProblemType (p. 373)

Returns the job's problem type.
Type: String

Valid Values: BinaryClassification | MulticlassClassification | Regression

ResolvedAttributes (p. 373)

Contains ProblemType, AutoMLJobObjective, and CompletionCriteria. If you do not provide these values, they are inferred.

Type: ResolvedAttributes (p. 1908) object

RoleArn (p. 373)

The Amazon Resource Name (ARN) of the Identity and Access Management (IAM) role that has read permission to the input data location and write permission to the output data location in Amazon S3.

Type: String


Pattern: ^arn:aws[a-z\-]*:iam::\d{12}:role/?[a-zA-Z_0-9+=,.@\-_/]+$

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceNotFound

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DescribeAutoMLJobV2

Service: Amazon SageMaker Service

Returns information about an AutoML job created by calling `CreateAutoMLJobV2` or `CreateAutoMLJob`.

Request Syntax

```json
{
    "AutoMLJobName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see [Common Parameters](p. 2178).

The request accepts the following data in JSON format.

**AutoMLJobName** *(p. 379)*

Requests information about an AutoML job V2 using its unique name.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,31}$

Required: Yes

Response Syntax

```json
{
    "AutoMLJobArn": "string",
    "AutoMLJobArtifacts": {
        "CandidateDefinitionNotebookLocation": "string",
        "DataExplorationNotebookLocation": "string"
    },
    "AutoMLJobInputDataConfig": [
        {
            "ChannelType": "string",
            "CompressionType": "string",
            "ContentType": "string",
            "DataSource": {
                "S3DataSource": {
                    "S3DataType": "string",
                    "S3Uri": "string"
                }
            }
        }
    ],
    "AutoMLJobName": "string",
    "AutoMLJobObjective": {
        "MetricName": "string"
    },
    "AutoMLJobSecondaryStatus": "string",
    "AutoMLJobStatus": "string",
    "AutoMLProblemTypeConfig": { ... }
}
```
"AutoMLProblemTypeConfigName": "string",
"BestCandidate": {
  "CandidateName": "string",
  "CandidateProperties": {
    "CandidateArtifactLocations": {
      "BacktestResults": "string",
      "Explainability": "string",
      "ModelInsights": "string"
    },
    "CandidateMetrics": {
      "MetricName": "string",
      "Set": "string",
      "StandardMetricName": "string",
      "Value": number
    }
  }
},
"CandidateStatus": "string",
"CandidateSteps": [
  {
    "CandidateStepArn": "string",
    "CandidateStepName": "string",
    "CandidateStepType": "string"
  }
],
"CreationTime": number,
"EndTime": number,
"FailureReason": "string",
"FinalAutoMLJobObjectiveMetric": {
  "MetricName": "string",
  "StandardMetricName": "string",
  "Type": "string",
  "Value": number
},
"InferenceContainerDefinitions": {
  "string": {
    "Environment": {
      "string": "string"
    },
    "Image": "string",
    "ModelDataUrl": "string"
  }
},
"InferenceContainers": {
  "string": {
    "Environment": {"string": "string"},
    "Image": "string",
    "ModelDataUrl": "string"
  }
},
"LastModifiedTime": number,
"ObjectiveStatus": "string"
},
"CreationTime": number,
"DataSplitConfig": {
  "ValidationFraction": number
},
"EndTime": number,
"FailureReason": "string",
"LastModifiedTime": number,
"ModelDeployConfig": {
  "string": {
    "Environment": {
      "string": "string"
    },
    "Image": "string",
    "ModelDataUrl": "string"
  }
},
"LastModifiedTime": number,
"ObjectiveStatus": "string"}
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**AutoMLJobArn (p. 379)**

Returns the Amazon Resource Name (ARN) of the AutoML job V2.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 256.

Pattern: `arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:automl-job/.*`

**AutoMLJobArtifacts (p. 379)**

The artifacts that are generated during an AutoML job.

Type: `AutoMLJobArtifacts (p. 1277)` object

**AutoMLJobInputDataConfig (p. 379)**

Returns an array of channel objects describing the input data and their location.

Type: Array of `AutoMLJobChannel (p. 1278)` objects
Array Members: Minimum number of 1 item. Maximum number of 2 items.

**AutoMLJobName (p. 379)**

Returns the name of the AutoML job V2.

Type: String


Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,31\}

**AutoMLJobObjective (p. 379)**

Returns the job's objective.

Type: `AutoMLJobObjective (p. 1284)` object

**AutoMLJobSecondaryStatus (p. 379)**

Returns the secondary status of the AutoML job V2.

Type: String


**AutoMLJobStatus (p. 379)**

Returns the status of the AutoML job V2.

Type: String

Valid Values: Completed | InProgress | Failed | Stopped | Stopping

**AutoMLProblemTypeConfig (p. 379)**

Returns the configuration settings of the problem type set for the AutoML job V2.

Type: `AutoMLProblemTypeConfig (p. 1291)` object

**Note:** This object is a Union. Only one member of this object can be specified or returned.

**AutoMLProblemTypeConfigName (p. 379)**

Returns the name of the problem type configuration set for the AutoML job V2.

Type: String

Valid Values: ImageClassification | TextClassification | Tabular | TimeSeriesForecasting | TextGeneration

**BestCandidate (p. 379)**

Information about the candidate produced by an AutoML training job V2, including its status, steps, and other properties.

Type: `AutoMLCandidate (p. 1265)` object

**CreationTime (p. 379)**

Returns the creation time of the AutoML job V2.
DescribeAutoMLJobV2

DataSplitConfig (p. 379)

Returns the configuration settings of how the data are split into train and validation datasets.

Type: AutoMLDataSplitConfig (p. 1276) object

EndTime (p. 379)

Returns the end time of the AutoML job V2.

Type: Timestamp

FailureReason (p. 379)

Returns the reason for the failure of the AutoML job V2, when applicable.

Type: String

Length Constraints: Maximum length of 1024.

LastModifiedTime (p. 379)

Returns the job's last modified time.

Type: Timestamp

ModelDeployConfig (p. 379)

Indicates whether the model was deployed automatically to an endpoint and the name of that endpoint if deployed automatically.

Type: ModelDeployConfig (p. 1689) object

ModelDeployResult (p. 379)

Provides information about endpoint for the model deployment.

Type: ModelDeployResult (p. 1690) object

OutputDataConfig (p. 379)

Returns the job's output data config.

Type: AutoMLOutputDataConfig (p. 1289) object

PartialFailureReasons (p. 379)

Returns a list of reasons for partial failures within an AutoML job V2.

Type: Array of AutoMLPartialFailureReason (p. 1290) objects

Array Members: Minimum number of 1 item. Maximum number of 5 items.

ResolvedAttributes (p. 379)

Returns the resolved attributes used by the AutoML job V2.

Type: AutoMLResolvedAttributes (p. 1294) object

RoleArn (p. 379)

The ARN of the Identity and Access Management role that has read permission to the input data location and write permission to the output data location in Amazon S3.

Type: String

Pattern: ^arn:aws[a-z\-]*:iam::\d{12}:role/?[a-zA-Z_0-9+=,.@\-_/]++$

SecurityConfig (p. 379)

Returns the security configuration for traffic encryption or Amazon VPC settings.

Type: AutoMLSecurityConfig (p. 1297) object

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceNotFoundException

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DescribeCluster

Service: Amazon SageMaker Service

Retrieves information of a SageMaker HyperPod cluster.

Request Syntax

```json
{
    "ClusterName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**ClusterName (p. 385)**

The string name or the Amazon Resource Name (ARN) of the SageMaker HyperPod cluster.

- **Type:** String
- **Length Constraints:** Maximum length of 256.
- **Pattern:** `^(arn:aws-[a-z-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:cluster/[a-z0-9]{12})|([a-zA-Z0-9](-*[a-zA-Z0-9]{0,62})$`
- **Required:** Yes

Response Syntax

```json
{
    "ClusterArn": "string",
    "ClusterName": "string",
    "ClusterStatus": "string",
    "CreationTime": number,
    "FailureMessage": "string",
    "InstanceGroups": [
        {
            "CurrentCount": number,
            "ExecutionRole": "string",
            "InstanceGroupName": "string",
            "InstanceType": "string",
            "LifeCycleConfig": {
                "OnCreate": "string",
                "SourceS3Uri": "string"
            },
            "TargetCount": number,
            "ThreadsPerCore": number
        }
    ],
    "VpcConfig": {
        "SecurityGroupIds": [ "string" ],
        "Subnets": [ "string" ]
    }
}
```
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**ClusterArn (p. 385)**

The Amazon Resource Name (ARN) of the SageMaker HyperPod cluster.

Type: String

Length Constraints: Maximum length of 256.

Pattern: `^arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:cluster/[a-z0-9]{12}$`

**ClusterName (p. 385)**

The name of the SageMaker HyperPod cluster.

Type: String


Pattern: `^[a-zA-Z0-9\-]*[a-zA-Z0-9\-]+$`

**ClusterStatus (p. 385)**

The status of the SageMaker HyperPod cluster.

Type: String

Valid Values: Creating | Deleting | Failed | InService | RollingBack | SystemUpdating | Updating

**CreationTime (p. 385)**

The time when the SageMaker Cluster is created.

Type: Timestamp

**FailureMessage (p. 385)**

The failure message of the SageMaker HyperPod cluster.

Type: String

**InstanceGroups (p. 385)**

The instance groups of the SageMaker HyperPod cluster.

Type: Array of `ClusterInstanceGroupDetails (p. 1341)` objects

**VpcConfig (p. 385)**

Specifies an Amazon Virtual Private Cloud (VPC) that your SageMaker jobs, hosted models, and compute resources have access to. You can control access to and from your resources by configuring a VPC. For more information, see Give SageMaker Access to Resources in your Amazon VPC.

Type: `VpcConfig (p. 2076)` object

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).
ResourceNotFound

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DescribeClusterNode
Service: Amazon SageMaker Service

Retrieves information of an instance (also called a node interchangeably) of a SageMaker HyperPod cluster.

Request Syntax

```json
{
    "ClusterName": "string",
    "NodeId": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**ClusterName (p. 388)**

The string name or the Amazon Resource Name (ARN) of the SageMaker HyperPod cluster in which the instance is.

Type: String

Length Constraints: Maximum length of 256.

Pattern: `^arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:cluster/[a-z0-9]{12}([a-zA-Z-A-Z0-9\-]*[a-zA-Z0-9\-]*){0,62}$`

Required: Yes

**NodeId (p. 388)**

The ID of the instance.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 256.

Pattern: `^[a-zA-Z\-][a-zA-Z0-9\-]*$`

Required: Yes

Response Syntax

```json
{
    "NodeDetails": {
        "InstanceGroupName": "string",
        "InstanceId": "string",
        "InstanceStatus": {
            "Message": "string",
            "Status": "string"
        },
        "InstanceType": "string",
        "LaunchTime": number
    }
}
```
"LifeCycleConfig": {
    "OnCreate": "string",
    "SourceS3Uri": "string"
},
"ThreadsPerCore": number
}

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

NodeDetails (p. 388)

The details of the instance.

Type: ClusterNodeDetails (p. 1347) object

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceNotFound

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DescribeCodeRepository

Service: Amazon SageMaker Service

Gets details about the specified Git repository.

Request Syntax

```json
{
   "CodeRepositoryName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CodeRepositoryName (p. 390)**

The name of the Git repository to describe.

Type: String


Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}$`

Required: Yes

Response Syntax

```json
{
   "CodeRepositoryArn": "string",
   "CodeRepositoryName": "string",
   "CreationTime": number,
   "GitConfig": {
      "Branch": "string",
      "RepositoryUrl": "string",
      "SecretArn": "string"
   },
   "LastModifiedTime": number
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**CodeRepositoryArn (p. 390)**

The Amazon Resource Name (ARN) of the Git repository.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.
Pattern: ^arn:aws(-cn|-us-gov)?:sagemaker:[a-z0-9\-]{9,16}:[0-9]\{12\}:code-repository/[\S]{1,2048}$

**CodeRepositoryName (p. 390)**

The name of the Git repository.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

**CreationTime (p. 390)**

The date and time that the repository was created.

Type: Timestamp

**GitConfig (p. 390)**

Configuration details about the repository, including the URL where the repository is located, the default branch, and the Amazon Resource Name (ARN) of the AWS Secrets Manager secret that contains the credentials used to access the repository.

Type: [GitConfig (p. 1505)] object

**LastModifiedTime (p. 390)**

The date and time that the repository was last changed.

Type: Timestamp

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)].

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
DescribeCompilationJob

Service: Amazon SageMaker Service

Returns information about a model compilation job.

To create a model compilation job, use CreateCompilationJob. To get information about multiple model compilation jobs, use ListCompilationJobs.

Request Syntax

```json
{
  "CompilationJobName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CompilationJobName (p. 392)**

The name of the model compilation job that you want information about.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

Required: Yes

Response Syntax

```json
{
  "CompilationEndTime": number,
  "CompilationJobArn": "string",
  "CompilationJobName": "string",
  "CompilationJobStatus": "string",
  "CompilationStartTime": number,
  "CreationTime": number,
  "DerivedInformation": {
    "DerivedDataInputConfig": "string"
  },
  "FailureReason": "string",
  "InferenceImage": "string",
  "InputConfig": {
    "DataInputConfig": "string",
    "Framework": "string",
    "FrameworkVersion": "string",
    "S3Uri": "string"
  },
  "LastModifiedTime": number,
  "ModelArtifacts": {
    "S3ModelArtifacts": "string"
  },
  "ModelDigests": {
```
"ArtifactDigest": "string",
"ModelPackageVersionArn": "string",
"OutputConfig": {
  "CompilerOptions": "string",
  "KmsKeyId": "string",
  "S3OutputLocation": "string",
  "TargetDevice": "string",
  "TargetPlatform": {
    "Accelerator": "string",
    "Arch": "string",
    "Os": "string"
  }
},
"RoleArn": "string",
"StoppingCondition": {
  "MaxPendingTimeInSeconds": number,
  "MaxRuntimeInSeconds": number,
  "MaxWaitTimeInSeconds": number
},
"VpcConfig": {
  "SecurityGroupIds": [ "string" ],
  "Subnets": [ "string" ]
}
}

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**CompilationEndTime (p. 392)**

The time when the model compilation job on a compilation job instance ended. For a successful or stopped job, this is when the job's model artifacts have finished uploading. For a failed job, this is when Amazon SageMaker detected that the job failed.

Type: Timestamp

**CompilationJobArn (p. 392)**

The Amazon Resource Name (ARN) of the model compilation job.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:compilation-job/.*

**CompilationJobName (p. 392)**

The name of the model compilation job.

Type: String


Pattern: ^[a-zA-Z-A-Z0-9]*(\-[a-zA-Z-A-Z0-9])\{0,62}$

**CompilationJobStatus (p. 392)**

The status of the model compilation job.

Type: String
Valid Values: INPROGRESS | COMPLETED | FAILED | STARTING | STOPPING | STOPPED

CompilationStartTime (p. 392)

The time when the model compilation job started the CompilationJob instances.

You are billed for the time between this timestamp and the timestamp in the CompilationEndTime field. In Amazon CloudWatch Logs, the start time might be later than this time. That's because it takes time to download the compilation job, which depends on the size of the compilation job container.

Type: Timestamp

CreationTime (p. 392)

The time that the model compilation job was created.

Type: Timestamp

DerivedInformation (p. 392)

Information that SageMaker Neo automatically derived about the model.

Type: DerivedInformation (p. 1409) object

FailureReason (p. 392)

If a model compilation job failed, the reason it failed.

Type: String

Length Constraints: Maximum length of 1024.

InferenceImage (p. 392)

The inference image to use when compiling a model. Specify an image only if the target device is a cloud instance.

Type: String

Length Constraints: Maximum length of 256.

InputConfig (p. 392)

Information about the location in Amazon S3 of the input model artifacts, the name and shape of the expected data inputs, and the framework in which the model was trained.

Type: InputConfig (p. 1605) object

LastModifiedTime (p. 392)

The time that the status of the model compilation job was last modified.

Type: Timestamp

ModelArtifacts (p. 392)

Information about the location in Amazon S3 that has been configured for storing the model artifacts used in the compilation job.

Type: ModelArtifacts (p. 1659) object

ModelDigests (p. 392)

Provides a BLAKE2 hash value that identifies the compiled model artifacts in Amazon S3.

Type: ModelDigests (p. 1691) object
**ModelPackageVersionArn (p. 392)**

The Amazon Resource Name (ARN) of the versioned model package that was provided to SageMaker Neo when you initiated a compilation job.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: ^arn:aws(-cn|-us-gov)?:sagemaker:[a-z0-9\-]{9,16}:[0-9]{12}:model-package/\[\S\]{1,2048}$

**OutputConfig (p. 392)**

Information about the output location for the compiled model and the target device that the model runs on.

Type: OutputConfig (p. 1788) object

**RoleArn (p. 392)**

The Amazon Resource Name (ARN) of an IAM role that Amazon SageMaker assumes to perform the model compilation job.

Type: String


Pattern: ^arn:aws[a-z\-]*:iam::\d{12}:role/?[a-zA-Z_0-9+=,.@\-_/>]+$

**StoppingCondition (p. 392)**

Specifies a limit to how long a model compilation job can run. When the job reaches the time limit, Amazon SageMaker ends the compilation job. Use this API to cap model training costs.

Type: StoppingCondition (p. 1968) object

**VpcConfig (p. 392)**

A VpcConfig object that specifies the VPC that you want your compilation job to connect to. Control access to your models by configuring the VPC. For more information, see [Protect Compilation Jobs by Using an Amazon Virtual Private Cloud](#).

Type: NeoVpcConfig (p. 1767) object

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)].

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
• AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
DescribeContext
Service: Amazon SageMaker Service

Describes a context.

Request Syntax

```json
{
   "ContextName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**ContextName (p. 397)**

The name of the context to describe.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 256.

Pattern: (arn:aws[a-z-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:(experiment|experiment-trial|experiment-trial-component|artifact|action|context)\//)?([a-zA-Z0-9](-*[a-zA-Z0-9]){0,119})

Required: Yes

Response Syntax

```json
{
   "ContextArn": "string",
   "ContextName": "string",
   "ContextType": "string",
   "CreatedBy": {
      "DomainId": "string",
      "IamIdentity": {
         "Arn": "string",
         "PrincipalId": "string",
         "SourceIdentity": "string"
      },
      "UserProfileArn": "string",
      "UserProfileName": "string"
   },
   "CreationTime": number,
   "Description": "string",
   "LastModifiedBy": {
      "DomainId": "string",
      "IamIdentity": {
         "Arn": "string",
         "PrincipalId": "string",
         "SourceIdentity": "string"
      },
      "UserProfileArn": "string",
   }
}
```
"UserProfileName": "string",
"LastModifiedTime": number,
"LineageGroupArn": "string",
"Properties": {
  "string": "string"
},
"Source": {
  "SourceId": "string",
  "SourceType": "string",
  "SourceUri": "string"
}

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**ContextArn (p. 397)**

The Amazon Resource Name (ARN) of the context.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:context/.*

**ContextName (p. 397)**

The name of the context.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: ^[a-zA-Z0-9\-]{0,119}$

**ContextType (p. 397)**

The type of the context.

Type: String

Length Constraints: Maximum length of 256.

**CreatedBy (p. 397)**

Information about the user who created or modified an experiment, trial, trial component, lineage group, project, or model card.

Type: UserContext (p. 2067) object

**CreationTime (p. 397)**

When the context was created.

Type: Timestamp

**Description (p. 397)**

The description of the context.

Type: String
Length Constraints: Maximum length of 3072.

Pattern: .*

**LastModifiedBy (p. 397)**

Information about the user who created or modified an experiment, trial, trial component, lineage group, project, or model card.

Type: [UserContext (p. 2067)](p. 2067) object

**LastModifiedTime (p. 397)**

When the context was last modified.

Type: Timestamp

**LineageGroupArn (p. 397)**

The Amazon Resource Name (ARN) of the lineage group.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:lineage-group/.*

**Properties (p. 397)**

A list of the context's properties.

Type: String to string map

Map Entries: Maximum number of 30 items.

Key Length Constraints: Maximum length of 256.

Key Pattern: . *

Value Length Constraints: Maximum length of 256.

Value Pattern: . *

**Source (p. 397)**

The source of the context.

Type: [ContextSource (p. 1369)](p. 1369) object

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](p. 2180).

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:
- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DescribeDataQualityJobDefinition
Service: Amazon SageMaker Service

Gets the details of a data quality monitoring job definition.

Request Syntax

```json
{
    "JobDefinitionName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**JobDefinitionName (p. 401)**

The name of the data quality monitoring job definition to describe.

Type: String


Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}\$`

Required: Yes

Response Syntax

```json
{
    "CreationTime": number,
    "DataQualityAppSpecification": {
        "ContainerArguments": [ "string" ],
        "ContainerEntrypoint": [ "string" ],
        "Environment": {
            "string": "string"
        },
        "ImageUri": "string",
        "PostAnalyticsProcessorSourceUri": "string",
        "RecordPreprocessorSourceUri": "string"
    },
    "DataQualityBaselineConfig": {
        "BaseliningJobName": "string",
        "ConstraintsResource": {
            "S3Uri": "string"
        },
        "StatisticsResource": {
            "S3Uri": "string"
        }
    },
    "DataQualityJobInput": {
        "BatchTransformInput": {
            "DataCapturedDestinationS3Uri": "string",
            "DatasetFormat": {
                "Csv": {
                    "Header": boolean
                }
            }
        }
    }
}
```
"Json": {  "Line": boolean  
},  "Parquet": {}  
},  "EndTimeOffset": "string",  "ExcludeFeaturesAttribute": "string",  "FeaturesAttribute": "string",  "InferenceAttribute": "string",  "LocalPath": "string",  "ProbabilityAttribute": "string",  "ProbabilityThresholdAttribute": number,  "S3DataDistributionType": "string",  "S3InputMode": "string",  "StartTimeOffset": "string"  
},  "EndpointInput": {  "EndpointName": "string",  "EndTimeOffset": "string",  "ExcludeFeaturesAttribute": "string",  "FeaturesAttribute": "string",  "InferenceAttribute": "string",  "LocalPath": "string",  "ProbabilityAttribute": "string",  "ProbabilityThresholdAttribute": number,  "S3DataDistributionType": "string",  "S3InputMode": "string",  "StartTimeOffset": "string"  
}  
"DataQualityJobOutputConfig": {  "KmsKeyId": "string",  "MonitoringOutputs": [  {  "S3Output": {  "LocalPath": "string",  "S3UploadMode": "string",  "S3Uri": "string"  }  }  
]  
},  "JobDefinitionArn": "string",  "JobDefinitionName": "string",  "JobResources": {  "ClusterConfig": {  "InstanceCount": number,  "InstanceType": "string",  "VolumeKmsKeyId": "string",  "VolumeSizeInGB": number  
}  
},  "NetworkConfig": {  "EnableInterContainerTrafficEncryption": boolean,  "EnableNetworkIsolation": boolean,  "VpcConfig": {  "SecurityGroupIds": [ "string" ],  "Subnets": [ "string" ]  
}  
},  "RoleArn": "string",  "StoppingCondition": {  "MaxRuntimeInSeconds": number  
}  
}
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**CreationTime (p. 401)**

The time that the data quality monitoring job definition was created.

Type: Timestamp

**DataQualityAppSpecification (p. 401)**

Information about the container that runs the data quality monitoring job.

Type: DataQualityAppSpecification (p. 1388) object

**DataQualityBaselineConfig (p. 401)**

The constraints and baselines for the data quality monitoring job definition.

Type: DataQualityBaselineConfig (p. 1390) object

**DataQualityJobInput (p. 401)**

The list of inputs for the data quality monitoring job. Currently endpoints are supported.

Type: DataQualityJobInput (p. 1391) object

**DataQualityJobOutputConfig (p. 401)**

The output configuration for monitoring jobs.

Type: MonitoringOutputConfig (p. 1754) object

**JobDefinitionArn (p. 401)**

The Amazon Resource Name (ARN) of the data quality monitoring job definition.

Type: String

Length Constraints: Maximum length of 256.

Pattern: .*

**JobDefinitionName (p. 401)**

The name of the data quality monitoring job definition.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

**JobResources (p. 401)**

Identifies the resources to deploy for a monitoring job.

Type: MonitoringResources (p. 1756) object

**NetworkConfig (p. 401)**

The networking configuration for the data quality monitoring job.
Type: MonitoringNetworkConfig (p. 1752) object

RoleArn (p. 401)

The Amazon Resource Name (ARN) of an IAM role that Amazon SageMaker can assume to perform tasks on your behalf.

Type: String


Pattern: ^arn:aws[a-z\-]*:iam::\d{12}:role/?[a-zA-Z_0-9+=,.@\-_]+$  

StoppingCondition (p. 401)

A time limit for how long the monitoring job is allowed to run before stopping.

Type: MonitoringStoppingCondition (p. 1765) object

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceNotFound

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DescribeDevice
Service: Amazon SageMaker Service

Describes the device.

Request Syntax

```json
{
  "DeviceFleetName": "string",
  "DeviceName": "string",
  "NextToken": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**DeviceFleetName (p. 405)**

The name of the fleet the devices belong to.

Type: String


Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}$`

Required: Yes

**DeviceName (p. 405)**

The unique ID of the device.

Type: String


Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}$`

Required: Yes

**NextToken (p. 405)**

Next token of device description.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: `.*`

Required: No

Response Syntax

```json
{
}
```
"AgentVersion": "string",
"Description": "string",
"DeviceArn": "string",
"DeviceFleetName": "string",
"DeviceName": "string",
"IotThingName": "string",
"LatestHeartbeat": number,
"MaxModels": number,
"Models": [
  {
    "LatestInference": number,
    "LatestSampleTime": number,
    "ModelName": "string",
    "ModelVersion": "string"
  }
],
"NextToken": "string",
"RegistrationTime": number
}

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**AgentVersion (p. 405)**

Edge Manager agent version.

Type: String


Pattern: \*[a-zA-Z0-9\\_\-\.]*

**Description (p. 405)**

A description of the device.

Type: String


Pattern: ^[-a-zA-Z0-9_.;!:]*$  

**DeviceArn (p. 405)**

The Amazon Resource Name (ARN) of the device.

Type: String


Pattern: ^arn:aws[a-z\-]*:[a-z\-]*:[a-z\-]*:d{12}:\[a-z\-]*?\[a-zA-Z_0-9+=,.@\-_\]+$  

**DeviceFleetName (p. 405)**

The name of the fleet the device belongs to.

Type: String

Pattern: ^[a-zA-Z0-9]{0,62}\$  

**DeviceName (p. 405)**

The unique identifier of the device.

Type: String


Pattern: ^[a-zA-Z0-9]{0,62}\$  

**IoTThingName (p. 405)**

The AWS Internet of Things (IoT) object thing name associated with the device.

Type: String

Length Constraints: Maximum length of 128.

Pattern: [a-zA-Z0-9:_.-]+  

**LatestHeartbeat (p. 405)**

The last heartbeat received from the device.

Type: Timestamp

**MaxModels (p. 405)**

The maximum number of models.

Type: Integer

**Models (p. 405)**

Models on the device.

Type: Array of **EdgeModel (p. 1440)** objects

**NextToken (p. 405)**

The response from the last list when returning a list large enough to need tokening.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

**RegistrationTime (p. 405)**

The timestamp of the last registration or de-reregistration.

Type: Timestamp

**Errors**

For information about the errors that are common to all actions, see **Common Errors (p. 2180)**.

**ResourceNot found**

Resource being access is not found.

HTTP Status Code: 400
See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
DescribeDeviceFleet

Service: Amazon SageMaker Service

A description of the fleet the device belongs to.

Request Syntax

```
{
   "DeviceFleetName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**DeviceFleetName (p. 409)**

The name of the fleet.

Type: String


Pattern: ^[a-zA-Z0-9\(-*\[a-zA-Z0-9\])\{0,62}\$

Required: Yes

Response Syntax

```
{
   "CreationTime": number,
   "Description": "string",
   "DeviceFleetArn": "string",
   "DeviceFleetName": "string",
   "IotRoleAlias": "string",
   "LastModifiedTime": number,
   "OutputConfig": {
      "KmsKeyId": "string",
      "PresetDeploymentConfig": "string",
      "PresetDeploymentType": "string",
      "S3OutputLocation": "string"
   },
   "RoleArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**CreationTime (p. 409)**

Timestamp of when the device fleet was created.
Type: Timestamp

**Description (p. 409)**

A description of the fleet.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 800.

Pattern: ^[-a-zA-Z0-9_,;!: ]*$

**DeviceFleetArn (p. 409)**

The The Amazon Resource Name (ARN) of the fleet.

Type: String

Pattern: ^arn:aws[a-z\-]*:iam::\d{12}:device-fleet/\?[a-zA-Z0-9=,\@\-_/]++$

**DeviceFleetName (p. 409)**

The name of the fleet.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}$

**IoTRoleAlias (p. 409)**

The Amazon Resource Name (ARN) alias created in AWS Internet of Things (IoT).

Type: String

Pattern: ^arn:aws[a-z\-]*:iam::\d{12}:rolealias/\?[a-zA-Z0-9=,\@\-_/]++$

**LastModifiedTime (p. 409)**

Timestamp of when the device fleet was last updated.

Type: Timestamp

**OutputConfig (p. 409)**

The output configuration for storing sampled data.

Type: EdgeOutputConfig (p. 1444) object

**RoleArn (p. 409)**

The Amazon Resource Name (ARN) that has access to AWS Internet of Things (IoT).

Type: String


Pattern: ^arn:aws[a-z\-]*:iam::\d{12}:role/?[a-zA-Z0-9=,\@\-_/]++$

**Errors**

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceNotFound**

Resource being access is not found.
HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DescribeDomain
Service: Amazon SageMaker Service
The description of the domain.

Request Syntax

```
{
  "DomainId": "string"
}
```

Request Parameters
For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**DomainId (p. 412)**

The domain ID.

Type: String

Length Constraints: Maximum length of 63.

Required: Yes

Response Syntax

```
{
  "AppNetworkAccessType": "string",
  "AppSecurityGroupManagement": "string",
  "AuthMode": "string",
  "CreationTime": number,
  "DefaultSpaceSettings": {
    "ExecutionRole": "string",
    "JupyterServerAppSettings": {
      "CodeRepositories": [
        {
          "RepositoryUrl": "string"
        }
      ],
      "DefaultResourceSpec": {
        "InstanceType": "string",
        "LifecycleConfigArn": "string",
        "SageMakerImageArn": "string",
        "SageMakerImageVersionAlias": "string",
        "SageMakerImageVersionArn": "string"
      }
    },
    "LifecycleConfigArns": [ "string" ]
  },
  "KernelGatewayAppSettings": {
    "CustomImages": [
      {
        "AppImageConfigName": "string",
        "ImageName": "string",
      }
    ]
  }
}
```
"ImageVersionNumber": number
}
],
"DefaultResourceSpec": {
  "InstanceType": "string",
  "LifecycleConfigArn": "string",
  "SageMakerImageArn": "string",
  "SageMakerImageVersionAlias": "string",
  "SageMakerImageVersionArn": "string"
},
"LifecycleConfigArns": [ "string" ]
],
"SecurityGroups": [ "string" ]
],
"DefaultUserSettings": {
  "CanvasAppSettings": {
    "DirectDeploySettings": {
      "Status": "string"
    },
    "IdentityProviderOAuthSettings": [
      {"DataSourceName": "string", "SecretArn": "string", "Status": "string"}
    ],
    "KendraSettings": {
      "Status": "string"
    },
    "ModelRegisterSettings": {
      "CrossAccountModelRegisterRoleArn": "string",
      "Status": "string"
    },
    "TimeSeriesForecastingSettings": {
      "AmazonForecastRoleArn": "string",
      "Status": "string"
    },
    "WorkspaceSettings": {
      "S3ArtifactPath": "string",
      "S3KmsKeyId": "string"
    }
  },
  "CodeEditorAppSettings": {
    "DefaultResourceSpec": {
      "InstanceType": "string",
      "LifecycleConfigArn": "string",
      "SageMakerImageArn": "string",
      "SageMakerImageVersionAlias": "string",
      "SageMakerImageVersionArn": "string"
    },
    "LifecycleConfigArns": [ "string" ]
  }
},
"CustomFileSystemConfigs": [ ...
],
"CustomPosixUserConfig": {
  "Gid": number,
  "Uid": number
},
"DefaultLandingUri": "string",
"ExecutionRole": "string",
"JupyterLabAppSettings": {
  "CodeRepositories": [
    {"RepositoryUrl": "string"}
  ]
}
"CustomImages": [
    {
        "AppImageConfigName": "string",
        "ImageName": "string",
        "ImageVersionNumber": number
    }
],
"DefaultResourceSpec": {
    "InstanceType": "string",
    "LifecycleConfigArn": "string",
    "SageMakerImageArn": "string",
    "SageMakerImageVersionAlias": "string",
    "SageMakerImageVersionArn": "string"
},
"LifecycleConfigArns": [ "string" ]
},
"JupyterServerAppSettings": {
    "DefaultRepositorySpec": {
        "InstanceType": "string",
        "LifecycleConfigArn": "string",
        "SageMakerImageArn": "string",
        "SageMakerImageVersionAlias": "string",
        "SageMakerImageVersionArn": "string"
    },
    "LifecycleConfigArns": [ "string" ]
},
"KernelGatewayAppSettings": {
    "CustomImages": [
        {
            "AppImageConfigName": "string",
            "ImageName": "string",
            "ImageVersionNumber": number
        }
    ],
    "DefaultResourceSpec": {
        "InstanceType": "string",
        "LifecycleConfigArn": "string",
        "SageMakerImageArn": "string",
        "SageMakerImageVersionAlias": "string",
        "SageMakerImageVersionArn": "string"
    },
    "LifecycleConfigArns": [ "string" ]
},
"RSessionAppSettings": {
    "CustomImages": [
        {
            "AppImageConfigName": "string",
            "ImageName": "string",
            "ImageVersionNumber": number
        }
    ],
    "DefaultResourceSpec": {
        "InstanceType": "string",
        "LifecycleConfigArn": "string",
        "SageMakerImageArn": "string",
        "SageMakerImageVersionAlias": "string",
        "SageMakerImageVersionArn": "string"
    },
    "LifecycleConfigArns": [ "string" ]
},
"RStudioServerProAppSettings": {
    "CustomImages": [
        {
            "AppImageConfigName": "string",
            "ImageName": "string",
            "ImageVersionNumber": number
        }
    ],
    "DefaultResourceSpec": {
        "InstanceType": "string",
        "LifecycleConfigArn": "string",
        "SageMakerImageArn": "string",
        "SageMakerImageVersionAlias": "string",
        "SageMakerImageVersionArn": "string"
    }
}
If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.
**AppNetworkAccessType (p. 412)**

Specifies the VPC used for non-EFS traffic. The default value is PublicInternetOnly.

- **PublicInternetOnly** - Non-EFS traffic is through a VPC managed by Amazon SageMaker, which allows direct internet access.
- **VpcOnly** - All traffic is through the specified VPC and subnets.

Type: String

Valid Values: PublicInternetOnly | VpcOnly

**AppSecurityGroupManagement (p. 412)**

The entity that creates and manages the required security groups for inter-app communication in VPCOnly mode. Required when CreateDomain.AppNetworkAccessType is VPCOnly and DomainSettings.RStudioServerProDomainSettings.DomainExecutionRoleArn is provided.

Type: String

Valid Values: Service | Customer

**AuthMode (p. 412)**

The domain's authentication mode.

Type: String

Valid Values: SSO | IAM

**CreationTime (p. 412)**

The creation time.

Type: Timestamp

**DefaultSpaceSettings (p. 412)**

The default settings used to create a space.

Type: DefaultSpaceSettings (p. 1402) object

**DefaultUserSettings (p. 412)**

Settings which are applied to UserProfiles in this domain if settings are not explicitly specified in a given UserProfile.

Type: UserSettings (p. 2070) object

**DomainArn (p. 412)**

The domain's Amazon Resource Name (ARN).

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:domain/.*

**DomainId (p. 412)**

The domain ID.

Type: String

Length Constraints: Maximum length of 63.
**DomainName (p. 412)**

The domain name.

Type: String

Length Constraints: Maximum length of 63.

Pattern: `^[a-zA-Z0-9]([-][a-zA-Z0-9])\{0,62\}`

**DomainSettings (p. 412)**

A collection of Domain settings.

Type: `DomainSettings (p. 1424)` object

**FailureReason (p. 412)**

The failure reason.

Type: String

Length Constraints: Maximum length of 1024.

**HomeEfsFileSystemId (p. 412)**

The ID of the Amazon Elastic File System (EFS) managed by this Domain.

Type: String

Length Constraints: Maximum length of 32.

**HomeEfsFileSystemKmsKeyId (p. 412)**

*This parameter has been deprecated.*

Use KmsKeyId.

Type: String

Length Constraints: Maximum length of 2048.

Pattern: `.*`

**KmsKeyId (p. 412)**

The AWS KMS customer managed key used to encrypt the EFS volume attached to the domain.

Type: String

Length Constraints: Maximum length of 2048.

Pattern: `.*`

**LastModifiedTime (p. 412)**

The last modified time.

Type: Timestamp

**SecurityGroupIdForDomainBoundary (p. 412)**

The ID of the security group that authorizes traffic between the RSessionGateway apps and the RStudioServerPro app.

Type: String
Length Constraints: Maximum length of 32.

Pattern: [-0-9a-zA-Z]+  

**SingleSignOnApplicationArn (p. 412)**

The ARN of the application managed by SageMaker in IAM Identity Center. This value is only returned for domains created after October 1, 2023.

Type: String


**SingleSignOnManagedApplicationInstanceId (p. 412)**

The IAM Identity Center managed application instance ID.

Type: String

Length Constraints: Maximum length of 256.

**Status (p. 412)**

The status.

Type: String

Valid Values: Deleting | Failed | InService | Pending | Updating | Update_Failed | Delete_Failed

**SubnetIds (p. 412)**

The VPC subnets that the domain uses for communication.

Type: Array of strings

Array Members: Minimum number of 1 item. Maximum number of 16 items.

Length Constraints: Maximum length of 32.

Pattern: [-0-9a-zA-Z]+  

**Url (p. 412)**

The domain's URL.

Type: String

Length Constraints: Maximum length of 1024.

**VpcId (p. 412)**

The ID of the Amazon Virtual Private Cloud (VPC) that the domain uses for communication.

Type: String

Length Constraints: Maximum length of 32.

Pattern: [-0-9a-zA-Z]+  

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)].
ResourceNotFound

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DescribeEdgeDeploymentPlan
Service: Amazon SageMaker Service

Describes an edge deployment plan with deployment status per stage.

Request Syntax

```json
{
    "EdgeDeploymentPlanName": "string",
    "MaxResults": number,
    "NextToken": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see [Common Parameters](p. 2178).

The request accepts the following data in JSON format.

**EdgeDeploymentPlanName (p. 420)**

The name of the deployment plan to describe.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

Required: Yes

**MaxResults (p. 420)**

The maximum number of results to select (50 by default).

Type: Integer

Valid Range: Maximum value of 10.

Required: No

**NextToken (p. 420)**

If the edge deployment plan has enough stages to require tokening, then this is the response from the last list of stages returned.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Required: No

Response Syntax

```json
{
}
```
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**CreationTime (p. 420)**

The time when the edge deployment plan was created.

Type: Timestamp

**DeviceFleetName (p. 420)**

The device fleet used for this edge deployment plan.

Type: String


Pattern: ^[a-zA-Z0-9\-]*[a-zA-Z0-9]\{0,62\}$

**EdgeDeploymentFailed (p. 420)**

The number of edge devices that failed the deployment.
Type: Integer

**EdgeDeploymentPending (p. 420)**

The number of edge devices yet to pick up deployment, or in progress.

Type: Integer

**EdgeDeploymentPlanArn (p. 420)**

The ARN of edge deployment plan.

Type: String


Pattern: `^arn:aws[a-z-]*:sagemaker:[a-z-]*:d{12}:edge-deployment/?[a-zA-Z_0-9+=,.@-_\/]+$`

**EdgeDeploymentPlanName (p. 420)**

The name of the edge deployment plan.

Type: String


Pattern: `^[a-zA-Z0-9-]*[a-zA-Z0-9-]{0,62}$`

**EdgeDeploymentSuccess (p. 420)**

The number of edge devices with the successful deployment.

Type: Integer

**LastModifiedTime (p. 420)**

The time when the edge deployment plan was last updated.

Type: Timestamp

**ModelConfigs (p. 420)**

List of models associated with the edge deployment plan.

Type: Array of **EdgeDeploymentModelConfig (p. 1435)** objects

**NextToken (p. 420)**

Token to use when calling the next set of stages in the edge deployment plan.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: `.*`

**Stages (p. 420)**

List of stages in the edge deployment plan.

Type: Array of **DeploymentStageStatusSummary (p. 1408)** objects

**Errors**

For information about the errors that are common to all actions, see **Common Errors (p. 2180)**.
ResourceNotFound

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DescribeEdgePackagingJob

Service: Amazon SageMaker Service

A description of edge packaging jobs.

Request Syntax

```json
{
   "EdgePackagingJobName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**EdgePackagingJobName (p. 424)**

The name of the edge packaging job.

Type: String


Pattern: `^[a-zA-Z0-9](-[a-zA-Z0-9]{0,62})\{0,62}$`

Required: Yes

Response Syntax

```json
{
   "CompilationJobName": "string",
   "CreationTime": number,
   "EdgePackagingJobArn": "string",
   "EdgePackagingJobName": "string",
   "EdgePackagingJobStatus": "string",
   "EdgePackagingJobStatusMessage": "string",
   "LastModifiedTime": number,
   "ModelArtifact": "string",
   "ModelName": "string",
   "ModelSignature": "string",
   "ModelVersion": "string",
   "OutputConfig": {
      "KmsKeyId": "string",
      "PresetDeploymentConfig": "string",
      "PresetDeploymentType": "string",
      "S3OutputLocation": "string"
   },
   "PresetDeploymentOutput": {
      "Artifact": "string",
      "Status": "string",
      "StatusMessage": "string",
      "Type": "string"
   },
   "ResourceKey": "string",
   "RoleArn": "string"
}
```
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**CompilationJobName (p. 424)**

The name of the SageMaker Neo compilation job that is used to locate model artifacts that are being packaged.

Type: String


Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}$`

**CreationTime (p. 424)**

The timestamp of when the packaging job was created.

Type: Timestamp

**EdgePackagingJobArn (p. 424)**

The Amazon Resource Name (ARN) of the edge packaging job.

Type: String


Pattern: `^arn:aws[a-z-]*:sagemaker:[a-z-]*:\d{12}:edge-packaging-job/?[a-zA-Z_0-9+=,.@-_/]+$`

**EdgePackagingJobName (p. 424)**

The name of the edge packaging job.

Type: String


Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}$`

**EdgePackagingJobStatus (p. 424)**

The current status of the packaging job.

Type: String

Valid Values: STARTING | INPROGRESS | COMPLETED | FAILED | STOPPING | STOPPED

**EdgePackagingJobStatusMessage (p. 424)**

Returns a message describing the job status and error messages.

Type: String

**LastModifiedTime (p. 424)**

The timestamp of when the job was last updated.

Type: Timestamp
ModelArtifact (p. 424)

The Amazon Simple Storage (S3) URI where model artifacts are stored.

Type: String

Length Constraints: Maximum length of 1024.

Pattern:\^(https|s3)://(\[^/]*)\/%(\^[^\^]+)\/?(\.*)$

ModelName (p. 424)

The name of the model.

Type: String


Pattern:\^[a-zA-Z0-9\-]*[a-zA-Z0-9]\{0,62}$

ModelSignature (p. 424)

The signature document of files in the model artifact.

Type: String

ModelVersion (p. 424)

The version of the model.

Type: String


Pattern: [a-zA-Z0-9\ \_\.]*

OutputConfig (p. 424)

The output configuration for the edge packaging job.

Type: EdgeOutputConfig (p. 1444) object

PresetDeploymentOutput (p. 424)

The output of a SageMaker Edge Manager deployable resource.

Type: EdgePresetDeploymentOutput (p. 1448) object

ResourceKey (p. 424)

The AWS KMS key to use when encrypting the EBS volume the job run on.

Type: String

Length Constraints: Maximum length of 2048.

Pattern: .*

RoleArn (p. 424)

The Amazon Resource Name (ARN) of an IAM role that enables Amazon SageMaker to download and upload the model, and to contact Neo.

Type: String

Pattern: ^arn:aws[a-z\-]*:iam::\d{12}:role/?[a-zA-Z_0-9+=,.@\-_/]++$

Errors
For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceNotFound
- Resource being access is not found.
- HTTP Status Code: 400

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DescribeEndpoint
Service: Amazon SageMaker Service

Returns the description of an endpoint.

Request Syntax

```json
{
  "EndpointName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see [Common Parameters (p. 2178)](#).

The request accepts the following data in JSON format.

**EndpointName (p. 428)**

- The name of the endpoint.
  - Type: String
  - Length Constraints: Maximum length of 63.
  - Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$
  - Required: Yes

Response Syntax

```json
{
  "AsyncInferenceConfig": {
    "ClientConfig": {
      "MaxConcurrentInvocationsPerInstance": number
    },
    "OutputConfig": {
      "KmsKeyId": "string",
      "NotificationConfig": {
        "ErrorTopic": "string",
        "IncludeInferenceResponseIn": [ "string" ],
        "SuccessTopic": "string"
      },
      "S3FailurePath": "string",
      "S3OutputPath": "string"
    }
  },
  "CreationTime": number,
  "DataCaptureConfig": {
    "CaptureStatus": "string",
    "CurrentSamplingPercentage": number,
    "DestinationS3Uri": "string",
    "EnableCapture": boolean,
    "KmsKeyId": "string"
  },
  "EndpointArn": "string",
  "EndpointConfigName": "string",
  "EndpointName": "string"
}
```
"EndpointStatus": "string",
"ExplainerConfig": {
  "ClarifyExplainerConfig": {
    "EnableExplanations": "string",
    "InferenceConfig": {
      "ContentType": "string",
      "FeatureHeaders": [ "string" ],
      "FeaturesAttribute": "string",
      "FeatureTypes": [ "string" ],
      "LabelAttribute": "string",
      "LabelHeaders": [ "string" ],
      "LabelIndex": number,
      "MaxPayloadInMB": number,
      "MaxRecordCount": number,
      "ProbabilityAttribute": "string",
      "ProbabilityIndex": number
    },
    "ShapConfig": {
      "NumberOfSamples": number,
      "Seed": number,
      "ShapBaselineConfig": {
        "MimeType": "string",
        "ShapBaseline": "string",
        "ShapBaselineUri": "string"
      },
      "TextConfig": {
        "Granularity": "string",
        "Language": "string"
      },
      "UseLogit": boolean
    }
  },
  "FailureReason": "string",
  "LastDeploymentConfig": {
    "AutoRollbackConfiguration": {
      "Alarms": [
        {"AlarmName": "string"
      ]
    },
    "BlueGreenUpdatePolicy": {
      "MaximumExecutionTimeoutInSeconds": number,
      "TerminationWaitInSeconds": number,
      "TrafficRoutingConfiguration": {
        "CanarySize": {
          "Type": "string",
          "Value": number
        },
        "LinearStepSize": {
          "Type": "string",
          "Value": number
        },
        "Type": "string",
        "WaitIntervalInSeconds": number
      }
    },
    "RollingUpdatePolicy": {
      "MaximumBatchSize": {
        "Type": "string",
        "Value": number
      },
      "MaximumExecutionTimeoutInSeconds": number,
      "RollbackMaximumBatchSize": {
        "Type": "string"
      }
    }
  }
}
"Value": number
"WaitIntervalInSeconds": number

"LastModifiedTime": number,
"PendingDeploymentSummary": {
  "EndpointConfigName": "string",
  "ProductionVariants": [
    {
      "AcceleratorType": "string",
      "CurrentInstanceCount": number,
      "CurrentServerlessConfig": {
        "MaxConcurrency": number,
        "MemorySizeInMB": number,
        "ProvisionedConcurrency": number
      },
      "CurrentWeight": number,
      "DeployedImages": [
        {
          "ResolutionTime": number,
          "ResolvedImage": "string",
          "SpecifiedImage": "string"
        }
      ],
      "DesiredInstanceCount": number,
      "DesiredServerlessConfig": {
        "MaxConcurrency": number,
        "MemorySizeInMB": number,
        "ProvisionedConcurrency": number
      },
      "DesiredWeight": number,
      "InstanceType": "string",
      "ManagedInstanceScaling": {
        "MaxInstanceCount": number,
        "MinInstanceCount": number,
        "Status": "string"
      },
      "RoutingConfig": {
        "RoutingStrategy": "string"
      },
      "VariantName": "string",
      "VariantStatus": [
        {
          "StartTime": number,
          "Status": "string",
          "StatusMessage": "string"
        }
      ],
      "ShadowProductionVariants": [
        {
          "AcceleratorType": "string",
          "CurrentInstanceCount": number,
          "CurrentServerlessConfig": {
            "MaxConcurrency": number,
            "MemorySizeInMB": number,
            "ProvisionedConcurrency": number
          },
          "CurrentWeight": number,
          "DeployedImages": [
            {
              "ResolutionTime": number,
              "ResolvedImage": "string",
              "SpecifiedImage": "string"
            }
          ]
        }
      ]
    }
  ]
}


```
{
    "DesiredInstanceCount": number,
    "DesiredServerlessConfig": {
        "MaxConcurrency": number,
        "MemorySizeInMB": number,
        "ProvisionedConcurrency": number
    },
    "DesiredWeight": number,
    "InstanceType": "string",
    "ManagedInstanceScaling": {
        "MaxInstanceCount": number,
        "MinInstanceCount": number,
        "Status": "string"
    },
    "RoutingConfig": {
        "RoutingStrategy": "string"
    },
    "VariantName": "string",
    "VariantStatus": [
        {
            "StartTime": number,
            "Status": "string",
            "StatusMessage": "string"
        }
    ]
},
"StartTime": number,
"ProductionVariants": [
    {
        "CurrentInstanceCount": number,
        "CurrentServerlessConfig": {
            "MaxConcurrency": number,
            "MemorySizeInMB": number,
            "ProvisionedConcurrency": number
        },
        "CurrentWeight": number,
        "DeployedImages": [
            {
                "ResolutionTime": number,
                "ResolvedImage": "string",
                "SpecifiedImage": "string"
            }
        ],
        "DesiredInstanceCount": number,
        "DesiredServerlessConfig": {
            "MaxConcurrency": number,
            "MemorySizeInMB": number,
            "ProvisionedConcurrency": number
        },
        "DesiredWeight": number,
        "ManagedInstanceScaling": {
            "MaxInstanceCount": number,
            "MinInstanceCount": number,
            "Status": "string"
        },
        "RoutingConfig": {
            "RoutingStrategy": "string"
        },
        "VariantName": "string",
        "VariantStatus": [
            {
                "StartTime": number,
                "Status": "string",
                "StatusMessage": "string"
            }
        ]
    }
}
```
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

AsyncInferenceConfig (p. 428)

Returns the description of an endpoint configuration created using the CreateEndpointConfig API.

Type: AsyncInferenceConfig (p. 1258) object

CreationTime (p. 428)

A timestamp that shows when the endpoint was created.
Type: Timestamp

**DataCaptureConfig (p. 428)**

The currently active data capture configuration used by your Endpoint.

Type: **DataCaptureConfigSummary (p. 1383)** object

**EndpointArn (p. 428)**

The Amazon Resource Name (ARN) of the endpoint.

Type: String


Pattern: arn:aws[a-z\-]*:sagemaker:[a-zA-Z0-9\-]*:[0-9]{12}:endpoint/.*

**EndpointConfigName (p. 428)**

The name of the endpoint configuration associated with this endpoint.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}

**EndpointName (p. 428)**

Name of the endpoint.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}

**EndpointStatus (p. 428)**

The status of the endpoint.

- **OutOfService**: Endpoint is not available to take incoming requests.
- **Creating**: CreateEndpoint is executing.
- **Updating**: UpdateEndpoint or UpdateEndpointWeightsAndCapacities is executing.
- **SystemUpdating**: Endpoint is undergoing maintenance and cannot be updated or deleted or re-scaled until it has completed. This maintenance operation does not change any customer-specified values such as VPC config, KMS encryption, model, instance type, or instance count.
- **RollingBack**: Endpoint fails to scale up or down or change its variant weight and is in the process of rolling back to its previous configuration. Once the rollback completes, endpoint returns to an InService status. This transitional status only applies to an endpoint that has autoscaling enabled and is undergoing variant weight or capacity changes as part of an UpdateEndpointWeightsAndCapacities call or when the UpdateEndpointWeightsAndCapacities operation is called explicitly.
- **InService**: Endpoint is available to process incoming requests.
- **Deleting**: DeleteEndpoint is executing.
- **Failed**: Endpoint could not be created, updated, or re-scaled. Use the FailureReason value returned by DescribeEndpoint for information about the failure. DeleteEndpoint is the only operation that can be performed on a failed endpoint.
- **UpdateRollbackFailed**: Both the rolling deployment and auto-rollback failed. Your endpoint is in service with a mix of the old and new endpoint configurations. For information about how to remedy this issue and restore the endpoint's status to InService, see Rolling Deployments.
DescribeEndpoint

Type: String

Valid Values: OutOfService | Creating | Updating | SystemUpdating | RollingBack | InService | Deleting | Failed | UpdateRollbackFailed

**ExplainerConfig (p. 428)**

The configuration parameters for an explainer.

Type: **ExplainerConfig (p. 1479)** object

**FailureReason (p. 428)**

If the status of the endpoint is Failed, the reason why it failed.

Type: String

Length Constraints: Maximum length of 1024.

**LastDeploymentConfig (p. 428)**

The most recent deployment configuration for the endpoint.

Type: **DeploymentConfig (p. 1405)** object

**LastModifiedTime (p. 428)**

A timestamp that shows when the endpoint was last modified.

Type: Timestamp

**PendingDeploymentSummary (p. 428)**

Returns the summary of an in-progress deployment. This field is only returned when the endpoint is creating or updating with a new endpoint configuration.

Type: **PendingDeploymentSummary (p. 1804)** object

**ProductionVariants (p. 428)**

An array of **ProductionVariantSummary** objects, one for each model hosted behind this endpoint.

Type: Array of **ProductionVariantSummary (p. 1858)** objects

Array Members: Minimum number of 1 item.

**ShadowProductionVariants (p. 428)**

An array of **ProductionVariantSummary** objects, one for each model that you want to host at this endpoint in shadow mode with production traffic replicated from the model specified on ProductionVariants.

Type: Array of **ProductionVariantSummary (p. 1858)** objects

Array Members: Minimum number of 1 item.

**Errors**

For information about the errors that are common to all actions, see **Common Errors (p. 2180)**.

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- **AWS Command Line Interface**
• AWS SDK for .NET
• AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
DescribeEndpointConfig
Service: Amazon SageMaker Service

Returns the description of an endpoint configuration created using the CreateEndpointConfig API.

Request Syntax

```json
{
   "EndpointConfigName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**EndpointConfigName (p. 436)**

The name of the endpoint configuration.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

Required: Yes

Response Syntax

```json
{
   "AsyncInferenceConfig": {
      "ClientConfig": {
         "MaxConcurrentInvocationsPerInstance": number
      },
      "OutputConfig": {
         "KmsKeyId": "string",
         "NotificationConfig": {
            "ErrorTopic": "string",
            "IncludeInferenceResponseIn": [ "string" ],
            "SuccessTopic": "string"
         },
         "S3FailurePath": "string",
         "S3OutputPath": "string"
      }
   },
   "CreationTime": number,
   "DataCaptureConfig": {
      "CaptureContentTypeHeader": {
         "CsvContentTypes": [ "string" ],
         "JsonContentTypes": [ "string" ]
      },
      "CaptureOptions": [
         "CaptureMode": "string"
      ]
   }
}
```
Amazon SageMaker Amazon Sagemaker API Reference

DescribeEndpointConfig

```json
{
  "DestinationS3Uri": "string",
  "EnableCapture": boolean,
  "InitialSamplingPercentage": number,
  "KmsKeyId": "string"
},
  "EnableNetworkIsolation": boolean,
  "EndpointConfigArn": "string",
  "EndpointConfigName": "string",
  "ExecutionRoleArn": "string",
  "ExplainerConfig": {
    "ClarifyExplainerConfig": {
      "EnableExplanations": "string",
      "InferenceConfig": {
        "ContentTemplate": "string",
        "FeatureHeaders": [ "string" ],
        "FeaturesAttribute": "string",
        "FeatureTypes": [ "string" ],
        "LabelAttribute": "string",
        "LabelHeaders": [ "string" ],
        "LabelIndex": number,
        "MaxPayloadInMB": number,
        "MaxRecordCount": number,
        "ProbabilityAttribute": "string",
        "ProbabilityIndex": number
      },
      "ShapConfig": {
        "NumberOfSamples": number,
        "Seed": number,
        "ShapBaselineConfig": {
          "MimeType": "string",
          "ShapBaseline": "string",
          "ShapBaselineUri": "string"
        },
        "TextConfig": {
          "Granularity": "string",
          "Language": "string"
        },
        "UseLogit": boolean
      }
    },
  "KmsKeyId": "string",
  "ProductionVariants": [
    {
      "AcceleratorType": "string",
      "ContainerStartupHealthCheckTimeoutInSeconds": number,
      "CoreDumpConfig": {
        "DestinationS3Uri": "string",
        "KmsKeyId": "string"
      },
      "EnableSSMAccess": boolean,
      "InitialInstanceCount": number,
      "InitialVariantWeight": number,
      "InstanceType": "string",
      "ManagedInstanceScaling": {
        "MaxInstanceCount": number,
        "MinInstanceCount": number,
        "Status": "string"
      },
      "ModelDataDownloadTimeoutInSeconds": number,
      "ModelName": "string",
      "RoutingConfig": {
        "RoutingStrategy": "string"
      }
    }
  ]
}
```
"ServerlessConfig": {  
  "MaxConcurrency": number,  
  "MemorySizeInMB": number,  
  "ProvisionedConcurrency": number  
},  
"VariantName": "string",  
"VolumeSizeInGB": number  
},  
"ShadowProductionVariants": [  
  {  
    "AcceleratorType": "string",  
    "ContainerStartupHealthCheckTimeoutInSeconds": number,  
    "CoreDumpConfig": {  
      "DestinationS3Uri": "string",  
      "KmsKeyId": "string"  
    },  
    "EnableSSMAccess": boolean,  
    "InitialInstanceCount": number,  
    "InitialVariantWeight": number,  
    "InstanceType": "string",  
    "ManagedInstanceScaling": {  
      "MaxInstanceCount": number,  
      "MinInstanceCount": number,  
      "Status": "string"  
    },  
    "ModelDataDownloadTimeoutInSeconds": number,  
    "ModelName": "string",  
    "RoutingConfig": {  
      "RoutingStrategy": "string"  
    },  
    "ServerlessConfig": {  
      "MaxConcurrency": number,  
      "MemorySizeInMB": number,  
      "ProvisionedConcurrency": number  
    },  
    "VariantName": "string",  
    "VolumeSizeInGB": number  
  }  
],  
"VpcConfig": {  
  "SecurityGroupIds": [ "string" ],  
  "Subnets": [ "string" ]  
}  
}  

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

AsyncInferenceConfig (p. 436)

Returns the description of an endpoint configuration created using the CreateEndpointConfig API.

Type: AsyncInferenceConfig (p. 1258) object

CreationTime (p. 436)

A timestamp that shows when the endpoint configuration was created.

Type: Timestamp
DataCaptureConfig (p. 436)

Configuration to control how SageMaker captures inference data.

Type: DataCaptureConfig (p. 1381) object

EnableNetworkIsolation (p. 436)

Indicates whether all model containers deployed to the endpoint are isolated. If they are, no inbound or outbound network calls can be made to or from the model containers.

Type: Boolean

EndpointConfigArn (p. 436)

The Amazon Resource Name (ARN) of the endpoint configuration.

Type: String


Pattern: arn:aws[a-z-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:endpoint-config/.*

EndpointConfigName (p. 436)

Name of the SageMaker endpoint configuration.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9][-]*[a-zA-Z0-9]{0,62}

ExecutionRoleArn (p. 436)

The Amazon Resource Name (ARN) of the IAM role that you assigned to the endpoint configuration.

Type: String


Pattern: ^arn:aws[a-z-]*:iam::\d{12}:role/?[a-zA-Z_0-9+=,.@-_\/%]*$ 

ExplainerConfig (p. 436)

The configuration parameters for an explainer.

Type: ExplainerConfig (p. 1479) object

KmsKeyId (p. 436)

AWS KMS key ID Amazon SageMaker uses to encrypt data when storing it on the ML storage volume attached to the instance.

Type: String

Length Constraints: Maximum length of 2048.

Pattern: .* 

ProductionVariants (p. 436)

An array of ProductionVariant objects, one for each model that you want to host at this endpoint.

Type: Array of ProductionVariant (p. 1847) objects
Array Members: Minimum number of 1 item. Maximum number of 10 items.

**ShadowProductionVariants** *(p. 436)*

An array of `ProductionVariant` objects, one for each model that you want to host at this endpoint in shadow mode with production traffic replicated from the model specified on `ProductionVariants`.

Type: Array of `ProductionVariant` *(p. 1847)* objects

Array Members: Minimum number of 1 item. Maximum number of 10 items.

**VpcConfig** *(p. 436)*

Specifies an Amazon Virtual Private Cloud (VPC) that your SageMaker jobs, hosted models, and compute resources have access to. You can control access to and from your resources by configuring a VPC. For more information, see [Give SageMaker Access to Resources in your Amazon VPC](#).

Type: `VpcConfig` *(p. 2076)* object

**Errors**

For information about the errors that are common to all actions, see [Common Errors](#).

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
DescribeExperiment
Service: Amazon SageMaker Service

Provides a list of an experiment's properties.

Request Syntax

```json
{
   "ExperimentName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**ExperimentName (p. 441)**

The name of the experiment to describe.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,119}$`

Required: Yes

Response Syntax

```json
{
   "CreatedBy": {
      "DomainId": "string",
      "IamIdentity": {
         "Arn": "string",
         "PrincipalId": "string",
         "SourceIdentity": "string"
      },
      "UserProfileArn": "string",
      "UserProfileName": "string"
   },
   "CreationTime": number,
   "Description": "string",
   "DisplayName": "string",
   "ExperimentArn": "string",
   "ExperimentName": "string",
   "LastModifiedBy": {
      "DomainId": "string",
      "IamIdentity": {
         "Arn": "string",
         "PrincipalId": "string",
         "SourceIdentity": "string"
      },
      "UserProfileArn": "string",
      "UserProfileName": "string"
   }
}
```
"LastModifiedTime": number,
"Source": {
    "SourceArn": "string",
    "SourceType": "string"
}

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

CreatedBy (p. 441)
Who created the experiment.
Type: UserContext (p. 2067) object

CreationTime (p. 441)
When the experiment was created.
Type: Timestamp

Description (p. 441)
The description of the experiment.
Type: String
Length Constraints: Maximum length of 3072.
Pattern: .*

DisplayName (p. 441)
The name of the experiment as displayed. If DisplayName isn't specified, ExperimentName is displayed.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 120.
Pattern: ^[a-zA-Z0-9](\-*[a-zA-Z0-9]){0,119}$

ExperimentArn (p. 441)
The Amazon Resource Name (ARN) of the experiment.
Type: String
Length Constraints: Maximum length of 256.
Pattern: arn:aws[a-zA-Z0-9\-]*:sagemaker:[a-zA-Z0-9\-]*:[0-9]{12}:experiment/.*

ExperimentName (p. 441)
The name of the experiment.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 120.
Pattern: ^[a-zA-Z0-9](\-*[a-zA-Z0-9]){0,119}$
**LastModifiedBy (p. 441)**

Who last modified the experiment.

Type: UserContext (p. 2067) object

**LastModifiedTime (p. 441)**

When the experiment was last modified.

Type: Timestamp

**Source (p. 441)**

The Amazon Resource Name (ARN) of the source and, optionally, the type.

Type: ExperimentSource (p. 1475) object

**Errors**

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DescribeFeatureGroup
Service: Amazon SageMaker Service

Use this operation to describe a FeatureGroup. The response includes information on the creation time, FeatureGroup name, the unique identifier for each FeatureGroup, and more.

Request Syntax

```json
{
  "FeatureGroupName": "string",
  "NextToken": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**FeatureGroupName (p. 444)**

The name or Amazon Resource Name (ARN) of the FeatureGroup you want described.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 256.

Pattern: (arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:feature-group/)?([a-zA-Z0-9\-_\*]{0,63})

Required: Yes

**NextToken (p. 444)**

A token to resume pagination of the list of Features (FeatureDefinitions). 2,500 Features are returned by default.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Required: No

Response Syntax

```json
{
  "CreationTime": number,
  "Description": "string",
  "EventTimeFeatureName": "string",
  "FailureReason": "string",
  "FeatureDefinitions": [
    {
      "CollectionConfig": { ... },
      "CollectionType": "string",
    }
  ]
}
```
"FeatureName": "string",
"FeatureType": "string"
},
"FeatureGroupArn": "string",
"FeatureGroupName": "string",
"FeatureGroupStatus": "string",
"LastModifiedTime": number,
"LastUpdateStatus": {
  "FailureReason": "string",
  "Status": "string"
},
"NextToken": "string",
"OfflineStoreConfig": {
  "DataCatalogConfig": {
    "Catalog": "string",
    "Database": "string",
    "TableName": "string"
  },
  "DisableGlueTableCreation": boolean,
  "S3StorageConfig": {
    "KmsKeyId": "string",
    "ResolvedOutputS3Uri": "string",
    "S3Uri": "string"
  },
  "TableFormat": "string"
},
"OfflineStoreStatus": {
  "BlockedReason": "string",
  "Status": "string"
},
"OnlineStoreConfig": {
  "EnableOnlineStore": boolean,
  "SecurityConfig": {
    "KmsKeyId": "string"
  },
  "StorageType": "string",
  "TtlDuration": {
    "Unit": "string",
    "Value": number
  }
},
"OnlineStoreTotalSizeBytes": number,
"RecordIdentifierFeatureName": "string",
"RoleArn": "string"
}

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

CreationTime (p. 444)

A timestamp indicating when SageMaker created the FeatureGroup.

Type: Timestamp

Description (p. 444)

A free form description of the feature group.

Type: String
Length Constraints: Maximum length of 128.

**EventTimeFeatureName (p. 444)**

The name of the feature that stores the EventTime of a Record in a FeatureGroup.

An EventTime is a point in time when a new event occurs that corresponds to the creation or update of a Record in a FeatureGroup. All Records in the FeatureGroup have a corresponding EventTime.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 64.

Pattern: ^[a-zA-Z0-9][_-][a-zA-Z0-9]{0,63}$

**FailureReason (p. 444)**

The reason that the FeatureGroup failed to be replicated in the OfflineStore. This failure can occur because:

- The FeatureGroup could not be created in the OfflineStore.
- The FeatureGroup could not be deleted from the OfflineStore.

Type: String

Length Constraints: Maximum length of 1024.

**FeatureDefinitions (p. 444)**

A list of the Features in the FeatureGroup. Each feature is defined by a FeatureName and FeatureType.

Type: Array of FeatureDefinition (p. 1481) objects

Array Members: Minimum number of 1 item. Maximum number of 2500 items.

**FeatureGroupArn (p. 444)**

The Amazon Resource Name (ARN) of the FeatureGroup.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-zA-Z\-]*:sagemaker:[a-zA-Z0-9\-]*:[0-9]{12}:feature-group/.*

**FeatureGroupName (p. 444)**

The name of the FeatureGroup.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 64.

Pattern: ^[a-zA-Z0-9][_-][a-zA-Z0-9]{0,63}$

**FeatureGroupStatus (p. 444)**

The status of the feature group.

Type: String

Valid Values: Creating | Created | CreateFailed | Deleting | DeleteFailed

**LastModifiedTime (p. 444)**

A timestamp indicating when the feature group was last updated.
Type: Timestamp

**LastUpdateStatus (p. 444)**

A value indicating whether the update made to the feature group was successful.

Type: `LastUpdateStatus (p. 1644)` object

**NextToken (p. 444)**

A token to resume pagination of the list of Features (FeatureDefinitions).

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

**OfflineStoreConfig (p. 444)**

The configuration of the offline store. It includes the following configurations:

- Amazon S3 location of the offline store.
- Configuration of the Glue data catalog.
- Table format of the offline store.
- Option to disable the automatic creation of a Glue table for the offline store.
- Encryption configuration.

Type: `OfflineStoreConfig (p. 1777)` object

**OfflineStoreStatus (p. 444)**

The status of the OfflineStore. Notifies you if replicating data into the OfflineStore has failed. Returns either: Active or Blocked

Type: `OfflineStoreStatus (p. 1778)` object

**OnlineStoreConfig (p. 444)**

The configuration for the OnlineStore.

Type: `OnlineStoreConfig (p. 1784)` object

**OnlineStoreTotalSizeBytes (p. 444)**

The size of the OnlineStore in bytes.

Type: Long

**RecordIdentifierFeatureName (p. 444)**

The name of the Feature used for RecordIdentifier, whose value uniquely identifies a record stored in the feature store.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 64.

Pattern: `^[a-zA-Z0-9](\[-_]*[a-zA-Z0-9]){0,63}`

**RoleArn (p. 444)**

The Amazon Resource Name (ARN) of the IAM execution role used to persist data into the OfflineStore if an OfflineStoreConfig is provided.

Type: String

Pattern: ^arn:aws[a-z\-]*:iam::\d{12}:role/?[a-zA-Z_0-9+=,.@-_\/]+$

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceNotFoundException

Resource being accessed is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DescribeFeatureMetadata
Service: Amazon SageMaker Service

Shows the metadata for a feature within a feature group.

Request Syntax

```
{
  "FeatureGroupName": "string",
  "FeatureName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**FeatureGroupName (p. 449)**

The name or Amazon Resource Name (ARN) of the feature group containing the feature.

- Type: String
- Pattern: (arn:aws[a-z\-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:feature-group\/)?(\[a-zA-Z0-9\]([-_]\[a-zA-Z0-9\]){0,63})
- Required: Yes

**FeatureName (p. 449)**

The name of the feature.

- Type: String
- Length Constraints: Minimum length of 1. Maximum length of 64.
- Pattern: ^[a-zA-Z0-9-\[\[-\]*[a-zA-Z0-9-]](0,63)
- Required: Yes

Response Syntax

```
{
  "CreationTime": number,
  "Description": "string",
  "FeatureGroupArn": "string",
  "FeatureGroupName": "string",
  "FeatureName": "string",
  "FeatureType": "string",
  "LastModifiedTime": number,
  "Parameters": [
    {
      "Key": "string",
      "Value": "string"
    }
  ]
}
```
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**CreationTime (p. 449)**

- A timestamp indicating when the feature was created.
  - Type: Timestamp

**Description (p. 449)**

- The description you added to describe the feature.
  - Type: String
  - Length Constraints: Minimum length of 0. Maximum length of 255.
  - Pattern: .*

**FeatureGroupArn (p. 449)**

- The Amazon Resource Number (ARN) of the feature group that contains the feature.
  - Type: String
  - Length Constraints: Maximum length of 256.
  - Pattern: arn:aws[\-a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:feature-group/.*

**FeatureGroupName (p. 449)**

- The name of the feature group that you've specified.
  - Type: String
  - Length Constraints: Minimum length of 1. Maximum length of 64.
  - Pattern: ^[a-zA-Z0-9][-]*[a-zA-Z0-9]{0,63}

**FeatureName (p. 449)**

- The name of the feature that you've specified.
  - Type: String
  - Length Constraints: Minimum length of 1. Maximum length of 64.
  - Pattern: ^[a-zA-Z0-9][-]*[a-zA-Z0-9]{0,63}

**FeatureType (p. 449)**

- The data type of the feature.
  - Type: String
  - Valid Values: Integral | Fractional | String
**LastModifiedTime (p. 449)**

A timestamp indicating when the metadata for the feature group was modified. For example, if you add a parameter describing the feature, the timestamp changes to reflect the last time you

Type: Timestamp

**Parameters (p. 449)**

The key-value pairs that you added to describe the feature.

Type: Array of FeatureParameter (p. 1491) objects

Array Members: Minimum number of 0 items. Maximum number of 25 items.

**Errors**

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
DescribeFlowDefinition

Service: Amazon SageMaker Service

Returns information about the specified flow definition.

Request Syntax

```
{
  "FlowDefinitionName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see [Common Parameters](p. 2178).

The request accepts the following data in JSON format.

**FlowDefinitionName (p. 452)**

The name of the flow definition.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

Required: Yes

Response Syntax

```
{
  "CreationTime": number,
  "FailureReason": "string",
  "FlowDefinitionArn": "string",
  "FlowDefinitionName": "string",
  "FlowDefinitionStatus": "string",
  "HumanLoopActivationConfig": {
    "HumanLoopActivationConditionsConfig": {
      "HumanLoopActivationConditions": "string"
    }
  },
  "HumanLoopConfig": {
    "HumanTaskUiArn": "string",
    "PublicWorkforceTaskPrice": {
      "AmountInUsd": {
        "Cents": number,
        "Dollars": number,
        "TenthFractionsOfACent": number
      }
    },
    "TaskAvailabilityLifetimeInSeconds": number,
    "TaskCount": number,
    "TaskDescription": "string",
    "TaskKeywords": [ "string" ],
    "TaskTimeLimitInSeconds": number,
    "TaskTitle": "string",
  }
}
```
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**CreationTime (p. 452)**

The timestamp when the flow definition was created.

Type: Timestamp

**FailureReason (p. 452)**

The reason your flow definition failed.

Type: String

Length Constraints: Maximum length of 1024.

**FlowDefinitionArn (p. 452)**

The Amazon Resource Name (ARN) of the flow definition.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: arn:aws[a-zA-Z-]+:sagemaker:[a-zA-Z0-9-]*:[0-9]{12}:flow-definition/.*

**FlowDefinitionName (p. 452)**

The Amazon Resource Name (ARN) of the flow definition.

Type: String


Pattern: ^[a-zA-Z0-9](\-*[a-zA-Z0-9])*$

**FlowDefinitionStatus (p. 452)**

The status of the flow definition. Valid values are listed below.

Type: String

Valid Values: Initializing | Active | Failed | Deleting

**HumanLoopActivationConfig (p. 452)**

An object containing information about what triggers a human review workflow.

Type: [HumanLoopActivationConfig (p. 1516)](p. 1516) object
**HumanLoopConfig (p. 452)**

An object containing information about who works on the task, the workforce task price, and other task details.

Type: HumanLoopConfig (p. 1517) object

**HumanLoopRequestSource (p. 452)**

Container for configuring the source of human task requests. Used to specify if Amazon Rekognition or Amazon Textract is used as an integration source.

Type: HumanLoopRequestSource (p. 1522) object

**OutputConfig (p. 452)**

An object containing information about the output file.

Type: FlowDefinitionOutputConfig (p. 1502) object

**RoleArn (p. 452)**

The Amazon Resource Name (ARN) of the AWS Identity and Access Management (IAM) execution role for the flow definition.

Type: String


Pattern: ^arn:aws[a-z\-]*:iam::\d{12}:role/?[a-zA-Z_0-9+=,.@\-_/]+$

**Errors**

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DescribeHub
Service: Amazon SageMaker Service

Describe a hub.

**Note**
Hub APIs are only callable through SageMaker Studio.

**Request Syntax**

```json
{
   "HubName": "string"
}
```

**Request Parameters**

For information about the parameters that are common to all actions, see [Common Parameters](p. 2178).

The request accepts the following data in JSON format.

**HubName (p. 455)**

- The name of the hub to describe.
- Type: String
- Length Constraints: Maximum length of 63.
- Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}`
- Required: Yes

**Response Syntax**

```json
{
   "CreationTime": number,
   "FailureReason": "string",
   "HubArn": "string",
   "HubDescription": "string",
   "HubDisplayName": "string",
   "HubName": "string",
   "HubSearchKeywords": [ "string" ],
   "HubStatus": "string",
   "LastModifiedTime": number,
   "S3StorageConfig": {
      "S3OutputPath": "string"
   }
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.
**CreationTime (p. 455)**

The date and time that the hub was created.

Type: Timestamp

**FailureReason (p. 455)**

The failure reason if importing hub content failed.

Type: String

Length Constraints: Maximum length of 1024.

**HubArn (p. 455)**

The Amazon Resource Name (ARN) of the hub.

Type: String

Length Constraints: Maximum length of 255.

Pattern: .*

**HubDescription (p. 455)**

A description of the hub.

Type: String

Length Constraints: Maximum length of 1023.

Pattern: .*

**HubDisplayName (p. 455)**

The display name of the hub.

Type: String

Length Constraints: Maximum length of 255.

Pattern: .*

**HubName (p. 455)**

The name of the hub.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}$

**HubSearchKeywords (p. 455)**

The searchable keywords for the hub.

Type: Array of strings

Array Members: Maximum number of 50 items.

Length Constraints: Maximum length of 255.

Pattern: ^[^A-Z]*$
HubStatus (p. 455)

The status of the hub.

Type: String

Valid Values: InService | Creating | Updating | Deleting | CreateFailed | UpdateFailed | DeleteFailed

LastModifiedTime (p. 455)

The date and time that the hub was last modified.

Type: Timestamp

S3StorageConfig (p. 455)

The Amazon S3 storage configuration for the hub.

Type: HubS3StorageConfig (p. 1514) object

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceNotFound

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DescribeHubContent

Service: Amazon SageMaker Service

Describe the content of a hub.

**Note**
Hub APIs are only callable through SageMaker Studio.

**Request Syntax**

```json
{
    "HubContentName": "string",
    "HubContentType": "string",
    "HubContentVersion": "string",
    "HubName": "string"
}
```

**Request Parameters**

For information about the parameters that are common to all actions, see [Common Parameters](p. 2178).

The request accepts the following data in JSON format.

**HubContentName** *(p. 458)*

The name of the content to describe.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

Required: Yes

**HubContentType** *(p. 458)*

The type of content in the hub.

Type: String

Valid Values: Model | Notebook

Required: Yes

**HubContentVersion** *(p. 458)*

The version of the content to describe.

Type: String


Pattern: ^\d{1,4}.\d{1,4}.\d{1,4}$

Required: No

**HubName** *(p. 458)*

The name of the hub that contains the content to describe.
Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

Required: Yes

Response Syntax

```json
{
    "CreationTime": number,
    "DocumentSchemaVersion": "string",
    "FailureReason": "string",
    "HubArn": "string",
    "HubContentArn": "string",
    "HubContentDependencies": [
        {
            "DependencyCopyPath": "string",
            "DependencyOriginPath": "string"
        }
    ],
    "HubContentDescription": "string",
    "HubContentDisplayName": "string",
    "HubContentDocument": "string",
    "HubContentMarkdown": "string",
    "HubContentName": "string",
    "HubContentSearchKeywords": [ "string" ],
    "HubContentStatus": "string",
    "HubContentType": "string",
    "HubContentVersion": "string",
    "HubName": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**CreationTime (p. 459)**

The date and time that hub content was created.

Type: Timestamp

**DocumentSchemaVersion (p. 459)**

The document schema version for the hub content.

Type: String


Pattern: ^\d{1,4}\.\d{1,4}\.\d{1,4}$

**FailureReason (p. 459)**

The failure reason if importing hub content failed.

Type: String
Length Constraints: Maximum length of 1024.

**HubArn (p. 459)**

The Amazon Resource Name (ARN) of the hub that contains the content.

Type: String

Length Constraints: Maximum length of 255.

Pattern: . *

**HubContentArn (p. 459)**

The Amazon Resource Name (ARN) of the hub content.

Type: String

Length Constraints: Maximum length of 255.

Pattern: . *

**HubContentDependencies (p. 459)**

The location of any dependencies that the hub content has, such as scripts, model artifacts, datasets, or notebooks.

Type: Array of **HubContentDependency (p. 1508)** objects

Array Members: Maximum number of 50 items.

**HubContentDescription (p. 459)**

A description of the hub content.

Type: String

Length Constraints: Maximum length of 1023.

Pattern: . *

**HubContentDisplayName (p. 459)**

The display name of the hub content.

Type: String

Length Constraints: Maximum length of 255.

Pattern: . *

**HubContentDocument (p. 459)**

The hub content document that describes information about the hub content such as type, associated containers, scripts, and more.

Type: String

Length Constraints: Maximum length of 65535.

Pattern: . *

**HubContentMarkdown (p. 459)**

A string that provides a description of the hub content. This string can include links, tables, and standard markdown formatting.

Type: String
Length Constraints: Maximum length of 65535.
Pattern: .*  

**HubContentName (p. 459)**

The name of the hub content.
Type: String
Length Constraints: Maximum length of 63.
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

**HubContentSearchKeywords (p. 459)**

The searchable keywords for the hub content.
Type: Array of strings
Array Members: Maximum number of 50 items.
Length Constraints: Maximum length of 255.
Pattern: ^[^A-Z]*$

**HubContentStatus (p. 459)**

The status of the hub content.
Type: String
Valid Values: Available | Importing | Deleting | ImportFailed | DeleteFailed

**HubContentType (p. 459)**

The type of hub content.
Type: String
Valid Values: Model | Notebook

**HubContentVersion (p. 459)**

The version of the hub content.
Type: String
Pattern: ^\d{1,4}.\d{1,4}.\d{1,4}$

**HubName (p. 459)**

The name of the hub that contains the content.
Type: String
Length Constraints: Maximum length of 63.
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

**Errors**

For information about the errors that are common to all actions, see Common Errors (p. 2180).
ResourceNotFound

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DescribeHumanTaskUi
Service: Amazon SageMaker Service

Returns information about the requested human task user interface (worker task template).

Request Syntax

```
{
    "HumanTaskUiName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see [Common Parameters](p. 2178).

The request accepts the following data in JSON format.

**HumanTaskUiName (p. 463)**

The name of the human task user interface (worker task template) you want information about.

Type: String
Pattern: ^[a-z0-9](-*[a-z0-9])*
Required: Yes

Response Syntax

```
{
    "CreationTime": number,
    "HumanTaskUiArn": "string",
    "HumanTaskUiName": "string",
    "HumanTaskUiStatus": "string",
    "UiTemplate": {
        "ContentSha256": "string",
        "Url": "string"
    }
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**CreationTime (p. 463)**

The timestamp when the human task user interface was created.

Type: Timestamp

**HumanTaskUiArn (p. 463)**

The Amazon Resource Name (ARN) of the human task user interface (worker task template).
Type: String
Length Constraints: Maximum length of 1024.
Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:human-task-ui/.*

**HumanTaskUiName (p. 463)**

The name of the human task user interface (worker task template).

Type: String
Pattern: ^[a-z0-9](-*[a-z0-9])*

**HumanTaskUiStatus (p. 463)**

The status of the human task user interface (worker task template). Valid values are listed below.

Type: String
Valid Values: Active | Deleting

**UiTemplate (p. 463)**

Container for user interface template information.

Type: `UiTemplateInfo (p. 2065)` object

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](#).

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
DescribeHyperParameterTuningJob

Service: Amazon SageMaker Service

Returns a description of a hyperparameter tuning job, depending on the fields selected. These fields can include the name, Amazon Resource Name (ARN), job status of your tuning job and more.

Request Syntax

```
{
   "HyperParameterTuningJobName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**HyperParameterTuningJobName (p. 465)**

The name of the tuning job.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,31}$

Required: Yes

Response Syntax

```
{
   "Autotune": {
      "Mode": "string"
   },
   "BestTrainingJob": {
      "CreationTime": number,
      "FailureReason": "string",
      "FinalHyperParameterTuningJobObjectiveMetric": {
         "MetricName": "string",
         "Type": "string",
         "Value": number
      },
      "ObjectiveStatus": "string",
      "TrainingEndTime": number,
      "TrainingJobArn": "string",
      "TrainingJobDefinitionName": "string",
      "TrainingJobName": "string",
      "TrainingJobStatus": "string",
      "TrainingStartTime": number,
      "TunedHyperParameters": {
         "string": "string"
      },
      "TuningJobName": "string"
   },
   "ConsumedResources": {
      "RuntimeInSeconds": number
   }
}
```

465


```
"CreationTime": number,
"FailureReason": "string",
"HyperParameterTuningEndTime": number,
"HyperParameterTuningJobArn": "string",
"HyperParameterTuningJobConfig": {
  "HyperParameterTuningJobObjective": {
    "MetricName": "string",
    "Type": "string"
  },
  "ParameterRanges": {
    "AutoParameters": [
      {
        "Name": "string",
        "ValueHint": "string"
      }
    ],
    "CategoricalParameterRanges": [
      {
        "Name": "string",
        "Values": ["string"]
      }
    ],
    "ContinuousParameterRanges": [
      {
        "MaxValue": "string",
        "MinValue": "string",
        "Name": "string",
        "ScalingType": "string"
      }
    ],
    "IntegerParameterRanges": [
      {
        "MaxValue": "string",
        "MinValue": "string",
        "Name": "string",
        "ScalingType": "string"
      }
    ]
  }
},
"RandomSeed": number,
"ResourceLimits": {
  "MaxNumberOfTrainingJobs": number,
  "MaxParallelTrainingJobs": number,
  "MaxRuntimeInSeconds": number
},
"Strategy": "string",
"StrategyConfig": {
  "HyperbandStrategyConfig": {
    "MaxResource": number,
    "MinResource": number
  }
},
"TrainingJobEarlyStoppingType": "string",
"TuningJobCompletionCriteria": {
  "BestObjectiveNotImproving": {
    "MaxNumberOfTrainingJobsNotImproving": number
  },
  "ConvergenceDetected": {
    "CompleteOnConvergence": "string"
  },
  "TargetObjectiveMetricValue": number
}
"HyperParameterTuningJobName": "string",
"HyperParameterTuningJobStatus": "string",
```

466
<table>
<thead>
<tr>
<th>Key</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>LastModifiedTime</td>
<td>number</td>
</tr>
<tr>
<td>ObjectiveStatusCounters</td>
<td></td>
</tr>
<tr>
<td>Failed</td>
<td>number</td>
</tr>
<tr>
<td>Pending</td>
<td>number</td>
</tr>
<tr>
<td>Succeeded</td>
<td>number</td>
</tr>
<tr>
<td>OverallBestTrainingJob</td>
<td></td>
</tr>
<tr>
<td>CreationTime</td>
<td>number</td>
</tr>
<tr>
<td>FailureReason</td>
<td>string</td>
</tr>
<tr>
<td>MetricName</td>
<td>string</td>
</tr>
<tr>
<td>Type</td>
<td>string</td>
</tr>
<tr>
<td>Value</td>
<td>number</td>
</tr>
<tr>
<td>ObjectiveStatus</td>
<td>string</td>
</tr>
<tr>
<td>TrainingEndTime</td>
<td>number</td>
</tr>
<tr>
<td>TrainingJobArn</td>
<td>string</td>
</tr>
<tr>
<td>TrainingJobDefinitionName</td>
<td>string</td>
</tr>
<tr>
<td>TrainingJobName</td>
<td>string</td>
</tr>
<tr>
<td>TrainingJobStatus</td>
<td>string</td>
</tr>
<tr>
<td>TrainingStartTime</td>
<td>number</td>
</tr>
<tr>
<td>TunedHyperParameters</td>
<td></td>
</tr>
<tr>
<td>TuningJobName</td>
<td>string</td>
</tr>
<tr>
<td>TrainingJobDefinition</td>
<td></td>
</tr>
<tr>
<td>AlgorithmSpecification</td>
<td></td>
</tr>
<tr>
<td>AlgorithmName</td>
<td>string</td>
</tr>
<tr>
<td>MetricDefinitions</td>
<td></td>
</tr>
<tr>
<td>Name</td>
<td>string</td>
</tr>
<tr>
<td>Regex</td>
<td>string</td>
</tr>
<tr>
<td>TrainingImage</td>
<td>string</td>
</tr>
<tr>
<td>TrainingInputMode</td>
<td>string</td>
</tr>
<tr>
<td>CheckpointConfig</td>
<td></td>
</tr>
<tr>
<td>LocalPath</td>
<td>string</td>
</tr>
<tr>
<td>S3Uri</td>
<td>string</td>
</tr>
<tr>
<td>DefinitionName</td>
<td>string</td>
</tr>
<tr>
<td>EnableInterContainerTrafficEncryption</td>
<td>boolean</td>
</tr>
<tr>
<td>EnableManagedSpotTraining</td>
<td>boolean</td>
</tr>
<tr>
<td>EnableNetworkIsolation</td>
<td>boolean</td>
</tr>
<tr>
<td>Environment</td>
<td>string</td>
</tr>
<tr>
<td>HyperParameterRanges</td>
<td></td>
</tr>
<tr>
<td>AutoParameters</td>
<td></td>
</tr>
<tr>
<td>Name</td>
<td>string</td>
</tr>
<tr>
<td>ValueHint</td>
<td>string</td>
</tr>
<tr>
<td>CategoricalParameterRanges</td>
<td></td>
</tr>
<tr>
<td>Name</td>
<td>string</td>
</tr>
<tr>
<td>Values</td>
<td>[ string ]</td>
</tr>
<tr>
<td>ContinuousParameterRanges</td>
<td></td>
</tr>
<tr>
<td>MaxValue</td>
<td>string</td>
</tr>
<tr>
<td>MinValue</td>
<td>string</td>
</tr>
</tbody>
</table>
"Name": "string",
"ScalingType": "string"
}
],
"IntegerParameterRanges": [
{
"MaxValue": "string",
"MinValue": "string",
"Name": "string",
"ScalingType": "string"
}
]
},
"HyperParameterTuningResourceConfig": {
"AllocationStrategy": "string",
"InstanceConfigs": [
{
"InstanceCount": number,
"InstanceType": "string",
"VolumeSizeInGB": number
}
],
"InstanceCount": number,
"InstanceType": "string",
"VolumeKmsKeyId": "string",
"VolumeSizeInGB": number
},
"InputDataConfig": [ 
{
"ChannelName": "string",
"CompressionType": "string",
"ContentType": "string",
"DataSource": {
"FileSystemDataSource": {
"DirectoryPath": "string",
"FileSystemAccessMode": "string",
"FileSystemId": "string",
"FileSystemType": "string"
},
"S3DataSource": {
"AttributeNames": [ "string" ],
"InstanceGroupNames": [ "string" ],
"S3DataDistributionType": "string",
"S3DataType": "string",
"S3Uri": "string"
}
},
"InputMode": "string",
"RecordWrapperType": "string",
"ShuffleConfig": { 
"Seed": number
}
}
],
"OutputDataConfig": {
"CompressionType": "string",
"KmsKeyId": "string",
"S3OutputPath": "string"
},
"ResourceConfig": {
"InstanceCount": number,
"InstanceGroups": [
{
"InstanceCount": number,
"InstanceGroupName": "string",
"InstanceType": "string"
}
]
"InstanceType": "string",
"Keep_alive_Period_In_Seconds": number,
"Volume_Kms_Key_Id": "string",
"Volume_Size_In_GB": number
},
"Retry_Strategy": {
  "Maximum_Retry_Attempts": number
},
"Role_Arn": "string",
"Static_Hyper_Parameters": {
  "string": "string"
},
"Stopping_Condition": {
  "Max_Pending_Time_In_Seconds": number,
  "Max_Runtime_In_Seconds": number,
  "Max.Wait_Time_In_Seconds": number
},
"Tuning_Objective": {
  "Metric_Name": "string",
  "Type": "string"
},
"Vpc_Config": {
  "Security_Group_Ids": [ "string" ],
  "Subnets": [ "string" ]
}
},
"Training_Job_Definitions": [
  {
    "Algorithm_Specification": {
      "Algorithm_Name": "string",
      "Metric_Definitions": [
        {
          "Name": "string",
          "Regex": "string"
        }
      ],
      "Training_Image": "string",
      "Training_Input_Mode": "string"
    },
    "Checkpoint_Config": {
      "Local_Path": "string",
      "S3_Uri": "string"
    },
    "Definition_Name": "string",
    "Enable_Interceptor_Traffic_Encryption": boolean,
    "Enable_Managed_Spot_Training": boolean,
    "Enable_Network_Isolation": boolean,
    "Environment": {
      "string": "string"
    },
    "Hyper_Parameter_Ranges": {
      "Auto_Parameters": [
        {
          "Name": "string",
          "Value_Hint": "string"
        }
      ],
      "Categorical_Parameter_Ranges": [
        {
          "Name": "string",
          "Values": [ "string" ]
        }
      ],
      "Continuous_Parameter_Ranges": [}
"MaxValue": "string",
"MinValue": "string",
"Name": "string",
"ScalingType": "string"
}],
"IntegerParameterRanges": [
{
"MaxValue": "string",
"MinValue": "string",
"Name": "string",
"ScalingType": "string"
}
],
"HyperParameterTuningResourceConfig": {
"AllocationStrategy": "string",
"InstanceConfig": [
{
"InstanceCount": number,
"InstanceType": "string",
"VolumeSizeInGB": number
}
],
"InstanceCount": number,
"InstanceType": "string",
"VolumeKmsKeyId": "string",
"VolumeSizeInGB": number
},
"InputDataConfig": [
{
"ChannelName": "string",
"CompressionType": "string",
"ContentType": "string",
"DataSource": {
"FileSystemDataSource": {
"DirectoryPath": "string",
"FileSystemAccessMode": "string",
"FileSystemId": "string",
"FileSystemType": "string"
},
"S3DataSource": {
"AttributeNames": [ "string" ],
"InstanceGroupNames": [ "string" ],
"S3DataDistributionType": "string",
"S3DataType": "string",
"S3Uri": "string"
}
},
"InputMode": "string",
"RecordWrapperType": "string",
"ShuffleConfig": {
"Seed": number
}
}
],
"OutputDataConfig": {
"CompressionType": "string",
"KmsKeyId": "string",
"S3OutputPath": "string"
},
"ResourceConfig": {
"InstanceCount": number,
"InstanceGroups": [ 
}
"InstanceCount": number,
"InstanceGroupName": "string",
"InstanceType": "string"
],
"InstanceType": "string",
"KeepAlivePeriodInSeconds": number,
"VolumeKmsKeyId": "string",
"VolumeSizeInGB": number
},
"RetryStrategy": {
  "MaximumRetryAttempts": number
},
"RoleArn": "string",
"StaticHyperParameters": {
  "string": "string"
},
"StoppingCondition": {
  "MaxPendingTimeInSeconds": number,
  "MaxRuntimeInSeconds": number,
  "MaxWaitTimeInSeconds": number
},
"TuningObjective": {
  "MetricName": "string",
  "Type": "string"
},
"VpcConfig": {
  "SecurityGroupIds": ["string"],
  "Subnets": ["string"]
}
]
},
"TrainingJobStatusCounters": {
  "Completed": number,
  "InProgress": number,
  "NonRetryableError": number,
  "RetryableError": number,
  "Stopped": number
},
"TuningJobCompletionDetails": {
  "ConvergenceDetectedTime": number,
  "NumberOfTrainingJobsObjectiveNotImproving": number
},
"WarmStartConfig": {
  "ParentHyperParameterTuningJobs": [
    {
      "HyperParameterTuningJobName": "string"
    }
  ],
  "WarmStartType": "string"
}
]

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**Autotune (p. 465)**

A flag to indicate if autotune is enabled for the hyperparameter tuning job.

Type: Autotune (p. 1300) object
**BestTrainingJob (p. 465)**

A TrainingJobSummary object that describes the training job that completed with the best current HyperParameterTuningJobObjective.

Type: HyperParameterTrainingJobSummary (p. 1548) object

**ConsumedResources (p. 465)**

The total resources consumed by your hyperparameter tuning job.

Type: HyperParameterTuningJobConsumedResources (p. 1556) object

**CreationTime (p. 465)**

The date and time that the tuning job started.

Type: Timestamp

**FailureReason (p. 465)**

If the tuning job failed, the reason it failed.

Type: String

Length Constraints: Maximum length of 1024.

**HyperParameterTuningEndTime (p. 465)**

The date and time that the tuning job ended.

Type: Timestamp

**HyperParameterTuningJobArn (p. 465)**

The Amazon Resource Name (ARN) of the tuning job.

Type: String

Length Constraints: Maximum length of 256.

Pattern: `arn:aws[a-zA-Z\-\]*:sagemaker:[a-zA-Z0-9\-]*:[0-9]{12}:hyper-parameter-tuning-job/.*`

**HyperParameterTuningJobConfig (p. 465)**

The HyperParameterTuningJobConfig object that specifies the configuration of the tuning job.

Type: HyperParameterTuningJobConfig (p. 1554) object

**HyperParameterTuningJobName (p. 465)**

The name of the hyperparameter tuning job.

Type: String


Pattern: `^[a-zA-ZA-Z0-9](-*[a-zA-ZA-Z0-9])\{0,31}`

**HyperParameterTuningJobStatus (p. 465)**

The status of the tuning job: InProgress, Completed, Failed, Stopping, or Stopped.

Type: String

Valid Values: Completed | InProgress | Failed | Stopped | Stopping
**LastModifiedTime (p. 465)**

The date and time that the status of the tuning job was modified.

Type: Timestamp

**ObjectiveStatusCounters (p. 465)**

The `ObjectiveStatusCounters` object that specifies the number of training jobs, categorized by the status of their final objective metric, that this tuning job launched.

Type: `ObjectiveStatusCounters (p. 1776)` object

**OverallBestTrainingJob (p. 465)**

If the hyperparameter tuning job is an warm start tuning job with a `WarmStartType` of `IDENTICAL_DATA_AND_ALGORITHM`, this is the `TrainingJobSummary` for the training job with the best objective metric value of all training jobs launched by this tuning job and all parent jobs specified for the warm start tuning job.

Type: `HyperParameterTrainingJobSummary (p. 1548)` object

**TrainingJobDefinition (p. 465)**

The `HyperParameterTrainingJobDefinition` object that specifies the definition of the training jobs that this tuning job launches.

Type: `HyperParameterTrainingJobDefinition (p. 1544)` object

**TrainingJobDefinitions (p. 465)**

A list of the `HyperParameterTrainingJobDefinition` objects launched for this tuning job.

Type: Array of `HyperParameterTrainingJobDefinition (p. 1544)` objects

Array Members: Minimum number of 1 item. Maximum number of 10 items.

**TrainingJobStatusCounters (p. 465)**

The `TrainingJobStatusCounters` object that specifies the number of training jobs, categorized by status, that this tuning job launched.

Type: `TrainingJobStatusCounters (p. 2011)` object

**TuningJobCompletionDetails (p. 465)**

Tuning job completion information returned as the response from a hyperparameter tuning job. This information tells if your tuning job has or has not converged. It also includes the number of training jobs that have not improved model performance as evaluated against the objective function.

Type: `HyperParameterTuningJobCompletionDetails (p. 1553)` object

**WarmStartConfig (p. 465)**

The configuration for starting the hyperparameter parameter tuning job using one or more previous tuning jobs as a starting point. The results of previous tuning jobs are used to inform which combinations of hyperparameters to search over in the new tuning job.

Type: `HyperParameterTuningJobWarmStartConfig (p. 1564)` object

**Errors**

For information about the errors that are common to all actions, see `Common Errors (p. 2180)`. 
**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DescribeImage
Service: Amazon SageMaker Service
Describes a SageMaker image.

Request Syntax

```
{
    "ImageName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

ImageName (p. 475)

  The name of the image to describe.

  Type: String


  Pattern: ^[a-zA-Z0-9][-._][a-zA-Z0-9]{0,62}$

  Required: Yes

Response Syntax

```
{
    "CreationTime": number,
    "Description": "string",
    "DisplayName": "string",
    "FailureReason": "string",
    "ImageArn": "string",
    "ImageName": "string",
    "ImageStatus": "string",
    "LastModifiedTime": number,
    "RoleArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

CreationTime (p. 475)

  When the image was created.

  Type: Timestamp
Description (p. 475)
The description of the image.
Type: String
Pattern: .*

DisplayName (p. 475)
The name of the image as displayed.
Type: String
Pattern: ^\S(.*\S)?$  

FailureReason (p. 475)
When a create, update, or delete operation fails, the reason for the failure.
Type: String
Length Constraints: Maximum length of 1024.

ImageArn (p. 475)
The ARN of the image.
Type: String
Length Constraints: Maximum length of 256.
Pattern: ^arn:aws(-\[-\w\]+)*:sagemaker:.+:[0-9][12]:image/[a-z0-9][\-\.][a-zA-Z0-9]{0,62}$

ImageName (p. 475)
The name of the image.
Type: String
Pattern: ^[a-zA-Z0-9][\-\.][a-zA-Z0-9]{0,62}$

ImageStatus (p. 475)
The status of the image.
Type: String
Valid Values: CREATING | CREATED | CREATE_FAILED | UPDATING | UPDATE_FAILED | DELETING | DELETE_FAILED

LastModifiedTime (p. 475)
When the image was last modified.
Type: Timestamp

RoleArn (p. 475)
The ARN of the IAM role that enables Amazon SageMaker to perform tasks on your behalf.
Type: String
Pattern: ^arn:aws[a-z\-]*:iam::\d{12}:role/?[a-zA-Z\-0-9+=,.@\-_]/+$

Errors
For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceNotFound
Resource being access is not found.
HTTP Status Code: 400

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DescribeImageVersion
Service: Amazon SageMaker Service
Describes a version of a SageMaker image.

Request Syntax

```
{
    "Alias": "string",
    "ImageName": "string",
    "Version": number
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**Alias (p. 478)**

The alias of the image version.

Type: String


Pattern: (?!^[.-])^([a-zA-Z0-9-_.]+)$

Required: No

**ImageName (p. 478)**

The name of the image.

Type: String


Pattern: ^[a-zA-Z0-9][-._]?[a-zA-Z0-9]{0,62}$

Required: Yes

**Version (p. 478)**

The version of the image. If not specified, the latest version is described.

Type: Integer

Valid Range: Minimum value of 0.

Required: No

Response Syntax

```
{
    "BaseImage": "string",
    "ContainerImage": "string",
}
```
"CreationTime": number,
"FailureReason": "string",
"Horovod": boolean,
"ImageArn": "string",
"ImageVersionArn": "string",
"ImageVersionStatus": "string",
"JobType": "string",
"LastModifiedTime": number,
"MLFramework": "string",
"Processor": "string",
"ProgrammingLang": "string",
"ReleaseNotes": "string",
"VendorGuidance": "string",
"Version": number
}

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**BaseImage (p. 478)**

The registry path of the container image on which this image version is based.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 255.

Pattern: .*

**ContainerImage (p. 478)**

The registry path of the container image that contains this image version.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 255.

**CreationTime (p. 478)**

When the version was created.

Type: Timestamp

**FailureReason (p. 478)**

When a create or delete operation fails, the reason for the failure.

Type: String

Length Constraints: Maximum length of 1024.

**Horovod (p. 478)**

Indicates Horovod compatibility.

Type: Boolean

**ImageArn (p. 478)**

The ARN of the image the version is based on.

Type: String
**ImageVersionArn (p. 478)**

The ARN of the version.

Type: String

Length Constraints: Maximum length of 256.

Pattern: `^arn:aws(-[\w]+)*:sagemaker:.+:[0-9]{12}:image/[a-z0-9](\.-)?[a-z0-9]+$`

**ImageVersionStatus (p. 478)**

The status of the version.

Type: String

Valid Values: CREATING | CREATED | CREATE_FAILED | DELETING | DELETE_FAILED

**JobType (p. 478)**

Indicates SageMaker job type compatibility.

- TRAINING: The image version is compatible with SageMaker training jobs.
- INFERENC: The image version is compatible with SageMaker inference jobs.
- NOTEBOOK_KERNEL: The image version is compatible with SageMaker notebook kernels.

Type: String

Valid Values: TRAINING | INFERENC | NOTEBOOK_KERNEL

**LastModifiedTime (p. 478)**

When the version was last modified.

Type: Timestamp

**MLFramework (p. 478)**

The machine learning framework vended in the image version.

Type: String


Pattern: `^[a-zA-Z]+ \d+\.\d+(\.\d+)?$`

**Processor (p. 478)**

Indicates CPU or GPU compatibility.

- CPU: The image version is compatible with CPU.
- GPU: The image version is compatible with GPU.

Type: String

Valid Values: CPU | GPU

**ProgrammingLang (p. 478)**

The supported programming language and its version.
Type: String


Pattern: ^[a-zA-Z]+ \d+\..\d+(\.\d+)?$

ReleaseNotes (p. 478)

The maintainer description of the image version.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 255.

Pattern: .*

VendorGuidance (p. 478)

The stability of the image version specified by the maintainer.

• NOT_PROVIDED: The maintainers did not provide a status for image version stability.
• STABLE: The image version is stable.
• TO_BE_ARCHIVED: The image version is set to be archived. Custom image versions that are set to be archived are automatically archived after three months.
• ARCHIVED: The image version is archived. Archived image versions are not searchable and are no longer actively supported.

Type: String

Valid Values: NOT_PROVIDED | STABLE | TO_BE_ARCHIVED | ARCHIVED

Version (p. 478)

The version number.

Type: Integer

Valid Range: Minimum value of 0.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceNotFound

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

• AWS Command Line Interface
• AWS SDK for .NET
• AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DescribeInferenceComponent

Service: Amazon SageMaker Service

Returns information about an inference component.

Request Syntax

```json
{
    "InferenceComponentName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**InferenceComponentName (p. 483)**

The name of the inference component.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](\-[a-zA-Z0-9]*[a-zA-Z0-9])?$

Required: Yes

Response Syntax

```json
[
    "CreationTime": number,
    "EndpointArn": "string",
    "EndpointName": "string",
    "FailureReason": "string",
    "InferenceComponentArn": "string",
    "InferenceComponentName": "string",
    "InferenceComponentStatus": "string",
    "LastModifiedTime": number,
    "RuntimeConfig": {
        "CurrentCopyCount": number,
        "DesiredCopyCount": number
    },
    "Specification": {
        "ComputeResourceRequirements": {
            "MaxMemoryRequiredInMb": number,
            "MinMemoryRequiredInMb": number,
            "NumberOfAcceleratorDevicesRequired": number,
            "NumberOfCpuCoresRequired": number
        },
        "Container": {
            "ArtifactUrl": "string",
            "DeployedImage": {
                "ResolutionTime": number,
                "ResolvedImage": "string",
                "SpecifiedImage": "string"
            }
        }
    }
]
```


},
  "Environment": {
    "string": "string"
  }
},
  "ModelName": "string",
  "StartupParameters": {
    "ContainerStartupHealthCheckTimeoutInSeconds": number,
    "ModelDataDownloadTimeoutInSeconds": number
  }
},
  "VariantName": "string"
}

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**CreationTime (p. 483)**

The time when the inference component was created.

Type: Timestamp

**EndpointArn (p. 483)**

The Amazon Resource Name (ARN) of the endpoint that hosts the inference component.

Type: String


Pattern: arn:aws[a-zA-Z-]*:sagemaker:[a-zA-Z0-9-]*:[0-9]{12}:endpoint/.*

**EndpointName (p. 483)**

The name of the endpoint that hosts the inference component.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z-0-9]{0,62}$

**FailureReason (p. 483)**

If the inference component status is Failed, the reason for the failure.

Type: String

Length Constraints: Maximum length of 1024.

**InferenceComponentArn (p. 483)**

The Amazon Resource Name (ARN) of the inference component.

Type: String


**InferenceComponentName (p. 483)**

The name of the inference component.
Type: String
Length Constraints: Maximum length of 63.
Pattern: ^[a-zA-Z0-9](\-*[a-zA-Z0-9])*[a-zA-Z0-9]?$

InferenceComponentStatus (p. 483)
The status of the inference component.
Type: String
Valid Values: InService | Creating | Updating | Failed | Deleting

LastModifiedTime (p. 483)
The time when the inference component was last updated.
Type: Timestamp

RuntimeConfig (p. 483)
Details about the runtime settings for the model that is deployed with the inference component.
Type: InferenceComponentRuntimeConfigSummary (p. 1583) object

Specification (p. 483)
Details about the resources that are deployed with this inference component.
Type: InferenceComponentSpecificationSummary (p. 1585) object

VariantName (p. 483)
The name of the production variant that hosts the inference component.
Type: String
Length Constraints: Maximum length of 63.
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])*\{0,62}$

Errors
For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DescribeInferenceExperiment
Service: Amazon SageMaker Service

Returns details about an inference experiment.

Request Syntax

```
{
   "Name": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**Name (p. 486)**

The name of the inference experiment to describe.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: ^[a-zA-Z0-9\-]*[a-zA-Z0-9]\{0,119\}

Required: Yes

Response Syntax

```
{
   "Arn": "string",
   "CompletionTime": number,
   "CreationTime": number,
   "DataStorageConfig": {
      "ContentType": {
         "CsvContentTypes": [ "string" ],
         "JsonContentTypes": [ "string" ]
      },
      "Destination": "string",
      "KmsKey": "string"
   },
   "Description": "string",
   "EndpointMetadata": {
      "EndpointConfigName": "string",
      "EndpointName": "string",
      "EndpointStatus": "string",
      "FailureReason": "string"
   },
   "KmsKey": "string",
   "LastModifiedTime": number,
   "ModelVariants": [
      {
         "InfrastructureConfig": {
            "InfrastructureType": "string",
            "RealTimeInferenceConfig": {
               "InstanceCount": number,
```
"InstanceType": "string",
},
"ModelName": "string",
"Status": "string",
"VariantName": "string"
}
],
"Name": "string",
"RoleArn": "string",
"Schedule": {
  "EndTime": number,
  "StartTime": number
},
"ShadowModeConfig": {
  "ShadowModelVariants": [
    {
      "SamplingPercentage": number,
      "ShadowModelVariantName": "string"
    }
  ],
  "SourceModelVariantName": "string"
},
"Status": "string",
"StatusReason": "string",
"Type": "string"
}

Response Elements

If the action is successful, the service sends back an HTTP 200 response. The following data is returned in JSON format by the service.

**Arn (p. 486)**

The ARN of the inference experiment being described.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:inference-experiment/.*

**CompletionTime (p. 486)**

The timestamp at which the inference experiment was completed.

Type: Timestamp

**CreationTime (p. 486)**

The timestamp at which you created the inference experiment.

Type: Timestamp

**DataStorageConfig (p. 486)**

The Amazon S3 location and configuration for storing inference request and response data.

Type: InferenceExperimentDataStorageConfig (p. 1590) object

**Description (p. 486)**

The description of the inference experiment.
Type: String
Length Constraints: Maximum length of 1024.
Pattern: .*

**EndpointMetadata (p. 486)**

The metadata of the endpoint on which the inference experiment ran.

Type: [EndpointMetadata (p. 1462)] object

**KmsKey (p. 486)**

The AWS Key Management Service (AWS KMS) key that Amazon SageMaker uses to encrypt data on the storage volume attached to the ML compute instance that hosts the endpoint. For more information, see [CreateInferenceExperiment](#).

Type: String
Length Constraints: Maximum length of 2048.
Pattern: .*

**LastModifiedTime (p. 486)**

The timestamp at which you last modified the inference experiment.

Type: Timestamp

**ModelVariants (p. 486)**

An array of [ModelVariantConfigSummary](#) objects. There is one for each variant in the inference experiment. Each ModelVariantConfigSummary object in the array describes the infrastructure configuration for deploying the corresponding variant.

Type: Array of [ModelVariantConfigSummary (p. 1729)] objects

**Name (p. 486)**

The name of the inference experiment.

Type: String
Length Constraints: Minimum length of 1. Maximum length of 120.
Pattern: ^[a-zA-Z0-9\-_]*[a-zA-Z0-9]{0,119}$

**RoleArn (p. 486)**

The ARN of the IAM role that Amazon SageMaker can assume to access model artifacts and container images, and manage Amazon SageMaker Inference endpoints for model deployment.

Type: String
Pattern: ^arn:aws[a-zA-Z\-]*:iam::\d{12}:role/\*[a-zA-Z\-0-9\+=,.@\-_/]+$
a percentage of the inference requests. For the shadow variant it also shows the percentage of requests that Amazon SageMaker replicates.

Type: ShadowModeConfig (p. 1950) object

**Status (p. 486)**

The status of the inference experiment. The following are the possible statuses for an inference experiment:
- **Creating** - Amazon SageMaker is creating your experiment.
- **Created** - Amazon SageMaker has finished the creation of your experiment and will begin the experiment at the scheduled time.
- **Updating** - When you make changes to your experiment, your experiment shows as updating.
- **Starting** - Amazon SageMaker is beginning your experiment.
- **Running** - Your experiment is in progress.
- **Stopping** - Amazon SageMaker is stopping your experiment.
- **Completed** - Your experiment has completed.
- **Cancelled** - When you conclude your experiment early using the StopInferenceExperiment API, or if any operation fails with an unexpected error, it shows as cancelled.

Type: String

Valid Values: Creating | Created | Updating | Running | Starting | Stopping | Completed | Cancelled

**StatusReason (p. 486)**

The error message or client-specified Reason from the StopInferenceExperiment API, that explains the status of the inference experiment.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: .*

**Type (p. 486)**

The type of the inference experiment.

Type: String

Valid Values: ShadowMode

**Errors**

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceNotFoundException**

Resource being access is not found.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:
- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DescribeInferenceRecommendationsJob

Service: Amazon SageMaker Service

Provides the results of the Inference Recommender job. One or more recommendation jobs are returned.

Request Syntax

```json
{
  "JobName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**JobName (p. 491)**

The name of the job. The name must be unique within an AWS Region in the AWS account.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 64.

Pattern: ^[a-zA-Z0-9][-.*[a-zA-Z0-9]]{0,63}

Required: Yes

Response Syntax

```json
{
  "CompletionTime": number,
  "CreationTime": number,
  "EndpointPerformances": [
    {
      "EndpointInfo": {
        "EndpointName": "string"
      },
      "Metrics": {
        "MaxInvocations": number,
        "ModelLatency": number
      }
    },
    "FailureReason": "string",
    "InferenceRecommendations": [
      {
        "EndpointConfiguration": {
          "EndpointName": "string",
          "InitialInstanceCount": number,
          "InstanceType": "string",
          "ServerlessConfig": {
            "MaxConcurrency": number,
            "MemorySizeInMB": number,
            "ProvisionedConcurrency": number
          },
          "VariantName": "string"
        }
      }
    ]
}
```
"InvocationEndTime": number,
"InvocationStartTime": number,
"Metrics": {
  "CostPerHour": number,
  "CostPerInference": number,
  "CpuUtilization": number,
  "InvocationCount": number,
  "MemoryUtilization": number,
  "ModelLatency": number,
  "ModelSetupTime": number
},
"ModelConfiguration": {
  "CompilationJobName": "string",
  "EnvironmentParameters": [
    {
      "Key": "string",
      "Value": "string",
      "ValueType": "string"
    }
  ],
  "InferenceSpecificationName": "string"
},
"RecommendationId": "string"
},
"InputConfig": {
  "ContainerConfig": {
    "DataInputConfig": "string",
    "Domain": "string",
    "Framework": "string",
    "FrameworkVersion": "string",
    "NearestModelName": "string",
    "PayloadConfig": {
      "SamplePayloadUrl": "string",
      "SupportedContentTypes": [ "string" ]
    },
    "SupportedEndpointType": "string",
    "SupportedInstanceTypes": [ "string" ],
    "SupportedResponseMIMETypes": [ "string" ],
    "Task": "string"
  },
  "EndpointConfigurations": [
    {
      "EnvironmentParameterRanges": {
        "CategoricalParameterRanges": [
          {
            "Name": "string",
            "Value": [ "string" ]
          }
        },
        "InferenceSpecificationName": "string",
        "InstanceType": "string",
        "ServerlessConfig": {
          "MaxConcurrency": number,
          "MemorySizeInMB": number,
          "ProvisionedConcurrency": number
        }
      },
      "Endpoints": [
        {
          "EndpointName": "string"
        }
      ]
    }
  ]
}
"JobDurationInSeconds": number,
"ModelName": "string",
"ModelPackageVersionArn": "string",
"ResourceLimit": {
  "MaxNumberOfTests": number,
  "MaxParallelOfTests": number
},
"TrafficPattern": {
  "Phases": [
    {
      "DurationInSeconds": number,
      "InitialNumberOfUsers": number,
      "SpawnRate": number
    }
  ],
  "Stairs": {
    "DurationInSeconds": number,
    "NumberOfSteps": number,
    "UsersPerStep": number
  },
  "TrafficType": "string"
},
"VolumeKmsKeyId": "string",
"VpcConfig": {
  "SecurityGroupIds": [ "string" ],
  "Subnets": [ "string" ]
}
"JobArn": "string",
"JobDescription": "string",
"JobName": "string",
"JobType": "string",
"LastModifiedTime": number,
"RoleArn": "string",
"Status": "string",
"StoppingConditions": {
  "FlatInvocations": "string",
  "MaxInvocations": number,
  "ModelLatencyThresholds": [
    {
      "Percentile": "string",
      "ValueInMilliseconds": number
    }
  ]
}]

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

CompletionTime (p. 491)

A timestamp that shows when the job completed.

Type: Timestamp

CreationTime (p. 491)

A timestamp that shows when the job was created.

Type: Timestamp
**EndpointPerformances (p. 491)**

The performance results from running an Inference Recommender job on an existing endpoint.

Type: Array of [EndpointPerformance (p. 1466)] objects

Array Members: Maximum number of 1 item.

**FailureReason (p. 491)**

If the job fails, provides information why the job failed.

Type: String

Length Constraints: Maximum length of 1024.

**InferenceRecommendations (p. 491)**

The recommendations made by Inference Recommender.

Type: Array of [InferenceRecommendation (p. 1595)] objects

Array Members: Minimum number of 1 item. Maximum number of 10 items.

**InputConfig (p. 491)**

Returns information about the versioned model package Amazon Resource Name (ARN), the traffic pattern, and endpoint configurations you provided when you initiated the job.

Type: [RecommendationJobInputConfig (p. 1892)] object

**JobArn (p. 491)**

The Amazon Resource Name (ARN) of the job.

Type: String

Length Constraints: Maximum length of 256.

Pattern: `arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:inference-recommendations-job/.*`

**JobDescription (p. 491)**

The job description that you provided when you initiated the job.

Type: String

Length Constraints: Maximum length of 128.

**JobName (p. 491)**

The name of the job. The name must be unique within an AWS Region in the AWS account.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 64.

Pattern: `^[a-zA-Z0-9-Z\-9](.*[a-zA-Z0-9-Z\-9])\{0,63}`

**JobType (p. 491)**

The job type that you provided when you initiated the job.

Type: String

Valid Values: Default | Advanced
**LastModifiedTime (p. 491)**

A timestamp that shows when the job was last modified.

Type: Timestamp

**RoleArn (p. 491)**

The Amazon Resource Name (ARN) of the AWS Identity and Access Management (IAM) role you provided when you initiated the job.

Type: String


Pattern: `^arn:aws[a-z\-]*:iam::\d{12}:role/\?a-zA-Z_0-9+=,.@\-_\/]+$`

**Status (p. 491)**

The status of the job.

Type: String

Valid Values: PENDING | IN_PROGRESS | COMPLETED | FAILED | STOPPING | STOPPED

**StoppingConditions (p. 491)**

The stopping conditions that you provided when you initiated the job.

Type: `RecommendationJobStoppingConditions (p. 1898)` object

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)].

**ResourceNotFoundException**

Resource being access is not found.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DescribeLabelingJob

Service: Amazon SageMaker Service

Gets information about a labeling job.

Request Syntax

```json
{
   "LabelingJobName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**LabelingJobName (p. 496)**

The name of the labeling job to return information for.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

Required: Yes

Response Syntax

```json
{
   "CreationTime": number,
   "FailureReason": "string",
   "HumanTaskConfig": {
      "AnnotationConsolidationConfig": {
         "AnnotationConsolidationLambdaArn": "string"
      },
      "MaxConcurrentTaskCount": number,
      "NumberOfHumanWorkersPerDataObject": number,
      "PreHumanTaskLambdaArn": "string",
      "PublicWorkforceTaskPrice": {
         "AmountInUsd": {
            "Cents": number,
            "Dollars": number,
            "TenthFractionsOfACent": number
         }
      },
      "TaskAvailabilityLifetimeInSeconds": number,
      "TaskDescription": "string",
      "TaskKeywords": [ "string" ],
      "TaskTimeLimitInSeconds": number,
      "TaskTitle": "string",
      "UiConfig": {
         "HumanTaskUiArn": "string",
         "UiTemplateS3Uri": "string"
      },
      "WorkteamArn": "string"
   }
}
```
Amazon SageMaker Amazon Sagemaker API Reference
DescribeLabelingJob

"InputConfig": {
    "DataAttributes": {
        "ContentClassifiers": [ "string" ]
    },
    "DataSource": {
        "S3DataSource": {
            "ManifestS3Uri": "string"
        },
        "SnsDataSource": {
            "SnsTopicArn": "string"
        }
    }
},
"JobReferenceCode": "string",
"LabelAttributeName": "string",
"LabelCategoryConfigS3Uri": "string",
"LabelCounters": {"FailedNonRetryableError": number, "HumanLabeled": number, "MachineLabeled": number, "TotalLabeled": number, "Unlabeled": number },
"LabelingJobAlgorithmsConfig": {
    "InitialActiveLearningModelArn": "string",
    "LabelingJobAlgorithmSpecificationArn": "string",
    "LabelingJobResourceConfig": {
        "VolumeKmsKeyId": "string",
        "VpcConfig": {
            "SecurityGroupIds": [ "string" ],
            "Subnets": [ "string" ]
        }
    }
},
"LabelingJobArn": "string",
"LabelingJobName": "string",
"LabelingJobOutput": {
    "FinalActiveLearningModelArn": "string",
    "OutputDatasetS3Uri": "string"
},
"LabelingJobStatus": "string",
"LastModifiedTime": number,
"OutputConfig": {
    "KmsKeyId": "string",
    "S3OutputPath": "string",
    "SnsTopicArn": "string"
},
"RoleArn": "string",
"StoppingConditions": {
    "MaxHumanLabeledObjectCount": number,
    "MaxPercentageOfInputDatasetLabeled": number
},
"Tags": [
    {"Key": "string",
    "Value": "string"
}
]}

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.
The following data is returned in JSON format by the service.

**CreationTime (p. 496)**

The date and time that the labeling job was created.

Type: Timestamp

**FailureReason (p. 496)**

If the job failed, the reason that it failed.

Type: String

Length Constraints: Maximum length of 1024.

**HumanTaskConfig (p. 496)**

Configuration information required for human workers to complete a labeling task.

Type: `HumanTaskConfig (p. 1523)` object

**InputConfig (p. 496)**

Input configuration information for the labeling job, such as the Amazon S3 location of the data objects and the location of the manifest file that describes the data objects.

Type: `LabelingJobInputConfig (p. 1632)` object

**JobReferenceCode (p. 496)**

A unique identifier for work done as part of a labeling job.

Type: String

Length Constraints: Minimum length of 1.

Pattern: .+

**LabelAttributeName (p. 496)**

The attribute used as the label in the output manifest file.

Type: String


Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,126}$`

**LabelCategoryConfigS3Uri (p. 496)**

The S3 location of the JSON file that defines the categories used to label data objects. Please note the following label-category limits:

- Semantic segmentation labeling jobs using automated labeling: 20 labels
- Box bounding labeling jobs (all): 10 labels

The file is a JSON structure in the following format:

```json
{
  "document-version": "2018-11-28",
  "labels": [
    
  ]
}
```
"label": "label 1"
},
{
"label": "label 2"
},
...
{
"label": "label n"
}
]

Type: String
Length Constraints: Maximum length of 1024.
Pattern: ^(https|s3)://([^/]+)/?(.*)$

**LabelCounters (p. 496)**

Provides a breakdown of the number of data objects labeled by humans, the number of objects labeled by machine, the number of objects that couldn’t be labeled, and the total number of objects labeled.

Type: LabelCounters (p. 1623) object

**LabelingJobAlgorithmsConfig (p. 496)**

Configuration information for automated data labeling.

Type: LabelingJobAlgorithmsConfig (p. 1626) object

**LabelingJobArn (p. 496)**

The Amazon Resource Name (ARN) of the labeling job.

Type: String
Length Constraints: Maximum length of 2048.
Pattern: arn:aws[a-zA-Z\-]*:sagemaker:[a-zA-Z0-9\-]*:[0-9]{12}:labeling-job/.*

**LabelingJobName (p. 496)**

The name assigned to the labeling job when it was created.

Type: String
Pattern: ^[a-zA-Z0-9-][\-\*[a-zA-Z0-9-]+[0, 62]

**LabelingJobOutput (p. 496)**

The location of the output produced by the labeling job.

Type: LabelingJobOutput (p. 1633) object
**LabelingJobStatus (p. 496)**

The processing status of the labeling job.

Type: String

Valid Values: Initializing | InProgress | Completed | Failed | Stopping | Stopped

**LastModifiedTime (p. 496)**

The date and time that the labeling job was last updated.

Type: Timestamp

**OutputConfig (p. 496)**

The location of the job's output data and the AWS Key Management Service key ID for the key used to encrypt the output data, if any.

Type: LabelingJobOutputConfig (p. 1634) object

**RoleArn (p. 496)**

The Amazon Resource Name (ARN) that SageMaker assumes to perform tasks on your behalf during data labeling.

Type: String


Pattern: ^arn:aws[a-z\-]*:iam::\d{12}:role/\?a-zA-Z0-9+=,\.@\-_/]+$

**StoppingConditions (p. 496)**

A set of conditions for stopping a labeling job. If any of the conditions are met, the job is automatically stopped.

Type: LabelingJobStoppingConditions (p. 1639) object

**Tags (p. 496)**

An array of key-value pairs. You can use tags to categorize your AWS resources in different ways, for example, by purpose, owner, or environment. For more information, see Tagging AWS Resources.

Type: Array of Tag (p. 1979) objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

**Errors**

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceNotFoundException**

Resource being access is not found.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:
- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DescribeLineageGroup

Service: Amazon SageMaker Service

Provides a list of properties for the requested lineage group. For more information, see Cross-Account Lineage Tracking in the Amazon SageMaker Developer Guide.

Request Syntax

```json
{
   "LineageGroupName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**LineageGroupName (p. 502)**

The name of the lineage group.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,119}$

Required: Yes

Response Syntax

```json
{
   "CreatedBy": {
      "DomainId": "string",
      "IamIdentity": {
         "Arn": "string",
         "PrincipalId": "string",
         "SourceIdentity": "string"
      },
      "UserProfileArn": "string",
      "UserProfileName": "string"
   },
   "CreationTime": number,
   "Description": "string",
   "DisplayName": "string",
   "LastModifiedBy": {
      "DomainId": "string",
      "IamIdentity": {
         "Arn": "string",
         "PrincipalId": "string",
         "SourceIdentity": "string"
      },
      "UserProfileArn": "string",
      "UserProfileName": "string"
   },
   "LastModifiedTime": number,
}
```
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

CreatedBy (p. 502)

Information about the user who created or modified an experiment, trial, trial component, lineage group, project, or model card.

Type: UserContext (p. 2067) object

CreationTime (p. 502)

The creation time of lineage group.

Type: Timestamp

Description (p. 502)

The description of the lineage group.

Type: String

Length Constraints: Maximum length of 3072.

Pattern: .* 

DisplayName (p. 502)

The display name of the lineage group.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: ^[a-zA-Z0-9]([-][a-zA-Z0-9]){0,119}$

LastModifiedBy (p. 502)

Information about the user who created or modified an experiment, trial, trial component, lineage group, project, or model card.

Type: UserContext (p. 2067) object

LastModifiedTime (p. 502)

The last modified time of the lineage group.

Type: Timestamp

LineageGroupArn (p. 502)

The Amazon Resource Name (ARN) of the lineage group.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-zA-Z-]*:sagemaker:[a-zA-Z0-9-]*:[0-9]{12}:lineage-group/.*
LineageGroupName (p. 502)
The name of the lineage group.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 120.
Pattern: ^[a-zA-Z0-9][a-zA-Z0-9-]*[a-zA-Z0-9] \(0,119\)

Errors
For information about the errors that are common to all actions, see Common Errors (p. 2180).
ResourceNotFoundException
Resource being accessed is not found.
HTTP Status Code: 400

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DescribeModel
Service: Amazon SageMaker Service

Describes a model that you created using the CreateModel API.

Request Syntax

{  
 "ModelName": "string"
}

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters. The request accepts the following data in JSON format.

**ModelName (p. 505)**

The name of the model.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]*)* Required: Yes

Response Syntax

{  
 "Containers": [  
 {  
 "ContainerHostname": "string",
 "Environment": {  
 "string": "string"
 },
 "Image": "string",
 "ImageConfig": {  
 "RepositoryAccessMode": "string",
 "RepositoryAuthConfig": {  
 "RepositoryCredentialsProviderArn": "string"
 }  
 },
 "InferenceSpecificationName": "string",
 "Mode": "string",
 "ModelDataSource": {  
 "S3DataSource": {  
 "CompressionType": "string",
 "ModelAccessConfig": {  
 "AcceptEula": boolean
 },
 "S3DataType": "string",
 "S3Uri": "string"
 }  
 },
 "ModelDataUrl": "string",
 "ModelPackageName": "string",
}  
}
"MultiModelConfig": {
  "ModelCacheSetting": "string"
}
],
"CreationTime": number,
"DeploymentRecommendation": {
  "RealTimeInferenceRecommendations": [
    {
      "Environment": {
        "string": "string"
      },
      "InstanceType": "string",
      "RecommendationId": "string"
    }
  ],
  "RecommendationStatus": "string"
},
"EnableNetworkIsolation": boolean,
"ExecutionRoleArn": "string",
"InferenceExecutionConfig": {
  "Mode": "string"
},
"ModelArn": "string",
"ModelName": "string",
"PrimaryContainer": {
  "ContainerHostname": "string",
  "Environment": {
    "string": "string"
  },
  "Image": "string",
  "ImageConfig": {
    "RepositoryAccessMode": "string",
    "RepositoryAuthConfig": {
      "RepositoryCredentialsProviderArn": "string"
    }
  },
  "InferenceSpecificationName": "string",
  "Mode": "string",
  "ModelDataSource": {
    "S3DataSource": {
      "CompressionType": "string",
      "ModelAccessConfig": {
        "AcceptEula": boolean
      },
      "S3DataType": "string",
      "S3Uri": "string"
    }
  },
  "ModelDataUrl": "string",
  "ModelPackageName": "string",
  "MultiModelConfig": {
    "ModelCacheSetting": "string"
  }
},
"VpcConfig": {
  "SecurityGroupIds": [ "string" ],
  "Subnets": [ "string" ]
}

Response Elements

If the action is successful, the service sends back an HTTP 200 response.
The following data is returned in JSON format by the service.

**Containers (p. 505)**

The containers in the inference pipeline.

Type: Array of [ContainerDefinition (p. 1366)] objects

Array Members: Maximum number of 15 items.

**CreationTime (p. 505)**

A timestamp that shows when the model was created.

Type: Timestamp

**DeploymentRecommendation (p. 505)**

A set of recommended deployment configurations for the model.

Type: [DeploymentRecommendation (p. 1406)] object

**EnableNetworkIsolation (p. 505)**

If True, no inbound or outbound network calls can be made to or from the model container.

Type: Boolean

**ExecutionRoleArn (p. 505)**

The Amazon Resource Name (ARN) of the IAM role that you specified for the model.

Type: String


Pattern: ^arn:aws\[a-z\-]*:iam::\d{12}:role/\?\[a-zA-Z\-\_\@-\_\-\+\=,\-\_\-\-\-\-\]/+$

**InferenceExecutionConfig (p. 505)**

Specifies details of how containers in a multi-container endpoint are called.

Type: [InferenceExecutionConfig (p. 1589)] object

**ModelArn (p. 505)**

The Amazon Resource Name (ARN) of the model.

Type: String


Pattern: ^arn:aws\[a-zA-Z\-\-\]*:sagemaker:\[a-zA-Z\-\-\-\-\-\-\]*\[0-9\]{12}\:model/\.*$

**ModelName (p. 505)**

Name of the SageMaker model.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[^a-zA-Z0-9-]*\[a-zA-Z0-9\-\-\-\-\-\-\-\]*$

**PrimaryContainer (p. 505)**

The location of the primary inference code, associated artifacts, and custom environment map that the inference code uses when it is deployed in production.
Type: ContainerDefinition (p. 1366) object

**VpcConfig (p. 505)**

A VpcConfig object that specifies the VPC that this model has access to. For more information, see Protect Endpoints by Using an Amazon Virtual Private Cloud

Type: VpcConfig (p. 2076) object

**Errors**

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DescribeModelBiasJobDefinition

Returns a description of a model bias job definition.

Request Syntax

```json
{
  "JobDefinitionName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**JobDefinitionName (p. 509)**

- The name of the model bias job definition. The name must be unique within an AWS Region in the AWS account.

  Type: String


  Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$`

  Required: Yes

Response Syntax

```json
{
  "CreationTime": number,
  "JobDefinitionArn": "string",
  "JobDefinitionName": "string",
  "JobResources": {
    "ClusterConfig": {
      "InstanceCount": number,
      "InstanceType": "string",
      "VolumeKmsKeyId": "string",
      "VolumeSizeInGB": number
    }
  },
  "ModelBiasAppSpecification": {
    "ConfigUri": "string",
    "Environment": {
      "string": "string"
    },
    "ImageUri": "string"
  },
  "ModelBiasBaselineConfig": {
    "BaseliningJobName": "string",
    "ConstraintsResource": {
      "S3Uri": "string"
    }
  }
}
```
Amazon SageMaker Amazon Sagemaker API Reference
DescribeModelBiasJobDefinition

"ModelBiasJobInput": {
  "BatchTransformInput": {
    "DataCapturedDestinationS3Uri": "string",
    "DatasetFormat": {
      "Csv": {
        "Header": boolean
      },
      "Json": {
        "Line": boolean
      },
      "Parquet": {
      }
    },
    "EndTimeOffset": "string",
    "ExcludeFeaturesAttribute": "string",
    "FeaturesAttribute": "string",
    "InferenceAttribute": "string",
    "LocalPath": "string",
    "ProbabilityAttribute": "string",
    "ProbabilityThresholdAttribute": number,
    "S3DataDistributionType": "string",
    "S3InputMode": "string",
    "StartTimeOffset": "string"
  },
  "EndpointInput": {
    "EndpointName": "string",
    "EndTimeOffset": "string",
    "ExcludeFeaturesAttribute": "string",
    "FeaturesAttribute": "string",
    "InferenceAttribute": "string",
    "LocalPath": "string",
    "ProbabilityAttribute": "string",
    "ProbabilityThresholdAttribute": number,
    "S3DataDistributionType": "string",
    "S3InputMode": "string",
    "StartTimeOffset": "string"
  },
  "GroundTruthS3Input": {
    "S3Uri": "string"
  }
},
"ModelBiasJobOutputConfig": {
  "KmsKeyId": "string",
  "MonitoringOutputs": [
    {
      "S3Output": {
        "LocalPath": "string",
        "S3UploadMode": "string",
        "S3Uri": "string"
      }
    }
  ]
},
"NetworkConfig": {
  "EnableInterContainerTrafficEncryption": boolean,
  "EnableNetworkIsolation": boolean,
  "VpcConfig": {
    "SecurityGroupIds": [ "string" ],
    "Subnets": [ "string" ]
  }
},
"RoleArn": "string",
"StoppingCondition": {
  "MaxRuntimeInSeconds": number
}
Response Elements

If the action is successful, the service sends back an HTTP 200 response. The following data is returned in JSON format by the service.

**CreationTime (p. 509)**

The time at which the model bias job was created.
Type: Timestamp

**JobDefinitionArn (p. 509)**

The Amazon Resource Name (ARN) of the model bias job.
Type: String

Length Constraints: Maximum length of 256.

Pattern: . *

**JobDefinitionName (p. 509)**

The name of the bias job definition. The name must be unique within an AWS Region in the AWS account.
Type: String


Pattern: ^[a-zA-Z0-9](\-*[a-zA-Z0-9])\{0,62}$

**JobResources (p. 509)**

Identifies the resources to deploy for a monitoring job.
Type: MonitoringResources (p. 1756) object

**ModelBiasAppSpecification (p. 509)**

Configures the model bias job to run a specified Docker container image.
Type: ModelBiasAppSpecification (p. 1660) object

**ModelBiasBaselineConfig (p. 509)**

The baseline configuration for a model bias job.
Type: ModelBiasBaselineConfig (p. 1661) object

**ModelBiasJobInput (p. 509)**

Inputs for the model bias job.
Type: ModelBiasJobInput (p. 1662) object

**ModelBiasJobOutputConfig (p. 509)**

The output configuration for monitoring jobs.
Type: MonitoringOutputConfig (p. 1754) object
NetworkConfig (p. 509)

Networking options for a model bias job.

Type: MonitoringNetworkConfig (p. 1752) object

RoleArn (p. 509)

The Amazon Resource Name (ARN) of the IAM role that has read permission to the input data location and write permission to the output data location in Amazon S3.

Type: String


Pattern: ^arn:aws[a-z\-]*:iam::\d{12}:role/?[a-zA-Z_0-9+=,.@\-_/]+$

StoppingCondition (p. 509)

A time limit for how long the monitoring job is allowed to run before stopping.

Type: MonitoringStoppingCondition (p. 1765) object

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceNotFound

Resource being accessed is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DescribeModelCard
Service: Amazon SageMaker Service

Describes the content, creation time, and security configuration of an Amazon SageMaker Model Card.

Request Syntax

```
{
  "ModelCardName": "string",
  "ModelCardVersion": number
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**ModelCardName (p. 513)**

The name or Amazon Resource Name (ARN) of the model card to describe.

- Type: String
- Pattern: (arn:aws[a-z-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:model-card/.*)?([a-zA-Z0-9](-*[a-zA-Z0-9]){0,62})
- Required: Yes

**ModelCardVersion (p. 513)**

The version of the model card to describe. If a version is not provided, then the latest version of the model card is described.

- Type: Integer
- Required: No

Response Syntax

```
{
  "Content": "string",
  "CreatedBy": {
    "DomainId": "string",
    "IamIdentity": {
      "Arn": "string",
      "PrincipalId": "string",
      "SourceIdentity": "string"
    },
    "UserProfileArn": "string",
    "UserProfileName": "string"
  },
  "CreationTime": number,
  "LastModifiedBy": {
    "DomainId": "string",
  }
}
```
"IamIdentity": {
  "Arn": "string",
  "PrincipalId": "string",
  "SourceIdentity": "string"
},
"UserProfileArn": "string",
"UserProfileName": "string"
},
"LastModifiedTime": number,
"ModelCardArn": "string",
"ModelCardName": "string",
"ModelCardProcessingStatus": "string",
"ModelCardStatus": "string",
"ModelCardVersion": number,
"SecurityConfig": {
  "KmsKeyId": "string"
}
}

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**Content (p. 513)**

The content of the model card.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 100000.

Pattern: .*  

**CreatedBy (p. 513)**

Information about the user who created or modified an experiment, trial, trial component, lineage group, project, or model card.

Type: UserContext (p. 2067) object

**CreationTime (p. 513)**

The date and time the model card was created.

Type: Timestamp

**LastModifiedBy (p. 513)**

Information about the user who created or modified an experiment, trial, trial component, lineage group, project, or model card.

Type: UserContext (p. 2067) object

**LastModifiedTime (p. 513)**

The date and time the model card was last modified.

Type: Timestamp

**ModelCardArn (p. 513)**

The Amazon Resource Name (ARN) of the model card.

Type: String
Length Constraints: Maximum length of 256.

Pattern: ^arn:aws[a-zA-Z\-]*:sagemaker:[a-zA-Z0-9\-]{9,16}:[0-9]{12}:model-card/[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}$

**ModelCardName (p. 513)**

The name of the model card.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}$

**ModelCardProcessingStatus (p. 513)**

The processing status of model card deletion. The ModelCardProcessingStatus updates throughout the different deletion steps.

- **DeletePending**: Model card deletion request received.
- **DeleteInProgress**: Model card deletion is in progress.
- **ContentDeleted**: Deleted model card content.
- **ExportJobsDeleted**: Deleted all export jobs associated with the model card.
- **DeleteCompleted**: Successfully deleted the model card.
- **DeleteFailed**: The model card failed to delete.

Type: String

Valid Values: DeleteInProgress | DeletePending | ContentDeleted | ExportJobsDeleted | DeleteCompleted | DeleteFailed

**ModelCardStatus (p. 513)**

The approval status of the model card within your organization. Different organizations might have different criteria for model card review and approval.

- **Draft**: The model card is a work in progress.
- **PendingReview**: The model card is pending review.
- **Approved**: The model card is approved.
- **Archived**: The model card is archived. No more updates should be made to the model card, but it can still be exported.

Type: String

Valid Values: Draft | PendingReview | Approved | Archived

**ModelCardVersion (p. 513)**

The version of the model card.

Type: Integer

**SecurityConfig (p. 513)**

The security configuration used to protect model card content.

Type: ModelCardSecurityConfig (p. 1670) object

**Errors**

For information about the errors that are common to all actions, see Common Errors (p. 2180).
ResourceNotFound

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DescribeModelCardExportJob
Service: Amazon SageMaker Service

Describes an Amazon SageMaker Model Card export job.

Request Syntax

```
{
   "ModelCardExportJobArn": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**ModelCardExportJobArn (p. 517)**

The Amazon Resource Name (ARN) of the model card export job to describe.

Type: String

Length Constraints: Maximum length of 256.

Pattern: ^arn:aws[a-z-]*:sagemaker:[a-z0-9-\-]{9,16}:[0-9]{12}:model-card/[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}/export-job/[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

Required: Yes

Response Syntax

```
{
   "CreatedAt": number,
   "ExportArtifacts": {
      "$S3ExportArtifacts": "string"
   },
   "FailureReason": "string",
   "LastModifiedAt": number,
   "ModelCardExportJobArn": "string",
   "ModelCardExportJobName": "string",
   "ModelCardName": "string",
   "ModelCardVersion": number,
   "OutputConfig": {
      "$S3OutputPath": "string"
   },
   "Status": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.
**CreatedAt (p. 517)**

The date and time that the model export job was created.

Type: Timestamp

**ExportArtifacts (p. 517)**

The exported model card artifacts.

Type: `ModelCardExportArtifacts (p. 1666)` object

**FailureReason (p. 517)**

The failure reason if the model export job fails.

Type: String

Length Constraints: Maximum length of 1024.

**LastModifiedAt (p. 517)**

The date and time that the model export job was last modified.

Type: Timestamp

**ModelCardExportJobArn (p. 517)**

The Amazon Resource Name (ARN) of the model card export job.

Type: String

Length Constraints: Maximum length of 256.

Pattern: `^arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]{9,16}:[0-9]{12}:model-card/[a-zA-Z0-9](\-*[a-zA-Z0-9])\{0,62}\/export-job/[a-zA-Z0-9](\-*[a-zA-Z0-9])\{0,62}$`

**ModelCardExportJobName (p. 517)**

The name of the model card export job to describe.

Type: String


Pattern: `^[a-zA-Z0-9](\-*[a-zA-Z0-9])\{0,62}$`

**ModelCardName (p. 517)**

The name or Amazon Resource Name (ARN) of the model card that the model export job exports.

Type: String


Pattern: `^[a-zA-Z0-9](\-*[a-zA-Z0-9])\{0,62}$`

**ModelCardVersion (p. 517)**

The version of the model card that the model export job exports.

Type: Integer

**OutputConfig (p. 517)**

The export output details for the model card.

Type: `ModelCardExportOutputConfig (p. 1669)` object
**Status (p. 517)**

The completion status of the model card export job.

- **InProgress**: The model card export job is in progress.
- **Completed**: The model card export job is complete.
- **Failed**: The model card export job failed. To see the reason for the failure, see the `FailureReason` field in the response to a `DescribeModelCardExportJob` call.

Type: String

Valid Values: InProgress | Completed | Failed

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)].

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
DescribeModelExplainabilityJobDefinition

Service: Amazon SageMaker Service

Returns a description of a model explainability job definition.

Request Syntax

```
{
  "JobDefinitionName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see [Common Parameters](p. 2178).

The request accepts the following data in JSON format.

**JobDefinitionName (p. 520)**

The name of the model explainability job definition. The name must be unique within an AWS Region in the AWS account.

- Type: String
- Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$`
- Required: Yes

Response Syntax

```
{
  "CreationTime": number,
  "JobDefinitionArn": "string",
  "JobDefinitionName": "string",
  "JobResources": {
    "ClusterConfig": {
      "InstanceCount": number,
      "InstanceType": "string",
      "VolumeKmsKeyId": "string",
      "VolumeSizeInGB": number
    }
  },
  "ModelExplainabilityAppSpecification": {
    "ConfigUri": "string",
    "Environment": {
      "string": "string"
    },
    "ImageUri": "string"
  },
  "ModelExplainabilityBaselineConfig": {
    "BaseliningJobName": "string",
    "ConstraintsResource": {
      "S3Uri": "string"
    }
  },
  "ModelExplainabilityJobInput": {
    "BatchTransformInput": {
```

520
"DataCapturedDestinationS3Uri": "string",
"DatasetFormat": {
  "Csv": {
    "Header": boolean
  },
  "Json": {
    "Line": boolean
  },
  "Parquet": {
  }
},
"EndTimeOffset": "string",
"ExcludeFeaturesAttribute": "string",
"FeaturesAttribute": "string",
"InferenceAttribute": "string",
"LocalPath": "string",
"ProbabilityAttribute": "string",
"ProbabilityThresholdAttribute": number,
"S3DataDistributionType": "string",
"S3InputMode": "string",
"StartTimeOffset": "string"
},
"EndpointInput": {
  "EndpointName": "string",
  "EndTimeOffset": "string",
  "ExcludeFeaturesAttribute": "string",
  "FeaturesAttribute": "string",
  "InferenceAttribute": "string",
  "LocalPath": "string",
  "ProbabilityAttribute": "string",
  "ProbabilityThresholdAttribute": number,
  "S3DataDistributionType": "string",
  "S3InputMode": "string",
  "StartTimeOffset": "string"
}
"ModelExplainabilityJobOutputConfig": {
  "KmsKeyId": "string",
  "MonitoringOutputs": [
    {
      "S3Output": {
        "LocalPath": "string",
        "S3UploadMode": "string",
        "S3Uri": "string"
      }
    }
  ],
  "NetworkConfig": {
    "EnableInterContainerTrafficEncryption": boolean,
    "EnableNetworkIsolation": boolean,
    "VpcConfig": {
      "SecurityGroupIds": [ "string" ],
      "Subnets": [ "string" ]
    }
  }},
  "RoleArn": "string",
  "StoppingCondition": {
    "MaxRuntimeInSeconds": number
  }
}
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**CreationTime (p. 520)**

The time at which the model explainability job was created.

Type: Timestamp

**JobDefinitionArn (p. 520)**

The Amazon Resource Name (ARN) of the model explainability job.

Type: String

Length Constraints: Maximum length of 256.

Pattern: .*

**JobDefinitionName (p. 520)**

The name of the explainability job definition. The name must be unique within an AWS Region in the AWS account.

Type: String


Pattern: ^[a-zA-Z0-9-][a-zA-Z0-9-]{0,62}$

**JobResources (p. 520)**

Identifies the resources to deploy for a monitoring job.

Type: MonitoringResources (p. 1756) object

**ModelExplainabilityAppSpecification (p. 520)**

Configures the model explainability job to run a specified Docker container image.

Type: ModelExplainabilityAppSpecification (p. 1692) object

**ModelExplainabilityBaselineConfig (p. 520)**

The baseline configuration for a model explainability job.

Type: ModelExplainabilityBaselineConfig (p. 1693) object

**ModelExplainabilityJobInput (p. 520)**

Inputs for the model explainability job.

Type: ModelExplainabilityJobInput (p. 1694) object

**ModelExplainabilityJobOutputConfig (p. 520)**

The output configuration for monitoring jobs.

Type: MonitoringOutputConfig (p. 1754) object

**NetworkConfig (p. 520)**

Networking options for a model explainability job.
Type: `MonitoringNetworkConfig (p. 1752)` object

**RoleArn (p. 520)**

The Amazon Resource Name (ARN) of the IAM role that has read permission to the input data location and write permission to the output data location in Amazon S3.

Type: String


Pattern: `^arn:aws[a-z\-]*:iam::\d{12}:role/?[a-zA-Z_0-9+=,.@-_\//]+$`

**StoppingCondition (p. 520)**

A time limit for how long the monitoring job is allowed to run before stopping.

Type: `MonitoringStoppingCondition (p. 1765)` object

Errors

For information about the errors that are common to all actions, see `Common Errors (p. 2180)`.

**ResourceNotFoundException**

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
DescribeModelPackage
Service: Amazon SageMaker Service

Returns a description of the specified model package, which is used to create SageMaker models or list them on AWS Marketplace.

To create models in SageMaker, buyers can subscribe to model packages listed on AWS Marketplace.

Request Syntax

```
{
    "ModelPackageName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**ModelPackageName (p. 524)**

The name or Amazon Resource Name (ARN) of the model package to describe.

When you specify a name, the name must have 1 to 63 characters. Valid characters are a-z, A-Z, 0-9, and - (hyphen).

Type: String


Pattern: (arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:[a-z\-]*\/)?(\[a-zA-Z0-9\[a-zA-Z0-9\-]{}0,62))\(\(?<!\-\)(\[0-9]{1,5})\)?$ Required: Yes

Response Syntax

```
{
    "AdditionalInferenceSpecifications": [
    {
        "Containers": [
        {
            "AdditionalS3DataSource": {
                "CompressionType": "string",
                "S3DataType": "string",
                "S3Uri": "string"
            },
            "ContainerHostname": "string",
            "Environment": {
                "string": "string"
            },
            "Framework": "string",
            "FrameworkVersion": "string",
            "Image": "string",
            "ImageDigest": "string",
            "ModelDataUrl": "string",
```
"ModelInput": {
  "DataInputConfig": "string",
  "NearestModelName": "string",
  "ProductId": "string"
},
"Description": "string",
"Name": "string",
"SupportedContentTypes": [ "string" ],
"SupportedRealtimeInferenceInstanceTypes": [ "string" ],
"SupportedResponseMIMETypes": [ "string" ],
"SupportedTransformInstanceTypes": [ "string" ]
},
"ApprovalDescription": "string",
"CertifyForMarketplace": boolean,
"CreatedBy": {
  "DomainId": "string",
  "IamIdentity": {
    "Arn": "string",
    "PrincipalId": "string",
    "SourceIdentity": "string"
  },
  "UserProfileArn": "string",
  "UserProfileName": "string"
},
"CreationTime": number,
"CustomerMetadataProperties": {
  "string": "string"
},
"Domain": "string",
"DriftCheckBaselines": {
  "Bias": {
    "ConfigFile": {
      "ContentDigest": "string",
      "ContentType": "string",
      "S3Url": "string"
    },
    "PostTrainingConstraints": {
      "ContentDigest": "string",
      "ContentType": "string",
      "S3Url": "string"
    },
    "PreTrainingConstraints": {
      "ContentDigest": "string",
      "ContentType": "string",
      "S3Url": "string"
    }
  },
  "Explainability": {
    "ConfigFile": {
      "ContentDigest": "string",
      "ContentType": "string",
      "S3Url": "string"
    },
    "Constraints": {
      "ContentDigest": "string",
      "ContentType": "string",
      "S3Url": "string"
    }},
    "ModelDataQuality": {
      "Constraints": {
        "ContentDigest": "string",
        "ContentType": "string",
        "S3Url": "string"
      }
    }
  },
"S3Uri": "string",
},
"Statistics": {
 "ContentDigest": "string",
 "ContentType": "string",
 "S3Uri": "string"
}
},
"ModelQuality": {
 "Constraints": {
 "ContentDigest": "string",
 "ContentType": "string",
 "S3Uri": "string"
 },
 "Statistics": {
 "ContentDigest": "string",
 "ContentType": "string",
 "S3Uri": "string"
}
},
"InferenceSpecification": {
 "Containers": [
 { "AdditionalS3DataSource": {
 "CompressionType": "string",
 "S3DataType": "string",
 "S3Uri": "string"
 },
 "ContainerHostname": "string",
 "Environment": {
 "string": "string"
 },
 "Framework": "string",
 "FrameworkVersion": "string",
 "Image": "string",
 "ImageDigest": "string",
 "ModelDataUrl": "string",
 "ModelInput": {
 "DataInputConfig": "string"
 },
 "NearestModelName": "string",
 "ProductId": "string"
 } ],
 "SupportedContentTypes": [ "string" ],
 "SupportedRealtimeInferenceInstanceTypes": [ "string" ],
 "SupportedResponseMIMETypes": [ "string" ],
 "SupportedTransformInstanceTypes": [ "string" ]
},
"LastModifiedBy": {
 "DomainId": "string",
 "IamIdentity": {
 "Arn": "string",
 "PrincipalId": "string",
 "SourceIdentity": "string"
 },
 "UserProfileArn": "string",
 "UserProfileName": "string"
 },
 "LastModifiedTime": number,
 "MetadataProperties": {
 "CommitId": "string",
 "GeneratedBy": "string",
 "ProjectId": "string",
 "Repository": "string"
}
"ModelApprovalStatus": "string",
"ModelMetrics": {
  "Bias": {
    "PostTrainingReport": {
      "ContentDigest": "string",
      "ContentType": "string",
      "S3Uri": "string"
    },
    "PreTrainingReport": {
      "ContentDigest": "string",
      "ContentType": "string",
      "S3Uri": "string"
    },
    "Report": {
      "ContentDigest": "string",
      "ContentType": "string",
      "S3Uri": "string"
    }
  },
  "Explainability": {
    "Report": {
      "ContentDigest": "string",
      "ContentType": "string",
      "S3Uri": "string"
    }
  },
  "ModelDataQuality": {
    "Constraints": {
      "ContentDigest": "string",
      "ContentType": "string",
      "S3Uri": "string"
    },
    "Statistics": {
      "ContentDigest": "string",
      "ContentType": "string",
      "S3Uri": "string"
    }
  },
  "ModelQuality": {
    "Constraints": {
      "ContentDigest": "string",
      "ContentType": "string",
      "S3Uri": "string"
    },
    "Statistics": {
      "ContentDigest": "string",
      "ContentType": "string",
      "S3Uri": "string"
    }
  }
},
"ModelPackageArn": "string",
"ModelPackageDescription": "string",
"ModelPackageGroupName": "string",
"ModelPackageName": "string",
"ModelPackageStatus": "string",
"ModelPackageStatusDetails": {
  "ImageScanStatuses": [
    {
      "FailureReason": "string",
      "Name": "string",
      "Status": "string"
    }
  ],
  "ValidationStatuses": [
    "string"
  ]
}
{ "FailureReason": "string", "Name": "string", "Status": "string" }

"ModelPackageVersion": number,
"SamplePayloadUrl": "string",
"SkipModelValidation": "string",
"SourceAlgorithmSpecification": { "SourceAlgorithms": [ { "AlgorithmName": "string", "ModelDataUrl": "string" } ] },
"Task": "string",
"ValidationSpecification": { "ValidationProfiles": [ { "ProfileName": "string", "TransformJobDefinition": { "BatchStrategy": "string", "Environment": { "string": "string" }, "MaxConcurrentTransforms": number, "MaxPayloadInMB": number, "TransformInput": { "CompressionType": "string", "ContentType": "string", "DataSource": { "S3DataSource": { "S3DataType": "string", "S3Uri": "string" } } }, "SplitType": "string" }, "TransformOutput": { "Accept": "string", "AssembleWith": "string", "KmsKeyId": "string", "S3OutputPath": "string" } }, "TransformResources": { "InstanceCount": number, "InstanceType": "string", "VolumeKmsKeyId": "string" } } ] },
"ValidationRole": "string" }

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.
**AdditionalInferenceSpecifications (p. 524)**

An array of additional Inference Specification objects. Each additional Inference Specification specifies artifacts based on this model package that can be used on inference endpoints. Generally used with SageMaker Neo to store the compiled artifacts.

Type: Array of [AdditionalInferenceSpecificationDefinition (p. 1219)] objects

Array Members: Minimum number of 1 item. Maximum number of 15 items.

**ApprovalDescription (p. 524)**

A description provided for the model approval.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: .*

**CertifyForMarketplace (p. 524)**

Whether the model package is certified for listing on AWS Marketplace.

Type: Boolean

**CreatedBy (p. 524)**

Information about the user who created or modified an experiment, trial, trial component, lineage group, project, or model card.

Type: [UserContext (p. 2067)] object

**CreationTime (p. 524)**

A timestamp specifying when the model package was created.

Type: Timestamp

**CustomerMetadataProperties (p. 524)**

The metadata properties associated with the model package versions.

Type: String to string map

Map Entries: Maximum number of 50 items.

Key Length Constraints: Minimum length of 1. Maximum length of 128.

Key Pattern: ^([\p{L}\p{Z}\p{N}_.:/=+\-@]*)${1,128}$

Value Length Constraints: Minimum length of 1. Maximum length of 256.

Value Pattern: ^([\p{L}\p{Z}\p{N}_.:/=+\-@]*)${1,256}$

**Domain (p. 524)**

The machine learning domain of the model package you specified. Common machine learning domains include computer vision and natural language processing.

Type: String

**DriftCheckBaselines (p. 524)**

Represents the drift check baselines that can be used when the model monitor is set using the model package. For more information, see the topic on [Drift Detection against Previous Baselines in SageMaker Pipelines](#) in the Amazon SageMaker Developer Guide.
Type: `DriftCheckBaselines (p. 1426)` object

**InferenceSpecification (p. 524)**

Details about inference jobs that can be run with models based on this model package.

Type: `InferenceSpecification (p. 1601)` object

**LastModifiedBy (p. 524)**

Information about the user who created or modified an experiment, trial, trial component, lineage group, project, or model card.

Type: `UserContext (p. 2067)` object

**LastModifiedTime (p. 524)**

The last time that the model package was modified.

Type: `Timestamp`

**MetadataProperties (p. 524)**

Metadata properties of the tracking entity, trial, or trial component.

Type: `MetadataProperties (p. 1648)` object

**ModelApprovalStatus (p. 524)**

The approval status of the model package.

Type: `String`

Valid Values: `Approved` | `Rejected` | `PendingManualApproval`

**ModelMetrics (p. 524)**

Metrics for the model.

Type: `ModelMetrics (p. 1701)` object

**ModelPackageArn (p. 524)**

The Amazon Resource Name (ARN) of the model package.

Type: `String`

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(-cn|-us-gov)?:sagemaker:[a-z0-9-]{9,16}:[0-9]{12}:model-package/[^\$]*$`

**ModelPackageDescription (p. 524)**

A brief summary of the model package.

Type: `String`

Length Constraints: Maximum length of 1024.

Pattern: `[^p{L}|p{M}|p{Z}|p{S}|p{N}|p{P}]`*
**ModelPackageName (p. 524)**

The name of the model package being described.

Type: String


Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$`

**ModelPackageStatus (p. 524)**

The current status of the model package.

Type: String

Valid Values: Pending | InProgress | Completed | Failed | Deleting

**ModelPackageStatusDetails (p. 524)**

Details about the current status of the model package.

Type: `ModelPackageStatusDetails (p. 1714)` object

**ModelPackageVersion (p. 524)**

The version of the model package.

Type: Integer

Valid Range: Minimum value of 1.

**SamplePayloadUrl (p. 524)**

The Amazon Simple Storage Service (Amazon S3) path where the sample payload are stored. This path points to a single gzip compressed tar archive (.tar.gz suffix).

Type: String

**SkipModelValidation (p. 524)**

Indicates if you want to skip model validation.

Type: String

Valid Values: All | None

**SourceAlgorithmSpecification (p. 524)**

Details about the algorithm that was used to create the model package.

Type: `SourceAlgorithmSpecification (p. 1955)` object

**Task (p. 524)**

The machine learning task you specified that your model package accomplishes. Common machine learning tasks include object detection and image classification.

Type: String

**ValidationSpecification (p. 524)**

Configurations for one or more transform jobs that SageMaker runs to test the model package.
Type: ModelPackageValidationSpecification (p. 1719) object

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DescribeModelPackageGroup
Service: Amazon SageMaker Service

Gets a description for the specified model group.

Request Syntax

```json
{
  "ModelPackageGroupName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**ModelPackageGroupName (p. 533)**

The name of the model group to describe.

- **Type:** String
- **Length Constraints:** Minimum length of 1. Maximum length of 170.
- **Pattern:** `(arn:aws[a-zA-Z-]*:sagemaker:[a-zA-Z-]*:[0-9]{12}:[a-zA-Z-]*\/)?([a-zA-Z0-9][a-zA-Z0-9-]*){0,62}(?!-)$`
- **Required:** Yes

Response Syntax

```json
{
  "CreatedBy": {
    "DomainId": "string",
    "IamIdentity": {
      "Arn": "string",
      "PrincipalId": "string",
      "SourceIdentity": "string"
    },
    "UserProfileArn": "string",
    "UserProfileName": "string"
  },
  "CreationTime": number,
  "ModelPackageGroupArn": "string",
  "ModelPackageGroupDescription": "string",
  "ModelPackageGroupName": "string",
  "ModelPackageGroupStatus": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.
CreatedBy (p. 533)
Information about the user who created or modified an experiment, trial, trial component, lineage group, project, or model card.
Type: UserContext (p. 2067) object

CreationTime (p. 533)
The time that the model group was created.
Type: Timestamp

ModelPackageGroupArn (p. 533)
The Amazon Resource Name (ARN) of the model group.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 2048.
Pattern: ^arn:aws(-cn|-us-gov)?:sagemaker:[a-z0-9\-]{9,16}:[0-9]\{12\}:model-package-group/\[\S\]{1,2048}$

ModelPackageGroupDescription (p. 533)
A description of the model group.
Type: String
Length Constraints: Maximum length of 1024.
Pattern: [\p{L}\p{M}\p{Z}\p{S}\p{N}\p{P}]*

ModelPackageGroupName (p. 533)
The name of the model group.
Type: String
Pattern: ^[a-zA-Z0-9\-0-9\-]{0,62}$

ModelPackageGroupStatus (p. 533)
The status of the model group.
Type: String
Valid Values: Pending | InProgress | Completed | Failed | Deleting | DeleteFailed

Errors
For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DescribeModelQualityJobDefinition

Returns a description of a model quality job definition.

Request Syntax

```
{
    "JobDefinitionName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**JobDefinitionName (p. 536)**

The name of the model quality job. The name must be unique within an AWS Region in the AWS account.

Type: String


Pattern: ^[a-zA-Z0-9-]*[a-zA-Z0-9]$ \{0,62\}$

Required: Yes

Response Syntax

```
{
    "CreationTime": number,
    "JobDefinitionArn": "string",
    "JobDefinitionName": "string",
    "JobResources": {
        "ClusterConfig": {
            "InstanceCount": number,
            "InstanceType": "string",
            "VolumeKmsKeyId": "string",
            "VolumeSizeInGB": number
        }
    },
    "ModelQualityAppSpecification": {
        "ContainerArguments": [ "string" ],
        "ContainerEntrypoint": [ "string" ],
        "Environment": {
            "string": "string"
        },
        "ImageUri": "string",
        "PostAnalyticsProcessorSourceUri": "string",
        "ProblemType": "string",
        "RecordPreprocessorSourceUri": "string"
    },
    "ModelQualityBaselineConfig": {
        "BaseliningJobName": "string",
```
"ConstraintsResource": {
  "S3Uri": "string"
},
"ModelQualityJobInput": {
  "BatchTransformInput": {
    "DataCapturedDestinationS3Uri": "string",
    "DatasetFormat": {
      "Csv": {
        "Header": boolean
      },
      "Json": {
        "Line": boolean
      },
      "Parquet": {
      }
    },
    "EndTimeOffset": "string",
    "ExcludeFeaturesAttribute": "string",
    "FeaturesAttribute": "string",
    "InferenceAttribute": "string",
    "LocalPath": "string",
    "ProbabilityAttribute": "string",
    "ProbabilityThresholdAttribute": number,
    "S3DataDistributionType": "string",
    "S3InputMode": "string",
    "StartTimeOffset": "string"
  },
  "EndpointInput": {
    "EndpointName": "string",
    "EndTimeOffset": "string",
    "ExcludeFeaturesAttribute": "string",
    "FeaturesAttribute": "string",
    "InferenceAttribute": "string",
    "LocalPath": "string",
    "ProbabilityAttribute": "string",
    "ProbabilityThresholdAttribute": number,
    "S3DataDistributionType": "string",
    "S3InputMode": "string",
    "StartTimeOffset": "string"
  },
  "GroundTruthS3Input": {
    "S3Uri": "string"
  }
},
"ModelQualityJobOutputConfig": {
  "KmsKeyId": "string",
  "MonitoringOutputs": [
    {
      "S3Output": {
        "LocalPath": "string",
        "S3UploadMode": "string",
        "S3Uri": "string"
      }
    }
  ],
  "NetworkConfig": {
    "EnableInterContainerTrafficEncryption": boolean,
    "EnableNetworkIsolation": boolean,
    "VpcConfig": {
      "SecurityGroupIds": [ "string" ],
      "Subnets": [ "string" ]
    }
  },
  "RoleArn": "string"}
"StoppingCondition": {
  "MaxRuntimeInSeconds": number
}

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**CreationTime (p. 536)**

The time at which the model quality job was created.

Type: Timestamp

**JobDefinitionArn (p. 536)**

The Amazon Resource Name (ARN) of the model quality job.

Type: String

Length Constraints: Maximum length of 256.

Pattern: .*

**JobDefinitionName (p. 536)**

The name of the quality job definition. The name must be unique within an AWS Region in the AWS account.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}$

**JobResources (p. 536)**

Identifies the resources to deploy for a monitoring job.

Type: `MonitoringResources (p. 1756)` object

**ModelQualityAppSpecification (p. 536)**

Configures the model quality job to run a specified Docker container image.

Type: `ModelQualityAppSpecification (p. 1721)` object

**ModelQualityBaselineConfig (p. 536)**

The baseline configuration for a model quality job.

Type: `ModelQualityBaselineConfig (p. 1723)` object

**ModelQualityJobInput (p. 536)**

Inputs for the model quality job.

Type: `ModelQualityJobInput (p. 1724)` object

**ModelQualityJobOutputConfig (p. 536)**

The output configuration for monitoring jobs.
Type: **MonitoringOutputConfig** *(p. 1754)* object

**NetworkConfig** *(p. 536)*

Networking options for a model quality job.

Type: **MonitoringNetworkConfig** *(p. 1752)* object

**RoleArn** *(p. 536)*

The Amazon Resource Name (ARN) of an IAM role that Amazon SageMaker can assume to perform tasks on your behalf.

Type: String


Pattern: `^arn:aws[a-zA-Z-]*:iam::\d{12}:role/?[a-zA-Z0-9+=,.@-/]+$`

**StoppingCondition** *(p. 536)*

A time limit for how long the monitoring job is allowed to run before stopping.

Type: **MonitoringStoppingCondition** *(p. 1765)* object

**Errors**

For information about the errors that are common to all actions, see *Common Errors* *(p. 2180).*

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DescribeMonitoringSchedule

Service: Amazon SageMaker Service

Describes the schedule for a monitoring job.

Request Syntax

```
{
    "MonitoringScheduleName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**MonitoringScheduleName (p. 540)**

Name of a previously created monitoring schedule.

Type: String


Pattern: ^[a-zA-Z0-9\-]*[a-zA-Z0-9]{0,62}$

Required: Yes

Response Syntax

```
{
    "CreationTime": number,
    "EndpointName": "string",
    "FailureReason": "string",
    "LastModifiedTime": number,
    "LastMonitoringExecutionSummary": {
        "CreationTime": number,
        "EndpointName": "string",
        "FailureReason": "string",
        "LastModifiedTime": number,
        "MonitoringExecutionStatus": "string",
        "MonitoringJobDefinitionName": "string",
        "MonitoringScheduleName": "string",
        "MonitoringType": "string",
        "ProcessingJobArn": "string",
        "ScheduledTime": number
    },
    "MonitoringScheduleArn": "string",
    "MonitoringScheduleConfig": {
        "BaselineConfig": {
            "BaselineJobName": "string",
            "ConstraintsResource": {
                "S3Uri": "string"
            },
            "StatisticsResource": {
                "S3Uri": "string"
            }
        }
    }
}
```
"Environment": {  
  "string": "string"
},
"MonitoringAppSpecification": {  
  "ContainerArguments": [ "string" ],  
  "ContainerEntrypoint": [ "string" ],  
  "ImageUri": "string",  
  "PostAnalyticsProcessorSourceUri": "string",  
  "RecordPreprocessorSourceUri": "string"
},
"MonitoringInputs": [
  {  
    "BatchTransformInput": {  
      "DataCapturedDestinationS3Uri": "string",  
      "DatasetFormat": {  
        "Csv": {  
          "Header": boolean
        },  
        "Json": {  
          "Line": boolean
        },  
        "Parquet": {  
        }
      },  
      "EndTimeOffset": "string",  
      "ExcludeFeaturesAttribute": "string",  
      "FeaturesAttribute": "string",  
      "InferenceAttribute": "string",  
      "LocalPath": "string",  
      "ProbabilityAttribute": "string",  
      "ProbabilityThresholdAttribute": number,  
      "S3DataDistributionType": "string",  
      "S3InputMode": "string",  
      "StartTimeOffset": "string"
    },  
    "EndpointInput": {  
      "EndpointName": "string",  
      "EndTimeOffset": "string",  
      "ExcludeFeaturesAttribute": "string",  
      "FeaturesAttribute": "string",  
      "InferenceAttribute": "string",  
      "LocalPath": "string",  
      "ProbabilityAttribute": "string",  
      "ProbabilityThresholdAttribute": number,  
      "S3DataDistributionType": "string",  
      "S3InputMode": "string",  
      "StartTimeOffset": "string"
    }
  }
},
"MonitoringOutputConfig": {  
  "KmsKeyId": "string",  
  "MonitoringOutputs": [
    {  
      "S3Output": {  
        "LocalPath": "string",  
        "S3UploadMode": "string",  
        "S3Uri": "string"
      }
    }
  ]
},
"MonitoringResources": {
  "ClusterConfig": {
}}
"InstanceCount": number,
"InstanceType": "string",
"VolumeKmsKeyId": "string",
"VolumeSizeInGB": number
}
"NetworkConfig": {
"EnableInterContainerTrafficEncryption": boolean,
"EnableNetworkIsolation": boolean,
"VpcConfig": {
"SecurityGroupIds": [ "string" ],
"Subnets": [ "string" ]
}
},
"RoleArn": "string",
"StoppingCondition": {
"MaxRuntimeInSeconds": number
}
},
"MonitoringJobDefinitionName": "string",
"MonitoringType": "string",
"ScheduleConfig": {
"DataAnalysisEndTime": "string",
"DataAnalysisStartTime": "string",
"ScheduleExpression": "string"
}
},
"MonitoringScheduleName": "string",
"MonitoringScheduleStatus": "string",
"MonitoringType": "string"
}

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**CreationTime (p. 540)**

The time at which the monitoring job was created.

Type: Timestamp

**EndpointName (p. 540)**

The name of the endpoint for the monitoring job.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

**FailureReason (p. 540)**

A string, up to one KB in size, that contains the reason a monitoring job failed, if it failed.

Type: String

Length Constraints: Maximum length of 1024.

**LastModifiedTime (p. 540)**

The time at which the monitoring job was last modified.
Type: Timestamp

**LastMonitoringExecutionSummary (p. 540)**

Describes metadata on the last execution to run, if there was one.

Type: **MonitoringExecutionSummary (p. 1743)** object

**MonitoringScheduleArn (p. 540)**

The Amazon Resource Name (ARN) of the monitoring schedule.

Type: String

Length Constraints: Maximum length of 256.

Pattern: .*

**MonitoringScheduleConfig (p. 540)**

The configuration object that specifies the monitoring schedule and defines the monitoring job.

Type: **MonitoringScheduleConfig (p. 1761)** object

**MonitoringScheduleName (p. 540)**

Name of the monitoring schedule.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

**MonitoringScheduleStatus (p. 540)**

The status of an monitoring job.

Type: String

Valid Values: Pending | Failed | Scheduled | Stopped

**MonitoringType (p. 540)**

The type of the monitoring job that this schedule runs. This is one of the following values.

- DATA_QUALITY - The schedule is for a data quality monitoring job.
- MODEL_QUALITY - The schedule is for a model quality monitoring job.
- MODEL_BIAS - The schedule is for a bias monitoring job.
- MODEL_EXPLAINABILITY - The schedule is for an explainability monitoring job.

Type: String

Valid Values: DataQuality | ModelQuality | ModelBias | ModelExplainability

**Errors**

For information about the errors that are common to all actions, see **Common Errors (p. 2180)**.

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400
See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DescribeNotebookInstance

Service: Amazon SageMaker Service

Returns information about a notebook instance.

Request Syntax

```
{
    "NotebookInstanceName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**NotebookInstanceName (p. 545)**

The name of the notebook instance that you want information about.

- Type: String
- Length Constraints: Maximum length of 63.
- Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])*
- Required: Yes

Response Syntax

```
{
    "AcceleratorTypes": [ "string" ],
    "AdditionalCodeRepositories": [ "string" ],
    "CreationTime": number,
    "DefaultCodeRepository": "string",
    "DirectInternetAccess": "string",
    "FailureReason": "string",
    "InstanceMetadataServiceConfiguration": {
        "MinimumInstanceMetadataServiceVersion": "string"
    },
    "InstanceType": "string",
    "KmsKeyId": "string",
    "LastModifiedTime": number,
    "NetworkInterfaceId": "string",
    "NotebookInstanceArn": "string",
    "NotebookInstanceLifecycleConfigName": "string",
    "NotebookInstanceName": "string",
    "NotebookInstanceStatus": "string",
    "PlatformIdentifier": "string",
    "RoleArn": "string",
    "RootAccess": "string",
    "SecurityGroups": [ "string" ],
    "SubnetId": "string",
    "Url": "string",
    "VolumeSizeInGB": number
}
```
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**AcceleratorTypes** (p. 545)

A list of the Elastic Inference (EI) instance types associated with this notebook instance. Currently only one EI instance type can be associated with a notebook instance. For more information, see Using Elastic Inference in Amazon SageMaker.

Type: Array of strings

Valid Values: ml.eia1.medium | ml.eia1.large | ml.eia1.xlarge | ml.eia2.medium | ml.eia2.large | ml.eia2.xlarge

**AdditionalCodeRepositories** (p. 545)

An array of up to three Git repositories associated with the notebook instance. These can be either the names of Git repositories stored as resources in your account, or the URL of Git repositories in AWS CodeCommit or in any other Git repository. These repositories are cloned at the same level as the default repository of your notebook instance. For more information, see Associating Git Repositories with SageMaker Notebook Instances.

Type: Array of strings

Array Members: Maximum number of 3 items.


Pattern: ^https://([^/]+)/?(.*)$|^[a-zA-Z0-9](-*[a-zA-Z0-9])*

**CreationTime** (p. 545)

A timestamp. Use this parameter to return the time when the notebook instance was created

Type: Timestamp

**DefaultCodeRepository** (p. 545)

The Git repository associated with the notebook instance as its default code repository. This can be either the name of a Git repository stored as a resource in your account, or the URL of a Git repository in AWS CodeCommit or in any other Git repository. When you open a notebook instance, it opens in the directory that contains this repository. For more information, see Associating Git Repositories with SageMaker Notebook Instances.

Type: String


Pattern: ^https://([^/]+)/?(.*)$|^[a-zA-Z0-9](-*[a-zA-Z0-9])*

**DirectInternetAccess** (p. 545)

Describes whether SageMaker provides internet access to the notebook instance. If this value is set to **Disabled**, the notebook instance does not have internet access, and cannot connect to SageMaker training and endpoint services.

For more information, see Notebook Instances Are Internet-Enabled by Default.

Type: String
Valid Values: Enabled | Disabled

**FailureReason (p. 545)**

If status is Failed, the reason it failed.

Type: String

Length Constraints: Maximum length of 1024.

**InstanceMetadataServiceConfiguration (p. 545)**

Information on the IMDS configuration of the notebook instance

Type: `InstanceMetadataServiceConfiguration (p. 1611)` object

**InstanceType (p. 545)**

The type of ML compute instance running on the notebook instance.

Type: String

Valid Values:
- `ml.t2.medium`
- `ml.t2.large`
- `ml.t2.xlarge`
- `ml.t2.2xlarge`
- `ml.t3.medium`
- `ml.t3.large`
- `ml.t3.xlarge`
- `ml.t3.2xlarge`
- `ml.m4.xlarge`
- `ml.m4.2xlarge`
- `ml.m4.4xlarge`
- `ml.m4.10xlarge`
- `ml.m4.16xlarge`
- `ml.m5.xlarge`
- `ml.m5.2xlarge`
- `ml.m5.xlarge`
- `ml.m5.2xlarge`
- `ml.m5.4xlarge`
- `ml.m5.12xlarge`
- `ml.m5.24xlarge`
- `ml.m5d.large`
- `ml.m5d.xlarge`
- `ml.m5d.2xlarge`
- `ml.m5d.4xlarge`
- `ml.m5d.8xlarge`
- `ml.m5d.12xlarge`
- `ml.m5d.16xlarge`
- `ml.m5d.24xlarge`
- `ml.c4.xlarge`
- `ml.c4.2xlarge`
- `ml.c4.4xlarge`
- `ml.c4.8xlarge`
- `ml.c5.xlarge`
- `ml.c5.2xlarge`
- `ml.c5.4xlarge`
- `ml.c5.9xlarge`
- `ml.c5.18xlarge`
- `ml.c5d.xlarge`
- `ml.c5d.2xlarge`
- `ml.c5d.4xlarge`
- `ml.c5d.9xlarge`
- `ml.c5d.18xlarge`
- `ml.c5d.32xlarge`
- `ml.p2.xlarge`
- `ml.p2.8xlarge`
- `ml.p3.2xlarge`
- `ml.p3.8xlarge`
- `ml.p3.16xlarge`
- `ml.p3dn.2xlarge`
- `ml.p3dn.8xlarge`
- `ml.p3dn.16xlarge`
- `ml.p4dn.8xlarge`
- `ml.p4dn.16xlarge`
- `ml.p4dn.32xlarge`
- `ml.p4d.16xlarge`
- `ml.p4d.32xlarge`
- `ml.p4d.64xlarge`
- `ml.g4dn.xlarge`
- `ml.g4dn.2xlarge`
- `ml.g4dn.4xlarge`
- `ml.g4dn.8xlarge`
- `ml.g4dn.16xlarge`
- `ml.g4dn.32xlarge`
- `ml.g5.xlarge`
- `ml.g5.2xlarge`
- `ml.g5.4xlarge`
- `ml.g5.8xlarge`
- `ml.g5.16xlarge`
- `ml.g5.24xlarge`
- `ml.g5.48xlarge`
- `ml.g5.96xlarge`
- `ml.g5.192xlarge`
- `ml.g5.384xlarge`
- `ml.g5d.xlarge`
- `ml.g5d.2xlarge`
- `ml.g5d.4xlarge`
- `ml.g5d.8xlarge`
- `ml.g5d.16xlarge`
- `ml.g5d.32xlarge`
- `ml.g5d.64xlarge`
- `ml.g5d.128xlarge`
- `ml.g5d.256xlarge`
- `ml.g5d.512xlarge`
- `ml.g5d.1024xlarge`
- `ml.g5d.2048xlarge`
- `ml.p2.xlarge`
- `ml.p2.8xlarge`
- `ml.p2.16xlarge`
- `ml.p2.32xlarge`
- `ml.p2.64xlarge`
- `ml.p2.128xlarge`
- `ml.p2.256xlarge`
- `ml.p2.512xlarge`
- `ml.p2.1024xlarge`
- `ml.p2.2048xlarge`
- `ml.p3.xlarge`
- `ml.p3.2xlarge`
- `ml.p3.8xlarge`
- `ml.p3.16xlarge`
- `ml.p3.32xlarge`
- `ml.p3.64xlarge`
- `ml.p3.128xlarge`
- `ml.p3.256xlarge`
- `ml.p3.512xlarge`
- `ml.p3.1024xlarge`
- `ml.p3.2048xlarge`
- `ml.p3dn.xlarge`
- `ml.p3dn.2xlarge`
- `ml.p3dn.8xlarge`
- `ml.p3dn.16xlarge`
- `ml.p3dn.32xlarge`
- `ml.p3dn.64xlarge`
- `ml.p3dn.128xlarge`
- `ml.p3dn.256xlarge`
- `ml.p3dn.512xlarge`
- `ml.p3dn.1024xlarge`
- `ml.p3dn.2048xlarge`
- `ml.g4.xlarge`
- `ml.g4.2xlarge`
- `ml.g4.4xlarge`
- `ml.g4.8xlarge`
- `ml.g4.16xlarge`
- `ml.g4.32xlarge`
- `ml.g4.64xlarge`
- `ml.g4.128xlarge`
- `ml.g4.256xlarge`
- `ml.g4.512xlarge`
- `ml.g4.1024xlarge`
- `ml.g4.2048xlarge`
- `ml.g4d.xlarge`
- `ml.g4d.2xlarge`
- `ml.g4d.4xlarge`
- `ml.g4d.8xlarge`
- `ml.g4d.16xlarge`
- `ml.g4d.32xlarge`
- `ml.g4d.64xlarge`
- `ml.g4d.128xlarge`
- `ml.g4d.256xlarge`
- `ml.g4d.512xlarge`
- `ml.g4d.1024xlarge`
- `ml.g4d.2048xlarge`
- `ml.g5.xlarge`
- `ml.g5.2xlarge`
- `ml.g5.4xlarge`
- `ml.g5.8xlarge`
- `ml.g5.16xlarge`
- `ml.g5.32xlarge`
- `ml.g5.64xlarge`
- `ml.g5.128xlarge`
- `ml.g5.256xlarge`
- `ml.g5.512xlarge`
- `ml.g5.1024xlarge`
- `ml.g5.2048xlarge`
- `ml.g5d.xlarge`
- `ml.g5d.2xlarge`
- `ml.g5d.4xlarge`
- `ml.g5d.8xlarge`
- `ml.g5d.16xlarge`
- `ml.g5d.32xlarge`
- `ml.g5d.64xlarge`
- `ml.g5d.128xlarge`
- `ml.g5d.256xlarge`
- `ml.g5d.512xlarge`
- `ml.g5d.1024xlarge`
- `ml.g5d.2048xlarge`
- `ml.p4.xlarge`
- `ml.p4.8xlarge`
- `ml.p4.16xlarge`
- `ml.p4.32xlarge`
- `ml.p4.64xlarge`
- `ml.p4.128xlarge`
- `ml.p4.256xlarge`
- `ml.p4.512xlarge`
- `ml.p4.1024xlarge`
- `ml.p4.2048xlarge`
- `ml.p4d.xlarge`
- `ml.p4d.8xlarge`
- `ml.p4d.16xlarge`
- `ml.p4d.32xlarge`
- `ml.p4d.64xlarge`
- `ml.p4d.128xlarge`
- `ml.p4d.256xlarge`
- `ml.p4d.512xlarge`
- `ml.p4d.1024xlarge`
- `ml.p4d.2048xlarge`

**KmsKeyId (p. 545)**

The AWS KMS key ID SageMaker uses to encrypt data when storing it on the ML storage volume attached to the instance.

Type: String

Length Constraints: Maximum length of 2048.

Pattern: .*

**LastModifiedTime (p. 545)**

A timestamp. Use this parameter to retrieve the time when the notebook instance was last modified.

Type: Timestamp

**NetworkInterfaceId (p. 545)**

The network interface IDs that SageMaker created at the time of creating the instance.

Type: String
**NotebookInstanceArn (p. 545)**

The Amazon Resource Name (ARN) of the notebook instance.

Type: String

Length Constraints: Maximum length of 256.

**NotebookInstanceLifecycleConfigName (p. 545)**

Returns the name of a notebook instance lifecycle configuration.

For information about notebook instance lifestyle configurations, see [Step 2.1: (Optional) Customize a Notebook Instance](#).  

Type: String

Length Constraints: Maximum length of 63.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9])*`  

**NotebookInstanceName (p. 545)**

The name of the SageMaker notebook instance.

Type: String

Length Constraints: Maximum length of 63.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9])*`  

**NotebookInstanceStatus (p. 545)**

The status of the notebook instance.

Type: String

Valid Values: Pending | InService | Stopping | Stopped | Failed | Deleting | Updating

**PlatformIdentifier (p. 545)**

The platform identifier of the notebook instance runtime environment.

Type: String

Length Constraints: Maximum length of 15.

Pattern: `^(notebook-al1-v1|notebook-al2-v1|notebook-al2-v2)$`  

**RoleArn (p. 545)**

The Amazon Resource Name (ARN) of the IAM role associated with the instance.

Type: String


Pattern: `^arn:aws[a-zA-Z\-]*:iam::\d{12}:role/?[a-zA-Z0-9+=,.@\-_\/]\+\+$`  

**RootAccess (p. 545)**

Whether root access is enabled or disabled for users of the notebook instance.

**Note**  
Lifecycle configurations need root access to be able to set up a notebook instance. Because of this, lifecycle configurations associated with a notebook instance always run with root access even if you disable root access for users.
Type: String
Valid Values: Enabled | Disabled

SecurityGroups (p. 545)
The IDs of the VPC security groups.
Type: Array of strings
Array Members: Maximum number of 5 items.
Length Constraints: Maximum length of 32.
Pattern: [-0-9a-zA-Z]+

SubnetId (p. 545)
The ID of the VPC subnet.
Type: String
Length Constraints: Maximum length of 32.
Pattern: [-0-9a-zA-Z]+

Url (p. 545)
The URL that you use to connect to the Jupyter notebook that is running in your notebook instance.
Type: String

VolumeSizeInGB (p. 545)
The size, in GB, of the ML storage volume attached to the notebook instance.
Type: Integer

Errors
For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DescribeNotebookInstanceLifecycleConfig

Service: Amazon SageMaker Service

Returns a description of a notebook instance lifecycle configuration.

For information about notebook instance lifestyle configurations, see Step 2.1: (Optional) Customize a Notebook Instance.

Request Syntax

```
{
   "NotebookInstanceLifecycleConfigName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**NotebookInstanceLifecycleConfigName (p. 550)**

The name of the lifecycle configuration to describe.

Type: String
Length Constraints: Maximum length of 63.
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])* Required: Yes

Response Syntax

```
{
   "CreationTime": number,
   "LastModifiedTime": number,
   "NotebookInstanceLifecycleConfigArn": "string",
   "NotebookInstanceLifecycleConfigName": "string",
   "OnCreate": [
      {
         "Content": "string"
      }
   ],
   "OnStart": [
      {
         "Content": "string"
      }
   ]
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.
**CreationTime (p. 550)**

A timestamp that tells when the lifecycle configuration was created.

Type: Timestamp

**LastModifiedTime (p. 550)**

A timestamp that tells when the lifecycle configuration was last modified.

Type: Timestamp

**NotebookInstanceLifecycleConfigArn (p. 550)**

The Amazon Resource Name (ARN) of the lifecycle configuration.

Type: String

Length Constraints: Maximum length of 256.

**NotebookInstanceLifecycleConfigName (p. 550)**

The name of the lifecycle configuration.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])*    

**OnCreate (p. 550)**

The shell script that runs only once, when you create a notebook instance.

Type: Array of **NotebookInstanceLifecycleHook (p. 1771)** objects

Array Members: Maximum number of 1 item.

**OnStart (p. 550)**

The shell script that runs every time you start a notebook instance, including when you create the notebook instance.

Type: Array of **NotebookInstanceLifecycleHook (p. 1771)** objects

Array Members: Maximum number of 1 item.

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](#).

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
DescribePipeline
Service: Amazon SageMaker Service

Describes the details of a pipeline.

Request Syntax

```
{
  "PipelineName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**PipelineName** (p. 553)

The name or Amazon Resource Name (ARN) of the pipeline to describe.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `(arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:pipeline/.*)?([a-zA-Z0-9][\-\d[a-zA-Z0-9]]{0,255})`

Required: Yes

Response Syntax

```
{
  "CreatedBy": {
    "DomainId": "string",
    "IamIdentity": {
      "Arn": "string",
      "PrincipalId": "string",
      "SourceIdentity": "string"
    },
    "UserProfileArn": "string",
    "UserProfileName": "string"
  },
  "CreationTime": number,
  "LastModifiedBy": {
    "DomainId": "string",
    "IamIdentity": {
      "Arn": "string",
      "PrincipalId": "string",
      "SourceIdentity": "string"
    },
    "UserProfileArn": "string",
    "UserProfileName": "string"
  },
  "LastModifiedTime": number,
  "LastRunTime": number,
  "ParallelismConfiguration": {
```
"MaxParallelExecutionSteps": number

"PipelineArn": "string",
"PipelineDefinition": "string",
"PipelineDescription": "string",
"PipelineDisplayName": "string",
"PipelineName": "string",
"PipelineStatus": "string",
"RoleArn": "string"

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

CreatedBy (p. 553)

Information about the user who created or modified an experiment, trial, trial component, lineage group, project, or model card.

Type: UserContext (p. 2067) object

CreationTime (p. 553)

The time when the pipeline was created.

Type: Timestamp

LastModifiedBy (p. 553)

Information about the user who created or modified an experiment, trial, trial component, lineage group, project, or model card.

Type: UserContext (p. 2067) object

LastModifiedTime (p. 553)

The time when the pipeline was last modified.

Type: Timestamp

LastRunTime (p. 553)

The time when the pipeline was last run.

Type: Timestamp

ParallelismConfiguration (p. 553)

Lists the parallelism configuration applied to the pipeline.

Type: ParallelismConfiguration (p. 1797) object

PipelineArn (p. 553)

The Amazon Resource Name (ARN) of the pipeline.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:pipeline/.*
**PipelineDefinition (p. 553)**

The JSON pipeline definition.

Type: String


Pattern: .*\[ \r\n\t\].*\)

**PipelineDescription (p. 553)**

The description of the pipeline.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 3072.

Pattern: .*

**PipelineDisplayName (p. 553)**

The display name of the pipeline.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 256.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,255}¥

**PipelineName (p. 553)**

The name of the pipeline.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 256.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,255}¥

**PipelineStatus (p. 553)**

The status of the pipeline execution.

Type: String

Valid Values: Active

**RoleArn (p. 553)**

The Amazon Resource Name (ARN) that the pipeline uses to execute.

Type: String


Pattern: ^arn:aws[a-zA-Z-}*:iam::\d{12}:role/\[a-zA-Z0-9+.,@\-_/]+$\}

**Errors**

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceNotFoundException**

Resource being access is not found.
HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DescribePipelineDefinitionForExecution

Service: Amazon SageMaker Service

Describes the details of an execution's pipeline definition.

Request Syntax

```json
{
  "PipelineExecutionArn": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**PipelineExecutionArn (p. 557)**

The Amazon Resource Name (ARN) of the pipeline execution.

Type: String

Length Constraints: Maximum length of 256.

Pattern: ^arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:pipeline\./.*\./execution\./.*$

Required: Yes

Response Syntax

```json
{
  "CreationTime": number,
  "PipelineDefinition": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**CreationTime (p. 557)**

The time when the pipeline was created.

Type: Timestamp

**PipelineDefinition (p. 557)**

The JSON pipeline definition.

Type: String

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceNotFound

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DescribePipelineExecution
Service: Amazon SageMaker Service
Describes the details of a pipeline execution.

Request Syntax

```
{
   "PipelineExecutionArn": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**PipelineExecutionArn (p. 559)**

The Amazon Resource Name (ARN) of the pipeline execution.

Type: String

Length Constraints: Maximum length of 256.

Pattern: ^arn:aws[a-z-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:pipeline\/[.*\]/execution\/[.*\]$  

Required: Yes

Response Syntax

```
{
   "CreatedBy": {
      "DomainId": "string",
      "IamIdentity": {
         "Arn": "string",
         "PrincipalId": "string",
         "SourceIdentity": "string"
      },
      "UserProfileArn": "string",
      "UserProfileName": "string"
   },
   "CreationTime": number,
   "FailureReason": "string",
   "LastModifiedBy": {
      "DomainId": "string",
      "IamIdentity": {
         "Arn": "string",
         "PrincipalId": "string",
         "SourceIdentity": "string"
      },
      "UserProfileArn": "string",
      "UserProfileName": "string"
   },
   "LastModifiedTime": number,
   "ParallelismConfiguration": {
```

559
"MaxParallelExecutionSteps": number
"PipelineArn": "string",
"PipelineExecutionArn": "string",
"PipelineExecutionDescription": "string",
"PipelineExecutionDisplayName": "string",
"PipelineExecutionStatus": "string",
"PipelineExperimentConfig": {
  "ExperimentName": "string",
  "TrialName": "string"
},
"SelectiveExecutionConfig": {
  "SelectedSteps": [
    {
      "StepName": "string"
    }
  ],
  "SourcePipelineExecutionArn": "string"
}

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

CreatedBy (p. 559)

Information about the user who created or modified an experiment, trial, trial component, lineage group, project, or model card.

Type: UserContext (p. 2067) object

CreationTime (p. 559)

The time when the pipeline execution was created.

Type: Timestamp

FailureReason (p. 559)

If the execution failed, a message describing why.

Type: String

Length Constraints: Maximum length of 1300.

Pattern: .*

LastModifiedBy (p. 559)

Information about the user who created or modified an experiment, trial, trial component, lineage group, project, or model card.

Type: UserContext (p. 2067) object

LastModifiedTime (p. 559)

The time when the pipeline execution was modified last.

Type: Timestamp

ParallellismConfiguration (p. 559)

The parallelism configuration applied to the pipeline.
Type: ParallelismConfiguration (p. 1797) object

**PipelineArn (p. 559)**

The Amazon Resource Name (ARN) of the pipeline.

Type: String

Length Constraints: Maximum length of 256.

Pattern: `arn:aws[a-z-\]*:sagemaker:[a-z0-9-]*:[0-9]{12}:pipeline/.*`

**PipelineExecutionArn (p. 559)**

The Amazon Resource Name (ARN) of the pipeline execution.

Type: String

Length Constraints: Maximum length of 256.

Pattern: `^arn:aws[a-z-\]*:sagemaker:[a-z0-9-]*:[0-9]{12}:pipeline/.*\ execution/.*$`

**PipelineExecutionDescription (p. 559)**

The description of the pipeline execution.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 3072.

Pattern: `.*`

**PipelineExecutionDisplayName (p. 559)**

The display name of the pipeline execution.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 82.

Pattern: `^[a-zA-Z0-9-]*$`

**PipelineExecutionStatus (p. 559)**

The status of the pipeline execution.

Type: String

Valid Values: Executing | Stopping | Stopped | Failed | Succeeded

**PipelineExperimentConfig (p. 559)**

Specifies the names of the experiment and trial created by a pipeline.

Type: PipelineExperimentConfig (p. 1825) object

**SelectiveExecutionConfig (p. 559)**

The selective execution configuration applied to the pipeline run.

Type: SelectiveExecutionConfig (p. 1944) object

**Errors**

For information about the errors that are common to all actions, see Common Errors (p. 2180).
ResourceNotFound

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DescribeProcessingJob

Service: Amazon SageMaker Service

Returns a description of a processing job.

Request Syntax

```
{
  "ProcessingJobName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**ProcessingJobName (p. 563)**

The name of the processing job. The name must be unique within an AWS Region in the AWS account.

Type: String


Pattern: ^[a-zA-Z0-9-]*[a-zA-Z0-9]$\{0,62\}

Required: Yes

Response Syntax

```
[
  "AppSpecification": {
    "ContainerArguments": [ "string" ],
    "ContainerEntrypoint": [ "string" ],
    "ImageUri": "string"
  },
  "AutoMLJobArn": "string",
  "CreationTime": number,
  "Environment": {
    "string": "string"
  },
  "ExitMessage": "string",
  "ExperimentConfig": {
    "ExperimentName": "string",
    "RunName": "string",
    "TrialComponentDisplayName": "string",
    "TrialName": "string"
  },
  "FailureReason": "string",
  "LastModifiedTime": number,
  "MonitoringScheduleArn": "string",
  "NetworkConfig": {
    "EnableInterContainerTrafficEncryption": boolean,
    "EnableNetworkIsolation": boolean,
    "VpcConfig": {
```

563
"SecurityGroupIds": [ "string" ],
"Subnets": [ "string" ]
},
"ProcessingEndTime": number,
"ProcessingInputs": [
  {
    "AppManaged": boolean,
    "DatasetDefinition": {
      "AthenaDatasetDefinition": {
        "Catalog": "string",
        "Database": "string",
        "KmsKeyId": "string",
        "OutputCompression": "string",
        "OutputFormat": "string",
        "OutputS3Uri": "string",
        "QueryString": "string",
        "WorkGroup": "string"
      },
      "DataDistributionType": "string",
      "InputMode": "string",
      "LocalPath": "string",
      "RedshiftDatasetDefinition": {
        "ClusterId": "string",
        "ClusterRoleArn": "string",
        "Database": "string",
        "DbUser": "string",
        "KmsKeyId": "string",
        "OutputCompression": "string",
        "OutputFormat": "string",
        "OutputS3Uri": "string",
        "QueryString": "string"
      }
    },
    "InputName": "string",
    "S3Input": {
      "LocalPath": "string",
      "S3CompressionType": "string",
      "S3DataDistributionType": "string",
      "S3DataType": "string",
      "S3InputMode": "string",
      "S3Uri": "string"
    }
  }
],
"ProcessingJobArn": "string",
"ProcessingJobName": "string",
"ProcessingJobStatus": "string",
"ProcessingOutputConfig": {
  "KmsKeyId": "string",
  "Outputs": [
    {
      "AppManaged": boolean,
      "FeatureStoreOutput": {
        "FeatureGroupName": "string"
      },
      "OutputName": "string",
      "S3Output": {
        "LocalPath": "string",
        "S3UploadMode": "string",
        "S3Uri": "string"
      }
    }
  ]
},
"ProcessingResources": {
"ClusterConfig": {
  "InstanceCount":  number,
  "InstanceType":  "string",
  "VolumeKmsKeyId":  "string",
  "VolumeSizeInGB":  number
},

"ProcessingStartTime":  number,
"RoleArn":  "string",
"StoppingCondition": {
  "MaxRuntimeInSeconds":  number
},

"TrainingJobArn":  "string"

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**AppSpecification (p. 563)**

Configures the processing job to run a specified container image.

Type: **AppSpecification (p. 1250)** object

**AutoMLJobArn (p. 563)**

The ARN of an AutoML job associated with this processing job.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 256.

Pattern: arn:aws[a-z-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:automl-job/.*

**CreationTime (p. 563)**

The time at which the processing job was created.

Type: Timestamp

**Environment (p. 563)**

The environment variables set in the Docker container.

Type: String to string map

Map Entries: Maximum number of 100 items.

Key Length Constraints: Maximum length of 256.

Key Pattern: [a-zA-Z_][a-zA-Z0-9_-]*

Value Length Constraints: Maximum length of 256.

Value Pattern: \S\s* 

**ExitMessage (p. 563)**

An optional string, up to one KB in size, that contains metadata from the processing container when the processing job exits.

Type: String
Length Constraints: Maximum length of 1024.

Pattern: [\S\s]*

**ExperimentConfig (p. 563)**

The configuration information used to create an experiment.

Type: [ExperimentConfig (p. 1473)](p. 1473) object

**FailureReason (p. 563)**

A string, up to one KB in size, that contains the reason a processing job failed, if it failed.

Type: String

Length Constraints: Maximum length of 1024.

**LastModifiedTime (p. 563)**

The time at which the processing job was last modified.

Type: Timestamp

**MonitoringScheduleArn (p. 563)**

The ARN of a monitoring schedule for an endpoint associated with this processing job.

Type: String

Length Constraints: Maximum length of 256.

Pattern: .*

**NetworkConfig (p. 563)**

Networking options for a processing job.

Type: [NetworkConfig (p. 1769)](p. 1769) object

**ProcessingEndTime (p. 563)**

The time at which the processing job completed.

Type: Timestamp

**ProcessingInputs (p. 563)**

The inputs for a processing job.

Type: Array of [ProcessingInput (p. 1832)](p. 1832) objects

Array Members: Minimum number of 0 items. Maximum number of 10 items.

**ProcessingJobArn (p. 563)**

The Amazon Resource Name (ARN) of the processing job.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:processing-job/.*

**ProcessingJobName (p. 563)**

The name of the processing job. The name must be unique within an AWS Region in the AWS account.
DescribeProcessingJob

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}$

ProcessingJobStatus (p. 563)

Provides the status of a processing job.

Type: String

Valid Values: InProgress | Completed | Failed | Stopping | Stopped

ProcessingOutputConfig (p. 563)

Output configuration for the processing job.

Type: ProcessingOutputConfig (p. 1841) object

ProcessingResources (p. 563)

Identifies the resources, ML compute instances, and ML storage volumes to deploy for a processing job. In distributed training, you specify more than one instance.

Type: ProcessingResources (p. 1842) object

ProcessingStartTime (p. 563)

The time at which the processing job started.

Type: Timestamp

RoleArn (p. 563)

The Amazon Resource Name (ARN) of an IAM role that Amazon SageMaker can assume to perform tasks on your behalf.

Type: String


Pattern: ^arn:aws[a-z\-]*:iam::\d{12}:role/\?([a-zA-Z0-9_-]+)\+$

StoppingCondition (p. 563)

The time limit for how long the processing job is allowed to run.

Type: ProcessingStoppingCondition (p. 1846) object

TrainingJobArn (p. 563)

The ARN of a training job associated with this processing job.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:training-job/.*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).
ResourceNotFound

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DescribeProject
Service: Amazon SageMaker Service
Describes the details of a project.

Request Syntax

```
{
  "ProjectName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**ProjectName (p. 569)**

- The name of the project to describe.
- Type: String
- Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,31}$`
- Required: Yes

Response Syntax

```
{
  "CreatedBy": {
    "DomainId": "string",
    "IamIdentity": {
      "Arn": "string",
      "PrincipalId": "string",
      "SourceIdentity": "string"
    },
    "UserProfileArn": "string",
    "UserProfileName": "string"
  },
  "CreationTime": number,
  "LastModifiedBy": {
    "DomainId": "string",
    "IamIdentity": {
      "Arn": "string",
      "PrincipalId": "string",
      "SourceIdentity": "string"
    },
    "UserProfileArn": "string",
    "UserProfileName": "string"
  },
  "LastModifiedTime": number,
  "ProjectArn": "string",
  "ProjectDescription": "string",
  "ProjectId": "string"
}
```

569
"ProjectName": "string",
"ProjectStatus": "string",
"ServiceCatalogProvisionedProductDetails": {
  "ProvisionedProductId": "string",
  "ProvisionedProductStatusMessage": "string"
},
"ServiceCatalogProvisioningDetails": {
  "PathId": "string",
  "ProductId": "string",
  "ProvisioningArtifactId": "string",
  "ProvisioningParameters": [
    {
      "Key": "string",
      "Value": "string"
    }
  ]
}

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**CreatedBy (p. 569)**

Information about the user who created or modified an experiment, trial, trial component, lineage group, project, or model card.

Type: UserContext (p. 2067) object

**CreationTime (p. 569)**

The time when the project was created.

Type: Timestamp

**LastModifiedBy (p. 569)**

Information about the user who created or modified an experiment, trial, trial component, lineage group, project, or model card.

Type: UserContext (p. 2067) object

**LastModifiedTime (p. 569)**

The timestamp when project was last modified.

Type: Timestamp

**ProjectArn (p. 569)**

The Amazon Resource Name (ARN) of the project.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: ^arn:aws(-cn|-us-gov)?:sagemaker:[a-z0-9-]{9,16}:[0-9]{12}:project/[^$]{1,2048}$

**ProjectDescription (p. 569)**

The description of the project.
DescribeProject

**Type:** String

**Length Constraints:** Maximum length of 1024.

**Pattern:** \p{L}\p{M}\p{Z}\p{S}\p{N}\p{P}]*

**Projectld (p. 569)**

The ID of the project.

**Type:** String

**Length Constraints:** Minimum length of 1. Maximum length of 20.

**Pattern:** ^[a-zA-Z0-9][-]*[a-zA-Z0-9]\]{$,20}$

**ProjectName (p. 569)**

The name of the project.

**Type:** String

**Length Constraints:** Minimum length of 1. Maximum length of 32.

**Pattern:** ^[a-zA-Z0-9][-]*[a-zA-Z0-9]\]{$,32}$

**ProjectStatus (p. 569)**

The status of the project.

**Type:** String

**Valid Values:** Pending | CreateInProgress | CreateCompleted | CreateFailed | DeleteInProgress | DeleteFailed | DeleteCompleted | UpdateInProgress | UpdateCompleted | UpdateFailed

**ServiceCatalogProvisionedProductDetails (p. 569)**

Information about a provisioned service catalog product.

**Type:** ServiceCatalogProvisionedProductDetails (p. 1946) object

**ServiceCatalogProvisioningDetails (p. 569)**

Information used to provision a service catalog product. For information, see What is AWS Service Catalog.

**Type:** ServiceCatalogProvisioningDetails (p. 1947) object

**Errors**

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
DescribeSpace
Service: Amazon SageMaker Service
Describes the space.

Request Syntax
```
{
    "DomainId": "string",
    "SpaceName": "string"
}
```

Request Parameters
For information about the parameters that are common to all actions, see [Common Parameters](p. 2178).

The request accepts the following data in JSON format.

**DomainId (p. 573)**
- The ID of the associated Domain.
  - Type: String
  - Length Constraints: Maximum length of 63.
  - Required: Yes

**SpaceName (p. 573)**
- The name of the space.
  - Type: String
  - Length Constraints: Maximum length of 63.
  - Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$`
  - Required: Yes

Response Syntax
```
{
    "CreationTime": number,
    "DomainId": "string",
    "FailureReason": "string",
    "HomeEfsFileSystemUid": "string",
    "LastModifiedTime": number,
    "OwnershipSettings": {
        "OwnerUserProfileName": "string"
    },
    "SpaceArn": "string",
    "SpaceDisplayName": "string",
    "SpaceName": "string",
    "SpaceSettings": {
        "AppType": "string",
        "CodeEditorAppSettings": {
            "DefaultResourceSpec": {}...
        }
    }
}
```
"InstanceType": "string",
"LifecycleConfigArn": "string",
"SageMakerImageArn": "string",
"SageMakerImageVersionAlias": "string",
"SageMakerImageVersionArn": "string"
}
},
"CustomFileSystems": [
(...)
],
"JupyterLabAppSettings": {
"CodeRepositories": [
{
"RepositoryUrl": "string"
}
],
"DefaultResourceSpec": {
"InstanceType": "string",
"LifecycleConfigArn": "string",
"SageMakerImageArn": "string",
"SageMakerImageVersionAlias": "string",
"SageMakerImageVersionArn": "string"
}
},
"JupyterServerAppSettings": {
"CodeRepositories": [
{
"RepositoryUrl": "string"
}
],
"DefaultResourceSpec": {
"InstanceType": "string",
"LifecycleConfigArn": "string",
"SageMakerImageArn": "string",
"SageMakerImageVersionAlias": "string",
"SageMakerImageVersionArn": "string"
}
},
"KernelGatewayAppSettings": {
"CustomImages": [
{
"AppImageConfigName": "string",
"ImageName": "string",
"ImageVersionNumber": number
}
],
"DefaultResourceSpec": {
"InstanceType": "string",
"LifecycleConfigArn": "string",
"SageMakerImageArn": "string",
"SageMakerImageVersionAlias": "string",
"SageMakerImageVersionArn": "string"
}
},
"LifecycleConfigArns": ["string"]
},
"SpaceStorageSettings": {
"EbsStorageSettings": {
"EbsVolumeSizeInGb": number
}
},
"SpaceSharingSettings": {
"SharingType": "string"
}
},
"Status": "string"}
Response Elements

If the action is successful, the service sends back an HTTP 200 response. The following data is returned in JSON format by the service.

**CreationTime (p. 573)**

The creation time.

Type: Timestamp

**DomainId (p. 573)**

The ID of the associated Domain.

Type: String

Length Constraints: Maximum length of 63.

**FailureReason (p. 573)**

The failure reason.

Type: String

Length Constraints: Maximum length of 1024.

**HomeEfsFileSystemUid (p. 573)**

The ID of the space's profile in the Amazon Elastic File System volume.

Type: String

Length Constraints: Maximum length of 10.

Pattern: \d+

**LastModifiedTime (p. 573)**

The last modified time.

Type: Timestamp

**OwnershipSettings (p. 573)**

The collection of ownership settings for a space.

Type: OwnershipSettings (p. 1795) object

**SpaceArn (p. 573)**

The space's Amazon Resource Name (ARN).

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:space/.*

**SpaceDisplayName (p. 573)**

The name of the space that appears in the Amazon SageMaker Studio UI.
Type: String

Length Constraints: Maximum length of 64.

Pattern: ^(?!\s*$).+

**SpaceName (p. 573)**

The name of the space.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}

**SpaceSettings (p. 573)**

A collection of space settings.

Type: SpaceSettings (p. 1961) object

**SpaceSharingSettings (p. 573)**

The collection of space sharing settings for a space.

Type: SpaceSharingSettings (p. 1964) object

**Status (p. 573)**

The status.

Type: String

Valid Values: Deleting | Failed | InService | Pending | Updating | Update_Failed | Delete_Failed

**Url (p. 573)**

Returns the URL of the space. If the space is created with AWS IAM Identity Center (Successor to AWS Single Sign-On) authentication, users can navigate to the URL after appending the respective redirect parameter for the application type to be federated through AWS IAM Identity Center.

The following application types are supported:
• Studio Classic: &redirect=JupyterServer
• JupyterLab: &redirect=JupyterLab
• Code Editor, based on Code-OSS, Visual Studio Code - Open Source: &redirect=CodeEditor

Type: String

Length Constraints: Maximum length of 1024.

**Errors**

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400
See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DescribeStudioLifecycleConfig
Service: Amazon SageMaker Service
Describes the Amazon SageMaker Studio Lifecycle Configuration.

Request Syntax

```json
{
   "StudioLifecycleConfigName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**StudioLifecycleConfigName** (p. 578)

The name of the Amazon SageMaker Studio Lifecycle Configuration to describe.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}

Required: Yes

Response Syntax

```json
{
   "CreationTime": number,
   "LastModifiedTime": number,
   "StudioLifecycleConfigAppType": "string",
   "StudioLifecycleConfigArn": "string",
   "StudioLifecycleConfigContent": "string",
   "StudioLifecycleConfigName": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**CreationTime** (p. 578)

The creation time of the Amazon SageMaker Studio Lifecycle Configuration.

Type: Timestamp

**LastModifiedTime** (p. 578)

This value is equivalent to CreationTime because Amazon SageMaker Studio Lifecycle Configurations are immutable.
Type: Timestamp

**StudioLifecycleConfigAppType (p. 578)**

The App type that the Lifecycle Configuration is attached to.

Type: String

Valid Values: JupyterServer | KernelGateway | JupyterLab | CodeEditor

**StudioLifecycleConfigArn (p. 578)**

The ARN of the Lifecycle Configuration to describe.

Type: String

Length Constraints: Maximum length of 256.

Pattern: `arn:aws[a-z\-]*:sagemaker:[a-z0-9\-\*:\[0-9\]{12}:studio-lifecycle-config/.*`

**StudioLifecycleConfigContent (p. 578)**

The content of your Amazon SageMaker Studio Lifecycle Configuration script.

Type: String


Pattern: `\[\S\s\]+`

**StudioLifecycleConfigName (p. 578)**

The name of the Amazon SageMaker Studio Lifecycle Configuration that is described.

Type: String

Length Constraints: Maximum length of 63.

Pattern: `^[a-zA-Z0-9]*(-*[a-zA-Z0-9]*)\[0,62]`

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](#).

**ResourceNot Found**

Resource being access is not found.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
DescribeSubscribedWorkteam

Service: Amazon SageMaker Service

Gets information about a work team provided by a vendor. It returns details about the subscription with a vendor in the AWS Marketplace.

Request Syntax

```json
{
   "WorkteamArn": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

WorkteamArn (p. 581)

The Amazon Resource Name (ARN) of the subscribed work team to describe.

Type: String

Length Constraints: Maximum length of 256.

Pattern: `arn:aws[a-z\-]*:sagemaker:[a-z0-9\-\*:][0-9]{12}:workteam/.*`

Required: Yes

Response Syntax

```json
{
   "SubscribedWorkteam": {
      "ListingId": "string",
      "MarketplaceDescription": "string",
      "MarketplaceTitle": "string",
      "SellerName": "string",
      "WorkteamArn": "string"
   }
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

SubscribedWorkteam (p. 581)

A Workteam instance that contains information about the work team.

Type: SubscribedWorkteam (p. 1972) object
Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DescribeTrainingJob
Service: Amazon SageMaker Service

Returns information about a training job.

Some of the attributes below only appear if the training job successfully starts. If the training job fails, TrainingJobStatus is Failed and, depending on the FailureReason, attributes like TrainingStartTime, TrainingTimeInSeconds, TrainingEndTime, and BillableTimeInSeconds may not be present in the response.

Request Syntax

```json
{
   "TrainingJobName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**TrainingJobName (p. 583)**

The name of the training job.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}

Required: Yes

Response Syntax

```json
{
   "AlgorithmSpecification": {
      "AlgorithmName": "string",
      "AlgorithmSpecificationConfig": {
         "TrainingRepositoryAccessMode": "string",
         "TrainingRepositoryAuthConfig": {
            "TrainingRepositoryCredentialsProviderArn": "string"
         }
      },
      "TrainingInputMode": "string"
   },
   "AutoMLJobArn": "string",
   "AutoMLJobMetrics": {
      "TrainingJobArn": "string",
      "TrainingJobStatus": "string",
      "FailureReason": "string",
      "TrainingStartTime": "string",
      "TrainingEndTime": "string",
      "TrainingTimeInSeconds": "string",
      "BillableTimeInSeconds": "string",
      "TrainingOutputConfig": {
         "S3OutputPath": "string",
         "KmsKeyId": "string"
      }
   }
}
```
"BillableTimeInSeconds": number,
"CheckpointConfig": {
  "LocalPath": "string",
  "S3Uri": "string"
},
"CreationTime": number,
"DebugHookConfig": {
  "CollectionConfigurations": [
    {
      "CollectioName": "string",
      "CollectionParameters": {
        "string": "string"
      }
    }
  ],
  "HookParameters": {
    "string": "string"
  },
  "LocalPath": "string",
  "S3OutputPath": "string"
},
"DebugRuleConfigurations": [
  {
    "InstanceType": "string",
    "LocalPath": "string",
    "RuleConfigurationName": "string",
    "RuleEvaluatorImage": "string",
    "RuleParameters": {
      "string": "string"
    },
    "S3OutputPath": "string",
    "VolumeSizeInGB": number
  }
],
"DebugRuleEvaluationStatuses": [
  {
    "LastModifiedTime": number,
    "RuleConfigurationName": "string",
    "RuleEvaluationJobArn": "string",
    "RuleEvaluationStatus": "string",
    "StatusDetails": "string"
  }
],
"EnableInterContainerTrafficEncryption": boolean,
"EnableManagedSpotTraining": boolean,
"EnableNetworkIsolation": boolean,
"Environment": {
  "string": "string"
},
"ExperimentConfig": {
  "ExperimentName": "string",
  "RunName": "string",
  "TrialComponentDisplayName": "string",
  "TrialName": "string"
},
"FailureReason": "string",
"FinalMetricDataList": [
  {
    "MetricName": "string",
    "Timestamp": number,
    "Value": number
  }
],
"HyperParameters": {
  "string": "string"
}
"InfraCheckConfig": { 
  "EnableInfraCheck": boolean
},
"InputDataConfig": [
  {
    "ChannelName": "string",
    "CompressionType": "string",
    "ContentType": "string",
    "DataSource": {
      "FileSystemDataSource": {
        "DirectoryPath": "string",
        "FileSystemAccessMode": "string",
        "FileSystemId": "string",
        "FileSystemType": "string"
      },
      "S3DataSource": {
        "AttributeNames": [ "string" ],
        "InstanceGroupNames": [ "string" ],
        "S3DataDistributionType": "string",
        "S3DataType": "string",
        "S3Uri": "string"
      }
    },
    "InputMode": "string",
    "RecordWrapperType": "string",
    "ShuffleConfig": {
      "Seed": number
    }
  }
],
"LabelingJobArn": "string",
"LastModifiedTime": number,
"ModelArtifacts": {
  "S3ModelArtifacts": "string"
},
"OutputDataConfig": {
  "CompressionType": "string",
  "KmsKeyId": "string",
  "S3OutputPath": "string"
},
"ProfilerConfig": {
  "DisableProfiler": boolean,
  "ProfilingIntervalInMilliseconds": number,
  "ProfilingParameters": {
    "string": "string"
  },
  "S3OutputPath": "string"
},
"ProfilerRuleConfigurations": [
  {
    "InstanceType": "string",
    "LocalPath": "string",
    "RuleConfigurationName": "string",
    "RuleEvaluatorImage": "string",
    "RuleParameters": {
      "string": "string"
    },
    "S3OutputPath": "string",
    "VolumeSizeInGB": number
  }
],
"ProfilerRuleEvaluationStatuses": [
  {
    "LastModifiedTime": number,
    "RuleConfigurationName": "string",
    "RuleEvaluationJobArn": "string",
    "RuleEvaluationStatus": "string"
  }
]
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.
AlgorithmSpecification (p. 583)
Information about the algorithm used for training, and algorithm metadata.
Type: AlgorithmSpecification (p. 1225) object

AutoMLJobArn (p. 583)
The Amazon Resource Name (ARN) of an AutoML job.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 256.
Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:automl-job/.*

BillableTimeInSeconds (p. 583)
The billable time in seconds. Billable time refers to the absolute wall-clock time.

Multiply BillableTimeInSeconds by the number of instances (InstanceCount) in your training cluster to get the total compute time SageMaker bills you if you run distributed training. The formula is as follows: BillableTimeInSeconds * InstanceCount.

You can calculate the savings from using managed spot training using the formula (1 - BillableTimeInSeconds / TrainingTimeInSeconds) * 100. For example, if BillableTimeInSeconds is 100 and TrainingTimeInSeconds is 500, the savings is 80%.

Type: Integer
Valid Range: Minimum value of 1.

CheckpointConfig (p. 583)
Contains information about the output location for managed spot training checkpoint data.
Type: CheckpointConfig (p. 1328) object

CreationTime (p. 583)
A timestamp that indicates when the training job was created.
Type: Timestamp

DebugHookConfig (p. 583)
Configuration information for the Amazon SageMaker Debugger hook parameters, metric and tensor collections, and storage paths. To learn more about how to configure the DebugHookConfig parameter, see Use the SageMaker and Debugger Configuration API Operations to Create, Update, and Debug Your Training Job.
Type: DebugHookConfig (p. 1395) object

DebugRuleConfigurations (p. 583)
Configuration information for Amazon SageMaker Debugger rules for debugging output tensors.
Type: Array of DebugRuleConfiguration (p. 1397) objects
Array Members: Minimum number of 0 items. Maximum number of 20 items.

DebugRuleEvaluationStatuses (p. 583)
Evaluation status of Amazon SageMaker Debugger rules for debugging on a training job.
Type: Array of DebugRuleEvaluationStatus (p. 1399) objects
Array Members: Minimum number of 0 items. Maximum number of 20 items.

**EnableInterContainerTrafficEncryption (p. 583)**

To encrypt all communications between ML compute instances in distributed training, choose True. Encryption provides greater security for distributed training, but training might take longer. How long it takes depends on the amount of communication between compute instances, especially if you use a deep learning algorithms in distributed training.

Type: Boolean

**EnableManagedSpotTraining (p. 583)**

A Boolean indicating whether managed spot training is enabled (True) or not (False).

Type: Boolean

**EnableNetworkIsolation (p. 583)**

If you want to allow inbound or outbound network calls, except for calls between peers within a training cluster for distributed training, choose True. If you enable network isolation for training jobs that are configured to use a VPC, SageMaker downloads and uploads customer data and model artifacts through the specified VPC, but the training container does not have network access.

Type: Boolean

**Environment (p. 583)**

The environment variables to set in the Docker container.

Type: String to string map

Map Entries: Maximum number of 100 items.

Key Length Constraints: Maximum length of 512.

Key Pattern: ![regular_expression](a-zA-Z_)[a-zA-Z0-9_]*

Value Length Constraints: Maximum length of 512.

Value Pattern: ![regular_expression](\S\s)*

**ExperimentConfig (p. 583)**

Associates a SageMaker job as a trial component with an experiment and trial. Specified when you call the following APIs:

- CreateProcessingJob
- CreateTrainingJob
- CreateTransformJob

Type: [ExperimentConfig (p. 1473)] object

**FailureReason (p. 583)**

If the training job failed, the reason it failed.

Type: String

Length Constraints: Maximum length of 1024.

**FinalMetricDataList (p. 583)**

A collection of MetricData objects that specify the names, values, and dates and times that the training algorithm emitted to Amazon CloudWatch.

Type: Array of [MetricData (p. 1650)] objects
Array Members: Minimum number of 0 items. Maximum number of 40 items.

**HyperParameters (p. 583)**

Algorithm-specific parameters.

Type: String to string map

Map Entries: Minimum number of 0 items. Maximum number of 100 items.

Key Length Constraints: Maximum length of 256.

Key Pattern: .*

Value Length Constraints: Maximum length of 2500.

Value Pattern: .*

**InfraCheckConfig (p. 583)**

Contains information about the infrastructure health check configuration for the training job.

Type: InfraCheckConfig (p. 1604) object

**InputDataConfig (p. 583)**

An array of Channel objects that describes each data input channel.

Type: Array of Channel (p. 1324) objects

Array Members: Minimum number of 1 item. Maximum number of 20 items.

**LabelingJobArn (p. 583)**

The Amazon Resource Name (ARN) of the SageMaker Ground Truth labeling job that created the transform or training job.

Type: String

Length Constraints: Maximum length of 2048.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:labeling-job/.*

**LastModifiedTime (p. 583)**

A timestamp that indicates when the status of the training job was last modified.

Type: Timestamp

**ModelArtifacts (p. 583)**

Information about the Amazon S3 location that is configured for storing model artifacts.

Type: ModelArtifacts (p. 1659) object

**OutputDataConfig (p. 583)**

The S3 path where model artifacts that you configured when creating the job are stored. SageMaker creates subfolders for model artifacts.

Type: OutputDataConfig (p. 1792) object

**ProfilerConfig (p. 583)**

Configuration information for Amazon SageMaker Debugger system monitoring, framework profiling, and storage paths.

Type: ProfilerConfig (p. 1860) object
ProfilerRuleConfigurations (p. 583)

Configuration information for Amazon SageMaker Debugger rules for profiling system and framework metrics.

Type: Array of ProfilerRuleConfiguration (p. 1864) objects

Array Members: Minimum number of 0 items. Maximum number of 20 items.

ProfilerRuleEvaluationStatuses (p. 583)

Evaluation status of Amazon SageMaker Debugger rules for profiling on a training job.

Type: Array of ProfilerRuleEvaluationStatus (p. 1866) objects

Array Members: Minimum number of 0 items. Maximum number of 20 items.

ProfilingStatus (p. 583)

Profiling status of a training job.

Type: String

Valid Values: Enabled | Disabled

ResourceConfig (p. 583)

Resources, including ML compute instances and ML storage volumes, that are configured for model training.

Type: ResourceConfig (p. 1910) object

RetryStrategy (p. 583)

The number of times to retry the job when the job fails due to an InternalServerError.

Type: RetryStrategy (p. 1918) object

RoleArn (p. 583)

The AWS Identity and Access Management (IAM) role configured for the training job.

Type: String


Pattern: ^arn:aws[a-z\-]*:iam::\d{12}:role/\?[a-zA-Z0-9=+,@\-/_]+$  

SecondaryStatus (p. 583)

Provides detailed information about the state of the training job. For detailed information on the secondary status of the training job, see StatusMessage under SecondaryStatusTransition.

SageMaker provides primary statuses and secondary statuses that apply to each of them:

- InProgress
  - Starting - Starting the training job.
  - Downloading - An optional stage for algorithms that support File training input mode. It indicates that data is being downloaded to the ML storage volumes.
  - Training - Training is in progress.
  - Interrupted - The job stopped because the managed spot training instances were interrupted.
  - Uploading - Training is complete and the model artifacts are being uploaded to the S3 location.
Completed
  • Completed - The training job has completed.

Failed
  • Failed - The training job has failed. The reason for the failure is returned in the FailureReason field of DescribeTrainingJobResponse.

Stopped
  • MaxRuntimeExceeded - The job stopped because it exceeded the maximum allowed runtime.
  • MaxWaitTimeExceeded - The job stopped because it exceeded the maximum allowed wait time.
  • Stopped - The training job has stopped.

Stopping
  • Stopping - Stopping the training job.

Important
  Valid values for SecondaryStatus are subject to change.

We no longer support the following secondary statuses:
• LaunchingMLInstances
• PreparingTraining
• DownloadingTrainingImage

Type: String

Valid Values: Starting | LaunchingMLInstances | PreparingTrainingStack | Downloading | DownloadingTrainingImage | Training | Uploading | Stopping | Stopped | MaxRuntimeExceeded | Completed | Failed | Interrupted | MaxWaitTimeExceeded | Updating | Restarting

SecondaryStatusTransitions (p. 583)

A history of all of the secondary statuses that the training job has transitioned through.

Type: Array of SecondaryStatusTransition (p. 1941) objects

StoppingCondition (p. 583)

Specifies a limit to how long a model training job can run. It also specifies how long a managed Spot training job has to complete. When the job reaches the time limit, SageMaker ends the training job. Use this API to cap model training costs.

To stop a job, SageMaker sends the algorithm the SIGTERM signal, which delays job termination for 120 seconds. Algorithms can use this 120-second window to save the model artifacts, so the results of training are not lost.

Type: StoppingCondition (p. 1968) object

TensorBoardOutputConfig (p. 583)

Configuration of storage locations for the Amazon SageMaker Debugger TensorBoard output data.

Type: TensorBoardOutputConfig (p. 1984) object

TrainingEndTime (p. 583)

Indicates the time when the training job ends on training instances. You are billed for the time interval between the value of TrainingStartTime and this time. For successful jobs and stopped jobs, this is the time after model artifacts are uploaded. For failed jobs, this is the time when SageMaker detects a job failure.
Type: Timestamp

**TrainingJobArn (p. 583)**

The Amazon Resource Name (ARN) of the training job.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:training-job/.*

**TrainingJobName (p. 583)**

Name of the model training job.

Type: String


Pattern: ^[a-zA-Z0-9][-]*[a-zA-Z0-9]{0,62}

**TrainingJobStatus (p. 583)**

The status of the training job.

SageMaker provides the following training job statuses:
- **InProgress** - The training is in progress.
- **Completed** - The training job has completed.
- **Failed** - The training job has failed. To see the reason for the failure, see the FailureReason field in the response to a DescribeTrainingJobResponse call.
- **Stopping** - The training job is stopping.
- **Stopped** - The training job has stopped.

For more detailed information, see SecondaryStatus.

Type: String

Valid Values: InProgress | Completed | Failed | Stopping | Stopped

**TrainingStartTime (p. 583)**

Indicates the time when the training job starts on training instances. You are billed for the time interval between this time and the value of TrainingEndTime. The start time in CloudWatch Logs might be later than this time. The difference is due to the time it takes to download the training data and to the size of the training container.

Type: Timestamp

**TrainingTimeInSeconds (p. 583)**

The training time in seconds.

Type: Integer

Valid Range: Minimum value of 1.

**TuningJobArn (p. 583)**

The Amazon Resource Name (ARN) of the associated hyperparameter tuning job if the training job was launched by a hyperparameter tuning job.

Type: String
DescribeTrainingJob

Length Constraints: Maximum length of 256.
Pattern: arn:aws[a-z-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:hyper-parameter-tuning-job/.*

VpcConfig (p. 583)

A VpcConfig object that specifies the VPC that this training job has access to. For more information, see Protect Training Jobs by Using an Amazon Virtual Private Cloud.

Type: VpcConfig (p. 2076) object

WarmPoolStatus (p. 583)

The status of the warm pool associated with the training job.

Type: WarmPoolStatus (p. 2077) object

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceNotFound

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DescribeTransformJob

Service: Amazon SageMaker Service

Returns information about a transform job.

Request Syntax

```
{
  "TransformJobName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**TransformJobName (p. 594)**

The name of the transform job that you want to view details of.

Type: String


Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$`

Required: Yes

Response Syntax

```
{
  "AutoMLJobArn": "string",
  "BatchStrategy": "string",
  "CreationTime": number,
  "DataCaptureConfig": {
    "DestinationS3Uri": "string",
    "GenerateInferenceId": boolean,
    "KmsKeyId": "string"
  },
  "DataProcessing": {
    "InputFilter": "string",
    "JoinSource": "string",
    "OutputFilter": "string"
  },
  "Environment": {
    "string": "string"
  },
  "ExperimentConfig": {
    "ExperimentName": "string",
    "RunName": "string",
    "TrialComponentDisplayName": "string",
    "TrialName": "string"
  },
  "FailureReason": "string",
  "LabelingJobArn": "string",
  "MaxConcurrentTransforms": number,
  "MaxPayloadInMB": number
}
```
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**AutoMLJobArn (p. 594)**

The Amazon Resource Name (ARN) of the AutoML transform job.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 256.

Pattern: `arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:automl-job/.*`

**BatchStrategy (p. 594)**

Specifies the number of records to include in a mini-batch for an HTTP inference request. A record is a single unit of input data that inference can be made on. For example, a single line in a CSV file is a record.

To enable the batch strategy, you must set `SplitType` to `Line`, `RecordIO`, or `TFRecord`.

Type: String

Valid Values: `MultiRecord` | `SingleRecord`

**CreationTime (p. 594)**

A timestamp that shows when the transform Job was created.
Type: Timestamp

**DataCaptureConfig (p. 594)**

Configuration to control how SageMaker captures inference data.

Type: BatchDataCaptureConfig (p. 1301) object

**DataProcessing (p. 594)**

The data structure used to specify the data to be used for inference in a batch transform job and to associate the data that is relevant to the prediction results in the output. The input filter provided allows you to exclude input data that is not needed for inference in a batch transform job. The output filter provided allows you to include input data relevant to interpreting the predictions in the output from the job. For more information, see [Associate Prediction Results with their Corresponding Input Records](#).

Type: DataProcessing (p. 1386) object

**Environment (p. 594)**

The environment variables to set in the Docker container. We support up to 16 key and values entries in the map.

Type: String to string map

Map Entries: Maximum number of 16 items.

Key Length Constraints: Maximum length of 1024.

Key Pattern: \[a-zA-Z_]\[a-zA-Z0-9_\]{0,1023}

Value Length Constraints: Maximum length of 10240.

Value Pattern: \[\S\s\]*

**ExperimentConfig (p. 594)**

 Associates a SageMaker job as a trial component with an experiment and trial. Specified when you call the following APIs:

- CreateProcessingJob
- CreateTrainingJob
- CreateTransformJob

Type: ExperimentConfig (p. 1473) object

**FailureReason (p. 594)**

If the transform job failed, FailureReason describes why it failed. A transform job creates a log file, which includes error messages, and stores it as an Amazon S3 object. For more information, see [Log Amazon SageMaker Events with Amazon CloudWatch](#).

Type: String

Length Constraints: Maximum length of 1024.

**LabelingJobArn (p. 594)**

The Amazon Resource Name (ARN) of the Amazon SageMaker Ground Truth labeling job that created the transform or training job.

Type: String

Length Constraints: Maximum length of 2048.
**MaxConcurrentTransforms (p. 594)**

The maximum number of parallel requests on each instance node that can be launched in a transform job. The default value is 1.

Type: Integer

Valid Range: Minimum value of 0.

**MaxPayloadInMB (p. 594)**

The maximum payload size, in MB, used in the transform job.

Type: Integer

Valid Range: Minimum value of 0.

**ModelClientConfig (p. 594)**

The timeout and maximum number of retries for processing a transform job invocation.

Type: `ModelClientConfig (p. 1675) object`

**ModelName (p. 594)**

The name of the model used in the transform job.

Type: String

Length Constraints: Maximum length of 63.

Pattern: `^[a-zA-Z0-9]{0,62}(-*[a-zA-Z0-9])*$`

**TransformEndTime (p. 594)**

Indicates when the transform job has been completed, or has stopped or failed. You are billed for the time interval between this time and the value of `TransformStartTime`.

Type: Timestamp

**TransformInput (p. 594)**

Describes the dataset to be transformed and the Amazon S3 location where it is stored.

Type: `TransformInput (p. 2020) object`

**TransformJobArn (p. 594)**

The Amazon Resource Name (ARN) of the transform job.

Type: String

Length Constraints: Maximum length of 256.

Pattern: `arn:aws[a-zA-Z0-9]*:sagemaker:[a-zA-Z0-9]*:[0-9]{12}:transform-job/.*`

**TransformJobName (p. 594)**

The name of the transform job.

Type: String


Pattern: `^[a-zA-Z0-9]{0,62}(-*[a-zA-Z0-9])*$`
**TransformJobStatus (p. 594)**

The status of the transform job. If the transform job failed, the reason is returned in the FailureReason field.

Type: String

Valid Values: InProgress | Completed | Failed | Stopping | Stopped

**TransformOutput (p. 594)**

Identifies the Amazon S3 location where you want Amazon SageMaker to save the results from the transform job.

Type: TransformOutput (p. 2032) object

**TransformResources (p. 594)**

Describes the resources, including ML instance types and ML instance count, to use for the transform job.

Type: TransformResources (p. 2034) object

**TransformStartTime (p. 594)**

Indicates when the transform job starts on ML instances. You are billed for the time interval between this time and the value of TransformEndTime.

Type: Timestamp

**Errors**

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DescribeTrial
Service: Amazon SageMaker Service
Provides a list of a trial's properties.

Request Syntax

```
{
    "TrialName": "string"
}
```

Request Parameters
For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**TrialName (p. 599)**

The name of the trial to describe.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: ^[a-zA-Z0-9][-\[a-zA-Z0-9]\{0,119}\}

Required: Yes

Response Syntax

```
{
    "CreatedBy": {
        "DomainId": "string",
        "IamIdentity": {
            "Arn": "string",
            "PrincipalId": "string",
            "SourceIdentity": "string"
        },
        "UserProfileArn": "string",
        "UserProfileName": "string"
    },
    "CreationTime": number,
    "DisplayName": "string",
    "ExperimentName": "string",
    "LastModifiedBy": {
        "DomainId": "string",
        "IamIdentity": {
            "Arn": "string",
            "PrincipalId": "string",
            "SourceIdentity": "string"
        },
        "UserProfileArn": "string",
        "UserProfileName": "string"
    },
    "LastModifiedTime": number,
    "MetadataProperties": {
```

599
"CommitId": "string",
"CreatedBy": "string",
"ProjectId": "string",
"Repository": "string"
},
"Source": {
  "SourceArn": "string",
  "SourceType": "string"
},
"TrialArn": "string",
"TrialName": "string"

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**CreatedBy (p. 599)**

Who created the trial.

Type: UserContext (p. 2067) object

**CreationTime (p. 599)**

When the trial was created.

Type: Timestamp

**DisplayName (p. 599)**

The name of the trial as displayed. If DisplayName isn't specified, TrialName is displayed.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: ^[a-zA-Z0-9][-]*[a-zA-Z0-9]{0,119}

**ExperimentName (p. 599)**

The name of the experiment the trial is part of.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: ^[a-zA-Z0-9][-]*[a-zA-Z0-9]{0,119}

**LastModifiedBy (p. 599)**

Who last modified the trial.

Type: UserContext (p. 2067) object

**LastModifiedTime (p. 599)**

When the trial was last modified.

Type: Timestamp

**MetadataProperties (p. 599)**

Metadata properties of the tracking entity, trial, or trial component.
Type: MetadataProperties (p. 1648) object

Source (p. 599)
The Amazon Resource Name (ARN) of the source and, optionally, the job type.

Type: TrialSource (p. 2056) object

TrialArn (p. 599)
The Amazon Resource Name (ARN) of the trial.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:experiment-trial/.*

TrialName (p. 599)
The name of the trial.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,119}

Errors
For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceNotFound
Resource being access is not found.

HTTP Status Code: 400

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DescribeTrialComponent

Service: Amazon SageMaker Service

Provides a list of a trials component's properties.

Request Syntax

```
{
  "TrialComponentName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see [Common Parameters](p. 2178).

The request accepts the following data in JSON format.

**TrialComponentName (p. 602)**

The name of the trial component to describe.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 256.

Pattern: (arn:aws[a-z-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:(experiment|experiment-trial|experiment-trial-component|artifact|action|context)\/)?([a-zA-Z0-9](-*[a-zA-Z0-9])\{0,119})

Required: Yes

Response Syntax

```
{
  "CreatedBy": {
    "DomainId": "string",
    "IamIdentity": { 
      "Arn": "string",
      "PrincipalId": "string",
      "SourceIdentity": "string"
    },
    "UserProfileArn": "string",
    "UserProfileName": "string"
  },
  "CreationTime": number,
  "DisplayName": "string",
  "EndTime": number,
  "InputArtifacts": {
    "string": {
      "MediaType": "string",
      "Value": "string"
    }
  },
  "LastModifiedBy": {
    "DomainId": "string",
    "IamIdentity": {
      "Arn": "string",
      "PrincipalId": "string",
      "SourceIdentity": "string"
    },
    "UserProfileArn": "string",
    "UserProfileName": "string"
  }
}
```
"Arn": "string",
"PrincipalId": "string",
"SourceIdentity": "string"
},
"UserProfileArn": "string",
"UserProfileName": "string"
},
"LastModifiedTime": number,
"LineageGroupArn": "string",
"MetadataProperties": {
  "CommitId": "string",
  "GeneratedBy": "string",
  "ProjectId": "string",
  "Repository": "string"
},
"Metrics": [
  {
    "Avg": number,
    "Count": number,
    "Last": number,
    "Max": number,
    "MetricName": "string",
    "Min": number,
    "SourceArn": "string",
    "StdDev": number,
    "TimeStamp": number
  }
],
"OutputArtifacts": {
  "string": {
    "MediaType": "string",
    "Value": "string"
  }
},
"Parameters": {
  "string": {
    "NumberValue": number,
    "StringValue": "string"
  }
},
"Source": {
  "SourceArn": "string",
  "SourceType": "string"
},
"Sources": [
  {
    "SourceArn": "string",
    "SourceType": "string"
  }
],
"StartTime": number,
"Status": {
  "Message": "string",
  "PrimaryStatus": "string"
},
"TrialComponentArn": "string",
"TrialComponentName": "string"

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.
**CreatedBy (p. 602)**

Who created the trial component.

Type: `UserContext (p. 2067)` object

**CreationTime (p. 602)**

When the component was created.

Type: Timestamp

**DisplayName (p. 602)**

The name of the component as displayed. If `DisplayName` isn't specified, `TrialComponentName` is displayed.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,119}$`

**EndTime (p. 602)**

When the component ended.

Type: Timestamp

**InputArtifacts (p. 602)**

The input artifacts of the component.

Type: String to `TrialComponentArtifact (p. 2045)` object map

Map Entries: Maximum number of 30 items.

Key Length Constraints: Maximum length of 64.

Key Pattern: `.*`

**LastModifiedBy (p. 602)**

Who last modified the component.

Type: `UserContext (p. 2067)` object

**LastModifiedTime (p. 602)**

When the component was last modified.

Type: Timestamp

**LineageGroupArn (p. 602)**

The Amazon Resource Name (ARN) of the lineage group.

Type: String

Length Constraints: Maximum length of 256.

Pattern: `arn:aws[a-zA-Z\-]*:sagemaker:[a-zA-Z0-9\-]*:[0-9]{12}:lineage-group/.*`
Type: **MetadataProperties (p. 1648)** object

**Metrics (p. 602)**

The metrics for the component.

Type: Array of **TrialComponentMetricSummary (p. 2046)** objects

**OutputArtifacts (p. 602)**

The output artifacts of the component.

Type: String to **TrialComponentArtifact (p. 2045)** object map

Map Entries: Maximum number of 30 items.

Key Length Constraints: Maximum length of 64.

Key Pattern: . *

**Parameters (p. 602)**

The hyperparameters of the component.

Type: String to **TrialComponentParameterValue (p. 2048)** object map

Map Entries: Maximum number of 150 items.

Key Length Constraints: Maximum length of 256.

Key Pattern: . *

**Source (p. 602)**

The Amazon Resource Name (ARN) of the source and, optionally, the job type.

Type: **TrialComponentSource (p. 2051)** object

**Sources (p. 602)**

A list of ARNs and, if applicable, job types for multiple sources of an experiment run.

Type: Array of **TrialComponentSource (p. 2051)** objects

**StartTime (p. 602)**

When the component started.

Type: Timestamp

**Status (p. 602)**

The status of the component. States include:

- InProgress
- Completed
- Failed

Type: **TrialComponentStatus (p. 2053)** object

**TrialComponentArn (p. 602)**

The Amazon Resource Name (ARN) of the trial component.

Type: String

Length Constraints: Maximum length of 256.
Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:experiment-trial-component/.*

**TrialComponentName (p. 602)**

The name of the trial component.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,119}$

**Errors**

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
DescribeUserProfile

Service: Amazon SageMaker Service

Describes a user profile. For more information, see CreateUserProfile.

Request Syntax

```json
{
  "DomainId": "string",
  "UserProfileName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**DomainId (p. 607)**

The domain ID.

Type: String

Length Constraints: Maximum length of 63.

Required: Yes

**UserProfileName (p. 607)**

The user profile name. This value is not case sensitive.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

Required: Yes

Response Syntax

```json
{
  "CreationTime": number,
  "DomainId": "string",
  "FailureReason": "string",
  "HomeEfsFileSystemUid": "string",
  "LastModifiedTime": number,
  "SingleSignOnUserIdentifier": "string",
  "SingleSignOnUserValue": "string",
  "Status": "string",
  "UserProfileArn": "string",
  "UserProfileName": "string",
  "UserSettings": {
    "CanvasAppSettings": {
      "DirectDeploySettings": {
        "Status": "string"
      }
    }
  }
}
```
},
"IdentityProviderOAuthSettings": [
  {
    "DataSourceName": "string",
    "SecretArn": "string",
    "Status": "string"
  }
],
"KendraSettings": {
  "Status": "string"
},
"ModelRegisterSettings": {
  "CrossAccountModelRegisterRoleArn": "string",
  "Status": "string"
},
"TimeSeriesForecastingSettings": {
  "AmazonForecastRoleArn": "string",
  "Status": "string"
},
"WorkspaceSettings": {
  "S3ArtifactPath": "string",
  "S3KmsKeyId": "string"
},
"CodeEditorAppSettings": {
  "DefaultResourceSpec": {
    "InstanceType": "string",
    "LifecycleConfigArn": "string",
    "SageMakerImageArn": "string",
    "SageMakerImageVersionAlias": "string",
    "SageMakerImageVersionArn": "string"
  },
  "LifecycleConfigArns": [ "string" ]
},
"CustomFileSystemConfigs": [ ...
],
"CustomPosixUserConfig": {
  "Gid": number,
  "Uid": number
},
"DefaultLandingUri": "string",
"ExecutionRole": "string",
"JupyterLabAppSettings": {
  "CodeRepositories": {
    "RepositoryUrl": "string"
  }
},
"CustomImages": [ 
  {
    "AppImageConfigName": "string",
    "ImageName": "string",
    "ImageVersionNumber": number
  }
],
"DefaultResourceSpec": {
  "InstanceType": "string",
  "LifecycleConfigArn": "string",
  "SageMakerImageArn": "string",
  "SageMakerImageVersionAlias": "string",
  "SageMakerImageVersionArn": "string"
},
"LifecycleConfigArns": [ "string" ]
},
"JupyterServerAppSettings": {

608
"CodeRepositories": [
  {
    "RepositoryUrl": "string"
  }
],
"DefaultResourceSpec": {
  "InstanceType": "string",
  "LifecycleConfigArn": "string",
  "SageMakerImageArn": "string",
  "SageMakerImageVersionAlias": "string",
  "SageMakerImageVersionArn": "string"
},
"LifecycleConfigArns": [ "string" ]
],
"KernelGatewayAppSettings": {
  "CustomImages": [
    {
      "AppImageConfigName": "string",
      "ImageName": "string",
      "ImageVersionNumber": number
    }
  ],
  "DefaultResourceSpec": {
    "InstanceType": "string",
    "LifecycleConfigArn": "string",
    "SageMakerImageArn": "string",
    "SageMakerImageVersionAlias": "string",
    "SageMakerImageVersionArn": "string"
  },
  "LifecycleConfigArns": [ "string" ]
},
"RSessionAppSettings": {
  "CustomImages": [
    {
      "AppImageConfigName": "string",
      "ImageName": "string",
      "ImageVersionNumber": number
    }
  ],
  "DefaultResourceSpec": {
    "InstanceType": "string",
    "LifecycleConfigArn": "string",
    "SageMakerImageArn": "string",
    "SageMakerImageVersionAlias": "string",
    "SageMakerImageVersionArn": "string"
  }
},
"RStudioServerProAppSettings": {
  "AccessStatus": "string",
  "UserGroup": "string"
},
"SecurityGroups": [ "string" ],
"SharingSettings": {
  "NotebookOutputOption": "string",
  "S3KmsKeyId": "string",
  "S3OutputPath": "string"
},
"SpaceStorageSettings": {
  "DefaultEbsStorageSettings": {
    "DefaultEbsVolumeSizeInGb": number,
    "MaximumEbsVolumeSizeInGb": number
  }
},
"StudioWebPortal": "string",
"TensorBoardAppSettings": {
  "DefaultResourceSpec": {
}
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

- **CreationTime (p. 607)**
  - The creation time.
  - Type: Timestamp

- **DomainId (p. 607)**
  - The ID of the domain that contains the profile.
  - Type: String
  - Length Constraints: Maximum length of 63.

- **FailureReason (p. 607)**
  - The failure reason.
  - Type: String
  - Length Constraints: Maximum length of 1024.

- **HomeEfsFileSystemUid (p. 607)**
  - The ID of the user's profile in the Amazon Elastic File System (EFS) volume.
  - Type: String
  - Pattern: \d+

- **LastModifiedTime (p. 607)**
  - The last modified time.
  - Type: Timestamp

- **SingleSignOnUserIdentifier (p. 607)**
  - The IAM Identity Center user identifier.
  - Type: String
  - Pattern: UserName

- **SingleSignOnUserValue (p. 607)**
  - The IAM Identity Center user value.
Type: String
Length Constraints: Maximum length of 256.

**Status (p. 607)**

The status.
Type: String
Valid Values: Deleting | Failed | InService | Pending | Updating | Update_Failed | Delete_Failed

**UserProfileArn (p. 607)**

The user profile Amazon Resource Name (ARN).
Type: String
Length Constraints: Maximum length of 256.
Pattern: arn:aws[a-zA-z\-]*:sagemaker:[a-zA-z0-9\-]*:[0-9]{12}:user-profile/.*

**UserProfileName (p. 607)**

The user profile name.
Type: String
Length Constraints: Maximum length of 63.
Pattern: ^[a-zA-z0-9](-*[a-zA-z0-9]){0,62}

**UserSettings (p. 607)**

A collection of settings.
Type: UserSettings (p. 2070) object

**Errors**

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceNotFound**

Resource being access is not found.
HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DescribeWorkforce

Service: Amazon SageMaker Service

Lists private workforce information, including workforce name, Amazon Resource Name (ARN), and, if applicable, allowed IP address ranges (CIDRs). Allowable IP address ranges are the IP addresses that workers can use to access tasks.

**Important**
This operation applies only to private workforces.

**Request Syntax**

```
{
   "WorkforceName": "string"
}
```

**Request Parameters**

For information about the parameters that are common to all actions, see [Common Parameters](p. 2178).

The request accepts the following data in JSON format.

**WorkforceName (p. 613)**

The name of the private workforce whose access you want to restrict. `WorkforceName` is automatically set to `default` when a workforce is created and cannot be modified.

Type: String


Pattern: `^[a-zA-Z0-9](^[a-zA-Z0-9-]\{0,62\})\{0,62\}$`

Required: Yes

**Response Syntax**

```
{
   "Workforce": {
      "CognitoConfig": {
         "ClientId": "string",
         "UserPool": "string"
      },
      "CreateDate": number,
      "FailureReason": "string",
      "LastUpdatedDate": number,
      "OidcConfig": {
         "AuthorizationEndpoint": "string",
         "ClientId": "string",
         "Issuer": "string",
         "JwksUri": "string",
         "LogoutEndpoint": "string",
         "TokenEndpoint": "string",
         "UserInfoEndpoint": "string"
      },
      "SourceIpConfig": {
         "Cidrs": [ "string" ]
      }
   }
}
```
},
"Status": "string",
"SubDomain": "string",
"WorkforceArn": "string",
"WorkforceName": "string",
"WorkforceVpcConfig": {
  "SecurityGroupId": [ "string" ],
  "Subnets": [ "string" ],
  "VpcEndpointId": "string",
  "VpcId": "string"
}
]

### Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**Workforce (p. 613)**

A single private workforce, which is automatically created when you create your first private work team. You can create one private work force in each AWS Region. By default, any workforce-related API operation used in a specific region will apply to the workforce created in that region. To learn how to create a private workforce, see Create a Private Workforce.

Type: [Workforce (p. 2079)] object

### Errors

For information about the errors that are common to all actions, see [Common Errors (p. 2180)].

### See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
DescribeWorkteam
Service: Amazon SageMaker Service

Gets information about a specific work team. You can see information such as the create date, the last updated date, membership information, and the work team's Amazon Resource Name (ARN).

Request Syntax

```
{
  "WorkteamName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**WorkteamName (p. 615)**

The name of the work team to return a description of.

Type: String


Pattern: ^[a-zA-Z0-9][-]*[a-zA-Z0-9]{0,62}

Required: Yes

Response Syntax

```
{
  "Workteam": {
    "CreateDate": number,
    "Description": "string",
    "LastUpdatedDate": number,
    "MemberDefinitions": [ {
      "CognitoMemberDefinition": {
        "ClientId": "string",
        "UserGroup": "string",
        "UserPool": "string"
      },
      "OidcMemberDefinition": {
        "Groups": [ "string" ]
      }
    } ],
    "NotificationConfiguration": { "NotificationTopicArn": "string" },
    "ProductListingIds": [ "string" ],
    "SubDomain": "string",
    "WorkforceArn": "string",
    "WorkteamArn": "string",
    "WorkteamName": "string"
  }
}
```
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**Workteam (p. 615)**

A `Workteam` instance that contains information about the work team.

Type: `Workteam (p. 2085)` object

Errors

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](#).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
DisableSagemakerServicecatalogPortfolio

Service: Amazon SageMaker Service

Disables using Service Catalog in SageMaker. Service Catalog is used to create SageMaker projects.

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
DisassociateTrialComponent
Service: Amazon SageMaker Service

Disassociates a trial component from a trial. This doesn't effect other trials the component is associated with. Before you can delete a component, you must disassociate the component from all trials it is associated with. To associate a trial component with a trial, call the AssociateTrialComponent API.

To get a list of the trials a component is associated with, use the Search API. Specify ExperimentTrialComponent for the Resource parameter. The list appears in the response under Results.TrialComponent.Parents.

Request Syntax

```json
{
   "TrialComponentName": "string",
   "TrialName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**TrialComponentName (p. 618)**

- The name of the component to disassociate from the trial.
- Type: String
- Length Constraints: Minimum length of 1. Maximum length of 120.
- Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,119}
- Required: Yes

**TrialName (p. 618)**

- The name of the trial to disassociate from.
- Type: String
- Length Constraints: Minimum length of 1. Maximum length of 120.
- Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,119}
- Required: Yes

Response Syntax

```json
{
   "TrialArn": "string",
   "TrialComponentArn": "string"
}
```
Response Elements

If the action is successful, the service sends back an HTTP 200 response. The following data is returned in JSON format by the service.

**TrialArn (p. 618)**

The Amazon Resource Name (ARN) of the trial.

- **Type:** String
- **Length Constraints:** Maximum length of 256.
- **Pattern:** `arn:aws[a-z-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:experiment-trial/.*`

**TrialComponentArn (p. 618)**

The Amazon Resource Name (ARN) of the trial component.

- **Type:** String
- **Length Constraints:** Maximum length of 256.
- **Pattern:** `arn:aws[a-z-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:experiment-trial-component/.*`

Errors

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](#).

**ResourceNotFound**

- Resource being access is not found.
- **HTTP Status Code:** 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
EnableSagemakerServicecatalogPortfolio

Service: Amazon SageMaker Service

Enables using Service Catalog in SageMaker. Service Catalog is used to create SageMaker projects.

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
GetDeviceFleetReport

Service: Amazon SageMaker Service

Describes a fleet.

Request Syntax

```
{
  "DeviceFleetName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**DeviceFleetName (p. 621)**

The name of the fleet.

Type: String


Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$`

Required: Yes

Response Syntax

```
{
  "AgentVersions": [
    {
      "AgentCount": number,
      "Version": "string"
    }
  ],
  "Description": "string",
  "DeviceFleetArn": "string",
  "DeviceFleetName": "string",
  "DeviceStats": {
    "ConnectedDeviceCount": number,
    "RegisteredDeviceCount": number
  },
  "ModelStats": [
    {
      "ActiveDeviceCount": number,
      "ConnectedDeviceCount": number,
      "ModelName": "string",
      "ModelVersion": "string",
      "OfflineDeviceCount": number,
      "SamplingDeviceCount": number
    }
  ],
  "OutputConfig": {
    "KmsKeyId": "string",
  }
}
```
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**AgentVersions (p. 621)**

The versions of Edge Manager agent deployed on the fleet.

Type: Array of AgentVersion (p. 1223) objects

**Description (p. 621)**

Description of the fleet.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 800.

Pattern: ^[-a-zA-Z0-9_.;:!]*$

**DeviceFleetArn (p. 621)**

The Amazon Resource Name (ARN) of the device.

Type: String

Pattern: ^arn:aws[\-a-zA-Z0-9\-]*:iam::d{12}:device-fleet/\?a-zA-Z_0-9+=,.@\-_/]+$

**DeviceFleetName (p. 621)**

The name of the fleet.

Type: String


Pattern: ^[a-zA-Z0-9][^-][a-zA-Z0-9]{0,62}$

**DeviceStats (p. 621)**

Status of devices.

Type: DeviceStats (p. 1418) object

**ModelStats (p. 621)**

Status of model on device.

Type: Array of EdgeModelStat (p. 1441) objects

**OutputConfig (p. 621)**

The output configuration for storing sample data collected by the fleet.

Type: EdgeOutputConfig (p. 1444) object
**ReportGenerated (p. 621)**

Timestamp of when the report was generated.

Type: Timestamp

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](##).

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](##)
- [AWS SDK for .NET](##)
- [AWS SDK for C++](##)
- [AWS SDK for Go](##)
- [AWS SDK for Java V2](##)
- [AWS SDK for JavaScript V3](##)
- [AWS SDK for PHP V3](##)
- [AWS SDK for Python](##)
- [AWS SDK for Ruby V3](##)
GetLineageGroupPolicy
Service: Amazon SageMaker Service

The resource policy for the lineage group.

Request Syntax

```json
{
    "LineageGroupName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**LineageGroupName (p. 624)**

The name or Amazon Resource Name (ARN) of the lineage group.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 256.

Pattern: (arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:lineage-group\/)?
([a-zA-Z0-9\-\]*[a-zA-Z0-9\-][0-9,119])

Required: Yes

Response Syntax

```json
{
    "LineageGroupArn": "string",
    "ResourcePolicy": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**LineageGroupArn (p. 624)**

The Amazon Resource Name (ARN) of the lineage group.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:lineage-group/.*

**ResourcePolicy (p. 624)**

The resource policy that gives access to the lineage group in another account.
Type: String
Length Constraints: Maximum length of 20480.
Pattern: .*(?:[\r\n\t].*)*

Errors
For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceNotFound
  Resource being access is not found.
  HTTP Status Code: 400

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
GetModelPackageGroupPolicy

Service: Amazon SageMaker Service

Gets a resource policy that manages access for a model group. For information about resource policies, see Identity-based policies and resource-based policies in the AWS Identity and Access Management User Guide.

Request Syntax

```json
{
  "ModelPackageGroupName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**ModelPackageName (p. 626)**

The name of the model group for which to get the resource policy.

- **Type**: String
- **Length Constraints**: Minimum length of 1. Maximum length of 63.
- **Pattern**: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$`
- **Required**: Yes

Response Syntax

```json
{
  "ResourcePolicy": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**ResourcePolicy (p. 626)**

The resource policy for the model group.

- **Type**: String
- **Length Constraints**: Minimum length of 1. Maximum length of 20480.
- **Pattern**: `.*`
Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
GetSagemakerServicecatalogPortfolioStatus

Service: Amazon SageMaker Service

Gets the status of Service Catalog in SageMaker. Service Catalog is used to create SageMaker projects.

Response Syntax

```json
{
   "Status": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**Status (p. 628)**

Whether Service Catalog is enabled or disabled in SageMaker.

Type: String

Valid Values: Enabled | Disabled

Errors

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](#).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
GetScalingConfigurationRecommendation

Service: Amazon SageMaker Service

Starts an Amazon SageMaker Inference Recommender autoscaling recommendation job. Returns recommendations for autoscaling policies that you can apply to your SageMaker endpoint.

Request Syntax

```json
{
  "EndpointName": "string",
  "InferenceRecommendationsJobName": "string",
  "RecommendationId": "string",
  "ScalingPolicyObjective": {
    "MaxInvocationsPerMinute": "number",
    "MinInvocationsPerMinute": "number"
  },
  "TargetCpuUtilizationPerCore": "number"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**EndpointName (p. 629)**

The name of an endpoint benchmarked during a previously completed inference recommendation job. This name should come from one of the recommendations returned by the job specified in the InferenceRecommendationsJobName field.

Specify either this field or the RecommendationId field.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9]+(-*[a-zA-Z0-9]+){0,62}

Required: No

**InferenceRecommendationsJobName (p. 629)**

The name of a previously completed Inference Recommender job.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 64.

Pattern: ^[a-zA-Z0-9][a-zA-Z0-9]{0,63}

Required: Yes

**RecommendationId (p. 629)**

The recommendation ID of a previously completed inference recommendation. This ID should come from one of the recommendations returned by the job specified in the InferenceRecommendationsJobName field.
Specify either this field or the EndpointName field.

Type: String

Required: No

### ScalingPolicyObjective (p. 629)

An object where you specify the anticipated traffic pattern for an endpoint.

Type: ScalingPolicyObjective (p. 1933) object

Required: No

### TargetCpuUtilizationPerCore (p. 629)

The percentage of how much utilization you want an instance to use before autoscaling. The default value is 50%.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No

## Response Syntax

```json
{
  "DynamicScalingConfiguration": {
    "MaxCapacity": number,
    "MinCapacity": number,
    "ScaleInCooldown": number,
    "ScaleOutCooldown": number,
    "ScalingPolicies": [...]
  },
  "EndpointName": "string",
  "InferenceRecommendationsJobName": "string",
  "Metric": {
    "InvocationsPerInstance": number,
    "ModelLatency": number
  },
  "RecommendationId": "string",
  "ScalingPolicyObjective": {
    "MaxInvocationsPerMinute": number,
    "MinInvocationsPerMinute": number
  },
  "TargetCpuUtilizationPerCore": number
}
```

## Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

### DynamicScalingConfiguration (p. 630)

An object with the recommended values for you to specify when creating an autoscaling policy.

Type: DynamicScalingConfiguration (p. 1431) object
**EndpointName (p. 630)**

The name of an endpoint benchmarked during a previously completed Inference Recommender job.

- **Type:** String
- **Length Constraints:** Maximum length of 63.
- **Pattern:** `^\[a-zA-Z0-9](\-*[a-zA-Z0-9]){0,62}$`

**InferenceRecommendationsJobName (p. 630)**

The name of a previously completed Inference Recommender job.

- **Type:** String
- **Length Constraints:** Minimum length of 1. Maximum length of 64.
- **Pattern:** `^\[a-zA-Z0-9](\-*[a-zA-Z0-9]){0,63}$`

**Metric (p. 630)**

An object with a list of metrics that were benchmarked during the previously completed Inference Recommender job.

- **Type:** `ScalingPolicyMetric (p. 1932)` object

**RecommendationId (p. 630)**

The recommendation ID of a previously completed inference recommendation.

- **Type:** String

**ScalingPolicyObjective (p. 630)**

An object representing the anticipated traffic pattern for an endpoint that you specified in the request.

- **Type:** `ScalingPolicyObjective (p. 1933)` object

**TargetCpuUtilizationPerCore (p. 630)**

The percentage of how much utilization you want an instance to use before autoscaling, which you specified in the request. The default value is 50%.

- **Type:** Integer
- **Valid Range:** Minimum value of 1. Maximum value of 100.

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)].

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:
- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
GetSearchSuggestions
Service: Amazon SageMaker Service

An auto-complete API for the search functionality in the SageMaker console. It returns suggestions of possible matches for the property name to use in Search queries. Provides suggestions for HyperParameters, Tags, and Metrics.

Request Syntax

```
{
  "Resource": "string",
  "SuggestionQuery": {
    "PropertyNameQuery": {
      "PropertyNameHint": "string"
    }
  }
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**Resource (p. 633)**

The name of the SageMaker resource to search for.

Type: String

Valid Values: TrainingJob | Experiment | ExperimentTrial | ExperimentTrialComponent | Endpoint | ModelPackage | ModelPackageGroup | Pipeline | PipelineExecution | FeatureGroup | Project | FeatureMetadata | HyperParameterTuningJob | ModelCard | Model

Required: Yes

**SuggestionQuery (p. 633)**

Limits the property names that are included in the response.

Type: SuggestionQuery (p. 1974) object

Required: No

Response Syntax

```
{
  "PropertyNameSuggestions": [
    {
      "PropertyName": "string"
    }
  ]
}
```
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**PropertyNameSuggestions (p. 633)**

A list of property names for a Resource that match a SuggestionQuery.

Type: Array of PropertyNameSuggestion (p. 1874) objects

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ImportHubContent
Service: Amazon SageMaker Service

Import hub content.

**Note**
Hub APIs are only callable through SageMaker Studio.

**Request Syntax**

```
{
    "DocumentSchemaVersion": "string",
    "HubContentDescription": "string",
    "HubContentDisplayName": "string",
    "HubContentDocument": "string",
    "HubContentMarkdown": "string",
    "HubContentSearchKeywords": [ "string" ],
    "HubContentType": "string",
    "HubContentVersion": "string",
    "HubName": "string",
    "Tags": [
        {
            "Key": "string",
            "Value": "string"
        }
    ]
}
```

**Request Parameters**

For information about the parameters that are common to all actions, see [Common Parameters (p. 2178)](p. 2178).

The request accepts the following data in JSON format.

**DocumentSchemaVersion (p. 635)**

The version of the hub content schema to import.

*Type: String*


*Pattern: ^\d{1,4}\.\d{1,4}\.\d{1,4}$*

*Required: Yes*

**HubContentDescription (p. 635)**

A description of the hub content to import.

*Type: String*

*Length Constraints: Maximum length of 1023.*

*Pattern: .**

*Required: No*
**HubContentDisplayName (p. 635)**

The display name of the hub content to import.

Type: String

Length Constraints: Maximum length of 255.

Pattern: `.*`

Required: No

**HubContentDocument (p. 635)**

The hub content document that describes information about the hub content such as type, associated containers, scripts, and more.

Type: String

Length Constraints: Maximum length of 65535.

Pattern: `.*`

Required: Yes

**HubContentMarkdown (p. 635)**

A string that provides a description of the hub content. This string can include links, tables, and standard markdown formatting.

Type: String

Length Constraints: Maximum length of 65535.

Pattern: `.*`

Required: No

**HubContentName (p. 635)**

The name of the hub content to import.

Type: String

Length Constraints: Maximum length of 63.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$`

Required: Yes

**HubContentSearchKeywords (p. 635)**

The searchable keywords of the hub content.

Type: Array of strings

Array Members: Maximum number of 50 items.

Length Constraints: Maximum length of 255.

Pattern: `^[^A-Z]*$`

Required: No

**HubContentType (p. 635)**

The type of hub content to import.
Type: String

Valid Values: Model | Notebook

Required: Yes

**HubContentVersion (p. 635)**

The version of the hub content to import.

Type: String


Pattern: ^\d{1,4}\d{1,4}\d{1,4}$

Required: No

**HubName (p. 635)**

The name of the hub to import content into.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](\-*[a-zA-Z0-9])\{0,62}$

Required: Yes

**Tags (p. 635)**

Any tags associated with the hub content.

Type: Array of [Tag (p. 1979)] objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

**Response Syntax**

```json
{
    "HubArn": "string",
    "HubContentArn": "string"
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**HubArn (p. 637)**

The ARN of the hub that the content was imported into.

Type: String

Length Constraints: Maximum length of 255.

Pattern: .*
**HubContentArn (p. 637)**

The ARN of the hub content that was imported.

Type: String

Length Constraints: Maximum length of 255.

Pattern: . *

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](#).

**ResourceInUse**

Resource being accessed is in use.

HTTP Status Code: 400

**ResourceLimitExceeded**

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
ListActions
Service: Amazon SageMaker Service
Lists the actions in your account and their properties.

Request Syntax

```
{
   "ActionType": "string",
   "CreatedAfter": number,
   "CreatedBefore": number,
   "MaxResults": number,
   "NextToken": "string",
   "SortBy": "string",
   "SortOrder": "string",
   "SourceUri": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**ActionType (p. 639)**

A filter that returns only actions of the specified type.

Type: String

Length Constraints: Maximum length of 256.

Required: No

**CreatedAfter (p. 639)**

A filter that returns only actions created on or after the specified time.

Type: Timestamp

Required: No

**CreatedBefore (p. 639)**

A filter that returns only actions created on or before the specified time.

Type: Timestamp

Required: No

**MaxResults (p. 639)**

The maximum number of actions to return in the response. The default value is 10.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No
NextToken (p. 639)

If the previous call to ListActions didn’t return the full set of actions, the call returns a token for getting the next set of actions.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Required: No

SortBy (p. 639)

The property used to sort results. The default value is CreationTime.

Type: String

Valid Values: Name | CreationTime

Required: No

SortOrder (p. 639)

The sort order. The default value is Descending.

Type: String

Valid Values: Ascending | Descending

Required: No

SourceUri (p. 639)

A filter that returns only actions with the specified source URI.

Type: String

Length Constraints: Maximum length of 2048.

Pattern: .*

Required: No

Response Syntax

```json
{
  "ActionSummaries": [
    {,
      "ActionArn": "string",
      "ActionName": "string",
      "ActionType": "string",
      "CreationTime": number,
      "LastModifiedTime": number,
      "Source": {
        "SourceId": "string",
        "SourceType": "string",
        "SourceUri": "string"
      },
      "Status": "string"
    }
  ],
}
```
"NextToken": "string"
}

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**ActionSummaries (p. 640)**

A list of actions and their properties.

Type: Array of [ActionSummary (p. 1217)] objects

**NextToken (p. 640)**

A token for getting the next set of actions, if there are any.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Errors

For information about the errors that are common to all actions, see [Common Errors (p. 2180)].

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListAlgorithms

Service: Amazon SageMaker Service

Lists the machine learning algorithms that have been created.

Request Syntax

```json
{
    "CreationTimeAfter": number,
    "CreationTimeBefore": number,
    "MaxResults": number,
    "NameContains": "string",
    "NextToken": "string",
    "SortBy": "string",
    "SortOrder": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CreationTimeAfter (p. 642)**

A filter that returns only algorithms created after the specified time (timestamp).

Type: Timestamp

Required: No

**CreationTimeBefore (p. 642)**

A filter that returns only algorithms created before the specified time (timestamp).

Type: Timestamp

Required: No

**MaxResults (p. 642)**

The maximum number of algorithms to return in the response.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No

**NameContains (p. 642)**

A string in the algorithm name. This filter returns only algorithms whose name contains the specified string.

Type: String

Length Constraints: Maximum length of 63.

Pattern: [a-zA-Z0-9\-\_]+
Required: No

**NextToken (p. 642)**

If the response to a previous ListAlgorithms request was truncated, the response includes a `NextToken`. To retrieve the next set of algorithms, use the token in the next request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Required: No

**SortBy (p. 642)**

The parameter by which to sort the results. The default is `CreationTime`.

Type: String

Valid Values: Name | CreationTime

Required: No

**SortOrder (p. 642)**

The sort order for the results. The default is Ascending.

Type: String

Valid Values: Ascending | Descending

Required: No

**Response Syntax**

```
{
   "AlgorithmSummaryList": [
      {
         "AlgorithmArn": "string",
         "AlgorithmDescription": "string",
         "AlgorithmName": "string",
         "AlgorithmStatus": "string",
         "CreationTime": number
      }
   ],
   "NextToken": "string"
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**AlgorithmSummaryList (p. 643)**

> An array of `AlgorithmSummary` objects, each of which lists an algorithm.

Type: Array of `AlgorithmSummary (p. 1230)` objects
NextToken (p. 643)

If the response is truncated, SageMaker returns this token. To retrieve the next set of algorithms, use it in the subsequent request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: . *

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

• AWS Command Line Interface
• AWS SDK for .NET
• AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
ListAliases
Service: Amazon SageMaker Service

Lists the aliases of a specified image or image version.

Request Syntax

```
{
    "Alias": "string",
    "ImageName": "string",
    "MaxResults": number,
    "NextToken": "string",
    "Version": number
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**Alias (p. 645)**

The alias of the image version.

Type: String


Pattern: (?![\-])^([a-zA-Z0-9-_.]+)$

Required: No

**ImageName (p. 645)**

The name of the image.

Type: String


Pattern: ^[a-zA-Z0-9][-._]?[a-zA-Z0-9]{0,62}$

Required: Yes

**MaxResults (p. 645)**

The maximum number of aliases to return.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No

**NextToken (p. 645)**

If the previous call to ListAliases didn't return the full set of aliases, the call returns a token for retrieving the next set of aliases.
Type: String
Length Constraints: Maximum length of 8192.
Pattern: .*
Required: No

Version (p. 645)
The version of the image. If image version is not specified, the aliases of all versions of the image are listed.
Type: Integer
Valid Range: Minimum value of 0.
Required: No

Response Syntax

```
{  
"NextToken": "string",
"SageMakerImageVersionAliases": [ "string" ]
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.
The following data is returned in JSON format by the service.

NextToken (p. 646)
A token for getting the next set of aliases, if more aliases exist.
Type: String
Length Constraints: Maximum length of 8192.
Pattern: .*

SageMakerImageVersionAliases (p. 646)
A list of SageMaker image version aliases.
Type: Array of strings
Pattern: (?![.\-])^([a-zA-Z0-9-._]+)$

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceNotFound
Resource being access is not found.
HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListAppImageConfigs
Service: Amazon SageMaker Service

Lists the AppImageConfigs in your account and their properties. The list can be filtered by creation time or modified time, and whether the AppImageConfig name contains a specified string.

Request Syntax

```json
{
  "CreationTimeAfter": number,
  "CreationTimeBefore": number,
  "MaxResults": number,
  "ModifiedTimeAfter": number,
  "ModifiedTimeBefore": number,
  "NameContains": "string",
  "NextToken": "string",
  "SortBy": "string",
  "SortOrder": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CreationTimeAfter (p. 648)**

A filter that returns only AppImageConfigs created on or after the specified time.

Type: Timestamp

Required: No

**CreationTimeBefore (p. 648)**

A filter that returns only AppImageConfigs created on or before the specified time.

Type: Timestamp

Required: No

**MaxResults (p. 648)**

The total number of items to return in the response. If the total number of items available is more than the value specified, a NextToken is provided in the response. To resume pagination, provide the NextToken value in the as part of a subsequent call. The default value is 10.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No

**ModifiedTimeAfter (p. 648)**

A filter that returns only AppImageConfigs modified on or after the specified time.

Type: Timestamp
ListAppImageConfigs

**ModifiedTimeBefore (p. 648)**
A filter that returns only AppImageConfigs modified on or before the specified time.
Type: Timestamp
Required: No

**NameContains (p. 648)**
A filter that returns only AppImageConfigs whose name contains the specified string.
Type: String
Length Constraints: Maximum length of 63.
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}$
Required: No

**NextToken (p. 648)**
If the previous call to ListImages didn't return the full set of AppImageConfigs, the call returns a token for getting the next set of AppImageConfigs.
Type: String
Length Constraints: Maximum length of 8192.
Pattern: .*
Required: No

**SortBy (p. 648)**
The property used to sort results. The default value is CreationTime.
Type: String
Valid Values: CreationTime | LastModifiedTime | Name
Required: No

**SortOrder (p. 648)**
The sort order. The default value is Descending.
Type: String
Valid Values: Ascending | Descending
Required: No

**Response Syntax**

```json
{
    "AppImageConfigs": [
        {
            "AppImageConfigArn": "string",
            "AppImageConfigName": "string",
            "CreationTime": number,
            "JupyterLabAppImageConfig": {
```
"ContainerConfig": {
  "ContainerArguments": [ "string" ],
  "ContainerEntrypoint": [ "string" ],
  "ContainerEnvironmentVariables": {
    "string": "string"
  }
},
"KernelGatewayImageConfig": {
  "FileSystemConfig": {
    "DefaultGid": number,
    "DefaultUid": number,
    "MountPath": "string"
  },
  "KernelSpecs": [
    {
      "DisplayName": "string",
      "Name": "string"
    }
  ],
  "LastModifiedTime": number
},
"NextToken": "string"

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**AppImageConfigs (p. 649)**

A list of AppImageConfigs and their properties.

Type: Array of AppImageConfigDetails (p. 1248) objects

**NextToken (p. 649)**

A token for getting the next set of AppImageConfigs, if there are any.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
ListApps
Service: Amazon SageMaker Service
Lists apps.

Request Syntax

```json
{
    "DomainIdEquals": "string",
    "MaxResults": number,
    "NextToken": "string",
    "SortBy": "string",
    "SortOrder": "string",
    "SpaceNameEquals": "string",
    "UserProfileNameEquals": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**DomainIdEquals (p. 652)**
A parameter to search for the domain ID.
Type: String
Length Constraints: Maximum length of 63.
Required: No

**MaxResults (p. 652)**
The total number of items to return in the response. If the total number of items available is more than the value specified, a NextToken is provided in the response. To resume pagination, provide the NextToken value in the as part of a subsequent call. The default value is 10.
Type: Integer
Valid Range: Minimum value of 1. Maximum value of 100.
Required: No

**NextToken (p. 652)**
If the previous response was truncated, you will receive this token. Use it in your next request to receive the next set of results.
Type: String
Length Constraints: Maximum length of 8192.
Pattern: .*
Required: No

**SortBy (p. 652)**
The parameter by which to sort the results. The default is CreationTime.
Type: String
Valid Values: CreationTime
Required: No

**SortOrder (p. 652)**

The sort order for the results. The default is Ascending.

Type: String
Valid Values: Ascending | Descending
Required: No

**SpaceNameEquals (p. 652)**

A parameter to search by space name. If UserProfileNameEquals is set, then this value cannot be set.

Type: String
Length Constraints: Maximum length of 63.
Pattern: \^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}\]
Required: No

**UserProfileNameEquals (p. 652)**

A parameter to search by user profile name. If SpaceNameEquals is set, then this value cannot be set.

Type: String
Length Constraints: Maximum length of 63.
Pattern: \^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}\]
Required: No

**Response Syntax**

```json
{
  "Apps": [
    {
      "AppName": "string",
      "AppType": "string",
      "CreationTime": number,
      "DomainId": "string",
      "ResourceSpec": {
        "InstanceType": "string",
        "LifecycleConfigArn": "string",
        "SageMakerImageArn": "string",
        "SageMakerImageVersionAlias": "string",
        "SageMakerImageVersionArn": "string"
      },
      "SpaceName": "string",
      "Status": "string",
      "UserProfileName": "string"
    }
  ]
}
```
"NextToken": "string"
}

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**Apps (p. 653)**

The list of apps.

Type: Array of [AppDetails (p. 1246)] objects

**NextToken (p. 653)**

If the previous response was truncated, you will receive this token. Use it in your next request to receive the next set of results.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)].

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
ListArtifacts

Service: Amazon SageMaker Service

Lists the artifacts in your account and their properties.

Request Syntax

```json
{
   "ArtifactType": "string",
   "CreatedAfter": number,
   "CreatedBefore": number,
   "MaxResults": number,
   "NextToken": "string",
   "SortBy": "string",
   "SortOrder": "string",
   "SourceUri": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**ArtifactType (p. 655)**

A filter that returns only artifacts of the specified type.

Type: String

Length Constraints: Maximum length of 256.

Required: No

**CreatedAfter (p. 655)**

A filter that returns only artifacts created on or after the specified time.

Type: Timestamp

Required: No

**CreatedBefore (p. 655)**

A filter that returns only artifacts created on or before the specified time.

Type: Timestamp

Required: No

**MaxResults (p. 655)**

The maximum number of artifacts to return in the response. The default value is 10.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No
**NextToken (p. 655)**

If the previous call to ListArtifacts didn't return the full set of artifacts, the call returns a token for getting the next set of artifacts.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Required: No

**SortBy (p. 655)**

The property used to sort results. The default value is CreationTime.

Type: String

Valid Values: CreationTime

Required: No

**SortOrder (p. 655)**

The sort order. The default value is Descending.

Type: String

Valid Values: Ascending | Descending

Required: No

**SourceUri (p. 655)**

A filter that returns only artifacts with the specified source URI.

Type: String

Length Constraints: Maximum length of 2048.

Pattern: .*

Required: No

**Response Syntax**

```json
{
    "ArtifactSummaries": [
        {
            "ArtifactArn": "string",
            "ArtifactName": "string",
            "ArtifactType": "string",
            "CreationTime": number,
            "LastModifiedTime": number,
            "Source": {
                "SourceTypes": [
                    {
                        "SourceIdType": "string",
                        "Value": "string"
                    }
                ],
                "SourceUri": "string"
            }
        }
    ]
}
```
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**ArtifactSummaries (p. 656)**

A list of artifacts and their properties.

Type: Array of [ArtifactSummary (p. 1253)] objects

**NextToken (p. 656)**

A token for getting the next set of artifacts, if there are any.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: . *

Errors

For information about the errors that are common to all actions, see [Common Errors (p. 2180)].

**ResourceNotFoundException**

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
ListAssociations
Service: Amazon SageMaker Service

Lists the associations in your account and their properties.

Request Syntax

```
{
    "AssociationType": "string",
    "CreatedAfter": number,
    "CreatedBefore": number,
    "DestinationArn": "string",
    "DestinationType": "string",
    "MaxResults": number,
    "NextToken": "string",
    "SortBy": "string",
    "SortOrder": "string",
    "SourceArn": "string",
    "SourceType": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**AssociationType (p. 658)**

A filter that returns only associations of the specified type.

Type: String

Valid Values: ContributedTo | AssociatedWith | DerivedFrom | Produced

Required: No

**CreatedAfter (p. 658)**

A filter that returns only associations created on or after the specified time.

Type: Timestamp

Required: No

**CreatedBefore (p. 658)**

A filter that returns only associations created on or before the specified time.

Type: Timestamp

Required: No

**DestinationArn (p. 658)**

A filter that returns only associations with the specified destination Amazon Resource Name (ARN).

Type: String

Length Constraints: Maximum length of 256.
Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:(experiment|experiment-trial-component|artifact|action|context)/.*

Required: No

**DestinationType (p. 658)**

A filter that returns only associations with the specified destination type.

Type: String

Length Constraints: Maximum length of 256.

Required: No

**MaxResults (p. 658)**

The maximum number of associations to return in the response. The default value is 10.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No

**NextToken (p. 658)**

If the previous call to ListAssociations didn't return the full set of associations, the call returns a token for getting the next set of associations.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Required: No

**SortBy (p. 658)**

The property used to sort results. The default value is CreationTime.

Type: String

Valid Values: SourceArn | DestinationArn | SourceType | DestinationType | CreationTime

Required: No

**SortOrder (p. 658)**

The sort order. The default value is Descending.

Type: String

Valid Values: Ascending | Descending

Required: No

**SourceArn (p. 658)**

A filter that returns only associations with the specified source ARN.

Type: String

Length Constraints: Maximum length of 256.
ListAssociations

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:(experiment|experiment-trial-component|artifact|action|context)/.*

Required: No

**SourceType (p. 658)**

A filter that returns only associations with the specified source type.

Type: String

Length Constraints: Maximum length of 256.

Required: No

**Response Syntax**

```json
{
"AssociationSummaries": [
{
"AssociationType": "string",
"CreatedBy": {
"DomainId": "string",
"IamIdentity": {
"Arn": "string",
"PrincipalId": "string",
"SourceIdentity": "string"
},
"UserProfileArn": "string",
"UserProfileName": "string"
},
"CreationTime": number,
"DestinationArn": "string",
"DestinationName": "string",
"DestinationType": "string",
"SourceArn": "string",
"SourceName": "string",
"SourceType": "string"
}
],
"NextToken": "string"
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**AssociationSummaries (p. 660)**

A list of associations and their properties.

Type: Array of **AssociationSummary (p. 1255)** objects

**NextToken (p. 660)**

A token for getting the next set of associations, if there are any.

Type: String

Length Constraints: Maximum length of 8192.
Pattern: . *

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceNotFound

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListAutoMLJobs
Service: Amazon SageMaker Service

Request a list of jobs.

Request Syntax

```
{
    "CreationTimeAfter": number,
    "CreationTimeBefore": number,
    "LastModifiedTimeAfter": number,
    "LastModifiedTimeBefore": number,
    "MaxResults": number,
    "NameContains": "string",
    "NextToken": "string",
    "SortBy": "string",
    "SortOrder": "string",
    "StatusEquals": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CreationTimeAfter (p. 662)**

Request a list of jobs, using a filter for time.

Type: Timestamp

Required: No

**CreationTimeBefore (p. 662)**

Request a list of jobs, using a filter for time.

Type: Timestamp

Required: No

**LastModifiedTimeAfter (p. 662)**

Request a list of jobs, using a filter for time.

Type: Timestamp

Required: No

**LastModifiedTimeBefore (p. 662)**

Request a list of jobs, using a filter for time.

Type: Timestamp

Required: No

**MaxResults (p. 662)**

Request a list of jobs up to a specified limit.
Type: Integer
Valid Range: Minimum value of 1. Maximum value of 100.
Required: No

**NameContains (p. 662)**
Request a list of jobs, using a search filter for name.
Type: String
Length Constraints: Maximum length of 63.
Pattern: [a-zA-Z0-9\-\-]*
Required: No

**NextToken (p. 662)**
If the previous response was truncated, you receive this token. Use it in your next request to receive the next set of results.
Type: String
Length Constraints: Maximum length of 8192.
Pattern: .*
Required: No

**SortBy (p. 662)**
The parameter by which to sort the results. The default is Name.
Type: String
Valid Values: Name | CreationTime | Status
Required: No

**SortOrder (p. 662)**
The sort order for the results. The default is Descending.
Type: String
Valid Values: Ascending | Descending
Required: No

**StatusEquals (p. 662)**
Request a list of jobs, using a filter for status.
Type: String
Valid Values: Completed | InProgress | Failed | Stopped | Stopping
Required: No

**Response Syntax**

```
"AutoMLJobSummaries": [
    {
      "AutoMLJobArn": "string",
      "AutoMLJobName": "string",
      "AutoMLJobSecondaryStatus": "string",
      "AutoMLJobStatus": "string",
      "CreationTime": number,
      "EndTime": number,
      "FailureReason": "string",
      "LastModifiedTime": number,
      "PartialFailureReasons": [
        {
          "PartialFailureMessage": "string"
        }
      ]
    }
  ],
"NextToken": "string"
}

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**AutoMLJobSummaries (p. 663)**

Returns a summary list of jobs.

Type: Array of **AutoMLJobSummary (p. 1287)** objects

**NextToken (p. 663)**

If the previous response was truncated, you receive this token. Use it in your next request to receive the next set of results.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Errors

For information about the errors that are common to all actions, see **Common Errors (p. 2180)**.

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
• AWS SDK for Python
• AWS SDK for Ruby V3
ListCandidatesForAutoMLJob

Service: Amazon SageMaker Service

List the candidates created for the job.

Request Syntax

```
{
  "AutoMLJobName": "string",
  "CandidateNameEquals": "string",
  "MaxResults": number,
  "NextToken": "string",
  "SortBy": "string",
  "SortOrder": "string",
  "StatusEquals": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**AutoMLJobName (p. 666)**

List the candidates created for the job by providing the job's name.

Type: String


Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,31}\]`  

Required: Yes

**CandidateNameEquals (p. 666)**

List the candidates for the job and filter by candidate name.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 64.

Required: No

**MaxResults (p. 666)**

List the job's candidates up to a specified limit.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No

**NextToken (p. 666)**

If the previous response was truncated, you receive this token. Use it in your next request to receive the next set of results.

Type: String
Length Constraints: Maximum length of 8192.
Pattern: .*
Required: No

**SortBy (p. 666)**
The parameter by which to sort the results. The default is Descending.
Type: String
Valid Values: CreationTime | Status | FinalObjectiveMetricValue
Required: No

**SortOrder (p. 666)**
The sort order for the results. The default is Ascending.
Type: String
Valid Values: Ascending | Descending
Required: No

**StatusEquals (p. 666)**
List the candidates for the job and filter by status.
Type: String
Valid Values: Completed | InProgress | Failed | Stopped | Stopping
Required: No

**Response Syntax**

```json
{
    "Candidates": [
        {
            "CandidateName": "string",
            "CandidateProperties": {
                "CandidateArtifactLocations": {
                    "BacktestResults": "string",
                    "Explainability": "string",
                    "ModelInsights": "string"
                },
                "CandidateMetrics": [
                    {
                        "MetricName": "string",
                        "Set": "string",
                        "StandardMetricName": "string",
                        "Value": "number"
                    }
                ],
                "CandidateStatus": "string",
                "CandidateSteps": [
                    {
                        "CandidateStepArn": "string",
                        "CandidateStepName": "string",
                        "CandidateStepType": "string"
                    }
                ]
            }
        }
    ]
}
```

667
"CreationTime": number,
"EndTime": number,
"FailureReason": "string",
"FinalAutoMLJobObjectiveMetric": {
  "MetricName": "string",
  "StandardMetricName": "string",
  "Type": "string",
  "Value": number
},
"InferenceContainerDefinitions": {
  "string": [
    {
      "Environment": {
        "string": "string"
      },
      "Image": "string",
      "ModelDataUrl": "string"
    }
  ]
},
"InferenceContainers": [
  {
    "Environment": {
      "string": "string"
    },
    "Image": "string",
    "ModelDataUrl": "string"
  }
],
"LastModifiedTime": number,
"ObjectiveStatus": "string"
],
"NextToken": "string"

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

Candidates (p. 667)

Summaries about the AutoMLCandidates.

Type: Array of AutoMLCandidate (p. 1265) objects

NextToken (p. 667)

If the previous response was truncated, you receive this token. Use it in your next request to receive the next set of results.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).
ResourceNotFound

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListClusterNodes
Service: Amazon SageMaker Service

 Retrieves the list of instances (also called nodes interchangeably) in a SageMaker HyperPod cluster.

Request Syntax

```
{
   "ClusterName": "string",
   "CreationTimeAfter": number,
   "CreationTimeBefore": number,
   "InstanceGroupNameContains": "string",
   "MaxResults": number,
   "NextToken": "string",
   "SortBy": "string",
   "SortOrder": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see [Common Parameters](p. 2178).

The request accepts the following data in JSON format.

**ClusterName (p. 670)**

The string name or the Amazon Resource Name (ARN) of the SageMaker HyperPod cluster in which you want to retrieve the list of nodes.

Type: String

Length Constraints: Maximum length of 256.

Pattern: ^arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:cluster/[a-zA-Z0-9]{12}$

Required: Yes

**CreationTimeAfter (p. 670)**

A filter that returns nodes in a SageMaker HyperPod cluster created after the specified time. Timestamps are formatted according to the ISO 8601 standard.

Acceptable formats include:

- YYYY-MM-DDThh:mm:ss.sssTZD (UTC), for example, 2014-10-01T20:30:00.000Z
- YYYY-MM-DDThh:mm:ss.sssTZD (with offset), for example, 2014-10-01T12:30:00.000-08:00
- YYYY-MM-DD, for example, 2014-10-01

Unix time in seconds, for example, 1412195400. This is also referred to as Unix Epoch time and represents the number of seconds since midnight, January 1, 1970 UTC.

For more information about the timestamp format, see [Timestamp](p. 2178) in the [AWS Command Line Interface User Guide](p. 2178).

Type: Timestamp

Required: No
**CreationTimeBefore (p. 670)**
A filter that returns nodes in a SageMaker HyperPod cluster created before the specified time. The acceptable formats are the same as the timestamp formats for CreationTimeAfter. For more information about the timestamp format, see Timestamp in the AWS Command Line Interface User Guide.

Type: Timestamp
Required: No

**InstanceGroupNameContains (p. 670)**
A filter that returns the instance groups whose name contain a specified string.

Type: String
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])*$
Required: No

**MaxResults (p. 670)**
The maximum number of nodes to return in the response.

Type: Integer
Valid Range: Minimum value of 1. Maximum value of 100.
Required: No

**NextToken (p. 670)**
If the result of the previous ListClusterNodes request was truncated, the response includes a NextToken. To retrieve the next set of cluster nodes, use the token in the next request.

Type: String
Length Constraints: Maximum length of 8192.
Pattern: .*
Required: No

**sortBy (p. 670)**
The field by which to sort results. The default value is CREATION_TIME.

Type: String
Valid Values: CREATION_TIME | NAME

Required: No

**sortOrder (p. 670)**
The sort order for results. The default value is Ascending.

Type: String
Valid Values: Ascending | Descending

Required: No
Response Syntax

```json
{
    "ClusterNodeSummaries": [
        {
            "InstanceGroupName": "string",
            "InstanceId": "string",
            "InstanceStatus": {
                "Message": "string",
                "Status": "string"
            },
            "InstanceType": "string",
            "LaunchTime": "number"
        }
    ],
    "NextToken": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**ClusterNodeSummaries (p. 672)**

The summaries of listed instances in a SageMaker HyperPod cluster

Type: Array of [ClusterNodeSummary (p. 1349)] objects

**NextToken (p. 672)**

The next token specified for listing instances in a SageMaker HyperPod cluster.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Errors

For information about the errors that are common to all actions, see [Common Errors (p. 2180)]

**ResourceNot Found**

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
Amazon SageMaker Amazon Sagemaker API Reference
ListClusterNodes

- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListClusters
Service: Amazon SageMaker Service
Retrieves the list of SageMaker HyperPod clusters.

Request Syntax

```
{
    "CreationTimeAfter": number,
    "CreationTimeBefore": number,
    "MaxResults": number,
    "NameContains": "string",
    "NextToken": "string",
    "SortBy": "string",
    "SortOrder": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CreationTimeAfter (p. 674)**

Set a start time for the time range during which you want to list SageMaker HyperPod clusters. Timestamps are formatted according to the ISO 8601 standard.

Acceptable formats include:
- YYYY-MM-DDThh:mm:ss.sssTZD (UTC), for example, 2014-10-01T20:30:00.000Z
- YYYY-MM-DDThh:mm:ss.sssTZD (with offset), for example, 2014-10-01T12:30:00.000-08:00
- YYYY-MM-DD, for example, 2014-10-01
- Unix time in seconds, for example, 1412195400. This is also referred to as Unix Epoch time and represents the number of seconds since midnight, January 1, 1970 UTC.

For more information about the timestamp format, see Timestamp in the AWS Command Line Interface User Guide.

Type: Timestamp

Required: No

**CreationTimeBefore (p. 674)**

Set an end time for the time range during which you want to list SageMaker HyperPod clusters. A filter that returns nodes in a SageMaker HyperPod cluster created before the specified time. The acceptable formats are the same as the timestamp formats for CreationTimeAfter. For more information about the timestamp format, see Timestamp in the AWS Command Line Interface User Guide.

Type: Timestamp

Required: No

**MaxResults (p. 674)**

Set the maximum number of SageMaker HyperPod clusters to list.
ListClusters

Type: Integer
Valid Range: Minimum value of 1. Maximum value of 100.
Required: No

NameContains (p. 674)
Set the maximum number of instances to print in the list.
Type: String
Length Constraints: Maximum length of 63.
Pattern: [a-zA-Z0-9-\-]+
Required: No

NextToken (p. 674)
Set the next token to retrieve the list of SageMaker HyperPod clusters.
Type: String
Length Constraints: Maximum length of 8192.
Pattern: .*
Required: No

SortBy (p. 674)
The field by which to sort results. The default value is CREATION_TIME.
Type: String
Valid Values: CREATION_TIME | NAME
Required: No

SortOrder (p. 674)
The sort order for results. The default value is Ascending.
Type: String
Valid Values: Ascending | Descending
Required: No

Response Syntax

```
{
   "ClusterSummaries": [
      {
         "ClusterArn": "string",
         "ClusterName": "string",
         "ClusterStatus": "string",
         "CreationTime": number
      }
   ],
   "NextToken": "string"
}
```
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**ClusterSummaries (p. 675)**

The summaries of listed SageMaker HyperPod clusters.

Type: Array of [ClusterSummary (p. 1351)] objects

**NextToken (p. 675)**

If the result of the previous ListClusters request was truncated, the response includes a NextToken. To retrieve the next set of clusters, use the token in the next request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Errors

For information about the errors that are common to all actions, see [Common Errors (p. 2180)].

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface]
- [AWS SDK for .NET]
- [AWS SDK for C++]
- [AWS SDK for Go]
- [AWS SDK for Java V2]
- [AWS SDK for JavaScript V3]
- [AWS SDK for PHP V3]
- [AWS SDK for Python]
- [AWS SDK for Ruby V3]
ListCodeRepositories

Service: Amazon SageMaker Service

Gets a list of the Git repositories in your account.

Request Syntax

```json
{
  "CreationTimeAfter": number,
  "CreationTimeBefore": number,
  "LastModifiedTimeAfter": number,
  "LastModifiedTimeBefore": number,
  "MaxResults": number,
  "NameContains": "string",
  "NextToken": "string",
  "SortBy": "string",
  "SortOrder": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see [Common Parameters](p. 2178).

The request accepts the following data in JSON format.

- **CreationTimeAfter (p. 677)**
  
  A filter that returns only Git repositories that were created after the specified time.
  
  Type: Timestamp
  
  Required: No

- **CreationTimeBefore (p. 677)**
  
  A filter that returns only Git repositories that were created before the specified time.
  
  Type: Timestamp
  
  Required: No

- **LastModifiedTimeAfter (p. 677)**
  
  A filter that returns only Git repositories that were last modified after the specified time.
  
  Type: Timestamp
  
  Required: No

- **LastModifiedTimeBefore (p. 677)**
  
  A filter that returns only Git repositories that were last modified before the specified time.
  
  Type: Timestamp
  
  Required: No

- **MaxResults (p. 677)**
  
  The maximum number of Git repositories to return in the response.
Type: Integer
Valid Range: Minimum value of 1. Maximum value of 100.
Required: No
NameContains (p. 677)
A string in the Git repositories name. This filter returns only repositories whose name contains the specified string.
Type: String
Length Constraints: Maximum length of 63.
Pattern: [a-zA-Z0-9-]+
Required: No
NextToken (p. 677)
If the result of a ListCodeRepositoriesOutput request was truncated, the response includes a NextToken. To get the next set of Git repositories, use the token in the next request.
Type: String
Length Constraints: Maximum length of 8192.
Pattern: .*
Required: No
SortBy (p. 677)
The field to sort results by. The default is Name.
Type: String
Valid Values: Name | CreationTime | LastModifiedTime
Required: No
SortOrder (p. 677)
The sort order for results. The default is Ascending.
Type: String
Valid Values: Ascending | Descending
Required: No

Response Syntax

```
{
    "CodeRepositorySummaryList": [
        {
            "CodeRepositoryArn": "string",
            "CodeRepositoryName": "string",
            "CreationTime": number,
            "GitConfig": {
                "Branch": "string",
                "RepositoryUrl": "string"
            }
        }
    ]
}
```
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**CodeRepositorySummaryList (p. 678)**

Gets a list of summaries of the Git repositories. Each summary specifies the following values for the repository:

- Name
- Amazon Resource Name (ARN)
- Creation time
- Last modified time
- Configuration information, including the URL location of the repository and the ARN of the AWS Secrets Manager secret that contains the credentials used to access the repository.

Type: Array of CodeRepositorySummary (p. 1355) objects

**NextToken (p. 678)**

If the result of a ListCodeRepositoriesOutput request was truncated, the response includes a NextToken. To get the next set of Git repositories, use the token in the next request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListCompilationJobs
Service: Amazon SageMaker Service

Lists model compilation jobs that satisfy various filters.

To create a model compilation job, use CreateCompilationJob. To get information about a particular model compilation job you have created, use DescribeCompilationJob.

Request Syntax

```json
{
    "CreationTimeAfter": number,
    "CreationTimeBefore": number,
    "LastModifiedTimeAfter": number,
    "LastModifiedTimeBefore": number,
    "MaxResults": number,
    "NameContains": "string",
    "NextToken": "string",
    "SortBy": "string",
    "SortOrder": "string",
    "StatusEquals": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

CreationTimeAfter (p. 681)

A filter that returns the model compilation jobs that were created after a specified time.

Type: Timestamp

Required: No

CreationTimeBefore (p. 681)

A filter that returns the model compilation jobs that were created before a specified time.

Type: Timestamp

Required: No

LastModifiedTimeAfter (p. 681)

A filter that returns the model compilation jobs that were modified after a specified time.

Type: Timestamp

Required: No

LastModifiedTimeBefore (p. 681)

A filter that returns the model compilation jobs that were modified before a specified time.

Type: Timestamp

Required: No
MaxResults (p. 681)
The maximum number of model compilation jobs to return in the response.
Type: Integer
Valid Range: Minimum value of 1. Maximum value of 100.
Required: No

NameContains (p. 681)
A filter that returns the model compilation jobs whose name contains a specified string.
Type: String
Length Constraints: Maximum length of 63.
Pattern: [a-zA-Z0-9\-]+
Required: No

NextToken (p. 681)
If the result of the previous ListCompilationJobs request was truncated, the response includes a NextToken. To retrieve the next set of model compilation jobs, use the token in the next request.
Type: String
Length Constraints: Maximum length of 8192.
Pattern: .*
Required: No

SortBy (p. 681)
The field by which to sort results. The default is CreationTime.
Type: String
Valid Values: Name | CreationTime | Status
Required: No

SortOrder (p. 681)
The sort order for results. The default is Ascending.
Type: String
Valid Values: Ascending | Descending
Required: No

StatusEquals (p. 681)
A filter that retrieves model compilation jobs with a specific CompilationJobStatus status.
Type: String
Valid Values: INPROGRESS | COMPLETED | FAILED | STARTING | STOPPING | STOPPED
Required: No
Response Syntax

```
{
    "CompilationJobSummaries": [
        {
            "CompilationEndTime": number,
            "CompilationJobArn": "string",
            "CompilationJobName": "string",
            "CompilationJobStatus": "string",
            "CompilationStartTime": number,
            "CompilationTargetDevice": "string",
            "CompilationTargetPlatformAccelerator": "string",
            "CompilationTargetPlatformArch": "string",
            "CompilationTargetPlatformOs": "string",
            "CreationTime": number,
            "LastModifiedTime": number
        }
    ],
    "NextToken": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**CompilationJobSummaries (p. 683)**

An array of [CompilationJobSummary](p. 1361) objects, each describing a model compilation job.

Type: Array of [CompilationJobSummary](p. 1361) objects

**NextToken (p. 683)**

If the response is truncated, Amazon SageMaker returns this NextToken. To retrieve the next set of model compilation jobs, use this token in the next request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Errors

For information about the errors that are common to all actions, see [Common Errors (p. 2180)].

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](p. 2180)
- [AWS SDK for .NET](p. 2180)
- [AWS SDK for C++](p. 2180)
- [AWS SDK for Go](p. 2180)
- [AWS SDK for Java V2](p. 2180)
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
**ListContexts**

Service: Amazon SageMaker Service

Lists the contexts in your account and their properties.

**Request Syntax**

```
{
    "ContextType": "string",
    "CreatedAfter": number,
    "CreatedBefore": number,
    "MaxResults": number,
    "NextToken": "string",
    "SortBy": "string",
    "SortOrder": "string",
    "SourceUri": "string"
}
```

**Request Parameters**

For information about the parameters that are common to all actions, see [Common Parameters](#).

The request accepts the following data in JSON format.

**ContextType (p. 685)**

A filter that returns only contexts of the specified type.

Type: String

Length Constraints: Maximum length of 256.

Required: No

**CreatedAfter (p. 685)**

A filter that returns only contexts created on or after the specified time.

Type: Timestamp

Required: No

**CreatedBefore (p. 685)**

A filter that returns only contexts created on or before the specified time.

Type: Timestamp

Required: No

**MaxResults (p. 685)**

The maximum number of contexts to return in the response. The default value is 10.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No
NextToken (p. 685)

If the previous call to ListContexts didn't return the full set of contexts, the call returns a token for getting the next set of contexts.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Required: No

SortBy (p. 685)

The property used to sort results. The default value is CreationTime.

Type: String

Valid Values: Name | CreationTime

Required: No

SortOrder (p. 685)

The sort order. The default value is Descending.

Type: String

Valid Values: Ascending | Descending

Required: No

SourceUri (p. 685)

A filter that returns only contexts with the specified source URI.

Type: String

Length Constraints: Maximum length of 2048.

Pattern: .*

Required: No

Response Syntax

```json
{
    "ContextSummaries": [
        {
            "ContextArn": "string",
            "ContextName": "string",
            "ContextType": "string",
            "CreationTime": number,
            "LastModifiedTime": number,
            "Source": {
                "SourceId": "string",
                "SourceType": "string",
                "SourceUri": "string"
            }
        }
    ],
    "NextToken": "string"
}
```
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**ContextSummaries (p. 686)**

A list of contexts and their properties.

Type: Array of [ContextSummary (p. 1370)] objects

**NextToken (p. 686)**

A token for getting the next set of contexts, if there are any.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Errors

For information about the errors that are common to all actions, see [Common Errors (p. 2180)].

**ResourceNotFoundException**

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- [AWS SDK for C++]
- AWS SDK for Go
- AWS SDK for Java V2
- [AWS SDK JavaScript V3]
- AWS SDK for PHP V3
- AWS SDK for Python
- [AWS SDK for Ruby V3]
ListDataQualityJobDefinitions

Service: Amazon SageMaker Service

Lists the data quality job definitions in your account.

Request Syntax

```json
{
   "CreationTimeAfter": number,
   "CreationTimeBefore": number,
   "EndpointName": "string",
   "MaxResults": number,
   "NameContains": "string",
   "NextToken": "string",
   "SortBy": "string",
   "SortOrder": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see [Common Parameters](p. 2178).

The request accepts the following data in JSON format.

**CreationTimeAfter (p. 688)**

A filter that returns only data quality monitoring job definitions created after the specified time.

Type: Timestamp

Required: No

**CreationTimeBefore (p. 688)**

A filter that returns only data quality monitoring job definitions created before the specified time.

Type: Timestamp

Required: No

**EndpointName (p. 688)**

A filter that lists the data quality job definitions associated with the specified endpoint.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

Required: No

**MaxResults (p. 688)**

The maximum number of data quality monitoring job definitions to return in the response.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.
Required: No

**NameContains (p. 688)**

A string in the data quality monitoring job definition name. This filter returns only data quality monitoring job definitions whose name contains the specified string.

Type: String

Length Constraints: Maximum length of 63.

Pattern: [a-zA-Z0-9\-\_]+

Required: No

**NextToken (p. 688)**

If the result of the previous ListDataQualityJobDefinitions request was truncated, the response includes a NextToken. To retrieve the next set of transform jobs, use the token in the next request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Required: No

**SortBy (p. 688)**

The field to sort results by. The default is CreationTime.

Type: String

Valid Values: Name  |  CreationTime

Required: No

**SortOrder (p. 688)**

Whether to sort the results in Ascending or Descending order. The default is Descending.

Type: String

Valid Values: Ascending  |  Descending

Required: No

**Response Syntax**

```
{
   "JobDefinitionSummaries": [
      {
         "CreationTime": number,
         "EndpointName": "string",
         "MonitoringJobDefinitionArn": "string",
         "MonitoringJobDefinitionName": "string"
      }
   ],
   "NextToken": "string"
}
```
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**JobDefinitionSummaries (p. 689)**

A list of data quality monitoring job definitions.

Type: Array of MonitoringJobDefinitionSummary (p. 1749) objects

**NextToken (p. 689)**

If the result of the previous ListDataQualityJobDefinitions request was truncated, the response includes a NextToken. To retrieve the next set of data quality monitoring job definitions, use the token in the next request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: . *

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListDeviceFleets
Service: Amazon SageMaker Service

Returns a list of devices in the fleet.

Request Syntax

```json
{
  "CreationTimeAfter": number,
  "CreationTimeBefore": number,
  "LastModifiedTimeAfter": number,
  "LastModifiedTimeBefore": number,
  "MaxResults": number,
  "NameContains": "string",
  "NextToken": "string",
  "SortBy": "string",
  "SortOrder": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CreationTimeAfter (p. 691)**

Filter fleets where packaging job was created after specified time.

Type: Timestamp

Required: No

**CreationTimeBefore (p. 691)**

Filter fleets where the edge packaging job was created before specified time.

Type: Timestamp

Required: No

**LastModifiedTimeAfter (p. 691)**

Select fleets where the job was updated after X

Type: Timestamp

Required: No

**LastModifiedTimeBefore (p. 691)**

Select fleets where the job was updated before X

Type: Timestamp

Required: No

**MaxResults (p. 691)**

The maximum number of results to select.
Type: Integer
Valid Range: Maximum value of 100.
Required: No
**NameContains (p. 691)**
Filter for fleets containing this name in their fleet device name.
Type: String
Length Constraints: Maximum length of 63.
Pattern: [a-zA-Z0-9\-]+
Required: No
**NextToken (p. 691)**
The response from the last list when returning a list large enough to need tokening.
Type: String
Length Constraints: Maximum length of 8192.
Pattern: .*
Required: No
**SortBy (p. 691)**
The column to sort by.
Type: String
Valid Values: NAME | CREATION_TIME | LAST_MODIFIED_TIME
Required: No
**SortOrder (p. 691)**
What direction to sort in.
Type: String
Valid Values: Ascending | Descending
Required: No

**Response Syntax**

```
{
    "DeviceFleetSummaries": [
        {
            "CreationTime": number,
            "DeviceFleetArn": "string",
            "DeviceFleetName": "string",
            "LastModifiedTime": number
        }
    ],
    "NextToken": "string"
}
```
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**DeviceFleetSummaries (p. 692)**

Summary of the device fleet.

Type: Array of DeviceFleetSummary (p. 1415) objects

**NextToken (p. 692)**

The response from the last list when returning a list large enough to need tokening.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListDevices
Service: Amazon SageMaker Service
A list of devices.

Request Syntax

```
{
  "DeviceFleetName": "string",
  "LatestHeartbeatAfter": number,
  "MaxResults": number,
  "ModelName": "string",
  "NextToken": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**DeviceFleetName (p. 694)**
 Filter for fleets containing this name in their device fleet name.

Type: String


Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$`

Required: No

**LatestHeartbeatAfter (p. 694)**
 Select fleets where the job was updated after X

Type: Timestamp

Required: No

**MaxResults (p. 694)**
 Maximum number of results to select.

Type: Integer

Valid Range: Maximum value of 100.

Required: No

**ModelName (p. 694)**
 A filter that searches devices that contains this name in any of their models.

Type: String

ListDevices

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62\}$

Required: No

NextToken (p. 694)

The response from the last list when returning a list large enough to need tokening.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Required: No

Response Syntax

```
{
    "DeviceSummaries": [
        {
            "AgentVersion": "string",
            "Description": "string",
            "DeviceArn": "string",
            "DeviceFleetName": "string",
            "DeviceName": "string",
            "IotThingName": "string",
            "LatestHeartbeat": number,
            "Models": [
                {
                    "ModelName": "string",
                    "ModelVersion": "string"
                }
            ],
            "RegistrationTime": number
        },
        {
            "NextToken": "string"
        }
    ]
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

DeviceSummaries (p. 695)

Summary of devices.

Type: Array of DeviceSummary (p. 1419) objects

NextToken (p. 695)

The response from the last list when returning a list large enough to need tokening.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*
Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
**ListDomains**

Service: Amazon SageMaker Service

Lists the domains.

**Request Syntax**

```json
{
  "MaxResults": number,
  "NextToken": "string"
}
```

**Request Parameters**

For information about the parameters that are common to all actions, see [Common Parameters](p. 2178).

The request accepts the following data in JSON format.

**MaxResults (p. 697)**

The total number of items to return in the response. If the total number of items available is more than the value specified, a NextToken is provided in the response. To resume pagination, provide the NextToken value in the as part of a subsequent call. The default value is 10.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No

**NextToken (p. 697)**

If the previous response was truncated, you will receive this token. Use it in your next request to receive the next set of results.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Required: No

**Response Syntax**

```json
{
  "Domains": [
    {
      "CreationTime": number,
      "DomainArn": "string",
      "DomainId": "string",
      "DomainName": "string",
      "LastModifiedTime": number,
      "Status": "string",
      "Url": "string"
    }
  ]
}
```
"NextToken": "string"
}

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

Domains (p. 697)

The list of domains.

Type: Array of DomainDetails (p. 1422) objects

NextToken (p. 697)

If the previous response was truncated, you will receive this token. Use it in your next request to receive the next set of results.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3

698
ListEdgeDeploymentPlans
Service: Amazon SageMaker Service

Lists all edge deployment plans.

Request Syntax

```json
{
   "CreationTimeAfter": number,
   "CreationTimeBefore": number,
   "DeviceFleetNameContains": "string",
   "LastModifiedTimeAfter": number,
   "LastModifiedTimeBefore": number,
   "MaxResults": number,
   "NameContains": "string",
   "NextToken": "string",
   "SortBy": "string",
   "SortOrder": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CreationTimeAfter (p. 699)**

Selects edge deployment plans created after this time.

Type: Timestamp

Required: No

**CreationTimeBefore (p. 699)**

Selects edge deployment plans created before this time.

Type: Timestamp

Required: No

**DeviceFleetNameContains (p. 699)**

Selects edge deployment plans with a device fleet name containing this name.

Type: String

Length Constraints: Maximum length of 63.

Pattern: [a-zA-Z0-9\-]+

Required: No

**LastModifiedTimeAfter (p. 699)**

Selects edge deployment plans that were last updated after this time.

Type: Timestamp

699
Required: No

**LastModifiedTimeBefore (p. 699)**

Selects edge deployment plans that were last updated before this time.

Type: Timestamp

Required: No

**MaxResults (p. 699)**

The maximum number of results to select (50 by default).

Type: Integer

Valid Range: Maximum value of 100.

Required: No

**NameContains (p. 699)**

Selects edge deployment plans with names containing this name.

Type: String

Length Constraints: Maximum length of 63.

Pattern: [a-zA-Z0-9-\-]+

Required: No

**NextToken (p. 699)**

The response from the last list when returning a list large enough to need tokening.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Required: No

**SortBy (p. 699)**

The column by which to sort the edge deployment plans. Can be one of NAME, DEVICE_FLEET_NAME, CREATION_TIME, LAST_MODIFIED_TIME.

Type: String

Valid Values: NAME | DEVICE_FLEET_NAME | CREATION_TIME | LAST_MODIFIED_TIME

Required: No

**SortOrder (p. 699)**

The direction of the sorting (ascending or descending).

Type: String

Valid Values: Ascending | Descending

Required: No
Response Syntax

```json
{
   "EdgeDeploymentPlanSummaries": [
      {
         "CreationTime": number,
         "DeviceFleetName": "string",
         "EdgeDeploymentFailed": number,
         "EdgeDeploymentPending": number,
         "EdgeDeploymentPlanArn": "string",
         "EdgeDeploymentPlanName": "string",
         "EdgeDeploymentSuccess": number,
         "LastModifiedTime": number
      }
   ],
   "NextToken": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**EdgeDeploymentPlanSummaries (p. 701)**

List of summaries of edge deployment plans.

Type: Array of EdgeDeploymentPlanSummary (p. 1436) objects

**NextToken (p. 701)**

The token to use when calling the next page of results.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
• AWS SDK for Ruby V3
ListEdgePackagingJobs
Service: Amazon SageMaker Service

Returns a list of edge packaging jobs.

Request Syntax

```
{
    "CreationTimeAfter": number,
    "CreationTimeBefore": number,
    "LastModifiedTimeAfter": number,
    "LastModifiedTimeBefore": number,
    "MaxResults": number,
    "ModelNameContains": "string",
    "NameContains": "string",
    "NextToken": "string",
    "SortBy": "string",
    "SortOrder": "string",
    "StatusEquals": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CreationTimeAfter (p. 703)**

Select jobs where the job was created after specified time.

Type: Timestamp

Required: No

**CreationTimeBefore (p. 703)**

Select jobs where the job was created before specified time.

Type: Timestamp

Required: No

**LastModifiedTimeAfter (p. 703)**

Select jobs where the job was updated after specified time.

Type: Timestamp

Required: No

**LastModifiedTimeBefore (p. 703)**

Select jobs where the job was updated before specified time.

Type: Timestamp

Required: No

**MaxResults (p. 703)**

Maximum number of results to select.
Type: Integer
Valid Range: Maximum value of 100.
Required: No

**ModelNameContains (p. 703)**
Filter for jobs where the model name contains this string.
Type: String
Length Constraints: Maximum length of 63.
Pattern: [a-zA-Z0-9\-\-]+
Required: No

**NameContains (p. 703)**
Filter for jobs containing this name in their packaging job name.
Type: String
Length Constraints: Maximum length of 63.
Pattern: [a-zA-Z0-9\-\-]+
Required: No

**NextToken (p. 703)**
The response from the last list when returning a list large enough to need tokening.
Type: String
Length Constraints: Maximum length of 8192.
Pattern: .*
Required: No

**SortBy (p. 703)**
Use to specify what column to sort by.
Type: String
Valid Values: NAME | MODEL_NAME | CREATION_TIME | LAST_MODIFIED_TIME | STATUS
Required: No

**SortOrder (p. 703)**
What direction to sort by.
Type: String
Valid Values: Ascending | Descending
Required: No

**StatusEquals (p. 703)**
The job status to filter for.
Type: String
Valid Values: STARTING | INPROGRESS | COMPLETED | FAILED | STOPPING | STOPPED

Required: No

Response Syntax

```
{
    "EdgePackagingJobSummaries": [
    {
        "CompilationJobName": "string",
        "CreationTime": number,
        "EdgePackagingJobArn": "string",
        "EdgePackagingJobName": "string",
        "EdgePackagingJobStatus": "string",
        "LastModifiedTime": number,
        "ModelName": "string",
        "ModelVersion": "string"
    }
    ],
    "NextToken": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**EdgePackagingJobSummaries (p. 705)**

Summaries of edge packaging jobs.

Type: Array of **EdgePackagingJobSummary (p. 1446)** objects

**NextToken (p. 705)**

Token to use when calling the next page of results.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Errors

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](#).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
ListEndpointConfigs

Service: Amazon SageMaker Service

Lists endpoint configurations.

Request Syntax

```json
{
  "CreationTimeAfter": number,
  "CreationTimeBefore": number,
  "MaxResults": number,
  "NameContains": "string",
  "NextToken": "string",
  "SortBy": "string",
  "SortOrder": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CreationTimeAfter (p. 707)**

A filter that returns only endpoint configurations with a creation time greater than or equal to the specified time (timestamp).

Type: Timestamp

Required: No

**CreationTimeBefore (p. 707)**

A filter that returns only endpoint configurations created before the specified time (timestamp).

Type: Timestamp

Required: No

**MaxResults (p. 707)**

The maximum number of training jobs to return in the response.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No

**NameContains (p. 707)**

A string in the endpoint configuration name. This filter returns only endpoint configurations whose name contains the specified string.

Type: String

Length Constraints: Maximum length of 63.
Pattern: [a-zA-Z0-9-]+  
Required: No  

**NextToken (p. 707)**  
If the result of the previous ListEndpointConfig request was truncated, the response includes a NextToken. To retrieve the next set of endpoint configurations, use the token in the next request.  
Type: String  
Length Constraints: Maximum length of 8192.  
Pattern: .*  
Required: No  

**SortBy (p. 707)**  
The field to sort results by. The default is CreationTime.  
Type: String  
Valid Values: Name | CreationTime  
Required: No  

**SortOrder (p. 707)**  
The sort order for results. The default is Descending.  
Type: String  
Valid Values: Ascending | Descending  
Required: No  

**Response Syntax**  
```json  
{  
  "EndpointConfigs": [  
    {  
      "CreationTime": number,  
      "EndpointConfigArn": "string",  
      "EndpointConfigName": "string"  
    },  
    {  
      "NextToken": "string"  
    }  
  ]  
}  
```

**Response Elements**  
If the action is successful, the service sends back an HTTP 200 response.  
The following data is returned in JSON format by the service.  

**EndpointConfigs (p. 708)**  
An array of endpoint configurations.  
Type: Array of EndpointConfigSummary (p. 1455) objects
NextToken (p. 708)

If the response is truncated, SageMaker returns this token. To retrieve the next set of endpoint configurations, use it in the subsequent request

Type: String

Length Constraints: Maximum length of 8192.

Pattern: . *

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListEndpoints
Service: Amazon SageMaker Service
Lists endpoints.

Request Syntax

```json
{
   "CreationTimeAfter": number,
   "CreationTimeBefore": number,
   "LastModifiedTimeAfter": number,
   "LastModifiedTimeBefore": number,
   "MaxResults": number,
   "NameContains": "string",
   "NextToken": "string",
   "SortBy": "string",
   "SortOrder": "string",
   "StatusEquals": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CreationTimeAfter (p. 710)**

A filter that returns only endpoints with a creation time greater than or equal to the specified time (timestamp).

Type: Timestamp

Required: No

**CreationTimeBefore (p. 710)**

A filter that returns only endpoints that were created before the specified time (timestamp).

Type: Timestamp

Required: No

**LastModifiedTimeAfter (p. 710)**

A filter that returns only endpoints that were modified after the specified timestamp.

Type: Timestamp

Required: No

**LastModifiedTimeBefore (p. 710)**

A filter that returns only endpoints that were modified before the specified timestamp.

Type: Timestamp

Required: No

**MaxResults (p. 710)**

The maximum number of endpoints to return in the response. This value defaults to 10.
Type: Integer
Valid Range: Minimum value of 1. Maximum value of 100.
Required: No

NameContains (p. 710)
A string in endpoint names. This filter returns only endpoints whose name contains the specified string.
Type: String
Length Constraints: Maximum length of 63.
Pattern: [a-zA-Z0-9-]+
Required: No

NextToken (p. 710)
If the result of a ListEndpoints request was truncated, the response includes a NextToken. To retrieve the next set of endpoints, use the token in the next request.
Type: String
Length Constraints: Maximum length of 8192.
Pattern: .*
Required: No

SortBy (p. 710)
Sorts the list of results. The default is CreationTime.
Type: String
Valid Values: Name | CreationTime | Status
Required: No

SortOrder (p. 710)
The sort order for results. The default is Descending.
Type: String
Valid Values: Ascending | Descending
Required: No

StatusEquals (p. 710)
A filter that returns only endpoints with the specified status.
Type: String
Valid Values: OutOfService | Creating | Updating | SystemUpdating | RollingBack | InService | Deleting | Failed | UpdateRollbackFailed
Required: No

Response Syntax
{

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

Endpoints (p. 711)

An array or endpoint objects.

Type: Array of EndpointSummary (p. 1467) objects

NextToken (p. 711)

If the response is truncated, SageMaker returns this token. To retrieve the next set of training jobs, use it in the subsequent request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListExperiments
Service: Amazon SageMaker Service

Lists all the experiments in your account. The list can be filtered to show only experiments that were created in a specific time range. The list can be sorted by experiment name or creation time.

Request Syntax

```
{
  "CreatedAfter": number,
  "CreatedBefore": number,
  "MaxResults": number,
  "NextToken": "string",
  "SortBy": "string",
  "SortOrder": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CreatedAfter (p. 713)**

A filter that returns only experiments created after the specified time.

Type: Timestamp

Required: No

**CreatedBefore (p. 713)**

A filter that returns only experiments created before the specified time.

Type: Timestamp

Required: No

**MaxResults (p. 713)**

The maximum number of experiments to return in the response. The default value is 10.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No

**NextToken (p. 713)**

If the previous call to ListExperiments didn't return the full set of experiments, the call returns a token for getting the next set of experiments.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*
Required: No

SortBy (p. 713)

The property used to sort results. The default value is CreationTime.

Type: String

Valid Values: Name | CreationTime

Required: No

SortOrder (p. 713)

The sort order. The default value is Descending.

Type: String

Valid Values: Ascending | Descending

Required: No

Response Syntax

```
{
"ExperimentSummaries": [
{
  "CreationTime": number,
  "DisplayName": "string",
  "ExperimentArn": "string",
  "ExperimentName": "string",
  "ExperimentSource": {
    "SourceArn": "string",
    "SourceType": "string"
  },
  "LastModifiedTime": number
},
"NextToken": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

ExperimentSummaries (p. 714)

A list of the summaries of your experiments.

Type: Array of ExperimentSummary (p. 1476) objects

NextToken (p. 714)

A token for getting the next set of experiments, if there are any.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*
Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListFeatureGroups
Service: Amazon SageMaker Service

List FeatureGroups based on given filter and order.

Request Syntax

```json
{
    "CreationTimeAfter": number,
    "CreationTimeBefore": number,
    "FeatureGroupStatusEquals": "string",
    "MaxResults": number,
    "NameContains": "string",
    "NextToken": "string",
    "OfflineStoreStatusEquals": "string",
    "SortBy": "string",
    "SortOrder": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CreationTimeAfter (p. 716)**

Use this parameter to search for FeatureGroupss created after a specific date and time.

Type: Timestamp

Required: No

**CreationTimeBefore (p. 716)**

Use this parameter to search for FeatureGroupss created before a specific date and time.

Type: Timestamp

Required: No

**FeatureGroupStatusEquals (p. 716)**

A FeatureGroup status. Filters by FeatureGroup status.

Type: String

Valid Values: Creating | Created | CreateFailed | Deleting | DeleteFailed

Required: No

**MaxResults (p. 716)**

The maximum number of results returned by ListFeatureGroups.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No
NameContains (p. 716)

A string that partially matches one or more FeatureGroups names. Filters FeatureGroups by name.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 64.

Required: No

NextToken (p. 716)

A token to resume pagination of ListFeatureGroups results.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Required: No

OfflineStoreStatusEquals (p. 716)

An OfflineStore status. Filters by OfflineStore status.

Type: String

Valid Values: Active | Blocked | Disabled

Required: No

SortBy (p. 716)

The value on which the feature group list is sorted.

Type: String

Valid Values: Name | FeatureGroupStatus | OfflineStoreStatus | CreationTime

Required: No

SortOrder (p. 716)

The order in which feature groups are listed.

Type: String

Valid Values: Ascending | Descending

Required: No

Response Syntax

```
{
  "FeatureGroupSummaries": [
    {
      "CreationTime": number,
      "FeatureGroupArn": "string",
      "FeatureGroupName": "string",
      "FeatureGroupStatus": "string",
      "OfflineStoreStatus": {
        "OfflineStoreStatus": "string"
      }
    }
  ]
}
```
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**FeatureGroupSummaries (p. 717)**

A summary of feature groups.

Type: Array of FeatureGroupSummary (p. 1487) objects

**NextToken (p. 717)**

A token to resume pagination of ListFeatureGroups results.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListFlowDefinitions
Service: Amazon SageMaker Service

Returns information about the flow definitions in your account.

Request Syntax

```
{
   "CreationTimeAfter": number,
   "CreationTimeBefore": number,
   "MaxResults": number,
   "NextToken": "string",
   "SortOrder": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CreationTimeAfter (p. 719)**

A filter that returns only flow definitions with a creation time greater than or equal to the specified timestamp.

Type: Timestamp

Required: No

**CreationTimeBefore (p. 719)**

A filter that returns only flow definitions that were created before the specified timestamp.

Type: Timestamp

Required: No

**MaxResults (p. 719)**

The total number of items to return. If the total number of available items is more than the value specified in MaxResults, then a NextToken will be provided in the output that you can use to resume pagination.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No

**NextToken (p. 719)**

A token to resume pagination.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*
SortOrder (p. 719)

An optional value that specifies whether you want the results sorted in Ascending or Descending order.

Type: String

Valid Values: Ascending | Descending

Required: No

Response Syntax

```
{
   "FlowDefinitionSummaries": [
      {
         "CreationTime": "number",
         "FailureReason": "string",
         "FlowDefinitionArn": "string",
         "FlowDefinitionName": "string",
         "FlowDefinitionStatus": "string"
      }
   ],
   "NextToken": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

FlowDefinitionSummaries (p. 720)

An array of objects describing the flow definitions.

Type: Array of FlowDefinitionSummary (p. 1503) objects

NextToken (p. 720)

A token to resume pagination.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
• AWS SDK for .NET
• AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
ListHubContents

Service: Amazon SageMaker Service

List the contents of a hub.

Note
Hub APIs are only callable through SageMaker Studio.

Request Syntax

```
{
    "CreationTimeAfter": number,
    "CreationTimeBefore": number,
    "HubContentType": "string",
    "HubName": "string",
    "MaxResults": number,
    "MaxSchemaVersion": "string",
    "NameContains": "string",
    "NextToken": "string",
    "SortBy": "string",
    "SortOrder": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CreationTimeAfter (p. 722)**

Only list hub content that was created after the time specified.

Type: Timestamp

Required: No

**CreationTimeBefore (p. 722)**

Only list hub content that was created before the time specified.

Type: Timestamp

Required: No

**HubContentType (p. 722)**

The type of hub content to list.

Type: String

Valid Values: Model  |  Notebook

Required: Yes

**HubName (p. 722)**

The name of the hub to list the contents of.

Type: String
Length Constraints: Maximum length of 63.
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}$
Required: Yes

**MaxResults (p. 722)**

The maximum amount of hub content to list.
Type: Integer
Valid Range: Minimum value of 1. Maximum value of 100.
Required: No

**MaxSchemaVersion (p. 722)**

The upper bound of the hub content schema version.
Type: String
Pattern: ^\d{1,4}.\d{1,4}.\d{1,4}$
Required: No

**NameContains (p. 722)**

Only list hub content if the name contains the specified string.
Type: String
Length Constraints: Maximum length of 63.
Pattern: [a-zA-Z0-9\-]+$
Required: No

**NextToken (p. 722)**

If the response to a previous `ListHubContents` request was truncated, the response includes a NextToken. To retrieve the next set of hub content, use the token in the next request.
Type: String
Length Constraints: Maximum length of 8192.
Pattern: .*
Required: No

**SortBy (p. 722)**

Sort hub content versions by either name or creation time.
Type: String
Valid Values: HubContentName | CreationTime | HubContentStatus
Required: No

**SortOrder (p. 722)**

Sort hubs by ascending or descending order.
Type: String
Valid Values: Ascending | Descending
Required: No

Response Syntax

```json
{
  "HubContentSummaries": [
    {
      "CreationTime": number,
      "DocumentSchemaVersion": "string",
      "HubContentArn": "string",
      "HubContentDescription": "string",
      "HubContentDisplayName": "string",
      "HubContentName": "string",
      "HubContentSearchKeywords": [ "string" ],
      "HubContentStatus": "string",
      "HubContentType": "string",
      "HubContentVersion": "string"
    }
  ],
  "NextToken": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**HubContentSummaries (p. 724)**

The summaries of the listed hub content.

Type: Array of HubContentInfo (p. 1509) objects

**NextToken (p. 724)**

If the response is truncated, SageMaker returns this token. To retrieve the next set of hub content, use it in the subsequent request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceNotFoundException**

Resource being access is not found.

HTTP Status Code: 400
See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListHubContentVersions

Service: Amazon SageMaker Service

List hub content versions.

**Note**

Hub APIs are only callable through SageMaker Studio.

**Request Syntax**

```json
{
    "CreationTimeAfter": number,
    "CreationTimeBefore": number,
    "HubContentName": "string",
    "HubContentType": "string",
    "HubName": "string",
    "MaxResults": number,
    "MaxSchemaVersion": "string",
    "MinVersion": "string",
    "NextToken": "string",
    "SortBy": "string",
    "SortOrder": "string"
}
```

**Request Parameters**

For information about the parameters that are common to all actions, see [Common Parameters](p. 2178).

The request accepts the following data in JSON format.

**CreationTimeAfter (p. 726)**

Only list hub content versions that were created after the time specified.

Type: Timestamp

Required: No

**CreationTimeBefore (p. 726)**

Only list hub content versions that were created before the time specified.

Type: Timestamp

Required: No

**HubContentName (p. 726)**

The name of the hub content.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9]( -*[a-zA-Z0-9] ){0,62}

Required: Yes

**HubContentType (p. 726)**

The type of hub content to list versions of.
**HubName (p. 726)**

The name of the hub to list the content versions of.

- **Type:** String
- **Length Constraints:** Maximum length of 63.
- **Pattern:** `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}`
- **Required:** Yes

**MaxResults (p. 726)**

The maximum number of hub content versions to list.

- **Type:** Integer
- **Valid Range:** Minimum value of 1. Maximum value of 100.
- **Required:** No

**MaxSchemaVersion (p. 726)**

The upper bound of the hub content schema version.

- **Type:** String
- **Length Constraints:** Minimum length of 5. Maximum length of 14.
- **Pattern:** `^\d{1,4}\d{1,4}\d{1,4}$`
- **Required:** No

**MinVersion (p. 726)**

The lower bound of the hub content versions to list.

- **Type:** String
- **Length Constraints:** Minimum length of 5. Maximum length of 14.
- **Pattern:** `^\d{1,4}\d{1,4}\d{1,4}$`
- **Required:** No

**NextToken (p. 726)**

If the response to a previous ListHubContentVersions request was truncated, the response includes a NextToken. To retrieve the next set of hub content versions, use the token in the next request.

- **Type:** String
- **Length Constraints:** Maximum length of 8192.
- **Pattern:** `.*`
- **Required:** No
SortBy (p. 726)
Sort hub content versions by either name or creation time.
Type: String
Valid Values: HubContentName | CreationTime | HubContentStatus
Required: No

SortOrder (p. 726)
Sort hub content versions by ascending or descending order.
Type: String
Valid Values: Ascending | Descending
Required: No

Response Syntax
```
{
    "HubContentSummaries": [
    {
        "CreationTime": number,
        "DocumentSchemaVersion": "string",
        "HubContentArn": "string",
        "HubContentDescription": "string",
        "HubContentDisplayName": "string",
        "HubContentName": "string",
        "HubContentSearchKeywords": [ "string" ],
        "HubContentStatus": "string",
        "HubContentType": "string",
        "HubContentVersion": "string"
    }
    ]
    "NextToken": "string"
}
```

Response Elements
If the action is successful, the service sends back an HTTP 200 response.
The following data is returned in JSON format by the service.

HubContentSummaries (p. 728)
The summaries of the listed hub content versions.
Type: Array of HubContentInfo (p. 1509) objects

NextToken (p. 728)
If the response is truncated, SageMaker returns this token. To retrieve the next set of hub content
versions, use it in the subsequent request.
Type: String
Length Constraints: Maximum length of 8192.
Pattern: .*
Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceNotFound

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListHubs
Service: Amazon SageMaker Service

List all existing hubs.

**Note**
Hub APIs are only callable through SageMaker Studio.

**Request Syntax**

```json
{
    "CreationTimeAfter": number,
    "CreationTimeBefore": number,
    "LastModifiedTimeAfter": number,
    "LastModifiedTimeBefore": number,
    "MaxResults": number,
    "NameContains": "string",
    "NextToken": "string",
    "SortBy": "string",
    "SortOrder": "string"
}
```

**Request Parameters**

For information about the parameters that are common to all actions, see [Common Parameters](#).

The request accepts the following data in JSON format.

**CreationTimeAfter (p. 730)**

Only list hubs that were created after the time specified.

Type: Timestamp

Required: No

**CreationTimeBefore (p. 730)**

Only list hubs that were created before the time specified.

Type: Timestamp

Required: No

**LastModifiedTimeAfter (p. 730)**

Only list hubs that were last modified after the time specified.

Type: Timestamp

Required: No

**LastModifiedTimeBefore (p. 730)**

Only list hubs that were last modified before the time specified.

Type: Timestamp

Required: No
MaxResults (p. 730)

The maximum number of hubs to list.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No

NameContains (p. 730)

Only list hubs with names that contain the specified string.

Type: String

Length Constraints: Maximum length of 63.

Pattern: [a-zA-Z0-9-]+

Required: No

NextToken (p. 730)

If the response to a previous ListHubs request was truncated, the response includes a NextToken. To retrieve the next set of hubs, use the token in the next request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Required: No

SortBy (p. 730)

Sort hubs by either name or creation time.

Type: String

Valid Values: HubName | CreationTime | HubStatus | AccountIdOwner

Required: No

SortOrder (p. 730)

Sort hubs by ascending or descending order.

Type: String

Valid Values: Ascending | Descending

Required: No

Response Syntax

```json
{
    "HubSummaries": [ 
    {
        "CreationTime": number,
        "HubArn": "string",
        "HubDescription": "string",
```
"HubDisplayName": "string",
"HubName": "string",
"HubSearchKeywords": [ "string" ],
"HubStatus": "string",
"LastModifiedTime": number
]
}]
"NextToken": "string"
}

### Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**HubSummaries (p. 731)**

The summaries of the listed hubs.

Type: Array of [HubInfo (p. 1512)] objects

**NextToken (p. 731)**

If the response is truncated, SageMaker returns this token. To retrieve the next set of hubs, use it in the subsequent request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

### Errors

For information about the errors that are common to all actions, see [Common Errors (p. 2180)]

### See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListHumanTaskUis
Service: Amazon SageMaker Service

Returns information about the human task user interfaces in your account.

Request Syntax

```json
{
   "CreationTimeAfter": number,
   "CreationTimeBefore": number,
   "MaxResults": number,
   "NextToken": "string",
   "SortOrder": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CreationTimeAfter (p. 733)**

A filter that returns only human task user interfaces with a creation time greater than or equal to the specified timestamp.

Type: Timestamp

Required: No

**CreationTimeBefore (p. 733)**

A filter that returns only human task user interfaces that were created before the specified timestamp.

Type: Timestamp

Required: No

**MaxResults (p. 733)**

The total number of items to return. If the total number of available items is more than the value specified in MaxResults, then a NextToken will be provided in the output that you can use to resume pagination.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No

**NextToken (p. 733)**

A token to resume pagination.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*
### SortOrder (p. 733)

An optional value that specifies whether you want the results sorted in Ascending or Descending order.

**Type:** String

**Valid Values:** Ascending | Descending

**Required:** No

### Response Syntax

```json
{
    "HumanTaskUiSummaries": [
        {
            "CreationTime": number,
            "HumanTaskUiArn": "string",
            "HumanTaskUiName": "string"
        }
    ],
    "NextToken": "string"
}
```

### Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**HumanTaskUiSummaries (p. 734)**

An array of objects describing the human task user interfaces.

**Type:** Array of [HumanTaskUiSummary (p. 1537)] objects

**NextToken (p. 734)**

A token to resume pagination.

**Type:** String

**Length Constraints:** Maximum length of 8192.

**Pattern:** .*

### Errors

For information about the errors that are common to all actions, see [Common Errors (p. 2180)].

### See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
• AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
ListHyperParameterTuningJobs
Service: Amazon SageMaker Service

Gets a list of HyperParameterTuningJobSummary objects that describe the hyperparameter tuning jobs launched in your account.

Request Syntax

```
{
  "CreationTimeAfter": number,
  "CreationTimeBefore": number,
  "LastModifiedTimeAfter": number,
  "LastModifiedTimeBefore": number,
  "MaxResults": number,
  "NameContains": "string",
  "NextToken": "string",
  "SortBy": "string",
  "SortOrder": "string",
  "StatusEquals": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CreationTimeAfter (p. 736)**

A filter that returns only tuning jobs that were created after the specified time.

Type: Timestamp

Required: No

**CreationTimeBefore (p. 736)**

A filter that returns only tuning jobs that were created before the specified time.

Type: Timestamp

Required: No

**LastModifiedTimeAfter (p. 736)**

A filter that returns only tuning jobs that were modified after the specified time.

Type: Timestamp

Required: No

**LastModifiedTimeBefore (p. 736)**

A filter that returns only tuning jobs that were modified before the specified time.

Type: Timestamp

Required: No
MaxResults (p. 736)

The maximum number of tuning jobs to return. The default value is 10.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No

NameContains (p. 736)

A string in the tuning job name. This filter returns only tuning jobs whose name contains the specified string.

Type: String

Length Constraints: Maximum length of 63.

Pattern: [a-zA-Z0-9-\-]*

Required: No

NextToken (p. 736)

If the result of the previous ListHyperParameterTuningJobs request was truncated, the response includes a NextToken. To retrieve the next set of tuning jobs, use the token in the next request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Required: No

SortBy (p. 736)

The field to sort results by. The default is Name.

Type: String

Valid Values: Name | Status | CreationTime

Required: No

SortOrder (p. 736)

The sort order for results. The default is Ascending.

Type: String

Valid Values: Ascending | Descending

Required: No

StatusEquals (p. 736)

A filter that returns only tuning jobs with the specified status.

Type: String

Valid Values: Completed | InProgress | Failed | Stopped | Stopping

Required: No
### Response Syntax

```json
{
  "HyperParameterTuningJobSummaries": [
    {
      "CreationTime": number,
      "HyperParameterTuningEndTime": number,
      "HyperParameterTuningJobArn": "string",
      "HyperParameterTuningJobName": "string",
      "HyperParameterTuningJobStatus": "string",
      "LastModifiedTime": number,
      "ObjectiveStatusCounters": {
        "Failed": number,
        "Pending": number,
        "Succeeded": number
      },
      "ResourceLimits": {
        "MaxNumberOfTrainingJobs": number,
        "MaxParallelTrainingJobs": number,
        "MaxRuntimeInSeconds": number
      },
      "Strategy": "string",
      "TrainingJobStatusCounters": {
        "Completed": number,
        "InProgress": number,
        "NonRetryableError": number,
        "RetryableError": number,
        "Stopped": number
      }
    }
  ]
}
```

### Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**HyperParameterTuningJobSummaries (p. 738)**

A list of HyperParameterTuningJobSummary objects that describe the tuning jobs that the ListHyperParameterTuningJobs request returned.

Type: Array of HyperParameterTuningJobSummary (p. 1562) objects

**NextToken (p. 738)**

If the result of this ListHyperParameterTuningJobs request was truncated, the response includes a NextToken. To retrieve the next set of tuning jobs, use the token in the next request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

### Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).
See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListImages
Service: Amazon SageMaker Service

Lists the images in your account and their properties. The list can be filtered by creation time or modified time, and whether the image name contains a specified string.

Request Syntax

```
{
  "CreationTimeAfter": number,
  "CreationTimeBefore": number,
  "LastModifiedTimeAfter": number,
  "LastModifiedTimeBefore": number,
  "MaxResults": number,
  "NameContains": "string",
  "NextToken": "string",
  "SortBy": "string",
  "SortOrder": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CreationTimeAfter (p. 740)**

A filter that returns only images created on or after the specified time.

Type: Timestamp

Required: No

**CreationTimeBefore (p. 740)**

A filter that returns only images created on or before the specified time.

Type: Timestamp

Required: No

**LastModifiedTimeAfter (p. 740)**

A filter that returns only images modified on or after the specified time.

Type: Timestamp

Required: No

**LastModifiedTimeBefore (p. 740)**

A filter that returns only images modified on or before the specified time.

Type: Timestamp

Required: No

**MaxResults (p. 740)**

The maximum number of images to return in the response. The default value is 10.
Type: Integer
Valid Range: Minimum value of 1. Maximum value of 100.
Required: No

**NameContains (p. 740)**
A filter that returns only images whose name contains the specified string.
Type: String
Length Constraints: Maximum length of 63.
Pattern: ^[a-zA-Z0-9\-\.]+$
Required: No

**NextToken (p. 740)**
If the previous call to ListImages didn't return the full set of images, the call returns a token for getting the next set of images.
Type: String
Length Constraints: Maximum length of 8192.
Pattern: .*
Required: No

**sortBy (p. 740)**
The property used to sort results. The default value is CREATION_TIME.
Type: String
Valid Values: CREATION_TIME | LAST_MODIFIED_TIME | IMAGE_NAME
Required: No

**SortOrder (p. 740)**
The sort order. The default value is DESCENDING.
Type: String
Valid Values: ASCENDING | DESCENDING
Required: No

**Response Syntax**

```json
{
    "Images": [
        {
            "CreationTime": number,
            "Description": "string",
            "DisplayName": "string",
            "FailureReason": "string",
            "ImageArn": "string",
            "ImageName": "string",
            "ImageStatus": "string"
        }
    ]
}
```
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

Images (p. 741)

A list of images and their properties.

Type: Array of Image (p. 1571) objects

NextToken (p. 741)

A token for getting the next set of images, if there are any.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListImageVersions

Service: Amazon SageMaker Service

Lists the versions of a specified image and their properties. The list can be filtered by creation time or modified time.

Request Syntax

```
{
  "CreationTimeAfter": number,
  "CreationTimeBefore": number,
  "ImageName": "string",
  "LastModifiedTimeAfter": number,
  "LastModifiedTimeBefore": number,
  "MaxResults": number,
  "NextToken": "string",
  "SortBy": "string",
  "SortOrder": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CreationTimeAfter (p. 743)**

A filter that returns only versions created on or after the specified time.

Type: Timestamp

Required: No

**CreationTimeBefore (p. 743)**

A filter that returns only versions created on or before the specified time.

Type: Timestamp

Required: No

**ImageName (p. 743)**

The name of the image to list the versions of.

Type: String


Pattern: `^[a-zA-Z0-9][-\.]?[a-zA-Z0-9]{0,62}$`

Required: Yes

**LastModifiedTimeAfter (p. 743)**

A filter that returns only versions modified on or after the specified time.

Type: Timestamp

Required: No
**LastModifiedTimeBefore (p. 743)**

A filter that returns only versions modified on or before the specified time.

Type: Timestamp

Required: No

**MaxResults (p. 743)**

The maximum number of versions to return in the response. The default value is 10.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No

**NextToken (p. 743)**

If the previous call to ListImageVersions didn't return the full set of versions, the call returns a token for getting the next set of versions.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Required: No

**SortBy (p. 743)**

The property used to sort results. The default value is CREATION_TIME.

Type: String

Valid Values: CREATION_TIME | LAST_MODIFIED_TIME | VERSION

Required: No

**SortOrder (p. 743)**

The sort order. The default value is DESCENDING.

Type: String

Valid Values: ASCENDING | DESCENDING

Required: No

**Response Syntax**

```
{
  "ImageVersions": [
    {
      "CreationTime": number,
      "FailureReason": "string",
      "ImageArn": "string",
      "ImageVersionArn": "string",
      "ImageVersionStatus": "string",
      "LastModifiedTime": number,
      "Version": number
    }
  ]
}
```
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**ImageVersions (p. 744)**

A list of versions and their properties.

Type: Array of *ImageVersion (p. 1575)* objects

**NextToken (p. 744)**

A token for getting the next set of versions, if there are any.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Errors

For information about the errors that are common to all actions, see *Common Errors (p. 2180)*.

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListInferenceComponents
Service: Amazon SageMaker Service

Lists the inference components in your account and their properties.

Request Syntax

```
{
  "CreationTimeAfter": number,
  "CreationTimeBefore": number,
  "EndpointNameEquals": "string",
  "LastModifiedTimeAfter": number,
  "LastModifiedTimeBefore": number,
  "MaxResults": number,
  "NameContains": "string",
  "NextToken": "string",
  "SortBy": "string",
  "SortOrder": "string",
  "StatusEquals": "string",
  "VariantNameEquals": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CreationTimeAfter** (p. 746)

Filters the results to only those inference components that were created after the specified time.

Type: Timestamp

Required: No

**CreationTimeBefore** (p. 746)

Filters the results to only those inference components that were created before the specified time.

Type: Timestamp

Required: No

**EndpointNameEquals** (p. 746)

An endpoint name to filter the listed inference components. The response includes only those inference components that are hosted at the specified endpoint.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}

Required: No

**LastModifiedTimeAfter** (p. 746)

Filters the results to only those inference components that were updated after the specified time.
Type: Timestamp
Required: No

**LastModifiedTimeBefore (p. 746)**
Filters the results to only those inference components that were updated before the specified time.
Type: Timestamp
Required: No

**MaxResults (p. 746)**
The maximum number of inference components to return in the response. This value defaults to 10.
Type: Integer
Valid Range: Minimum value of 1. Maximum value of 100.
Required: No

**NameContains (p. 746)**
Filters the results to only those inference components with a name that contains the specified string.
Type: String
Length Constraints: Maximum length of 63.
Pattern: [a-zA-Z0-9-]+
Required: No

**NextToken (p. 746)**
A token that you use to get the next set of results following a truncated response. If the response to the previous request was truncated, that response provides the value for this token.
Type: String
Length Constraints: Maximum length of 8192.
Pattern: .*
Required: No

**SortBy (p. 746)**
The field by which to sort the inference components in the response. The default is CreationTime.
Type: String
Valid Values: Name | CreationTime | Status
Required: No

**SortOrder (p. 746)**
The sort order for results. The default is Descending.
Type: String
Valid Values: Ascending | Descending
Required: No
**StatusEquals (p. 746)**

Filters the results to only those inference components with the specified status.

Type: String

Valid Values: InService | Creating | Updating | Failed | Deleting

Required: No

**VariantNameEquals (p. 746)**

A production variant name to filter the listed inference components. The response includes only those inference components that are hosted at the specified variant.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

Required: No

**Response Syntax**

```json
{
"InferenceComponents": [
{
"CreationTime": number,
"EndpointArn": "string",
"EndpointName": "string",
"InferenceComponentArn": "string",
"InferenceComponentName": "string",
"InferenceComponentStatus": "string",
"LastModifiedTime": number,
"VariantName": "string"
}
], "NextToken": "string"
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**InferenceComponents (p. 748)**

A list of inference components and their properties that matches any of the filters you specified in the request.

Type: Array of InferenceComponentSummary (p. 1587) objects

**NextToken (p. 748)**

The token to use in a subsequent request to get the next set of results following a truncated response.

Type: String
Length Constraints: Maximum length of 8192.

Pattern: . *

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListInferenceExperiments
Service: Amazon SageMaker Service
Returns the list of all inference experiments.

Request Syntax

```
{
  "CreationTimeAfter": number,
  "CreationTimeBefore": number,
  "LastModifiedTimeAfter": number,
  "LastModifiedTimeBefore": number,
  "MaxResults": number,
  "NameContains": "string",
  "NextToken": "string",
  "SortBy": "string",
  "SortOrder": "string",
  "StatusEquals": "string",
  "Type": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CreationTimeAfter (p. 750)**

Selects inference experiments which were created after this timestamp.

Type: Timestamp

Required: No

**CreationTimeBefore (p. 750)**

Selects inference experiments which were created before this timestamp.

Type: Timestamp

Required: No

**LastModifiedTimeAfter (p. 750)**

Selects inference experiments which were last modified after this timestamp.

Type: Timestamp

Required: No

**LastModifiedTimeBefore (p. 750)**

Selects inference experiments which were last modified before this timestamp.

Type: Timestamp

Required: No

**MaxResults (p. 750)**

The maximum number of results to select.
Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No

**NameContains (p. 750)**

Selects inference experiments whose names contain this name.

Type: String

Length Constraints: Maximum length of 63.

Pattern: [a-zA-Z0-9-]+

Required: No

**NextToken (p. 750)**

The response from the last list when returning a list large enough to need tokening.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Required: No

**SortBy (p. 750)**

The column by which to sort the listed inference experiments.

Type: String

Valid Values: Name | CreationTime | Status

Required: No

**SortOrder (p. 750)**

The direction of sorting (ascending or descending).

Type: String

Valid Values: Ascending | Descending

Required: No

**StatusEquals (p. 750)**

Selects inference experiments which are in this status. For the possible statuses, see `DescribeInferenceExperiment`.

Type: String

Valid Values: Creating | Created | Updating | Running | Starting | Stopping | Completed | Cancelled

Required: No

**Type (p. 750)**

Selects inference experiments of this type. For the possible types of inference experiments, see `CreateInferenceExperiment`.
Type: String
Valid Values: ShadowMode
Required: No

Response Syntax

```json
{
"InferenceExperiments": [
{
"CompletionTime": number,
"CreationTime": number,
"Description": "string",
"LastModifiedTime": number,
"Name": "string",
"RoleArn": "string",
"Schedule": {
  "EndTime": number,
  "StartTime": number
},
"Status": "string",
"StatusReason": "string",
"Type": "string"
}
],
"NextToken": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**InferenceExperiments (p. 752)**

List of inference experiments.

Type: Array of InferenceExperimentSummary (p. 1592) objects

**NextToken (p. 752)**

The token to use when calling the next page of results.

Type: String
Length Constraints: Maximum length of 8192.
Pattern: .*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:
• AWS Command Line Interface
• AWS SDK for .NET
• AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
ListInferenceRecommendationsJobs
Service: Amazon SageMaker Service

Lists recommendation jobs that satisfy various filters.

Request Syntax

```
{
    "CreationTimeAfter": number,
    "CreationTimeBefore": number,
    "LastModifiedTimeAfter": number,
    "LastModifiedTimeBefore": number,
    "MaxResults": number,
    "ModelNameEquals": "string",
    "ModelPackageVersionArnEquals": "string",
    "NameContains": "string",
    "NextToken": "string",
    "SortBy": "string",
    "SortOrder": "string",
    "StatusEquals": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CreationTimeAfter (p. 754)**

A filter that returns only jobs created after the specified time (timestamp).

Type: Timestamp

Required: No

**CreationTimeBefore (p. 754)**

A filter that returns only jobs created before the specified time (timestamp).

Type: Timestamp

Required: No

**LastModifiedTimeAfter (p. 754)**

A filter that returns only jobs that were last modified after the specified time (timestamp).

Type: Timestamp

Required: No

**LastModifiedTimeBefore (p. 754)**

A filter that returns only jobs that were last modified before the specified time (timestamp).

Type: Timestamp

Required: No
MaxResults (p. 754)

The maximum number of recommendations to return in the response.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No

ModelNameEquals (p. 754)

A filter that returns only jobs that were created for this model.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])* 

Required: No

ModelPackageVersionArnEquals (p. 754)

A filter that returns only jobs that were created for this versioned model package.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: ^arn:aws(-cn|-us-gov)?:sagemaker:[a-zA-Z0-9\-]{9,16}:[0-9]{12}:model-package/\S{1,2048} 

Required: No

NameContains (p. 754)

A string in the job name. This filter returns only recommendations whose name contains the specified string.

Type: String

Length Constraints: Maximum length of 63.

Pattern: [a-zA-Z0-9-\-]+ 

Required: No

NextToken (p. 754)

If the response to a previous ListInferenceRecommendationsJobsRequest request was truncated, the response includes a NextToken. To retrieve the next set of recommendations, use the token in the next request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .* 

Required: No

SortBy (p. 754)

The parameter by which to sort the results.

Type: String
Valid Values: Name | CreationTime | Status
Required: No

**SortOrder (p. 754)**

The sort order for the results.
Type: String
Valid Values: Ascending | Descending
Required: No

**StatusEquals (p. 754)**

A filter that retrieves only inference recommendations jobs with a specific status.
Type: String
Valid Values: PENDING | IN_PROGRESS | COMPLETED | FAILED | STOPPING | STOPPED
Required: No

**Response Syntax**

```json
{
  "InferenceRecommendationsJobs": [
    {
      "CompletionTime": number,
      "CreationTime": number,
      "FailureReason": "string",
      "JobArn": "string",
      "JobDescription": "string",
      "JobName": "string",
      "JobType": "string",
      "LastModifiedTime": number,
      "ModelName": "string",
      "ModelPackageVersionArn": "string",
      "RoleArn": "string",
      "SamplePayloadUrl": "string",
      "Status": "string"
    }
  ],
  "NextToken": "string"
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**InferenceRecommendationsJobs (p. 756)**

The recommendations created from the Amazon SageMaker Inference Recommender job.
Type: Array of InferenceRecommendationsJob (p. 1597) objects

**NextToken (p. 756)**

A token for getting the next set of recommendations, if there are any.
Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
**ListInferenceRecommendationsJobSteps**

Service: Amazon SageMaker Service

Returns a list of the subtasks for an Inference Recommender job.

The supported subtasks are benchmarks, which evaluate the performance of your model on different instance types.

**Request Syntax**

```json
{
   "JobName": "string",
   "MaxResults": number,
   "NextToken": "string",
   "Status": "string",
   "StepType": "string"
}
```

**Request Parameters**

For information about the parameters that are common to all actions, see [Common Parameters](p. 2178).

The request accepts the following data in JSON format.

**JobName (p. 758)**

The name for the Inference Recommender job.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 64.

Pattern: \^[a-zA-Z0-9](\-*[a-zA-Z0-9])\{0,63\}

Required: Yes

**MaxResults (p. 758)**

The maximum number of results to return.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No

**NextToken (p. 758)**

A token that you can specify to return more results from the list. Specify this field if you have a token that was returned from a previous request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Required: No
**Status (p. 758)**

A filter to return benchmarks of a specified status. If this field is left empty, then all benchmarks are returned.

Type: String

Valid Values: PENDING | IN_PROGRESS | COMPLETED | FAILED | STOPPING | STOPPED

Required: No

**StepType (p. 758)**

A filter to return details about the specified type of subtask.

BENCHMARK: Evaluate the performance of your model on different instance types.

Type: String

Valid Values: BENCHMARK

Required: No

**Response Syntax**

```
{
  "NextToken": "string",
  "Steps": [
    {
      "InferenceBenchmark": {
        "EndpointConfiguration": {
          "EndpointName": "string",
          "InitialInstanceCount": number,
          "InstanceType": "string",
          "ServerlessConfig": {
            "MaxConcurrency": number,
            "MemorySizeInMB": number,
            "ProvisionedConcurrency": number
          },
          "VariantName": "string"
        },
        "EndpointMetrics": {
          "MaxInvocations": number,
          "ModelLatency": number
        },
        "FailureReason": "string",
        "InvocationEndTime": number,
        "InvocationStartTime": number,
        "Metrics": {
          "CostPerHour": number,
          "CostPerInference": number,
          "CpuUtilization": number,
          "MaxInvocations": number,
          "MemoryUtilization": number,
          "ModelLatency": number,
          "ModelSetupTime": number
        },
        "ModelConfiguration": {
          "CompilationJobName": "string",
          "EnvironmentParameters": [
            {
              "Key": "string",
              "Value": "string",
            }
          ]
        }
      }
    }
  ]
}
```
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**NextToken (p. 759)**

A token that you can specify in your next request to return more results from the list.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

**Steps (p. 759)**

A list of all subtask details in Inference Recommender.

Type: Array of `InferenceRecommendationsJobStep (p. 1600)` objects

Errors

For information about the errors that are common to all actions, see `Common Errors (p. 2180)`.

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](https://docs.aws.amazon.com/cli/latest/index.html)
- [AWS SDK for .NET](https://docs.aws.amazon.com/sdk-for-net/v3/)
- [AWS SDK for C++](https://docs.aws.amazon.com/sdk-for-cpp/v1/)
- [AWS SDK for Go](https://docs.aws.amazon.com/sdk-for-go/v1/)
- [AWS SDK for Java V2](https://docs.aws.amazon.com/java/latest/index.html)
- [AWS SDK for JavaScript V3](https://docs.aws.amazon.com/sdk-for-javascript/v3/)
- [AWS SDK for PHP V3](https://docs.aws.amazon.com/sdk-for-php/v3/)
- [AWS SDK for Python](https://docs.aws.amazon.com/sdk-for-python/v3/)
• AWS SDK for Ruby V3
ListLabelingJobs
Service: Amazon SageMaker Service

Gets a list of labeling jobs.

Request Syntax

```json
{
  "CreationTimeAfter": number,
  "CreationTimeBefore": number,
  "LastModifiedTimeAfter": number,
  "LastModifiedTimeBefore": number,
  "MaxResults": number,
  "NameContains": "string",
  "NextToken": "string",
  "SortBy": "string",
  "SortOrder": "string",
  "StatusEquals": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CreationTimeAfter (p. 762)**

A filter that returns only labeling jobs created after the specified time (timestamp).

Type: Timestamp

Required: No

**CreationTimeBefore (p. 762)**

A filter that returns only labeling jobs created before the specified time (timestamp).

Type: Timestamp

Required: No

**LastModifiedTimeAfter (p. 762)**

A filter that returns only labeling jobs modified after the specified time (timestamp).

Type: Timestamp

Required: No

**LastModifiedTimeBefore (p. 762)**

A filter that returns only labeling jobs modified before the specified time (timestamp).

Type: Timestamp

Required: No

**MaxResults (p. 762)**

The maximum number of labeling jobs to return in each page of the response.
Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No

**NameContains (p. 762)**

A string in the labeling job name. This filter returns only labeling jobs whose name contains the specified string.

Type: String

Length Constraints: Maximum length of 63.

Pattern: [a-zA-Z0-9-\-]+

Required: No

**NextToken (p. 762)**

If the result of the previous ListLabelingJobs request was truncated, the response includes a NextToken. To retrieve the next set of labeling jobs, use the token in the next request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Required: No

**SortBy (p. 762)**

The field to sort results by. The default is CreationTime.

Type: String

Valid Values: Name | CreationTime | Status

Required: No

**SortOrder (p. 762)**

The sort order for results. The default is Ascending.

Type: String

Valid Values: Ascending | Descending

Required: No

**StatusEquals (p. 762)**

A filter that retrieves only labeling jobs with a specific status.

Type: String

Valid Values: Initializing | InProgress | Completed | Failed | Stopping | Stopped

Required: No
Response Syntax

```json
{
  "LabelingJobSummaryList": [
    {
      "AnnotationConsolidationLambdaArn": "string",
      "CreationTime": number,
      "FailureReason": "string",
      "InputConfig": {
        "DataAttributes": {
          "ContentClassifiers": [ "string" ]
        },
        "DataSource": {
          "S3DataSource": {
            "ManifestS3Uri": "string"
          },
          "SnsDataSource": {
            "SnsTopicArn": "string"
          }
        }
      },
      "LabelCounters": {
        "FailedNonRetryableError": number,
        "HumanLabeled": number,
        "MachineLabeled": number,
        "TotalLabeled": number,
        "Unlabeled": number
      },
      "LabelingJobArn": "string",
      "LabelingJobName": "string",
      "LabelingJobOutput": {
        "FinalActiveLearningModelArn": "string",
        "OutputDatasetS3Uri": "string"
      },
      "LabelingJobStatus": "string",
      "LastModifiedTime": number,
      "PreHumanTaskLambdaArn": "string",
      "WorkteamArn": "string"
    }
  ],
  "NextToken": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**LabelingJobSummaryList** *(p. 764)*

An array of LabelingJobSummary objects, each describing a labeling job.

Type: Array of LabelingJobSummary *(p. 1640)* objects

**NextToken** *(p. 764)*

If the response is truncated, SageMaker returns this token. To retrieve the next set of labeling jobs, use it in the subsequent request.

Type: String

Length Constraints: Maximum length of 8192.
Pattern: .*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListLabelingJobsForWorkteam

Service: Amazon SageMaker Service

Gets a list of labeling jobs assigned to a specified work team.

Request Syntax

```
{
    "CreationTimeAfter": number,
    "CreationTimeBefore": number,
    "JobReferenceCodeContains": "string",
    "MaxResults": number,
    "NextToken": "string",
    "SortBy": "string",
    "SortOrder": "string",
    "WorkteamArn": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CreationTimeAfter (p. 766)**

A filter that returns only labeling jobs created after the specified time (timestamp).

Type: Timestamp

Required: No

**CreationTimeBefore (p. 766)**

A filter that returns only labeling jobs created before the specified time (timestamp).

Type: Timestamp

Required: No

**JobReferenceCodeContains (p. 766)**

A filter the limits jobs to only the ones whose job reference code contains the specified string.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 255.

Pattern: .*

Required: No

**MaxResults (p. 766)**

The maximum number of labeling jobs to return in each page of the response.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No
NextToken (p. 766)

If the result of the previous ListLabelingJobsForWorkteam request was truncated, the response includes a NextToken. To retrieve the next set of labeling jobs, use the token in the next request.

Type: String
Length Constraints: Maximum length of 8192.
Pattern: .*
Required: No

SortBy (p. 766)

The field to sort results by. The default is CreationTime.

Type: String
Valid Values: CreationTime
Required: No

SortOrder (p. 766)

The sort order for results. The default is Ascending.

Type: String
Valid Values: Ascending | Descending
Required: No

WorkteamArn (p. 766)

The Amazon Resource Name (ARN) of the work team for which you want to see labeling jobs for.

Type: String
Length Constraints: Maximum length of 256.
Pattern: arn:aws[a-z-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:workteam/.
Required: Yes

Response Syntax

```json
{
    "LabelingJobSummaryList": [
        {
            "CreationTime": "number",
            "JobReferenceCode": "string",
            "LabelCounters": {
                "HumanLabeled": "number",
                "PendingHuman": "number",
                "Total": "number"
            },
            "LabelingJobName": "string",
            "NumberOfHumanWorkersPerDataObject": "number",
            "WorkRequesterAccountId": "string"
        }
    ],
    "NextToken": "string"
}
```
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**LabelingJobSummaryList (p. 767)**

An array of LabelingJobSummary objects, each describing a labeling job.

Type: Array of LabelingJobForWorkteamSummary (p. 1630) objects

**NextToken (p. 767)**

If the response is truncated, SageMaker returns this token. To retrieve the next set of labeling jobs, use it in the subsequent request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Errors

For information about the errors that are common to all actions, see [Common Errors (p. 2180)].

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListLineageGroups

Service: Amazon SageMaker Service

A list of lineage groups shared with your AWS account. For more information, see Cross-Account Lineage Tracking in the Amazon SageMaker Developer Guide.

Request Syntax

```
{
    "CreatedAfter": number,
    "CreatedBefore": number,
    "MaxResults": number,
    "NextToken": "string",
    "SortBy": "string",
    "SortOrder": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CreatedAfter (p. 769)**

A timestamp to filter against lineage groups created after a certain point in time.

Type: Timestamp

Required: No

**CreatedBefore (p. 769)**

A timestamp to filter against lineage groups created before a certain point in time.

Type: Timestamp

Required: No

**MaxResults (p. 769)**

The maximum number of endpoints to return in the response. This value defaults to 10.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No

**NextToken (p. 769)**

If the response is truncated, SageMaker returns this token. To retrieve the next set of algorithms, use it in the subsequent request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Required: No
SortBy (p. 769)
The parameter by which to sort the results. The default is CreationTime.

Type: String

Valid Values: Name | CreationTime

Required: No

SortOrder (p. 769)
The sort order for the results. The default is Ascending.

Type: String

Valid Values: Ascending | Descending

Required: No

Response Syntax

```
{
  "LineageGroupSummaries": [
    {
      "CreationTime": number,
      "DisplayName": "string",
      "LastModifiedTime": number,
      "LineageGroupArn": "string",
      "LineageGroupName": "string"
    }
  ],
  "NextToken": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

LineageGroupSummaries (p. 770)
A list of lineage groups and their properties.

Type: Array of LineageGroupSummary (p. 1645) objects

NextToken (p. 770)
If the response is truncated, SageMaker returns this token. To retrieve the next set of algorithms, use it in the subsequent request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).
See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListModelBiasJobDefinitions
Service: Amazon SageMaker Service
Lists model bias jobs definitions that satisfy various filters.

Request Syntax

```
{
    "CreationTimeAfter": number,
    "CreationTimeBefore": number,
    "EndpointName": "string",
    "MaxResults": number,
    "NameContains": "string",
    "NextToken": "string",
    "SortBy": "string",
    "SortOrder": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CreationTimeAfter** (p. 772)

A filter that returns only model bias jobs created after a specified time.

Type: Timestamp

Required: No

**CreationTimeBefore** (p. 772)

A filter that returns only model bias jobs created before a specified time.

Type: Timestamp

Required: No

**EndpointName** (p. 772)

Name of the endpoint to monitor for model bias.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9]*[-\[a-zA-Z0-9]*\{0,62

Required: No

**MaxResults** (p. 772)

The maximum number of model bias jobs to return in the response. The default value is 10.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No
**NameContains (p. 772)**

Filter for model bias jobs whose name contains a specified string.

Type: String

Length Constraints: Maximum length of 63.

Pattern: [a-zA-Z0-9-\-]+

Required: No

**NextToken (p. 772)**

The token returned if the response is truncated. To retrieve the next set of job executions, use it in the next request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Required: No

**SortBy (p. 772)**

Whether to sort results by the Name or CreationTime field. The default is CreationTime.

Type: String

Valid Values: Name | CreationTime

Required: No

**SortOrder (p. 772)**

Whether to sort the results in Ascending or Descending order. The default is Descending.

Type: String

Valid Values: Ascending | Descending

Required: No

**Response Syntax**

```json
{
  "JobDefinitionSummaries": [
    {
      "CreationTime": number,
      "EndpointName": "string",
      "MonitoringJobDefinitionArn": "string",
      "MonitoringJobDefinitionName": "string"
    }
  ],
  "NextToken": "string"
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.
The following data is returned in JSON format by the service.

**JobDefinitionSummaries (p. 773)**

A JSON array in which each element is a summary for a model bias jobs.

Type: Array of MonitoringJobDefinitionSummary (p. 1749) objects

**NextToken (p. 773)**

The token returned if the response is truncated. To retrieve the next set of job executions, use it in the next request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

**Errors**

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
**ListModelCardExportJobs**

Service: Amazon SageMaker Service

List the export jobs for the Amazon SageMaker Model Card.

**Request Syntax**

```json
{
    "CreationTimeAfter": number,
    "CreationTimeBefore": number,
    "MaxResults": number,
    "ModelCardExportJobNameContains": "string",
    "ModelCardName": "string",
    "ModelCardVersion": number,
    "NextToken": "string",
    "SortBy": "string",
    "SortOrder": "string",
    "StatusEquals": "string"
}
```

**Request Parameters**

For information about the parameters that are common to all actions, see [Common Parameters](p. 2178).

The request accepts the following data in JSON format.

**CreationTimeAfter (p. 775)**

Only list model card export jobs that were created after the time specified.

Type: Timestamp

Required: No

**CreationTimeBefore (p. 775)**

Only list model card export jobs that were created before the time specified.

Type: Timestamp

Required: No

**MaxResults (p. 775)**

The maximum number of model card export jobs to list.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No

**ModelCardExportJobNameContains (p. 775)**

Only list model card export jobs with names that contain the specified string.

Type: String


Pattern: `^[a-zA-Z0-9](\-*[a-zA-Z0-9])\{0,62}\$`
ListModelCardExportJobs

List export jobs for the model card with the specified name.

**ModelCardName (p. 775)**

List export jobs for the model card with the specified name.

- Type: String
- Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62\}$`
- Required: Yes

**ModelCardVersion (p. 775)**

List export jobs for the model card with the specified version.

- Type: Integer
- Required: No

**NextToken (p. 775)**

If the response to a previous ListModelCardExportJobs request was truncated, the response includes a NextToken. To retrieve the next set of model card export jobs, use the token in the next request.

- Type: String
- Length Constraints: Maximum length of 8192.
- Pattern: `.*`
- Required: No

**SortBy (p. 775)**

Sort model card export jobs by either name or creation time. Sorts by creation time by default.

- Type: String
- Valid Values: Name | CreationTime | Status
- Required: No

**SortOrder (p. 775)**

Sort model card export jobs by ascending or descending order.

- Type: String
- Valid Values: Ascending | Descending
- Required: No

**StatusEquals (p. 775)**

Only list model card export jobs with the specified status.

- Type: String
- Valid Values: InProgress | Completed | Failed
- Required: No
Response Syntax

```
{
    "ModelCardExportJobSummaries": [
    {
        "CreatedAt": number,
        "LastModifiedAt": number,
        "ModelCardExportJobArn": "string",
        "ModelCardExportJobName": "string",
        "ModelCardName": "string",
        "ModelCardVersion": number,
        "Status": "string"
    }
    ],
    "NextToken": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response. The following data is returned in JSON format by the service.

**ModelCardExportJobSummaries (p. 777)**

The summaries of the listed model card export jobs.

Type: Array of ModelCardExportJobSummary (p. 1667) objects

**NextToken (p. 777)**

If the response is truncated, SageMaker returns this token. To retrieve the next set of model card export jobs, use it in the subsequent request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListModelCards

Service: Amazon SageMaker Service

List existing model cards.

Request Syntax

```json
{
    "CreationTimeAfter": number,
    "CreationTimeBefore": number,
    "MaxResults": number,
    "ModelCardStatus": "string",
    "NameContains": "string",
    "NextToken": "string",
    "SortBy": "string",
    "SortOrder": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see [Common Parameters](p. 2178).

The request accepts the following data in JSON format.

**CreationTimeAfter (p. 779)**

Only list model cards that were created after the time specified.

Type: Timestamp

Required: No

**CreationTimeBefore (p. 779)**

Only list model cards that were created before the time specified.

Type: Timestamp

Required: No

**MaxResults (p. 779)**

The maximum number of model cards to list.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No

**ModelCardStatus (p. 779)**

Only list model cards with the specified approval status.

Type: String

Valid Values: Draft | PendingReview | Approved | Archived

Required: No
NameContains (p. 779)

Only list model cards with names that contain the specified string.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}\$

Required: No

NextToken (p. 779)

If the response to a previous ListModelCards request was truncated, the response includes a NextToken. To retrieve the next set of model cards, use the token in the next request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Required: No

SortBy (p. 779)

Sort model cards by either name or creation time. Sorts by creation time by default.

Type: String

Valid Values: Name | CreationTime

Required: No

SortOrder (p. 779)

Sort model cards by ascending or descending order.

Type: String

Valid Values: Ascending | Descending

Required: No

Response Syntax

```json
{
    "ModelCardSummaries": [
        {
            "CreationTime": number,
            "LastModifiedTime": number,
            "ModelCardArn": "string",
            "ModelCardName": "string",
            "ModelCardStatus": "string"
        }
    ],
    "NextToken": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.
The following data is returned in JSON format by the service.

**ModelCardSummaries (p. 780)**

The summaries of the listed model cards.

Type: Array of [ModelCardSummary (p. 1671)] objects

**NextToken (p. 780)**

If the response is truncated, SageMaker returns this token. To retrieve the next set of model cards, use it in the subsequent request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: . *

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)].

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
ListModelCardVersions

Service: Amazon SageMaker Service

List existing versions of an Amazon SageMaker Model Card.

Request Syntax

```
{
  "CreationTimeAfter": number,
  "CreationTimeBefore": number,
  "MaxResults": number,
  "ModelCardName": "string",
  "ModelCardStatus": "string",
  "NextToken": "string",
  "SortBy": "string",
  "SortOrder": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CreationTimeAfter (p. 782)**

Only list model card versions that were created after the time specified.

Type: Timestamp

Required: No

**CreationTimeBefore (p. 782)**

Only list model card versions that were created before the time specified.

Type: Timestamp

Required: No

**MaxResults (p. 782)**

The maximum number of model card versions to list.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No

**ModelCardName (p. 782)**

List model card versions for the model card with the specified name or Amazon Resource Name (ARN).

Type: String

Length Constraints: Minimum length of 1. Maximum length of 256.

Pattern: (arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]\*:\[0-9]\{12\}:model-card/.*?([a-zA-Z0-9\-]*[^a-zA-Z0-9\-])\{0,62\})
Required: Yes

**ModelCardStatus (p. 782)**

Only list model card versions with the specified approval status.

Type: String

Valid Values: Draft | PendingReview | Approved | Archived

Required: No

**NextToken (p. 782)**

If the response to a previous ListModelCardVersions request was truncated, the response includes a NextToken. To retrieve the next set of model card versions, use the token in the next request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Required: No

**SortBy (p. 782)**

Sort listed model card versions by version. Sorts by version by default.

Type: String

Valid Values: Version

Required: No

**SortOrder (p. 782)**

Sort model card versions by ascending or descending order.

Type: String

Valid Values: Ascending | Descending

Required: No

**Response Syntax**

```
{
    "ModelCardVersionSummaryList": [
    {
        "CreationTime": number,
        "LastModifiedTime": number,
        "ModelCardArn": "string",
        "ModelCardName": "string",
        "ModelCardStatus": "string",
        "ModelCardVersion": number
    }
    ],
    "NextToken": "string"
}
```
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**ModelCardVersionSummaryList (p. 783)**

The summaries of the listed versions of the model card.

Type: Array of `ModelCardVersionSummary (p. 1673)` objects

**NextToken (p. 783)**

If the response is truncated, SageMaker returns this token. To retrieve the next set of model card versions, use it in the subsequent request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: . *

Errors

For information about the errors that are common to all actions, see [Common Errors (p. 2180)].

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
ListModelExplainabilityJobDefinitions

Service: Amazon SageMaker Service

Lists model explainability job definitions that satisfy various filters.

Request Syntax

```json
{
  "CreationTimeAfter": number,
  "CreationTimeBefore": number,
  "EndpointName": "string",
  "MaxResults": number,
  "NameContains": "string",
  "NextToken": "string",
  "SortBy": "string",
  "SortOrder": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CreationTimeAfter (p. 785)**

A filter that returns only model explainability jobs created after a specified time.

Type: Timestamp

Required: No

**CreationTimeBefore (p. 785)**

A filter that returns only model explainability jobs created before a specified time.

Type: Timestamp

Required: No

**EndpointName (p. 785)**

Name of the endpoint to monitor for model explainability.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9\-]*[a-zA-Z0-9]{0,62}

Required: No

**MaxResults (p. 785)**

The maximum number of jobs to return in the response. The default value is 10.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No
**NameContains (p. 785)**

Filter for model explainability jobs whose name contains a specified string.

Type: String

Length Constraints: Maximum length of 63.

Pattern: [a-zA-Z0-9\-\_]+

Required: No

**NextToken (p. 785)**

The token returned if the response is truncated. To retrieve the next set of job executions, use it in the next request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Required: No

**SortBy (p. 785)**

Whether to sort results by the Name or CreationTime field. The default is CreationTime.

Type: String

Valid Values: Name | CreationTime

Required: No

**SortOrder (p. 785)**

Whether to sort the results in Ascending or Descending order. The default is Descending.

Type: String

Valid Values: Ascending | Descending

Required: No

**Response Syntax**

```
{"JobDefinitionSummaries": [
  {
    "CreationTime": number,
    "EndpointName": "string",
    "MonitoringJobDefinitionArn": "string",
    "MonitoringJobDefinitionName": "string"
  }
], "NextToken": "string"}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.
The following data is returned in JSON format by the service.

**JobDefinitionSummaries (p. 786)**

A JSON array in which each element is a summary for a explainability bias jobs.

Type: Array of MonitoringJobDefinitionSummary (p. 1749) objects

**NextToken (p. 786)**

The token returned if the response is truncated. To retrieve the next set of job executions, use it in the next request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: . *

**Errors**

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListModelMetadata

Service: Amazon SageMaker Service

Lists the domain, framework, task, and model name of standard machine learning models found in common model zoos.

Request Syntax

```
{
  "MaxResults": number,
  "NextToken": "string",
  "SearchExpression": {
    "Filters": [
      {
        "Name": "string",
        "Value": "string"
      }
    ]
  }
}
```

Request Parameters

For information about the parameters that are common to all actions, see [Common Parameters (p. 2178)](CommonParameters).

The request accepts the following data in JSON format.

**MaxResults (p. 788)**

The maximum number of models to return in the response.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No

**NextToken (p. 788)**

If the response to a previous ListModelMetadataResponse request was truncated, the response includes a NextToken. To retrieve the next set of model metadata, use the token in the next request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Required: No

**SearchExpression (p. 788)**

One or more filters that searches for the specified resource or resources in a search. All resource objects that satisfy the expression's condition are included in the search results. Specify the Framework, FrameworkVersion, Domain or Task to filter supported. Filter names and values are case-sensitive.

Type: [ModelMetadataSearchExpression (p. 1699)](ModelMetadataSearchExpression) object

Required: No
Response Syntax

```json
{
  "ModelMetadataSummaries": [
    {
      "Domain": "string",
      "Framework": "string",
      "FrameworkVersion": "string",
      "Model": "string",
      "Task": "string"
    }
  ],
  "NextToken": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**ModelMetadataSummaries (p. 789)**

A structure that holds model metadata.

Type: Array of ModelMetadataSummary (p. 1700) objects

**NextToken (p. 789)**

A token for getting the next set of recommendations, if there are any.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListModelPackageGroups

Service: Amazon SageMaker Service

Gets a list of the model groups in your AWS account.

Request Syntax

```
{
  "CreationTimeAfter": number,
  "CreationTimeBefore": number,
  "MaxResults": number,
  "NameContains": "string",
  "NextToken": "string",
  "SortBy": "string",
  "SortOrder": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see [Common Parameters](p. 2178).

The request accepts the following data in JSON format.

**CreationTimeAfter (p. 790)**

A filter that returns only model groups created after the specified time.

Type: Timestamp

Required: No

**CreationTimeBefore (p. 790)**

A filter that returns only model groups created before the specified time.

Type: Timestamp

Required: No

**MaxResults (p. 790)**

The maximum number of results to return in the response.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No

**NameContains (p. 790)**

A string in the model group name. This filter returns only model groups whose name contains the specified string.

Type: String

Length Constraints: Maximum length of 63.

Pattern: [a-zA-Z0-9\-\_]+
ListModelPackageGroups

Required: No

**NextToken (p. 790)**

If the result of the previous ListModelPackageGroups request was truncated, the response includes a NextToken. To retrieve the next set of model groups, use the token in the next request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Required: No

**SortBy (p. 790)**

The field to sort results by. The default is CreationTime.

Type: String

Valid Values: Name | CreationTime

Required: No

**SortOrder (p. 790)**

The sort order for results. The default is Ascending.

Type: String

Valid Values: Ascending | Descending

Required: No

Response Syntax

```
{
  "ModelPackageGroupSummaryList": [
    {
      "CreationTime": number,
      "ModelPackageGroupArn": "string",
      "ModelPackageGroupName": "string",
      "ModelPackageGroupDescription": "string",
      "ModelPackageGroupStatus": "string"
    }
  ],
  "NextToken": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**ModelPackageGroupSummaryList (p. 791)**

A list of summaries of the model groups in your AWS account.

Type: Array of ModelPackageGroupSummary (p. 1712) objects
**NextToken (p. 791)**

If the response is truncated, SageMaker returns this token. To retrieve the next set of model groups, use it in the subsequent request.

- **Type:** String
- **Length Constraints:** Maximum length of 8192.
- **Pattern:** . *

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](#).

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
**ListModelPackages**

Service: Amazon SageMaker Service

Lists the model packages that have been created.

**Request Syntax**

```
{
    "CreationTimeAfter": number,
    "CreationTimeBefore": number,
    "MaxResults": number,
    "ModelApprovalStatus": "string",
    "ModelPackageGroupName": "string",
    "ModelPackageType": "string",
    "NameContains": "string",
    "NextToken": "string",
    "SortBy": "string",
    "SortOrder": "string"
}
```

**Request Parameters**

For information about the parameters that are common to all actions, see [Common Parameters](p. 2178).

The request accepts the following data in JSON format.

**CreationTimeAfter (p. 793)**

A filter that returns only model packages created after the specified time (timestamp).

Type: Timestamp

Required: No

**CreationTimeBefore (p. 793)**

A filter that returns only model packages created before the specified time (timestamp).

Type: Timestamp

Required: No

**MaxResults (p. 793)**

The maximum number of model packages to return in the response.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No

**ModelApprovalStatus (p. 793)**

A filter that returns only the model packages with the specified approval status.

Type: String

Valid Values: Approved | Rejected | PendingManualApproval

Required: No
ModelPackageName (p. 793)

A filter that returns only model versions that belong to the specified model group.

Type: String


Pattern: (arn:aws[a-z\-]*:sagemaker:[a-zA-Z0-9\-]*:[0-9]{12}:[a-zA-Z\-]*\/)?(a-zA-Z0-9\-[0-9-]([a-zA-Z0-9-]{0,62})(?!-)$

Required: No

ModelPackageType (p. 793)

A filter that returns only the model packages of the specified type. This can be one of the following values.

- UNVERSIONED - List only unversioned models. This is the default value if no ModelPackageType is specified.
- VERSIONED - List only versioned models.
- BOTH - List both versioned and unversioned models.

Type: String

Valid Values: Versioned | Unversioned | Both

Required: No

NameContains (p. 793)

A string in the model package name. This filter returns only model packages whose name contains the specified string.

Type: String

Length Constraints: Maximum length of 63.

Pattern: [a-zA-Z0-9\-]+

Required: No

NextToken (p. 793)

If the response to a previous ListModelPackages request was truncated, the response includes a NextToken. To retrieve the next set of model packages, use the token in the next request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Required: No

SortBy (p. 793)

The parameter by which to sort the results. The default is CreationTime.

Type: String

Valid Values: Name | CreationTime

Required: No
**SortOrder (p. 793)**

The sort order for the results. The default is Ascending.

Type: String

Valid Values: Ascending | Descending

Required: No

### Response Syntax

```json
{
    "ModelPackageSummaryList": [
        {
            "CreationTime": "number",
            "ModelApprovalStatus": "string",
            "ModelPackageArn": "string",
            "ModelPackageDescription": "string",
            "ModelPackageGroupName": "string",
            "ModelPackageName": "string",
            "ModelPackageStatus": "string",
            "ModelPackageVersion": "number"
        }
    ],
    "NextToken": "string"
}
```

### Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**ModelPackageSummaryList (p. 795)**

An array of ModelPackageSummary objects, each of which lists a model package.

Type: Array of ModelPackageSummary (p. 1716) objects

**NextToken (p. 795)**

If the response is truncated, SageMaker returns this token. To retrieve the next set of model packages, use it in the subsequent request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

### Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

### See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:
• AWS Command Line Interface
• AWS SDK for .NET
• AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
ListModelQualityJobDefinitions

Service: Amazon SageMaker Service

Gets a list of model quality monitoring job definitions in your account.

Request Syntax

```
{
   "CreationTimeAfter": number,
   "CreationTimeBefore": number,
   "EndpointName": "string",
   "MaxResults": number,
   "NameContains": "string",
   "NextToken": "string",
   "SortBy": "string",
   "SortOrder": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CreationTimeAfter (p. 797)**

A filter that returns only model quality monitoring job definitions created after the specified time.

Type: Timestamp

Required: No

**CreationTimeBefore (p. 797)**

A filter that returns only model quality monitoring job definitions created before the specified time.

Type: Timestamp

Required: No

**EndpointName (p. 797)**

A filter that returns only model quality monitoring job definitions that are associated with the specified endpoint.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}$

Required: No

**MaxResults (p. 797)**

The maximum number of results to return in a call to ListModelQualityJobDefinitions.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.
Amazon SageMaker Amazon Sagemaker API Reference
ListModelQualityJobDefinitions

Required: No

**NameContains (p. 797)**

A string in the transform job name. This filter returns only model quality monitoring job definitions whose name contains the specified string.

Type: String

Length Constraints: Maximum length of 63.

Pattern: [a-zA-Z0-9-\-]+

Required: No

**NextToken (p. 797)**

If the result of the previous ListModelQualityJobDefinitions request was truncated, the response includes a NextToken. To retrieve the next set of model quality monitoring job definitions, use the token in the next request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Required: No

**SortBy (p. 797)**

The field to sort results by. The default is CreationTime.

Type: String

Valid Values: Name | CreationTime

Required: No

**SortOrder (p. 797)**

Whether to sort the results in Ascending or Descending order. The default is Descending.

Type: String

Valid Values: Ascending | Descending

Required: No

**Response Syntax**

```json
{
  "JobDefinitionSummaries": [
    {
      "CreationTime": number,
      "EndpointName": "string",
      "MonitoringJobDefinitionArn": "string",
      "MonitoringJobDefinitionName": "string"
    }
  ],
  "NextToken": "string"
}
```
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**JobDefinitionSummaries (p. 798)**

A list of summaries of model quality monitoring job definitions.

Type: Array of MonitoringJobDefinitionSummary (p. 1749) objects

**NextToken (p. 798)**

If the response is truncated, Amazon SageMaker returns this token. To retrieve the next set of model quality monitoring job definitions, use it in the next request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListModels
Service: Amazon SageMaker Service

Lists models created with the CreateModel API.

Request Syntax

```
{
  "CreationTimeAfter": number,
  "CreationTimeBefore": number,
  "MaxResults": number,
  "NameContains": "string",
  "NextToken": "string",
  "SortBy": "string",
  "SortOrder": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CreationTimeAfter (p. 800)**

A filter that returns only models with a creation time greater than or equal to the specified time (timestamp).

Type: Timestamp

Required: No

**CreationTimeBefore (p. 800)**

A filter that returns only models created before the specified time (timestamp).

Type: Timestamp

Required: No

**MaxResults (p. 800)**

The maximum number of models to return in the response.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No

**NameContains (p. 800)**

A string in the model name. This filter returns only models whose name contains the specified string.

Type: String

Length Constraints: Maximum length of 63.

Pattern: [a-zA-Z0-9-]+

Required: No
NextToken (p. 800)

If the response to a previous ListModels request was truncated, the response includes a NextToken. To retrieve the next set of models, use the token in the next request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Required: No

SortBy (p. 800)

Sorts the list of results. The default is CreationTime.

Type: String

Valid Values: Name | CreationTime

Required: No

SortOrder (p. 800)

The sort order for results. The default is Descending.

Type: String

Valid Values: Ascending | Descending

Required: No

Response Syntax

```
{
  "Models": [
    {
      "CreationTime": number,
      "ModelArn": "string",
      "ModelName": "string"
    }
  ],
  "NextToken": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

Models (p. 801)

An array of ModelSummary objects, each of which lists a model.

Type: Array of ModelSummary (p. 1727) objects

NextToken (p. 801)

If the response is truncated, SageMaker returns this token. To retrieve the next set of models, use it in the subsequent request.
Type: String
Length Constraints: Maximum length of 8192.
Pattern: .*

Errors
For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListMonitoringAlertHistory

Service: Amazon SageMaker Service

 Gets a list of past alerts in a model monitoring schedule.

Request Syntax

{
    "CreationTimeAfter": number,
    "CreationTimeBefore": number,
    "MaxResults": number,
    "MonitoringAlertName": "string",
    "MonitoringScheduleName": "string",
    "NextToken": "string",
    "SortBy": "string",
    "SortOrder": "string",
    "StatusEquals": "string"
}

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CreationTimeAfter (p. 803)**

A filter that returns only alerts created on or after the specified time.

Type: Timestamp

Required: No

**CreationTimeBefore (p. 803)**

A filter that returns only alerts created on or before the specified time.

Type: Timestamp

Required: No

**MaxResults (p. 803)**

The maximum number of results to display. The default is 100.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No

**MonitoringAlertName (p. 803)**

The name of a monitoring alert.

Type: String


Pattern: ^[a-zA-Z0-9]([-][a-zA-Z0-9]){0,62}$
Required: No

**MonitoringScheduleName (p. 803)**

The name of a monitoring schedule.

Type: String


Pattern: \^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}\$

Required: No

**NextToken (p. 803)**

If the result of the previous ListMonitoringAlertHistory request was truncated, the response includes a NextToken. To retrieve the next set of alerts in the history, use the token in the next request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Required: No

**SortBy (p. 803)**

The field used to sort results. The default is CreationTime.

Type: String

Valid Values: CreationTime | Status

Required: No

**SortOrder (p. 803)**

The sort order, whether Ascending or Descending, of the alert history. The default is Descending.

Type: String

Valid Values: Ascending | Descending

Required: No

**StatusEquals (p. 803)**

A filter that retrieves only alerts with a specific status.

Type: String

Valid Values: InAlert | OK

Required: No

**Response Syntax**

```
{
  "MonitoringAlertHistory": [  
```

804
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**MonitoringAlertHistory (p. 804)**

An alert history for a model monitoring schedule.

Type: Array of [MonitoringAlertHistorySummary (p. 1732)] objects

**NextToken (p. 804)**

If the response is truncated, SageMaker returns this token. To retrieve the next set of alerts, use it in the subsequent request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Errors

For information about the errors that are common to all actions, see [Common Errors (p. 2180)].

**ResourceNotFoundException**

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListMonitoringAlerts
Service: Amazon SageMaker Service

Gets the alerts for a single monitoring schedule.

Request Syntax

```json
{
  "MaxResults": number,
  "MonitoringScheduleName": "string",
  "NextToken": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

MaxResults (p. 807)

The maximum number of results to display. The default is 100.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No

MonitoringScheduleName (p. 807)

The name of a monitoring schedule.

Type: String


Pattern: ^[a-zA-Z0-9](\-*[a-zA-Z0-9])\{0,62}$

Required: Yes

NextToken (p. 807)

If the result of the previous ListMonitoringAlerts request was truncated, the response includes a NextToken. To retrieve the next set of alerts in the history, use the token in the next request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Required: No

Response Syntax

```json
{
}
```
"MonitoringAlertSummaries": [
    {
        "Actions": {
            "ModelDashboardIndicator": {
                "Enabled": boolean
            },
        },
        "AlertStatus": "string",
        "CreationTime": number,
        "DatapointsToAlert": number,
        "EvaluationPeriod": number,
        "LastModifiedTime": number,
        "MonitoringAlertName": "string"
    }
],
"NextToken": "string"

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**MonitoringAlertSummaries (p. 807)**

A JSON array where each element is a summary for a monitoring alert.

Type: Array of **MonitoringAlertSummary (p. 1733)** objects

Array Members: Minimum number of 1 item. Maximum number of 100 items.

**NextToken (p. 807)**

If the response is truncated, SageMaker returns this token. To retrieve the next set of alerts, use it in the subsequent request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Errors

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](#).

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
• AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
ListMonitoringExecutions

Service: Amazon SageMaker Service

Returns list of all monitoring job executions.

Request Syntax

```json
{
    "CreationTimeAfter": number,
    "CreationTimeBefore": number,
    "EndpointName": "string",
    "LastModifiedTimeAfter": number,
    "LastModifiedTimeBefore": number,
    "MaxResults": number,
    "MonitoringJobDefinitionName": "string",
    "MonitoringScheduleName": "string",
    "MonitoringTypeEquals": "string",
    "NextToken": "string",
    "ScheduledTimeAfter": number,
    "ScheduledTimeBefore": number,
    "sortBy": "string",
    "sortOrder": "string",
    "StatusEquals": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CreationTimeAfter (p. 810)**

A filter that returns only jobs created after a specified time.

Type: Timestamp

Required: No

**CreationTimeBefore (p. 810)**

A filter that returns only jobs created before a specified time.

Type: Timestamp

Required: No

**EndpointName (p. 810)**

Name of a specific endpoint to fetch jobs for.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}\}

Required: No

**LastModifiedTimeAfter (p. 810)**

A filter that returns only jobs modified before a specified time.
Type: Timestamp
Required: No

**LastModifiedTimeBefore (p. 810)**
A filter that returns only jobs modified after a specified time.
Type: Timestamp
Required: No

**MaxResults (p. 810)**
The maximum number of jobs to return in the response. The default value is 10.
Type: Integer
Valid Range: Minimum value of 1. Maximum value of 100.
Required: No

**MonitoringJobDefinitionName (p. 810)**
Gets a list of the monitoring job runs of the specified monitoring job definitions.
Type: String
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])*\{0,62}$
Required: No

**MonitoringScheduleName (p. 810)**
Name of a specific schedule to fetch jobs for.
Type: String
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])*\{0,62}$
Required: No

**MonitoringTypeEquals (p. 810)**
A filter that returns only the monitoring job runs of the specified monitoring type.
Type: String
Valid Values: DataQuality | ModelQuality | ModelBias | ModelExplainability
Required: No

**NextToken (p. 810)**
The token returned if the response is truncated. To retrieve the next set of job executions, use it in the next request.
Type: String
Length Constraints: Maximum length of 8192.
Pattern: .*
ListMonitoringExecutions

**Required:** No

### ScheduledTimeAfter (p. 810)
Filter for jobs scheduled after a specified time.

- **Type:** Timestamp
- **Required:** No

### ScheduledTimeBefore (p. 810)
Filter for jobs scheduled before a specified time.

- **Type:** Timestamp
- **Required:** No

### SortBy (p. 810)
Whether to sort the results by the Status, CreationTime, or ScheduledTime field. The default is CreationTime.

- **Type:** String
- **Valid Values:** CreationTime | ScheduledTime | Status
- **Required:** No

### SortOrder (p. 810)
Whether to sort the results in Ascending or Descending order. The default is Descending.

- **Type:** String
- **Valid Values:** Ascending | Descending
- **Required:** No

### StatusEquals (p. 810)
A filter that retrieves only jobs with a specific status.

- **Type:** String
- **Valid Values:** Pending | Completed | CompletedWithViolations | InProgress | Failed | Stopping | Stopped
- **Required:** No

### Response Syntax

```json
{
  "MonitoringExecutionSummaries": [
    {
      "CreationTime": number,
      "EndpointName": "string",
      "FailureReason": "string",
      "LastModifiedTime": number,
      "MonitoringExecutionStatus": "string",
      "MonitoringJobDefinitionName": "string",
      "MonitoringScheduleName": "string",
      "MonitoringType": "string",
      "ProcessingJobArn": "string",
    }
  ]
}
```
"ScheduledTime": number
]
, "NextToken": "string"
}

Response Elements

If the action is successful, the service sends back an HTTP 200 response. The following data is returned in JSON format by the service.

**MonitoringExecutionSummaries (p. 812)**

A JSON array in which each element is a summary for a monitoring execution.

Type: Array of MonitoringExecutionSummary (p. 1743) objects

**NextToken (p. 812)**

The token returned if the response is truncated. To retrieve the next set of job executions, use it in the next request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListMonitoringSchedules

Service: Amazon SageMaker Service

Returns list of all monitoring schedules.

Request Syntax

```json
{
    "CreationTimeAfter": number,
    "CreationTimeBefore": number,
    "EndpointName": "string",
    "LastModifiedTimeAfter": number,
    "LastModifiedTimeBefore": number,
    "MaxResults": number,
    "MonitoringJobDefinitionName": "string",
    "MonitoringTypeEquals": "string",
    "NameContains": "string",
    "NextToken": "string",
    "SortBy": "string",
    "SortOrder": "string",
    "StatusEquals": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CreationTimeAfter (p. 814)**

A filter that returns only monitoring schedules created after a specified time.

Type: Timestamp

Required: No

**CreationTimeBefore (p. 814)**

A filter that returns only monitoring schedules created before a specified time.

Type: Timestamp

Required: No

**EndpointName (p. 814)**

Name of a specific endpoint to fetch schedules for.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}

Required: No

**LastModifiedTimeAfter (p. 814)**

A filter that returns only monitoring schedules modified after a specified time.
Type: Timestamp
Required: No

**LastModifiedTimeBefore** *(p. 814)*
A filter that returns only monitoring schedules modified before a specified time.

Type: Timestamp
Required: No

**MaxResults** *(p. 814)*
The maximum number of jobs to return in the response. The default value is 10.

Type: Integer
Valid Range: Minimum value of 1. Maximum value of 100.
Required: No

**MonitoringJobDefinitionName** *(p. 814)*
Gets a list of the monitoring schedules for the specified monitoring job definition.

Type: String
Pattern: ^[a-zA-Z0-9\-\[\]*\-]*[a-zA-Z0-9\-]{0,62}$
Required: No

**MonitoringTypeEquals** *(p. 814)*
A filter that returns only the monitoring schedules for the specified monitoring type.

Type: String
Valid Values: DataQuality | ModelQuality | ModelBias | ModelExplainability
Required: No

**NameContains** *(p. 814)*
Filter for monitoring schedules whose name contains a specified string.

Type: String
Length Constraints: Maximum length of 63.
Pattern: [a-zA-Z0-9\-\-]*
Required: No

**NextToken** *(p. 814)*
The token returned if the response is truncated. To retrieve the next set of job executions, use it in the next request.

Type: String
Length Constraints: Maximum length of 8192.
Pattern: .*
ListMonitoringSchedules

Required: No

**SortBy (p. 814)**

Whether to sort the results by the Status, CreationTime, or ScheduledTime field. The default is CreationTime.

Type: String

Valid Values: Name | CreationTime | Status

Required: No

**SortOrder (p. 814)**

Whether to sort the results in Ascending or Descending order. The default is Descending.

Type: String

Valid Values: Ascending | Descending

Required: No

**StatusEquals (p. 814)**

A filter that returns only monitoring schedules modified before a specified time.

Type: String

Valid Values: Pending | Failed | Scheduled | Stopped

Required: No

**Response Syntax**

```json
{
    "MonitoringScheduleSummaries": [
        {
            "CreationTime": number,
            "EndpointName": "string",
            "LastModifiedTime": number,
            "MonitoringJobDefinitionName": "string",
            "MonitoringScheduleArn": "string",
            "MonitoringScheduleName": "string",
            "MonitoringScheduleStatus": "string",
            "MonitoringType": "string"
        }
    ],
    "NextToken": "string"
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**MonitoringScheduleSummaries (p. 816)**

A JSON array in which each element is a summary for a monitoring schedule.

Type: Array of MonitoringScheduleSummary (p. 1762) objects
**NextToken (p. 816)**

The token returned if the response is truncated. To retrieve the next set of job executions, use it in the next request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)].

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListNotebookInstanceLifecycleConfigs

Service: Amazon SageMaker Service

Lists notebook instance lifestyle configurations created with the CreateNotebookInstanceLifecycleConfig API.

Request Syntax

```json
{
  "CreationTimeAfter": number,
  "CreationTimeBefore": number,
  "LastModifiedTimeAfter": number,
  "LastModifiedTimeBefore": number,
  "MaxResults": number,
  "NameContains": "string",
  "NextToken": "string",
  "SortBy": "string",
  "SortOrder": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CreationTimeAfter (p. 818)**

A filter that returns only lifecycle configurations that were created after the specified time (timestamp).

Type: Timestamp

Required: No

**CreationTimeBefore (p. 818)**

A filter that returns only lifecycle configurations that were created before the specified time (timestamp).

Type: Timestamp

Required: No

**LastModifiedTimeAfter (p. 818)**

A filter that returns only lifecycle configurations that were modified after the specified time (timestamp).

Type: Timestamp

Required: No

**LastModifiedTimeBefore (p. 818)**

A filter that returns only lifecycle configurations that were modified before the specified time (timestamp).

Type: Timestamp
Required: No

**MaxResults (p. 818)**

The maximum number of lifecycle configurations to return in the response.  
Type: Integer  
Valid Range: Minimum value of 1. Maximum value of 100.  
Required: No

**NameContains (p. 818)**

A string in the lifecycle configuration name. This filter returns only lifecycle configurations whose name contains the specified string.  
Type: String  
Length Constraints: Maximum length of 63.  
Pattern: [a-zA-Z0-9-]+  
Required: No

**NextToken (p. 818)**

If the result of a ListNotebookInstanceLifecycleConfigs request was truncated, the response includes a NextToken. To get the next set of lifecycle configurations, use the token in the next request.  
Type: String  
Length Constraints: Maximum length of 8192.  
Pattern: .*  
Required: No

**SortBy (p. 818)**

Sorts the list of results. The default is CreationTime.  
Type: String  
Valid Values: Name | CreationTime | LastModifiedTime  
Required: No

**SortOrder (p. 818)**

The sort order for results.  
Type: String  
Valid Values: Ascending | Descending  
Required: No

**Response Syntax**

```json
{
    "NextToken": "string",
}
```
**ListNotebookInstanceLifecycleConfigs**

```json
"NotebookInstanceLifecycleConfigs": [
  {
    "CreationTime": number,
    "LastModifiedTime": number,
    "NotebookInstanceLifecycleConfigArn": "string",
    "NotebookInstanceLifecycleConfigName": "string"
  }
]
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**NextToken (p. 819)**

If the response is truncated, SageMaker returns this token. To get the next set of lifecycle configurations, use it in the next request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

**NotebookInstanceLifecycleConfigs (p. 819)**

An array of `NotebookInstanceLifecycleConfiguration` objects, each listing a lifecycle configuration.

Type: Array of `NotebookInstanceLifecycleConfigSummary (p. 1770)` objects

**Errors**

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListNotebookInstances
Service: Amazon SageMaker Service

Returns a list of the SageMaker notebook instances in the requester's account in an AWS Region.

Request Syntax

```
{
  "AdditionalCodeRepositoryEquals": "string",
  "CreationTimeAfter": number,
  "CreationTimeBefore": number,
  "DefaultCodeRepositoryContains": "string",
  "LastModifiedTimeAfter": number,
  "LastModifiedTimeBefore": number,
  "MaxResults": number,
  "NameContains": "string",
  "NextToken": "string",
  "NotebookInstanceLifecycleConfigNameContains": "string",
  "SortBy": "string",
  "SortOrder": "string",
  "StatusEquals": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**AdditionalCodeRepositoryEquals (p. 821)**

A filter that returns only notebook instances with associated with the specified git repository.

Type: String


Pattern: ^https://([^/]+)/?(.*)$|^a-zA-Z0-9(-*[a-zA-Z0-9])*

Required: No

**CreationTimeAfter (p. 821)**

A filter that returns only notebook instances that were created after the specified time (timestamp).

Type: Timestamp

Required: No

**CreationTimeBefore (p. 821)**

A filter that returns only notebook instances that were created before the specified time (timestamp).

Type: Timestamp

Required: No
**DefaultCodeRepositoryContains (p. 821)**

A string in the name or URL of a Git repository associated with this notebook instance. This filter returns only notebook instances associated with a git repository with a name that contains the specified string.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: `[a-zA-Z0-9-]+`

Required: No

**LastModifiedTimeAfter (p. 821)**

A filter that returns only notebook instances that were modified after the specified time (timestamp).

Type: Timestamp

Required: No

**LastModifiedTimeBefore (p. 821)**

A filter that returns only notebook instances that were modified before the specified time (timestamp).

Type: Timestamp

Required: No

**MaxResults (p. 821)**

The maximum number of notebook instances to return.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No

**NameContains (p. 821)**

A string in the notebook instances' name. This filter returns only notebook instances whose name contains the specified string.

Type: String

Length Constraints: Maximum length of 63.

Pattern: `[a-zA-Z0-9-]+`

Required: No

**NextToken (p. 821)**

If the previous call to the `ListNotebookInstances` is truncated, the response includes a NextToken. You can use this token in your subsequent `ListNotebookInstances` request to fetch the next set of notebook instances.

**Note**

You might specify a filter or a sort order in your request. When response is truncated, you must use the same values for the filter and sort order in the next request.

Type: String
Length Constraints: Maximum length of 8192.
Pattern: .*
Required: No

**NotebookInstanceLifecycleConfigNameContains (p. 821)**
A string in the name of a notebook instances lifecycle configuration associated with this notebook instance. This filter returns only notebook instances associated with a lifecycle configuration with a name that contains the specified string.
Type: String
Length Constraints: Maximum length of 63.
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])*$
Required: No

**SortBy (p. 821)**
The field to sort results by. The default is Name.
Type: String
Valid Values: Name  |  CreationTime  |  Status
Required: No

**SortOrder (p. 821)**
The sort order for results.
Type: String
Valid Values: Ascending  |  Descending
Required: No

**StatusEquals (p. 821)**
A filter that returns only notebook instances with the specified status.
Type: String
Valid Values: Pending  |  InService  |  Stopping  |  Stopped  |  Failed  |  Deleting  |  Updating
Required: No

**Response Syntax**
```
{
  "NextToken": "string",
  "NotebookInstances": [
    {
      "AdditionalCodeRepositories": [ "string" ],
      "CreationTime": number,
      "DefaultCodeRepository": "string",
      "InstanceType": "string",
      "LastModifiedTime": number,
      "NotebookInstanceArn": "string",
      "NotebookInstanceLifecycleConfigArn": "string",
    }
  ]
}
```
"NotebookInstanceLifecycleConfigName": "string",
"NotebookInstanceName": "string",
"NotebookInstanceStatus": "string",
"Url": "string"
]}
}

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**NextToken (p. 823)**

If the response to the previous ListNotebookInstances request was truncated, SageMaker returns this token. To retrieve the next set of notebook instances, use the token in the next request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

**NotebookInstances (p. 823)**

An array of NotebookInstanceSummary objects, one for each notebook instance.

Type: Array of NotebookInstanceSummary (p. 1772) objects

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
**ListPipelineExecutions**  
Service: Amazon SageMaker Service  

Gets a list of the pipeline executions.

**Request Syntax**

```json
{
  "CreatedAfter": number,
  "CreatedBefore": number,
  "MaxResults": number,
  "NextToken": "string",
  "PipelineName": "string",
  "SortBy": "string",
  "SortOrder": "string"
}
```

**Request Parameters**

For information about the parameters that are common to all actions, see [Common Parameters](p. 2178).

The request accepts the following data in JSON format.

- **CreatedAfter (p. 825)**
  A filter that returns the pipeline executions that were created after a specified time.
  - Type: Timestamp
  - Required: No

- **CreatedBefore (p. 825)**
  A filter that returns the pipeline executions that were created before a specified time.
  - Type: Timestamp
  - Required: No

- **MaxResults (p. 825)**
  The maximum number of pipeline executions to return in the response.
  - Type: Integer
  - Valid Range: Minimum value of 1. Maximum value of 100.
  - Required: No

- **NextToken (p. 825)**
  If the result of the previous ListPipelineExecutions request was truncated, the response includes a NextToken. To retrieve the next set of pipeline executions, use the token in the next request.
  - Type: String
  - Length Constraints: Maximum length of 8192.
  - Pattern: .*
Required: No

**PipelineName (p. 825)**

The name or Amazon Resource Name (ARN) of the pipeline.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: (arn:aws[\a-zA-Z-]*:sagemaker:[\a-zA-Z0-9-]*:[0-9]{12}:pipeline/.*)?([\a-zA-Z0-9\-]*([\a-zA-Z0-9\-]{0,255}))

Required: Yes

**SortBy (p. 825)**

The field by which to sort results. The default is *CreatedTime*.

Type: String

Valid Values: *CreationTime* | *PipelineExecutionArn*

Required: No

**SortOrder (p. 825)**

The sort order for results.

Type: String

Valid Values: *Ascending* | *Descending*

Required: No

**Response Syntax**

```json
{
  "NextToken": "string",
  "PipelineExecutionSummaries": [
    {
      "PipelineExecutionArn": "string",
      "PipelineExecutionDescription": "string",
      "PipelineExecutionDisplayName": "string",
      "PipelineExecutionFailureReason": "string",
      "PipelineExecutionStatus": "string",
      "StartTime": "number"
    }
  ]
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**NextToken (p. 826)**

If the result of the previous *ListPipelineExecutions* request was truncated, the response includes a *NextToken*. To retrieve the next set of pipeline executions, use the token in the next request.
ListPipelineExecutions

Type: String

Length Constraints: Maximum length of 8192.

Pattern: . *

**PipelineExecutionSummaries (p. 826)**

Contains a sorted list of pipeline execution summary objects matching the specified filters. Each run summary includes the Amazon Resource Name (ARN) of the pipeline execution, the run date, and the status. This list can be empty.

Type: Array of **PipelineExecutionSummary (p. 1823)** objects

Array Members: Minimum number of 0 items. Maximum number of 100 items.

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)].

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
ListPipelineExecutionSteps
Service: Amazon SageMaker Service

Gets a list of PipeLineExecutionStep objects.

Request Syntax

```
{
   "MaxResults": number,
   "NextToken": "string",
   "PipelineExecutionArn": "string",
   "SortOrder": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**MaxResults (p. 828)**

The maximum number of pipeline execution steps to return in the response.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No

**NextToken (p. 828)**

If the result of the previous ListPipelineExecutionSteps request was truncated, the response includes a NextToken. To retrieve the next set of pipeline execution steps, use the token in the next request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Required: No

**PipelineExecutionArn (p. 828)**

The Amazon Resource Name (ARN) of the pipeline execution.

Type: String

Length Constraints: Maximum length of 256.

Pattern: ^arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:pipeline\/*/.*\ execution\/*/.*$

Required: No
**SortOrder (p. 828)**

The field by which to sort results. The default is `CreatedTime`.

Type: String

Valid Values: Ascending | Descending

Required: No

**Response Syntax**

```json
{
  "NextToken": "string",
  "PipelineExecutionSteps": [
    {
      "AttemptCount": "number",
      "CacheHitResult": {
        "SourcePipelineExecutionArn": "string"
      },
      "EndTime": "number",
      "FailureReason": "string",
      "Metadata": {
        "AutoMLJob": {
          "Arn": "string"
        },
        "Callback": {
          "CallbackToken": "string",
          "OutputParameters": [
            {
              "Name": "string",
              "Value": "string"
            }
          ],
          "SqsQueueUrl": "string"
        },
        "ClarifyCheck": {
          "BaselineUsedForDriftCheckConstraints": "string",
          "CalculatedBaselineConstraints": "string",
          "CheckJobArn": "string",
          "CheckType": "string",
          "ModelPackageGroupName": "string",
          "RegisterNewBaseline": "boolean",
          "SkipCheck": "boolean",
          "ViolationReport": "string"
        },
        "Condition": {
          "Outcome": "string"
        },
        "EMR": {
          "ClusterId": "string",
          "LogFilePath": "string",
          "StepId": "string",
          "StepName": "string"
        },
        "Fail": {
          "ErrorMessage": "string"
        },
        "Lambda": {
          "Arn": "string",
          "OutputParameters": [
            {
              "Name": "string"
            }
          ]
        }
    }
  ]
}
```
"Value": "string"
]}

"Model": {
  "Arn": "string"
},

"ProcessingJob": {
  "Arn": "string"
},

"QualityCheck": {
  "BaselineUsedForDriftCheckConstraints": "string",
  "BaselineUsedForDriftCheckStatistics": "string",
  "CalculatedBaselineConstraints": "string",
  "CalculatedBaselineStatistics": "string",
  "CheckJobArn": "string",
  "CheckType": "string",
  "ModelPackageGroupName": "string",
  "RegisterNewBaseline": boolean,
  "SkipCheck": boolean,
  "ViolationReport": "string"
},

"RegisterModel": {
  "Arn": "string"
},

"TrainingJob": {
  "Arn": "string"
},

"TransformJob": {
  "Arn": "string"
},

"TuningJob": {
  "Arn": "string"
},

"SelectiveExecutionResult": {
  "SourcePipelineExecutionArn": "string"
},

"StartTime": number,
"StepDescription": "string",
"StepDisplayName": "string",
"StepName": "string",
"StepStatus": "string"
}

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**NextToken (p. 829)**

If the result of the previous ListPipelineExecutionSteps request was truncated, the response includes a NextToken. To retrieve the next set of pipeline execution steps, use the token in the next request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*
**PipelineExecutionSteps (p. 829)**

A list of PipelineExecutionStep objects. Each PipelineExecutionStep consists of StepName, StartTime, EndTime, StepStatus, and Metadata. Metadata is an object with properties for each job that contains relevant information about the job created by the step.

Type: Array of PipelineExecutionStep (p. 1818) objects

Array Members: Minimum number of 0 items. Maximum number of 100 items.

**Errors**

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListPipelineParametersForExecution
Service: Amazon SageMaker Service

Gets a list of parameters for a pipeline execution.

Request Syntax

```json
{
   "MaxResults": number,
   "NextToken": "string",
   "PipelineExecutionArn": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**MaxResults (p. 832)**

The maximum number of parameters to return in the response.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No

**NextToken (p. 832)**

If the result of the previous ListPipelineParametersForExecution request was truncated, the response includes a NextToken. To retrieve the next set of parameters, use the token in the next request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Required: No

**PipelineExecutionArn (p. 832)**

The Amazon Resource Name (ARN) of the pipeline execution.

Type: String

Length Constraints: Maximum length of 256.

Pattern: ^arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:pipeline\[/.*\]/.*$/

Required: Yes

Response Syntax

```json
{
}
```
"NextToken": "string",
"PipelineParameters": [ 
  { 
    "Name": "string",
    "Value": "string"
  }
]}

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**NextToken (p. 832)**

If the result of the previous ListPipelineParametersForExecution request was truncated, the response includes a NextToken. To retrieve the next set of parameters, use the token in the next request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

**PipelineParameters (p. 832)**

Contains a list of pipeline parameters. This list can be empty.

Type: Array of Parameter (p. 1798) objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

**Errors**

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
ListPipelines
Service: Amazon SageMaker Service

Gets a list of pipelines.

Request Syntax

```
{
  "CreatedAfter": number,
  "CreatedBefore": number,
  "MaxResults": number,
  "NextToken": "string",
  "PipelineNamePrefix": "string",
  "SortBy": "string",
  "SortOrder": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CreatedAfter (p. 835)**

A filter that returns the pipelines that were created after a specified time.

Type: Timestamp

Required: No

**CreatedBefore (p. 835)**

A filter that returns the pipelines that were created before a specified time.

Type: Timestamp

Required: No

**MaxResults (p. 835)**

The maximum number of pipelines to return in the response.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No

**NextToken (p. 835)**

If the result of the previous ListPipelines request was truncated, the response includes a NextToken. To retrieve the next set of pipelines, use the token in the next request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*
Required: No

**PipelineNamePrefix (p. 835)**

The prefix of the pipeline name.

- **Type:** String
- **Length Constraints:** Minimum length of 1. Maximum length of 256.
- **Pattern:** `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,255}$`

Required: No

**SortBy (p. 835)**

The field by which to sort results. The default is `CreatedTime`.

- **Type:** String
- **Valid Values:** Name | CreationTime

Required: No

**SortOrder (p. 835)**

The sort order for results.

- **Type:** String
- **Valid Values:** Ascending | Descending

Required: No

### Response Syntax

```json
{
  "NextToken": "string",
  "PipelineSummaries": [
    {
      "CreationTime": number,
      "LastExecutionTime": number,
      "LastModifiedTime": number,
      "PipelineArn": "string",
      "PipelineDescription": "string",
      "PipelineDisplayName": "string",
      "PipelineName": "string",
      "RoleArn": "string"
    }
  ]
}
```

### Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**NextToken (p. 836)**

If the result of the previous ListPipelines request was truncated, the response includes a NextToken. To retrieve the next set of pipelines, use the token in the next request.
ListPipelines

Type: String
Length Constraints: Maximum length of 8192.

Pattern: .*

PipelineSummaries

Contains a sorted list of PipelineSummary objects matching the specified filters. Each PipelineSummary consists of PipelineArn, PipelineName, ExperimentName, PipelineDescription, CreationTime, LastModifiedTime, LastRunTime, and RoleArn. This list can be empty.

Type: Array of PipelineSummary objects
Array Members: Minimum number of 0 items. Maximum number of 100 items.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListProcessingJobs
Service: Amazon SageMaker Service
Lists processing jobs that satisfy various filters.

Request Syntax

```json
{
    "CreationTimeAfter": number,
    "CreationTimeBefore": number,
    "LastModifiedTimeAfter": number,
    "LastModifiedTimeBefore": number,
    "MaxResults": number,
    "NameContains": "string",
    "NextToken": "string",
    "SortBy": "string",
    "SortOrder": "string",
    "StatusEquals": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CreationTimeAfter (p. 838)**

A filter that returns only processing jobs created after the specified time.

Type: Timestamp

Required: No

**CreationTimeBefore (p. 838)**

A filter that returns only processing jobs created after the specified time.

Type: Timestamp

Required: No

**LastModifiedTimeAfter (p. 838)**

A filter that returns only processing jobs modified after the specified time.

Type: Timestamp

Required: No

**LastModifiedTimeBefore (p. 838)**

A filter that returns only processing jobs modified before the specified time.

Type: Timestamp

Required: No

**MaxResults (p. 838)**

The maximum number of processing jobs to return in the response.
Type: Integer
Valid Range: Minimum value of 1. Maximum value of 100.
Required: No

**NameContains (p. 838)**
A string in the processing job name. This filter returns only processing jobs whose name contains the specified string.
Type: String
Required: No

**NextToken (p. 838)**
If the result of the previous ListProcessingJobs request was truncated, the response includes a NextToken. To retrieve the next set of processing jobs, use the token in the next request.
Type: String
Length Constraints: Maximum length of 8192.
Pattern: .*
Required: No

**SortBy (p. 838)**
The field to sort results by. The default is CreationTime.
Type: String
Valid Values: Name | CreationTime | Status
Required: No

**SortOrder (p. 838)**
The sort order for results. The default is Ascending.
Type: String
Valid Values: Ascending | Descending
Required: No

**StatusEquals (p. 838)**
A filter that retrieves only processing jobs with a specific status.
Type: String
Valid Values: InProgress | Completed | Failed | Stopping | Stopped
Required: No

**Response Syntax**
```json
{
    "NextToken": "string",
    "ProcessingJobSummaries": [
```
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**NextToken (p. 839)**

If the response is truncated, Amazon SageMaker returns this token. To retrieve the next set of processing jobs, use it in the subsequent request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*  

**ProcessingJobSummaries (p. 839)**

An array of ProcessingJobSummary objects, each listing a processing job.

Type: Array of ProcessingJobSummary (p. 1838) objects

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListProjects
Service: Amazon SageMaker Service

Gets a list of the projects in an AWS account.

Request Syntax

```
{
  "CreationTimeAfter": number,
  "CreationTimeBefore": number,
  "MaxResults": number,
  "NameContains": "string",
  "NextToken": "string",
  "SortBy": "string",
  "SortOrder": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CreationTimeAfter (p. 841)**

A filter that returns the projects that were created after a specified time.

Type: Timestamp

Required: No

**CreationTimeBefore (p. 841)**

A filter that returns the projects that were created before a specified time.

Type: Timestamp

Required: No

**MaxResults (p. 841)**

The maximum number of projects to return in the response.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No

**NameContains (p. 841)**

A filter that returns the projects whose name contains a specified string.

Type: String


Pattern: `^[a-zA-Z\-0-9]*(\-*[a-zA-Z\-0-9])\{0,31}\]`

Required: No
NextToken (p. 841)

If the result of the previous ListProjects request was truncated, the response includes a NextToken. To retrieve the next set of projects, use the token in the next request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Required: No

SortBy (p. 841)

The field by which to sort results. The default is CreationTime.

Type: String

Valid Values: Name | CreationTime

Required: No

SortOrder (p. 841)

The sort order for results. The default is Ascending.

Type: String

Valid Values: Ascending | Descending

Required: No

Response Syntax

```
{
  "NextToken": "string",
  "ProjectSummaryList": [
    {
      "CreationTime": number,
      "ProjectArn": "string",
      "ProjectDescription": "string",
      "ProjectId": "string",
      "ProjectName": "string",
      "ProjectStatus": "string"
    }
  ]
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

NextToken (p. 842)

If the result of the previous ListCompilationJobs request was truncated, the response includes a NextToken. To retrieve the next set of model compilation jobs, use the token in the next request.

Type: String
Length Constraints: Maximum length of 8192.

Pattern: . *

**ProjectSummaryList (p. 842)**

A list of summaries of projects.

Type: Array of ProjectSummary (p. 1871) objects

**Errors**

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListResourceCatalogs
Service: Amazon SageMaker Service

Lists Amazon SageMaker Catalogs based on given filters and orders. The maximum number of ResourceCatalogs viewable is 1000.

Request Syntax

```json
{
    "CreationTimeAfter": number,
    "CreationTimeBefore": number,
    "MaxResults": number,
    "NameContains": "string",
    "NextToken": "string",
    "SortBy": "string",
    "SortOrder": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CreationTimeAfter (p. 844)**

Use this parameter to search for ResourceCatalogs created after a specific date and time.

Type: Timestamp

Required: No

**CreationTimeBefore (p. 844)**

Use this parameter to search for ResourceCatalogs created before a specific date and time.

Type: Timestamp

Required: No

**MaxResults (p. 844)**

The maximum number of results returned by ListResourceCatalogs.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No

**NameContains (p. 844)**

A string that partially matches one or more ResourceCatalogs names. Filters ResourceCatalog by name.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 64.

Required: No
**NextToken (p. 844)**

A token to resume pagination of ListResourceCatalogs results.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Required: No

**SortBy (p. 844)**

The value on which the resource catalog list is sorted.

Type: String

Valid Values: CreationTime

Required: No

**SortOrder (p. 844)**

The order in which the resource catalogs are listed.

Type: String

Valid Values: Ascending | Descending

Required: No

**Response Syntax**

```
{
   "NextToken": "string",
   "ResourceCatalogs": [
   {
      "CreationTime": number,
      "Description": "string",
      "ResourceCatalogArn": "string",
      "ResourceCatalogName": "string"
   }
   ]
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**NextToken (p. 845)**

A token to resume pagination of ListResourceCatalogs results.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*
ResourceCatalogs (p. 845)

A list of the requested ResourceCatalogs.

Type: Array of ResourceCatalog (p. 1909) objects

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListSpaces

Service: Amazon SageMaker Service

Lists spaces.

Request Syntax

```
{
  "DomainIdEquals": "string",
  "MaxResults": number,
  "NextToken": "string",
  "SortBy": "string",
  "SortOrder": "string",
  "SpaceNameContains": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

DomainIdEquals (p. 847)

A parameter to search for the Domain ID.

Type: String

Length Constraints: Maximum length of 63.

Required: No

MaxResults (p. 847)

The total number of items to return in the response. If the total number of items available is more than the value specified, a NextToken is provided in the response. To resume pagination, provide the NextToken value in the as part of a subsequent call. The default value is 10.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No

NextToken (p. 847)

If the previous response was truncated, you will receive this token. Use it in your next request to receive the next set of results.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Required: No

SortBy (p. 847)

The parameter by which to sort the results. The default is CreationTime.
Type: String

Valid Values: CreationTime | LastModifiedTime

Required: No

**SortOrder (p. 847)**

The sort order for the results. The default is Ascending.

Type: String

Valid Values: Ascending | Descending

Required: No

**SpaceNameContains (p. 847)**

A parameter by which to filter the results.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

Required: No

**Response Syntax**

```json
{
    "NextToken": "string",
    "Spaces": [
        {
            "CreationTime": number,
            "DomainId": "string",
            "LastModifiedTime": number,
            "OwnershipSettingsSummary": {
                "OwnerUserProfileName": "string"
            },
            "SpaceDisplayName": "string",
            "SpaceName": "string",
            "SpaceSettingsSummary": {
                "AppType": "string",
                "SpaceStorageSettings": {
                    "EbsStorageSettings": {
                        "EbsVolumeSizeInGb": number
                    }
                }
            },
            "SpaceSharingSettingsSummary": {
                "SharingType": "string"
            },
            "Status": "string"
        }
    ]
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.
The following data is returned in JSON format by the service.

**NextToken (p. 848)**

If the previous response was truncated, you will receive this token. Use it in your next request to receive the next set of results.

Type: String
Length Constraints: Maximum length of 8192.
Pattern: . *

**Spaces (p. 848)**

The list of spaces.
Type: Array of [SpaceDetails (p. 1958)] objects

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)]

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
ListStageDevices
Service: Amazon SageMaker Service

Lists devices allocated to the stage, containing detailed device information and deployment status.

Request Syntax

```json
{
    "EdgeDeploymentPlanName": "string",
    "ExcludeDevicesDeployedInOtherStage": boolean,
    "MaxResults": number,
    "NextToken": "string",
    "StageName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**EdgeDeploymentPlanName (p. 850)**

The name of the edge deployment plan.

- Type: String
- Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$`
- Required: Yes

**ExcludeDevicesDeployedInOtherStage (p. 850)**

Toggle for excluding devices deployed in other stages.

- Type: Boolean
- Required: No

**MaxResults (p. 850)**

The maximum number of requests to select.

- Type: Integer
- Valid Range: Maximum value of 100.
- Required: No

**NextToken (p. 850)**

The response from the last list when returning a list large enough to need tokening.

- Type: String
- Length Constraints: Maximum length of 8192.
- Pattern: `.*`
Required: No

**StageName (p. 850)**

The name of the stage in the deployment.

Type: String


Pattern: \^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}\$

Required: Yes

**Response Syntax**

```
{
    "DeviceDeploymentSummaries": [
        {
            "DeployedStageName": "string",
            "DeploymentStartTime": number,
            "Description": "string",
            "DeviceArn": "string",
            "DeviceDeploymentStatus": "string",
            "DeviceDeploymentStatusMessage": "string",
            "DeviceFleetName": "string",
            "DeviceName": "string",
            "EdgeDeploymentPlanArn": "string",
            "EdgeDeploymentPlanName": "string",
            "StageName": "string"
        }
    ],
    "NextToken": "string"
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**DeviceDeploymentSummaries (p. 851)**

List of summaries of devices allocated to the stage.

Type: Array of [DeviceDeploymentSummary (p. 1412)] objects

**NextToken (p. 851)**

The token to use when calling the next page of results.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)].
See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListStudioLifecycleConfigs
Service: Amazon SageMaker Service

Lists the Amazon SageMaker Studio Lifecycle Configurations in your AWS Account.

Request Syntax

```
{
  "AppTypeEquals": "string",
  "CreationTimeAfter": number,
  "CreationTimeBefore": number,
  "MaxResults": number,
  "ModifiedTimeAfter": number,
  "ModifiedTimeBefore": number,
  "NameContains": "string",
  "NextToken": "string",
  "SortBy": "string",
  "SortOrder": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**AppTypeEquals (p. 853)**

A parameter to search for the App Type to which the Lifecycle Configuration is attached.

Type: String

Valid Values: JupyterServer | KernelGateway | JupyterLab | CodeEditor

Required: No

**CreationTimeAfter (p. 853)**

A filter that returns only Lifecycle Configurations created on or after the specified time.

Type: Timestamp

Required: No

**CreationTimeBefore (p. 853)**

A filter that returns only Lifecycle Configurations created on or before the specified time.

Type: Timestamp

Required: No

**MaxResults (p. 853)**

The total number of items to return in the response. If the total number of items available is more than the value specified, a NextToken is provided in the response. To resume pagination, provide the NextToken value in the as part of a subsequent call. The default value is 10.

Type: Integer
Valid Range: Minimum value of 1. Maximum value of 100.

Required: No

**ModifiedTimeAfter (p. 853)**

A filter that returns only Lifecycle Configurations modified after the specified time.

Type: Timestamp

Required: No

**ModifiedTimeBefore (p. 853)**

A filter that returns only Lifecycle Configurations modified before the specified time.

Type: Timestamp

Required: No

**NameContains (p. 853)**

A string in the Lifecycle Configuration name. This filter returns only Lifecycle Configurations whose name contains the specified string.

Type: String

Length Constraints: Maximum length of 63.

Pattern: \^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62\}

Required: No

**NextToken (p. 853)**

If the previous call to ListStudioLifecycleConfigs didn't return the full set of Lifecycle Configurations, the call returns a token for getting the next set of Lifecycle Configurations.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Required: No

**SortBy (p. 853)**

The property used to sort results. The default value is CreationTime.

Type: String

Valid Values: CreationTime | LastModifiedTime | Name

Required: No

**SortOrder (p. 853)**

The sort order. The default value is Descending.

Type: String

Valid Values: Ascending | Descending

Required: No
Response Syntax

```json
{
  "NextToken": "string",
  "StudioLifecycleConfigs": [
    {
      "CreationTime": number,
      "LastModifiedTime": number,
      "StudioLifecycleConfigAppType": "string",
      "StudioLifecycleConfigArn": "string",
      "StudioLifecycleConfigName": "string"
    }
  ]
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**NextToken (p. 855)**

If the previous response was truncated, you will receive this token. Use it in your next request to receive the next set of results.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

**StudioLifecycleConfigs (p. 855)**

A list of Lifecycle Configurations and their properties.

Type: Array of `StudioLifecycleConfigDetails (p. 1970)` objects

Errors

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](#).

**ResourceInUse**

Resource being accessed is in use.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
ListSubscribedWorkteams
Service: Amazon SageMaker Service

Gets a list of the work teams that you are subscribed to in the AWS Marketplace. The list may be empty if no work team satisfies the filter specified in the NameContains parameter.

Request Syntax

```json
{
  "MaxResults": number,
  "NameContains": "string",
  "NextToken": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**MaxResults (p. 857)**

The maximum number of work teams to return in each page of the response.

- Type: Integer
- Valid Range: Minimum value of 1. Maximum value of 100.
- Required: No

**NameContains (p. 857)**

A string in the work team name. This filter returns only work teams whose name contains the specified string.

- Type: String
- Pattern: ^[a-zA-Z0-9][-]*[a-zA-Z0-9]{0,62}
- Required: No

**NextToken (p. 857)**

If the result of the previous ListSubscribedWorkteams request was truncated, the response includes a NextToken. To retrieve the next set of labeling jobs, use the token in the next request.

- Type: String
- Length Constraints: Maximum length of 8192.
- Pattern: .*
- Required: No

Response Syntax

```json
{
}
```
"NextToken": "string",
"SubscribedWorkteams": [
  {
    "ListingId": "string",
    "MarketplaceDescription": "string",
    "MarketplaceTitle": "string",
    "SellerName": "string",
    "WorkteamArn": "string"
  }
]

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

NextToken (p. 857)

If the response is truncated, Amazon SageMaker returns this token. To retrieve the next set of work teams, use it in the subsequent request.

Type: String
Length Constraints: Maximum length of 8192.
Pattern: .*

SubscribedWorkteams (p. 857)

An array of Workteam objects, each describing a work team.

Type: Array of SubscribedWorkteam (p. 1972) objects

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListTags
Service: Amazon SageMaker Service

Returns the tags for the specified SageMaker resource.

Request Syntax

```
{
    "MaxResults": number,
    "NextToken": "string",
    "ResourceArn": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

MaxResults (p. 859)

- Maximum number of tags to return.
- Type: Integer
- Valid Range: Minimum value of 50.
- Required: No

NextToken (p. 859)

- If the response to the previous ListTags request is truncated, SageMaker returns this token. To retrieve the next set of tags, use it in the subsequent request.
- Type: String
- Length Constraints: Maximum length of 8192.
- Pattern: .*
- Required: No

ResourceArn (p. 859)

- The Amazon Resource Name (ARN) of the resource whose tags you want to retrieve.
- Type: String
- Length Constraints: Maximum length of 256.
- Pattern: arn:aws[a-z-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:.+
- Required: Yes

Response Syntax

```
{
}
```
Response Elements

If the action is successful, the service sends back an HTTP 200 response. The following data is returned in JSON format by the service.

NextToken (p. 859)

If response is truncated, SageMaker includes a token in the response. You can use this token in your subsequent request to fetch next set of tokens.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: . *

Tags (p. 859)

An array of Tag objects, each with a tag key and a value.

Type: Array of Tag (p. 1979) objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
**ListTrainingJobs**

Service: Amazon SageMaker Service

Lists training jobs.

**Note**

When StatusEquals and MaxResults are set at the same time, the MaxResults number of training jobs are first retrieved ignoring the StatusEquals parameter and then they are filtered by the StatusEquals parameter, which is returned as a response. For example, if ListTrainingJobs is invoked with the following parameters:

```json
{ ... MaxResults: 100, StatusEquals: InProgress ... }
```

First, 100 trainings jobs with any status, including those other than InProgress, are selected (sorted according to the creation time, from the most current to the oldest). Next, those with a status of InProgress are returned.

You can quickly test the API using the following AWS CLI code:

```bash
aws sagemaker list-training-jobs --max-results 100 --status-equals InProgress
```

**Request Syntax**

```json
{
    "CreationTimeAfter": number,
    "CreationTimeBefore": number,
    "LastModifiedTimeAfter": number,
    "LastModifiedTimeBefore": number,
    "MaxResults": number,
    "NameContains": "string",
    "NextToken": "string",
    "SortBy": "string",
    "SortOrder": "string",
    "StatusEquals": "string",
    "WarmPoolStatusEquals": "string"
}
```

**Request Parameters**

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CreationTimeAfter (p. 861)**

A filter that returns only training jobs created after the specified time (timestamp).

Type: Timestamp

Required: No

**CreationTimeBefore (p. 861)**

A filter that returns only training jobs created before the specified time (timestamp).

Type: Timestamp

Required: No

**LastModifiedTimeAfter (p. 861)**

A filter that returns only training jobs modified after the specified time (timestamp).
Type: Timestamp
Required: No

**LastModifiedTimeBefore (p. 861)**
A filter that returns only training jobs modified before the specified time (timestamp).

Type: Timestamp
Required: No

**MaxResults (p. 861)**
The maximum number of training jobs to return in the response.

Type: Integer
Valid Range: Minimum value of 1. Maximum value of 100.
Required: No

**NameContains (p. 861)**
A string in the training job name. This filter returns only training jobs whose name contains the specified string.

Type: String
Length Constraints: Maximum length of 63.
Pattern: [a-zA-Z0-9-\-]+
Required: No

**NextToken (p. 861)**
If the result of the previous `ListTrainingJobs` request was truncated, the response includes a `NextToken`. To retrieve the next set of training jobs, use the token in the next request.

Type: String
Length Constraints: Maximum length of 8192.
Pattern: .*
Required: No

**SortBy (p. 861)**
The field to sort results by. The default is `CreationTime`.

Type: String
Valid Values: Name | CreationTime | Status
Required: No

**SortOrder (p. 861)**
The sort order for results. The default is `Ascending`.

Type: String
Valid Values: Ascending | Descending
Required: No
**StatusEquals (p. 861)**

A filter that retrieves only training jobs with a specific status.

Type: String

Valid Values: InProgress | Completed | Failed | Stopping | Stopped

Required: No

**WarmPoolStatusEquals (p. 861)**

A filter that retrieves only training jobs with a specific warm pool status.

Type: String

Valid Values: Available | Terminated | Reused | InUse

Required: No

**Response Syntax**

```json
{
   "NextToken": "string",
   "TrainingJobSummaries": [
      {
         "CreationTime": number,
         "LastModifiedTime": number,
         "TrainingEndTime": number,
         "TrainingJobArn": "string",
         "TrainingJobName": "string",
         "TrainingJobStatus": "string",
         "WarmPoolStatus": {
            "ResourceRetainedBillableTimeInSeconds": number,
            "ReusedByJob": "string",
            "Status": "string"
         }
      }
   ]
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**NextToken (p. 863)**

If the response is truncated, SageMaker returns this token. To retrieve the next set of training jobs, use it in the subsequent request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

**TrainingJobSummaries (p. 863)**

An array of TrainingJobSummary objects, each listing a training job.
Type: Array of TrainingJobSummary (p. 2014) objects

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListTrainingJobsForHyperParameterTuningJob

Service: Amazon SageMaker Service

Gets a list of TrainingJobSummary objects that describe the training jobs that a hyperparameter tuning job launched.

Request Syntax

```json
{
    "HyperParameterTuningJobName": "string",
    "MaxResults": number,
    "NextToken": "string",
    "SortBy": "string",
    "SortOrder": "string",
    "StatusEquals": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**HyperParameterTuningJobName (p. 865)**

The name of the tuning job whose training jobs you want to list.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,31}

Required: Yes

**MaxResults (p. 865)**

The maximum number of training jobs to return. The default value is 10.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No

**NextToken (p. 865)**

If the result of the previous ListTrainingJobsForHyperParameterTuningJob request was truncated, the response includes a NextToken. To retrieve the next set of training jobs, use the token in the next request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Required: No

**SortBy (p. 865)**

The field to sort results by. The default is Name.
If the value of this field is `FinalObjectiveMetricValue`, any training jobs that did not return an objective metric are not listed.

Type: String

Valid Values: Name | CreationTime | Status | FinalObjectiveMetricValue

Required: No

**SortOrder (p. 865)**

The sort order for results. The default is Ascending.

Type: String

Valid Values: Ascending | Descending

Required: No

**StatusEquals (p. 865)**

A filter that returns only training jobs with the specified status.

Type: String

Valid Values: InProgress | Completed | Failed | Stopping | Stopped

Required: No

**Response Syntax**

```
{
    "NextToken": "string",
    "TrainingJobSummaries": [ 
        {
            "CreationTime": number,
            "FailureReason": "string",
            "FinalHyperParameterTuningJobObjectiveMetric": {
                "MetricName": "string",
                "Type": "string",
                "Value": number
            },
            "ObjectiveStatus": "string",
            "TrainingEndTime": number,
            "TrainingJobArn": "string",
            "TrainingJobDefinitionName": "string",
            "TrainingJobName": "string",
            "TrainingJobStatus": "string",
            "TrainingStart_time": number,
            "TunedHyperParameters": {
                "string": "string"
            },
            "TuningJobName": "string"
        }
    ]
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.
**NextToken (p. 866)**

If the result of this `ListTrainingJobsForHyperParameterTuningJob` request was truncated, the response includes a `NextToken`. To retrieve the next set of training jobs, use the token in the next request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

**TrainingJobSummaries (p. 866)**

A list of `TrainingJobSummary` objects that describe the training jobs that the `ListTrainingJobsForHyperParameterTuningJob` request returned.

Type: Array of `HyperParameterTrainingJobSummary (p. 1548)` objects

**Errors**

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceNotFoundException**

Resource being access is not found.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
ListTransformJobs

Service: Amazon SageMaker Service

Lists transform jobs.

Request Syntax

```json
{
    "CreationTimeAfter": number,
    "CreationTimeBefore": number,
    "LastModifiedTimeAfter": number,
    "LastModifiedTimeBefore": number,
    "MaxResults": number,
    "NameContains": "string",
    "NextToken": "string",
    "SortBy": "string",
    "SortOrder": "string",
    "StatusEquals": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CreationTimeAfter (p. 868)**

A filter that returns only transform jobs created after the specified time.

Type: Timestamp

Required: No

**CreationTimeBefore (p. 868)**

A filter that returns only transform jobs created before the specified time.

Type: Timestamp

Required: No

**LastModifiedTimeAfter (p. 868)**

A filter that returns only transform jobs modified after the specified time.

Type: Timestamp

Required: No

**LastModifiedTimeBefore (p. 868)**

A filter that returns only transform jobs modified before the specified time.

Type: Timestamp

Required: No

**MaxResults (p. 868)**

The maximum number of transform jobs to return in the response. The default value is 10.

---

868
Type: Integer
Valid Range: Minimum value of 1. Maximum value of 100.
Required: No

**NameContains (p. 868)**
A string in the transform job name. This filter returns only transform jobs whose name contains the specified string.
Type: String
Length Constraints: Maximum length of 63.
Pattern: [a-zA-Z0-9-\-]*
Required: No

**NextToken (p. 868)**
If the result of the previous ListTransformJobs request was truncated, the response includes a NextToken. To retrieve the next set of transform jobs, use the token in the next request.
Type: String
Length Constraints: Maximum length of 8192.
Pattern: .*
Required: No

**SortBy (p. 868)**
The field to sort results by. The default is CreationTime.
Type: String
Valid Values: Name | CreationTime | Status
Required: No

**SortOrder (p. 868)**
The sort order for results. The default is Descending.
Type: String
Valid Values: Ascending | Descending
Required: No

**StatusEquals (p. 868)**
A filter that retrieves only transform jobs with a specific status.
Type: String
Valid Values: InProgress | Completed | Failed | Stopping | Stopped
Required: No

**Response Syntax**

```json
{
}
```
"NextToken": "string",
"TransformJobSummaries": [ 
    
    "CreationTime": "number",
    "FailureReason": "string",
    "LastModifiedTime": "number",
    "TransformEndTime": "number",
    "TransformJobArn": "string",
    "TransformJobName": "string",
    "TransformJobStatus": "string"
]

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**NextToken (p. 869)**

If the response is truncated, Amazon SageMaker returns this token. To retrieve the next set of transform jobs, use it in the next request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

**TransformJobSummaries (p. 869)**

An array of TransformJobSummary objects.

Type: Array of TransformJobSummary (p. 2030) objects

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
### ListTrialComponents

**Service:** Amazon SageMaker Service

Lists the trial components in your account. You can sort the list by trial component name or creation time. You can filter the list to show only components that were created in a specific time range. You can also filter on one of the following:

- `ExperimentName`
- `SourceArn`
- `TrialName`

#### Request Syntax

```json
{
  "CreatedAfter": number,
  "CreatedBefore": number,
  "ExperimentName": "string",
  "MaxResults": number,
  "NextToken": "string",
  "SortBy": "string",
  "SortOrder": "string",
  "SourceArn": "string",
  "TrialName": "string"
}
```

#### Request Parameters

For information about the parameters that are common to all actions, see [Common Parameters](p. 2178).

The request accepts the following data in JSON format.

**CreatedAfter (p. 871)**

A filter that returns only components created after the specified time.

Type: Timestamp

Required: No

**CreatedBefore (p. 871)**

A filter that returns only components created before the specified time.

Type: Timestamp

Required: No

**ExperimentName (p. 871)**

A filter that returns only components that are part of the specified experiment. If you specify `ExperimentName`, you can't filter by `SourceArn` or `TrialName`.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: ^[a-zA-Z0-9](\-?[a-zA-Z0-9])\{0,119}$
Required: No

MaxResults (p. 871)

The maximum number of components to return in the response. The default value is 10.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No

NextToken (p. 871)

If the previous call to ListTrialComponents didn't return the full set of components, the call returns a token for getting the next set of components.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Required: No

SortBy (p. 871)

The property used to sort results. The default value is CreationTime.

Type: String

Valid Values: Name | CreationTime

Required: No

SortOrder (p. 871)

The sort order. The default value is Descending.

Type: String

Valid Values: Ascending | Descending

Required: No

SourceArn (p. 871)

A filter that returns only components that have the specified source Amazon Resource Name (ARN). If you specify SourceArn, you can't filter by ExperimentName or TrialName.

Type: String

Length Constraints: Maximum length of 256.

Required: No

TrialName (p. 871)

A filter that returns only components that are part of the specified trial. If you specify TrialName, you can't filter by ExperimentName or SourceArn.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: ^[a-zA-Z0-9]([-]*[a-zA-Z0-9])\{0,119\}
ListTrialComponents

Required: No

Response Syntax

```
{
   "NextToken": "string",
   "TrialComponentSummaries": [ 
      {
         "CreatedBy": {
            "DomainId": "string",
            "IamIdentity": {
               "Arn": "string",
               "PrincipalId": "string",
               "SourceIdentity": "string"
            },
            "UserProfileArn": "string",
            "UserProfileName": "string"
         },
         "CreationTime": number,
         "DisplayName": "string",
         "EndTime": number,
         "LastModifiedBy": {
            "DomainId": "string",
            "IamIdentity": {
               "Arn": "string",
               "PrincipalId": "string",
               "SourceIdentity": "string"
            },
            "UserProfileArn": "string",
            "UserProfileName": "string"
         },
         "LastModifiedTime": number,
         "StartTime": number,
         "Status": {
            "Message": "string",
            "PrimaryStatus": "string"
         },
         "TrialComponentArn": "string",
         "TrialComponentName": "string",
         "TrialComponentSource": {
            "SourceArn": "string",
            "SourceType": "string"
         }
      }
   ]
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**NextToken (p. 873)**

A token for getting the next set of components, if there are any.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*
**TrialComponentSummaries (p. 873)**

A list of the summaries of your trial components.

Type: Array of TrialComponentSummary (p. 2054) objects

**Errors**

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListTrials

Service: Amazon SageMaker Service

Lists the trials in your account. Specify an experiment name to limit the list to the trials that are part of that experiment. Specify a trial component name to limit the list to the trials that associated with that trial component. The list can be filtered to show only trials that were created in a specific time range. The list can be sorted by trial name or creation time.

Request Syntax

```
{
    "CreatedAfter": number,
    "CreatedBefore": number,
    "ExperimentName": "string",
    "MaxResults": number,
    "NextToken": "string",
    "SortBy": "string",
    "SortOrder": "string",
    "TrialComponentName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

CreatedAfter (p. 875)

A filter that returns only trials created after the specified time.

Type: Timestamp

Required: No

CreatedBefore (p. 875)

A filter that returns only trials created before the specified time.

Type: Timestamp

Required: No

ExperimentName (p. 875)

A filter that returns only trials that are part of the specified experiment.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: ^[a-zA-Z0-9]{0,119}$

Required: No

MaxResults (p. 875)

The maximum number of trials to return in the response. The default value is 10.

Type: Integer
Valid Range: Minimum value of 1. Maximum value of 100.

Required: No

**NextToken (p. 875)**

If the previous call to `ListTrials` didn't return the full set of trials, the call returns a token for getting the next set of trials.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Required: No

**SortBy (p. 875)**

The property used to sort results. The default value is `CreationTime`.

Type: String

Valid Values: `Name` | `CreationTime`

Required: No

**SortOrder (p. 875)**

The sort order. The default value is `Descending`.

Type: String

Valid Values: `Ascending` | `Descending`

Required: No

**TrialComponentName (p. 875)**

A filter that returns only trials that are associated with the specified trial component.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: `^[a-zA-Z0-9][-*[a-zA-Z0-9]]{0,119}$`

Required: No

**Response Syntax**

```json
{
   "NextToken": "string",
   "TrialSummaries": [
   {
   "CreationTime": number,
   "DisplayName": "string",
   "LastModifiedTime": number,
   "TrialArn": "string",
   "TrialName": "string",
   "TrialSource": {"SourceArn": "string", "SourceType": "string"}

   }]
}
```
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

NextToken (p. 876)
A token for getting the next set of trials, if there are any.

Type: String
Length Constraints: Maximum length of 8192.
Pattern: .*

TrialSummaries (p. 876)
A list of the summaries of your trials.

Type: Array of TrialSummary (p. 2057) objects

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceNotFound
Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListUserProfiles
Service: Amazon SageMaker Service
Lists user profiles.

Request Syntax

```json
{
    "DomainIdEquals": "string",
    "MaxResults": number,
    "NextToken": "string",
    "SortBy": "string",
    "SortOrder": "string",
    "UserProfileNameContains": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**DomainIdEquals (p. 878)**

A parameter by which to filter the results.

Type: String

Length Constraints: Maximum length of 63.

Required: No

**MaxResults (p. 878)**

The total number of items to return in the response. If the total number of items available is more than the value specified, a NextToken is provided in the response. To resume pagination, provide the NextToken value in the as part of a subsequent call. The default value is 10.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No

**NextToken (p. 878)**

If the previous response was truncated, you will receive this token. Use it in your next request to receive the next set of results.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Required: No

**SortBy (p. 878)**

The parameter by which to sort the results. The default is CreationTime.
ListUserProfiles

Type: String
Valid Values: CreationTime | LastModifiedTime
Required: No

**SortOrder (p. 878)**

The sort order for the results. The default is Ascending.

Type: String
Valid Values: Ascending | Descending
Required: No

**UserProfileNameContains (p. 878)**

A parameter by which to filter the results.

Type: String
Length Constraints: Maximum length of 63.
Pattern: ^[a-zA-Z0-9\-]*[a-zA-Z0-9]\{0,62\}
Required: No

**Response Syntax**

```
{
   "NextToken": "string",
   "UserProfiles": [ 
      { 
         "CreationTime": number,
         "DomainId": "string",
         "LastModifiedTime": number,
         "Status": "string",
         "UserProfileName": "string"
      } 
   ]
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**NextToken (p. 879)**

If the previous response was truncated, you will receive this token. Use it in your next request to receive the next set of results.

Type: String
Length Constraints: Maximum length of 8192.
Pattern: .*

**UserProfiles (p. 879)**

The list of user profiles.
Type: Array of [UserProfileDetails (p. 2068)] objects

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
ListWorkforces
Service: Amazon SageMaker Service

Use this operation to list all private and vendor workforces in an AWS Region. Note that you can only have one private workforce per AWS Region.

Request Syntax

```
{
    "MaxResults": number,
    "NameContains": "string",
    "NextToken": "string",
    "SortBy": "string",
    "SortOrder": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

MaxResults (p. 881)

The maximum number of workforces returned in the response.

*Type*: Integer

*Valid Range*: Minimum value of 1. Maximum value of 100.

*Required*: No

NameContains (p. 881)

A filter you can use to search for workforces using part of the workforce name.

*Type*: String


*Pattern*: `^[a-zA-Z0-9](\[[a-zA-Z0-9-\-]\]{0,62}\)$`

*Required*: No

NextToken (p. 881)

A token to resume pagination.

*Type*: String

*Length Constraints*: Maximum length of 8192.

*Pattern*: `.\*`

*Required*: No

SortBy (p. 881)

Sort workforces using the workforce name or creation date.

*Type*: String
Valid Values: Name | CreateDate  
Required: No  
**SortOrder (p. 881)**  
Sort workforces in ascending or descending order.  
Type: String  
Valid Values: Ascending | Descending  
Required: No  

**Response Syntax**
```json
{
    "NextToken": "string",
    "Workforces": [
        {
            "CognitoConfig": {
                "ClientId": "string",
                "UserPool": "string"
            },
            "CreateDate": number,
            "FailureReason": "string",
            "LastUpdatedDate": number,
            "OidcConfig": {
                "AuthorizationEndpoint": "string",
                "ClientId": "string",
                "Issuer": "string",
                "JwksUri": "string",
                "LogoutEndpoint": "string",
                "TokenEndpoint": "string",
                "UserInfoEndpoint": "string"
            },
            "SourceIpConfig": {
                "Cidrs": [ "string" ]
            },
            "Status": "string",
            "SubDomain": "string",
            "WorkforceArn": "string",
            "WorkforceName": "string",
            "WorkforceVpcConfig": {
                "SecurityGroupIds": ["string"],
                "Subnets": ["string"],
                "VpcEndpointId": "string",
                "VpcId": "String"
            }
        }
    ]
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**NextToken (p. 882)**

A token to resume pagination.
Type: String

Length Constraints: Maximum length of 8192.

Pattern: . *

**Workforces (p. 882)**

A list containing information about your workforce.

Type: Array of [Workforce (p. 2079)] objects

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)].

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
ListWorkteams

Service: Amazon SageMaker Service

Gets a list of private work teams that you have defined in a region. The list may be empty if no work team satisfies the filter specified in the NameContains parameter.

Request Syntax

```json
{
   "MaxResults": number,
   "NameContains": "string",
   "NextToken": "string",
   "SortBy": "string",
   "SortOrder": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see [Common Parameters](p. 2178).

The request accepts the following data in JSON format.

**MaxResults (p. 884)**

The maximum number of work teams to return in each page of the response.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: No

**NameContains (p. 884)**

A string in the work team's name. This filter returns only work teams whose name contains the specified string.

Type: String


Pattern: ^[a-zA-Z0-9\-]*[a-zA-Z0-9]{0,62}

Required: No

**NextToken (p. 884)**

If the result of the previous ListWorkteams request was truncated, the response includes a NextToken. To retrieve the next set of labeling jobs, use the token in the next request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*

Required: No

**SortBy (p. 884)**

The field to sort results by. The default is CreationTime.
Type: String

Valid Values: Name | CreateDate

Required: No

**SortOrder (p. 884)**

The sort order for results. The default is Ascending.

Type: String

Valid Values: Ascending | Descending

Required: No

### Response Syntax

```json
{
    "NextToken": "string",
    "Workteams": [
        {
            "CreateDate": number,
            "Description": "string",
            "LastUpdatedDate": number,
            "MemberDefinitions": [
                {
                    "CognitoMemberDefinition": {
                        "ClientId": "string",
                        "UserGroup": "string",
                        "UserPool": "string"
                    },
                    "OidcMemberDefinition": {
                        "Groups": [ "string" ]
                    }
                }
            ],
            "NotificationConfiguration": {
                "NotificationTopicArn": "string"
            },
            "ProductListingIds": [ "string" ],
            "SubDomain": "string",
            "WorkforceArn": "string",
            "WorkteamArn": "string",
            "WorkteamName": "string"
        }
    ]
}
```

### Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**NextToken (p. 885)**

If the response is truncated, Amazon SageMaker returns this token. To retrieve the next set of work teams, use it in the subsequent request.

Type: String
Length Constraints: Maximum length of 8192.
Pattern: .*

**Workteams (p. 885)**

An array of Workteam objects, each describing a work team.
Type: Array of Workteam objects

**Errors**

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
PutModelPackageGroupPolicy

Service: Amazon SageMaker Service

Adds a resource policy to control access to a model group. For information about resource policies, see Identity-based policies and resource-based policies in the AWS Identity and Access Management User Guide.

Request Syntax

```
{
    "ModelPackageGroupName": "string",
    "ResourcePolicy": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**ModelPackageName (p. 887)**

The name of the model group to add a resource policy to.

Type: String


Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$`

Required: Yes

**ResourcePolicy (p. 887)**

The resource policy for the model group.

Type: String


Pattern: `.*`

Required: Yes

Response Syntax

```
{
    "ModelPackageGroupArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.
**ModelPackageGroupArn (p. 887)**

The Amazon Resource Name (ARN) of the model package group.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: ^arn:aws(-cn|-us-gov)?:sagemaker:[a-z0-9\-]{9,16}:\d{12}:model-package-group/[^\s]{1,2048}$

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](https://docs.aws.amazon.com/AmazonSageMaker/latest/APIReference/error-response.html).

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](https://docs.aws.amazon.com/cli/index.html)
- [AWS SDK for .NET](https://docs.aws.amazon.com/sdk-for-net/v3/)
- [AWS SDK for C++](https://docs.aws.amazon.com/sdk-for-cpp/v1/)
- [AWS SDK for Go](https://docs.aws.amazon.com/sdk-for-golang/v1/)
- [AWS SDK for Java V2](https://docs.aws.amazon.com/AmazonSageMaker/latest/dg/sdk-v2-jvm.html)
- [AWS SDK for JavaScript V3](https://docs.aws.amazon.com/AmazonSageMaker/latest/dg/sdk-v3-nodejs.html)
- [AWS SDK for PHP V3](https://docs.aws.amazon.com/AmazonSageMaker/latest/dg/sdk-v3-php.html)
- [AWS SDK for Python](https://docs.aws.amazon.com/AmazonSageMaker/latest/dg/sdk-v3-python.html)
- [AWS SDK for Ruby V3](https://docs.aws.amazon.com/AmazonSageMaker/latest/dg/sdk-v3-ruby.html)
QueryLineage
Service: Amazon SageMaker Service

Use this action to inspect your lineage and discover relationships between entities. For more information, see Querying Lineage Entities in the Amazon SageMaker Developer Guide.

Request Syntax

```
{
  "Direction": "string",
  "Filters": {
    "CreatedAfter": number,
    "CreatedBefore": number,
    "LineageTypes": [ "string" ],
    "ModifiedAfter": number,
    "ModifiedBefore": number,
    "Properties": {
      "string": "string"
    },
    "Types": [ "string" ]
  },
  "IncludeEdges": boolean,
  "MaxDepth": number,
  "MaxResults": number,
  "NextToken": "string",
  "StartArns": [ "string" ]
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**Direction (p. 889)**

Associations between lineage entities have a direction. This parameter determines the direction from the StartArn(s) that the query traverses.

Type: String

Valid Values: Both | Ascendants | Descendants

Required: No

**Filters (p. 889)**

A set of filtering parameters that allow you to specify which entities should be returned.

- Properties - Key-value pairs to match on the lineage entities' properties.
- LineageTypes - A set of lineage entity types to match on. For example: TrialComponent, Artifact, or Context.
- CreatedBefore - Filter entities created before this date.
- ModifiedBefore - Filter entities modified before this date.
- ModifiedAfter - Filter entities modified after this date.

Type: QueryFilters (p. 1881) object

Required: No
**IncludeEdges (p. 889)**

Setting this value to True retrieves not only the entities of interest but also the Associations and lineage entities on the path. Set to False to only return lineage entities that match your query.

Type: Boolean
Required: No

**MaxDepth (p. 889)**

The maximum depth in lineage relationships from the StartArns that are traversed. Depth is a measure of the number of Associations from the StartArn entity to the matched results.

Type: Integer
Valid Range: Maximum value of 10.
Required: No

**MaxResults (p. 889)**

Limits the number of vertices in the results. Use the NextToken in a response to to retrieve the next page of results.

Type: Integer
Valid Range: Maximum value of 50.
Required: No

**NextToken (p. 889)**

Limits the number of vertices in the request. Use the NextToken in a response to to retrieve the next page of results.

Type: String
Length Constraints: Maximum length of 8192.
Required: No

**StartArns (p. 889)**

A list of resource Amazon Resource Name (ARN) that represent the starting point for your lineage query.

Type: Array of strings
Array Members: Minimum number of 0 items. Maximum number of 1 item.
Length Constraints: Maximum length of 256.
Pattern: arn:aws[\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:(experiment|experiment-trial-component|artifact|action|context)/.*
Required: No

**Response Syntax**

```
{
    "Edges": [
    ]
```
"AssociationType": "string",
"DestinationArn": "string",
"SourceArn": "string"
]

"NextToken": "string",
"Vertices": [ {
  "Arn": "string",
  "LineageType": "string",
  "Type": "string"
}
]

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**Edges (p. 890)**

A list of edges that connect vertices in the response.

Type: Array of Edge (p. 1433) objects

**NextToken (p. 890)**

Limits the number of vertices in the response. Use the NextToken in a response to retrieve the next page of results.

Type: String

Length Constraints: Maximum length of 8192.

**Vertices (p. 890)**

A list of vertices connected to the start entity(ies) in the lineage graph.

Type: Array of Vertex (p. 2075) objects

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
RegisterDevices
Service: Amazon SageMaker Service

Register devices.

Request Syntax

```json
{
    "DeviceFleetName": "string",
    "Devices": [  
        {  
            "Description": "string",
            "DeviceName": "string",
            "IotThingName": "string"
        }
    ],
    "Tags": [
        {  
            "Key": "string",
            "Value": "string"
        }
    ]
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**DeviceFleetName (p. 893)**

The name of the fleet.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

Required: Yes

**Devices (p. 893)**

A list of devices to register with SageMaker Edge Manager.

Type: Array of Device (p. 1411) objects

Required: Yes

**Tags (p. 893)**

The tags associated with devices.

Type: Array of Tag (p. 1979) objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No
Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceLimitExceeded

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
RenderUiTemplate
Service: Amazon SageMaker Service

Renders the UI template so that you can preview the worker's experience.

Request Syntax

```json
{
  "HumanTaskUiArn": "string",
  "RoleArn": "string",
  "Task": {
    "Input": "string"
  },
  "UiTemplate": {
    "Content": "string"
  }
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**HumanTaskUiArn (p. 895)**

TheHumanTaskUiArn of the worker UI that you want to render. Do not provide a HumanTaskUiArn if you use the UiTemplate parameter.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: `arn:aws[a-z-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:human-task-ui/.*`

Required: No

**RoleArn (p. 895)**

The Amazon Resource Name (ARN) that has access to the S3 objects that are used by the template.

Type: String


Pattern: `^arn:aws[a-z-]*:iam::\d{12}:role/?[a-zA-Z0-9+=,.@\-_]+$/`

Required: Yes

**Task (p. 895)**

A RenderableTask object containing a representative task to render.

Type: RenderableTask (p. 1905) object

Required: Yes
**UiTemplate (p. 895)**

A Template object containing the worker UI template to render.

Type: **UiTemplate (p. 2064)** object

Required: No

**Response Syntax**

```json
{
    "Errors": [
        {
            "Code": "string",
            "Message": "string"
        }
    ],
    "RenderedContent": "string"
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response. The following data is returned in JSON format by the service.

**Errors (p. 896)**

A list of one or more `RenderingError` objects if any were encountered while rendering the template. If there were no errors, the list is empty.

Type: Array of **RenderingError (p. 1906)** objects

**RenderedContent (p. 896)**

A Liquid template that renders the HTML for the worker UI.

Type: String

**Errors**

For information about the errors that are common to all actions, see **Common Errors (p. 2180)**.

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- **AWS Command Line Interface**
- **AWS SDK for .NET**
- **AWS SDK for C++**
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
RetryPipelineExecution
Service: Amazon SageMaker Service
Retry the execution of the pipeline.

Request Syntax

```json
{
    "ClientRequestToken": "string",
    "ParallelismConfiguration": {
        "MaxParallelExecutionSteps": number
    },
    "PipelineExecutionArn": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**ClientRequestToken (p. 898)**

A unique, case-sensitive identifier that you provide to ensure the idempotency of the operation. An idempotent operation completes no more than once.

Type: String


Required: Yes

**ParallelismConfiguration (p. 898)**

This configuration, if specified, overrides the parallelism configuration of the parent pipeline.

Type: ParallelismConfiguration (p. 1797) object

Required: No

**PipelineExecutionArn (p. 898)**

The Amazon Resource Name (ARN) of the pipeline execution.

Type: String

Length Constraints: Maximum length of 256.

Pattern: ^arn:aws[a-z-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:pipeline\/*\/**

Required: Yes

Response Syntax

```json
{
    "PipelineExecutionArn": "string"
}
```
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**PipelineExecutionArn (p. 898)**

The Amazon Resource Name (ARN) of the pipeline execution.

Type: String

Length Constraints: Maximum length of 256.

Pattern: ^arn:aws[a-z-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:pipeline\/*\/execution\/*$

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ConflictException**

There was a conflict when you attempted to modify a SageMaker entity such as an Experiment or Artifact.

HTTP Status Code: 400

**ResourceLimitExceeded**

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
Search

Service: Amazon SageMaker Service

Finds SageMaker resources that match a search query. Matching resources are returned as a list of SearchRecord objects in the response. You can sort the search results by any resource property in an ascending or descending order.

You can query against the following value types: numeric, text, Boolean, and timestamp.

**Note**
The Search API may provide access to otherwise restricted data. See Amazon SageMaker API Permissions: Actions, Permissions, and Resources Reference for more information.

## Request Syntax

```
{
  "CrossAccountFilterOption": "string",
  "MaxResults": number,
  "NextToken": "string",
  "Resource": "string",
  "SearchExpression": {
    "Filters": [
      {
        "Name": "string",
        "Operator": "string",
        "Value": "string"
      }
    ],
    "NestedFilters": [
      {
        "Filters": [
          {
            "Name": "string",
            "Operator": "string",
            "Value": "string"
          }
        ],
        "NestedPropertyName": "string"
      }
    ],
    "Operator": "string",
    "SubExpressions": [
      "SearchExpression"
    ],
    "SortBy": "string",
    "SortOrder": "string"
  }
}
```

## Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CrossAccountFilterOption (p. 901)**

A cross account filter option. When the value is "CrossAccount" the search results will only include resources made discoverable to you from other accounts. When the value is
"SameAccount" or null the search results will only include resources from your account. Default is null. For more information on searching for resources made discoverable to your account, see Search discoverable resources in the SageMaker Developer Guide. The maximum number of ResourceCatalogs viewable is 1000.

Type: String
Valid Values: SameAccount | CrossAccount
Required: No

MaxResults (p. 901)

The maximum number of results to return.
Type: Integer
Valid Range: Minimum value of 1. Maximum value of 100.
Required: No

NextToken (p. 901)

If more than MaxResults resources match the specified SearchExpression, the response includes a NextToken. The NextToken can be passed to the next SearchRequest to continue retrieving results.
Type: String
Length Constraints: Maximum length of 8192.
Pattern: .*
Required: No

Resource (p. 901)

The name of the SageMaker resource to search for.
Type: String
Valid Values: TrainingJob | Experiment | ExperimentTrial | ExperimentTrialComponent | Endpoint | ModelPackage | ModelPackageGroup | Pipeline | PipelineExecution | FeatureGroup | Project | FeatureMetadata | HyperParameterTuningJob | ModelCard | Model
Required: Yes

SearchExpression (p. 901)

A Boolean conditional statement. Resources must satisfy this condition to be included in search results. You must provide at least one subexpression, filter, or nested filter. The maximum number of recursive SubExpressions, NestedFilters, and Filters that can be included in a SearchExpression object is 50.
Type: SearchExpression (p. 1936) object
Required: No

SortBy (p. 901)

The name of the resource property used to sort the SearchResults. The default is LastModifiedTime.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 255.
Pattern: .+
Required: No

**SortOrder (p. 901)**

How SearchResults are ordered. Valid values are Ascending or Descending. The default is Descending.

Type: String
Valid Values: Ascending | Descending
Required: No

**Response Syntax**

```json
{
    "NextToken": "string",
    "Results": [
        {
            "Endpoint": {
                "CreationTime": number,
                "DataCaptureConfig": {
                    "CaptureStatus": "string",
                    "CurrentSamplingPercentage": number,
                    "DestinationS3Uri": "string",
                    "EnableCapture": boolean,
                    "KmsKeyId": "string"
                },
                "EndpointArn": "string",
                "EndpointConfigName": "string",
                "EndpointName": "string",
                "EndpointStatus": "string",
                "FailureReason": "string",
                "LastModifiedTime": number,
                "MonitoringSchedules": [
                    {
                        "CreationTime": number,
                        "EndpointName": "string",
                        "FailureReason": "string",
                        "LastModifiedTime": number,
                        "LastMonitoringExecutionSummary": {
                            "CreationTime": number,
                            "EndTime": "string",
                            "FailureReason": "string",
                            "LastModifiedTime": number,
                            "MonitoringExecutionStatus": "string",
                            "MonitoringJobDefinitionName": "string",
                            "MonitoringScheduleName": "string",
                            "MonitoringType": "string",
                            "ProcessingJobArn": "string",
                            "ScheduledTime": number
                        },
                        "MonitoringScheduleArn": "string",
                        "MonitoringScheduleConfig": {
                            "BaselineConfig": {
                                "BaselineJobName": "string",
                                "BaselineJobVersion": "string"
                            },
                            "MonitoringJobDefinitionName": "string",
                            "MonitoringScheduleName": "string",
                            "MonitoringType": "string"  
                        }  
                    }
                ]
            }
        }
    ]
}
```
"ConstraintsResource": {
  "S3Uri": "string"
},
"StatisticsResource": {
  "S3Uri": "string"
}
},
"Environment": {
  "string": "string"
},
"MonitoringAppSpecification": {
  "ContainerArguments": [ "string" ],
  "ContainerEntrypoint": [ "string" ],
  "ImageUri": "string",
  "PostAnalyticsProcessorSourceUri": "string",
  "RecordPreprocessorSourceUri": "string"
},
"MonitoringInputs": [
  {
    "BatchTransformInput": {
      "DataCapturedDestinationS3Uri": "string",
      "DatasetFormat": {
        "Csv": {
          "Header": boolean
        },
        "Json": {
          "Line": boolean
        },
        "Parquet": {
        }
      },
      "EndTimeOffset": "string",
      "ExcludeFeaturesAttribute": "string",
      "FeaturesAttribute": "string",
      "InferenceAttribute": "string",
      "LocalPath": "string",
      "ProbabilityAttribute": "string",
      "ProbabilityThresholdAttribute": number,
      "S3DataDistributionType": "string",
      "S3InputMode": "string",
      "StartTimeOffset": "string"
    },
    "EndpointInput": {
      "EndpointName": "string",
      "EndTimeOffset": "string",
      "ExcludeFeaturesAttribute": "string",
      "FeaturesAttribute": "string",
      "InferenceAttribute": "string",
      "LocalPath": "string",
      "ProbabilityAttribute": "string",
      "ProbabilityThresholdAttribute": number,
      "S3DataDistributionType": "string",
      "S3InputMode": "string",
      "StartTimeOffset": "string"
    }
  }
},
"MonitoringOutputConfig": {
  "KmsKeyId": "string",
  "MonitoringOutputs": [
    {
      "S3Output": {
        "LocalPath": "string",
        "S3UploadMode": "string",
        "S3Uri": "string"
      }
    }
  ]
}
{
    "MonitoringResources": {
        "ClusterConfig": {
            "InstanceCount": number,
            "InstanceType": string,
            "VolumeKmsKeyId": string,
            "VolumeSizeInGB": number
        },
        "NetworkConfig": {
            "EnableInterContainerTrafficEncryption": boolean,
            "EnableNetworkIsolation": boolean,
            "VpcConfig": {
                "SecurityGroupIds": [ "string" ],
                "Subnets": [ "string" ]
            }
        },
        "RoleArn": "string",
        "StoppingCondition": {
            "MaxRuntimeInSeconds": number
        },
        "MonitoringJobDefinitionName": "string",
        "MonitoringType": "string",
        "ScheduleConfig": {
            "DataAnalysisEndTime": "string",
            "DataAnalysisStartTime": "string",
            "ScheduleExpression": "string"
        },
        "MonitoringScheduleName": "string",
        "MonitoringScheduleStatus": "string",
        "MonitoringType": "string",
        "Tags": [
            { "Key": "string",
              "Value": "string" }
        ]
    },
    "ProductionVariants": [
        {
            "CurrentInstanceCount": number,
            "CurrentServerlessConfig": {
                "MaxConcurrency": number,
                "MemorySizeInMB": number,
                "ProvisionedConcurrency": number
            },
            "CurrentWeight": number,
            "DeployedImages": [
                { "ResolutionTime": number,
                  "ResolvedImage": "string",
                  "SpecifiedImage": "string"
                }
            ],
            "DesiredInstanceCount": number,
            "DesiredServerlessConfig": {
                "MaxConcurrency": number,
                "MemorySizeInMB": number,
                "ProvisionedConcurrency": number
            },
            "DesiredWeight": number
        }
    ]
}
"ManagedInstanceScaling": {
  "MaxInstanceCount": number,
  "MinInstanceCount": number,
  "Status": "string"
},
"RoutingConfig": {
  "RoutingStrategy": "string"
},
"VariantName": "string",
"VariantStatus": [
  {
    "StartTime": number,
    "Status": "string",
    "StatusMessage": "string"
  }
],
"ShadowProductionVariants": [
  {
    "CurrentInstanceCount": number,
    "CurrentServerlessConfig": {
      "MaxConcurrency": number,
      "MemorySizeInMB": number,
      "ProvisionedConcurrency": number
    },
    "CurrentWeight": number,
    "DeployedImages": [
      {
        "ResolutionTime": number,
        "ResolvedImage": "string",
        "SpecifiedImage": "string"
      }
    ],
    "DesiredInstanceCount": number,
    "DesiredServerlessConfig": {
      "MaxConcurrency": number,
      "MemorySizeInMB": number,
      "ProvisionedConcurrency": number
    },
    "DesiredWeight": number,
    "ManagedInstanceScaling": {
      "MaxInstanceCount": number,
      "MinInstanceCount": number,
      "Status": "string"
    },
    "RoutingConfig": {
      "RoutingStrategy": "string"
    },
    "VariantName": "string",
    "VariantStatus": [
      {
        "StartTime": number,
        "Status": "string",
        "StatusMessage": "string"
      }
    ]
  }
],
"Tags": [
  {
    "Key": "string",
    "Value": "string"
  }
]
"Experiment": {
  "CreatedBy": {
    "DomainId": "string",
    "IamIdentity": {
      "Arn": "string",
      "PrincipalId": "string",
      "SourceIdentity": "string"
    },
    "UserProfileArn": "string",
    "UserProfileName": "string"
  },
  "CreationTime": number,
  "Description": "string",
  "DisplayName": "string",
  "ExperimentArn": "string",
  "ExperimentName": "string",
  "LastModifiedBy": {
    "DomainId": "string",
    "IamIdentity": {
      "Arn": "string",
      "PrincipalId": "string",
      "SourceIdentity": "string"
    },
    "UserProfileArn": "string",
    "UserProfileName": "string"
  },
  "LastModifiedTime": number,
  "Source": {
    "SourceArn": "string",
    "SourceType": "string"
  },
  "Tags": [
    {
      "Key": "string",
      "Value": "string"
    }
  ]
},
"FeatureGroup": {
  "CreationTime": number,
  "Description": "string",
  "EventTimeFeatureName": "string",
  "FailureReason": "string",
  "FeatureDefinitions": [
    {
      "CollectionConfig": { ... },
      "CollectionType": "string",
      "FeatureName": "string",
      "FeatureType": "string"
    }
  ],
  "FeatureGroupArn": "string",
  "FeatureGroupName": "string",
  "FeatureGroupStatus": "string",
  "LastModifiedTime": number,
  "LastUpdateStatus": {
    "FailureReason": "string",
    "Status": "string"
  },
  "OfflineStoreConfig": {
    "DataCatalogConfig": {
      "Catalog": "string",
      "Database": "string",
      "TableName": "string"
    },
    "DisableGlueTableCreation": boolean,
  }
"S3StorageConfig": { 
  "KmsKeyId": "string",
  "ResolvedOutputS3Uri": "string",
  "S3Uri": "string"
},
"TableFormat": "string"
},
"OfflineStoreStatus": { 
  "BlockedReason": "string",
  "Status": "string"
},
"OnlineStoreConfig": { 
  "EnableOnlineStore": boolean,
  "SecurityConfig": { 
    "KmsKeyId": "string"
  },
  "StorageType": "string",
  "TtlDuration": { 
    "Unit": "string",
    "Value": number
  }
},
"RecordIdentifierFeatureName": "string",
"RoleArn": "string",
"Tags": [
  {
    "Key": "string",
    "Value": "string"
  }
],
"FeatureMetadata": { 
  "CreationTime": number,
  "Description": "string",
  "FeatureGroupArn": "string",
  "FeatureGroupName": "string",
  "FeatureName": "string",
  "FeatureType": "string",
  "LastModifiedTime": number,
  "Parameters": [
    {
      "Key": "string",
      "Value": "string"
    }
  ],
"HyperParameterTuningJob": { 
  "BestTrainingJob": { 
    "CreationTime": number,
    "FailureReason": "string",
    "FinalHyperParameterTuningJobObjectiveMetric": { 
      "MetricName": "string",
      "Type": "string",
      "Value": number
    },
    "ObjectiveStatus": "string",
    "TrainingEndTime": number,
    "TrainingJobArn": "string",
    "TrainingJobDefinitionName": "string",
    "TrainingJobName": "string",
    "TrainingJobStatus": "string",
    "TrainingStartTime": number,
    "TunedHyperParameters": [
      "string": "string"
    ],
    "TuningJobName": "string"
"ConsumedResources": {
    "RuntimeInSeconds": number
},
"CreationTime": number,
"FailureReason": "string",
"HyperParameterTuningEndTime": number,
"HyperParameterTuningJobArn": "string",
"HyperParameterTuningJobConfig": {
    "HyperParameterTuningJobObjective": {
        "MetricName": "string",
        "Type": "string"
    },
    "ParameterRanges": {
        "AutoParameters": [
            {
                "Name": "string",
                "ValueHint": "string"
            }
        ],
        "CategoricalParameterRanges": [
            {
                "Name": "string",
                "Values": ["string"]
            }
        ],
        "ContinuousParameterRanges": [
            {
                "MaxValue": "string",
                "MinValue": "string",
                "Name": "string",
                "ScalingType": "string"
            }
        ],
        "IntegerParameterRanges": [
            {
                "MaxValue": "string",
                "MinValue": "string",
                "Name": "string",
                "ScalingType": "string"
            }
        ]
    },
    "RandomSeed": number,
    "ResourceLimits": {
        "MaxNumberOfTrainingJobs": number,
        "MaxParallelTrainingJobs": number,
        "MaxRuntimeInSeconds": number
    },
    "Strategy": "string",
    "StrategyConfig": {
        "HyperbandStrategyConfig": {
            "MaxResource": number,
            "MinResource": number
        }
    },
    "TrainingJobEarlyStoppingType": "string",
    "TuningJobCompletionCriteria": {
        "BestObjectiveNotImproving": {
            "MaxNumberOfTrainingJobsNotImproving": number
        },
        "ConvergenceDetected": {
            "CompleteOnConvergence": "string"
        },
        "TargetObjectiveMetricValue": number
    }
}
"HyperParameterTuningJobName": "string",
"HyperParameterTuningJobStatus": "string",
"LastModifiedTime": "number",
"ObjectiveStatusCounters": {
  "Failed": "number",
  "Pending": "number",
  "Succeeded": "number"
},
"OverallBestTrainingJob": {
  "CreationTime": "number",
  "FailureReason": "string",
  "FinalHyperParameterTuningJobObjectiveMetric": {
    "MetricName": "string",
    "Type": "string",
    "Value": "number"
  },
  "ObjectiveStatus": "string",
  "TrainingEndTime": "number",
  "TrainingJobArn": "string",
  "TrainingJobDefinitionName": "string",
  "TrainingJobName": "string",
  "TrainingJobStatus": "string",
  "TrainingStart_time": "number",
  "TunedHyperParameters": {
    "string": "string"
  },
  "TuningJobName": "string"
},
"Tags": [
  {
    "Key": "string",
    "Value": "string"
  }
],
"TrainingJobDefinition": {
  "AlgorithmSpecification": {
    "AlgorithmName": "string",
    "MetricDefinitions": [
      {
        "Name": "string",
        "Regex": "string"
      }
    ],
    "TrainingImage": "string",
    "TrainingInputMode": "string"
  },
  "CheckpointConfig": {
    "LocalPath": "string",
    "S3Uri": "string"
  },
  "DefinitionName": "string",
  "EnableInterContainerTrafficEncryption": "boolean",
  "EnableManagedSpotTraining": "boolean",
  "EnableNetworkIsolation": "boolean",
  "Environment": {
    "string": "string"
  },
  "HyperParameterRanges": {
    "AutoParameters": [
      {
        "Name": "string",
        "ValueHint": "string"
      }
    ],
    "CategoricalParameterRanges": ["string"]
  }
}
{
  "Name": "string",
  "Values": [ "string" ]
}
],
"ContinuousParameterRanges": [
{
  "MaxValue": "string",
  "MinValue": "string",
  "Name": "string",
  "ScalingType": "string"
}
],
"IntegerParameterRanges": [
{
  "MaxValue": "string",
  "MinValue": "string",
  "Name": "string",
  "ScalingType": "string"
}
]
},
"HyperParameterTuningResourceConfig": {
  "AllocationStrategy": "string",
  "InstanceConfigs": [
  {
    "InstanceCount": number,
    "InstanceType": "string",
    "VolumeSizeInGB": number
  }
  ],
  "InstanceCount": number,
  "InstanceType": "string",
  "VolumeKmsKeyId": "string",
  "VolumeSizeInGB": number
},
"InputDataConfig": [
{
  "ChannelName": "string",
  "CompressionType": "string",
  "ContentType": "string",
  "DataSource": {
    "FileSystemDataSource": {
      "DirectoryPath": "string",
      "FileSystemAccessMode": "string",
      "FileSystemId": "string",
      "FileSystemType": "string"
    },
    "S3DataSource": {
      "AttributeNames": [ "string" ],
      "InstanceGroupNames": [ "string" ],
      "S3DataDistributionType": "string",
      "S3DataType": "string",
      "S3Uri": "string"
    }
  },
  "InputMode": "string",
  "RecordWrapperType": "string",
  "ShuffleConfig": {
    "Seed": number
  }
}
],
"OutputDataConfig": {
  "CompressionType": "string",
  "KmsKeyId": "string",
  "OutputMode": "string",
  "RecordWrapperType": "string",
  "ShuffleConfig": {
    "Seed": number
  }
}
"S3OutputPath": "string",
"ResourceConfig": {
  "InstanceCount": number,
  "InstanceGroups": [
    {
      "InstanceCount": number,
      "InstanceGroupName": "string",
      "InstanceType": "string"
    }
  ],
  "InstanceType": "string",
  "KeepAlivePeriodInSeconds": number,
  "VolumeKmsKeyId": "string",
  "VolumeSizeInGB": number
},
"RetryStrategy": {
  "MaximumRetryAttempts": number
},
"RoleArn": "string",
"StaticHyperParameters": {
  "string": "string"
},
"StoppingCondition": {
  "MaxPendingTimeInSeconds": number,
  "MaxRuntimeInSeconds": number,
  "MaxWaitTimeInSeconds": number
},
"TuningObjective": {
  "MetricName": "string",
  "Type": "string"
},
"VpcConfig": {
  "SecurityGroupIds": [ "string" ],
  "Subnets": [ "string" ]
},
"TrainingJobDefinitions": [
  {
    "AlgorithmSpecification": {
      "AlgorithmName": "string",
      "MetricDefinitions": [
        {
          "Name": "string",
          "Regex": "string"
        }
      ],
      "TrainingImage": "string",
      "TrainingInputMode": "string"
    },
    "CheckpointConfig": {
      "LocalPath": "string",
      "S3Uri": "string"
    },
    "DefinitionName": "string",
    "EnableInterContainerTrafficEncryption": boolean,
    "EnableManagedSpotTraining": boolean,
    "EnableNetworkIsolation": boolean,
    "Environment": {
      "string": "string"
    },
    "HyperParameterRanges": {
      "AutoParameters": [
        {
          "Name": "string",
          "ValueHint": "string"
        }
      ]
    }
  }
]


`CategoricalParameterRanges`: [
```
{
  "Name": "string",
  "Values": [ "string" ]
}
```
],
`ContinuousParameterRanges`: [
```
{
  "MaxValue": "string",
  "MinValue": "string",
  "Name": "string",
  "ScalingType": "string"
}
```
],
`IntegerParameterRanges`: [
```
{
  "MaxValue": "string",
  "MinValue": "string",
  "Name": "string",
  "ScalingType": "string"
}
```
]
`HyperParameterTuningResourceConfig": {
```
"AllocationStrategy": "string",
"InstanceConfigs": [
```
{
  "InstanceCount": number,
  "InstanceType": "string",
  "VolumeSizeInGB": number
}
```
],
"InstanceCount": number,
"InstanceType": "string",
"VolumeKmsKeyId": "string",
"VolumeSizeInGB": number
```
]
`InputDataConfig": [
```
{
  "ChannelName": "string",
  "CompressionType": "string",
  "ContentType": "string",
  "DataSource": {
  "FileSystemDataSource": {
    "DirectoryPath": "string",
    "FileSystemAccessMode": "string",
    "FileSystemId": "string",
    "FileSystemType": "string"
  },
  "S3DataSource": {
    "AttributeNames": [ "string" ],
    "InstanceGroupNames": [ "string" ],
    "S3DataDistributionType": "string",
    "S3DataType": "string",
    "S3Uri": "string"
  }
  },
  "InputMode": "string",
  "RecordWrapperType": "string",
  "ShuffleConfig": {
    "Seed": number
  }
}
```
]
"OutputDataConfig": {
  "CompressionType": "string",
  "KmsKeyId": "string",
  "S3OutputPath": "string"
},
"ResourceConfig": {
  "InstanceCount": number,
  "InstanceGroups": [
    {
      "InstanceCount": number,
      "InstanceGroupName": "string",
      "InstanceType": "string"
    }
  ],
  "InstanceType": "string",
  "KeepAlivePeriodInSeconds": number,
  "VolumeKmsKeyId": "string",
  "VolumeSizeInGB": number
},
"RetryStrategy": {
  "MaximumRetryAttempts": number
},
"RoleArn": "string",
"StaticHyperParameters": {
  "string": "string"
},
"StoppingCondition": {
  "MaxPendingTimeInSeconds": number,
  "MaxRuntimeInSeconds": number,
  "MaxWaitTimeInSeconds": number
},
"TuningObjective": {
  "MetricName": "string",
  "Type": "string"
},
"VpcConfig": {
  "SecurityGroupIds": [ "string" ],
  "Subnets": [ "string" ]
},
"TrainingJobStatusCounters": {
  "Completed": number,
  "InProgress": number,
  "NonRetryableError": number,
  "RetryableError": number,
  "Stopped": number
},
"TuningJobCompletionDetails": {
  "ConvergenceDetectedTime": number,
  "NumberOfTrainingJobsObjectiveNotImproving": number
},
"WarmStartConfig": {
  "ParentHyperParameterTuningJobs": [
    {
      "HyperParameterTuningJobName": "string"
    }
  ],
  "WarmStartType": "string"
},
"Model": {
  "Endpoints": [
    {
      "CreationTime": number,
      "EndpointArn": "string"
    }
  ]
}
"EndpointName": "string",
"EndpointStatus": "string",
"LastModifiedTime": number
],
"LastBatchTransformJob": {
  "AutoMLJobArn": "string",
  "BatchStrategy": "string",
  "CreationTime": number,
  "DataCaptureConfig": {
    "DestinationS3Uri": "string",
    "GenerateInferenceId": boolean,
    "KmsKeyId": "string"
  },
  "DataProcessing": {
    "InputFilter": "string",
    "JoinSource": "string",
    "OutputFilter": "string"
  },
  "Environment": {
    "string": "string"
  },
  "ExperimentConfig": {
    "ExperimentName": "string",
    "RunName": "string",
    "TrialComponentDisplayName": "string",
    "TrialName": "string"
  },
  "FailureReason": "string",
  "LabelingJobArn": "string",
  "MaxConcurrentTransforms": number,
  "MaxPayloadInMB": number,
  "ModelClientConfig": {
    "InvocationsMaxRetries": number,
    "InvocationsTimeoutInSeconds": number
  },
  "ModelName": "string",
  "Tags": [
    {
      "Key": "string",
      "Value": "string"
    }
  ],
  "TransformEndTime": number,
  "TransformInput": {
    "CompressionType": "string",
    "ContentType": "string",
    "DataSource": {
      "S3DataSource": {
        "S3DataType": "string",
        "S3Uri": "string"
      }
    },
    "SplitType": "string"
  },
  "TransformJobArn": "string",
  "TransformJobName": "string",
  "TransformJobStatus": "string",
  "TransformOutput": {
    "Accept": "string",
    "AssembleWith": "string",
    "KmsKeyId": "string",
    "S3OutputPath": "string"
  },
  "TransformResources": {
    "InstanceCount": number,
    ""}
"InstanceType": "string",
"VolumeKmsKeyId": "string"
},
"TransformStartTime": number
},
"Model": {
"Containers": [
{
"ContainerHostname": "string",
"Environment": {
"string": "string"
},
"Image": "string",
"ImageConfig": {
"RepositoryAccessMode": "string",
"RepositoryAuthConfig": {
"RepositoryCredentialsProviderArn": "string"
}
},
"InferenceSpecificationName": "string",
"Mode": "string",
"ModelDataSource": {
"S3DataSource": {
"CompressionType": "string",
"ModelAccessConfig": {
"AcceptEula": boolean
},
"S3DataType": "string",
"S3Uri": "string"
}
},
"ModelDataUrl": "string",
"ModelPackageName": "string",
"MultiModelConfig": {
"ModelCacheSetting": "string"
}
}
],
"CreationTime": number,
"DeploymentRecommendation": {
"RealTimeInferenceRecommendations": [
{
"Environment": {
"string": "string"
},
"InstanceType": "string",
"RecommendationId": "string"
}
],
"RecommendationStatus": "string"
},
"EnableNetworkIsolation": boolean,
"ExecutionRoleArn": "string",
"InferenceExecutionConfig": {
"Mode": "string"
}
},
"ModelArn": "string",
"ModelName": "string",
"PrimaryContainer": {
"ContainerHostname": "string",
"Environment": {
"string": "string"
},
"Image": "string",
"ImageConfig": {
"RepositoryAccessMode": "string",
"RepositoryAuthConfig": {
  "RepositoryCredentialsProviderArn": "string"
},
"InferenceSpecificationName": "string",
"Mode": "string",
"ModelDataSource": {
  "S3DataSource": {
    "CompressionType": "string",
    "ModelAccessConfig": {
      "AcceptEula": boolean
    },
    "S3DataType": "string",
    "S3Uri": "string"
  }
},
"ModelDataUrl": "string",
"ModelPackageName": "string",
"MultiModelConfig": {
  "ModelCacheSetting": "string"
},
"Tags": [
  {
    "Key": "string",
    "Value": "string"
  }
],
"VpcConfig": {
  "SecurityGroupIds": [ "string" ],
  "Subnets": [ "string" ]
},
"ModelCard": {
  "CreatedBy": {
    "DomainId": "string",
    "IamIdentity": {
      "Arn": "string",
      "PrincipalId": "string",
      "SourceIdentity": "string"
    },
    "UserProfileArn": "string",
    "UserProfileName": "string"
  },
  "CreationTime": number,
  "LastModifiedBy": {
    "DomainId": "string",
    "IamIdentity": {
      "Arn": "string",
      "PrincipalId": "string",
      "SourceIdentity": "string"
    },
    "UserProfileArn": "string",
    "UserProfileName": "string"
  },
  "LastModifiedTime": number,
  "ModelCardArn": "string",
  "ModelCardName": "string",
  "ModelCardStatus": "string",
  "ModelCardVersion": number,
  "ModelId": "string",
  "RiskRating": "string",
  "SecurityConfig": {
    "KmsKeyId": "string"
  },
  "Tags": [
    }

```json
{
  "Key": "string",
  "Value": "string"
}
]
"MonitoringSchedules": [
{
  "BatchTransformInput": {
    "DataCapturedDestinationS3Uri": "string",
    "DatasetFormat": {
      "Csv": { "Header": boolean },
      "Json": { "Line": boolean },
      "Parquet": { }
    },
    "EndTimeOffset": "string",
    "ExcludeFeaturesAttribute": "string",
    "FeaturesAttribute": "string",
    "InferenceAttribute": "string",
    "LocalPath": "string",
    "ProbabilityAttribute": "string",
    "ProbabilityThresholdAttribute": number,
    "S3DataDistributionType": "string",
    "S3InputMode": "string",
    "StartTimeOffset": "string"
  },
  "CreationTime": number,
  "EndpointName": "string",
  "FailureReason": "string",
  "LastModifiedTime": number,
  "LastMonitoringExecutionSummary": {
    "CreationTime": number,
    "EndpointName": "string",
    "FailureReason": "string",
    "LastModifiedTime": number,
    "MonitoringExecutionStatus": "string",
    "MonitoringJobDefinitionName": "string",
    "MonitoringScheduleName": "string",
    "MonitoringType": "string",
    "ProcessingJobArn": "string",
    "ScheduledTime": number
  },
  "MonitoringAlertSummaries": [
    {
      "Actions": {
        "ModelDashboardIndicator": { "Enabled": boolean }
      },
      "AlertStatus": "string",
      "CreationTime": number,
      "DatapointsToAlert": number,
      "EvaluationPeriod": number,
      "LastModifiedTime": number,
      "MonitoringAlertName": "string"
    }
  ],
  "MonitoringScheduleArn": "string",
  "MonitoringScheduleConfig": {
    "MonitoringJobDefinition": {
      "BaselineConfig": {
      }
    }
  }
}```
"BaseliningJobName": "string",
"ConstraintsResource": {
  "S3Uri": "string"
},
"StatisticsResource": {
  "S3Uri": "string"
},
"Environment": {
  "string": "string"
},
"MonitoringAppSpecification": {
  "ContainerArguments": [ "string" ],
  "ContainerEntrypoint": [ "string" ],
  "ImageUri": "string",
  "PostAnalyticsProcessorSourceUri": "string",
  "RecordPreprocessorSourceUri": "string"
},
"MonitoringInputs": [
  {
    "BatchTransformInput": {
      "DataCapturedDestinationS3Uri": "string",
      "DatasetFormat": {
        "Csv": {
          "Header": boolean
        },
        "Json": {
          "Line": boolean
        },
        "Parquet": {
        }
      },
      "EndTimeOffset": "string",
      "ExcludeFeaturesAttribute": "string",
      "FeaturesAttribute": "string",
      "InferenceAttribute": "string",
      "LocalPath": "string",
      "ProbabilityAttribute": "string",
      "ProbabilityThresholdAttribute": number,
      "S3DataDistributionType": "string",
      "S3InputMode": "string",
      "StartTimeOffset": "string"
    },
    "EndpointInput": {
      "EndpointName": "string",
      "EndTimeOffset": "string",
      "ExcludeFeaturesAttribute": "string",
      "FeaturesAttribute": "string",
      "InferenceAttribute": "string",
      "LocalPath": "string",
      "ProbabilityAttribute": "string",
      "ProbabilityThresholdAttribute": number,
      "S3DataDistributionType": "string",
      "S3InputMode": "string",
      "StartTimeOffset": "string"
    }
  }
},
"MonitoringOutputConfig": {
  "KmsKeyId": "string",
  "MonitoringOutputs": [
    {
      "S3Output": {
        "LocalPath": "string",
        "S3UploadMode": "string",
        "S3Uri": "string"
      }
    }
  ]
}
"MonitoringResources": {
"ClusterConfig": {
"InstanceCount": number,
"InstanceType": "string",
"VolumeKmsKeyId": "string",
"VolumeSizeInGB": number
}
},
"NetworkConfig": {
"EnableInterContainerTrafficEncryption": boolean,
"EnableNetworkIsolation": boolean,
"VpcConfig": {
"SecurityGroupIds": [ "string" ],
"Subnets": [ "string" ]
}
},
"RoleArn": "string",
"StoppingCondition": {
"MaxRuntimeInSeconds": number
}
},
"MonitoringJobDefinitionName": "string",
"MonitoringType": "string",
"ScheduleConfig": {
"DataAnalysisEndTime": "string",
"DataAnalysisStartTime": "string",
"ScheduleExpression": "string"
}
},
"MonitoringScheduleName": "string",
"MonitoringScheduleStatus": "string",
"MonitoringType": "string"
}
],
"ModelCard": {
"Content": "string",
"CreatedBy": {
"DomainId": "string",
"IamIdentity": {
"Arn": "string",
"PrincipalId": "string",
"SourceIdentity": "string"
},
"UserProfileArn": "string",
"UserProfileName": "string"
},
"CreationTime": number,
"LastModifiedBy": {
"DomainId": "string",
"IamIdentity": {
"Arn": "string",
"PrincipalId": "string",
"SourceIdentity": "string"
},
"UserProfileArn": "string",
"UserProfileName": "string"
},
"LastModifiedTime": number,
"ModelCardArn": "string",
"ModelCardName": "string",
"ModelCardStatus": "string"}
"ModelCardVersion": number,
"ModelId": "string",
"ModelPackageGroupName": "string",
"RiskRating": "string",
"SecurityConfig": {
    "KmsKeyId": "string"
},
"Tags": [
    {
        "Key": "string",
        "Value": "string"
    }
],
"ModelPackage": {
    "AdditionalInferenceSpecifications": [
        {
            "Containers": [
                {
                    "AdditionalS3DataSource": {
                        "CompressionType": "string",
                        "S3DataType": "string",
                        "S3Uri": "string"
                    },
                    "ContainerHostname": "string",
                    "Environment": {
                        "string": "string"
                    },
                    "Framework": "string",
                    "FrameworkVersion": "string",
                    "Image": "string",
                    "ImageDigest": "string",
                    "ModelDataURL": "string",
                    "ModelInput": {
                        "DataInputConfig": "string"
                    },
                    "NearestModelName": "string",
                    "ProductId": "string"
                }
            ],
            "Description": "string",
            "Name": "string",
            "SupportedContentTypes": [ "string" ],
            "SupportedRealtimeInferenceInstanceTypes": [ "string" ],
            "SupportedResponseMIMETypes": [ "string" ],
            "SupportedTransformInstanceTypes": [ "string" ]
        }
    ],
    "ApprovalDescription": "string",
    "CertifyForMarketplace": boolean,
    "CreatedBy": {
        "DomainId": "string",
        "IamIdentity": {
            "Arn": "string",
            "PrincipalId": "string",
            "SourceIdentity": "string"
        },
        "UserProfileArn": "string",
        "UserProfileName": "string"
    },
    "CreationTime": number,
    "CustomerMetadataProperties": {
        "string": "string"
    }
},
"Domain": "string",
"DriftCheckBaselines": ["string"]
"Bias": { 
  "ConfigFile": { 
    "ContentDigest": "string",
    "ContentType": "string",
    "S3Uri": "string"
  },
  "PostTrainingConstraints": { 
    "ContentDigest": "string",
    "ContentType": "string",
    "S3Uri": "string"
  },
  "PreTrainingConstraints": { 
    "ContentDigest": "string",
    "ContentType": "string",
    "S3Uri": "string"
  }
},
"Explainability": { 
  "ConfigFile": { 
    "ContentDigest": "string",
    "ContentType": "string",
    "S3Uri": "string"
  },
  "Constraints": { 
    "ContentDigest": "string",
    "ContentType": "string",
    "S3Uri": "string"
  }
},
"ModelDataQuality": { 
  "Constraints": { 
    "ContentDigest": "string",
    "ContentType": "string",
    "S3Uri": "string"
  },
  "Statistics": { 
    "ContentDigest": "string",
    "ContentType": "string",
    "S3Uri": "string"
  }
},
"ModelQuality": { 
  "Constraints": { 
    "ContentDigest": "string",
    "ContentType": "string",
    "S3Uri": "string"
  },
  "Statistics": { 
    "ContentDigest": "string",
    "ContentType": "string",
    "S3Uri": "string"
  }
}
},
"InferenceSpecification": { 
  "Containers": [ 
    { 
      "AdditionalS3DataSource": { 
        "CompressionType": "string",
        "S3DataType": "string",
        "S3Uri": "string"
      },
      "ContainerHostname": "string",
      "Environment": { 
        "string": "string"
      }
    }
  ]
}
"Framework": "string",
"FrameworkVersion": "string",
"Image": "string",
"ImageDigest": "string",
"ModelDataUrl": "string",
"ModelInput": {
  "DataInputConfig": "string"
},
"NearestModelName": "string",
"ProductId": "string"
},
"SupportedContentTypes": [ "string" ],
"SupportedRealtimeInferenceInstanceTypes": [ "string" ],
"SupportedResponseMIMETypes": [ "string" ],
"SupportedTransformInstanceTypes": [ "string" ]
},
"LastModifiedBy": {
  "DomainId": "string",
  "IamIdentity": {
    "Arn": "string",
    "PrincipalId": "string",
    "SourceIdentity": "string"
  },
  "UserProfileArn": "string",
  "UserProfileName": "string"
},
"LastModifiedTime": number,
"MetadataProperties": {
  "CommitId": "string",
  "GeneratedBy": "string",
  "ProjectId": "string",
  "Repository": "string"
},
"ModelApprovalStatus": "string",
"ModelMetrics": {
  "Bias": {
    "PostTrainingReport": {
      "ContentDigest": "string",
      "ContentType": "string",
      "S3Uri": "string"
    },
    "PreTrainingReport": {
      "ContentDigest": "string",
      "ContentType": "string",
      "S3Uri": "string"
    },
    "Report": {
      "ContentDigest": "string",
      "ContentType": "string",
      "S3Uri": "string"
    }
  },
  "Explainability": {
    "Report": {
      "ContentDigest": "string",
      "ContentType": "string",
      "S3Uri": "string"
    }
  },
  "ModelDataQuality": {
    "Constraints": {
      "ContentDigest": "string",
      "ContentType": "string",
      "S3Uri": "string"
    }
  }
},
"ModelQuality": {
  "Constraints": {
    "ContentDigest": "string",
    "ContentType": "string",
    "S3Uri": "string"
  },
  "Statistics": {
    "ContentDigest": "string",
    "ContentType": "string",
    "S3Uri": "string"
  }
},
"ModelPackageArn": "string",
"ModelPackageDescription": "string",
"ModelPackageGroupName": "string",
"ModelPackageName": "string",
"ModelPackageStatus": "string",
"ModelPackageStatusDetails": {
  "ImageScanStatuses": [
    {
      "FailureReason": "string",
      "Name": "string",
      "Status": "string"
    }
  ],
  "ValidationStatuses": [
    {
      "FailureReason": "string",
      "Name": "string",
      "Status": "string"
    }
  ]
},
"ModelPackageVersion": number,
"SamplePayloadUrl": "string",
"SkipModelValidation": "string",
"SourceAlgorithmSpecification": {
  "SourceAlgorithms": [
    {
      "AlgorithmName": "string",
      "ModelDataUrl": "string"
    }
  ]
},
"tags": [
  {
    "Key": "string",
    "Value": "string"
  }
],
"Task": "string",
"ValidationSpecification": {
  "ValidationProfiles": [
    {
      "ProfileName": "string",
      "TransformJobDefinition": {
        "BatchStrategy": "string",
        "Environment": {
          "string": "string"
        }
      }
    }
  ]
}
],
"MaxConcurrentTransforms": number,
"MaxPayloadInMB": number,
"TransformInput": {
  "CompressionType": "string",
  "ContentType": "string",
  "DataSource": {
    "S3DataSource": {
      "S3DataType": "string",
      "S3Uri": "string"
    }
  },
  "SplitType": "string"
},
"TransformOutput": {
  "Accept": "string",
  "AssembleWith": "string",
  "KmsKeyId": "string",
  "S3OutputPath": "string"
},
"TransformResources": {
  "InstanceCount": number,
  "InstanceType": "string",
  "VolumeKmsKeyId": "string"
}
},
"ValidationRole": "string"
},
"ModelPackageGroup": {
  "CreatedBy": {
    "DomainId": "string",
    "IamIdentity": {
      "Arn": "string",
      "PrincipalId": "string",
      "SourceIdentity": "string"
    },
    "UserProfileArn": "string",
    "UserProfileName": "string"
  },
  "CreationTime": number,
  "ModelPackageGroupArn": "string",
  "ModelPackageGroupDescription": "string",
  "ModelPackageGroupName": "string",
  "ModelPackageGroupStatus": "string",
  "Tags": [
    {
      "Key": "string",
      "Value": "string"
    }
  ]
},
"Pipeline": {
  "CreatedBy": {
    "DomainId": "string",
    "IamIdentity": {
      "Arn": "string",
      "PrincipalId": "string",
      "SourceIdentity": "string"
    },
    "UserProfileArn": "string",
    "UserProfileName": "string"
  },
  "CreationTime": number,
  "PipelineArn": "string",
  "PipelineDescription": "string",
  "PipelineName": "string",
  "PipelineStatus": "string",
  "Tags": [
    {
      "Key": "string",
      "Value": "string"
    }
  ]
}}
"LastModifiedBy": {  
  "DomainId": "string",  
  "IamIdentity": {  
    "Arn": "string",  
    "PrincipalId": "string",  
    "SourceIdentity": "string"  
  },  
  "UserProfileArn": "string",  
  "UserProfileName": "string"  
},  
"LastModifiedTime": number,  
"LastRunTime": number,  
"ParallelismConfiguration": {  
  "MaxParallelExecutionSteps": number  
},  
"PipelineArn": "string",  
"PipelineDescription": "string",  
"PipelineDisplayName": "string",  
"PipelineName": "string",  
"PipelineStatus": "string",  
"RoleArn": "string",  
"Tags": [  
  {  
    "Key": "string",  
    "Value": "string"  
  }  
],  
"PipelineExecution": {  
  "CreatedBy": {  
    "DomainId": "string",  
    "IamIdentity": {  
      "Arn": "string",  
      "PrincipalId": "string",  
      "SourceIdentity": "string"  
    },  
    "UserProfileArn": "string",  
    "UserProfileName": "string"  
  },  
  "CreationTime": number,  
  "FailureReason": "string",  
  "LastModifiedBy": {  
    "DomainId": "string",  
    "IamIdentity": {  
      "Arn": "string",  
      "PrincipalId": "string",  
      "SourceIdentity": "string"  
    },  
    "UserProfileArn": "string",  
    "UserProfileName": "string"  
  },  
  "LastModifiedTime": number,  
  "ParallelismConfiguration": {  
    "MaxParallelExecutionSteps": number  
  },  
  "PipelineArn": "string",  
  "PipelineExecutionArn": "string",  
  "PipelineExecutionDescription": "string",  
  "PipelineExecutionDisplayName": "string",  
  "PipelineExecutionStatus": "string",  
  "PipelineExperimentConfig": {  
    "ExperimentName": "string",  
    "TrialName": "string"  
  },  
  "PipelineParameters": [  
    
  ]  
}
"Name": "string",
"Value": "string"
],
"SelectiveExecutionConfig": {
  "SelectedSteps": [
    {
      "StepName": "string"
    }
  ],
  "SourcePipelineExecutionArn": "string"
},
"Project": {
  "CreatedBy": {
    "DomainId": "string",
    "IamIdentity": {
      "Arn": "string",
      "PrincipalId": "string",
      "SourceIdentity": "string"
    },
    "UserProfileArn": "string",
    "UserProfileName": "string"
  },
  "CreationTime": number,
  "LastModifiedBy": {
    "DomainId": "string",
    "IamIdentity": {
      "Arn": "string",
      "PrincipalId": "string",
      "SourceIdentity": "string"
    },
    "UserProfileArn": "string",
    "UserProfileName": "string"
  },
  "LastModifiedTime": number,
  "ProjectArn": "string",
  "ProjectDescription": "string",
  "ProjectId": "string",
  "ProjectName": "string",
  "ProjectStatus": "string",
  "ServiceCatalogProvisionedProductDetails": {
    "ProvisionedProductId": "string",
    "ProvisionedProductStatusMessage": "string"
  },
  "ServiceCatalogProvisioningDetails": {
    "PathId": "string",
    "ProductId": "string",
    "ProvisioningArtifactId": "string",
    "ProvisioningParameters": [
      {
        "Key": "string",
        "Value": "string"
      }
    ],
    "Tags": [
      {
        "Key": "string",
        "Value": "string"
      }
    ],
    "TrainingJob": {
      "AlgorithmSpecification": {
        "AlgorithmName": "string",
        "AlgorithmArn": "string"
      }
    }
  }
}
"ContainerArguments": [ "string" ],
"ContainerEntrypoint": [ "string" ],
"EnableSageMakerMetricsTimeSeries": boolean,
"MetricDefinitions": [ 
  { 
    "Name": "string",
    "Regex": "string"
  }
],
"TrainingImage": "string",
"TrainingImageConfig": { 
  "TrainingRepositoryAccessMode": "string",
  "TrainingRepositoryAuthConfig": { 
    "TrainingRepositoryCredentialsProviderArn": "string"
  }
},
"TrainingInputMode": "string"
},
"AutoMLJobArn": "string",
"BillableTimeInSeconds": number,
"CheckpointConfig": { 
  "LocalPath": "string",
  "S3Uri": "string"
},
"CreationTime": number,
"DebugHookConfig": { 
  "CollectionConfigurations": [ 
    { 
      "CollectionName": "string",
      "CollectionParameters": { 
        "string": "string" 
      }
    }
  ],
  "HookParameters": { 
    "string": "string"
  },
  "LocalPath": "string",
  "S3OutputPath": "string"
},
"DebugRuleConfigurations": [ 
  { 
    "InstanceType": "string",
    "LocalPath": "string",
    "RuleConfigurationName": "string",
    "RuleEvaluatorImage": "string",
    "RuleParameters": { 
      "string": "string"
    },
    "S3OutputPath": "string",
    "VolumeSizeInGB": number
  }
],
"DebugRuleEvaluationStatuses": [ 
  { 
    "LastModifiedTime": number,
    "RuleConfigurationName": "string",
    "RuleEvaluationJobArn": "string",
    "RuleEvaluationStatus": "string",
    "StatusDetails": "string"
  }
],
"EnableInterContainerTrafficEncryption": boolean,
"EnableManagedSpotTraining": boolean,
"EnableNetworkIsolation": boolean,
"Environment": { 
  "string": "string" 
}
"string" : "string"
],
"ExperimentConfig": {
  "ExperimentName": "string",
  "RunName": "string",
  "TrialComponentDisplayName": "string",
  "TrialName": "string"
},
"FailureReason": "string",
"FinalMetricDataList": [
  {
    "MetricName": "string",
    "Timestamp": number,
    "Value": number
  }
],
"HyperParameters": {
  "string": "string"
},
"InputDataConfig": [
  {
    "ChannelName": "string",
    "CompressionType": "string",
    "ContentType": "string",
    "DataSource": {
      "FileSystemDataSource": {
        "DirectoryPath": "string",
        "FileSystemAccessMode": "string",
        "FileSystemId": "string",
        "FileSystemType": "string"
      },
      "S3DataSource": {
        "AttributeNames": [ "string" ],
        "InstanceGroupNames": [ "string" ],
        "S3DataDistributionType": "string",
        "S3DataType": "string",
        "S3Uri": "string"
      }
    },
    "InputMode": "string",
    "RecordWrapperType": "string",
    "ShuffleConfig": {
      "Seed": number
    }
  }
],
"LabelingJobArn": "string",
"LastModifiedTime": number,
"ModelArtifacts": {
  "S3ModelArtifacts": "string"
},
"OutputDataConfig": {
  "CompressionType": "string",
  "KmsKeyId": "string",
  "S3OutputPath": "string"
},
"ProfilerConfig": {
  "DisableProfiler": boolean,
  "ProfilingIntervalInMilliseconds": number,
  "ProfilingParameters": {
    "string": "string"
  },
  "S3OutputPath": "string"
},
"ResourceConfig": {
  "InstanceCount": number,
"InstanceGroups": [
  {
    "InstanceCount": number,
    "InstanceGroupName": "string",
    "InstanceType": "string"
  }
],
"InstanceType": "string",
"KeepAlivePeriodInSeconds": number,
"VolumeKmsKeyId": "string",
"VolumeSizeInGB": number
],
"RetryStrategy": {
  "MaximumRetryAttempts": number
},
"RoleArn": "string",
"SecondaryStatus": "string",
"SecondaryStatusTransitions": [
  {
    "EndTime": number,
    "StartTime": number,
    "Status": "string",
    "StatusMessage": "string"
  }
],
"StoppingCondition": {
  "MaxPendingTimeInSeconds": number,
  "MaxRuntimeInSeconds": number,
  "MaxWaitTimeInSeconds": number
},
"Tags": [
  {
    "Key": "string",
    "Value": "string"
  }
],
"TensorBoardOutputConfig": {
  "LocalPath": "string",
  "S3OutputPath": "string"
},
"TrainingEndTime": number,
"TrainingJobArn": "string",
"TrainingJobName": "string",
"TrainingJobStatus": "string",
"TrainingStartTime": number,
"TrainingTimeInSeconds": number,
"TuningJobArn": "string",
"VpcConfig": {
  "SecurityGroupIds": [ "string" ],
  "Subnets": [ "string"
}]
],
"Trial": {
  "CreatedBy": {
    "DomainId": "string",
    "IamIdentity": {
      "Arn": "string",
      "PrincipalId": "string",
      "SourceIdentity": "string"
    },
    "UserProfileArn": "string",
    "UserProfileName": "string"
  },
  "CreationTime": number,
  "DisplayName": "string",
  "ExperimentName": "string"
"LastModifiedBy": {
  "DomainId": "string",
  "IamIdentity": {
    "Arn": "string",
    "PrincipalId": "string",
    "SourceIdentity": "string"
  },
  "UserProfileArn": "string",
  "UserProfileName": "string"
},
"LastModifiedTime": number,
"MetadataProperties": {
  "CommitId": "string",
  "CreatedBy": "string",
  "ProjectId": "string",
  "Repository": "string"
},
"Source": {
  "SourceArn": "string",
  "SourceType": "string"
},
"Tags": [
  {
    "Key": "string",
    "Value": "string"
  }
],
"TrialArn": "string",
"TrialComponentSummaries": [
  {
    "CreatedBy": {
      "DomainId": "string",
      "IamIdentity": {
        "Arn": "string",
        "PrincipalId": "string",
        "SourceIdentity": "string"
      },
      "UserProfileArn": "string",
      "UserProfileName": "string"
    },
    "CreationTime": number,
    "TrialComponentArn": "string",
    "TrialComponentName": "string",
    "TrialComponentSource": {
      "SourceArn": "string",
      "SourceType": "string"
    }
  }
],
"TrialName": "string"
},
"TrialComponent": {
  "CreatedBy": {
    "DomainId": "string",
    "IamIdentity": {
      "Arn": "string",
      "PrincipalId": "string",
      "SourceIdentity": "string"
    },
    "UserProfileArn": "string",
    "UserProfileName": "string"
  },
  "CreationTime": number,
  "DisplayName": "string",
  "EndTime": number,
  "InputArtifacts": {
"string" : { 
    "MediaType": "string",
    "Value": "string"
  }
},
"LastModifiedBy": { 
    "DomainId": "string",
    "IamIdentity": { 
        "Arn": "string",
        "PrincipalId": "string",
        "SourceIdentity": "string"
    },
    "UserProfileArn": "string",
    "UserProfileName": "string"
},
"LastModifiedTime": number,
"LineageGroupArn": "string",
"MetadataProperties": { 
    "CommitId": "string",
    "GeneratedBy": "string",
    "ProjectId": "string",
    "Repository": "string"
},
"Metrics": [ 
    { 
        "Avg": number,
        "Count": number,
        "Last": number,
        "Max": number,
        "MetricName": "string",
        "Min": number,
        "SourceArn": "string",
        "StdDev": number,
        "TimeStamp": number
    }
],
"OutputArtifacts": { 
    "string" : { 
        "MediaType": "string",
        "Value": "string"
    }
},
"Parameters": { 
    "string" : { 
        "NumberValue": number,
        "StringValue": "string"
    }
},
"Parents": [ 
    { 
        "ExperimentName": "string",
        "TrialName": "string"
    }
],
"RunName": "string",
"Source": { 
    "SourceArn": "string",
    "SourceType": "string"
},
"SourceDetail": { 
    "ProcessingJob": { 
        "AppSpecification": { 
            "ContainerArguments": [ "string" ],
            "ContainerEntrypoint": [ "string" ],
            "ImageUri": "string"
        },
        "AppSpecification": { 
            "ContainerArguments": [ "string" ],
            "ContainerEntrypoint": [ "string" ],
            "ImageUri": "string"
        }
    }
}
"AutoMLJobArn": "string",
"CreationTime": number,
"Environment": {
"string": "string"
},
"ExitMessage": "string",
"ExperimentConfig": {
"ExperimentName": "string",
"RunName": "string",
"TrialComponentDisplayName": "string",
"TrialName": "string"
},
"FailureReason": "string",
"LastModifiedTime": number,
"MonitoringScheduleArn": "string",
"NetworkConfig": {
"EnableInterContainerTrafficEncryption": boolean,
"EnableNetworkIsolation": boolean,
"VpcConfig": {
"SecurityGroupIds": [ "string" ],
"Subnets": [ "string" ]
}
},
"ProcessingEndTime": number,
"ProcessingInputs": [
{
"AppManaged": boolean,
"DatasetDefinition": {
"AthenaDatasetDefinition": {
"Catalog": "string",
"Database": "string",
"KmsKeyId": "string",
"OutputCompression": "string",
"OutputFormat": "string",
"OutputS3Uri": "string",
"QueryString": "string",
"WorkGroup": "string"
}
},
"DataDistributionType": "string",
"InputMode": "string",
"LocalPath": "string",
"RedshiftDatasetDefinition": {
"ClusterId": "string",
"ClusterRoleArn": "string",
"Database": "string",
"DbUser": "string",
"KmsKeyId": "string",
"OutputCompression": "string",
"OutputFormat": "string",
"OutputS3Uri": "string",
"QueryString": "string",
"WorkGroup": "string"
}
},
"InputName": "string",
"S3Input": {
"LocalPath": "string",
"S3CompressionType": "string",
"S3DataDistributionType": "string",
"S3DataType": "string",
"S3InputMode": "string",
"S3Uri": "string"
}
}
],
"ProcessingJobArn": "string",
"ProcessingJobName": "string",
"string": "string"
"ProcessingJobStatus": "string",
"ProcessingOutputConfig": {
  "KmsKeyId": "string",
  "Outputs": [
    {
      "AppManaged": boolean,
      "FeatureStoreOutput": {
        "FeatureGroupName": "string"
      },
      "OutputName": "string",
      "S3Output": {
        "LocalPath": "string",
        "S3UploadMode": "string",
        "S3Uri": "string"
      }
    }
  ]
},
"ProcessingResources": {
  "ClusterConfig": {
    "InstanceCount": number,
    "InstanceType": "string",
    "VolumeKmsKeyId": "string",
    "VolumeSizeInGB": number
  },
  "ProcessingStartTime": number,
  "RoleArn": "string",
  "StoppingCondition": {
    "MaxRuntimeInSeconds": number
  },
  "Tags": [
    {
      "Key": "string",
      "Value": "string"
    }
  ],
  "TrainingJobArn": "string"
},
"SourceArn": "string",
"TrainingJob": {
  "AlgorithmSpecification": {
    "AlgorithmName": "string",
    "ContainerArguments": [ "string" ],
    "ContainerEntrypoint": [ "string" ],
    "EnableSageMakerMetricsTimeSeries": boolean,
    "MetricDefinitions": [ ]
  },
  "TrainingImage": "string",
  "TrainingImageConfig": {
    "TrainingRepositoryAccessMode": "string",
    "TrainingRepositoryAuthConfig": {
      "TrainingRepositoryCredentialsProviderArn": "string"
    }
  },
  "TrainingInputMode": "string"
},
"AutoMLJobArn": "string",
"BillableTimeInSeconds": number,
"CheckpointConfig": {
  "LocalPath": "string",
  "S3Uri": "string"
},
"CreationTime": number,
"DebugHookConfig": {
  "CollectionConfigurations": [
    {
      "CollectionName": "string",
      "CollectionParameters": {
        "string": "string"
      }
    }
  ],
  "HookParameters": {
    "string": "string"
  },
  "LocalPath": "string",
  "S3OutputPath": "string"
},
"DebugRuleConfigurations": [
  {
    "InstanceType": "string",
    "LocalPath": "string",
    "RuleConfigurationName": "string",
    "RuleEvaluatorImage": "string",
    "RuleParameters": {
      "string": "string"
    },
    "S3OutputPath": "string",
    "VolumeSizeInGB": number
  }
],
"DebugRuleEvaluationStatuses": [
  {
    "LastModifiedTime": number,
    "RuleConfigurationName": "string",
    "RuleEvaluationJobArn": "string",
    "RuleEvaluationStatus": "string",
    "StatusDetails": "string"
  }
],
"EnableInterContainerTrafficEncryption": boolean,
"EnableManagedSpotTraining": boolean,
"EnableNetworkIsolation": boolean,
"Environment": {
  "string": "string"
},
"ExperimentConfig": {
  "ExperimentName": "string",
  "RunName": "string",
  "TrialComponentDisplayName": "string",
  "TrialName": "string"
},
"FailureReason": "string",
"FinalMetricDataList": [
  {
    "MetricName": "string",
    "Timestamp": number,
    "Value": number
  }
],
"HyperParameters": {
  "string": "string"
},
"InputDataConfig": [
  {
    "ChannelName": "string",
    "CompressionType": "string",
    "CompressionType": "string"
  }
]
"ContentType": "string",
"DataSource": {
  "FileSystemDataSource": {
    "DirectoryPath": "string",
    "FileSystemAccessMode": "string",
    "FileSystemId": "string",
    "FileSystemType": "string"
  },
  "S3DataSource": {
    "AttributeNames": [ "string" ],
    "InstanceGroupNames": [ "string" ],
    "S3DistributionType": "string",
    "S3DataType": "string",
    "S3Uri": "string"
  }
},
"InputMode": "string",
"RecordWrapperType": "string",
"ShuffleConfig": {
  "Seed": number
}
],
"LabelingJobArn": "string",
"LastModifiedTime": number,
"ModelArtifacts": {
  "S3ModelArtifacts": "string"
},
"OutputDataConfig": {
  "CompressionType": "string",
  "KmsKeyId": "string",
  "S3OutputPath": "string"
},
"ProfilerConfig": {
  "DisableProfiler": boolean,
  "ProfilingIntervalInMilliseconds": number,
  "ProfilingParameters": {
    "string": "string"
  },
  "S3OutputPath": "string"
},
"ResourceConfig": {
  "InstanceCount": number,
  "InstanceGroups": [
    {
      "InstanceCount": number,
      "InstanceGroupName": "string",
      "InstanceType": "string"
    }
  ],
  "InstanceType": "string",
  "KeepAlivePeriodInSeconds": number,
  "VolumeKmsKeyId": "string",
  "VolumeSizeInGB": number
},
"RetryStrategy": {
  "MaximumRetryAttempts": number
},
"RoleArn": "string",
"SecondaryStatus": "string",
"SecondaryStatusTransitions": [
  {
    "EndTime": number,
    "StartTime": number,
    "Status": "string",
    "StatusMessage": "string"
  }
]
"StoppingCondition": {
  "MaxPendingTimeInSeconds": number,
  "MaxRuntimeInSeconds": number,
  "MaxWaitTimeInSeconds": number
},
"Tags": [
  {
    "Key": "string",
    "Value": "string"
  }
],
"TensorBoardOutputConfig": {
  "LocalPath": "string",
  "S3OutputPath": "string"
},
"TrainingEndTime": number,
"TrainingJobArn": "string",
"TrainingJobName": "string",
"TrainingJobStatus": "string",
"TrainingStartTime": number,
"TrainingTimeInSeconds": number,
"TuningJobArn": "string",
"VpcConfig": {
  "SecurityGroupIds": [ "string" ],
  "Subnets": [ "string" ]
},
"TransformJob": {
  "AutoMLJobArn": "string",
  "BatchStrategy": "string",
  "CreationTime": number,
  "DataCaptureConfig": {
    "DestinationS3Uri": "string",
    "GenerateInferenceId": boolean,
    "KmsKeyId": "string"
  },
  "DataProcessing": {
    "InputFilter": "string",
    "JoinSource": "string",
    "OutputFilter": "string"
  },
  "Environment": {
    "string": "string"
  },
  "ExperimentConfig": {
    "ExperimentName": "string",
    "RunName": "string",
    "TrialComponentDisplayName": "string",
    "TrialName": "string"
  },
  "FailureReason": "string",
  "LabelingJobArn": "string",
  "MaxConcurrentTransforms": number,
  "MaxPayloadInMB": number,
  "ModelClientConfig": {
    "InvocationsMaxRetries": number,
    "InvocationsTimeoutInSeconds": number
  },
  "ModelName": "string",
  "Tags": [
    {
      "Key": "string",
      "Value": "string"
    }
  ]
}
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**NextToken (p. 903)**

If the result of the previous `Search` request was truncated, the response includes a NextToken. To retrieve the next set of results, use the token in the next request.

Type: String

Length Constraints: Maximum length of 8192.

Pattern: .*
Results (p. 903)

A list of SearchRecord objects.

Type: Array of SearchRecord (p. 1938) objects

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
SendPipelineExecutionStepFailure

Service: Amazon SageMaker Service

Notifies the pipeline that the execution of a callback step failed, along with a message describing why. When a callback step is run, the pipeline generates a callback token and includes the token in a message sent to Amazon Simple Queue Service (Amazon SQS).

Request Syntax

```
{
   "CallbackToken": "string",
   "ClientRequestToken": "string",
   "FailureReason": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CallbackToken (p. 940)**

The pipeline generated token from the Amazon SQS queue.

Type: String

Length Constraints: Fixed length of 10.

Pattern: ^[a-zA-Z0-9]+$

Required: Yes

**ClientRequestToken (p. 940)**

A unique, case-sensitive identifier that you provide to ensure the idempotency of the operation. An idempotent operation completes no more than one time.

Type: String


Required: No

**FailureReason (p. 940)**

A message describing why the step failed.

Type: String

Length Constraints: Maximum length of 256.

Required: No

Response Syntax

```
{
}
```
"PipelineExecutionArn": "string"
}

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

### PipelineExecutionArn (p. 940)

The Amazon Resource Name (ARN) of the pipeline execution.

Type: String

Length Constraints: Maximum length of 256.

Pattern: ^arn:aws[a-z-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:pipeline\/[a-z0-9-]+\/*/execution\/[a-z0-9-]+$

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceLimitExceeded**

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
SendPipelineExecutionStepSuccess

Service: Amazon SageMaker Service

Notifies the pipeline that the execution of a callback step succeeded and provides a list of the step's output parameters. When a callback step is run, the pipeline generates a callback token and includes the token in a message sent to Amazon Simple Queue Service (Amazon SQS).

Request Syntax

```json
{
    "CallbackToken": "string",
    "ClientRequestToken": "string",
    "OutputParameters": [
        {
            "Name": "string",
            "Value": "string"
        }
    ]
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CallbackToken (p. 942)**

The pipeline generated token from the Amazon SQS queue.

Type: String

Length Constraints: Fixed length of 10.

Pattern: ^[a-zA-Z0-9-]+$  

Required: Yes

**ClientRequestToken (p. 942)**

A unique, case-sensitive identifier that you provide to ensure the idempotency of the operation. An idempotent operation completes no more than one time.

Type: String


Required: No

**OutputParameters (p. 942)**

A list of the output parameters of the callback step.

Type: Array of OutputParameter (p. 1794) objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No
Response Syntax

```json
{
   "PipelineExecutionArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**PipelineExecutionArn (p. 943)**

The Amazon Resource Name (ARN) of the pipeline execution.

Type: String

Length Constraints: Maximum length of 256.

Pattern: `^arn:aws[a-z-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:pipeline\./.*\/execution\./.*$`

Errors

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](CommonErrors).

**ResourceLimitExceeded**

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](AWSCLI)
- [AWS SDK for .NET](AWSSDKNET)
- [AWS SDK for C++](AWSSDKCPP)
- [AWS SDK for Go](AWSSDKGO)
- [AWS SDK for Java V2](AWSSDKJAVA)
- [AWS SDK for JavaScript V3](AWSSDKJAVASCRIPT)
- [AWS SDK for PHP V3](AWSSDKPHP)
- [AWS SDK for Python](AWSSDKPYTHON)
- [AWS SDK for Ruby V3](AWSSDKRUBY)
StartEdgeDeploymentStage
Service: Amazon SageMaker Service
Starts a stage in an edge deployment plan.

Request Syntax

```
{
    "EdgeDeploymentPlanName": "string",
    "StageName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**EdgeDeploymentPlanName** (p. 944)

The name of the edge deployment plan to start.

Type: String


Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$`

Required: Yes

**StageName** (p. 944)

The name of the stage to start.

Type: String


Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$`

Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
• AWS SDK for .NET
• AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
StartInferenceExperiment
Service: Amazon SageMaker Service
Starts an inference experiment.

Request Syntax

```json
{
   "Name": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

Name (p. 946)

The name of the inference experiment to start.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,119}$

Required: Yes

Response Syntax

```json
{
   "InferenceExperimentArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

InferenceExperimentArn (p. 946)

The ARN of the started inference experiment to start.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:inference-experiment/.*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).
ConflictException

There was a conflict when you attempted to modify a SageMaker entity such as an Experiment or Artifact.

HTTP Status Code: 400

ResourceNotFoundException

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
StartMonitoringSchedule

Service: Amazon SageMaker Service

Starts a previously stopped monitoring schedule.

**Note**

By default, when you successfully create a new schedule, the status of a monitoring schedule is scheduled.

**Request Syntax**

```json
{
  "MonitoringScheduleName": "string"
}
```

**Request Parameters**

For information about the parameters that are common to all actions, see [Common Parameters](p. 2178).

The request accepts the following data in JSON format.

**MonitoringScheduleName (p. 948)**

The name of the schedule to start.

Type: String


Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62\}$`

Required: Yes

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)].

**ResourceNotFoundException**

Resource being access is not found.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
StartNotebookInstance
Service: Amazon SageMaker Service

Launches an ML compute instance with the latest version of the libraries and attaches your ML storage volume. After configuring the notebook instance, SageMaker sets the notebook instance status to InService. A notebook instance's status must be InService before you can connect to your Jupyter notebook.

Request Syntax

```
{
    "NotebookInstanceName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**NotebookInstanceName** (p. 950)

The name of the notebook instance to start.

Type: String

Length Constraints: Maximum length of 63.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9])*$`

Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceLimitExceeded**

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
• AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
StartPipelineExecution

Service: Amazon SageMaker Service

Starts a pipeline execution.

Request Syntax

```json
{
    "ClientRequestToken": "string",
    "ParallelismConfiguration": {
        "MaxParallelExecutionSteps": number
    },
    "PipelineExecutionDescription": "string",
    "PipelineExecutionDisplayName": "string",
    "PipelineName": "string",
    "PipelineParameters": [ {
        "Name": "string",
        "Value": "string"
    } ],
    "SelectiveExecutionConfig": {
        "SelectedSteps": [
            { "StepName": "string"
        }
    ],
    "SourcePipelineExecutionArn": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**ClientRequestToken (p. 952)**

A unique, case-sensitive identifier that you provide to ensure the idempotency of the operation. An idempotent operation completes no more than once.

- Type: String
- Required: Yes

**ParallelismConfiguration (p. 952)**

This configuration, if specified, overrides the parallelism configuration of the parent pipeline for this specific run.

- Type: ParallelismConfiguration (p. 1797) object
- Required: No

**PipelineExecutionDescription (p. 952)**

The description of the pipeline execution.
Type: String
Length Constraints: Minimum length of 0. Maximum length of 3072.
Pattern: .*
Required: No

**PipelineExecutionDisplayName (p. 952)**

The display name of the pipeline execution.

Type: String
Length Constraints: Minimum length of 1. Maximum length of 82.
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,81}
Required: No

**PipelineName (p. 952)**

The name or Amazon Resource Name (ARN) of the pipeline.

Type: String
Length Constraints: Minimum length of 1. Maximum length of 2048.
Pattern: (arn:aws[\-a-z]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:pipeline/.*)?([a-zA-Z0-9](-*[a-zA-Z0-9])\{0,255\})
Required: Yes

**PipelineParameters (p. 952)**

Contains a list of pipeline parameters. This list can be empty.

Type: Array of [Parameter (p. 1798)] objects
Array Members: Minimum number of 0 items. Maximum number of 50 items.
Required: No

**SelectiveExecutionConfig (p. 952)**

The selective execution configuration applied to the pipeline run.

Type: [SelectiveExecutionConfig (p. 1944)] object
Required: No

**Response Syntax**

```json
{
  "PipelineExecutionArn": "string"
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.
PipelineExecutionArn (p. 953)

The Amazon Resource Name (ARN) of the pipeline execution.

Type: String

Length Constraints: Maximum length of 256.

Pattern: ^arn:aws[a-z-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:pipeline\./.*\./execution\./.*$

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceLimitExceeded

You have exceeded a SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

ResourceNotFound

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
StopAutoMLJob
Service: Amazon SageMaker Service

A method for forcing a running job to shut down.

Request Syntax

```json
{
    "AutoMLJobName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**AutoMLJobName (p. 955)**

The name of the object you are requesting.

Type: String


Pattern: ^[a-zA-Z0-9-]*[a-zA-Z0-9]{0,31}$

Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
StopCompilationJob

Service: Amazon SageMaker Service

Stops a model compilation job.

To stop a job, Amazon SageMaker sends the algorithm the SIGTERM signal. This gracefully shuts the job down. If the job hasn't stopped, it sends the SIGKILL signal.

When it receives a StopCompilationJob request, Amazon SageMaker changes the CompilationJobStatus of the job to Stopping. After Amazon SageMaker stops the job, it sets the CompilationJobStatus to Stopped.

Request Syntax

```
{
  "CompilationJobName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CompilationJobName (p. 957)**

The name of the model compilation job to stop.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:
• **AWS Command Line Interface**
• **AWS SDK for .NET**
• **AWS SDK for C++**
• **AWS SDK for Go**
• **AWS SDK for Java V2**
• **AWS SDK for JavaScript V3**
• **AWS SDK for PHP V3**
• **AWS SDK for Python**
• **AWS SDK for Ruby V3**
StopEdgeDeploymentStage

Service: Amazon SageMaker Service

Stops a stage in an edge deployment plan.

Request Syntax

```json
{
  "EdgeDeploymentPlanName": "string",
  "StageName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**EdgeDeploymentPlanName** (p. 959)

The name of the edge deployment plan to stop.

Type: String


Pattern: ^[a-zA-Z0-9-]*[a-zA-Z0-9]{0,62}$

Required: Yes

**StageName** (p. 959)

The name of the stage to stop.

Type: String


Pattern: ^[a-zA-Z0-9-]*[a-zA-Z0-9]{0,62}$

Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- **AWS SDK for .NET**
- **AWS SDK for C++**
- **AWS SDK for Go**
- **AWS SDK for Java V2**
- **AWS SDK for JavaScript V3**
- **AWS SDK for PHP V3**
- **AWS SDK for Python**
- **AWS SDK for Ruby V3**
StopEdgePackagingJob

Service: Amazon SageMaker Service

Request to stop an edge packaging job.

Request Syntax

```json
{  "EdgePackagingJobName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**EdgePackagingJobName (p. 961)**

The name of the edge packaging job.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
StopHyperParameterTuningJob

Service: Amazon SageMaker Service

Stops a running hyperparameter tuning job and all running training jobs that the tuning job launched.

All model artifacts output from the training jobs are stored in Amazon Simple Storage Service (Amazon S3). All data that the training jobs write to Amazon CloudWatch Logs are still available in CloudWatch. After the tuning job moves to the Stopped state, it releases all reserved resources for the tuning job.

Request Syntax

```json
{
    "HyperParameterTuningJobName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**HyperParameterTuningJobName (p. 962)**

  The name of the tuning job to stop.

  Type: String


  Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,31}$

  Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceNotFoundException**

  Resource being access is not found.

  HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- **AWS SDK for C++**
- **AWS SDK for Go**
- **AWS SDK for Java V2**
- **AWS SDK for JavaScript V3**
- **AWS SDK for PHP V3**
- **AWS SDK for Python**
- **AWS SDK for Ruby V3**
StopInferenceExperiment

Service: Amazon SageMaker Service

Stops an inference experiment.

Request Syntax

```json
{
   "DesiredModelVariants": [
      {
         "InfrastructureConfig": {
            "InfrastructureType": "string",
            "RealTimeInferenceConfig": {
               "InstanceCount": number,
               "InstanceType": "string"
            }
         },
         "ModelName": "string",
         "VariantName": "string"
      },
      "DesiredState": "string",
      "ModelVariantActions": {
         "string": "string"
      },
      "Name": "string",
      "Reason": "string"
   }
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**DesiredModelVariants (p. 964)**

An array of ModelVariantConfig objects. There is one for each variant that you want to deploy after the inference experiment stops. Each ModelVariantConfig describes the infrastructure configuration for deploying the corresponding variant.

Type: Array of ModelVariantConfig (p. 1728) objects

Array Members: Minimum number of 1 item. Maximum number of 2 items.

Required: No

**DesiredState (p. 964)**

The desired state of the experiment after stopping. The possible states are the following:

- Completed: The experiment completed successfully
- Cancelled: The experiment was canceled

Type: String

Valid Values: Completed | Cancelled

Required: No
**ModelVariantActions (p. 964)**

Array of key-value pairs, with names of variants mapped to actions. The possible actions are the following:
- **Promote** - Promote the shadow variant to a production variant
- **Remove** - Delete the variant
- **Retain** - Keep the variant as it is

Type: String to string map

Map Entries: Maximum number of 2 items.

Key Length Constraints: Maximum length of 63.

Key Pattern: `^[a-zA-Z0-9](\-[a-zA-Z0-9]{0,119})?$`

Valid Values: Retain | Remove | Promote

Required: Yes

**Name (p. 964)**

The name of the inference experiment to stop.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,119}\}0,119`

Required: Yes

**Reason (p. 964)**

The reason for stopping the experiment.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: `\.*`

Required: No

**Response Syntax**

```
{
   "InferenceExperimentArn": "string"
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**InferenceExperimentArn (p. 965)**

The ARN of the stopped inference experiment.
Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:inference-experiment/.*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ConflictException

There was a conflict when you attempted to modify a SageMaker entity such as an Experiment or Artifact.

HTTP Status Code: 400

ResourceNotFoundException

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
StopInferenceRecommendationsJob
Service: Amazon SageMaker Service

Stops an Inference Recommender job.

Request Syntax

```json
{
  "JobName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**JobName (p. 967)**

The name of the job you want to stop.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 64.

Pattern: ^[a-zA-Z0-9](\-*[a-zA-Z0-9])\{0, 63\}

Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceNotFoundException

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
StopLabelingJob
Service: Amazon SageMaker Service

Stops a running labeling job. A job that is stopped cannot be restarted. Any results obtained before the job is stopped are placed in the Amazon S3 output bucket.

Request Syntax

```json
{
    "LabelingJobName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**LabelingJobName** *(p. 969)*

The name of the labeling job to stop.

Type: String


Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$`

Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceNotFoundException**

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
StopLabelingJob

- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
StopMonitoringSchedule

Service: Amazon SageMaker Service

Stops a previously started monitoring schedule.

Request Syntax

```
{
   "MonitoringScheduleName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**MonitoringScheduleName (p. 971)**

The name of the schedule to stop.

Type: String


Pattern: ^[a-zA-Z0-9](\-?[a-zA-Z0-9]){0,62}$

Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceNotFoundException**

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
StopNotebookInstance

Service: Amazon SageMaker Service

Terminates the ML compute instance. Before terminating the instance, SageMaker disconnects the ML storage volume from it. SageMaker preserves the ML storage volume. SageMaker stops charging you for the ML compute instance when you call StopNotebookInstance.

To access data on the ML storage volume for a notebook instance that has been terminated, call the StartNotebookInstance API. StartNotebookInstance launches another ML compute instance, configures it, and attaches the preserved ML storage volume so you can continue your work.

Request Syntax

```
{
    "NotebookInstanceName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**NotebookInstanceName (p. 973)**

The name of the notebook instance to terminate.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])*  

Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
StopPipelineExecution

Service: Amazon SageMaker Service

Stops a pipeline execution.

Callback Step

A pipeline execution won't stop while a callback step is running. When you call StopPipelineExecution on a pipeline execution with a running callback step, SageMaker Pipelines sends an additional Amazon SQS message to the specified SQS queue. The body of the SQS message contains a "Status" field which is set to "Stopping".

You should add logic to your Amazon SQS message consumer to take any needed action (for example, resource cleanup) upon receipt of the message followed by a call to SendPipelineExecutionStepSuccess or SendPipelineExecutionStepFailure. Only when SageMaker Pipelines receives one of these calls will it stop the pipeline execution.

Lambda Step

A pipeline execution can't be stopped while a lambda step is running because the Lambda function invoked by the lambda step can't be stopped. If you attempt to stop the execution while the Lambda function is running, the pipeline waits for the Lambda function to finish or until the timeout is hit, whichever occurs first, and then stops. If the Lambda function finishes, the pipeline execution status is Stopped. If the timeout is hit the pipeline execution status is Failed.

Request Syntax

{
   "ClientRequestToken": "string",
   "PipelineExecutionArn": "string"
}

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

ClientRequestToken (p. 975)

A unique, case-sensitive identifier that you provide to ensure the idempotency of the operation. An idempotent operation completes no more than once.

Type: String


Required: Yes

PipelineExecutionArn (p. 975)

The Amazon Resource Name (ARN) of the pipeline execution.

Type: String

Length Constraints: Maximum length of 256.
Pattern: ^arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:pipeline\./.*\execution\./.*$

Required: Yes

Response Syntax

```
{
   "PipelineExecutionArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**PipelineExecutionArn** *(p. 976)*

The Amazon Resource Name (ARN) of the pipeline execution.

Type: String

Length Constraints: Maximum length of 256.

Pattern: ^arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:pipeline\./.*\execution\./.*$

Errors

For information about the errors that are common to all actions, see `Common Errors (p. 2180)`.

**ResourceNotFoundException**

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
StopProcessingJob
Service: Amazon SageMaker Service
Stops a processing job.

Request Syntax

```json
{
   "ProcessingJobName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**ProcessingJobName (p. 977)**

The name of the processing job to stop.

Type: String


Pattern: ^[a-zA-Z0-9]-*[a-zA-Z0-9]{0,62}

Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
StopTrainingJob
Service: Amazon SageMaker Service

Stops a training job. To stop a job, SageMaker sends the algorithm the SIGTERM signal, which delays job termination for 120 seconds. Algorithms might use this 120-second window to save the model artifacts, so the results of the training is not lost.

When it receives a StopTrainingJob request, SageMaker changes the status of the job to Stopping. After SageMaker stops the job, it sets the status to Stopped.

Request Syntax

```json
{
   "TrainingJobName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**TrainingJobName (p. 979)**

The name of the training job to stop.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}

Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
StopTransformJob

Service: Amazon SageMaker Service

Stops a batch transform job.

When Amazon SageMaker receives a StopTransformJob request, the status of the job changes to Stopping. After Amazon SageMaker stops the job, the status is set to Stopped. When you stop a batch transform job before it is completed, Amazon SageMaker doesn't store the job's output in Amazon S3.

Request Syntax

{  "TransformJobName": "string"}

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

TransformJobName (p. 981)

The name of the batch transform job to stop.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}$

Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceNotFound

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
<table>
<thead>
<tr>
<th>Library</th>
</tr>
</thead>
<tbody>
<tr>
<td>AWS SDK for C++</td>
</tr>
<tr>
<td>AWS SDK for Go</td>
</tr>
<tr>
<td>AWS SDK for Java V2</td>
</tr>
<tr>
<td>AWS SDK for JavaScript V3</td>
</tr>
<tr>
<td>AWS SDK for PHP V3</td>
</tr>
<tr>
<td>AWS SDK for Python</td>
</tr>
<tr>
<td>AWS SDK for Ruby V3</td>
</tr>
</tbody>
</table>
UpdateAction
Service: Amazon SageMaker Service
Updates an action.

Request Syntax

```json
{
   "ActionName": "string",
   "Description": "string",
   "Properties": {
      "string": "string"
   },
   "PropertiesToRemove": [ "string" ],
   "Status": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**ActionName (p. 983)**

The name of the action to update.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,119}$

Required: Yes

**Description (p. 983)**

The new description for the action.

Type: String

Length Constraints: Maximum length of 3072.

Pattern: .*

Required: No

**Properties (p. 983)**

The new list of properties. Overwrites the current property list.

Type: String to string map

Map Entries: Maximum number of 30 items.

Key Length Constraints: Maximum length of 256.

Key Pattern: .*

Value Length Constraints: Maximum length of 256.
**Value Pattern:** .*

Required: No  

**PropertiesToRemove (p. 983)**

A list of properties to remove.

Type: Array of strings

Length Constraints: Maximum length of 256.

Pattern: .*

Required: No  

**Status (p. 983)**

The new status for the action.

Type: String

Valid Values: Unknown | InProgress | Completed | Failed | Stopping | Stopped

Required: No

**Response Syntax**

```
{
   "ActionArn": "string"
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.  
The following data is returned in JSON format by the service.

**ActionArn (p. 984)**

The Amazon Resource Name (ARN) of the action.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:action/.*

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](https://docs.aws.amazon.com/sagemaker/latest/dg/API_CommonErrors.html).

**ConflictException**

There was a conflict when you attempted to modify a SageMaker entity such as an Experiment or Artifact.

HTTP Status Code: 400
ResourceNotFound

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
UpdateAppImageConfig
Service: Amazon SageMaker Service

Updates the properties of an AppImageConfig.

Request Syntax

```
{
    "AppImageConfigName": "string",
    "JupyterLabAppImageConfig": {
        "ContainerConfig": {
            "ContainerArguments": [ "string" ],
            "ContainerEntrypoint": [ "string" ],
            "ContainerEnvironmentVariables": {
                "string": "string"
            }
        }
    },
    "KernelGatewayImageConfig": {
        "FileSystemConfig": {
            "DefaultGid": number,
            "DefaultUid": number,
            "MountPath": "string"
        },
        "KernelSpecs": [
            {
                "DisplayName": "string",
                "Name": "string"
            }
        ]
    }
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**AppImageConfigName (p. 986)**

The name of the AppImageConfig to update.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62\}

Required: Yes

**JupyterLabAppImageConfig (p. 986)**

The JupyterLab app running on the image.

Type: JupyterLabAppImageConfig (p. 1615) object

Required: No

**KernelGatewayImageConfig (p. 986)**

The new KernelGateway app to run on the image.
Type: KernelGatewayImageConfig (p. 1621) object

Required: No

Response Syntax

```json
{
  "AppImageConfigArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**AppImageConfigArn** *(p. 987)*

The Amazon Resource Name (ARN) for the AppImageConfig.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:app-image-config/.*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
UpdateArtifact
Service: Amazon SageMaker Service
Updates an artifact.

Request Syntax

```
{
    "ArtifactArn": "string",
    "ArtifactName": "string",
    "Properties": {
        "string": "string"
    },
    "PropertiesToRemove": [ "string" ]
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

ArtifactArn (p. 988)
The Amazon Resource Name (ARN) of the artifact to update.
Type: String
Length Constraints: Maximum length of 256.
Pattern: arn:aws[a-z\-]*:sagemaker:[a-zA-Z0-9\-]*:\d{12}:artifact/.*
Required: Yes

ArtifactName (p. 988)
The new name for the artifact.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 120.
Pattern: ^[a-zA-Z0-9-](\-*[a-zA-Z0-9]\{0,119\})
Required: No

Properties (p. 988)
The new list of properties. Overwrites the current property list.
Type: String to string map
Map Entries: Maximum number of 30 items.
Key Length Constraints: Maximum length of 256.
Key Pattern: .*
Value Length Constraints: Maximum length of 256.
Value Pattern: .*
Required: No

**PropertiesToRemove (p. 988)**
A list of properties to remove.
Type: Array of strings
Length Constraints: Maximum length of 256.
Pattern: .*
Required: No

**Response Syntax**

```
{
    "ArtifactArn": "string"
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.
The following data is returned in JSON format by the service.

**ArtifactArn (p. 989)**
The Amazon Resource Name (ARN) of the artifact.
Type: String
Length Constraints: Maximum length of 256.
Pattern: arn:aws[a-z-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:artifact/.*

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)].

**ConflictException**
There was a conflict when you attempted to modify a SageMaker entity such as an Experiment or Artifact.
HTTP Status Code: 400

**ResourceNotFoundException**
Resource being access is not found.
HTTP Status Code: 400

**See Also**
For more information about using this API in one of the language-specific AWS SDKs, see the following:
- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
UpdateCluster
Service: Amazon SageMaker Service
Update a SageMaker HyperPod cluster.

Request Syntax

```
{
  "ClusterName": "string",
  "InstanceGroups": [
    {
      "ExecutionRole": "string",
      "InstanceCount": number,
      "InstanceGroupName": "string",
      "InstanceType": "string",
      "LifecycleConfig": {
        "OnCreate": "string",
        "SourceS3Uri": "string"
      },
      "ThreadsPerCore": number
    }
  ]
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**ClusterName (p. 991)**

Specify the name of the SageMaker HyperPod cluster you want to update.

Type: String

Length Constraints: Maximum length of 256.

Pattern: `^(arn:aws[a-zA-Z-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:cluster/[a-z0-9]{12})|([a-zA-Z0-9]+-[-*][a-zA-Z0-9]{12}){0,62}$`

Required: Yes

**InstanceGroups (p. 991)**

Specify the instance groups to update.

Type: Array of ClusterInstanceGroupSpecification (p. 1343) objects

Array Members: Minimum number of 1 item. Maximum number of 5 items.

Required: Yes

Response Syntax

```
{
  "ClusterArn": "string"
}
```
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

ClusterArn (p. 991)

The Amazon Resource Name (ARN) of the updated SageMaker HyperPod cluster.

Type: String

Length Constraints: Maximum length of 256.

Pattern: ^arn:aws[a-z-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:cluster/[a-z0-9]{12}$

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ConflictException

There was a conflict when you attempted to modify a SageMaker entity such as an Experiment or Artifact.

HTTP Status Code: 400

ResourceLimitExceeded

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

ResourceNotFoundException

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
UpdateCodeRepository
Service: Amazon SageMaker Service
Updates the specified Git repository with the specified values.

Request Syntax

```json
{
    "CodeRepositoryName": "string",
    "GitConfig": {
        "SecretArn": "string"
    }
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**CodeRepositoryName (p. 994)**

The name of the Git repository to update.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

Required: Yes

**GitConfig (p. 994)**

The configuration of the git repository, including the URL and the Amazon Resource Name (ARN) of the AWS Secrets Manager secret that contains the credentials used to access the repository. The secret must have a staging label of AWSCURRENT and must be in the following format:

{"username": Username, "password": Password}

Type: GitConfigForUpdate (p. 1506) object

Required: No

Response Syntax

```json
{
    "CodeRepositoryArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.
**CodeRepositoryArn (p. 994)**

The ARN of the Git repository.

**Type:** String

Length Constraints: Minimum length of 1. Maximum length of 2048.

**Pattern:** ^arn:aws(-cn|-us-gov)?:sagemaker:[a-z0-9\-]{9,16}:\[0-9]{12}:code-repository/[^\$]{1,2048}$

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](p. 2180).

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](aws-cli)
- [AWS SDK for .NET](aws-sdk-net)
- [AWS SDK for C++](aws-sdk-cpp)
- [AWS SDK for Go](aws-sdk-go)
- [AWS SDK for Java V2](aws-sdk-java)
- [AWS SDK for JavaScript V3](aws-sdk-js)
- [AWS SDK for PHP V3](aws-sdk-php)
- [AWS SDK for Python](aws-sdk-python)
- [AWS SDK for Ruby V3](aws-sdk-ruby)

995
**UpdateContext**

Service: Amazon SageMaker Service

Updates a context.

**Request Syntax**

```json
{
    "ContextName": "string",
    "Description": "string",
    "Properties": {
        "string": "string"
    },
    "PropertiesToRemove": [ "string" ]
}
```

**Request Parameters**

For information about the parameters that are common to all actions, see [Common Parameters](p. 2178).

The request accepts the following data in JSON format.

**ContextName (p. 996)**

The name of the context to update.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,119}$

Required: Yes

**Description (p. 996)**

The new description for the context.

Type: String

Length Constraints: Maximum length of 3072.

Pattern: .*

Required: No

**Properties (p. 996)**

The new list of properties. Overwrites the current property list.

Type: String to string map

Map Entries: Maximum number of 30 items.

Key Length Constraints: Maximum length of 256.

Key Pattern: .*

Value Length Constraints: Maximum length of 256.
UpdateContext

Value Pattern: .*

Required: No

**PropertiesToRemove (p. 996)**

A list of properties to remove.

Type: Array of strings

Length Constraints: Maximum length of 256.

Pattern: .*

Required: No

**Response Syntax**

```
{
  "ContextArn": "string"
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**ContextArn (p. 997)**

The Amazon Resource Name (ARN) of the context.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:context/.*

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](#).

**ConflictException**

There was a conflict when you attempted to modify a SageMaker entity such as an Experiment or Artifact.

HTTP Status Code: 400

**ResourceNotFoundException**

Resource being access is not found.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:
• AWS Command Line Interface
• AWS SDK for .NET
• AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
UpdateDeviceFleet

Service: Amazon SageMaker Service

Updates a fleet of devices.

Request Syntax

```
{
  "Description": "string",
  "DeviceFleetName": "string",
  "EnableIotRoleAlias": boolean,
  "OutputConfig": {
    "KmsKeyId": "string",
    "PresetDeploymentConfig": "string",
    "PresetDeploymentType": "string",
    "S3OutputLocation": "string"
  },
  "RoleArn": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see [Common Parameters](p. 2178).

The request accepts the following data in JSON format.

**Description (p. 999)**

Description of the fleet.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 800.

Pattern: ^[-a-zA-Z0-9_.,;:! ]*$

Required: No

**DeviceFleetName (p. 999)**

The name of the fleet.

Type: String


Pattern: ^[a-zA-Z0-9\-\_\.,;\:\! ]{0,62}$

Required: Yes

**EnableIotRoleAlias (p. 999)**

Whether to create an AWS IoT Role Alias during device fleet creation. The name of the role alias generated will match this pattern: "SageMakerEdge-{DeviceFleetName}".

For example, if your device fleet is called "demo-fleet", the name of the role alias will be "SageMakerEdge-demo-fleet".

Type: Boolean
**OutputConfig (p. 999)**

Output configuration for storing sample data collected by the fleet.

Type: [EdgeOutputConfig (p. 1444)](#) object

**RoleArn (p. 999)**

The Amazon Resource Name (ARN) of the device.

Type: String


Pattern: `arn:aws[a-z\-]*:iam::\d{12}:role/?[a-zA-Z_0-9+=,.@-_/]+$`

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](#).

**ResourceInUse**

Resource being accessed is in use.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
UpdateDevices
Service: Amazon SageMaker Service
Updates one or more devices in a fleet.

Request Syntax

```
{
  "DeviceFleetName": "string",
  "Devices": [
    {
      "Description": "string",
      "DeviceName": "string",
      "IotThingName": "string"
    }
  ]
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**DeviceFleetName (p. 1001)**

The name of the fleet the devices belong to.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]{0,62})$  

Required: Yes

**Devices (p. 1001)**

List of devices to register with Edge Manager agent.

Type: Array of Device (p. 1411) objects

Required: Yes

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:
- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
**UpdateDomain**

**Service:** Amazon SageMaker Service

Updates the default settings for new user profiles in the domain.

**Request Syntax**

```json
{
    "AppNetworkAccessType": "string",
    "AppSecurityGroupManagement": "string",
    "DefaultSpaceSettings": {
        "ExecutionRole": "string",
        "JupyterServerAppSettings": {
            "CodeRepositories": [
                { "RepositoryUrl": "string"
            ]
        },
        "DefaultResourceSpec": {
            "InstanceType": "string",
            "LifecycleConfigArn": "string",
            "SageMakerImageArn": "string",
            "SageMakerImageVersionAlias": "string",
            "SageMakerImageVersionArn": "string"
        },
        "LifecycleConfigArns": [ "string" ]
    },
    "KernelGatewayAppSettings": {
        "CustomImages": [
            { "AppImageConfigName": "string",
            "ImageName": "string",
            "ImageVersionNumber": number
        ]
    },
    "DefaultResourceSpec": {
        "InstanceType": "string",
        "LifecycleConfigArn": "string",
        "SageMakerImageArn": "string",
        "SageMakerImageVersionAlias": "string",
        "SageMakerImageVersionArn": "string"
    },
    "LifecycleConfigArns": [ "string" ]
},
"DefaultUserSettings": {"string"}
},
"CanvasAppSettings": {
    "DirectDeploySettings": {  "Status": "string"
},
"IdentityProviderOAuthSettings": [  
    { "DataSourceName": "string",
    "SecretArn": "string",
    "Status": "string"
    }
],
"KendraSettings": {  "Status": "string"
},
"ModelRegisterSettings": {  "CrossAccountModelRegisterRoleArn": "string",
..."string"
```
"Status": "string",
"TimeSeriesForecastingSettings": {
  "AmazonForecastRoleArn": "string",
  "Status": "string"
},
"WorkspaceSettings": {
  "S3ArtifactPath": "string",
  "$S3KmsKeyId": "string"
}
],
"CodeEditorAppSettings": {
  "DefaultResourceSpec": {
    "InstanceType": "string",
    "LifecycleConfigArn": "string",
    "SageMakerImageArn": "string",
    "SageMakerImageVersionAlias": "string",
    "SageMakerImageVersionArn": "string"
  },
  "LifecycleConfigArns": [ "string" ]
},
"CustomFileSystemConfigs": [
  { ... }
],
"CustomPosixUserConfig": {
  "Gid": number,
  "Uid": number
},
"DefaultLandingUri": "string",
"ExecutionRole": "string",
"JupyterLabAppSettings": {
  "CodeRepositories": [
    { "RepositoryUrl": "string" }
  ],
  "CustomImages": [
    { "AppImageConfigName": "string",
     "ImageName": "string",
     "ImageVersionNumber": number
    }
  ],
  "DefaultResourceSpec": {
    "InstanceType": "string",
    "LifecycleConfigArn": "string",
    "SageMakerImageArn": "string",
    "SageMakerImageVersionAlias": "string",
    "SageMakerImageVersionArn": "string"
  },
  "LifecycleConfigArns": [ "string" ]
},
"JupyterServerAppSettings": {
  "CodeRepositories": [
    { "RepositoryUrl": "string" }
  ],
  "DefaultResourceSpec": {
    "InstanceType": "string",
    "LifecycleConfigArn": "string",
    "SageMakerImageArn": "string",
    "SageMakerImageVersionAlias": "string",
    "SageMakerImageVersionArn": "string"
  },
  "LifecycleConfigArns": [ "string" ]
}
"KernelGatewayAppSettings": {
  "CustomImages": [
    {
      "AppImageConfigName": "string",
      "ImageName": "string",
      "ImageVersionNumber": number
    }
  ],
  "DefaultResourceSpec": {
    "InstanceType": "string",
    "LifecycleConfigArn": "string",
    "SageMakerImageArn": "string",
    "SageMakerImageVersionAlias": "string",
    "SageMakerImageVersionArn": "string"
  },
  "LifecycleConfigArns": [ "string" ]
},
"RSessionAppSettings": {
  "CustomImages": [
    {
      "AppImageConfigName": "string",
      "ImageName": "string",
      "ImageVersionNumber": number
    }
  ],
  "DefaultResourceSpec": {
    "InstanceType": "string",
    "LifecycleConfigArn": "string",
    "SageMakerImageArn": "string",
    "SageMakerImageVersionAlias": "string",
    "SageMakerImageVersionArn": "string"
  }
},
"RStudioServerProAppSettings": {
  "AccessStatus": "string",
  "UserGroup": "string"
},
"SecurityGroups": [ "string" ],
"SharingSettings": {
  "NotebookOutputOption": "string",
  "S3KmsKeyId": "string",
  "S3OutputPath": "string"
},
"SpaceStorageSettings": {
  "DefaultEbsStorageSettings": {
    "DefaultEbsVolumeSizeInGb": number,
    "MaximumEbsVolumeSizeInGb": number
  }
},
"StudioWebPortal": "string",
"TensorBoardAppSettings": {
  "DefaultResourceSpec": {
    "InstanceType": "string",
    "LifecycleConfigArn": "string",
    "SageMakerImageArn": "string",
    "SageMakerImageVersionAlias": "string",
    "SageMakerImageVersionArn": "string"
  }
},
"DomainId": "string",
"DomainSettingsForUpdate": {
  "ExecutionRoleIdentityConfig": "string",
  "RStudioServerProDomainSettingsForUpdate": {
    "DefaultResourceSpec": {
      "InstanceType": "string",
      "LifecycleConfigArn": "string",
      "SageMakerImageArn": "string",
      "SageMakerImageVersionAlias": "string",
      "SageMakerImageVersionArn": "string"
    }
  }
}
"InstanceType": "string",
"LifecycleConfigArn": "string",
"SageMakerImageArn": "string",
"SageMakerImageVersionAlias": "string",
"SageMakerImageVersionArn": "string"
},
"DomainExecutionRoleArn": "string",
"RStudioConnectUrl": "string",
"RStudioPackageManagerUrl": "string"
},
"SecurityGroupIds": [ "string" ]
},
"SubnetIds": [ "string" ]
}

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**AppNetworkAccessType (p. 1003)**

Specifies the VPC used for non-EFS traffic.
- **PublicInternetOnly** - Non-EFS traffic is through a VPC managed by Amazon SageMaker, which allows direct internet access.
- **VpcOnly** - All Studio traffic is through the specified VPC and subnets.

This configuration can only be modified if there are no apps in the InService, Pending, or Deleting state. The configuration cannot be updated if DomainSettings.RStudioServerProDomainSettings.DomainExecutionRoleArn is already set or DomainSettings.RStudioServerProDomainSettings.DomainExecutionRoleArn is provided as part of the same request.

Type: String

Valid Values: PublicInternetOnly | VpcOnly

Required: No

**AppSecurityGroupManagement (p. 1003)**

The entity that creates and manages the required security groups for inter-app communication in VpcOnly mode. Required when CreateDomain.AppNetworkAccessType is VpcOnly and DomainSettings.RStudioServerProDomainSettings.DomainExecutionRoleArn is provided. If setting up the domain for use with RStudio, this value must be set to Service.

Type: String

Valid Values: Service | Customer

Required: No

**DefaultSpaceSettings (p. 1003)**

The default settings used to create a space within the Domain.

Type: DefaultSpaceSettings (p. 1402) object

Required: No
**DefaultUserSettings (p. 1003)**

A collection of settings.

Type: UserSettings (p. 2070) object

Required: No

**DomainId (p. 1003)**

The ID of the domain to be updated.

Type: String

Length Constraints: Maximum length of 63.

Required: Yes

**DomainSettingsForUpdate (p. 1003)**

A collection of DomainSettings configuration values to update.

Type: DomainSettingsForUpdate (p. 1425) object

Required: No

**SubnetIds (p. 1003)**

The VPC subnets that Studio uses for communication.

If removing subnets, ensure there are no apps in the InService, Pending, or Deleting state.

Type: Array of strings

Array Members: Minimum number of 1 item. Maximum number of 16 items.

Length Constraints: Maximum length of 32.

Pattern: [-0-9a-zA-Z]+

Required: No

### Response Syntax

```json
{
  "DomainArn": "string"
}
```

### Response Elements

If the action is successful, the service sends back an HTTP 200 response. The following data is returned in JSON format by the service.

**DomainArn (p. 1007)**

The Amazon Resource Name (ARN) of the domain.

Type: String

Length Constraints: Maximum length of 256.
Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceInUse

Resource being accessed is in use.

HTTP Status Code: 400

ResourceLimitExceeded

You have exceeded a SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

ResourceNotFound

Resource being accessed is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
UpdateEndpoint
Service: Amazon SageMaker Service

Deploys the new EndpointConfig specified in the request, switches to using newly created endpoint, and then deletes resources provisioned for the endpoint using the previous EndpointConfig (there is no availability loss).

When SageMaker receives the request, it sets the endpoint status to Updating. After updating the endpoint, it sets the status to InService. To check the status of an endpoint, use the DescribeEndpoint API.

**Note**
You must not delete an EndpointConfig in use by an endpoint that is live or while the UpdateEndpoint or CreateEndpoint operations are being performed on the endpoint. To update an endpoint, you must create a new EndpointConfig. If you delete the EndpointConfig of an endpoint that is active or being created or updated you may lose visibility into the instance type the endpoint is using. The endpoint must be deleted in order to stop incurring charges.

**Request Syntax**

```json
{
  "DeploymentConfig": {
    "AutoRollbackConfiguration": {
      "Alarms": [
        {
          "AlarmName": "string"
        }
      ],
      "BlueGreenUpdatePolicy": {
        "MaximumExecutionTimeoutIn Seconds": number,
        "TerminationWaitInSeconds": number,
        "TrafficRoutingConfiguration": {
          "CanarySize": {
            "Type": "string",
            "Value": number
          },
          "LinearStepSize": {
            "Type": "string",
            "Value": number
          },
          "Type": "string",
          "WaitIntervalInSeconds": number
        }
      },
      "RollingUpdatePolicy": {
        "MaximumBatchSize": {
          "Type": "string",
          "Value": number
        },
        "MaximumExecutionTimeoutIn Seconds": number,
        "RollbackMaximumBatchSize": {
          "Type": "string",
          "Value": number
        },
        "WaitIntervalInSeconds": number
      }
    },
    "EndpointConfigName": "string",
    "EndpointName": "string",
    "ExcludeRetainedVariantProperties": [
  
```
UpdateEndpoint

```json
{
   "VariantPropertyType": "string",
}
,
"RetainAllVariantProperties": boolean,
"RetainDeploymentConfig": boolean
}
```

**Request Parameters**

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**DeploymentConfig (p. 1009)**

The deployment configuration for an endpoint, which contains the desired deployment strategy and rollback configurations.

Type: DeploymentConfig (p. 1405) object

Required: No

**EndpointConfigName (p. 1009)**

The name of the new endpoint configuration.

Type: String

Length Constraints: Maximum length of 63.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}`

Required: Yes

**EndpointName (p. 1009)**

The name of the endpoint whose configuration you want to update.

Type: String

Length Constraints: Maximum length of 63.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}`

Required: Yes

**ExcludeRetainedVariantProperties (p. 1009)**

When you are updating endpoint resources with RetainAllVariantProperties, whose value is set to true, ExcludeRetainedVariantProperties specifies the list of type VariantProperty to override with the values provided by EndpointConfig. If you don't specify a value for ExcludeRetainedVariantProperties, no variant properties are overridden.

Type: Array of VariantProperty (p. 2073) objects

Array Members: Minimum number of 0 items. Maximum number of 3 items.

Required: No

**RetainAllVariantProperties (p. 1009)**

When updating endpoint resources, enables or disables the retention of variant properties, such as the instance count or the variant weight. To retain the variant properties of an endpoint when
updating it, set `RetainAllVariantProperties` to `true`. To use the variant properties specified in a new `EndpointConfig` call when updating an endpoint, set `RetainAllVariantProperties` to `false`. The default is `false`.

Type: Boolean
Required: No

**RetainDeploymentConfig (p. 1009)**

Specifies whether to reuse the last deployment configuration. The default value is `false` (the configuration is not reused).

Type: Boolean
Required: No

### Response Syntax

```json
{
  "EndpointArn": "string"
}
```

### Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**EndpointArn (p. 1011)**

The Amazon Resource Name (ARN) of the endpoint.

Type: String


Pattern: `arn:aws[a-z-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:endpoint/.*`

### Errors

For information about the errors that are common to all actions, see [Common Errors (p. 2180)].

**ResourceLimitExceeded**

You have exceeded a SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

### See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
• AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
UpdateEndpointWeightsAndCapacities

Service: Amazon SageMaker Service

Updates variant weight of one or more variants associated with an existing endpoint, or capacity of one variant associated with an existing endpoint. When it receives the request, SageMaker sets the endpoint status to Updating. After updating the endpoint, it sets the status to InService. To check the status of an endpoint, use the DescribeEndpoint API.

Request Syntax

```json
{
    "DesiredWeightsAndCapacities": [
        {
            "DesiredInstanceCount": number,
            "DesiredWeight": number,
            "ServerlessUpdateConfig": {
                "MaxConcurrency": number,
                "ProvisionedConcurrency": number
            },
            "VariantName": "string"
        }
    ],
    "EndpointName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**DesiredWeightsAndCapacities (p. 1013)**

An object that provides new capacity and weight values for a variant.

Type: Array of DesiredWeightAndCapacity (p. 1410) objects

Array Members: Minimum number of 1 item.

Required: Yes

**EndpointName (p. 1013)**

The name of an existing SageMaker endpoint.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

Required: Yes

Response Syntax

```json
{
    "EndpointArn": "string"
}
```
Response Elements

If the action is successful, the service sends back an HTTP 200 response. The following data is returned in JSON format by the service.

**EndpointArn (p. 1013)**

The Amazon Resource Name (ARN) of the updated endpoint.

- Type: String
- Pattern: `arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:\[0-9\]{12}:endpoint/.*`

Errors

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](#).

**ResourceLimitExceeded**

You have exceeded a SageMaker resource limit. For example, you might have too many training jobs created.

- HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
UpdateExperiment
Service: Amazon SageMaker Service

Adds, updates, or removes the description of an experiment. Updates the display name of an experiment.

Request Syntax

```json
{
    "Description": "string",
    "DisplayName": "string",
    "ExperimentName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see [Common Parameters](p. 2178).

The request accepts the following data in JSON format.

**Description (p. 1015)**

The description of the experiment.

- Type: String
- Length Constraints: Maximum length of 3072.
- Pattern: .*
- Required: No

**DisplayName (p. 1015)**

The name of the experiment as displayed. The name doesn't need to be unique. If DisplayName isn't specified, ExperimentName is displayed.

- Type: String
- Length Constraints: Minimum length of 1. Maximum length of 120.
- Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,119}
- Required: No

**ExperimentName (p. 1015)**

The name of the experiment to update.

- Type: String
- Length Constraints: Minimum length of 1. Maximum length of 120.
- Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,119}
- Required: Yes

Response Syntax

```json
{
}
```
# UpdateExperiment

"ExperimentArn": "string"
}

## Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**ExperimentArn (p. 1015)**

The Amazon Resource Name (ARN) of the experiment.

- **Type:** String
- **Length Constraints:** Maximum length of 256.
- **Pattern:** arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:experiment/.*

## Errors

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](https://docs.aws.amazon.com/AmazonSageMaker/latest/APIReference/AmazonSageMaker.html).

**ConflictException**

There was a conflict when you attempted to modify a SageMaker entity such as an Experiment or Artifact.

- **HTTP Status Code:** 400

**ResourceNotFoundException**

Resource being access is not found.

- **HTTP Status Code:** 400

## See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](https://docs.aws.amazon.com/cli/index.html)
- [AWS SDK for .NET](https://aws.amazon.com/sdk-for-net/)
- [AWS SDK for C++](https://aws.amazon.com/sdk-for-cpp/)
- [AWS SDK for Go](https://aws.amazon.com/sdk-for-golang/)
- [AWS SDK for Java V2](https://aws.amazon.com/sdk-for-java/)
- [AWS SDK for JavaScript V3](https://aws.amazon.com/sdk-for-javascript/)
- [AWS SDK for PHP V3](https://aws.amazon.com/sdk-for-php/)
- [AWS SDK for Python](https://aws.amazon.com/sdk-for-python/)
- [AWS SDK for Ruby V3](https://aws.amazon.com/sdk-for-ruby/)

---

1016
UpdateFeatureGroup

Service: Amazon SageMaker Service

Updates the feature group by either adding features or updating the online store configuration. Use one of the following request parameters at a time while using the UpdateFeatureGroup API.

You can add features for your feature group using the FeatureAdditions request parameter. Features cannot be removed from a feature group.

You can update the online store configuration by using the OnlineStoreConfig request parameter. If a TtlDuration is specified, the default TtlDuration applies for all records added to the feature group after the feature group is updated. If a record level TtlDuration exists from using the PutRecord API, the record level TtlDuration applies to that record instead of the default TtlDuration.

Request Syntax

```json
{
  "FeatureAdditions": [
    {
      "CollectionConfig": { ... },
      "CollectionType": "string",
      "FeatureName": "string",
      "FeatureType": "string"
    }
  ],
  "FeatureGroupName": "string",
  "OnlineStoreConfig": {
    "TtlDuration": {
      "Unit": "string",
      "Value": number
    }
  }
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**FeatureAdditions (p. 1017)**

Updates the feature group. Updating a feature group is an asynchronous operation. When you get an HTTP 200 response, you've made a valid request. It takes some time after you've made a valid request for Feature Store to update the feature group.

Type: Array of FeatureDefinition (p. 1481) objects

Array Members: Minimum number of 1 item. Maximum number of 100 items.

Required: No

**FeatureGroupName (p. 1017)**

The name or Amazon Resource Name (ARN) of the feature group that you're updating.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 256.
UpdateFeatureGroup

**Pattern:** (arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:feature-group/)?(\[a-zA-Z0-9](\[_-]*[a-zA-Z0-9]){0,63})

Required: Yes

**OnlineStoreConfig (p. 1017)**

Updates the feature group online store configuration.

Type: **OnlineStoreConfigUpdate (p. 1786)** object

Required: No

**Response Syntax**

```
{
   "FeatureGroupArn": "string"
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**FeatureGroupArn (p. 1018)**

The Amazon Resource Number (ARN) of the feature group that you're updating.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:feature-group/.*

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](Common_Errors). **ResourceNotFoundException**

Resource being access is not found.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](AWS_Command_Line_Interface)
- [AWS SDK for .NET](AWS_SDK_for_DOT_NET)
- [AWS SDK for C++](AWS_SDK_for_C++)
- [AWS SDK for Go](AWS_SDK_for_Go)
- [AWS SDK for Java V2](AWS_SDK_for_Java_V2)
- [AWS SDK for JavaScript V3](AWS_SDK_for_JavaScript_V3)
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
UpdateFeatureMetadata
Service: Amazon SageMaker Service

Updates the description and parameters of the feature group.

Request Syntax

```json
{
    "Description": "string",
    "FeatureGroupName": "string",
    "FeatureName": "string",
    "ParameterAdditions": [
        {
            "Key": "string",
            "Value": "string"
        }
    ],
    "ParameterRemovals": [ "string" ]
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**Description (p. 1020)**

A description that you can write to better describe the feature.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 255.

Pattern: .*

Required: No

**FeatureGroupName (p. 1020)**

The name or Amazon Resource Name (ARN) of the feature group containing the feature that you're updating.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 256.

Pattern: (arn:aws[\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:feature-group\//)? ([a-zA-Z0-9\-]*\[\-]*[a-zA-Z0-9\-]\{0,63\})

Required: Yes

**FeatureName (p. 1020)**

The name of the feature that you're updating.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 64.
Pattern: ^[a-zA-Z0-9](\[-\]*[a-zA-Z0-9])\{0,63}\n
Required: Yes

**ParameterAdditions (p. 1020)**

A list of key-value pairs that you can add to better describe the feature.

Type: Array of FeatureParameter (p. 1491) objects

Array Members: Maximum number of 25 items.

Required: No

**ParameterRemovals (p. 1020)**

A list of parameter keys that you can specify to remove parameters that describe your feature.

Type: Array of strings

Array Members: Maximum number of 25 items.

Length Constraints: Minimum length of 1. Maximum length of 255.

Pattern: ^([\p{L}\p{Z}\p{N}\p{P}_.:/=+\-]*)$

Required: No

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

**Errors**

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
UpdateHub

Service: Amazon SageMaker Service

Update a hub.

**Note**

Hub APIs are only callable through SageMaker Studio.

**Request Syntax**

```json
{
    "HubDescription": "string",
    "HubDisplayName": "string",
    "HubName": "string",
    "HubSearchKeywords": [ "string" ]
}
```

**Request Parameters**

For information about the parameters that are common to all actions, see [Common Parameters](p. 2178).

The request accepts the following data in JSON format.

**HubDescription (p. 1022)**

A description of the updated hub.

Type: String

Length Constraints: Maximum length of 1023.

Pattern: .*

Required: No

**HubDisplayName (p. 1022)**

The display name of the hub.

Type: String

Length Constraints: Maximum length of 255.

Pattern: .*

Required: No

**HubName (p. 1022)**

The name of the hub to update.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9-]*[a-zA-Z0-9]\{0,62}$

Required: Yes
HubSearchKeywords (p. 1022)

The searchable keywords for the hub.

Type: Array of strings

Array Members: Maximum number of 50 items.

Length Constraints: Maximum length of 255.

Pattern: `^[^A-Z]*$`

Required: No

Response Syntax

```
{
  "HubArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

HubArn (p. 1023)

The Amazon Resource Name (ARN) of the updated hub.

Type: String

Length Constraints: Maximum length of 255.

Pattern: `.*`

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceNotFound

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
UpdateImage

Service: Amazon SageMaker Service

Updates the properties of a SageMaker image. To change the image's tags, use the AddTags and DeleteTags APIs.

Request Syntax

```
{
    "DeleteProperties": [ "string" ],
    "Description": "string",
    "DisplayName": "string",
    "ImageName": "string",
    "RoleArn": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**DeleteProperties (p. 1025)**

A list of properties to delete. Only the Description and DisplayName properties can be deleted.

- Type: Array of strings
- Array Members: Maximum number of 2 items.
- Pattern: \(^\text{DisplayName}\)|\(^\text{Description}\)

- Required: No

**Description (p. 1025)**

The new description for the image.

- Type: String
- Pattern: .*

- Required: No

**DisplayName (p. 1025)**

The new display name for the image.

- Type: String
- Pattern: \(^\backslash S(.^{*}\backslash S)?\$

- Required: No
**ImageName (p. 1025)**

The name of the image to update.

Type: String


Pattern: ^[a-zA-Z0-9][-._]?[a-zA-Z0-9]{0,62}$

Required: Yes

**RoleArn (p. 1025)**

The new ARN for the IAM role that enables Amazon SageMaker to perform tasks on your behalf.

Type: String


Pattern: ^arn:aws[a-zA-Z\-]+:iam::\d{12}:role/\?[a-zA-Z0-9-]+,-@\-\//+$

Required: No

**Response Syntax**

```json
{
  "ImageArn": "string"
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**ImageArn (p. 1026)**

The ARN of the image.

Type: String

Length Constraints: Maximum length of 256.

Pattern: ^arn:aws(-[\w]+)*:sagemaker:.+:image/[a-zA-Z0-9][0-9][12]:image/[a-zA-Z0-9][-._]?[a-zA-Z0-9]+$ $

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](link).

**ResourceInUse**

Resource being accessed is in use.

HTTP Status Code: 400

**ResourceNotFound**

Resource being access is not found.
HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
UpdateImageVersion
Service: Amazon SageMaker Service

Updates the properties of a SageMaker image version.

Request Syntax

```
{
    "Alias": "string",
    "AliasesToAdd": [ "string" ],
    "AliasesToDelete": [ "string" ],
    "Horovod": boolean,
    "ImageName": "string",
    "JobType": "string",
    "MLFramework": "string",
    "Processor": "string",
    "ProgrammingLang": "string",
    "ReleaseNotes": "string",
    "VendorGuidance": "string",
    "Version": number
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

Alias (p. 1028)

The alias of the image version.

Type: String


Pattern: (?![^\[-\]])^([a-zA-Z0-9-_.]+)$

Required: No

AliasesToAdd (p. 1028)

A list of aliases to add.

Type: Array of strings


Pattern: (?![^\[-\]])^([a-zA-Z0-9-_.]+)$

Required: No

AliasesToDelete (p. 1028)

A list of aliases to delete.

Type: Array of strings


Pattern: (?![^\[-\]])^([a-zA-Z0-9-_.]+)$
Required: No

**Horovod (p. 1028)**

Indicates Horovod compatibility.

Type: Boolean

Required: No

**ImageName (p. 1028)**

The name of the image.

Type: String


Pattern: `^[a-zA-Z0-9][-._]?[a-zA-Z0-9]{0,62}$`

Required: Yes

**JobType (p. 1028)**

Indicates SageMaker job type compatibility.

- **TRAINING**: The image version is compatible with SageMaker training jobs.
- **INFERENCED**: The image version is compatible with SageMaker inference jobs.
- **NOTEBOOK_KERNEL**: The image version is compatible with SageMaker notebook kernels.

Type: String

Valid Values: TRAINING | INFERENCED | NOTEBOOK_KERNEL

Required: No

**MLFramework (p. 1028)**

The machine learning framework vended in the image version.

Type: String


Pattern: `^[a-zA-Z]+ \d+.(\d+)?$`

Required: No

**Processor (p. 1028)**

Indicates CPU or GPU compatibility.

- **CPU**: The image version is compatible with CPU.
- **GPU**: The image version is compatible with GPU.

Type: String

Valid Values: CPU | GPU

Required: No

**ProgrammingLang (p. 1028)**

The supported programming language and its version.

Type: String

Pattern: ^[a-zA-Z]+ \d+\.\d+(\.\d+)?$  
Required: No

**ReleaseNotes (p. 1028)**

The maintainer description of the image version.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 255.

Pattern: .*

Required: No

**VendorGuidance (p. 1028)**

The availability of the image version specified by the maintainer.

- **NOT_PROVIDED**: The maintainers did not provide a status for image version stability.
- **STABLE**: The image version is stable.
- **TO_BE_ARCHIVED**: The image version is set to be archived. Custom image versions that are set to be archived are automatically archived after three months.
- **ARCHIVED**: The image version is archived. Archived image versions are not searchable and are no longer actively supported.

Type: String

Valid Values: NOT_PROVIDED | STABLE | TO_BE_ARCHIVED | ARCHIVED

Required: No

**Version (p. 1028)**

The version of the image.

Type: Integer

Valid Range: Minimum value of 0.

Required: No

**Response Syntax**

```json
{
   "ImageVersionArn": "string"
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**ImageVersionArn (p. 1030)**

The ARN of the image version.
Type: String

Length Constraints: Maximum length of 256.

Pattern: ^arn:aws(-[\w]+)*:sagemaker:.+:[0-9]{12}:image-version/[a-z0-9][0-9]+$/

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceInUse

Resource being accessed is in use.

HTTP Status Code: 400

ResourceNotFound

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
UpdateInferenceComponent

Service: Amazon SageMaker Service

Updates an inference component.

Request Syntax

```json
{
  "InferenceComponentName": "string",
  "RuntimeConfig": {
    "CopyCount": number
  },
  "Specification": {
    "ComputeResourceRequirements": {
      "MaxMemoryRequiredInMb": number,
      "MinMemoryRequiredInMb": number,
      "NumberOfAcceleratorDevicesRequired": number,
      "NumberOfCpuCoresRequired": number
    },
    "Container": {
      "ArtifactUrl": "string",
      "Environment": {
        "string": "string"
      },
      "Image": "string"
    },
    "ModelName": "string",
    "StartupParameters": {
      "ContainerStartupHealthCheckTimeoutInSeconds": number,
      "ModelDataDownloadTimeoutInSeconds": number
    }
  }
}
```

Request Parameters

For information about the parameters that are common to all actions, see [Common Parameters](p. 2178).

The request accepts the following data in JSON format.

**InferenceComponentName (p. 1032)**

The name of the inference component.

Type: String

Length Constraints: Maximum length of 63.

Pattern: `^[a-zA-Z0-9-]*[a-zA-Z0-9]$`

Required: Yes

**RuntimeConfig (p. 1032)**

Runtime settings for a model that is deployed with an inference component.

Type: [InferenceComponentRuntimeConfig (p. 1582)] object

Required: No
**Specification (p. 1032)**

Details about the resources to deploy with this inference component, including the model, container, and compute resources.

Type: [InferenceComponentSpecification (p. 1584)]

Required: No

**Response Syntax**

```
{
    "InferenceComponentArn": "string"
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**InferenceComponentArn (p. 1033)**

The Amazon Resource Name (ARN) of the inference component.

Type: String


**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)].

**ResourceLimitExceeded**

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- [AWS SDK for .NET](https://aws.amazon.com/sdk-for-net/)
- [AWS SDK for C++](https://aws.amazon.com/sdk-for-cpp/)
- [AWS SDK for Go](https://aws.amazon.com/sdk-for-go/)
- [AWS SDK for Java V2](https://aws.amazon.com/sdk-for-java/)
- [AWS SDK for JavaScript V3](https://aws.amazon.com/sdk-for-node-js/)
- [AWS SDK for PHP V3](https://aws.amazon.com/sdk-for-php/)
- [AWS SDK for Python](https://aws.amazon.com/sdk-for-python/)
- [AWS SDK for Ruby V3](https://aws.amazon.com/sdk-for-ruby/)

1033
UpdateInferenceComponentRuntimeConfig

Service: Amazon SageMaker Service

Runtime settings for a model that is deployed with an inference component.

Request Syntax

```json
{
    "DesiredRuntimeConfig": {
        "CopyCount": number,
        "InferenceComponentName": "string"
    }
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**DesiredRuntimeConfig (p. 1035)**

Runtime settings for a model that is deployed with an inference component.

Type: InferenceComponentRuntimeConfig (p. 1582) object

Required: Yes

**InferenceComponentName (p. 1035)**

The name of the inference component to update.

Type: String

Length Constraints: Maximum length of 63.

Pattern: `^[a-zA-Z0-9](\-[a-zA-Z0-9]*[a-zA-Z0-9])?$`

Required: Yes

Response Syntax

```json
{
    "InferenceComponentArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**InferenceComponentArn (p. 1035)**

The Amazon Resource Name (ARN) of the inference component.
Type: String

Errors
For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceLimitExceeded
You have exceeded a SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
UpdateInferenceExperiment

Service: Amazon SageMaker Service

Updates an inference experiment that you created. The status of the inference experiment has to be either Created, Running. For more information on the status of an inference experiment, see DescribeInferenceExperiment.

Request Syntax

```
{
  "DataStorageConfig": {
    "ContentType": {
      "CsvContentTypes": [ "string" ],
      "JsonContentTypes": [ "string" ]
    },
    "Destination": "string",
    "KmsKey": "string"
  },
  "Description": "string",
  "ModelVariants": [
    {
      "InfrastructureConfig": {
        "InfrastructureType": "string",
        "RealTimeInferenceConfig": {
          "InstanceCount": number,
          "InstanceType": "string"
        }
      },
      "ModelName": "string",
      "VariantName": "string"
    }
  ],
  "Name": "string",
  "Schedule": {
    "EndTime": number,
    "StartTime": number
  },
  "ShadowModeConfig": {
    "ShadowModelVariants": [
      {
        "SamplingPercentage": number,
        "ShadowModelVariantName": "string"
      }
    ],
    "SourceModelVariantName": "string"
  }
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**DataStorageConfig (p. 1037)**

The Amazon S3 location and configuration for storing inference request and response data.

Type: InferenceExperimentDataStorageConfig (p. 1590) object
Required: No

**Description (p. 1037)**

The description of the inference experiment.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: .*

Required: No

**ModelVariants (p. 1037)**

An array of ModelVariantConfig objects. There is one for each variant, whose infrastructure configuration you want to update.

Type: Array of ModelVariantConfig (p. 1728) objects

Array Members: Minimum number of 1 item. Maximum number of 2 items.

Required: No

**Name (p. 1037)**

The name of the inference experiment to be updated.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,119}$

Required: Yes

**Schedule (p. 1037)**

The duration for which the inference experiment will run. If the status of the inference experiment is Created, then you can update both the start and end dates. If the status of the inference experiment is Running, then you can update only the end date.

Type: InferenceExperimentSchedule (p. 1591) object

Required: No

**ShadowModeConfig (p. 1037)**

The configuration of ShadowMode inference experiment type. Use this field to specify a production variant which takes all the inference requests, and a shadow variant to which Amazon SageMaker replicates a percentage of the inference requests. For the shadow variant also specify the percentage of requests that Amazon SageMaker replicates.

Type: ShadowModeConfig (p. 1950) object

Required: No

**Response Syntax**

```json
{
    "InferenceExperimentArn": "string"
}
```
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**InferenceExperimentArn (p. 1038)**

The ARN of the updated inference experiment.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:inference-experiment/.*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ConflictException**

There was a conflict when you attempted to modify a SageMaker entity such as an Experiment or Artifact.

HTTP Status Code: 400

**ResourceNotFoundException**

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
UpdateModelCard

Service: Amazon SageMaker Service

Update an Amazon SageMaker Model Card.

**Important**
You cannot update both model card content and model card status in a single call.

**Request Syntax**

```
{
    "Content": "string",
    "ModelCardName": "string",
    "ModelCardStatus": "string"
}
```

**Request Parameters**

For information about the parameters that are common to all actions, see [Common Parameters](p. 2178).

The request accepts the following data in JSON format.

**Content (p. 1040)**

The updated model card content. Content must be in model card JSON schema and provided as a string.

When updating model card content, be sure to include the full content and not just updated content.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 100000.

Pattern: .*

Required: No

**ModelCardName (p. 1040)**

The name or Amazon Resource Name (ARN) of the model card to update.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 256.

Pattern: (arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:model-card/.*)?([a-zA-Z0-9-]*([a-zA-Z0-9]{0,62}))

Required: Yes

**ModelCardStatus (p. 1040)**

The approval status of the model card within your organization. Different organizations might have different criteria for model card review and approval.

- Draft: The model card is a work in progress.
- PendingReview: The model card is pending review.
• Approved: The model card is approved.
• Archived: The model card is archived. No more updates should be made to the model card, but it can still be exported.

Type: String

Valid Values: Draft | PendingReview | Approved | Archived

Required: No

Response Syntax

```json
{
  "ModelCardArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**ModelCardArn (p. 1041)**

The Amazon Resource Name (ARN) of the updated model card.

Type: String

Length Constraints: Maximum length of 256.

Pattern: ^arn:aws[a-z-]*:sagemaker:[a-z0-9\-]{9,16}:[0-9]{12}:model-card/[a-zA-Z0-9\-\]*[a-zA-Z0-9]{0,62}$

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ConflictException**

There was a conflict when you attempted to modify a SageMaker entity such as an Experiment or Artifact.

HTTP Status Code: 400

**ResourceLimitExceeded**

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

**ResourceNotFoundException**

Resource being access is not found.

HTTP Status Code: 400
See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
UpdateModelPackage
Service: Amazon SageMaker Service
Updates a versioned model.

Request Syntax

```
{
    "AdditionalInferenceSpecificationsToAdd": [
        {
            "Containers": [
                {
                    "AdditionalS3DataSource": {
                        "CompressionType": "string",
                        "S3DataType": "string",
                        "S3Uri": "string"
                    },
                    "ContainerHostname": "string",
                    "Environment": {
                        "string": "string"
                    },
                    "Framework": "string",
                    "FrameworkVersion": "string",
                    "Image": "string",
                    "ImageDigest": "string",
                    "ModelDataUrl": "string",
                    "ModelInput": {
                        "DataInputConfig": "string"
                    },
                    "NearestModelName": "string",
                    "ProductId": "string"
                }
            ],
            "Description": "string",
            "Name": "string",
            "SupportedContentTypes": [ "string" ],
            "SupportedRealtimeInferenceInstanceTypes": [ "string" ],
            "SupportedResponseMIMETypes": [ "string" ],
            "SupportedTransformInstanceTypes": [ "string" ]
        }
    ],
    "ApprovalDescription": "string",
    "CustomerMetadataProperties": {
        "string": "string"
    },
    "CustomerMetadataPropertiesToRemove": [ "string" ],
    "ModelApprovalStatus": "string",
    "ModelPackageArn": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see [Common Parameters](p. 2178).

The request accepts the following data in JSON format.

**AdditionalInferenceSpecificationsToAdd (p. 1043)**

An array of additional Inference Specification objects to be added to the existing array additional Inference Specification. Total number of additional Inference Specifications can not exceed 15. Each
additional Inference Specification specifies artifacts based on this model package that can be used on inference endpoints. Generally used with SageMaker Neo to store the compiled artifacts.

Type: Array of AdditionalInferenceSpecificationDefinition (p. 1219) objects

Array Members: Minimum number of 1 item. Maximum number of 15 items.

Required: No

ApprovalDescription (p. 1043)

A description for the approval status of the model.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: .*

Required: No

CustomerMetadataProperties (p. 1043)

The metadata properties associated with the model package versions.

Type: String to string map

Map Entries: Maximum number of 50 items.

Key Length Constraints: Minimum length of 1. Maximum length of 128.

Key Pattern: ^([\p{L}\p{Z}\p{N}_.:\/=\*\-@]*){1,128}$

Value Length Constraints: Minimum length of 1. Maximum length of 256.

Value Pattern: ^([\p{L}\p{Z}\p{N}_.:\/=\*\-@]*){1,256}$

Required: No

CustomerMetadataPropertiesToRemove (p. 1043)

The metadata properties associated with the model package versions to remove.

Type: Array of strings


Pattern: ^([\p{L}\p{Z}\p{N}_.:\/=\*\-@]*){1,128}$

Required: No

ModelApprovalStatus (p. 1043)

The approval status of the model.

Type: String

Valid Values: Approved | Rejected | PendingManualApproval

Required: No

ModelPackageArn (p. 1043)

The Amazon Resource Name (ARN) of the model package.

Type: String
Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: ^arn:aws(-cn|-us-gov)?:sagemaker:[a-z0-9\-]{9,16}:[0-9]{12}:model-package/\S{1,2048}$

Required: Yes

**Response Syntax**

```json
{
  "ModelPackageArn": "string"
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**ModelPackageArn (p. 1045)**

  The Amazon Resource Name (ARN) of the model.

  Type: String

  Length Constraints: Minimum length of 1. Maximum length of 2048.

  Pattern: ^arn:aws(-cn|-us-gov)?:sagemaker:[a-z0-9\-]{9,16}:[0-9]{12}:model-package/\S{1,2048}$

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](#).

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
UpdateMonitoringAlert
Service: Amazon SageMaker Service
Update the parameters of a model monitor alert.

Request Syntax

```json
{
   "DatapointsToAlert": number,
   "EvaluationPeriod": number,
   "MonitoringAlertName": "string",
   "MonitoringScheduleName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**DatapointsToAlert (p. 1046)**

Within EvaluationPeriod, how many execution failures will raise an alert.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: Yes

**EvaluationPeriod (p. 1046)**

The number of most recent monitoring executions to consider when evaluating alert status.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: Yes

**MonitoringAlertName (p. 1046)**

The name of a monitoring alert.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

Required: Yes

**MonitoringScheduleName (p. 1046)**

The name of a monitoring schedule.

Type: String

UpdateMonitoringAlert

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}$

Required: Yes

Response Syntax

```
{
  "MonitoringAlertName": "string",
  "MonitoringScheduleArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**MonitoringAlertName (p. 1047)**

The name of a monitoring alert.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}$

**MonitoringScheduleArn (p. 1047)**

The Amazon Resource Name (ARN) of the monitoring schedule.

Type: String

Length Constraints: Maximum length of 256.

Pattern: .*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceLimitExceeded**

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:
- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
UpdateMonitoringSchedule

Service: Amazon SageMaker Service

Updates a previously created schedule.

Request Syntax

```json
{
  "MonitoringScheduleConfig": {
    "MonitoringJobDefinition": {
      "BaselineConfig": {
        "BaseliningJobName": "string",
        "ConstraintsResource": {
          "S3Uri": "string"
        },
        "StatisticsResource": {
          "S3Uri": "string"
        }
      },
      "Environment": {
        "string": "string"
      },
      "MonitoringAppSpecification": {
        "ContainerArguments": ["string"],
        "ContainerEntrypoint": ["string"],
        "ImageUri": "string",
        "PostAnalyticsProcessorSourceUri": "string",
        "RecordPreprocessorSourceUri": "string"
      },
      "MonitoringInputs": [
        {
          "BatchTransformInput": {
            "DataCapturedDestinationS3Uri": "string",
            "DatasetFormat": {
              "Csv": {
                "Header": boolean
              },
              "Json": {
                "Line": boolean
              },
              "Parquet": {
              }
            },
            "EndTimeOffset": "string",
            "ExcludeFeaturesAttribute": "string",
            "FeaturesAttribute": "string",
            "InferenceAttribute": "string",
            "LocalPath": "string",
            "ProbabilityAttribute": "string",
            "ProbabilityThresholdAttribute": number,
            "S3DataDistributionType": "string",
            "S3InputMode": "string",
            "StartTimeOffset": "string"
          },
          "EndpointInput": {
            "EndpointName": "string",
            "EndTimeOffset": "string",
            "ExcludeFeaturesAttribute": "string",
            "FeaturesAttribute": "string",
            "InferenceAttribute": "string",
            "LocalPath": "string",
            "ProbabilityAttribute": "string",
            "ProbabilityThresholdAttribute": number,
```
Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**MonitoringScheduleConfig (p. 1049)**

The configuration object that specifies the monitoring schedule and defines the monitoring job.

Type: MonitoringScheduleConfig (p. 1761) object

Required: Yes
MonitoringScheduleName (p. 1049)

The name of the monitoring schedule. The name must be unique within an AWS Region within an AWS account.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}\$

Required: Yes

Response Syntax

```
{
  "MonitoringScheduleArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

MonitoringScheduleArn (p. 1051)

The Amazon Resource Name (ARN) of the monitoring schedule.

Type: String

Length Constraints: Maximum length of 256.

Pattern: .*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceLimitExceeded

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

ResourceNotFound

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
UpdateNotebookInstance
Service: Amazon SageMaker Service

Updates a notebook instance. NotebookInstance updates include upgrading or downgrading the ML compute instance used for your notebook instance to accommodate changes in your workload requirements.

Request Syntax

```
{
"AcceleratorTypes": [ "string" ],
"AdditionalCodeRepositories": [ "string" ],
"DefaultCodeRepository": "string",
"DisassociateAcceleratorTypes": boolean,
"DisassociateAdditionalCodeRepositories": boolean,
"DisassociateDefaultCodeRepository": boolean,
"DisassociateLifecycleConfig": boolean,
"InstanceMetadataServiceConfiguration": {
  "MinimumInstanceMetadataServiceVersion": "string"
},
"InstanceType": "string",
"LifecycleConfigName": "string",
"NotebookInstanceName": "string",
"RoleArn": "string",
"RootAccess": "string",
"VolumeSizeInGB": number
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**AcceleratorTypes (p. 1053)**

A list of the Elastic Inference (EI) instance types to associate with this notebook instance. Currently only one EI instance type can be associated with a notebook instance. For more information, see Using Elastic Inference in Amazon SageMaker.

Type: Array of strings

Valid Values: ml.eia1.medium | ml.eia1.large | ml.eia1.xlarge | ml.eia2.medium | ml.eia2.large | ml.eia2.xlarge

Required: No

**AdditionalCodeRepositories (p. 1053)**

An array of up to three Git repositories to associate with the notebook instance. These can be either the names of Git repositories stored as resources in your account, or the URL of Git repositories in AWS CodeCommit or in any other Git repository. These repositories are cloned at the same level as the default repository of your notebook instance. For more information, see Associating Git Repositories with SageMaker Notebook Instances.

Type: Array of strings

Array Members: Maximum number of 3 items.
DefaultCodeRepository (p. 1053)

The Git repository to associate with the notebook instance as its default code repository. This can be either the name of a Git repository stored as a resource in your account, or the URL of a Git repository in AWS CodeCommit or in any other Git repository. When you open a notebook instance, it opens in the directory that contains this repository. For more information, see Associating Git Repositories with SageMaker Notebook Instances.

Type: String


Pattern: ^https://([^/]+)/(.*)$|^[a-zA-Z0-9]+\-[a-zA-Z0-9]*

Required: No

DisassociateAcceleratorTypes (p. 1053)

A list of the Elastic Inference (EI) instance types to remove from this notebook instance. This operation is idempotent. If you specify an accelerator type that is not associated with the notebook instance when you call this method, it does not throw an error.

Type: Boolean

Required: No

DisassociateAdditionalCodeRepositories (p. 1053)

A list of names or URLs of the default Git repositories to remove from this notebook instance. This operation is idempotent. If you specify a Git repository that is not associated with the notebook instance when you call this method, it does not throw an error.

Type: Boolean

Required: No

DisassociateDefaultCodeRepository (p. 1053)

The name or URL of the default Git repository to remove from this notebook instance. This operation is idempotent. If you specify a Git repository that is not associated with the notebook instance when you call this method, it does not throw an error.

Type: Boolean

Required: No

DisassociateLifecycleConfig (p. 1053)

Set to true to remove the notebook instance lifecycle configuration currently associated with the notebook instance. This operation is idempotent. If you specify a lifecycle configuration that is not associated with the notebook instance when you call this method, it does not throw an error.

Type: Boolean

Required: No

InstanceMetadataServiceConfiguration (p. 1053)

Information on the IMDS configuration of the notebook instance

Type: InstanceMetadataServiceConfiguration (p. 1611) object
UpdateNotebookInstance

Required: No

**InstanceType (p. 1053)**

The Amazon ML compute instance type.

Type: String

Valid Values: ml.t2.medium | ml.t2.large | ml.t2.xlarge | ml.t2.2xlarge | ml.t3.medium | ml.t3.large | ml.t3.xlarge | ml.t3.2xlarge | ml.m4.xlarge | ml.m4.2xlarge | ml.m4.4xlarge | ml.m4.10xlarge | ml.m4.16xlarge | ml.m5.xlarge | ml.m5.2xlarge | ml.m5.4xlarge | ml.m5.12xlarge | ml.m5.24xlarge | ml.m5d.large | ml.m5d.xlarge | ml.m5d.2xlarge | ml.m5d.4xlarge | ml.m5d.8xlarge | ml.m5d.12xlarge | ml.m5d.16xlarge | ml.m5d.24xlarge | ml.c4.large | ml.c4.xlarge | ml.c4.2xlarge | ml.c4.4xlarge | ml.c4.8xlarge | ml.c5.large | ml.c5.xlarge | ml.c5.2xlarge | ml.c5.4xlarge | ml.c5.9xlarge | ml.c5d.large | ml.c5d.xlarge | ml.c5d.2xlarge | ml.c5d.4xlarge | ml.c5d.9xlarge | ml.c5d.18xlarge | ml.p2.large | ml.p2.xlarge | ml.p3.large | ml.p3.xlarge | ml.p3.2xlarge | ml.p3.8xlarge | ml.p3.16xlarge | ml.p3dn.xlarge | ml.p4dlarge.xlarge | ml.p4xlarge | ml.p4.large | ml.p4xlarge | ml.p4dn.xlarge | ml.g4dn.large | ml.g4dn.xlarge | ml.g4d.large | ml.g4d.xlarge | ml.g4d.2xlarge | ml.g4d.4xlarge | ml.g4d.8xlarge | ml.g4d.12xlarge | ml.g4d.16xlarge | ml.r5.large | ml.r5.xlarge | ml.r5.2xlarge | ml.r5.4xlarge | ml.r5.8xlarge | ml.r5.12xlarge | ml.r5.16xlarge | ml.r5.24xlarge | ml.g5.large | ml.g5.xlarge | ml.g5.2xlarge | ml.g5.4xlarge | ml.g5.8xlarge | ml.g5.16xlarge | ml.g5.24xlarge | ml.g5.48xlarge | ml.inf1.large | ml.inf1.xlarge | ml.inf1.2xlarge | ml.inf1.6xlarge | ml.inf1.24xlarge | ml.p4.large | ml.p4xlarge | ml.p4d.xlarge | ml.p4de.xlarge

Required: No

**LifecycleConfigName (p. 1053)**

The name of a lifecycle configuration to associate with the notebook instance. For information about lifecycle configurations, see Step 2.1: (Optional) Customize a Notebook Instance.

Type: String

Length Constraints: Maximum length of 63.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9])*$`

Required: No

**NotebookInstanceName (p. 1053)**

The name of the notebook instance to update.

Type: String

Length Constraints: Maximum length of 63.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9])*$`

Required: Yes

**RoleArn (p. 1053)**

The Amazon Resource Name (ARN) of the IAM role that SageMaker can assume to access the notebook instance. For more information, see SageMaker Roles.

**Note**

To be able to pass this role to SageMaker, the caller of this API must have the `iam:PassRole` permission.
UpdateNotebookInstance

Type: String


Pattern: ^arn:aws[a-z-]+:iam::\d{12}:role/?[a-zA-Z_0-9+=,.@-_]/+$

Required: No

RootAccess (p. 1053)

Whether root access is enabled or disabled for users of the notebook instance. The default value is Enabled.

Note
If you set this to Disabled, users don't have root access on the notebook instance, but lifecycle configuration scripts still run with root permissions.

Type: String

Valid Values: Enabled | Disabled

Required: No

VolumeSizeInGB (p. 1053)

The size, in GB, of the ML storage volume to attach to the notebook instance. The default value is 5 GB. ML storage volumes are encrypted, so SageMaker can't determine the amount of available free space on the volume. Because of this, you can increase the volume size when you update a notebook instance, but you can't decrease the volume size. If you want to decrease the size of the ML storage volume in use, create a new notebook instance with the desired size.

Type: Integer


Required: No

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceLimitExceeded

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
UpdateNotebookInstanceLifecycleConfig

Service: Amazon SageMaker Service

Updates a notebook instance lifecycle configuration created with the CreateNotebookInstanceLifecycleConfig API.

Request Syntax

```json
{
  "NotebookInstanceLifecycleConfigName": "string",
  "OnCreate": [
    {
      "Content": "string"
    }
  ],
  "OnStart": [
    {
      "Content": "string"
    }
  ]
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**NotebookInstanceLifecycleConfigName (p. 1058)**

The name of the lifecycle configuration.

Type: String

Length Constraints: Maximum length of 63.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]*)*`

Required: Yes

**OnCreate (p. 1058)**

The shell script that runs only once, when you create a notebook instance. The shell script must be a base64-encoded string.

Type: Array of NotebookInstanceLifecycleHook (p. 1771) objects

Array Members: Maximum number of 1 item.

Required: No

**OnStart (p. 1058)**

The shell script that runs every time you start a notebook instance, including when you create the notebook instance. The shell script must be a base64-encoded string.

Type: Array of NotebookInstanceLifecycleHook (p. 1771) objects

Array Members: Maximum number of 1 item.
Required: No

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceLimitExceeded

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
UpdatePipeline
Service: Amazon SageMaker Service
Updates a pipeline.

Request Syntax

```json
{
   "ParallelismConfiguration": {
      "MaxParallelExecutionSteps": number
   },
   "PipelineDefinition": "string",
   "PipelineDefinitionS3Location": {
      "Bucket": "string",
      "ObjectKey": "string",
      "VersionId": "string"
   },
   "PipelineDescription": "string",
   "PipelineDisplayName": "string",
   "PipelineName": "string",
   "RoleArn": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

ParallelismConfiguration (p. 1060)

If specified, it applies to all executions of this pipeline by default.

Type: ParallelismConfiguration (p. 1797) object

Required: No

PipelineDefinition (p. 1060)

The JSON pipeline definition.

Type: String


Pattern: .*(?::\r\n\t).*

Required: No

PipelineDefinitionS3Location (p. 1060)

The location of the pipeline definition stored in Amazon S3. If specified, SageMaker will retrieve the pipeline definition from this location.

Type: PipelineDefinitionS3Location (p. 1814) object

Required: No

PipelineDescription (p. 1060)

The description of the pipeline.
**UpdatePipeline**

Type: String

Length Constraints: Minimum length of 0. Maximum length of 3072.

Pattern: .*

Required: No

**PipelineDisplayName (p. 1060)**

The display name of the pipeline.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 256.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,255}$

Required: No

**PipelineName (p. 1060)**

The name of the pipeline to update.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 256.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,255}$

Required: Yes

**RoleArn (p. 1060)**

The Amazon Resource Name (ARN) that the pipeline uses to execute.

Type: String


Pattern: ^arn:aws[a-z\-]*:iam::\d{12}:role/\?[a-zA-Z0-9=+,\@\-_]/+$

Required: No

**Response Syntax**

```
{
   "PipelineArn": "string"
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**PipelineArn (p. 1061)**

The Amazon Resource Name (ARN) of the updated pipeline.

Type: String
Length Constraints: Maximum length of 256.
Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]\{12\}:pipeline/.*

Errors
For information about the errors that are common to all actions, see Common Errors (p. 2180).

ResourceNotFound
Resource being access is not found.
HTTP Status Code: 400

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
UpdatePipelineExecution

Service: Amazon SageMaker Service

Updates a pipeline execution.

Request Syntax

```
{
  "ParallelismConfiguration": {
    "MaxParallelExecutionSteps": number
  },
  "PipelineExecutionArn": "string",
  "PipelineExecutionDescription": "string",
  "PipelineExecutionDisplayName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**ParallelismConfiguration (p. 1063)**

This configuration, if specified, overrides the parallelism configuration of the parent pipeline for this specific run.

Type: ParallelismConfiguration (p. 1797) object

Required: No

**PipelineExecutionArn (p. 1063)**

The Amazon Resource Name (ARN) of the pipeline execution.

Type: String

Length Constraints: Maximum length of 256.

Pattern: ^arn:aws[a-z-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:pipeline\/[a-z0-9-]*$/

Required: Yes

**PipelineExecutionDescription (p. 1063)**

The description of the pipeline execution.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 3072.

Pattern: .*

Required: No

**PipelineExecutionDisplayName (p. 1063)**

The display name of the pipeline execution.
Type: String

Length Constraints: Minimum length of 1. Maximum length of 82.

Pattern: ^[a-zA-Z0-9][-]*[a-zA-Z0-9]{0,81}$

Required: No

Response Syntax

```json
{
  "PipelineExecutionArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**PipelineExecutionArn (p. 1064)**

The Amazon Resource Name (ARN) of the updated pipeline execution.

Type: String

Length Constraints: Maximum length of 256.

Pattern: ^arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:pipeline\./.*\./execution\./.*$

Errors

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](#).

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
• AWS SDK for Ruby V3
UpdateProject
Service: Amazon SageMaker Service

Updates a machine learning (ML) project that is created from a template that sets up an ML pipeline from training to deploying an approved model.

Note
You must not update a project that is in use. If you update the ServiceCatalogProvisioningUpdateDetails of a project that is active or being created, or updated, you may lose resources already created by the project.

Request Syntax

```
{
  "ProjectDescription": "string",
  "ProjectName": "string",
  "ServiceCatalogProvisioningUpdateDetails": {
    "ProvisioningArtifactId": "string",
    "ProvisioningParameters": [
      {
        "Key": "string",
        "Value": "string"
      }
    ],
  },
  "Tags": [
    {
      "Key": "string",
      "Value": "string"
    }
  ]
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**ProjectDescription (p. 1066)**

The description for the project.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: \[\p{L}\p{M}\p{Z}\p{S}\p{N}\p{P}]*

Required: No

**ProjectName (p. 1066)**

The name of the project.

Type: String


Pattern: ^[a-zA-Z0-9-9](\-*[a-zA-Z0-9-9])\{0,31\}
UpdateProject

Required: Yes

**ServiceCatalogProvisioningUpdateDetails (p. 1066)**

The product ID and provisioning artifact ID to provision a service catalog. The provisioning artifact ID will default to the latest provisioning artifact ID of the product, if you don't provide the provisioning artifact ID. For more information, see [What is AWS Service Catalog](#).

Type: [ServiceCatalogProvisioningUpdateDetails (p. 1949)](#) object

Required: No

**Tags (p. 1066)**

An array of key-value pairs. You can use tags to categorize your AWS resources in different ways, for example, by purpose, owner, or environment. For more information, see [Tagging AWS Resources](#). In addition, the project must have tag update constraints set in order to include this parameter in the request. For more information, see [AWS Service Catalog Tag Update Constraints](#).

Type: Array of [Tag (p. 1979)](#) objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

**Response Syntax**

```json
{
   "ProjectArn": "string"
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**ProjectArn (p. 1067)**

The Amazon Resource Name (ARN) of the project.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: ^arn:aws(-cn|-us-gov)?:sagemaker:[a-z0-9-]{9,16}:[0-9]{12}:project/\S{1,2048}$_

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](#).

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
UpdateSpace
Service: Amazon SageMaker Service

Updates the settings of a space.

Request Syntax

```
{
    "DomainId": "string",
    "SpaceDisplayName": "string",
    "SpaceName": "string",
    "SpaceSettings": {
        "AppType": "string",
        "CodeEditorAppSettings": {
            "DefaultResourceSpec": {
                "InstanceType": "string",
                "LifecycleConfigArn": "string",
                "SageMakerImageArn": "string",
                "SageMakerImageVersionAlias": "string",
                "SageMakerImageVersionArn": "string"
            }
        },
        "CustomFileSystems": [
            ...
        ],
        "JupyterLabAppSettings": {
            "CodeRepositories": [
                {
                    "RepositoryUrl": "string"
                }
            ],
            "DefaultResourceSpec": {
                "InstanceType": "string",
                "LifecycleConfigArn": "string",
                "SageMakerImageArn": "string",
                "SageMakerImageVersionAlias": "string",
                "SageMakerImageVersionArn": "string"
            }
        },
        "JupyterServerAppSettings": {
            "CodeRepositories": [
                {
                    "RepositoryUrl": "string"
                }
            ],
            "DefaultResourceSpec": {
                "InstanceType": "string",
                "LifecycleConfigArn": "string",
                "SageMakerImageArn": "string",
                "SageMakerImageVersionAlias": "string",
                "SageMakerImageVersionArn": "string"
            }
        },
        "KernelGatewayAppSettings": {
            "CustomImages": [
                {
                    "AppImageConfigName": "string",
                    "ImageName": "string",
                    "ImageVersionNumber": number
                }
            ],
            "DefaultResourceSpec": {
```
Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**DomainId (p. 1069)**

The ID of the associated Domain.

Type: String

Length Constraints: Maximum length of 63.

Required: Yes

**SpaceDisplayName (p. 1069)**

The name of the space that appears in the Amazon SageMaker Studio UI.

Type: String

Length Constraints: Maximum length of 64.

Pattern: ^(?![\s*$]).+$

Required: No

**SpaceName (p. 1069)**

The name of the space.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9-]{0,62}$

Required: Yes

**SpaceSettings (p. 1069)**

A collection of space settings.

Type: SpaceSettings (p. 1961) object

Required: No
Response Syntax

```json
{
    "SpaceArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**SpaceArn (p. 1071)**

The space's Amazon Resource Name (ARN).

Type: String

Length Constraints: Maximum length of 256.

Pattern: `arn:aws[a-z-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:space/.*`

Errors

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](https://docs.aws.amazon.com/code/sagemaker-api/latest/ug/common-errors.html).

**ResourceInUse**

Resource being accessed is in use.

HTTP Status Code: 400

**ResourceLimitExceeded**

You have exceeded a SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

**ResourceNotFound**

Resource being accessed is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](https://docs.aws.amazon.com/code/sagemaker-cli/latest/userguide/sagemaker-cli.html)
- [AWS SDK for .NET](https://docs.aws.amazon.com/code/sagemaker-sdk-dotnet.html)
- [AWS SDK for C++](https://docs.aws.amazon.com/code/sagemaker-sdk-c.html)
- [AWS SDK for Go](https://docs.aws.amazon.com/code/sagemaker-sdk-go.html)
- [AWS SDK for Java V2](https://docs.aws.amazon.com/code/sagemaker-sdk-java.html)
- [AWS SDK for JavaScript V3](https://docs.aws.amazon.com/code/sagemaker-sdk-node.html)
- [AWS SDK for PHP V3](https://docs.aws.amazon.com/code/sagemaker-sdk-php.html)
• AWS SDK for Python
• AWS SDK for Ruby V3
UpdateTrainingJob

Service: Amazon SageMaker Service

Update a model training job to request a new Debugger profiling configuration or to change warm pool retention length.

Request Syntax

```
{
    "ProfilerConfig": {
        "DisableProfiler": boolean,
        "ProfilingIntervalInMilliseconds": number,
        "ProfilingParameters": {
            "string": "string"
        },
        "S3OutputPath": "string"
    },
    "ProfilerRuleConfigurations": [
        {
            "InstanceType": "string",
            "LocalPath": "string",
            "RuleConfigurationName": "string",
            "RuleEvaluatorImage": "string",
            "RuleParameters": {
                "string": "string"
            },
            "S3OutputPath": "string",
            "VolumeSizeInGB": number
        }
    ],
    "ResourceConfig": {
        "KeepAlivePeriodInSeconds": number
    },
    "TrainingJobName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**ProfilerConfig (p. 1073)**

Configuration information for Amazon SageMaker Debugger system monitoring, framework profiling, and storage paths.

Type: ProfilerConfigForUpdate (p. 1862) object

Required: No

**ProfilerRuleConfigurations (p. 1073)**

Configuration information for Amazon SageMaker Debugger rules for profiling system and framework metrics.

Type: Array of ProfilerRuleConfiguration (p. 1864) objects

Array Members: Minimum number of 0 items. Maximum number of 20 items.
Required: No

**ResourceConfig (p. 1073)**

The training job ResourceConfig to update warm pool retention length.

Type: **ResourceConfigForUpdate (p. 1913)** object

Required: No

**TrainingJobName (p. 1073)**

The name of a training job to update the Debugger profiling configuration.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}

Required: Yes

**Response Syntax**

```json
{
  "TrainingJobArn": "string"
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**TrainingJobArn (p. 1074)**

The Amazon Resource Name (ARN) of the training job.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:training-job/.*

**Errors**

For information about the errors that are common to all actions, see **Common Errors (p. 2180)**.

**ResourceNotFoundException**

Resource being accessed is not found.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:
- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
**UpdateTrial**

Service: Amazon SageMaker Service

Updates the display name of a trial.

**Request Syntax**

```
{
  "DisplayName": "string",
  "TrialName": "string"
}
```

**Request Parameters**

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**DisplayName (p. 1076)**

The name of the trial as displayed. The name doesn't need to be unique. If DisplayName isn't specified, TrialName is displayed.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,119}$

Required: No

**TrialName (p. 1076)**

The name of the trial to update.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,119}$

Required: Yes

**Response Syntax**

```
{
  "TrialArn": "string"
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.
TrialArn (p. 1076)

The Amazon Resource Name (ARN) of the trial.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:experiment-trial/.*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

ConflictException

There was a conflict when you attempted to modify a SageMaker entity such as an Experiment or Artifact.

HTTP Status Code: 400

ResourceNotFoundException

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
UpdateTrialComponent
Service: Amazon SageMaker Service

Updates one or more properties of a trial component.

Request Syntax

```
{
    "DisplayName": "string",
    "EndTime": number,
    "InputArtifacts": {
        "string": {
            "MediaType": "string",
            "Value": "string"
        }
    },
    "InputArtifactsToRemove": [ "string" ],
    "OutputArtifacts": {
        "string": {
            "MediaType": "string",
            "Value": "string"
        }
    },
    "OutputArtifactsToRemove": [ "string" ],
    "Parameters": {
        "string": {
            "NumberValue": number,
            "StringValue": "string"
        }
    },
    "ParametersToRemove": [ "string" ],
    "StartTime": number,
    "Status": {
        "Message": "string",
        "PrimaryStatus": "string"
    },
    "TrialComponentName": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**DisplayName (p. 1078)**

The name of the component as displayed. The name doesn't need to be unique. If DisplayName isn't specified, TrialComponentName is displayed.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,119}$

Required: No

**EndTime (p. 1078)**

When the component ended.
Type: Timestamp
Required: No

**InputArtifacts (p. 1078)**

Replaces all of the component's input artifacts with the specified artifacts or adds new input artifacts. Existing input artifacts are replaced if the trial component is updated with an identical input artifact key.

Type: String to [TrialComponentArtifact](#) object map

Map Entries: Maximum number of 30 items.

Key Length Constraints: Maximum length of 64.

Key Pattern: . *

Required: No

**InputArtifactsToRemove (p. 1078)**

The input artifacts to remove from the component.

Type: Array of strings

Length Constraints: Maximum length of 256.

Pattern: . *

Required: No

**OutputArtifacts (p. 1078)**

Replaces all of the component's output artifacts with the specified artifacts or adds new output artifacts. Existing output artifacts are replaced if the trial component is updated with an identical output artifact key.

Type: String to [TrialComponentArtifact](#) object map

Map Entries: Maximum number of 30 items.

Key Length Constraints: Maximum length of 64.

Key Pattern: . *

Required: No

**OutputArtifactsToRemove (p. 1078)**

The output artifacts to remove from the component.

Type: Array of strings

Length Constraints: Maximum length of 256.

Pattern: . *

Required: No

**Parameters (p. 1078)**

Replaces all of the component's hyperparameters with the specified hyperparameters or add new hyperparameters. Existing hyperparameters are replaced if the trial component is updated with an identical hyperparameter key.
Type: String to TrialComponentParameterValue (p. 2048) object map

Map Entries: Maximum number of 150 items.
Key Length Constraints: Maximum length of 256.
Key Pattern: \.*
Required: No

ParametersToRemove (p. 1078)

The hyperparameters to remove from the component.
Type: Array of strings
Length Constraints: Maximum length of 256.
Pattern: \.*
Required: No

StartTime (p. 1078)

When the component started.
Type: Timestamp
Required: No

Status (p. 1078)

The new status of the component.
Type: TrialComponentStatus (p. 2053) object
Required: No

TrialComponentName (p. 1078)

The name of the component to update.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 120.
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,119}$
Required: Yes

Response Syntax

```
{
  "TrialComponentArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.
**TrialComponentArn (p. 1080)**

The Amazon Resource Name (ARN) of the trial component.

Type: String

Length Constraints: Maximum length of 256.

Pattern: `arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:experiment-trial-component/.*`

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](https://docs.aws.amazon.com/auditreport/latest/userguide/).

**ConflictException**

There was a conflict when you attempted to modify a SageMaker entity such as an Experiment or Artifact.

HTTP Status Code: 400

**ResourceNotFoundException**

Resource being access is not found.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](https://docs.aws.amazon.com/auditreport/latest/userguide/)
- [AWS SDK for .NET](https://docs.aws.amazon.com/auditreport/latest/userguide/)
- [AWS SDK for C++](https://docs.aws.amazon.com/auditreport/latest/userguide/)
- [AWS SDK for Go](https://docs.aws.amazon.com/auditreport/latest/userguide/)
- [AWS SDK for Java V2](https://docs.aws.amazon.com/auditreport/latest/userguide/)
- [AWS SDK for JavaScript V3](https://docs.aws.amazon.com/auditreport/latest/userguide/)
- [AWS SDK for PHP V3](https://docs.aws.amazon.com/auditreport/latest/userguide/)
- [AWS SDK for Python](https://docs.aws.amazon.com/auditreport/latest/userguide/)
- [AWS SDK for Ruby V3](https://docs.aws.amazon.com/auditreport/latest/userguide/)
## UpdateUserProfile

Service: Amazon SageMaker Service

Updates a user profile.

### Request Syntax

```json
{
  "DomainId": "string",
  "UserProfileName": "string",
  "UserSettings": {
    "CanvasAppSettings": {
      "DirectDeploySettings": {
        "Status": "string"
      },
      "IdentityProviderOAuthSettings": [
        { "DataSourceName": "string", "SecretArn": "string", "Status": "string" }
      ],
      "KendraSettings": {
        "Status": "string"
      },
      "ModelRegisterSettings": {
        "CrossAccountModelRegisterRoleArn": "string",
        "Status": "string"
      },
      "TimeSeriesForecastingSettings": {
        "AmazonForecastRoleArn": "string",
        "Status": "string"
      },
      "WorkspaceSettings": {
        "S3ArtifactPath": "string",
        "S3KmsKeyId": "string"
      }
    },
    "CodeEditorAppSettings": {
      "DefaultResourceSpec": {
        "InstanceType": "string",
        "LifecycleConfigArn": "string",
        "SageMakerImageArn": "string",
        "SageMakerImageVersionAlias": "string",
        "SageMakerImageVersionArn": "string"
      },
      "LifecycleConfigArns": [ "string" ]
    },
    "CustomFileSystemConfigs": [ ...
    ],
    "CustomPosixUserConfig": {
      "Gid": number,
      "Uid": number
    },
    "DefaultLandingUri": "string",
    "ExecutionRole": "string",
    "JupyterLabAppSettings": {
      "CodeRepositories": [
        { "RepositoryUrl": "string"
      }]
    }
  }
}
```
"CustomImages": [
  {  
    "AppImageConfigName": "string",
    "ImageName": "string",
    "ImageVersionNumber": number 
  }
],
"DefaultResourceSpec": {  
  "InstanceType": "string",
  "LifecycleConfigArn": "string",
  "SageMakerImageArn": "string",
  "SageMakerImageVersionAlias": "string",
  "SageMakerImageVersionArn": "string"
},
"LifecycleConfigArns": [ "string" ]
],
"JupyterServerAppSettings": {  
  "CodeRepositories": [ {  
    "RepositoryUrl": "string"  
  }  
  ],
  "DefaultResourceSpec": {  
   "InstanceType": "string",
   "LifecycleConfigArn": "string",
   "SageMakerImageArn": "string",
   "SageMakerImageVersionAlias": "string",
   "SageMakerImageVersionArn": "string"
  },
  "LifecycleConfigArns": [ "string" ]
},
"KernelGatewayAppSettings": {  
  "CustomImages": [  
    {  
      "AppImageConfigName": "string",
      "ImageName": "string",
      "ImageVersionNumber": number 
    }
  ],
  "DefaultResourceSpec": {  
    "InstanceType": "string",
    "LifecycleConfigArn": "string",
    "SageMakerImageArn": "string",
    "SageMakerImageVersionAlias": "string",
    "SageMakerImageVersionArn": "string"
  },
  "LifecycleConfigArns": [ "string" ]
},
"RSessionAppSettings": {  
  "CustomImages": [  
    {  
      "AppImageConfigName": "string",
      "ImageName": "string",
      "ImageVersionNumber": number 
    }
  ],
  "DefaultResourceSpec": {  
    "InstanceType": "string",
    "LifecycleConfigArn": "string",
    "SageMakerImageArn": "string",
    "SageMakerImageVersionAlias": "string",
    "SageMakerImageVersionArn": "string"
  }
},
"RStudioServerProAppSettings": {  
  "AccessStatus": "string"  
}
"UserGroup": "string",
"SecurityGroups": [ "string" ],
"SharingSettings": {
  "NotebookOutputOption": "string",
  "S3KmsKeyId": "string",
  "S3OutputPath": "string"
},
"SpaceStorageSettings": {
  "DefaultEbsStorageSettings": {
    "DefaultEbsVolumeSizeInGb": number,
    "MaximumEbsVolumeSizeInGb": number
  }
},
"StudioWebPortal": "string",
"TensorBoardAppSettings": {
  "DefaultResourceSpec": {
    "InstanceType": "string",
    "LifecycleConfigArn": "string",
    "SageMakerImageArn": "string",
    "SageMakerImageVersionAlias": "string",
    "SageMakerImageVersionArn": "string"
  }
}

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).

The request accepts the following data in JSON format.

**DomainId (p. 1082)**

The domain ID.

Type: String

Length Constraints: Maximum length of 63.

Required: Yes

**UserProfileName (p. 1082)**

The user profile name.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9\-]*[a-zA-Z0-9]\{0,62}$

Required: Yes

**UserSettings (p. 1082)**

A collection of settings.

Type: UserSettings (p. 2070) object

Required: No
Response Syntax

```json
{
   "UserProfileArn": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**UserProfileArn (p. 1085)**

The user profile Amazon Resource Name (ARN).

Type: String

Length Constraints: Maximum length of 256.

Pattern: `arn:aws[a-z-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:user-profile/*`

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceInUse**

Resource being accessed is in use.

HTTP Status Code: 400

**ResourceLimitExceeded**

You have exceeded a SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

**ResourceNotFound**

Resource being access is not found.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
UpdateWorkforce

Service: Amazon SageMaker Service

Use this operation to update your workforce. You can use this operation to require that workers use specific IP addresses to work on tasks and to update your OpenID Connect (OIDC) Identity Provider (IdP) workforce configuration.

The worker portal is now supported in VPC and public internet.

Use SourceIpConfig to restrict worker access to tasks to a specific range of IP addresses. You specify allowed IP addresses by creating a list of up to ten CIDRs. By default, a workforce isn't restricted to specific IP addresses. If you specify a range of IP addresses, workers who attempt to access tasks using any IP address outside the specified range are denied and get a Not Found error message on the worker portal.

To restrict access to all the workers in public internet, add the SourceIpConfig CIDR value as "10.0.0.0/16".

**Important**
Amazon SageMaker does not support Source Ip restriction for worker portals in VPC.

Use OidcConfig to update the configuration of a workforce created using your own OIDC IdP.

**Important**
You can only update your OIDC IdP configuration when there are no work teams associated with your workforce. You can delete work teams using the DeleteWorkteam operation.

After restricting access to a range of IP addresses or updating your OIDC IdP configuration with this operation, you can view details about your update workforce using the DescribeWorkforce operation.

**Important**
This operation only applies to private workforces.

Request Syntax

```json
{
    "OidcConfig": {
        "AuthorizationEndpoint": "string",
        "ClientId": "string",
        "ClientSecret": "string",
        "Issuer": "string",
        "JwksUri": "string",
        "LogoutEndpoint": "string",
        "TokenEndpoint": "string",
        "UserInfoEndpoint": "string"
    },
    "SourceIpConfig": {
        "Cidrs": [ "string" ]
    },
    "WorkforceName": "string",
    "WorkforceVpcConfig": {
        "SecurityGroupIds": [ "string" ],
        "Subnets": [ "string" ],
        "VpcId": "string"
    }
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 2178).
The request accepts the following data in JSON format.

**OidcConfig (p. 1087)**

Use this parameter to update your OIDC Identity Provider (IdP) configuration for a workforce made using your own IdP.

Type: **OidcConfig (p. 1779)** object

Required: No

**SourceIpConfig (p. 1087)**

A list of one to ten worker IP address ranges (CIDRs) that can be used to access tasks assigned to this workforce.

Maximum: Ten CIDR values

Type: **SourceIpConfig (p. 1956)** object

Required: No

**WorkforceName (p. 1087)**

The name of the private workforce that you want to update. You can find your workforce name by using the **ListWorkforces** operation.

Type: String


Pattern: `^[a-zA-Z0-9][-]{0,62}$`

Required: Yes

**WorkforceVpcConfig (p. 1087)**

Use this parameter to update your VPC configuration for a workforce.

Type: **WorkforceVpcConfigRequest (p. 2081)** object

Required: No

**Response Syntax**

```json
{
    "Workforce": {
        "CognitoConfig": {
            "ClientId": "string",
            "UserPool": "string"
        },
        "CreateDate": number,
        "FailureReason": "string",
        "LastUpdatedDate": number,
        "OidcConfig": {
            "AuthorizationEndpoint": "string",
            "ClientId": "string",
            "Issuer": "string",
            "JwksUri": "string",
            "LogoutEndpoint": "string",
            "TokenEndpoint": "string",
            "UserInfoEndpoint": "string"
        }
    }
}
```
"SourceIpConfig": {  "Cidrs": [ "string" ] },  "Status": "string",  "SubDomain": "string",  "WorkforceArn": "string",  "WorkforceName": "string",  "WorkforceVpcConfig": {  "SecurityGroupIds": [ "string" ],  "Subnets": [ "string" ],  "VpcEndpointId": "string",  "VpcId": "string"  }  }

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**Workforce (p. 1088)**

A single private workforce. You can create one private work force in each AWS Region. By default, any workforce-related API operation used in a specific region will apply to the workforce created in that region. To learn how to create a private workforce, see [Create a Private Workforce](#).

Type: [Workforce](p. 2079) object

Errors

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](#).

**ConflictException**

There was a conflict when you attempted to modify a SageMaker entity such as an Experiment or Artifact.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
**UpdateWorkteam**

Service: Amazon SageMaker Service

Updates an existing work team with new member definitions or description.

**Request Syntax**

```json
{
   "Description": "string",
   "MemberDefinitions": [
      {
         "CognitoMemberDefinition": {
            "ClientId": "string",
            "UserGroup": "string",
            "UserPool": "string"
         },
         "OidcMemberDefinition": {
            "Groups": [ "string" ]
         }
      }
   ],
   "NotificationConfiguration": {
      "NotificationTopicArn": "string"
   },
   "WorkteamName": "string"
}
```

**Request Parameters**

For information about the parameters that are common to all actions, see [Common Parameters](p. 2178).

The request accepts the following data in JSON format.

**Description (p. 1090)**

An updated description for the work team.

Type: String


Pattern: .+

Required: No

**MemberDefinitions (p. 1090)**

A list of `MemberDefinition` objects that contains objects that identify the workers that make up the work team.

Workforces can be created using Amazon Cognito or your own OIDC Identity Provider (IdP). For private workforces created using Amazon Cognito use `CognitoMemberDefinition`. For workforces created using your own OIDC identity provider (IdP) use `OidcMemberDefinition`. You should not provide input for both of these parameters in a single request.

For workforces created using Amazon Cognito, private work teams correspond to Amazon Cognito user groups within the user pool used to create a workforce. All of the `CognitoMemberDefinition` objects that make up the member definition must have the same `ClientId` and `UserPool` values.
To add a Amazon Cognito user group to an existing worker pool, see Adding groups to a User Pool. For more information about user pools, see Amazon Cognito User Pools.

For workforces created using your own OIDC IdP, specify the user groups that you want to include in your private work team in OidcMemberDefinition by listing those groups in Groups. Be aware that user groups that are already in the work team must also be listed in Groups when you make this request to remain on the work team. If you do not include these user groups, they will no longer be associated with the work team you update.

Type: Array of MemberDefinition (p. 1647) objects

Array Members: Minimum number of 1 item. Maximum number of 10 items.

Required: No

NotificationConfiguration (p. 1090)

Configures SNS topic notifications for available or expiring work items

Type: NotificationConfiguration (p. 1775) object

Required: No

WorkteamName (p. 1090)

The name of the work team to update.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

Required: Yes

Response Syntax

```
{
  "Workteam": {
    "CreateDate": "number",
    "Description": "string",
    "LastUpdatedDate": "number",
    "MemberDefinitions": [
      {
        "CognitoMemberDefinition": {
          "ClientId": "string",
          "UserGroup": "string",
          "UserPool": "string"
        },
        "OidcMemberDefinition": {
          "Groups": [ "string" ]
        }
      }]
    },
    "NotificationConfiguration": {
      "NotificationTopicArn": "string"
    },
    "ProductListingIds": [ "string" ],
    "SubDomain": "string",
    "WorkforceArn": "string",
    "WorkteamArn": "string",
    "WorkteamName": "string"
  }
}
```
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**Workteam (p. 1091)**

A Workteam object that describes the updated work team.

Type: Workteam (p. 2085) object

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**ResourceLimitExceeded**

You have exceeded an SageMaker resource limit. For example, you might have too many training jobs created.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

Amazon SageMaker Runtime

The following actions are supported by Amazon SageMaker Runtime:

- [InvokeEndpoint (p. 1093)](#)
- [InvokeEndpointAsync (p. 1099)](#)
- [InvokeEndpointWithResponseStream (p. 1103)](#)
InvokeEndpoint
Service: Amazon SageMaker Runtime

After you deploy a model into production using Amazon SageMaker hosting services, your client applications use this API to get inferences from the model hosted at the specified endpoint.

For an overview of Amazon SageMaker, see [How It Works](#).

Amazon SageMaker strips all POST headers except those supported by the API. Amazon SageMaker might add additional headers. You should not rely on the behavior of headers outside those enumerated in the request syntax.

Calls to InvokeEndpoint are authenticated by using AWS Signature Version 4. For information, see [Authenticating Requests (AWS Signature Version 4)](#) in the [Amazon S3 API Reference](#).

A customer's model containers must respond to requests within 60 seconds. The model itself can have a maximum processing time of 60 seconds before responding to invocations. If your model is going to take 50-60 seconds of processing time, the SDK socket timeout should be set to be 70 seconds.

**Note**
Endpoints are scoped to an individual account, and are not public. The URL does not contain the account ID, but Amazon SageMaker determines the account ID from the authentication token that is supplied by the caller.

Request Syntax

```
POST /endpoints/EndpointName/invocations HTTP/1.1
Content-Type: Content-Type
Accept: Accept
X-Amzn-SageMaker-Custom-Attributes:CustomAttributes
X-Amzn-SageMaker-Target-Model: TargetModel
X-Amzn-SageMaker-Target-Variant: TargetVariant
X-Amzn-SageMaker-Target-Container-Hostname: TargetContainerHostname
X-Amzn-SageMaker-Inference-Id: InferenceId
X-Amzn-SageMaker-Enable-Explanations: EnableExplanations
X-Amzn-SageMaker-Inference-Component: InferenceComponentName

Body
```

URI Request Parameters

The request uses the following URI parameters.

**Accept (p. 1093)**

The desired MIME type of the inference response from the model container.

Length Constraints: Maximum length of 1024.

Pattern: \p{ASCII}*

**ContentType (p. 1093)**

The MIME type of the input data in the request body.

Length Constraints: Maximum length of 1024.

Pattern: \p{ASCII}**
**CustomAttributes (p. 1093)**

Provides additional information about a request for an inference submitted to a model hosted at an Amazon SageMaker endpoint. The information is an opaque value that is forwarded verbatim. You could use this value, for example, to provide an ID that you can use to track a request or to provide other metadata that a service endpoint was programmed to process. The value must consist of no more than 1024 visible US-ASCII characters as specified in Section 3.3.6. Field Value Components of the Hypertext Transfer Protocol (HTTP/1.1).

The code in your model is responsible for setting or updating any custom attributes in the response. If your code does not set this value in the response, an empty value is returned. For example, if a custom attribute represents the trace ID, your model can prepend the custom attribute with Trace ID: in your post-processing function.

This feature is currently supported in the AWS SDKs but not in the Amazon SageMaker Python SDK.

Length Constraints: Maximum length of 1024.

Pattern: \p{ASCII}*  

**EnableExplanations (p. 1093)**

An optional JMESPath expression used to override the EnableExplanations parameter of the ClarifyExplainerConfig API. See the EnableExplanations section in the developer guide for more information.

Length Constraints: Minimum length of 1. Maximum length of 64.

Pattern: .*

**EndpointName (p. 1093)**

The name of the endpoint that you specified when you created the endpoint using the CreateEndpoint API.

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])*$  

Required: Yes

**InferenceComponentName (p. 1093)**

If the endpoint hosts one or more inference components, this parameter specifies the name of inference component to invoke.

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](\[-a-zA-Z0-9]*[a-zA-Z0-9])?$

**InferenceId (p. 1093)**

If you provide a value, it is added to the captured data when you enable data capture on the endpoint. For information about data capture, see Capture Data.

Length Constraints: Minimum length of 1. Maximum length of 64.

Pattern: \A\S\[\p{Print}]*\z

**TargetContainerHostname (p. 1093)**

If the endpoint hosts multiple containers and is configured to use direct invocation, this parameter specifies the host name of the container to invoke.
Length Constraints: Maximum length of 63.
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])*

**TargetModel (p. 1093)**

The model to request for inference when invoking a multi-model endpoint.

Pattern: \A\S[\p{Print}]\*\z

**TargetVariant (p. 1093)**

Specify the production variant to send the inference request to when invoking an endpoint that is running two or more variants. Note that this parameter overrides the default behavior for the endpoint, which is to distribute the invocation traffic based on the variant weights.

For information about how to use variant targeting to perform a/b testing, see Test models in production

Length Constraints: Maximum length of 63.
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])*

**Request Body**

The request accepts the following binary data.

**Body (p. 1093)**

Provides input data, in the format specified in the ContentType request header. Amazon SageMaker passes all of the data in the body to the model.

For information about the format of the request body, see Common Data Formats-Inference.

Length Constraints: Maximum length of 6291456.
Required: Yes

**Response Syntax**

<table>
<thead>
<tr>
<th>HTTP/1.1 200</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content-Type: <strong>ContentType</strong></td>
</tr>
<tr>
<td>x-Amzn-Invoked-Production-Variant: <strong>InvokedProductionVariant</strong></td>
</tr>
<tr>
<td>X-Amzn-SageMaker-Custom-Attributes: <strong>CustomAttributes</strong></td>
</tr>
</tbody>
</table>

**Body**

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The response returns the following HTTP headers.

**ContentType (p. 1095)**

The MIME type of the inference returned from the model container.
Length Constraints: Maximum length of 1024.
Pattern: \p{ASCII}*

**CustomAttributes (p. 1095)**

Provides additional information in the response about the inference returned by a model hosted at an Amazon SageMaker endpoint. The information is an opaque value that is forwarded verbatim. You could use this value, for example, to return an ID received in the CustomAttributes header of a request or other metadata that a service endpoint was programmed to produce. The value must consist of no more than 1024 visible US-ASCII characters as specified in [Section 3.3.6. Field Value Components](https://www.w3.org/Protocols/rfc2616/rfc2616-sec3.3.html) of the Hypertext Transfer Protocol (HTTP/1.1). If the customer wants the custom attribute returned, the model must set the custom attribute to be included on the way back.

The code in your model is responsible for setting or updating any custom attributes in the response. If your code does not set this value in the response, an empty value is returned. For example, if a custom attribute represents the trace ID, your model can prepend the custom attribute with `Trace ID:` in your post-processing function.

This feature is currently supported in the AWS SDKs but not in the Amazon SageMaker Python SDK.

Length Constraints: Maximum length of 1024.
Pattern: \p{ASCII}*

**InvokedProductionVariant (p. 1095)**

Identifies the production variant that was invoked.

Length Constraints: Maximum length of 1024.
Pattern: \p{ASCII}*

The response returns the following as the HTTP body.

**Body (p. 1095)**

Includes the inference provided by the model.

For information about the format of the response body, see [Common Data Formats-Inference](https://docs.aws.amazon.com/sagemaker/latest/dg/inference-data-format.html).

If the explainer is activated, the body includes the explanations provided by the model. For more information, see the Response section under [Invoke the Endpoint](https://docs.aws.amazon.com/sagemaker/latest/dg/invoke-endpoint.html) in the Developer Guide.

Length Constraints: Maximum length of 6291456.

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](https://docs.aws.amazon.com/sagemaker/latest/dg/common-errors.html).

**InternalDependencyException**

Your request caused an exception with an internal dependency. Contact customer support.

HTTP Status Code: 530

**InternalFailure**

An internal failure occurred.

HTTP Status Code: 500
ModelError

Model (owned by the customer in the container) returned 4xx or 5xx error code.

HTTP Status Code: 424

ModelError

ModelNotReadyException

Either a serverless endpoint variant's resources are still being provisioned, or a multi-model endpoint is still downloading or loading the target model. Wait and try your request again.

HTTP Status Code: 429

ServiceUnavailable

The service is unavailable. Try your call again.

HTTP Status Code: 503

ValidationError

Inspect your request and try again.

HTTP Status Code: 400

Examples

Pass a trace ID in the CustomAttribute of a request and return it in the CustomAttribute of the response.

In this example a trace ID is passed to the service endpoint in the CustomAttributes header of the request and then retrieved and returned in the CustomAttributes header of the response.

Sample Request

```python
import boto3
client = boto3.client('sagemaker-runtime')
custom_attributes = "c000b4f9-df62-4c85-a0bf-7c525f9104a4"  # An example of a trace ID.
endpoint_name = "..."  # Your endpoint name.
content_type = "..."  # The MIME type of the input data in the request body.
accept = "..."  # The desired MIME type of the inference in the response.
payload = "..."  # Payload for inference.
response = client.invoke_endpoint(
    EndpointName=endpoint_name,
   CustomAttributes=custom_attributes,
   ContentType=content_type,
   Accept=accept,
   Body=payload
)
print(response['CustomAttributes'])  # If model receives and updates the custom_attributes header in front of custom_attributes in the request, becomes "Trace ID: c000b4f9-"
Sample Response

Trace ID: c000b4f9-df62-4c85-a0bf-7c525f9104a4

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
InvokeEndpointAsync

Service: Amazon SageMaker Runtime

After you deploy a model into production using Amazon SageMaker hosting services, your client applications use this API to get inferences from the model hosted at the specified endpoint in an asynchronous manner.

Inference requests sent to this API are enqueued for asynchronous processing. The processing of the inference request may or may not complete before you receive a response from this API. The response from this API will not contain the result of the inference request but contain information about where you can locate it.

Amazon SageMaker strips all POST headers except those supported by the API. Amazon SageMaker might add additional headers. You should not rely on the behavior of headers outside those enumerated in the request syntax.

Calls to InvokeEndpointAsync are authenticated by using AWS Signature Version 4. For information, see Authenticating Requests (AWS Signature Version 4) in the Amazon S3 API Reference.

Request Syntax

<table>
<thead>
<tr>
<th>POST /endpoints/EndpointName/async-invocations HTTP/1.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>X-Amzn-SageMaker-Content-Type: ContentType</td>
</tr>
<tr>
<td>X-Amzn-SageMaker-Accept: Accept</td>
</tr>
<tr>
<td>X-Amzn-SageMaker-Custom-Attributes: CustomAttributes</td>
</tr>
<tr>
<td>X-Amzn-SageMaker-Inference-Id: InferenceId</td>
</tr>
<tr>
<td>X-Amzn-SageMaker-InputLocation: InputLocation</td>
</tr>
<tr>
<td>X-Amzn-SageMaker-RequestTTLSeconds: RequestTTLSeconds</td>
</tr>
<tr>
<td>X-Amzn-SageMaker-InvocationTimeoutSeconds: InvocationTimeoutSeconds</td>
</tr>
</tbody>
</table>

URI Request Parameters

The request uses the following URI parameters.

Accept (p. 1099)

The desired MIME type of the inference response from the model container.

Length Constraints: Maximum length of 1024.

Pattern: \p{ASCII}*

ContentType (p. 1099)

The MIME type of the input data in the request body.

Length Constraints: Maximum length of 1024.

Pattern: \p{ASCII}*

CustomAttributes (p. 1099)

Provides additional information about a request for an inference submitted to a model hosted at an Amazon SageMaker endpoint. The information is an opaque value that is forwarded verbatim. You could use this value, for example, to provide an ID that you can use to track a request or to provide other metadata that a service endpoint was programmed to process. The value must consist of no more than 1024 visible US-ASCII characters as specified in Section 3.3.6. Field Value Components of the Hypertext Transfer Protocol (HTTP/1.1).
The code in your model is responsible for setting or updating any custom attributes in the response. If your code does not set this value in the response, an empty value is returned. For example, if a custom attribute represents the trace ID, your model can prepend the custom attribute with Trace ID: in your post-processing function.

This feature is currently supported in the AWS SDKs but not in the Amazon SageMaker Python SDK.

Length Constraints: Maximum length of 1024.

Pattern: \p{ASCII}*  

**EndpointName (p. 1099)**  
The name of the endpoint that you specified when you created the endpoint using the CreateEndpoint API.  
Length Constraints: Maximum length of 63.  
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])*  
Required: Yes  

**InferenceId (p. 1099)**  
The identifier for the inference request. Amazon SageMaker will generate an identifier for you if none is specified.  
Length Constraints: Minimum length of 1. Maximum length of 64.  
Pattern: \A\S[\p{Print}]*/z  

**InputLocation (p. 1099)**  
The Amazon S3 URI where the inference request payload is stored.  
Pattern: ^(https|s3)://(/[\^/]+)/?(.*)$  
Required: Yes  

**InvocationTimeoutSeconds (p. 1099)**  
Maximum amount of time in seconds a request can be processed before it is marked as expired. The default is 15 minutes, or 900 seconds.  

**RequestTTLSeconds (p. 1099)**  
Maximum age in seconds a request can be in the queue before it is marked as expired. The default is 6 hours, or 21,600 seconds.  
Valid Range: Minimum value of 60. Maximum value of 21600.

**Request Body**  
The request does not have a request body.  

**Response Syntax**

```
HTTP/1.1 202
X-Amzn-SageMaker-OutputLocation: OutputLocation
```
Response Elements

If the action is successful, the service sends back an HTTP 202 response. The response returns the following HTTP headers.

**FailureLocation (p. 1100)**

The Amazon S3 URI where the inference failure response payload is stored.

Length Constraints: Maximum length of 1024.

Pattern: \p{ASCII}*

**OutputLocation (p. 1100)**

The Amazon S3 URI where the inference response payload is stored.

Length Constraints: Maximum length of 1024.

Pattern: \p{ASCII}*

The following data is returned in JSON format by the service.

**InferenceId (p. 1100)**

Identifier for an inference request. This will be the same as the InferenceId specified in the input. Amazon SageMaker will generate an identifier for you if you do not specify one.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: \p{ASCII}*

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**InternalFailure**

An internal failure occurred.

HTTP Status Code: 500

**ServiceUnavailable**

The service is unavailable. Try your call again.

HTTP Status Code: 503

**ValidationError**

Inspect your request and try again.
HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
InvokeEndpointWithResponseStream

Service: Amazon SageMaker Runtime

Invokes a model at the specified endpoint to return the inference response as a stream. The inference stream provides the response payload incrementally as a series of parts. Before you can get an inference stream, you must have access to a model that's deployed using Amazon SageMaker hosting services, and the container for that model must support inference streaming.

For more information that can help you use this API, see the following sections in the Amazon SageMaker Developer Guide:

- For information about how to add streaming support to a model, see How Containers Serve Requests.
- For information about how to process the streaming response, see Invoke real-time endpoints.

Before you can use this operation, your IAM permissions must allow the sagemaker:InvokeEndpoint action. For more information about Amazon SageMaker actions for IAM policies, see Actions, resources, and condition keys for Amazon SageMaker in the IAM Service Authorization Reference.

Amazon SageMaker strips all POST headers except those supported by the API. Amazon SageMaker might add additional headers. You should not rely on the behavior of headers outside those enumerated in the request syntax.

Calls to InvokeEndpointWithResponseStream are authenticated by using AWS Signature Version 4. For information, see Authenticating Requests (AWS Signature Version 4) in the Amazon S3 API Reference.

Request Syntax

POST /endpoints/EndpointName/invocations-response-stream HTTP/1.1
Content-Type: Content-Type
X-Amzn-SageMaker-Accept: Accept
X-Amzn-SageMaker-Custom-Attributes: CustomAttributes
X-Amzn-SageMaker-Target-Variant: TargetVariant
X-Amzn-SageMaker-Target-Container-Hostname: TargetContainerHostname
X-Amzn-SageMaker-Inference-Id: InferenceId
X-Amzn-SageMaker-Inference-Component: InferenceComponentName

Body

URI Request Parameters

The request uses the following URI parameters.

Accept (p. 1103)

The desired MIME type of the inference response from the model container.

Length Constraints: Maximum length of 1024.

Pattern: \p{ASCII}*

ContentType (p. 1103)

The MIME type of the input data in the request body.

Length Constraints: Maximum length of 1024.

Pattern: \p{ASCII}*

1103
CustomAttributes (p. 1103)

Provides additional information about a request for an inference submitted to a model hosted at an Amazon SageMaker endpoint. The information is an opaque value that is forwarded verbatim. You could use this value, for example, to provide an ID that you can use to track a request or to provide other metadata that a service endpoint was programmed to process. The value must consist of no more than 1024 visible US-ASCII characters as specified in Section 3.3.6. Field Value Components of the Hypertext Transfer Protocol (HTTP/1.1).

The code in your model is responsible for setting or updating any custom attributes in the response. If your code does not set this value in the response, an empty value is returned. For example, if a custom attribute represents the trace ID, your model can prepend the custom attribute with Trace ID: in your post-processing function.

This feature is currently supported in the AWS SDKs but not in the Amazon SageMaker Python SDK.

Length Constraints: Maximum length of 1024.

Pattern: \p{ASCII}*

EndpointName (p. 1103)

The name of the endpoint that you specified when you created the endpoint using the CreateEndpoint API.

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])*$

InferenceComponentName (p. 1103)

If the endpoint hosts one or more inference components, this parameter specifies the name of inference component to invoke for a streaming response.

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9]([-a-zA-Z0-9]*[a-zA-Z0-9])?$

InferenceId (p. 1103)

An identifier that you assign to your request.

Length Constraints: Minimum length of 1. Maximum length of 64.

Pattern: \A\S\p{Print}]*\z

TargetContainerHostname (p. 1103)

If the endpoint hosts multiple containers and is configured to use direct invocation, this parameter specifies the host name of the container to invoke.

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9]([-a-zA-Z0-9]*[a-zA-Z0-9])*$

TargetVariant (p. 1103)

Specify the production variant to send the inference request to when invoking an endpoint that is running two or more variants. Note that this parameter overrides the default behavior for the endpoint, which is to distribute the invocation traffic based on the variant weights.

For information about how to use variant targeting to perform a/b testing, see Test models in production.
Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])*

**Request Body**

The request accepts the following binary data.

**Body (p. 1103)**

Provides input data, in the format specified in the `ContentType` request header. Amazon SageMaker passes all of the data in the body to the model.

For information about the format of the request body, see [Common Data Formats-Inference](#).

Length Constraints: Maximum length of 6291456.

Required: Yes

**Response Syntax**

```plaintext
HTTP/1.1 200
X-Amzn-SageMaker-Content-Type: Content_Type
x-Amzn-Invoked-Production-Variant: Invoked_Production_Variant
X-Amzn-SageMaker-Custom-Attributes: Custom_Attributes
Content-type: application/json
{
  "InternalStreamFailure": {
  },
  "ModelError": {
  },
  "PayloadPart": {
    "Bytes": blob
  }
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The response returns the following HTTP headers.

**ContentType (p. 1105)**

The MIME type of the inference returned from the model container.

Length Constraints: Maximum length of 1024.

Pattern: \p{ASCII}*

**CustomAttributes (p. 1105)**

Provides additional information in the response about the inference returned by a model hosted at an Amazon SageMaker endpoint. The information is an opaque value that is forwarded verbatim. You could use this value, for example, to return an ID received in the CustomAttributes header of a request or other metadata that a service endpoint was programmed to produce. The value must consist of no more than 1024 visible US-ASCII characters as specified in [Section 3.3.6. Field](#).
Value Components of the Hypertext Transfer Protocol (HTTP/1.1). If the customer wants the custom attribute returned, the model must set the custom attribute to be included on the way back.

The code in your model is responsible for setting or updating any custom attributes in the response. If your code does not set this value in the response, an empty value is returned. For example, if a custom attribute represents the trace ID, your model can prepend the custom attribute with Trace ID: in your post-processing function.

This feature is currently supported in the AWS SDKs but not in the Amazon SageMaker Python SDK.

Length Constraints: Maximum length of 1024.

InvokedProductionVariant

Identifies the production variant that was invoked.

Length Constraints: Maximum length of 1024.

Pattern: \p{ASCII}*

The following data is returned in JSON format by the service.

InternalStreamFailure

The stream processing failed because of an unknown error, exception or failure. Try your request again.

Type: Exception

HTTP Status Code:

ModelError

An error occurred while streaming the response body. This error can have the following error codes:

ModelInvocationTimeExceeded

The model failed to finish sending the response within the timeout period allowed by Amazon SageMaker.

StreamBroken

The Transmission Control Protocol (TCP) connection between the client and the model was reset or closed.

Type: Exception

HTTP Status Code:

PayloadPart

A wrapper for pieces of the payload that's returned in response to a streaming inference request. A streaming inference response consists of one or more payload parts.

Type: PayloadPart object

Errors

For information about the errors that are common to all actions, see Common Errors.

InternalFailure

An internal failure occurred.
HTTP Status Code: 500

**InternalStreamFailure**

The stream processing failed because of an unknown error, exception or failure. Try your request again.

HTTP Status Code: 500

**ModelError**

Model (owned by the customer in the container) returned 4xx or 5xx error code.

HTTP Status Code: 424

**ModelStreamError**

An error occurred while streaming the response body. This error can have the following error codes:

- **ModelInvocationTimeExceeded**
  
  The model failed to finish sending the response within the timeout period allowed by Amazon SageMaker.

- **StreamBroken**
  
  The Transmission Control Protocol (TCP) connection between the client and the model was reset or closed.

HTTP Status Code: 400

**ServiceUnavailable**

The service is unavailable. Try your call again.

HTTP Status Code: 503

**Validation>Error**

Inspect your request and try again.

HTTP Status Code: 400

### See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3

### Amazon Sagemaker Edge Manager

The following actions are supported by Amazon Sagemaker Edge Manager:
• GetDeployments (p. 1109)
• GetDeviceRegistration (p. 1111)
• SendHeartbeat (p. 1113)
GetDeployments
Service: Amazon Sagemaker Edge Manager
Use to get the active deployments from a device.

Request Syntax
```plaintext
POST /GetDeployments HTTP/1.1
Content-type: application/json
{
  "DeviceFleetName": "string",
  "DeviceName": "string"
}
```

URI Request Parameters
The request does not use any URI parameters.

Request Body
The request accepts the following data in JSON format.

DeviceFleetName (p. 1109)
The name of the fleet to which the device belongs.
Type: String
Pattern: ^[a-zA-Z0-9](-*_*[a-zA-Z0-9])*$
Required: Yes

DeviceName (p. 1109)
The unique name of the device from which you want to get the configuration of active deployments.
Type: String
Pattern: ^[a-zA-Z0-9](-*_*[a-zA-Z0-9])*$
Required: Yes

Response Syntax
```plaintext
HTTP/1.1 200
Content-type: application/json
{
  "Deployments": [
    {
      "Definitions": [
        {
          "Checksum": {
          
```
GetDeployments

```
{ "Sum": "string",
  "Type": "string"
},
{ "ModelHandle": "string",
  "S3Url": "string",
  "State": "string"
},
{ "DeploymentName": "string",
  "FailureHandlingPolicy": "string",
  "Type": "string"
}
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**Deployments (p. 1109)**

Returns a list of the configurations of the active deployments on the device.

Type: Array of [EdgeDeployment (p. 2097)] objects

Errors

For information about the errors that are common to all actions, see [Common Errors (p. 2180)].

**InternalServiceException**

An internal failure occurred. Try your request again. If the problem persists, contact AWS customer support.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
GetDeviceRegistration

Service: Amazon Sagemaker Edge Manager

Use to check if a device is registered with SageMaker Edge Manager.

Request Syntax

```plaintext
POST /GetDeviceRegistration HTTP/1.1
Content-type: application/json

{
   "DeviceFleetName": "string",
   "DeviceName": "string"
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

**DeviceFleetName (p. 1111)**

The name of the fleet to which the device belongs.

- Type: String
- Pattern: `^[a-zA-Z0-9](-*_*[a-zA-Z0-9])*$`
- Required: Yes

**DeviceName (p. 1111)**

The unique name of the device from which you want to get the registration status.

- Type: String
- Pattern: `^[a-zA-Z0-9](-*_*[a-zA-Z0-9])*$`
- Required: Yes

Response Syntax

```plaintext
HTTP/1.1 200
Content-type: application/json

{
   "CacheTTL": "string",
   "DeviceRegistration": "string"
}
```
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**CacheTTL (p. 1111)**

The amount of time, in seconds, that the registration status is stored on the device's cache before it is refreshed.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 1000.

**DeviceRegistration (p. 1111)**

Describes if the device is currently registered with SageMaker Edge Manager.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 1000.

Errors

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](#).

**InternalServiceException**

An internal failure occurred. Try your request again. If the problem persists, contact AWS customer support.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
SendHeartbeat
Service: Amazon Sagemaker Edge Manager

Use to get the current status of devices registered on SageMaker Edge Manager.

Request Syntax

POST /SendHeartbeat HTTP/1.1
Content-type: application/json

{
    "AgentMetrics": [
        {
            "Dimension": "string",
            "MetricName": "string",
            "Timestamp": number,
            "Value": number
        }
    ],
    "AgentVersion": "string",
    "DeploymentResult": {
        "DeploymentEndTime": number,
        "DeploymentModels": [
            {
                "DesiredState": "string",
                "ModelHandle": "string",
                "ModelName": "string",
                "ModelVersion": "string",
                "RollbackFailureReason": "string",
                "State": "string",
                "Status": "string",
                "StatusReason": "string"
            }
        ],
        "DeploymentName": "string",
        "DeploymentStartTime": number,
        "DeploymentStatus": "string",
        "DeploymentStatusMessage": "string"
    },
    "DeviceFleetName": "string",
    "DeviceName": "string",
    "Models": [
        {
            "LatestInference": number,
            "LatestSampleTime": number,
            "ModelMetrics": [
                {
                    "Dimension": "string",
                    "MetricName": "string",
                    "Timestamp": number,
                    "Value": number
                }
            ],
            "ModelName": "string",
            "ModelVersion": "string"
        }
    ]
}

URI Request Parameters

The request does not use any URI parameters.
Request Body

The request accepts the following data in JSON format.

**AgentMetrics (p. 1113)**

For internal use. Returns a list of SageMaker Edge Manager agent operating metrics.

Type: Array of [EdgeMetric (p. 2098)] objects

Required: No

**AgentVersion (p. 1113)**

Returns the version of the agent.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 64.

Pattern: [a-zA-Z0-9\_\-\.]*

Required: Yes

**DeploymentResult (p. 1113)**

Returns the result of a deployment on the device.

Type: [DeploymentResult (p. 2095)] object

Required: No

**DeviceFleetName (p. 1113)**

The name of the fleet to which the device belongs.

Type: String


Pattern: ^[a-zA-Z0-9\_\-\.]($\_\-\-\_\-\_\-\_\-\_\-\_]\.*[a-zA-Z0-9]$)

Required: Yes

**DeviceName (p. 1113)**

The unique name of the device.

Type: String


Pattern: ^[a-zA-Z0-9\_\-\.]($\_\-\-\_\-\_\-\_\-\_\-\_]\.*[a-zA-Z0-9]$)

Required: Yes

**Models (p. 1113)**

Returns a list of models deployed on the device.

Type: Array of [Model (p. 2099)] objects

Required: No
**Response Syntax**

HTTP/1.1 200

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](#).

**InternalServiceException**

An internal failure occurred. Try your request again. If the problem persists, contact AWS customer support.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)

---

**Amazon SageMaker Feature Store Runtime**

The following actions are supported by Amazon SageMaker Feature Store Runtime:

- [BatchGetRecord (p. 1116)](#)
- [DeleteRecord (p. 1119)](#)
- [GetRecord (p. 1122)](#)
- [PutRecord (p. 1125)](#)
BatchGetRecord
Service: Amazon SageMaker Feature Store Runtime
Retrieves a batch of Records from a FeatureGroup.

Request Syntax

POST /BatchGetRecord HTTP/1.1
Content-type: application/json

{
   "ExpirationTimeResponse": "string",
   "Identifiers": [
      {
         "FeatureGroupName": "string",
         "FeatureNames": [ "string" ],
         "RecordIdentifiersValueAsString": [ "string" ]
      }
   ]
}

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

ExpirationTimeResponse (p. 1116)

Parameter to request ExpiresAt in response. If Enabled, BatchGetRecord will return the value of ExpiresAt, if it is not null. If Disabled and null, BatchGetRecord will return null.

Type: String

Valid Values: Enabled | Disabled

Required: No

Identifiers (p. 1116)

A list containing the name or Amazon Resource Name (ARN) of the FeatureGroup, the list of names of Features to be retrieved, and the corresponding RecordIdentifier values as strings.

Type: Array of BatchGetRecordIdentifier (p. 2103) objects

Array Members: Minimum number of 1 item. Maximum number of 10 items.

Required: Yes

Response Syntax

HTTP/1.1 200
Content-type: application/json

{
   "Errors": [
      {}
   ]
}
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

Errors (p. 1116)

A list of errors that have occurred when retrieving a batch of Records.

Type: Array of BatchGetRecordError (p. 2101) objects

Array Members: Minimum number of 0 items.

Records (p. 1116)

A list of Records you requested to be retrieved in batch.

Type: Array of BatchGetRecordResultDetail (p. 2105) objects

Array Members: Minimum number of 0 items.

UnprocessedIdentifiers (p. 1116)

A unprocessed list of FeatureGroup names, with their corresponding RecordIdentifier value, and Feature name.

Type: Array of BatchGetRecordIdentifier (p. 2103) objects

Array Members: Minimum number of 0 items.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).
AccessForbidden

You do not have permission to perform an action.

HTTP Status Code: 403

InternalFailure

An internal failure occurred. Try your request again. If the problem persists, contact AWS customer support.

HTTP Status Code: 500

ServiceUnavailable

The service is currently unavailable.

HTTP Status Code: 503

ValidationError

There was an error validating your request.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
DeleteRecord
Service: Amazon SageMaker Feature Store Runtime

Deletes a Record from a FeatureGroup in the OnlineStore. Feature Store supports both SoftDelete and HardDelete. For SoftDelete (default), feature columns are set to null and the record is no longer retrievable by GetRecord or BatchGetRecord. For HardDelete, the complete Record is removed from the OnlineStore. In both cases, Feature Store appends the deleted record marker to the OfflineStore. The deleted record marker is a record with the same RecordIdentifier as the original, but with is_deleted value set to True, EventTime set to the delete input EventTime, and other feature values set to null.

Note that the EventTime specified in DeleteRecord should be set later than the EventTime of the existing record in the OnlineStore for that RecordIdentifier. If it is not, the deletion does not occur:

- For SoftDelete, the existing (not deleted) record remains in the OnlineStore, though the delete record marker is still written to the OfflineStore.
- HardDelete returns EventTime: 400 ValidationException to indicate that the delete operation failed. No delete record marker is written to the OfflineStore.

When a record is deleted from the OnlineStore, the deleted record marker is appended to the OfflineStore. If you have the Iceberg table format enabled for your OfflineStore, you can remove all history of a record from the OfflineStore using Amazon Athena or Apache Spark. For information on how to hard delete a record from the OfflineStore with the Iceberg table format enabled, see Delete records from the offline store.

Request Syntax

```
DELETE /FeatureGroup/FeatureGroupName?
DeletionMode=DeletionMode&EventTime=EventTime&RecordIdentifierValueAsString=RecordIdentifierValueAsString&TargetStores=TargetStores
HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

**DeletionMode (p. 1119)**

The name of the deletion mode for deleting the record. By default, the deletion mode is set to SoftDelete.

Valid Values: SoftDelete | HardDelete

**EventTime (p. 1119)**

Timestamp indicating when the deletion event occurred. EventTime can be used to query data at a certain point in time.

Length Constraints: Maximum length of 358400.

Pattern: .*

Required: Yes

**FeatureGroupName (p. 1119)**

The name or Amazon Resource Name (ARN) of the feature group to delete the record from.

Length Constraints: Minimum length of 1. Maximum length of 150.
DeleteRecord

Pattern: (arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:feature-group/)?([a-zA-Z0-9][-_]*[a-zA-Z0-9]{0,63})

Required: Yes

RecordIdentifierValueAsString (p. 1119)

The value for the RecordIdentifier that uniquely identifies the record, in string format.

Length Constraints: Maximum length of 358400.

Pattern: .*

Required: Yes

TargetStores (p. 1119)

A list of stores from which you're deleting the record. By default, Feature Store deletes the record from all of the stores that you're using for the FeatureGroup.

Array Members: Minimum number of 1 item. Maximum number of 2 items.

Valid Values: OnlineStore | OfflineStore

Request Body

The request does not have a request body.

Response Syntax

HTTP/1.1 200

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

AccessForbidden

You do not have permission to perform an action.

HTTP Status Code: 403

InternalFailure

An internal failure occurred. Try your request again. If the problem persists, contact AWS customer support.

HTTP Status Code: 500

ServiceUnavailable

The service is currently unavailable.

HTTP Status Code: 503

ValidationErrors

There was an error validating your request.
HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
GetRecord
Service: Amazon SageMaker Feature Store Runtime

Use for OnlineStore serving from a FeatureStore. Only the latest records stored in the OnlineStore can be retrieved. If no Record with RecordIdentifierValue is found, then an empty result is returned.

Request Syntax

```
GET /FeatureGroup/FeatureGroupName?
ExpirationTimeResponse=ExpirationTimeResponse&FeatureName=FeatureNames&RecordIdentifierValueAsString=RecordIdentifierValueAsString
HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

**ExpirationTimeResponse (p. 1122)**
Parameter to request ExpiresAt in response. If Enabled, GetRecord will return the value of ExpiresAt, if it is not null. If Disabled and null, GetRecord will return null.

Valid Values: Enabled | Disabled

**FeatureGroupName (p. 1122)**

The name or Amazon Resource Name (ARN) of the feature group from which you want to retrieve a record.

Length Constraints: Minimum length of 1. Maximum length of 150.

Pattern: (arn:aws[a-zA-Z-]*:sagemaker:[a-zA-Z0-9-]*:[0-9]{12}:feature-group/)?([a-zA-Z0-9_]*[a-zA-Z0-9])

Required: Yes

**FeatureNames (p. 1122)**

List of names of Features to be retrieved. If not specified, the latest value for all the Features are returned.

Array Members: Minimum number of 1 item.

Length Constraints: Minimum length of 1. Maximum length of 64.

Pattern: ^[a-zA-Z0-9][__]*[a-zA-Z0-9]{0,63}

**RecordIdentifierValueAsString (p. 1122)**

The value that corresponds to RecordIdentifier type and uniquely identifies the record in the FeatureGroup.

Length Constraints: Maximum length of 358400.

Pattern: .*

Required: Yes

Request Body

The request does not have a request body.
Response Syntax

HTTP/1.1 200
Content-type: application/json

{
  "ExpiresAt": "string",
  "Record": [
    {
      "FeatureName": "string",
      "ValueAsString": "string",
      "ValueAsStringList": [ "string" ]
    }
  ]
}

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

ExpiresAt (p. 1123)

The ExpiresAt ISO string of the requested record.

Type: String

Record (p. 1123)

The record you requested. A list of FeatureValues.

Type: Array of FeatureValue (p. 2106) objects

Array Members: Minimum number of 1 item.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

AccessForbidden

You do not have permission to perform an action.

HTTP Status Code: 403

InternalFailure

An internal failure occurred. Try your request again. If the problem persists, contact AWS customer support.

HTTP Status Code: 500

ResourceNotFoundException

A resource that is required to perform an action was not found.

HTTP Status Code: 404

ServiceUnavailable

The service is currently unavailable.
HTTP Status Code: 503

**ValidationError**

There was an error validating your request.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
**PutRecord**
Service: Amazon SageMaker Feature Store Runtime

The PutRecord API is used to ingest a list of Records into your feature group.

If a new record's EventTime is greater, the new record is written to both the OnlineStore and OfflineStore. Otherwise, the record is a historic record and it is written only to the OfflineStore.

You can specify the ingestion to be applied to the OnlineStore, OfflineStore, or both by using the TargetStores request parameter.

You can set the ingested record to expire at a given time to live (TTL) duration after the record's event time, ExpiresAt = EventTime + TtlDuration, by specifying the TtlDuration parameter. A record level TtlDuration is set when specifying the TtlDuration parameter using the PutRecord API call. If the input TtlDuration is null or unspecified, TtlDuration is set to the default feature group level TtlDuration. A record level TtlDuration supersedes the group level TtlDuration.

**Request Syntax**

```
PUT /FeatureGroup/FeatureGroupName HTTP/1.1
Content-type: application/json

{
  "Record": [
    {
      "FeatureName": "string",
      "ValueAsString": "string",
      "ValueAsStringList": [ "string" ]
    }
  ],
  "TargetStores": [ "string" ],
  "TtlDuration": {
    "Unit": "string",
    "Value": number
  }
}
```

**URI Request Parameters**

The request uses the following URI parameters.

**FeatureGroupName (p. 1125)**

The name or Amazon Resource Name (ARN) of the feature group that you want to insert the record into.

Length Constraints: Minimum length of 1. Maximum length of 150.

Pattern: (arn:aws[a-z-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:feature-group/)?([a-zA-Z0-9][-_]*[a-zA-Z0-9]{0,63})

Required: Yes

**Request Body**

The request accepts the following data in JSON format.
**Record (p. 1125)**

List of FeatureValues to be inserted. This will be a full over-write. If you only want to update few of the feature values, do the following:

- Use `GetRecord` to retrieve the latest record.
- Update the record returned from `GetRecord`.
- Use `PutRecord` to update feature values.

Type: Array of [FeatureValue (p. 2106)] objects

Array Members: Minimum number of 1 item.

Required: Yes

**TargetStores (p. 1125)**

A list of stores to which you're adding the record. By default, Feature Store adds the record to all of the stores that you're using for the [FeatureGroup](#).

Type: Array of strings

Array Members: Minimum number of 1 item. Maximum number of 2 items.

Valid Values: OnlineStore | OfflineStore

Required: No

**TtlDuration (p. 1125)**

Time to live duration, where the record is hard deleted after the expiration time is reached; ExpiresAt = EventTime + TtlDuration. For information on HardDelete, see the DeleteRecord API in the Amazon SageMaker API Reference guide.

Type: [TtlDuration (p. 2107)] object

Required: No

### Response Syntax

```
HTTP/1.1 200
```

### Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

### Errors

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](#).

**AccessForbidden**

You do not have permission to perform an action.

HTTP Status Code: 403

**InternalFailure**

An internal failure occurred. Try your request again. If the problem persists, contact AWS customer support.
HTTP Status Code: 500

**ServiceUnavailable**

The service is currently unavailable.

HTTP Status Code: 503

**ValidationError**

There was an error validating your request.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3

### Amazon SageMaker geospatial capabilities

The following actions are supported by Amazon SageMaker geospatial capabilities:

- [DeleteEarthObservationJob](#) (p. 1129)
- [DeleteVectorEnrichmentJob](#) (p. 1131)
- [ExportEarthObservationJob](#) (p. 1133)
- [ExportVectorEnrichmentJob](#) (p. 1137)
- [GetEarthObservationJob](#) (p. 1141)
- [GetRasterDataCollection](#) (p. 1145)
- [GetTile](#) (p. 1148)
- [GetVectorEnrichmentJob](#) (p. 1151)
- [ListEarthObservationJobs](#) (p. 1155)
- [ListRasterDataCollections](#) (p. 1158)
- [ListTagsForResource](#) (p. 1161)
- [ListVectorEnrichmentJobs](#) (p. 1163)
- [SearchRasterDataCollection](#) (p. 1166)
- [StartEarthObservationJob](#) (p. 1169)
- [StartVectorEnrichmentJob](#) (p. 1174)
- [StopEarthObservationJob](#) (p. 1179)
- [StopVectorEnrichmentJob](#) (p. 1181)
- [TagResource](#) (p. 1183)
• **UntagResource (p. 1185)**
DeleteEarthObservationJob

Service: Amazon SageMaker geospatial capabilities

Use this operation to delete an Earth Observation job.

Request Syntax

DELETE /earth-observation-jobs/<Arn> HTTP/1.1

URI Request Parameters

The request uses the following URI parameters.

Arn (p. 1129)

The Amazon Resource Name (ARN) of the Earth Observation job being deleted.

Pattern: ^arn:aws[a-z-]{0,12}:sagemaker-geospatial:[a-z0-9-]{1,25}:[0-9]{12}:earth-observation-job/[a-z0-9]{12}\$

Required: Yes

Request Body

The request does not have a request body.

Response Syntax

HTTP/1.1 200

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

AccessDeniedException

You do not have sufficient access to perform this action.

HTTP Status Code: 403

ConflictException

Updating or deleting a resource can cause an inconsistent state.

HTTP Status Code: 409

InternalServerException

The request processing has failed because of an unknown error, exception, or failure.

HTTP Status Code: 500
ResourceNotFoundException

The request references a resource which does not exist.
HTTP Status Code: 404

ThrottlingException

The request was denied due to request throttling.
HTTP Status Code: 429

ValidationException

The input fails to satisfy the constraints specified by an AWS service.
HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

• AWS Command Line Interface
• AWS SDK for .NET
• AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
DeleteVectorEnrichmentJob
Service: Amazon SageMaker geospatial capabilities
Use this operation to delete a Vector Enrichment job.

Request Syntax

DELETE /vector-enrichment-jobs/<Arn> HTTP/1.1

URI Request Parameters

The request uses the following URI parameters.

Arn (p. 1131)

The Amazon Resource Name (ARN) of the Vector Enrichment job being deleted.

Pattern: ^arn:aws[a-z-]{0,12}:sagemaker-geospatial:[a-z0-9-]{1,25}:[0-9]{12}:vector-enrichment-job/[a-z0-9]{12},$

Required: Yes

Request Body

The request does not have a request body.

Response Syntax

HTTP/1.1 200

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

AccessDeniedException

You do not have sufficient access to perform this action.

HTTP Status Code: 403

ConflictException

Updating or deleting a resource can cause an inconsistent state.

HTTP Status Code: 409

InternalServerException

The request processing has failed because of an unknown error, exception, or failure.

HTTP Status Code: 500
**ResourceNotFoundException**

The request references a resource which does not exist.

HTTP Status Code: 404

**ThrottlingException**

The request was denied due to request throttling.

HTTP Status Code: 429

**ValidationException**

The input fails to satisfy the constraints specified by an AWS service.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ExportEarthObservationJob
Service: Amazon SageMaker geospatial capabilities

Use this operation to export results of an Earth Observation job and optionally source images used as input to the EOJ to an Amazon S3 location.

Request Syntax

```javascript
POST /export-earth-observation-job HTTP/1.1
Content-type: application/json

{
   "Arn": "string",
   "ClientToken": "string",
   "ExecutionRoleArn": "string",
   "ExportSourceImages": boolean,
   "OutputConfig": {
      "S3Data": {
         "KmsKeyId": "string",
         "S3Uri": "string"
      }
   }
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

**Arn (p. 1133)**

The input Amazon Resource Name (ARN) of the Earth Observation job being exported.

Type: String

Pattern: `arn:aws[a-z-][0,12]:sagemaker-geospatial:[a-z0-9-]{1,25}:[0-9]{12}:earth-observation-job/[a-z0-9]{12}$`

Required: Yes

**ClientToken (p. 1133)**

A unique token that guarantees that the call to this API is idempotent.

Type: String

Length Constraints: Minimum length of 36. Maximum length of 64.

Required: No

**ExecutionRoleArn (p. 1133)**

The Amazon Resource Name (ARN) of the IAM role that you specified for the job.

Type: String

ExportEarthObservationJob

Pattern: ^arn:(aws[a-z-]*):iam::([0-9]{12}):role/[a-zA-Z0-9+=,.@_/-]+$ 
Required: Yes

**ExportSourceImages (p. 1133)**

The source images provided to the Earth Observation job being exported.

Type: Boolean
Required: No

**OutputConfig (p. 1133)**

An object containing information about the output file.

Type: [OutputConfigInput (p. 2141)] object
Required: Yes

**Response Syntax**

```
HTTP/1.1 200
Content-type: application/json

{
    "Arn": "string",
    "CreationTime": "string",
    "ExecutionRoleArn": "string",
    "ExportSourceImages": boolean,
    "ExportStatus": "string",
    "OutputConfig": {
        "S3Data": {
            "KmsKeyId": "string",
            "S3Uri": "string"
        }
    }
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**Arn (p. 1134)**

The output Amazon Resource Name (ARN) of the Earth Observation job being exported.

Type: String

Pattern: ^arn:aws[a-z-]{0,12}:sagemaker-geospatial:[a-z0-9-]{1,25}:[0-9]{12}:earth-observation-job/[a-z0-9]{12}$

**CreationTime (p. 1134)**

The creation time.

Type: Timestamp

**ExecutionRoleArn (p. 1134)**

The Amazon Resource Name (ARN) of the IAM role that you specified for the job.
Type: String


Pattern: ^arn:(aws-[a-z-]*)\:iam::([0-9]{12})\:role/[a-zA-Z0-9+=,.@_/-]+$

**ExportSourceImages (p. 1134)**

The source images provided to the Earth Observation job being exported.

Type: Boolean

**ExportStatus (p. 1134)**

The status of the results of the Earth Observation job being exported.

Type: String

Valid Values: IN_PROGRESS | SUCCEEDED | FAILED

**OutputConfig (p. 1134)**

An object containing information about the output file.

Type: [OutputConfigInput (p. 2141)](p. 2141) object

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](p. 2180).

**AccessDeniedException**

You do not have sufficient access to perform this action.

HTTP Status Code: 403

**ConflictException**

Updating or deleting a resource can cause an inconsistent state.

HTTP Status Code: 409

**InternalServerException**

The request processing has failed because of an unknown error, exception, or failure.

HTTP Status Code: 500

**ResourceNotFoundException**

The request references a resource which does not exist.

HTTP Status Code: 404

**ServiceQuotaExceededException**

You have exceeded the service quota.

HTTP Status Code: 402

**ThrottlingException**

The request was denied due to request throttling.

HTTP Status Code: 429
ValidationException

The input fails to satisfy the constraints specified by an AWS service.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ExportVectorEnrichmentJob

Service: Amazon SageMaker geospatial capabilities

Use this operation to copy results of a Vector Enrichment job to an Amazon S3 location.

Request Syntax

```
POST /export-vector-enrichment-jobs HTTP/1.1
Content-type: application/json

{
   "Arn": "string",
   "ClientToken": "string",
   "ExecutionRoleArn": "string",
   "OutputConfig": {
      "$S3Data": {
         "KmsKeyId": "string",
         "$S3Uri": "string"
      }
   }
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

**Arn (p. 1137)**

The Amazon Resource Name (ARN) of the Vector Enrichment job.

Type: String

Pattern: ^arn:aws[a-z-]{0,12}:sagemaker-geospatial:[a-z0-9-]{1,25}:[0-9]{12}:vector-enrichment-job/[a-z0-9]{12},$  

Required: Yes

**ClientToken (p. 1137)**

A unique token that guarantees that the call to this API is idempotent.

Type: String

Length Constraints: Minimum length of 36. Maximum length of 64.

Required: No

**ExecutionRoleArn (p. 1137)**

The Amazon Resource Name (ARN) of the IAM role with permission to upload to the location in OutputConfig.

Type: String

Pattern: ^arn:(aws[a-z-]*):iam::([0-9]{12}):role/[a-zA-Z0-9-+=,.@_/-]+$  
Required: Yes

**OutputConfig (p. 1137)**

Output location information for exporting Vector Enrichment Job results.

Type: ExportVectorEnrichmentJobOutputConfig (p. 2122) object

Required: Yes

**Response Syntax**

```json
HTTP/1.1 200
Content-type: application/json

{
  "Arn": "string",
  "CreationTime": "string",
  "ExecutionRoleArn": "string",
  "ExportStatus": "string",
  "OutputConfig": {
    "S3Data": {
      "KmsKeyId": "string",
      "S3Uri": "string"
    }
  }
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**Arn (p. 1138)**

The Amazon Resource Name (ARN) of the Vector Enrichment job being exported.

Type: String

Pattern: ^arn:aws[a-z-][0,12]:sagemaker-geospatial:[a-z0-9-]{1,25}:[0-9]{12}:vector-enrichment-job/[a-z0-9]{{12,}}$

**CreationTime (p. 1138)**

The creation time.

Type: Timestamp

**ExecutionRoleArn (p. 1138)**

The Amazon Resource Name (ARN) of the IAM role with permission to upload to the location in OutputConfig.

Type: String


Pattern: ^arn:(aws[a-z-]*):iam::([0-9]{12}):role/[a-zA-Z0-9-+=,.@_/-]+$
ExportStatus (p. 1138)

The status of the results the Vector Enrichment job being exported.

Type: String

Valid Values: IN_PROGRESS | SUCCEEDED | FAILED

OutputConfig (p. 1138)

Output location information for exporting Vector Enrichment Job results.

Type: ExportVectorEnrichmentJobOutputConfig (p. 2122) object

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

AccessDeniedException

You do not have sufficient access to perform this action.

HTTP Status Code: 403

ConflictException

Updating or deleting a resource can cause an inconsistent state.

HTTP Status Code: 409

InternalServerErrorException

The request processing has failed because of an unknown error, exception, or failure.

HTTP Status Code: 500

ResourceNotFoundException

The request references a resource which does not exist.

HTTP Status Code: 404

ServiceQuotaExceededException

You have exceeded the service quota.

HTTP Status Code: 402

ThrottlingException

The request was denied due to request throttling.

HTTP Status Code: 429

ValidationException

The input fails to satisfy the constraints specified by an AWS service.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:
- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
GetEarthObservationJob
Service: Amazon SageMaker geospatial capabilities
Get the details for a previously initiated Earth Observation job.

Request Syntax

GET /earth-observation-jobs/{Arn} HTTP/1.1

URI Request Parameters
The request uses the following URI parameters.

Arn (p. 1141)
The Amazon Resource Name (ARN) of the Earth Observation job.
Pattern: ^arn:aws[a-z-]\{0,12\}:sagemaker-geospatial:[a-zA-Z-]\{1,25\}:[0-9] {12}:earth-observation-job/[a-z0-9]\{12\}$
Required: Yes

Request Body
The request does not have a request body.

Response Syntax

HTTP/1.1 200
Content-type: application/json

{  
  "Arn": "string",
  "CreationTime": "string",
  "DurationInSeconds": number,
  "ErrorDetails": {
    "Message": "string",
    "Type": "string"
  },
  "ExecutionRoleArn": "string",
  "ExportErrorDetails": {
    "ExportResults": {
      "Message": "string",
      "Type": "string"
    },
    "ExportSourceImages": {
      "Message": "string",
      "Type": "string"
    }
  },
  "ExportStatus": "string",
  "InputConfig": {
    "PreviousEarthObservationJobArn": "string",
    "RasterDataCollectionQuery": {
      "AreaOfInterest": {... },
      "PropertyFilters": {
        "LogicalOperator": "string",
        "Properties": [  
      ]
    }
  }
}
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**Arn (p. 1141)**

The Amazon Resource Name (ARN) of the Earth Observation job.

Type: String

**CreationTime (p. 1141)**

The creation time of the initiated Earth Observation job.

Type: Timestamp

**DurationInSeconds (p. 1141)**

The duration of Earth Observation job, in seconds.

Type: Integer

**ErrorDetails (p. 1141)**

Details about the errors generated during the Earth Observation job.

Type: EarthObservationJobErrorDetails (p. 2117) object

**ExecutionRoleArn (p. 1141)**

The Amazon Resource Name (ARN) of the IAM role that you specified for the job.

Type: String

Pattern: ^arn:(aws[a-z-]*)iam::([0-9]{12}):role/[a-zA-Z0-9+=,.@_/-]+$  

**ExportErrorDetails (p. 1141)**  
Details about the errors generated during ExportEarthObservationJob.  
Type: `ExportErrorDetails (p. 2119)` object  

**ExportStatus (p. 1141)**  
The status of the Earth Observation job.  
Type: String  
Valid Values: IN_PROGRESS | SUCCEEDED | FAILED  

**InputConfig (p. 1141)**  
Input data for the Earth Observation job.  
Type: `InputConfigOutput (p. 2127)` object  

**JobConfig (p. 1141)**  
An object containing information about the job configuration.  
Type: `JobConfigInput (p. 2129)` object  

**Note:** This object is a Union. Only one member of this object can be specified or returned.  

**KmsKeyId (p. 1141)**  
The Key Management Service key ID for server-side encryption.  
Type: String  
Length Constraints: Minimum length of 0. Maximum length of 2048.  

**Name (p. 1141)**  
The name of the Earth Observation job.  
Type: String  

**OutputBands (p. 1141)**  
Bands available in the output of an operation.  
Type: Array of `OutputBand (p. 2140)` objects  

**Status (p. 1141)**  
The status of a previously initiated Earth Observation job.  
Type: String  
Valid Values: INITIALIZING | IN_PROGRESS | STOPPING | COMPLETED | STopped | FAILED | DELETING | DELETED  

**Tags (p. 1141)**  
Each tag consists of a key and a value.  
Type: String to string map  

**Errors**  
For information about the errors that are common to all actions, see `Common Errors (p. 2180)`.
AccessDeniedException

You do not have sufficient access to perform this action.

HTTP Status Code: 403

InternalServerException

The request processing has failed because of an unknown error, exception, or failure.

HTTP Status Code: 500

ResourceNotFoundException

The request references a resource which does not exist.

HTTP Status Code: 404

ThrottlingException

The request was denied due to request throttling.

HTTP Status Code: 429

ValidationException

The input fails to satisfy the constraints specified by an AWS service.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
GetRasterDataCollection
Service: Amazon SageMaker geospatial capabilities

Use this operation to get details of a specific raster data collection.

Request Syntax

GET /raster-data-collection/Arn HTTP/1.1

URI Request Parameters

The request uses the following URI parameters.

Arn (p. 1145)

The Amazon Resource Name (ARN) of the raster data collection.

Pattern: ^arn:aws[a-z-]{0,12}:sagemaker-geospatial:[a-z0-9-]{1,25}:[0-9]{12}:raster-data-collection/(public|premium|user)/[a-z0-9]{12,}$

Required: Yes

Request Body

The request does not have a request body.

Response Syntax

HTTP/1.1 200
Content-type: application/json

{
    "Arn": "string",
    "Description": "string",
    "DescriptionPageUrl": "string",
    "ImageSourceBands": [ "string" ],
    "Name": "string",
    "SupportedFilters": [ {
        "Maximum": number,
        "Minimum": number,
        "Name": "string",
        "Type": "string"
    } ],
    "Tags": {
        "string" : "string"
    },
    "Type": "string"
}

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.
Arn (p. 1145)
The Amazon Resource Name (ARN) of the raster data collection.

Type: String

Pattern: ^arn:aws[a-z-]{0,12}:sagemaker-geospatial:[a-z0-9-]{1,25}:[0-9]{12}:raster-data-collection/(public|premium|user)/[a-z0-9]{12,}$

Description (p. 1145)
A description of the raster data collection.

Type: String

DescriptionPageUrl (p. 1145)
The URL of the description page.

Type: String

ImageSourceBands (p. 1145)
The list of image source bands in the raster data collection.

Type: Array of strings

Name (p. 1145)
The name of the raster data collection.

Type: String

SupportedFilters (p. 1145)
The filters supported by the raster data collection.

Type: Array of Filter (p. 2123) objects

Tags (p. 1145)
Each tag consists of a key and a value.

Type: String to string map

Type (p. 1145)
The raster data collection type.

Type: String

Valid Values: PUBLIC | PREMIUM | USER

Errors
For information about the errors that are common to all actions, see Common Errors (p. 2180).

AccessDeniedException
You do not have sufficient access to perform this action.

HTTP Status Code: 403

InternalServerException
The request processing has failed because of an unknown error, exception, or failure.
HTTP Status Code: 500

**ResourceNotFoundException**

The request references a resource which does not exist.

HTTP Status Code: 404

**ThrottlingException**

The request was denied due to request throttling.

HTTP Status Code: 429

**ValidationException**

The input fails to satisfy the constraints specified by an AWS service.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
GetTile
Service: Amazon SageMaker geospatial capabilities

Gets a web mercator tile for the given Earth Observation job.

Request Syntax

```
GET /tile/\{\z\}/\{\x\}/\{\y\}?
Arn=Arn&ExecutionRoleArn=ExecutionRoleArn&ImageAssets=ImageAssets&ImageMask=ImageMask&OutputDataType=OutputDataType&OutputFormat=OutputFormat&PropertyFilters=PropertyFilters&Target=Target&TimeRangeFilter=TimeRangeFilter
```

HTTP/1.1

URI Request Parameters

The request uses the following URI parameters.

**Arn (p. 1148)**

The Amazon Resource Name (ARN) of the tile operation.

Pattern: `arn:aws[a-z][0,12]:sagemaker-geospatial:[a-z0-9-][1,25]:[0-9] {12}:earth-observation-job/[a-z0-9]{12,}$`

Required: Yes

**ExecutionRoleArn (p. 1148)**

The Amazon Resource Name (ARN) of the IAM role that you specify.


Pattern: `arn:(aws[a-z-]*):iam::([0-9]{12}):role/[a-zA-Z0-9-+,.@_/-]+$`

**ImageAssets (p. 1148)**

The particular assets or bands to tile.

Array Members: Minimum number of 1 item.

Required: Yes

**ImageMask (p. 1148)**

Determines whether or not to return a valid data mask.

**OutputDataType (p. 1148)**

The output data type of the tile operation.

Valid Values: INT32 | FLOAT32 | INT16 | FLOAT64 | UINT16

**OutputFormat (p. 1148)**

The data format of the output tile. The formats include .npy, .png and .jpg.

**PropertyFilters (p. 1148)**

Property filters for the imagery to tile.

**Target (p. 1148)**

Determines what part of the Earth Observation job to tile. 'INPUT' or 'OUTPUT' are the valid options.

Valid Values: INPUT | OUTPUT
GetTile

Required: Yes

**TimeRangeFilter (p. 1148)**

Time range filter applied to imagery to find the images to tile.

**x (p. 1148)**

The x coordinate of the tile input.

Required: Yes

**y (p. 1148)**

The y coordinate of the tile input.

Required: Yes

**z (p. 1148)**

The z coordinate of the tile input.

Required: Yes

**Request Body**

The request does not have a request body.

**Response Syntax**

```
HTTP/1.1 200

BinaryFile
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The response returns the following as the HTTP body.

**BinaryFile (p. 1149)**

The output binary file.

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](https://aws.amazon.com/documentation/sagemaker/api-reference/).

**AccessDeniedException**

You do not have sufficient access to perform this action.

HTTP Status Code: 403

**InternalServerException**

The request processing has failed because of an unknown error, exception, or failure.

HTTP Status Code: 500
ResourceNotFoundException

The request references a resource which does not exist.

HTTP Status Code: 404

ThrottlingException

The request was denied due to request throttling.

HTTP Status Code: 429

ValidationException

The input fails to satisfy the constraints specified by an AWS service.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
GetVectorEnrichmentJob

Service: Amazon SageMaker geospatial capabilities

Retrieves details of a Vector Enrichment Job for a given job Amazon Resource Name (ARN).

Request Syntax

GET /vector-enrichment-jobs/Arn HTTP/1.1

URI Request Parameters

The request uses the following URI parameters.

Arn (p. 1151)

The Amazon Resource Name (ARN) of the Vector Enrichment job.

Pattern: `arn:aws[a-z-]{0,12}:sagemaker-geospatial:[a-z0-9-]{1,25}:[0-9]\{12\}:vector-enrichment-job/[a-z0-9]\{12\}$`

Required: Yes

Request Body

The request does not have a request body.

Response Syntax

HTTP/1.1 200
Content-type: application/json

```json
{
   "Arn": "string",
   "CreationTime": "string",
   "DurationInSeconds": number,
   "ErrorDetails": {
      "ErrorMessage": "string",
      "ErrorType": "string"
   },
   "ExecutionRoleArn": "string",
   "ExportErrorDetails": {
      "Message": "string",
      "Type": "string"
   },
   "ExportStatus": "string",
   "InputConfig": {
      "DataSourceConfig": { ... },
      "DocumentType": "string"
   },
   "JobConfig": { ... },
   "KmsKeyId": "string",
   "Name": "string",
   "Status": "string",
   "Tags": {
      "string": "string"
   },
   "Type": "string"
}
```
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

Arn (p. 1151)

The Amazon Resource Name (ARN) of the Vector Enrichment job.

Type: String

CreationTime (p. 1151)

The creation time.

Type: Timestamp

DurationInSeconds (p. 1151)

The duration of the Vector Enrichment job, in seconds.

Type: Integer

ErrorDetails (p. 1151)

Details about the errors generated during the Vector Enrichment job.

Type: VectorEnrichmentJobErrorDetails (p. 2167) object

ExecutionRoleArn (p. 1151)

The Amazon Resource Name (ARN) of the IAM role that you specified for the job.

Type: String


Pattern: ^arn:(aws[a-z-]*):iam::([0-9]{12}):role/[a-zA-Z0-9-9+,.@_/-]+$ ExportErrorDetails (p. 1151)

Details about the errors generated during the ExportVectorEnrichmentJob.

Type: VectorEnrichmentJobExportErrorDetails (p. 2168) object

ExportStatus (p. 1151)

The export status of the Vector Enrichment job being initiated.

Type: String

Valid Values: IN_PROGRESS | SUCCEEDED | FAILED

InputConfig (p. 1151)

Input configuration information for the Vector Enrichment job.

Type: VectorEnrichmentJobInputConfig (p. 2169) object

JobConfig (p. 1151)

An object containing information about the job configuration.

Type: VectorEnrichmentJobConfig (p. 2165) object
Note: This object is a Union. Only one member of this object can be specified or returned.

KmsKeyId (p. 1151)

The Key Management Service key ID for server-side encryption.
Type: String
Length Constraints: Minimum length of 0. Maximum length of 2048.

Name (p. 1151)

The name of the Vector Enrichment job.
Type: String

Status (p. 1151)

The status of the initiated Vector Enrichment job.
Type: String
Valid Values: INITIALIZING | IN_PROGRESS | STOPPING | STOPPED | COMPLETED | FAILED | DELETING | DELETED

Tags (p. 1151)

Each tag consists of a key and a value.
Type: String to string map

Type (p. 1151)

The type of the Vector Enrichment job being initiated.
Type: String
Valid Values: REVERSE_GEOCODING | MAP_MATCHING

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

AccessDeniedException

You do not have sufficient access to perform this action.
HTTP Status Code: 403

InternalServerException

The request processing has failed because of an unknown error, exception, or failure.
HTTP Status Code: 500

ResourceNotFoundException

The request references a resource which does not exist.
HTTP Status Code: 404

ThrottlingException

The request was denied due to request throttling.
HTTP Status Code: 429
ValidationException

The input fails to satisfy the constraints specified by an AWS service.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListEarthObservationJobs
Service: Amazon SageMaker geospatial capabilities

Use this operation to get a list of the Earth Observation jobs associated with the calling AWS account.

Request Syntax

```plaintext
POST /list-earth-observation-jobs HTTP/1.1
Content-type: application/json

{
    "MaxResults": number,
    "NextToken": "string",
    "SortBy": "string",
    "SortOrder": "string",
    "StatusEquals": "string"
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

MaxResults (p. 1155)

The total number of items to return.

Type: Integer


Required: No

NextToken (p. 1155)

If the previous response was truncated, you receive this token. Use it in your next request to receive the next set of results.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 8192.

Required: No

SortBy (p. 1155)

The parameter by which to sort the results.

Type: String

Required: No

SortOrder (p. 1155)

An optional value that specifies whether you want the results sorted in Ascending or Descending order.

Type: String
Valid Values: ASCENDING | DESCENDING

Required: No

**StatusEquals (p. 1155)**

A filter that retrieves only jobs with a specific status.

Type: String

Valid Values: INITIALIZING | IN_PROGRESS | STOPPING | COMPLETED | STOPPED | FAILED | DELETING | DELETED

Required: No

### Response Syntax

HTTP/1.1 200
Content-type: application/json

```json
{
  "EarthObservationJobSummaries": [
    {
      "Arn": "string",
      "CreationTime": "string",
      "DurationInSeconds": number,
      "Name": "string",
      "OperationType": "string",
      "Status": "string",
      "Tags": {
        "string": "string"
      }
    }
  ],
  "NextToken": "string"
}
```

### Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**EarthObservationJobSummaries (p. 1156)**

Contains summary information about the Earth Observation jobs.

Type: Array of ListEarthObservationJobOutputConfig (p. 2133) objects

**NextToken (p. 1156)**

If the previous response was truncated, you receive this token. Use it in your next request to receive the next set of results.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 8192.

### Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).
AccessDeniedException

You do not have sufficient access to perform this action.

HTTP Status Code: 403

InternalServerException

The request processing has failed because of an unknown error, exception, or failure.

HTTP Status Code: 500

ResourceNotFoundException

The request references a resource which does not exist.

HTTP Status Code: 404

ThrottlingException

The request was denied due to request throttling.

HTTP Status Code: 429

ValidationException

The input fails to satisfy the constraints specified by an AWS service.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListRasterDataCollections
Service: Amazon SageMaker geospatial capabilities

Use this operation to get raster data collections.

Request Syntax

GET /raster-data-collections?MaxResults=MaxResults&NextToken=NextToken HTTP/1.1

URI Request Parameters

The request uses the following URI parameters.

**MaxResults (p. 1158)**

The total number of items to return.


**NextToken (p. 1158)**

If the previous response was truncated, you receive this token. Use it in your next request to receive the next set of results.

Length Constraints: Minimum length of 0. Maximum length of 8192.

Request Body

The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
    "NextToken": "string",
    "RasterDataCollectionSummaries": [
        {
            "Arn": "string",
            "Description": "string",
            "DescriptionPageUrl": "string",
            "Name": "string",
            "SupportedFilters": [
                {
                    "Maximum": number,
                    "Minimum": number,
                    "Name": "string",
                    "Type": "string"
                }
            ],
            "Tags": {
                "string": "string"
            },
            "Type": "string"
        }
    ]
}
```
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**NextToken (p. 1158)**

If the previous response was truncated, you receive this token. Use it in your next request to receive the next set of results.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 8192.

**RasterDataCollectionSummaries (p. 1158)**

Contains summary information about the raster data collection.

Type: Array of **RasterDataCollectionMetadata (p. 2152)** objects

Errors

For information about the errors that are common to all actions, see [Common Errors (p. 2180)].

**AccessDeniedException**

You do not have sufficient access to perform this action.

HTTP Status Code: 403

**InternalServerException**

The request processing has failed because of an unknown error, exception, or failure.

HTTP Status Code: 500

**ResourceNotFoundException**

The request references a resource which does not exist.

HTTP Status Code: 404

**ThrottlingException**

The request was denied due to request throttling.

HTTP Status Code: 429

**ValidationException**

The input fails to satisfy the constraints specified by an AWS service.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
• AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for JavaScript V3
• AWS SDK for PHP V3
• AWS SDK for Python
• AWS SDK for Ruby V3
ListTagsForResource

Service: Amazon SageMaker geospatial capabilities
Lists the tags attached to the resource.

Request Syntax

GET /tags/{ResourceArn} HTTP/1.1

URI Request Parameters

The request uses the following URI parameters.

ResourceArn (p. 1161)

The Amazon Resource Name (ARN) of the resource you want to tag.

Length Constraints: Minimum length of 1. Maximum length of 2048.

Required: Yes

Request Body

The request does not have a request body.

Response Syntax

HTTP/1.1 200
Content-type: application/json

{
    "Tags": {
        "String": "string"
    }
}

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

Tags (p. 1161)

Each tag consists of a key and a value.

Type: String to string map

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

AccessDeniedException

You do not have sufficient access to perform this action.
HTTP Status Code: 403
*InternalServerException*

The request processing has failed because of an unknown error, exception, or failure.

HTTP Status Code: 500
*ResourceNotFoundException*

The request references a resource which does not exist.

HTTP Status Code: 404
*ThrottlingException*

The request was denied due to request throttling.

HTTP Status Code: 429
*ValidationException*

The input fails to satisfy the constraints specified by an AWS service.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
ListVectorEnrichmentJobs
Service: Amazon SageMaker geospatial capabilities

Retrieves a list of vector enrichment jobs.

Request Syntax

POST /list-vector-enrichment-jobs HTTP/1.1
Content-type: application/json

{
   "MaxResults": number,
   "NextToken": "string",
   "SortBy": "string",
   "SortOrder": "string",
   "StatusEquals": "string"
}

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

MaxResults (p. 1163)

The maximum number of items to return.

Type: Integer


Required: No

NextToken (p. 1163)

If the previous response was truncated, you receive this token. Use it in your next request to receive the next set of results.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 8192.

Required: No

SortBy (p. 1163)

The parameter by which to sort the results.

Type: String

Required: No

SortOrder (p. 1163)

An optional value that specifies whether you want the results sorted in Ascending or Descending order.
Type: String

Valid Values: ASCENDING | DESCENDING

Required: No

**StatusEquals (p. 1163)**

A filter that retrieves only jobs with a specific status.

Type: String

Required: No

**Response Syntax**

```
HTTP/1.1 200
Content-type: application/json

{
    "NextToken": "string",
    "VectorEnrichmentJobSummaries": [
        {
            "Arn": "string",
            "CreationTime": "string",
            "DurationInSeconds": number,
            "Name": "string",
            "Status": "string",
            "Tags": {
                "string": "string"
            },
            "Type": "string"
        }
    ]
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**NextToken (p. 1164)**

If the previous response was truncated, you receive this token. Use it in your next request to receive the next set of results.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 8192.

**VectorEnrichmentJobSummaries (p. 1164)**

Contains summary information about the Vector Enrichment jobs.

Type: Array of ListVectorEnrichmentJobOutputConfig (p. 2135) objects

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)](Common Errors (p. 2180)).
AccessDeniedException
You do not have sufficient access to perform this action.
HTTP Status Code: 403

InternalServerException
The request processing has failed because of an unknown error, exception, or failure.
HTTP Status Code: 500

ResourceNotFoundException
The request references a resource which does not exist.
HTTP Status Code: 404

ThrottlingException
The request was denied due to request throttling.
HTTP Status Code: 429

ValidationException
The input fails to satisfy the constraints specified by an AWS service.
HTTP Status Code: 400

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
**SearchRasterDataCollection**

Service: Amazon SageMaker geospatial capabilities

Allows you run image query on a specific raster data collection to get a list of the satellite imagery matching the selected filters.

**Request Syntax**

```
POST /search-raster-data-collection HTTP/1.1
Content-type: application/json

{
    "Arn": "string",
    "NextToken": "string",
    "RasterDataCollectionQuery": {
        "AreaOfInterest": { ... },
        "BandFilter": [ "string" ],
        "PropertyFilters": {
            "LogicalOperator": "string",
            "Properties": [
                { "Property": { ... } }
            ]
        },
        "TimeRangeFilter": {
            "EndTime": number,
            "StartTime": number
        }
    }
}
```

**URI Request Parameters**

The request does not use any URI parameters.

**Request Body**

The request accepts the following data in JSON format.

**Arn (p. 1166)**

The Amazon Resource Name (ARN) of the raster data collection.

Type: String

Pattern: ^arn:aws[a-z-][0,12]:sagemaker-geospatial:[a-z0-9-]{1,25}:[0-9]{12}:raster-data-collection/(public|premium|user)/[a-z0-9]{12}$

Required: Yes

**NextToken (p. 1166)**

If the previous response was truncated, you receive this token. Use it in your next request to receive the next set of results.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 8192.

Required: No
**Response Syntax**

```
HTTP/1.1 200
Content-type: application/json

{
  "ApproximateResultCount": number,
  "Items": [
    {
      "Assets": {
        "Href": "string"
      },
      "DateTime": number,
      "Geometry": {
        "Coordinates": [
          [ number ]
        ],
        "Type": "string"
      },
      "Id": "string",
      "Properties": {
        "EoCloudCover": number,
        "LandsatCloudCoverLand": number,
        "Platform": "string",
        "ViewOffNadir": number,
        "ViewSunAzimuth": number,
        "ViewSunElevation": number
      }
    }
  ],
  "NextToken": "string"
}
```

**Response Elements**

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**ApproximateResultCount (p. 1167)**

Approximate number of results in the response.

Type: Integer

**Items (p. 1167)**

List of items matching the RasterDataCollectionQuery.
Type: Array of `ItemSource (p. 2128)` objects

**NextToken (p. 1167)**

If the previous response was truncated, you receive this token. Use it in your next request to receive the next set of results.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 8192.

**Errors**

For information about the errors that are common to all actions, see [Common Errors (p. 2180)].

**AccessDeniedException**

You do not have sufficient access to perform this action.

HTTP Status Code: 403

**InternalServerException**

The request processing has failed because of an unknown error, exception, or failure.

HTTP Status Code: 500

**ResourceNotFoundException**

The request references a resource which does not exist.

HTTP Status Code: 404

**ThrottlingException**

The request was denied due to request throttling.

HTTP Status Code: 429

**ValidationException**

The input fails to satisfy the constraints specified by an AWS service.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
StartEarthObservationJob

Service: Amazon SageMaker geospatial capabilities

Use this operation to create an Earth observation job.

Request Syntax

POST /earth-observation-jobs HTTP/1.1
Content-type: application/json

{
   "ClientToken": "string",
   "ExecutionRoleArn": "string",
   "InputConfig": {
      "PreviousEarthObservationJobArn": "string",
      "RasterDataCollectionQuery": {
         "AreaOfInterest": { ... },
         "PropertyFilters": {
            "LogicalOperator": "string",
            "Properties": [
               "Property": { ... }
            ]
         },
         "RasterDataCollectionArn": "string",
         "TimeRangeFilter": {
            "EndTime": number,
            "StartTime": number
         }
      }
   },
   "JobConfig": { ... },
   "KmsKeyId": "string",
   "Name": "string",
   "Tags": {
      "string": "string"
   }
}

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

ClientToken (p. 1169)

A unique token that guarantees that the call to this API is idempotent.

Type: String

Length Constraints: Minimum length of 36. Maximum length of 64.

Required: No

ExecutionRoleArn (p. 1169)

The Amazon Resource Name (ARN) of the IAM role that you specified for the job.
Type: String
Pattern: ^arn:(aws[a-z-]+):iam::([0-9]{12}):role/([a-zA-Z0-9+=,.@_/-]+)$
Required: Yes

**InputConfig (p. 1169)**
Input configuration information for the Earth Observation job.
Type: `InputConfigInput (p. 2126)` object
Required: Yes

**JobConfig (p. 1169)**
An object containing information about the job configuration.
Type: `JobConfigInput (p. 2129)` object

Note: This object is a Union. Only one member of this object can be specified or returned.
Required: Yes

**KmsKeyId (p. 1169)**
The Key Management Service key ID for server-side encryption.
Type: String
Length Constraints: Minimum length of 0. Maximum length of 2048.
Required: No

**Name (p. 1169)**
The name of the Earth Observation job.
Type: String
Length Constraints: Minimum length of 0. Maximum length of 200.
Required: Yes

**Tags (p. 1169)**
Each tag consists of a key and a value.
Type: String to string map
Required: No

**Response Syntax**

```json
HTTP/1.1 200
Content-type: application/json

{
    "Arn": "string",
    "CreationTime": "string",
    "DurationInSeconds": number,
    "ExecutionRoleArn": "string",
    "InputConfig": {
```
"PreviousEarthObservationJobArn": "string",
"RasterDataCollectionQuery": {
   "AreaOfInterest": { ... },
   "PropertyFilters": {
      "LogicalOperator": "string",
      "Properties": [
         { "Property": { ... } }
      ]
   },
   "RasterDataCollectionArn": "string",
   "RasterDataCollectionName": "string",
   "TimeRangeFilter": {
      "EndTime": "string",
      "StartTime": "string"
   }
},
"JobConfig": { ... },
"KmsKeyId": "string",
"Name": "string",
"Status": "string",
"Tags": {
   "string": "string"
}
}

Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

Arn (p. 1170)

The Amazon Resource Name (ARN) of the Earth Observation job.

Type: String

CreationTime (p. 1170)

The creation time.

Type: Timestamp

DurationInSeconds (p. 1170)

The duration of the session, in seconds.

Type: Integer

ExecutionRoleArn (p. 1170)

The Amazon Resource Name (ARN) of the IAM role that you specified for the job.

Type: String


Pattern: ^arn:(aws[a-z-]*):iam::([0-9]{12}):role/[a-zA-Z0-9-+=,.@_/-]+$

InputConfig (p. 1170)

Input configuration information for the Earth Observation job.
Type: \textit{InputConfigOutput (p. 2127)} object

\textbf{JobConfig (p. 1170)}

An object containing information about the job configuration.

Type: \textit{JobConfigInput (p. 2129)} object

\textbf{Note}: This object is a Union. Only one member of this object can be specified or returned.

\textbf{KmsKeyId (p. 1170)}

The Key Management Service key ID for server-side encryption.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

\textbf{Name (p. 1170)}

The name of the Earth Observation job.

Type: String

\textbf{Status (p. 1170)}

The status of the Earth Observation job.

Type: String

Valid Values: \texttt{INITIALIZING | IN\_PROGRESS | STOPPING | COMPLETED | STOPPED | FAILED | DELETING | DELETED}

\textbf{Tags (p. 1170)}

Each tag consists of a key and a value.

Type: String to string map

\section*{Errors}

For information about the errors that are common to all actions, see \textit{Common Errors (p. 2180)}.

\textbf{AccessDeniedException}

You do not have sufficient access to perform this action.

HTTP Status Code: 403

\textbf{ConflictException}

Updating or deleting a resource can cause an inconsistent state.

HTTP Status Code: 409

\textbf{InternalServerException}

The request processing has failed because of an unknown error, exception, or failure.

HTTP Status Code: 500

\textbf{ResourceNotFoundException}

The request references a resource which does not exist.

HTTP Status Code: 404
ServiceQuotaExceededException

You have exceeded the service quota.

HTTP Status Code: 402

ThrottlingException

The request was denied due to request throttling.

HTTP Status Code: 429

ValidationException

The input fails to satisfy the constraints specified by an AWS service.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
StartVectorEnrichmentJob

Service: Amazon SageMaker geospatial capabilities

Creates a Vector Enrichment job for the supplied job type. Currently, there are two supported job types: reverse geocoding and map matching.

Request Syntax

```
POST /vector-enrichment-jobs HTTP/1.1
Content-type: application/json

{
   "ClientToken": "string",
   "ExecutionRoleArn": "string",
   "InputConfig": {
      "DataSourceConfig": { ... },
      "DocumentType": "string"
   },
   "JobConfig": { ... },
   "KmsKeyId": "string",
   "Name": "string",
   "Tags": {
      "string": "string"
   }
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

**ClientToken (p. 1174)**

A unique token that guarantees that the call to this API is idempotent.

Type: String

Length Constraints: Minimum length of 36. Maximum length of 64.

Required: No

**ExecutionRoleArn (p. 1174)**

The Amazon Resource Name (ARN) of the IAM role that you specified for the job.

Type: String


Pattern: `^arn:(aws[a-z-]*):iam::([0-9]{12}):role/[a-zA-Z0-9-+=,.@_/\-]+$`

Required: Yes

**InputConfig (p. 1174)**

Input configuration information for the Vector Enrichment job.
Type: `VectorEnrichmentJobInputConfig (p. 2169)` object

Required: Yes

**JobConfig (p. 1174)**

An object containing information about the job configuration.

Type: `VectorEnrichmentJobConfig (p. 2165)` object

Note: This object is a Union. Only one member of this object can be specified or returned.

Required: Yes

**KmsKeyId (p. 1174)**

The Key Management Service key ID for server-side encryption.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Required: No

**Name (p. 1174)**

The name of the Vector Enrichment job.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 200.

Required: Yes

**Tags (p. 1174)**

Each tag consists of a key and a value.

Type: String to string map

Required: No

**Response Syntax**

HTTP/1.1 200
Content-type: application/json

```json
{
    "Arn": "string",
    "CreationTime": "string",
    "DurationInSeconds": number,
    "ExecutionRoleArn": "string",
    "InputConfig": {
        "DataSourceConfig": { ... },
        "DocumentType": "string"
    },
    "JobConfig": { ... },
    "KmsKeyId": "string",
    "Name": "string",
    "Status": "string",
    "Tags": {
        "string" : "string"
    },
    "Type": "string"
}
```
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**Arn (p. 1175)**

The Amazon Resource Name (ARN) of the Vector Enrichment job.

Type: String

Pattern: ^arn:aws[a-z-]*sagemaker-geospatial:[a-z0-9-]*[0-9]{12}:vector-enrichment-job/[a-z0-9]{12}$

**CreationTime (p. 1175)**

The creation time.

Type: Timestamp

**DurationInSeconds (p. 1175)**

The duration of the Vector Enrichment job, in seconds.

Type: Integer

**ExecutionRoleArn (p. 1175)**

The Amazon Resource Name (ARN) of the IAM role that you specified for the job.

Type: String


Pattern: ^arn:aws[a-z-]*:iam::*[0-9]{12}:role/[a-zA-Z0-9-]+,.,@/_]+$
**Status** (p. 1175)

The status of the Vector Enrichment job being started.

Type: String

Valid Values: INITIALIZING | IN_PROGRESS | STOPPING | STOPPED | COMPLETED | FAILED | DELETING | DELETED

**Tags** (p. 1175)

Each tag consists of a key and a value.

Type: String to string map

**Type** (p. 1175)

The type of the Vector Enrichment job.

Type: String

Valid Values: REVERSE_GEOCODING | MAP_MATCHING

**Errors**

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**AccessDeniedException**

You do not have sufficient access to perform this action.

HTTP Status Code: 403

**ConflictException**

Updating or deleting a resource can cause an inconsistent state.

HTTP Status Code: 409

**InternalServerException**

The request processing has failed because of an unknown error, exception, or failure.

HTTP Status Code: 500

**ResourceNotFoundException**

The request references a resource which does not exist.

HTTP Status Code: 404

**ServiceQuotaExceededException**

You have exceeded the service quota.

HTTP Status Code: 402

**ThrottlingException**

The request was denied due to request throttling.

HTTP Status Code: 429

**ValidationException**

The input fails to satisfy the constraints specified by an AWS service.
HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3
StopEarthObservationJob
Service: Amazon SageMaker geospatial capabilities

Use this operation to stop an existing earth observation job.

Request Syntax

```
POST /earth-observation-jobs/stop HTTP/1.1
Content-type: application/json

{
  "Arn": "string"
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

**Arn (p. 1179)**

The Amazon Resource Name (ARN) of the Earth Observation job being stopped.

Type: String

Pattern: `^arn:aws[a-z-][0,12]:sagemaker-geospatial:[a-z0-9-][1,25]:[0-9]{12}:earth-observation-job/[a-z0-9]{12},\$`

Required: Yes

Response Syntax

```
HTTP/1.1 200
```

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

**AccessDeniedException**

You do not have sufficient access to perform this action.

HTTP Status Code: 403

**ConflictException**

Updating or deleting a resource can cause an inconsistent state.

HTTP Status Code: 409
InternalServerException

The request processing has failed because of an unknown error, exception, or failure.

HTTP Status Code: 500

ResourceNotFoundException

The request references a resource which does not exist.

HTTP Status Code: 404

ThrottlingException

The request was denied due to request throttling.

HTTP Status Code: 429

ValidationException

The input fails to satisfy the constraints specified by an AWS service.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
StopVectorEnrichmentJob
Service: Amazon SageMaker geospatial capabilities

Stops the Vector Enrichment job for a given job ARN.

Request Syntax

POST /vector-enrichment-jobs/stop HTTP/1.1
Content-type: application/json
{
   "Arn": "string"
}

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

Arn (p. 1181)
The Amazon Resource Name (ARN) of the Vector Enrichment job.
Type: String
Pattern: `^arn:aws[a-z-][0,12]:sagemaker-geospatial:[a-z0-9-]{1,25}:[0-9]{12}:vector-enrichment-job/[a-z0-9]{12},]$
Required: Yes

Response Syntax

HTTP/1.1 200

Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors

For information about the errors that are common to all actions, see Common Errors (p. 2180).

AccessDeniedException
You do not have sufficient access to perform this action.
HTTP Status Code: 403

ConflictException
Updating or deleting a resource can cause an inconsistent state.
HTTP Status Code: 409
InternalServerException

The request processing has failed because of an unknown error, exception, or failure.

HTTP Status Code: 500

ResourceNotFoundException

The request references a resource which does not exist.

HTTP Status Code: 404

ThrottlingException

The request was denied due to request throttling.

HTTP Status Code: 429

ValidationException

The input fails to satisfy the constraints specified by an AWS service.

HTTP Status Code: 400

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
TagResource
Service: Amazon SageMaker geospatial capabilities
The resource you want to tag.

Request Syntax

```
PUT /tags/{ResourceArn} HTTP/1.1
Content-type: application/json

{
  "Tags": {
    "string": "string"
  }
}
```

URI Request Parameters
The request uses the following URI parameters.

**ResourceArn (p. 1183)**
- The Amazon Resource Name (ARN) of the resource you want to tag.
- Required: Yes

Request Body
The request accepts the following data in JSON format.

**Tags (p. 1183)**
- Each tag consists of a key and a value.
- Type: String to string map
- Required: Yes

Response Syntax

```
HTTP/1.1 200
```

Response Elements
If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors
For information about the errors that are common to all actions, see Common Errors (p. 2180).

**AccessDeniedException**
You do not have sufficient access to perform this action.
HTTP Status Code: 403

**InternalServerErrorException**

The request processing has failed because of an unknown error, exception, or failure.

HTTP Status Code: 500

**ResourceNotFoundException**

The request references a resource which does not exist.

HTTP Status Code: 404

**ThrottlingException**

The request was denied due to request throttling.

HTTP Status Code: 429

**ValidationException**

The input fails to satisfy the constraints specified by an AWS service.

HTTP Status Code: 400

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for JavaScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
UntagResource
Service: Amazon SageMaker geospatial capabilities
The resource you want to untag.

Request Syntax

```
DELETE /tags/ResourceArn?tagKeys=TagKeys HTTP/1.1
```

URI Request Parameters
The request uses the following URI parameters.

**ResourceArn (p. 1185)**

The Amazon Resource Name (ARN) of the resource you want to untag.

Length Constraints: Minimum length of 1. Maximum length of 2048.

Required: Yes

**TagKeys (p. 1185)**

Keys of the tags you want to remove.

Array Members: Minimum number of 1 item. Maximum number of 50 items.

Required: Yes

Request Body
The request does not have a request body.

Response Syntax

```
HTTP/1.1 200
```

Response Elements
If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

Errors
For information about the errors that are common to all actions, see Common Errors (p. 2180).

**AccessDeniedException**

You do not have sufficient access to perform this action.

HTTP Status Code: 403

**InternalServerException**

The request processing has failed because of an unknown error, exception, or failure.

HTTP Status Code: 500
ResourceNotFoundException
The request references a resource which does not exist.
HTTP Status Code: 404

ThrottlingException
The request was denied due to request throttling.
HTTP Status Code: 429

ValidationException
The input fails to satisfy the constraints specified by an AWS service.
HTTP Status Code: 400

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS Command Line Interface
- AWS SDK for .NET
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V3

Amazon SageMaker Metrics Service

The following actions are supported by Amazon SageMaker Metrics Service:

- BatchPutMetrics (p. 1187)
BatchPutMetrics

Service: Amazon SageMaker Metrics Service

Used to ingest training metrics into SageMaker. These metrics can be visualized in SageMaker Studio.

Request Syntax

PUT /BatchPutMetrics HTTP/1.1
Content-type: application/json

```json
{
   "MetricData": [
      {
         "MetricName": "string",
         "Step": number,
         "Timestamp": number,
         "Value": number
      }
   ],
   "TrialComponentName": "string"
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

**MetricData (p. 1187)**

A list of raw metric values to put.

Type: Array of RawMetricData (p. 2177) objects

Array Members: Minimum number of 1 item. Maximum number of 10 items.

Required: Yes

**TrialComponentName (p. 1187)**

The name of the Trial Component to associate with the metrics. The Trial Component name must be entirely lowercase.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: ^[a-zA-Z0-9](\-*[a-zA-Z0-9])?{0,119}

Required: Yes

Response Syntax

```
HTTP/1.1 200
Content-type: application/json
```
Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

**Errors** *(p. 1187)*

Lists any errors that occur when inserting metric data.

Type: Array of **BatchPutMetricsError** *(p. 2176)* objects

Array Members: Minimum number of 1 item. Maximum number of 10 items.

**Errors**

For information about the errors that are common to all actions, see **Common Errors** *(p. 2180).*

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for TypeScript V3](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V3](#)
Data Types

The following data types are supported by Amazon SageMaker Service:

- ActionSource (p. 1216)
- ActionSummary (p. 1217)
- AdditionalInferenceSpecificationDefinition (p. 1219)
- AdditionalS3DataSource (p. 1222)
- AgentVersion (p. 1223)
- Alarm (p. 1224)
- AlgorithmSpecification (p. 1225)
- AlgorithmStatusDetails (p. 1228)
- AlgorithmStatusItem (p. 1229)
- AlgorithmSummary (p. 1230)
- AlgorithmValidationProfile (p. 1232)
- AlgorithmValidationSpecification (p. 1233)
- AnnotationConsolidationConfig (p. 1234)
- AppDetails (p. 1246)
- AppImageConfigDetails (p. 1248)
- AppSpecification (p. 1250)
- ArtifactSource (p. 1251)
- ArtifactSourceType (p. 1252)
- ArtifactSummary (p. 1253)
- AssociationSummary (p. 1255)
- AsyncInferenceClientConfig (p. 1257)
- AsyncInferenceConfig (p. 1258)
- AsyncInferenceNotificationConfig (p. 1259)
- AsyncInferenceOutputConfig (p. 1260)
- AthenaDatasetDefinition (p. 1262)
- AutoMLAlgorithmConfig (p. 1264)
- AutoMLCandidate (p. 1265)
- AutoMLCandidateGenerationConfig (p. 1268)
- AutoMLCandidateStep (p. 1270)
- AutoMLChannel (p. 1271)
- AutoMLContainerDefinition (p. 1273)
- AutoMLDataSource (p. 1275)
- AutoMLDataSplitConfig (p. 1276)
- AutoMLJobArtifacts (p. 1277)
- AutoMLJobChannel (p. 1278)
- AutoMLJobCompletionCriteria (p. 1280)
- AutoMLJobConfig (p. 1282)
- AutoMLJobObjective (p. 1284)
- AutoMLJobStepMetadata (p. 1286)
- AutoMLJobSummary (p. 1287)
• AutoMLOutputDataConfig (p. 1289)
• AutoMLPartialFailureReason (p. 1290)
• AutoMLProblemTypeConfig (p. 1291)
• AutoMLProblemTypeResolvedAttributes (p. 1293)
• AutoMLResolvedAttributes (p. 1294)
• AutoMLS3DataSource (p. 1295)
• AutoMLSecurityConfig (p. 1297)
• AutoParameter (p. 1298)
• AutoRollbackConfig (p. 1299)
• Autotune (p. 1300)
• BatchDataCaptureConfig (p. 1301)
• BatchDescribeModelPackageError (p. 1302)
• BatchDescribeModelPackageSummary (p. 1303)
• BatchTransformInput (p. 1305)
• BestObjectiveNotImproving (p. 1308)
• Bias (p. 1309)
• BlueGreenUpdatePolicy (p. 1310)
• CacheHitResult (p. 1311)
• CallbackStepMetadata (p. 1312)
• CandidateArtifactLocations (p. 1313)
• CandidateGenerationConfig (p. 1314)
• CandidateProperties (p. 1315)
• CanvasAppSettings (p. 1316)
• CapacitySize (p. 1318)
• CaptureContentTypeHeader (p. 1319)
• CaptureOption (p. 1320)
• CategoricalParameter (p. 1321)
• CategoricalParameterRange (p. 1322)
• CategoricalParameterRangeSpecification (p. 1323)
• Channel (p. 1324)
• ChannelSpecification (p. 1326)
• CheckpointConfig (p. 1328)
• ClarifyCheckStepMetadata (p. 1329)
• ClarifyExplainerConfig (p. 1331)
• ClarifyInferenceConfig (p. 1332)
• ClarifyShapBaselineConfig (p. 1336)
• ClarifyShapConfig (p. 1338)
• ClarifyTextConfig (p. 1340)
• ClusterInstanceGroupDetails (p. 1341)
• ClusterInstanceGroupSpecification (p. 1343)
• ClusterInstanceStatusDetails (p. 1345)
• ClusterLifeCycleConfig (p. 1346)
• ClusterNodeDetails (p. 1347)
• ClusterNodeSummary (p. 1349)
• ClusterSummary (p. 1351)
• CodeEditorAppSettings (p. 1353)
- CodeRepository (p. 1354)
- CodeRepositorySummary (p. 1355)
- CognitoConfig (p. 1357)
- CognitoMemberDefinition (p. 1358)
- CollectionConfig (p. 1359)
- CollectionConfiguration (p. 1360)
- CompilationJobSummary (p. 1361)
- ConditionStepMetadata (p. 1363)
- ContainerConfig (p. 1364)
- ContainerDefinition (p. 1366)
- ContextSource (p. 1369)
- ContextSummary (p. 1370)
- ContinuousParameterRange (p. 1372)
- ContinuousParameterRangeSpecification (p. 1374)
- ConvergenceDetected (p. 1375)
- CustomFileSystem (p. 1376)
- CustomFileSystemConfig (p. 1377)
- CustomImage (p. 1378)
- CustomizedMetricSpecification (p. 1379)
- CustomPosixUserConfig (p. 1380)
- DataCaptureConfig (p. 1381)
- DataCaptureConfigSummary (p. 1383)
- DataCatalogConfig (p. 1385)
- DataProcessing (p. 1386)
- DataQualityAppSpecification (p. 1388)
- DataQualityBaselineConfig (p. 1390)
- DataQualityJobInput (p. 1391)
- DatasetDefinition (p. 1392)
- DataSource (p. 1394)
- DebugHookConfig (p. 1395)
- DebugRuleConfiguration (p. 1397)
- DebugRuleEvaluationStatus (p. 1399)
- DefaultEbsStorageSettings (p. 1401)
- DefaultSpaceSettings (p. 1402)
- DefaultSpaceStorageSettings (p. 1403)
- DeployedImage (p. 1404)
- DeploymentConfig (p. 1405)
- DeploymentRecommendation (p. 1406)
- DeploymentStage (p. 1407)
- DeploymentStageStatusSummary (p. 1408)
- DerivedInformation (p. 1409)
- DesiredWeightAndCapacity (p. 1410)
- Device (p. 1411)
- DeviceDeploymentSummary (p. 1412)
- DeviceFleetSummary (p. 1415)
- DeviceSelectionConfig (p. 1416)
• **DeviceStats** (p. 1418)
  • **DeviceSummary** (p. 1419)
  • **DirectDeploySettings** (p. 1421)
  • **DomainDetails** (p. 1422)
  • **DomainSettings** (p. 1424)
  • **DomainSettingsForUpdate** (p. 1425)
  • **DriftCheckBaselines** (p. 1426)
  • **DriftCheckBias** (p. 1427)
  • **DriftCheckExplainability** (p. 1428)
  • **DriftCheckModelDataQuality** (p. 1429)
  • **DriftCheckModelQuality** (p. 1430)
  • **DynamicScalingConfiguration** (p. 1431)
  • **EbsStorageSettings** (p. 1432)
  • **Edge** (p. 1433)
    • **EdgeDeploymentConfig** (p. 1434)
    • **EdgeDeploymentModelConfig** (p. 1435)
    • **EdgeDeploymentPlanSummary** (p. 1436)
    • **EdgeDeploymentStatus** (p. 1438)
  • **EdgeModel** (p. 1440)
  • **EdgeModelStat** (p. 1441)
  • **EdgeModelSummary** (p. 1443)
  • **EdgeOutputConfig** (p. 1444)
  • **EdgePackagingJobSummary** (p. 1446)
  • **EdgePresetDeploymentOutput** (p. 1448)
  • **EFSFileSystem** (p. 1449)
  • **EFSFileSystemConfig** (p. 1450)
  • **EMRStepMetadata** (p. 1451)
  • **Endpoint** (p. 1452)
    • **EndpointConfigSummary** (p. 1455)
    • **EndpointInfo** (p. 1456)
    • **EndpointInput** (p. 1457)
    • **EndpointInputConfiguration** (p. 1460)
    • **EndpointMetadata** (p. 1462)
    • **EndpointOutputConfiguration** (p. 1464)
    • **EndpointPerformance** (p. 1466)
    • **EndpointSummary** (p. 1467)
  • **EnvironmentParameter** (p. 1469)
  • **EnvironmentParameterRanges** (p. 1470)
  • **Experiment** (p. 1471)
    • **ExperimentConfig** (p. 1473)
    • **ExperimentSource** (p. 1475)
    • **ExperimentSummary** (p. 1476)
  • **Explainability** (p. 1478)
  • **ExplainerConfig** (p. 1479)
  • **FailStepMetadata** (p. 1480)
  • **FeatureDefinition** (p. 1481)
• FeatureGroup (p. 1483)
• FeatureGroupSummary (p. 1487)
• FeatureMetadata (p. 1489)
• FeatureParameter (p. 1491)
• FileSource (p. 1492)
• FileSystemConfig (p. 1493)
• FileSystemDataSource (p. 1494)
• Filter (p. 1496)
• FinalAutoMLJobObjectiveMetric (p. 1499)
• FinalHyperParameterTuningJobObjectiveMetric (p. 1501)
• FlowDefinitionOutputConfig (p. 1502)
• FlowDefinitionSummary (p. 1503)
• GitConfig (p. 1505)
• GitConfigForUpdate (p. 1506)
• HolidayConfigAttributes (p. 1507)
• HubContentDependency (p. 1508)
• HubContentInfo (p. 1509)
• HubInfo (p. 1512)
• HubS3StorageConfig (p. 1514)
• HumanLoopActivationConditionsConfig (p. 1515)
• HumanLoopActivationConfig (p. 1516)
• HumanLoopConfig (p. 1517)
• HumanLoopRequestSource (p. 1522)
• HumanTaskConfig (p. 1523)
• HumanTaskUiSummary (p. 1537)
• HyperbandStrategyConfig (p. 1538)
• HyperParameterAlgorithmSpecification (p. 1540)
• HyperParameterSpecification (p. 1542)
• HyperParameterTrainingJobDefinition (p. 1544)
• HyperParameterTrainingJobSummary (p. 1548)
• HyperParameterTuningInstanceConfig (p. 1551)
• HyperParameterTuningJobCompletionDetails (p. 1553)
• HyperParameterTuningJobConfig (p. 1554)
• HyperParameterTuningJobConsumedResources (p. 1556)
• HyperParameterTuningJobObjective (p. 1557)
• HyperParameterTuningJobSearchEntity (p. 1558)
• HyperParameterTuningJobStrategyConfig (p. 1561)
• HyperParameterTuningJobSummary (p. 1562)
• HyperParameterTuningJobWarmStartConfig (p. 1564)
• HyperParameterTuningResourceConfig (p. 1566)
• iamIdentity (p. 1569)
• identityProviderOAuthSetting (p. 1570)
• image (p. 1571)
• ImageClassificationJobConfig (p. 1573)
• ImageConfig (p. 1574)
• ImageVersion (p. 1575)
• InferenceComponentComputeResourceRequirements (p. 1577)
• InferenceComponentContainerSpecification (p. 1578)
• InferenceComponentContainerSpecificationSummary (p. 1580)
• InferenceComponentRuntimeConfig (p. 1582)
• InferenceComponentRuntimeConfigSummary (p. 1583)
• InferenceComponentSpecification (p. 1584)
• InferenceComponentSpecificationSummary (p. 1585)
• InferenceComponentStartupParameters (p. 1586)
• InferenceComponentSummary (p. 1587)
• InferenceExecutionConfig (p. 1589)
• InferenceExperimentDataStorageConfig (p. 1590)
• InferenceExperimentSchedule (p. 1591)
• InferenceExperimentSummary (p. 1592)
• InferenceMetrics (p. 1594)
• InferenceRecommendation (p. 1595)
• InferenceRecommendationsJob (p. 1597)
• InferenceRecommendationsJobStep (p. 1600)
• InferenceSpecification (p. 1601)
• InfraCheckConfig (p. 1604)
• InputConfig (p. 1605)
• InstanceGroup (p. 1609)
• InstanceMetadataServiceConfiguration (p. 1611)
• IntegerParameterRange (p. 1612)
• IntegerParameterRangeSpecification (p. 1614)
• JupyterLabAppImageConfig (p. 1615)
• JupyterLabAppSettings (p. 1616)
• JupyterServerAppSettings (p. 1618)
• KendraSettings (p. 1619)
• KernelGatewayAppSettings (p. 1620)
• KernelGatewayImageConfig (p. 1621)
• KernelSpec (p. 1622)
• LabelCounters (p. 1623)
• LabelCountersForWorkteam (p. 1625)
• LabelingJobAlgorithmsConfig (p. 1626)
• LabelingJobDataAttributes (p. 1628)
• LabelingJobDataSource (p. 1629)
• LabelingJobForWorkteamSummary (p. 1630)
• LabelingJobInputConfig (p. 1632)
• LabelingJobOutput (p. 1633)
• LabelingJobOutputConfig (p. 1634)
• LabelingJobResourceConfig (p. 1636)
• LabelingJobS3DataSource (p. 1637)
• LabelingJobSnsDataSource (p. 1638)
• LabelingJobStoppingConditions (p. 1639)
• LabelingJobSummary (p. 1640)
• LambdaStepMetadata (p. 1643)
- LastUpdateStatus (p. 1644)
- LineageGroupSummary (p. 1645)
- MemberDefinition (p. 1647)
- MetadataProperties (p. 1648)
- MetricData (p. 1650)
- MetricDatum (p. 1651)
- MetricDefinition (p. 1653)
- MetricSpecification (p. 1654)
- MetricsSource (p. 1655)
- Model (p. 1656)
- ModelAccessConfig (p. 1658)
- ModelArtifacts (p. 1659)
- ModelBiasAppSpecification (p. 1660)
- ModelBiasBaselineConfig (p. 1661)
- ModelBiasJobInput (p. 1662)
- ModelCard (p. 1663)
- ModelCardExportArtifacts (p. 1666)
- ModelCardExportJobSummary (p. 1667)
- ModelCardExportOutputConfig (p. 1669)
- ModelCardSecurityConfig (p. 1670)
- ModelCardSummary (p. 1671)
- ModelCardVersionSummary (p. 1673)
- ModelClientConfig (p. 1675)
- ModelConfiguration (p. 1676)
- ModelDashboardEndpoint (p. 1677)
- ModelDashboardIndicatorAction (p. 1679)
- ModelDashboardModel (p. 1680)
- ModelDashboardModelCard (p. 1681)
- ModelDashboardMonitoringSchedule (p. 1684)
- ModelDataQuality (p. 1687)
- ModelDataSource (p. 1688)
- ModelDeployConfig (p. 1689)
- ModelDeployResult (p. 1690)
- ModelDigests (p. 1691)
- ModelExplainabilityAppSpecification (p. 1692)
- ModelExplainabilityBaselineConfig (p. 1693)
- ModelExplainabilityJobInput (p. 1694)
- ModelInfrastructureConfig (p. 1695)
- ModelInput (p. 1696)
- ModelLatencyThreshold (p. 1697)
- ModelMetadataFilter (p. 1698)
- ModelMetadataSearchExpression (p. 1699)
- ModelMetadataSummary (p. 1700)
- ModelMetrics (p. 1701)
- ModelPackage (p. 1702)
- ModelPackageContainerDefinition (p. 1707)
• ModelPackageGroup (p. 1710)
• ModelPackageGroupSummary (p. 1712)
• ModelPackageStatusDetails (p. 1714)
• ModelPackageStatusItem (p. 1715)
• ModelPackageSummary (p. 1716)
• ModelPackageValidationProfile (p. 1718)
• ModelPackageValidationSpecification (p. 1719)
• ModelQuality (p. 1720)
• ModelQualityAppSpecification (p. 1721)
• ModelQualityBaselineConfig (p. 1723)
• ModelQualityJobInput (p. 1724)
• ModelRegisterSettings (p. 1725)
• ModelStepMetadata (p. 1726)
• ModelSummary (p. 1727)
• ModelVariantConfig (p. 1728)
• ModelVariantConfigSummary (p. 1729)
• MonitoringAlertActions (p. 1731)
• MonitoringAlertHistorySummary (p. 1732)
• MonitoringAlertSummary (p. 1733)
• MonitoringAppSpecification (p. 1735)
• MonitoringBaselineConfig (p. 1737)
• MonitoringClusterConfig (p. 1738)
• MonitoringConstraintsResource (p. 1740)
• MonitoringCsvDatasetFormat (p. 1741)
• MonitoringDatasetFormat (p. 1742)
• MonitoringExecutionSummary (p. 1743)
• MonitoringGroundTruthS3Input (p. 1745)
• MonitoringInput (p. 1746)
• MonitoringJobDefinition (p. 1747)
• MonitoringJobDefinitionSummary (p. 1749)
• MonitoringJsonDatasetFormat (p. 1751)
• MonitoringNetworkConfig (p. 1752)
• MonitoringOutput (p. 1753)
• MonitoringOutputConfig (p. 1754)
• MonitoringParquetDatasetFormat (p. 1755)
• MonitoringResources (p. 1756)
• MonitoringS3Output (p. 1757)
• MonitoringSchedule (p. 1758)
• MonitoringScheduleConfig (p. 1761)
• MonitoringScheduleSummary (p. 1762)
• MonitoringStatisticsResource (p. 1764)
• MonitoringStoppingCondition (p. 1765)
• MultiModelConfig (p. 1766)
• NeoVpcConfig (p. 1767)
• NestedFilters (p. 1768)
• NetworkConfig (p. 1769)

1196
• **NotebookInstanceLifecycleConfigSummary** (p. 1770)
• **NotebookInstanceLifecycleHook** (p. 1771)
• **NotebookInstanceSummary** (p. 1772)
• **NotificationConfiguration** (p. 1775)
• **ObjectiveStatusCounters** (p. 1776)
• **OfflineStoreConfig** (p. 1777)
• **OfflineStoreStatus** (p. 1778)
• **OidcConfig** (p. 1779)
• **OidcConfigForResponse** (p. 1781)
• **OidcMemberDefinition** (p. 1783)
• **OnlineStoreConfig** (p. 1784)
• **OnlineStoreConfigUpdate** (p. 1786)
• **OnlineStoreSecurityConfig** (p. 1787)
• **OutputConfig** (p. 1788)
• **OutputDataConfig** (p. 1792)
• **OutputParameter** (p. 1794)
• **OwnershipSettings** (p. 1795)
• **OwnershipSettingsSummary** (p. 1796)
• **ParallelismConfiguration** (p. 1797)
• **Parameter** (p. 1798)
• **ParameterRange** (p. 1799)
• **ParameterRanges** (p. 1800)
• **Parent** (p. 1802)
• **ParentHyperParameterTuningJob** (p. 1803)
• **PendingDeploymentSummary** (p. 1804)
• **PendingProductionVariantSummary** (p. 1806)
• **Phase** (p. 1810)
• **Pipeline** (p. 1811)
• **PipelineDefinitionS3Location** (p. 1814)
• **PipelineExecution** (p. 1815)
• **PipelineExecutionStep** (p. 1818)
• **PipelineExecutionStepMetadata** (p. 1820)
• **PipelineExecutionSummary** (p. 1823)
• **PipelineExperimentConfig** (p. 1825)
• **PipelineSummary** (p. 1826)
• **PredefinedMetricSpecification** (p. 1828)
• **ProcessingClusterConfig** (p. 1829)
• **ProcessingFeatureStoreOutput** (p. 1831)
• **ProcessingInput** (p. 1832)
• **ProcessingJob** (p. 1833)
• **ProcessingJobStepMetadata** (p. 1837)
• **ProcessingJobSummary** (p. 1838)
• **ProcessingOutput** (p. 1840)
• **ProcessingOutputConfig** (p. 1841)
• **ProcessingResources** (p. 1842)
• **ProcessingS3Input** (p. 1843)
• ProcessingS3Output (p. 1845)
• ProcessingStoppingCondition (p. 1846)
• ProductionVariant (p. 1847)
• ProductionVariantCoreDumpConfig (p. 1851)
• ProductionVariantManagedInstanceScaling (p. 1853)
• ProductionVariantRoutingConfig (p. 1854)
• ProductionVariantServerlessConfig (p. 1855)
• ProductionVariantServerlessUpdateConfig (p. 1856)
• ProductionVariantStatus (p. 1857)
• ProductionVariantSummary (p. 1858)
• ProfilerConfig (p. 1860)
• ProfilerConfigForUpdate (p. 1862)
• ProfilerRuleConfiguration (p. 1864)
• ProfilerRuleEvaluationStatus (p. 1866)
• Project (p. 1868)
• ProjectSummary (p. 1871)
• PropertyNameQuery (p. 1873)
• PropertyNameSuggestion (p. 1874)
• ProvisioningParameter (p. 1875)
• PublicWorkforceTaskPrice (p. 1876)
• QualityCheckStepMetadata (p. 1879)
• QueryFilters (p. 1881)
• RealTimeInferenceConfig (p. 1883)
• RealTimeInferenceRecommendation (p. 1884)
• RecommendationJobCompiledOutputConfig (p. 1886)
• RecommendationJobContainerConfig (p. 1887)
• RecommendationJobInferenceBenchmark (p. 1890)
• RecommendationJobInputConfig (p. 1892)
• RecommendationJobOutputConfig (p. 1895)
• RecommendationJobPayloadConfig (p. 1896)
• RecommendationJobResourceLimit (p. 1897)
• RecommendationJobStoppingConditions (p. 1898)
• RecommendationJobVpcConfig (p. 1899)
• RecommendationMetrics (p. 1900)
• RedshiftDatasetDefinition (p. 1902)
• RegisterModelStepMetadata (p. 1904)
• RenderableTask (p. 1905)
• RenderingError (p. 1906)
• RepositoryAuthConfig (p. 1907)
• ResolvedAttributes (p. 1908)
• ResourceCatalog (p. 1909)
• ResourceConfig (p. 1910)
• ResourceConfigForUpdate (p. 1913)
• ResourceLimits (p. 1914)
• ResourceSpec (p. 1915)
• RetentionPolicy (p. 1917)
- **RetryStrategy** (p. 1918)
- **RollingUpdatePolicy** (p. 1919)
- **RSessionAppSettings** (p. 1920)
- **RStudioServerProAppSettings** (p. 1921)
- **RStudioServerProDomainSettings** (p. 1922)
- **RStudioServerProDomainSettingsForUpdate** (p. 1923)
- **S3DataSource** (p. 1924)
- **S3ModelDataSource** (p. 1927)
- **S3StorageConfig** (p. 1929)
- **ScalingPolicy** (p. 1931)
- **ScalingPolicyMetric** (p. 1932)
- **ScalingPolicyObjective** (p. 1933)
- **ScheduleConfig** (p. 1934)
- **SearchExpression** (p. 1936)
- **SearchRecord** (p. 1938)
- **SecondaryStatusTransition** (p. 1941)
- **SelectedStep** (p. 1943)
- **SelectiveExecutionConfig** (p. 1944)
- **SelectiveExecutionResult** (p. 1945)
- **ServiceCatalogProvisionedProductDetails** (p. 1946)
- **ServiceCatalogProvisioningDetails** (p. 1947)
- **ServiceCatalogProvisioningUpdateDetails** (p. 1949)
- **ShadowModeConfig** (p. 1950)
- **ShadowModelVariantConfig** (p. 1951)
- **SharingSettings** (p. 1952)
- **ShuffleConfig** (p. 1953)
- **SourceAlgorithm** (p. 1954)
- **SourceAlgorithmSpecification** (p. 1955)
- **SourceIpConfig** (p. 1956)
- **SpaceCodeEditorAppSettings** (p. 1957)
- **SpaceDetails** (p. 1958)
- **SpaceJupyterLabAppSettings** (p. 1960)
- **SpaceSettings** (p. 1961)
- **SpaceSettingsSummary** (p. 1963)
- **SpaceSharingSettings** (p. 1964)
- **SpaceSharingSettingsSummary** (p. 1965)
- **SpaceStorageSettings** (p. 1966)
- **Stairs** (p. 1967)
- **StoppingCondition** (p. 1968)
- **StudioLifecycleConfigDetails** (p. 1970)
- **SubscribedWorkteam** (p. 1972)
- **SuggestionQuery** (p. 1974)
- **TabularJobConfig** (p. 1975)
- **TabularResolvedAttributes** (p. 1978)
- **Tag** (p. 1979)
- **TargetPlatform** (p. 1980)
• TargetTrackingScalingPolicyConfiguration (p. 1982)
• TensorBoardAppSettings (p. 1983)
• TensorBoardOutputConfig (p. 1984)
• TextClassificationJobConfig (p. 1985)
• TextGenerationJobConfig (p. 1986)
• TextGenerationResolvedAttributes (p. 1988)
• TimeSeriesConfig (p. 1989)
• TimeSeriesForecastingJobConfig (p. 1991)
• TimeSeriesForecastingSettings (p. 1994)
• TimeSeriesTransformations (p. 1995)
• TrafficPattern (p. 1997)
• TrafficRoutingConfig (p. 1998)
• TrainingImageConfig (p. 2000)
• TrainingJob (p. 2001)
• TrainingJobDefinition (p. 2009)
• TrainingJobStatusCounters (p. 2011)
• TrainingJobStepMetadata (p. 2013)
• TrainingJobSummary (p. 2014)
• TrainingRepositoryAuthConfig (p. 2016)
• TrainingSpecification (p. 2017)
• TransformDataSource (p. 2019)
• TransformInput (p. 2020)
• TransformJob (p. 2022)
• TransformJobDefinition (p. 2027)
• TransformJobStepMetadata (p. 2029)
• TransformJobSummary (p. 2030)
• TransformOutput (p. 2032)
• TransformResources (p. 2034)
• TransformS3DataSource (p. 2036)
• Trial (p. 2038)
• TrialComponent (p. 2041)
• TrialComponentArtifact (p. 2045)
• TrialComponentMetricSummary (p. 2046)
• TrialComponentParameterValue (p. 2048)
• TrialComponentSimpleSummary (p. 2049)
• TrialComponentSource (p. 2051)
• TrialComponentSourceDetail (p. 2052)
• TrialComponentStatus (p. 2053)
• TrialComponentSummary (p. 2054)
• TrialSource (p. 2056)
• TrialSummary (p. 2057)
• TtlDuration (p. 2059)
• TuningJobCompletionCriteria (p. 2060)
• TuningJobStepMetaData (p. 2061)
• UiConfig (p. 2062)
• UiTemplate (p. 2064)
The following data types are supported by Amazon SageMaker Runtime:

- **PayloadPart (p. 2088)**
- **ResponseStream (p. 2089)**

The following data types are supported by Amazon SageMaker Edge Manager:

- **Checksum (p. 2091)**
- **Definition (p. 2092)**
- **DeploymentModel (p. 2093)**
- **DeploymentResult (p. 2095)**
- **EdgeDeployment (p. 2097)**
- **EdgeMetric (p. 2098)**
- **Model (p. 2099)**

The following data types are supported by Amazon SageMaker Feature Store Runtime:

- **BatchGetRecordError (p. 2101)**
- **BatchGetRecordIdentifier (p. 2103)**
- **BatchGetRecordResultDetail (p. 2105)**
- **FeatureValue (p. 2106)**
- **TtlDuration (p. 2107)**

The following data types are supported by Amazon SageMaker geospatial capabilities:

- **AreaOfInterest (p. 2110)**
- **AreaOfInterestGeometry (p. 2111)**
- **AssetValue (p. 2112)**
- **BandMathConfigInput (p. 2113)**
- **CloudMaskingConfigInput (p. 2114)**
- **CloudRemovalConfigInput (p. 2115)**
• CustomIndicesInput (p. 2116)
• EarthObservationJobErrorDetails (p. 2117)
• EoCloudCoverInput (p. 2118)
• ExportErrorDetails (p. 2119)
• ExportErrorDetailsOutput (p. 2120)
• ExportS3DataInput (p. 2121)
• ExportVectorEnrichmentJobOutputConfig (p. 2122)
• Filter (p. 2123)
• Geometry (p. 2124)
• GeoMosaicConfigInput (p. 2125)
• InputConfigInput (p. 2126)
• InputConfigOutput (p. 2127)
• ItemSource (p. 2128)
• JobConfigInput (p. 2129)
• LandCoverSegmentationConfigInput (p. 2131)
• LandsatCloudCoverLandInput (p. 2132)
• ListEarthObservationJobOutputConfig (p. 2133)
• ListVectorEnrichmentJobOutputConfig (p. 2135)
• MapMatchingConfig (p. 2137)
• MultiPolygonGeometryInput (p. 2138)
• Operation (p. 2139)
• OutputBand (p. 2140)
• OutputConfigInput (p. 2141)
• OutputResolutionResamplingInput (p. 2142)
• OutputResolutionStackInput (p. 2143)
• PlatformInput (p. 2144)
• PolygonGeometryInput (p. 2145)
• Properties (p. 2146)
• Property (p. 2148)
• PropertyFilter (p. 2150)
• PropertyFilters (p. 2151)
• RasterDataCollectionMetadata (p. 2152)
• RasterDataCollectionQueryInput (p. 2154)
• RasterDataCollectionQueryOutput (p. 2155)
• RasterDataCollectionQueryWithBandFilterInput (p. 2157)
• ResamplingConfigInput (p. 2158)
• ReverseGeocodingConfig (p. 2159)
• StackConfigInput (p. 2160)
• TemporalStatisticsConfigInput (p. 2161)
• TimeRangeFilterInput (p. 2162)
• TimeRangeFilterOutput (p. 2163)
• UserDefined (p. 2164)
• VectorEnrichmentJobConfig (p. 2165)
• VectorEnrichmentJobDataSourceConfigInput (p. 2166)
• VectorEnrichmentJobErrorDetails (p. 2167)
• VectorEnrichmentJobExportErrorDetails (p. 2168)
Amazon SageMaker Service

The following data types are supported by Amazon SageMaker Service:

- ActionSource (p. 1216)
- ActionSummary (p. 1217)
- AdditionalInferenceSpecificationDefinition (p. 1219)
- AdditionalS3DataSource (p. 1222)
- AgentVersion (p. 1223)
- Alarm (p. 1224)
- AlgorithmSpecification (p. 1225)
- AlgorithmStatusDetails (p. 1228)
- AlgorithmStatusItem (p. 1229)
- AlgorithmSummary (p. 1230)
- AlgorithmValidationProfile (p. 1232)
- AlgorithmValidationSpecification (p. 1233)
- AnnotationConsolidationConfig (p. 1234)
- AppDetails (p. 1246)
- AppImageConfigDetails (p. 1248)
- AppSpecification (p. 1250)
- ArtifactSource (p. 1251)
- ArtifactSourceType (p. 1252)
- ArtifactSummary (p. 1253)
- AssociationSummary (p. 1255)
- AsyncInferenceClientConfig (p. 1257)
- AsyncInferenceConfig (p. 1258)
- AsyncInferenceNotificationConfig (p. 1259)
- AsyncInferenceOutputConfig (p. 1260)
- AthenaDatasetDefinition (p. 1262)
- AutoMLAlgorithmConfig (p. 1264)
- AutoMLCandidate (p. 1265)
- AutoMLCandidateGenerationConfig (p. 1268)
- AutoMLCandidateStep (p. 1270)
- AutoMLChannel (p. 1271)
• **AutoMLContainerDefinition** (p. 1273)
• **AutoMLDataSource** (p. 1275)
• **AutoMLDataSplitConfig** (p. 1276)
• **AutoMLJobArtifacts** (p. 1277)
• **AutoMLJobChannel** (p. 1278)
• **AutoMLJobCompletionCriteria** (p. 1280)
• **AutoMLJobConfig** (p. 1282)
• **AutoMLJobObjective** (p. 1284)
• **AutoMLJobStepMetadata** (p. 1286)
• **AutoMLJobSummary** (p. 1287)
• **AutoMLOutputDataConfig** (p. 1289)
• **AutoMLPartialFailureReason** (p. 1290)
• **AutoMLProblemTypeConfig** (p. 1291)
• **AutoMLProblemTypeResolvedAttributes** (p. 1293)
• **AutoMLResolvedAttributes** (p. 1294)
• **AutoMLSS3DataSource** (p. 1295)
• **AutoMLSecurityConfig** (p. 1297)
• **AutoParameter** (p. 1298)
• **AutoRollbackConfig** (p. 1299)
• **Autotune** (p. 1300)
• **BatchDataCaptureConfig** (p. 1301)
• **BatchDescribeModelPackageError** (p. 1302)
• **BatchDescribeModelPackageSummary** (p. 1303)
• **BatchTransformInput** (p. 1305)
• **BestObjectiveNotImproving** (p. 1308)
• **Bias** (p. 1309)
• **BlueGreenUpdatePolicy** (p. 1310)
• **CacheHitResult** (p. 1311)
• **CallbackStepMetadata** (p. 1312)
• **CandidateArtifactLocations** (p. 1313)
• **CandidateGenerationConfig** (p. 1314)
• **CandidateProperties** (p. 1315)
• **CanvasAppSettings** (p. 1316)
• **CapacitySize** (p. 1318)
• **CaptureContentTypeHeader** (p. 1319)
• **CaptureOption** (p. 1320)
• **CategoricalParameter** (p. 1321)
• **CategoricalParameterRange** (p. 1322)
• **CategoricalParameterRangeSpecification** (p. 1323)
• **Channel** (p. 1324)
• **ChannelSpecification** (p. 1326)
• **CheckpointConfig** (p. 1328)
• **ClarifyCheckStepMetadata** (p. 1329)
• **ClarifyExplainerConfig** (p. 1331)
• **ClarifyInferenceConfig** (p. 1332)
• **ClarifyShapBaselineConfig** (p. 1336)
• ClarifyShapConfig (p. 1338)
• ClarifyTextConfig (p. 1340)
• ClusterInstanceGroupDetails (p. 1341)
• ClusterInstanceGroupSpecification (p. 1343)
• ClusterInstanceStatusDetails (p. 1345)
• ClusterLifeCycleConfig (p. 1346)
• ClusterNodeDetails (p. 1347)
• ClusterNodeSummary (p. 1349)
• ClusterSummary (p. 1351)
• CodeEditorAppSettings (p. 1353)
• CodeRepository (p. 1354)
• CodeRepositorySummary (p. 1355)
• CognitoConfig (p. 1357)
• CognitoMemberDefinition (p. 1358)
• CollectionConfig (p. 1359)
• CollectionConfiguration (p. 1360)
• CompilationJobSummary (p. 1361)
• ConditionStepMetadata (p. 1363)
• ContainerConfig (p. 1364)
• ContainerDefinition (p. 1366)
• ContextSource (p. 1369)
• ContextSummary (p. 1370)
• ContinuousParameterRange (p. 1372)
• ContinuousParameterRangeSpecification (p. 1374)
• ConvergenceDetected (p. 1375)
• CustomFileSystem (p. 1376)
• CustomFileSystemConfig (p. 1377)
• CustomImage (p. 1378)
• CustomizedMetricSpecification (p. 1379)
• CustomPosixUserConfig (p. 1380)
• DataCaptureConfig (p. 1381)
• DataCaptureConfigSummary (p. 1383)
• DataCatalogConfig (p. 1385)
• DataProcessing (p. 1386)
• DataQualityAppSpecification (p. 1388)
• DataQualityBaselineConfig (p. 1390)
• DataQualityJobInput (p. 1391)
• DatasetDefinition (p. 1392)
• DataSource (p. 1394)
• DebugHookConfig (p. 1395)
• DebugRuleConfiguration (p. 1397)
• DebugRuleEvaluationStatus (p. 1399)
• DefaultEbsStorageSettings (p. 1401)
• DefaultSpaceSettings (p. 1402)
• DefaultSpaceStorageSettings (p. 1403)
• DeployedImage (p. 1404)
• **DeploymentConfig (p. 1405)**
• **DeploymentRecommendation (p. 1406)**
• **DeploymentStage (p. 1407)**
• **DeploymentStageStatusSummary (p. 1408)**
• **DerivedInformation (p. 1409)**
• **DesiredWeightAndCapacity (p. 1410)**
• **Device (p. 1411)**
• **DeviceDeploymentSummary (p. 1412)**
• **DeviceFleetSummary (p. 1415)**
• **DeviceSelectionConfig (p. 1416)**
• **DeviceStats (p. 1418)**
• **DeviceSummary (p. 1419)**
• **DirectDeploySettings (p. 1421)**
• **DomainDetails (p. 1422)**
• **DomainSettings (p. 1424)**
• **DomainSettingsForUpdate (p. 1425)**
• **DriftCheckBaselines (p. 1426)**
• **DriftCheckBias (p. 1427)**
• **DriftCheckExplainability (p. 1428)**
• **DriftCheckModelDataQuality (p. 1429)**
• **DriftCheckModelQuality (p. 1430)**
• **DynamicScalingConfiguration (p. 1431)**
• **EbsStorageSettings (p. 1432)**
• **Edge (p. 1433)**
• **EdgeDeploymentConfig (p. 1434)**
• **EdgeDeploymentModelConfig (p. 1435)**
• **EdgeDeploymentPlanSummary (p. 1436)**
• **EdgeDeploymentStatus (p. 1438)**
• **EdgeModel (p. 1440)**
• **EdgeModelStat (p. 1441)**
• **EdgeModelSummary (p. 1443)**
• **EdgeOutputConfig (p. 1444)**
• **EdgePackagingJobSummary (p. 1446)**
• **EdgePresetDeploymentOutput (p. 1448)**
• **EFSFileSystem (p. 1449)**
• **EFSFileSystemConfig (p. 1450)**
• **EMRStepMetadata (p. 1451)**
• **Endpoint (p. 1452)**
• **EndpointConfigSummary (p. 1455)**
• **EndpointInfo (p. 1456)**
• **EndpointInput (p. 1457)**
• **EndpointInputConfiguration (p. 1460)**
• **EndpointMetadata (p. 1462)**
• **EndpointOutputConfiguration (p. 1464)**
• **EndpointPerformance (p. 1466)**
• **EndpointSummary (p. 1467)**
• EnvironmentParameter (p. 1469)
• EnvironmentParameterRanges (p. 1470)
• Experiment (p. 1471)
• ExperimentConfig (p. 1473)
• ExperimentSource (p. 1475)
• ExperimentSummary (p. 1476)
• Explainability (p. 1478)
• ExplainerConfig (p. 1479)
• FailStepMetadata (p. 1480)
• FeatureDefinition (p. 1481)
• FeatureGroup (p. 1483)
• FeatureGroupSummary (p. 1487)
• FeatureMetadata (p. 1489)
• FeatureParameter (p. 1491)
• FileSource (p. 1492)
• FileSystemConfig (p. 1493)
• FileSystemDataSource (p. 1494)
• Filter (p. 1496)
• FinalAutoMLJobObjectiveMetric (p. 1499)
• FinalHyperParameterTuningJobObjectiveMetric (p. 1501)
• FlowDefinitionOutputConfig (p. 1502)
• FlowDefinitionSummary (p. 1503)
• GitConfig (p. 1505)
• GitConfigForUpdate (p. 1506)
• HolidayConfigAttributes (p. 1507)
• HubContentDependency (p. 1508)
• HubContentInfo (p. 1509)
• HubInfo (p. 1512)
• HubS3StorageConfig (p. 1514)
• HumanLoopActivationConditionsConfig (p. 1515)
• HumanLoopActivationConfig (p. 1516)
• HumanLoopConfig (p. 1517)
• HumanLoopRequestSource (p. 1522)
• HumanTaskConfig (p. 1523)
• HumanTaskUiSummary (p. 1537)
• HyperbandStrategyConfig (p. 1538)
• HyperParameterAlgorithmSpecification (p. 1540)
• HyperParameterSpecification (p. 1542)
• HyperParameterTrainingJobDefinition (p. 1544)
• HyperParameterTrainingJobSummary (p. 1548)
• HyperParameterTuningInstanceConfig (p. 1551)
• HyperParameterTuningJobCompletionDetails (p. 1553)
• HyperParameterTuningJobConfig (p. 1554)
• HyperParameterTuningJobConsumedResources (p. 1556)
• HyperParameterTuningJobObjective (p. 1557)
• HyperParameterTuningJobSearchEntity (p. 1558)
• HyperParameterTuningJobStrategyConfig (p. 1561)
• HyperParameterTuningJobSummary (p. 1562)
• HyperParameterTuningJobWarmStartConfig (p. 1564)
• HyperParameterTuningResourceConfig (p. 1566)
• iamIdentity (p. 1569)
• IdentityProviderOAuthSetting (p. 1570)
• Image (p. 1571)
• ImageClassificationJobConfig (p. 1573)
• ImageConfig (p. 1574)
• ImageVersion (p. 1575)
• InferenceComponentComputeResourceRequirements (p. 1577)
• InferenceComponentContainerSpecification (p. 1578)
• InferenceComponentContainerSpecificationSummary (p. 1580)
• InferenceComponentRuntimeConfig (p. 1582)
• InferenceComponentRuntimeConfigSummary (p. 1583)
• InferenceComponentSpecification (p. 1584)
• InferenceComponentSpecificationSummary (p. 1585)
• InferenceComponentStartupParameters (p. 1586)
• InferenceComponentSummary (p. 1587)
• InferenceExecutionConfig (p. 1589)
• InferenceExperimentDataStorageConfig (p. 1590)
• InferenceExperimentSchedule (p. 1591)
• InferenceExperimentSummary (p. 1592)
• InferenceMetrics (p. 1594)
• InferenceRecommendation (p. 1595)
• InferenceRecommendationsJob (p. 1597)
• InferenceRecommendationsJobStep (p. 1600)
• InferenceSpecification (p. 1601)
• InfraCheckConfig (p. 1604)
• InputConfig (p. 1605)
• InstanceGroup (p. 1609)
• InstanceMetadataServiceConfiguration (p. 1611)
• IntegerParameterRange (p. 1612)
• IntegerParameterRangeSpecification (p. 1614)
• JupyterLabAppImageConfig (p. 1615)
• JupyterLabAppSettings (p. 1616)
• JupyterServerAppSettings (p. 1618)
• KendraSettings (p. 1619)
• KernelGatewayAppSettings (p. 1620)
• KernelGatewayImageConfig (p. 1621)
• KernelSpec (p. 1622)
• LabelCounters (p. 1623)
• LabelCountersForWorkteam (p. 1625)
• LabelingJobAlgorithmsConfig (p. 1626)
• LabelingJobDataAttributes (p. 1628)
• LabelingJobDataSource (p. 1629)
• LabelingJobForWorkteamSummary (p. 1630)
• LabelingJobInputConfig (p. 1632)
• LabelingJobOutput (p. 1633)
• LabelingJobOutputConfig (p. 1634)
• LabelingJobResourceConfig (p. 1636)
• LabelingJobS3DataSource (p. 1637)
• LabelingJobSnsDataSource (p. 1638)
• LabelingJobStoppingConditions (p. 1639)
• LabelingJobSummary (p. 1640)
• LambdaStepMetadata (p. 1643)
• LastUpdateStatus (p. 1644)
• LineageGroupSummary (p. 1645)
• MemberDefinition (p. 1647)
• MetadataProperties (p. 1648)
• MetricData (p. 1650)
• MetricDatum (p. 1651)
• MetricDefinition (p. 1653)
• MetricSpecification (p. 1654)
• MetricsSource (p. 1655)
• Model (p. 1656)
• ModelAccessConfig (p. 1658)
• ModelArtifacts (p. 1659)
• ModelBiasAppSpecification (p. 1660)
• ModelBiasBaselineConfig (p. 1661)
• ModelBiasJobInput (p. 1662)
• ModelCard (p. 1663)
• ModelCardExportArtifacts (p. 1666)
• ModelCardExportJobSummary (p. 1667)
• ModelCardExportOutputConfig (p. 1669)
• ModelCardSecurityConfig (p. 1670)
• ModelCardSummary (p. 1671)
• ModelCardVersionSummary (p. 1673)
• ModelClientConfig (p. 1675)
• ModelConfiguration (p. 1676)
• ModelDashboardEndpoint (p. 1677)
• ModelDashboardIndicatorAction (p. 1679)
• ModelDashboardModel (p. 1680)
• ModelDashboardModelCard (p. 1681)
• ModelDashboardMonitoringSchedule (p. 1684)
• ModelDataQuality (p. 1687)
• ModelDataSource (p. 1688)
• ModelDeployConfig (p. 1689)
• ModelDeployResult (p. 1690)
• ModelDigests (p. 1691)
• ModelExplainabilityAppSpecification (p. 1692)
• ModelExplainabilityBaselineConfig (p. 1693)
- ModelExplainabilityJobInput (p. 1694)
- ModelInfrastructureConfig (p. 1695)
- ModelInput (p. 1696)
- ModelLatencyThreshold (p. 1697)
- ModelMetadataFilter (p. 1698)
- ModelMetadataSearchExpression (p. 1699)
- ModelMetadataSummary (p. 1700)
- ModelMetrics (p. 1701)
- ModelPackage (p. 1702)
- ModelPackageContainerDefinition (p. 1707)
- ModelPackageGroup (p. 1710)
- ModelPackageGroupId (p. 1712)
- ModelPackageStatusDetails (p. 1714)
- ModelPackageStatusItem (p. 1715)
- ModelPackageSummary (p. 1716)
- ModelPackageValidationProfile (p. 1718)
- ModelPackageValidationSpecification (p. 1719)
- ModelQuality (p. 1720)
- ModelQualityAppSpecification (p. 1721)
- ModelQualityBaselineConfig (p. 1723)
- ModelQualityJobInput (p. 1724)
- ModelRegisterSettings (p. 1725)
- ModelStepMetadata (p. 1726)
- ModelSummary (p. 1727)
- ModelVariantConfig (p. 1728)
- ModelVariantConfigSummary (p. 1729)
- MonitoringAlertActions (p. 1731)
- MonitoringAlertHistorySummary (p. 1732)
- MonitoringAlertSummary (p. 1733)
- MonitoringAppSpecification (p. 1735)
- MonitoringBaselineConfig (p. 1737)
- MonitoringClusterConfig (p. 1738)
- MonitoringConstraintsResource (p. 1740)
- MonitoringCsvDatasetFormat (p. 1741)
- MonitoringDatasetFormat (p. 1742)
- MonitoringExecutionSummary (p. 1743)
- MonitoringGroundTruthS3Input (p. 1745)
- MonitoringInput (p. 1746)
- MonitoringJobDefinition (p. 1747)
- MonitoringJobDefinitionSummary (p. 1749)
- MonitoringJsonDatasetFormat (p. 1751)
- MonitoringNetworkConfig (p. 1752)
- MonitoringOutput (p. 1753)
- MonitoringOutputConfig (p. 1754)
- MonitoringParquetDatasetFormat (p. 1755)
- MonitoringResources (p. 1756)
- `MonitoringS3Output` (p. 1757)
- `MonitoringSchedule` (p. 1758)
- `MonitoringScheduleConfig` (p. 1761)
- `MonitoringScheduleSummary` (p. 1762)
- `MonitoringStatisticsResource` (p. 1764)
- `MonitoringStoppingCondition` (p. 1765)
- `MultiModelConfig` (p. 1766)
- `NeoVpcConfig` (p. 1767)
- `NestedFilters` (p. 1768)
- `NetworkConfig` (p. 1769)
- `NotebookInstanceLifecycleConfigSummary` (p. 1770)
- `NotebookInstanceLifecycleHook` (p. 1771)
- `NotificationConfiguration` (p. 1775)
- `ObjectiveStatusCounters` (p. 1776)
- `OfflineStoreConfig` (p. 1777)
- `OfflineStoreStatus` (p. 1778)
- `OidcConfig` (p. 1779)
- `OidcConfigForResponse` (p. 1781)
- `OidcMemberDefinition` (p. 1783)
- `OnlineStoreConfig` (p. 1784)
- `OnlineStoreConfigUpdate` (p. 1786)
- `OnlineStoreSecurityConfig` (p. 1787)
- `OutputConfig` (p. 1788)
- `OutputDataConfig` (p. 1792)
- `OutputParameter` (p. 1794)
- `OwnershipSettings` (p. 1795)
- `OwnershipSettingsSummary` (p. 1796)
- `ParallelismConfiguration` (p. 1797)
- `Parameter` (p. 1798)
- `ParameterRange` (p. 1799)
- `ParameterRanges` (p. 1800)
- `Parent` (p. 1802)
- `ParentHyperParameterTuningJob` (p. 1803)
- `PendingDeploymentSummary` (p. 1804)
- `PendingProductionVariantSummary` (p. 1806)
- `Phase` (p. 1810)
- `Pipeline` (p. 1811)
- `PipelineDefinitionS3Location` (p. 1814)
- `PipelineExecution` (p. 1815)
- `PipelineExecutionStep` (p. 1818)
- `PipelineExecutionStepMetadata` (p. 1820)
- `PipelineExecutionSummary` (p. 1823)
- `PipelineExperimentConfig` (p. 1825)
- `PipelineSummary` (p. 1826)
- `PredefinedMetricSpecification` (p. 1828)
• ProcessingClusterConfig (p. 1829)
• ProcessingFeatureStoreOutput (p. 1831)
• ProcessingInput (p. 1832)
• ProcessingJob (p. 1833)
• ProcessingJobStepMetadata (p. 1837)
• ProcessingJobSummary (p. 1838)
• ProcessingOutput (p. 1840)
• ProcessingOutputConfig (p. 1841)
• ProcessingResources (p. 1842)
• ProcessingS3Input (p. 1843)
• ProcessingS3Output (p. 1845)
• ProcessingStoppingCondition (p. 1846)
• ProductionVariant (p. 1847)
• ProductionVariantCoreDumpConfig (p. 1851)
• ProductionVariantManagedInstanceScaling (p. 1853)
• ProductionVariantRoutingConfig (p. 1854)
• ProductionVariantServerlessConfig (p. 1855)
• ProductionVariantServerlessUpdateConfig (p. 1856)
• ProductionVariantStatus (p. 1857)
• ProductionVariantSummary (p. 1858)
• ProfilerConfig (p. 1860)
• ProfilerConfigForUpdate (p. 1862)
• ProfilerRuleConfiguration (p. 1864)
• ProfilerRuleEvaluationStatus (p. 1866)
• Project (p. 1868)
• ProjectSummary (p. 1871)
• PropertyNameQuery (p. 1873)
• PropertyNameSuggestion (p. 1874)
• ProvisioningParameter (p. 1875)
• PublicWorkforceTaskPrice (p. 1876)
• QualityCheckStepMetadata (p. 1879)
• QueryFilters (p. 1881)
• RealTimeInferenceConfig (p. 1883)
• RealTimeInferenceRecommendation (p. 1884)
• RecommendationJobCompiledOutputConfig (p. 1886)
• RecommendationJobContainerConfig (p. 1887)
• RecommendationJobInferenceBenchmark (p. 1890)
• RecommendationJobInputConfig (p. 1892)
• RecommendationJobOutputConfig (p. 1895)
• RecommendationJobPayloadConfig (p. 1896)
• RecommendationJobResourceLimit (p. 1897)
• RecommendationJobStoppingConditions (p. 1898)
• RecommendationJobVpcConfig (p. 1899)
• RecommendationMetrics (p. 1900)
• RedshiftDatasetDefinition (p. 1902)
• RegisterModelStepMetadata (p. 1904)
- RenderableTask (p. 1905)
- RenderingError (p. 1906)
- RepositoryAuthConfig (p. 1907)
- ResolvedAttributes (p. 1908)
- ResourceCatalog (p. 1909)
- ResourceConfig (p. 1910)
- ResourceConfigForUpdate (p. 1913)
- ResourceLimits (p. 1914)
- ResourceSpec (p. 1915)
- RetentionPolicy (p. 1917)
- RetryStrategy (p. 1918)
- RollingUpdatePolicy (p. 1919)
- RSessionAppSettings (p. 1920)
- RStudioServerProAppSettings (p. 1921)
- RStudioServerProDomainSettings (p. 1922)
- RStudioServerProDomainSettingsForUpdate (p. 1923)
- S3DataSource (p. 1924)
- S3ModelDataSource (p. 1927)
- S3StorageConfig (p. 1929)
- ScalingPolicy (p. 1931)
- ScalingPolicyMetric (p. 1932)
- ScalingPolicyObjective (p. 1933)
- ScheduleConfig (p. 1934)
- SearchExpression (p. 1936)
- SearchRecord (p. 1938)
- SecondaryStatusTransition (p. 1941)
- SelectedStep (p. 1943)
- SelectiveExecutionConfig (p. 1944)
- SelectiveExecutionResult (p. 1945)
- ServiceCatalogProvisionedProductDetails (p. 1946)
- ServiceCatalogProvisioningDetails (p. 1947)
- ServiceCatalogProvisioningUpdateDetails (p. 1949)
- ShadowModeConfig (p. 1950)
- ShadowModelVariantConfig (p. 1951)
- SharingSettings (p. 1952)
- ShuffleConfig (p. 1953)
- SourceAlgorithm (p. 1954)
- SourceAlgorithmSpecification (p. 1955)
- SourceIpConfig (p. 1956)
- SpaceCodeEditorAppSettings (p. 1957)
- SpaceDetails (p. 1958)
- SpaceJupyterLabAppSettings (p. 1960)
- SpaceSettings (p. 1961)
- SpaceSettingsSummary (p. 1963)
- SpaceSharingSettings (p. 1964)
- SpaceSharingSettingsSummary (p. 1965)
• **SpaceStorageSettings** (p. 1966)
• **Stairs** (p. 1967)
• **StoppingCondition** (p. 1968)
• **StudioLifecycleConfigDetails** (p. 1970)
• **SubscribedWorkteam** (p. 1972)
• **SuggestionQuery** (p. 1974)
• **TabularJobConfig** (p. 1975)
• **TabularResolvedAttributes** (p. 1978)
• **Tag** (p. 1979)
• **TargetPlatform** (p. 1980)
• **TargetTrackingScalingPolicyConfiguration** (p. 1982)
• **TensorBoardAppSettings** (p. 1983)
• **TensorBoardOutputConfig** (p. 1984)
• **TextClassificationJobConfig** (p. 1985)
• **TextGenerationJobConfig** (p. 1986)
• **TextGenerationResolvedAttributes** (p. 1988)
• **TimeSeriesConfig** (p. 1989)
• **TimeSeriesForecastingJobConfig** (p. 1991)
• **TimeSeriesForecastingSettings** (p. 1994)
• **TimeSeriesTransformations** (p. 1995)
• **TrafficPattern** (p. 1997)
• **TrafficRoutingConfig** (p. 1998)
• **TrainingImageConfig** (p. 2000)
• **TrainingJob** (p. 2001)
• **TrainingJobDefinition** (p. 2009)
• **TrainingJobStatusCounters** (p. 2011)
• **TrainingJobStepMetadata** (p. 2013)
• **TrainingJobSummary** (p. 2014)
• **TrainingRepositoryAuthConfig** (p. 2016)
• **TrainingSpecification** (p. 2017)
• **TransformDataSource** (p. 2019)
• **TransformInput** (p. 2020)
• **TransformJob** (p. 2022)
• **TransformJobDefinition** (p. 2027)
• **TransformJobStepMetadata** (p. 2029)
• **TransformJobSummary** (p. 2030)
• **TransformOutput** (p. 2032)
• **TransformResources** (p. 2034)
• **TransformS3DataSource** (p. 2036)
• **Trial** (p. 2038)
• **TrialComponent** (p. 2041)
• **TrialComponentArtifact** (p. 2045)
• **TrialComponentMetricSummary** (p. 2046)
• **TrialComponentParameterValue** (p. 2048)
• **TrialComponentSimpleSummary** (p. 2049)
• **TrialComponentSource** (p. 2051)
• TrialComponentSourceDetail (p. 2052)
• TrialComponentStatus (p. 2053)
• TrialComponentSummary (p. 2054)
• TrialSource (p. 2056)
• TrialSummary (p. 2057)
• TtlDuration (p. 2059)
• TuningJobCompletionCriteria (p. 2060)
• TuningJobStepMetaData (p. 2061)
• UiConfig (p. 2062)
• UiTemplate (p. 2064)
• UiTemplateInfo (p. 2065)
• USD (p. 2066)
• UserContext (p. 2067)
• UserProfileDetails (p. 2068)
• UserSettings (p. 2070)
• VariantProperty (p. 2073)
• VectorConfig (p. 2074)
• Vertex (p. 2075)
• VpcConfig (p. 2076)
• WarmPoolStatus (p. 2077)
• Workforce (p. 2079)
• WorkforceVpcConfigRequest (p. 2081)
• WorkforceVpcConfigResponse (p. 2082)
• WorkspaceSettings (p. 2084)
• Workteam (p. 2085)
**ActionSource**

Service: Amazon SageMaker Service

A structure describing the source of an action.

**Contents**

**SourceUri**

The URI of the source.

Type: String

Length Constraints: Maximum length of 2048.

Required: Yes

**SourceId**

The ID of the source.

Type: String

Length Constraints: Maximum length of 256.

Required: No

**SourceType**

The type of the source.

Type: String

Length Constraints: Maximum length of 256.

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
ActionSummary

Service: Amazon SageMaker Service

Lists the properties of an action. An action represents an action or activity. Some examples are a workflow step and a model deployment. Generally, an action involves at least one input artifact or output artifact.

Contents

ActionArn

The Amazon Resource Name (ARN) of the action.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-zA-Z-]*:sagemaker:[a-zA-Z0-9-]*:[0-9]{12}:action/.*

Required: No

ActionName

The name of the action.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: ^[a-zA-Z-]{0,128}([-][a-zA-Z0-9]{0,128}){0,119}$

Required: No

ActionType

The type of the action.

Type: String

Length Constraints: Maximum length of 64.

Required: No

CreationTime

When the action was created.

Type: Timestamp

Required: No

LastModifiedTime

When the action was last modified.

Type: Timestamp

Required: No

Source

The source of the action.

Type: ActionSource (p. 1216) object
ActionSummary

Required: No

**Status**

The status of the action.

Type: String

Valid Values: Unknown | InProgress | Completed | Failed | Stopping | Stopped

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
**AdditionalInferenceSpecificationDefinition**

Service: Amazon SageMaker Service

A structure of additional Inference Specification. Additional Inference Specification specifies details about inference jobs that can be run with models based on this model package.

**Contents**

**Containers**

The Amazon ECR registry path of the Docker image that contains the inference code.

Type: Array of [ModelPackageContainerDefinition (p. 1707)] objects

Array Members: Minimum number of 1 item. Maximum number of 15 items.

Required: Yes

**Name**

A unique name to identify the additional inference specification. The name must be unique within the list of your additional inference specifications for a particular model package.

Type: String


Pattern: \^[a-zA-Z0-9](\-*[a-zA-Z0-9])\{0,62\}$

Required: Yes

**Description**

A description of the additional Inference specification

Type: String

Length Constraints: Maximum length of 1024.

Pattern: [\p{L}\p{M}\p{Z}\p{S}\p{N}\p{P}]*

Required: No

**SupportedContentTypes**

The supported MIME types for the input data.

Type: Array of strings

Length Constraints: Maximum length of 256.

Pattern: .*

Required: No

**SupportedRealtimeInferenceInstanceTypes**

A list of the instance types that are used to generate inferences in real-time.

Type: Array of strings

Valid Values: `ml.t2.medium` | `ml.t2.large` | `ml.t2.xlarge` | `ml.t2.2xlarge` | `ml.m4.xlarge` | `ml.m4.2xlarge` | `ml.m4.4xlarge` | `ml.m4.10xlarge` | `ml.m4.16xlarge` | `ml.m5.large` | `ml.m5.xlarge` | `ml.m5.2xlarge` | `ml.m5.4xlarge`
| ml.m5.12xlarge | ml.m5.24xlarge | ml.m5.d.large | ml.m5d.xlarge |
| ml.m5d.2xlarge | ml.m5d.4xlarge | ml.m5d.12xlarge | ml.m5d.24xlarge |
| ml.c4.large | ml.c4.xlarge | ml.c4.2xlarge | ml.c4.4xlarge | ml.c4.8xlarge |
| ml.p2.xlarge | ml.p2.8xlarge | ml.p2.16xlarge | ml.p3.2xlarge |
| ml.p3.8xlarge | ml.p3.16xlarge | ml.c5.large | ml.c5.xlarge | ml.c5.2xlarge |
| ml.c5.4xlarge | ml.c5.9xlarge | ml.c5.18xlarge | ml.c5d.large |
| ml.c5d.xlarge | ml.c5d.2xlarge | ml.c5d.4xlarge | ml.c5d.9xlarge |
| ml.g4dn.xlarge | ml.g4dn.2xlarge | ml.g4dn.12xlarge | ml.g4dn.16xlarge | ml.r5.large |
| ml.r5.xlarge | ml.r5.2xlarge | ml.r5.4xlarge | ml.r5.12xlarge |
| ml.r5.24xlarge | ml.r5d.large | ml.r5d.xlarge | ml.r5d.2xlarge |
| ml.r5d.4xlarge | ml.r5d.12xlarge | ml.r5d.24xlarge | ml.inf1.xlarge |
| ml.inf1.xlarge | ml.inf1.6xlarge | ml.inf1.24xlarge | ml.c6i.large |
| ml.c6i.large | ml.c6i.12xlarge | ml.c6i.16xlarge | ml.c6i.24xlarge |
| ml.c6i.4xlarge | ml.c6i.8xlarge | ml.c6i.32xlarge |
| ml.g5.xlarge | ml.g5.2xlarge | ml.g5.4xlarge | ml.g5.8xlarge |
| ml.g5.12xlarge | ml.g5.16xlarge | ml.g5.24xlarge | ml.g5.48xlarge |
| ml.p4d.24xlarge | ml.c7g.large | ml.c7g.xlarge | ml.c7g.2xlarge |
| ml.c7g.4xlarge | ml.c7g.8xlarge | ml.c7g.12xlarge | ml.c7g.16xlarge |
| ml.m6g.large | ml.m6g.xlarge | ml.m6g.2xlarge | ml.m6g.4xlarge |
| ml.m6g.8xlarge | ml.m6g.12xlarge | ml.m6g.16xlarge | ml.m6gd.large |
| ml.m6gd.xlarge | ml.m6gd.2xlarge | ml.m6gd.4xlarge | ml.m6gd.8xlarge |
| ml.m6gd.12xlarge | ml.m6gd.16xlarge | ml.c6g.large | ml.c6g.xlarge |
| ml.c6g.2xlarge | ml.c6g.4xlarge | ml.c6g.8xlarge | ml.c6g.12xlarge |
| ml.c6g.16xlarge | ml.c6gd.large | ml.c6gd.xlarge | ml.c6gd.2xlarge |
| ml.c6gd.4xlarge | ml.c6gd.8xlarge | ml.c6gd.12xlarge | ml.c6gd.16xlarge |
| ml.c6gn.large | ml.c6gn.xlarge | ml.c6gn.2xlarge | ml.c6gn.4xlarge |
| ml.c6gn.8xlarge | ml.c6gn.12xlarge | ml.c6gn.16xlarge | ml.r6g.large |
| ml.r6g.xlarge | ml.r6g.2xlarge | ml.r6g.4xlarge | ml.r6g.8xlarge |
| ml.r6g.12xlarge | ml.r6g.16xlarge | ml.r6gd.large | ml.r6gd.xlarge |
| ml.r6gd.2xlarge | ml.r6gd.4xlarge | ml.r6gd.8xlarge | ml.r6gd.12xlarge |
| ml.r6gd.16xlarge | ml.p4de.24xlarge | ml.trn1.2xlarge | ml.trn1.32xlarge |
| ml.inf2.xlarge | ml.inf2.8xlarge | ml.inf2.24xlarge | ml.inf2.48xlarge |
| ml.p5.48xlarge |

**SupportedResponseMIMETypes**

The supported MIME types for the output data.

**Type:** Array of strings

**Length Constraints:** Maximum length of 1024.

**Pattern:** `^[-\w]+/\w+$`

**Required:** No

**SupportedTransformInstanceTypes**

A list of the instance types on which a transformation job can be run or on which an endpoint can be deployed.

**Type:** Array of strings

**Array Members:** Minimum number of 1 item.

**Valid Values:**
- ml.m4.xlarge
- ml.m4.2xlarge
- ml.m4.4xlarge
- ml.m4.10xlarge
- ml.m4.16xlarge
- ml.c4.xlarge
- ml.c4.2xlarge
- ml.c4.4xlarge
- ml.p2.xlarge
- ml.p2.8xlarge
- ml.p2.16xlarge
- ml.p3.2xlarge
- ml.p3.8xlarge
- ml.p3.16xlarge
- ml.c5.large
- ml.c5.xlarge
- ml.c5.2xlarge
- ml.c5.4xlarge
- ml.c5.9xlarge
- ml.c5.18xlarge
- ml.c5d.large
- ml.c5d.xlarge
- ml.c5d.2xlarge
- ml.c5d.4xlarge
- ml.c5d.9xlarge
- ml.c5d.18xlarge
- ml.g4dn.xlarge
- ml.g4dn.2xlarge
- ml.g4dn.12xlarge
- ml.g4dn.16xlarge
- ml.g4dn.4xlarge
- ml.g4dn.8xlarge
- ml.r5.large
- ml.r5.xlarge
- ml.r5.2xlarge
- ml.r5.4xlarge
- ml.r5.12xlarge
- ml.r5.24xlarge
- ml.r5d.large
- ml.r5d.xlarge
- ml.r5d.2xlarge
- ml.r5d.4xlarge
- ml.r5d.12xlarge
- ml.r5d.24xlarge
- ml.inf1.xlarge
- ml.inf1.2xlarge
- ml.inf1.6xlarge
- ml.inf1.4xlarge
- ml.inf1.24xlarge
- ml.c6i.large
- ml.c6i.xlarge
- ml.c6i.12xlarge
- ml.c6i.16xlarge
- ml.c6i.24xlarge
- ml.c6i.4xlarge
- ml.c6i.8xlarge
- ml.c6i.32xlarge
- ml.g5.xlarge
- ml.g5.2xlarge
- ml.g5.4xlarge
- ml.g5.8xlarge
- ml.g5.12xlarge
- ml.g5.16xlarge
- ml.g5.24xlarge
- ml.g5.48xlarge
- ml.p4d.24xlarge
- ml.c7g.large
- ml.c7g.xlarge
- ml.c7g.2xlarge
- ml.c7g.4xlarge
- ml.c7g.8xlarge
- ml.c7g.12xlarge
- ml.c7g.16xlarge
- ml.m6g.large
- ml.m6g.xlarge
- ml.m6g.2xlarge
- ml.m6g.4xlarge
- ml.m6g.8xlarge
- ml.m6g.12xlarge
- ml.m6g.16xlarge
- ml.m6gd.large
- ml.m6gd.xlarge
- ml.m6gd.2xlarge
- ml.m6gd.4xlarge
- ml.m6gd.8xlarge
- ml.m6gd.12xlarge
- ml.m6gd.16xlarge
- ml.c6g.large
- ml.c6g.xlarge
- ml.c6g.2xlarge
- ml.c6g.4xlarge
- ml.c6g.8xlarge
- ml.c6g.12xlarge
- ml.c6g.16xlarge
- ml.c6gd.large
- ml.c6gd.xlarge
- ml.c6gd.2xlarge
- ml.c6gd.4xlarge
- ml.c6gd.8xlarge
- ml.c6gd.12xlarge
- ml.c6gd.16xlarge
- ml.c6gn.large
- ml.c6gn.xlarge
- ml.c6gn.2xlarge
- ml.c6gn.4xlarge
- ml.c6gn.8xlarge
- ml.c6gn.12xlarge
- ml.c6gn.16large
- ml.r6g.large
- ml.r6g.xlarge
- ml.r6g.2xlarge
- ml.r6g.4xlarge
- ml.r6g.8xlarge
- ml.r6g.12xlarge
- ml.r6g.16xlarge
- ml.r6gd.large
- ml.r6gd.xlarge
- ml.r6gd.2large
- ml.r6gd.4large
- ml.r6gd.8large
- ml.r6gd.12large
- ml.r6gd.16large
- ml.p4de.24large
- ml.trn1.2large
- ml.trn1.3large
- ml.inf2.xlarge
- ml.inf2.8large
- ml.inf2.4large
- ml.inf2.24large
- ml.inf2.48large
- ml.p5.48large
- ml.p4de.24large
- ml.trn1.2large
- ml.trn1.3large
- ml.inf2.xlarge
- ml.inf2.8large
- ml.inf2.4large
- ml.inf2.24large
- ml.inf2.48large
- ml.p5.48large
| ml.c4.8xlarge | ml.p2.xlarge | ml.p2.8xlarge | ml.p2.16xlarge |
| ml.p3.2xlarge | ml.p3.8xlarge | ml.p3.16xlarge | ml.c5.xlarge |
| ml.c5.2xlarge | ml.c5.4xlarge | ml.c5.9xlarge | ml.c5.18xlarge |
| ml.m5.xlarge | ml.m5.2xlarge | ml.m5.4xlarge | ml.m5.12xlarge |
| ml.m5.24xlarge | ml.g4dn.xlarge | ml.g4dn.2xlarge | ml.g4dn.4xlarge |
| ml.g4dn.8xlarge | ml.g4dn.12xlarge | ml.g4dn.16xlarge |

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
AdditionalS3DataSource

Service: Amazon SageMaker Service

A data source used for training or inference that is in addition to the input dataset or model data.

Contents

S3DataType

The data type of the additional data source that you specify for use in inference or training.

Type: String

Valid Values: S3Object

Required: Yes

S3Uri

The uniform resource identifier (URI) used to identify an additional data source used in inference or training.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: ^((https|s3):/(/[^/]+)\?([^]*)$)

Required: Yes

CompressionType

The type of compression used for an additional data source used in inference or training. Specify None if your additional data source is not compressed.

Type: String

Valid Values: None | Gzip

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
AgentVersion
Service: Amazon SageMaker Service

Edge Manager agent version.

Contents

AgentCount
The number of Edge Manager agents.
Type: Long
Required: Yes

Version
Version of the agent.
Type: String
Pattern: [a-zA-Z0-9\ _\ \_\ .]+
Required: Yes

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
Alarm

Service: Amazon SageMaker Service

An Amazon CloudWatch alarm configured to monitor metrics on an endpoint.

Contents

**AlarmName**

The name of a CloudWatch alarm in your account.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 255.

Pattern: ^(?![\s$]).+

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
AlgorithmSpecification
Service: Amazon SageMaker Service

Specifies the training algorithm to use in a CreateTrainingJob request.

For more information about algorithms provided by SageMaker, see Algorithms. For information about using your own algorithms, see Using Your Own Algorithms with Amazon SageMaker.

Contents

TrainingInputMode

The training input mode that the algorithm supports. For more information about input modes, see Algorithms.

Pipe mode

If an algorithm supports Pipe mode, Amazon SageMaker streams data directly from Amazon S3 to the container.

File mode

If an algorithm supports File mode, SageMaker downloads the training data from S3 to the provisioned ML storage volume, and mounts the directory to the Docker volume for the training container.

You must provision the ML storage volume with sufficient capacity to accommodate the data downloaded from S3. In addition to the training data, the ML storage volume also stores the output model. The algorithm container uses the ML storage volume to also store intermediate information, if any.

For distributed algorithms, training data is distributed uniformly. Your training duration is predictable if the input data objects sizes are approximately the same. SageMaker does not split the files any further for model training. If the object sizes are skewed, training won't be optimal as the data distribution is also skewed when one host in a training cluster is overloaded, thus becoming a bottleneck in training.

FastFile mode

If an algorithm supports FastFile mode, SageMaker streams data directly from S3 to the container with no code changes, and provides file system access to the data. Users can author their training script to interact with these files as if they were stored on disk.

FastFile mode works best when the data is read sequentially. Augmented manifest files aren't supported. The startup time is lower when there are fewer files in the S3 bucket provided.

Type: String
Valid Values: Pipe | File | FastFile
Required: Yes

AlgorithmName

The name of the algorithm resource to use for the training job. This must be an algorithm resource that you created or subscribe to on AWS Marketplace.

Note
You must specify either the algorithm name to the AlgorithmName parameter or the image URI of the algorithm container to the TrainingImage parameter.
Note that the AlgorithmName parameter is mutually exclusive with the TrainingImage parameter. If you specify a value for the AlgorithmName parameter, you can’t specify a value for TrainingImage, and vice versa.

If you specify values for both parameters, the training job might break; if you don’t specify any value for both parameters, the training job might raise a null error.

**Type:** String

**Length Constraints:** Minimum length of 1. Maximum length of 170.

**Pattern:** `(arn:aws[\w\-]*:sagemaker:[\w\-]*:[0-9]{12}:\w*/)?([\w\-]*\d+([-]*\d+){62}(?!-)$

**Required:** No

### ContainerArguments

The arguments for a container used to run a training job. See [How Amazon SageMaker Runs Your Training Image](#) for additional information.

**Type:** Array of strings

**Array Members:** Minimum number of 1 item. Maximum number of 100 items.

**Length Constraints:** Maximum length of 256.

**Pattern:** .*

**Required:** No

### ContainerEntrypoint

The **entrypoint script for a Docker container** used to run a training job. This script takes precedence over the default train processing instructions. See [How Amazon SageMaker Runs Your Training Image](#) for more information.

**Type:** Array of strings

**Array Members:** Minimum number of 1 item. Maximum number of 100 items.

**Length Constraints:** Maximum length of 256.

**Pattern:** .*

**Required:** No

### EnableSageMakerMetricsTimeSeries

To generate and save time-series metrics during training, set to `true`. The default is `false` and time-series metrics aren’t generated except in the following cases:

- You use one of the SageMaker built-in algorithms
- You use one of the following [Prebuilt SageMaker Docker Images](#):
  - Tensorflow (version >= 1.15)
  - MXNet (version >= 1.6)
  - PyTorch (version >= 1.3)
- You specify at least one [MetricDefinition](#)

**Type:** Boolean

**Required:** No
MetricDefinitions

A list of metric definition objects. Each object specifies the metric name and regular expressions used to parse algorithm logs. SageMaker publishes each metric to Amazon CloudWatch.

Type: Array of MetricDefinition (p. 1653) objects

Array Members: Minimum number of 0 items. Maximum number of 40 items.

Required: No

TrainingImage

The registry path of the Docker image that contains the training algorithm. For information about docker registry paths for SageMaker built-in algorithms, see Docker Registry Paths and Example Code in the Amazon SageMaker developer guide. SageMaker supports both registry/repository[:tag] and registry/repository[@digest] image path formats. For more information about using your custom training container, see Using Your Own Algorithms with Amazon SageMaker.

Note
You must specify either the algorithm name to the AlgorithmName parameter or the image URI of the algorithm container to the TrainingImage parameter.

For more information, see the note in the AlgorithmName parameter description.

Type: String

Length Constraints: Maximum length of 255.

Pattern: .*

Required: No

TrainingImageConfig

The configuration to use an image from a private Docker registry for a training job.

Type: TrainingImageConfig (p. 2000) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
AlgorithmStatusDetails

Service: Amazon SageMaker Service

Specifies the validation and image scan statuses of the algorithm.

Contents

**ImageScanStatuses**

The status of the scan of the algorithm’s Docker image container.

Type: Array of [AlgorithmStatusItem (p. 1229)] objects

Required: No

**ValidationStatuses**

The status of algorithm validation.

Type: Array of [AlgorithmStatusItem (p. 1229)] objects

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
AlgorithmStatusItem
Service: Amazon SageMaker Service
Represents the overall status of an algorithm.

Contents

Name
The name of the algorithm for which the overall status is being reported.
Type: String
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}\$
Required: Yes

Status
The current status.
Type: String
Valid Values: NotStarted | InProgress | Completed | Failed
Required: Yes

FailureReason
if the overall status is Failed, the reason for the failure.
Type: String
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
AlgorithmSummary
Service: Amazon SageMaker Service

Provides summary information about an algorithm.

Contents
AlgorithmArn
The Amazon Resource Name (ARN) of the algorithm.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 2048.
Pattern: ^arn:aws(-cn|-us-gov)?:sagemaker:[a-z0-9\-]{9,16}:[0-9]\{12\}:algorithm/[\S]{1,2048}$
Required: Yes

AlgorithmName
The name of the algorithm that is described by the summary.
Type: String
Pattern: ^[a-zA-Z0-9\-\(\*\)[a-zA-Z0-9]\{0,62\}$
Required: Yes

AlgorithmStatus
The overall status of the algorithm.
Type: String
Valid Values: Pending | InProgress | Completed | Failed | Deleting
Required: Yes

CreationTime
A timestamp that shows when the algorithm was created.
Type: Timestamp
Required: Yes

AlgorithmDescription
A brief description of the algorithm.
Type: String
Length Constraints: Maximum length of 1024.
Pattern: \[\p{L}\p{M}\p{Z}\p{S}\p{N}\p{P}]\*
Required: No
See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
AlgorithmValidationProfile

Service: Amazon SageMaker Service

Defines a training job and a batch transform job that SageMaker runs to validate your algorithm.

The data provided in the validation profile is made available to your buyers on AWS Marketplace.

**Contents**

**ProfileName**

The name of the profile for the algorithm. The name must have 1 to 63 characters. Valid characters are a-z, A-Z, 0-9, and - (hyphen).

Type: String


Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$`

Required: Yes

**TrainingJobDefinition**

The TrainingJobDefinition object that describes the training job that SageMaker runs to validate your algorithm.

Type: TrainingJobDefinition (p. 2009) object

Required: Yes

**TransformJobDefinition**

The TransformJobDefinition object that describes the transform job that SageMaker runs to validate your algorithm.

Type: TransformJobDefinition (p. 2027) object

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
AlgorithmValidationSpecification
Service: Amazon SageMaker Service
Specifies configurations for one or more training jobs that SageMaker runs to test the algorithm.

Contents

ValidationProfiles
An array of AlgorithmValidationProfile objects, each of which specifies a training job and batch transform job that SageMaker runs to validate your algorithm.
Type: Array of AlgorithmValidationProfile objects
Array Members: Fixed number of 1 item.
Required: Yes

ValidationRole
The IAM roles that SageMaker uses to run the training jobs.
Type: String
Pattern: ^arn:aws[a-z\-]*:iam::\d{12}:role/?[a-zA-Z_0-9+=,.@\-_]+$
Required: Yes

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
AnnotationConsolidationConfig
Service: Amazon SageMaker Service

Configures how labels are consolidated across human workers and processes output data.

Contents

AnnotationConsolidationLambdaArn

The Amazon Resource Name (ARN) of a Lambda function implements the logic for annotation consolidation and to process output data.

This parameter is required for all labeling jobs. For built-in task types, use one of the following Amazon SageMaker Ground Truth Lambda function ARNs for AnnotationConsolidationLambdaArn. For custom labeling workflows, see Post-annotation Lambda.

**Bounding box** - Finds the most similar boxes from different workers based on the Jaccard index of the boxes.

- \texttt{arn:aws:lambda:us-east-1:432418664414:function:ACS-BoundingBox}
- \texttt{arn:aws:lambda:us-east-2:266458841044:function:ACS-BoundingBox}
- \texttt{arn:aws:lambda:us-west-2:081040173940:function:ACS-BoundingBox}
- \texttt{arn:aws:lambda:eu-west-1:568282634449:function:ACS-BoundingBox}
- \texttt{arn:aws:lambda:ap-northeast-1:477331159723:function:ACS-BoundingBox}
- \texttt{arn:aws:lambda:ap-south-1:565803892007:function:ACS-BoundingBox}
- \texttt{arn:aws:lambda:eu-central-1:203001061592:function:ACS-BoundingBox}
- \texttt{arn:aws:lambda:ap-northeast-2:845288260483:function:ACS-BoundingBox}
- \texttt{arn:aws:lambda:eu-west-2:487402164563:function:ACS-BoundingBox}
- \texttt{arn:aws:lambda:ap-south-1:377565633583:function:ACS-BoundingBox}
- \texttt{arn:aws:lambda:ca-central-1:918755190332:function:ACS-BoundingBox}

**Image classification** - Uses a variant of the Expectation Maximization approach to estimate the true class of an image based on annotations from individual workers.

- \texttt{arn:aws:lambda:us-east-1:432418664414:function:ACS-ImageMultiClass}
- \texttt{arn:aws:lambda:us-east-2:266458841044:function:ACS-ImageMultiClass}
- \texttt{arn:aws:lambda:us-west-2:081040173940:function:ACS-ImageMultiClass}
- \texttt{arn:aws:lambda:eu-west-1:568282634449:function:ACS-ImageMultiClass}
- \texttt{arn:aws:lambda:ap-northeast-2:454466003867:function:ACS-ImageMultiClass}
- \texttt{arn:aws:lambda:ap-south-1:565803892007:function:ACS-ImageMultiClass}
- \texttt{arn:aws:lambda:eu-central-1:203001061592:function:ACS-ImageMultiClass}
- \texttt{arn:aws:lambda:ap-northeast-2:845288260483:function:ACS-ImageMultiClass}
- \texttt{arn:aws:lambda:eu-west-2:487402164563:function:ACS-ImageMultiClass}
- \texttt{arn:aws:lambda:ap-south-1:377565633583:function:ACS-ImageMultiClass}
- \texttt{arn:aws:lambda:ca-central-1:918755190332:function:ACS-ImageMultiClass}

**Multi-label image classification** - Uses a variant of the Expectation Maximization approach to estimate the true classes of an image based on annotations from individual workers.
• arn:aws:lambda:us-east-1:432418664414:function:ACS-ImageMultiClassMultiLabel
• arn:aws:lambda:eu-west-1:568282634449:function:ACS-ImageMultiClassMultiLabel
• arn:aws:lambda:ap-northeast-1:477331159723:function:ACS-ImageMultiClassMultiLabel
• arn:aws:lambda:ap-south-1:565803892007:function:ACS-ImageMultiClassMultiLabel
• arn:aws:lambda:eu-central-1:203001061592:function:ACS-ImageMultiClassMultiLabel
• arn:aws:lambda:eu-west-2:487402164563:function:ACS-ImageMultiClassMultiLabel
• arn:aws:lambda:ap-southeast-1:377565633583:function:ACS-ImageMultiClassMultiLabel
• arn:aws:lambda:ca-central-1:918755190332:function:ACS-ImageMultiClassMultiLabel

**Semantic segmentation** - Treats each pixel in an image as a multi-class classification and treats pixel annotations from workers as "votes" for the correct label.
• arn:aws:lambda:us-east-1:432418664414:function:ACS-SemanticSegmentation
• arn:aws:lambda:eu-west-1:568282634449:function:ACS-SemanticSegmentation
• arn:aws:lambda:ap-northeast-1:477331159723:function:ACS-SemanticSegmentation
• arn:aws:lambda:ap-south-1:565803892007:function:ACS-SemanticSegmentation
• arn:aws:lambda:eu-central-1:203001061592:function:ACS-SemanticSegmentation
• arn:aws:lambda:eu-west-2:487402164563:function:ACS-SemanticSegmentation
• arn:aws:lambda:ap-southeast-1:377565633583:function:ACS-SemanticSegmentation

**Text classification** - Uses a variant of the Expectation Maximization approach to estimate the true class of text based on annotations from individual workers.
• arn:aws:lambda:us-east-1:432418664414:function:ACS-TextMultiClass
• arn:aws:lambda:us-east-2:266458841044:function:ACS-TextMultiClass
• arn:aws:lambda:us-west-2:081040173940:function:ACS-TextMultiClass
• arn:aws:lambda:eu-west-1:568282634449:function:ACS-TextMultiClass
Multi-label text classification - Uses a variant of the Expectation Maximization approach to estimate the true classes of text based on annotations from individual workers.

Named entity recognition - Groups similar selections and calculates aggregate boundaries, resolving to most-assigned label.

**Video Classification** - Use this task type when you need workers to classify videos using predefined labels that you specify. Workers are shown videos and are asked to choose one label for each video.

• arn:aws:lambda:us-east-1:432418664414:function:ACS-VideoMultiClass
• arn:aws:lambda:us-east-2:266458841044:function:ACS-VideoMultiClass
• arn:aws:lambda:us-west-2:081040173940:function:ACS-VideoMultiClass
• arn:aws:lambda:eu-west-1:568282634449:function:ACS-VideoMultiClass
• arn:aws:lambda:ap-northeast-1:477331159723:function:ACS-VideoMultiClass
• arn:aws:lambda:ap-southeast-2:454466003867:function:ACS-VideoMultiClass
• arn:aws:lambda:ap-south-1:565803892007:function:ACS-VideoMultiClass
• arn:aws:lambda:eu-central-1:203001061592:function:ACS-VideoMultiClass
• arn:aws:lambda:ap-northeast-2:845288260483:function:ACS-VideoMultiClass
• arn:aws:lambda:eu-west-2:487402164563:function:ACS-VideoMultiClass
• arn:aws:lambda:ap-southeast-1:377565633583:function:ACS-VideoMultiClass
• arn:aws:lambda:ca-central-1:918755190332:function:ACS-VideoMultiClass

**Video Frame Object Detection** - Use this task type to have workers identify and locate objects in a sequence of video frames (images extracted from a video) using bounding boxes. For example, you can use this task to ask workers to identify and localize various objects in a series of video frames, such as cars, bikes, and pedestrians.

• arn:aws:lambda:us-east-1:432418664414:function:ACS-VideoObjectDetection
• arn:aws:lambda:eu-west-1:568282634449:function:ACS-VideoObjectDetection
• arn:aws:lambda:ap-northeast-1:477331159723:function:ACS-VideoObjectDetection
• arn:aws:lambda:ap-south-1:565803892007:function:ACS-VideoObjectDetection
• arn:aws:lambda:eu-central-1:203001061592:function:ACS-VideoObjectDetection
• arn:aws:lambda:eu-west-2:487402164563:function:ACS-VideoObjectDetection
• arn:aws:lambda:ap-southeast-1:377565633583:function:ACS-VideoObjectDetection
• arn:aws:lambda:ca-central-1:918755190332:function:ACS-VideoObjectDetection

**Video Frame Object Tracking** - Use this task type to have workers track the movement of objects in a sequence of video frames (images extracted from a video) using bounding boxes. For example, you can use this task to ask workers to track the movement of objects, such as cars, bikes, and pedestrians.

• arn:aws:lambda:us-east-1:432418664414:function:ACS-VideoObjectTracking

1237
3D Point Cloud Object Detection - Use this task type when you want workers to classify objects in a 3D point cloud by drawing 3D cuboids around objects. For example, you can use this task type to ask workers to identify different types of objects in a point cloud, such as cars, bikes, and pedestrians.

- `arn:aws:lambda:eu-west-1:568282634449:function:ACS-3DPointCloudObjectDetection`

3D Point Cloud Object Tracking - Use this task type when you want workers to draw 3D cuboids around objects that appear in a sequence of 3D point cloud frames. For example, you can use this task type to ask workers to track the movement of vehicles across multiple point cloud frames.

- `arn:aws:lambda:us-east-1:432418664414:function:ACS-3DPointCloudObjectTracking`
• arn:aws:lambda:us-west-2:081040173940:function:ACS-3DPointCloudObjectTracking
• arn:aws:lambda:eu-west-1:568282634449:function:ACS-3DPointCloudObjectTracking
• arn:aws:lambda:ap-northeast-1:477331159723:function:ACS-3DPointCloudObjectTracking
• arn:aws:lambda:ap-southeast-2:454466003867:function:ACS-3DPointCloudObjectTracking
• arn:aws:lambda:ap-south-1:565803892007:function:ACS-3DPointCloudObjectTracking
• arn:aws:lambda:eu-central-1:203001061592:function:ACS-3DPointCloudObjectTracking
• arn:aws:lambda:ap-northeast-2:845288260483:function:ACS-3DPointCloudObjectTracking
• arn:aws:lambda:eu-west-2:487402164563:function:ACS-3DPointCloudObjectTracking
• arn:aws:lambda:ap-southeast-1:377565633583:function:ACS-3DPointCloudObjectTracking
• arn:aws:lambda:ca-central-1:918755190332:function:ACS-3DPointCloudObjectTracking

3D Point Cloud Semantic Segmentation - Use this task type when you want workers to create a point-level semantic segmentation masks by painting objects in a 3D point cloud using different colors where each color is assigned to one of the classes you specify.

• arn:aws:lambda:us-east-1:432418664414:function:ACS-3DPointCloudSemanticSegmentation
• arn:aws:lambda:us-east-2:266458841044:function:ACS-3DPointCloudSemanticSegmentation
• arn:aws:lambda:eu-west-1:568282634449:function:ACS-3DPointCloudSemanticSegmentation
• arn:aws:lambda:ap-northeast-1:477331159723:function:ACS-3DPointCloudSemanticSegmentation
• arn:aws:lambda:ap-south-1:565803892007:function:ACS-3DPointCloudSemanticSegmentation
• arn:aws:lambda:eu-central-1:203001061592:function:ACS-3DPointCloudSemanticSegmentation
• arn:aws:lambda:eu-west-2:487402164563:function:ACS-3DPointCloudSemanticSegmentation
• arn:aws:lambda:ap-southeast-1:377565633583:function:ACS-3DPointCloudSemanticSegmentation
• arn:aws:lambda:ca-central-1:918755190332:function:ACS-3DPointCloudSemanticSegmentation

Use the following ARNs for Label Verification and Adjustment Jobs
Use label verification and adjustment jobs to review and adjust labels. To learn more, see Verify and Adjust Labels.

**Semantic Segmentation Adjustment** - Treats each pixel in an image as a multi-class classification and treats pixel adjusted annotations from workers as "votes" for the correct label.


**Semantic Segmentation Verification** - Uses a variant of the Expectation Maximization approach to estimate the true class of verification judgment for semantic segmentation labels based on annotations from individual workers.

• arn:aws:lambda:ap-southeast-1:377565633583:function:ACS-VerificationSemanticSegmentation

Bounding Box Adjustment - Finds the most similar boxes from different workers based on the Jaccard index of the adjusted annotations.
• arn:aws:lambda:us-east-1:432418664414:function:ACS-AdjustmentBoundingBox
• arn:aws:lambda:us-east-2:266458841044:function:ACS-AdjustmentBoundingBox
• arn:aws:lambda:us-west-2:081040173940:function:ACS-AdjustmentBoundingBox
• arn:aws:lambda:eu-west-1:568282634449:function:ACS-AdjustmentBoundingBox
• arn:aws:lambda:ap-northeast-1:477331159723:function:ACS-AdjustmentBoundingBox
• arn:aws:lambda:ap-southeast-2:454466003867:function:ACS-AdjustmentBoundingBox
• arn:aws:lambda:ap-south-1:565803892007:function:ACS-AdjustmentBoundingBox
• arn:aws:lambda:eu-central-1:203001061592:function:ACS-AdjustmentBoundingBox
• arn:aws:lambda:ap-northeast-2:2845288260483:function:ACS-AdjustmentBoundingBox
• arn:aws:lambda:eu-west-2:487402164563:function:ACS-AdjustmentBoundingBox
• arn:aws:lambda:ap-southeast-1:377565633583:function:ACS-AdjustmentBoundingBox
• arn:aws:lambda:ca-central-1:918755190332:function:ACS-AdjustmentBoundingBox

Bounding Box Verification - Uses a variant of the Expectation Maximization approach to estimate the true class of verification judgement for bounding box labels based on annotations from individual workers.
• arn:aws:lambda:us-east-1:432418664414:function:ACS-VerificationBoundingBox
• arn:aws:lambda:us-east-2:266458841044:function:ACS-VerificationBoundingBox
• arn:aws:lambda:us-west-2:081040173940:function:ACS-VerificationBoundingBox
• arn:aws:lambda:eu-west-1:568282634449:function:ACS-VerificationBoundingBox
• arn:aws:lambda:ap-northeast-1:477331159723:function:ACS-VerificationBoundingBox
• arn:aws:lambda:ap-southeast-2:454466003867:function:ACS-VerificationBoundingBox
• arn:aws:lambda:ap-south-1:565803892007:function:ACS-VerificationBoundingBox
• arn:aws:lambda:eu-central-1:203001061592:function:ACS-VerificationBoundingBox
• arn:aws:lambda:ap-northeast-2:2845288260483:function:ACS-VerificationBoundingBox
• arn:aws:lambda:eu-west-2:487402164563:function:ACS-VerificationBoundingBox
• arn:aws:lambda:ap-southeast-1:377565633583:function:ACS-VerificationBoundingBox
• arn:aws:lambda:ca-central-1:918755190332:function:ACS-VerificationBoundingBox
**Video Frame Object Detection Adjustment** - Use this task type when you want workers to adjust bounding boxes that workers have added to video frames to classify and localize objects in a sequence of video frames.


**Video Frame Object Tracking Adjustment** - Use this task type when you want workers to adjust bounding boxes that workers have added to video frames to track object movement across a sequence of video frames.

- arn:aws:lambda:eu-west-1:568282634449:function:ACS-AdjustmentVideoObjectTracking
• arn:aws:lambda:ap-southeast-1:377565633583:function:ACS-AdjustmentVideoObjectTracking
• arn:aws:lambda:ca-central-1:918755190332:function:ACS-AdjustmentVideoObjectTracking

3D Point Cloud Object Detection Adjustment - Use this task type when you want workers to adjust 3D cuboids around objects in a 3D point cloud.
• arn:aws:lambda:us-east-1:432418664414:function:ACS-Adjustment3DPointCloudObjectDetection
• arn:aws:lambda:eu-west-1:568282634449:function:ACS-Adjustment3DPointCloudObjectDetection
• arn:aws:lambda:ap-northeast-1:477331159723:function:ACS-Adjustment3DPointCloudObjectDetection
• arn:aws:lambda:ap-south-1:565803892007:function:ACS-Adjustment3DPointCloudObjectDetection
• arn:aws:lambda:eu-central-1:203001061592:function:ACS-Adjustment3DPointCloudObjectDetection
• arn:aws:lambda:eu-west-2:487402164563:function:ACS-Adjustment3DPointCloudObjectDetection
• arn:aws:lambda:ap-southeast-1:377565633583:function:ACS-Adjustment3DPointCloudObjectDetection

3D Point Cloud Object Tracking Adjustment - Use this task type when you want workers to adjust 3D cuboids around objects that appear in a sequence of 3D point cloud frames.
• arn:aws:lambda:us-east-1:432418664414:function:ACS-Adjustment3DPointCloudObjectTracking
• arn:aws:lambda:us-east-2:266458841044:function:ACS-Adjustment3DPointCloudObjectTracking
• arn:aws:lambda:us-west-2:081040173940:function:ACS-Adjustment3DPointCloudObjectTracking
• arn:aws:lambda:eu-west-1:568282634449:function:ACS-Adjustment3DPointCloudObjectTracking
• arn:aws:lambda:ap-northeast-1:477331159723:function:ACS-Adjustment3DPointCloudObjectTracking
• arn:aws:lambda:ap-south-1:565803892007:function:ACS-Adjustment3DPointCloudObjectTracking
• arn:aws:lambda:eu-central-1:203001061592:function:ACS-Adjustment3DPointCloudObjectTracking
• arn:aws:lambda:ap-northeast-2:845288260483:function:ACS-Adjustment3DPointCloudObjectTracking
• arn:aws:lambda:eu-west-2:487402164563:function:ACS-Adjustment3DPointCloudObjectTracking
• arn:aws:lambda:ap-southeast-1:377565633583:function:ACS-Adjustment3DPointCloudObjectTracking
• arn:aws:lambda:ca-central-1:918755190332:function:ACS-Adjustment3DPointCloudObjectTracking

3D Point Cloud Semantic Segmentation Adjustment - Use this task type when you want workers to adjust a point-level semantic segmentation masks using a paint tool.
• arn:aws:lambda:us-east-1:432418664414:function:ACS-3DPointCloudSemanticSegmentation
• arn:aws:lambda:us-east-1:432418664414:function:ACS-Adjustment3DPointCloudSemanticSegmentation
• arn:aws:lambda:eu-west-1:568282634449:function:ACS-Adjustment3DPointCloudSemanticSegmentation
• arn:aws:lambda:ap-northeast-1:477331159723:function:ACS-Adjustment3DPointCloudSemanticSegmentation
• arn:aws:lambda:ap-south-1:565803892007:function:ACS-Adjustment3DPointCloudSemanticSegmentation
• arn:aws:lambda:eu-central-1:203001061592:function:ACS-Adjustment3DPointCloudSemanticSegmentation
• arn:aws:lambda:eu-west-2:487402164563:function:ACS-Adjustment3DPointCloudSemanticSegmentation
• arn:aws:lambda:ap-southeast-1:377565633583:function:ACS-Adjustment3DPointCloudSemanticSegmentation
• arn:aws:lambda:ca-central-1:918755190332:function:ACS-Adjustment3DPointCloudSemanticSegmentation

Type: String
Length Constraints: Maximum length of 2048.
Pattern: arn:[a-z-]*:lambda:[a-z0-9-]*:[0-9]{12}:function:.*
Required: Yes

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:
• AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for Ruby V3
AppDetails

Service: Amazon SageMaker Service

Details about an Amazon SageMaker app.

Contents

AppName

The name of the app.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62\}

Required: No

AppType

The type of app.

Type: String

Valid Values: JupyterServer | KernelGateway | TensorBoard | RStudioServerPro | RSessionGateway | JupyterLab | CodeEditor

Required: No

CreationTime

The creation time.

Type: Timestamp

Required: No

DomainId

The domain ID.

Type: String

Length Constraints: Maximum length of 63.

Required: No

ResourceSpec

Specifies the ARN's of a SageMaker image and SageMaker image version, and the instance type that the version runs on.

Type: ResourceSpec (p. 1915) object

Required: No

SpaceName

The name of the space.

Type: String

Length Constraints: Maximum length of 63.
AppDetails

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}
Required: No

Status

The status.
Type: String
Valid Values: Deleted | Deleting | Failed | InService | Pending
Required: No

UserProfileName

The user profile name.
Type: String
Length Constraints: Maximum length of 63.
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
AppImageConfigDetails

Service: Amazon SageMaker Service

The configuration for running a SageMaker image as a KernelGateway app.

Contents

AppImageConfigArn

The Amazon Resource Name (ARN) of the AppImageConfig.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:app-image-config/.*

Required: No

AppImageConfigName

The name of the AppImageConfig. Must be unique to your account.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9\-](-*[a-zA-Z0-9]\{0,62\}$

Required: No

CreationTime

When the AppImageConfig was created.

Type: Timestamp

Required: No

JupyterLabAppImageConfig

The configuration for the file system and the runtime, such as the environment variables and entry point.

Type: JupyterLabAppImageConfig (p. 1615) object

Required: No

KernelGatewayImageConfig

The configuration for the file system and kernels in the SageMaker image.

Type: KernelGatewayImageConfig (p. 1621) object

Required: No

LastModifiedTime

When the AppImageConfig was last modified.

Type: Timestamp

Required: No
See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
AppSpecification

Service: Amazon SageMaker Service

Configuration to run a processing job in a specified container image.

Contents

ImageUri

The container image to be run by the processing job.

Type: String

Length Constraints: Maximum length of 255.

Pattern: . *

Required: Yes

ContainerArguments

The arguments for a container used to run a processing job.

Type: Array of strings

Array Members: Minimum number of 1 item. Maximum number of 100 items.

Length Constraints: Maximum length of 256.

Pattern: . *

Required: No

ContainerEntrypoint

The entrypoint for a container used to run a processing job.

Type: Array of strings

Array Members: Minimum number of 1 item. Maximum number of 100 items.

Length Constraints: Maximum length of 256.

Pattern: . *

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ArtifactSource
Service: Amazon SageMaker Service
A structure describing the source of an artifact.

Contents

SourceUri
The URI of the source.
Type: String
Length Constraints: Maximum length of 2048.
Required: Yes

SourceTypes
A list of source types.
Type: Array of ArtifactSourceType (p. 1252) objects
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ArtifactSourceType
Service: Amazon SageMaker Service
The ID and ID type of an artifact source.

Contents

SourceldType
The type of ID.
Type: String
Valid Values: MD5Hash | S3ETag | S3Version | Custom
Required: Yes

Value
The ID.
Type: String
Length Constraints: Maximum length of 256.
Required: Yes

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
ArtifactSummary
Service: Amazon SageMaker Service

Lists a summary of the properties of an artifact. An artifact represents a URI addressable object or data. Some examples are a dataset and a model.

Contents

ArtifactArn
The Amazon Resource Name (ARN) of the artifact.
Type: String
Length Constraints: Maximum length of 256.
Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:artifact/.*
Required: No

ArtifactName
The name of the artifact.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 120.
Pattern: ^[a-zA-Z0-9](-{2}[a-zA-Z0-9])\{0,119\}
Required: No

ArtifactType
The type of the artifact.
Type: String
Length Constraints: Maximum length of 256.
Required: No

CreationTime
When the artifact was created.
Type: Timestamp
Required: No

LastModifiedTime
When the artifact was last modified.
Type: Timestamp
Required: No

Source
The source of the artifact.
Type: ArtifactSource (p. 1251) object
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
AssociationSummary
Service: Amazon SageMaker Service

Lists a summary of the properties of an association. An association is an entity that links other lineage or experiment entities. An example would be an association between a training job and a model.

Contents

**AssociationType**

The type of the association.

Type: String

Valid Values: ContributedTo | AssociatedWith | DerivedFrom | Produced

Required: No

**CreatedBy**

Information about the user who created or modified an experiment, trial, trial component, lineage group, project, or model card.

Type: UserContext (p. 2067) object

Required: No

**CreationTime**

When the association was created.

Type: Timestamp

Required: No

**DestinationArn**

The Amazon Resource Name (ARN) of the destination.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-zA-Z\-]*:sagemaker:[a-zA-Z0-9\-]*:[0-9]{12}:(experiment|experiment-trial-component|artifact|action|context)/.*

Required: No

**DestinationName**

The name of the destination.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: ^[a-zA-Z0-9](\*[^a-zA-Z0-9])\{0,119}$

Required: No

**DestinationType**

The destination type.
Type: String
Length Constraints: Maximum length of 256.
Required: No

SourceArn
The ARN of the source.
Type: String
Length Constraints: Maximum length of 256.
Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:\{(experiment|experiment-trial-component|artifact|action|context)/.*
Required: No

SourceName
The name of the source.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 120.
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,119
Required: No

SourceType
The source type.
Type: String
Length Constraints: Maximum length of 256.
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
AsyncInferenceClientConfig

Service: Amazon SageMaker Service

Configures the behavior of the client used by SageMaker to interact with the model container during asynchronous inference.

Contents

MaxConcurrentInvocationsPerInstance

The maximum number of concurrent requests sent by the SageMaker client to the model container. If no value is provided, SageMaker chooses an optimal value.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 1000.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
AsyncInferenceConfig

Service: Amazon SageMaker Service

Specifies configuration for how an endpoint performs asynchronous inference.

Contents

OutputConfig

Specifies the configuration for asynchronous inference invocation outputs.

Type: AsyncInferenceOutputConfig (p. 1260) object

Required: Yes

ClientConfig

Configures the behavior of the client used by SageMaker to interact with the model container during asynchronous inference.

Type: AsyncInferenceClientConfig (p. 1257) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
AsyncInferenceNotificationConfig

Service: Amazon SageMaker Service

Specifies the configuration for notifications of inference results for asynchronous inference.

Contents

ErrorTopic

Amazon SNS topic to post a notification to when inference fails. If no topic is provided, no notification is sent on failure.

Type: String

Length Constraints: Maximum length of 2048.

Pattern: arn:aws[a-z\-]*:sns:[a-z0-9\-]*:[0-9]{12}:[a-zA-Z0-9_.-]+

Required: No

IncludeInferenceResponseIn

The Amazon SNS topics where you want the inference response to be included.

Note

The inference response is included only if the response size is less than or equal to 128 KB.

Type: Array of strings

Array Members: Minimum number of 0 items. Maximum number of 2 items.

Valid Values: SUCCESS_NOTIFICATION_TOPIC | ERROR_NOTIFICATION_TOPIC

Required: No

SuccessTopic

Amazon SNS topic to post a notification to when inference completes successfully. If no topic is provided, no notification is sent on success.

Type: String

Length Constraints: Maximum length of 2048.

Pattern: arn:aws[a-z\-]*:sns:[a-z0-9\-]*:[0-9]{12}:[a-zA-Z0-9_.-]+

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
AsyncInferenceOutputConfig

Service: Amazon SageMaker Service

Specifies the configuration for asynchronous inference invocation outputs.

Contents

KmsKeyId

The AWS Key Management Service (AWS KMS) key that SageMaker uses to encrypt the asynchronous inference output in Amazon S3.

Type: String

Length Constraints: Maximum length of 2048.

Pattern: . *

Required: No

NotificationConfig

Specifies the configuration for notifications of inference results for asynchronous inference.

Type: AsyncInferenceNotificationConfig (p. 1259) object

Required: No

S3FailurePath

The Amazon S3 location to upload failure inference responses to.

Type: String

Length Constraints: Maximum length of 512.

Pattern: ^((https|s3)://([/])?([^/])?)$?

Required: No

S3OutputPath

The Amazon S3 location to upload inference responses to.

Type: String

Length Constraints: Maximum length of 512.

Pattern: ^((https|s3)://([/])?([^/])?)$?

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
• AWS SDK for Ruby V3
AthenaDatasetDefinition
Service: Amazon SageMaker Service
Configuration for Athena Dataset Definition input.

Contents

Catalog
The name of the data catalog used in Athena query execution.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 256.
Pattern: [\u0020-\uD7FF\uE000-\uFFE\uD800-\u\uDBFF\u\uFFF\t]*
Required: Yes

Database
The name of the database used in the Athena query execution.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 255.
Pattern: .*
Required: Yes

OutputFormat
The data storage format for Athena query results.
Type: String
Valid Values: PARQUET | ORC | AVRO | JSON | TEXTFILE
Required: Yes

OutputS3Uri
The location in Amazon S3 where Athena query results are stored.
Type: String
Length Constraints: Maximum length of 1024.
Pattern: ^(https|s3://[\w\./-]+)$
Required: Yes

QueryString
The SQL query statements, to be executed.
Type: String
Pattern: [\s\S]+
Required: Yes
**KmsKeyId**

The AWS Key Management Service (AWS KMS) key that Amazon SageMaker uses to encrypt data generated from an Athena query execution.

Type: String

Length Constraints: Maximum length of 2048.

Pattern: .*

Required: No

**OutputCompression**

The compression used for Athena query results.

Type: String

Valid Values: GZIP | SNAPPY | ZLIB

Required: No

**WorkGroup**

The name of the workgroup in which the Athena query is being started.

Type: String


Pattern: [a-zA-Z0-9-.]+

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
AutoMLAlgorithmConfig
Service: Amazon SageMaker Service

The collection of algorithms run on a dataset for training the model candidates of an Autopilot job.

Contents

AutoMLAlgorithms

The selection of algorithms run on a dataset to train the model candidates of an Autopilot job.

Note
Selected algorithms must belong to the list corresponding to the training mode set in `AutoMLJobConfig.Mode` (ENSEMBLING or HYPERPARAMETER_TUNING). Choose a minimum of 1 algorithm.

- In ENSEMBLING mode:
  - "catboost"
  - "extra-trees"
  - "fastai"
  - "lightgbm"
  - "linear-learner"
  - "nn-torch"
  - "randomforest"
  - "xgboost"
- In HYPERPARAMETER_TUNING mode:
  - "linear-learner"
  - "mlp"
  - "xgboost"

Type: Array of strings

Array Members: Maximum number of 11 items.

Valid Values: xgboost | linear-learner | mlp | lightgbm | catboost | randomforest | extra-trees | nn-torch | fastai

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
AutoMLCandidate

Service: Amazon SageMaker Service

Information about a candidate produced by an AutoML training job, including its status, steps, and other properties.

Contents

CandidateName

The name of the candidate.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 64.
Required: Yes

CandidateStatus

The candidate's status.
Type: String
Valid Values: Completed | InProgress | Failed | Stopped | Stopping
Required: Yes

CandidateSteps

Information about the candidate's steps.
Type: Array of AutoMLCandidateStep (p. 1270) objects
Required: Yes

CreationTime

The creation time.
Type: Timestamp
Required: Yes

LastModifiedTime

The last modified time.
Type: Timestamp
Required: Yes

ObjectiveStatus

The objective's status.
Type: String
Valid Values: Succeeded | Pending | Failed
Required: Yes

CandidateProperties

The properties of an AutoML candidate job.
Type: `CandidateProperties (p. 1315)` object

Required: No

**EndTime**

The end time.

Type: `Timestamp`

Required: No

**FailureReason**

The failure reason.

Type: `String`

Length Constraints: Maximum length of 1024.

Required: No

**FinalAutoMLJobObjectiveMetric**

The best candidate result from an AutoML training job.

Type: `FinalAutoMLJobObjectiveMetric (p. 1499)` object

Required: No

**InferenceContainerDefinitions**

The mapping of all supported processing unit (CPU, GPU, etc...) to inference container definitions for the candidate. This field is populated for the AutoML jobs V2 (for example, for jobs created by calling `CreateAutoMLJobV2`) related to image or text classification problem types only.

Type: String to array of `AutoMLContainerDefinition (p. 1273)` objects map

Map Entries: Maximum number of 2 items.

Valid Keys: `CPU` | `GPU`

Array Members: Maximum number of 5 items.

Required: No

**InferenceContainers**

Information about the recommended inference container definitions.

Type: Array of `AutoMLContainerDefinition (p. 1273)` objects

Array Members: Maximum number of 5 items.

Required: No

### See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
• AWS SDK for Ruby V3
AutoMLCandidateGenerationConfig
Service: Amazon SageMaker Service

Stores the configuration information for how a candidate is generated (optional).

Contents

AlgorithmsConfig
Stores the configuration information for the selection of algorithms used to train the model candidates.

The list of available algorithms to choose from depends on the training mode set in AutoMLJobConfig.Mode.

- AlgorithmsConfig should not be set in AUTO training mode.
- When AlgorithmsConfig is provided, one AutoMLAlgorithms attribute must be set and one only.

If the list of algorithms provided as values for AutoMLAlgorithms is empty, AutoMLCandidateGenerationConfig uses the full set of algorithms for the given training mode.
- When AlgorithmsConfig is not provided, AutoMLCandidateGenerationConfig uses the full set of algorithms for the given training mode.

For the list of all algorithms per training mode, see AutoMLAlgorithmConfig.

For more information on each algorithm, see the Algorithm support section in Autopilot developer guide.

Type: Array of AutoMLAlgorithmConfig (p. 1264) objects

Array Members: Maximum number of 1 item.

Required: No

FeatureSpecificationS3Uri
A URL to the Amazon S3 data source containing selected features from the input data source to run an Autopilot job. You can input FeatureAttributeNames (optional) in JSON format as shown below:

{ "FeatureAttributeNames":["col1", "col2", ...] }.

You can also specify the data type of the feature (optional) in the format shown below:

{ "FeatureDataTypes":{"col1":"numeric", "col2":"categorical" ... } }

Note
These column keys may not include the target column.

In ensembling mode, Autopilot only supports the following data types: numeric, categorical, text, and datetime. In HPO mode, Autopilot can support numeric, categorical, text, datetime, and sequence.

If only FeatureDataTypes is provided, the column keys (col1, col2,..) should be a subset of the column names in the input data.

If both FeatureDataTypes and FeatureAttributeNames are provided, then the column keys should be a subset of the column names provided in FeatureAttributeNames.
The key name `FeatureAttributeNames` is fixed. The values listed in `"col1", "col2", ...` are case sensitive and should be a list of strings containing unique values that are a subset of the column names in the input data. The list of columns provided must not include the target column.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: `^((https|s3):/\/(^[^/]*)/\?(.*))$`

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
AutoMLCandidateStep
Service: Amazon SageMaker Service

Information about the steps for a candidate and what step it is working on.

Contents

CandidateStepArn
The ARN for the candidate's step.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 256.
Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:.*
Required: Yes

CandidateStepName
The name for the candidate's step.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 64.
Required: Yes

CandidateStepType
Whether the candidate is at the transform, training, or processing step.
Type: String
Valid Values: AWS::SageMaker::TrainingJob | AWS::SageMaker::TransformJob | AWS::SageMaker::ProcessingJob
Required: Yes

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
AutoMLChannel
Service: Amazon SageMaker Service

A channel is a named input source that training algorithms can consume. The validation dataset size is limited to less than 2 GB. The training dataset size must be less than 100 GB. For more information, see Channel.

**Note**
A validation dataset must contain the same headers as the training dataset.

**Contents**

**DataSource**
The data source for an AutoML channel.

Type: AutoMLDataSource (p. 1275) object

Required: Yes

**TargetAttributeName**
The name of the target variable in supervised learning, usually represented by 'y'.

Type: String

Length Constraints: Minimum length of 1.

Required: Yes

**ChannelType**
The channel type (optional) is an enum string. The default value is `training`. Channels for training and validation must share the same `ContentType` and `TargetAttributeName`. For information on specifying training and validation channel types, see How to specify training and validation datasets.

Type: String

Valid Values: `training` | `validation`

Required: No

**CompressionType**
You can use Gzip or None. The default value is None.

Type: String

Valid Values: None  |  Gzip

Required: No

**ContentType**
The content type of the data from the input source. You can use text/csv;header=present or x-application/vnd.amazon+parquet. The default value is text/csv;header=present.

Type: String

Length Constraints: Maximum length of 256.
Pattern: .* 
Required: No  

**SampleWeightAttributeName**

If specified, this column name indicates which column of the dataset should be treated as sample weights for use by the objective metric during the training, evaluation, and the selection of the best model. This column is not considered as a predictive feature. For more information on Autopilot metrics, see Metrics and validation.

Sample weights should be numeric, non-negative, with larger values indicating which rows are more important than others. Data points that have invalid or no weight value are excluded.

Support for sample weights is available in Ensembling mode only.

Type: String 
Length Constraints: Minimum length of 1. Maximum length of 256.

Pattern: ^[a-zA-Z0-9_-]+$ 
Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
AutoMLContainerDefinition
Service: Amazon SageMaker Service

A list of container definitions that describe the different containers that make up an AutoML candidate. For more information, see ContainerDefinition.

Contents

Image

The Amazon Elastic Container Registry (Amazon ECR) path of the container. For more information, see ContainerDefinition.

Type: String
Length Constraints: Maximum length of 255.
Pattern: \S+ 
Required: Yes

ModelDataUrl

The location of the model artifacts. For more information, see ContainerDefinition.

Type: String
Length Constraints: Maximum length of 1024.
Pattern: ^(https|s3)://(\[^/\]+)/(.*)$
Required: Yes

Environment

The environment variables to set in the container. For more information, see ContainerDefinition.

Type: String to string map
Map Entries: Maximum number of 16 items.
Key Length Constraints: Maximum length of 1024.
Key Pattern: [a-zA-Z_][a-zA-Z0-9_]*
Value Length Constraints: Maximum length of 1024.
Value Pattern: \S+ 
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
AutoMLDataSource
Service: Amazon SageMaker Service
The data source for the Autopilot job.

Contents

S3DataSource
The Amazon S3 location of the input data.
Type: AutoMLS3DataSource (p. 1295) object
Required: Yes

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
AutoMLDataSplitConfig

Service: Amazon SageMaker Service

This structure specifies how to split the data into train and validation datasets.

The validation and training datasets must contain the same headers. For jobs created by calling `CreateAutoMLJob`, the validation dataset must be less than 2 GB in size.

Contents

**ValidationFraction**

The validation fraction (optional) is a float that specifies the portion of the training dataset to be used for validation. The default value is 0.2, and values must be greater than 0 and less than 1. We recommend setting this value to be less than 0.5.

Type: Float

Valid Range: Minimum value of 0. Maximum value of 1.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
AutoMLJobArtifacts
Service: Amazon SageMaker Service
The artifacts that are generated during an AutoML job.

Contents

CandidateDefinitionNotebookLocation
The URL of the notebook location.
Type: String
Length Constraints: Minimum length of 1.
Required: No

DataExplorationNotebookLocation
The URL of the notebook location.
Type: String
Length Constraints: Minimum length of 1.
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
AutoMLJobChannel

Service: Amazon SageMaker Service

A channel is a named input source that training algorithms can consume. This channel is used for AutoML jobs V2 (jobs created by calling `CreateAutoMLJobV2`).

Contents

**ChannelType**

The type of channel. Defines whether the data are used for training or validation. The default value is `training`. Channels for training and validation must share the same `ContentType`.

**Note**

The type of channel defaults to `training` for the time-series forecasting problem type.

Type: String

Valid Values: `training` | `validation`

Required: No

**CompressionType**

The allowed compression types depend on the input format and problem type. We allow the compression type Gzip for `S3Prefix` inputs on tabular data only. For all other inputs, the compression type should be `None`. If no compression type is provided, we default to `None`.

Type: String

Valid Values: `None` | `Gzip`

Required: No

**ContentType**

The content type of the data from the input source. The following are the allowed content types for different problems:

- For tabular problem types: `text/csv;header=present` or `x-application/vnd.amazon+parquet`. The default value is `text/csv;header=present`.
- For image classification: `image/png`, `image/jpeg`, or `image/*`. The default value is `image/*`.
- For text classification: `text/csv;header=present` or `x-application/vnd.amazon+parquet`. The default value is `text/csv;header=present`.
- For time-series forecasting: `text/csv;header=present` or `x-application/vnd.amazon+parquet`. The default value is `text/csv;header=present`.
- For text generation (LLMs fine-tuning): `text/csv;header=present` or `x-application/vnd.amazon+parquet`. The default value is `text/csv;header=present`.

Type: String

Length Constraints: Maximum length of 256.

Pattern: `.=`

Required: No

**DataSource**

The data source for an AutoML channel (Required).
Type: AutoMLDataSource (p. 1275) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
AutoMLJobCompletionCriteria
Service: Amazon SageMaker Service

How long a job is allowed to run, or how many candidates a job is allowed to generate.

Contents

MaxAutoMLJobRuntimeInSeconds

The maximum runtime, in seconds, an AutoML job has to complete.

If an AutoML job exceeds the maximum runtime, the job is stopped automatically and its processing is ended gracefully. The AutoML job identifies the best model whose training was completed and marks it as the best-performing model. Any unfinished steps of the job, such as automatic one-click Autopilot model deployment, are not completed.

Type: Integer

Valid Range: Minimum value of 1.

Required: No

MaxCandidates

The maximum number of times a training job is allowed to run.

For text and image classification, time-series forecasting, as well as text generation (LLMs fine-tuning) problem types, the supported value is 1. For tabular problem types, the maximum value is 750.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 750.

Required: No

MaxRuntimePerTrainingJobInSeconds

The maximum time, in seconds, that each training job executed inside hyperparameter tuning is allowed to run as part of a hyperparameter tuning job. For more information, see the StoppingCondition used by the CreateHyperParameterTuningJob action.

For job V2s (jobs created by calling CreateAutoMLJobV2), this field controls the runtime of the job candidate.

For TextGenerationJobConfig problem types, the maximum time defaults to 72 hours (259200 seconds).

Type: Integer

Valid Range: Minimum value of 1.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
AutoMLJobConfig

Service: Amazon SageMaker Service

A collection of settings used for an AutoML job.

Contents

CandidateGenerationConfig

The configuration for generating a candidate for an AutoML job (optional).

Type: AutoMLCandidateGenerationConfig (p. 1268) object

Required: No

CompletionCriteria

How long an AutoML job is allowed to run, or how many candidates a job is allowed to generate.

Type: AutoMLJobCompletionCriteria (p. 1280) object

Required: No

DataSplitConfig

The configuration for splitting the input training dataset.

Type: AutoMLDataSplitConfig

Type: AutoMLDataSplitConfig (p. 1276) object

Required: No

Mode

The method that Autopilot uses to train the data. You can either specify the mode manually or let Autopilot choose for you based on the dataset size by selecting AUTO. In AUTO mode, Autopilot chooses ENSEMBLING for datasets smaller than 100 MB, and HYPERPARAMETER_TUNING for larger ones.

The ENSEMBLING mode uses a multi-stack ensemble model to predict classification and regression tasks directly from your dataset. This machine learning mode combines several base models to produce an optimal predictive model. It then uses a stacking ensemble method to combine predictions from contributing members. A multi-stack ensemble model can provide better performance over a single model by combining the predictive capabilities of multiple models. See Autopilot algorithm support for a list of algorithms supported by ENSEMBLING mode.

The HYPERPARAMETER_TUNING (HPO) mode uses the best hyperparameters to train the best version of a model. HPO automatically selects an algorithm for the type of problem you want to solve. Then HPO finds the best hyperparameters according to your objective metric. See Autopilot algorithm support for a list of algorithms supported by HYPERPARAMETER_TUNING mode.

Type: String

Valid Values: AUTO | ENSEMBLING | HYPERPARAMETER_TUNING

Required: No

SecurityConfig

The security configuration for traffic encryption or Amazon VPC settings.

Type: AutoMLSecurityConfig (p. 1297) object
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
AutoMLJobObjective

Service: Amazon SageMaker Service

Specifies a metric to minimize or maximize as the objective of an AutoML job.

Contents

MetricName

The name of the objective metric used to measure the predictive quality of a machine learning system. During training, the model's parameters are updated iteratively to optimize its performance based on the feedback provided by the objective metric when evaluating the model on the validation dataset.

The list of available metrics supported by Autopilot and the default metric applied when you do not specify a metric name explicitly depend on the problem type:

- For tabular problem types:
  - List of available metrics:
    - Regression: InferenceLatency, MAE, MSE, R2, RMSE
    - Binary classification: Accuracy, AUC, BalancedAccuracy, F1, InferenceLatency, LogLoss, Precision, Recall
    - Multiclass classification: Accuracy, BalancedAccuracy, F1macro, InferenceLatency, LogLoss, PrecisionMacro, RecallMacro
  
  For a description of each metric, see Autopilot metrics for classification and regression.

  - Default objective metrics:
    - Regression: MSE.
    - Binary classification: F1.
    - Multiclass classification: Accuracy.

- For image or text classification problem types:
  - List of available metrics: Accuracy
  
  For a description of each metric, see Autopilot metrics for text and image classification.

- For time-series forecasting problem types:
  - List of available metrics: RMSE, wQL, Average wQL, MASE, MAPE, WAPE
  
  For a description of each metric, see Autopilot metrics for time-series forecasting.

  - Default objective metrics: AverageWeightedQuantileLoss

- For text generation problem types (LLMs fine-tuning): Fine-tuning language models in Autopilot does not require setting the AutoMLJobObjective field. Autopilot fine-tunes LLMs without requiring multiple candidates to be trained and evaluated. Instead, using your dataset, Autopilot directly fine-tunes your target model to enhance a default objective metric, the cross-entropy loss. After fine-tuning a language model, you can evaluate the quality of its generated text using different metrics. For a list of the available metrics, see Metrics for fine-tuning LLMs in Autopilot.

Type: String

Valid Values: Accuracy | MSE | F1 | F1macro | AUC | RMSE | MAE | R2 | BalancedAccuracy | Precision | PrecisionMacro | Recall | RecallMacro | MAPE | MASE | WAPE | AverageWeightedQuantileLoss

Required: Yes
See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
AutoMLJobStepMetadata
Service: Amazon SageMaker Service
Metadata for an AutoML job step.

Contents

Arn
The Amazon Resource Name (ARN) of the AutoML job.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 256.
Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:automl-job/.*
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
AutoMLJobSummary

Service: Amazon SageMaker Service

Provides a summary about an AutoML job.

Contents

AutoMLJobArn

The ARN of the AutoML job.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:automl-job/.*

Required: Yes

AutoMLJobName

The name of the AutoML job you are requesting.

Type: String


Pattern: ^[a-zA-Z0-9\-\_*\[a-zA-Z0-9\-\]*\[a-zA-Z0-9\-\]*\[0,31]}

Required: Yes

AutoMLJobSecondaryStatus

The secondary status of the AutoML job.

Type: String


Required: Yes

AutoMLJobStatus

The status of the AutoML job.

Type: String

Valid Values: Completed | InProgress | Failed | Stopped | Stopping

Required: Yes

CreationTime

When the AutoML job was created.

Type: Timestamp

Required: Yes
LastModifiedTime
When the AutoML job was last modified.
Type: Timestamp
Required: Yes

EndTime
The end time of an AutoML job.
Type: Timestamp
Required: No

FailureReason
The failure reason of an AutoML job.
Type: String
Length Constraints: Maximum length of 1024.
Required: No

PartialFailureReasons
The list of reasons for partial failures within an AutoML job.
Type: Array of AutoMLPartialFailureReason (p. 1290) objects
Array Members: Minimum number of 1 item. Maximum number of 5 items.
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
AutoMLOutputDataConfig

Service: Amazon SageMaker Service

The output data configuration.

Contents

S3OutputPath

The Amazon S3 output path. Must be 128 characters or less.

Type: String
Length Constraints: Maximum length of 1024.
Pattern: ^(https|s3)://([^/]+)(([^/]+)?/.*)$
Required: Yes

KmsKeyId

The Key Management Service (KMS) encryption key ID.

Type: String
Length Constraints: Maximum length of 2048.
Pattern: .*
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
AutoMLPartialFailureReason
Service: Amazon SageMaker Service

The reason for a partial failure of an AutoML job.

Contents

PartialFailureMessage

The message containing the reason for a partial failure of an AutoML job.

Type: String

Length Constraints: Maximum length of 1024.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
AutoMLProblemTypeConfig

Service: Amazon SageMaker Service

A collection of settings specific to the problem type used to configure an AutoML job V2. There must be one and only one config of the following type.

Contents

Important
This data type is a UNION, so only one of the following members can be specified when used or returned.

ImageClassificationJobConfig
Settings used to configure an AutoML job V2 for the image classification problem type.
Type: ImageClassificationJobConfig (p. 1573) object
Required: No

TabularJobConfig
Settings used to configure an AutoML job V2 for the tabular problem type (regression, classification).
Type: TabularJobConfig (p. 1975) object
Required: No

TextClassificationJobConfig
Settings used to configure an AutoML job V2 for the text classification problem type.
Type: TextClassificationJobConfig (p. 1985) object
Required: No

TextGenerationJobConfig
Settings used to configure an AutoML job V2 for the text generation (LLMs fine-tuning) problem type.
Note
The text generation models that support fine-tuning in Autopilot are currently accessible exclusively in regions supported by Canvas. Refer to the documentation of Canvas for the full list of its supported Regions.
Type: TextGenerationJobConfig (p. 1986) object
Required: No

TimeSeriesForecastingJobConfig
Settings used to configure an AutoML job V2 for the time-series forecasting problem type.
Type: TimeSeriesForecastingJobConfig (p. 1991) object
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:
• AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for Ruby V3
AutoMLProblemTypeResolvedAttributes

Service: Amazon SageMaker Service

Stores resolved attributes specific to the problem type of an AutoML job V2.

Contents

**Important**

This data type is a UNION, so only one of the following members can be specified when used or returned.

**TabularResolvedAttributes**

The resolved attributes for the tabular problem type.

Type: [TabularResolvedAttributes](p. 1978) object

Required: No

**TextGenerationResolvedAttributes**

The resolved attributes for the text generation problem type.

Type: [TextGenerationResolvedAttributes](p. 1988) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](aws-sdk-c++)
- [AWS SDK for Go](aws-sdk-go)
- [AWS SDK for Java V2](aws-sdk-java)
- [AWS SDK for Ruby V3](aws-sdk-ruby)

1293
AutoMLResolvedAttributes

Service: Amazon SageMaker Service

The resolved attributes used to configure an AutoML job V2.

Contents

AutoMLJobObjective

Specifies a metric to minimize or maximize as the objective of an AutoML job.

Type: AutoMLJobObjective (p. 1284) object

Required: No

AutoMLProblemTypeResolvedAttributes

Defines the resolved attributes specific to a problem type.

Type: AutoMLProblemTypeResolvedAttributes (p. 1293) object

Note: This object is a Union. Only one member of this object can be specified or returned.

Required: No

CompletionCriteria

How long a job is allowed to run, or how many candidates a job is allowed to generate.

Type: AutoMLJobCompletionCriteria (p. 1280) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
AutoMLS3DataSource

Service: Amazon SageMaker Service

Describes the Amazon S3 data source.

Contents

S3DataType

The data type.

- If you choose S3Prefix, S3Uri identifies a key name prefix. SageMaker uses all objects that match the specified key name prefix for model training.

  The S3Prefix should have the following format:

  s3://DOC-EXAMPLE-BUCKET/DOC-EXAMPLE-FOLDER-OR-FOLDER

- If you choose ManifestFile, S3Uri identifies an object that is a manifest file containing a list of object keys that you want SageMaker to use for model training.

  A ManifestFile should have the format shown below:

  [ {"prefix": "s3://DOC-EXAMPLE-BUCKET/DOC-EXAMPLE-FOLDER/DOC-EXAMPLE-PREFIX/"},
    "DOC-EXAMPLE-RELATIVE-PATH/DOC-EXAMPLE-FOLDER/MODEL-EXAMPLE-RELATIVE-PATH/DOC-EXAMPLE-FOLDER/DATA-1",
    "DOC-EXAMPLE-RELATIVE-PATH/DOC-EXAMPLE-FOLDER/DATA-2",
    ...
    "DOC-EXAMPLE-RELATIVE-PATH/DOC-EXAMPLE-FOLDER/MODEL-EXAMPLE-RELATIVE-PATH/DOC-EXAMPLE-FOLDER/DATA-N" }
  ]

- If you choose AugmentedManifestFile, S3Uri identifies an object that is an augmented manifest file in JSON lines format. This file contains the data you want to use for model training. AugmentedManifestFile is available for V2 API jobs only (for example, for jobs created by calling CreateAutoMLJobV2).

  Here is a minimal, single-record example of an AugmentedManifestFile:

  {"source-ref": "s3://DOC-EXAMPLE-BUCKET/DOC-EXAMPLE-FOLDER/DOC-EXAMPLE-PREFIX/cats/cat.jpg",
   "label-metadata": {"class-name": "cat"}
  }

  For more information on AugmentedManifestFile, see Provide Dataset Metadata to Training Jobs with an Augmented Manifest File.

Type: String

Valid Values: ManifestFile | S3Prefix | AugmentedManifestFile

Required: Yes

S3Uri

The URL to the Amazon S3 data source. The Uri refers to the Amazon S3 prefix or ManifestFile depending on the data type.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: ^(https|s3)://([^/]+)/?([^/]*$
Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
AutoMLSecurityConfig
Service: Amazon SageMaker Service
Security options.

Contents

EnableInterContainerTrafficEncryption
Whether to use traffic encryption between the container layers.
Type: Boolean
Required: No

VolumeKmsKeyId
The key used to encrypt stored data.
Type: String
Length Constraints: Maximum length of 2048.
Pattern: . *
Required: No

VpcConfig
The VPC configuration.
Type: VpcConfig (p. 2076) object
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
AutoParameter
Service: Amazon SageMaker Service

The name and an example value of the hyperparameter that you want to use in Autotune. If Automatic model tuning (AMT) determines that your hyperparameter is eligible for Autotune, an optimal hyperparameter range is selected for you.

Contents

Name
The name of the hyperparameter to optimize using Autotune.
Type: String
Length Constraints: Maximum length of 256.
Pattern: .*
Required: Yes

ValueHint
An example value of the hyperparameter to optimize using Autotune.
Type: String
Length Constraints: Maximum length of 256.
Pattern: .*
Required: Yes

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
AutoRollbackConfig
Service: Amazon SageMaker Service

Automatic rollback configuration for handling endpoint deployment failures and recovery.

Contents

Alarms

List of CloudWatch alarms in your account that are configured to monitor metrics on an endpoint. If any alarms are tripped during a deployment, SageMaker rolls back the deployment.

Type: Array of Alarm (p. 1224) objects

Array Members: Minimum number of 1 item. Maximum number of 10 items.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
Autotune
Service: Amazon SageMaker Service

A flag to indicate if you want to use Autotune to automatically find optimal values for the following fields:

- **ParameterRanges**: The names and ranges of parameters that a hyperparameter tuning job can optimize.
- **ResourceLimits**: The maximum resources that can be used for a training job. These resources include the maximum number of training jobs, the maximum runtime of a tuning job, and the maximum number of training jobs to run at the same time.
- **TrainingJobEarlyStoppingType**: A flag that specifies whether or not to use early stopping for training jobs launched by a hyperparameter tuning job.
- **RetryStrategy**: The number of times to retry a training job.
- **Strategy**: Specifies how hyperparameter tuning chooses the combinations of hyperparameter values to use for the training jobs that it launches.
- **ConvergenceDetected**: A flag to indicate that Automatic model tuning (AMT) has detected model convergence.

Contents

**Mode**

Set Mode to Enabled if you want to use Autotune.

Type: String

Valid Values: Enabled

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
BatchDataCaptureConfig

Service: Amazon SageMaker Service

Configuration to control how SageMaker captures inference data for batch transform jobs.

Contents

DestinationS3Uri

The Amazon S3 location being used to capture the data.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: ^(https|s3)://(\^[^/]+)/(.*$)

Required: Yes

GenerateInferenceId

Flag that indicates whether to append inference id to the output.

Type: Boolean

Required: No

KmsKeyId

The Amazon Resource Name (ARN) of a AWS Key Management Service key that SageMaker uses to encrypt data on the storage volume attached to the ML compute instance that hosts the batch transform job.

The KmsKeyId can be any of the following formats:
- Key ID: 1234abcd-12ab-34cd-56ef-1234567890ab
- Key ARN: arn:aws:kms:us-west-2:111122223333:key/1234abcd-12ab-34cd-56ef-1234567890ab
- Alias name: alias/ExampleAlias

Type: String

Length Constraints: Maximum length of 2048.

Pattern: .*

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
BatchDescribeModelPackageError

Service: Amazon SageMaker Service

The error code and error description associated with the resource.

Contents

**ErrorCode**

Type: String

Required: Yes

**ErrorResponse**

Type: String

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
BatchDescribeModelPackageSummary

Service: Amazon SageMaker Service

Provides summary information about the model package.

Contents

**CreationTime**

The creation time of the mortgage package summary.

Type: Timestamp

Required: Yes

**InferenceSpecification**

Defines how to perform inference generation after a training job is run.

Type: `InferenceSpecification (p. 1601)` object

Required: Yes

**ModelPackageArn**

The Amazon Resource Name (ARN) of the model package.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(-cn|-us-gov)?:sagemaker:[a-z0-9\-]{9,16}:[0-9]{12}:model-package/[^\$]+\{1,2048}\$`

Required: Yes

**ModelPackageGroupName**

The group name for the model package.

Type: String


Pattern: `^[a-zA-Z0-9\-]*[a-zA-Z0-9\-]\{0,62\}$`

Required: Yes

**ModelPackageStatus**

The status of the mortgage package.

Type: String

Valid Values: Pending | InProgress | Completed | Failed | Deleting

Required: Yes

**ModelApprovalStatus**

The approval status of the model.

Type: String
Valid Values: Approved | Rejected | PendingManualApproval

Required: No

ModelPackageDescription

The description of the model package.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: [\p{L}\p{M}\p{Z}\p{S}\p{N}\p{P}]*

Required: No

ModelPackageVersion

The version number of a versioned model.

Type: Integer

Valid Range: Minimum value of 1.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
BatchTransformInput
Service: Amazon SageMaker Service

Input object for the batch transform job.

Contents

DataCapturedDestinationS3Uri

The Amazon S3 location being used to capture the data.

Type: String

Length Constraints: Maximum length of 512.

Pattern: ^https:\/\/(^/)?[^/]?(.*)$

Required: Yes

DatasetFormat

The dataset format for your batch transform job.

Type: MonitoringDatasetFormat (p. 1742) object

Required: Yes

LocalPath

Path to the filesystem where the batch transform data is available to the container.

Type: String

Length Constraints: Maximum length of 256.

Pattern: .*

Required: Yes

EndTimeOffset

If specified, monitoring jobs subtract this time from the end time. For information about using offsets for scheduling monitoring jobs, see Schedule Model Quality Monitoring Jobs.

Type: String


Pattern: ^\.?P.*$

Required: No

ExcludeFeaturesAttribute

The attributes of the input data to exclude from the analysis.

Type: String

Length Constraints: Maximum length of 100.

Required: No
FeaturesAttribute

The attributes of the input data that are the input features.

Type: String

Required: No

InferenceAttribute

The attribute of the input data that represents the ground truth label.

Type: String

Required: No

ProbabilityAttribute

In a classification problem, the attribute that represents the class probability.

Type: String

Required: No

ProbabilityThresholdAttribute

The threshold for the class probability to be evaluated as a positive result.

Type: Double

Required: No

S3DataDistributionType

Whether input data distributed in Amazon S3 is fully replicated or sharded by an S3 key. Defaults to FullyReplicated.

Type: String

Valid Values: FullyReplicated | ShardedByS3Key

Required: No

S3InputMode

Whether the Pipe or File is used as the input mode for transferring data for the monitoring job. Pipe mode is recommended for large datasets. File mode is useful for small files that fit in memory. Defaults to File.

Type: String

Valid Values: Pipe | File

Required: No

StartTimeOffset

If specified, monitoring jobs subtract this time from the start time. For information about using offsets for scheduling monitoring jobs, see Schedule Model Quality Monitoring Jobs.

Type: String


Pattern: ^.*P.*
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
BestObjectiveNotImproving
Service: Amazon SageMaker Service

A structure that keeps track of which training jobs launched by your hyperparameter tuning job are not improving model performance as evaluated against an objective function.

Contents

MaxNumberOfTrainingJobsNotImproving

The number of training jobs that have failed to improve model performance by 1% or greater over prior training jobs as evaluated against an objective function.

Type: Integer

Valid Range: Minimum value of 3.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
Bias
Service: Amazon SageMaker Service
Contains bias metrics for a model.

Contents

PostTrainingReport
The post-training bias report for a model.
Type: MetricsSource (p. 1655) object
Required: No

PreTrainingReport
The pre-training bias report for a model.
Type: MetricsSource (p. 1655) object
Required: No

Report
The bias report for a model
Type: MetricsSource (p. 1655) object
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
BlueGreenUpdatePolicy
Service: Amazon SageMaker Service

Update policy for a blue/green deployment. If this update policy is specified, SageMaker creates a new fleet during the deployment while maintaining the old fleet. SageMaker flips traffic to the new fleet according to the specified traffic routing configuration. Only one update policy should be used in the deployment configuration. If no update policy is specified, SageMaker uses a blue/green deployment strategy with all at once traffic shifting by default.

Contents

TrafficRoutingConfiguration

Defines the traffic routing strategy to shift traffic from the old fleet to the new fleet during an endpoint deployment.

Type: TrafficRoutingConfig (p. 1998) object

Required: Yes

MaximumExecutionTimeoutInSeconds

Maximum execution timeout for the deployment. Note that the timeout value should be larger than the total waiting time specified in TerminationWaitInSeconds and WaitIntervalInSeconds.

Type: Integer


Required: No

TerminationWaitInSeconds

Additional waiting time in seconds after the completion of an endpoint deployment before terminating the old endpoint fleet. Default is 0.

Type: Integer

Valid Range: Minimum value of 0. Maximum value of 3600.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
CacheHitResult
Service: Amazon SageMaker Service
Details on the cache hit of a pipeline execution step.

Contents

SourcePipelineExecutionArn

The Amazon Resource Name (ARN) of the pipeline execution.

Type: String

Length Constraints: Maximum length of 256.

Pattern: ^arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:pipeline\/[a-z0-9\-]*$ 

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
CallbackStepMetadata
Service: Amazon SageMaker Service
Metadata about a callback step.

Contents

**CallbackToken**

The pipeline generated token from the Amazon SQS queue.
Type: String
Length Constraints: Fixed length of 10.
Pattern: `^[a-zA-Z0-9]+$`
Required: No

**OutputParameters**

A list of the output parameters of the callback step.
Type: Array of `OutputParameter` objects
Array Members: Minimum number of 0 items. Maximum number of 50 items.
Required: No

**SqsQueueUrl**

The URL of the Amazon Simple Queue Service (Amazon SQS) queue used by the callback step.
Type: String
Length Constraints: Maximum length of 256.
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
CandidateArtifactLocations

Service: Amazon SageMaker Service

The location of artifacts for an AutoML candidate job.

Contents

**Explainability**

The Amazon S3 prefix to the explainability artifacts generated for the AutoML candidate.

Type: String

Length Constraints: Minimum length of 1.

Required: Yes

**BacktestResults**

The Amazon S3 prefix to the accuracy metrics and the inference results observed over the testing window. Available only for the time-series forecasting problem type.

Type: String

Length Constraints: Minimum length of 1.

Required: No

**ModelInsights**

The Amazon S3 prefix to the model insight artifacts generated for the AutoML candidate.

Type: String

Length Constraints: Minimum length of 1.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
CandidateGenerationConfig

Service: Amazon SageMaker Service

Stores the configuration information for how model candidates are generated using an AutoML job V2.

Contents

AlgorithmsConfig

Stores the configuration information for the selection of algorithms used to train model candidates on tabular data.

The list of available algorithms to choose from depends on the training mode set in TabularJobConfig.Mode.

- AlgorithmsConfig should not be set in AUTO training mode.
- When AlgorithmsConfig is provided, one AutoMLAlgorithms attribute must be set and one only.

If the list of algorithms provided as values for AutoMLAlgorithms is empty, CandidateGenerationConfig uses the full set of algorithms for the given training mode.
- When AlgorithmsConfig is not provided, CandidateGenerationConfig uses the full set of algorithms for the given training mode.

For the list of all algorithms per problem type and training mode, see AutoMLAlgorithmConfig.

For more information on each algorithm, see the Algorithm support section in Autopilot developer guide.

Type: Array of AutoMLAlgorithmConfig (p. 1264) objects

Array Members: Maximum number of 1 item.
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
CandidateProperties

Service: Amazon SageMaker Service

The properties of an AutoML candidate job.

Contents

CandidateArtifactLocations

The Amazon S3 prefix to the artifacts generated for an AutoML candidate.

Type: CandidateArtifactLocations (p. 1313) object

Required: No

CandidateMetrics

Information about the candidate metrics for an AutoML job.

Type: Array of MetricDatum (p. 1651) objects

Array Members: Minimum number of 0 items. Maximum number of 40 items.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
CanvasAppSettings

Service: Amazon SageMaker Service

The SageMaker Canvas application settings.

Contents

DirectDeploySettings

The model deployment settings for the SageMaker Canvas application.

Type: DirectDeploySettings (p. 1421) object

Required: No

IdentityProviderOAuthSettings

The settings for connecting to an external data source with OAuth.

Type: Array of IdentityProviderOAuthSetting (p. 1570) objects

Array Members: Maximum number of 20 items.

Required: No

KendraSettings

The settings for document querying.

Type: KendraSettings (p. 1619) object

Required: No

ModelRegisterSettings

The model registry settings for the SageMaker Canvas application.

Type: ModelRegisterSettings (p. 1725) object

Required: No

TimeSeriesForecastingSettings

Time series forecast settings for the SageMaker Canvas application.

Type: TimeSeriesForecastingSettings (p. 1994) object

Required: No

WorkspaceSettings

The workspace settings for the SageMaker Canvas application.

Type: WorkspaceSettings (p. 2084) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for Ruby V3
CapacitySize

Service: Amazon SageMaker Service

Specifies the type and size of the endpoint capacity to activate for a blue/green deployment, a rolling deployment, or a rollback strategy. You can specify your batches as either instance count or the overall percentage or your fleet.

For a rollback strategy, if you don't specify the fields in this object, or if you set the Value to 100%, then SageMaker uses a blue/green rollback strategy and rolls all traffic back to the blue fleet.

Contents

Type

Specifies the endpoint capacity type.

- INSTANCE_COUNT: The endpoint activates based on the number of instances.
- CAPACITY_PERCENT: The endpoint activates based on the specified percentage of capacity.

Type: String

Valid Values: INSTANCE_COUNT | CAPACITY_PERCENT

Required: Yes

Value

Defines the capacity size, either as a number of instances or a capacity percentage.

Type: Integer

Valid Range: Minimum value of 1.

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
CaptureContentTypeHeader

Configuration specifying how to treat different headers. If no headers are specified Amazon SageMaker will by default base64 encode when capturing the data.

Contents

CsvContentTypes

The list of all content type headers that Amazon SageMaker will treat as CSV and capture accordingly.

Type: Array of strings

Array Members: Minimum number of 1 item. Maximum number of 10 items.

Length Constraints: Minimum length of 1. Maximum length of 256.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])*\/[a-zA-Z0-9](-*[a-zA-Z0-9.])* 

Required: No

JsonContentTypes

The list of all content type headers that SageMaker will treat as JSON and capture accordingly.

Type: Array of strings

Array Members: Minimum number of 1 item. Maximum number of 10 items.

Length Constraints: Minimum length of 1. Maximum length of 256.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])*\/[a-zA-Z0-9](-*[a-zA-Z0-9.])* 

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
CaptureOption
Service: Amazon SageMaker Service
Specifies data Model Monitor will capture.

Contents

CaptureMode

Specify the boundary of data to capture.

Type: String

Valid Values: Input | Output

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
CategoricalParameter
Service: Amazon SageMaker Service
Environment parameters you want to benchmark your load test against.

Contents

Name
The Name of the environment variable.
Type: String
Length Constraints: Maximum length of 64.
Required: Yes

Value
The list of values you can pass.
Type: Array of strings
Array Members: Minimum number of 1 item. Maximum number of 3 items.
Length Constraints: Maximum length of 128.
Required: Yes

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
CategoricalParameterRange
Service: Amazon SageMaker Service
A list of categorical hyperparameters to tune.

Contents

Name
The name of the categorical hyperparameter to tune.
Type: String
Length Constraints: Maximum length of 256.
Pattern: . *
Required: Yes

Values
A list of the categories for the hyperparameter.
Type: Array of strings
Array Members: Minimum number of 1 item. Maximum number of 30 items.
Length Constraints: Maximum length of 256.
Pattern: . *
Required: Yes

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
CategoricalParameterRangeSpecification
Service: Amazon SageMaker Service
Defines the possible values for a categorical hyperparameter.

Contents

Values

The allowed categories for the hyperparameter.
Type: Array of strings
Array Members: Minimum number of 1 item. Maximum number of 30 items.
Length Constraints: Maximum length of 256.
Pattern: .*
Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
Channel
Service: Amazon SageMaker Service

A channel is a named input source that training algorithms can consume.

Contents

ChannelName

The name of the channel.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 64.
Pattern: [A-Za-z0-9\-\._]+
Required: Yes

DataSource

The location of the channel data.
Type: DataSource (p. 1394) object
Required: Yes

CompressionType

If training data is compressed, the compression type. The default value is None. CompressionType is used only in Pipe input mode. In File mode, leave this field unset or set it to None.
Type: String
Valid Values: None | Gzip
Required: No

ContentType

The MIME type of the data.
Type: String
Length Constraints: Maximum length of 256.
Pattern: .*
Required: No

InputMode

(Optional) The input mode to use for the data channel in a training job. If you don't set a value for InputMode, SageMaker uses the value set for TrainingInputMode. Use this parameter to override the TrainingInputMode setting in a AlgorithmSpecification request when you have a channel that needs a different input mode from the training job's general setting. To download the data from Amazon Simple Storage Service (Amazon S3) to the provisioned ML storage volume, and mount the directory to a Docker volume, use File input mode. To stream data directly from Amazon S3 to the container, choose Pipe input mode.
To use a model for incremental training, choose File input model.
Type: String
Valid Values: Pipe | File | FastFile

Required: No

RecordWrapperType

Specify RecordIO as the value when input data is in raw format but the training algorithm requires the RecordIO format. In this case, SageMaker wraps each individual S3 object in a RecordIO record. If the input data is already in RecordIO format, you don't need to set this attribute. For more information, see Create a Dataset Using RecordIO.

In File mode, leave this field unset or set it to None.

Type: String

Valid Values: None | RecordIO

Required: No

ShuffleConfig

A configuration for a shuffle option for input data in a channel. If you use S3Prefix for S3DataType, this shuffles the results of the S3 key prefix matches. If you use ManifestFile, the order of the S3 object references in the ManifestFile is shuffled. If you use AugmentedManifestFile, the order of the JSON lines in the AugmentedManifestFile is shuffled. The shuffling order is determined using the Seed value.

For Pipe input mode, shuffling is done at the start of every epoch. With large datasets this ensures that the order of the training data is different for each epoch, it helps reduce bias and possible overfitting. In a multi-node training job when ShuffleConfig is combined with S3DataDistributionType of ShardedByS3Key, the data is shuffled across nodes so that the content sent to a particular node on the first epoch might be sent to a different node on the second epoch.

Type: ShuffleConfig (p. 1953) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ChannelSpecification
Service: Amazon SageMaker Service

Defines a named input source, called a channel, to be used by an algorithm.

Contents

Name
The name of the channel.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 64.
Pattern: [A-Za-z0-9\.\-_]+
Required: Yes

SupportedContentTypes
The supported MIME types for the data.
Type: Array of strings
Length Constraints: Maximum length of 256.
Pattern: .*
Required: Yes

SupportedInputModes
The allowed input mode, either FILE or PIPE.

In FILE mode, Amazon SageMaker copies the data from the input source onto the local Amazon Elastic Block Store (Amazon EBS) volumes before starting your training algorithm. This is the most commonly used input mode.

In PIPE mode, Amazon SageMaker streams input data from the source directly to your algorithm without using the EBS volume.

Type: Array of strings
Array Members: Minimum number of 1 item.
Valid Values: Pipe | File | FastFile
Required: Yes

Description
A brief description of the channel.
Type: String
Length Constraints: Maximum length of 1024.
Pattern: [\p{L}\p{M}\p{Z}\p{S}\p{N}\p{P}]*
Required: No
IsRequired

Indicates whether the channel is required by the algorithm.

Type: Boolean

Required: No

SupportedCompressionTypes

The allowed compression types, if data compression is used.

Type: Array of strings

Valid Values: None | Gzip

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
**CheckpointConfig**

Service: Amazon SageMaker Service

Contains information about the output location for managed spot training checkpoint data.

**Contents**

**S3Uri**

Identifies the S3 path where you want SageMaker to store checkpoints. For example, s3://bucket-name/key-name-prefix.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: ^((https|s3):/(/[^/]+/)?(.*)$

Required: Yes

**LocalPath**

(Optional) The local directory where checkpoints are written. The default directory is /opt/ml/checkpoints/.

Type: String

Length Constraints: Maximum length of 4096.

Pattern: .*

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
ClarifyCheckStepMetadata
Service: Amazon SageMaker Service

The container for the metadata for the ClarifyCheck step. For more information, see the topic on ClarifyCheck step in the Amazon SageMaker Developer Guide.

Contents

BaselineUsedForDriftCheckConstraints

The Amazon S3 URI of baseline constraints file to be used for the drift check.

Type: String

Length Constraints: Maximum length of 1024.

Required: No

CalculatedBaselineConstraints

The Amazon S3 URI of the newly calculated baseline constraints file.

Type: String

Length Constraints: Maximum length of 1024.

Required: No

CheckJobArn

The Amazon Resource Name (ARN) of the check processing job that was run by this step's execution.

Type: String

Length Constraints: Maximum length of 256.

Required: No

CheckType

The type of the Clarify Check step

Type: String

Length Constraints: Maximum length of 256.

Required: No

ModelPackageName

The model package group name.

Type: String

Length Constraints: Maximum length of 256.

Required: No

RegisterNewBaseline

This flag indicates if a newly calculated baseline can be accessed through step properties BaselineUsedForDriftCheckConstraints and BaselineUsedForDriftCheckStatistics. If it is set to False, the previous baseline of the configured check type must also be available. These can be accessed through the BaselineUsedForDriftCheckConstraints property.
Type: Boolean
Required: No

SkipCheck

This flag indicates if the drift check against the previous baseline will be skipped or not. If it is set to False, the previous baseline of the configured check type must be available.

Type: Boolean
Required: No

ViolationReport

The Amazon S3 URI of the violation report if violations are detected.

Type: String
Length Constraints: Maximum length of 1024.
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
ClarifyExplainerConfig

Service: Amazon SageMaker Service

The configuration parameters for the SageMaker Clarify explainer.

Contents

**ShapConfig**

The configuration for SHAP analysis.

Type: `ClarifyShapConfig (p. 1338)` object

Required: Yes

**EnableExplanations**

A JMESPath boolean expression used to filter which records to explain. Explanations are activated by default. See `EnableExplanations` for additional information.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 64.

Pattern: .*

Required: No

**InferenceConfig**

The inference configuration parameter for the model container.

Type: `ClarifyInferenceConfig (p. 1332)` object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ClarifyInferenceConfig

Service: Amazon SageMaker Service

The inference configuration parameter for the model container.

Contents

ContentTemplate

A template string used to format a JSON record into an acceptable model container input. For example, a ContentTemplate string '{"myfeatures":$features}' will format a list of features [1,2,3] into the record string '{"myfeatures":[1,2,3]}'. Required only when the model container input is in JSON Lines format.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 64.

Pattern: .*

Required: No

FeatureHeaders

The names of the features. If provided, these are included in the endpoint response payload to help readability of the InvokeEndpoint output. See the Response section under Invoke the endpoint in the Developer Guide for more information.

Type: Array of strings

Array Members: Minimum number of 1 item. Maximum number of 256 items.

Length Constraints: Minimum length of 1. Maximum length of 64.

Pattern: .*

Required: No

FeaturesAttribute

Provides the JMESPath expression to extract the features from a model container input in JSON Lines format. For example, if FeaturesAttribute is the JMESPath expression 'myfeatures', it extracts a list of features [1,2,3] from request data '{"myfeatures":[1,2,3]}'.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 64.

Pattern: .*

Required: No

FeatureTypes

A list of data types of the features (optional). Applicable only to NLP explainability. If provided, FeatureTypes must have at least one 'text' string (for example, ['text']). If FeatureTypes is not provided, the explainer infers the feature types based on the baseline data. The feature types are included in the endpoint response payload. For additional information see the response section under Invoke the endpoint in the Developer Guide for more information.

Type: Array of strings
Array Members: Minimum number of 1 item. Maximum number of 256 items.

Valid Values: numerical | categorical | text

Required: No

LabelAttribute

A JMESPath expression used to locate the list of label headers in the model container output.

Example: If the model container output of a batch request is '{"labels":
["cat","dog","fish"],"probability":[0.6,0.3,0.1]}', then set LabelAttribute to 'labels' to extract the list of label headers ['cat', 'dog', 'fish']

Type: String

Length Constraints: Minimum length of 1. Maximum length of 64.

Pattern: .*

Required: No

LabelHeaders

For multiclass classification problems, the label headers are the names of the classes. Otherwise, the label header is the name of the predicted label. These are used to help readability for the output of the InvokeEndpoint API. See the response section under Invoke the endpoint in the Developer Guide for more information. If there are no label headers in the model container output, provide them manually using this parameter.

Type: Array of strings

Array Members: Minimum number of 1 item. Maximum number of 16 items.

Length Constraints: Minimum length of 1. Maximum length of 64.

Pattern: .*

Required: No

LabelIndex

A zero-based index used to extract a label header or list of label headers from model container output in CSV format.

Example for a multiclass model: If the model container output consists of label headers followed by probabilities: '"[\'cat\',\'dog\',\'fish\']","[0.1,0.6,0.3]"', set LabelIndex to 0 to select the label headers ['cat', 'dog', 'fish'].

Type: Integer

Valid Range: Minimum value of 0.

Required: No

MaxPayloadInMB

The maximum payload size (MB) allowed of a request from the explainer to the model container. Defaults to 6 MB.

Type: Integer

**MaxRecordCount**

The maximum number of records in a request that the model container can process when querying the model container for the predictions of a synthetic dataset. A record is a unit of input data that inference can be made on, for example, a single line in CSV data. If MaxRecordCount is 1, the model container expects one record per request. A value of 2 or greater means that the model expects batch requests, which can reduce overhead and speed up the inferencing process. If this parameter is not provided, the explainer will tune the record count per request according to the model container’s capacity at runtime.

Type: Integer

Valid Range: Minimum value of 1.

Required: No

**ProbabilityAttribute**

A JMESPath expression used to extract the probability (or score) from the model container output if the model container is in JSON Lines format.

**Example**: If the model container output of a single request is

```
{"predicted_label":1,"probability":0.6}
```

then set ProbabilityAttribute to ‘probability’.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 64.

Pattern: .*

Required: No

**ProbabilityIndex**

A zero-based index used to extract a probability value (score) or list from model container output in CSV format. If this value is not provided, the entire model container output will be treated as a probability value (score) or list.

**Example for a single class model**: If the model container output consists of a string-formatted prediction label followed by its probability: '1,0.6', set ProbabilityIndex to 1 to select the probability value 0.6.

**Example for a multiclass model**: If the model container output consists of a string-formatted prediction label followed by its probability: '

```
[\'cat\',\'dog\',\'fish\']
```

then set ProbabilityIndex to 1 to select the probability values [0.1,0.6,0.3].

Type: Integer

Valid Range: Minimum value of 0.

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for Ruby V3
ClarifyShapBaselineConfig

Service: Amazon SageMaker Service

The configuration for the SHAP baseline (also called the background or reference dataset) of the Kernal SHAP algorithm.

**Note**

- The number of records in the baseline data determines the size of the synthetic dataset, which has an impact on latency of explainability requests. For more information, see the Synthetic data of [Configure and create an endpoint](#).
- ShapBaseline and ShapBaselineUri are mutually exclusive parameters. One or the other is required to configure a SHAP baseline.

**Contents**

**MimeType**

The MIME type of the baseline data. Choose from 'text/csv' or 'application/jsonlines'. Defaults to 'text/csv'.

- **Type**: String
- **Length Constraints**: Maximum length of 255.
- **Pattern**: `^[a-zA-Z0-9-]*[a-zA-Z0-9-]*\/[a-zA-Z0-9-]*[a-zA-Z0-9-]*\/[a-zA-Z0-9-]*[a-zA-Z0-9-]*\/[a-zA-Z0-9-]*$`
- **Required**: No

**ShapBaseline**

The inline SHAP baseline data in string format. ShapBaseline can have one or multiple records to be used as the baseline dataset. The format of the SHAP baseline file should be the same format as the training dataset. For example, if the training dataset is in CSV format and each record contains four features, and all features are numerical, then the format of the baseline data should also share these characteristics. For natural language processing (NLP) of text columns, the baseline value should be the value used to replace the unit of text specified by the Granularity of the TextConfig parameter. The size limit for ShapBaseline is 4 KB. Use the ShapBaselineUri parameter if you want to provide more than 4 KB of baseline data.

- **Type**: String
- **Length Constraints**: Minimum length of 1. Maximum length of 4096.
- **Pattern**: `[^s\S]+`
- **Required**: No

**ShapBaselineUri**

The uniform resource identifier (URI) of the S3 bucket where the SHAP baseline file is stored. The format of the SHAP baseline file should be the same format as the format of the training dataset. For example, if the training dataset is in CSV format, and each record in the training dataset has four features, and all features are numerical, then the baseline file should also have this same format. Each record should contain only the features. If you are using a virtual private cloud (VPC), the ShapBaselineUri should be accessible to the VPC. For more information about setting up endpoints with Amazon Virtual Private Cloud, see [Give SageMaker access to Resources in your Amazon Virtual Private Cloud](#).

- **Type**: String
Length Constraints: Maximum length of 1024.
Pattern: ^(https|s3)://([\^/]+)/(.*$)

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ClarifyShapConfig

Service: Amazon SageMaker Service

The configuration for SHAP analysis using SageMaker Clarify Explainer.

Contents

ShapBaselineConfig

The configuration for the SHAP baseline of the Kernal SHAP algorithm.

Type: ClarifyShapBaselineConfig (p. 1336) object

Required: Yes

NumberOfSamples

The number of samples to be used for analysis by the Kernal SHAP algorithm.

Note

The number of samples determines the size of the synthetic dataset, which has an impact on latency of explainability requests. For more information, see the Synthetic data of Configure and create an endpoint.

Type: Integer

Valid Range: Minimum value of 1.

Required: No

Seed

The starting value used to initialize the random number generator in the explainer. Provide a value for this parameter to obtain a deterministic SHAP result.

Type: Integer

Required: No

TextConfig

A parameter that indicates if text features are treated as text and explanations are provided for individual units of text. Required for natural language processing (NLP) explainability only.

Type: ClarifyTextConfig (p. 1340) object

Required: No

UseLogit

A Boolean toggle to indicate if you want to use the logit function (true) or log-odds units (false) for model predictions. Defaults to false.

Type: Boolean

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

* AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ClarifyTextConfig
Service: Amazon SageMaker Service

A parameter used to configure the SageMaker Clarify explainer to treat text features as text so that explanations are provided for individual units of text. Required only for natural language processing (NLP) explainability.

Contents

Granularity

The unit of granularity for the analysis of text features. For example, if the unit is 'token', then each token (like a word in English) of the text is treated as a feature. SHAP values are computed for each unit/feature.

Type: String

Valid Values: token | sentence | paragraph

Required: Yes

Language

Specifies the language of the text features in ISO 639-1 or ISO 639-3 code of a supported language.

Note
For a mix of multiple languages, use code 'xx'.

Type: String

Valid Values: af | sq | ar | hy | eu | bn | bg | ca | zh | hr | cs | da | nl | en | et | fi | fr | de | el | gu | he | hi | hu | is | id | ga | it | kn | ky | lv | lt | lb | mk | ml | mr | ne | nb | fa | pl | pt | ro | ru | sa | sr | tn | si | sk | sl | es | sv | tl | ta | tt | te | tr | uk | ur | yo | lij | xx

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ClusterInstanceGroupDetails
Service: Amazon SageMaker Service
Details of an instance group in a SageMaker HyperPod cluster.

Contents

CurrentCount
The number of instances that are currently in the instance group of a SageMaker HyperPod cluster.
Type: Integer
Valid Range: Minimum value of 0.
Required: No

ExecutionRole
The execution role for the instance group to assume.
Type: String
Pattern: ^arn:aws[a-z-]*:iam::\d{12}:role/?[a-zA-Z_0-9+=,.@-/]+$ Required: No

InstanceGroupName
The name of the instance group of a SageMaker HyperPod cluster.
Type: String
Pattern: ^[a-zA-Z0-9\-\[\]_\]()]\*\^[a-zA-Z0-9\-\[\]_\]()]*$ Required: No

InstanceType
The instance type of the instance group of a SageMaker HyperPod cluster.
Type: String
Valid Values: ml.p4d.24xlarge | ml.p4de.24xlarge | ml.p5.48xlarge | ml.trn1.32xlarge | ml.trn1n.32xlarge | ml.g5.xlarge | ml.g5.2xlarge | ml.g5.4xlarge | ml.g5.8xlarge | ml.g5.12xlarge | ml.g5.16xlarge | ml.g5.24xlarge | ml.g5.48xlarge | ml.c5.large | ml.c5.xlarge | ml.c5.2xlarge | ml.c5.4xlarge | ml.c5.9xlarge | ml.c5.12xlarge | ml.c5.18xlarge | ml.c5.24xlarge | ml.c5n.large | ml.c5n.2xlarge | ml.c5n.4xlarge | ml.c5n.9xlarge | ml.c5n.18xlarge | ml.m5.large | ml.m5.xlarge | ml.m5.2xlarge | ml.m5.4xlarge | ml.m5.8xlarge | ml.m5.12xlarge | ml.m5.16xlarge | ml.m5.24xlarge | ml.t3.medium | ml.t3.large | ml.t3.xlarge | ml.t3.2xlarge
Required: No

LifeCycleConfig
Details of LifeCycle configuration for the instance group.
Type: `ClusterLifeCycleConfig (p. 1346)` object

Required: No

**TargetCount**

The number of instances you specified to add to the instance group of a SageMaker HyperPod cluster.

Type: Integer

Valid Range: Minimum value of 1.

Required: No

**ThreadsPerCore**

The number you specified to `TreadsPerCore` in `CreateCluster` for enabling or disabling multithreading. For instance types that support multithreading, you can specify 1 for disabling multithreading and 2 for enabling multithreading. For more information, see the reference table of [CPU cores and threads per CPU core per instance type](https://docs.aws.amazon.com/AmazonElasticComputeCloud/latest/UserGuide/index.html) in the `Amazon Elastic Compute Cloud User Guide`.

Type: Integer


Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](https://docs.aws.amazon.com/sdk-for-cpp/v1/developer-guide/com Moreno-1343.html)
- [AWS SDK for Java V2](https://docs.aws.amazon.com/aws-java-sdk/latest/com Moreno-1343.html)
- [AWS SDK for Ruby V3](https://docs.aws.amazon.com/aws-ruby-sdk/latest/com Moreno-1343.html)
ClusterInstanceGroupSpecification

Service: Amazon SageMaker Service

The specifications of an instance group that you need to define.

Contents

ExecutionRole

Specifies an IAM execution role to be assumed by the instance group.

Type: String


Pattern: ^arn:aws[a-z\-]*:iam::\d{12}:role/\|a-zA-Z0-9\-\+=,.@-\_/]+$

Required: Yes

InstanceCount

Specifies the number of instances to add to the instance group of a SageMaker HyperPod cluster.

Type: Integer

Valid Range: Minimum value of 1.

Required: Yes

InstanceGroupName

Specifies the name of the instance group.

Type: String


Pattern: ^[a-zA-Z0-9\-\]+\-*[a-zA-Z0-9\-\]+$

Required: Yes

InstanceType

Specifies the instance type of the instance group.

Type: String

Valid Values:
- ml.p4d.24xlarge
- ml.p4de.24xlarge
- ml.p5.48xlarge
- ml.p4d.24xlarge
- ml.p4de.24xlarge
- ml.p5.48xlarge
- ml.g5.4xlarge
- ml.g5.8xlarge
- ml.g5.12xlarge
- ml.g5.16xlarge
- ml.g5.24xlarge
- ml.g5.48xlarge
- ml.m5.large
- ml.m5.xlarge
- ml.m5.2xlarge
- ml.m5.4xlarge
- ml.m5.8xlarge
- ml.m5.12xlarge
- ml.m5.16xlarge
- ml.m5.24xlarge
- ml.m5.48xlarge
- ml.m5.9xlarge
- ml.m5.18xlarge
- ml.m5.large
- ml.m5.xlarge
- ml.m5.2xlarge
- ml.m5.4xlarge
- ml.m5.8xlarge
- ml.m5.12xlarge
- ml.m5.16xlarge
- ml.m5.24xlarge
- ml.t3.medium
- ml.t3.large
- ml.t3.xlarge
- ml.t3.2xlarge

Required: Yes

LifeCycleConfig

Specifies the LifeCycle configuration for the instance group.
Type: ClusterLifeCycleConfig (p. 1346) object

Required: Yes

**ThreadsPerCore**

Specifies the value for **Threads per core**. For instance types that support multithreading, you can specify 1 for disabling multithreading and 2 for enabling multithreading. For instance types that don't support multithreading, specify 1. For more information, see the reference table of CPU cores and threads per CPU core per instance type in the Amazon Elastic Compute Cloud User Guide.

Type: Integer


Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
ClusterInstanceStatusDetails
Service: Amazon SageMaker Service

Details of an instance in a SageMaker HyperPod cluster.

Contents

Status
The status of an instance in a SageMaker HyperPod cluster.

Type: String

Valid Values: Running | Failure | Pending | ShuttingDown | SystemUpdating

Required: Yes

Message
The message from an instance in a SageMaker HyperPod cluster.

Type: String

Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ClusterLifeCycleConfig

Service: Amazon SageMaker Service

The LifeCycle configuration for a SageMaker HyperPod cluster.

Contents

OnCreate

The directory of the LifeCycle script under SourceS3Uri. This LifeCycle script runs during cluster creation.

Type: String


Pattern: ^\S+\$

Required: Yes

SourceS3Uri

An Amazon S3 bucket path where your LifeCycle scripts are stored.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: ^(https|s3)://[^/]+/(.*)$

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ClusterNodeDetails

Service: Amazon SageMaker Service

Details of an instance (also called a node interchangeably) in a SageMaker HyperPod cluster.

Contents

**InstanceGroupName**

The instance group name in which the instance is.

Type: String


Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9])*$`

Required: No

**InstanceId**

The ID of the instance.

Type: String

Required: No

**InstanceStatus**

The status of the instance.

Type: ClusterInstanceStatusDetails (p. 1345) object

Required: No

**InstanceType**

The type of the instance.

Type: String

Valid Values: ml.p4d.24xlarge | ml.p4de.24xlarge | ml.p5.48xlarge | ml.trn1.32xlarge | ml.trn1n.32xlarge | ml.g5.xlarge | ml.g5.2xlarge | ml.g5.4xlarge | ml.g5.8xlarge | ml.g5.12xlarge | ml.g5.16xlarge | ml.g5.24xlarge | ml.g5.48xlarge | ml.c5.large | ml.c5.xlarge | ml.c5.2xlarge | ml.c5.4xlarge | ml.c5.9xlarge | ml.c5.12xlarge | ml.c5.18xlarge | ml.c5.24xlarge | ml.c5n.large | ml.c5n.xlarge | ml.c5n.2xlarge | ml.c5n.4xlarge | ml.c5n.9xlarge | ml.c5n.18xlarge | ml.m5.large | ml.m5.xlarge | ml.m5.2xlarge | ml.m5.4xlarge | ml.m5.8xlarge | ml.m5.12xlarge | ml.m5.16xlarge | ml.m5.24xlarge | ml.t3.medium | ml.t3.large | ml.t3.xlarge | ml.t3.2xlarge

Required: No

**LaunchTime**

The time when the instance is launched.

Type: Timestamp

Required: No
LifeCycleConfig

The LifeCycle configuration applied to the instance.

Type: `ClusterLifeCycleConfig (p. 1346)` object

Required: No

ThreadsPerCore

The number of threads per CPU core you specified under CreateCluster.

Type: Integer


Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
ClusterNodeSummary

Service: Amazon SageMaker Service

Lists a summary of the properties of an instance (also called a node interchangeably) of a SageMaker HyperPod cluster.

Contents

InstanceGroupName

The name of the instance group in which the instance is.

Type: String


Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9])*`$ 

Required: Yes

InstanceId

The ID of the instance.

Type: String

Required: Yes

InstanceStatus

The status of the instance.

Type: ClusterInstanceStatusDetails (p. 1345) object

Required: Yes

InstanceType

The type of the instance.

Type: String

Valid Values: ml.p4d.24xlarge | ml.p4de.24xlarge | ml.p5.48xlarge | ml.ptrn1.32xlarge | ml.p5n.32xlarge | ml.g5.xlarge | ml.g5.2xlarge | ml.g5.4xlarge | ml.g5.8xlarge | ml.g5.12xlarge | ml.g5.16xlarge | ml.g5.24xlarge | ml.g5.48xlarge | ml.c5.large | ml.c5.xlarge | ml.c5.2xlarge | ml.c5.4xlarge | ml.c5.9xlarge | ml.c5.12xlarge | ml.c5.18xlarge | ml.c5.24xlarge | ml.c5n.large | ml.c5n.2xlarge | ml.c5n.4xlarge | ml.c5n.9xlarge | ml.c5n.18xlarge | ml.m5.large | ml.m5.xlarge | ml.m5.2xlarge | ml.m5.4xlarge | ml.m5.8xlarge | ml.m5.12xlarge | ml.m5.16xlarge | ml.m5.24xlarge | ml.t3.medium | ml.t3.large | ml.t3.xlarge | ml.t3.2xlarge

Required: Yes

LaunchTime

The time when the instance is launched.

Type: Timestamp

Required: Yes
See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](https://aws.amazon.com/cplusplus-sdk/)
- [AWS SDK for Go](https://aws.amazon.com/go-sdk/)
- [AWS SDK for Java V2](https://aws.amazon.com/java-sdk/)
- [AWS SDK for Ruby V3](https://aws.amazon.com/ruby-sdk/)
ClusterSummary
Service: Amazon SageMaker Service

Lists a summary of the properties of a SageMaker HyperPod cluster.

Contents

ClusterArn
The Amazon Resource Name (ARN) of the SageMaker HyperPod cluster.
Type: String
Length Constraints: Maximum length of 256.
Pattern: ^arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:cluster/[a-z0-9]{12}$
Required: Yes

ClusterName
The name of the SageMaker HyperPod cluster.
Type: String
Pattern: ^[a-zA-Z0-9\-](-*[a-zA-Z0-9])*$
Required: Yes

ClusterStatus
The status of the SageMaker HyperPod cluster.
Type: String
Valid Values: Creating | Deleting | Failed | InService | RollingBack | SystemUpdating | Updating
Required: Yes

CreationTime
The time when the SageMaker HyperPod cluster is created.
Type: Timestamp
Required: Yes

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
CodeEditorAppSettings
Service: Amazon SageMaker Service
The Code Editor application settings.
For more information about Code Editor, see Get started with Code Editor in Amazon SageMaker.

Contents

DefaultResourceSpec
Specifies the ARN's of a SageMaker image and SageMaker image version, and the instance type that the version runs on.
Type: ResourceSpec (p. 1915) object
Required: No

LifecycleConfigArns
The Amazon Resource Name (ARN) of the Code Editor application lifecycle configuration.
Type: Array of strings
Length Constraints: Maximum length of 256.
Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:studio-lifecycle-config/.*
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
**CodeRepository**

Service: Amazon SageMaker Service

A Git repository that SageMaker automatically displays to users for cloning in the JupyterServer application.

**Contents**

**RepositoryUrl**

The URL of the Git repository.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: `^https://[.\-_a-zA-Z0-9]+/?{3,1016}$`

Required: Yes

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
CodeRepositorySummary

Service: Amazon SageMaker Service

Specifies summary information about a Git repository.

Contents

CodeRepositoryArn

The Amazon Resource Name (ARN) of the Git repository.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: ^arn:aws(-cn|-us-gov)?:sagemaker:[a-z0-9\-]{9,16}:[0-9]{12}:code-repository/[\S]{1,2048}$

Required: Yes

CodeRepositoryName

The name of the Git repository.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}$

Required: Yes

CreationTime

The date and time that the Git repository was created.

Type: Timestamp

Required: Yes

LastModifiedTime

The date and time that the Git repository was last modified.

Type: Timestamp

Required: Yes

GitConfig

Configuration details for the Git repository, including the URL where it is located and the ARN of the AWS Secrets Manager secret that contains the credentials used to access the repository.

Type: GitConfig (p. 1505) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for Ruby V3
CognitoConfig
Service: Amazon SageMaker Service

Use this parameter to configure your Amazon Cognito workforce. A single Cognito workforce is created using and corresponds to a single Amazon Cognito user pool.

Contents

ClientId

The client ID for your Amazon Cognito user pool.

Type: String


Pattern: [ -~]+

Required: Yes

UserPool

A user pool is a user directory in Amazon Cognito. With a user pool, your users can sign in to your web or mobile app through Amazon Cognito. Your users can also sign in through social identity providers like Google, Facebook, Amazon, or Apple, and through SAML identity providers.

Type: String


Pattern: [\w-]+_[0-9a-zA-Z]+

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
CognitoMemberDefinition
Service: Amazon SageMaker Service

Identifies a Amazon Cognito user group. A user group can be used in on or more work teams.

Contents

ClientId

An identifier for an application client. You must create the app client ID using Amazon Cognito.

Type: String
Pattern: [ -~]+
Required: Yes

UserGroup

An identifier for a user group.

Type: String
Pattern: [\p{L}\p{M}\p{S}\p{N}\p{P}]+
Required: Yes

UserPool

An identifier for a user pool. The user pool must be in the same region as the service that you are calling.

Type: String
Pattern: [\w-]+[\0-9a-zA-Z]+
Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
CollectionConfig
Service: Amazon SageMaker Service
Configuration for your collection.

Contents

**Important**
This data type is a UNION, so only one of the following members can be specified when used or returned.

**VectorConfig**
Configuration for your vector collection type.
- **Dimension**: The number of elements in your vector.

Type: [VectorConfig](#) object
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
CollectionConfiguration
Service: Amazon SageMaker Service

Configuration information for the Amazon SageMaker Debugger output tensor collections.

Contents

**CollectionName**

The name of the tensor collection. The name must be unique relative to other rule configuration names.

Type: String
Length Constraints: Minimum length of 1. Maximum length of 256.
Pattern: .*
Required: No

**CollectionParameters**

Parameter values for the tensor collection. The allowed parameters are "name", "include_regex", "reduction_config", "save_config", "tensor_names", and "save_histogram".

Type: String to string map
Map Entries: Minimum number of 0 items. Maximum number of 20 items.
Key Length Constraints: Minimum length of 1. Maximum length of 256.
Key Pattern: .*
Value Length Constraints: Maximum length of 256.
Value Pattern: .*
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
CompilationJobSummary
Service: Amazon SageMaker Service
A summary of a model compilation job.

Contents

CompilationJobArn
The Amazon Resource Name (ARN) of the model compilation job.
Type: String
Length Constraints: Maximum length of 256.
Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:compilation-job/.*
Required: Yes

CompilationJobName
The name of the model compilation job that you want a summary for.
Type: String
Pattern: ^[a-zA-Z0-9\-](\*\[a-zA-Z0-9\-]\{0,62}\}$
Required: Yes

CompilationJobStatus
The status of the model compilation job.
Type: String
Valid Values: INPROGRESS | COMPLETED | FAILED | STARTING | STOPPING | STOPPED
Required: Yes

CreationTime
The time when the model compilation job was created.
Type: Timestamp
Required: Yes

CompilationEndTime
The time when the model compilation job completed.
Type: Timestamp
Required: No

CompilationStartTime
The time when the model compilation job started.
Type: Timestamp
Required: No
CompilationTargetDevice

The type of device that the model will run on after the compilation job has completed.

Type: String

Valid Values: lambda | ml_m4 | ml_m5 | ml_c4 | ml_c5 | ml_p2 | ml_p3 | ml_g4dn
| ml_inf1 | ml_inf2 | ml_trn1 | ml_eia2 | jetson_tx1 | jetson_tx2 | jetson_nano | jetson_xavier | rasp3b | imx8qm | deeplens | rk3399 | rk3288
| aisage | sbe_c | qcs605 | qcs603 | sitara_am57x | amba_cv2 | amba_cv22 | amba_cv25 | x86_win32 | x86_win64 | coreml | jacinto_tda4vm | imx8mpplus

Required: No

CompilationTargetPlatformAccelerator

The type of accelerator that the model will run on after the compilation job has completed.

Type: String

Valid Values: INTEL_GRAPHICS | MALI | NVIDIA | NNA

Required: No

CompilationTargetPlatformArch

The type of architecture that the model will run on after the compilation job has completed.

Type: String

Valid Values: X86_64 | X86 | ARM64 | ARM_EABI | ARM_EABIHF

Required: No

CompilationTargetPlatformOs

The type of OS that the model will run on after the compilation job has completed.

Type: String

Valid Values: ANDROID | LINUX

Required: No

LastModifiedTime

The time when the model compilation job was last modified.

Type: Timestamp

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ConditionStepMetadata
Service: Amazon SageMaker Service

Metadata for a Condition step.

Contents

Outcome

The outcome of the Condition step evaluation.

Type: String

Valid Values: True | False

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
**ContainerConfig**

Service: Amazon SageMaker Service

The configuration used to run the application image container.

**Contents**

**ContainerArguments**

The arguments for the container when you're running the application.

- Type: Array of strings
- Array Members: Maximum number of 50 items.
- Length Constraints: Maximum length of 64.
- Pattern: `^(?!\s*$).+`
- Required: No

**ContainerEntrypoint**

The entrypoint used to run the application in the container.

- Type: Array of strings
- Array Members: Maximum number of 1 item.
- Length Constraints: Maximum length of 256.
- Pattern: `^(?!\s*$).+`
- Required: No

**ContainerEnvironmentVariables**

The environment variables to set in the container

- Type: String to string map
- Map Entries: Maximum number of 25 items.
- Key Length Constraints: Maximum length of 256.
- Key Pattern: `^(?!\s*$).+`
- Value Length Constraints: Maximum length of 256.
- Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
**ContainerDefinition**

Service: Amazon SageMaker Service

Describes the container, as part of model definition.

**Contents**

**ContainerHostname**

This parameter is ignored for models that contain only a PrimaryContainer.

When a ContainerDefinition is part of an inference pipeline, the value of the parameter uniquely identifies the container for the purposes of logging and metrics. For information, see [Use Logs and Metrics to Monitor an Inference Pipeline](https://docs.aws.amazon.com/sagemaker/latest/dg/monitor-inference-pipeline.html). If you don't specify a value for this parameter for a ContainerDefinition that is part of an inference pipeline, a unique name is automatically assigned based on the position of the ContainerDefinition in the pipeline. If you specify a value for the ContainerHostname for any ContainerDefinition that is part of an inference pipeline, you must specify a value for the ContainerHostname parameter of every ContainerDefinition in that pipeline.

Type: String

Length Constraints: Maximum length of 63.

Pattern: `^[a-zA-Z0-9-]*[a-zA-Z0-9-]{0,62}$`

Required: No

**Environment**

The environment variables to set in the Docker container. Each key and value in the Environment string to string map can have length of up to 1024. We support up to 16 entries in the map.

Type: String to string map

Map Entries: Maximum number of 16 items.

Key Length Constraints: Maximum length of 1024.

Key Pattern: `[a-zA-Z_][a-zA-Z0-9_]*`

Value Length Constraints: Maximum length of 1024.

Value Pattern: `[^\S\s]*`

Required: No

**Image**

The path where inference code is stored. This can be either in Amazon EC2 Container Registry or in a Docker registry that is accessible from the same VPC that you configure for your endpoint. If you are using your own custom algorithm instead of an algorithm provided by SageMaker, the inference code must meet SageMaker requirements. SageMaker supports both `registry/repository[:tag]` and `registry/repository[@digest]` image path formats. For more information, see [Using Your Own Algorithms with Amazon SageMaker](https://docs.aws.amazon.com/sagemaker/latest/dg/using-your-own-algorithm.html).

**Note**

The model artifacts in an Amazon S3 bucket and the Docker image for inference container in Amazon EC2 Container Registry must be in the same region as the model or endpoint you are creating.
Type: String

Length Constraints: Maximum length of 255.

Pattern: \[\S\]+

Required: No

**ImageConfig**

Specifies whether the model container is in Amazon ECR or a private Docker registry accessible from your Amazon Virtual Private Cloud (VPC). For information about storing containers in a private Docker registry, see Use a Private Docker Registry for Real-Time Inference Containers.

**Note**

The model artifacts in an Amazon S3 bucket and the Docker image for inference container in Amazon EC2 Container Registry must be in the same region as the model or endpoint you are creating.

Type: ImageConfig (p. 1574) object

Required: No

**InferenceSpecificationName**

The inference specification name in the model package version.

Type: String


Pattern: ^[a-zA-Z0-9][-]*[a-zA-Z0-9]{0,62}$

Required: No

**Mode**

Whether the container hosts a single model or multiple models.

Type: String

Valid Values: SingleModel | MultiModel

Required: No

**ModelDataSource**

Specifies the location of ML model data to deploy.

**Note**


Type: ModelDataSource (p. 1688) object

Required: No

**ModelDataUrl**

The S3 path where the model artifacts, which result from model training, are stored. This path must point to a single gzip compressed tar archive (.tar.gz suffix). The S3 path is required for SageMaker built-in algorithms, but not if you use your own algorithms. For more information on built-in algorithms, see Common Parameters.
**Note**
The model artifacts must be in an S3 bucket that is in the same region as the model or endpoint you are creating.

If you provide a value for this parameter, SageMaker uses AWS Security Token Service to download model artifacts from the S3 path you provide. AWS STS is activated in your AWS account by default. If you previously deactivated AWS STS for a region, you need to reactivate AWS STS for that region. For more information, see *Activating and Deactivating AWS STS in an AWS Region* in the *AWS Identity and Access Management User Guide*.

**Important**
If you use a built-in algorithm to create a model, SageMaker requires that you provide a S3 path to the model artifacts in `ModelDataUrl`.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: `^(https|s3)://([^/]+)/?([^]*)$`

Required: No

**ModelPackageName**

The name or Amazon Resource Name (ARN) of the model package to use to create the model.

Type: String


Pattern: `(arn:aws[a-zA-Z-]*:sagemaker:[a-zA-Z-]*:[0-9]{12}:[a-zA-Z-]*\[/][\[0-9\]{1,5})?([a-zA-Z0-9-a-zA-Z0-9]*\[0,62])\[\[0-9\]{1,5}]$`

Required: No

**MultiModelConfig**

Specifies additional configuration for multi-model endpoints.

Type: `MultiModelConfig (p. 1766)` object

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ContextSource
Service: Amazon SageMaker Service
A structure describing the source of a context.

Contents

SourceUri
The URI of the source.
Type: String
Length Constraints: Maximum length of 2048.
Required: Yes

SourceId
The ID of the source.
Type: String
Length Constraints: Maximum length of 256.
Required: No

SourceType
The type of the source.
Type: String
Length Constraints: Maximum length of 256.
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ContextSummary

Service: Amazon SageMaker Service

Lists a summary of the properties of a context. A context provides a logical grouping of other entities.

Contents

ContextArn

The Amazon Resource Name (ARN) of the context.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:context/.*

Required: No

ContextName

The name of the context.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,119}$

Required: No

ContextType

The type of the context.

Type: String

Length Constraints: Maximum length of 256.

Required: No

CreationTime

When the context was created.

Type: Timestamp

Required: No

LastModifiedTime

When the context was last modified.

Type: Timestamp

Required: No

Source

The source of the context.

Type: ContextSource (p. 1369) object

Required: No
See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ContinuousParameterRange
Service: Amazon SageMaker Service

A list of continuous hyperparameters to tune.

Contents

MaxValue

The maximum value for the hyperparameter. The tuning job uses floating-point values between MinValue value and this value for tuning.

Type: String
Length Constraints: Maximum length of 256.
Pattern: .*
Required: Yes

MinValue

The minimum value for the hyperparameter. The tuning job uses floating-point values between this value and MaxValue for tuning.

Type: String
Length Constraints: Maximum length of 256.
Pattern: .*
Required: Yes

Name

The name of the continuous hyperparameter to tune.

Type: String
Length Constraints: Maximum length of 256.
Pattern: .*
Required: Yes

ScalingType

The scale that hyperparameter tuning uses to search the hyperparameter range. For information about choosing a hyperparameter scale, see Hyperparameter Scaling. One of the following values:

Auto

SageMaker hyperparameter tuning chooses the best scale for the hyperparameter.

Linear

Hyperparameter tuning searches the values in the hyperparameter range by using a linear scale.

Logarithmic

Hyperparameter tuning searches the values in the hyperparameter range by using a logarithmic scale.

Logarithmic scaling works only for ranges that have only values greater than 0.
ReverseLogarithmic

Hyperparameter tuning searches the values in the hyperparameter range by using a reverse logarithmic scale.

Reverse logarithmic scaling works only for ranges that are entirely within the range 0<=x<1.0.

Type: String

Valid Values: Auto | Linear | Logarithmic | ReverseLogarithmic

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ContinuousParameterRangeSpecification
Service: Amazon SageMaker Service

Defines the possible values for a continuous hyperparameter.

Contents

MaxValue
The maximum floating-point value allowed.
Type: String
Length Constraints: Maximum length of 256.
Pattern: .*
Required: Yes

MinValue
The minimum floating-point value allowed.
Type: String
Length Constraints: Maximum length of 256.
Pattern: .*
Required: Yes

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ConvergenceDetected

Service: Amazon SageMaker Service

A flag to indicating that automatic model tuning (AMT) has detected model convergence, defined as a lack of significant improvement (1% or less) against an objective metric.

Contents

CompleteOnConvergence

A flag to stop a tuning job once AMT has detected that the job has converged.

Type: String

Valid Values: Disabled | Enabled

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
**CustomFileSystem**

**Service:** Amazon SageMaker Service

A file system, created by you, that you assign to a user profile or space for an Amazon SageMaker Domain. Permitted users can access this file system in Amazon SageMaker Studio.

**Contents**

**Important**

This data type is a UNION, so only one of the following members can be specified when used or returned.

**EFSFileSystem**

A custom file system in Amazon EFS.

Type: [EFSFileSystem](#) object

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
CustomFileSystemConfig
Service: Amazon SageMaker Service

The settings for assigning a custom file system to a user profile or space for an Amazon SageMaker Domain. Permitted users can access this file system in Amazon SageMaker Studio.

Contents

Important
This data type is a UNION, so only one of the following members can be specified when used or returned.

EFSFileSystemConfig
The settings for a custom Amazon EFS file system.

Type: EFSFileSystemConfig (p. 1450) object
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
CustomImage
Service: Amazon SageMaker Service

A custom SageMaker image. For more information, see Bring your own SageMaker image.

Contents

AppImageConfigName

The name of the AppImageConfig.
Type: String
Length Constraints: Maximum length of 63.
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$
Required: Yes

ImageName

The name of the CustomImage. Must be unique to your account.
Type: String
Pattern: ^[a-zA-Z0-9][-._]?[a-zA-Z0-9]{0,62}$
Required: Yes

ImageVersionNumber

The version number of the CustomImage.
Type: Integer
Valid Range: Minimum value of 0.
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
CustomizedMetricSpecification
Service: Amazon SageMaker Service

A customized metric.

Contents

MetricName
   The name of the customized metric.
   Type: String
   Required: No

Namespace
   The namespace of the customized metric.
   Type: String
   Required: No

Statistic
   The statistic of the customized metric.
   Type: String
   Valid Values: Average | Minimum | Maximum | SampleCount | Sum
   Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
CustomPosixUserConfig
Service: Amazon SageMaker Service

Details about the POSIX identity that is used for file system operations.

Contents

Gid

The POSIX group ID.
Type: Long
Required: Yes

Uid

The POSIX user ID.
Type: Long
Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
DataCaptureConfig
Service: Amazon SageMaker Service

Configuration to control how SageMaker captures inference data.

Contents

CaptureOptions

Specifies data Model Monitor will capture. You can configure whether to collect only input, only output, or both.

Type: Array of CaptureOption (p. 1320) objects

Array Members: Minimum number of 1 item. Maximum number of 2 items.

Required: Yes

DestinationS3Uri

The Amazon S3 location used to capture the data.

Type: String

Length Constraints: Maximum length of 512.

Pattern: ^(https|s3)://(\[^/\]/?[^/\)/?'\]+$)

Required: Yes

InitialSamplingPercentage

The percentage of requests SageMaker will capture. A lower value is recommended for Endpoints with high traffic.

Type: Integer

Valid Range: Minimum value of 0. Maximum value of 100.

Required: Yes

CaptureContentTypeHeader

Configuration specifying how to treat different headers. If no headers are specified SageMaker will by default base64 encode when capturing the data.

Type: CaptureContentTypeHeader (p. 1319) object

Required: No

EnableCapture

Whether data capture should be enabled or disabled (defaults to enabled).

Type: Boolean

Required: No

KmsKeyId

The Amazon Resource Name (ARN) of an AWS Key Management Service key that SageMaker uses to encrypt the captured data at rest using Amazon S3 server-side encryption.

The KmsKeyId can be any of the following formats:
• Key ID: 1234abcd-12ab-34cd-56ef-1234567890ab
• Key ARN: arn:aws:kms:us-west-2:111122223333:key/1234abcd-12ab-34cd-56ef-1234567890ab
• Alias name: alias/ExampleAlias
• Alias name ARN: arn:aws:kms:us-west-2:111122223333:alias/ExampleAlias

Type: String
Length Constraints: Maximum length of 2048.
Pattern: .*
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

• AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for Ruby V3
DataCaptureConfigSummary

Service: Amazon SageMaker Service

The currently active data capture configuration used by your Endpoint.

Contents

CaptureStatus

Whether data capture is currently functional.

Type: String

Valid Values: Started | Stopped

Required: Yes

CurrentSamplingPercentage

The percentage of requests being captured by your Endpoint.

Type: Integer

Valid Range: Minimum value of 0. Maximum value of 100.

Required: Yes

DestinationS3Uri

The Amazon S3 location being used to capture the data.

Type: String

Length Constraints: Maximum length of 512.

Pattern: ^(https|s3)://([^/]+)/(.*)$

Required: Yes

EnableCapture

Whether data capture is enabled or disabled.

Type: Boolean

Required: Yes

KmsKeyId

The KMS key being used to encrypt the data in Amazon S3.

Type: String

Length Constraints: Maximum length of 2048.

Pattern: .*

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:
• AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for Ruby V3
DataCatalogConfig
Service: Amazon SageMaker Service

The meta data of the Glue table which serves as data catalog for the OfflineStore.

Contents

Catalog
The name of the Glue table catalog.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 255.
Pattern: \[\u0020-\uD7FF\uE000-\uFFFD\uD800-\uDBFF\uDFFF\t]*
Required: Yes

Database
The name of the Glue table database.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 255.
Pattern: \[\u0020-\uD7FF\uE000-\uFFFD\uD800-\uDBFF\uDFFF\t]*
Required: Yes

TableName
The name of the Glue table.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 255.
Pattern: \[\u0020-\uD7FF\uE000-\uFFFD\uD800-\uDBFF\uDFFF\t]*
Required: Yes

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
DataProcessing

Service: Amazon SageMaker Service

The data structure used to specify the data to be used for inference in a batch transform job and to associate the data that is relevant to the prediction results in the output. The input filter provides you to exclude input data that is not needed for inference in a batch transform job. The output filter provides you to include input data relevant to interpreting the predictions in the output from the job. For more information, see Associate Prediction Results with their Corresponding Input Records.

Contents

InputFilter

A JSONPath expression used to select a portion of the input data to pass to the algorithm. Use the InputFilter parameter to exclude fields, such as an ID column, from the input. If you want SageMaker to pass the entire input dataset to the algorithm, accept the default value $.

Examples: "$", "$[1:]", "$.features"

Type: String

Length Constraints: Minimum length of 0. Maximum length of 63.

Required: No

JoinSource

Specifies the source of the data to join with the transformed data. The valid values are None and Input. The default value is None, which specifies not to join the input with the transformed data. If you want the batch transform job to join the original input data with the transformed data, set JoinSource to Input. You can specify OutputFilter as an additional filter to select a portion of the joined dataset and store it in the output file.

For JSON or JSONLines objects, such as a JSON array, SageMaker adds the transformed data to the input JSON object in an attribute called SageMakerOutput. The joined result for JSON must be a key-value pair object. If the input is not a key-value pair object, SageMaker creates a new JSON file. In the new JSON file, and the input data is stored under the SageMakerInput key and the results are stored in SageMakerOutput.

For CSV data, SageMaker takes each row as a JSON array and joins the transformed data with the input by appending each transformed row to the end of the input. The joined data has the original input data followed by the transformed data and the output is a CSV file.

For information on how joining in applied, see Workflow for Associating Inferences with Input Records.

Type: String

Valid Values: Input | None

Required: No

OutputFilter

A JSONPath expression used to select a portion of the joined dataset to save in the output file for a batch transform job. If you want SageMaker to store the entire input dataset in the output file, leave the default value, $. If you specify indexes that aren't within the dimension size of the joined dataset, you get an error.

Examples: "$", "[$0,5:]", "["id","SageMakerOutput"]"
Type: String

Length Constraints: Minimum length of 0. Maximum length of 63.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
DataQualityAppSpecification

Service: Amazon SageMaker Service

Information about the container that a data quality monitoring job runs.

Contents

ImageUri

The container image that the data quality monitoring job runs.

Type: String

Length Constraints: Maximum length of 255.

Pattern: .*

Required: Yes

ContainerArguments

The arguments to send to the container that the monitoring job runs.

Type: Array of strings

Array Members: Minimum number of 1 item. Maximum number of 50 items.

Length Constraints: Maximum length of 256.

Pattern: .*

Required: No

ContainerEntrypoint

The entrypoint for a container used to run a monitoring job.

Type: Array of strings

Array Members: Minimum number of 1 item. Maximum number of 100 items.

Length Constraints: Maximum length of 256.

Pattern: .*

Required: No

Environment

Sets the environment variables in the container that the monitoring job runs.

Type: String to string map

Map Entries: Maximum number of 50 items.

Key Length Constraints: Maximum length of 256.

Key Pattern: [a-zA-Z_][a-zA-Z0-9_.]*

Value Length Constraints: Maximum length of 256.

Value Pattern: [\S\s]*
**PostAnalyticsProcessorSourceUri**

An Amazon S3 URI to a script that is called after analysis has been performed. Applicable only for the built-in (first party) containers.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: ^(https|s3)://([^/]+)/(.*)$

Required: No

**RecordPreprocessorSourceUri**

An Amazon S3 URI to a script that is called per row prior to running analysis. It can base64 decode the payload and convert it into a flattened JSON so that the built-in container can use the converted data. Applicable only for the built-in (first party) containers.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: ^(https|s3)://([^/]+)/(.*)$

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
DataQualityBaselineConfig
Service: Amazon SageMaker Service

Configuration for monitoring constraints and monitoring statistics. These baseline resources are compared against the results of the current job from the series of jobs scheduled to collect data periodically.

Contents

BaseliningJobName
The name of the job that performs baselining for the data quality monitoring job.
Type: String
Pattern: ^[a-zA-Z0-9\-\[\]*a-zA-Z0-9]{0,62}$
Required: No

ConstraintsResource
The constraints resource for a monitoring job.
Type: MonitoringConstraintsResource (p. 1740) object
Required: No

StatisticsResource
The statistics resource for a monitoring job.
Type: MonitoringStatisticsResource (p. 1764) object
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
DataQualityJobInput

Service: Amazon SageMaker Service

The input for the data quality monitoring job. Currently endpoints are supported for input.

Contents

BatchTransformInput

Input object for the batch transform job.

Type: BatchTransformInput (p. 1305) object

Required: No

EndpointInput

Input object for the endpoint

Type: EndpointInput (p. 1457) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
DatasetDefinition
Service: Amazon SageMaker Service

Configuration for Dataset Definition inputs. The Dataset Definition input must specify exactly one of either AthenaDatasetDefinition or RedshiftDatasetDefinition types.

Contents

AthenaDatasetDefinition
Configuration for Athena Dataset Definition input.

Type: AthenaDatasetDefinition (p. 1262) object

Required: No

DataDistributionType
Whether the generated dataset is FullyReplicated or ShardedByS3Key (default).

Type: String

Valid Values: FullyReplicated | ShardedByS3Key

Required: No

InputMode
Whether to use File or Pipe input mode. In File (default) mode, Amazon SageMaker copies the data from the input source onto the local Amazon Elastic Block Store (Amazon EBS) volumes before starting your training algorithm. This is the most commonly used input mode. In Pipe mode, Amazon SageMaker streams input data from the source directly to your algorithm without using the EBS volume.

Type: String

Valid Values: Pipe | File

Required: No

LocalPath
The local path where you want Amazon SageMaker to download the Dataset Definition inputs to run a processing job. LocalPath is an absolute path to the input data. This is a required parameter when AppManaged is False (default).

Type: String

Length Constraints: Maximum length of 256.

Pattern: .*

Required: No

RedshiftDatasetDefinition
Configuration for Redshift Dataset Definition input.

Type: RedshiftDatasetDefinition (p. 1902) object

Required: No
See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
**DataSource**

Service: Amazon SageMaker Service

Describes the location of the channel data.

**Contents**

**FileSystemDataSource**

The file system that is associated with a channel.

Type: [FileSystemDataSource](p. 1494) object

Required: No

**S3DataSource**

The S3 location of the data source that is associated with a channel.

Type: [S3DataSource](p. 1924) object

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](AWS SDK for C++)
- [AWS SDK for Go](AWS SDK for Go)
- [AWS SDK for Java V2](AWS SDK for Java V2)
- [AWS SDK for Ruby V3](AWS SDK for Ruby V3)
DebugHookConfig

Service: Amazon SageMaker Service

Configuration information for the Amazon SageMaker Debugger hook parameters, metric and tensor collections, and storage paths. To learn more about how to configure the DebugHookConfig parameter, see Use the SageMaker and Debugger Configuration API Operations to Create, Update, and Debug Your Training Job.

Contents

S3OutputPath

Path to Amazon S3 storage location for metrics and tensors.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: ^(https|s3)://([^/]+)/(.*)$

Required: Yes

CollectionConfigurations

Configuration information for Amazon SageMaker Debugger tensor collections. To learn more about how to configure the CollectionConfiguration parameter, see Use the SageMaker and Debugger Configuration API Operations to Create, Update, and Debug Your Training Job.

Type: Array of CollectionConfiguration (p. 1360) objects

Array Members: Minimum number of 0 items. Maximum number of 20 items.

Required: No

HookParameters

Configuration information for the Amazon SageMaker Debugger hook parameters.

Type: String to string map

Map Entries: Minimum number of 0 items. Maximum number of 20 items.

Key Length Constraints: Minimum length of 1. Maximum length of 256.

Key Pattern: .*

Value Length Constraints: Maximum length of 256.

Value Pattern: .*

Required: No

LocalPath

Path to local storage location for metrics and tensors. Defaults to /opt/ml/output/tensors/.

Type: String

Length Constraints: Maximum length of 4096.

Pattern: .*
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
DebugRuleConfiguration

Service: Amazon SageMaker Service

Configuration information for SageMaker Debugger rules for debugging. To learn more about how to configure the DebugRuleConfiguration parameter, see Use the SageMaker and Debugger Configuration API Operations to Create, Update, and Debug Your Training Job.

Contents

RuleConfigurationName

The name of the rule configuration. It must be unique relative to other rule configuration names.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 256.

Pattern: .*

Required: Yes

RuleEvaluatorImage

The Amazon Elastic Container (ECR) Image for the managed rule evaluation.

Type: String

Length Constraints: Maximum length of 255.

Pattern: .*

Required: Yes

InstanceType

The instance type to deploy a custom rule for debugging a training job.

Type: String

Valid Values: ml.t3.medium | ml.t3.large | ml.t3.xlarge | ml.t3.2xlarge | ml.m4.xlarge | ml.m4.2xlarge | ml.m4.4xlarge | ml.m4.10xlarge | ml.m4.16xlarge | ml.m4.2xlarge | ml.m4.4xlarge | ml.m4.8xlarge | ml.m4.16xlarge | ml.p2.xlarge | ml.p2.8xlarge | ml.p2.16xlarge | ml.p3.2xlarge | ml.p3.8xlarge | ml.p3.16xlarge | ml.c5.xlarge | ml.c5.2xlarge | ml.c5.4xlarge | ml.c5.9xlarge | ml.c5.18xlarge | ml.m5.large | ml.m5.xlarge | ml.m5.2xlarge | ml.m5.4xlarge | ml.m5.12xlarge | ml.m5.24xlarge | ml.r5.large | ml.r5.xlarge | ml.r5.2xlarge | ml.r5.4xlarge | ml.r5.8xlarge | ml.r5.12xlarge | ml.r5.16xlarge | ml.r5.24xlarge | ml.g4dn.xlarge | ml.g4dn.2xlarge | ml.g4dn.4xlarge | ml.g4dn.8xlarge | ml.g4dn.12xlarge | ml.g4dn.16xlarge

Required: No

LocalPath

Path to local storage location for output of rules. Defaults to /opt/ml/processing/output/rule/

Type: String

Length Constraints: Maximum length of 4096.
Pattern: . *
Required: No

**RuleParameters**

Runtime configuration for rule container.

Type: String to string map

Map Entries: Minimum number of 0 items. Maximum number of 100 items.

Key Length Constraints: Minimum length of 1. Maximum length of 256.

Key Pattern: . *

Value Length Constraints: Maximum length of 256.

Value Pattern: . *

Required: No

**S3OutputPath**

Path to Amazon S3 storage location for rules.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: ^((https|s3)://([^/]+)/*(.*))$

Required: No

**VolumeSizeInGB**

The size, in GB, of the ML storage volume attached to the processing instance.

Type: Integer

Valid Range: Minimum value of 0.

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
DebugRuleEvaluationStatus
Service: Amazon SageMaker Service

Information about the status of the rule evaluation.

Contents

LastModifiedTime
Timestamp when the rule evaluation status was last modified.
Type: Timestamp
Required: No

RuleConfigurationName
The name of the rule configuration.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 256.
Pattern: .*
Required: No

RuleEvaluationJobArn
The Amazon Resource Name (ARN) of the rule evaluation job.
Type: String
Length Constraints: Maximum length of 256.
Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:processing-job/.*
Required: No

RuleEvaluationStatus
Status of the rule evaluation.
Type: String
Valid Values: InProgress | NoIssuesFound | IssuesFound | Error | Stopping | Stopped
Required: No

StatusDetails
Details from the rule evaluation.
Type: String
Length Constraints: Maximum length of 1024.
Pattern: .*
Required: No
See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
DefaultEbsStorageSettings
Service: Amazon SageMaker Service
A collection of default EBS storage settings that applies to private spaces created within a domain or user profile.

Contents

DefaultEbsVolumeSizeInGb
The default size of the EBS storage volume for a private space.
Type: Integer
Required: Yes

MaximumEbsVolumeSizeInGb
The maximum size of the EBS storage volume for a private space.
Type: Integer
Required: Yes

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
DefaultSpaceSettings

Service: Amazon SageMaker Service

A collection of settings that apply to spaces created in the Domain.

Contents

ExecutionRole

The ARN of the execution role for the space.

Type: String


Pattern: ^arn:aws[a-z\-]*:iam::\d{12}:role/\?a-zA-Z_0-9+=,.@\-_/]+$

Required: No

JupyterServerAppSettings

The JupyterServer app settings.

Type: JupyterServerAppSettings (p. 1618) object

Required: No

KernelGatewayAppSettings

The KernelGateway app settings.

Type: KernelGatewayAppSettings (p. 1620) object

Required: No

SecurityGroups

The security group IDs for the Amazon Virtual Private Cloud that the space uses for communication.

Type: Array of strings

Array Members: Maximum number of 5 items.

Length Constraints: Maximum length of 32.

Pattern: [-0-9a-zA-Z]+

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
DefaultSpaceStorageSettings
Service: Amazon SageMaker Service

The default storage settings for a private space.

Contents

DefaultEbsStorageSettings

The default EBS storage settings for a private space.

Type: DefaultEbsStorageSettings (p. 1401) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
DeployedImage
Service: Amazon SageMaker Service

Gets the Amazon EC2 Container Registry path of the docker image of the model that is hosted in this ProductionVariant.

If you used the registry/repository[:tag] form to specify the image path of the primary container when you created the model hosted in this ProductionVariant, the path resolves to a path of the form registry/repository[@digest]. A digest is a hash value that identifies a specific version of an image. For information about Amazon ECR paths, see Pulling an Image in the Amazon ECR User Guide.

Contents

ResolutionTime
The date and time when the image path for the model resolved to the ResolvedImage
Type: Timestamp
Required: No

ResolvedImage
The specific digest path of the image hosted in this ProductionVariant.
Type: String
Length Constraints: Maximum length of 255.
Pattern: [\S]+
Required: No

SpecifiedImage
The image path you specified when you created the model.
Type: String
Length Constraints: Maximum length of 255.
Pattern: [\S]+
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
DeploymentConfig
Service: Amazon SageMaker Service

The deployment configuration for an endpoint, which contains the desired deployment strategy and rollback configurations.

Contents

AutoRollbackConfiguration

Automatic rollback configuration for handling endpoint deployment failures and recovery.

Type: AutoRollbackConfig (p. 1299) object

Required: No

BlueGreenUpdatePolicy

Update policy for a blue/green deployment. If this update policy is specified, SageMaker creates a new fleet during the deployment while maintaining the old fleet. SageMaker flips traffic to the new fleet according to the specified traffic routing configuration. Only one update policy should be used in the deployment configuration. If no update policy is specified, SageMaker uses a blue/green deployment strategy with all at once traffic shifting by default.

Type: BlueGreenUpdatePolicy (p. 1310) object

Required: No

RollingUpdatePolicy

Specifies a rolling deployment strategy for updating a SageMaker endpoint.

Type: RollingUpdatePolicy (p. 1919) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
DeploymentRecommendation

Service: Amazon SageMaker Service

A set of recommended deployment configurations for the model. To get more advanced recommendations, see [CreateInferenceRecommendationsJob](#) to create an inference recommendation job.

Contents

**RecommendationStatus**

Status of the deployment recommendation. The status NOT_APPLICABLE means that SageMaker is unable to provide a default recommendation for the model using the information provided. If the deployment status is IN_PROGRESS, retry your API call after a few seconds to get a COMPLETED deployment recommendation.

Type: String

Valid Values: IN_PROGRESS | COMPLETED | FAILED | NOT_APPLICABLE

Required: Yes

**RealTimeInferenceRecommendations**

A list of [RealTimeInferenceRecommendation](#) items.

Type: Array of [RealTimeInferenceRecommendation](#) objects

Array Members: Minimum number of 0 items. Maximum number of 3 items.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
DeploymentStage

Service: Amazon SageMaker Service

Contains information about a stage in an edge deployment plan.

Contents

DeviceSelectionConfig

Configuration of the devices in the stage.

Type: 

Required: Yes

StageName

The name of the stage.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])*\{0,62\}$

Required: Yes

DeploymentConfig

Configuration of the deployment details.

Type: 

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
DeploymentStageStatusSummary
Service: Amazon SageMaker Service
Contains information summarizing the deployment stage results.

Contents

DeploymentConfig
Configuration of the deployment details.
Type: EdgeDeploymentConfig (p. 1434) object
Required: Yes

DeploymentStatus
General status of the current state.
Type: EdgeDeploymentStatus (p. 1438) object
Required: Yes

DeviceSelectionConfig
Configuration of the devices in the stage.
Type: DeviceSelectionConfig (p. 1416) object
Required: Yes

StageName
The name of the stage.
Type: String
Pattern: ^[a-zA-Z0-9](\-[a-zA-Z0-9]{0,62})*$
Required: Yes

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
DerivedInformation
Service: Amazon SageMaker Service

Information that SageMaker Neo automatically derived about the model.

Contents

DerivedDataInputConfig

The data input configuration that SageMaker Neo automatically derived for the model. When SageMaker Neo derives this information, you don't need to specify the data input configuration when you create a compilation job.

Type: String


Pattern: \S\s+  

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
DesiredWeightAndCapacity

Service: Amazon SageMaker Service

Specifies weight and capacity values for a production variant.

Contents

**VariantName**

The name of the variant to update.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}

Required: Yes

**DesiredInstanceCount**

The variant's capacity.

Type: Integer

Valid Range: Minimum value of 0.

Required: No

**DesiredWeight**

The variant's weight.

Type: Float

Valid Range: Minimum value of 0.

Required: No

**ServerlessUpdateConfig**

Specifies the serverless update concurrency configuration for an endpoint variant.

Type: ProductionVariantServerlessUpdateConfig (p. 1856) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
Device
Service: Amazon SageMaker Service
Information of a particular device.

Contents

DeviceName
The name of the device.
Type: String
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$
Required: Yes

Description
Description of the device.
Type: String
Pattern: ^([-a-zA-Z0-9\_.,;!:\s]+)$
Required: No

IotThingName
AWS Internet of Things (IoT) object name.
Type: String
Length Constraints: Maximum length of 128.
Pattern: [a-zA-Z0-9:_.\-]+$
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
DeviceDeploymentSummary

Service: Amazon SageMaker Service

Contains information summarizing device details and deployment status.

Contents

DeviceArn

The ARN of the device.

Type: String


Pattern: ^arn:aws[a-z\-]*:[a-z\-]*:[a-z\-]*:\d{12}:\[a-z\-]*[/\-_\[\]]+$/

Required: Yes

DeviceName

The name of the device.

Type: String


Pattern: ^[a-zA-Z0-9\-\*\]([a-zA-Z0-9\-\]*){0,62}$

Required: Yes

EdgeDeploymentPlanArn

The ARN of the edge deployment plan.

Type: String


Pattern: ^arn:aws[a-z\-]*:sagemaker:[a-z\-]*:\d{12}:edge-deployment/\[a-zA-Z0-9\-\*\]([a-zA-Z0-9\-\*\]%){0,62}$

Required: Yes

EdgeDeploymentPlanName

The name of the edge deployment plan.

Type: String


Pattern: ^[a-zA-Z0-9\-\*\]([a-zA-Z0-9\-\*\]%){0,62}$

Required: Yes

StageName

The name of the stage in the edge deployment plan.

Type: String
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}$
Required: Yes

**DeployedStageName**
The name of the deployed stage.
Type: String
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}$
Required: No

**DeploymentStartTime**
The time when the deployment on the device started.
Type: Timestamp
Required: No

**Description**
The description of the device.
Type: String
Pattern: ^[-a-zA-Z0-9_.;!]*$
Required: No

**DeviceDeploymentStatus**
The deployment status of the device.
Type: String
Valid Values: READYTODEPLOY | INPROGRESS | DEPLOYED | FAILED | STOPPING | STOPPED
Required: No

**DeviceDeploymentStatusMessage**
The detailed error message for the deployment status result.
Type: String
Required: No

**DeviceFleetName**
The name of the fleet to which the device belongs to.
Type: String
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}$
See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
DeviceFleetSummary

Service: Amazon SageMaker Service

Summary of the device fleet.

Contents

DeviceFleetArn

Amazon Resource Name (ARN) of the device fleet.

Type: String

Pattern: ^arn:aws[a-z\-]\*:iam::\d{12}:device-fleet/?[a-zA-Z_0-9+=,.@\-_]+$

Required: Yes

DeviceFleetName

Name of the device fleet.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}$

Required: Yes

CreationTime

Timestamp of when the device fleet was created.

Type: Timestamp

Required: No

LastModifiedTime

Timestamp of when the device fleet was last updated.

Type: Timestamp

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
DeviceSelectionConfig

Service: Amazon SageMaker Service

Contains information about the configurations of selected devices.

Contents

DeviceSubsetType

Type of device subsets to deploy to the current stage.

Type: String

Valid Values: PERCENTAGE | SELECTION | NAMECONTAINS

Required: Yes

DeviceNameContains

A filter to select devices with names containing this name.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}\$

Required: No

DeviceNames

List of devices chosen to deploy.

Type: Array of strings


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}\$

Required: No

Percentage

Percentage of devices in the fleet to deploy to the current stage.

Type: Integer

Valid Range: Maximum value of 100.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
DeviceStats
Service: Amazon SageMaker Service
Status of devices.

Contents

*ConnectedDeviceCount*

The number of devices connected with a heartbeat.

Type: Long
Required: Yes

*RegisteredDeviceCount*

The number of registered devices.

Type: Long
Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
DeviceSummary
Service: Amazon SageMaker Service
Summary of the device.

Contents

DeviceArn
Amazon Resource Name (ARN) of the device.
Type: String
Pattern: ^arn:aws[a-z\-]*:[a-z\-]*:[a-z\-]*:\d{12}:[a-z\-]*+/?\[a-zA-Z_0-9+=,.@\-_]/+\$
Required: Yes

DeviceName
The unique identifier of the device.
Type: String
Pattern: ^[a-zA-Z0-9\-_\(\)\[\]\{\}\_\\:\:\]+\$
Required: Yes

AgentVersion
Edge Manager agent version.
Type: String
Pattern: [a-zA-Z0-9\-_\\.\]+$
Required: No

Description
A description of the device.
Type: String
Pattern: ^[-a-zA-Z0-9._,;::! \]+$
Required: No

DeviceFleetName
The name of the fleet the device belongs to.
Type: String
DeviceSummary

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$
Required: No

IotThingName
The AWS Internet of Things (IoT) object thing name associated with the device.
Type: String
Length Constraints: Maximum length of 128.
Pattern: [a-zA-Z0-9:_-]+
Required: No

LatestHeartbeat
The last heartbeat received from the device.
Type: Timestamp
Required: No

Models
Models on the device.
Type: Array of EdgeModelSummary (p. 1443) objects
Required: No

RegistrationTime
The timestamp of the last registration or de-reregistration.
Type: Timestamp
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
DirectDeploySettings

Service: Amazon SageMaker Service

The model deployment settings for the SageMaker Canvas application.

**Note**
In order to enable model deployment for Canvas, the SageMaker Domain's or user profile's AWS IAM execution role must have the AmazonSageMakerCanvasDirectDeployAccess policy attached. You can also turn on model deployment permissions through the SageMaker Domain's or user profile's settings in the SageMaker console.

**Contents**

**Status**

Describes whether model deployment permissions are enabled or disabled in the Canvas application.

- Type: String
- Valid Values: ENABLED | DISABLED
- Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
DomainDetails

Service: Amazon SageMaker Service

The domain's details.

Contents

**CreationTime**

The creation time.

Type: Timestamp

Required: No

**DomainArn**

The domain's Amazon Resource Name (ARN).

Type: String

Length Constraints: Maximum length of 256.

Pattern: `arn:aws[a-zA-Z-]*:sagemaker:[a-zA-Z0-9-]*:[0-9]{12}:domain/.*`

Required: No

**DomainId**

The domain ID.

Type: String

Length Constraints: Maximum length of 63.

Required: No

**DomainName**

The domain name.

Type: String

Length Constraints: Maximum length of 63.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}`

Required: No

**LastModifiedTime**

The last modified time.

Type: Timestamp

Required: No

**Status**

The status.

Type: String

Valid Values: Deleting | Failed | InService | Pending | Updating | Update_Failed | Delete_Failed
DomainDetails

Required: No

Url

The domain's URL.

Type: String

Length Constraints: Maximum length of 1024.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
DomainSettings
Service: Amazon SageMaker Service

A collection of settings that apply to the SageMaker Domain. These settings are specified through the CreateDomain API call.

Contents

ExecutionRoleIdentityConfig

The configuration for attaching a SageMaker user profile name to the execution role as a sts:SourceIdentity key.

Type: String

Valid Values: USER_PROFILE_NAME | DISABLED

Required: No

RStudioServerProDomainSettings

A collection of settings that configure the RStudioServerPro Domain-level app.

Type: RStudioServerProDomainSettings (p. 1922) object

Required: No

SecurityGroupIds

The security groups for the Amazon Virtual Private Cloud that the Domain uses for communication between Domain-level apps and user apps.

Type: Array of strings

Array Members: Maximum number of 3 items.

Length Constraints: Maximum length of 32.

Pattern: [-0-9a-zA-Z]+

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
DomainSettingsForUpdate
Service: Amazon SageMaker Service

A collection of Domain configuration settings to update.

Contents

ExecutionRoleIdentityConfig

The configuration for attaching a SageMaker user profile name to the execution role as a sts:SourceIdentity key. This configuration can only be modified if there are no apps in the InService or Pending state.

Type: String
Valid Values: USER_PROFILE_NAME | DISABLED
Required: No

RStudioServerProDomainSettingsForUpdate

A collection of RStudioServerPro Domain-level app settings to update. A single RStudioServerPro application is created for a domain.

Type: RStudioServerProDomainSettingsForUpdate (p. 1923) object
Required: No

SecurityGroupIds

The security groups for the Amazon Virtual Private Cloud that the Domain uses for communication between Domain-level apps and user apps.

Type: Array of strings
Array Members: Maximum number of 3 items.
Length Constraints: Maximum length of 32.
Pattern: [-0-9a-zA-Z]+
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
DriftCheckBaselines

Service: Amazon SageMaker Service

Represents the drift check baselines that can be used when the model monitor is set using the model package.

Contents

Bias

Represents the drift check bias baselines that can be used when the model monitor is set using the model package.

Type: DriftCheckBias (p. 1427) object

Required: No

Explainability

Represents the drift check explainability baselines that can be used when the model monitor is set using the model package.

Type: DriftCheckExplainability (p. 1428) object

Required: No

ModelDataQuality

Represents the drift check model data quality baselines that can be used when the model monitor is set using the model package.

Type: DriftCheckModelDataQuality (p. 1429) object

Required: No

ModelQuality

Represents the drift check model quality baselines that can be used when the model monitor is set using the model package.

Type: DriftCheckModelQuality (p. 1430) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
DriftCheckBias
Service: Amazon SageMaker Service

Represents the drift check bias baselines that can be used when the model monitor is set using the model package.

Contents

ConfigFile

The bias config file for a model.
Type: FileSource (p. 1492) object
Required: No

PostTrainingConstraints

The post-training constraints.
Type: MetricsSource (p. 1655) object
Required: No

PreTrainingConstraints

The pre-training constraints.
Type: MetricsSource (p. 1655) object
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
DriftCheckExplainability
Service: Amazon SageMaker Service

Represents the drift check explainability baselines that can be used when the model monitor is set using the model package.

Contents

ConfigFile

The explainability config file for the model.

Type: FileSource (p. 1492) object

Required: No

Constraints

The drift check explainability constraints.

Type: MetricsSource (p. 1655) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
DriftCheckModelDataQuality
Service: Amazon SageMaker Service

Represents the drift check data quality baselines that can be used when the model monitor is set using the model package.

Contents

Constraints

The drift check model data quality constraints.
Type: MetricsSource (p. 1655) object
Required: No

Statistics

The drift check model data quality statistics.
Type: MetricsSource (p. 1655) object
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
DriftCheckModelQuality

Service: Amazon SageMaker Service

Represents the drift check model quality baselines that can be used when the model monitor is set using the model package.

Contents

Constraints

The drift check model quality constraints.

Type: MetricsSource (p. 1655) object

Required: No

Statistics

The drift check model quality statistics.

Type: MetricsSource (p. 1655) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
DynamicScalingConfiguration

Service: Amazon SageMaker Service

An object with the recommended values for you to specify when creating an autoscaling policy.

Contents

MaxCapacity

The recommended maximum capacity to specify for your autoscaling policy.

Type: Integer

Required: No

MinCapacity

The recommended minimum capacity to specify for your autoscaling policy.

Type: Integer

Required: No

ScaleInCooldown

The recommended scale in cooldown time for your autoscaling policy.

Type: Integer

Required: No

ScaleOutCooldown

The recommended scale out cooldown time for your autoscaling policy.

Type: Integer

Required: No

ScalingPolicies

An object of the scaling policies for each metric.

Type: Array of `ScalingPolicy` objects

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
EbsStorageSettings
Service: Amazon SageMaker Service
A collection of EBS storage settings that applies to private spaces.

Contents

EbsVolumeSizeInGb
The size of an EBS storage volume for a private space.
Type: Integer
Required: Yes

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
**Edge**

Service: Amazon SageMaker Service

A directed edge connecting two lineage entities.

**Contents**

**AssociationType**

The type of the Association(Edge) between the source and destination. For example ContributedTo, Produced, or DerivedFrom.

Type: String

Valid Values: ContributedTo | AssociatedWith | DerivedFrom | Produced

Required: No

**DestinationArn**

The Amazon Resource Name (ARN) of the destination lineage entity of the directed edge.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:(experiment|experiment-trial-component|artifact|action|context)/.*

Required: No

**SourceArn**

The Amazon Resource Name (ARN) of the source lineage entity of the directed edge.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:(experiment|experiment-trial-component|artifact|action|context)/.*

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
EdgeDeploymentConfig
Service: Amazon SageMaker Service

Contains information about the configuration of a deployment.

Contents

FailureHandlingPolicy

Toggle that determines whether to rollback to previous configuration if the current deployment fails. By default this is turned on. You may turn this off if you want to investigate the errors yourself.

Type: String

Valid Values: ROLLBACK_ON_FAILURE | DO NOTHING

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
EdgeDeploymentModelConfig
Service: Amazon SageMaker Service
Contains information about the configuration of a model in a deployment.

Contents

EdgePackagingJobName
The edge packaging job associated with this deployment.
Type: String
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$
Required: Yes

ModelHandle
The name the device application uses to reference this model.
Type: String
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$
Required: Yes

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
**EdgeDeploymentPlanSummary**

Service: Amazon SageMaker Service

Contains information summarizing an edge deployment plan.

**Contents**

**DeviceFleetName**

The name of the device fleet used for the deployment.

Type: String


Pattern: `^[a-zA-Z0-9][-][a-zA-Z0-9]{0,62}$`

Required: Yes

**EdgeDeploymentFailed**

The number of edge devices that failed the deployment.

Type: Integer

Required: Yes

**EdgeDeploymentPending**

The number of edge devices yet to pick up the deployment, or in progress.

Type: Integer

Required: Yes

**EdgeDeploymentPlanArn**

The ARN of the edge deployment plan.

Type: String


Pattern: `^arn:aws[a-zA-Z]+:sagemaker::\d{12}:edge-deployment/?[a-zA-Z0-9=,.@/-_]+$`

Required: Yes

**EdgeDeploymentPlanName**

The name of the edge deployment plan.

Type: String


Pattern: `^[a-zA-Z0-9][-][a-zA-Z0-9]{0,62}$`

Required: Yes

**EdgeDeploymentSuccess**

The number of edge devices with the successful deployment.
Type: Integer
Required: Yes

**CreationTime**

The time when the edge deployment plan was created.

Type: Timestamp
Required: No

**LastModifiedTime**

The time when the edge deployment plan was last updated.

Type: Timestamp
Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
**EdgeDeploymentStatus**

Service: Amazon SageMaker Service

Contains information summarizing the deployment stage results.

**Contents**

**EdgeDeploymentFailedInStage**

- The number of edge devices that failed the deployment in current stage.
- Type: Integer
- Required: Yes

**EdgeDeploymentPendingInStage**

- The number of edge devices yet to pick up the deployment in current stage, or in progress.
- Type: Integer
- Required: Yes

**EdgeDeploymentSuccessInStage**

- The number of edge devices with the successful deployment in the current stage.
- Type: Integer
- Required: Yes

**StageStatus**

- The general status of the current stage.
- Type: String
- Valid Values: CREATING | READYTODEPLOY | STARTING | INPROGRESS | DEPLOYED | FAILED | STOPPING | STOPPED
- Required: Yes

**EdgeDeploymentStageStartTime**

- The time when the deployment API started.
- Type: Timestamp
- Required: No

**EdgeDeploymentStatusMessage**

- A detailed message about deployment status in current stage.
- Type: String
- Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](https://aws.amazon.com/sdk-for-cpp/)

1438
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for Ruby V3
EdgeModel
Service: Amazon SageMaker Service

The model on the edge device.

Contents

ModelName
The name of the model.
Type: String
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}\$
Required: Yes

ModelVersion
The model version.
Type: String
Pattern: [a-zA-Z0-9\ \_\.]+
Required: Yes

LatestInference
The timestamp of the last inference that was made.
Type: Timestamp
Required: No

LatestSampleTime
The timestamp of the last data sample taken.
Type: Timestamp
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
EdgeModelStat
Service: Amazon SageMaker Service

Status of edge devices with this model.

Contents

**ActiveDeviceCount**

The number of devices that have this model version, a heart beat, and are currently running.

Type: Long

Required: Yes

**ConnectedDeviceCount**

The number of devices that have this model version and have a heart beat.

Type: Long

Required: Yes

**ModelName**

The name of the model.

Type: String


Pattern: ^[a-zA-Z0-9\-_\\ ]{0,62}$

Required: Yes

**ModelVersion**

The model version.

Type: String


Pattern: [a-zA-Z0-9\-_\\ ]+

Required: Yes

**OfflineDeviceCount**

The number of devices that have this model version and do not have a heart beat.

Type: Long

Required: Yes

**SamplingDeviceCount**

The number of devices with this model version and are producing sample data.

Type: Long

Required: Yes
See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
EdgeModelSummary
Service: Amazon SageMaker Service
Summary of model on edge device.

Contents

ModelName
The name of the model.
Type: String
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$
Required: Yes

ModelVersion
The version model.
Type: String
Pattern: [a-zA-Z0-9\ _\.]+
Required: Yes

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
EdgeOutputConfig

Service: Amazon SageMaker Service

The output configuration.

Contents

S3OutputLocation

The Amazon Simple Storage (S3) bucket URI.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: ^(https|s3)://([^/]+)/(.*)$

Required: Yes

KmsKeyId

The AWS Key Management Service (AWS KMS) key that Amazon SageMaker uses to encrypt data on the storage volume after compilation job. If you don't provide a KMS key ID, Amazon SageMaker uses the default KMS key for Amazon S3 for your role's account.

Type: String

Length Constraints: Maximum length of 2048.

Pattern: .*

Required: No

PresetDeploymentConfig

The configuration used to create deployment artifacts. Specify configuration options with a JSON string. The available configuration options for each type are:

- ComponentName (optional) - Name of the GreenGrass V2 component. If not specified, the default name generated consists of “SagemakerEdgeManager” and the name of your SageMaker Edge Manager packaging job.
- ComponentDescription (optional) - Description of the component.
- ComponentVersion (optional) - The version of the component.

Note

AWS IoT Greengrass uses semantic versions for components. Semantic versions follow a major.minor.patch number system. For example, version 1.0.0 represents the first major release for a component. For more information, see the semantic version specification.

- PlatformOS (optional) - The name of the operating system for the platform. Supported platforms include Windows and Linux.
- PlatformArchitecture (optional) - The processor architecture for the platform.

Supported architectures Windows include: Windows32_x86, Windows64_x64.

Supported architectures for Linux include: Linux x86_64, Linux ARMV8.

Type: String

Required: No
PresetDeploymentType

The deployment type SageMaker Edge Manager will create. Currently only supports AWS IoT Greengrass Version 2 components.

Type: String

Valid Values: GreengrassV2Component

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
EdgePackagingJobSummary
Service: Amazon SageMaker Service
Summary of edge packaging job.

Contents

EdgePackagingJobArn
The Amazon Resource Name (ARN) of the edge packaging job.
Type: String
Pattern: ^arn:aws[a-z\-]*:sagemaker:[a-z\-]*:d\d*:edge-packaging-job/?[a-zA-Z_0-9+=,.@-_\/]+$
Required: Yes

EdgePackagingJobName
The name of the edge packaging job.
Type: String
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$
Required: Yes

EdgePackagingJobStatus
The status of the edge packaging job.
Type: String
Valid Values: STARTING | INPROGRESS | COMPLETED | FAILED | STOPPING | STOPPED
Required: Yes

CompilationJobName
The name of the SageMaker Neo compilation job.
Type: String
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$
Required: No

CreationTime
The timestamp of when the job was created.
Type: Timestamp
Required: No

LastModifiedTime
The timestamp of when the edge packaging job was last updated.
EdgePackagingJobSummary

Type: Timestamp
Required: No

**ModelName**
The name of the model.
Type: String
Pattern: ^[a-zA-Z0-9][-]*[a-zA-Z0-9]{0,62}$
Required: No

**ModelVersion**
The version of the model.
Type: String
Pattern: [a-zA-Z0-9\ \._\-]+
Required: No

**See Also**
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
**EdgePresetDeploymentOutput**

Service: Amazon SageMaker Service

The output of a SageMaker Edge Manager deployable resource.

**Contents**

**Type**

The deployment type created by SageMaker Edge Manager. Currently only supports AWS IoT Greengrass Version 2 components.

Type: String

Valid Values: GreengrassV2Component

Required: Yes

**Artifact**

The Amazon Resource Name (ARN) of the generated deployable resource.

Type: String


Required: No

**Status**

The status of the deployable resource.

Type: String

Valid Values: COMPLETED | FAILED

Required: No

**StatusMessage**

Returns a message describing the status of the deployed resource.

Type: String

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
**EFSFileSystem**

**Service: Amazon SageMaker Service**

A file system, created by you in Amazon EFS, that you assign to a user profile or space for an Amazon SageMaker Domain. Permitted users can access this file system in Amazon SageMaker Studio.

**Contents**

**FileSystemId**

The ID of your Amazon EFS file system.

Type: String

Length Constraints: Minimum length of 11.

Pattern: `^(fs-[0-9a-f]{8,})$

Required: Yes

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
EFSFileSystemConfig

Service: Amazon SageMaker Service

The settings for assigning a custom Amazon EFS file system to a user profile or space for an Amazon SageMaker Domain.

Contents

FileSystemId

The ID of your Amazon EFS file system.

Type: String

Length Constraints: Minimum length of 11.

Pattern: ^(fs-[0-9a-f]{8,})$

Required: Yes

FileSystemPath

The path to the file system directory that is accessible in Amazon SageMaker Studio. Permitted users can access only this directory and below.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 256.

Pattern: ^\S*$

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
EMRStepMetadata

Service: Amazon SageMaker Service

The configurations and outcomes of an Amazon EMR step execution.

Contents

ClusterId

The identifier of the EMR cluster.

Type: String

Length Constraints: Maximum length of 256.

Required: No

LogFilePath

The path to the log file where the cluster step's failure root cause is recorded.

Type: String

Length Constraints: Maximum length of 1024.

Required: No

StepId

The identifier of the EMR cluster step.

Type: String

Length Constraints: Maximum length of 256.

Required: No

StepName

The name of the EMR cluster step.

Type: String

Length Constraints: Maximum length of 256.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
Endpoint

Service: Amazon SageMaker Service

A hosted endpoint for real-time inference.

Contents

CreationTime

The time that the endpoint was created.

Type: Timestamp

Required: Yes

EndpointArn

The Amazon Resource Name (ARN) of the endpoint.

Type: String


Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:endpoint/.*

Required: Yes

EndpointConfigName

The endpoint configuration associated with the endpoint.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}

Required: Yes

EndpointName

The name of the endpoint.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}

Required: Yes

EndpointStatus

The status of the endpoint.

Type: String

Valid Values: OutOfService | Creating | Updating | SystemUpdating | RollingBack | InService | Deleting | Failed | UpdateRollbackFailed

Required: Yes

LastModifiedTime

The last time the endpoint was modified.
Type: Timestamp
Required: Yes

**DataCaptureConfig**

The currently active data capture configuration used by your Endpoint.

Type: [DataCaptureConfigSummary](p. 1383) object

Required: No

**FailureReason**

If the endpoint failed, the reason it failed.

Type: String

Length Constraints: Maximum length of 1024.

Required: No

**MonitoringSchedules**

A list of monitoring schedules for the endpoint. For information about model monitoring, see [Amazon SageMaker Model Monitor](#).

Type: Array of [MonitoringSchedule](p. 1758) objects

Required: No

**ProductionVariants**

A list of the production variants hosted on the endpoint. Each production variant is a model.

Type: Array of [ProductionVariantSummary](p. 1858) objects

Array Members: Minimum number of 1 item.

Required: No

**ShadowProductionVariants**

A list of the shadow variants hosted on the endpoint. Each shadow variant is a model in shadow mode with production traffic replicated from the production variant.

Type: Array of [ProductionVariantSummary](p. 1858) objects

Array Members: Minimum number of 1 item.

Required: No

**Tags**

A list of the tags associated with the endpoint. For more information, see [Tagging AWS resources](#) in the [AWS General Reference Guide](#).

Type: Array of [Tag](p. 1979) objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
EndpointConfigSummary
Service: Amazon SageMaker Service
Provides summary information for an endpoint configuration.

Contents

CreationTime
A timestamp that shows when the endpoint configuration was created.
Type: Timestamp
Required: Yes

EndpointConfigArn
The Amazon Resource Name (ARN) of the endpoint configuration.
Type: String
Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:endpoint-config/.*
Required: Yes

EndpointConfigName
The name of the endpoint configuration.
Type: String
Length Constraints: Maximum length of 63.
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}
Required: Yes

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
EndpointInfo
Service: Amazon SageMaker Service

Details about a customer endpoint that was compared in an Inference Recommender job.

Contents

EndpointName

The name of a customer's endpoint.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9]-*[^a-zA-Z0-9]{0,62}

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
EndpointInput
Service: Amazon SageMaker Service
Input object for the endpoint

Contents

EndpointName
An endpoint in customer’s account which has enabled DataCaptureConfig enabled.
Type: String
Length Constraints: Maximum length of 63.
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$
Required: Yes

LocalPath
Path to the filesystem where the endpoint data is available to the container.
Type: String
Length Constraints: Maximum length of 256.
Pattern: .*
Required: Yes

EndTimeOffset
If specified, monitoring jobs substract this time from the end time. For information about using offsets for scheduling monitoring jobs, see Schedule Model Quality Monitoring Jobs.
Type: String
Pattern: ^\d+P.*
Required: No

ExcludeFeaturesAttribute
The attributes of the input data to exclude from the analysis.
Type: String
Length Constraints: Maximum length of 100.
Required: No

FeaturesAttribute
The attributes of the input data that are the input features.
Type: String
Required: No

InferenceAttribute
The attribute of the input data that represents the ground truth label.
Type: String  
Required: No

**ProbabilityAttribute**

In a classification problem, the attribute that represents the class probability.

Type: String  
Required: No

**ProbabilityThresholdAttribute**

The threshold for the class probability to be evaluated as a positive result.

Type: Double  
Required: No

**S3DataDistributionType**

Whether input data distributed in Amazon S3 is fully replicated or sharded by an Amazon S3 key. Defaults to FullyReplicated

Type: String  
Valid Values: FullyReplicated | ShardedByS3Key  
Required: No

**S3InputMode**

Whether the Pipe or File is used as the input mode for transferring data for the monitoring job. Pipe mode is recommended for large datasets. File mode is useful for small files that fit in memory. Defaults to File.

Type: String  
Valid Values: Pipe | File  
Required: No

**StartTimeOffset**

If specified, monitoring jobs subtract this time from the start time. For information about using offsets for scheduling monitoring jobs, see [Schedule Model Quality Monitoring Jobs](#).

Type: String  
Pattern: ^\.?P\.*  
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- AWS SDK for Ruby V3
EndpointInputConfiguration

Service: Amazon SageMaker Service

The endpoint configuration for the load test.

Contents

EnvironmentParameterRanges

The parameter you want to benchmark against.

Type: EnvironmentParameterRanges (p. 1470) object

Required: No

InferenceSpecificationName

The inference specification name in the model package version.

Type: String


Pattern: ^[a-zA-Z0-9-]*[a-zA-Z0-9]${0,62}$

Required: No

InstanceType

The instance types to use for the load test.

Type: String

Valid Values: ml.t2.medium | ml.t2.large | ml.t2.xlarge | ml.t2.2xlarge | ml.m4.xlarge | ml.m4.2xlarge | ml.m4.4xlarge | ml.m4.10xlarge | ml.m4.16xlarge | ml.m5.large | ml.m5.xlarge | ml.m5.2xlarge | ml.m5.4xlarge | ml.m5.12xlarge | ml.m5.24xlarge | ml.m5d.large | ml.m5d.xlarge | ml.m5d.2xlarge | ml.m5d.4xlarge | ml.m5d.12xlarge | ml.m5d.24xlarge | ml.c4.large | ml.c4.xlarge | ml.c4.2xlarge | ml.c4.8xlarge | ml.p2.xlarge | ml.p2.8xlarge | ml.p2.16xlarge | ml.p3.2xlarge | ml.p3.16xlarge | ml.p3.8xlarge | ml.c5.large | ml.c5.xlarge | ml.c5.2xlarge | ml.c5.4xlarge | ml.c5.9xlarge | ml.c5.18xlarge | ml.c5d.large | ml.c5d.xlarge | ml.c5d.2xlarge | ml.c5d.4xlarge | ml.c5d.9xlarge | ml.c5d.18xlarge | ml.g4dn.xlarge | ml.g4dn.2xlarge | ml.g4dn.4xlarge | ml.g4dn.8xlarge | ml.g4dn.12xlarge | ml.g4dn.16xlarge | ml.r5.large | ml.r5.xlarge | ml.r5.2xlarge | ml.r5.4xlarge | ml.r5.12xlarge | ml.r5.24xlarge | ml.r5d.large | ml.r5d.xlarge | ml.r5d.2xlarge | ml.r5d.4xlarge | ml.r5d.12xlarge | ml.r5d.24xlarge | ml.inf1.xlarge | ml.inf1.2xlarge | ml.inf1.6xlarge | ml.inf1.4xlarge | ml.c6i.large | ml.c6i.xlarge | ml.c6i.2xlarge | ml.c6i.4xlarge | ml.c6i.8xlarge | ml.c6i.12xlarge | ml.c6i.16xlarge | ml.c6i.24xlarge | ml.c6i.32xlarge | ml.g5.large | ml.g5.xlarge | ml.g5.2xlarge | ml.g5.4xlarge | ml.g5.8xlarge | ml.g5.12xlarge | ml.g5.16xlarge | ml.g5.24xlarge | ml.g5.48xlarge | ml.p4d.xlarge | ml.c7g.large | ml.c7g.xlarge | ml.c7g.2xlarge | ml.c7g.4xlarge | ml.c7g.8xlarge | ml.c7g.12xlarge | ml.c7g.16xlarge | ml.m6g.large | ml.m6g.xlarge | ml.m6g.2xlarge | ml.m6g.4xlarge | ml.m6g.8xlarge | ml.m6g.12xlarge | ml.m6g.16xlarge | ml.m6gd.large | ml.m6gd.xlarge | ml.m6gd.2xlarge | ml.m6gd.4xlarge | ml.m6gd.8xlarge | ml.m6gd.12xlarge | ml.m6gd.16xlarge | ml.c6g.large | ml.c6g.xlarge

1460
| ml.c6g.2xlarge | ml.c6g.4xlarge | ml.c6g.8xlarge | ml.c6g.12xlarge |
| ml.c6g.16xlarge | ml.c6gd.large | ml.c6gd.xlarge | ml.c6gd.2xlarge |
| ml.c6gd.4xlarge | ml.c6gd.8xlarge | ml.c6gd.12xlarge | ml.c6gd.16xlarge |
| ml.c6gn.large | ml.c6gn.xlarge | ml.c6gn.2xlarge | ml.c6gn.4xlarge |
| ml.c6gn.8xlarge | ml.c6gn.12xlarge | ml.c6gn.16xlarge | ml.r6g.large |
| ml.r6g.xlarge | ml.r6g.2xlarge | ml.r6g.4xlarge | ml.r6g.8xlarge |
| ml.r6g.12xlarge | ml.r6g.16xlarge | ml.r6gd.large | ml.r6gd.xlarge |
| ml.r6gd.2xlarge | ml.r6gd.4xlarge | ml.r6gd.8xlarge | ml.r6gd.12xlarge |
| ml.r6gd.16xlarge | ml.p4de.24xlarge | ml.trn1.2xlarge | ml.trn1.32xlarge |
| ml.inf2.xlarge | ml.inf2.8xlarge | ml.inf2.24xlarge | ml.inf2.48xlarge |
| ml.p5.48xlarge |

Required: No

**ServerlessConfig**

Specifies the serverless configuration for an endpoint variant.

Type: `ProductionVariantServerlessConfig (p. 1855)` object

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
EndpointMetadata
Service: Amazon SageMaker Service

The metadata of the endpoint.

Contents

EndpointName

The name of the endpoint.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62\}

Required: Yes

EndpointConfigName

The name of the endpoint configuration.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62\}

Required: No

EndpointStatus

The status of the endpoint. For possible values of the status of an endpoint, see EndpointSummary.

Type: String

Valid Values: OutOfService | Creating | Updating | SystemUpdating | RollingBack | InService | Deleting | Failed | UpdateRollbackFailed

Required: No

FailureReason

If the status of the endpoint is Failed, or the status is InService but update operation fails, this provides the reason why it failed.

Type: String

Length Constraints: Maximum length of 1024.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
• AWS SDK for Ruby V3
**EndpointOutputConfiguration**

Service: Amazon SageMaker Service

The endpoint configuration made by Inference Recommender during a recommendation job.

**Contents**

**EndpointName**

The name of the endpoint made during a recommendation job.

Type: String

Required: Yes

**VariantName**

The name of the production variant (deployed model) made during a recommendation job.

Type: String

Required: Yes

**InitialInstanceCount**

The number of instances recommended to launch initially.

Type: Integer

Valid Range: Minimum value of 1.

Required: No

**InstanceType**

The instance type recommended by Amazon SageMaker Inference Recommender.

Type: String

Valid Values: ml.t2.medium | ml.t2.large | ml.t2.xlarge | ml.t2.2xlarge | ml.m4.xlarge | ml.m4.2xlarge | ml.m4.4xlarge | ml.m4.10xlarge | ml.m4.16xlarge | ml.m5.large | ml.m5.xlarge | ml.m5.2xlarge | ml.m5.4xlarge | ml.m5.12xlarge | ml.m5.24xlarge | ml.m5d.large | ml.m5d.xlarge | ml.m5d.2xlarge | ml.m5d.4xlarge | ml.m5d.12xlarge | ml.m5d.24xlarge | ml.c4.large | ml.c4.xlarge | ml.c4.2xlarge | ml.c4.4xlarge | ml.c4.8xlarge | ml.p2.xlarge | ml.p2.8xlarge | ml.p2.16xlarge | ml.p3.2xlarge | ml.p3.8xlarge | ml.p3.16xlarge | ml.c5.large | ml.c5.xlarge | ml.c5.2xlarge | ml.c5.4xlarge | ml.c5.9xlarge | ml.c5.18xlarge | ml.c5d.large | ml.c5d.xlarge | ml.c5d.2xlarge | ml.c5d.4xlarge | ml.c5d.9xlarge | ml.c5d.18xlarge | ml.g4dn.xlarge | ml.g4dn.2xlarge | ml.g4dn.4xlarge | ml.g4dn.8xlarge | ml.g4dn.12xlarge | ml.g4dn.16xlarge | ml.r5.large | ml.r5.xlarge | ml.r5.2xlarge | ml.r5.4xlarge | ml.r5.12xlarge | ml.r5.24xlarge | ml.r5d.large | ml.r5d.xlarge | ml.r5d.2xlarge | ml.r5d.4xlarge | ml.r5d.12xlarge | ml.r5d.24xlarge | ml.r5d.48xlarge | ml.inf1.xlarge | ml.inf1.2xlarge | ml.inf1.6xlarge | ml.inf1.24xlarge | ml.inf1.xlarge | ml.c6i.large | ml.c6i.xlarge | ml.c6i.2xlarge | ml.c6i.4xlarge | ml.c6i.8xlarge | ml.c6i.12xlarge | ml.c6i.16xlarge | ml.c6i.24xlarge | ml.c6i.48xlarge | ml.g5.large | ml.g5.xlarge | ml.g5.2xlarge | ml.g5.4xlarge | ml.g5.8xlarge | ml.g5.12xlarge | ml.g5.16xlarge | ml.g5.24xlarge | ml.g5.48xlarge | ml.p4d.24xlarge | ml.c7g.large | ml.c7g.xlarge | ml.c7g.2xlarge |
ServerlessConfig

Specifies the serverless configuration for an endpoint variant.

Type: ProductionVariantServerlessConfig (p. 1855) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
EndpointPerformance
Service: Amazon SageMaker Service

The performance results from running an Inference Recommender job on an existing endpoint.

Contents

EndpointInfo

Details about a customer endpoint that was compared in an Inference Recommender job.

Type: EndpointInfo (p. 1456) object

Required: Yes

Metrics

The metrics for an existing endpoint.

Type: InferenceMetrics (p. 1594) object

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
EndpointSummary

Service: Amazon SageMaker Service

Provides summary information for an endpoint.

Contents

CreationTime

A timestamp that shows when the endpoint was created.

Type: Timestamp

Required: Yes

EndpointArn

The Amazon Resource Name (ARN) of the endpoint.

Type: String


Pattern: `arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:endpoint/.*`

Required: Yes

EndpointName

The name of the endpoint.

Type: String

Length Constraints: Maximum length of 63.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}`

Required: Yes

EndpointStatus

The status of the endpoint.

- OutOfService: Endpoint is not available to take incoming requests.
- Creating: CreateEndpoint is executing.
- Updating: UpdateEndpoint or UpdateEndpointWeightsAndCapacities is executing.
- SystemUpdating: Endpoint is undergoing maintenance and cannot be updated or deleted or re-scaled until it has completed. This maintenance operation does not change any customer-specified values such as VPC config, KMS encryption, model, instance type, or instance count.
- RollingBack: Endpoint fails to scale up or down or change its variant weight and is in the process of rolling back to its previous configuration. Once the rollback completes, endpoint returns to an InService status. This transitional status only applies to an endpoint that has autoscaling enabled and is undergoing variant weight or capacity changes as part of an UpdateEndpointWeightsAndCapacities call or when the UpdateEndpointWeightsAndCapacities operation is called explicitly.
- InService: Endpoint is available to process incoming requests.
- Deleting: DeleteEndpoint is executing.
- Failed: Endpoint could not be created, updated, or re-scaled. Use DescribeEndpointOutput $FailureReason for information about the failure. DeleteEndpoint is the only operation that can be performed on a failed endpoint.
To get a list of endpoints with a specified status, use the `StatusEquals` filter with a call to `ListEndpoints`.

**Type:** String

**Valid Values:** OutOfService | Creating | Updating | SystemUpdating | RollingBack | InService | Deleting | Failed | UpdateRollbackFailed

**Required:** Yes

**LastModifiedTime**

A timestamp that shows when the endpoint was last modified.

**Type:** Timestamp

**Required:** Yes

### See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++]
- [AWS SDK for Go]
- [AWS SDK for Java V2]
- [AWS SDK for Ruby V3]
EnvironmentParameter

Service: Amazon SageMaker Service

A list of environment parameters suggested by the Amazon SageMaker Inference Recommender.

Contents

Key

The environment key suggested by the Amazon SageMaker Inference Recommender.

Type: String

Required: Yes

Value

The value suggested by the Amazon SageMaker Inference Recommender.

Type: String

Required: Yes

ValueType

The value type suggested by the Amazon SageMaker Inference Recommender.

Type: String

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
EnvironmentParameterRanges
Service: Amazon SageMaker Service
Specifies the range of environment parameters

Contents

CategoricalParameterRanges
Specified a list of parameters for each category.
Type: Array of CategoricalParameter (p. 1321) objects
Array Members: Minimum number of 1 item. Maximum number of 5 items.
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
Experiment
Service: Amazon SageMaker Service

The properties of an experiment as returned by the `Search` API.

Contents

**CreatedBy**
Who created the experiment.
Type: `UserContext (p. 2067)` object
Required: No

**CreationTime**
When the experiment was created.
Type: `Timestamp`
Required: No

**Description**
The description of the experiment.
Type: `String`
Length Constraints: Maximum length of 3072.
Pattern: `.*`
Required: No

**DisplayName**
The name of the experiment as displayed. If `DisplayName` isn't specified, `ExperimentName` is displayed.
Type: `String`
Length Constraints: Minimum length of 1. Maximum length of 120.
Pattern: `^[a-zA-Z0-9\-]*[a-zA-Z0-9]{0,119}`
Required: No

**ExperimentArn**
The Amazon Resource Name (ARN) of the experiment.
Type: `String`
Length Constraints: Maximum length of 256.
Pattern: `arn:aws\-[a-z-]*:sagemaker\-[a-z0-9-]*:[0-9]{12}:experiment/\.*`
Required: No

**ExperimentName**
The name of the experiment.
Experiment

Type: String
Length Constraints: Minimum length of 1. Maximum length of 120.
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,119\}
Required: No

LastModifiedBy
Information about the user who created or modified an experiment, trial, trial component, lineage group, project, or model card.
Type: UserContext (p. 2067) object
Required: No

LastModifiedTime
When the experiment was last modified.
Type: Timestamp
Required: No

Source
The source of the experiment.
Type: ExperimentSource (p. 1475) object
Required: No

Tags
The list of tags that are associated with the experiment. You can use Search API to search on the tags.
Type: Array of Tag (p. 1979) objects
Array Members: Minimum number of 0 items. Maximum number of 50 items.
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ExperimentConfig

Service: Amazon SageMaker Service

Associates a SageMaker job as a trial component with an experiment and trial. Specified when you call the following APIs:

- CreateProcessingJob
- CreateTrainingJob
- CreateTransformJob

Contents

**ExperimentName**

The name of an existing experiment to associate with the trial component.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: ^[a-zA-Z0-9][-]*[a-zA-Z0-9]{0,119}$

Required: No

**RunName**

The name of the experiment run to associate with the trial component.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: ^[a-zA-Z0-9][-]*[a-zA-Z0-9]{0,119}$

Required: No

**TrialComponentDisplayName**

The display name for the trial component. If this key isn't specified, the display name is the trial component name.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: ^[a-zA-Z0-9][-]*[a-zA-Z0-9]{0,119}$

Required: No

**TrialName**

The name of an existing trial to associate the trial component with. If not specified, a new trial is created.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: ^[a-zA-Z0-9][-]*[a-zA-Z0-9]{0,119}$

Required: No
See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ExperimentSource

Service: Amazon SageMaker Service

The source of the experiment.

Contents

SourceArn

The Amazon Resource Name (ARN) of the source.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:.*

Required: Yes

SourceType

The source type.

Type: String

Length Constraints: Maximum length of 128.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
**ExperimentSummary**

Service: Amazon SageMaker Service

A summary of the properties of an experiment. To get the complete set of properties, call the DescribeExperiment API and provide the ExperimentName.

**Contents**

**CreationTime**

When the experiment was created.

Type: Timestamp

Required: No

**DisplayName**

The name of the experiment as displayed. If DisplayName isn't specified, ExperimentName is displayed.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,119}`

Required: No

**ExperimentArn**

The Amazon Resource Name (ARN) of the experiment.

Type: String

Length Constraints: Maximum length of 256.

Pattern: `arn:aws[a-z\-]*:sagemaker:[a-zA-Z0-9\-]*:[0-9]{12}:experiment/.*`

Required: No

**ExperimentName**

The name of the experiment.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,119}`

Required: No

**ExperimentSource**

The source of the experiment.

Type: [ExperimentSource](p. 1475) object

Required: No

**LastModifiedTime**

When the experiment was last modified.
Type: Timestamp
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
Explainability
Service: Amazon SageMaker Service

Contains explainability metrics for a model.

Contents

Report

The explainability report for a model.

Type: MetricsSource (p. 1655) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ExplainerConfig

Service: Amazon SageMaker Service

A parameter to activate explainers.

Contents

ClarifyExplainerConfig

A member of ExplainerConfig that contains configuration parameters for the SageMaker Clarify explainer.

Type: ClarifyExplainerConfig (p. 1331) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
FailStepMetadata
Service: Amazon SageMaker Service

The container for the metadata for Fail step.

Contents

ErrorMessage

A message that you define and then is processed and rendered by the Fail step when the error occurs.

Type: String

Length Constraints: Maximum length of 3072.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
FeatureDefinition
Service: Amazon SageMaker Service

A list of features. You must include FeatureName and FeatureType. Valid feature FeatureTypes are Integral, Fractional and String.

Contents

CollectionConfig

Configuration for your collection.

Type: CollectionConfig (p. 1359) object

Note: This object is a Union. Only one member of this object can be specified or returned.

Required: No

CollectionType

A grouping of elements where each element within the collection must have the same feature type (String, Integral, or Fractional).

- List: An ordered collection of elements.
- Set: An unordered collection of unique elements.
- Vector: A specialized list that represents a fixed-size array of elements. The vector dimension is determined by you. Must have elements with fractional feature types.

Type: String

Valid Values: List | Set | Vector

Required: No

FeatureName

The name of a feature. The type must be a string. FeatureName cannot be any of the following: is_deleted, write_time, api_invocation_time.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 64.

Pattern: ^[a-zA-Z0-9][-_]*[a-zA-Z0-9]{0,63}

Required: No

FeatureType

The value type of a feature. Valid values are Integral, Fractional, or String.

Type: String

Valid Values: Integral | Fractional | String

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
FeatureGroup
Service: Amazon SageMaker Service

Amazon SageMaker Feature Store stores features in a collection called Feature Group. A Feature Group can be visualized as a table which has rows, with a unique identifier for each row where each column in the table is a feature. In principle, a Feature Group is composed of features and values per features.

Contents

**CreationTime**

The time a FeatureGroup was created.

Type: Timestamp

Required: No

**Description**

A free form description of a FeatureGroup.

Type: String

Length Constraints: Maximum length of 128.

Required: No

**EventTimeFeatureName**

The name of the feature that stores the EventTime of a Record in a FeatureGroup.

A EventTime is point in time when a new event occurs that corresponds to the creation or update of a Record in FeatureGroup. All Records in the FeatureGroup must have a corresponding EventTime.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 64.

Pattern: \^[a-zA-Z0-9]([-_]*[a-zA-Z0-9]){0,63}\$

Required: No

**FailureReason**

The reason that the FeatureGroup failed to be replicated in the OfflineStore. This is failure may be due to a failure to create a FeatureGroup in or delete a FeatureGroup from the OfflineStore.

Type: String

Length Constraints: Maximum length of 1024.

Required: No

**FeatureDefinitions**

A list of Features. Each Feature must include a FeatureName and a FeatureType.

Valid FeatureTypes are Integral, Fractional and String.

FeatureNames cannot be any of the following: is_deleted, write_time, api_invocation_time.
You can create up to 2,500 FeatureDefinitions per FeatureGroup.

Type: Array of FeatureDefinition (p. 1481) objects

Array Members: Minimum number of 1 item. Maximum number of 2500 items.

Required: No

**FeatureGroupArn**

The Amazon Resource Name (ARN) of a FeatureGroup.

Type: String

Length Constraints: Maximum length of 256.

Pattern: `arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:feature-group/.*`

Required: No

**FeatureGroupName**

The name of the FeatureGroup.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 64.

Pattern: `^[a-zA-Z0-9][-_]*[a-zA-Z0-9]{0,63}`

Required: No

**FeatureGroupStatus**

A FeatureGroup status.

Type: String

Valid Values: Creating | Created | CreateFailed | Deleting | DeleteFailed

Required: No

**LastModifiedTime**

A timestamp indicating the last time you updated the feature group.

Type: Timestamp

Required: No

**LastUpdateStatus**

A value that indicates whether the feature group was updated successfully.

Type: LastUpdateStatus (p. 1644) object

Required: No

**OfflineStoreConfig**

The configuration of an OfflineStore.

Provide an OfflineStoreConfig in a request to CreateFeatureGroup to create an OfflineStore.

To encrypt an OfflineStore using at rest data encryption, specify AWS Key Management Service (KMS) key ID, or KMSKeyId, in S3StorageConfig.
FeatureGroup

Type: OfflineStoreConfig (p. 1777) object

Required: No

OfflineStoreStatus

The status of OfflineStore.

Type: OfflineStoreStatus (p. 1778) object

Required: No

OnlineStoreConfig

Use this to specify the AWS Key Management Service (KMS) Key ID, or KMSKeyId, for at rest data encryption. You can turn OnlineStore on or off by specifying the EnableOnlineStore flag at General Assembly.

The default value is False.

Type: OnlineStoreConfig (p. 1784) object

Required: No

RecordIdentifierFeatureName

The name of the Feature whose value uniquely identifies a Record defined in the FeatureGroup FeatureDefinitions.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 64.

Pattern: ^[a-zA-Z0-9][-_]*[a-zA-Z0-9]{0,63}

Required: No

RoleArn

The Amazon Resource Name (ARN) of the IAM execution role used to create the feature group.

Type: String


Pattern: ^arn:aws[a-z-]*:iam::\d{12}:role/\?[a-zA-Z0-9+=,.@-_/]+$

Required: No

Tags

Tags used to define a FeatureGroup.

Type: Array of Tag (p. 1979) objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
**FeatureGroupSummary**

Service: Amazon SageMaker Service

The name, ARN, CreationTime, FeatureGroup values, LastUpdatedTime and EnableOnlineStorage status of a FeatureGroup.

**Contents**

**CreationTime**

A timestamp indicating the time of creation time of the FeatureGroup.

Type: Timestamp

Required: Yes

**FeatureGroupArn**

Unique identifier for the FeatureGroup.

Type: String

Length Constraints: Maximum length of 256.

Pattern: `arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:feature-group/.*`

Required: Yes

**FeatureGroupName**

The name of FeatureGroup.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 64.

Pattern: `^[a-zA-Z0-9]([-\_]*[a-zA-Z0-9])\{0,63\}`

Required: Yes

**FeatureGroupStatus**

The status of a FeatureGroup. The status can be any of the following: Creating, Created, CreateFail, Deleting or DetailFail.

Type: String

Valid Values: Creating | Created | CreateFailed | Deleting | DeleteFailed

Required: No

**OfflineStoreStatus**

Notifies you if replicating data into the OfflineStore has failed. Returns either: Active or Blocked.

Type: OfflineStoreStatus (p. 1778) object

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:
• AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for Ruby V3
FeatureMetadata

Service: Amazon SageMaker Service

The metadata for a feature. It can either be metadata that you specify, or metadata that is updated automatically.

Contents

CreationTime

A timestamp indicating when the feature was created.

Type: Timestamp

Required: No

Description

An optional description that you specify to better describe the feature.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 255.

Pattern: .*

Required: No

FeatureGroupArn

The Amazon Resource Number (ARN) of the feature group.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-zA-Z-]*:sagemaker:[a-zA-Z-]*:[0-9]{12}:feature-group/.*

Required: No

FeatureGroupName

The name of the feature group containing the feature.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 64.

Pattern: ^[a-zA-Z0-9\-\_]\*[a-zA-Z0-9\-\_]\{0,63}\$

Required: No

FeatureName

The name of feature.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 64.

Pattern: ^[a-zA-Z0-9\-\_]\*[a-zA-Z0-9\-\_]\{0,63}\$

Required: No
**FeatureType**

The data type of the feature.

Type: String

Valid Values: Integral | Fractional | String

Required: No

**LastModifiedTime**

A timestamp indicating when the feature was last modified.

Type: Timestamp

Required: No

**Parameters**

Optional key-value pairs that you specify to better describe the feature.

Type: Array of FeatureParameter (p. 1491) objects

Array Members: Minimum number of 0 items. Maximum number of 25 items.

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
FeatureParameter
Service: Amazon SageMaker Service
A key-value pair that you specify to describe the feature.

Contents

Key
A key that must contain a value to describe the feature.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 255.
Pattern: ^([\p{L}\p{Z}\p{N}\_\.:+=\-]+)\$
Required: No

Value
The value that belongs to a key.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 255.
Pattern: ^([\p{L}\p{Z}\p{N}\_\.:+=\-]+)\$
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
FileSource
Service: Amazon SageMaker Service
Contains details regarding the file source.

Contents

S3Uri
The Amazon S3 URI for the file source.
Type: String
Length Constraints: Maximum length of 1024.
Pattern: ^(https|s3)://([^/]+)/(.*)$
Required: Yes

ContentDigest
The digest of the file source.
Type: String
Length Constraints: Maximum length of 72.
Pattern: ^[Ss][Hh][Aa]256:[0-9a-fA-F]{64}$
Required: No

ContentType
The type of content stored in the file source.
Type: String
Length Constraints: Maximum length of 256.
Pattern: .*$
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
FileSystemConfig
Service: Amazon SageMaker Service

The Amazon Elastic File System (EFS) storage configuration for a SageMaker image.

Contents

DefaultGid

The default POSIX group ID (GID). If not specified, defaults to 100.

Type: Integer

Valid Range: Minimum value of 0. Maximum value of 65535.

Required: No

DefaultUid

The default POSIX user ID (UID). If not specified, defaults to 1000.

Type: Integer

Valid Range: Minimum value of 0. Maximum value of 65535.

Required: No

MountPath

The path within the image to mount the user's EFS home directory. The directory should be empty. If not specified, defaults to /home/sagemaker-user.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: ^\/.*$

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
FileSystemDataSource
Service: Amazon SageMaker Service

Specifies a file system data source for a channel.

Contents

DirectoryPath

The full path to the directory to associate with the channel.

Type: String

Length Constraints: Maximum length of 4096.

Pattern: .*

Required: Yes

FileSystemAccessMode

The access mode of the mount of the directory associated with the channel. A directory can be mounted either in `ro` (read-only) or `rw` (read-write) mode.

Type: String

Valid Values: `rw` | `ro`

Required: Yes

FileSystemId

The file system id.

Type: String

Length Constraints: Minimum length of 11.

Pattern: ^fs-[0-9a-f]{8,}$

Required: Yes

FileSystemType

The file system type.

Type: String

Valid Values: `EFS` | `FSxLustre`

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
Filter
Service: Amazon SageMaker Service

A conditional statement for a search expression that includes a resource property, a Boolean operator, and a value. Resources that match the statement are returned in the results from the Search API.

If you specify a Value, but not an Operator, SageMaker uses the equals operator.

In search, there are several property types:

Metrics
To define a metric filter, enter a value using the form "Metrics.<name>" where <name> is a metric name. For example, the following filter searches for training jobs with an "accuracy" metric greater than "0.9":

```
{
    "Name": "Metrics.accuracy",
    "Operator": "GreaterThan",
    "Value": "0.9"
}
```

HyperParameters
To define a hyperparameter filter, enter a value with the form "HyperParameters.<name>". Decimal hyperparameter values are treated as a decimal in a comparison if the specified Value is also a decimal value. If the specified Value is an integer, the decimal hyperparameter values are treated as integers. For example, the following filter is satisfied by training jobs with a "learning_rate" hyperparameter that is less than "0.5":

```
{
    "Name": "HyperParameters.learning_rate",
    "Operator": "LessThan",
    "Value": "0.5"
}
```

Tags
To define a tag filter, enter a value with the form Tags.<key>.

Contents

Name
A resource property name. For example, TrainingJobName. For valid property names, see SearchRecord. You must specify a valid property for the resource.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 255.

Pattern: .+
Required: Yes

**Operator**

A Boolean binary operator that is used to evaluate the filter. The operator field contains one of the following values:

- **Equals**
  The value of Name equals Value.

- **NotEquals**
  The value of Name doesn't equal Value.

- **Exists**
  The Name property exists.

- **NotExists**
  The Name property does not exist.

- **GreaterThan**
  The value of Name is greater than Value. Not supported for text properties.

- **GreaterThanOrEqualTo**
  The value of Name is greater than or equal to Value. Not supported for text properties.

- **LessThan**
  The value of Name is less than Value. Not supported for text properties.

- **LessThanOrEqualTo**
  The value of Name is less than or equal to Value. Not supported for text properties.

- **In**
  The value of Name is one of the comma delimited strings in Value. Only supported for text properties.

- **Contains**
  The value of Name contains the string Value. Only supported for text properties.

A SearchExpression can include the Contains operator multiple times when the value of Name is one of the following:
- `Experiment.DisplayName`
- `Experiment.ExperimentName`
- `Experiment.Tags`
- `Trial.DisplayName`
- `Trial.TrialName`
- `Trial.Tags`
- `TrialComponent.DisplayName`
- `TrialComponent.TrialComponentName`
- `TrialComponent.Tags`
- `TrialComponent.InputArtifacts`
- `TrialComponent.OutputArtifacts`

A SearchExpression can include only one Contains operator for all other values of Name. In these cases, if you include multiple Contains operators in the SearchExpression, the result is the following error message: "'CONTAINS' operator usage limit of 1 exceeded."
Filter

**Type:** String

Valid Values: Equals | NotEquals | GreaterThan | GreaterThanOrEqualTo | LessThan | LessThanOrEqualTo | Contains | Exists | NotExists | In

Required: No

**Value**

A value used with Name and Operator to determine which resources satisfy the filter's condition. For numerical properties, Value must be an integer or floating-point decimal. For timestamp properties, Value must be an ISO 8601 date-time string of the following format: YYYY-mm-dd'T'HH:MM:SS.

**Type:** String


Pattern: .+

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
FinalAutoMLJobObjectiveMetric

Service: Amazon SageMaker Service

The best candidate result from an AutoML training job.

Contents

**MetricName**

The name of the metric with the best result. For a description of the possible objective metrics, see `AutoMLJobObjective$MetricName`.

Type: String

Valid Values: Accuracy | MSE | F1 | F1macro | AUC | RMSE | MAE | R2 | BalancedAccuracy | Precision | PrecisionMacro | Recall | RecallMacro | MAPE | MASE | WAPE | AverageWeightedQuantileLoss

Required: Yes

**Value**

The value of the metric with the best result.

Type: Float

Required: Yes

**StandardMetricName**

The name of the standard metric. For a description of the standard metrics, see [Autopilot candidate metrics](#).

Type: String

Valid Values: Accuracy | MSE | F1 | F1macro | AUC | RMSE | MAE | R2 | BalancedAccuracy | Precision | PrecisionMacro | Recall | RecallMacro | MAPE | MASE | WAPE | AverageWeightedQuantileLoss

Required: No

**Type**

The type of metric with the best result.

Type: String

Valid Values: Maximize | Minimize

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
FinalHyperParameterTuningJobObjectiveMetric

Service: Amazon SageMaker Service

Shows the latest objective metric emitted by a training job that was launched by a hyperparameter tuning job. You define the objective metric in the HyperParameterTuningJobObjective parameter of HyperParameterTuningJobConfig.

Contents

MetricName

The name of the objective metric. For SageMaker built-in algorithms, metrics are defined per algorithm. See the metrics for XGBoost as an example. You can also use a custom algorithm for training and define your own metrics. For more information, see Define metrics and environment variables.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 255.

Pattern: .+

Required: Yes

Value

The value of the objective metric.

Type: Float

Required: Yes

Type

Select if you want to minimize or maximize the objective metric during hyperparameter tuning.

Type: String

Valid Values: Maximize | Minimize

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
FlowDefinitionOutputConfig

Service: Amazon SageMaker Service

Contains information about where human output will be stored.

Contents

S3OutputPath

The Amazon S3 path where the object containing human output will be made available.

To learn more about the format of Amazon A2I output data, see Amazon A2I Output Data.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: ^(https|s3)://([^/]*)/?(.*)$

Required: Yes

KmsKeyId

The Amazon Key Management Service (KMS) key ID for server-side encryption.

Type: String

Length Constraints: Maximum length of 2048.

Pattern: .*

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
FlowDefinitionSummary
Service: Amazon SageMaker Service
Contains summary information about the flow definition.

Contents

CreationTime
The timestamp when SageMaker created the flow definition.
Type: Timestamp
Required: Yes

FlowDefinitionArn
The Amazon Resource Name (ARN) of the flow definition.
Type: String
Length Constraints: Maximum length of 1024.
Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:flow-definition/.*
Required: Yes

FlowDefinitionName
The name of the flow definition.
Type: String
Pattern: ^[a-z0-9](-*[a-z0-9])\{0,62
Required: Yes

FlowDefinitionStatus
The status of the flow definition. Valid values:
Type: String
Valid Values: Initializing | Active | Failed | Deleting
Required: Yes

FailureReason
The reason why the flow definition creation failed. A failure reason is returned only when the flow
definition status is Failed.
Type: String
Length Constraints: Maximum length of 1024.
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:
• AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for Ruby V3
GitConfig
Service: Amazon SageMaker Service

Specifies configuration details for a Git repository in your AWS account.

Contents

RepositoryUrl

The URL where the Git repository is located.

Type: String


Pattern: ^https://([^/]+)/?.{3,1016}$

Required: Yes

Branch

The default branch for the Git repository.

Type: String


Pattern: [^ ~^:?*\[]+

Required: No

SecretArn

The Amazon Resource Name (ARN) of the AWS Secrets Manager secret that contains the credentials used to access the git repository. The secret must have a staging label of AWSCURRENT and must be in the following format:

{"username": UserName, "password": Password}

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: arn:aws[a-z\-]*:secretsmanager:[a-z0-9\-]*:[0-9]{12}:secret:.*

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
GitConfigForUpdate

Service: Amazon SageMaker Service

Specifies configuration details for a Git repository when the repository is updated.

Contents

SecretArn

The Amazon Resource Name (ARN) of the AWS Secrets Manager secret that contains the credentials used to access the git repository. The secret must have a staging label of AWSCURRENT and must be in the following format:

{"username": UserName, "password": Password}

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: arn:aws[a-z\-]*:secretsmanager:[a-z0-9\-]*:[0-9]{12}:secret:.*

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
HolidayConfigAttributes

Service: Amazon SageMaker Service

Stores the holiday featurization attributes applicable to each item of time-series datasets during the training of a forecasting model. This allows the model to identify patterns associated with specific holidays.

Contents

CountryCode

The country code for the holiday calendar.

For the list of public holiday calendars supported by AutoML job V2, see Country Codes. Use the country code corresponding to the country of your choice.

Type: String

Length Constraints: Fixed length of 2.

Pattern: \[A-Z\]\{2\}

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
HubContentDependency
Service: Amazon SageMaker Service

Any dependencies related to hub content, such as scripts, model artifacts, datasets, or notebooks.

Contents

**DependencyCopyPath**

The hub content dependency copy path.

- **Type:** String
- **Length Constraints:** Maximum length of 1023.
- **Pattern:** .*
- **Required:** No

**DependencyOriginPath**

The hub content dependency origin path.

- **Type:** String
- **Length Constraints:** Maximum length of 1023.
- **Pattern:** .*
- **Required:** No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](https://aws.amazon.com/sdk-for-cpp/)
- [AWS SDK for Go](https://aws.amazon.com/sdk-for-go/)
- [AWS SDK for Java V2](https://aws.amazon.com/sdk-for-java/)
- [AWS SDK for Ruby V3](https://aws.amazon.com/sdk-for-ruby/)
**HubContentInfo**

Service: Amazon SageMaker Service

Information about hub content.

**Contents**

**CreationTime**

The date and time that the hub content was created.

Type: Timestamp

Required: Yes

**DocumentSchemaVersion**

The version of the hub content document schema.

Type: String


Pattern: ^\d{1,4}.\d{1,4}.\d{1,4}$

Required: Yes

**HubContentArn**

The Amazon Resource Name (ARN) of the hub content.

Type: String

Length Constraints: Maximum length of 255.

Pattern: .*

Required: Yes

**HubContentName**

The name of the hub content.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](?=.*[a-zA-Z0-9])\{0,62}$

Required: Yes

**HubContentStatus**

The status of the hub content.

Type: String

Valid Values: Available | Importing | Deleting | ImportFailed | DeleteFailed

Required: Yes

**HubContentType**

The type of hub content.
Type: String
Valid Values: Model | Notebook
Required: Yes

**HubContentVersion**

The version of the hub content.

Type: String


Pattern: `^\d{1,4}\.\d{1,4}\.\d{1,4}$`

Required: Yes

**HubContentDescription**

A description of the hub content.

Type: String

Length Constraints: Maximum length of 1023.

Pattern: `. *

Required: No

**HubContentDisplayName**

The display name of the hub content.

Type: String

Length Constraints: Maximum length of 255.

Pattern: `. *

Required: No

**HubContentSearchKeywords**

The searchable keywords for the hub content.

Type: Array of strings

Array Members: Maximum number of 50 items.

Length Constraints: Maximum length of 255.

Pattern: `^[^A-Z]*$

Required: No

### See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
• AWS SDK for Ruby V3
HubInfo
Service: Amazon SageMaker Service

Information about a hub.

Contents

CreationTime
The date and time that the hub was created.
Type: Timestamp
Required: Yes

HubArn
The Amazon Resource Name (ARN) of the hub.
Type: String
Length Constraints: Maximum length of 255.
Pattern: .*
Required: Yes

HubName
The name of the hub.
Type: String
Length Constraints: Maximum length of 63.
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62\}
Required: Yes

HubStatus
The status of the hub.
Type: String
Valid Values: InService | Creating | Updating | Deleting | CreateFailed | UpdateFailed | DeleteFailed
Required: Yes

LastModifiedTime
The date and time that the hub was last modified.
Type: Timestamp
Required: Yes

HubDescription
A description of the hub.
Type: String
Length Constraints: Maximum length of 1023.
Pattern: . *
Required: No

**HubDisplayName**

The display name of the hub.
Type: String
Length Constraints: Maximum length of 255.
Pattern: . *
Required: No

**HubSearchKeywords**

The searchable keywords for the hub.
Type: Array of strings
Array Members: Maximum number of 50 items.
Length Constraints: Maximum length of 255.
Pattern: ^[^A-Z]*$
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
HubS3StorageConfig
Service: Amazon SageMaker Service
The Amazon S3 storage configuration of a hub.

Contents

S3OutputPath

The Amazon S3 bucket prefix for hosting hub content.
Type: String
Length Constraints: Maximum length of 1024.
Pattern: ^((https|s3)://([^/]+)/)?(.*$)
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
HumanLoopActivationConditionsConfig

Service: Amazon SageMaker Service

Defines under what conditions SageMaker creates a human loop. Used within `CreateFlowDefinition`. See `HumanLoopActivationConditionsConfig` for the required format of activation conditions.

Contents

**HumanLoopActivationConditions**

JSON expressing use-case specific conditions declaratively. If any condition is matched, atomic tasks are created against the configured work team. The set of conditions is different for Rekognition and Textract. For more information about how to structure the JSON, see JSON Schema for Human Loop Activation Conditions in Amazon Augmented AI in the Amazon SageMaker Developer Guide.

Type: String

Length Constraints: Maximum length of 10240.

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
HumanLoopActivationConfig
Service: Amazon SageMaker Service

Provides information about how and under what conditions SageMaker creates a human loop. If HumanLoopActivationConfig is not given, then all requests go to humans.

Contents

HumanLoopActivationConditionsConfig

Container structure for defining under what conditions SageMaker creates a human loop.

Type: HumanLoopActivationConditionsConfig (p. 1515) object

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
HumanLoopConfig

Service: Amazon SageMaker Service

Describes the work to be performed by human workers.

Contents

HumanTaskUiArn

The Amazon Resource Name (ARN) of the human task user interface.

You can use standard HTML and Crowd HTML Elements to create a custom worker task template. You use this template to create a human task UI.

To learn how to create a custom HTML template, see Create Custom Worker Task Template.

To learn how to create a human task UI, which is a worker task template that can be used in a flow definition, see Create and Delete a Worker Task Templates.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: `arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:human-task-ui/.*`

Required: Yes

TaskCount

The number of distinct workers who will perform the same task on each object. For example, if TaskCount is set to 3 for an image classification labeling job, three workers will classify each input image. Increasing TaskCount can improve label accuracy.

Type: Integer


Required: Yes

TaskDescription

A description for the human worker task.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 255.

Pattern: .+

Required: Yes

TaskTitle

A title for the human worker task.

Type: String


Pattern: `^[\t\n\r\x00-\xD7FF\xE000-\xFFFD]*$`

Required: Yes
WorkteamArn

Amazon Resource Name (ARN) of a team of workers. To learn more about the types of workforces and work teams you can create and use with Amazon A2I, see Create and Manage Workforces.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:workteam/.*

Required: Yes

PublicWorkforceTaskPrice

Defines the amount of money paid to an Amazon Mechanical Turk worker for each task performed.

Use one of the following prices for bounding box tasks. Prices are in US dollars and should be based on the complexity of the task; the longer it takes in your initial testing, the more you should offer.

- 0.036
- 0.048
- 0.060
- 0.072
- 0.120
- 0.240
- 0.360
- 0.480
- 0.600
- 0.720
- 0.840
- 0.960
- 1.080
- 1.200

Use one of the following prices for image classification, text classification, and custom tasks. Prices are in US dollars.

- 0.012
- 0.024
- 0.036
- 0.048
- 0.060
- 0.072
- 0.120
- 0.240
- 0.360
- 0.480
- 0.600
- 0.720
- 0.840
- 0.960
- 1.080
• 1.200

Use one of the following prices for semantic segmentation tasks. Prices are in US dollars.

• 0.840
• 0.960
• 1.080
• 1.200

Use one of the following prices for Textract AnalyzeDocument Important Form Key Amazon Augmented AI review tasks. Prices are in US dollars.

• 2.400
• 2.280
• 2.160
• 2.040
• 1.920
• 1.800
• 1.680
• 1.560
• 1.440
• 1.320
• 1.200
• 1.080
• 0.960
• 0.840
• 0.720
• 0.600
• 0.480
• 0.360
• 0.240
• 0.120
• 0.072
• 0.060
• 0.048
• 0.036
• 0.024
• 0.012

Use one of the following prices for Rekognition DetectModerationLabels Amazon Augmented AI review tasks. Prices are in US dollars.

• 1.200
• 1.080
• 0.960
• 0.840
• 0.720
• 0.600
• 0.480
• 0.360
Use one of the following prices for Amazon Augmented AI custom human review tasks. Prices are in US dollars.

- 1.200
- 1.080
- 0.960
- 0.840
- 0.720
- 0.600
- 0.480
- 0.360
- 0.240
- 0.120
- 0.072
- 0.060
- 0.048
- 0.036
- 0.024
- 0.012

Type: `PublicWorkforceTaskPrice (p. 1876)` object

Required: No

**TaskAvailabilityLifetimeInSeconds**

The length of time that a task remains available for review by human workers.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 864000.

Required: No

**TaskKeywords**

Keywords used to describe the task so that workers can discover the task.

Type: Array of strings

Array Members: Minimum number of 1 item. Maximum number of 5 items.


Pattern: `^[A-Za-z0-9]+( [A-Za-z0-9]+)*$`

Required: No
TaskTimeLimitInSeconds

The amount of time that a worker has to complete a task. The default value is 3,600 seconds (1 hour).

Type: Integer
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
HumanLoopRequestSource

Service: Amazon SageMaker Service

Container for configuring the source of human task requests.

Contents

AwsManagedHumanLoopRequestSource

Specifies whether Amazon Rekognition or Amazon Textract are used as the integration source. The default field settings and JSON parsing rules are different based on the integration source. Valid values:

Type: String

Valid Values: AWS/Rekognition/DetectModerationLabels/Image/V3 | AWS/Textract/AnalyzeDocument/Forms/V1

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
HumanTaskConfig

Service: Amazon SageMaker Service

Information required for human workers to complete a labeling task.

Contents

AnnotationConsolidationConfig

Configures how labels are consolidated across human workers.

Type: AnnotationConsolidationConfig (p. 1234) object

Required: Yes

NumberOfHumanWorkersPerDataObject

The number of human workers that will label an object.

Type: Integer


Required: Yes

PreHumanTaskLambdaArn

The Amazon Resource Name (ARN) of a Lambda function that is run before a data object is sent to a human worker. Use this function to provide input to a custom labeling job.

For built-in task types, use one of the following Amazon SageMaker Ground Truth Lambda function ARNs for PreHumanTaskLambdaArn. For custom labeling workflows, see Pre-annotation Lambda.

- **Bounding box** - Finds the most similar boxes from different workers based on the Jaccard index of the boxes.
  - arn:aws:lambda:us-east-1:432418664414:function:PRE-BoundingBox
  - arn:aws:lambda:ca-central-1:918755190332:function:PRE-BoundingBox
  - arn:aws:lambda:eu-west-1:568282634449:function:PRE-BoundingBox
  - arn:aws:lambda:eu-west-2:487402164563:function:PRE-BoundingBox
  - arn:aws:lambda:eu-central-1:203001061592:function:PRE-BoundingBox
  - arn:aws:lambda:ap-northeast-1:477331159723:function:PRE-BoundingBox
  - arn:aws:lambda:ap-south-1:565803892007:function:PRE-BoundingBox
  - arn:aws:lambda:ap-southeast-1:377565633583:function:PRE-BoundingBox

- **Image classification** - Uses a variant of the Expectation Maximization approach to estimate the true class of an image based on annotations from individual workers.
Multi-label image classification - Uses a variant of the Expectation Maximization approach to estimate the true classes of an image based on annotations from individual workers.


Semantic segmentation - Treats each pixel in an image as a multi-class classification and treats pixel annotations from workers as "votes" for the correct label.

Text classification - Uses a variant of the Expectation Maximization approach to estimate the true class of text based on annotations from individual workers.

Multi-label text classification - Uses a variant of the Expectation Maximization approach to estimate the true classes of text based on annotations from individual workers.

Named entity recognition - Groups similar selections and calculates aggregate boundaries, resolving to most-assigned label.
Video Classification - Use this task type when you need workers to classify videos using predefined labels that you specify. Workers are shown videos and are asked to choose one label for each video.

Video Frame Object Detection - Use this task type to have workers identify and locate objects in a sequence of video frames (images extracted from a video) using bounding boxes. For example, you can use this task to ask workers to identify and localize various objects in a series of video frames, such as cars, bikes, and pedestrians.
HumanTaskConfig

Video Frame Object Tracking - Use this task type to have workers track the movement of objects in a sequence of video frames (images extracted from a video) using bounding boxes. For example, you can use this task to ask workers to track the movement of objects, such as cars, bikes, and pedestrians.


3D Point Cloud Modalities

Use the following pre-annotation lambdas for 3D point cloud labeling modality tasks. See 3D Point Cloud Task types to learn more.

3D Point Cloud Object Detection - Use this task type when you want workers to classify objects in a 3D point cloud by drawing 3D cuboids around objects. For example, you can use this task type to ask workers to identify different types of objects in a point cloud, such as cars, bikes, and pedestrians.

3D Point Cloud Object Detection - Use this task type when you want workers to identify objects in a sequence of 3D point cloud frames. For example, you can use this task type to ask workers to detect objects in a video sequence.


3D Point Cloud Object Tracking - Use this task type when you want workers to draw 3D cuboids around objects that appear in a sequence of 3D point cloud frames. For example, you can use this task type to ask workers to track the movement of objects across multiple point cloud frames.

- arn:aws:lambda:us-east-1:432418664414:function:PRE-3DPointCloudObjectTracking
- arn:aws:lambda:eu-west-1:568282634449:function:PRE-3DPointCloudObjectTracking
- arn:aws:lambda:ap-south-1:565803892007:function:PRE-3DPointCloudObjectTracking

3D Point Cloud Semantic Segmentation - Use this task type when you want workers to create a point-level semantic segmentation masks by painting objects in a 3D point cloud using different colors where each color is assigned to one of the classes you specify.

- arn:aws:lambda:eu-west-1:568282634449:function:PRE-3DPointCloudSemanticSegmentation
Use the following ARNs for Label Verification and Adjustment Jobs

Use label verification and adjustment jobs to review and adjust labels. To learn more, see Verify and Adjust Labels.

**Bounding box verification** - Uses a variant of the Expectation Maximization approach to estimate the true class of verification judgement for bounding box labels based on annotations from individual workers.

- arn:aws:lambda:us-east-1:432418664414:function:PRE-VerificationBoundingBox
- arn:aws:lambda:eu-west-1:568282634449:function:PRE-VerificationBoundingBox
- arn:aws:lambda:ap-northeast-1:477331159723:function:PRE-VerificationBoundingBox
- arn:aws:lambda:ap-southeast-1:2:454466003867:function:PRE-VerificationBoundingBox
- arn:aws:lambda:ap-south-1:565803892007:function:PRE-VerificationBoundingBox
- arn:aws:lambda:eu-central-1:203001061592:function:PRE-VerificationBoundingBox
- arn:aws:lambda:eu-west-2:487402164563:function:PRE-VerificationBoundingBox
- arn:aws:lambda:ap-southeast-1:377565633583:function:PRE-VerificationBoundingBox
- arn:aws:lambda:ca-central-1:918755190332:function:PRE-VerificationBoundingBox

**Bounding box adjustment** - Finds the most similar boxes from different workers based on the Jaccard index of the adjusted annotations.

- arn:aws:lambda:us-east-1:432418664414:function:PRE-AdjustmentBoundingBox
- arn:aws:lambda:ca-central-1:918755190332:function:PRE-AdjustmentBoundingBox
- arn:aws:lambda:eu-west-1:568282634449:function:PRE-AdjustmentBoundingBox
- arn:aws:lambda:eu-west-2:487402164563:function:PRE-AdjustmentBoundingBox
- arn:aws:lambda:ap-southeast-1:377565633583:function:PRE-AdjustmentBoundingBox
- arn:aws:lambda:ca-central-1:918755190332:function:PRE-AdjustmentBoundingBox
Semantic segmentation verification - Uses a variant of the Expectation Maximization approach to estimate the true class of verification judgment for semantic segmentation labels based on annotations from individual workers.

Semantic segmentation adjustment - Treats each pixel in an image as a multi-class classification and treats pixel adjusted annotations from workers as "votes" for the correct label.
Video Frame Object Detection Adjustment - Use this task type when you want workers to adjust bounding boxes that workers have added to video frames to classify and localize objects in a sequence of video frames.

Video Frame Object Tracking Adjustment - Use this task type when you want workers to adjust bounding boxes that workers have added to video frames to track object movement across a sequence of video frames.
• arn:aws:lambda:eu-west-1:568282634449:function:PRE-AdjustmentVideoObjectTracking
• arn:aws:lambda:ap-northeast-1:477331159723:function:PRE-AdjustmentVideoObjectTracking
• arn:aws:lambda:ap-south-1:565803892007:function:PRE-AdjustmentVideoObjectTracking
• arn:aws:lambda:eu-central-1:203001061592:function:PRE-AdjustmentVideoObjectTracking
• arn:aws:lambda:eu-west-2:487402164563:function:PRE-AdjustmentVideoObjectTracking
• arn:aws:lambda:ap-southeast-1:377565633583:function:PRE-AdjustmentVideoObjectTracking
• arn:aws:lambda:eu-central-1:918755190332:function:PRE-AdjustmentVideoObjectTracking

3D point cloud object detection adjustment - Adjust 3D cuboids in a point cloud frame.
• arn:aws:lambda:us-east-1:432418664414:function:PRE-Adjustment3DPointCloudObjectDetection
• arn:aws:lambda:eu-west-1:568282634449:function:PRE-Adjustment3DPointCloudObjectDetection
• arn:aws:lambda:ap-northeast-1:477331159723:function:PRE-Adjustment3DPointCloudObjectDetection
• arn:aws:lambda:ap-south-1:565803892007:function:PRE-Adjustment3DPointCloudObjectDetection
• arn:aws:lambda:eu-central-1:203001061592:function:PRE-Adjustment3DPointCloudObjectDetection
• arn:aws:lambda:ap-southeast-1:377565633583:function:PRE-Adjustment3DPointCloudObjectDetection

3D point cloud object tracking adjustment - Adjust 3D cuboids across a sequence of point cloud frames.
• arn:aws:lambda:us-east-1:432418664414:function:PRE-Adjustment3DPointCloudObjectTracking
• arn:aws:lambda:us-east-2:266458841044:function:PRE-Adjustment3DPointCloudObjectTracking
• arn:aws:lambda:eu-west-1:568282634449:function:PRE-Adjustment3DPointCloudObjectTracking
• arn:aws:lambda:ap-northeast-1:477331159723:function:PRE-Adjustment3DPointCloudObjectTracking
• arn:aws:lambda:ap-south-1:565803892007:function:PRE-Adjustment3DPointCloudObjectTracking
• arn:aws:lambda:eu-central-1:203001061592:function:PRE-Adjustment3DPointCloudObjectTracking
• arn:aws:lambda:ap-southeast-1:377565633583:function:PRE-Adjustment3DPointCloudObjectTracking
• arn:aws:lambda:ca-central-1:918755190332:function:PRE-Adjustment3DPointCloudObjectTracking

3D point cloud semantic segmentation adjustment - Adjust semantic segmentation masks in a 3D point cloud.
• arn:aws:lambda:us-east-1:432418664414:function:PRE-Adjustment3DPointCloudSemanticSegmentation
• arn:aws:lambda:eu-west-1:568282634449:function:PRE-Adjustment3DPointCloudSemanticSegmentation
• arn:aws:lambda:ap-northeast-1:477331159723:function:PRE-Adjustment3DPointCloudSemanticSegmentation
• arn:aws:lambda:ap-south-1:565803892007:function:PRE-Adjustment3DPointCloudSemanticSegmentation
• arn:aws:lambda:eu-central-1:203001061592:function:PRE-Adjustment3DPointCloudSemanticSegmentation
• arn:aws:lambda:eu-west-1:487402164563:function:PRE-Adjustment3DPointCloudSemanticSegmentation
• arn:aws:lambda:ap-southeast-1:377565633583:function:PRE-Adjustment3DPointCloudSemanticSegmentation
• arn:aws:lambda:ca-central-1:918755190332:function:PRE-Adjustment3DPointCloudSemanticSegmentation

Type: String
Length Constraints: Maximum length of 2048.

Pattern: arn:aws[a-z\-]*:lambda:[a-z0-9\-]*:[0-9]{12}:function:.*

Required: Yes

**TaskDescription**

A description of the task for your human workers.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 255.

Pattern: .+

Required: Yes

**TaskTimeLimitInSeconds**

The amount of time that a worker has to complete a task.

If you create a custom labeling job, the maximum value for this parameter is 8 hours (28,800 seconds).

If you create a labeling job using a built-in task type the maximum for this parameter depends on the task type you use:

- For image and text labeling jobs, the maximum is 8 hours (28,800 seconds).
- For 3D point cloud and video frame labeling jobs, the maximum is 30 days (2952,000 seconds) for non-AL mode. For most users, the maximum is also 30 days.

Type: Integer

Valid Range: Minimum value of 30.

Required: Yes

**TaskTitle**

A title for the task for your human workers.

Type: String


Pattern: ^[\t\n\r -\u0D7FF\uE000-\uFFFF]*$

Required: Yes

**UiConfig**

Information about the user interface that workers use to complete the labeling task.

Type: UiConfig (p. 2062) object

Required: Yes

**WorkteamArn**

The Amazon Resource Name (ARN) of the work team assigned to complete the tasks.

Type: String

Length Constraints: Maximum length of 256.
Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:workteam/.*

Required: Yes

**MaxConcurrentTaskCount**

Defines the maximum number of data objects that can be labeled by human workers at the same time. Also referred to as batch size. Each object may have more than one worker at one time. The default value is 1000 objects. To increase the maximum value to 5000 objects, contact AWS Support.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 5000.

Required: No

**PublicWorkforceTaskPrice**

The price that you pay for each task performed by an Amazon Mechanical Turk worker.

Type: `PublicWorkforceTaskPrice (p. 1876)` object

Required: No

**TaskAvailabilityLifetimeInSeconds**

The length of time that a task remains available for labeling by human workers. The default and maximum values for this parameter depend on the type of workforce you use.

- If you choose the Amazon Mechanical Turk workforce, the maximum is 12 hours (43,200 seconds). The default is 6 hours (21,600 seconds).
- If you choose a private or vendor workforce, the default value is 30 days (2592,000 seconds) for non-AL mode. For most users, the maximum is also 30 days.

Type: Integer

Valid Range: Minimum value of 60.

Required: No

**TaskKeywords**

Keywords used to describe the task so that workers on Amazon Mechanical Turk can discover the task.

Type: Array of strings

Array Members: Minimum number of 1 item. Maximum number of 5 items.


Pattern: `^[A-Za-z0-9]+( [A-Za-z0-9]+)*$`

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
• AWS SDK for Ruby V3
HumanTaskUiSummary

Service: Amazon SageMaker Service

Container for human task user interface information.

Contents

**CreationTime**

A timestamp when SageMaker created the human task user interface.

Type: Timestamp

Required: Yes

**HumanTaskUiArn**

The Amazon Resource Name (ARN) of the human task user interface.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: `arn:aws[a-z\-\*]:sagemaker:[a-z0-9\-\*]:[0-9]{12}:human-task-ui/.*`

Required: Yes

**HumanTaskUiName**

The name of the human task user interface.

Type: String


Pattern: `^[a-z0-9\-\*]$`

Required: Yes

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](https://docs.aws.amazon.com/sdk-for-cpp/v1/)
- [AWS SDK for Go](https://docs.aws.amazon.com/sdk-for-go/v1/)
- [AWS SDK for Java V2](https://docs.aws.amazon.com/sdk-for-java/v2/)
- [AWS SDK for Ruby V3](https://docs.aws.amazon.com/sdk-for-ruby/v3/)

1537
HyperbandStrategyConfig

Service: Amazon SageMaker Service

The configuration for Hyperband, a multi-fidelity based hyperparameter tuning strategy. Hyperband uses the final and intermediate results of a training job to dynamically allocate resources to utilized hyperparameter configurations while automatically stopping under-performing configurations. This parameter should be provided only if Hyperband is selected as the StrategyConfig under the HyperParameterTuningJobConfig API.

Contents

MaxResource

The maximum number of resources (such as epochs) that can be used by a training job launched by a hyperparameter tuning job. Once a job reaches the MaxResource value, it is stopped. If a value for MaxResource is not provided, and Hyperband is selected as the hyperparameter tuning strategy, HyperbandTrainingJob attempts to infer MaxResource from the following keys (if present) in StaticsHyperParameters:

- epochs
- numepochs
- n-epochs
- n_epochs
- num_epochs

If HyperbandStrategyConfig is unable to infer a value for MaxResource, it generates a validation error. The maximum value is 20,000 epochs. All metrics that correspond to an objective metric are used to derive early stopping decisions. For distributive training jobs, ensure that duplicate metrics are not printed in the logs across the individual nodes in a training job. If multiple nodes are publishing duplicate or incorrect metrics, training jobs may make an incorrect stopping decision and stop the job prematurely.

Type: Integer

Valid Range: Minimum value of 1.

Required: No

MinResource

The minimum number of resources (such as epochs) that can be used by a training job launched by a hyperparameter tuning job. If the value for MinResource has not been reached, the training job is not stopped by Hyperband.

Type: Integer

Valid Range: Minimum value of 1.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
• AWS SDK for Ruby V3
HyperParameterAlgorithmSpecification

Service: Amazon SageMaker Service

Specifies which training algorithm to use for training jobs that a hyperparameter tuning job launches and the metrics to monitor.

Contents

TrainingInputMode

The training input mode that the algorithm supports. For more information about input modes, see Algorithms.

Pipe mode

If an algorithm supports Pipe mode, Amazon SageMaker streams data directly from Amazon S3 to the container.

File mode

If an algorithm supports File mode, SageMaker downloads the training data from S3 to the provisioned ML storage volume, and mounts the directory to the Docker volume for the training container.

You must provision the ML storage volume with sufficient capacity to accommodate the data downloaded from S3. In addition to the training data, the ML storage volume also stores the output model. The algorithm container uses the ML storage volume to also store intermediate information, if any.

For distributed algorithms, training data is distributed uniformly. Your training duration is predictable if the input data objects sizes are approximately the same. SageMaker does not split the files any further for model training. If the object sizes are skewed, training won't be optimal as the data distribution is also skewed when one host in a training cluster is overloaded, thus becoming a bottleneck in training.

FastFile mode

If an algorithm supports FastFile mode, SageMaker streams data directly from S3 to the container with no code changes, and provides file system access to the data. Users can author their training script to interact with these files as if they were stored on disk.

FastFile mode works best when the data is read sequentially. Augmented manifest files aren't supported. The startup time is lower when there are fewer files in the S3 bucket provided.

Type: String

Valid Values: Pipe | File | FastFile

Required: Yes

AlgorithmName

The name of the resource algorithm to use for the hyperparameter tuning job. If you specify a value for this parameter, do not specify a value for TrainingImage.

Type: String


Pattern: (arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:[a-z\-]*\/)??([a-zA-Z0-9\-]+)[a-zA-Z0-9\-]+\{0,62}\(\<\?-\)?$
MetricDefinitions

An array of MetricDefinition objects that specify the metrics that the algorithm emits.

Type: Array of MetricDefinition (p. 1653) objects

Array Members: Minimum number of 0 items. Maximum number of 40 items.

TrainingImage

The registry path of the Docker image that contains the training algorithm. For information about Docker registry paths for built-in algorithms, see Algorithms Provided by Amazon SageMaker: Common Parameters. SageMaker supports both registry/repository[:tag] and registry/repository[@digest] image path formats. For more information, see Using Your Own Algorithms with Amazon SageMaker.

Type: String

Length Constraints: Maximum length of 255.

Pattern: .*

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
HyperParameterSpecification

Service: Amazon SageMaker Service

Defines a hyperparameter to be used by an algorithm.

Contents

Name

The name of this hyperparameter. The name must be unique.

Type: String

Length Constraints: Maximum length of 256.

Pattern: [\p{L}\p{M}\p{Z}\p{S}\p{N}\p{P}]*

Required: Yes

Type

The type of this hyperparameter. The valid types are Integer, Continuous, Categorical, and FreeText.

Type: String

Valid Values: Integer | Continuous | Categorical | FreeText

Required: Yes

DefaultValue

The default value for this hyperparameter. If a default value is specified, a hyperparameter cannot be required.

Type: String

Length Constraints: Maximum length of 2500.

Pattern: .*

Required: No

Description

A brief description of the hyperparameter.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: [\p{L}\p{M}\p{Z}\p{S}\p{N}\p{P}]*

Required: No

IsRequired

Indicates whether this hyperparameter is required.

Type: Boolean

Required: No
IsTunable

Indicates whether this hyperparameter is tunable in a hyperparameter tuning job.

Type: Boolean

Required: No

Range

The allowed range for this hyperparameter.

Type: ParameterRange (p. 1799) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
HyperParameterTrainingJobDefinition

Service: Amazon SageMaker Service

Defines the training jobs launched by a hyperparameter tuning job.

Contents

AlgorithmSpecification

The HyperParameterAlgorithmSpecification object that specifies the resource algorithm to use for the training jobs that the tuning job launches.

Type: HyperParameterAlgorithmSpecification (p. 1540) object

Required: Yes

OutputDataConfig

Specifies the path to the Amazon S3 bucket where you store model artifacts from the training jobs that the tuning job launches.

Type: OutputDataConfig (p. 1792) object

Required: Yes

RoleArn

The Amazon Resource Name (ARN) of the IAM role associated with the training jobs that the tuning job launches.

Type: String


Pattern: ^arn:aws[a-z\-]*:iam::\d{12}:role/?[a-zA-Z_0-9+=,.@\-_\/]+$

Required: Yes

StoppingCondition

Specifies a limit to how long a model hyperparameter training job can run. It also specifies how long a managed spot training job has to complete. When the job reaches the time limit, SageMaker ends the training job. Use this API to cap model training costs.

Type: StoppingCondition (p. 1968) object

Required: Yes

CheckpointConfig

Contains information about the output location for managed spot training checkpoint data.

Type: CheckpointConfig (p. 1328) object

Required: No

DefinitionName

The job definition name.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 64.
HyperParameterTrainingJobDefinition

Pattern: ^[a-zA-Z0-9]{0,63}$

Required: No

EnableInterContainerTrafficEncryption

To encrypt all communications between ML compute instances in distributed training, choose True. Encryption provides greater security for distributed training, but training might take longer. How long it takes depends on the amount of communication between compute instances, especially if you use a deep learning algorithm in distributed training.

Type: Boolean

Required: No

EnableManagedSpotTraining

A Boolean indicating whether managed spot training is enabled (True) or not (False).

Type: Boolean

Required: No

EnableNetworkIsolation

Isolates the training container. No inbound or outbound network calls can be made, except for calls between peers within a training cluster for distributed training. If network isolation is used for training jobs that are configured to use a VPC, SageMaker downloads and uploads customer data and model artifacts through the specified VPC, but the training container does not have network access.

Type: Boolean

Required: No

Environment

An environment variable that you can pass into the SageMaker CreateTrainingJob API. You can use an existing environment variable from the training container or use your own. See Define metrics and variables for more information.

**Note**

The maximum number of items specified for Map Entries refers to the maximum number of environment variables for each TrainingJobDefinition and also the maximum for the hyperparameter tuning job itself. That is, the sum of the number of environment variables for all the training job definitions can't exceed the maximum number specified.

Type: String to string map

Map Entries: Maximum number of 48 items.

Key Length Constraints: Maximum length of 512.

Key Pattern: [a-zA-Z\_\-]*

Value Length Constraints: Maximum length of 512.

Value Pattern: \S\s*

Required: No

HyperParameterRanges

Specifies ranges of integer, continuous, and categorical hyperparameters that a hyperparameter tuning job searches. The hyperparameter tuning job launches training jobs with hyperparameter
values within these ranges to find the combination of values that result in the training job with the best performance as measured by the objective metric of the hyperparameter tuning job.

**Note**
The maximum number of items specified for `Array Members` refers to the maximum number of hyperparameters for each range and also the maximum for the hyperparameter tuning job itself. That is, the sum of the number of hyperparameters for all the ranges can't exceed the maximum number specified.

Type: [ParameterRanges](p. 1800) object

Required: No

**HyperParameterTuningResourceConfig**

The configuration for the hyperparameter tuning resources, including the compute instances and storage volumes, used for training jobs launched by the tuning job. By default, storage volumes hold model artifacts and incremental states. Choose `File` for `TrainingInputMode` in the `AlgorithmSpecification` parameter to additionally store training data in the storage volume (optional).

Type: [HyperParameterTuningResourceConfig](p. 1566) object

Required: No

**InputDataConfig**

An array of `Channel` objects that specify the input for the training jobs that the tuning job launches.

Type: Array of [Channel](p. 1324) objects

Array Members: Minimum number of 1 item. Maximum number of 20 items.

Required: No

**ResourceConfig**

The resources, including the compute instances and storage volumes, to use for the training jobs that the tuning job launches.

Storage volumes store model artifacts and incremental states. Training algorithms might also use storage volumes for scratch space. If you want SageMaker to use the storage volume to store the training data, choose `File` as the `TrainingInputMode` in the algorithm specification. For distributed training algorithms, specify an instance count greater than 1.

**Note**
If you want to use hyperparameter optimization with instance type flexibility, use `HyperParameterTuningResourceConfig` instead.

Type: [ResourceConfig](p. 1910) object

Required: No

**RetryStrategy**

The number of times to retry the job when the job fails due to an `InternalServerError`.

Type: [RetryStrategy](p. 1918) object

Required: No

**StaticHyperParameters**

Specifies the values of hyperparameters that do not change for the tuning job.
Type: String to string map

Map Entries: Minimum number of 0 items. Maximum number of 100 items.

Key Length Constraints: Maximum length of 256.

Key Pattern: . *

Value Length Constraints: Maximum length of 2500.

Value Pattern: . *

Required: No

**TuningObjective**

Defines the objective metric for a hyperparameter tuning job. Hyperparameter tuning uses the value of this metric to evaluate the training jobs it launches, and returns the training job that results in either the highest or lowest value for this metric, depending on the value you specify for the Type parameter. If you want to define a custom objective metric, see [Define metrics and environment variables](#).

Type: `HyperParameterTuningJobObjective` (p. 1557) object

Required: No

**VpcConfig**

The `VpcConfig` object that specifies the VPC that you want the training jobs that this hyperparameter tuning job launches to connect to. Control access to and from your training container by configuring the VPC. For more information, see [Protect Training Jobs by Using an Amazon Virtual Private Cloud](#).

Type: `VpcConfig` (p. 2076) object

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
HyperParameterTrainingJobSummary

Service: Amazon SageMaker Service

The container for the summary information about a training job.

Contents

CreationTime

The date and time that the training job was created.
Type: Timestamp
Required: Yes

TrainingJobArn

The Amazon Resource Name (ARN) of the training job.
Type: String
Length Constraints: Maximum length of 256.
Pattern: \barn:aws[a-z-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:training-job/.*
Required: Yes

TrainingJobName

The name of the training job.
Type: String
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}
Required: Yes

TrainingJobStatus

The status of the training job.
Type: String
Valid Values: InProgress | Completed | Failed | Stopping | Stopped
Required: Yes

TunedHyperParameters

A list of the hyperparameters for which you specified ranges to search.
Type: String to string map
Map Entries: Minimum number of 0 items. Maximum number of 100 items.
Key Length Constraints: Maximum length of 256.
Key Pattern: .*
Value Length Constraints: Maximum length of 2500.
Value Pattern: .*
HyperParameterTrainingJobSummary

Required: Yes

**FailureReason**

The reason that the training job failed.

Type: String

Length Constraints: Maximum length of 1024.

Required: No

**FinalHyperParameterTuningJobObjectiveMetric**

The [FinalHyperParameterTuningJobObjectiveMetric](p. 1501) object that specifies the value of the objective metric of the tuning job that launched this training job.

Type: [FinalHyperParameterTuningJobObjectiveMetric](p. 1501) object

Required: No

**ObjectiveStatus**

The status of the objective metric for the training job:

- Succeeded: The final objective metric for the training job was evaluated by the hyperparameter tuning job and used in the hyperparameter tuning process.
- Pending: The training job is in progress and evaluation of its final objective metric is pending.
- Failed: The final objective metric for the training job was not evaluated, and was not used in the hyperparameter tuning process. This typically occurs when the training job failed or did not emit an objective metric.

Type: String

Valid Values: Succeeded | Pending | Failed

Required: No

**TrainingEndTime**

Specifies the time when the training job ends on training instances. You are billed for the time interval between the value of `TrainingStartTime` and this time. For successful jobs and stopped jobs, this is the time after model artifacts are uploaded. For failed jobs, this is the time when SageMaker detects a job failure.

Type: Timestamp

Required: No

**TrainingJobDefinitionName**

The training job definition name.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 64.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,63}$`

Required: No

**TrainingStartTime**

The date and time that the training job started.

Type: Timestamp
Required: No

**TuningJobName**

The HyperParameter tuning job that launched the training job.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,31\}

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
HyperParameterTuningInstanceConfig

Service: Amazon SageMaker Service

The configuration for hyperparameter tuning resources for use in training jobs launched by the tuning job. These resources include compute instances and storage volumes. Specify one or more compute instance configurations and allocation strategies to select resources (optional).

Contents

InstanceCount

The number of instances of the type specified by InstanceType. Choose an instance count larger than 1 for distributed training algorithms. See Step 2: Launch a SageMaker Distributed Training Job Using the SageMaker Python SDK for more information.

Type: Integer

Valid Range: Minimum value of 0.

Required: Yes

InstanceType

The instance type used for processing of hyperparameter optimization jobs. Choose from general purpose (no GPUs) instance types: ml.m5.xlarge, ml.m5.2xlarge, and ml.m5.4xlarge or compute optimized (no GPUs) instance types: ml.c5.xlarge and ml.c5.2xlarge. For more information about instance types, see instance type descriptions.

Type: String

Valid Values: ml.m4.xlarge | ml.m4.2xlarge | ml.m4.4xlarge | ml.m4.10xlarge | ml.m4.16xlarge | ml.g4dn.xlarge | ml.g4dn.2xlarge | ml.g4dn.4xlarge | ml.g4dn.8xlarge | ml.g4dn.12xlarge | ml.g4dn.16xlarge | ml.m5.large | ml.m5.xlarge | ml.m5.2xlarge | ml.m5.4xlarge | ml.m5.12xlarge | ml.m5.24xlarge | ml.c4.xlarge | ml.c4.2xlarge | ml.c4.4xlarge | ml.c4.8xlarge | ml.p2.xlarge | ml.p2.8xlarge | ml.p2.16xlarge | ml.p3.2xlarge | ml.p3.8xlarge | ml.p3.16xlarge | ml.p3dn.24xlarge | ml.p4d.24xlarge | ml.c5.xlarge | ml.c5.2xlarge | ml.c5.4xlarge | ml.c5.9xlarge | ml.c5.18xlarge | ml.c5n.xlarge | ml.c5n.2xlarge | ml.c5n.4xlarge | ml.c5n.9xlarge | ml.c5n.18xlarge | ml.g5.xlarge | ml.g5.2xlarge | ml.g5.4xlarge | ml.g5.8xlarge | ml.g5.16xlarge | ml.g5.32xlarge | ml.g5.64xlarge | ml.g5.96xlarge | ml.g5.12xlarge | ml.g5.24xlarge | ml.g5.48xlarge | ml.g5.96xlarge | ml.trn1.2xlarge | ml.trn1.32xlarge | ml.trn1n.32xlarge | ml.trn1.64xlarge | ml.trn1n.64xlarge | ml.p4.8xlarge | ml.p4.16xlarge | ml.p4.32xlarge | ml.p4.64xlarge | ml.p4.96xlarge | ml.p4.12xlarge | ml.p4.24xlarge | ml.p4.48xlarge | ml.p4.96xlarge | ml.p4.128xlarge | ml.p4n.128xlarge

Required: Yes

VolumeSizeInGB

The volume size in GB of the data to be processed for hyperparameter optimization (optional).

Type: Integer

Valid Range: Minimum value of 1.

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
HyperParameterTuningJobCompletionDetails
Service: Amazon SageMaker Service

A structure that contains runtime information about both current and completed hyperparameter tuning jobs.

Contents

ConvergenceDetectedTime

The time in timestamp format that AMT detected model convergence, as defined by a lack of significant improvement over time based on criteria developed over a wide range of diverse benchmarking tests.

Type: Timestamp
Required: No

NumberOfTrainingJobsObjectiveNotImproving

The number of training jobs launched by a tuning job that are not improving (1% or less) as measured by model performance evaluated against an objective function.

Type: Integer
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
**HyperParameterTuningJobConfig**
Service: Amazon SageMaker Service

Configures a hyperparameter tuning job.

**Contents**

**ResourceLimits**

The `ResourceLimits` object that specifies the maximum number of training and parallel training jobs that can be used for this hyperparameter tuning job.

Type: `ResourceLimits` (p. 1914) object

Required: Yes

**Strategy**

Specifies how hyperparameter tuning chooses the combinations of hyperparameter values to use for the training job it launches. For information about search strategies, see [How Hyperparameter Tuning Works](#).

Type: String

Valid Values: Bayesian | Random | Hyperband | Grid

Required: Yes

**HyperParameterTuningJobObjective**

The `HyperParameterTuningJobObjective` specifies the objective metric used to evaluate the performance of training jobs launched by this tuning job.

Type: `HyperParameterTuningJobObjective` (p. 1557) object

Required: No

**ParameterRanges**

The `ParameterRanges` object that specifies the ranges of hyperparameters that this tuning job searches over to find the optimal configuration for the highest model performance against your chosen objective metric.

Type: `ParameterRanges` (p. 1800) object

Required: No

**RandomSeed**

A value used to initialize a pseudo-random number generator. Setting a random seed and using the same seed later for the same tuning job will allow hyperparameter optimization to find more a consistent hyperparameter configuration between the two runs.

Type: Integer

Valid Range: Minimum value of 0.

Required: No

**StrategyConfig**

The configuration for the Hyperband optimization strategy. This parameter should be provided only if Hyperband is selected as the strategy for `HyperParameterTuningJobConfig`. 
Type: `HyperParameterTuningJobStrategyConfig` (p. 1561) object

Required: No

**TrainingJobEarlyStoppingType**

Specifies whether to use early stopping for training jobs launched by the hyperparameter tuning job. Because the Hyperband strategy has its own advanced internal early stopping mechanism, `TrainingJobEarlyStoppingType` must be `OFF` to use Hyperband. This parameter can take on one of the following values (the default value is `OFF`):

OFF

Training jobs launched by the hyperparameter tuning job do not use early stopping.

AUTO

SageMaker stops training jobs launched by the hyperparameter tuning job when they are unlikely to perform better than previously completed training jobs. For more information, see [Stop Training Jobs Early](#).

Type: String

Valid Values: Off | Auto

Required: No

**TuningJobCompletionCriteria**

The tuning job's completion criteria.

Type: `TuningJobCompletionCriteria` (p. 2060) object

Required: No

---

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
HyperParameterTuningJobConsumedResources
Service: Amazon SageMaker Service

The total resources consumed by your hyperparameter tuning job.

Contents

RuntimeInSeconds

The wall clock runtime in seconds used by your hyperparameter tuning job.

Type: Integer

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
HyperParameterTuningJobObjective

Service: Amazon SageMaker Service

Defines the objective metric for a hyperparameter tuning job. Hyperparameter tuning uses the value of this metric to evaluate the training jobs it launches, and returns the training job that results in either the highest or lowest value for this metric, depending on the value you specify for the Type parameter. If you want to define a custom objective metric, see Define metrics and environment variables.

Contents

MetricName

The name of the metric to use for the objective metric.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 255.

Pattern: .+

Required: Yes

Type

Whether to minimize or maximize the objective metric.

Type: String

Valid Values: Maximize | Minimize

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
HyperParameterTuningJobSearchEntity

Service: Amazon SageMaker Service

An entity returned by the SearchRecord API containing the properties of a hyperparameter tuning job.

Contents

BestTrainingJob

The container for the summary information about a training job.

Type: HyperParameterTrainingJobSummary (p. 1548) object

Required: No

ConsumedResources

The total amount of resources consumed by a hyperparameter tuning job.

Type: HyperParameterTuningJobConsumedResources (p. 1556) object

Required: No

CreationTime

The time that a hyperparameter tuning job was created.

Type: Timestamp

Required: No

FailureReason

The error that was created when a hyperparameter tuning job failed.

Type: String

Length Constraints: Maximum length of 1024.

Required: No

HyperParameterTuningEndTime

The time that a hyperparameter tuning job ended.

Type: Timestamp

Required: No

HyperParameterTuningJobArn

The Amazon Resource Name (ARN) of a hyperparameter tuning job.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:hyper-parameter-tuning-job/.*

Required: No

HyperParameterTuningJobConfig

Configures a hyperparameter tuning job.
Type: `HyperParameterTuningJobConfig (p. 1554)` object

Required: No

**HyperParameterTuningJobName**

The name of a hyperparameter tuning job.

Type: String


Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,31}$`

Required: No

**HyperParameterTuningJobStatus**

The status of a hyperparameter tuning job.

Type: String

Valid Values: `Completed` | `InProgress` | `Failed` | `Stopped` | `Stopping`

Required: No

**LastModifiedTime**

The time that a hyperparameter tuning job was last modified.

Type: Timestamp

Required: No

**ObjectiveStatusCounters**

Specifies the number of training jobs that this hyperparameter tuning job launched, categorized by the status of their objective metric. The objective metric status shows whether the final objective metric for the training job has been evaluated by the tuning job and used in the hyperparameter tuning process.

Type: `ObjectiveStatusCounters (p. 1776)` object

Required: No

**OverallBestTrainingJob**

The container for the summary information about a training job.

Type: `HyperParameterTrainingJobSummary (p. 1548)` object

Required: No

**Tags**

The tags associated with a hyperparameter tuning job. For more information see `Tagging AWS resources`.

Type: Array of `Tag (p. 1979)` objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

**TrainingJobDefinition**

Defines the training jobs launched by a hyperparameter tuning job.
Type: HyperParameterTrainingJobDefinition (p. 1544) object

Required: No

**TrainingJobDefinitions**

The job definitions included in a hyperparameter tuning job.

Type: Array of HyperParameterTrainingJobDefinition (p. 1544) objects

Array Members: Minimum number of 1 item. Maximum number of 10 items.

Required: No

**TrainingJobStatusCounters**

The numbers of training jobs launched by a hyperparameter tuning job, categorized by status.

Type: TrainingJobStatusCounters (p. 2011) object

Required: No

**TuningJobCompletionDetails**

Information about either a current or completed hyperparameter tuning job.

Type: HyperParameterTuningJobCompletionDetails (p. 1553) object

Required: No

**WarmStartConfig**

Specifies the configuration for a hyperparameter tuning job that uses one or more previous hyperparameter tuning jobs as a starting point. The results of previous tuning jobs are used to inform which combinations of hyperparameters to search over in the new tuning job.

All training jobs launched by the new hyperparameter tuning job are evaluated by using the objective metric, and the training job that performs the best is compared to the best training jobs from the parent tuning jobs. From these, the training job that performs the best as measured by the objective metric is returned as the overall best training job.

**Note**

All training jobs launched by parent hyperparameter tuning jobs and the new hyperparameter tuning jobs count against the limit of training jobs for the tuning job.

Type: HyperParameterTuningJobWarmStartConfig (p. 1564) object

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
**HyperParameterTuningJobStrategyConfig**

Service: Amazon SageMaker Service

The configuration for a training job launched by a hyperparameter tuning job. Choose Bayesian for Bayesian optimization, and Random for random search optimization. For more advanced use cases, use Hyperband, which evaluates objective metrics for training jobs after every epoch. For more information about strategies, see [How Hyperparameter Tuning Works](#).

### Contents

**HyperbandStrategyConfig**

The configuration for the object that specifies the Hyperband strategy. This parameter is only supported for the Hyperband selection for Strategy within the HyperParameterTuningJobConfig API.

Type: [HyperbandStrategyConfig](#) object

Required: No

### See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
HyperParameterTuningJobSummary

Service: Amazon SageMaker Service

Provides summary information about a hyperparameter tuning job.

Contents

CreationTime

The date and time that the tuning job was created.

Type: Timestamp

Required: Yes

HyperParameterTuningJobArn

The Amazon Resource Name (ARN) of the tuning job.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:hyper-parameter-tuning-job/.*

Required: Yes

HyperParameterTuningJobName

The name of the tuning job.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,31}

Required: Yes

HyperParameterTuningJobStatus

The status of the tuning job.

Type: String

Valid Values: Completed | InProgress | Failed | Stopped | Stopping

Required: Yes

ObjectiveStatusCounters

The ObjectiveStatusCounters object that specifies the numbers of training jobs, categorized by objective metric status, that this tuning job launched.

Type: ObjectiveStatusCounters (p. 1776) object

Required: Yes

Strategy

Specifies the search strategy hyperparameter tuning uses to choose which hyperparameters to evaluate at each iteration.
Type: String

Valid Values: Bayesian | Random | Hyperband | Grid

Required: Yes

**TrainingJobStatusCounters**

The *TrainingJobStatusCounters* object that specifies the numbers of training jobs, categorized by status, that this tuning job launched.

Type: *TrainingJobStatusCounters (p. 2011)* object

Required: Yes

**HyperParameterTuningEndTime**

The date and time that the tuning job ended.

Type: Timestamp

Required: No

**LastModifiedTime**

The date and time that the tuning job was modified.

Type: Timestamp

Required: No

**ResourceLimits**

The *ResourceLimits* object that specifies the maximum number of training jobs and parallel training jobs allowed for this tuning job.

Type: *ResourceLimits (p. 1914)* object

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
HyperParameterTuningJobWarmStartConfig

Service: Amazon SageMaker Service

Specifies the configuration for a hyperparameter tuning job that uses one or more previous hyperparameter tuning jobs as a starting point. The results of previous tuning jobs are used to inform which combinations of hyperparameters to search over in the new tuning job.

All training jobs launched by the new hyperparameter tuning job are evaluated by using the objective metric, and the training job that performs the best is compared to the best training jobs from the parent tuning jobs. From these, the training job that performs the best as measured by the objective metric is returned as the overall best training job.

**Note**
All training jobs launched by parent hyperparameter tuning jobs and the new hyperparameter tuning jobs count against the limit of training jobs for the tuning job.

**Contents**

**ParentHyperParameterTuningJobs**

An array of hyperparameter tuning jobs that are used as the starting point for the new hyperparameter tuning job. For more information about warm starting a hyperparameter tuning job, see [Using a Previous Hyperparameter Tuning Job as a Starting Point](#).

Hyperparameter tuning jobs created before October 1, 2018 cannot be used as parent jobs for warm start tuning jobs.

Type: Array of [ParentHyperParameterTuningJob](#) objects

Array Members: Minimum number of 1 item. Maximum number of 5 items.

Required: Yes

**WarmStartType**

Specifies one of the following:

**IDENTICAL_DATA_AND_ALGORITHM**

The new hyperparameter tuning job uses the same input data and training image as the parent tuning jobs. You can change the hyperparameter ranges to search and the maximum number of training jobs that the hyperparameter tuning job launches. You cannot use a new version of the training algorithm, unless the changes in the new version do not affect the algorithm itself. For example, changes that improve logging or adding support for a different data format are allowed. You can also change hyperparameters from tunable to static, and from static to tunable, but the total number of static plus tunable hyperparameters must remain the same as it is in all parent jobs. The objective metric for the new tuning job must be the same as for all parent jobs.

**TRANSFER_LEARNING**

The new hyperparameter tuning job can include input data, hyperparameter ranges, maximum number of concurrent training jobs, and maximum number of training jobs that are different than those of its parent hyperparameter tuning jobs. The training image can also be a different version from the version used in the parent hyperparameter tuning job. You can also change hyperparameters from tunable to static, and from static to tunable, but the total number of static plus tunable hyperparameters must remain the same as it is in all parent jobs. The objective metric for the new tuning job must be the same as for all parent jobs.

Type: String
Valid Values: IdenticalDataAndAlgorithm | TransferLearning

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
HyperParameterTuningResourceConfig

Service: Amazon SageMaker Service

The configuration of resources, including compute instances and storage volumes for use in training jobs launched by hyperparameter tuning jobs. HyperParameterTuningResourceConfig is similar to ResourceConfig, but has the additional InstanceConfigs and AllocationStrategy fields to allow for flexible instance management. Specify one or more instance types, count, and the allocation strategy for instance selection.

Note

HyperParameterTuningResourceConfig supports the capabilities of ResourceConfig with the exception of KeepAlivePeriodInSeconds. Hyperparameter tuning jobs use warm pools by default, which reuse clusters between training jobs.

Contents

AllocationStrategy

The strategy that determines the order of preference for resources specified in InstanceConfigs used in hyperparameter optimization.

Type: String

Valid Values: Prioritized

Required: No

InstanceConfigs

A list containing the configuration(s) for one or more resources for processing hyperparameter jobs. These resources include compute instances and storage volumes to use in model training jobs launched by hyperparameter tuning jobs. The AllocationStrategy controls the order in which multiple configurations provided in InstanceConfigs are used.

Note

If you only want to use a single instance configuration inside the HyperParameterTuningResourceConfig API, do not provide a value for InstanceConfigs. Instead, use InstanceType, VolumeSizeInGB and InstanceCount. If you use InstanceConfigs, do not provide values for InstanceType, VolumeSizeInGB or InstanceCount.

Type: Array of HyperParameterTuningInstanceConfig (p. 1551) objects

Array Members: Minimum number of 1 item. Maximum number of 6 items.

Required: No

InstanceCount

The number of compute instances of type InstanceType to use. For distributed training, select a value greater than 1.

Type: Integer

Valid Range: Minimum value of 0.

Required: No

InstanceType

The instance type used to run hyperparameter optimization tuning jobs. See descriptions of instance types for more information.
HyperParameterTuningResourceConfig

Type: String

Valid Values:
- `ml.m4.xlarge`
- `ml.m4.2xlarge`
- `ml.m4.4xlarge`
- `ml.m4.10xlarge`
- `ml.m4.16xlarge`
- `ml.g4dn.xlarge`
- `ml.g4dn.2xlarge`
- `ml.g4dn.4xlarge`
- `ml.g4dn.8xlarge`
- `ml.g4dn.12xlarge`
- `ml.g4dn.16xlarge`
- `ml.m5.large`
- `ml.m5.xlarge`
- `ml.m5.2xlarge`
- `ml.m5.4xlarge`
- `ml.m5.12xlarge`
- `ml.m5.24xlarge`
- `ml.c4.xlarge`
- `ml.c4.2xlarge`
- `ml.c4.4xlarge`
- `ml.c4.8xlarge`
- `ml.p2.xlarge`
- `ml.p2.8xlarge`
- `ml.p2.16xlarge`
- `ml.p3.xlarge`
- `ml.p3.8xlarge`
- `ml.p3.16xlarge`
- `ml.p3dn.24xlarge`
- `ml.c5.xlarge`
- `ml.c5.2xlarge`
- `ml.c5.4xlarge`
- `ml.c5.8xlarge`
- `ml.c5.16xlarge`
- `ml.c5.32xlarge`
- `ml.c5n.xlarge`
- `ml.c5n.2xlarge`
- `ml.c5n.4xlarge`
- `ml.c5n.9xlarge`
- `ml.c5n.18xlarge`
- `ml.g4.xlarge`
- `ml.g4.4xlarge`
- `ml.g4.8xlarge`
- `ml.g4.16xlarge`
- `ml.g4.32xlarge`
- `ml.g4.64xlarge`
- `ml.g5.xlarge`
- `ml.g5.2xlarge`
- `ml.g5.4xlarge`
- `ml.g5.8xlarge`
- `ml.g5.16xlarge`
- `ml.g5.32xlarge`
- `ml.g5.64xlarge`
- `ml.trn1.2xlarge`
- `ml.trn1.32xlarge`
- `ml.trn1n.32xlarge`
- `ml.p2dn.4xlarge`
- `ml.p2dn.8xlarge`
- `ml.p2dn.16xlarge`
- `ml.p2dn.32xlarge`
- `ml.p2dn.64xlarge`
- `ml.p2dn.96xlarge`
- `ml.p3dn.12xlarge`
- `ml.p3dn.24xlarge`
- `ml.p3dn.48xlarge`
- `ml.p3dn.96xlarge`
- `ml.p4d.24xlarge`
- `ml.p4d.4xlarge`
- `ml.p4d.8xlarge`
- `ml.p4d.16xlarge`
- `ml.p4d.32xlarge`
- `ml.p4d.64xlarge`
- `ml.p4d.96xlarge`
- `ml.m1.2xlarge`
- `ml.m1.4xlarge`
- `ml.m1.8xlarge`
- `ml.m1.16xlarge`
- `ml.m1.32xlarge`
- `ml.m1.64xlarge`
- `ml.m1.large`
- `ml.m1.xlarge`
- `ml.c1.xlarge`
- `ml.c1.2xlarge`
- `ml.c1.4xlarge`
- `ml.c1.8xlarge`
- `ml.c1.16xlarge`
- `ml.c1.32xlarge`
- `ml.c1n.2xlarge`
- `ml.c1n.4xlarge`
- `ml.c1n.8xlarge`
- `ml.c1n.16xlarge`
- `ml.c1n.32xlarge`
- `ml.c2.xlarge`
- `ml.c2.2xlarge`
- `ml.c2.4xlarge`
- `ml.c2.8xlarge`
- `ml.c2.16xlarge`
- `ml.c2.32xlarge`
- `ml.c3.4xlarge`
- `ml.c3.8xlarge`
- `ml.c3.16xlarge`
- `ml.c3.32xlarge`
- `ml.c4.1xlarge`
- `ml.c4.2xlarge`
- `ml.c4.4xlarge`
- `ml.c4.8xlarge`
- `ml.c4.16xlarge`
- `ml.c4.32xlarge`
- `ml.c5.1xlarge`
- `ml.c5.2xlarge`
- `ml.c5.4xlarge`
- `ml.c5.8xlarge`
- `ml.c5.16xlarge`
- `ml.c5.32xlarge`
- `ml.c5n.1xlarge`
- `ml.c5n.2xlarge`
- `ml.c5n.4xlarge`
- `ml.c5n.8xlarge`
- `ml.c5n.16xlarge`
- `ml.c5n.32xlarge`
- `ml.p1.2xlarge`
- `ml.p1.4xlarge`
- `ml.p1.8xlarge`
- `ml.p1.16xlarge`
- `ml.p1.32xlarge`
- `ml.p1.64xlarge`
- `ml.p2.2xlarge`
- `ml.p2.4xlarge`
- `ml.p2.8xlarge`
- `ml.p2.16xlarge`
- `ml.p2.32xlarge`
- `ml.p2.64xlarge`
- `ml.p3.2xlarge`
- `ml.p3.4xlarge`
- `ml.p3.8xlarge`
- `ml.p3.16xlarge`
- `ml.p3.32xlarge`
- `ml.p3.64xlarge`
- `ml.p4.2xlarge`
- `ml.p4.4xlarge`
- `ml.p4.8xlarge`
- `ml.p4.16xlarge`
- `ml.p4.32xlarge`
- `ml.p4.64xlarge`
- `ml.p100.v2.xlarge`
- `ml.p100.v2.2xlarge`
- `ml.p100.v2.4xlarge`
- `ml.p100.v2.6xlarge`
- `ml.p100.v2.8xlarge`
- `ml.p100.v2.12xlarge`
- `ml.p100.v2.16xlarge`
- `ml.p100.v2.24xlarge`
- `ml.p100.v2.32xlarge`
- `ml.p100.v2.48xlarge`
- `ml.p100.v2.64xlarge`
- `ml.p100.v2.72xlarge`
- `ml.p100.v2.96xlarge`
- `ml.p100.v2.112xlarge`
- `ml.p100.v2.160xlarge`
- `ml.trn1.2xlarge`
- `ml.trn1.32xlarge`
- `ml.trn1n.32xlarge`
- `ml.trn2.2xlarge`
- `ml.trn2.4xlarge`
- `ml.trn2.8xlarge`
- `ml.trn2.16xlarge`
- `ml.trn2.32xlarge`
- `ml.trn2n.32xlarge`

Required: No

**VolumeKmsKeyId**

A key used by AWS Key Management Service to encrypt data on the storage volume attached to the compute instances used to run the training job. You can use either of the following formats to specify a key.

KMS Key ID:

"1234abcd-12ab-34cd-56ef-1234567890ab"

Amazon Resource Name (ARN) of a KMS key:

"arn:aws:kms:us-west-2:111122223333:key/1234abcd-12ab-34cd-56ef-1234567890ab"

Some instances use local storage, which use a hardware module to encrypt storage volumes. If you choose one of these instance types, you cannot request a VolumeKmsKeyId. For a list of instance types that use local storage, see instance store volumes. For more information about AWS Key Management Service, see KMS encryption for more information.

Type: String

Length Constraints: Maximum length of 2048.

Pattern: .*

Required: No

**VolumeSizeInGB**

The volume size in GB for the storage volume to be used in processing hyperparameter optimization jobs (optional). These volumes store model artifacts, incremental states and optionally, scratch space for training algorithms. Do not provide a value for this parameter if a value for InstanceConfigs is also specified.

Some instance types have a fixed total local storage size. If you select one of these instances for training, VolumeSizeInGB cannot be greater than this total size. For a list of instance types with local instance storage and their sizes, see instance store volumes.

**Note**

SageMaker supports only the General Purpose SSD (gp2) storage volume type.

Type: Integer
Valid Range: Minimum value of 0.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
IamIdentity
Service: Amazon SageMaker Service

The IAM Identity details associated with the user. These details are associated with model package groups, model packages and project entities only.

Contents

Arn
The Amazon Resource Name (ARN) of the IAM identity.
Type: String
Required: No

PrincipalId
The ID of the principal that assumes the IAM identity.
Type: String
Required: No

SourceIdentity
The person or application which assumes the IAM identity.
Type: String
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](https://aws.amazon.com/sdk-for-cpp/)
- [AWS SDK for Go](https://aws.amazon.com/sdk-for-golang/)
- [AWS SDK for Java V2](https://aws.amazon.com/sdk-for-java/)
- [AWS SDK for Ruby V3](https://aws.amazon.com/sdk-for-ruby/)
IdentityProviderOAuthSetting

Service: Amazon SageMaker Service

The Amazon SageMaker Canvas application setting where you configure OAuth for connecting to an external data source, such as Snowflake.

Contents

DataSourceName

The name of the data source that you're connecting to. Canvas currently supports OAuth for Snowflake and Salesforce Data Cloud.

Type: String

Valid Values: SalesforceGenie | Snowflake

Required: No

SecretArn

The ARN of an AWS Secrets Manager secret that stores the credentials from your identity provider, such as the client ID and secret, authorization URL, and token URL.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: arn:aws[a-z\-]*:secretsmanager:[a-z0-9\-]*:[0-9]{12}:secret:.*

Required: No

Status

Describes whether OAuth for a data source is enabled or disabled in the Canvas application.

Type: String

Valid Values: ENABLED | DISABLED

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
Image
Service: Amazon SageMaker Service

A SageMaker image. A SageMaker image represents a set of container images that are derived from a common base container image. Each of these container images is represented by a SageMaker ImageVersion.

Contents

CreationTime
When the image was created.
Type: Timestamp
Required: Yes

ImageArn
The ARN of the image.
Type: String
Length Constraints: Maximum length of 256.
Pattern: ^arn:aws(-[\w]+)*:sagemaker:.+:[0-9]{12}:image/[a-z0-9][\-][a-zA-Z0-9]*$
Required: Yes

ImageName
The name of the image.
Type: String
Pattern: ^[a-zA-Z0-9][\-][a-zA-Z0-9][0-9]{62}$
Required: Yes

ImageStatus
The status of the image.
Type: String
Valid Values: CREATING | CREATED | CREATE_FAILED | UPDATING | UPDATE_FAILED | DELETING | DELETE_FAILED
Required: Yes

LastModifiedTime
When the image was last modified.
Type: Timestamp
Required: Yes

Description
The description of the image.
Type: String
Pattern: .*
Required: No

**DisplayName**

The name of the image as displayed.

Type: String
Pattern: ^\S( .*\S)?$^
Required: No

**FailureReason**

When a create, update, or delete operation fails, the reason for the failure.

Type: String
Length Constraints: Maximum length of 1024.
Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
ImageClassificationJobConfig

Service: Amazon SageMaker Service

The collection of settings used by an AutoML job V2 for the image classification problem type.

Contents

CompletionCriteria

How long a job is allowed to run, or how many candidates a job is allowed to generate.

Type: AutoMLJobCompletionCriteria (p. 1280) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ImageConfig

Service: Amazon SageMaker Service

Specifies whether the model container is in Amazon ECR or a private Docker registry accessible from your Amazon Virtual Private Cloud (VPC).

Contents

RepositoryAccessMode

Set this to one of the following values:

- Platform - The model image is hosted in Amazon ECR.
- Vpc - The model image is hosted in a private Docker registry in your VPC.

Type: String

Valid Values: Platform | Vpc

Required: Yes

RepositoryAuthConfig

(Optional) Specifies an authentication configuration for the private docker registry where your model image is hosted. Specify a value for this property only if you specified Vpc as the value for the RepositoryAccessMode field, and the private Docker registry where the model image is hosted requires authentication.

Type: RepositoryAuthConfig (p. 1907) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ImageVersion
Service: Amazon SageMaker Service
A version of a SageMaker Image. A version represents an existing container image.

Contents

CreationTime
When the version was created.
Type: Timestamp
Required: Yes

ImageArn
The ARN of the image the version is based on.
Type: String
Length Constraints: Maximum length of 256.
Pattern: ^arn:aws(-[\w]+)*:sagemaker:.+:[0-9]{12}:image/[a-z0-9]([-.]?[a-z0-9])*$
Required: Yes

ImageVersionArn
The ARN of the version.
Type: String
Length Constraints: Maximum length of 256.
Pattern: ^arn:aws(-[\w]+)*:sagemaker:.+:[0-9]{12}:image-version/[a-z0-9]([-.]?[a-z0-9])*[/0-9]+$
Required: Yes

ImageVersionStatus
The status of the version.
Type: String
Valid Values: CREATING | CREATED | CREATE_FAILED | DELETING | DELETE_FAILED
Required: Yes

LastModifiedTime
When the version was last modified.
Type: Timestamp
Required: Yes

Version
The version number.
Type: Integer
Valid Range: Minimum value of 0.
Required: Yes

**FailureReason**

When a create or delete operation fails, the reason for the failure.

Type: String
Length Constraints: Maximum length of 1024.
Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
InferenceComponentComputeResourceRequirements

Service: Amazon SageMaker Service

Defines the compute resources to allocate to run a model that you assign to an inference component. These resources include CPU cores, accelerators, and memory.

Contents

MinMemoryRequiredInMb

The minimum MB of memory to allocate to run a model that you assign to an inference component.

Type: Integer

Valid Range: Minimum value of 128.

Required: Yes

MaxMemoryRequiredInMb

The maximum MB of memory to allocate to run a model that you assign to an inference component.

Type: Integer

Valid Range: Minimum value of 128.

Required: No

NumberOfAcceleratorDevicesRequired

The number of accelerators to allocate to run a model that you assign to an inference component. Accelerators include GPUs and AWS Inferentia.

Type: Float

Valid Range: Minimum value of 1.

Required: No

NumberOfCpuCoresRequired

The number of CPU cores to allocate to run a model that you assign to an inference component.

Type: Float

Valid Range: Minimum value of 0.25.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
InferenceComponentContainerSpecification

Service: Amazon SageMaker Service

Defines a container that provides the runtime environment for a model that you deploy with an inference component.

Contents

ArtifactUrl

The Amazon S3 path where the model artifacts, which result from model training, are stored. This path must point to a single gzip compressed tar archive (.tar.gz suffix).

Type: String

Length Constraints: Maximum length of 1024.

Pattern: ^(https|s3)://([^/]+)/?(.*)$

Required: No

Environment

The environment variables to set in the Docker container. Each key and value in the Environment string-to-string map can have length of up to 1024. We support up to 16 entries in the map.

Type: String to string map

Map Entries: Maximum number of 16 items.

Key Length Constraints: Maximum length of 1024.

Key Pattern: [a-zA-Z_]\[a-zA-Z0-9_]*

Value Length Constraints: Maximum length of 1024.

Value Pattern: \[\S\s\]*

Required: No

Image

The Amazon Elastic Container Registry (Amazon ECR) path where the Docker image for the model is stored.

Type: String

Length Constraints: Maximum length of 255.

Pattern: \[\S\]+

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for Ruby V3
InferenceComponentContainerSpecificationSummary

Service: Amazon SageMaker Service

Details about the resources that are deployed with this inference component.

Contents

ArtifactUrl

The Amazon S3 path where the model artifacts are stored.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: ^(https|s3)://([^/]+)/?(.*)$

Required: No

DeployedImage

Gets the Amazon EC2 Container Registry path of the docker image of the model that is hosted in this ProductionVariant.

If you used the registry/repository[::tag] form to specify the image path of the primary container when you created the model hosted in this ProductionVariant, the path resolves to a path of the form registry/repository[@digest]. A digest is a hash value that identifies a specific version of an image. For information about Amazon ECR paths, see Pulling an Image in the Amazon ECR User Guide.

Type: DeployedImage (p. 1404) object

Required: No

Environment

The environment variables to set in the Docker container.

Type: String to string map

Map Entries: Maximum number of 16 items.

Key Length Constraints: Maximum length of 1024.

Key Pattern: [a-zA-Z_]([a-zA-Z0-9_]*)

Value Length Constraints: Maximum length of 1024.

Value Pattern: [\S\s]*

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
InferenceComponentRuntimeConfig

Service: Amazon SageMaker Service

Runtime settings for a model that is deployed with an inference component.

Contents

**CopyCount**

The number of runtime copies of the model container to deploy with the inference component. Each copy can serve inference requests.

Type: Integer

Valid Range: Minimum value of 0.

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
InferenceComponentRuntimeConfigSummary

Service: Amazon SageMaker Service

Details about the runtime settings for the model that is deployed with the inference component.

Contents

CurrentCopyCount

The number of runtime copies of the model container that are currently deployed.

Type: Integer

Valid Range: Minimum value of 0.

Required: No

DesiredCopyCount

The number of runtime copies of the model container that you requested to deploy with the inference component.

Type: Integer

Valid Range: Minimum value of 0.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
InferenceComponentSpecification

Service: Amazon SageMaker Service

Details about the resources to deploy with this inference component, including the model, container, and compute resources.

Contents

ComputeResourceRequirements

The compute resources allocated to run the model assigned to the inference component.

Type: InferenceComponentComputeResourceRequirements (p. 1577) object

Required: Yes

Container

Defines a container that provides the runtime environment for a model that you deploy with an inference component.

Type: InferenceComponentContainerSpecification (p. 1578) object

Required: No

ModelName

The name of an existing SageMaker model object in your account that you want to deploy with the inference component.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])* 

Required: No

StartupParameters

Settings that take effect while the model container starts up.

Type: InferenceComponentStartupParameters (p. 1586) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
InferenceComponentSpecificationSummary
Service: Amazon SageMaker Service

Details about the resources that are deployed with this inference component.

Contents

ComputeResourceRequirements

The compute resources allocated to run the model assigned to the inference component.

Type: InferenceComponentComputeResourceRequirements (p. 1577) object

Required: No

Container

Details about the container that provides the runtime environment for the model that is deployed with the inference component.

Type: InferenceComponentContainerSpecificationSummary (p. 1580) object

Required: No

ModelName

The name of the SageMaker model object that is deployed with the inference component.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]*)*

Required: No

StartupParameters

Settings that take effect while the model container starts up.

Type: InferenceComponentStartupParameters (p. 1586) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
InferenceComponentStartupParameters
Service: Amazon SageMaker Service

Settings that take effect while the model container starts up.

Contents

ContainerStartupHealthCheckTimeoutInSeconds
The timeout value, in seconds, for your inference container to pass health check by Amazon S3 Hosting. For more information about health check, see How Your Container Should Respond to Health Check (Ping) Requests.

Type: Integer

Valid Range: Minimum value of 60. Maximum value of 3600.

Required: No

ModelDataDownloadTimeoutInSeconds
The timeout value, in seconds, to download and extract the model that you want to host from Amazon S3 to the individual inference instance associated with this inference component.

Type: Integer

Valid Range: Minimum value of 60. Maximum value of 3600.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
InferenceComponentSummary

Service: Amazon SageMaker Service

A summary of the properties of an inference component.

Contents

**CreationTime**

The time when the inference component was created.

Type: Timestamp

Required: Yes

**EndpointArn**

The Amazon Resource Name (ARN) of the endpoint that hosts the inference component.

Type: String


Pattern: `arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:endpoint/.*`

Required: Yes

**EndpointName**

The name of the endpoint that hosts the inference component.

Type: String

Length Constraints: Maximum length of 63.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9])?\{0,62}`

Required: Yes

**InferenceComponentArn**

The Amazon Resource Name (ARN) of the inference component.

Type: String


Required: Yes

**InferenceComponentName**

The name of the inference component.

Type: String

Length Constraints: Maximum length of 63.

Pattern: `^[a-zA-Z0-9]\([-a-zA-Z0-9]*[a-zA-Z0-9]\){0,62}`

Required: Yes

**LastModifiedTime**

The time when the inference component was last updated.
InferenceComponentSummary

Type: Timestamp
Required: Yes

**VariantName**

The name of the production variant that hosts the inference component.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

Required: Yes

**InferenceComponentStatus**

The status of the inference component.

Type: String

Valid Values: InService | Creating | Updating | Failed | Deleting

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
InferenceExecutionConfig
Service: Amazon SageMaker Service

Specifies details about how containers in a multi-container endpoint are run.

Contents

Mode

- How containers in a multi-container are run. The following values are valid.
  - SERIAL - Containers run as a serial pipeline.
  - DIRECT - Only the individual container that you specify is run.

Type: String

Valid Values: Serial | Direct

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
InferenceExperimentDataStorageConfig
Service: Amazon SageMaker Service

The Amazon S3 location and configuration for storing inference request and response data.

Contents

Destination
The Amazon S3 bucket where the inference request and response data is stored.
Type: String
Length Constraints: Maximum length of 512.
Pattern: \^((https|s3):///([^/])/*\s+[^/])$/
Required: Yes

ContentType
Configuration specifying how to treat different headers. If no headers are specified Amazon SageMaker will by default base64 encode when capturing the data.

Type: CaptureContentTypeHeader (p. 1319) object
Required: No

KmsKey

The AWS Key Management Service key that Amazon SageMaker uses to encrypt captured data at rest using Amazon S3 server-side encryption.

Type: String
Length Constraints: Maximum length of 2048.
Pattern: ^.*$
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
InferenceExperimentSchedule

Service: Amazon SageMaker Service

The start and end times of an inference experiment.

The maximum duration that you can set for an inference experiment is 30 days.

Contents

EndTime

The timestamp at which the inference experiment ended or will end.

Type: Timestamp

Required: No

StartTime

The timestamp at which the inference experiment started or will start.

Type: Timestamp

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
**InferenceExperimentSummary**

Service: Amazon SageMaker Service

Lists a summary of properties of an inference experiment.

**Contents**

**CreationTime**

The timestamp at which the inference experiment was created.

Type: Timestamp

Required: Yes

**LastModifiedTime**

The timestamp when you last modified the inference experiment.

Type: Timestamp

Required: Yes

**Name**

The name of the inference experiment.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,119}$

Required: Yes

**Status**

The status of the inference experiment.

Type: String

Valid Values: Creating | Created | Updating | Running | Starting | Stopping | Completed | Cancelled

Required: Yes

**Type**

The type of the inference experiment.

Type: String

Valid Values: ShadowMode

Required: Yes

**CompletionTime**

The timestamp at which the inference experiment was completed.

Type: Timestamp

Required: No
Description

The description of the inference experiment.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: .*

Required: No

RoleArn

The ARN of the IAM role that Amazon SageMaker can assume to access model artifacts and container images, and manage Amazon SageMaker Inference endpoints for model deployment.

Type: String


Pattern: ^arn:aws[a-z\-]*:iam::\d{12}:role/?[a-zA-Z_0-9+=,.@\-_]+$

Required: No

Schedule

The duration for which the inference experiment ran or will run.

The maximum duration that you can set for an inference experiment is 30 days.

Type: InferenceExperimentSchedule (p. 1591) object

Required: No

StatusReason

The error message for the inference experiment status result.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: .*

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
**InferenceMetrics**
Service: Amazon SageMaker Service

The metrics for an existing endpoint compared in an Inference Recommender job.

**Contents**

**MaxInvocations**

The expected maximum number of requests per minute for the instance.

Type: Integer

Required: Yes

**ModelLatency**

The expected model latency at maximum invocations per minute for the instance.

Type: Integer

Required: Yes

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
InferenceRecommendation
Service: Amazon SageMaker Service
A list of recommendations made by Amazon SageMaker Inference Recommender.

Contents

EndpointConfiguration
Defines the endpoint configuration parameters.
Type: EndpointOutputConfiguration (p. 1464) object
Required: Yes

Metrics
The metrics used to decide what recommendation to make.
Type: RecommendationMetrics (p. 1900) object
Required: Yes

ModelConfiguration
Defines the model configuration.
Type: ModelConfiguration (p. 1676) object
Required: Yes

InvocationEndTime
A timestamp that shows when the benchmark completed.
Type: Timestamp
Required: No

InvocationStartTime
A timestamp that shows when the benchmark started.
Type: Timestamp
Required: No

RecommendationId
The recommendation ID which uniquely identifies each recommendation.
Type: String
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
• *AWS SDK for Ruby V3*
InferenceRecommendationsJob
Service: Amazon SageMaker Service
A structure that contains a list of recommendation jobs.

Contents

CreationTime
A timestamp that shows when the job was created.
Type: Timestamp
Required: Yes

JobArn
The Amazon Resource Name (ARN) of the recommendation job.
Type: String
Length Constraints: Maximum length of 256.
Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:inference-recommendations-job/.*
Required: Yes

JobDescription
The job description.
Type: String
Length Constraints: Maximum length of 128.
Required: Yes

JobName
The name of the job.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 64.
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,63
Required: Yes

JobType
The recommendation job type.
Type: String
Valid Values: Default | Advanced
Required: Yes

LastModifiedTime
A timestamp that shows when the job was last modified.
InferenceRecommendationsJob

Type: Timestamp
Required: Yes

RoleArn
The Amazon Resource Name (ARN) of an IAM role that enables Amazon SageMaker to perform tasks on your behalf.

Type: String


Pattern: ^arn:aws[a-z\-]*:iam::\d{12}:role/?[a-zA-Z_0-9+=,.@\-_]/+$
Required: Yes

Status
The status of the job.

Type: String

Valid Values: PENDING | IN_PROGRESS | COMPLETED | FAILED | STOPPING | STOPPED
Required: Yes

CompletionTime
A timestamp that shows when the job completed.

Type: Timestamp
Required: No

FailureReason
If the job fails, provides information why the job failed.

Type: String

Length Constraints: Maximum length of 1024.
Required: No

ModelName
The name of the created model.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])*
Required: No

ModelPackageVersionArn
The Amazon Resource Name (ARN) of a versioned model package.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: ^arn:aws(-cn|-us-gov):sagemaker:[a-zA-Z0-9\-]{9,16}:[0-9]{12}:model-package/[\S]{1,2048}$
SamplePayloadUrl

The Amazon Simple Storage Service (Amazon S3) path where the sample payload is stored. This path must point to a single gzip compressed tar archive (.tar.gz suffix).

Type: String

Length Constraints: Maximum length of 1024.

Pattern: ^(https|s3)://([^/]*/)?(.*)$

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
InferenceRecommendationsJobStep

Service: Amazon SageMaker Service

A returned array object for the Steps response field in the ListInferenceRecommendationsJobSteps API command.

Contents

JobName

The name of the Inference Recommender job.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 64.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]{0,63})

Required: Yes

Status

The current status of the benchmark.

Type: String

Valid Values: PENDING | IN_PROGRESS | COMPLETED | FAILED | STOPPING | STOPPED

Required: Yes

StepType

The type of the subtask.

BENCHMARK: Evaluate the performance of your model on different instance types.

Type: String

Valid Values: BENCHMARK

Required: Yes

InferenceBenchmark

The details for a specific benchmark.

Type: RecommendationJobInferenceBenchmark (p. 1890) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
InferenceSpecification
Service: Amazon SageMaker Service

Defines how to perform inference generation after a training job is run.

Contents

Containers

The Amazon ECR registry path of the Docker image that contains the inference code.

Type: Array of ModelPackageContainerDefinition objects

Array Members: Minimum number of 1 item. Maximum number of 15 items.

Required: Yes

SupportedContentTypes

The supported MIME types for the input data.

Type: Array of strings

Length Constraints: Maximum length of 256.

Pattern: .*

Required: No

SupportedRealtimeInferenceInstanceTypes

A list of the instance types that are used to generate inferences in real-time.

This parameter is required for unversioned models, and optional for versioned models.

Type: Array of strings

Valid Values: ml.t2.medium | ml.t2.large | ml.t2.xlarge | ml.t2.2xlarge | ml.m4.xlarge | ml.m4.2xlarge | ml.m4.4xlarge | ml.m4.10xlarge | ml.m4.16xlarge | ml.m5.large | ml.m5.xlarge | ml.m5.2xlarge | ml.m5.4xlarge | ml.m5.12xlarge | ml.m5.24xlarge | ml.m5d.large | ml.m5d.xlarge | ml.m5d.2xlarge | ml.m5d.4xlarge | ml.m5d.12xlarge | ml.m5d.24xlarge | ml.c4.large | ml.c4.xlarge | ml.c4.2xlarge | ml.c4.4xlarge | ml.c4.8xlarge | ml.p2.xlarge | ml.p2.8xlarge | ml.p2.16xlarge | ml.p3.2xlarge | ml.p3.8xlarge | ml.p3.16xlarge | ml.c5.large | ml.c5.xlarge | ml.c5.2xlarge | ml.c5.4xlarge | ml.c5.9xlarge | ml.c5.18xlarge | ml.c5d.large | ml.c5d.xlarge | ml.c5d.2xlarge | ml.c5d.4xlarge | ml.c5d.9xlarge | ml.c5d.18xlarge | ml.g4dn.xlarge | ml.g4dn.2xlarge | ml.g4dn.4xlarge | ml.g4dn.8xlarge | ml.g4dn.12xlarge | ml.g4dn.16xlarge | ml.r5.large | ml.r5.xlarge | ml.r5.2xlarge | ml.r5.4xlarge | ml.r5.12xlarge | ml.r5.24xlarge | ml.r5d.large | ml.r5d.xlarge | ml.r5d.2xlarge | ml.r5d.4xlarge | ml.r5d.12xlarge | ml.r5d.24xlarge | ml.inf1.xlarge | ml.inf1.2xlarge | ml.inf1.6xlarge | ml.inf1.24xlarge | ml.c6i.large | ml.c6i.xlarge | ml.c6i.2xlarge | ml.c6i.4xlarge | ml.c6i.8xlarge | ml.c6i.12xlarge | ml.c6i.16xlarge | ml.c6i.24xlarge | ml.c6i.32xlarge | ml.g5.large | ml.g5.xlarge | ml.g5.2xlarge | ml.g5.4xlarge | ml.g5.8xlarge | ml.g5.12xlarge | ml.g5.16xlarge | ml.g5.24xlarge | ml.g5.48xlarge | ml.p4d.2xlarge | ml.p4d.4xlarge | ml.c7.large | ml.c7.xlarge | ml.c7g.large | ml.c7g.xlarge | ml.c7g.4xlarge | ml.c7g.8xlarge | ml.c7g.12xlarge | ml.c7g.16xlarge | ml.m6g.large | ml.m6g.xlarge | ml.m6g.2xlarge | ml.m6g.4xlarge |
ml.m6g.8xlarge | ml.m6g.12xlarge | ml.m6g.16xlarge | ml.m6gd.large | ml.m6gd.xlarge | ml.m6gd.2xlarge | ml.m6gd.4xlarge | ml.m6gd.8xlarge | ml.m6gd.12xlarge | ml.m6gd.16xlarge | ml.m6g.xlarge | ml.m6g.2xlarge | ml.m6g.4xlarge | ml.m6g.8xlarge | ml.m6g.12xlarge | ml.m6g.16xlarge | ml.c6g.large | ml.c6g.xlarge | ml.c6g.2xlarge | ml.c6g.4xlarge | ml.c6g.8xlarge | ml.c6g.12xlarge | ml.c6g.16xlarge | ml.c6gd.large | ml.c6gd.xlarge | ml.c6gd.2xlarge | ml.c6gd.4xlarge | ml.c6gd.8xlarge | ml.c6gd.12xlarge | ml.c6gd.16xlarge | ml.c6gn.large | ml.c6gn.xlarge | ml.c6gn.2xlarge | ml.c6gn.4xlarge | ml.c6gn.8xlarge | ml.c6gn.12xlarge | ml.r6g.xlarge | ml.r6g.2xlarge | ml.r6g.4xlarge | ml.r6g.8xlarge | ml.r6g.12xlarge | ml.r6g.16xlarge | ml.r6gd.large | ml.r6gd.xlarge | ml.r6gd.2xlarge | ml.r6gd.4xlarge | ml.r6gd.8xlarge | ml.r6gd.12xlarge | ml.r6gd.16xlarge | ml.p4de.24xlarge | ml.trn1.2xlarge | ml.trn1.32xlarge | ml.inf2.xlarge | ml.inf2.8xlarge | ml.inf2.24xlarge | ml.inf2.48xlarge | ml.p5.48xlarge

Required: No

**SupportedResponseMIMETypes**

The supported MIME types for the output data.

Type: Array of strings

Length Constraints: Maximum length of 1024.

Pattern: ^[-\w]+/\.+$

Required: No

**SupportedTransformInstanceTypes**

A list of the instance types on which a transformation job can be run or on which an endpoint can be deployed.

This parameter is required for unversioned models, and optional for versioned models.

Type: Array of strings

Array Members: Minimum number of 1 item.

Valid Values: ml.m4.xlarge | ml.m4.2xlarge | ml.m4.4xlarge | ml.m4.10xlarge | ml.m4.16xlarge | ml.c4.xlarge | ml.c4.2xlarge | ml.c4.4xlarge | ml.c4.8xlarge | ml.p2.xlarge | ml.p2.8xlarge | ml.p2.16xlarge | ml.p2.2xlarge | ml.p3.8xlarge | ml.p3.16xlarge | ml.c5.xlarge | ml.c5.2xlarge | ml.c5.4xlarge | ml.c5.9xlarge | ml.c5.18xlarge | ml.c5.16xlarge | ml.m5.xlarge | ml.m5.2xlarge | ml.m5.4xlarge | ml.m5.8xlarge | ml.m5.12xlarge | ml.m5.24xlarge | ml.g4dn.xlarge | ml.g4dn.2xlarge | ml.g4dn.4xlarge | ml.g4dn.8xlarge | ml.g4dn.12xlarge | ml.g4dn.16xlarge

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
InfraCheckConfig
Service: Amazon SageMaker Service

Configuration information for the infrastructure health check of a training job. A SageMaker-provided health check tests the health of instance hardware and cluster network connectivity.

Contents

EnableInfraCheck
- Enables an infrastructure health check.
  - Type: Boolean
  - Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
InputConfig

Service: Amazon SageMaker Service

Contains information about the location of input model artifacts, the name and shape of the expected data inputs, and the framework in which the model was trained.

Contents

Framework

Identifies the framework in which the model was trained. For example: TENSORFLOW.

Type: String

Valid Values: TENSORFLOW | KERAS | MXNET | ONNX | PYTORCH | XGBOOST | TFLITE | DARKNET | SKLEARN

Required: Yes

S3Uri

The S3 path where the model artifacts, which result from model training, are stored. This path must point to a single gzip compressed tar archive (.tar.gz suffix).

Type: String

Length Constraints: Maximum length of 1024.

Pattern: ^(https|s3)//([^/]+)/(.*)$

Required: Yes

DataInputConfig

Specifies the name and shape of the expected data inputs for your trained model with a JSON dictionary form. The data inputs are Framework specific.

- TensorFlow: You must specify the name and shape (NHWC format) of the expected data inputs using a dictionary format for your trained model. The dictionary formats required for the console and CLI are different.
  - Examples for one input:
    - If using the console, {"input": [1,1024,1024,3]}
    - If using the CLI, {"input": [1,1024,1024,3]}
  - Examples for two inputs:
    - If using the console, {"data1": [1,28,28,1], "data2": [1,28,28,1]}
    - If using the CLI, {"data1": [1,28,28,1], "data2": [1,28,28,1]}

- Keras: You must specify the name and shape (NCHW format) of expected data inputs using a dictionary format for your trained model. Note that while Keras model artifacts should be uploaded in NHWC (channel-last) format, DataInputConfig should be specified in NCHW (channel-first) format. The dictionary formats required for the console and CLI are different.
  - Examples for one input:
    - If using the console, {"input_1": [1,3,224,224]}
    - If using the CLI, {"input_1": [1,3,224,224]}
  - Examples for two inputs:
    - If using the console, {"input_1": [1,3,224,224], "input_2": [1,3,224,224]}
    - If using the CLI, {"input_1": [1,3,224,224], "input_2": [1,3,224,224]}
- **MXNET/ONNX/DARKNET**: You must specify the name and shape (NCHW format) of the expected data inputs in order using a dictionary format for your trained model. The dictionary formats required for the console and CLI are different.
  - Examples for one input:
    - If using the console, `{"data":[1,3,1024,1024]}
    - If using the CLI, `{"data":[1,3,1024,1024]}
  - Examples for two inputs:
    - If using the console, `{"var1": [1,1,28,28], "var2": [1,1,28,28]}
    - If using the CLI, `{"var1": [1,1,28,28], "var2": [1,1,28,28]}
- **PyTorch**: You can either specify the name and shape (NCHW format) of expected data inputs in order using a dictionary format for your trained model or you can specify the shape only using a list format. The dictionary formats required for the console and CLI are different. The list formats for the console and CLI are the same.
  - Examples for one input in dictionary format:
    - If using the console, `{"input0": [1,3,224,224]}
    - If using the CLI, `{"input0": [1,3,224,224]}
  - Example for one input in list format: [[1,3,224,224]]
  - Examples for two inputs in dictionary format:
    - If using the console, `{"input0": [1,3,224,224], "input1": [1,3,224,224]}
    - If using the CLI, `{"input0": [1,3,224,224], "input1": [1,3,224,224]}
  - Example for two inputs in list format: [[1,3,224,224], [1,3,224,224]]
- **XGBOOST**: input data name and shape are not needed.

**DataInputConfig** supports the following parameters for CoreML TargetDevice (ML Model format):
- **shape**: Input shape, for example `{"input_1": {"shape": [1,224,224,3]}}`. In addition to static input shapes, CoreML converter supports Flexible input shapes:
  - **Range Dimension**: You can use the Range Dimension feature if you know the input shape will be within some specific interval in that dimension, for example: `{"input_1": {"shape": ["1..10", 224, 224, 3]}}
  - **Enumerated shapes**: Sometimes, the models are trained to work only on a select set of inputs. You can enumerate all supported input shapes, for example: `{"input_1": {"shape": [[1, 224, 224, 3], [1, 160, 160, 3]]}}
- **default_shape**: Default input shape. You can set a default shape during conversion for both Range Dimension and Enumerated Shapes. For example `{"input_1": {"shape": ["1..10", 224, 224, 3], "default_shape": [1, 224, 224, 3]}}
- **type**: Input type. Allowed values: Image and Tensor. By default, the converter generates an ML Model with inputs of type Tensor (MultiArray). User can set input type to be Image. Image input type requires additional input parameters such as bias and scale.
- **bias**: If the input type is an Image, you need to provide the bias vector.
- **scale**: If the input type is an Image, you need to provide a scale factor.

CoreML ClassifierConfig parameters can be specified using **OutputConfig** CompilerOptions. CoreML converter supports Tensorflow and PyTorch models. CoreML conversion examples:
- **Tensor type input**:
  - "DataInputConfig": "{"input_1": {"shape": [[1,224,224,3], [1,160,160,3]], "default_shape": [1,224,224,3]}"
- **Tensor type input without input name (PyTorch)**:
  - "DataInputConfig": "{"shape": [[1,3,224,224], [1,3,160,160]], "default_shape": [1,3,224,224]}"
• Image type input:
  • "DataInputConfig": {
    "input_1": {
      "shape": [[1,224,224,3], [1,160,160,3]],
      "default_shape": [1,224,224,3], "type": "Image", "bias": [-1,-1,-1],
      "scale": 0.007843137255}
    }
  }
  "CompilerOptions": {
    "class_labels": "imagenet_labels_1000.txt"
  }

• Image type input without input name (PyTorch):
  • "DataInputConfig": [
    {
      "shape": [[1,3,224,224], [1,3,160,160]],
      "default_shape": [1,3,224,224], "type": "Image", "bias": [-1,-1,-1],
      "scale": 0.007843137255}
    ]
  }
  "CompilerOptions": {
    "class_labels": "imagenet_labels_1000.txt"
  }

Depending on the model format, DataInputConfig requires the following parameters for ml_eia2 OutputConfig:TargetDevice.

• For TensorFlow models saved in the SavedModel format, specify the input names from signature_def_key and the input model shapes for DataInputConfig. Specify the signature_def_key in OutputConfig:CompilerOptions if the model does not use TensorFlow's default signature def key. For example:
  • "DataInputConfig": {
    "inputs": [1, 224, 224, 3]
  }
  "CompilerOptions": {
    "signature_def_key": "serving_custom"
  }

• For TensorFlow models saved as a frozen graph, specify the input tensor names and shapes in DataInputConfig and the output tensor names for output_names in OutputConfig:CompilerOptions. For example:
  • "DataInputConfig": {
    "input_tensor:0": [1, 224, 224, 3]
  }
  "CompilerOptions": {
    "output_names": ["output_tensor:0"]
  }

Type: String


Pattern: [\S\s]+

Required: No

FrameworkVersion

Specifies the framework version to use. This API field is only supported for the MXNet, PyTorch, TensorFlow and TensorFlow Lite frameworks.

For information about framework versions supported for cloud targets and edge devices, see Cloud Supported Instance Types and Frameworks and Edge Supported Frameworks.

Type: String


Pattern: [0-9]\.[A-Za-z0-9.]+

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

• AWS SDK for C++
• AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
**InstanceGroup**

Service: Amazon SageMaker Service

Defines an instance group for heterogeneous cluster training. When requesting a training job using the [CreateTrainingJob](#) API, you can configure multiple instance groups.

**Contents**

**InstanceCount**

Specifies the number of instances of the instance group.

Type: Integer

Valid Range: Minimum value of 0.

Required: Yes

**InstanceGroupName**

Specifies the name of the instance group.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 64.

Pattern: .+

Required: Yes

**InstanceType**

Specifies the instance type of the instance group.

Type: String

Valid Values: ml.m4.xlarge | ml.m4.2xlarge | ml.m4.4xlarge | ml.m4.10xlarge | ml.m4.16xlarge | ml.m4.24xlarge | ml.m4.48xlarge | ml.m4.96xlarge | ml.m4.128xlarge | ml.m4.256xlarge | ml.m4.512xlarge | ml.m4.1024xlarge | ml.m4.2048xlarge | ml.m4.4096xlarge | ml.m4.8192xlarge | ml.m4.16384xlarge | ml.m4.32768xlarge | ml.m4.65536xlarge | ml.m4.131072xlarge | ml.m5.large | ml.m5.xlarge | ml.m5.2xlarge | ml.m5.4xlarge | ml.m5.8xlarge | ml.m5.16xlarge | ml.m5.32xlarge | ml.m5.64xlarge | ml.m5.128xlarge | ml.m5.256xlarge | ml.m5.512xlarge | ml.m5.1024xlarge | ml.m5.2048xlarge | ml.m5.4096xlarge | ml.m5.8192xlarge | ml.m5.16384xlarge | ml.m5.32768xlarge | ml.m5.65536xlarge | ml.m5.131072xlarge | ml.g4dn.xlarge | ml.g4dn.2xlarge | ml.g4dn.4xlarge | ml.g4dn.8xlarge | ml.g4dn.16xlarge | ml.g4dn.32xlarge | ml.g4dn.64xlarge | ml.g4dn.96xlarge | ml.g4dn.128xlarge | ml.g4dn.256xlarge | ml.g4dn.512xlarge | ml.g4dn.1024xlarge | ml.g4dn.2048xlarge | ml.g4dn.4096xlarge | ml.g4dn.8192xlarge | ml.g4dn.16384xlarge | ml.g4dn.32768xlarge | ml.g4dn.65536xlarge | ml.g4dn.131072xlarge | ml.g4n.xlarge | ml.g4n.2xlarge | ml.g4n.4xlarge | ml.g4n.8xlarge | ml.g4n.16xlarge | ml.g4n.32xlarge | ml.g4n.64xlarge | ml.g4n.96xlarge | ml.g4n.128xlarge | ml.g4n.256xlarge | ml.g4n.512xlarge | ml.g4n.1024xlarge | ml.g4n.2048xlarge | ml.g4n.4096xlarge | ml.g4n.8192xlarge | ml.g4n.16384xlarge | ml.g4n.32768xlarge | ml.g4n.65536xlarge | ml.g4n.131072xlarge | ml.c4.xlarge | ml.c4.2xlarge | ml.c4.4xlarge | ml.c4.8xlarge | ml.c4.16xlarge | ml.c4.32xlarge | ml.c4.64xlarge | ml.c4.128xlarge | ml.c4.256xlarge | ml.c4.512xlarge | ml.c4.1024xlarge | ml.c4.2048xlarge | ml.c4.4096xlarge | ml.c4.8192xlarge | ml.c4.16384xlarge | ml.c4.32768xlarge | ml.c4.65536xlarge | ml.c4.131072xlarge | ml.c5.xlarge | ml.c5.2xlarge | ml.c5.4xlarge | ml.c5.8xlarge | ml.c5.16xlarge | ml.c5.32xlarge | ml.c5.64xlarge | ml.c5.128xlarge | ml.c5.256xlarge | ml.c5.512xlarge | ml.c5.1024xlarge | ml.c5.2048xlarge | ml.c5.4096xlarge | ml.c5.8192xlarge | ml.c5.16384xlarge | ml.c5.32768xlarge | ml.c5.65536xlarge | ml.c5.131072xlarge | ml.c5n.xlarge | ml.c5n.2xlarge | ml.c5n.4xlarge | ml.c5n.8xlarge | ml.c5n.16xlarge | ml.c5n.32xlarge | ml.c5n.64xlarge | ml.c5n.128xlarge | ml.c5n.256xlarge | ml.c5n.512xlarge | ml.c5n.1024xlarge | ml.c5n.2048xlarge | ml.c5n.4096xlarge | ml.c5n.8192xlarge | ml.c5n.16384xlarge | ml.c5n.32768xlarge | ml.c5n.65536xlarge | ml.c5n.131072xlarge | ml.p2.xlarge | ml.p2.8xlarge | ml.p2.16xlarge | ml.p2.32xlarge | ml.p2.64xlarge | ml.p2.128xlarge | ml.p2.256xlarge | ml.p2.512xlarge | ml.p2.1024xlarge | ml.p2.2048xlarge | ml.p2.4096xlarge | ml.p2.8192xlarge | ml.p2.16384xlarge | ml.p2.32768xlarge | ml.p2.65536xlarge | ml.p2.131072xlarge | ml.p3.xlarge | ml.p3.8xlarge | ml.p3.16xlarge | ml.p3.32xlarge | ml.p3.64xlarge | ml.p3.128xlarge | ml.p3.256xlarge | ml.p3.512xlarge | ml.p3.1024xlarge | ml.p3.2048xlarge | ml.p3.4096xlarge | ml.p3.8192xlarge | ml.p3.16384xlarge | ml.p3.32768xlarge | ml.p3.65536xlarge | ml.p3.131072xlarge | ml.p4.xlarge | ml.p4.8xlarge | ml.p4.16xlarge | ml.p4.32xlarge | ml.p4.64xlarge | ml.p4.128xlarge | ml.p4.256xlarge | ml.p4.512xlarge | ml.p4.1024xlarge | ml.p4.2048xlarge | ml.p4.4096xlarge | ml.p4.8192xlarge | ml.p4.16384xlarge | ml.p4.32768xlarge | ml.p4.65536xlarge | ml.p4.131072xlarge | ml.trn1.xlarge | ml.trn1.2xlarge | ml.trn1.4xlarge | ml.trn1.8xlarge | ml.trn1.16xlarge | ml.trn1.32xlarge | ml.trn1n.32xlarge | ml.p3dn.2xlarge | ml.p3dn.4xlarge | ml.p3dn.8xlarge | ml.p3dn.16xlarge | ml.p3dn.32xlarge | ml.p3dn.64xlarge | ml.p3dn.128xlarge | ml.p3dn.256xlarge | ml.p3dn.512xlarge | ml.p3dn.1024xlarge | ml.p3dn.2048xlarge | ml.p3dn.4096xlarge | ml.p3dn.8192xlarge | ml.p3dn.16384xlarge | ml.p3dn.32768xlarge | ml.p3dn.65536xlarge | ml.p3dn.131072xlarge | ml.g5.xlarge | ml.g5.2xlarge | ml.g5.4xlarge | ml.g5.8xlarge | ml.g5.16xlarge | ml.g5.32xlarge | ml.g5.64xlarge | ml.g5.128xlarge | ml.g5.256xlarge | ml.g5.512xlarge | ml.g5.1024xlarge | ml.g5.2048xlarge | ml.g5.4096xlarge | ml.g5.8192xlarge | ml.g5.16384xlarge | ml.g5.32768xlarge | ml.g5.65536xlarge | ml.g5.131072xlarge | ml.g5n.xlarge | ml.g5n.2xlarge | ml.g5n.4xlarge | ml.g5n.8xlarge | ml.g5n.16xlarge | ml.g5n.32xlarge | ml.g5n.64xlarge | ml.g5n.128xlarge | ml.g5n.256xlarge | ml.g5n.512xlarge | ml.g5n.1024xlarge | ml.g5n.2048xlarge | ml.g5n.4096xlarge | ml.g5n.8192xlarge | ml.g5n.16384xlarge | ml.g5n.32768xlarge | ml.g5n.65536xlarge | ml.g5n.131072xlarge | ml.xlarge | ml.2xlarge | ml.4xlarge | ml.8xlarge | ml.16xlarge | ml.32xlarge | ml.64xlarge | ml.128xlarge | ml.256xlarge | ml.512xlarge | ml.1024xlarge | ml.2048xlarge | ml.4096xlarge | ml.8192xlarge | ml.16384xlarge | ml.32768xlarge | ml.65536xlarge | ml.131072xlarge

Required: Yes

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- AWS SDK for Java V2
- AWS SDK for Ruby V3
InstanceMetadataServiceConfiguration
Service: Amazon SageMaker Service

Information on the IMDS configuration of the notebook instance

Contents

MinimumInstanceMetadataServiceVersion

Indicates the minimum IMDS version that the notebook instance supports. When passed as part of CreateNotebookInstance, if no value is selected, then it defaults to IMDSv1. This means that both IMDSv1 and IMDSv2 are supported. If passed as part of UpdateNotebookInstance, there is no default.

Type: String

Length Constraints: Maximum length of 1.

Pattern: 1 | 2

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
**IntegerParameterRange**

Service: Amazon SageMaker Service

For a hyperparameter of the integer type, specifies the range that a hyperparameter tuning job searches.

**Contents**

**MaxValue**

The maximum value of the hyperparameter to search.

Type: String

Length Constraints: Maximum length of 256.

Pattern: . *

Required: Yes

**MinValue**

The minimum value of the hyperparameter to search.

Type: String

Length Constraints: Maximum length of 256.

Pattern: . *

Required: Yes

**Name**

The name of the hyperparameter to search.

Type: String

Length Constraints: Maximum length of 256.

Pattern: . *

Required: Yes

**ScalingType**

The scale that hyperparameter tuning uses to search the hyperparameter range. For information about choosing a hyperparameter scale, see [Hyperparameter Scaling](#). One of the following values:

- **Auto**
  
  SageMaker hyperparameter tuning chooses the best scale for the hyperparameter.

- **Linear**
  
  Hyperparameter tuning searches the values in the hyperparameter range by using a linear scale.

- **Logarithmic**
  
  Hyperparameter tuning searches the values in the hyperparameter range by using a logarithmic scale.

  Logarithmic scaling works only for ranges that have only values greater than 0.

Type: String
Valid Values: Auto | Linear | Logarithmic | ReverseLogarithmic

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
IntegerParameterRangeSpecification

Service: Amazon SageMaker Service

Defines the possible values for an integer hyperparameter.

Contents

MaxValue

The maximum integer value allowed.

Type: String

Length Constraints: Maximum length of 256.

Pattern: .*

Required: Yes

MinValue

The minimum integer value allowed.

Type: String

Length Constraints: Maximum length of 256.

Pattern: .*

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
JupyterLabAppImageConfig
Service: Amazon SageMaker Service

The configuration for the file system and kernels in a SageMaker image running as a JupyterLab app.

Contents

ContainerConfig

The configuration used to run the application image container.

Type: ContainerConfig (p. 1364) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
JupyterLabAppSettings

Service: Amazon SageMaker Service

The settings for the JupyterLab application.

Contents

**CodeRepositories**

A list of Git repositories that SageMaker automatically displays to users for cloning in the JupyterLab application.

Type: Array of [CodeRepository](p. 1354) objects

Array Members: Maximum number of 10 items.

Required: No

**CustomImages**

A list of custom SageMaker images that are configured to run as a JupyterLab app.

Type: Array of [CustomImage](p. 1378) objects

Array Members: Maximum number of 200 items.

Required: No

**DefaultResourceSpec**

Specifies the ARN's of a SageMaker image and SageMaker image version, and the instance type that the version runs on.

Type: [ResourceSpec](p. 1915) object

Required: No

**LifecycleConfigArns**

The Amazon Resource Name (ARN) of the lifecycle configurations attached to the user profile or domain. To remove a lifecycle config, you must set `LifecycleConfigArns` to an empty list.

Type: Array of strings

Length Constraints: Maximum length of 256.

Pattern: `arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:studio-lifecycle-config/.*`

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
JupyterServerAppSettings

Service: Amazon SageMaker Service

The JupyterServer app settings.

Contents

CodeRepositories

A list of Git repositories that SageMaker automatically displays to users for cloning in the JupyterServer application.

Type: Array of CodeRepository (p. 1354) objects

Array Members: Maximum number of 10 items.

Required: No

DefaultResourceSpec

The default instance type and the Amazon Resource Name (ARN) of the default SageMaker image used by the JupyterServer app. If you use the LifecycleConfigArns parameter, then this parameter is also required.

Type: ResourceSpec (p. 1915) object

Required: No

LifecycleConfigArns

The Amazon Resource Name (ARN) of the Lifecycle Configurations attached to the JupyterServerApp. If you use this parameter, the DefaultResourceSpec parameter is also required.

Note
To remove a Lifecycle Config, you must set LifecycleConfigArns to an empty list.

Type: Array of strings

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:studio-lifecycle-config/.*

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
KendraSettings
Service: Amazon SageMaker Service

The Amazon SageMaker Canvas application setting where you configure document querying.

Contents

Status

Describes whether the document querying feature is enabled or disabled in the Canvas application.

Type: String

Valid Values: ENABLED | DISABLED

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
KernelGatewayAppSettings
Service: Amazon SageMaker Service

The KernelGateway app settings.

Contents

CustomImages
A list of custom SageMaker images that are configured to run as a KernelGateway app.

Type: Array of CustomImage (p. 1378) objects

Array Members: Maximum number of 200 items.

Required: No

DefaultResourceSpec
The default instance type and the Amazon Resource Name (ARN) of the default SageMaker image used by the KernelGateway app.

Note
The Amazon SageMaker Studio UI does not use the default instance type value set here. The default instance type set here is used when Apps are created using the AWS Command Line Interface or AWS CloudFormation and the instance type parameter value is not passed.

Type: ResourceSpec (p. 1915) object

Required: No

LifecycleConfigArns
The Amazon Resource Name (ARN) of the Lifecycle Configurations attached to the the user profile or domain.

Note
To remove a Lifecycle Config, you must set LifecycleConfigArns to an empty list.

Type: Array of strings

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:studio-lifecycle-config/.*

Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
KernelGatewayImageConfig

Service: Amazon SageMaker Service

The configuration for the file system and kernels in a SageMaker image running as a KernelGateway app.

Contents

KernelSpecs

The specification of the Jupyter kernels in the image.

Type: Array of KernelSpec (p. 1622) objects

Array Members: Fixed number of 1 item.

Required: Yes

FileSystemConfig

The Amazon Elastic File System (EFS) storage configuration for a SageMaker image.

Type: FileSystemConfig (p. 1493) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
KernelSpec
Service: Amazon SageMaker Service
The specification of a Jupyter kernel.

Contents

Name
The name of the Jupyter kernel in the image. This value is case sensitive.
Type: String
Length Constraints: Maximum length of 1024.
Required: Yes

DisplayName
The display name of the kernel.
Type: String
Length Constraints: Maximum length of 1024.
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
LabelCounters
Service: Amazon SageMaker Service
Provides a breakdown of the number of objects labeled.

Contents

FailedNonRetryableError
The total number of objects that could not be labeled due to an error.
Type: Integer
Valid Range: Minimum value of 0.
Required: No

HumanLabeled
The total number of objects labeled by a human worker.
Type: Integer
Valid Range: Minimum value of 0.
Required: No

MachineLabeled
The total number of objects labeled by automated data labeling.
Type: Integer
Valid Range: Minimum value of 0.
Required: No

TotalLabeled
The total number of objects labeled.
Type: Integer
Valid Range: Minimum value of 0.
Required: No

Unlabeled
The total number of objects not yet labeled.
Type: Integer
Valid Range: Minimum value of 0.
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for Ruby V3
LabelCountersForWorkteam

Service: Amazon SageMaker Service

Provides counts for human-labeled tasks in the labeling job.

Contents

HumanLabeled

The total number of data objects labeled by a human worker.

Type: Integer
Valid Range: Minimum value of 0.
Required: No

PendingHuman

The total number of data objects that need to be labeled by a human worker.

Type: Integer
Valid Range: Minimum value of 0.
Required: No

Total

The total number of tasks in the labeling job.

Type: Integer
Valid Range: Minimum value of 0.
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
LabelingJobAlgorithmsConfig

Service: Amazon SageMaker Service

Provides configuration information for auto-labeling of your data objects. A LabelingJobAlgorithmsConfig object must be supplied in order to use auto-labeling.

Contents

LabelingJobAlgorithmSpecificationArn

Specifies the Amazon Resource Name (ARN) of the algorithm used for auto-labeling. You must select one of the following ARNs:

- **Image classification**
  

- **Text classification**
  

- **Object detection**
  

- **Semantic Segmentation**
  

Type: String

Length Constraints: Maximum length of 2048.

Pattern: `arn:*`

Required: Yes

InitialActiveLearningModelArn

At the end of an auto-label job Ground Truth sends the Amazon Resource Name (ARN) of the final model used for auto-labeling. You can use this model as the starting point for subsequent similar jobs by providing the ARN of the model here.

Type: String


Pattern: `arn:aws[a-z-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:model/.*`

Required: No

LabelingJobResourceConfig

Provides configuration information for a labeling job.

Type: `LabelingJobResourceConfig (p. 1636)` object

Required: No
See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
LabelingJobDataAttributes
Service: Amazon SageMaker Service

Attributes of the data specified by the customer. Use these to describe the data to be labeled.

Contents

ContentClassifiers

Declares that your content is free of personally identifiable information or adult content. SageMaker may restrict the Amazon Mechanical Turk workers that can view your task based on this information.

Type: Array of strings

Array Members: Maximum number of 256 items.

Valid Values: FreeOfPersonallyIdentifiableInformation | FreeOfAdultContent

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
LabelingJobDataSource

Service: Amazon SageMaker Service

Provides information about the location of input data.

You must specify at least one of the following: S3DataSource or SnsDataSource.

Use SnsDataSource to specify an SNS input topic for a streaming labeling job. If you do not specify and SNS input topic ARN, Ground Truth will create a one-time labeling job.

Use S3DataSource to specify an input manifest file for both streaming and one-time labeling jobs. Adding an S3DataSource is optional if you use SnsDataSource to create a streaming labeling job.

Contents

S3DataSource

The Amazon S3 location of the input data objects.

Type: LabelingJobS3DataSource (p. 1637) object

Required: No

SnsDataSource

An Amazon SNS data source used for streaming labeling jobs. To learn more, see Send Data to a Streaming Labeling Job.

Type: LabelingJobSnsDataSource (p. 1638) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
LabelingJobForWorkteamSummary
Service: Amazon SageMaker Service
Provides summary information for a work team.

Contents

**CreationTime**
The date and time that the labeling job was created.
Type: Timestamp
Required: Yes

**JobReferenceCode**
A unique identifier for a labeling job. You can use this to refer to a specific labeling job.
Type: String
Length Constraints: Minimum length of 1.
Pattern: .+
Required: Yes

**WorkRequesterAccountId**
The AWS account ID of the account used to start the labeling job.
Type: String
Pattern: ^\d+$
Required: Yes

**LabelCounters**
Provides information about the progress of a labeling job.
Type: `LabelCountersForWorkteam (p. 1625)` object
Required: No

**LabelingJobName**
The name of the labeling job that the work team is assigned to.
Type: String
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}$
Required: No

**NumberOfHumanWorkersPerDataObject**
The configured number of workers per data object.
Type: Integer
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
LabelingJobInputConfig
Service: Amazon SageMaker Service

Input configuration information for a labeling job.

Contents

**DataSource**

The location of the input data.

Type: LabelingJobDataSource (p. 1629) object

Required: Yes

**DataAttributes**

Attributes of the data specified by the customer.

Type: LabelingJobDataAttributes (p. 1628) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
LabelingJobOutput
Service: Amazon SageMaker Service

Specifies the location of the output produced by the labeling job.

Contents

OutputDatasetS3Uri

The Amazon S3 bucket location of the manifest file for labeled data.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: ^(https|s3)://([^/]+)?(\*)$

Required: Yes

FinalActiveLearningModelArn

The Amazon Resource Name (ARN) for the most recent SageMaker model trained as part of automated data labeling.

Type: String


Pattern: arn:aws[a-z\-]*:sagemaker:[a-z\-0-9\-]*:[0-9]{12}:model/\.*

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
LabelingJobOutputConfig

Service: Amazon SageMaker Service

Output configuration information for a labeling job.

Contents

S3OutputPath

The Amazon S3 location to write output data.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: ^(https|s3):/(/[^/]+?)(.*)$

Required: Yes

KmsKeyId

The AWS Key Management Service ID of the key used to encrypt the output data, if any.

If you provide your own KMS key ID, you must add the required permissions to your KMS key described in Encryption Output Data and Storage Volume with AWS KMS.

If you don't provide a KMS key ID, Amazon SageMaker uses the default AWS KMS key for Amazon S3 for your role's account to encrypt your output data.

If you use a bucket policy with an s3:PutObject permission that only allows objects with server-side encryption, set the condition key of s3:x-amz-server-side-encryption to “aws:kms”. For more information, see KMS-Managed Encryption Keys in the Amazon Simple Storage Service Developer Guide.

Type: String

Length Constraints: Maximum length of 2048.

Pattern: .*

Required: No

SnsTopicArn

An Amazon Simple Notification Service (Amazon SNS) output topic ARN. Provide a SnsTopicArn if you want to do real time chaining to another streaming job and receive an Amazon SNS notifications each time a data object is submitted by a worker.

If you provide an SnsTopicArn in OutputConfig, when workers complete labeling tasks, Ground Truth will send labeling task output data to the SNS output topic you specify here.

To learn more, see Receive Output Data from a Streaming Labeling Job.

Type: String

Length Constraints: Maximum length of 2048.

Pattern: arn:aws[a-z\-]*:sns:[a-z0-9\-]*:[0-9]{12}:[a-zA-Z0-9_.-]+

Required: No
See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
LabelingJobResourceConfig

Service: Amazon SageMaker Service

Configure encryption on the storage volume attached to the ML compute instance used to run automated data labeling model training and inference.

Contents

VolumeKmsKeyId

The AWS Key Management Service (AWS KMS) key that Amazon SageMaker uses to encrypt data on the storage volume attached to the ML compute instance(s) that run the training and inference jobs used for automated data labeling.

You can only specify a VolumeKmsKeyId when you create a labeling job with automated data labeling enabled using the API operation CreateLabelingJob. You cannot specify an AWS KMS key to encrypt the storage volume used for automated data labeling model training and inference when you create a labeling job using the console. To learn more, see Output Data and Storage Volume Encryption.

The VolumeKmsKeyId can be any of the following formats:
- KMS Key ID
  "1234abcd-12ab-34cd-56ef-1234567890ab"
- Amazon Resource Name (ARN) of a KMS Key
  "arn:aws:kms:us-west-2:111122223333:key/1234abcd-12ab-34cd-56ef-1234567890ab"

Type: String
Length Constraints: Maximum length of 2048.
Pattern: .*
Required: No

VpcConfig

Specifies an Amazon Virtual Private Cloud (VPC) that your SageMaker jobs, hosted models, and compute resources have access to. You can control access to and from your resources by configuring a VPC. For more information, see Give SageMaker Access to Resources in your Amazon VPC.

Type: VpcConfig (p. 2076) object
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
LabelingJobS3DataSource

The Amazon S3 location of the input data objects.

Contents

ManifestS3Uri

The Amazon S3 location of the manifest file that describes the input data objects.

The input manifest file referenced in ManifestS3Uri must contain one of the following keys: source-ref or source. The value of the keys are interpreted as follows:

- source-ref: The source of the object is the Amazon S3 object specified in the value. Use this value when the object is a binary object, such as an image.
- source: The source of the object is the value. Use this value when the object is a text value.

If you are a new user of Ground Truth, it is recommended you review Use an Input Manifest File in the Amazon SageMaker Developer Guide to learn how to create an input manifest file.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: ^(https|s3)://([^/]+)/(.*)$

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
LabelingJobSnsDataSource

Service: Amazon SageMaker Service

An Amazon SNS data source used for streaming labeling jobs.

Contents

SnsTopicArn

The Amazon SNS input topic Amazon Resource Name (ARN). Specify the ARN of the input topic you will use to send new data objects to a streaming labeling job.

Type: String

Length Constraints: Maximum length of 2048.

Pattern: \*\*\*:sns:\*:\*:\*:\*:\*:\*\*

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
LabelingJobStoppingConditions

Service: Amazon SageMaker Service

A set of conditions for stopping a labeling job. If any of the conditions are met, the job is automatically stopped. You can use these conditions to control the cost of data labeling.

Note
Labeling jobs fail after 30 days with an appropriate client error message.

Contents

MaxHumanLabeledObjectCount

The maximum number of objects that can be labeled by human workers.

Type: Integer
Valid Range: Minimum value of 1.
Required: No

MaxPercentageOfInputDatasetLabeled

The maximum number of input data objects that should be labeled.

Type: Integer
Valid Range: Minimum value of 1. Maximum value of 100.
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
**LabelingJobSummary**  
Service: Amazon SageMaker Service

Provides summary information about a labeling job.

**Contents**

**CreationTime**

The date and time that the job was created (timestamp).

Type: Timestamp

Required: Yes

**LabelCounters**

Counts showing the progress of the labeling job.

Type: LabelCounters object

Required: Yes

**LabelingJobArn**

The Amazon Resource Name (ARN) assigned to the labeling job when it was created.

Type: String

Length Constraints: Maximum length of 2048.

Pattern: `arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:labeling-job/.*`

Required: Yes

**LabelingJobName**

The name of the labeling job.

Type: String


Pattern: `^[a-zA-Z0-9](\-[a-zA-Z0-9])\{0,62}`

Required: Yes

**LabelingJobStatus**

The current status of the labeling job.

Type: String

Valid Values: Initializing | InProgress | Completed | Failed | Stopping | Stopped

Required: Yes

**LastModifiedTime**

The date and time that the job was last modified (timestamp).

Type: Timestamp
**PreHumanTaskLambdaArn**

The Amazon Resource Name (ARN) of a Lambda function. The function is run before each data object is sent to a worker.

Type: String

Length Constraints: Maximum length of 2048.

Pattern: `arn:aws[a-z\-]*:lambda:[a-z0-9\-]*:[0-9]{12}:function:.`

Required: Yes

**WorkteamArn**

The Amazon Resource Name (ARN) of the work team assigned to the job.

Type: String

Length Constraints: Maximum length of 256.

Pattern: `arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:workteam/.*`

Required: Yes

**AnnotationConsolidationLambdaArn**

The Amazon Resource Name (ARN) of the Lambda function used to consolidate the annotations from individual workers into a label for a data object. For more information, see [Annotation Consolidation](#).

Type: String

Length Constraints: Maximum length of 2048.

Pattern: `arn:aws[a-z\-]*:lambda:[a-z0-9\-]*:[0-9]{12}:function:.`

Required: No

**FailureReason**

If the `LabelingJobStatus` field is Failed, this field contains a description of the error.

Type: String

Length Constraints: Maximum length of 1024.

Required: No

**InputConfig**

Input configuration for the labeling job.

Type: `LabelingJobInputConfig (p. 1632)` object

Required: No

**LabelingJobOutput**

The location of the output produced by the labeling job.

Type: `LabelingJobOutput (p. 1633)` object

Required: No
See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
LambdaStepMetadata
Service: Amazon SageMaker Service
Metadata for a Lambda step.

Contents

Arn
The Amazon Resource Name (ARN) of the Lambda function that was run by this step execution.
Type: String
Length Constraints: Maximum length of 256.
Required: No

OutputParameters
A list of the output parameters of the Lambda step.
Type: Array of OutputParameter (p. 1794) objects
Array Members: Minimum number of 0 items. Maximum number of 50 items.
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
**LastUpdateStatus**

*Service: Amazon SageMaker Service*

A value that indicates whether the update was successful.

**Contents**

**Status**

A value that indicates whether the update was made successful.

Type: String

Valid Values: Successful | Failed | InProgress

Required: Yes

**FailureReason**

If the update wasn't successful, indicates the reason why it failed.

Type: String

Length Constraints: Maximum length of 1024.

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
LineageGroupSummary

Service: Amazon SageMaker Service

Lists a summary of the properties of a lineage group. A lineage group provides a group of shareable lineage entity resources.

Contents

CreationTime

The creation time of the lineage group summary.

Type: Timestamp

Required: No

DisplayName

The display name of the lineage group summary.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,119}

Required: No

LastModifiedTime

The last modified time of the lineage group summary.

Type: Timestamp

Required: No

LineageGroupArn

The Amazon Resource Name (ARN) of the lineage group resource.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[\a-z\-]*:sagemaker:\[a-zA-Z0-9\-]*:[0-9]{12}:lineage-group/.*

Required: No

LineageGroupName

The name or Amazon Resource Name (ARN) of the lineage group.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,119}

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:
• AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for Ruby V3
MemberDefinition
Service: Amazon SageMaker Service

Defines an Amazon Cognito or your own OIDC IdP user group that is part of a work team.

Contents

CognitoMemberDefinition

The Amazon Cognito user group that is part of the work team.

Type: CognitoMemberDefinition (p. 1358) object

Required: No

OidcMemberDefinition

A list user groups that exist in your OIDC Identity Provider (IdP). One to ten groups can be used to create a single private work team. When you add a user group to the list of Groups, you can add that user group to one or more private work teams. If you add a user group to a private work team, all workers in that user group are added to the work team.

Type: OidcMemberDefinition (p. 1783) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
**MetadataProperties**

Service: Amazon SageMaker Service

Metadata properties of the tracking entity, trial, or trial component.

**Contents**

**CommitId**

The commit ID.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: . *

Required: No

**GeneratedBy**

The entity this entity was generated by.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: . *

Required: No

**ProjectId**

The project ID.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: . *

Required: No

**Repository**

The repository.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: . *

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for Ruby V3
MetricData

Service: Amazon SageMaker Service

The name, value, and date and time of a metric that was emitted to Amazon CloudWatch.

Contents

MetricName

The name of the metric.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 255.
Pattern: . +
Required: No

Timestamp

The date and time that the algorithm emitted the metric.
Type: Timestamp
Required: No

Value

The value of the metric.
Type: Float
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
MetricDatum
Service: Amazon SageMaker Service

Information about the metric for a candidate produced by an AutoML job.

Contents

MetricName
The name of the metric.

Type: String

Valid Values: Accuracy | MSE | F1 | F1macro | AUC | RMSE | MAE | R2 | BalancedAccuracy | Precision | PrecisionMacro | Recall | RecallMacro | MAPE | MASE | WAPE | AverageWeightedQuantileLoss

Required: No

Set
The dataset split from which the AutoML job produced the metric.

Type: String

Valid Values: Train | Validation | Test

Required: No

StandardMetricName
The name of the standard metric.

Note
For definitions of the standard metrics, see Autopilot candidate metrics.

Type: String

Valid Values: Accuracy | MSE | F1 | F1macro | AUC | RMSE | MAE | R2 | BalancedAccuracy | Precision | PrecisionMacro | Recall | RecallMacro | LogLoss | InferenceLatency | MAPE | MASE | WAPE | AverageWeightedQuantileLoss | Rouge1 | Rouge2 | RougeL | RougeLSum | Perplexity | ValidationLoss | TrainingLoss

Required: No

Value
The value of the metric.

Type: Float

Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for Ruby V3
MetricDefinition

Service: Amazon SageMaker Service

Specifies a metric that the training algorithm writes to stderr or stdout. You can view these logs to understand how your training job performs and check for any errors encountered during training. SageMaker hyperparameter tuning captures all defined metrics. Specify one of the defined metrics to use as an objective metric using the TuningObjective parameter in the HyperParameterTrainingJobDefinition API to evaluate job performance during hyperparameter tuning.

Contents

Name

The name of the metric.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 255.

Pattern: .+

Required: Yes

Regex

A regular expression that searches the output of a training job and gets the value of the metric. For more information about using regular expressions to define metrics, see Defining metrics and environment variables.

Type: String


Pattern: .+

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
MetricSpecification

Service: Amazon SageMaker Service

An object containing information about a metric.

Contents

**Important**
This data type is a UNION, so only one of the following members can be specified when used or returned.

**Customized**

Information about a customized metric.

Type: CustomizedMetricSpecification (p. 1379) object

Required: No

**Predefined**

Information about a predefined metric.

Type: PredefinedMetricSpecification (p. 1828) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
MetricsSource
Service: Amazon SageMaker Service

Details about the metrics source.

Contents

**ContentType**

The metric source content type.

Type: String

Length Constraints: Maximum length of 256.

Pattern: .*

Required: Yes

**S3Uri**

The S3 URI for the metrics source.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: ^(https|s3)://([^/]+)/(.*$)

Required: Yes

**ContentDigest**

The hash key used for the metrics source.

Type: String

Length Constraints: Maximum length of 72.

Pattern: ^[Ss][Hh][Aa]256:[0-9a-fA-F]{64}$

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
Model
Service: Amazon SageMaker Service

The properties of a model as returned by the Search API.

Contents

Containers
The containers in the inference pipeline.
Type: Array of ContainerDefinition (p. 1366) objects
Array Members: Maximum number of 15 items.
Required: No

CreationTime
A timestamp that indicates when the model was created.
Type: Timestamp
Required: No

DeploymentRecommendation
A set of recommended deployment configurations for the model.
Type: DeploymentRecommendation (p. 1406) object
Required: No

EnableNetworkIsolation
Isolates the model container. No inbound or outbound network calls can be made to or from the model container.
Type: Boolean
Required: No

ExecutionRoleArn
The Amazon Resource Name (ARN) of the IAM role that you specified for the model.
Type: String
Pattern: ^arn:aws[\-]*:iam::\d{12}:role/\?\{[a-zA-Z0-9+=,.@\-_]+\}$
Required: No

InferenceExecutionConfig
Specifies details about how containers in a multi-container endpoint are run.
Type: InferenceExecutionConfig (p. 1589) object
Required: No

ModelArn
The Amazon Resource Name (ARN) of the model.
Type: String


Pattern: arn:aws[a-z-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:model/.*

Required: No

**ModelName**

The name of the model.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9][-]*[a-zA-Z0-9]*)

Required: No

**PrimaryContainer**

Describes the container, as part of model definition.

Type: ContainerDefinition (p. 1366) object

Required: No

**Tags**

A list of key-value pairs associated with the model. For more information, see Tagging AWS resources in the AWS General Reference Guide.

Type: Array of Tag (p. 1979) objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

**VpcConfig**

Specifies an Amazon Virtual Private Cloud (VPC) that your SageMaker jobs, hosted models, and compute resources have access to. You can control access to and from your resources by configuring a VPC. For more information, see Give SageMaker Access to Resources in your Amazon VPC.

Type: VpcConfig (p. 2076) object

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ModelAccessConfig
Service: Amazon SageMaker Service

The access configuration file for the ML model. You can explicitly accept the model end-user license agreement (EULA) within the ModelAccessConfig. For more information, see End-user license agreements.

Contents

AcceptEula

Specifies agreement to the model end-user license agreement (EULA). The AcceptEula value must be explicitly defined as True in order to accept the EULA that this model requires. You are responsible for reviewing and complying with any applicable license terms and making sure they are acceptable for your use case before downloading or using a model.

Type: Boolean
Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ModelArtifacts

Service: Amazon SageMaker Service

Provides information about the location that is configured for storing model artifacts.

Model artifacts are the output that results from training a model, and typically consist of trained parameters, a model definition that describes how to compute inferences, and other metadata.

Contents

S3ModelArtifacts

The path of the S3 object that contains the model artifacts. For example, s3://bucket-name/keynameprefix/model.tar.gz.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: ^(/https|s3)://([^/]+)/*(.*)$

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ModelBiasAppSpecification

Service: Amazon SageMaker Service

Docker container image configuration object for the model bias job.

Contents

ConfigUri

JSON formatted S3 file that defines bias parameters. For more information on this JSON configuration file, see Configure bias parameters.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: ^([https|s3])://([^/]+)\?\([^/\s]+\)$

Required: Yes

ImageUri

The container image to be run by the model bias job.

Type: String

Length Constraints: Maximum length of 255.

Pattern: .*

Required: Yes

Environment

Sets the environment variables in the Docker container.

Type: String to string map

Map Entries: Maximum number of 50 items.

Key Length Constraints: Maximum length of 256.

Key Pattern: [a-zA-Z_-][a-zA-Z0-9_-]*

Value Length Constraints: Maximum length of 256.

Value Pattern: [\S\s]*

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ModelBiasBaselineConfig
Service: Amazon SageMaker Service
The configuration for a baseline model bias job.

Contents

BaseliningJobName
The name of the baseline model bias job.
Type: String
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62\}
Required: No

ConstraintsResource
The constraints resource for a monitoring job.
Type: MonitoringConstraintsResource (p. 1740) object
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ModelBiasJobInput
Service: Amazon SageMaker Service
Inputs for the model bias job.

Contents

GroundTruthS3Input
Location of ground truth labels to use in model bias job.
Type: MonitoringGroundTruthS3Input (p. 1745) object
Required: Yes

BatchTransformInput
Input object for the batch transform job.
Type: BatchTransformInput (p. 1305) object
Required: No

EndpointInput
Input object for the endpoint
Type: EndpointInput (p. 1457) object
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ModelCard

Service: Amazon SageMaker Service

An Amazon SageMaker Model Card.

Contents

Content

The content of the model card. Content uses the model card JSON schema and provided as a string.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 100000.

Pattern: .*

Required: No

CreatedBy

Information about the user who created or modified an experiment, trial, trial component, lineage group, project, or model card.

Type: UserContext (p. 2067) object

Required: No

CreationTime

The date and time that the model card was created.

Type: Timestamp

Required: No

LastModifiedBy

Information about the user who created or modified an experiment, trial, trial component, lineage group, project, or model card.

Type: UserContext (p. 2067) object

Required: No

LastModifiedTime

The date and time that the model card was last modified.

Type: Timestamp

Required: No

ModelCardArn

The Amazon Resource Name (ARN) of the model card.

Type: String

Length Constraints: Maximum length of 256.

Pattern: ^arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]{9,16}:[0-9]{12}:model-card/[a-zA-Z0-9]*-[a-zA-Z0-9]{0,62}$
ModelCard

ModelCardName

The unique name of the model card.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

Required: No

ModelCardStatus

The approval status of the model card within your organization. Different organizations might have different criteria for model card review and approval.

- Draft: The model card is a work in progress.
- PendingReview: The model card is pending review.
- Approved: The model card is approved.
- Archived: The model card is archived. No more updates should be made to the model card, but it can still be exported.

Type: String

Valid Values: Draft | PendingReview | Approved | Archived

Required: No

ModelCardVersion

The version of the model card.

Type: Integer

Required: No

ModelId

The unique name (ID) of the model.

Type: String

Required: No

ModelPackageGroupName

The model package group that contains the model package. Only relevant for model cards created for model packages in the Amazon SageMaker Model Registry.

Type: String

Required: No

RiskRating

The risk rating of the model. Different organizations might have different criteria for model card risk ratings. For more information, see Risk ratings.

Type: String

Required: No
SecurityConfig

The security configuration used to protect model card data.

Type: ModelCardSecurityConfig (p. 1670) object

Required: No

Tags

Key-value pairs used to manage metadata for the model card.

Type: Array of Tag (p. 1979) objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ModelCardExportArtifacts

Service: Amazon SageMaker Service

The artifacts of the model card export job.

Contents

S3ExportArtifacts

The Amazon S3 URI of the exported model artifacts.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: ^((https|s3)://([^/]+)/?)\?.*$

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ModelCardExportJobSummary

Service: Amazon SageMaker Service

The summary of the Amazon SageMaker Model Card export job.

Contents

CreatedAt

The date and time that the model card export job was created.

Type: Timestamp

Required: Yes

LastModifiedAt

The date and time that the model card export job was last modified.

Type: Timestamp

Required: Yes

ModelCardExportJobArn

The Amazon Resource Name (ARN) of the model card export job.

Type: String

Length Constraints: Maximum length of 256.

Pattern: ^arn:aws[a-z-]*:sagemaker:[a-z0-9-]+[0-9]+[a-zA-Z0-9]+([0-9]+[a-zA-Z0-9]+)\[0-62]/export-job/[a-zA-Z0-9]+[0-62]$

Required: Yes

ModelCardExportJobName

The name of the model card export job.

Type: String


Pattern: ^[a-zA-Z0-9]+[0-62]$

Required: Yes

ModelCardName

The name of the model card that the export job exports.

Type: String


Pattern: ^[a-zA-Z0-9]+[0-62]$

Required: Yes

ModelCardVersion

The version of the model card that the export job exports.
Type: Integer
Required: Yes

**Status**

The completion status of the model card export job.

Type: String

Valid Values: InProgress | Completed | Failed

Required: Yes

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](https://aws.amazon.com/sdk For C++)
- [AWS SDK for Go](https://aws.amazon.com/sdk For Go)
- [AWS SDK for Java V2](https://aws.amazon.com/sdk For Java V2)
- [AWS SDK for Ruby V3](https://aws.amazon.com/sdk For Ruby V3)
ModelCardExportOutputConfig

Service: Amazon SageMaker Service

Configure the export output details for an Amazon SageMaker Model Card.

Contents

S3OutputPath

The Amazon S3 output path to export your model card PDF.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: ^((https|s3):/\([^/]+\)/?)\/(.*)$

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ModelCardSecurityConfig

Service: Amazon SageMaker Service

Configure the security settings to protect model card data.

Contents

KmsKeyId

A Key Management Service key ID to use for encrypting a model card.

Type: String

Length Constraints: Maximum length of 2048.

Pattern: . *

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ModelCardSummary

Service: Amazon SageMaker Service

A summary of the model card.

Contents

CreationTime

The date and time that the model card was created.

Type: Timestamp

Required: Yes

ModelCardArn

The Amazon Resource Name (ARN) of the model card.

Type: String

Length Constraints: Maximum length of 256.

Pattern: ^arn:aws[a-z-]*:sagemaker:[a-z0-9-]{9,16}:[0-9]{12}:model-card/[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

Required: Yes

ModelCardName

The name of the model card.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

Required: Yes

ModelCardStatus

The approval status of the model card within your organization. Different organizations might have different criteria for model card review and approval.

- Draft: The model card is a work in progress.
- PendingReview: The model card is pending review.
- Approved: The model card is approved.
- Archived: The model card is archived. No more updates should be made to the model card, but it can still be exported.

Type: String

Valid Values: Draft | PendingReview | Approved | Archived

Required: Yes

LastModifiedTime

The date and time that the model card was last modified.

Type: Timestamp
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ModelCardVersionSummary

Service: Amazon SageMaker Service

A summary of a specific version of the model card.

Contents

CreationTime

The date and time that the model card version was created.

Type: Timestamp

Required: Yes

ModelCardArn

The Amazon Resource Name (ARN) of the model card.

Type: String

Length Constraints: Maximum length of 256.

Pattern: ^arn:aws[a-z-]*:sagemaker:[a-z0-9-]{9,16}:[0-9]{12}:model-card/[a-zA-Z0-9]*[a-zA-Z0-9]{0,62}$

Required: Yes

ModelCardName

The name of the model card.

Type: String


Pattern: ^[a-zA-Z0-9]*[a-zA-Z0-9]{0,62}$

Required: Yes

ModelCardStatus

The approval status of the model card version within your organization. Different organizations might have different criteria for model card review and approval.

- Draft: The model card is a work in progress.
- PendingReview: The model card is pending review.
- Approved: The model card is approved.
- Archived: The model card is archived. No more updates should be made to the model card, but it can still be exported.

Type: String

Valid Values: Draft | PendingReview | Approved | Archived

Required: Yes

ModelCardVersion

A version of the model card.

Type: Integer
Required: Yes

**LastModifiedTime**

The time date and time that the model card version was last modified.

Type: Timestamp

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
ModelClientConfig
Service: Amazon SageMaker Service

Configures the timeout and maximum number of retries for processing a transform job invocation.

Contents

InvocationsMaxRetries
The maximum number of retries when invocation requests are failing. The default value is 3.
Type: Integer
Valid Range: Minimum value of 0. Maximum value of 3.
Required: No

InvocationsTimeoutInSeconds
The timeout value in seconds for an invocation request. The default value is 600.
Type: Integer
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ModelConfiguration
Service: Amazon SageMaker Service

Defines the model configuration. Includes the specification name and environment parameters.

Contents

CompilationJobName

The name of the compilation job used to create the recommended model artifacts.

Type: String


Pattern: ^[a-zA-Z0-9][-*[a-zA-Z0-9]]{0,62}$

Required: No

EnvironmentParameters

Defines the environment parameters that includes key, value types, and values.

Type: Array of EnvironmentParameter (p. 1469) objects

Array Members: Minimum number of 1 item. Maximum number of 10 items.

Required: No

InferenceSpecificationName

The inference specification name in the model package version.

Type: String


Pattern: ^[a-zA-Z0-9][-*[a-zA-Z0-9]]{0,62}$

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ModelDashboardEndpoint

Service: Amazon SageMaker Service

An endpoint that hosts a model displayed in the Amazon SageMaker Model Dashboard.

Contents

**CreationTime**

A timestamp that indicates when the endpoint was created.

Type: Timestamp

Required: Yes

**EndpointArn**

The Amazon Resource Name (ARN) of the endpoint.

Type: String


Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:endpoint/.*

Required: Yes

**EndpointName**

The endpoint name.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9-](\-*[a-zA-Z0-9]){0,62}

Required: Yes

**EndpointStatus**

The endpoint status.

Type: String

Valid Values: OutOfService | Creating | Updating | SystemUpdating | RollingBack | InService | Deleting | Failed | UpdateRollbackFailed

Required: Yes

**LastModifiedTime**

The last time the endpoint was modified.

Type: Timestamp

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:
• AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for Ruby V3
ModelDashboardIndicatorAction

Service: Amazon SageMaker Service

An alert action taken to light up an icon on the Amazon SageMaker Model Dashboard when an alert goes into InAlert status.

Contents

Enabled

Indicates whether the alert action is turned on.

Type: Boolean

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ModelDashboardModel
Service: Amazon SageMaker Service

A model displayed in the Amazon SageMaker Model Dashboard.

Contents

Endpoints
The endpoints that host a model.
Type: Array of ModelDashboardEndpoint (p. 1677) objects
Required: No

LastBatchTransformJob
A batch transform job. For information about SageMaker batch transform, see Use Batch Transform.
Type: TransformJob (p. 2022) object
Required: No

Model
A model displayed in the Model Dashboard.
Type: Model (p. 1656) object
Required: No

ModelCard
The model card for a model.
Type: ModelDashboardModelCard (p. 1681) object
Required: No

MonitoringSchedules
The monitoring schedules for a model.
Type: Array of ModelDashboardMonitoringSchedule (p. 1684) objects
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ModelDashboardModelCard

Service: Amazon SageMaker Service

The model card for a model displayed in the Amazon SageMaker Model Dashboard.

Contents

CreatedBy

Information about the user who created or modified an experiment, trial, trial component, lineage group, project, or model card.

Type: UserContext (p. 2067) object

Required: No

CreationTime

A timestamp that indicates when the model card was created.

Type: Timestamp

Required: No

LastModifiedBy

Information about the user who created or modified an experiment, trial, trial component, lineage group, project, or model card.

Type: UserContext (p. 2067) object

Required: No

LastModifiedTime

A timestamp that indicates when the model card was last updated.

Type: Timestamp

Required: No

ModelCardArn

The Amazon Resource Name (ARN) for a model card.

Type: String

Length Constraints: Maximum length of 256.

Pattern: ^arn:aws[a-z-]+:sagemaker:[a-zA-Z0-9-]-[0-9]{12}:model-card/[a-zA-Z0-9]-[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}$

Required: No

ModelCardName

The name of a model card.

Type: String


Pattern: ^[a-zA-Z0-9]-*[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}$
Required: No

**ModelCardStatus**

The model card status.

Type: String

Valid Values: Draft | PendingReview | Approved | Archived

Required: No

**ModelCardVersion**

The model card version.

Type: Integer

Required: No

**ModelId**

For models created in SageMaker, this is the model ARN. For models created outside of SageMaker, this is a user-customized string.

Type: String

Required: No

**RiskRating**

A model card's risk rating. Can be low, medium, or high.

Type: String

Required: No

**SecurityConfig**

The KMS Key ID (KMSKeyId) for encryption of model card information.

Type: `ModelCardSecurityConfig (p. 1670)` object

Required: No

**Tags**

The tags associated with a model card.

Type: Array of `Tag (p. 1979)` objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

### See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
ModelDashboardMonitoringSchedule

Service: Amazon SageMaker Service

A monitoring schedule for a model displayed in the Amazon SageMaker Model Dashboard.

Contents

**BatchTransformInput**

Input object for the batch transform job.

Type: [BatchTransformInput](p. 1305) object

Required: No

**CreationTime**

A timestamp that indicates when the monitoring schedule was created.

Type: Timestamp

Required: No

**EndpointName**

The endpoint which is monitored.

Type: String

Length Constraints: Maximum length of 63.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}`

Required: No

**FailureReason**

If a monitoring job failed, provides the reason.

Type: String

Length Constraints: Maximum length of 1024.

Required: No

**LastModifiedTime**

A timestamp that indicates when the monitoring schedule was last updated.

Type: Timestamp

Required: No

**LastMonitoringExecutionSummary**

Summary of information about the last monitoring job to run.

Type: [MonitoringExecutionSummary](p. 1743) object

Required: No

**MonitoringAlertSummaries**

A JSON array where each element is a summary for a monitoring alert.
ModelDashboardMonitoringSchedule

Type: Array of MonitoringAlertSummary (p. 1733) objects

Array Members: Minimum number of 1 item. Maximum number of 100 items.

Required: No

**MonitoringScheduleArn**

The Amazon Resource Name (ARN) of a monitoring schedule.

Type: String

Length Constraints: Maximum length of 256.

Pattern: .*

Required: No

**MonitoringScheduleConfig**

Configures the monitoring schedule and defines the monitoring job.

Type: MonitoringScheduleConfig (p. 1761) object

Required: No

**MonitoringScheduleName**

The name of a monitoring schedule.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

Required: No

**MonitoringScheduleStatus**

The status of the monitoring schedule.

Type: String

Valid Values: Pending  |  Failed  |  Scheduled  |  Stopped

Required: No

**MonitoringType**

The monitor type of a model monitor.

Type: String

Valid Values: DataQuality  |  ModelQuality  |  ModelBias  |  ModelExplainability

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ModelDataQuality
Service: Amazon SageMaker Service

Data quality constraints and statistics for a model.

Contents

Constraints
Data quality constraints for a model.
Type: MetricsSource (p. 1655) object
Required: No

Statistics
Data quality statistics for a model.
Type: MetricsSource (p. 1655) object
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ModelDataSource
Service: Amazon SageMaker Service

Specifies the location of ML model data to deploy. If specified, you must specify one and only one of the available data sources.

Contents

S3DataSource

Specifies the S3 location of ML model data to deploy.

Type: S3ModelDataSource (p. 1927) object

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ModelDeployConfig

Service: Amazon SageMaker Service

Specifies how to generate the endpoint name for an automatic one-click Autopilot model deployment.

Contents

**AutoGenerateEndpointName**

Set to True to automatically generate an endpoint name for a one-click Autopilot model deployment; set to False otherwise. The default value is False.

*Note*
If you set AutoGenerateEndpointName to True, do not specify the EndpointName; otherwise a 400 error is thrown.

Type: Boolean

Required: No

**EndpointName**

Specifies the endpoint name to use for a one-click Autopilot model deployment if the endpoint name is not generated automatically.

*Note*
Specify the EndpointName if and only if you set AutoGenerateEndpointName to False; otherwise a 400 error is thrown.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ModelDeployResult
Service: Amazon SageMaker Service
Provides information about the endpoint of the model deployment.

Contents

EndpointName
The name of the endpoint to which the model has been deployed.

Note
If model deployment fails, this field is omitted from the response.

Type: String
Length Constraints: Maximum length of 63.
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62\}
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ModelDigests
Service: Amazon SageMaker Service
Provides information to verify the integrity of stored model artifacts.

Contents

ArtifactDigest
Provides a hash value that uniquely identifies the stored model artifacts.

Type: String
Pattern: `^[\w:]+$`
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ModelExplainabilityAppSpecification
Service: Amazon SageMaker Service
Docker container image configuration object for the model explainability job.

Contents

**ConfigUri**
JSON formatted Amazon S3 file that defines explainability parameters. For more information on this JSON configuration file, see Configure model explainability parameters.

Type: String
Length Constraints: Maximum length of 1024.

Pattern: `^(https|s3)://([^[/]+)/(.*\))$`

Required: Yes

**ImageUri**
The container image to be run by the model explainability job.

Type: String
Length Constraints: Maximum length of 255.

Pattern: `.\n
Required: Yes

**Environment**
Sets the environment variables in the Docker container.

Type: String to string map
Map Entries: Maximum number of 50 items.
Key Length Constraints: Maximum length of 256.
Key Pattern: `[a-zA-z\-\_][a-zA-z0-9\-\_]*`
Value Length Constraints: Maximum length of 256.
Value Pattern: `[\S\s]*`

Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ModelExplainabilityBaselineConfig

Service: Amazon SageMaker Service

The configuration for a baseline model explainability job.

Contents

**BaseliningJobName**

The name of the baseline model explainability job.

Type: String


Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}`

Required: No

**ConstraintsResource**

The constraints resource for a monitoring job.

Type: MonitoringConstraintsResource (p. 1740) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
ModelExplainabilityJobInput
Service: Amazon SageMaker Service
Inputs for the model explainability job.

Contents

BatchTransformInput
Input object for the batch transform job.
Type: BatchTransformInput (p. 1305) object
Required: No

EndpointInput
Input object for the endpoint
Type: EndpointInput (p. 1457) object
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ModelInfrastructureConfig

Service: Amazon SageMaker Service

The configuration for the infrastructure that the model will be deployed to.

Contents

InfrastructureType

The inference option to which to deploy your model. Possible values are the following:

- RealTime: Deploy to real-time inference.

  Type: String
  Valid Values: RealTimeInference
  Required: Yes

RealTimeInferenceConfig

The infrastructure configuration for deploying the model to real-time inference.

  Type: RealTimeInferenceConfig (p. 1883) object

  Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ModelInput
Service: Amazon SageMaker Service
Input object for the model.

Contents

DataInputConfig
The input configuration object for the model.
Type: String
Pattern: [\S\s]+
Required: Yes

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ModelLatencyThreshold

Service: Amazon SageMaker Service

The model latency threshold.

Contents

Percentile

The model latency percentile threshold. Acceptable values are P95 and P99. For custom load tests, specify the value as P95.

Type: String

Length Constraints: Maximum length of 64.

Required: No

ValueInMilliseconds

The model latency percentile value in milliseconds.

Type: Integer

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ModelMetadataFilter

Service: Amazon SageMaker Service

Part of the search expression. You can specify the name and value (domain, task, framework, framework version, task, and model).

Contents

Name

The name of the of the model to filter by.

Type: String

Valid Values: Domain | Framework | Task | FrameworkVersion

Required: Yes

Value

The value to filter the model metadata.

Type: String

Length Constraints: Maximum length of 256.

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ModelMetadataSearchExpression

Service: Amazon SageMaker Service

One or more filters that searches for the specified resource or resources in a search. All resource objects that satisfy the expression's condition are included in the search results.

Contents

Filters

A list of filter objects.

Type: Array of ModelMetadataFilter (p. 1698) objects

Array Members: Minimum number of 1 item. Maximum number of 4 items.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ModelMetadataSummary
Service: Amazon SageMaker Service

A summary of the model metadata.

Contents

Domain
The machine learning domain of the model.
Type: String
Required: Yes

Framework
The machine learning framework of the model.
Type: String
Required: Yes

FrameworkVersion
The framework version of the model.
Type: String
Required: Yes

Model
The name of the model.
Type: String
Required: Yes

Task
The machine learning task of the model.
Type: String
Required: Yes

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ModelMetrics
Service: Amazon SageMaker Service
Contains metrics captured from a model.

Contents

Bias
Metrics that measure bias in a model.
Type: Bias (p. 1309) object
Required: No

Explainability
Metrics that help explain a model.
Type: Explainability (p. 1478) object
Required: No

ModelDataQuality
Metrics that measure the quality of the input data for a model.
Type: ModelDataQuality (p. 1687) object
Required: No

ModelQuality
Metrics that measure the quality of a model.
Type: ModelQuality (p. 1720) object
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
**ModelPackage**

Service: Amazon SageMaker Service

A versioned model that can be deployed for SageMaker inference.

**Contents**

**AdditionalInferenceSpecifications**

An array of additional Inference Specification objects.

- Type: Array of `AdditionalInferenceSpecificationDefinition` objects
- Array Members: Minimum number of 1 item. Maximum number of 15 items.
- Required: No

**ApprovalDescription**

A description provided when the model approval is set.

- Type: String
- Length Constraints: Maximum length of 1024.
- Pattern: `.*`
- Required: No

**CertifyForMarketplace**

Whether the model package is to be certified to be listed on AWS Marketplace. For information about listing model packages on AWS Marketplace, see [List Your Algorithm or Model Package on AWS Marketplace](#).

- Type: Boolean
- Required: No

**CreatedBy**

Information about the user who created or modified an experiment, trial, trial component, lineage group, or project.

- Type: `UserContext` object
- Required: No

**CreationTime**

The time that the model package was created.

- Type: Timestamp
- Required: No

**CustomerMetadataProperties**

The metadata properties for the model package.

- Type: String to string map
- Map Entries: Maximum number of 50 items.
Key Length Constraints: Minimum length of 1. Maximum length of 128.
Key Pattern: ^([\p{L}\p{Z}\p{N}_.:/=+-@]*)${1,128}$
Value Length Constraints: Minimum length of 1. Maximum length of 256.
Value Pattern: ^([\p{L}\p{Z}\p{N}_.:/=+-@]*)${1,256}$
Required: No

**Domain**

The machine learning domain of your model package and its components. Common machine learning domains include computer vision and natural language processing.

Type: String
Required: No

**DriftCheckBaselines**

Represents the drift check baselines that can be used when the model monitor is set using the model package.

Type: DriftCheckBaselines (p. 1426) object
Required: No

**InferenceSpecification**

Defines how to perform inference generation after a training job is run.

Type: InferenceSpecification (p. 1601) object
Required: No

**LastModifiedBy**

Information about the user who created or modified an experiment, trial, trial component, lineage group, or project.

Type: UserContext (p. 2067) object
Required: No

**LastModifiedTime**

The last time the model package was modified.

Type: Timestamp
Required: No

**MetadataProperties**

Metadata properties of the tracking entity, trial, or trial component.

Type: MetadataProperties (p. 1648) object
Required: No

**ModelApprovalStatus**

The approval status of the model. This can be one of the following values.
- APPROVED - The model is approved

1703
• **REJECTED** - The model is rejected.
• **PENDING_MANUAL_APPROVAL** - The model is waiting for manual approval.

**Type:** String

**Valid Values:** Approved | Rejected | PendingManualApproval

**Required:** No

**ModelMetrics**

Metrics for the model.

**Type:** ModelMetrics (p. 1701) object

**Required:** No

**ModelPackageArn**

The Amazon Resource Name (ARN) of the model package.

**Type:** String

**Length Constraints:** Minimum length of 1. Maximum length of 2048.

**Pattern:** `^arn:aws(-cn|-us-gov)?:sagemaker:[a-z0-9\-]{9,16}:[0-9]{12}:model-package/[^\S]{1,2048}$`

**Required:** No

**ModelPackageDescription**

The description of the model package.

**Type:** String

**Length Constraints:** Maximum length of 1024.

**Pattern:** `[^\p{L}\p{M}\p{Z}\p{S}\p{N}\p{P}]*`

**Required:** No

**ModelPackageName**

The name of the model.

**Type:** String

**Length Constraints:** Minimum length of 1. Maximum length of 63.

**Pattern:** `^[a-zA-Z0-9]\(-*[a-zA-Z0-9]\){0,62}$`

**Required:** No
ModelPackageStatus

The status of the model package. This can be one of the following values.
- PENDING - The model package is pending being created.
- IN_PROGRESS - The model package is in the process of being created.
- COMPLETED - The model package was successfully created.
- FAILED - The model package failed.
- DELETING - The model package is in the process of being deleted.

Type: String

Valid Values: Pending | InProgress | Completed | Failed | Deleting

Required: No

ModelPackageStatusDetails

 Specifies the validation and image scan statuses of the model package.

Type: ModelPackageStatusDetails (p. 1714) object

Required: No

ModelPackageVersion

The version number of a versioned model.

Type: Integer

Valid Range: Minimum value of 1.

Required: No

SamplePayloadUrl

The Amazon Simple Storage Service path where the sample payload are stored. This path must point to a single gzip compressed tar archive (.tar.gz suffix).

Type: String

Required: No

SkipModelValidation

Indicates if you want to skip model validation.

Type: String

Valid Values: All | None

Required: No

SourceAlgorithmSpecification

A list of algorithms that were used to create a model package.

Type: SourceAlgorithmSpecification (p. 1955) object

Required: No

Tags

A list of the tags associated with the model package. For more information, see Tagging AWS resources in the AWS General Reference Guide.
Type: Array of [Tag](p. 1979) objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

**Task**

The machine learning task your model package accomplishes. Common machine learning tasks include object detection and image classification.

Type: String

Required: No

**ValidationSpecification**

Specifies batch transform jobs that SageMaker runs to validate your model package.

Type: [ModelPackageValidationSpecification](p. 1719) object

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](aws-sdk-cpp)
- [AWS SDK for Go](aws-sdk-go)
- [AWS SDK for Java V2](aws-sdk-java-v2)
- [AWS SDK for Ruby V3](aws-sdk-ruby-v3)
ModelPackageContainerDefinition
Service: Amazon SageMaker Service

Describes the Docker container for the model package.

Contents

Image

The Amazon EC2 Container Registry (Amazon ECR) path where inference code is stored.

If you are using your own custom algorithm instead of an algorithm provided by SageMaker, the inference code must meet SageMaker requirements. SageMaker supports both registry/repository[:tag] and registry/repository[@digest] image path formats. For more information, see Using Your Own Algorithms with Amazon SageMaker.

Type: String

Length Constraints: Maximum length of 255.

Pattern: \S+

Required: Yes

AdditionalS3DataSource

The additional data source that is used during inference in the Docker container for your model package.

Type: AdditionalS3DataSource (p. 1222) object

Required: No

ContainerHostname

The DNS host name for the Docker container.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9][-]*[a-zA-Z0-9]{0,62}$

Required: No

Environment

The environment variables to set in the Docker container. Each key and value in the Environment string to string map can have length of up to 1024. We support up to 16 entries in the map.

Type: String to string map

Map Entries: Maximum number of 16 items.

Key Length Constraints: Maximum length of 1024.

Key Pattern: [a-zA-Z_] [a-zA-Z0-9_] *

Value Length Constraints: Maximum length of 1024.

Value Pattern: \S\s*
ModelPackageContainerDefinition

- **Framework**
  The machine learning framework of the model package container image.
  
  **Type:** String
  
  **Required:** No

- **FrameworkVersion**
  The framework version of the Model Package Container Image.
  
  **Type:** String
  
  **Length Constraints:** Minimum length of 3. Maximum length of 10.
  
  **Pattern:** \[0-9\].[A-Za-z0-9.-]+
  
  **Required:** No

- **ImageDigest**
  An MD5 hash of the training algorithm that identifies the Docker image used for training.
  
  **Type:** String
  
  **Length Constraints:** Maximum length of 72.
  
  **Pattern:** ^[Ss][Hh][Aa][256]:[0-9a-fA-F]{64}$
  
  **Required:** No

- **ModelDataUrl**
  The Amazon S3 path where the model artifacts, which result from model training, are stored. This path must point to a single gzip compressed tar archive (.tar.gz suffix).
  
  **Note**
  The model artifacts must be in an S3 bucket that is in the same region as the model package.
  
  **Type:** String
  
  **Length Constraints:** Maximum length of 1024.
  
  **Pattern:** ^(https|s3)://([^/]+)/(.*)$
  
  **Required:** No

- **ModelInput**
  A structure with Model Input details.
  
  **Type:** ModelInput (p. 1696) object
  
  **Required:** No

- **NearestModelName**
  The name of a pre-trained machine learning benchmarked by Amazon SageMaker Inference Recommender model that matches your model. You can find a list of benchmarked models by calling ListModelMetadata.
  
  **Type:** String
Required: No

**Productld**

The AWS Marketplace product ID of the model package.

Type: String

Length Constraints: Maximum length of 256.

Pattern: `^[a-zA-Z0-9][-][a-zA-Z0-9]*$`

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
ModelPackageGroup

Service: Amazon SageMaker Service

A group of versioned models in the model registry.

Contents

**CreatedBy**

Information about the user who created or modified an experiment, trial, trial component, lineage group, project, or model card.

Type: `UserContext (p. 2067)` object

Required: No

**CreationTime**

The time that the model group was created.

Type: Timestamp

Required: No

**ModelPackageGroupArn**

The Amazon Resource Name (ARN) of the model group.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(-cn|-us-gov)?:sagemaker:[a-z0-9\-]{9,16}:\[0-9\]{12}:model-package-group/\[\S\]{1,2048}$`

Required: No

**ModelPackageGroupDescription**

The description for the model group.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: `\[p\{L\}\p\{M\}\p\{Z\}\p\{S\}\p\{N\}\p\{P\}]`* 

Required: No

**ModelPackageGroupName**

The name of the model group.

Type: String


Pattern: `^[a-zA-Z0-9\-]{0,62}$`

Required: No

**ModelPackageGroupStatus**

The status of the model group. This can be one of the following values.
• PENDING - The model group is pending being created.
• IN_PROGRESS - The model group is in the process of being created.
• COMPLETED - The model group was successfully created.
• FAILED - The model group failed.
• DELETING - The model group is in the process of being deleted.
• DELETE_FAILED - SageMaker failed to delete the model group.

Type: String

Valid Values: Pending | InProgress | Completed | Failed | Deleting | DeleteFailed

Required: No

Tags

A list of the tags associated with the model group. For more information, see Tagging AWS resources in the AWS General Reference Guide.

Type: Array of Tag (p. 1979) objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

• AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for Ruby V3
ModelPackageGroupSummary
Service: Amazon SageMaker Service

Summary information about a model group.

Contents

CreationTime
  The time that the model group was created.
  Type: Timestamp
  Required: Yes

ModelPackageGroupArn
  The Amazon Resource Name (ARN) of the model group.
  Type: String
  Length Constraints: Minimum length of 1. Maximum length of 2048.
  Pattern: ^arn:aws(-cn|-us-gov)?:sagemaker:[a-z0-9-]+[9,16]:[0-9]+:model-package-group/[\S]{1,2048}$
  Required: Yes

ModelPackageGroupName
  The name of the model group.
  Type: String
  Pattern: ^[a-zA-Z0-9-\d\-\.]\d*[a-zA-Z0-9-]*$[0,62]$
  Required: Yes

ModelPackageGroupStatus
  The status of the model group.
  Type: String
  Valid Values: Pending | InProgress | Completed | Failed | Deleting | DeleteFailed
  Required: Yes

ModelPackageGroupDescription
  A description of the model group.
  Type: String
  Length Constraints: Maximum length of 1024.
  Pattern: [\p{L}\p{M}\p{Z}\p{S}\p{N}\p{P}]*
  Required: No
See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ModelPackageStatusDetails

Service: Amazon SageMaker Service

Specifies the validation and image scan statuses of the model package.

Contents

ValidationStatuses

The validation status of the model package.

Type: Array of ModelPackageStatusItem (p. 1715) objects

Required: Yes

ImageScanStatuses

The status of the scan of the Docker image container for the model package.

Type: Array of ModelPackageStatusItem (p. 1715) objects

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ModelPackageStatusItem

Service: Amazon SageMaker Service

Represents the overall status of a model package.

Contents

Name

The name of the model package for which the overall status is being reported.

Type: String


Pattern: ^[a-zA-Z0-9]( -*[a-zA-Z0-9])\{0,62}$

Required: Yes

Status

The current status.

Type: String

Valid Values: NotStarted | InProgress | Completed | Failed

Required: Yes

FailureReason

if the overall status is Failed, the reason for the failure.

Type: String

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ModelPackageSummary
Service: Amazon SageMaker Service
Provides summary information about a model package.

Contents

CreationTime
A timestamp that shows when the model package was created.
Type: Timestamp
Required: Yes

ModelPackageArn
The Amazon Resource Name (ARN) of the model package.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 2048.
Pattern: ^arn:aws(-cn|-us-gov)?:sagemaker:[a-z0-9\-]{9,16}:[0-9]{12}:model-package/\[\S\]{1,2048}\$
Required: Yes

ModelPackageStatus
The overall status of the model package.
Type: String
Valid Values: Pending | InProgress | Completed | Failed | Deleting
Required: Yes

ModelApprovalStatus
The approval status of the model. This can be one of the following values.
• APPROVED - The model is approved
• REJECTED - The model is rejected.
• PENDING_MANUAL_APPROVAL - The model is waiting for manual approval.
Type: String
Valid Values: Approved | Rejected | PendingManualApproval
Required: No

ModelPackageDescription
A brief description of the model package.
Type: String
Length Constraints: Maximum length of 1024.
Pattern: \[p\{L\}\p\{M\}\p\{Z\}\p\{S\}\p\{N\}\p\{P\}\]*
Required: No
ModelPackageGroupName

If the model package is a versioned model, the model group that the versioned model belongs to.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

Required: No

ModelPackageName

The name of the model package.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

Required: No

ModelPackageVersion

If the model package is a versioned model, the version of the model.

Type: Integer

Valid Range: Minimum value of 1.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ModelPackageValidationProfile

Service: Amazon SageMaker Service

Contains data, such as the inputs and targeted instance types that are used in the process of validating the model package.

The data provided in the validation profile is made available to your buyers on AWS Marketplace.

Contents

ProfileName

The name of the profile for the model package.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]{0,62})$

Required: Yes

TransformJobDefinition

The TransformJobDefinition object that describes the transform job used for the validation of the model package.

Type: TransformJobDefinition (p. 2027) object

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ModelPackageValidationSpecification

Service: Amazon SageMaker Service

Specifies batch transform jobs that SageMaker runs to validate your model package.

Contents

ValidationProfiles

An array of ModelPackageValidationProfile objects, each of which specifies a batch transform job that SageMaker runs to validate your model package.

Type: Array of ModelPackageValidationProfile (p. 1718) objects

Array Members: Fixed number of 1 item.

Required: Yes

ValidationRole

The IAM roles to be used for the validation of the model package.

Type: String


Pattern: ^arn:aws[a-z\-]*:iam::\d{12}:role/?[a-zA-Z_0-9+=,.@\-_/]+$

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ModelQuality
Service: Amazon SageMaker Service

Model quality statistics and constraints.

Contents

Constraints
Model quality constraints.
Type: MetricsSource (p. 1655) object
Required: No

Statistics
Model quality statistics.
Type: MetricsSource (p. 1655) object
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ModelQualityAppSpecification

Service: Amazon SageMaker Service

Container image configuration object for the monitoring job.

Contents

ImageUri

The address of the container image that the monitoring job runs.

Type: String

Length Constraints: Maximum length of 255.

Pattern: .*

Required: Yes

ContainerArguments

An array of arguments for the container used to run the monitoring job.

Type: Array of strings

Array Members: Minimum number of 1 item. Maximum number of 50 items.

Length Constraints: Maximum length of 256.

Pattern: .*

Required: No

ContainerEntrypoint

Specifies the entrypoint for a container that the monitoring job runs.

Type: Array of strings

Array Members: Minimum number of 1 item. Maximum number of 100 items.

Length Constraints: Maximum length of 256.

Pattern: .*

Required: No

Environment

Sets the environment variables in the container that the monitoring job runs.

Type: String to string map

Map Entries: Maximum number of 50 items.

Key Length Constraints: Maximum length of 256.

Key Pattern: [a-zA-Z_][a-zA-Z0-9_]*

Value Length Constraints: Maximum length of 256.

Value Pattern: [\S\s]*
**PostAnalyticsProcessorSourceUri**

An Amazon S3 URI to a script that is called after analysis has been performed. Applicable only for the built-in (first party) containers.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: `^(https|s3)://([\^/]+)/?(.*)$

Required: No

**ProblemType**

The machine learning problem type of the model that the monitoring job monitors.

Type: String

Valid Values: BinaryClassification | MulticlassClassification | Regression

Required: No

**RecordPreprocessorSourceUri**

An Amazon S3 URI to a script that is called per row prior to running analysis. It can base64 decode the payload and convert it into a flattened JSON so that the built-in container can use the converted data. Applicable only for the built-in (first party) containers.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: `^(https|s3)://([\^/]+)/?(.*)$

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](https://aws-sdk_cpp.amazonaws.com)
- [AWS SDK for Go](https://aws-sdk_golang.amazonaws.com)
- [AWS SDK for Java V2](https://aws-sdk_java2.amazonaws.com)
- [AWS SDK for Ruby V3](https://aws-sdk_ruby3.amazonaws.com)
ModelQualityBaselineConfig

Service: Amazon SageMaker Service

Configuration for monitoring constraints and monitoring statistics. These baseline resources are compared against the results of the current job from the series of jobs scheduled to collect data periodically.

Contents

BaseliningJobName

The name of the job that performs baselining for the monitoring job.

Type: String


Pattern: ^[a-zA-Z0-9]( -*[a-zA-Z0-9])\{0,62}\$

Required: No

ConstraintsResource

The constraints resource for a monitoring job.

Type: MonitoringConstraintsResource (p. 1740) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ModelQualityJobInput

Service: Amazon SageMaker Service

The input for the model quality monitoring job. Currently endpoints are supported for input for model quality monitoring jobs.

Contents

GroundTruthS3Input

The ground truth label provided for the model.

Type: MonitoringGroundTruthS3Input (p. 1745) object

Required: Yes

BatchTransformInput

Input object for the batch transform job.

Type: BatchTransformInput (p. 1305) object

Required: No

EndpointInput

Input object for the endpoint

Type: EndpointInput (p. 1457) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
**ModelRegisterSettings**

**Service**: Amazon SageMaker Service

The model registry settings for the SageMaker Canvas application.

**Contents**

**CrossAccountModelRegisterRoleArn**

The Amazon Resource Name (ARN) of the SageMaker model registry account. Required only to register model versions created by a different SageMaker Canvas AWS account than the AWS account in which SageMaker model registry is set up.

Type: String


Pattern: \^arn:aws[a-z\-]*:iam::\d{12}:role/?[a-zA-Z0-9\+=,\@\-_/]+\$

Required: No

**Status**

Describes whether the integration to the model registry is enabled or disabled in the Canvas application.

Type: String

Valid Values: ENABLED | DISABLED

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ModelStepMetadata
Service: Amazon SageMaker Service
Metadata for Model steps.

Contents

Arn
The Amazon Resource Name (ARN) of the created model.
Type: String
Length Constraints: Maximum length of 256.
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ModelSummary

Service: Amazon SageMaker Service

Provides summary information about a model.

Contents

CreationTime

A timestamp that indicates when the model was created.
Type: Timestamp
Required: Yes

ModelArn

The Amazon Resource Name (ARN) of the model.
Type: String
Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:model/.*
Required: Yes

ModelName

The name of the model that you want a summary for.
Type: String
Length Constraints: Maximum length of 63.
Pattern: ^[a-zA-Z0-9](\-[a-zA-Z0-9]*)*
Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ModelVariantConfig

Service: Amazon SageMaker Service

Contains information about the deployment options of a model.

Contents

InfrastructureConfig

The configuration for the infrastructure that the model will be deployed to.

Type: ModelInfrastructureConfig (p. 1695) object

Required: Yes

ModelName

The name of the Amazon SageMaker Model entity.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])*|

Required: Yes

VariantName

The name of the variant.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](?([-a-zA-Z0-9]*[a-zA-Z0-9])?)|

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ModelVariantConfigSummary
Service: Amazon SageMaker Service

Summary of the deployment configuration of a model.

Contents

**InfrastructureConfig**

The configuration of the infrastructure that the model has been deployed to.

Type: ModelInfrastructureConfig (p. 1695) object

Required: Yes

**ModelName**

The name of the Amazon SageMaker Model entity.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])*

Required: Yes

**Status**

The status of deployment for the model variant on the hosted inference endpoint.

- Creating - Amazon SageMaker is preparing the model variant on the hosted inference endpoint.
- InService - The model variant is running on the hosted inference endpoint.
- Updating - Amazon SageMaker is updating the model variant on the hosted inference endpoint.
- Deleting - Amazon SageMaker is deleting the model variant on the hosted inference endpoint.
- Deleted - The model variant has been deleted on the hosted inference endpoint. This can only happen after stopping the experiment.

Type: String

Valid Values: Creating | Updating | InService | Deleting | Deleted

Required: Yes

**VariantName**

The name of the variant.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9]([-a-zA-Z0-9]*[a-zA-Z0-9])?

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:
• AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for Ruby V3
MonitoringAlertActions
Service: Amazon SageMaker Service

A list of alert actions taken in response to an alert going into InAlert status.

Contents

ModelDashboardIndicator

An alert action taken to light up an icon on the Model Dashboard when an alert goes into InAlert status.

Type: ModelDashboardIndicatorAction (p. 1679) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
MonitoringAlertHistorySummary
Service: Amazon SageMaker Service
Provides summary information of an alert's history.

Contents

AlertStatus
The current alert status of an alert.
Type: String
Valid Values: InAlert | OK
Required: Yes

CreationTime
A timestamp that indicates when the first alert transition occurred in an alert history. An alert transition can be from status InAlert to OK, or from OK to InAlert.
Type: Timestamp
Required: Yes

MonitoringAlertName
The name of a monitoring alert.
Type: String
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}\$
Required: Yes

MonitoringScheduleName
The name of a monitoring schedule.
Type: String
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}\$
Required: Yes

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
MonitoringAlertSummary
Service: Amazon SageMaker Service
Provides summary information about a monitor alert.

Contents

Actions
A list of alert actions taken in response to an alert going into InAlert status.
Type: MonitoringAlertActions (p. 1731) object
Required: Yes

AlertStatus
The current status of an alert.
Type: String
Valid Values: InAlert | OK
Required: Yes

CreationTime
A timestamp that indicates when a monitor alert was created.
Type: Timestamp
Required: Yes

DatapointsToAlert
Within EvaluationPeriod, how many execution failures will raise an alert.
Type: Integer
Valid Range: Minimum value of 1. Maximum value of 100.
Required: Yes

EvaluationPeriod
The number of most recent monitoring executions to consider when evaluating alert status.
Type: Integer
Valid Range: Minimum value of 1. Maximum value of 100.
Required: Yes

LastModifiedTime
A timestamp that indicates when a monitor alert was last updated.
Type: Timestamp
Required: Yes

MonitoringAlertName
The name of a monitoring alert.
Type: String
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}$
Required: Yes

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
MonitoringAppSpecification

Service: Amazon SageMaker Service

Container image configuration object for the monitoring job.

Contents

ImageUri

The container image to be run by the monitoring job.

Type: String

Length Constraints: Maximum length of 255.

Pattern: .*

Required: Yes

ContainerArguments

An array of arguments for the container used to run the monitoring job.

Type: Array of strings

Array Members: Minimum number of 1 item. Maximum number of 50 items.

Length Constraints: Maximum length of 256.

Pattern: .*

Required: No

ContainerEntrypoint

Specifies the entrypoint for a container used to run the monitoring job.

Type: Array of strings

Array Members: Minimum number of 1 item. Maximum number of 100 items.

Length Constraints: Maximum length of 256.

Pattern: .*

Required: No

PostAnalyticsProcessorSourceUri

An Amazon S3 URI to a script that is called after analysis has been performed. Applicable only for the built-in (first party) containers.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: ^(https|s3)://([^/]+)/(.*)$

Required: No

RecordPreprocessorSourceUri

An Amazon S3 URI to a script that is called per row prior to running analysis. It can base64 decode the payload and convert it into a flattened JSON so that the built-in container can use the converted data. Applicable only for the built-in (first party) containers.
Type: String

Length Constraints: Maximum length of 1024.

Pattern: ^((https|s3):/[/^/]+)?(.*$)

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
MonitoringBaselineConfig
Service: Amazon SageMaker Service

Configuration for monitoring constraints and monitoring statistics. These baseline resources are compared against the results of the current job from the series of jobs scheduled to collect data periodically.

Contents

BaseliningJobName
The name of the job that performs baselining for the monitoring job.
Type: String
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]{0,62})
Required: No

ConstraintsResource
The baseline constraint file in Amazon S3 that the current monitoring job should validated against.
Type: MonitoringConstraintsResource (p. 1740) object
Required: No

StatisticsResource
The baseline statistics file in Amazon S3 that the current monitoring job should be validated against.
Type: MonitoringStatisticsResource (p. 1764) object
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
MonitoringClusterConfig

Service: Amazon SageMaker Service

Configuration for the cluster used to run model monitoring jobs.

Contents

InstanceCount

The number of ML compute instances to use in the model monitoring job. For distributed processing jobs, specify a value greater than 1. The default value is 1.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 100.

Required: Yes

InstanceType

The ML compute instance type for the processing job.

Type: String

Valid Values: ml.t3.medium | ml.t3.large | ml.t3.xlarge | ml.t3.2xlarge | ml.m4.xlarge | ml.m4.2xlarge | ml.m4.4xlarge | ml.m4.10xlarge | ml.m4.16xlarge | ml.c4.xlarge | ml.c4.2xlarge | ml.c4.4xlarge | ml.c4.8xlarge | ml.p2.xlarge | ml.p2.8xlarge | ml.p2.16xlarge | ml.p3.2xlarge | ml.p3.8xlarge | ml.p3.16xlarge | ml.c5.xlarge | ml.c5.2xlarge | ml.c5.4xlarge | ml.c5.9xlarge | ml.c5.18xlarge | ml.m5.large | ml.m5.xlarge | ml.m5.2xlarge | ml.m5.4xlarge | ml.m5.12xlarge | ml.m5.24xlarge | ml.r5.large | ml.r5.xlarge | ml.r5.2xlarge | ml.r5.4xlarge | ml.r5.8xlarge | ml.r5.12xlarge | ml.r5.16xlarge | ml.r5.24xlarge | ml.g4dn.xlarge | ml.g4dn.2xlarge | ml.g4dn.4xlarge | ml.g4dn.8xlarge | ml.g4dn.12xlarge | ml.g4dn.16xlarge

Required: Yes

VolumeSizeInGB

The size of the ML storage volume, in gigabytes, that you want to provision. You must specify sufficient ML storage for your scenario.

Type: Integer


Required: Yes

VolumeKmsKeyId

The AWS Key Management Service (AWS KMS) key that Amazon SageMaker uses to encrypt data on the storage volume attached to the ML compute instance(s) that run the model monitoring job.

Type: String

Length Constraints: Maximum length of 2048.

Pattern: .*

Required: No
See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
MonitoringConstraintsResource

Service: Amazon SageMaker Service

The constraints resource for a monitoring job.

Contents

S3Uri

The Amazon S3 URI for the constraints resource.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: ^(https|s3)://([^/]+)/?([^/]+)$

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
MonitoringCsvDatasetFormat
Service: Amazon SageMaker Service

Represents the CSV dataset format used when running a monitoring job.

Contents

**Header**

Indicates if the CSV data has a header.

Type: Boolean

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
MonitoringDatasetFormat
Service: Amazon SageMaker Service

 Represents the dataset format used when running a monitoring job.

Contents

Csv
The CSV dataset used in the monitoring job.
Type: MonitoringCsvDatasetFormat (p. 1741) object
Required: No

Json
The JSON dataset used in the monitoring job
Type: MonitoringJsonDatasetFormat (p. 1751) object
Required: No

Parquet
The Parquet dataset used in the monitoring job
Type: MonitoringParquetDatasetFormat (p. 1755) object
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
MonitoringExecutionSummary
Service: Amazon SageMaker Service

Summary of information about the last monitoring job to run.

Contents

**CreationTime**

The time at which the monitoring job was created.

Type: Timestamp

Required: Yes

**LastModifiedTime**

A timestamp that indicates the last time the monitoring job was modified.

Type: Timestamp

Required: Yes

**MonitoringExecutionStatus**

The status of the monitoring job.

Type: String

Valid Values: Pending | Completed | CompletedWithViolations | InProgress | Failed | Stopping | Stopped

Required: Yes

**MonitoringScheduleName**

The name of the monitoring schedule.

Type: String


Pattern: ^[a-zA-Z0-9][-\[a-zA-Z0-9]{0,62}$

Required: Yes

**ScheduledTime**

The time the monitoring job was scheduled.

Type: Timestamp

Required: Yes

**EndpointName**

The name of the endpoint used to run the monitoring job.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9][-\[a-zA-Z0-9]{0,62}
Required: No

**FailureReason**

Contains the reason a monitoring job failed, if it failed.

Type: String

Length Constraints: Maximum length of 1024.

Required: No

**MonitoringJobDefinitionName**

The name of the monitoring job.

Type: String


Pattern: ^[a-zA-Z0-9]-*[a-zA-Z0-9]([0-9]62)$

Required: No

**MonitoringType**

The type of the monitoring job.

Type: String

Valid Values: DataQuality | ModelQuality | ModelBias | ModelExplainability

Required: No

**ProcessingJobArn**

The Amazon Resource Name (ARN) of the monitoring job.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z-]*:sagemaker:[a-z0-9-]*:[0-9][12]:processing-job/.*

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
MonitoringGroundTruthS3Input

Service: Amazon SageMaker Service

The ground truth labels for the dataset used for the monitoring job.

Contents

**S3Uri**

The address of the Amazon S3 location of the ground truth labels.

Type: String

Length Constraints: Maximum length of 512.

Pattern: `^((https|s3)://([^/]+)/?)\([^/\]*\)$`

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
MonitoringInput
Service: Amazon SageMaker Service
The inputs for a monitoring job.

Contents

BatchTransformInput
Input object for the batch transform job.
Type: BatchTransformInput (p. 1305) object
Required: No

EndpointInput
The endpoint for a monitoring job.
Type: EndpointInput (p. 1457) object
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
MonitoringJobDefinition
Service: Amazon SageMaker Service
Defines the monitoring job.

Contents

MonitoringAppSpecification
Configures the monitoring job to run a specified Docker container image.
Type: MonitoringAppSpecification (p. 1735) object
Required: Yes

MonitoringInputs
The array of inputs for the monitoring job. Currently we support monitoring an Amazon SageMaker Endpoint.
Type: Array of MonitoringInput (p. 1746) objects
Array Members: Fixed number of 1 item.
Required: Yes

MonitoringOutputConfig
The array of outputs from the monitoring job to be uploaded to Amazon S3.
Type: MonitoringOutputConfig (p. 1754) object
Required: Yes

MonitoringResources
Identifies the resources, ML compute instances, and ML storage volumes to deploy for a monitoring job. In distributed processing, you specify more than one instance.
Type: MonitoringResources (p. 1756) object
Required: Yes

RoleArn
The Amazon Resource Name (ARN) of an IAM role that Amazon SageMaker can assume to perform tasks on your behalf.
Type: String
Pattern: ^arn:aws[a-z\-]*:iam::\d{12}:role/?[a-zA-Z0-9+=,.@\-_/]++$ Required: Yes

BaselineConfig
Baseline configuration used to validate that the data conforms to the specified constraints and statistics
Type: MonitoringBaselineConfig (p. 1737) object
Required: No
Environment

Sets the environment variables in the Docker container.

Type: String to string map

Map Entries: Maximum number of 50 items.

Key Length Constraints: Maximum length of 256.

Key Pattern: [a-zA-Z_] [a-zA-Z0-9_] *

Value Length Constraints: Maximum length of 256.

Value Pattern: [\S\s]*

Required: No

NetworkConfig

Specifies networking options for an monitoring job.

Type: NetworkConfig (p. 1769) object

Required: No

StoppingCondition

Specifies a time limit for how long the monitoring job is allowed to run.

Type: MonitoringStoppingCondition (p. 1765) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
MonitoringJobDefinitionSummary

Service: Amazon SageMaker Service

Summary information about a monitoring job.

Contents

CreationTime

The time that the monitoring job was created.

Type: Timestamp

Required: Yes

EndpointName

The name of the endpoint that the job monitors.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0, 62}$

Required: Yes

MonitoringJobDefinitionArn

The Amazon Resource Name (ARN) of the monitoring job.

Type: String

Length Constraints: Maximum length of 256.

Pattern: .*

Required: Yes

MonitoringJobDefinitionName

The name of the monitoring job.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0, 62}$

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
MonitoringJsonDatasetFormat
Service: Amazon SageMaker Service

Represents the JSON dataset format used when running a monitoring job.

Contents

Line
Indicates if the file should be read as a JSON object per line.

Type: Boolean
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
MonitoringNetworkConfig

Service: Amazon SageMaker Service

The networking configuration for the monitoring job.

Contents

EnableInterContainerTrafficEncryption

Whether to encrypt all communications between the instances used for the monitoring jobs. Choose True to encrypt communications. Encryption provides greater security for distributed jobs, but the processing might take longer.

Type: Boolean
Required: No

EnableNetworkIsolation

Whether to allow inbound and outbound network calls to and from the containers used for the monitoring job.

Type: Boolean
Required: No

VpcConfig

Specifies an Amazon Virtual Private Cloud (VPC) that your SageMaker jobs, hosted models, and compute resources have access to. You can control access to and from your resources by configuring a VPC. For more information, see Give SageMaker Access to Resources in your Amazon VPC.

Type: VpcConfig (p. 2076) object
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
MonitoringOutput

Service: Amazon SageMaker Service

The output object for a monitoring job.

Contents

S3Output

The Amazon S3 storage location where the results of a monitoring job are saved.

Type: MonitoringS3Output (p. 1757) object

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
MonitoringOutputConfig

Service: Amazon SageMaker Service

The output configuration for monitoring jobs.

Contents

MonitoringOutputs

Monitoring outputs for monitoring jobs. This is where the output of the periodic monitoring jobs is uploaded.

Type: Array of MonitoringOutput objects

Array Members: Fixed number of 1 item.

Required: Yes

KmsKeyId

The AWS Key Management Service (AWS KMS) key that Amazon SageMaker uses to encrypt the model artifacts at rest using Amazon S3 server-side encryption.

Type: String

Length Constraints: Maximum length of 2048.

Pattern: .*

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
MonitoringParquetDatasetFormat

Service: Amazon SageMaker Service

Represents the Parquet dataset format used when running a monitoring job.

Contents

The members of this exception structure are context-dependent.

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
MonitoringResources
Service: Amazon SageMaker Service

Identifies the resources to deploy for a monitoring job.

Contents

ClusterConfig

The configuration for the cluster resources used to run the processing job.

Type: MonitoringClusterConfig (p. 1738) object

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
MonitoringS3Output
Service: Amazon SageMaker Service

Information about where and how you want to store the results of a monitoring job.

Contents

LocalPath

The local path to the Amazon S3 storage location where Amazon SageMaker saves the results of a monitoring job. LocalPath is an absolute path for the output data.

Type: String

Length Constraints: Maximum length of 256.

Pattern: .*

Required: Yes

S3Uri

A URI that identifies the Amazon S3 storage location where Amazon SageMaker saves the results of a monitoring job.

Type: String

Length Constraints: Maximum length of 512.

Pattern: ^((https|s3)://([^/]+)?([^/]*)$?

Required: Yes

S3UploadMode

Whether to upload the results of the monitoring job continuously or after the job completes.

Type: String

Valid Values: Continuous | EndOfJob

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
MonitoringSchedule
Service: Amazon SageMaker Service

A schedule for a model monitoring job. For information about model monitor, see Amazon SageMaker Model Monitor.

Contents

CreationTime
The time that the monitoring schedule was created.
Type: Timestamp
Required: No

EndpointName
The endpoint that hosts the model being monitored.
Type: String
Length Constraints: Maximum length of 63.
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$
Required: No

FailureReason
If the monitoring schedule failed, the reason it failed.
Type: String
Length Constraints: Maximum length of 1024.
Required: No

LastModifiedTime
The last time the monitoring schedule was changed.
Type: Timestamp
Required: No

LastMonitoringExecutionSummary
Summary of information about the last monitoring job to run.
Type: MonitoringExecutionSummary (p. 1743) object
Required: No

MonitoringScheduleArn
The Amazon Resource Name (ARN) of the monitoring schedule.
Type: String
Length Constraints: Maximum length of 256.
Pattern: .*
MonitoringScheduleConfig

Configures the monitoring schedule and defines the monitoring job.

Type: MonitoringScheduleConfig (p. 1761) object

MonitoringScheduleName

The name of the monitoring schedule.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

MonitoringScheduleStatus

The status of the monitoring schedule. This can be one of the following values.
- PENDING - The schedule is pending being created.
- FAILED - The schedule failed.
- SCHEDULED - The schedule was successfully created.
- STOPPED - The schedule was stopped.

Type: String

Valid Values: Pending | Failed | Scheduled | Stopped

MonitoringType

The type of the monitoring job definition to schedule.

Type: String

Valid Values: DataQuality | ModelQuality | ModelBias | ModelExplainability

Tags

A list of the tags associated with the monitoring schedule. For more information, see Tagging AWS resources in the AWS General Reference Guide.

Type: Array of Tag (p. 1979) objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for Ruby V3
MonitoringScheduleConfig

Service: Amazon SageMaker Service

Configures the monitoring schedule and defines the monitoring job.

Contents

MonitoringJobDefinition

Defines the monitoring job.

Type: MonitoringJobDefinition (p. 1747) object

Required: No

MonitoringJobDefinitionName

The name of the monitoring job definition to schedule.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])0,62$\n
Required: No

MonitoringType

The type of the monitoring job definition to schedule.

Type: String

Valid Values: DataQuality | ModelQuality | ModelBias | ModelExplainability

Required: No

ScheduleConfig

Configures the monitoring schedule.

Type: ScheduleConfig (p. 1934) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
MonitoringScheduleSummary
Service: Amazon SageMaker Service
Summarizes the monitoring schedule.

Contents

CreationTime
The creation time of the monitoring schedule.
Type: Timestamp
Required: Yes

LastModifiedTime
The last time the monitoring schedule was modified.
Type: Timestamp
Required: Yes

MonitoringScheduleArn
The Amazon Resource Name (ARN) of the monitoring schedule.
Type: String
Length Constraints: Maximum length of 256.
Pattern: .*
Required: Yes

MonitoringScheduleName
The name of the monitoring schedule.
Type: String
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$
Required: Yes

MonitoringScheduleStatus
The status of the monitoring schedule.
Type: String
Valid Values: Pending | Failed | Scheduled | Stopped
Required: Yes

EndpointName
The name of the endpoint using the monitoring schedule.
Type: String
Length Constraints: Maximum length of 63.
MonitoringScheduleSummary

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}$
Required: No

**MonitoringJobDefinitionName**

The name of the monitoring job definition that the schedule is for.

Type: String

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}$
Required: No

**MonitoringType**

The type of the monitoring job definition that the schedule is for.

Type: String

Valid Values: DataQuality | ModelQuality | ModelBias | ModelExplainability
Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
MonitoringStatisticsResource

Service: Amazon SageMaker Service

The statistics resource for a monitoring job.

Contents

S3Uri

The Amazon S3 URI for the statistics resource.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: ^(https|s3)://([^/]+)/?(:[^/]+)?(/[^/])*$

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
MonitoringStoppingCondition

Service: Amazon SageMaker Service

A time limit for how long the monitoring job is allowed to run before stopping.

Contents

MaxRuntimeInSeconds

The maximum runtime allowed in seconds.

Note

The MaxRuntimeInSeconds cannot exceed the frequency of the job. For data quality and model explainability, this can be up to 3600 seconds for an hourly schedule. For model bias and model quality hourly schedules, this can be up to 1800 seconds.

Type: Integer

Valid Range: Minimum value of 1. Maximum value of 86400.

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
**MultiModelConfig**

Service: Amazon SageMaker Service

Specifies additional configuration for hosting multi-model endpoints.

**Contents**

**ModelCacheSetting**

Whether to cache models for a multi-model endpoint. By default, multi-model endpoints cache models so that a model does not have to be loaded into memory each time it is invoked. Some use cases do not benefit from model caching. For example, if an endpoint hosts a large number of models that are each invoked infrequently, the endpoint might perform better if you disable model caching. To disable model caching, set the value of this parameter to `Disabled`.

Type: String

Valid Values: Enabled | Disabled

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
NeoVpcConfig

Service: Amazon SageMaker Service

The **VpcConfig** configuration object that specifies the VPC that you want the compilation jobs to connect to. For more information on controlling access to your Amazon S3 buckets used for compilation job, see [Give Amazon SageMaker Compilation Jobs Access to Resources in Your Amazon VPC](https://docs.aws.amazon.com/sagemaker/latest/dg/how-to-control-access.html).

## Contents

### SecurityGroupIds

The VPC security group IDs. IDs have the form of `sg-xxxxxxxx`. Specify the security groups for the VPC that is specified in the *Subnets* field.

- **Type:** Array of strings
- **Array Members:** Minimum number of 1 item. Maximum number of 5 items.
- **Length Constraints:** Maximum length of 32.
- **Pattern:** `\-[0-9a-zA-Z]+`
- **Required:** Yes

### Subnets

The ID of the subnets in the VPC that you want to connect the compilation job to for accessing the model in Amazon S3.

- **Type:** Array of strings
- **Array Members:** Minimum number of 1 item. Maximum number of 16 items.
- **Length Constraints:** Maximum length of 32.
- **Pattern:** `\-[0-9a-zA-Z]+`
- **Required:** Yes

## See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](https://docs.aws.amazon.com/sdk-for-cpp/v1/developer-guide/api-reference-sagemaker.html)
- [AWS SDK for Go](https://github.com/aws/aws-sdk-go-v2)
- [AWS SDK for Java V2](https://docs.aws.amazon.com/sdk-for-java/v1/developer-guide/sagemaker-api-index.html)
- [AWS SDK for Ruby V3](https://docs.aws.amazon.com/sdk-for-ruby-v3/api/sagemaker/1.0.0/index.html)
NestedFilters

Service: Amazon SageMaker Service

A list of nested Filter objects. A resource must satisfy the conditions of all filters to be included in the results returned from the Search API.

For example, to filter on a training job's InputDataConfig property with a specific channel name and S3Uri prefix, define the following filters:

- '{Name:"InputDataConfig.ChannelName", "Operator":"Equals", "Value":"train"}',
- '{Name:"InputDataConfig.DataSource.S3DataSource.S3Uri", "Operator":"Contains", "Value":"mybucket/catdata"}'

Contents

Filters

A list of filters. Each filter acts on a property. Filters must contain at least one Filters value. For example, a NestedFilters call might include a filter on the PropertyName parameter of the InputDataConfig property: InputDataConfig.DataSource.S3DataSource.S3Uri.

Type: Array of Filter (p. 1496) objects

Array Members: Minimum number of 1 item. Maximum number of 20 items.

Required: Yes

NestedPropertyName

The name of the property to use in the nested filters. The value must match a listed property name, such as InputDataConfig.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 255.

Pattern: .+

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
NetworkConfig
Service: Amazon SageMaker Service

Networking options for a job, such as network traffic encryption between containers, whether to allow
inbound and outbound network calls to and from containers, and the VPC subnets and security groups to
use for VPC-enabled jobs.

Contents

EnableInterContainerTrafficEncryption
Whether to encrypt all communications between distributed processing jobs. Choose True to
crypt communications. Encryption provides greater security for distributed processing jobs, but
the processing might take longer.

Type: Boolean
Required: No

EnableNetworkIsolation
Whether to allow inbound and outbound network calls to and from the containers used for the
processing job.

Type: Boolean
Required: No

VpcConfig
Specifies an Amazon Virtual Private Cloud (VPC) that your SageMaker jobs, hosted models, and
compute resources have access to. You can control access to and from your resources by configuring
a VPC. For more information, see Give SageMaker Access to Resources in your Amazon VPC.

Type: VpcConfig (p. 2076) object
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
NotebookInstanceLifecycleConfigSummary

Service: Amazon SageMaker Service

Provides a summary of a notebook instance lifecycle configuration.

Contents

NotebookInstanceLifecycleConfigArn

The Amazon Resource Name (ARN) of the lifecycle configuration.

Type: String

Length Constraints: Maximum length of 256.

Required: Yes

NotebookInstanceLifecycleConfigName

The name of the lifecycle configuration.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9\-]*[a-zA-Z0-9\-]*

Required: Yes

CreationTime

A timestamp that tells when the lifecycle configuration was created.

Type: Timestamp

Required: No

LastModifiedTime

A timestamp that tells when the lifecycle configuration was last modified.

Type: Timestamp

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
**NotebookInstanceLifecycleHook**

Service: Amazon SageMaker Service

Contains the notebook instance lifecycle configuration script.

Each lifecycle configuration script has a limit of 16384 characters.

The value of the $PATH environment variable that is available to both scripts is /sbin:bin:/usr/sbin:/usr/bin.

View CloudWatch Logs for notebook instance lifecycle configurations in log group /aws/sagemaker/NotebookInstances in log stream [notebook-instance-name]/[LifecycleConfigHook].

Lifecycle configuration scripts cannot run for longer than 5 minutes. If a script runs for longer than 5 minutes, it fails and the notebook instance is not created or started.

For information about notebook instance lifestyle configurations, see [Step 2.1: (Optional) Customize a Notebook Instance](#).

**Contents**

**Content**

A base64-encoded string that contains a shell script for a notebook instance lifecycle configuration.

Type: String


Pattern: [\S\s]+

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
NotebookInstanceSummary

Service: Amazon SageMaker Service

Provides summary information for an SageMaker notebook instance.

Contents

**NotebookInstanceArn**

The Amazon Resource Name (ARN) of the notebook instance.

Type: String

Length Constraints: Maximum length of 256.

Required: Yes

**NotebookInstanceName**

The name of the notebook instance that you want a summary for.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]*)*

Required: Yes

**AdditionalCodeRepositories**

An array of up to three Git repositories associated with the notebook instance. These can be either the names of Git repositories stored as resources in your account, or the URL of Git repositories in AWS CodeCommit or in any other Git repository. These repositories are cloned at the same level as the default repository of your notebook instance. For more information, see [Associating Git Repositories with SageMaker Notebook Instances](#).

Type: Array of strings

Array Members: Maximum number of 3 items.


Pattern: ^https://[^/]+/(.*)|^\[[a-zA-Z0-9](-*[a-zA-Z0-9]*)*$

Required: No

**CreationTime**

A timestamp that shows when the notebook instance was created.

Type: Timestamp

Required: No

**DefaultCodeRepository**

The Git repository associated with the notebook instance as its default code repository. This can be either the name of a Git repository stored as a resource in your account, or the URL of a Git repository in AWS CodeCommit or in any other Git repository. When you open a notebook instance, it opens in the directory that contains this repository. For more information, see [Associating Git Repositories with SageMaker Notebook Instances](#).
<table>
<thead>
<tr>
<th>Field</th>
<th>Type</th>
<th>Constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>NotebookInstanceSummary</td>
<td>Type: String</td>
<td>Length Constraints: Minimum length of 1. Maximum length of 1024.</td>
</tr>
<tr>
<td></td>
<td>Pattern: ^https://([/]+)/(.*$</td>
<td>^<a href="-%5Ba-zA-Z0-9%5D*">a-zA-Z0-9</a>* $</td>
</tr>
<tr>
<td></td>
<td>Required: No</td>
<td></td>
</tr>
<tr>
<td>InstanceType</td>
<td>Type: String</td>
<td>Valid Values: ml.t2.medium</td>
</tr>
<tr>
<td></td>
<td>Required: No</td>
<td></td>
</tr>
<tr>
<td>LastModifiedTime</td>
<td>Type: Timestamp</td>
<td>A timestamp that shows when the notebook instance was last modified.</td>
</tr>
<tr>
<td></td>
<td>Required: No</td>
<td></td>
</tr>
<tr>
<td>NotebookInstanceLifecycleConfigName</td>
<td>Type: String</td>
<td>The name of a notebook instance lifecycle configuration associated with this notebook instance.</td>
</tr>
<tr>
<td></td>
<td>Required: No</td>
<td></td>
</tr>
<tr>
<td>NotebookInstanceStatus</td>
<td>Type: String</td>
<td>Length Constraints: Maximum length of 63.</td>
</tr>
<tr>
<td></td>
<td>Pattern: ^<a href="-%5Ba-zA-Z0-9%5D*">a-zA-Z0-9</a>*</td>
<td>Required: No</td>
</tr>
<tr>
<td></td>
<td>NotebooksInstanceStatus</td>
<td>Type: String</td>
</tr>
</tbody>
</table>
Type: String

Valid Values: Pending | InService | Stopping | Stopped | Failed | Deleting | Updating

Required: No

Url

The URL that you use to connect to the Jupyter notebook running in your notebook instance.

Type: String

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
NotificationConfiguration
Service: Amazon SageMaker Service

Configures Amazon SNS notifications of available or expiring work items for work teams.

Contents

NotificationTopicArn

The ARN for the Amazon SNS topic to which notifications should be published.

Type: String

Pattern: arn:aws[a-z\-]*:sns:[a-z0-9\-]*:[0-9]{12}:a-zA-Z0-9_.-]*

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ObjectiveStatusCounters

Service: Amazon SageMaker Service

Specifies the number of training jobs that this hyperparameter tuning job launched, categorized by the status of their objective metric. The objective metric status shows whether the final objective metric for the training job has been evaluated by the tuning job and used in the hyperparameter tuning process.

Contents

Failed

The number of training jobs whose final objective metric was not evaluated and used in the hyperparameter tuning process. This typically occurs when the training job failed or did not emit an objective metric.

Type: Integer

Valid Range: Minimum value of 0.

Required: No

Pending

The number of training jobs that are in progress and pending evaluation of their final objective metric.

Type: Integer

Valid Range: Minimum value of 0.

Required: No

Succeeded

The number of training jobs whose final objective metric was evaluated by the hyperparameter tuning job and used in the hyperparameter tuning process.

Type: Integer

Valid Range: Minimum value of 0.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
OfflineStoreConfig

Service: Amazon SageMaker Service

The configuration of an OfflineStore.

Provide an OfflineStoreConfig in a request to CreateFeatureGroup to create an OfflineStore.

To encrypt an OfflineStore using at rest data encryption, specify AWS Key Management Service (KMS) key ID, or KMSKeyId, in S3StorageConfig.

Contents

S3StorageConfig

The Amazon Simple Storage (Amazon S3) location of OfflineStore.

Type: S3StorageConfig (p. 1929) object

Required: Yes

DataCatalogConfig

The meta data of the Glue table that is autogenerated when an OfflineStore is created.

Type: DataCatalogConfig (p. 1385) object

Required: No

DisableGlueTableCreation

Set to True to disable the automatic creation of an AWS Glue table when configuring an OfflineStore. If set to False, Feature Store will name the OfflineStore Glue table following Athena's naming recommendations.

The default value is False.

Type: Boolean

Required: No

TableFormat

Format for the offline store table. Supported formats are Glue (Default) and Apache Iceberg.

Type: String

Valid Values: Glue | Iceberg

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
OfflineStoreStatus
Service: Amazon SageMaker Service

The status of OfflineStore.

Contents

Status

An OfflineStore status.
Type: String
Valid Values: Active | Blocked | Disabled
Required: Yes

BlockedReason

The justification for why the OfflineStoreStatus is Blocked (if applicable).
Type: String
Length Constraints: Maximum length of 1024.
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
OidcConfig
Service: Amazon SageMaker Service

Use this parameter to configure your OIDC Identity Provider (IdP).

Contents

AuthorizationEndpoint

The OIDC IdP authorization endpoint used to configure your private workforce.

Type: String
Length Constraints: Maximum length of 500.
Pattern: https://\S+
Required: Yes

ClientId

The OIDC IdP client ID used to configure your private workforce.

Type: String
Pattern: \[ -~\]+ 
Required: Yes

ClientSecret

The OIDC IdP client secret used to configure your private workforce.

Type: String
Pattern: \[ -~\]+ 
Required: Yes

Issuer

The OIDC IdP issuer used to configure your private workforce.

Type: String
Length Constraints: Maximum length of 500.
Pattern: https://\S+
Required: Yes

JwksUri

The OIDC IdP JSON Web Key Set (Jwks) URI used to configure your private workforce.

Type: String
Length Constraints: Maximum length of 500.
Pattern: https://\S+
**Required: Yes**

**LogoutEndpoint**

The OIDC IdP logout endpoint used to configure your private workforce.

*Type: String*

*Length Constraints: Maximum length of 500.*

*Pattern: https://\S+

**Required: Yes**

**TokenEndpoint**

The OIDC IdP token endpoint used to configure your private workforce.

*Type: String*

*Length Constraints: Maximum length of 500.*

*Pattern: https://\S+

**Required: Yes**

**UserInfoEndpoint**

The OIDC IdP user information endpoint used to configure your private workforce.

*Type: String*

*Length Constraints: Maximum length of 500.*

*Pattern: https://\S+

**Required: Yes**

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
OidcConfigForResponse

Service: Amazon SageMaker Service

Your OIDC IdP workforce configuration.

Contents

AuthorizationEndpoint

The OIDC IdP authorization endpoint used to configure your private workforce.

Type: String

Length Constraints: Maximum length of 500.

Pattern: https://\S+

Required: No

ClientId

The OIDC IdP client ID used to configure your private workforce.

Type: String


Pattern: [ -~]+

Required: No

Issuer

The OIDC IdP issuer used to configure your private workforce.

Type: String

Length Constraints: Maximum length of 500.

Pattern: https://\S+

Required: No

JwksUri

The OIDC IdP JSON Web Key Set (Jwks) URI used to configure your private workforce.

Type: String

Length Constraints: Maximum length of 500.

Pattern: https://\S+

Required: No

LogoutEndpoint

The OIDC IdP logout endpoint used to configure your private workforce.

Type: String

Length Constraints: Maximum length of 500.

Pattern: https://\S+
Required: No

**TokenEndpoint**

The OIDC IdP token endpoint used to configure your private workforce.

Type: String

Length Constraints: Maximum length of 500.

Pattern: https://\S+

Required: No

**UserInfoEndpoint**

The OIDC IdP user information endpoint used to configure your private workforce.

Type: String

Length Constraints: Maximum length of 500.

Pattern: https://\S+

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
OidcMemberDefinition
Service: Amazon SageMaker Service

A list of user groups that exist in your OIDC Identity Provider (IdP). One to ten groups can be used to create a single private work team. When you add a user group to the list of Groups, you can add that user group to one or more private work teams. If you add a user group to a private work team, all workers in that user group are added to the work team.

Contents

Groups

A list of comma separated strings that identifies user groups in your OIDC IdP. Each user group is made up of a group of private workers.

Type: Array of strings

Array Members: Minimum number of 1 item. Maximum number of 10 items.


Pattern: [\p{L}\p{M}\p{S}\p{N}\p{P}]+

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
OnlineStoreConfig
Service: Amazon SageMaker Service

Use this to specify the AWS Key Management Service (KMS) Key ID, or KMSKeyId, for at rest data encryption. You can turn OnlineStore on or off by specifying the EnableOnlineStore flag at General Assembly.

The default value is False.

Contents

EnableOnlineStore

Turn OnlineStore off by specifying False for the EnableOnlineStore flag. Turn OnlineStore on by specifying True for the EnableOnlineStore flag.

The default value is False.
Type: Boolean
Required: No

SecurityConfig

Use to specify KMS Key ID (KMSKeyId) for at-rest encryption of your OnlineStore.

Type: OnlineStoreSecurityConfig (p. 1787) object
Required: No

StorageType

Option for different tiers of low latency storage for real-time data retrieval.
• Standard: A managed low latency data store for feature groups.
• InMemory: A managed data store for feature groups that supports very low latency retrieval.

Type: String
Valid Values: Standard | InMemory
Required: No

TtlDuration

Time to live duration, where the record is hard deleted after the expiration time is reached; ExpiresAt = EventTime + TtlDuration. For information on HardDelete, see the DeleteRecord API in the Amazon SageMaker API Reference guide.

Type: TtlDuration (p. 2059) object
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

• AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for Ruby V3
OnlineStoreConfigUpdate

Service: Amazon SageMaker Service

Updates the feature group online store configuration.

Contents

TtlDuration

Time to live duration, where the record is hard deleted after the expiration time is reached; ExpiresAt = EventTime + TtlDuration. For information on HardDelete, see the DeleteRecord API in the Amazon SageMaker API Reference guide.

Type: TtlDuration (p. 2059) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
OnlineStoreSecurityConfig
Service: Amazon SageMaker Service

The security configuration for OnlineStore.

Contents

KmsKeyId
The AWS Key Management Service (KMS) key ARN that SageMaker Feature Store uses to encrypt the Amazon S3 objects at rest using Amazon S3 server-side encryption.

The caller (either user or IAM role) of CreateFeatureGroup must have below permissions to the OnlineStore KmsKeyId:
• "kms:Encrypt"
• "kms:Decrypt"
• "kms:DescribeKey"
• "kms:CreateGrant"
• "kms:RetireGrant"
• "kms:ReEncryptFrom"
• "kms:ReEncryptTo"
• "kms:GenerateDataKey"
• "kms:ListAliases"
• "kms:ListGrants"
• "kms:RevokeGrant"

The caller (either user or IAM role) to all DataPlane operations (PutRecord, GetRecord, DeleteRecord) must have the following permissions to the KmsKeyId:
• "kms:Decrypt"

Type: String
Length Constraints: Maximum length of 2048.
Pattern: .*
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:
• AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for Ruby V3
OutputConfig

Service: Amazon SageMaker Service

Contains information about the output location for the compiled model and the target device that the model runs on. TargetDevice and TargetPlatform are mutually exclusive, so you need to choose one between the two to specify your target device or platform. If you cannot find your device you want to use from the TargetDevice list, use TargetPlatform to describe the platform of your edge device and CompilerOptions if there are specific settings that are required or recommended to use for particular TargetPlatform.

Contents

S3OutputLocation

Identifies the S3 bucket where you want Amazon SageMaker to store the model artifacts. For example, s3://bucket-name/key-name-prefix.

Type: String
Length Constraints: Maximum length of 1024.
Pattern: ^(https|s3)://([^/]+)/?(.*)$
Required: Yes

CompilerOptions

Specifies additional parameters for compiler options in JSON format. The compiler options are TargetPlatform specific. It is required for NVIDIA accelerators and highly recommended for CPU compilations. For any other cases, it is optional to specify CompilerOptions.

- **DTYPE**: Specifies the data type for the input. When compiling for ml_* (except for ml_inf) instances using PyTorch framework, provide the data type (dtype) of the model's input. "float32" is used if "DTYPE" is not specified. Options for data type are:
  - float32: Use either "float" or "float32".
  - int64: Use either "int64" or "long".

  For example, {"dtype" : "float32"}.

- **CPU**: Compilation for CPU supports the following compiler options.
  - **mcpu**: CPU micro-architecture. For example, {'mcpu': 'skylake-avx512'}
  - **mattr**: CPU flags. For example, {'mattr': ['+neon', '+vfpv4']}

- **ARM**: Details of ARM CPU compilations.
  - **NEON**: NEON is an implementation of the Advanced SIMD extension used in ARMv7 processors.

    For example, add {'mattr': ['+neon']} to the compiler options if compiling for ARM 32-bit platform with the NEON support.

- **NVIDIA**: Compilation for NVIDIA GPU supports the following compiler options.
  - **gpu_code**: Specifies the targeted architecture.
  - **trt-ver**: Specifies the TensorRT versions in x.y.z format.
  - **cuda-ver**: Specifies the CUDA version in x.y format.

    For example, {'gpu-code': 'sm_72', 'trt-ver': '6.0.1', 'cuda-ver': '10.1'}

- **ANDROID**: Compilation for the Android OS supports the following compiler options:
  - **ANDROID_PLATFORM**: Specifies the Android API levels. Available levels range from 21 to 29. For example, {'ANDROID_PLATFORM': 28}.
• **mattr**: Add `{'mattr': ['+neon']}` to compiler options if compiling for ARM 32-bit platform with NEON support.

• **INFERENTIA**: Compilation for target ml_inf1 uses compiler options passed in as a JSON string. For example, "CompilerOptions": "\"--verbose 1 --num-neuroncores 2 -O2\"".

For information about supported compiler options, see [Neuron Compiler CLI Reference Guide](#).

• **CoreML**: Compilation for the CoreML **OutputConfig** TargetDevice supports the following compiler options:

  • **class_labels**: Specifies the classification labels file name inside input tar.gz file. For example, `{"class_labels": "imagenet_labels_1000.txt"}`. Labels inside the txt file should be separated by newlines.

• **EIA**: Compilation for the Elastic Inference Accelerator supports the following compiler options:

  • **precision_mode**: Specifies the precision of compiled artifacts. Supported values are "FP16" and "FP32". Default is "FP32".

  • **signature_def_key**: Specifies the signature to use for models in SavedModel format. Defaults is TensorFlow's default signature def key.

  • **output_names**: Specifies a list of output tensor names for models in FrozenGraph format. Set at most one API field, either: signature_def_key or output_names.

For example: `{"precision_mode": "FP32", "output_names": ["output:0"]}

**Type**: String

**Length Constraints**: Minimum length of 3. Maximum length of 1024.

**Pattern**: .*

**Required**: No

---

**KmsKeyld**

The AWS Key Management Service key (AWS KMS) that Amazon SageMaker uses to encrypt your output models with Amazon S3 server-side encryption after compilation job. If you don't provide a KMS key ID, Amazon SageMaker uses the default KMS key for Amazon S3 for your role's account. For more information, see [KMS-Managed Encryption Keys](#) in the [Amazon Simple Storage Service Developer Guide](#).

The KmsKeyld can be any of the following formats:

- **Key ID**: 1234abcd-12ab-34cd-56ef-1234567890ab
- **Key ARN**: arn:aws:kms:us-west-2:111122223333:key/1234abcd-12ab-34cd-56ef-1234567890ab
- **Alias name**: alias/ExampleAlias
- **Alias name ARN**: arn:aws:kms:us-west-2:111122223333:alias/ExampleAlias

**Type**: String

**Length Constraints**: Maximum length of 2048.

**Pattern**: .*

**Required**: No

---

**TargetDevice**

Identifies the target device or the machine learning instance that you want to run your model on after the compilation has completed. Alternatively, you can specify OS, architecture, and accelerator using **TargetPlatform** fields. It can be used instead of **TargetPlatform**.
Note
Currently ml_trn1 is available only in US East (N. Virginia) Region, and ml_inf2 is available only in US East (Ohio) Region.

Type: String

Valid Values: lambda | ml_m4 | ml_m5 | ml_c4 | ml_c5 | ml_p2 | ml_p3 | ml_g4dn | ml_inf1 | ml_inf2 | ml_trn1 | ml_eia2 | jetson_tx1 | jetson_tx2 | jetson纳米 | jetson_xavier | rasp3b | imx8qm | deeplens | rk3399 | rk3288 | sbe_c | qcs605 | qcs603 | sitara_am57x | amba_cv2 | amba_cv22 | amba_cv25 | x86_win32 | x86_win64 | coreml | jacinto_tda4vm | imx8mplus

Required: No

TargetPlatform

Contains information about a target platform that you want your model to run on, such as OS, architecture, and accelerators. It is an alternative of TargetDevice.

The following examples show how to configure the TargetPlatform and CompilerOptions JSON strings for popular target platforms:

- Raspberry Pi 3 Model B+
  
  "TargetPlatform": {"Os": "LINUX", "Arch": "ARM_EABIHF"},
  "CompilerOptions": {'mattr': ['+neon']}

- Jetson TX2
  
  "TargetPlatform": {"Os": "LINUX", "Arch": "ARM64", "Accelerator": "NVIDIA"},
  "CompilerOptions": {'gpu-code': 'sm_62', 'trt-ver': '6.0.1', 'cuda-ver': '10.0'}

- EC2 m5.2xlarge instance OS
  
  "TargetPlatform": {"Os": "LINUX", "Arch": "X86_64", "Accelerator": "NVIDIA"},
  "CompilerOptions": {'mcpu': 'skylake-avx512'}

- RK3399
  
  "TargetPlatform": {"Os": "LINUX", "Arch": "ARM64", "Accelerator": "MALI"}

- ARMv7 phone (CPU)
  
  "TargetPlatform": {"Os": "ANDROID", "Arch": "ARM_EABI"},
  "CompilerOptions": {'ANDROID_PLATFORM': 25, 'mattr': ['+neon']}

- ARMv8 phone (CPU)
  
  "TargetPlatform": {"Os": "ANDROID", "Arch": "ARM64"},
  "CompilerOptions": {'ANDROID_PLATFORM': 29}

Type: TargetPlatform object

Required: No
See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
OutputDataConfig

Service: Amazon SageMaker Service

Provides information about how to store model training results (model artifacts).

Contents

S3OutputPath

Identifies the S3 path where you want SageMaker to store the model artifacts. For example, s3://bucket-name/key-name-prefix.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: ^(https|s3)://([^/]+)/?(.*)$

Required: Yes

CompressionType

The model output compression type. Select None to output an uncompressed model, recommended for large model outputs. Defaults to gzip.

Type: String

Valid Values: GZIP | NONE

Required: No

KmsKeyId

The AWS Key Management Service (AWS KMS) key that SageMaker uses to encrypt the model artifacts at rest using Amazon S3 server-side encryption. The KmsKeyId can be any of the following formats:

• // KMS Key ID

"1234abcd-12ab-34cd-56ef-1234567890ab"

• // Amazon Resource Name (ARN) of a KMS Key

"arn:aws:kms:us-west-2:111122223333:key/1234abcd-12ab-34cd-56ef-1234567890ab"

• // KMS Key Alias

"alias/ExampleAlias"

• // Amazon Resource Name (ARN) of a KMS Key Alias


If you use a KMS key ID or an alias of your KMS key, the SageMaker execution role must include permissions to call kms:Encrypt. If you don't provide a KMS key ID, SageMaker uses the default KMS key for Amazon S3 for your role's account. For more information, see KMS-Managed Encryption Keys in the Amazon Simple Storage Service Developer Guide. If the output data is stored in Amazon S3 Express One Zone, it is encrypted with server-side encryption with Amazon S3 managed keys (SSE-S3). KMS key is not supported for Amazon S3 Express One Zone

The KMS key policy must grant permission to the IAM role that you specify in your CreateTrainingJob, CreateTransformJob, or CreateHyperParameterTuningJob requests.
For more information, see [Using Key Policies in AWS KMS](https://docs.aws.amazon.com/kms/latest/developerguide/) in the *AWS Key Management Service Developer Guide*.

Type: String

Length Constraints: Maximum length of 2048.

Pattern: .*

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](https://docs.aws.amazon.com/sdk-for-cpp/v1/developer-guide/getting-started.html)
- [AWS SDK for Go](https://docs.aws.amazon.com/sdk-for-go/v1/developer-guide/getting-started.html)
- [AWS SDK for Java V2](https://docs.aws.amazon.com/java-sdk/latest/developer-guide/getting-started.html)
- [AWS SDK for Ruby V3](https://docs.aws.amazon.com/ruby-sdk/latest/developer-guide/getting-started.html)
OutputParameter
Service: Amazon SageMaker Service
An output parameter of a pipeline step.

Contents

Name
The name of the output parameter.
Type: String
Length Constraints: Maximum length of 256.
Required: Yes

Value
The value of the output parameter.
Type: String
Length Constraints: Maximum length of 1024.
Required: Yes

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
OwnershipSettings
Service: Amazon SageMaker Service

The collection of ownership settings for a space.

Contents

OwnerUserProfileName

The user profile who is the owner of the private space.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
OwnershipSettingsSummary

Service: Amazon SageMaker Service

Specifies summary information about the ownership settings.

Contents

OwnerUserProfileName

The user profile who is the owner of the private space.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ParallelismConfiguration
Service: Amazon SageMaker Service

Configuration that controls the parallelism of the pipeline. By default, the parallelism configuration specified applies to all executions of the pipeline unless overridden.

Contents

MaxParallelExecutionSteps

The max number of steps that can be executed in parallel.

Type: Integer

Valid Range: Minimum value of 1.

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
Parameter
Service: Amazon SageMaker Service
Assigns a value to a named Pipeline parameter.

Contents

Name
The name of the parameter to assign a value to. This parameter name must match a named parameter in the pipeline definition.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 256.
Pattern: ^[A-Za-z0-9-\_]*$
Required: Yes

Value
The literal value for the parameter.
Type: String
Length Constraints: Maximum length of 1024.
Required: Yes

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ParameterRange
Service: Amazon SageMaker Service

Defines the possible values for categorical, continuous, and integer hyperparameters to be used by an algorithm.

Contents

CategoricalParameterRangeSpecification

A CategoricalParameterRangeSpecification object that defines the possible values for a categorical hyperparameter.

Type: CategoricalParameterRangeSpecification (p. 1323) object

Required: No

ContinuousParameterRangeSpecification

A ContinuousParameterRangeSpecification object that defines the possible values for a continuous hyperparameter.

Type: ContinuousParameterRangeSpecification (p. 1374) object

Required: No

IntegerParameterRangeSpecification

A IntegerParameterRangeSpecification object that defines the possible values for an integer hyperparameter.

Type: IntegerParameterRangeSpecification (p. 1614) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ParameterRanges

Service: Amazon SageMaker Service

Specifies ranges of integer, continuous, and categorical hyperparameters that a hyperparameter tuning job searches. The hyperparameter tuning job launches training jobs with hyperparameter values within these ranges to find the combination of values that result in the training job with the best performance as measured by the objective metric of the hyperparameter tuning job.

Note
The maximum number of items specified for Array Members refers to the maximum number of hyperparameters for each range and also the maximum for the hyperparameter tuning job itself. That is, the sum of the number of hyperparameters for all the ranges can't exceed the maximum number specified.

Contents

AutoParameters

A list containing hyperparameter names and example values to be used by Autotune to determine optimal ranges for your tuning job.

Type: Array of AutoParameter (p. 1298) objects

Array Members: Minimum number of 0 items. Maximum number of 100 items.

Required: No

CategoricalParameterRanges

The array of CategoricalParameterRange objects that specify ranges of categorical hyperparameters that a hyperparameter tuning job searches.

Type: Array of CategoricalParameterRange (p. 1322) objects

Array Members: Minimum number of 0 items. Maximum number of 30 items.

Required: No

ContinuousParameterRanges

The array of ContinuousParameterRange objects that specify ranges of continuous hyperparameters that a hyperparameter tuning job searches.

Type: Array of ContinuousParameterRange (p. 1372) objects

Array Members: Minimum number of 0 items. Maximum number of 30 items.

Required: No

IntegerParameterRanges

The array of IntegerParameterRange objects that specify ranges of integer hyperparameters that a hyperparameter tuning job searches.

Type: Array of IntegerParameterRange (p. 1612) objects

Array Members: Minimum number of 0 items. Maximum number of 30 items.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:
• AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for Ruby V3
Parent

Service: Amazon SageMaker Service

The trial that a trial component is associated with and the experiment the trial is part of. A component might not be associated with a trial. A component can be associated with multiple trials.

Contents

ExperimentName

The name of the experiment.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 120.
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,119}
Required: No

TrialName

The name of the trial.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 120.
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,119}
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ParentHyperParameterTuningJob

Service: Amazon SageMaker Service

A previously completed or stopped hyperparameter tuning job to be used as a starting point for a new hyperparameter tuning job.

Contents

HyperParameterTuningJobName

The name of the hyperparameter tuning job to be used as a starting point for a new hyperparameter tuning job.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,31}$

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
PendingDeploymentSummary

Service: Amazon SageMaker Service

The summary of an in-progress deployment when an endpoint is creating or updating with a new endpoint configuration.

Contents

EndpointConfigName

The name of the endpoint configuration used in the deployment.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

Required: Yes

ProductionVariants

An array of PendingProductionVariantSummary objects, one for each model hosted behind this endpoint for the in-progress deployment.

Type: Array of PendingProductionVariantSummary objects

Array Members: Minimum number of 1 item.

Required: No

ShadowProductionVariants

An array of PendingProductionVariantSummary objects, one for each model hosted behind this endpoint in shadow mode with production traffic replicated from the model specified on ProductionVariants for the in-progress deployment.

Type: Array of PendingProductionVariantSummary objects

Array Members: Minimum number of 1 item.

Required: No

StartTime

The start time of the deployment.

Type: Timestamp

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
PendingProductionVariantSummary

Service: Amazon SageMaker Service

The production variant summary for a deployment when an endpoint is creating or updating with the CreateEndpoint or UpdateEndpoint operations. Describes the VariantStatus, weight and capacity for a production variant associated with an endpoint.

Contents

**VariantName**

The name of the variant.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]{0,62}){0,62}

Required: Yes

**AcceleratorType**

The size of the Elastic Inference (EI) instance to use for the production variant. EI instances provide on-demand GPU computing for inference. For more information, see [Using Elastic Inference in Amazon SageMaker](#).

Type: String

Valid Values: ml.eia1.medium | ml.eia1.large | ml.eia1.xlarge | ml.eia2.medium | ml.eia2.large | ml.eia2.xlarge

Required: No

**CurrentInstanceCount**

The number of instances associated with the variant.

Type: Integer

Valid Range: Minimum value of 0.

Required: No

**CurrentServerlessConfig**

The serverless configuration for the endpoint.

Type: [ProductionVariantServerlessConfig](#) object

Required: No

**CurrentWeight**

The weight associated with the variant.

Type: Float

Valid Range: Minimum value of 0.

Required: No
**DeployedImages**

An array of DeployedImage objects that specify the Amazon EC2 Container Registry paths of the inference images deployed on instances of this ProductionVariant.

Type: Array of DeployedImage (p. 1404) objects

Required: No

**DesiredInstanceCount**

The number of instances requested in this deployment, as specified in the endpoint configuration for the endpoint. The value is taken from the request to the CreateEndpointConfig operation.

Type: Integer

Valid Range: Minimum value of 0.

Required: No

**DesiredServerlessConfig**

The serverless configuration requested for this deployment, as specified in the endpoint configuration for the endpoint.

Type: ProductionVariantServerlessConfig (p. 1855) object

Required: No

**DesiredWeight**

The requested weight for the variant in this deployment, as specified in the endpoint configuration for the endpoint. The value is taken from the request to the CreateEndpointConfig operation.

Type: Float

Valid Range: Minimum value of 0.

Required: No

**InstanceType**

The type of instances associated with the variant.

Type: String

Valid Values: ml.t2.medium | ml.t2.large | ml.t2.xlarge | ml.t2.2xlarge | ml.m4.xlarge | ml.m4.2xlarge | ml.m4.4xlarge | ml.m4.10xlarge | ml.m4.16xlarge | ml.m5.large | ml.m5.xlarge | ml.m5.2xlarge | ml.m5.4xlarge | ml.m5.12xlarge | ml.m5.24xlarge | ml.c4.large | ml.c4.xlarge | ml.c4.2xlarge | ml.c4.4xlarge | ml.c4.8xlarge | ml.p2.xlarge | ml.p2.8xlarge | ml.p2.16xlarge | ml.p3.2xlarge | ml.p3.8xlarge | ml.p3.16xlarge | ml.c5.large | ml.c5.xlarge | ml.c5.2xlarge | ml.c5.4xlarge | ml.c5.9xlarge | ml.c5.18xlarge | ml.c5d.large | ml.c5d.xlarge | ml.c5d.2xlarge | ml.c5d.4xlarge | ml.c5d.9xlarge | ml.c5d.18xlarge | ml.g4dn.xlarge | ml.g4dn.2xlarge | ml.g4dn.4xlarge | ml.g4dn.8xlarge | ml.g4dn.12xlarge | ml.g4dn.16xlarge | ml.r5.large | ml.r5.xlarge | ml.r5.2xlarge | ml.r5.4xlarge | ml.r5.12xlarge | ml.r5.24xlarge | ml.r5d.large | ml.r5d.xlarge | ml.r5d.2xlarge | ml.r5d.4xlarge | ml.r5d.12xlarge | ml.r5d.24xlarge | ml.inf1.xlarge | ml.inf1.2xlarge | ml.inf1.6xlarge | ml.inf1.24xlarge | ml.c6i.large | ml.c6i.xlarge | ml.c6i.2xlarge | ml.c6i.4xlarge | ml.c6i.8xlarge |
### Amazon SageMaker API Reference

#### PendingProductionVariantSummary

| ml.c6i.12xlarge | ml.c6i.16xlarge | ml.c6i.24xlarge | ml.c6i.32xlarge |
| ml.g5.xlarge | ml.g5.2xlarge | ml.g5.4xlarge | ml.g5.8xlarge |
| ml.g5.12xlarge | ml.g5.16xlarge | ml.g5.24xlarge | ml.g5.48xlarge |
| ml.p4d.24xlarge | ml.c7g.large | ml.c7g.xlarge | ml.c7g.2xlarge |
| ml.c7g.4xlarge | ml.c7g.8xlarge | ml.c7g.12xlarge | ml.c7g.16xlarge |
| ml.m6g.large | ml.m6g.xlarge | ml.m6g.2xlarge | ml.m6g.4xlarge |
| ml.m6g.8xlarge | ml.m6g.12xlarge | ml.m6g.16xlarge | ml.m6g.48xlarge |
| ml.p4de.24xlarge | ml.trn1.2xlarge | ml.trn1.32xlarge |
| ml.inf2.xlarge | ml.inf2.8xlarge | ml.inf2.24xlarge | ml.inf2.48xlarge |
| ml.p5.48xlarge |

Required: No

#### ManagedInstanceScaling

Settings that control the range in the number of instances that the endpoint provisions as it scales up or down to accommodate traffic.

Type: `ProductionVariantManagedInstanceScaling` (p. 1853) object

Required: No

#### RoutingConfig

Settings that control how the endpoint routes incoming traffic to the instances that the endpoint hosts.

Type: `ProductionVariantRoutingConfig` (p. 1854) object

Required: No

#### VariantStatus

The endpoint variant status which describes the current deployment stage status or operational status.

Type: Array of `ProductionVariantStatus` (p. 1857) objects

Array Members: Minimum number of 0 items. Maximum number of 5 items.

Required: No

### See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
• AWS SDK for Ruby V3
Phase
Service: Amazon SageMaker Service
Defines the traffic pattern.

Contents

**DurationInSeconds**
- Specifies how long a traffic phase should be. For custom load tests, the value should be between 120 and 3600. This value should not exceed `JobDurationInSeconds`.
- Type: Integer
- Valid Range: Minimum value of 1.
- Required: No

**InitialNumberOfUsers**
- Specifies how many concurrent users to start with. The value should be between 1 and 3.
- Type: Integer
- Valid Range: Minimum value of 1.
- Required: No

**SpawnRate**
- Specified how many new users to spawn in a minute.
- Type: Integer
- Valid Range: Minimum value of 0.
- Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
Pipeline

Service: Amazon SageMaker Service

A SageMaker Model Building Pipeline instance.

Contents

**CreatedBy**

Information about the user who created or modified an experiment, trial, trial component, lineage group, project, or model card.

Type: UserContext (p. 2067) object

Required: No

**CreationTime**

The creation time of the pipeline.

Type: Timestamp

Required: No

**LastModifiedBy**

Information about the user who created or modified an experiment, trial, trial component, lineage group, project, or model card.

Type: UserContext (p. 2067) object

Required: No

**LastModifiedTime**

The time that the pipeline was last modified.

Type: Timestamp

Required: No

**LastRunTime**

The time when the pipeline was last run.

Type: Timestamp

Required: No

**ParallelismConfiguration**

The parallelism configuration applied to the pipeline.

Type: ParallelismConfiguration (p. 1797) object

Required: No

**PipelineArn**

The Amazon Resource Name (ARN) of the pipeline.

Type: String

Length Constraints: Maximum length of 256.
Pipeline

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:pipeline/.*
Required: No

**PipelineDescription**

The description of the pipeline.
Type: String
Length Constraints: Minimum length of 0. Maximum length of 3072.
Pattern: .*
Required: No

**PipelineDisplayName**

The display name of the pipeline.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 256.
Pattern: ^[a-zA-Z0-9\-][^a-zA-Z0-9\-]{0,255}$
Required: No

**PipelineName**

The name of the pipeline.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 256.
Pattern: ^[a-zA-Z0-9\-][^a-zA-Z0-9\-]{0,255}$
Required: No

**PipelineStatus**

The status of the pipeline.
Type: String
Valid Values: Active
Required: No

**RoleArn**

The Amazon Resource Name (ARN) of the role that created the pipeline.
Type: String
Pattern: ^arn:aws[a-zA-Z\-]+:iam::\d{12}:role/\^[a-zA-Z_0-9+=,.@\-_]+/+$
Required: No

**Tags**

A list of tags that apply to the pipeline.
Type: Array of **Tag (p. 1979)** objects
Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
PipelineDefinitionS3Location

Service: Amazon SageMaker Service

The location of the pipeline definition stored in Amazon S3.

Contents

Bucket

Name of the S3 bucket.

Type: String


Pattern: \[a-z0-9\\.\\-a-z0-9][1,61][a-z0-9]

Required: Yes

ObjectKey

The object key (or key name) uniquely identifies the object in an S3 bucket.

Type: String


Pattern: .+

Required: Yes

VersionId

Version Id of the pipeline definition file. If not specified, Amazon SageMaker will retrieve the latest version.

Type: String


Pattern: .+

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
PipelineExecution
Service: Amazon SageMaker Service
An execution of a pipeline.

Contents

CreatedBy
Information about the user who created or modified an experiment, trial, trial component, lineage group, project, or model card.
Type: UserContext (p. 2067) object
Required: No

CreationTime
The creation time of the pipeline execution.
Type: Timestamp
Required: No

FailureReason
If the execution failed, a message describing why.
Type: String
Length Constraints: Maximum length of 1300.
Pattern: .*
Required: No

LastModifiedBy
Information about the user who created or modified an experiment, trial, trial component, lineage group, project, or model card.
Type: UserContext (p. 2067) object
Required: No

LastModifiedTime
The time that the pipeline execution was last modified.
Type: Timestamp
Required: No

ParallelismConfiguration
The parallelism configuration applied to the pipeline execution.
Type: ParallelismConfiguration (p. 1797) object
Required: No

PipelineArn
The Amazon Resource Name (ARN) of the pipeline that was executed.
Type: String
Length Constraints: Maximum length of 256.
Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:pipeline/.*
Required: No

**PipelineExecutionArn**

The Amazon Resource Name (ARN) of the pipeline execution.
Type: String
Length Constraints: Maximum length of 256.
Pattern: ^arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:pipeline\./.*\%/execution\./.*$
Required: No

**PipelineExecutionDescription**

The description of the pipeline execution.
Type: String
Length Constraints: Minimum length of 0. Maximum length of 3072.
Pattern: .*
Required: No

**PipelineExecutionDisplayName**

The display name of the pipeline execution.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 82.
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,81}$
Required: No

**PipelineExecutionStatus**

The status of the pipeline status.
Type: String
Valid Values: Executing | Stopping | Stopped | Failed | Succeeded
Required: No

**PipelineExperimentConfig**

Specifies the names of the experiment and trial created by a pipeline.
Type: PipelineExperimentConfig (p. 1825) object
Required: No

**PipelineParameters**

Contains a list of pipeline parameters. This list can be empty.
Type: Array of Parameter (p. 1798) objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

**SelectiveExecutionConfig**

The selective execution configuration applied to the pipeline run.

Type: SelectiveExecutionConfig (p. 1944) object

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
PipelineExecutionStep
Service: Amazon SageMaker Service

An execution of a step in a pipeline.

Contents

**AttemptCount**
The current attempt of the execution step. For more information, see Retry Policy for SageMaker Pipelines steps.

Type: Integer
Required: No

**CacheHitResult**
If this pipeline execution step was cached, details on the cache hit.

Type: CacheHitResult (p. 1311) object
Required: No

**EndTime**
The time that the step stopped executing.

Type: Timestamp
Required: No

**FailureReason**
The reason why the step failed execution. This is only returned if the step failed its execution.

Type: String
Length Constraints: Maximum length of 1024.
Required: No

**Metadata**
Metadata to run the pipeline step.

Type: PipelineExecutionStepMetadata (p. 1820) object
Required: No

**SelectiveExecutionResult**
The ARN from an execution of the current pipeline from which results are reused for this step.

Type: SelectiveExecutionResult (p. 1945) object
Required: No

**StartTime**
The time that the step started executing.

Type: Timestamp
Required: No
**StepDescription**

The description of the step.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 3072.

Pattern: . *

Required: No

**StepDisplayName**

The display name of the step.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

Pattern: . *

Required: No

**StepName**

The name of the step that is executed.

Type: String

Length Constraints: Maximum length of 64.

Pattern: ^[A-Za-z0-9\-\_]*$

Required: No

**StepStatus**

The status of the step execution.

Type: String

Valid Values: Starting | Executing | Stopping | Stopped | Failed | Succeeded

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
PipelineExecutionStepMetadata
Service: Amazon SageMaker Service
Metadata for a step execution.

Contents

AutoMLJob
The Amazon Resource Name (ARN) of the AutoML job that was run by this step.
Type: AutoMLJobStepMetadata (p. 1286) object
Required: No

Callback
The URL of the Amazon SQS queue used by this step execution, the pipeline generated token, and a list of output parameters.
Type: CallbackStepMetadata (p. 1312) object
Required: No

ClarifyCheck
Container for the metadata for a Clarify check step. The configurations and outcomes of the check step execution. This includes:
• The type of the check conducted,
• The Amazon S3 URIs of baseline constraints and statistics files to be used for the drift check.
• The Amazon S3 URIs of newly calculated baseline constraints and statistics.
• The model package group name provided.
• The Amazon S3 URI of the violation report if violations detected.
• The Amazon Resource Name (ARN) of check processing job initiated by the step execution.
• The boolean flags indicating if the drift check is skipped.
• If step property BaselineUsedForDriftCheck is set the same as CalculatedBaseline.
Type: ClarifyCheckStepMetadata (p. 1329) object
Required: No

Condition
The outcome of the condition evaluation that was run by this step execution.
Type: ConditionStepMetadata (p. 1363) object
Required: No

EMR
The configurations and outcomes of an Amazon EMR step execution.
Type: EMRStepMetadata (p. 1451) object
Required: No

Fail
The configurations and outcomes of a Fail step execution.
Type: **FailStepMetadata** *(p. 1480)* object

Required: No

**Lambda**

The Amazon Resource Name (ARN) of the Lambda function that was run by this step execution and a list of output parameters.

Type: **LambdaStepMetadata** *(p. 1643)* object

Required: No

**Model**

The Amazon Resource Name (ARN) of the model that was created by this step execution.

Type: **ModelStepMetadata** *(p. 1726)* object

Required: No

**ProcessingJob**

The Amazon Resource Name (ARN) of the processing job that was run by this step execution.

Type: **ProcessingJobStepMetadata** *(p. 1837)* object

Required: No

**QualityCheck**

The configurations and outcomes of the check step execution. This includes:

- The type of the check conducted.
- The Amazon S3 URIs of baseline constraints and statistics files to be used for the drift check.
- The Amazon S3 URIs of newly calculated baseline constraints and statistics.
- The model package group name provided.
- The Amazon S3 URI of the violation report if violations detected.
- The Amazon Resource Name (ARN) of check processing job initiated by the step execution.
- The Boolean flags indicating if the drift check is skipped.
- If step property `BaselineUsedForDriftCheck` is set the same as `CalculatedBaseline`.

Type: **QualityCheckStepMetadata** *(p. 1879)* object

Required: No

**RegisterModel**

The Amazon Resource Name (ARN) of the model package that the model was registered to by this step execution.

Type: **RegisterModelStepMetadata** *(p. 1904)* object

Required: No

**TrainingJob**

The Amazon Resource Name (ARN) of the training job that was run by this step execution.

Type: **TrainingJobStepMetadata** *(p. 2013)* object

Required: No

**TransformJob**

The Amazon Resource Name (ARN) of the transform job that was run by this step execution.
Type: TransformJobStepMetadata (p. 2029) object

Required: No

**TuningJob**

The Amazon Resource Name (ARN) of the tuning job that was run by this step execution.

Type: TuningJobStepMetaData (p. 2061) object

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
PipelineExecutionSummary
Service: Amazon SageMaker Service
A pipeline execution summary.

Contents

PipelineExecutionArn
The Amazon Resource Name (ARN) of the pipeline execution.
Type: String
Length Constraints: Maximum length of 256.
Pattern: ^arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:pipeline\./.*\execution\./.*$
Required: No

PipelineExecutionDescription
The description of the pipeline execution.
Type: String
Length Constraints: Minimum length of 0. Maximum length of 3072.
Pattern: .*
Required: No

PipelineExecutionDisplayName
The display name of the pipeline execution.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 82.
Pattern: \^[a-zA-Z0-9](-*[a-zA-Z0-9]{0,81})$.
Required: No

PipelineExecutionFailureReason
A message generated by SageMaker Pipelines describing why the pipeline execution failed.
Type: String
Length Constraints: Maximum length of 3072.
Required: No

PipelineExecutionStatus
The status of the pipeline execution.
Type: String
Valid Values: Executing | Stopping | Stopped | Failed | Succeeded
Required: No
**StartTime**

The start time of the pipeline execution.

Type: Timestamp

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
PipelineExperimentConfig

Specifies the names of the experiment and trial created by a pipeline.

Contents

**ExperimentName**

The name of the experiment.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,119}$`

Required: No

**TrialName**

The name of the trial.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,119}$`

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
**PipelineSummary**

Service: Amazon SageMaker Service

A summary of a pipeline.

**Contents**

**CreationTime**

The creation time of the pipeline.

Type: Timestamp

Required: No

**LastExecutionTime**

The last time that a pipeline execution began.

Type: Timestamp

Required: No

**LastModifiedTime**

The time that the pipeline was last modified.

Type: Timestamp

Required: No

**PipelineArn**

The Amazon Resource Name (ARN) of the pipeline.

Type: String

Length Constraints: Maximum length of 256.

Pattern: `arn:aws[a-zA-z\-]*:sagemaker:[a-zA-z0-9\-]*:[0-9]{12}:pipeline/.*`

Required: No

**PipelineDescription**

The description of the pipeline.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 3072.

Pattern: `.*`

Required: No

**PipelineDisplayName**

The display name of the pipeline.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 256.

Pattern: `^[a-zA-zA-Z0-9](\-*[a-zA-zA-Z0-9])\{0,255}`
PipelineSummary

**Required:** No

**PipelineName**

The name of the pipeline.

*Type:* String

*Length Constraints:* Minimum length of 1. Maximum length of 256.

*Pattern:* `^[a-zA-Z0-9](-*[a-zA-Z0-9])*\{0,255}$`

**RoleArn**

The Amazon Resource Name (ARN) that the pipeline used to execute.

*Type:* String


*Pattern:* `^arn:aws[a-z\-]*:iam::\d{12}:role/?[a-zA-Z0-9\-=,.@\-_\/]+$`

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
**PredefinedMetricSpecification**

Service: Amazon SageMaker Service

A specification for a predefined metric.

## Contents

**PredefinedMetricType**

The metric type. You can only apply SageMaker metric types to SageMaker endpoints.

Type: String

Required: No

## See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
ProcessingClusterConfig
Service: Amazon SageMaker Service
Configuration for the cluster used to run a processing job.

Contents

InstanceCount
The number of ML compute instances to use in the processing job. For distributed processing jobs, specify a value greater than 1. The default value is 1.

Type: Integer
Valid Range: Minimum value of 1. Maximum value of 100.
Required: Yes

InstanceType
The ML compute instance type for the processing job.

Type: String
Valid Values: ml.t3.medium | ml.t3.large | ml.t3.xlarge | ml.t3.2xlarge | ml.m4.xlarge | ml.m4.2xlarge | ml.m4.xlarge | ml.m4.10xlarge | ml.m4.16xlarge | ml.c4.xlarge | ml.c4.2xlarge | ml.c4.4xlarge | ml.c4.8xlarge | ml.p2.xlarge | ml.p2.8xlarge | ml.p2.16xlarge | ml.p3.2xlarge | ml.p3.8xlarge | ml.p3.16xlarge | ml.c5.9xlarge | ml.c5.18xlarge | ml.m5.large | ml.m5.xlarge | ml.m5.2xlarge | ml.m5.4xlarge | ml.m5.12xlarge | ml.r5.large | ml.r5.xlarge | ml.r5.2xlarge | ml.r5.4xlarge | ml.r5.8xlarge | ml.r5.12xlarge | ml.r5.16xlarge | ml.r5.24xlarge | ml.g4dn.xlarge | ml.g4dn.2xlarge | ml.g4dn.4xlarge | ml.g4dn.8xlarge | ml.g4dn.12xlarge | ml.g4dn.16xlarge

Required: Yes

VolumeSizeInGB
The size of the ML storage volume in gigabytes that you want to provision. You must specify sufficient ML storage for your scenario.

Note
Certain Nitro-based instances include local storage with a fixed total size, dependent on the instance type. When using these instances for processing, Amazon SageMaker mounts the local instance storage instead of Amazon EBS gp2 storage. You can't request a VolumeSizeInGB greater than the total size of the local instance storage.

For a list of instance types that support local instance storage, including the total size per instance type, see Instance Store Volumes.

Type: Integer
Required: Yes

VolumeKmsKeyId
The AWS Key Management Service (AWS KMS) key that Amazon SageMaker uses to encrypt data on the storage volume attached to the ML compute instance(s) that run the processing job.
Note
Certain Nitro-based instances include local storage, dependent on the instance type. Local storage volumes are encrypted using a hardware module on the instance. You can't request a VolumeKmsKeyId when using an instance type with local storage. For a list of instance types that support local instance storage, see Instance Store Volumes. For more information about local instance storage encryption, see SSD Instance Store Volumes.

Type: String
Length Constraints: Maximum length of 2048.
Pattern: .*
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ProcessingFeatureStoreOutput

Service: Amazon SageMaker Service

Configuration for processing job outputs in Amazon SageMaker Feature Store.

Contents

FeatureGroupName

The name of the Amazon SageMaker FeatureGroup to use as the destination for processing job output. Note that your processing script is responsible for putting records into your Feature Store.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 64.

Pattern: ^[a-zA-Z0-9]([_-]*[a-zA-Z0-9])\[0,63]\)

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ProcessingInput

Service: Amazon SageMaker Service

The inputs for a processing job. The processing input must specify exactly one of either S3Input or DatasetDefinition types.

Contents

**InputName**

The name for the processing job input.

Type: String

Required: Yes

**AppManaged**

When True, input operations such as data download are managed natively by the processing job application. When False (default), input operations are managed by Amazon SageMaker.

Type: Boolean

Required: No

**DatasetDefinition**

Configuration for a Dataset Definition input.

Type: [DatasetDefinition](p. 1392) object

Required: No

**S3Input**

Configuration for downloading input data from Amazon S3 into the processing container.

Type: [ProcessingS3Input](p. 1843) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
ProcessingJob

Service: Amazon SageMaker Service

An Amazon SageMaker processing job that is used to analyze data and evaluate models. For more information, see Process Data and Evaluate Models.

Contents

AppSpecification

    Configuration to run a processing job in a specified container image.

    Type: AppSpecification (p. 1250) object

    Required: No

AutoMLJobArn

    The Amazon Resource Name (ARN) of the AutoML job associated with this processing job.

    Type: String

    Length Constraints: Minimum length of 1. Maximum length of 256.

    Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:automl-job/.*

    Required: No

CreationTime

    The time the processing job was created.

    Type: Timestamp

    Required: No

Environment

    Sets the environment variables in the Docker container.

    Type: String to string map

    Map Entries: Maximum number of 100 items.

    Key Length Constraints: Maximum length of 256.

    Key Pattern: [\a-zA-Z\_]\[a-zA-Z0-9\-\]*

    Value Length Constraints: Maximum length of 256.

    Value Pattern: [\S\s]*

    Required: No

ExitMessage

    A string, up to one KB in size, that contains metadata from the processing container when the processing job exits.

    Type: String

    Length Constraints: Maximum length of 1024.

    Pattern: [\S\s]*
Required: No

**ExperimentConfig**

Associates a SageMaker job as a trial component with an experiment and trial. Specified when you call the following APIs:

- CreateProcessingJob
- CreateTrainingJob
- CreateTransformJob

Type: ExperimentConfig (p. 1473) object

Required: No

**FailureReason**

A string, up to one KB in size, that contains the reason a processing job failed, if it failed.

Type: String

Length Constraints: Maximum length of 1024.

Required: No

**LastModifiedTime**

The time the processing job was last modified.

Type: Timestamp

Required: No

**MonitoringScheduleArn**

The ARN of a monitoring schedule for an endpoint associated with this processing job.

Type: String

Length Constraints: Maximum length of 256.

Pattern: . *

Required: No

**NetworkConfig**

Networking options for a job, such as network traffic encryption between containers, whether to allow inbound and outbound network calls to and from containers, and the VPC subnets and security groups to use for VPC-enabled jobs.

Type: NetworkConfig (p. 1769) object

Required: No

**ProcessingEndTime**

The time that the processing job ended.

Type: Timestamp

Required: No

**ProcessingInputs**

List of input configurations for the processing job.
Type: Array of ProcessingInput (p. 1832) objects

Array Members: Minimum number of 0 items. Maximum number of 10 items.

Required: No
ProcessingJobArn

The ARN of the processing job.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:a-z0-9\-*:0-9\{12\}:processing-job/.*

Required: No
ProcessingJobName

The name of the processing job.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}

Required: No
ProcessingJobStatus

The status of the processing job.

Type: String

Valid Values: InProgress | Completed | Failed | Stopping | Stopped

Required: No
ProcessingOutputConfig

Configuration for uploading output from the processing container.

Type: ProcessingOutputConfig (p. 1841) object

Required: No
ProcessingResources

Identifies the resources, ML compute instances, and ML storage volumes to deploy for a processing job. In distributed training, you specify more than one instance.

Type: ProcessingResources (p. 1842) object

Required: No
ProcessingStartTime

The time that the processing job started.

Type: Timestamp

Required: No
RoleArn

The ARN of the role used to create the processing job.
ProcessingJob

Type: String


Pattern: ^arn:aws[a-z\-]*:iam::\d{12}:role/?[a-zA-Z_0-9+=,.@\-_\/]+$

Required: No

StoppingCondition

Configures conditions under which the processing job should be stopped, such as how long the processing job has been running. After the condition is met, the processing job is stopped.

Type: ProcessingStoppingCondition (p. 1846) object

Required: No

Tags

An array of key-value pairs. For more information, see Using Cost Allocation Tags in the AWS Billing and Cost Management User Guide.

Type: Array of Tag (p. 1979) objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

TrainingJobArn

The ARN of the training job associated with this processing job.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-\*:\[0-9]\{12\}:training-job/.*

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ProcessingJobStepMetadata
Service: Amazon SageMaker Service
Metadata for a processing job step.

Contents

Arn
The Amazon Resource Name (ARN) of the processing job.
Type: String
Length Constraints: Maximum length of 256.
Pattern: \[arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:processing-job/\.*
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ProcessingJobSummary

Service: Amazon SageMaker Service

Summary of information about a processing job.

Contents

**CreationTime**

The time at which the processing job was created.

Type: Timestamp

Required: Yes

**ProcessingJobArn**

The Amazon Resource Name (ARN) of the processing job.

Type: String

Length Constraints: Maximum length of 256.

Pattern: `arn:aws[a-zA-Z-]*:sagemaker:[a-zA-Z0-9-]*:[0-9]{12}:processing-job/.*`

Required: Yes

**ProcessingJobName**

The name of the processing job.

Type: String


Pattern: `^[a-zA-Z0-9-]*[^a-zA-Z0-9]{0,62}`

Required: Yes

**ProcessingJobStatus**

The status of the processing job.

Type: String

Valid Values: InProgress | Completed | Failed | Stopping | Stopped

Required: Yes

**ExitMessage**

An optional string, up to one KB in size, that contains metadata from the processing container when the processing job exits.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: `[\S\s]*`

Required: No

**FailureReason**

A string, up to one KB in size, that contains the reason a processing job failed, if it failed.
Type: String
Length Constraints: Maximum length of 1024.
Required: No

**LastModifiedTime**
A timestamp that indicates the last time the processing job was modified.
Type: Timestamp
Required: No

**ProcessingEndTime**
The time at which the processing job completed.
Type: Timestamp
Required: No

**See Also**
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](https://aws.amazon.com/sdk-for-cpp/
- [AWS SDK for Go](https://aws.amazon.com/sdk-for-go/
- [AWS SDK for Java V2](https://aws.amazon.com/sdk-for-java/
- [AWS SDK for Ruby V3](https://aws.amazon.com/sdk-for-ruby/
ProcessingOutput
Service: Amazon SageMaker Service

Describes the results of a processing job. The processing output must specify exactly one of either S3Output or FeatureStoreOutput types.

Contents

OutputName
The name for the processing job output.
Type: String
Required: Yes

AppManaged
When True, output operations such as data upload are managed natively by the processing job application. When False (default), output operations are managed by Amazon SageMaker.
Type: Boolean
Required: No

FeatureStoreOutput
Configuration for processing job outputs in Amazon SageMaker Feature Store. This processing output type is only supported when AppManaged is specified.
Type: ProcessingFeatureStoreOutput (p. 1831) object
Required: No

S3Output
Configuration for processing job outputs in Amazon S3.
Type: ProcessingS3Output (p. 1845) object
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ProcessingOutputConfig

Service: Amazon SageMaker Service

Configuration for uploading output from the processing container.

Contents

Outputs

An array of outputs configuring the data to upload from the processing container.

Type: Array of ProcessingOutput (p. 1840) objects

Array Members: Minimum number of 0 items. Maximum number of 10 items.

Required: Yes

KmsKeyId

The AWS Key Management Service (AWS KMS) key that Amazon SageMaker uses to encrypt the processing job output. KmsKeyId can be an ID of a KMS key, ARN of a KMS key, alias of a KMS key, or alias of a KMS key. The KmsKeyId is applied to all outputs.

Type: String

Length Constraints: Maximum length of 2048.

Pattern: .*

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
ProcessingResources
Service: Amazon SageMaker Service

Identifies the resources, ML compute instances, and ML storage volumes to deploy for a processing job. In distributed training, you specify more than one instance.

Contents

ClusterConfig

The configuration for the resources in a cluster used to run the processing job.

Type: ProcessingClusterConfig (p. 1829) object

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ProcessingS3Input
Service: Amazon SageMaker Service

Configuration for downloading input data from Amazon S3 into the processing container.

Contents

S3DataType

Whether you use an S3Prefix or a ManifestFile for the data type. If you choose S3Prefix, S3Uri identifies a key name prefix. Amazon SageMaker uses all objects with the specified key name prefix for the processing job. If you choose ManifestFile, S3Uri identifies an object that is a manifest file containing a list of object keys that you want Amazon SageMaker to use for the processing job.

Type: String

Valid Values: ManifestFile | S3Prefix

Required: Yes

S3Uri

The URI of the Amazon S3 prefix Amazon SageMaker downloads data required to run a processing job.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: ^(https|s3)://([^/]+)/(.*)$

Required: Yes

LocalPath

The local path in your container where you want Amazon SageMaker to write input data to. LocalPath is an absolute path to the input data and must begin with /opt/ml/processing/. LocalPath is a required parameter when AppManaged is False (default).

Type: String

Length Constraints: Maximum length of 256.

Pattern: .*

Required: No

S3CompressionType

Whether to GZIP-decompress the data in Amazon S3 as it is streamed into the processing container. Gzip can only be used when Pipe mode is specified as the S3InputMode. In Pipe mode, Amazon SageMaker streams input data from the source directly to your container without using the EBS volume.

Type: String

Valid Values: None | Gzip

Required: No
**S3DataDistributionType**

Whether to distribute the data from Amazon S3 to all processing instances with FullyReplicated, or whether the data from Amazon S3 is shared by Amazon S3 key, downloading one shard of data to each processing instance.

Type: String

Valid Values: FullyReplicated | ShardedByS3Key

Required: No

**S3InputMode**

Whether to use File or Pipe input mode. In File mode, Amazon SageMaker copies the data from the input source onto the local ML storage volume before starting your processing container. This is the most commonly used input mode. In Pipe mode, Amazon SageMaker streams input data from the source directly to your processing container into named pipes without using the ML storage volume.

Type: String

Valid Values: Pipe | File

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
ProcessingS3Output
Service: Amazon SageMaker Service

Configuration for uploading output data to Amazon S3 from the processing container.

Contents

LocalPath

The local path of a directory where you want Amazon SageMaker to upload its contents to Amazon S3. LocalPath is an absolute path to a directory containing output files. This directory will be created by the platform and exist when your container's entrypoint is invoked.

Type: String
Length Constraints: Maximum length of 256.
Pattern: .*
Required: Yes

S3UploadMode

Whether to upload the results of the processing job continuously or after the job completes.

Type: String
Valid Values: Continuous | EndOfJob
Required: Yes

S3Uri

A URI that identifies the Amazon S3 bucket where you want Amazon SageMaker to save the results of a processing job.

Type: String
Length Constraints: Maximum length of 1024.
Pattern: ^(https|s3)://([^/]+)/?([^/]+)$
Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ProcessingStoppingCondition
Service: Amazon SageMaker Service

Configures conditions under which the processing job should be stopped, such as how long the processing job has been running. After the condition is met, the processing job is stopped.

Contents

MaxRuntimeInSeconds

Specifies the maximum runtime in seconds.

Type: Integer


Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ProductionVariant
Service: Amazon SageMaker Service

Identifies a model that you want to host and the resources chosen to deploy for hosting it. If you are deploying multiple models, tell SageMaker how to distribute traffic among the models by specifying variant weights. For more information on production variants, check Production variants.

Contents

VariantName

The name of the production variant.
Type: String
Length Constraints: Maximum length of 63.
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$
Required: Yes

AcceleratorType

The size of the Elastic Inference (EI) instance to use for the production variant. EI instances provide on-demand GPU computing for inference. For more information, see Using Elastic Inference in Amazon SageMaker.
Type: String
Valid Values: ml.eia1.medium | ml.eia1.large | ml.eia1.xlarge | ml.eia2.medium | ml.eia2.large | ml.eia2.xlarge
Required: No

ContainerStartupHealthCheckTimeoutInSeconds

The timeout value, in seconds, for your inference container to pass health check by SageMaker Hosting. For more information about health check, see How Your Container Should Respond to Health Check (Ping) Requests.
Type: Integer
Valid Range: Minimum value of 60. Maximum value of 3600.
Required: No

CoreDumpConfig

Specifies configuration for a core dump from the model container when the process crashes.
Type: ProductionVariantCoreDumpConfig (p. 1851) object
Required: No

EnableSSMAccess

You can use this parameter to turn on native AWS Systems Manager (SSM) access for a production variant behind an endpoint. By default, SSM access is disabled for all production variants behind an endpoint. You can turn on or turn off SSM access for a production variant behind an existing endpoint by creating a new endpoint configuration and calling UpdateEndpoint.
Type: Boolean
**InitialInstanceCount**

Number of instances to launch initially.

Type: Integer

Valid Range: Minimum value of 1.

Required: No

**InitialVariantWeight**

Determines initial traffic distribution among all of the models that you specify in the endpoint configuration. The traffic to a production variant is determined by the ratio of the VariantWeight to the sum of all VariantWeight values across all ProductionVariants. If unspecified, it defaults to 1.0.

Type: Float

Valid Range: Minimum value of 0.

Required: No

**InstanceType**

The ML compute instance type.

Type: String

Valid Values: `ml.t2.medium`, `ml.t2.large`, `ml.t2.xlarge`, `ml.t2.2xlarge`, `ml.m4.xlarge`, `ml.m4.2xlarge`, `ml.m4.4xlarge`, `ml.m4.10xlarge`, `ml.m4.16xlarge`, `ml.m5.large`, `ml.m5.xlarge`, `ml.m5.2xlarge`, `ml.m5.4xlarge`, `ml.m5.12xlarge`, `ml.m5.24xlarge`, `ml.m5d.large`, `ml.m5d.xlarge`, `ml.m5d.2xlarge`, `ml.m5d.4xlarge`, `ml.m5d.12xlarge`, `ml.m5d.24xlarge`, `ml.c4.large`, `ml.c4.xlarge`, `ml.c4.2xlarge`, `ml.c4.4xlarge`, `ml.c4.8xlarge`, `ml.p2.xlarge`, `ml.p2.8xlarge`, `ml.p2.16xlarge`, `ml.p3.2xlarge`, `ml.p3.8xlarge`, `ml.p3.16xlarge`, `ml.c5.large`, `ml.c5.xlarge`, `ml.c5.2xlarge`, `ml.c5.4xlarge`, `ml.c5.9xlarge`, `ml.c5.18xlarge`, `ml.c5d.large`, `ml.c5d.xlarge`, `ml.c5d.2xlarge`, `ml.c5d.4xlarge`, `ml.c5d.9xlarge`, `ml.c5d.18xlarge`, `ml.g4dn.xlarge`, `ml.g4dn.2xlarge`, `ml.g4dn.4xlarge`, `ml.g4dn.8xlarge`, `ml.g4dn.12xlarge`, `ml.g4dn.16xlarge`, `ml.r5.large`, `ml.r5.xlarge`, `ml.r5.2xlarge`, `ml.r5.4xlarge`, `ml.r5.12xlarge`, `ml.r5.24xlarge`, `ml.r5d.large`, `ml.r5d.xlarge`, `ml.r5d.2xlarge`, `ml.r5d.4xlarge`, `ml.r5d.12xlarge`, `ml.r5d.24xlarge`, `ml.inf1.xlarge`, `ml.inf1.2xlarge`, `ml.inf1.6xlarge`, `ml.inf1.24xlarge`, `ml.c6i.large`, `ml.c6i.xlarge`, `ml.c6i.2xlarge`, `ml.c6i.4xlarge`, `ml.c6i.8xlarge`, `ml.c6i.12xlarge`, `ml.c6i.16xlarge`, `ml.c6i.24xlarge`, `ml.c6i.32xlarge`, `ml.g5.large`, `ml.g5.xlarge`, `ml.g5.2xlarge`, `ml.g5.4xlarge`, `ml.g5.8xlarge`, `ml.g5.12xlarge`, `ml.g5.16xlarge`, `ml.g5.24xlarge`, `ml.g5.48xlarge`, `ml.p4d.24xlarge`, `ml.p4d.4xlarge`, `ml.p4d.8xlarge`, `ml.p4d.16xlarge`, `ml.m6g.large`, `ml.m6g.xlarge`, `ml.m6g.2xlarge`, `ml.m6g.4xlarge`, `ml.m6g.8xlarge`, `ml.m6g.12xlarge`, `ml.m6g.16xlarge`, `ml.m6g.32xlarge`, `ml.m6gd.large`, `ml.m6gd.xlarge`, `ml.m6gd.2xlarge`, `ml.m6gd.4xlarge`, `ml.m6gd.8xlarge`, `ml.m6gd.12xlarge`, `ml.m6gd.16xlarge`, `ml.m6gd.32xlarge`, `ml.m6gn.large`, `ml.m6gn.xlarge`, `ml.m6gn.2xlarge`, `ml.m6gn.4xlarge`
| ml.c6gn.8xlarge | ml.c6gn.12xlarge | ml.c6gn.16xlarge | ml.r6g.large |
| ml.r6g.xlarge | ml.r6g.2xlarge | ml.r6g.4xlarge | ml.r6g.8xlarge |
| ml.r6g.12xlarge | ml.r6g.16xlarge | ml.r6gd.large | ml.r6gd.xlarge |
| ml.r6gd.2xlarge | ml.r6gd.4xlarge | ml.r6gd.8xlarge | ml.r6gd.12xlarge |
| ml.r6gd.16xlarge | ml.p4de.24xlarge | ml.trn1.2xlarge | ml.trn1.32xlarge |
| ml.inf2.xlarge | ml.inf2.8xlarge | ml.inf2.24xlarge | ml.inf2.48xlarge |
| ml.p5.48xlarge |

Required: No

**ManagedInstanceScaling**

Settings that control the range in the number of instances that the endpoint provisions as it scales up or down to accommodate traffic.

Type: `ProductionVariantManagedInstanceScaling (p. 1853)` object

Required: No

**ModelDataDownloadTimeoutInSeconds**

The timeout value, in seconds, to download and extract the model that you want to host from Amazon S3 to the individual inference instance associated with this production variant.

Type: Integer

Valid Range: Minimum value of 60. Maximum value of 3600.

Required: No

**ModelName**

The name of the model that you want to host. This is the name that you specified when creating the model.

Type: String

Length Constraints: Maximum length of 63.

Pattern: `^[a-zA-Z0-9][-]*[a-zA-Z0-9-]*`

Required: No

**RoutingConfig**

Settings that control how the endpoint routes incoming traffic to the instances that the endpoint hosts.

Type: `ProductionVariantRoutingConfig (p. 1854)` object

Required: No

**ServerlessConfig**

The serverless configuration for an endpoint. Specifies a serverless endpoint configuration instead of an instance-based endpoint configuration.

Type: `ProductionVariantServerlessConfig (p. 1855)` object

Required: No

**VolumeSizeInGB**

The size, in GB, of the ML storage volume attached to individual inference instance associated with the production variant. Currently only Amazon EBS gp2 storage volumes are supported.
Type: Integer


Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ProductionVariantCoreDumpConfig

Service: Amazon SageMaker Service

Specifies configuration for a core dump from the model container when the process crashes.

Contents

DestinationS3Uri

The Amazon S3 bucket to send the core dump to.

Type: String

Length Constraints: Maximum length of 512.

Pattern: ^(https|s3)://\([^/\\]()/?\(.*\)$

Required: Yes

KmsKeyId

The AWS Key Management Service (AWS KMS) key that SageMaker uses to encrypt the core dump data at rest using Amazon S3 server-side encryption. The KmsKeyId can be any of the following formats:

- // KMS Key ID
  "1234abcd-12ab-34cd-56ef-1234567890ab"
- // Amazon Resource Name (ARN) of a KMS Key
  "arn:aws:kms:us-west-2:111122223333:key/1234abcd-12ab-34cd-56ef-1234567890ab"
- // KMS Key Alias
  "alias/ExampleAlias"
- // Amazon Resource Name (ARN) of a KMS Key Alias

If you use a KMS key ID or an alias of your KMS key, the SageMaker execution role must include permissions to call kms:Encrypt. If you don't provide a KMS key ID, SageMaker uses the default KMS key for Amazon S3 for your role's account. SageMaker uses server-side encryption with KMS-managed keys for OutputDataConfig. If you use a bucket policy with an s3:PutObject permission that only allows objects with server-side encryption, set the condition key of s3:x-amz-server-side-encryption to "aws:kms". For more information, see KMS-Managed Encryption Keys in the Amazon Simple Storage Service Developer Guide.

The KMS key policy must grant permission to the IAM role that you specify in your CreateEndpoint and UpdateEndpoint requests. For more information, see Using Key Policies in AWS KMS in the AWS Key Management Service Developer Guide.

Type: String

Length Constraints: Maximum length of 2048.

Pattern: .*

Required: No
See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ProductionVariantManagedInstanceScaling

Service: Amazon SageMaker Service

Settings that control the range in the number of instances that the endpoint provisions as it scales up or down to accommodate traffic.

Contents

MaxInstanceCount

The maximum number of instances that the endpoint can provision when it scales up to accommodate an increase in traffic.

Type: Integer

Valid Range: Minimum value of 1.

Required: No

MinInstanceCount

The minimum number of instances that the endpoint must retain when it scales down to accommodate a decrease in traffic.

Type: Integer

Valid Range: Minimum value of 1.

Required: No

Status

Indicates whether managed instance scaling is enabled.

Type: String

Valid Values: ENABLED | DISABLED

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ProductionVariantRoutingConfig

Service: Amazon SageMaker Service

Settings that control how the endpoint routes incoming traffic to the instances that the endpoint hosts.

Contents

RoutingStrategy

Sets how the endpoint routes incoming traffic:

- LEAST_OUTSTANDING_REQUESTS: The endpoint routes requests to the specific instances that have more capacity to process them.
- RANDOM: The endpoint routes each request to a randomly chosen instance.

Type: String

Valid Values: LEAST_OUTSTANDING_REQUESTS | RANDOM

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ProductionVariantServerlessConfig

Specifies the serverless configuration for an endpoint variant.

Contents

MaxConcurrency

The maximum number of concurrent invocations your serverless endpoint can process.

Type: Integer


Required: Yes

MemorySizeInMB

The memory size of your serverless endpoint. Valid values are in 1 GB increments: 1024 MB, 2048 MB, 3072 MB, 4096 MB, 5120 MB, or 6144 MB.

Type: Integer


Required: Yes

ProvisionedConcurrency

The amount of provisioned concurrency to allocate for the serverless endpoint. Should be less than or equal to MaxConcurrency.

Note

This field is not supported for serverless endpoint recommendations for Inference Recommender jobs. For more information about creating an Inference Recommender job, see CreateInferenceRecommendationsJobs.

Type: Integer


Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ProductionVariantServerlessUpdateConfig

Service: Amazon SageMaker Service

Specifies the serverless update concurrency configuration for an endpoint variant.

Contents

MaxConcurrency

The updated maximum number of concurrent invocations your serverless endpoint can process.

Type: Integer


Required: No

ProvisionedConcurrency

The updated amount of provisioned concurrency to allocate for the serverless endpoint. Should be less than or equal to MaxConcurrency.

Type: Integer


Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ProductionVariantStatus
Service: Amazon SageMaker Service
Describes the status of the production variant.

Contents

Status
The endpoint variant status which describes the current deployment stage status or operational status.
- Creating: Creating inference resources for the production variant.
- Deleting: Terminating inference resources for the production variant.
- Updating: Updating capacity for the production variant.
- ActivatingTraffic: Turning on traffic for the production variant.
- Baking: Waiting period to monitor the CloudWatch alarms in the automatic rollback configuration.

Type: String
Valid Values: Creating | Updating | Deleting | ActivatingTraffic | Baking
Required: Yes

StartTime
The start time of the current status change.
Type: Timestamp
Required: No

StatusMessage
A message that describes the status of the production variant.
Type: String
Length Constraints: Maximum length of 1024.
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ProductionVariantSummary

Service: Amazon SageMaker Service

Describes weight and capacities for a production variant associated with an endpoint. If you sent a request to the UpdateEndpointWeightsAndCapacities API and the endpoint status is Updating, you get different desired and current values.

Contents

**VariantName**

The name of the variant.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}$

Required: Yes

**CurrentInstanceCount**

The number of instances associated with the variant.

Type: Integer

Valid Range: Minimum value of 0.

Required: No

**CurrentServerlessConfig**

The serverless configuration for the endpoint.

Type: ProductionVariantServerlessConfig (p. 1855) object

Required: No

**CurrentWeight**

The weight associated with the variant.

Type: Float

Valid Range: Minimum value of 0.

Required: No

**DeployedImages**

An array of DeployedImage objects that specify the Amazon EC2 Container Registry paths of the inference images deployed on instances of this ProductionVariant.

Type: Array of DeployedImage (p. 1404) objects

Required: No

**DesiredInstanceCount**

The number of instances requested in the UpdateEndpointWeightsAndCapacities request.

Type: Integer
Valid Range: Minimum value of 0.
Required: No

**DesiredServerlessConfig**

The serverless configuration requested for the endpoint update.

Type: `ProductionVariantServerlessConfig (p. 1855)` object

Required: No

**DesiredWeight**

The requested weight, as specified in the `UpdateEndpointWeightsAndCapacities` request.

Type: Float

Valid Range: Minimum value of 0.

Required: No

**ManagedInstanceScaling**

Settings that control the range in the number of instances that the endpoint provisions as it scales up or down to accommodate traffic.

Type: `ProductionVariantManagedInstanceScaling (p. 1853)` object

Required: No

**RoutingConfig**

Settings that control how the endpoint routes incoming traffic to the instances that the endpoint hosts.

Type: `ProductionVariantRoutingConfig (p. 1854)` object

Required: No

**VariantStatus**

The endpoint variant status which describes the current deployment stage status or operational status.

Type: Array of `ProductionVariantStatus (p. 1857)` objects

Array Members: Minimum number of 0 items. Maximum number of 5 items.

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ProfilerConfig

Service: Amazon SageMaker Service

Configuration information for Amazon SageMaker Debugger system monitoring, framework profiling, and storage paths.

Contents

DisableProfiler

Configuration to turn off Amazon SageMaker Debugger's system monitoring and profiling functionality. To turn it off, set to True.

Type: Boolean
Required: No

ProfilingIntervalInMilliseconds

A time interval for capturing system metrics in milliseconds. Available values are 100, 200, 500, 1000 (1 second), 5000 (5 seconds), and 60000 (1 minute) milliseconds. The default value is 500 milliseconds.

Type: Long
Required: No

ProfilingParameters

Configuration information for capturing framework metrics. Available key strings for different profiling options are DetailedProfilingConfig, PythonProfilingConfig, and DataLoaderProfilingConfig. The following codes are configuration structures for the ProfilingParameters parameter. To learn more about how to configure the ProfilingParameters parameter, see Use the SageMaker and Debugger Configuration API Operations to Create, Update, and Debug Your Training Job.

Type: String to string map
Map Entries: Minimum number of 0 items. Maximum number of 20 items.
Key Length Constraints: Minimum length of 1. Maximum length of 256.
Key Pattern: .*
Value Length Constraints: Maximum length of 256.
Value Pattern: .*
Required: No

S3OutputPath

Path to Amazon S3 storage location for system and framework metrics.

Type: String
Length Constraints: Maximum length of 1024.
Pattern: ^(https|s3)://([^/]+)?(.*$)
Required: No
See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ProfilerConfigForUpdate

Service: Amazon SageMaker Service

Configuration information for updating the Amazon SageMaker Debugger profile parameters, system and framework metrics configurations, and storage paths.

Contents

DisableProfiler

To turn off Amazon SageMaker Debugger monitoring and profiling while a training job is in progress, set to True.

Type: Boolean

Required: No

ProfilingIntervalInMilliseconds

A time interval for capturing system metrics in milliseconds. Available values are 100, 200, 500, 1000 (1 second), 5000 (5 seconds), and 60000 (1 minute) milliseconds. The default value is 500 milliseconds.

Type: Long

Required: No

ProfilingParameters

Configuration information for capturing framework metrics. Available key strings for different profiling options are DetailedProfilingConfig, PythonProfilingConfig, and DataLoaderProfilingConfig. The following codes are configuration structures for the ProfilingParameters parameter. To learn more about how to configure the ProfilingParameters parameter, see Use the SageMaker and Debugger Configuration API Operations to Create, Update, and Debug Your Training Job.

Type: String to string map

Map Entries: Minimum number of 0 items. Maximum number of 20 items.

Key Length Constraints: Minimum length of 1. Maximum length of 256.

Key Pattern: .*

Value Length Constraints: Maximum length of 256.

Value Pattern: .*

Required: No

S3OutputPath

Path to Amazon S3 storage location for system and framework metrics.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: ^(https|s3):/(/[^/]*)/?(.*$)

Required: No
See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ProfilerRuleConfiguration

Service: Amazon SageMaker Service

Configuration information for profiling rules.

Contents

**RuleConfigurationName**

The name of the rule configuration. It must be unique relative to other rule configuration names.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 256.

Pattern: .*

Required: Yes

**RuleEvaluatorImage**

The Amazon Elastic Container Registry Image for the managed rule evaluation.

Type: String

Length Constraints: Maximum length of 255.

Pattern: .*

Required: Yes

**InstanceType**

The instance type to deploy a custom rule for profiling a training job.

Type: String

Valid Values: ml.t3.medium | ml.t3.large | ml.t3.xlarge | ml.t3.2xlarge | ml.m4.xlarge | ml.m4.2xlarge | ml.m4.4xlarge | ml.m4.10xlarge | ml.m4.16xlarge | ml.m4.2xlarge | ml.m4.4xlarge | ml.c4.xlarge | ml.c4.2xlarge | ml.c4.4xlarge | ml.c4.8xlarge | ml.p2.xlarge | ml.p2.8xlarge | ml.p2.16xlarge | ml.p3.2xlarge | ml.p3.8xlarge | ml.p3.16xlarge | ml.c5.xlarge | ml.c5.2xlarge | ml.c5.4xlarge | ml.c5.9xlarge | ml.c5.18xlarge | ml.m5.large | ml.m5.xlarge | ml.m5.2xlarge | ml.m5.4xlarge | ml.m5.12xlarge | ml.m5.24xlarge | ml.r5.large | ml.r5.xlarge | ml.r5.2xlarge | ml.r5.4xlarge | ml.r5.8xlarge | ml.r5.12xlarge | ml.r5.16xlarge | ml.r5.24xlarge | ml.g4dn.xlarge | ml.g4dn.2xlarge | ml.g4dn.4xlarge | ml.g4dn.8xlarge | ml.g4dn.12xlarge | ml.g4dn.16xlarge

Required: No

**LocalPath**

Path to local storage location for output of rules. Defaults to /opt/ml/processing/output/rule/.

Type: String

Length Constraints: Maximum length of 4096.

Pattern: .*
Required: No

**RuleParameters**

Runtime configuration for rule container.

Type: String to string map

Map Entries: Minimum number of 0 items. Maximum number of 100 items.

Key Length Constraints: Minimum length of 1. Maximum length of 256.

Key Pattern: . *

Value Length Constraints: Maximum length of 256.

Value Pattern: . *

Required: No

**S3OutputPath**

Path to Amazon S3 storage location for rules.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: ^(https|s3)://([^/]+)/?(.*)$

Required: No

**VolumeSizeInGB**

The size, in GB, of the ML storage volume attached to the processing instance.

Type: Integer

Valid Range: Minimum value of 0.

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
ProfilerRuleEvaluationStatus
Service: Amazon SageMaker Service

Information about the status of the rule evaluation.

Contents

LastModifiedTime

Timestamp when the rule evaluation status was last modified.

Type: Timestamp

Required: No

RuleConfigurationName

The name of the rule configuration.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 256.

Pattern: .*

Required: No

RuleEvaluationJobArn

The Amazon Resource Name (ARN) of the rule evaluation job.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:processing-job/.*

Required: No

RuleEvaluationStatus

Status of the rule evaluation.

Type: String

Valid Values: InProgress | NoIssuesFound | IssuesFound | Error | Stopping | Stopped

Required: No

StatusDetails

Details from the rule evaluation.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: .*

Required: No
See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
Project
Service: Amazon SageMaker Service
The properties of a project as returned by the Search API.

Contents

CreatedBy
Who created the project.
Type: UserContext (p. 2067) object
Required: No

CreationTime
A timestamp specifying when the project was created.
Type: Timestamp
Required: No

LastModifiedBy
Information about the user who created or modified an experiment, trial, trial component, lineage group, project, or model card.
Type: UserContext (p. 2067) object
Required: No

LastModifiedTime
A timestamp container for when the project was last modified.
Type: Timestamp
Required: No

ProjectArn
The Amazon Resource Name (ARN) of the project.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: ^arn:aws(-cn|-us-gov)?:sagemaker:[a-z0-9\-]{9,16}:\[0-9\]{12}:project/\[\S\]{1,2048}$
Required: No

ProjectDescription
The description of the project.
Type: String
Length Constraints: Maximum length of 1024.

Pattern: ^\p{L}\p{M}\p{Z}\p{S}\p{N}\p{P}]*$
Required: No
**ProjectId**

The ID of the project.

Type: String


Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9])*$`

Required: No

**ProjectName**

The name of the project.

Type: String


Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,31\}`

Required: No

**ProjectStatus**

The status of the project.

Type: String

Valid Values: Pending | CreateInProgress | CreateCompleted | CreateFailed | DeleteInProgress | DeleteFailed | DeleteCompleted | UpdateInProgress | UpdateCompleted | UpdateFailed

Required: No

**ServiceCatalogProvisionedProductDetails**

Details of a provisioned service catalog product. For information about service catalog, see [What is AWS Service Catalog](https://docs.aws.amazon.com/servicecatalog/latest/dg/index.html).

Type: [ServiceCatalogProvisionedProductDetails](#) object

Required: No

**ServiceCatalogProvisioningDetails**

Details that you specify to provision a service catalog product. For information about service catalog, see [What is AWS Service Catalog](https://docs.aws.amazon.com/servicecatalog/latest/dg/index.html).

Type: [ServiceCatalogProvisioningDetails](#) object

Required: No

**Tags**

An array of key-value pairs. You can use tags to categorize your AWS resources in different ways, for example, by purpose, owner, or environment. For more information, see [Tagging AWS Resources](https://docs.aws.amazon.com/general/latest/gr/aws- abbiamo-taxonomy-tags.html).

Type: Array of [Tag](#) objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No
See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
**ProjectSummary**

**Service**: Amazon SageMaker Service

Information about a project.

**Contents**

**CreationTime**

The time that the project was created.

Type: Timestamp

Required: Yes

**ProjectArn**

The Amazon Resource Name (ARN) of the project.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `^arn:aws(-cn|-us-gov)?:sagemaker:[a-z0-9\-\[9,16]\[0-9]\{12\}:project/\S\{1,2048\}$`

Required: Yes

**ProjectId**

The ID of the project.

Type: String


Pattern: `^[a-zA-Z0-9\-]*[a-zA-Z0-9]$`

Required: Yes

**ProjectName**

The name of the project.

Type: String


Pattern: `^[a-zA-Z0-9\-]*[a-zA-Z0-9]$`

Required: Yes

**ProjectStatus**

The status of the project.

Type: String

Valid Values: Pending | CreateInProgress | CreateCompleted | CreateFailed | DeleteInProgress | DeleteFailed | DeleteCompleted | UpdateInProgress | UpdateCompleted | UpdateFailed

Required: Yes
ProjectDescription

The description of the project.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: [\p{L}\p{M}\p{Z}\p{S}\p{N}\p{P}]*

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
PropertyNameQuery
Service: Amazon SageMaker Service

Part of the SuggestionQuery type. Specifies a hint for retrieving property names that begin with the specified text.

Contents

PropertyNameHint
- Text that begins a property's name.
- Type: String
- Length Constraints: Minimum length of 0. Maximum length of 100.
- Pattern: .*
- Required: Yes

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
PropertyNameSuggestion

Service: Amazon SageMaker Service

A property name returned from a GetSearchSuggestions call that specifies a value in the PropertyNameQuery field.

Contents

PropertyName

A suggested property name based on what you entered in the search textbox in the SageMaker console.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 255.

Pattern: .+

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ProvisioningParameter
Service: Amazon SageMaker Service

A key value pair used when you provision a project as a service catalog product. For information, see What is AWS Service Catalog.

Contents

Key

The key that identifies a provisioning parameter.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 1000.

Pattern: .*

Required: No

Value

The value of the provisioning parameter.

Type: String

Length Constraints: Maximum length of 4096.

Pattern: .*

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
PublicWorkforceTaskPrice

Service: Amazon SageMaker Service

Defines the amount of money paid to an Amazon Mechanical Turk worker for each task performed.

Use one of the following prices for bounding box tasks. Prices are in US dollars and should be based on the complexity of the task; the longer it takes in your initial testing, the more you should offer.

- 0.036
- 0.048
- 0.060
- 0.072
- 0.120
- 0.240
- 0.360
- 0.480
- 0.600
- 0.720
- 0.840
- 0.960
- 1.080
- 1.200

Use one of the following prices for image classification, text classification, and custom tasks. Prices are in US dollars.

- 0.012
- 0.024
- 0.036
- 0.048
- 0.060
- 0.072
- 0.120
- 0.240
- 0.360
- 0.480
- 0.600
- 0.720
- 0.840
- 0.960
- 1.080
- 1.200

Use one of the following prices for semantic segmentation tasks. Prices are in US dollars.

- 0.840
- 0.960
- 1.080
Use one of the following prices for Textract AnalyzeDocument Important Form Key Amazon Augmented AI review tasks. Prices are in US dollars.

- 1.200
- 2.400
- 2.280
- 2.160
- 2.040
- 1.920
- 1.800
- 1.680
- 1.560
- 1.440
- 1.320
- 1.200
- 1.080
- 0.960
- 0.840
- 0.720
- 0.600
- 0.480
- 0.360
- 0.240
- 0.120
- 0.072
- 0.060
- 0.048
- 0.036
- 0.024
- 0.012

Use one of the following prices for Rekognition DetectModerationLabels Amazon Augmented AI review tasks. Prices are in US dollars.

- 1.200
- 1.080
- 0.960
- 0.840
- 0.720
- 0.600
- 0.480
- 0.360
- 0.240
- 0.120
- 0.072
- 0.060
• 0.048  
• 0.036  
• 0.024  
• 0.012

Use one of the following prices for Amazon Augmented AI custom human review tasks. Prices are in US dollars.
• 1.200  
• 1.080  
• 0.960  
• 0.840  
• 0.720  
• 0.600  
• 0.480  
• 0.360  
• 0.240  
• 0.120  
• 0.072  
• 0.060  
• 0.048  
• 0.036  
• 0.024  
• 0.012

**Contents**

**AmountInUsd**

Defines the amount of money paid to an Amazon Mechanical Turk worker in United States dollars.

Type: [USD](p. 2066) object

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

• [AWS SDK for C++](#)  
• [AWS SDK for Go](#)  
• [AWS SDK for Java V2](#)  
• [AWS SDK for Ruby V3](#)
QualityCheckStepMetadata

Service: Amazon SageMaker Service

Container for the metadata for a Quality check step. For more information, see the topic on QualityCheck step in the Amazon SageMaker Developer Guide.

Contents

BaselineUsedForDriftCheckConstraints

The Amazon S3 URI of the baseline constraints file used for the drift check.

Type: String

Length Constraints: Maximum length of 1024.

Required: No

BaselineUsedForDriftCheckStatistics

The Amazon S3 URI of the baseline statistics file used for the drift check.

Type: String

Length Constraints: Maximum length of 1024.

Required: No

CalculatedBaselineConstraints

The Amazon S3 URI of the newly calculated baseline constraints file.

Type: String

Length Constraints: Maximum length of 1024.

Required: No

CalculatedBaselineStatistics

The Amazon S3 URI of the newly calculated baseline statistics file.

Type: String

Length Constraints: Maximum length of 1024.

Required: No

CheckJobArn

The Amazon Resource Name (ARN) of the Quality check processing job that was run by this step execution.

Type: String

Length Constraints: Maximum length of 256.

Required: No

CheckType

The type of the Quality check step.

Type: String
Length Constraints: Maximum length of 256.

Required: No

**ModelPackageGroupName**

The model package group name.

Type: String

Length Constraints: Maximum length of 256.

Required: No

**RegisterNewBaseline**

This flag indicates if a newly calculated baseline can be accessed through step properties `BaselineUsedForDriftCheckConstraints` and `BaselineUsedForDriftCheckStatistics`. If it is set to `False`, the previous baseline of the configured check type must also be available. These can be accessed through the `BaselineUsedForDriftCheckConstraints` and `BaselineUsedForDriftCheckStatistics` properties.

Type: Boolean

Required: No

**SkipCheck**

This flag indicates if the drift check against the previous baseline will be skipped or not. If it is set to `False`, the previous baseline of the configured check type must be available.

Type: Boolean

Required: No

**ViolationReport**

The Amazon S3 URI of violation report if violations are detected.

Type: String

Length Constraints: Maximum length of 1024.

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
QueryFilters
Service: Amazon SageMaker Service

A set of filters to narrow the set of lineage entities connected to the StartArn(s) returned by the QueryLineage API action.

Contents

CreatedAfter
  Filter the lineage entities connected to the StartArn(s) after the create date.
  Type: Timestamp
  Required: No

CreatedBefore
  Filter the lineage entities connected to the StartArn(s) by created date.
  Type: Timestamp
  Required: No

LineageTypes
  Filter the lineage entities connected to the StartArn(s) by the type of the lineage entity.
  Type: Array of strings
  Array Members: Maximum number of 4 items.
  Valid Values: TrialComponent | Artifact | Context | Action
  Required: No

ModifiedAfter
  Filter the lineage entities connected to the StartArn(s) after the last modified date.
  Type: Timestamp
  Required: No

ModifiedBefore
  Filter the lineage entities connected to the StartArn(s) before the last modified date.
  Type: Timestamp
  Required: No

Properties
  Filter the lineage entities connected to the StartArn(s) by a set if property key value pairs. If multiple pairs are provided, an entity is included in the results if it matches any of the provided pairs.
  Type: String to string map
  Map Entries: Maximum number of 5 items.
  Key Length Constraints: Maximum length of 256.
  Value Length Constraints: Maximum length of 256.
QueryFilters

Required: No

Types

Filter the lineage entities connected to the StartArn by type. For example: DataSet, Model, Endpoint, or ModelDeployment.

Type: Array of strings

Array Members: Maximum number of 5 items.

Length Constraints: Maximum length of 40.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
RealTimeInferenceConfig

Service: Amazon SageMaker Service

The infrastructure configuration for deploying the model to a real-time inference endpoint.

Contents

InstanceCount

The number of instances of the type specified by InstanceType.

Type: Integer

Valid Range: Minimum value of 0.

Required: Yes

InstanceType

The instance type the model is deployed to.

Type: String

Valid Values: ml.t2.medium | ml.t2.large | ml.t2.xlarge | ml.t2.2xlarge | ml.t3.medium | ml.t3.large | ml.t3.xlarge | ml.t3.2xlarge | ml.m4.xlarge | ml.m4.2xlarge | ml.m4.4xlarge | ml.m4.10xlarge | ml.m4.16xlarge | ml.m5.xlarge | ml.m5.2xlarge | ml.m5.4xlarge | ml.m5.12xlarge | ml.m5.24xlarge | ml.m5d.large | ml.m5d.xlarge | ml.m5d.2xlarge | ml.m5d.4xlarge | ml.m5d.8xlarge | ml.m5d.12xlarge | ml.m5d.16xlarge | ml.m5d.24xlarge | ml.c4.xlarge | ml.c4.2xlarge | ml.c4.4xlarge | ml.c4.8xlarge | ml.c5.xlarge | ml.c5.2xlarge | ml.c5.4xlarge | ml.c5.9xlarge | ml.c5d.9xlarge | ml.c5d.18xlarge | ml.c5d.2xlarge | ml.c5d.4xlarge | ml.p2.16xlarge | ml.p2.2xlarge | ml.p2.8xlarge | ml.p3.16xlarge | ml.p3.2xlarge | ml.p3.8xlarge | ml.p3.16xlarge | ml.p3dn.24xlarge | ml.g4dn.xlarge | ml.g4dn.2xlarge | ml.g4dn.4xlarge | ml.g4dn.8xlarge | ml.g4dn.12xlarge | ml.g4dn.16xlarge | ml.r5.large | ml.r5.xlarge | ml.r5.2xlarge | ml.r5.4xlarge | ml.r5.8xlarge | ml.g5.large | ml.g5.2xlarge | ml.g5.4xlarge | ml.g5.8xlarge | ml.g5.16xlarge | ml.g5.24xlarge | ml.g5.48xlarge | ml.g5.12xlarge | ml.g5.24xlarge | ml.g5.48xlarge | ml.inf1.xlarge | ml.inf1.2xlarge | ml.inf1.6xlarge | ml.inf1.24xlarge | ml.p4d.24xlarge | ml.p4de.24xlarge

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
RealTimeInferenceRecommendation

Service: Amazon SageMaker Service

The recommended configuration to use for Real-Time Inference.

Contents

InstanceType

The recommended instance type for Real-Time Inference.

Type: String

Valid Values: ml.t2.medium | ml.t2.large | ml.t2.xlarge | ml.t2.2xlarge | ml.m4.xlarge | ml.m4.2xlarge | ml.m4.4xlarge | ml.m4.10xlarge | ml.m4.16xlarge | ml.m5.large | ml.m5.xlarge | ml.m5.2xlarge | ml.m5.4xlarge | ml.m5.12xlarge | ml.m5.24xlarge | ml.m5d.large | ml.m5d.xlarge | ml.m5d.2xlarge | ml.m5d.4xlarge | ml.m5d.12xlarge | ml.m5d.24xlarge | ml.c4.large | ml.c4.xlarge | ml.c4.2xlarge | ml.c4.4xlarge | ml.c4.8xlarge | ml.p2.xlarge | ml.p2.8xlarge | ml.p2.16xlarge | ml.p3.xlarge | ml.p3.8xlarge | ml.p3.16xlarge | ml.c5.large | ml.c5.xlarge | ml.c5.2xlarge | ml.c5.4xlarge | ml.c5.8xlarge | ml.c5.16xlarge | ml.c5.32xlarge | ml.c5d.large | ml.c5d.xlarge | ml.c5d.2xlarge | ml.c5d.4xlarge | ml.c5d.8xlarge | ml.c5d.16xlarge | ml.c5d.32xlarge | ml.g4dn.large | ml.g4dn.xlarge | ml.g4dn.2xlarge | ml.g4dn.4xlarge | ml.g4dn.8xlarge | ml.g4dn.16xlarge | ml.g4dn.32xlarge | ml.m6g.large | ml.m6g.xlarge | ml.m6g.2xlarge | ml.m6g.4xlarge | ml.m6g.8xlarge | ml.m6g.16xlarge | ml.m6g.32xlarge | ml.m6gd.large | ml.m6gd.xlarge | ml.m6gd.2xlarge | ml.m6gd.4xlarge | ml.m6gd.8xlarge | ml.m6gd.16xlarge | ml.m6gd.32xlarge | ml.m6gdn.large | ml.m6gdn.xlarge | ml.m6gdn.2xlarge | ml.m6gdn.4xlarge | ml.m6gdn.8xlarge | ml.m6gdn.16xlarge | ml.m6gdn.32xlarge | ml.r6g.large | ml.r6g.xlarge | ml.r6g.2xlarge | ml.r6g.4xlarge | ml.r6g.8xlarge | ml.r6g.16xlarge | ml.r6g.32xlarge | ml.r6gd.large | ml.r6gd.xlarge | ml.r6gd.2xlarge | ml.r6gd.4xlarge | ml.r6gd.8xlarge | ml.r6gd.16xlarge | ml.r6gd.32xlarge | ml.p4de.large | ml.p4de.xlarge | ml.p4de.2xlarge | ml.p4de.4xlarge | ml.p4de.8xlarge | ml.p4de.16xlarge | ml.p4de.32xlarge | ml.trn1.large | ml.trn1.xlarge | ml.trn1.2xlarge | ml.trn1.32xlarge | ml.inf2.large | ml.inf2.xlarge | ml.inf2.8xlarge | ml.inf2.16xlarge | ml.inf2.24xlarge | ml.inf2.48xlarge | ml.p5.48xlarge

Required: Yes

RecommendationId

The recommendation ID which uniquely identifies each recommendation.

Type: String
Required: Yes

**Environment**

The recommended environment variables to set in the model container for Real-Time Inference.

Type: String to string map

Map Entries: Maximum number of 16 items.

Key Length Constraints: Maximum length of 1024.

Key Pattern: `[a-zA-Z_] [a-zA-Z0-9_] *`

Value Length Constraints: Maximum length of 1024.

Value Pattern: `[^\s\s] *`

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
RecommendationJobCompiledOutputConfig

Service: Amazon SageMaker Service

Provides information about the output configuration for the compiled model.

Contents

S3OutputUri

Identifies the Amazon S3 bucket where you want SageMaker to store the compiled model artifacts.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: ^(https|s3)://([^/]+)/(.*$

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
RecommendationJobContainerConfig

Service: Amazon SageMaker Service

Specifies mandatory fields for running an Inference Recommender job directly in the CreateInferenceRecommendationsJob API. The fields specified in ContainerConfig override the corresponding fields in the model package. Use ContainerConfig if you want to specify these fields for the recommendation job but don't want to edit them in your model package.

Contents

DataInputConfig

Specifies the name and shape of the expected data inputs for your trained model with a JSON dictionary form. This field is used for optimizing your model using SageMaker Neo. For more information, see DataInputConfig.

Type: String


Pattern: [\S\s]+

Required: No

Domain

The machine learning domain of the model and its components.

Valid Values: COMPUTER_VISION | NATURAL_LANGUAGE_PROCESSING | MACHINE_LEARNING

Type: String

Required: No

Framework

The machine learning framework of the container image.

Valid Values: TENSORFLOW | PYTORCH | XGBOOST | SAGEMAKER-SCIKIT-LEARN

Type: String

Required: No

FrameworkVersion

The framework version of the container image.

Type: String

Required: No

NearestModelName

The name of a pre-trained machine learning model benchmarked by Amazon SageMaker Inference Recommender that matches your model.

Valid Values: efficientnetb7 | unet | xgboost | faster-rcnn-resnet101 | nasnetlarge | vgg16 | inception-v3 | mask-rcnn | sagemaker-scikit-learn | densenet201-gluon | resnet18v2-gluon | xception | densenet201 | yolov4 | resnet152 | bert-base-cased | xceptionV1-keras | resnet50 | retinanet

Type: String
PayloadConfig

Specifies the SamplePayloadUrl and all other sample payload-related fields.

Type: RecommendationJobPayloadConfig (p. 1896) object

Required: No

SupportedEndpointType

The endpoint type to receive recommendations for. By default this is null, and the results of the inference recommendation job return a combined list of both real-time and serverless benchmarks. By specifying a value for this field, you can receive a longer list of benchmarks for the desired endpoint type.

Type: String

Valid Values: RealTime | Serverless

Required: No

SupportedInstanceTypes

A list of the instance types that are used to generate inferences in real-time.

Type: Array of strings

Required: No

SupportedResponseMIMETypes

The supported MIME types for the output data.

Type: Array of strings

Length Constraints: Maximum length of 1024.

Pattern: ^[-\w]+\/[.]+$

Required: No

Task

The machine learning task that the model accomplishes.

Valid Values: IMAGE_CLASSIFICATION | OBJECT_DETECTION | TEXT_GENERATION | IMAGE_SEGMENTATION | FILL_MASK | CLASSIFICATION | REGRESSION | OTHER

Type: String

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
RecommendationJobInferenceBenchmark

Service: Amazon SageMaker Service

The details for a specific benchmark from an Inference Recommender job.

Contents

ModelConfiguration

Defines the model configuration. Includes the specification name and environment parameters.

Type: ModelConfiguration (p. 1676) object

Required: Yes

EndpointConfiguration

The endpoint configuration made by Inference Recommender during a recommendation job.

Type: EndpointOutputConfiguration (p. 1464) object

Required: No

EndpointMetrics

The metrics for an existing endpoint compared in an Inference Recommender job.

Type: InferenceMetrics (p. 1594) object

Required: No

FailureReason

The reason why a benchmark failed.

Type: String

Required: No

InvocationEndTime

A timestamp that shows when the benchmark completed.

Type: Timestamp

Required: No

InvocationStartTime

A timestamp that shows when the benchmark started.

Type: Timestamp

Required: No

Metrics

The metrics of recommendations.

Type: RecommendationMetrics (p. 1900) object

Required: No
See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
RecommendationJobInputConfig

Service: Amazon SageMaker Service

The input configuration of the recommendation job.

Contents

ContainerConfig

Specifies mandatory fields for running an Inference Recommender job. The fields specified in ContainerConfig override the corresponding fields in the model package.

Type: RecommendationJobContainerConfig (p. 1887) object

Required: No

EndpointConfigurations

Specifies the endpoint configuration to use for a job.

Type: Array of EndpointInputConfiguration (p. 1460) objects

Array Members: Minimum number of 1 item. Maximum number of 10 items.

Required: No

Endpoints

Existing customer endpoints on which to run an Inference Recommender job.

Type: Array of EndpointInfo (p. 1456) objects

Array Members: Maximum number of 1 item.

Required: No

JobDurationInSeconds

Specifies the maximum duration of the job, in seconds. The maximum value is 18,000 seconds.

Type: Integer

Valid Range: Minimum value of 1.

Required: No

ModelName

The name of the created model.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9-]*[a-zA-Z0-9]*$

Required: No

ModelPackageVersionArn

The Amazon Resource Name (ARN) of a versioned model package.

Type: String
RecommendationJobInputConfig

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: ^arn:aws(-cn|-us-gov)?:sagemaker:[a-z0-9-]{9,16}:[0-9]{12}:model-package/\S{1,2048}$

Required: No

**ResourceLimit**

Defines the resource limit of the job.

Type: [RecommendationJobResourceLimit](p. 1897) object

Required: No

**TrafficPattern**

Specifies the traffic pattern of the job.

Type: [TrafficPattern](p. 1997) object

Required: No

**VolumeKmsKeyId**

The Amazon Resource Name (ARN) of a AWS Key Management Service (AWS KMS) key that Amazon SageMaker uses to encrypt data on the storage volume attached to the ML compute instance that hosts the endpoint. This key will be passed to SageMaker Hosting for endpoint creation.

The SageMaker execution role must have `kms:CreateGrant` permission in order to encrypt data on the storage volume of the endpoints created for inference recommendation. The inference recommendation job will fail asynchronously during endpoint configuration creation if the role passed does not have `kms:CreateGrant` permission.

The KmsKeyId can be any of the following formats:

- **// KMS Key ID**
  ```
  "1234abcd-12ab-34cd-56ef-1234567890ab"
  ```

- **// Amazon Resource Name (ARN) of a KMS Key**
  ```
  "arn:aws:kms:<region>:<account>::key/<key-id-12ab-34cd-56ef-1234567890ab>"
  ```

- **// KMS Key Alias**
  ```
  "alias/ExampleAlias"
  ```

- **// Amazon Resource Name (ARN) of a KMS Key Alias**
  ```
  "arn:aws:kms:<region>:<account>::alias/<ExampleAlias>"
  ```

For more information about key identifiers, see [Key identifiers (KeyID)](p. 1897) in the AWS Key Management Service (AWS KMS) documentation.

Type: String

Length Constraints: Maximum length of 2048.

Pattern: .*

Required: No

**VpcConfig**

Inference Recommender provisions SageMaker endpoints with access to VPC in the inference recommendation job.
Type: RecommendationJobVpcConfig (p. 1899) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
**RecommendationJobOutputConfig**

Service: Amazon SageMaker Service

Provides information about the output configuration for the compiled model.

**Contents**

**CompiledOutputConfig**

Provides information about the output configuration for the compiled model.

Type: `RecommendationJobCompiledOutputConfig (p. 1886)` object

Required: No

**KmsKeyId**

The Amazon Resource Name (ARN) of a AWS Key Management Service (AWS KMS) key that Amazon SageMaker uses to encrypt your output artifacts with Amazon S3 server-side encryption. The SageMaker execution role must have `kms:GenerateDataKey` permission.

The `KmsKeyId` can be any of the following formats:

- // KMS Key ID
  
  "1234abcd-12ab-34cd-56ef-1234567890ab"

- // Amazon Resource Name (ARN) of a KMS Key
  
  "arn:aws:kms:<region>:<account>:key/<key-id-12ab-34cd-56ef-1234567890ab>"

- // KMS Key Alias
  
  "alias/ExampleAlias"

- // Amazon Resource Name (ARN) of a KMS Key Alias
  
  "arn:aws:kms:<region>:<account>:alias/<ExampleAlias>"

For more information about key identifiers, see [Key identifiers (KeyID)](Key identifiers (KeyID)) in the AWS Key Management Service (AWS KMS) documentation.

Type: String

Length Constraints: Maximum length of 2048.

Pattern: .*

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](AWS SDK for C++)
- [AWS SDK for Go](AWS SDK for Go)
- [AWS SDK for Java V2](AWS SDK for Java V2)
- [AWS SDK for Ruby V3](AWS SDK for Ruby V3)
RecommendationJobPayloadConfig

Service: Amazon SageMaker Service

The configuration for the payload for a recommendation job.

Contents

**SamplePayloadUrl**

The Amazon Simple Storage Service (Amazon S3) path where the sample payload is stored. This path must point to a single gzip compressed tar archive (.tar.gz suffix).

Type: String

Required: No

**SupportedContentTypes**

The supported MIME types for the input data.

Type: Array of strings

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
RecommendationJobResourceLimit

Service: Amazon SageMaker Service

Specifies the maximum number of jobs that can run in parallel and the maximum number of jobs that can run.

Contents

MaxNumberOfTests
- Defines the maximum number of load tests.
  - Type: Integer
  - Valid Range: Minimum value of 1.
  - Required: No

MaxParallelOfTests
- Defines the maximum number of parallel load tests.
  - Type: Integer
  - Valid Range: Minimum value of 1.
  - Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
RecommendationJobStoppingConditions
Service: Amazon SageMaker Service

Specifies conditions for stopping a job. When a job reaches a stopping condition limit, SageMaker ends the job.

Contents

FlatInvocations

Stops a load test when the number of invocations (TPS) peaks and flattens, which means that the instance has reached capacity. The default value is Stop. If you want the load test to continue after invocations have flattened, set the value to Continue.

Type: String

Valid Values: Continue | Stop

Required: No

MaxInvocations

The maximum number of requests per minute expected for the endpoint.

Type: Integer

Required: No

ModelLatencyThresholds

The interval of time taken by a model to respond as viewed from SageMaker. The interval includes the local communication time taken to send the request and to fetch the response from the container of a model and the time taken to complete the inference in the container.

Type: Array of ModelLatencyThreshold (p. 1697) objects

Array Members: Fixed number of 1 item.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
RecommendationJobVpcConfig

Service: Amazon SageMaker Service

Inference Recommender provisions SageMaker endpoints with access to VPC in the inference recommendation job.

Contents

SecurityGroupIds

The VPC security group IDs. IDs have the form of sg-xxxxxxxxx. Specify the security groups for the VPC that is specified in the Subnets field.

Type: Array of strings
Array Members: Minimum number of 1 item. Maximum number of 5 items.
Length Constraints: Maximum length of 32.
Pattern: [-0-9a-zA-Z]+
Required: Yes

Subnets

The ID of the subnets in the VPC to which you want to connect your model.

Type: Array of strings
Array Members: Minimum number of 1 item. Maximum number of 16 items.
Length Constraints: Maximum length of 32.
Pattern: [-0-9a-zA-Z]+
Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
RecommendationMetrics
Service: Amazon SageMaker Service

The metrics of recommendations.

Contents

CostPerHour

Defines the cost per hour for the instance.

Type: Float

Required: Yes

CostPerInference

Defines the cost per inference for the instance.

Type: Float

Required: Yes

MaxInvocations

The expected maximum number of requests per minute for the instance.

Type: Integer

Required: Yes

ModelLatency

The expected model latency at maximum invocation per minute for the instance.

Type: Integer

Required: Yes

CpuUtilization

The expected CPU utilization at maximum invocations per minute for the instance.

NaN indicates that the value is not available.

Type: Float

Valid Range: Minimum value of 0.0.

Required: No

MemoryUtilization

The expected memory utilization at maximum invocations per minute for the instance.

NaN indicates that the value is not available.

Type: Float

Valid Range: Minimum value of 0.0.

Required: No
ModelSetupTime

The time it takes to launch new compute resources for a serverless endpoint. The time can vary
depending on the model size, how long it takes to download the model, and the start-up time of the
container.

NaN indicates that the value is not available.

Type: Integer

Valid Range: Minimum value of 0.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
RedshiftDatasetDefinition
Service: Amazon SageMaker Service
Configuration for Redshift Dataset Definition input.

Contents

ClusterId
The Redshift cluster Identifier.
Type: String
Pattern: .*
Required: Yes

ClusterRoleArn
The IAM role attached to your Redshift cluster that Amazon SageMaker uses to generate datasets.
Type: String
Pattern: ^arn:aws[a-z\-]*:iam::\d{12}:role/\?[a-zA-Z0-9\-_]+]+$
Required: Yes

Database
The name of the Redshift database used in Redshift query execution.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 64.
Pattern: .*
Required: Yes

DbUser
The database user name used in Redshift query execution.
Type: String
Pattern: .*
Required: Yes

OutputFormat
The data storage format for Redshift query results.
Type: String
Valid Values: PARQUET | CSV
Required: Yes
**OutputS3Uri**

The location in Amazon S3 where the Redshift query results are stored.

Type: String

- Length Constraints: Maximum length of 1024.
- Pattern: ^(https|s3)://([^/]+)?([^/]+)?(.*)$

Required: Yes

**QueryString**

The SQL query statements to be executed.

Type: String

- Pattern: \s\S+"}

Required: Yes

**KmsKeyId**

The AWS Key Management Service (AWS KMS) key that Amazon SageMaker uses to encrypt data from a Redshift execution.

Type: String

- Length Constraints: Maximum length of 2048.
- Pattern: *.*

Required: No

**OutputCompression**

The compression used for Redshift query results.

Type: String

- Valid Values: None | GZIP | BZIP2 | ZSTD | SNAPPY

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
RegisterModelStepMetadata
Service: Amazon SageMaker Service

Metadata for a register model job step.

Contents

Arn

The Amazon Resource Name (ARN) of the model package.

Type: String

Length Constraints: Maximum length of 256.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
RenderableTask
Service: Amazon SageMaker Service

Contains input values for a task.

Contents

Input

A JSON object that contains values for the variables defined in the template. It is made available to the template under the substitution variable `task.input`. For example, if you define a variable `task.input.text` in your template, you can supply the variable in the JSON object as "text": "sample text".

Type: String


Pattern: `[\S\s]+`

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
RenderingError
Service: Amazon SageMaker Service
A description of an error that occurred while rendering the template.

Contents

Code
A unique identifier for a specific class of errors.
Type: String
Required: Yes

Message
A human-readable message describing the error.
Type: String
Required: Yes

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
RepositoryAuthConfig
Service: Amazon SageMaker Service

Specifies an authentication configuration for the private docker registry where your model image is hosted. Specify a value for this property only if you specified Vpc as the value for the RepositoryAccessMode field of the ImageConfig object that you passed to a call to CreateModel and the private Docker registry where the model image is hosted requires authentication.

Contents

RepositoryCredentialsProviderArn

The Amazon Resource Name (ARN) of an AWS Lambda function that provides credentials to authenticate to the private Docker registry where your model image is hosted. For information about how to create an AWS Lambda function, see Create a Lambda function with the console in the AWS Lambda Developer Guide.

Type: String
Length Constraints: Minimum length of 1. Maximum length of 2048.
Pattern: .*
Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ResolvedAttributes
Service: Amazon SageMaker Service
The resolved attributes.

Contents

AutoMLJobObjective
Specifies a metric to minimize or maximize as the objective of an AutoML job.
Type: AutoMLJobObjective (p. 1284) object
Required: No

CompletionCriteria
How long a job is allowed to run, or how many candidates a job is allowed to generate.
Type: AutoMLJobCompletionCriteria (p. 1280) object
Required: No

ProblemType
The problem type.
Type: String
Valid Values: BinaryClassification | MulticlassClassification | Regression
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ResourceCatalog

Service: Amazon SageMaker Service

A resource catalog containing all of the resources of a specific resource type within a resource owner account. For an example on sharing the Amazon SageMaker Feature Store DefaultFeatureGroupCatalog, see Share Amazon SageMaker Catalog resource type in the Amazon SageMaker Developer Guide.

Contents

CreationTime

The time the ResourceCatalog was created.

Type: Timestamp

Required: Yes

Description

A free form description of the ResourceCatalog.

Type: String

Length Constraints: Maximum length of 256.

Required: Yes

ResourceCatalogArn

The Amazon Resource Name (ARN) of the ResourceCatalog.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-\:]*:sagemaker:[a-z0-9\-\:]*:[0-9]{12}:sagemaker-catalog/.*

Required: Yes

ResourceCatalogName

The name of the ResourceCatalog.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 64.

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ResourceConfig
Service: Amazon SageMaker Service

Describes the resources, including machine learning (ML) compute instances and ML storage volumes, to use for model training.

Contents

VolumeSizeInGB

The size of the ML storage volume that you want to provision.

ML storage volumes store model artifacts and incremental states. Training algorithms might also use the ML storage volume for scratch space. If you want to store the training data in the ML storage volume, choose File as the TrainingInputMode in the algorithm specification.

When using an ML instance with NVMe SSD volumes, SageMaker doesn't provision Amazon EBS General Purpose SSD (gp2) storage. Available storage is fixed to the NVMe-type instance's storage capacity. SageMaker configures storage paths for training datasets, checkpoints, model artifacts, and outputs to use the entire capacity of the instance storage. For example, ML instance families with the NVMe-type instance storage include ml.p4d, ml.g4dn, and ml.g5.

When using an ML instance with the EBS-only storage option and without instance storage, you must define the size of EBS volume through VolumeSizeInGB in the ResourceConfig API. For example, ML instance families that use EBS volumes include ml.c5 and ml.p2.

To look up instance types and their instance storage types and volumes, see Amazon EC2 Instance Types.

To find the default local paths defined by the SageMaker training platform, see Amazon SageMaker Training Storage Folders for Training Datasets, Checkpoints, Model Artifacts, and Outputs.

Type: Integer
Valid Range: Minimum value of 1.
Required: Yes

InstanceCount

The number of ML compute instances to use. For distributed training, provide a value greater than 1.

Type: Integer
Valid Range: Minimum value of 0.
Required: No

InstanceGroups

The configuration of a heterogeneous cluster in JSON format.

Type: Array of InstanceGroup (p. 1609) objects
Array Members: Maximum number of 5 items.
Required: No

InstanceType

The ML compute instance type.
Note
SageMaker Training on Amazon Elastic Compute Cloud (EC2) P4de instances is in preview release starting December 9th, 2022. Amazon EC2 P4de instances (currently in preview) are powered by 8 NVIDIA A100 GPUs with 80GB high-performance HBM2e GPU memory, which accelerate the speed of training ML models that need to be trained on large datasets of high-resolution data. In this preview release, Amazon SageMaker supports ML training jobs on P4de instances (ml.p4de.24xlarge) to reduce model training time. The ml.p4de.24xlarge instances are available in the following AWS Regions.

- US East (N. Virginia) (us-east-1)
- US West (Oregon) (us-west-2)

To request quota limit increase and start using P4de instances, contact the SageMaker Training service team through your account team.

Type: String

Valid Values:

- ml.m4.xlarge
- ml.m4.2xlarge
- ml.m4.4xlarge
- ml.m4.10xlarge
- ml.m4.16xlarge
- ml.g4dn.xlarge
- ml.g4dn.2xlarge
- ml.g4dn.4xlarge
- ml.g4dn.8xlarge
- ml.g4dn.12xlarge
- ml.g4dn.16xlarge
- ml.m5.large
- ml.m5.xlarge
- ml.m5.2xlarge
- ml.m5.4xlarge
- ml.m5.12xlarge
- ml.m5.24xlarge
- ml.c4.xlarge
- ml.c4.2xlarge
- ml.c4.4xlarge
- ml.c4.8xlarge
- ml.p2.xlarge
- ml.p2.8xlarge
- ml.p2.16xlarge
- ml.p3.2xlarge
- ml.p3.8xlarge
- ml.p3.16xlarge
- ml.p3dn.24xlarge
- ml.p4d.24xlarge
- ml.c5.xlarge
- ml.c5.2xlarge
- ml.c5.4xlarge
- ml.c5.9xlarge
- ml.c5.18xlarge
- ml.c5n.xlarge
- ml.c5n.2xlarge
- ml.c5n.4xlarge
- ml.c5n.9xlarge
- ml.c5n.18xlarge
- ml.g5.xlarge
- ml.g5.2xlarge
- ml.g5.4xlarge
- ml.g5.8xlarge
- ml.g5.16xlarge
- ml.g5.12xlarge
- ml.g5.24xlarge
- ml.g5.48xlarge
- ml.trn1.2xlarge
- ml.trn1.32xlarge
- ml.trn1n.32xlarge
- ml.p5.48xlarge

Required: No

KeepAlivePeriodInSeconds

The duration of time in seconds to retain configured resources in a warm pool for subsequent training jobs.

Type: Integer

Valid Range: Minimum value of 0. Maximum value of 3600.

Required: No

VolumeKmsKeyId

The AWS KMS key that SageMaker uses to encrypt data on the storage volume attached to the ML compute instance(s) that run the training job.

Note
Certain Nitro-based instances include local storage, dependent on the instance type. Local storage volumes are encrypted using a hardware module on the instance. You can't request a VolumeKmsKeyId when using an instance type with local storage. For a list of instance types that support local instance storage, see Instance Store Volumes. For more information about local instance storage encryption, see SSD Instance Store Volumes.

The VolumeKmsKeyId can be in any of the following formats:

- // KMS Key ID
  "1234abcd-12ab-34cd-56ef-1234567890ab"
ResourceConfig

- // Amazon Resource Name (ARN) of a KMS Key
  "arn:aws:kms:us-west-2:111122223333:key/1234abcd-12ab-34cd-56ef-1234567890ab"

  Type: String
  Length Constraints: Maximum length of 2048.
  Pattern: .*
  Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ResourceConfigForUpdate
Service: Amazon SageMaker Service

The ResourceConfig to update KeepAlivePeriodInSeconds. Other fields in the ResourceConfig cannot be updated.

Contents

KeepAlivePeriodInSeconds

The KeepAlivePeriodInSeconds value specified in the ResourceConfig to update.

Type: Integer

Valid Range: Minimum value of 0. Maximum value of 3600.

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ResourceLimits
Service: Amazon SageMaker Service

Specifies the maximum number of training jobs and parallel training jobs that a hyperparameter tuning job can launch.

Contents

MaxParallelTrainingJobs
The maximum number of concurrent training jobs that a hyperparameter tuning job can launch.
Type: Integer
Valid Range: Minimum value of 1.
Required: Yes

MaxNumberOfTrainingJobs
The maximum number of training jobs that a hyperparameter tuning job can launch.
Type: Integer
Valid Range: Minimum value of 1.
Required: No

MaxRuntimeInSeconds
The maximum time in seconds that a hyperparameter tuning job can run.
Type: Integer
Valid Range: Minimum value of 120. Maximum value of 15768000.
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ResourceSpec

Service: Amazon SageMaker Service

Specifies the ARN's of a SageMaker image and SageMaker image version, and the instance type that the version runs on.

Contents

InstanceType

The instance type that the image version runs on.

Note

JupyterServer apps only support the system value.

For KernelGateway apps, the system value is translated to ml.t3.medium. KernelGateway apps also support all other values for available instance types.

Type: String

Valid Values: system | ml.t3.micro | ml.t3.small | ml.t3.medium | ml.t3.large | ml.t3.xlarge | ml.t3.2xlarge | ml.m5.large | ml.m5.xlarge | ml.m5.2xlarge
| ml.m5.4xlarge | ml.m5.8xlarge | ml.m5.12xlarge | ml.m5.16xlarge | ml.m5.24xlarge | ml.m5d.large | ml.m5d.xlarge | ml.m5d.2xlarge | ml.m5d.4xlarge | ml.m5d.8xlarge | ml.m5d.12xlarge | ml.m5d.16xlarge | ml.m5d.24xlarge | ml.c5.large | ml.c5.xlarge | ml.c5.2xlarge | ml.c5.4xlarge | ml.c5.9xlarge | ml.c5.12xlarge | ml.c5.18xlarge | ml.c5.24xlarge | ml.p3.2xlarge | ml.p3.8xlarge | ml.p3.16xlarge | ml.p3dn.24xlarge | ml.g4dn.xlarge | ml.g4dn.2xlarge | ml.g4dn.4xlarge | ml.g4dn.8xlarge | ml.g4dn.12xlarge | ml.g4dn.16xlarge | ml.r5.large | ml.r5.xlarge | ml.r5.2xlarge | ml.r5.4xlarge | ml.r5.8xlarge | ml.r5.12xlarge | ml.g5.large | ml.g5.xlarge | ml.g5.2xlarge | ml.g5.4xlarge | ml.g5.8xlarge | ml.g5.16xlarge | ml.g5.12xlarge | ml.g5.24xlarge | ml.g5.48xlarge | ml.geospatial.interactive | ml.p4d.24xlarge | ml.p4de.24xlarge | ml.trn1.2xlarge | ml.trn1.32xlarge | ml.trn1.32xlarge | ml.trn1.64xlarge

Required: No

LifecycleConfigArn

The Amazon Resource Name (ARN) of the Lifecycle Configuration attached to the Resource.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:studio-lifecycle-config/.*

Required: No

SageMakerImageArn

The ARN of the SageMaker image that the image version belongs to.

Type: String

Length Constraints: Maximum length of 256.

Pattern: ^arn:aws(-[\w]+)*:sagemaker:.+:[0-9]{12}:image/[a-z0-9]([-.]?[a-z0-9])*$
Required: No

**SageMakerImageVersionAlias**

The SageMakerImageVersionAlias of the image to launch with. This value is in SemVer 2.0.0 versioning format.

Type: String


Pattern: (\d*)((0|\[1-9]\d*)\.(0|\[1-9]\d*)\.(0|\[1-9]\d*)(-((?:0|\[1-9]\d*|\d*[a-zA-Z-][0-9a-zA-Z-]*)(?:\.(?:0|\[1-9]\d*|\d*[a-zA-Z-][0-9a-zA-Z-]*))*)*)((?:0|\[1-9]\d*|\d*[a-zA-Z-][0-9a-zA-Z-]*))*)?

Required: No

**SageMakerImageVersionArn**

The ARN of the image version created on the instance.

Type: String

Length Constraints: Maximum length of 256.

Pattern: ^arn:aws(-\w+)*:sagemaker:.+:\[0-9\]{12}:image-version/[a-zA-Z0-9\[\-\]]+[a-zA-Z0-9\[\-\]]+/\[0-9\]+$

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
RetentionPolicy
Service: Amazon SageMaker Service

The retention policy for data stored on an Amazon Elastic File System (EFS) volume.

Contents

HomeEfsFileSystem

The default is Retain, which specifies to keep the data stored on the EFS volume.

Specify Delete to delete the data stored on the EFS volume.

Type: String

Valid Values: Retain | Delete

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
RetryStrategy
Service: Amazon SageMaker Service

The retry strategy to use when a training job fails due to an InternalServerError. RetryStrategy is specified as part of the CreateTrainingJob and CreateHyperParameterTuningJob requests. You can add the StoppingCondition parameter to the request to limit the training time for the complete job.

Contents

MaximumRetryAttempts

The number of times to retry the job. When the job is retried, it's SecondaryStatus is changed to STARTING.

Type: Integer


Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
RollingUpdatePolicy

Service: Amazon SageMaker Service

Specifies a rolling deployment strategy for updating a SageMaker endpoint.

Contents

MaximumBatchSize

Batch size for each rolling step to provision capacity and turn on traffic on the new endpoint fleet, and terminate capacity on the old endpoint fleet. Value must be between 5% to 50% of the variant's total instance count.

Type: CapacitySize (p. 1318) object

Required: Yes

WaitIntervalInSeconds

The length of the baking period, during which SageMaker monitors alarms for each batch on the new fleet.

Type: Integer

Valid Range: Minimum value of 0. Maximum value of 3600.

Required: Yes

MaximumExecutionTimeoutInSeconds

The time limit for the total deployment. Exceeding this limit causes a timeout.

Type: Integer


Required: No

RollbackMaximumBatchSize

Batch size for rollback to the old endpoint fleet. Each rolling step to provision capacity and turn on traffic on the old endpoint fleet, and terminate capacity on the new endpoint fleet. If this field is absent, the default value will be set to 100% of total capacity which means to bring up the whole capacity of the old fleet at once during rollback.

Type: CapacitySize (p. 1318) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
RSessionAppSettings
Service: Amazon SageMaker Service
A collection of settings that apply to an RSessionGateway app.

Contents

CustomImages
A list of custom SageMaker images that are configured to run as a RSession app.
Type: Array of CustomImage (p. 1378) objects
Array Members: Maximum number of 200 items.
Required: No

DefaultResourceSpec
Specifies the ARN's of a SageMaker image and SageMaker image version, and the instance type that the version runs on.
Type: ResourceSpec (p. 1915) object
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
RStudioServerProAppSettings

Service: Amazon SageMaker Service

A collection of settings that configure user interaction with the RStudioServerPro app.

Contents

AccessStatus

Indicates whether the current user has access to the RStudioServerPro app.

Type: String

Valid Values: ENABLED | DISABLED

Required: No

UserGroup

The level of permissions that the user has within the RStudioServerPro app. This value defaults to `User`. The `Admin` value allows the user access to the RStudio Administrative Dashboard.

Type: String

Valid Values: R_STUDIO_ADMIN | R_STUDIO_USER

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
RStudioServerProDomainSettings
Service: Amazon SageMaker Service

A collection of settings that configure the RStudioServerPro Domain-level app.

Contents

DomainExecutionRoleArn
The ARN of the execution role for the RStudioServerPro Domain-level app.
Type: String
Pattern: ^arn:aws[a-z\-]*:iam::\d{12}:role/?[a-zA-Z_0-9+=,.@\-_]+$
Required: Yes

DefaultResourceSpec
Specifies the ARN's of a SageMaker image and SageMaker image version, and the instance type that the version runs on.
Type: ResourceSpec (p. 1915) object
Required: No

RStudioConnectUrl
A URL pointing to an RStudio Connect server.
Type: String
Required: No

RStudioPackageManagerUrl
A URL pointing to an RStudio Package Manager server.
Type: String
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
RStudioServerProDomainSettingsForUpdate
Service: Amazon SageMaker Service

A collection of settings that update the current configuration for the RStudioServerPro Domain-level app.

Contents

DomainExecutionRoleArn

The execution role for the RStudioServerPro Domain-level app.

Type: String


Pattern: ^arn:aws[a-z\-]*:iam::\d{12}:role/\?a-zA-Z0-9+=,.@\-_/]+$

Required: Yes

DefaultResourceSpec

Specifies the ARN's of a SageMaker image and SageMaker image version, and the instance type that the version runs on.

Type: ResourceSpec (p. 1915) object

Required: No

RStudioConnectUrl

A URL pointing to an RStudio Connect server.

Type: String

Required: No

RStudioPackageManagerUrl

A URL pointing to an RStudio Package Manager server.

Type: String

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
S3DataSource
Service: Amazon SageMaker Service

Describes the S3 data source.

Your input bucket must be in the same AWS region as your training job.

Contents

S3DataType

- If you choose S3Prefix, S3Uri identifies a key name prefix. SageMaker uses all objects that match the specified key name prefix for model training.

- If you choose ManifestFile, S3Uri identifies an object that is a manifest file containing a list of object keys that you want SageMaker to use for model training.

- If you choose AugmentedManifestFile, S3Uri identifies an object that is an augmented manifest file in JSON lines format. This file contains the data you want to use for model training. AugmentedManifestFile can only be used if the Channel's input mode is Pipe.

Type: String
Valid Values: ManifestFile | S3Prefix | AugmentedManifestFile
Required: Yes

S3Uri

Depending on the value specified for the S3DataType, identifies either a key name prefix or a manifest. For example:

- A key name prefix might look like this: s3://bucketname/exampleprefix
- A manifest might look like this: s3://bucketname/example.manifest

A manifest is an S3 object which is a JSON file consisting of an array of elements. The first element is a prefix which is followed by one or more suffixes. SageMaker appends the suffix elements to the prefix to get a full set of S3Uri. Note that the prefix must be a valid non-empty S3Uri that precludes users from specifying a manifest whose individual S3Uri is sourced from different S3 buckets.

The following code example shows a valid manifest format:

```
[ {"prefix": "s3://customer_bucket/some/prefix/"},
  "relative/path/to/custdata-1",
  "relative/path/to/custdata-2",
  ...
  "relative/path/to/custdata-N"
]
```

This JSON is equivalent to the following S3Uri list:

- s3://customer_bucket/some/prefix/relative/path/to/custdata-1
- s3://customer_bucket/some/prefix/relative/path/to/custdata-2
... 

s3://customer_bucket/some/prefix/relative/path/custdata-N

The complete set of S3Uri in this manifest is the input data for the channel for this data source. The object that each S3Uri points to must be readable by the IAM role that SageMaker uses to perform tasks on your behalf.

Your input bucket must be located in same AWS region as your training job.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: ^(https|s3)://([^/]+)/(.*)$

Required: Yes

**AttributeNames**

A list of one or more attribute names to use that are found in a specified augmented manifest file.

Type: Array of strings

Array Members: Maximum number of 16 items.

Length Constraints: Minimum length of 1. Maximum length of 256.

Pattern: .+

Required: No

**InstanceGroupNames**

A list of names of instance groups that get data from the S3 data source.

Type: Array of strings

Array Members: Maximum number of 5 items.

Length Constraints: Minimum length of 1. Maximum length of 64.

Pattern: .+

Required: No

**S3DataDistributionType**

If you want SageMaker to replicate the entire dataset on each ML compute instance that is launched for model training, specify FullyReplicated.

If you want SageMaker to replicate a subset of data on each ML compute instance that is launched for model training, specify ShardedByS3Key. If there are n ML compute instances launched for a training job, each instance gets approximately 1/n of the number of S3 objects. In this case, model training on each machine uses only the subset of training data.

Don't choose more ML compute instances for training than available S3 objects. If you do, some nodes won't get any data and you will pay for nodes that aren't getting any training data. This applies in both File and Pipe modes. Keep this in mind when developing algorithms.

In distributed training, where you use multiple ML compute EC2 instances, you might choose ShardedByS3Key. If the algorithm requires copying training data to the ML storage volume (when TrainingInputMode is set to File), this copies 1/n of the number of objects.
Type: String

Valid Values: FullyReplicated | ShardedByS3Key

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
S3ModelDataSource
Service: Amazon SageMaker Service

Specifies the S3 location of ML model data to deploy.

Contents

CompressionType

Specifies how the ML model data is prepared.

If you choose Gzip and choose S3Object as the value of S3DataType, S3Uri identifies an object that is a gzip-compressed TAR archive. SageMaker will attempt to decompress and untar the object during model deployment.

If you choose None and choose S3Object as the value of S3DataType, S3Uri identifies an object that represents an uncompressed ML model to deploy.

If you choose None and choose S3Prefix as the value of S3DataType, S3Uri identifies a key name prefix, under which all objects represents the uncompressed ML model to deploy.

If you choose None, then SageMaker will follow rules below when creating model data files under /opt/ml/model directory for use by your inference code:
• If you choose S3Object as the value of S3DataType, then SageMaker will split the key of the S3 object referenced by S3Uri by slash (/), and use the last part as the filename of the file holding the content of the S3 object.
• If you choose S3Prefix as the value of S3DataType, then for each S3 object under the key name prefix referenced by S3Uri, SageMaker will trim its key by the prefix, and use the remainder as the path (relative to /opt/ml/model) of the file holding the content of the S3 object. SageMaker will split the remainder by slash (/), using intermediate parts as directory names and the last part as filename of the file holding the content of the S3 object.
• Do not use any of the following as file names or directory names:
  • An empty or blank string
  • A string which contains null bytes
  • A string longer than 255 bytes
  • A single dot (.)
  • A double dot (..)
• Ambiguous file names will result in model deployment failure. For example, if your uncompressed ML model consists of two S3 objects s3://mybucket/model/weights and s3://mybucket/model/weights/part1 and you specify s3://mybucket/model/ as the value of S3Uri and S3Prefix as the value of S3DataType, then it will result in name clash between /opt/ml/model/weights (a regular file) and /opt/ml/model/weights/ (a directory).
• Do not organize the model artifacts in S3 console using folders. When you create a folder in S3 console, S3 creates a 0-byte object with a key set to the folder name you provide. They key of the 0-byte object ends with a slash (/) which violates SageMaker restrictions on model artifact file names, leading to model deployment failure.

Type: String

Valid Values: None | Gzip

Required: Yes

S3DataType

Specifies the type of ML model data to deploy.
If you choose S3Prefix, S3Uri identifies a key name prefix. SageMaker uses all objects that match the specified key name prefix as part of the ML model data to deploy. A valid key name prefix identified by S3Uri always ends with a forward slash (/).

If you choose S3Object, S3Uri identifies an object that is the ML model data to deploy.

Type: String

Valid Values: S3Prefix | S3Object

Required: Yes

S3Uri

Specifies the S3 path of ML model data to deploy.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: ^(https|s3)://([^/]+)/?(.*)$

Required: Yes

ModelAccessConfig

Specifies the access configuration file for the ML model. You can explicitly accept the model end-user license agreement (EULA) within the ModelAccessConfig. You are responsible for reviewing and complying with any applicable license terms and making sure they are acceptable for your use case before downloading or using a model.

Type: \[ModelAccessConfig (p. 1658)\] object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
S3StorageConfig
Service: Amazon SageMaker Service

The Amazon Simple Storage (Amazon S3) location and security configuration for OfflineStore.

Contents

S3Uri
The S3 URI, or location in Amazon S3, of OfflineStore.
S3 URIs have a format similar to the following: s3://example-bucket/prefix/
Type: String
Length Constraints: Maximum length of 1024.
Pattern: ^((https|s3)://([\^/]+)/?)+$  
Required: Yes

KmsKeyId
The AWS Key Management Service (KMS) key ARN of the key used to encrypt any objects written into the OfflineStore S3 location.
The IAM roleARN that is passed as a parameter to CreateFeatureGroup must have below permissions to the KmsKeyId:
• "kms:GenerateDataKey"
Type: String
Length Constraints: Maximum length of 2048.
Pattern: .*
Required: No

ResolvedOutputS3Uri
The S3 path where offline records are written.
Type: String
Length Constraints: Maximum length of 1024.
Pattern: ^((https|s3)://([\^/]+)/?)+$  
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ScalingPolicy
Service: Amazon SageMaker Service

An object containing a recommended scaling policy.

Contents

Important
This data type is a UNION, so only one of the following members can be specified when used or returned.

TargetTracking
A target tracking scaling policy. Includes support for predefined or customized metrics.

Type: TargetTrackingScalingPolicyConfiguration (p. 1982) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ScalingPolicyMetric
Service: Amazon SageMaker Service

The metric for a scaling policy.

Contents

InvocationsPerInstance

The number of invocations sent to a model, normalized by InstanceCount in each ProductionVariant. \( \frac{1}{\text{numberOfInstances}} \) is sent as the value on each request, where numberOfInstances is the number of active instances for the ProductionVariant behind the endpoint at the time of the request.

Type: Integer
Required: No

ModelLatency

The interval of time taken by a model to respond as viewed from SageMaker. This interval includes the local communication times taken to send the request and to fetch the response from the container of a model and the time taken to complete the inference in the container.

Type: Integer
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ScalingPolicyObjective
Service: Amazon SageMaker Service

An object where you specify the anticipated traffic pattern for an endpoint.

Contents

MaxInvocationsPerMinute

The maximum number of expected requests to your endpoint per minute.

Type: Integer
Required: No

MinInvocationsPerMinute

The minimum number of expected requests to your endpoint per minute.

Type: Integer
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ScheduleConfig
Service: Amazon SageMaker Service

Configuration details about the monitoring schedule.

Contents

ScheduleExpression

A cron expression that describes details about the monitoring schedule.

The supported cron expressions are:

- If you want to set the job to start every hour, use the following:
  
  Hourly: \texttt{cron(0 * ? * * *)}

- If you want to start the job daily:
  
  \texttt{cron(0 [00-23] ? * * *)}

- If you want to run the job one time, immediately, use the following keyword:
  
  \texttt{NOW}

For example, the following are valid cron expressions:

- Daily at noon UTC: \texttt{cron(0 12 ? * * *)}
- Daily at midnight UTC: \texttt{cron(0 0 ? * * *)}

To support running every 6, 12 hours, the following are also supported:

\texttt{cron(0 [00-23]/[01-24] ? * * *)}

For example, the following are valid cron expressions:

- Every 12 hours, starting at 5pm UTC: \texttt{cron(0 17/12 ? * * *)}
- Every two hours starting at midnight: \texttt{cron(0 0/2 ? * * *)}

Note

- Even though the cron expression is set to start at 5PM UTC, note that there could be a delay of 0-20 minutes from the actual requested time to run the execution.

- We recommend that if you would like a daily schedule, you do not provide this parameter. Amazon SageMaker will pick a time for running every day.

You can also specify the keyword \texttt{NOW} to run the monitoring job immediately, one time, without recurring.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 256.

Required: Yes

DataAnalysisEndTime

Sets the end time for a monitoring job window. Express this time as an offset to the times that you schedule your monitoring jobs to run. You schedule monitoring jobs with the ScheduleExpression parameter. Specify this offset in ISO 8601 duration format. For example, if you want to end the window one hour before the start of each monitoring job, you would specify: "-PT1H".
The end time that you specify must not follow the start time that you specify by more than 24 hours. You specify the start time with the DataAnalysisStartTime parameter.

If you set ScheduleExpression to NOW, this parameter is required.

Type: String
Required: No

**DataAnalysisStartTime**

Sets the start time for a monitoring job window. Express this time as an offset to the times that you schedule your monitoring jobs to run. You schedule monitoring jobs with the ScheduleExpression parameter. Specify this offset in ISO 8601 duration format. For example, if you want to monitor the five hours of data in your dataset that precede the start of each monitoring job, you would specify: "-PT5H".

The start time that you specify must not precede the end time that you specify by more than 24 hours. You specify the end time with the DataAnalysisEndTime parameter.

If you set ScheduleExpression to NOW, this parameter is required.

Type: String
Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
SearchExpression

Service: Amazon SageMaker Service

A multi-expression that searches for the specified resource or resources in a search. All resource objects that satisfy the expression's condition are included in the search results. You must specify at least one subexpression, filter, or nested filter. A SearchExpression can contain up to twenty elements.

A SearchExpression contains the following components:

- A list of Filter objects. Each filter defines a simple Boolean expression comprised of a resource property name, Boolean operator, and value.
- A list of NestedFilter objects. Each nested filter defines a list of Boolean expressions using a list of resource properties. A nested filter is satisfied if a single object in the list satisfies all Boolean expressions.
- A list of SearchExpression objects. A search expression object can be nested in a list of search expression objects.
- A Boolean operator: And or Or.

Contents

Filters

A list of filter objects.

Type: Array of Filter (p. 1496) objects

Array Members: Minimum number of 1 item. Maximum number of 20 items.

Required: No

NestedFilters

A list of nested filter objects.

Type: Array of NestedFilters (p. 1768) objects

Array Members: Minimum number of 1 item. Maximum number of 20 items.

Required: No

Operator

A Boolean operator used to evaluate the search expression. If you want every conditional statement in all lists to be satisfied for the entire search expression to be true, specify And. If only a single conditional statement needs to be true for the entire search expression to be true, specify Or. The default value is And.

Type: String

Valid Values: And | Or

Required: No

SubExpressions

A list of search expression objects.

Type: Array of SearchExpression (p. 1936) objects

Array Members: Minimum number of 1 item. Maximum number of 20 items.
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
**SearchRecord**

**Service:** Amazon SageMaker Service

A single resource returned as part of the [Search](#) API response.

**Contents**

**Endpoint**

A hosted endpoint for real-time inference.

Type: [Endpoint](#) object

Required: No

**Experiment**

The properties of an experiment.

Type: [Experiment](#) object

Required: No

**FeatureGroup**

Amazon SageMaker Feature Store stores features in a collection called Feature Group. A Feature Group can be visualized as a table which has rows, with a unique identifier for each row where each column in the table is a feature. In principle, a Feature Group is composed of features and values per features.

Type: [FeatureGroup](#) object

Required: No

**FeatureMetadata**

The feature metadata used to search through the features.

Type: [FeatureMetadata](#) object

Required: No

**HyperParameterTuningJob**

The properties of a hyperparameter tuning job.

Type: [HyperParameterTuningJob](#) object

Required: No

**Model**

A model displayed in the Amazon SageMaker Model Dashboard.

Type: [ModelDashboardModel](#) object

Required: No

**ModelCard**

An Amazon SageMaker Model Card that documents details about a machine learning model.

Type: [ModelCard](#) object

Required: No
ModelPackage

A versioned model that can be deployed for SageMaker inference.

Type: ModelPackage (p. 1702) object

Required: No

ModelPackageGroup

A group of versioned models in the model registry.

Type: ModelPackageGroup (p. 1710) object

Required: No

Pipeline

A SageMaker Model Building Pipeline instance.

Type: Pipeline (p. 1811) object

Required: No

PipelineExecution

An execution of a pipeline.

Type: PipelineExecution (p. 1815) object

Required: No

Project

The properties of a project.

Type: Project (p. 1868) object

Required: No

TrainingJob

The properties of a training job.

Type: TrainingJob (p. 2001) object

Required: No

Trial

The properties of a trial.

Type: Trial (p. 2038) object

Required: No

TrialComponent

The properties of a trial component.

Type: TrialComponent (p. 2041) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:
• AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for Ruby V3
SecondaryStatusTransition

Service: Amazon SageMaker Service

An array element of SecondaryStatusTransitions for DescribeTrainingJob. It provides additional details about a status that the training job has transitioned through. A training job can be in one of several states, for example, starting, downloading, training, or uploading. Within each state, there are a number of intermediate states. For example, within the starting state, SageMaker could be starting the training job or launching the ML instances. These transitional states are referred to as the job's secondary status.

Contents

StartTime

A timestamp that shows when the training job transitioned to the current secondary status state.

Type: Timestamp

Required: Yes

Status

Contains a secondary status information from a training job.

Status might be one of the following secondary statuses:

InProgress

• Starting - Starting the training job.
• Downloading - An optional stage for algorithms that support File training input mode. It indicates that data is being downloaded to the ML storage volumes.
• Training - Training is in progress.
• Uploading - Training is complete and the model artifacts are being uploaded to the S3 location.

Completed

• Completed - The training job has completed.

Failed

• Failed - The training job has failed. The reason for the failure is returned in the FailureReason field of DescribeTrainingJobResponse.

Stopped

• MaxRuntimeExceeded - The job stopped because it exceeded the maximum allowed runtime.
• Stopped - The training job has stopped.

Stopping

• Stopping - Stopping the training job.

We no longer support the following secondary statuses:

• LaunchingMLInstances
• PreparingTrainingStack
• DownloadingTrainingImage

Type: String

Valid Values: Starting | LaunchingMLInstances | PreparingTrainingStack | Downloading | DownloadingTrainingImage | Training | Uploading | Stopping
SecondaryStatusTransition

| Stopped | MaxRuntimeExceeded | Completed | Failed | Interrupted | MaxWaitTimeExceeded | Updating | Restarting

Required: Yes

**EndTime**

A timestamp that shows when the training job transitioned out of this secondary status state into another secondary status state or when the training job has ended.

Type: Timestamp

Required: No

**StatusMessage**

A detailed description of the progress within a secondary status.

SageMaker provides secondary statuses and status messages that apply to each of them:

Starting
- Starting the training job.
- Launching requested ML instances.
- Insufficient capacity error from EC2 while launching instances, retrying!
- Launched instance was unhealthy, replacing it!
- Preparing the instances for training.

Training
- Downloading the training image.
- Training image download completed. Training in progress.

**Important**

Status messages are subject to change. Therefore, we recommend not including them in code that programatically initiates actions. For examples, don't use status messages in if statements.

To have an overview of your training job's progress, view TrainingJobStatus and SecondaryStatus in DescribeTrainingJob, and StatusMessage together. For example, at the start of a training job, you might see the following:

- TrainingJobStatus - InProgress
- SecondaryStatus - Training
- StatusMessage - Downloading the training image

Type: String

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
SelectedStep
Service: Amazon SageMaker Service
A step selected to run in selective execution mode.

Contents

StepName
The name of the pipeline step.
Type: String
Length Constraints: Maximum length of 256.
Required: Yes

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
SelectiveExecutionConfig
Service: Amazon SageMaker Service

The selective execution configuration applied to the pipeline run.

Contents

SelectedSteps

A list of pipeline steps to run. All step(s) in all path(s) between two selected steps should be included.

Type: Array of `SelectedStep (p. 1943)` objects
Array Members: Minimum number of 1 item. Maximum number of 50 items.
Required: Yes

SourcePipelineExecutionArn

The ARN from a reference execution of the current pipeline. Used to copy input collaterals needed for the selected steps to run. The execution status of the pipeline can be either Failed or Success.

This field is required if the steps you specify for SelectedSteps depend on output collaterals from any non-specified pipeline steps. For more information, see Selective Execution for Pipeline Steps.

Type: String
Length Constraints: Maximum length of 256.
Pattern: ^arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:pipeline\./.*\./\.*$ Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
SelectiveExecutionResult
Service: Amazon SageMaker Service

The ARN from an execution of the current pipeline.

Contents

SourcePipelineExecutionArn

  The ARN from an execution of the current pipeline.
  
  Type: String
  
  Length Constraints: Maximum length of 256.

  Pattern: ^arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:pipeline\/.\/*\execute\/.\/*$

  Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ServiceCatalogProvisionedProductDetails

Service: Amazon SageMaker Service

Details of a provisioned service catalog product. For information about service catalog, see What is AWS Service Catalog.

Contents

ProvisionedProductId

The ID of the provisioned product.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 100.

Pattern: ^[a-zA-Z0-9_\-]*

Required: No

ProvisionedProductStatusMessage

The current status of the product.

- AVAILABLE - Stable state, ready to perform any operation. The most recent operation succeeded and completed.
- UNDER_CHANGE - Transitive state. Operations performed might not have valid results. Wait for an AVAILABLE status before performing operations.
- TAINTED - Stable state, ready to perform any operation. The stack has completed the requested operation but is not exactly what was requested. For example, a request to update to a new version failed and the stack rolled back to the current version.
- ERROR - An unexpected error occurred. The provisioned product exists but the stack is not running. For example, CloudFormation received a parameter value that was not valid and could not launch the stack.
- PLAN_IN_PROGRESS - Transitive state. The plan operations were performed to provision a new product, but resources have not yet been created. After reviewing the list of resources to be created, execute the plan. Wait for an AVAILABLE status before performing operations.

Type: String

Pattern: .*

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ServiceCatalogProvisioningDetails
Service: Amazon SageMaker Service

Details that you specify to provision a service catalog product. For information about service catalog, see What is AWS Service Catalog.

Contents

ProductId
The ID of the product to provision.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 100.
Pattern: ^[a-zA-Z0-9_\-]*
Required: Yes

PathId
The path identifier of the product. This value is optional if the product has a default path, and required if the product has more than one path.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 100.
Pattern: ^[a-zA-Z0-9_\-]*
Required: No

ProvisioningArtifactId
The ID of the provisioning artifact.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 100.
Pattern: ^[a-zA-Z0-9_\-]*
Required: No

ProvisioningParameters
A list of key value pairs that you specify when you provision a product.
Type: Array of ProvisioningParameter (p. 1875) objects
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
• AWS SDK for Ruby V3
ServiceCatalogProvisioningUpdateDetails
Service: Amazon SageMaker Service

Details that you specify to provision a service catalog product. For information about service catalog, see What is AWS Service Catalog.

Contents

ProvisioningArtifactId

The ID of the provisioning artifact.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 100.
Pattern: ^[a-zA-Z0-9-\_\-]*
Required: No

ProvisioningParameters

A list of key value pairs that you specify when you provision a product.
Type: Array of ProvisioningParameter (p. 1875) objects
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

• AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for Ruby V3
ShadowModeConfig
Service: Amazon SageMaker Service

The configuration of ShadowMode inference experiment type, which specifies a production variant to take all the inference requests, and a shadow variant to which Amazon SageMaker replicates a percentage of the inference requests. For the shadow variant it also specifies the percentage of requests that Amazon SageMaker replicates.

Contents

ShadowModelVariants
List of shadow variant configurations.
Type: Array of ShadowModelVariantConfig objects
Array Members: Fixed number of 1 item.
Required: Yes

SourceModelVariantName
The name of the production variant, which takes all the inference requests.
Type: String
Length Constraints: Maximum length of 63.
Pattern: ^[a-zA-Z0-9](\-[a-zA-Z0-9]*)[a-zA-Z0-9]?
Required: Yes

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ShadowModelVariantConfig
Service: Amazon SageMaker Service

The name and sampling percentage of a shadow variant.

Contents

SamplingPercentage

The percentage of inference requests that Amazon SageMaker replicates from the production variant to the shadow variant.

Type: Integer

Valid Range: Maximum value of 100.

Required: Yes

ShadowModelVariantName

The name of the shadow variant.

Type: String

Length Constraints: Maximum length of 63.

Pattern: ^[a-zA-Z0-9](\-[a-zA-Z0-9]*[a-zA-Z0-9])?

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
SharingSettings
Service: Amazon SageMaker Service

Specifies options for sharing Amazon SageMaker Studio notebooks. These settings are specified as part of DefaultUserSettings when the CreateDomain API is called, and as part of UserSettings when the CreateUserProfile API is called. When SharingSettings is not specified, notebook sharing isn't allowed.

Contents

NotebookOutputOption

Whether to include the notebook cell output when sharing the notebook. The default is Disabled.

Type: String

Valid Values: Allowed | Disabled

Required: No

S3KmsKeyId

When NotebookOutputOption is Allowed, the AWS Key Management Service (KMS) encryption key ID used to encrypt the notebook cell output in the Amazon S3 bucket.

Type: String

Length Constraints: Maximum length of 2048.

Pattern: .*

Required: No

S3OutputPath

When NotebookOutputOption is Allowed, the Amazon S3 bucket used to store the shared notebook snapshots.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: ^(https|s3)//([^/]+)/(.*)$

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ShuffleConfig

Service: Amazon SageMaker Service

A configuration for a shuffle option for input data in a channel. If you use S3Prefix for S3DataType, the results of the S3 key prefix matches are shuffled. If you use ManifestFile, the order of the S3 object references in the ManifestFile is shuffled. If you use AugmentedManifestFile, the order of the JSON lines in the AugmentedManifestFile is shuffled. The shuffling order is determined using the Seed value.

For Pipe input mode, when ShuffleConfig is specified shuffling is done at the start of every epoch. With large datasets, this ensures that the order of the training data is different for each epoch, and it helps reduce bias and possible overfitting. In a multi-node training job when ShuffleConfig is combined with S3DataDistributionType of ShardedByS3Key, the data is shuffled across nodes so that the content sent to a particular node on the first epoch might be sent to a different node on the second epoch.

Contents

Seed

Determines the shuffling order in ShuffleConfig value.

Type: Long

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
SourceAlgorithm

Service: Amazon SageMaker Service

Specifies an algorithm that was used to create the model package. The algorithm must be either an algorithm resource in your SageMaker account or an algorithm in AWS Marketplace that you are subscribed to.

Contents

AlgorithmName

The name of an algorithm that was used to create the model package. The algorithm must be either an algorithm resource in your SageMaker account or an algorithm in AWS Marketplace that you are subscribed to.

Type: String


Pattern: (arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:[a-z\-]*\//)?([a-zA-Z0-9](\[a-zA-Z0-9\-\]){0,62})(?!-)$

Required: Yes

ModelDataUrl

The Amazon S3 path where the model artifacts, which result from model training, are stored. This path must point to a single gzip compressed tar archive (.tar.gz suffix).

Note

The model artifacts must be in an S3 bucket that is in the same AWS region as the algorithm.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: ^(https|s3)://([^/\]+)/(.*)$

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
**SourceAlgorithmSpecification**

Service: Amazon SageMaker Service

A list of algorithms that were used to create a model package.

**Contents**

**SourceAlgorithms**

A list of the algorithms that were used to create a model package.

Type: Array of `SourceAlgorithm` objects

Array Members: Fixed number of 1 item.

Required: Yes

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
**SourceIpConfig**

Service: Amazon SageMaker Service

A list of IP address ranges (CIDRs). Used to create an allow list of IP addresses for a private workforce. Workers will only be able to login to their worker portal from an IP address within this range. By default, a workforce isn’t restricted to specific IP addresses.

**Contents**

**Cidrs**

A list of one to ten *Classless Inter-Domain Routing* (CIDR) values.

Maximum: Ten CIDR values

**Note**

The following Length Constraints apply to individual CIDR values in the CIDR value list.

Type: Array of strings

Length Constraints: Minimum length of 4. Maximum length of 64.

Pattern: `(\^((0-9)|[1-9][0-9]*|[0-9][1-9]([0-9]|0-9)|2[0-4][0-9]|25[0-5])\d(0-9)|[1-9][0-9]*|[0-9][1-9]([0-9]|0-9)|2[0-4][0-9]|25[0-5])(\%/((30-2)|12))(0-9)|[0-9]*\d(0-9)|[1-9]0-9|2[0-4]|25[0-5])$)

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
SpaceCodeEditorAppSettings
Service: Amazon SageMaker Service

The application settings for a Code Editor space.

Contents

DefaultResourceSpec
Specifies the ARN's of a SageMaker image and SageMaker image version, and the instance type that the version runs on.

Type: ResourceSpec (p. 1915) object

Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
SpaceDetails
Service: Amazon SageMaker Service
The space's details.

Contents

CreationTime
The creation time.
Type: Timestamp
Required: No

DomainId
The ID of the associated Domain.
Type: String
Length Constraints: Maximum length of 63.
Required: No

LastModifiedTime
The last modified time.
Type: Timestamp
Required: No

OwnershipSettingsSummary
Specifies summary information about the ownership settings.
Type: OwnershipSettingsSummary (p. 1796) object
Required: No

SpaceDisplayName
The name of the space that appears in the Studio UI.
Type: String
Length Constraints: Maximum length of 64.
Pattern: ^(?!\s*$).+
Required: No

SpaceName
The name of the space.
Type: String
Length Constraints: Maximum length of 63.
Pattern: ^[a-zA-Z0-9]( -*[a-zA-Z0-9])\{0,62}$
Required: No
SpaceSettingsSummary

Specifies summary information about the space settings.

Type: SpaceSettingsSummary (p. 1963) object

Required: No

SpaceSharingSettingsSummary

Specifies summary information about the space sharing settings.

Type: SpaceSharingSettingsSummary (p. 1965) object

Required: No

Status

The status.

Type: String

Valid Values: Deleting | Failed | InService | Pending | Updating | Update_Failed | Delete_Failed

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
SpaceJupyterLabAppSettings
Service: Amazon SageMaker Service

The settings for the JupyterLab application within a space.

Contents

CodeRepositories
A list of Git repositories that SageMaker automatically displays to users for cloning in the JupyterLab application.

Type: Array of CodeRepository (p. 1354) objects

Array Members: Maximum number of 10 items.

Required: No

DefaultResourceSpec
Specifies the ARN's of a SageMaker image and SageMaker image version, and the instance type that the version runs on.

Type: ResourceSpec (p. 1915) object

Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
SpaceSettings
Service: Amazon SageMaker Service
A collection of space settings.

Contents

AppType
The type of app created within the space.
Type: String
Valid Values: JupyterServer | KernelGateway | TensorBoard | RStudioServerPro | RSessionGateway | JupyterLab | CodeEditor
Required: No

CodeEditorAppSettings
The Code Editor application settings.
Type: SpaceCodeEditorAppSettings (p. 1957) object
Required: No

CustomFileSystems
A file system, created by you, that you assign to a space for an Amazon SageMaker Domain. Permitted users can access this file system in Amazon SageMaker Studio.
Type: Array of CustomFileSystem (p. 1376) objects
Array Members: Maximum number of 1 item.
Required: No

JupyterLabAppSettings
The settings for the JupyterLab application.
Type: SpaceJupyterLabAppSettings (p. 1960) object
Required: No

JupyterServerAppSettings
The JupyterServer app settings.
Type: JupyterServerAppSettings (p. 1618) object
Required: No

KernelGatewayAppSettings
The KernelGateway app settings.
Type: KernelGatewayAppSettings (p. 1620) object
Required: No

SpaceStorageSettings
The storage settings for a private space.
Type: `SpaceStorageSettings (p. 1966)` object

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
SpaceSettingsSummary
Service: Amazon SageMaker Service

Specifies summary information about the space settings.

Contents

AppType

The type of app created within the space.

Type: String

Valid Values: JupyterServer | KernelGateway | TensorBoard | RStudioServerPro | RSessionGateway | JupyterLab | CodeEditor

Required: No

SpaceStorageSettings

The storage settings for a private space.

Type: SpaceStorageSettings (p. 1966) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
SpaceSharingSettings
Service: Amazon SageMaker Service

A collection of space sharing settings.

Contents

SharingType

Specifies the sharing type of the space.

Type: String

Valid Values: Private | Shared

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
SpaceSharingSettingsSummary
Service: Amazon SageMaker Service

Specifies summary information about the space sharing settings.

Contents

SharingType

Specifies the sharing type of the space.

Type: String

Valid Values: Private | Shared

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
SpaceStorageSettings
Service: Amazon SageMaker Service
The storage settings for a private space.

Contents

EbsStorageSettings
A collection of EBS storage settings for a private space.
Type: EbsStorageSettings (p. 1432) object
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
**Stairs**
Service: Amazon SageMaker Service

Defines the stairs traffic pattern for an Inference Recommender load test. This pattern type consists of multiple steps where the number of users increases at each step.

Specify either the stairs or phases traffic pattern.

**Contents**

**DurationInSeconds**

Defines how long each traffic step should be.

Type: Integer

Valid Range: Minimum value of 1.

Required: No

**NumberOfSteps**

Specifies how many steps to perform during traffic.

Type: Integer

Valid Range: Minimum value of 1.

Required: No

**UsersPerStep**

Specifies how many new users to spawn in each step.

Type: Integer


Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
StoppingCondition

Service: Amazon SageMaker Service

Specifies a limit to how long a model training job or model compilation job can run. It also specifies how long a managed spot training job has to complete. When the job reaches the time limit, SageMaker ends the training or compilation job. Use this API to cap model training costs.

To stop a training job, SageMaker sends the algorithm the SIGTERM signal, which delays job termination for 120 seconds. Algorithms can use this 120-second window to save the model artifacts, so the results of training are not lost.

The training algorithms provided by SageMaker automatically save the intermediate results of a model training job when possible. This attempt to save artifacts is only a best effort case as model might not be in a state from which it can be saved. For example, if training has just started, the model might not be ready to save. When saved, this intermediate data is a valid model artifact. You can use it to create a model with CreateModel.

Note
The Neural Topic Model (NTM) currently does not support saving intermediate model artifacts. When training NTMs, make sure that the maximum runtime is sufficient for the training job to complete.

Contents

MaxPendingTimeInSeconds

The maximum length of time, in seconds, that a training or compilation job can be pending before it is stopped.

Type: Integer


Required: No

MaxRuntimeInSeconds

The maximum length of time, in seconds, that a training or compilation job can run before it is stopped.

For compilation jobs, if the job does not complete during this time, a TimeOut error is generated. We recommend starting with 900 seconds and increasing as necessary based on your model.

For all other jobs, if the job does not complete during this time, SageMaker ends the job. When RetryStrategy is specified in the job request, MaxRuntimeInSeconds specifies the maximum time for all of the attempts in total, not each individual attempt. The default value is 1 day. The maximum value is 28 days.

The maximum time that a TrainingJob can run in total, including any time spent publishing metrics or archiving and uploading models after it has been stopped, is 30 days.

Type: Integer

Valid Range: Minimum value of 1.

Required: No

MaxWaitTimeInSeconds

The maximum length of time, in seconds, that a managed Spot training job has to complete. It is the amount of time spent waiting for Spot capacity plus the amount of time the job can run. It must
StoppingCondition

be equal to or greater than MaxRuntimeInSeconds. If the job does not complete during this time, SageMaker ends the job.

When RetryStrategy is specified in the job request, MaxWaitTimeInSeconds specifies the maximum time for all of the attempts in total, not each individual attempt.

Type: Integer

Valid Range: Minimum value of 1.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
StudioLifecycleConfigDetails
Service: Amazon SageMaker Service
Details of the Amazon SageMaker Studio Lifecycle Configuration.

Contents

CreationTime
The creation time of the Amazon SageMaker Studio Lifecycle Configuration.
Type: Timestamp
Required: No

LastModifiedTime
This value is equivalent to CreationTime because Amazon SageMaker Studio Lifecycle Configurations are immutable.
Type: Timestamp
Required: No

StudioLifecycleConfigAppType
The App type to which the Lifecycle Configuration is attached.
Type: String
Valid Values: JupyterServer | KernelGateway | JupyterLab | CodeEditor
Required: No

StudioLifecycleConfigArn
The Amazon Resource Name (ARN) of the Lifecycle Configuration.
Type: String
Length Constraints: Maximum length of 256.
Pattern: \barn:aws[a-zA-Z\-]*:sagemaker:[a-zA-Z0-9\-]*:[0-9]{12}:studio-lifecycle-config/.*\b
Required: No

StudioLifecycleConfigName
The name of the Amazon SageMaker Studio Lifecycle Configuration.
Type: String
Length Constraints: Maximum length of 63.
Pattern: ^[a-zA-Z0-9\-](\-*[a-zA-Z0-9]*){0,62}
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:
• AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for Ruby V3
SubscribedWorkteam
Service: Amazon SageMaker Service

Describes a work team of a vendor that does the a labelling job.

Contents

WorkteamArn

The Amazon Resource Name (ARN) of the vendor that you have subscribed.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:workteam/.*

Required: Yes

ListingId

Marketplace product listing ID.

Type: String

Required: No

MarketplaceDescription

The description of the vendor from the Amazon Marketplace.

Type: String


Pattern: .+

Required: No

MarketplaceTitle

The title of the service provided by the vendor in the Amazon Marketplace.

Type: String


Pattern: .+

Required: No

SellerName

The name of the vendor in the Amazon Marketplace.

Type: String

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
SuggestionQuery

Service: Amazon SageMaker Service

Specified in the `GetSearchSuggestions` request. Limits the property names that are included in the response.

Contents

PropertyNameQuery

Defines a property name hint. Only property names that begin with the specified hint are included in the response.

Type: `PropertyNameQuery (p. 1873)` object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
**TabularJobConfig**

Service: Amazon SageMaker Service

The collection of settings used by an AutoML job V2 for the tabular problem type.

**Contents**

**TargetAttributeName**

The name of the target variable in supervised learning, usually represented by 'y'.

Type: String

Length Constraints: Minimum length of 1.

Required: Yes

**CandidateGenerationConfig**

The configuration information of how model candidates are generated.

Type: `CandidateGenerationConfig (p. 1314)` object

Required: No

**CompletionCriteria**

How long a job is allowed to run, or how many candidates a job is allowed to generate.

Type: `AutoMLJobCompletionCriteria (p. 1280)` object

Required: No

**FeatureSpecificationS3Uri**

A URL to the Amazon S3 data source containing selected features from the input data source to run an Autopilot job V2. You can input `FeatureAttributeNames` (optional) in JSON format as shown below:

```json
{ "FeatureAttributeNames": ["col1", "col2", ... ] }
```

You can also specify the data type of the feature (optional) in the format shown below:

```json
{ "FeatureDataTypes": {"col1":"numeric", "col2":"categorical" ... } }
```

**Note**

These column keys may not include the target column.

In ensembling mode, Autopilot only supports the following data types: numeric, categorical, text, and datetime. In HPO mode, Autopilot can support numeric, categorical, text, datetime, and sequence.

If only `FeatureDataTypes` is provided, the column keys (col1, col2,..) should be a subset of the column names in the input data.

If both `FeatureDataTypes` and `FeatureAttributeNames` are provided, then the column keys should be a subset of the column names provided in `FeatureAttributeNames`.

The key name `FeatureAttributeNames` is fixed. The values listed in ["col1", "col2", ...] are case sensitive and should be a list of strings containing unique values that are a subset of the column names in the input data. The list of columns provided must not include the target column.
TabularJobConfig

**Type:** String

**Length Constraints:** Maximum length of 1024.

**Pattern:** ^https://\([^/]+\)/\([^/]+\)/.*$

**Required:** No

**GenerateCandidateDefinitionsOnly**

Generates possible candidates without training the models. A model candidate is a combination of data preprocessors, algorithms, and algorithm parameter settings.

**Type:** Boolean

**Required:** No

**Mode**

The method that Autopilot uses to train the data. You can either specify the mode manually or let Autopilot choose for you based on the dataset size by selecting AUTO. In AUTO mode, Autopilot chooses ENSEMBLING for datasets smaller than 100 MB, and HYPERPARAMETER_TUNING for larger ones.

The ENSEMBLING mode uses a multi-stack ensemble model to predict classification and regression tasks directly from your dataset. This machine learning mode combines several base models to produce an optimal predictive model. It then uses a stacking ensemble method to combine predictions from contributing members. A multi-stack ensemble model can provide better performance over a single model by combining the predictive capabilities of multiple models. See Autopilot algorithm support for a list of algorithms supported by ENSEMBLING mode.

The HYPERPARAMETER_TUNING (HPO) mode uses the best hyperparameters to train the best version of a model. HPO automatically selects an algorithm for the type of problem you want to solve. Then HPO finds the best hyperparameters according to your objective metric. See Autopilot algorithm support for a list of algorithms supported by HYPERPARAMETER_TUNING mode.

**Type:** String

**Valid Values:** AUTO | ENSEMBLING | HYPERPARAMETER_TUNING

**Required:** No

**ProblemType**

The type of supervised learning problem available for the model candidates of the AutoML job V2. For more information, see Amazon SageMaker Autopilot problem types.  

**Note**

You must either specify the type of supervised learning problem in ProblemType and provide the AutoMLJobObjective metric, or none at all.

**Type:** String

**Valid Values:** BinaryClassification | MulticlassClassification | Regression

**Required:** No

**SampleWeightAttributeName**

If specified, this column name indicates which column of the dataset should be treated as sample weights for use by the objective metric during the training, evaluation, and the selection of the best model. This column is not considered as a predictive feature. For more information on Autopilot metrics, see Metrics and validation.
Sample weights should be numeric, non-negative, with larger values indicating which rows are more important than others. Data points that have invalid or no weight value are excluded.

Support for sample weights is available in Ensembling mode only.

Type: String
Length Constraints: Minimum length of 1. Maximum length of 256.
Pattern: ^[a-zA-Z0-9_\-\.]+$
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
TabularResolvedAttributes
Service: Amazon SageMaker Service

The resolved attributes specific to the tabular problem type.

Contents

ProblemType

The type of supervised learning problem available for the model candidates of the AutoML job V2 (Binary Classification, Multiclass Classification, Regression). For more information, see Amazon SageMaker Autopilot problem types.

Type: String

Valid Values: BinaryClassification | MulticlassClassification | Regression

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
Tag
Service: Amazon SageMaker Service

A tag object that consists of a key and an optional value, used to manage metadata for SageMaker AWS resources.

You can add tags to notebook instances, training jobs, hyperparameter tuning jobs, batch transform jobs, models, labeling jobs, work teams, endpoint configurations, and endpoints. For more information on adding tags to SageMaker resources, see AddTags.

For more information on adding metadata to your AWS resources with tagging, see Tagging AWS resources. For advice on best practices for managing AWS resources with tagging, see Tagging Best Practices: Implement an Effective AWS Resource Tagging Strategy.

Contents

Key

The tag key. Tag keys must be unique per resource.

Type: String


Pattern: ^([\p{L}\p{Z}\p{N}_./:=+-@]*)$

Required: Yes

Value

The tag value.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

Pattern: ^([\p{L}\p{Z}\p{N}_./:=+-@]*)$

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
TargetPlatform
Service: Amazon SageMaker Service

Contains information about a target platform that you want your model to run on, such as OS, architecture, and accelerators. It is an alternative of TargetDevice.

Contents

Arch
Specifies a target platform architecture.

- X86_64: 64-bit version of the x86 instruction set.
- X86: 32-bit version of the x86 instruction set.
- ARM64: ARMv8 64-bit CPU.
- ARM_EABIHF: ARMv7 32-bit, Hard Float.

Type: String
Valid Values: X86_64 | X86 | ARM64 | ARM_EABI | ARM_EABIHF

Required: Yes

Os
Specifies a target platform OS.

- LINUX: Linux-based operating systems.
- ANDROID: Android operating systems. Android API level can be specified using the ANDROID_PLATFORM compiler option. For example, "CompilerOptions":
  {'ANDROID_PLATFORM': 28}

Type: String
Valid Values: ANDROID | LINUX

Required: Yes

Accelerator
Specifies a target platform accelerator (optional).

- NVIDIA: Nvidia graphics processing unit. It also requires gpu-code, trt-ver, cuda-ver compiler options
- MALI: ARM Mali graphics processor
- INTEL_GRAPHICS: Integrated Intel graphics

Type: String
Valid Values: INTEL_GRAPHICS | MALI | NVIDIA | NNA

Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for Ruby V3
TargetTrackingScalingPolicyConfiguration

Service: Amazon SageMaker Service

A target tracking scaling policy. Includes support for predefined or customized metrics.

When using the `PutScalingPolicy` API, this parameter is required when you are creating a policy with the policy type `TargetTrackingScaling`.

Contents

MetricSpecification

An object containing information about a metric.

Type: `MetricSpecification` object

**Note:** This object is a Union. Only one member of this object can be specified or returned.

Required: No

TargetValue

The recommended target value to specify for the metric when creating a scaling policy.

Type: Double

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
TensorBoardAppSettings
Service: Amazon SageMaker Service

The TensorBoard app settings.

Contents

DefaultResourceSpec

The default instance type and the Amazon Resource Name (ARN) of the SageMaker image created on the instance.

Type: ResourceSpec (p. 1915) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
TensorBoardOutputConfig

Service: Amazon SageMaker Service

Configuration of storage locations for the Amazon SageMaker Debugger TensorBoard output data.

Contents

S3OutputPath

Path to Amazon S3 storage location for TensorBoard output.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: ^(https|s3)://([^/]+)/(.*)$

Required: Yes

LocalPath

Path to local storage location for tensorBoard output. Defaults to /opt/ml/output/tensorboard.

Type: String

Length Constraints: Maximum length of 4096.

Pattern: .*

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
TextClassificationJobConfig
Service: Amazon SageMaker Service

The collection of settings used by an AutoML job V2 for the text classification problem type.

Contents

ContentColumn

The name of the column used to provide the sentences to be classified. It should not be the same as the target column.

Type: String
Length Constraints: Minimum length of 1. Maximum length of 256.
Required: Yes

TargetLabelColumn

The name of the column used to provide the class labels. It should not be same as the content column.

Type: String
Length Constraints: Minimum length of 1. Maximum length of 256.
Required: Yes

CompletionCriteria

How long a job is allowed to run, or how many candidates a job is allowed to generate.

Type: AutoMLJobCompletionCriteria (p. 1280) object
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
TextGenerationJobConfig
Service: Amazon SageMaker Service

The collection of settings used by an AutoML job V2 for the text generation problem type.

**Note**
The text generation models that support fine-tuning in Autopilot are currently accessible exclusively in regions supported by Canvas. Refer to the documentation of Canvas for the full list of its supported Regions.

**Contents**

**BaseModelName**
The name of the base model to fine-tune. Autopilot supports fine-tuning a variety of large language models. For information on the list of supported models, see [Text generation models supporting fine-tuning in Autopilot](#). If no BaseModelName is provided, the default model used is **Falcon7BInstruct**.

Type: String
Length Constraints: Minimum length of 1. Maximum length of 256.
Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9])*$`
Required: No

**CompletionCriteria**
How long a fine-tuning job is allowed to run. For TextGenerationJobConfig problem types, the MaxRuntimePerTrainingJobInSeconds attribute of AutoMLJobCompletionCriteria defaults to 72h (259200s).

Type: [AutoMLJobCompletionCriteria](#) object
Required: No

**TextGenerationHyperParameters**
The hyperparameters used to configure and optimize the learning process of the base model. You can set any combination of the following hyperparameters for all base models. For more information on each supported hyperparameter, see [Optimize the learning process of your text generation models with hyperparameters](#).

- **"epochCount"**: The number of times the model goes through the entire training dataset. Its value should be a string containing an integer value within the range of "1" to "10".
- **"batchSize"**: The number of data samples used in each iteration of training. Its value should be a string containing an integer value within the range of "1" to "64".
- **"learningRate"**: The step size at which a model's parameters are updated during training. Its value should be a string containing a floating-point value within the range of "0" to "1".
- **"learningRateWarmupSteps"**: The number of training steps during which the learning rate gradually increases before reaching its target or maximum value. Its value should be a string containing an integer value within the range of "0" to "250".

Here is an example where all four hyperparameters are configured.

```
{ "epochCount":"5", "learningRate":"0.5", "batchSize": "32", "learningRateWarmupSteps": "10" }
```

Type: String to string map
Map Entries: Minimum number of 0 items. Maximum number of 30 items.

Key Length Constraints: Maximum length of 32.

Key Pattern: ^[a-zA-Z0-9_.-]+$

Value Length Constraints: Maximum length of 16.

Value Pattern: ^[a-zA-Z0-9_.-]+$

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
TextGenerationResolvedAttributes
Service: Amazon SageMaker Service
The resolved attributes specific to the text generation problem type.

Contents

BaseModelName
The name of the base model to fine-tune.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 256.
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]*)*
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
**TimeSeriesConfig**

Service: Amazon SageMaker Service

The collection of components that defines the time-series.

**Contents**

**ItemIdentifierAttributeName**

The name of the column that represents the set of item identifiers for which you want to predict the target value.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 256.

Required: Yes

**TargetAttributeName**

The name of the column representing the target variable that you want to predict for each item in your dataset. The data type of the target variable must be numerical.

Type: String

Length Constraints: Minimum length of 1.

Required: Yes

**TimestampAttributeName**

The name of the column indicating a point in time at which the target value of a given item is recorded.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 256.

Required: Yes

**GroupingAttributeNames**

A set of columns names that can be grouped with the item identifier column to create a composite key for which a target value is predicted.

Type: Array of strings

Array Members: Minimum number of 1 item. Maximum number of 5 items.

Length Constraints: Minimum length of 1. Maximum length of 256.

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- AWS SDK for Ruby V3
TimeSeriesForecastingJobConfig

Service: Amazon SageMaker Service

The collection of settings used by an AutoML job V2 for the time-series forecasting problem type.

Contents

ForecastFrequency

The frequency of predictions in a forecast.

Valid intervals are an integer followed by Y (Year), M (Month), W (Week), D (Day), H (Hour), and min (Minute). For example, 1D indicates every day and 15min indicates every 15 minutes. The value of a frequency must not overlap with the next larger frequency. For example, you must use a frequency of 1H instead of 60min.

The valid values for each frequency are the following:
- Minute - 1-59
- Hour - 1-23
- Day - 1-6
- Week - 1-4
- Month - 1-11
- Year - 1

Type: String


Pattern: ^1Y|Y|([1-9]|1[0-1])M|M|([1-9]|1[0-9]|2[0-3])H|H|([1-5]?)W|W|([1-6])D|D|([1-9]|1[0-9]|2[0-3])min$

Required: Yes

ForecastHorizon

The number of time-steps that the model predicts. The forecast horizon is also called the prediction length. The maximum forecast horizon is the lesser of 500 time-steps or 1/4 of the time-steps in the dataset.

Type: Integer

Valid Range: Minimum value of 1.

Required: Yes

TimeSeriesConfig

The collection of components that defines the time-series.

Type: TimeSeriesConfig (p. 1989) object

Required: Yes

CompletionCriteria

How long a job is allowed to run, or how many candidates a job is allowed to generate.

Type: AutoMLJobCompletionCriteria (p. 1280) object

Required: No
FeatureSpecificationS3Uri

A URL to the Amazon S3 data source containing additional selected features that complement the target, itemID, timestamp, and grouped columns set in TimeSeriesConfig. When not provided, the AutoML job V2 includes all the columns from the original dataset that are not already declared in TimeSeriesConfig. If provided, the AutoML job V2 only considers these additional columns as a complement to the ones declared in TimeSeriesConfig.

You can input FeatureAttributeNames (optional) in JSON format as shown below:

{ "FeatureAttributeNames": ["col1", "col2", ...] }.

You can also specify the data type of the feature (optional) in the format shown below:

{ "FeatureDataTypes": {"col1":"numeric", "col2":"categorical" ... } }.

Autopilot supports the following data types: numeric, categorical, text, and datetime.

Note

These column keys must not include any column set in TimeSeriesConfig.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: ^(https|s3)://([^/]+)?([^/]+)$

Required: No

ForecastQuantiles

The quantiles used to train the model for forecasts at a specified quantile. You can specify quantiles from 0.01 (p1) to 0.99 (p99), by increments of 0.01 or higher. Up to five forecast quantiles can be specified. When ForecastQuantiles is not provided, the AutoML job uses the quantiles p10, p50, and p90 as default.

Type: Array of strings

Array Members: Minimum number of 1 item. Maximum number of 5 items.


Pattern: (^p[1-9]\d?$)

Required: No

HolidayConfig

The collection of holiday featurization attributes used to incorporate national holiday information into your forecasting model.

Type: Array of HolidayConfigAttributes (p. 1507) objects

Array Members: Fixed number of 1 item.

Required: No

Transformations

The transformations modifying specific attributes of the time-series, such as filling strategies for missing values.

Type: TimeSeriesTransformations (p. 1995) object
See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
TimeSeriesForecastingSettings

Service: Amazon SageMaker Service

Time series forecast settings for the SageMaker Canvas application.

Contents

AmazonForecastRoleArn

The IAM role that Canvas passes to Amazon Forecast for time series forecasting. By default, Canvas uses the execution role specified in the UserProfile that launches the Canvas application. If an execution role is not specified in the UserProfile, Canvas uses the execution role specified in the Domain that owns the UserProfile. To allow time series forecasting, this IAM role should have the AmazonSageMakerCanvasForecastAccess policy attached and forecast.amazonaws.com added in the trust relationship as a service principal.

Type: String


Pattern: ^arn:aws[a-z\-]*:iam::\d{12}:role/?[a-zA-Z_0-9+=,.@\-_\/]+$

Required: No

Status

Describes whether time series forecasting is enabled or disabled in the Canvas application.

Type: String

Valid Values: ENABLED | DISABLED

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
TimeSeriesTransformations
Service: Amazon SageMaker Service

Transformations allowed on the dataset. Supported transformations are Filling and Aggregation. Filling specifies how to add values to missing values in the dataset. Aggregation defines how to aggregate data that does not align with forecast frequency.

Contents

Aggregation

A key value pair defining the aggregation method for a column, where the key is the column name and the value is the aggregation method.

The supported aggregation methods are sum (default), avg, first, min, max.

Note
Aggregation is only supported for the target column.

Type: String to string map
Map Entries: Maximum number of 50 items.
Key Length Constraints: Minimum length of 1. Maximum length of 256.
Valid Values: sum | avg | first | min | max
Required: No

Filling

A key value pair defining the filling method for a column, where the key is the column name and the value is an object which defines the filling logic. You can specify multiple filling methods for a single column.

The supported filling methods and their corresponding options are:
- frontfill: none (Supported only for target column)
- middlefill: zero, value, median, mean, min, max
- backfill: zero, value, median, mean, min, max
- futurefill: zero, value, median, mean, min, max

To set a filling method to a specific value, set the fill parameter to the chosen filling method value (for example "backfill" : "value"), and define the filling value in an additional parameter prefixed with ",value". For example, to set backfill to a value of 2, you must include two parameters: "backfill": "value" and "backfill_value":"2".

Type: String to string to string map map
Map Entries: Maximum number of 50 items.
Key Length Constraints: Minimum length of 1. Maximum length of 256.
Valid Keys: frontfill | middlefill | backfill | futurefill | frontfill_value | middlefill_value | backfill_value | futurefill_value
Value Length Constraints: Minimum length of 1. Maximum length of 256.
Value Pattern: ^[a-zA-Z0-9\-_]+$
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
TrafficPattern
Service: Amazon SageMaker Service
Defines the traffic pattern of the load test.

Contents

Phases
Defines the phases traffic specification.
Type: Array of Phase (p. 1810) objects
Array Members: Minimum number of 1 item.
Required: No

Stairs
Defines the stairs traffic pattern.
Type: Stairs (p. 1967) object
Required: No

TrafficType
Defines the traffic patterns. Choose either PHASES or STAIRS.
Type: String
Valid Values: PHASES | STAIRS
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
TrafficRoutingConfig

Service: Amazon SageMaker Service

Defines the traffic routing strategy during an endpoint deployment to shift traffic from the old fleet to the new fleet.

Contents

Type

Traffic routing strategy type.
- ALL_AT_ONCE: Endpoint traffic shifts to the new fleet in a single step.
- CANARY: Endpoint traffic shifts to the new fleet in two steps. The first step is the canary, which is a small portion of the traffic. The second step is the remainder of the traffic.
- LINEAR: Endpoint traffic shifts to the new fleet in n steps of a configurable size.

Type: String
Valid Values: ALL_AT_ONCE | CANARY | LINEAR
Required: Yes

WaitIntervalInSeconds

The waiting time (in seconds) between incremental steps to turn on traffic on the new endpoint fleet.

Type: Integer
Valid Range: Minimum value of 0. Maximum value of 3600.
Required: Yes

CanarySize

Batch size for the first step to turn on traffic on the new endpoint fleet. Value must be less than or equal to 50% of the variant's total instance count.

Type: CapacitySize (p. 1318) object
Required: No

LinearStepSize

Batch size for each step to turn on traffic on the new endpoint fleet. Value must be 10-50% of the variant's total instance count.

Type: CapacitySize (p. 1318) object
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
• AWS SDK for Ruby V3
TrainingImageConfig
Service: Amazon SageMaker Service

The configuration to use an image from a private Docker registry for a training job.

Contents

TrainingRepositoryAccessMode

The method that your training job will use to gain access to the images in your private Docker registry. For access to an image in a private Docker registry, set to Vpc.

Type: String

Valid Values: Platform | Vpc

Required: Yes

TrainingRepositoryAuthConfig

An object containing authentication information for a private Docker registry containing your training images.

Type: TrainingRepositoryAuthConfig (p. 2016) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
TrainingJob
Service: Amazon SageMaker Service
Contains information about a training job.

Contents

AlgorithmSpecification
Information about the algorithm used for training, and algorithm metadata.
Type: AlgorithmSpecification (p. 1225) object
Required: No

AutoMLJobArn
The Amazon Resource Name (ARN) of the job.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 256.
Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:automl-job/.*
Required: No

BillableTimeInSeconds
The billable time in seconds.
Type: Integer
Valid Range: Minimum value of 1.
Required: No

CheckpointConfig
Contains information about the output location for managed spot training checkpoint data.
Type: CheckpointConfig (p. 1328) object
Required: No

CreationTime
A timestamp that indicates when the training job was created.
Type: Timestamp
Required: No

DebugHookConfig
Configuration information for the Amazon SageMaker Debugger hook parameters, metric and tensor collections, and storage paths. To learn more about how to configure the DebugHookConfig parameter, see Use the SageMaker and Debugger Configuration API Operations to Create, Update, and Debug Your Training Job.
Type: DebugHookConfig (p. 1395) object
Required: No
**DebugRuleConfigurations**

Information about the debug rule configuration.

Type: Array of **DebugRuleConfiguration (p. 1397)** objects

Array Members: Minimum number of 0 items. Maximum number of 20 items.

Required: No

**DebugRuleEvaluationStatuses**

Information about the evaluation status of the rules for the training job.

Type: Array of **DebugRuleEvaluationStatus (p. 1399)** objects

Array Members: Minimum number of 0 items. Maximum number of 20 items.

Required: No

**EnableInterContainerTrafficEncryption**

To encrypt all communications between ML compute instances in distributed training, choose `True`. Encryption provides greater security for distributed training, but training might take longer. How long it takes depends on the amount of communication between compute instances, especially if you use a deep learning algorithm in distributed training.

Type: Boolean

Required: No

**EnableManagedSpotTraining**

When true, enables managed spot training using Amazon EC2 Spot instances to run training jobs instead of on-demand instances. For more information, see **Managed Spot Training**.

Type: Boolean

Required: No

**EnableNetworkIsolation**

If the `TrainingJob` was created with network isolation, the value is set to `true`. If network isolation is enabled, nodes can't communicate beyond the VPC they run in.

Type: Boolean

Required: No

**Environment**

The environment variables to set in the Docker container.

Type: String to string map

Map Entries: Maximum number of 100 items.

Key Length Constraints: Maximum length of 512.

Key Pattern: `[a-zA-Z_][a-zA-Z0-9_]*`

Value Length Constraints: Maximum length of 512.

Value Pattern: `\S\s*`
Required: No

**ExperimentConfig**

Associates a SageMaker job as a trial component with an experiment and trial. Specified when you call the following APIs:

- `CreateProcessingJob`
- `CreateTrainingJob`
- `CreateTransformJob`

Type: `ExperimentConfig (p. 1473)` object

Required: No

**FailureReason**

If the training job failed, the reason it failed.

Type: String

Length Constraints: Maximum length of 1024.

Required: No

**FinalMetricDataList**

A list of final metric values that are set when the training job completes. Used only if the training job was configured to use metrics.

Type: Array of `MetricData (p. 1650)` objects

Array Members: Minimum number of 0 items. Maximum number of 40 items.

Required: No

**HyperParameters**

Algorithm-specific parameters.

Type: String to string map

Map Entries: Minimum number of 0 items. Maximum number of 100 items.

Key Length Constraints: Maximum length of 256.

Key Pattern: . *

Value Length Constraints: Maximum length of 2500.

Value Pattern: . *

Required: No

**InputDataConfig**

An array of `Channel` objects that describes each data input channel.

Your input must be in the same AWS region as your training job.

Type: Array of `Channel (p. 1324)` objects

Array Members: Minimum number of 1 item. Maximum number of 20 items.

Required: No
LabelingJobArn

The Amazon Resource Name (ARN) of the labeling job.
Type: String
Length Constraints: Maximum length of 2048.
Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:labeling-job/.*
Required: No

LastModifiedTime

A timestamp that indicates when the status of the training job was last modified.
Type: Timestamp
Required: No

ModelArtifacts

Information about the Amazon S3 location that is configured for storing model artifacts.
Type: ModelArtifacts (p. 1659) object
Required: No

OutputDataConfig

The S3 path where model artifacts that you configured when creating the job are stored. SageMaker creates subfolders for model artifacts.
Type: OutputDataConfig (p. 1792) object
Required: No

ProfilerConfig

Configuration information for Amazon SageMaker Debugger system monitoring, framework profiling, and storage paths.
Type: ProfilerConfig (p. 1860) object
Required: No

ResourceConfig

Resources, including ML compute instances and ML storage volumes, that are configured for model training.
Type: ResourceConfig (p. 1910) object
Required: No

RetryStrategy

The number of times to retry the job when the job fails due to an InternalServerError.
Type: RetryStrategy (p. 1918) object
Required: No

RoleArn

The AWS Identity and Access Management (IAM) role configured for the training job.
Type: String


Pattern: ^arn:aws[a-z\-]*:iam::\d{12}:role/?[a-zA-Z_0-9+=,.@-\_/]+$

Required: No

SecondaryStatus

Provides detailed information about the state of the training job. For detailed information about the secondary status of the training job, see StatusMessage under SecondaryStatusTransition.

SageMaker provides primary statuses and secondary statuses that apply to each of them:

InProgress
- Starting - Starting the training job.
- Downloading - An optional stage for algorithms that support File training input mode. It indicates that data is being downloaded to the ML storage volumes.
- Training - Training is in progress.
- Uploading - Training is complete and the model artifacts are being uploaded to the S3 location.

Completed
- Completed - The training job has completed.

Failed
- Failed - The training job has failed. The reason for the failure is returned in the FailureReason field of DescribeTrainingJobResponse.

Stopped
- MaxRuntimeExceeded - The job stopped because it exceeded the maximum allowed runtime.
- Stopped - The training job has stopped.

Stopping
- Stopping - Stopping the training job.

Important
Valid values for SecondaryStatus are subject to change.

We no longer support the following secondary statuses:
- LaunchingMLInstances
- PreparingTrainingStack
- DownloadingTrainingImage

Type: String

Valid Values: Starting | LaunchingMLInstances | PreparingTrainingStack | Downloading | DownloadingTrainingImage | Training | Uploading | Stopping | Stopped | MaxRuntimeExceeded | Completed | Failed | Interrupted | MaxWaitTimeExceeded | Updating | Restarting

Required: No

SecondaryStatusTransitions

A history of all of the secondary statuses that the training job has transitioned through.

Type: Array of SecondaryStatusTransition (p. 1941) objects

Required: No
StoppingCondition

Specifies a limit to how long a model training job can run. It also specifies how long a managed Spot training job has to complete. When the job reaches the time limit, SageMaker ends the training job. Use this API to cap model training costs.

To stop a job, SageMaker sends the algorithm the SIGTERM signal, which delays job termination for 120 seconds. Algorithms can use this 120-second window to save the model artifacts, so the results of training are not lost.

Type: StoppingCondition (p. 1968) object

Required: No

Tags

An array of key-value pairs. You can use tags to categorize your AWS resources in different ways, for example, by purpose, owner, or environment. For more information, see Tagging AWS Resources.

Type: Array of Tag (p. 1979) objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

TensorBoardOutputConfig

Configuration of storage locations for the Amazon SageMaker Debugger TensorBoard output data.

Type: TensorBoardOutputConfig (p. 1984) object

Required: No

TrainingEndTime

Indicates the time when the training job ends on training instances. You are billed for the time interval between the value of TrainingStartTime and this time. For successful jobs and stopped jobs, this is the time after model artifacts are uploaded. For failed jobs, this is the time when SageMaker detects a job failure.

Type: Timestamp

Required: No

TrainingJobArn

The Amazon Resource Name (ARN) of the training job.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:training-job/.*

Required: No

TrainingJobName

The name of the training job.

Type: String


Pattern: ^[a-zA-Z0-9-](\-*[a-zA-Z0-9])\{0,62}$
TrainingJobStatus

The status of the training job.

Training job statuses are:
- InProgress - The training is in progress.
- Completed - The training job has completed.
- Failed - The training job has failed. To see the reason for the failure, see the FailureReason field in the response to a DescribeTrainingJobResponse call.
- Stopping - The training job is stopping.
- Stopped - The training job has stopped.

For more detailed information, see SecondaryStatus.

Type: String

Valid Values: InProgress | Completed | Failed | Stopping | Stopped

Required: No

TrainingStartTime

Indicates the time when the training job starts on training instances. You are billed for the time interval between this time and the value of TrainingEndTime. The start time in CloudWatch Logs might be later than this time. The difference is due to the time it takes to download the training data and to the size of the training container.

Type: Timestamp

Required: No

TrainingTimeInSeconds

The training time in seconds.

Type: Integer

Valid Range: Minimum value of 1.

Required: No

TuningJobArn

The Amazon Resource Name (ARN) of the associated hyperparameter tuning job if the training job was launched by a hyperparameter tuning job.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:hyper-parameter-tuning-job/.*

Required: No

VpcConfig

A VpcConfig object that specifies the VPC that this training job has access to. For more information, see Protect Training Jobs by Using an Amazon Virtual Private Cloud.

Type: VpcConfig (p. 2076) object
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
TrainingJobDefinition
Service: Amazon SageMaker Service
Defines the input needed to run a training job using the algorithm.

Contents

InputDataConfig
An array of Channel objects, each of which specifies an input source.
Type: Array of Channel (p. 1324) objects
Array Members: Minimum number of 1 item. Maximum number of 20 items.
Required: Yes

OutputDataConfig
The path to the S3 bucket where you want to store model artifacts. SageMaker creates subfolders for the artifacts.
Type: OutputDataConfig (p. 1792) object
Required: Yes

ResourceConfig
The resources, including the ML compute instances and ML storage volumes, to use for model training.
Type: ResourceConfig (p. 1910) object
Required: Yes

StoppingCondition
Specifies a limit to how long a model training job can run. It also specifies how long a managed Spot training job has to complete. When the job reaches the time limit, SageMaker ends the training job. Use this API to cap model training costs.

To stop a job, SageMaker sends the algorithm the SIGTERM signal, which delays job termination for 120 seconds. Algorithms can use this 120-second window to save the model artifacts.
Type: StoppingCondition (p. 1968) object
Required: Yes

TrainingInputMode
The training input mode that the algorithm supports. For more information about input modes, see Algorithms.

Pipe mode
If an algorithm supports Pipe mode, Amazon SageMaker streams data directly from Amazon S3 to the container.

File mode
If an algorithm supports File mode, SageMaker downloads the training data from S3 to the provisioned ML storage volume, and mounts the directory to the Docker volume for the training container.
You must provision the ML storage volume with sufficient capacity to accommodate the data downloaded from S3. In addition to the training data, the ML storage volume also stores the output model. The algorithm container uses the ML storage volume to also store intermediate information, if any.

For distributed algorithms, training data is distributed uniformly. Your training duration is predictable if the input data objects sizes are approximately the same. SageMaker does not split the files any further for model training. If the object sizes are skewed, training won't be optimal as the data distribution is also skewed when one host in a training cluster is overloaded, thus becoming a bottleneck in training.

**FastFile mode**

If an algorithm supports FastFile mode, SageMaker streams data directly from S3 to the container with no code changes, and provides file system access to the data. Users can author their training script to interact with these files as if they were stored on disk.

FastFile mode works best when the data is read sequentially. Augmented manifest files aren't supported. The startup time is lower when there are fewer files in the S3 bucket provided.

Type: String

Valid Values: Pipe | File | FastFile

Required: Yes

**HyperParameters**

The hyperparameters used for the training job.

Type: String to string map

Map Entries: Minimum number of 0 items. Maximum number of 100 items.

Key Length Constraints: Maximum length of 256.

Key Pattern: .*

Value Length Constraints: Maximum length of 2500.

Value Pattern: .*

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](https://aws.amazon.com/sdk-for-cpp)
- [AWS SDK for Go](https://aws.amazon.com/sdk-for-go)
- [AWS SDK for Java V2](https://aws.amazon.com/sdk-for-java)
- [AWS SDK for Ruby V3](https://aws.amazon.com/sdk-for-ruby)
TrainingJobStatusCounters
Service: Amazon SageMaker Service
The numbers of training jobs launched by a hyperparameter tuning job, categorized by status.

Contents

Completed
The number of completed training jobs launched by the hyperparameter tuning job.
Type: Integer
Valid Range: Minimum value of 0.
Required: No

InProgress
The number of in-progress training jobs launched by a hyperparameter tuning job.
Type: Integer
Valid Range: Minimum value of 0.
Required: No

NonRetryableError
The number of training jobs that failed and can't be retried. A failed training job can't be retried if it failed because a client error occurred.
Type: Integer
Valid Range: Minimum value of 0.
Required: No

RetryableError
The number of training jobs that failed, but can be retried. A failed training job can be retried only if it failed because an internal service error occurred.
Type: Integer
Valid Range: Minimum value of 0.
Required: No

Stopped
The number of training jobs launched by a hyperparameter tuning job that were manually stopped.
Type: Integer
Valid Range: Minimum value of 0.
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
TrainingJobStepMetadata
Service: Amazon SageMaker Service
Metadata for a training job step.

Contents

Arn

The Amazon Resource Name (ARN) of the training job that was run by this step execution.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:training-job/.*

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
TrainingJobSummary

Service: Amazon SageMaker Service

Provides summary information about a training job.

Contents

CreationTime

A timestamp that shows when the training job was created.

Type: Timestamp

Required: Yes

TrainingJobArn

The Amazon Resource Name (ARN) of the training job.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:training-job/.*

Required: Yes

TrainingJobName

The name of the training job that you want a summary for.

Type: String


Pattern: ^[a-zA-Z0-9][-]*$[a-zA-Z0-9]{0,62}$

Required: Yes

TrainingJobStatus

The status of the training job.

Type: String

Valid Values: InProgress | Completed | Failed | Stopping | Stopped

Required: Yes

LastModifiedTime

Timestamp when the training job was last modified.

Type: Timestamp

Required: No

TrainingEndTime

A timestamp that shows when the training job ended. This field is set only if the training job has one of the terminal statuses (Completed, Failed, or Stopped).

Type: Timestamp
Required: No

**WarmPoolStatus**

The status of the warm pool associated with the training job.

Type: *WarmPoolStatus (p. 2077)* object

Required: No

### See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
TrainingRepositoryAuthConfig
Service: Amazon SageMaker Service

An object containing authentication information for a private Docker registry.

Contents

TrainingRepositoryCredentialsProviderArn

The Amazon Resource Name (ARN) of an AWS Lambda function used to give SageMaker access credentials to your private Docker registry.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: arn:[\p{Alnum}-]+:lambda:[\p{Alnum}-]+[0-9]{12}:function:.*

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
**TrainingSpecification**

Service: Amazon SageMaker Service

Defines how the algorithm is used for a training job.

**Contents**

**SupportedTrainingInstanceTypes**

A list of the instance types that this algorithm can use for training.

Type: Array of strings

Valid Values: ml.m4.xlarge | ml.m4.2xlarge | ml.m4.4xlarge | ml.m4.10xlarge | ml.m4.16xlarge | ml.g4dn.xlarge | ml.g4dn.2xlarge | ml.g4dn.4xlarge | ml.g4dn.8xlarge | ml.g4dn.12xlarge | ml.g4dn.16xlarge | ml.m5.large | ml.m5.xlarge | ml.m5.2xlarge | ml.m5.4xlarge | ml.m5.12xlarge | ml.m5.24xlarge | ml.c4.xlarge | ml.c4.2xlarge | ml.c4.4xlarge | ml.c4.8xlarge | ml.p2.xlarge | ml.p2.8xlarge | ml.p2.16xlarge | ml.p3.xlarge | ml.p3.8xlarge | ml.p3.16xlarge | ml.p3dn.24xlarge | ml.p4d.24xlarge | ml.c5.xlarge | ml.c5.2xlarge | ml.c5.4xlarge | ml.c5.9xlarge | ml.c5.18xlarge | ml.c5n.xlarge | ml.c5n.2xlarge | ml.c5n.4xlarge | ml.c5n.9xlarge | ml.c5n.18xlarge | ml.g5.xlarge | ml.g5.2xlarge | ml.g5.4xlarge | ml.g5.8xlarge | ml.g5.16xlarge | ml.g5.12xlarge | ml.g5.24xlarge | ml.g5.48xlarge | ml.trn1.2xlarge | ml.trn1n.32xlarge | ml.p5.48xlarge

Required: Yes

**TrainingChannels**

A list of ChannelSpecification objects, which specify the input sources to be used by the algorithm.

Type: Array of ChannelSpecification (p. 1326) objects

Array Members: Minimum number of 1 item. Maximum number of 8 items.

Required: Yes

**TrainingImage**

The Amazon ECR registry path of the Docker image that contains the training algorithm.

Type: String

Length Constraints: Maximum length of 255.

Pattern: \[\S]+

Required: Yes

**AdditionalS3DataSource**

The additional data source used during the training job.

Type: AdditionalS3DataSource (p. 1222) object

Required: No
MetricDefinitions

A list of MetricDefinition objects, which are used for parsing metrics generated by the algorithm.

Type: Array of MetricDefinition (p. 1653) objects

Array Members: Minimum number of 0 items. Maximum number of 40 items.

Required: No

SupportedHyperParameters

A list of the HyperParameterSpecification objects, that define the supported hyperparameters. This is required if the algorithm supports automatic model tuning.

Type: Array of HyperParameterSpecification (p. 1542) objects

Array Members: Minimum number of 0 items. Maximum number of 100 items.

Required: No

SupportedTuningJobObjectiveMetrics

A list of the metrics that the algorithm emits that can be used as the objective metric in a hyperparameter tuning job.

Type: Array of HyperParameterTuningJobObjective (p. 1557) objects

Required: No

SupportsDistributedTraining

Indicates whether the algorithm supports distributed training. If set to false, buyers can't request more than one instance during training.

Type: Boolean

Required: No

TrainingImageDigest

An MD5 hash of the training algorithm that identifies the Docker image used for training.

Type: String

Length Constraints: Maximum length of 72.

Pattern: ^[Ss][Hh][Aa]256:[0-9a-fA-F]{64}$

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3

2018
TransformDataSource
Service: Amazon SageMaker Service

Describes the location of the channel data.

Contents

S3DataSource

The S3 location of the data source that is associated with a channel.

Type: TransformS3DataSource (p. 2036) object

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
TransformInput

Service: Amazon SageMaker Service

Describes the input source of a transform job and the way the transform job consumes it.

Contents

**DataSource**

Describes the location of the channel data, which is, the S3 location of the input data that the model can consume.

Type: TransformDataSource (p. 2019) object

Required: Yes

**CompressionType**

If your transform data is compressed, specify the compression type. Amazon SageMaker automatically decompresses the data for the transform job accordingly. The default value is None.

Type: String

Valid Values: None | Gzip

Required: No

**ContentType**

The multipurpose internet mail extension (MIME) type of the data. Amazon SageMaker uses the MIME type with each http call to transfer data to the transform job.

Type: String

Length Constraints: Maximum length of 256.

Pattern: .*

Required: No

**SplitType**

The method to use to split the transform job’s data files into smaller batches. Splitting is necessary when the total size of each object is too large to fit in a single request. You can also use data splitting to improve performance by processing multiple concurrent mini-batches. The default value for SplitType is None, which indicates that input data files are not split, and request payloads contain the entire contents of an input object. Set the value of this parameter to Line to split records on a newline character boundary. SplitType also supports a number of record-oriented binary data formats. Currently, the supported record formats are:

- RecordIO
- TFRecord

When splitting is enabled, the size of a mini-batch depends on the values of the BatchStrategy and MaxPayloadInMB parameters. When the value of BatchStrategy is MultiRecord, Amazon SageMaker sends the maximum number of records in each request, up to the MaxPayloadInMB limit. If the value of BatchStrategy is SingleRecord, Amazon SageMaker sends individual records in each request.

**Note**

Some data formats represent a record as a binary payload wrapped with extra padding bytes. When splitting is applied to a binary data format, padding is removed if the value
of BatchStrategy is set to SingleRecord. Padding is not removed if the value of BatchStrategy is set to MultiRecord. For more information about RecordIO, see Create a Dataset Using RecordIO in the MXNet documentation. For more information about TFRecord, see Consuming TFRecord data in the TensorFlow documentation.

Type: String

Valid Values: None | Line | RecordIO | TFRecord

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
TransformJob
Service: Amazon SageMaker Service

A batch transform job. For information about SageMaker batch transform, see Use Batch Transform.

Contents

AutoMLJobArn
The Amazon Resource Name (ARN) of the AutoML job that created the transform job.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 256.
Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:automl-job/.*
Required: No

BatchStrategy
Specifies the number of records to include in a mini-batch for an HTTP inference request. A record is a single unit of input data that inference can be made on. For example, a single line in a CSV file is a record.
Type: String
Valid Values: MultiRecord | SingleRecord
Required: No

CreationTime
A timestamp that shows when the transform Job was created.
Type: Timestamp
Required: No

DataCaptureConfig
Configuration to control how SageMaker captures inference data for batch transform jobs.
Type: BatchDataCaptureConfig (p. 1301) object
Required: No

DataProcessing
The data structure used to specify the data to be used for inference in a batch transform job and to associate the data that is relevant to the prediction results in the output. The input filter provided allows you to exclude input data that is not needed for inference in a batch transform job. The output filter provided allows you to include input data relevant to interpreting the predictions in the output from the job. For more information, see Associate Prediction Results with their Corresponding Input Records.
Type: DataProcessing (p. 1386) object
Required: No

Environment
The environment variables to set in the Docker container. We support up to 16 key and values entries in the map.
**TransformJob**

Type: String to string map

Map Entries: Maximum number of 16 items.

Key Length Constraints: Maximum length of 1024.

Key Pattern: [a-zA-Z_][a-zA-Z0-9_]{0,1023}

Value Length Constraints: Maximum length of 10240.

Value Pattern: \S\s*

Required: No

**ExperimentConfig**

Associates a SageMaker job as a trial component with an experiment and trial. Specified when you call the following APIs:

- CreateProcessingJob
- CreateTrainingJob
- CreateTransformJob

Type: ExperimentConfig (p. 1473) object

Required: No

**FailureReason**

If the transform job failed, the reason it failed.

Type: String

Length Constraints: Maximum length of 1024.

Required: No

**LabelingJobArn**

The Amazon Resource Name (ARN) of the labeling job that created the transform job.

Type: String

Length Constraints: Maximum length of 2048.

Pattern: arn:aws[a-z-]*:sagemaker:[a-z0-9-]*:[0-9]{12}:labeling-job/.*

Required: No

**MaxConcurrentTransforms**

The maximum number of parallel requests that can be sent to each instance in a transform job. If MaxConcurrentTransforms is set to 0 or left unset, SageMaker checks the optional execution-parameters to determine the settings for your chosen algorithm. If the execution-parameters endpoint is not enabled, the default value is 1. For built-in algorithms, you don't need to set a value for MaxConcurrentTransforms.

Type: Integer

Valid Range: Minimum value of 0.

Required: No
MaxPayloadInMB

The maximum allowed size of the payload, in MB. A payload is the data portion of a record (without metadata). The value in MaxPayloadInMB must be greater than, or equal to, the size of a single record. To estimate the size of a record in MB, divide the size of your dataset by the number of records. To ensure that the records fit within the maximum payload size, we recommend using a slightly larger value. The default value is 6 MB. For cases where the payload might be arbitrarily large and is transmitted using HTTP chunked encoding, set the value to 0. This feature works only in supported algorithms. Currently, SageMaker built-in algorithms do not support HTTP chunked encoding.

Type: Integer
Valid Range: Minimum value of 0.
Required: No

ModelClientConfig

Configures the timeout and maximum number of retries for processing a transform job invocation.

Type: ModelClientConfig (p. 1675) object
Required: No

ModelName

The name of the model associated with the transform job.

Type: String
Length Constraints: Maximum length of 63.
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])*
Required: No

Tags

A list of tags associated with the transform job.

Type: Array of Tag (p. 1979) objects
Array Members: Minimum number of 0 items. Maximum number of 50 items.
Required: No

TransformEndTime

Indicates when the transform job has been completed, or has stopped or failed. You are billed for the time interval between this time and the value of TransformStartTime.

Type: Timestamp
Required: No

TransformInput

Describes the input source of a transform job and the way the transform job consumes it.

Type: TransformInput (p. 2020) object
Required: No

TransformJobArn

The Amazon Resource Name (ARN) of the transform job.
Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:transform-job/.*

Required: No

**TransformJobName**

The name of the transform job.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,62}

Required: No

**TransformJobStatus**

The status of the transform job.

Transform job statuses are:

- **InProgress** - The job is in progress.
- **Completed** - The job has completed.
- **Failed** - The transform job has failed. To see the reason for the failure, see the FailureReason field in the response to a DescribeTransformJob call.
- **Stopping** - The transform job is stopping.
- **Stopped** - The transform job has stopped.

Type: String

Valid Values: InProgress | Completed | Failed | Stopping | Stopped

Required: No

**TransformOutput**

Describes the results of a transform job.

Type: `TransformOutput (p. 2032)` object

Required: No

**TransformResources**

Describes the resources, including ML instance types and ML instance count, to use for transform job.

Type: `TransformResources (p. 2034)` object

Required: No

**TransformStartTime**

Indicates when the transform job starts on ML instances. You are billed for the time interval between this time and the value of TransformEndTime.

Type: Timestamp

Required: No
See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
TransformJobDefinition
Service: Amazon SageMaker Service

Defines the input needed to run a transform job using the inference specification specified in the algorithm.

Contents

TransformInput
A description of the input source and the way the transform job consumes it.

Type: TransformInput (p. 2020) object

Required: Yes

TransformOutput
Identifies the Amazon S3 location where you want Amazon SageMaker to save the results from the transform job.

Type: TransformOutput (p. 2032) object

Required: Yes

TransformResources
Identifies the ML compute instances for the transform job.

Type: TransformResources (p. 2034) object

Required: Yes

BatchStrategy
A string that determines the number of records included in a single mini-batch.

SingleRecord means only one record is used per mini-batch. MultiRecord means a mini-batch is set to contain as many records that can fit within the MaxPayloadInMB limit.

Type: String

Valid Values: MultiRecord | SingleRecord

Required: No

Environment
The environment variables to set in the Docker container. We support up to 16 key and values entries in the map.

Type: String to string map

Map Entries: Maximum number of 16 items.

Key Length Constraints: Maximum length of 1024.

Key Pattern: [a-zA-Z_][a-zA-Z0-9_]{0,1023}

Value Length Constraints: Maximum length of 10240.

Value Pattern: \S\s*
**MaxConcurrentTransforms**

The maximum number of parallel requests that can be sent to each instance in a transform job. The default value is 1.

Type: Integer

Valid Range: Minimum value of 0.

Required: No

**MaxPayloadInMB**

The maximum payload size allowed, in MB. A payload is the data portion of a record (without metadata).

Type: Integer

Valid Range: Minimum value of 0.

Required: No

### See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](https://aws.amazon.com/sdk-for-cpp/)
- [AWS SDK for Go](https://github.com/aws/aws-sdk-go)
- [AWS SDK for Java V2](https://aws.amazon.com/sdk-for-java/)
- [AWS SDK for Ruby V3](https://github.com/aws/aws-sdk-ruby)
TransformJobStepMetadata
Service: Amazon SageMaker Service

Metadata for a transform job step.

Contents

Arn
The Amazon Resource Name (ARN) of the transform job that was run by this step execution.
Type: String
Length Constraints: Maximum length of 256.
Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:transform-job/.*
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

• AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for Ruby V3
TransformJobSummary

Service: Amazon SageMaker Service

Provides a summary of a transform job. Multiple TransformJobSummary objects are returned as a list after in response to a ListTransformJobs call.

Contents

CreationTime

A timestamp that shows when the transform Job was created.

Type: Timestamp

Required: Yes

TransformJobArn

The Amazon Resource Name (ARN) of the transform job.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:transform-job/.*

Required: Yes

TransformJobName

The name of the transform job.

Type: String


Pattern: ^[a-zA-Z0-9\-](-*[a-zA-Z0-9]\{0,62}$

Required: Yes

TransformJobStatus

The status of the transform job.

Type: String

Valid Values: InProgress | Completed | Failed | Stopping | Stopped

Required: Yes

FailureReason

If the transform job failed, the reason it failed.

Type: String

Length Constraints: Maximum length of 1024.

Required: No

LastModifiedTime

Indicates when the transform job was last modified.
Type: Timestamp
Required: No

**TransformEndTime**

Indicates when the transform job ends on compute instances. For successful jobs and stopped jobs, this is the exact time recorded after the results are uploaded. For failed jobs, this is when Amazon SageMaker detected that the job failed.

Type: Timestamp
Required: No

### See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
**TransformOutput**

Service: Amazon SageMaker Service

Describes the results of a transform job.

## Contents

### S3OutputPath

The Amazon S3 path where you want Amazon SageMaker to store the results of the transform job. For example, s3://bucket-name/key-name-prefix.

For every S3 object used as input for the transform job, batch transform stores the transformed data with an .out suffix in a corresponding subfolder in the location in the output prefix. For example, for the input data stored at s3://bucket-name/input-name-prefix/dataset01/data.csv, batch transform stores the transformed data at s3://bucket-name/output-name-prefix/input-name-prefix/data.csv.out. Batch transform doesn't upload partially processed objects. For an input S3 object that contains multiple records, it creates an .out file only if the transform job succeeds on the entire file. When the input contains multiple S3 objects, the batch transform job processes the listed S3 objects and uploads only the output for successfully processed objects. If any object fails in the transform job batch transform marks the job as failed to prompt investigation.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: ^(https|s3)://([^/]+)/(.*)$

Required: Yes

### Accept

The MIME type used to specify the output data. Amazon SageMaker uses the MIME type with each http call to transfer data from the transform job.

Type: String

Length Constraints: Maximum length of 256.

Pattern: .*

Required: No

### AssembleWith

Defines how to assemble the results of the transform job as a single S3 object. Choose a format that is most convenient to you. To concatenate the results in binary format, specify None. To add a newline character at the end of every transformed record, specify Line.

Type: String

Valid Values: None | Line

Required: No

### KmsKeyId

The AWS Key Management Service (AWS KMS) key that Amazon SageMaker uses to encrypt the model artifacts at rest using Amazon S3 server-side encryption. The KmsKeyId can be any of the following formats:

- Key ID: 1234abcd-12ab-34cd-56ef-1234567890ab
• Key ARN: arn:aws:kms:us-west-2:111122223333:key/1234abcd-12ab-34cd-56ef-1234567890ab
• Alias name: alias/ExampleAlias
• Alias name ARN: arn:aws:kms:us-west-2:111122223333:alias/ExampleAlias

If you don't provide a KMS key ID, Amazon SageMaker uses the default KMS key for Amazon S3 for your role's account. For more information, see KMS-Managed Encryption Keys in the Amazon Simple Storage Service Developer Guide.

The KMS key policy must grant permission to the IAM role that you specify in your CreateModel request. For more information, see Using Key Policies in AWS KMS in the AWS Key Management Service Developer Guide.

Type: String
Length Constraints: Maximum length of 2048.
Pattern: .*
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

• AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for Ruby V3
TransformResources

Service: Amazon SageMaker Service

Describes the resources, including ML instance types and ML instance count, to use for transform job.

Contents

InstanceCount

The number of ML compute instances to use in the transform job. The default value is 1, and the maximum is 100. For distributed transform jobs, specify a value greater than 1.

Type: Integer

Valid Range: Minimum value of 1.

Required: Yes

InstanceType

The ML compute instance type for the transform job. If you are using built-in algorithms to transform moderately sized datasets, we recommend using ml.m4.xlarge or ml.m5.large instance types.

Type: String

Valid Values: ml.m4.xlarge | ml.m4.2xlarge | ml.m4.4xlarge | ml.m4.10xlarge
| ml.m4.16xlarge | ml.m5.xlarge | ml.m5.2xlarge | ml.m5.4xlarge | ml.m5.12xlarge |
| ml.m5.24xlarge | ml.g4dn.xlarge | ml.g4dn.2xlarge | ml.g4dn.4xlarge | ml.g4dn.16xlarge | ml.g4dn.12xlarge | ml.g4dn.8xlarge | ml.g4dn.16xlarge

Required: Yes

VolumeKmsKeyId

The AWS Key Management Service (AWS KMS) key that Amazon SageMaker uses to encrypt model data on the storage volume attached to the ML compute instance(s) that run the batch transform job.

Note

Certain Nitro-based instances include local storage, dependent on the instance type. Local storage volumes are encrypted using a hardware module on the instance. You can't request a VolumeKmsKeyId when using an instance type with local storage.

For a list of instance types that support local instance storage, see Instance Store Volumes.

For more information about local instance storage encryption, see SSD Instance Store Volumes.

The VolumeKmsKeyId can be any of the following formats:

- Key ID: 1234abcd-12ab-34cd-56ef-1234567890ab
- Key ARN: arn:aws:kms:us-west-2:111122223333:key/1234abcd-12ab-34cd-56ef-1234567890ab
- Alias name: alias/ExampleAlias

Type: String
Length Constraints: Maximum length of 2048.
Pattern: . *
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
TransformS3DataSource
Service: Amazon SageMaker Service

Describes the S3 data source.

Contents

S3DataType

If you choose S3Prefix, S3Uri identifies a key name prefix. Amazon SageMaker uses all objects with the specified key name prefix for batch transform.

If you choose ManifestFile, S3Uri identifies an object that is a manifest file containing a list of object keys that you want Amazon SageMaker to use for batch transform.

The following values are compatible: ManifestFile, S3Prefix

The following value is not compatible: AugmentedManifestFile

Type: String

Valid Values: ManifestFile | S3Prefix | AugmentedManifestFile

Required: Yes

S3Uri

Depending on the value specified for the S3DataType, identifies either a key name prefix or a manifest. For example:

- A key name prefix might look like this: s3://bucketname/exampleprefix.
- A manifest might look like this: s3://bucketname/example.manifest

The manifest is an S3 object which is a JSON file with the following format:

```
[ {"prefix": "s3://customer_bucket/some/prefix/"},
  "relative/path/to/custdata-1",
  "relative/path/custdata-2",
  ...
  "relative/path/custdata-N"
]
```

The preceding JSON matches the following S3Uris:

- s3://customer_bucket/some/prefix/relative/path/to/custdata-1
- s3://customer_bucket/some/prefix/relative/path/custdata-2
- ...
- s3://customer_bucket/some/prefix/relative/path/custdata-N

The complete set of S3Uris in this manifest constitutes the input data for the channel for this datasource. The object that each S3Uris points to must be readable by the IAM role that Amazon SageMaker uses to perform tasks on your behalf.
Type: String

Length Constraints: Maximum length of 1024.

Pattern: ^https:\/\/s3:////([^/]+)\/?.*$

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
Trial
Service: Amazon SageMaker Service

The properties of a trial as returned by the Search API.

Contents

CreatedBy
Who created the trial.
Type: UserContext (p. 2067) object
Required: No

CreationTime
When the trial was created.
Type: Timestamp
Required: No

DisplayName
The name of the trial as displayed. If DisplayName isn't specified, TrialName is displayed.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 120.
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,119\}
Required: No

ExperimentName
The name of the experiment the trial is part of.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 120.
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,119\}
Required: No

LastModifiedBy
Information about the user who created or modified an experiment, trial, trial component, lineage group, project, or model card.
Type: UserContext (p. 2067) object
Required: No

LastModifiedTime
Who last modified the trial.
Type: Timestamp
Required: No
MetadataProperties

Metadata properties of the tracking entity, trial, or trial component.

Type: MetadataProperties (p. 1648) object

Required: No

Source

The source of the trial.

Type: TrialSource (p. 2056) object

Required: No

Tags

The list of tags that are associated with the trial. You can use Search API to search on the tags.

Type: Array of Tag (p. 1979) objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

TrialArn

The Amazon Resource Name (ARN) of the trial.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:experiment-trial/.*

Required: No

TrialComponentSummaries

A list of the components associated with the trial. For each component, a summary of the component's properties is included.

Type: Array of TrialComponentSimpleSummary (p. 2049) objects

Required: No

TrialName

The name of the trial.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: ^[a-zA-Z0-9-]\([-]*[a-zA-Z0-9]\){0,119}

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for Ruby V3
**TrialComponent**

Service: Amazon SageMaker Service

The properties of a trial component as returned by the Search API.

## Contents

### CreatedBy

Who created the trial component.

Type: UserContext (p. 2067) object

Required: No

### CreationTime

When the component was created.

Type: Timestamp

Required: No

### DisplayName

The name of the component as displayed. If DisplayName isn't specified, TrialComponentName is displayed.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,119}$

Required: No

### EndTime

When the component ended.

Type: Timestamp

Required: No

### InputArtifacts

The input artifacts of the component.

Type: String to TrialComponentArtifact (p. 2045) object map

Map Entries: Maximum number of 30 items.

Key Length Constraints: Maximum length of 64.

Key Pattern: .*

Required: No

### LastModifiedBy

Information about the user who created or modified an experiment, trial, trial component, lineage group, project, or model card.

Type: UserContext (p. 2067) object
**Required:** No

**LastModifiedTime**

When the component was last modified.

*Type:* Timestamp

**LineageGroupArn**

The Amazon Resource Name (ARN) of the lineage group resource.

*Type:* String

*Length Constraints:* Maximum length of 256.

*Pattern:* `arn:aws[a-zA-Z\-]*:sagemaker:[a-zA-Z0-9\-]*:[0-9]{12}:lineage-group/.*`

**MetadataProperties**

Metadata properties of the tracking entity, trial, or trial component.

*Type:* [MetadataProperties](p. 1648) object

**Metrics**

The metrics for the component.

*Type:* Array of [TrialComponentMetricSummary](p. 2046) objects

**OutputArtifacts**

The output artifacts of the component.

*Type:* String to [TrialComponentArtifact](p. 2045) object map

*Map Entries:* Maximum number of 30 items.

*Key Length Constraints:* Maximum length of 64.

*Key Pattern:* .`

**Parameters**

The hyperparameters of the component.

*Type:* String to [TrialComponentParameterValue](p. 2048) object map

*Map Entries:* Maximum number of 150 items.

*Key Length Constraints:* Maximum length of 256.

*Key Pattern:* .`
**Parents**

An array of the parents of the component. A parent is a trial the component is associated with and the experiment the trial is part of. A component might not have any parents.

Type: Array of [Parent](p. 1802) objects

Required: No

**RunName**

The name of the experiment run.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,119}$`

Required: No

**Source**

The Amazon Resource Name (ARN) and job type of the source of the component.

Type: [TrialComponentSource](p. 2051) object

Required: No

**SourceDetail**

Details of the source of the component.

Type: [TrialComponentSourceDetail](p. 2052) object

Required: No

**StartTime**

When the component started.

Type: Timestamp

Required: No

**Status**

The status of the trial component.

Type: [TrialComponentStatus](p. 2053) object

Required: No

**Tags**

The list of tags that are associated with the component. You can use [Search](enzyme) API to search on the tags.

Type: Array of [Tag](p. 1979) objects

Array Members: Minimum number of 0 items. Maximum number of 50 items.

Required: No

**TrialComponentArn**

The Amazon Resource Name (ARN) of the trial component.
TrialComponent

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:experiment-trial-component/.*

Required: No

TrialComponentName

The name of the trial component.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,119}$

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
**TrialComponentArtifact**

Service: Amazon SageMaker Service

Represents an input or output artifact of a trial component. You specify `TrialComponentArtifact` as part of the `InputArtifacts` and `OutputArtifacts` parameters in the `CreateTrialComponent` request.

Examples of input artifacts are datasets, algorithms, hyperparameters, source code, and instance types. Examples of output artifacts are metrics, snapshots, logs, and images.

**Contents**

**Value**

The location of the artifact.

Type: String

Length Constraints: Maximum length of 2048.

Pattern: .*

Required: Yes

**MediaType**

The media type of the artifact, which indicates the type of data in the artifact file. The media type consists of a type and a subtype concatenated with a slash (/) character, for example, text/csv, image/jpeg, and s3/uri. The type specifies the category of the media. The subtype specifies the kind of data.

Type: String

Length Constraints: Maximum length of 64.

Pattern: ^[-\w]+\/[\-\w]+$

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
TrialComponentMetricSummary

Service: Amazon SageMaker Service

A summary of the metrics of a trial component.

**Contents**

**Avg**

The average value of the metric.

Type: Double

Required: No

**Count**

The number of samples used to generate the metric.

Type: Integer

Required: No

**Last**

The most recent value of the metric.

Type: Double

Required: No

**Max**

The maximum value of the metric.

Type: Double

Required: No

**MetricName**

The name of the metric.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 255.

Pattern: .+

Required: No

**Min**

The minimum value of the metric.

Type: Double

Required: No

**SourceArn**

The Amazon Resource Name (ARN) of the source.

Type: String
Length Constraints: Maximum length of 256.
Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}::*
Required: No

StdDev
The standard deviation of the metric.
Type: Double
Required: No

TimeStamp
When the metric was last updated.
Type: Timestamp
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
TrialComponentParameterValue

Service: Amazon SageMaker Service

The value of a hyperparameter. Only one of NumberValue or StringValue can be specified.

This object is specified in the CreateTrialComponent request.

Contents

NumberValue

The numeric value of a numeric hyperparameter. If you specify a value for this parameter, you can't specify the StringValue parameter.

Type: Double

Required: No

StringValue

The string value of a categorical hyperparameter. If you specify a value for this parameter, you can't specify the NumberValue parameter.

Type: String

Length Constraints: Maximum length of 256.

Pattern: .*

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
**TrialComponentSimpleSummary**

Service: Amazon SageMaker Service

A short summary of a trial component.

**Contents**

**CreatedBy**

Information about the user who created or modified an experiment, trial, trial component, lineage group, project, or model card.

Type: *UserContext* (p. 2067) object

Required: No

**CreationTime**

When the component was created.

Type: Timestamp

Required: No

**TrialComponentArn**

The Amazon Resource Name (ARN) of the trial component.

Type: String

Length Constraints: Maximum length of 256.

Pattern: `arn:aws[a-z\-]*:sagemaker:[a-zA-Z0-9\-]*:[0-9]{12}:experiment-trial-component/.*`

Required: No

**TrialComponentName**

The name of the trial component.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: `^[a-zA-Z0-9\-]*[^a-zA-Z0-9\-]{0,119}`

Required: No

**TrialComponentSource**

The Amazon Resource Name (ARN) and job type of the source of a trial component.

Type: *TrialComponentSource* (p. 2051) object

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
TrialComponentSource

Service: Amazon SageMaker Service

The Amazon Resource Name (ARN) and job type of the source of a trial component.

Contents

SourceArn

The source Amazon Resource Name (ARN).

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:.*

Required: Yes

SourceType

The source job type.

Type: String

Length Constraints: Maximum length of 128.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](https://aws.amazon.com/sdk-for-cpp/)
- [AWS SDK for Go](https://aws.amazon.com/sdk-for-go/)
- [AWS SDK for Java V2](https://aws.amazon.com/sdk-for-java/)
- [AWS SDK for Ruby V3](https://aws.amazon.com/sdk-for-ruby/)

2051
TrialComponentSourceDetail

Service: Amazon SageMaker Service

Detailed information about the source of a trial component. Either ProcessingJob or TrainingJob is returned.

Contents

ProcessingJob

Information about a processing job that's the source of a trial component.

Type: ProcessingJob (p. 1833) object

Required: No

SourceArn

The Amazon Resource Name (ARN) of the source.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-\]*:sagemaker:[a-z0-9\-\]*:[0-9]{12}::*

Required: No

TrainingJob

Information about a training job that's the source of a trial component.

Type: TrainingJob (p. 2001) object

Required: No

TransformJob

Information about a transform job that's the source of a trial component.

Type: TransformJob (p. 2022) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
TrialComponentStatus
Service: Amazon SageMaker Service

The status of the trial component.

Contents

Message
If the component failed, a message describing why.
Type: String
Length Constraints: Maximum length of 1024.
Pattern: .*
Required: No

PrimaryStatus
The status of the trial component.
Type: String
Valid Values: InProgress | Completed | Failed | Stopping | Stopped
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
**TrialComponentSummary**

Service: Amazon SageMaker Service

A summary of the properties of a trial component. To get all the properties, call the
[DescribeTrialComponent](https://docs.aws.amazon.com/sagemaker/latest/dg/API_DescribeTrialComponent.html) API and provide the `TrialComponentName`.

**Contents**

**CreatedBy**

Who created the trial component.

Type: [UserContext](https://docs.aws.amazon.com/sagemaker/latest/dg/API_UserContext.html) object

Required: No

**CreationTime**

When the component was created.

Type: Timestamp

Required: No

**DisplayName**

The name of the component as displayed. If `DisplayName` isn't specified, `TrialComponentName` is displayed.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: `^[a-zA-Z0-9](-*[a-zA-Z0-9]){0,119}$`

Required: No

**EndTime**

When the component ended.

Type: Timestamp

Required: No

**LastModifiedBy**

Who last modified the component.

Type: [UserContext](https://docs.aws.amazon.com/sagemaker/latest/dg/API_UserContext.html) object

Required: No

**LastModifiedTime**

When the component was last modified.

Type: Timestamp

Required: No

**StartTime**

When the component started.
Type: Timestamp
Required: No

**Status**

The status of the component. States include:
- InProgress
- Completed
- Failed

Type: [TrialComponentStatus](p. 2053) object
Required: No

**TrialComponentArn**

The Amazon Resource Name (ARN) of the trial component.

Type: String

Length Constraints: Maximum length of 256.

Pattern: `arn:aws[a-z\-]*:sagemaker:[a-zA-Z0-9\-]*:[0-9]{12}:experiment-trial-component/.*`
Required: No

**TrialComponentName**

The name of the trial component.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 120.

Pattern: `^[a-zA-Z0-9](\*.?-[a-zA-Z0-9]{0,119}){0,119}`
Required: No

**TrialComponentSource**

The Amazon Resource Name (ARN) and job type of the source of a trial component.

Type: [TrialComponentSource](p. 2051) object
Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
TrialSource
Service: Amazon SageMaker Service

The source of the trial.

Contents

SourceArn

The Amazon Resource Name (ARN) of the source.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[\-a-z0-9\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}::*

Required: Yes

SourceType

The source job type.

Type: String

Length Constraints: Maximum length of 128.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
TrialSummary
Service: Amazon SageMaker Service

A summary of the properties of a trial. To get the complete set of properties, call the DescribeTrial API and provide the TrialName.

Contents

CreationTime
When the trial was created.
Type: Timestamp
Required: No

DisplayName
The name of the trial as displayed. If DisplayName isn't specified, TrialName is displayed.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 120.
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,119\}
Required: No

LastModifiedTime
When the trial was last modified.
Type: Timestamp
Required: No

TrialArn
The Amazon Resource Name (ARN) of the trial.
Type: String
Length Constraints: Maximum length of 256.
Pattern: arn:aws[a-zA-Z\-]*:sagemaker:[a-zA-Z0-9\-]*:[0-9]{12}:experiment-trial/.*
Required: No

TrialName
The name of the trial.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 120.
Pattern: ^[a-zA-Z0-9]$-*[a-zA-Z0-9]\}\{0,119\}
Required: No

TrialSource
The source of the trial.
Type: TrialSource (p. 2056) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
TtlDuration
Service: Amazon SageMaker Service

Time to live duration, where the record is hard deleted after the expiration time is reached; ExpiresAt = EventTime + TtlDuration. For information on HardDelete, see the DeleteRecord API in the Amazon SageMaker API Reference guide.

Contents

Unit

TtlDuration time unit.
Type: String
Valid Values: Seconds | Minutes | Hours | Days | Weeks
Required: No

Value

TtlDuration time value.
Type: Integer
Valid Range: Minimum value of 1.
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
TuningJobCompletionCriteria
Service: Amazon SageMaker Service

The job completion criteria.

Contents

BestObjectiveNotImproving
A flag to stop your hyperparameter tuning job if model performance fails to improve as evaluated against an objective function.

Type: BestObjectiveNotImproving (p. 1308) object

Required: No

ConvergenceDetected
A flag to top your hyperparameter tuning job if automatic model tuning (AMT) has detected that your model has converged as evaluated against your objective function.

Type: ConvergenceDetected (p. 1375) object

Required: No

TargetObjectiveMetricValue
The value of the objective metric.

Type: Float

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
TuningJobStepMetaData

Service: Amazon SageMaker Service

Metadata for a tuning step.

Contents

Arn

The Amazon Resource Name (ARN) of the tuning job that was run by this step execution.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:hyper-parameter-tuning-job/.*

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
UiConfig

Service: Amazon SageMaker Service

Provided configuration information for the worker UI for a labeling job. Provide either HumanTaskUiArn or UiTemplateS3Uri.

For named entity recognition, 3D point cloud and video frame labeling jobs, use HumanTaskUiArn.

For all other Ground Truth built-in task types and custom task types, use UiTemplateS3Uri to specify the location of a worker task template in Amazon S3.

Contents

HumanTaskUiArn

The ARN of the worker task template used to render the worker UI and tools for labeling job tasks.

Use this parameter when you are creating a labeling job for named entity recognition, 3D point cloud and video frame labeling jobs. Use your labeling job task type to select one of the following ARNs and use it with this parameter when you create a labeling job. Replace aws-region with the AWS Region you are creating your labeling job in. For example, replace aws-region with us-west-1 if you create a labeling job in US West (N. California).

Named Entity Recognition

Use the following HumanTaskUiArn for named entity recognition labeling jobs:


3D Point Cloud HumanTaskUiArns

Use this HumanTaskUiArn for 3D point cloud object detection and 3D point cloud object detection adjustment labeling jobs.


Use this HumanTaskUiArn for 3D point cloud object tracking and 3D point cloud object tracking adjustment labeling jobs.

- arn:aws:sagemaker:aws-region:394669845002:human-task-ui/PointCloudObjectTracking

Use this HumanTaskUiArn for 3D point cloud semantic segmentation and 3D point cloud semantic segmentation adjustment labeling jobs.


Video Frame HumanTaskUiArns

Use this HumanTaskUiArn for video frame object detection and video frame object detection adjustment labeling jobs.


Use this HumanTaskUiArn for video frame object tracking and video frame object tracking adjustment labeling jobs.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:human-task-ui/.*

Required: No

UiTemplateS3Uri

The Amazon S3 bucket location of the UI template, or worker task template. This is the template used to render the worker UI and tools for labeling job tasks. For more information about the contents of a UI template, see Creating Your Custom Labeling Task Template.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: ^(https|s3)://([^/]+)/(.*$)

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
UiTemplate

Service: Amazon SageMaker Service

The Liquid template for the worker user interface.

Contents

Content

The content of the Liquid template for the worker user interface.

Type: String


Pattern: [\S\s]+

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
UiTemplateInfo
Service: Amazon SageMaker Service
Container for user interface template information.

Contents

ContentSha256
  The SHA-256 digest of the contents of the template.
  Type: String
  Required: No

Url
  The URL for the user interface template.
  Type: String
  Length Constraints: Minimum length of 1. Maximum length of 2048.
  Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
USD
Service: Amazon SageMaker Service
Represents an amount of money in United States dollars.

Contents

Cents
The fractional portion, in cents, of the amount.
Type: Integer
Valid Range: Minimum value of 0. Maximum value of 99.
Required: No

Dollars
The whole number of dollars in the amount.
Type: Integer
Valid Range: Minimum value of 0. Maximum value of 2.
Required: No

TenthFractionsOfACent
Fractions of a cent, in tenths.
Type: Integer
Valid Range: Minimum value of 0. Maximum value of 9.
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
UserContext

Service: Amazon SageMaker Service

Information about the user who created or modified an experiment, trial, trial component, lineage group, project, or model card.

Contents

DomainId
- The domain associated with the user.
- Type: String
- Required: No

IamIdentity
- The IAM Identity details associated with the user. These details are associated with model package groups, model packages, and project entities only.
- Type: IamIdentity (p. 1569) object
- Required: No

UserProfileArn
- The Amazon Resource Name (ARN) of the user's profile.
- Type: String
- Required: No

UserProfileName
- The name of the user's profile.
- Type: String
- Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
UserProfileDetails
Service: Amazon SageMaker Service
The user profile details.

Contents

CreationTime
The creation time.
Type: Timestamp
Required: No

DomainId
The domain ID.
Type: String
  Length Constraints: Maximum length of 63.
  Required: No

LastModifiedTime
The last modified time.
Type: Timestamp
Required: No

Status
The status.
Type: String
  Valid Values: Deleting | Failed | InService | Pending | Updating | Update_Failed | Delete_Failed
  Required: No

UserProfileName
The user profile name.
Type: String
  Length Constraints: Maximum length of 63.
  Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])\{0,62}$
  Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:
  • AWS SDK for C++
• AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for Ruby V3
UserSettings

Service: Amazon SageMaker Service

A collection of settings that apply to users in a domain. These settings are specified when the CreateUserProfile API is called, and as DefaultUserSettings when the CreateDomain API is called.

SecurityGroups is aggregated when specified in both calls. For all other settings in UserSettings, the values specified in CreateUserProfile take precedence over those specified in CreateDomain.

Contents

CanvasAppSettings

The Canvas app settings.

Type: CanvasAppSettings (p. 1316) object

Required: No

CodeEditorAppSettings

The Code Editor application settings.

Type: CodeEditorAppSettings (p. 1353) object

Required: No

CustomFileSystemConfigs

The settings for assigning a custom file system to a user profile. Permitted users can access this file system in Amazon SageMaker Studio.

Type: Array of CustomFileSystemConfig (p. 1377) objects

Array Members: Maximum number of 2 items.

Required: No

CustomPosixUserConfig

Details about the POSIX identity that is used for file system operations.

Type: CustomPosixUserConfig (p. 1380) object

Required: No

DefaultLandingUri

The default experience that the user is directed to when accessing the domain. The supported values are:

- `studio::`: Indicates that Studio is the default experience. This value can only be passed if StudioWebPortal is set to ENABLED.
- `app:JupyterServer::`: Indicates that Studio Classic is the default experience.

Type: String

Length Constraints: Maximum length of 1023.

Required: No

ExecutionRole

The execution role for the user.
UserSettings

Type: String


Pattern: ^arn:aws[a-z\-]*:iam::\d{12}:role/?[a-zA-Z_0-9+=,.@\-_]+$  

Required: No

JupyterLabAppSettings

The settings for the JupyterLab application.

Type: JupyterLabAppSettings (p. 1616) object

Required: No

JupyterServerAppSettings

The Jupyter server's app settings.

Type: JupyterServerAppSettings (p. 1618) object

Required: No

KernelGatewayAppSettings

The kernel gateway app settings.

Type: KernelGatewayAppSettings (p. 1620) object

Required: No

RSessionAppSettings

A collection of settings that configure the RSessionGateway app.

Type: RSessionAppSettings (p. 1920) object

Required: No

RStudioServerProAppSettings

A collection of settings that configure user interaction with the RStudioServerPro app.

Type: RStudioServerProAppSettings (p. 1921) object

Required: No

SecurityGroups

The security groups for the Amazon Virtual Private Cloud (VPC) that the domain uses for communication.

Optional when the CreateDomain.AppNetworkAccessType parameter is set to PublicInternetOnly.

Required when the CreateDomain.AppNetworkAccessType parameter is set to VpcOnly, unless specified as part of the DefaultUserSettings for the domain.

Amazon SageMaker adds a security group to allow NFS traffic from Amazon SageMaker Studio. Therefore, the number of security groups that you can specify is one less than the maximum number shown.

Type: Array of strings

Array Members: Maximum number of 5 items.
UserSettings

Length Constraints: Maximum length of 32.
Pattern: [-0-9a-zA-Z]+
Required: No

SharingSettings

Specifies options for sharing Amazon SageMaker Studio notebooks.
Type: SharingSettings (p. 1952) object
Required: No

SpaceStorageSettings

The storage settings for a private space.
Type: DefaultSpaceStorageSettings (p. 1403) object
Required: No

StudioWebPortal

Whether the user can access Studio. If this value is set to DISABLED, the user cannot access Studio, even if that is the default experience for the domain.
Type: String
Valid Values: ENABLED | DISABLED
Required: No

TensorBoardAppSettings

The TensorBoard app settings.
Type: TensorBoardAppSettings (p. 1983) object
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
**VariantProperty**

Service: Amazon SageMaker Service

Specifies a production variant property type for an Endpoint.

If you are updating an endpoint with the `RetainAllVariantProperties` option of `UpdateEndpointInput` set to `true`, the `VariantProperty` objects listed in the `ExcludeRetainedVariantProperties` parameter of `UpdateEndpointInput` override the existing variant properties of the endpoint.

**Contents**

**VariantPropertyType**

The type of variant property. The supported values are:

- DesiredInstanceCount: Overrides the existing variant instance counts using the `InitialInstanceCount` values in the `ProductionVariants` of `CreateEndpointConfig`.
- DesiredWeight: Overrides the existing variant weights using the `InitialVariantWeight` values in the `ProductionVariants` of `CreateEndpointConfig`.
- DataCaptureConfig: (Not currently supported.)

Type: String

Valid Values: DesiredInstanceCount | DesiredWeight | DataCaptureConfig

Required: Yes

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
VectorConfig
Service: Amazon SageMaker Service
Configuration for your vector collection type.

Contents

Dimension
The number of elements in your vector.
Type: Integer
Required: Yes

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
Vertex
Service: Amazon SageMaker Service
A lineage entity connected to the starting entity(ies).

Contents

Arn
The Amazon Resource Name (ARN) of the lineage entity resource.
Type: String
Length Constraints: Maximum length of 256.
Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}):(experiment|experiment-trial-component|artifact|action|context)/.*
Required: No

LineageType
The type of resource of the lineage entity.
Type: String
Valid Values: TrialComponent | Artifact | Context | Action
Required: No

Type
The type of the lineage entity resource. For example: DataSet, Model, Endpoint, etc...
Type: String
Length Constraints: Maximum length of 40.
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
VpcConfig

Service: Amazon SageMaker Service

Specifies an Amazon Virtual Private Cloud (VPC) that your SageMaker jobs, hosted models, and compute resources have access to. You can control access to and from your resources by configuring a VPC. For more information, see Give SageMaker Access to Resources in your Amazon VPC.

Contents

SecurityGroupIds

The VPC security group IDs, in the form sg-xxxxxxxx. Specify the security groups for the VPC that is specified in the Subnets field.

Type: Array of strings

Array Members: Minimum number of 1 item. Maximum number of 5 items.

Length Constraints: Maximum length of 32.

Pattern: [-0-9a-zA-Z]+

Required: Yes

Subnets

The ID of the subnets in the VPC to which you want to connect your training job or model. For information about the availability of specific instance types, see Supported Instance Types and Availability Zones.

Type: Array of strings

Array Members: Minimum number of 1 item. Maximum number of 16 items.

Length Constraints: Maximum length of 32.

Pattern: [-0-9a-zA-Z]+

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](https://aws.amazon.com/sdk-for-cpp/)
- [AWS SDK for Go](https://github.com/aws/aws-sdk-go)
- [AWS SDK for Java V2](https://docs.aws.amazon.com/sdk-for-java/guide/)
- [AWS SDK for Ruby V3](https://aws.amazon.com/sdk-for-ruby/)

2076
WarmPoolStatus
Service: Amazon SageMaker Service

Status and billing information about the warm pool.

Contents

**Status**

The status of the warm pool.

- **InUse**: The warm pool is in use for the training job.
- **Available**: The warm pool is available to reuse for a matching training job.
- **Reused**: The warm pool moved to a matching training job for reuse.
- **Terminated**: The warm pool is no longer available. Warm pools are unavailable if they are terminated by a user, terminated for a patch update, or terminated for exceeding the specified KeepAlivePeriodInSeconds.

Type: String

Valid Values: Available | Terminated | Reused | InUse

Required: Yes

**ResourceRetainedBillableTimeInSeconds**

The billable time in seconds used by the warm pool. Billable time refers to the absolute wall-clock time.

Multiply **ResourceRetainedBillableTimeInSeconds** by the number of instances (**InstanceCount**) in your training cluster to get the total compute time SageMaker bills you if you run warm pool training. The formula is as follows: `ResourceRetainedBillableTimeInSeconds * InstanceCount`.

Type: Integer

Valid Range: Minimum value of 0.

Required: No

**ReusedByJob**

The name of the matching training job that reused the warm pool.

Type: String


Pattern: `^[a-zA-Z0-9-]*[a-zA-Z0-9]{0,62}`

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2

2077
• AWS SDK for Ruby V3
Workforce
Service: Amazon SageMaker Service

A single private workforce, which is automatically created when you create your first private work team. You can create one private work force in each AWS Region. By default, any workforce-related API operation used in a specific region will apply to the workforce created in that region. To learn how to create a private workforce, see Create a Private Workforce.

Contents

WorkforceArn
The Amazon Resource Name (ARN) of the private workforce.
Type: String
Length Constraints: Maximum length of 256.
Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:workforce/.*
Required: Yes

WorkforceName
The name of the private workforce.
Type: String
Pattern: ^[a-zA-Z0-9\-\((a-zA-Z0-9\-\)]\{0,62}$
Required: Yes

CognitoConfig
The configuration of an Amazon Cognito workforce. A single Cognito workforce is created using and corresponds to a single Amazon Cognito user pool.
Type: CognitoConfig (p. 1357) object
Required: No

CreateDate
The date that the workforce is created.
Type: Timestamp
Required: No

FailureReason
The reason your workforce failed.
Type: String
Pattern: .+
Required: No
LastUpdatedDate

The most recent date that UpdateWorkforce was used to successfully add one or more IP address ranges (CIDRs) to a private workforce's allow list.

Type: Timestamp
Required: No

OidcConfig

The configuration of an OIDC Identity Provider (IdP) private workforce.

Type: OidcConfigForResponse object
Required: No

SourceIpConfig

A list of one to ten IP address ranges (CIDRs) to be added to the workforce allow list. By default, a workforce isn't restricted to specific IP addresses.

Type: SourceIpConfig object
Required: No

Status

The status of your workforce.

Type: String
Valid Values: Initializing | Updating | Deleting | Failed | Active
Required: No

SubDomain

The subdomain for your OIDC Identity Provider.

Type: String
Required: No

WorkforceVpcConfig

The configuration of a VPC workforce.

Type: WorkforceVpcConfigResponse object
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
WorkforceVpcConfigRequest
Service: Amazon SageMaker Service
The VPC object you use to create or update a workforce.

Contents

SecurityGroupIds
The VPC security group IDs, in the form sg-xxxxxxxx. The security groups must be for the same VPC as specified in the subnet.
Type: Array of strings
Array Members: Minimum number of 1 item. Maximum number of 5 items.
Length Constraints: Maximum length of 32.
Pattern: ^sg-[0-9a-z]+$
Required: No

Subnets
The ID of the subnets in the VPC that you want to connect.
Type: Array of strings
Array Members: Minimum number of 1 item. Maximum number of 16 items.
Length Constraints: Maximum length of 32.
Pattern: ^subnet-[0-9a-z]+$
Required: No

VpcId
The ID of the VPC that the workforce uses for communication.
Type: String
Length Constraints: Maximum length of 32.
Pattern: ^vpc-[0-9a-z]+$
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
WorkforceVpcConfigResponse
Service: Amazon SageMaker Service
A VpcConfig object that specifies the VPC that you want your workforce to connect to.

Contents

SecurityGroupIds
The VPC security group IDs, in the form sg-xxxxxxxx. The security groups must be for the same VPC as specified in the subnet.
Type: Array of strings
Array Members: Minimum number of 1 item. Maximum number of 5 items.
Length Constraints: Maximum length of 32.
Pattern: ^sg-[0-9a-z]+$
Required: Yes

Subnets
The ID of the subnets in the VPC that you want to connect.
Type: Array of strings
Array Members: Minimum number of 1 item. Maximum number of 16 items.
Length Constraints: Maximum length of 32.
Pattern: ^subnet-[0-9a-z]+$
Required: Yes

VpcId
The ID of the VPC that the workforce uses for communication.
Type: String
Length Constraints: Maximum length of 32.
Pattern: ^vpc-[0-9a-z]+$
Required: Yes

VpcEndpointId
The IDs for the VPC service endpoints of your VPC workforce when it is created and updated.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 255.
Pattern: ^vpce-[0-9a-z]+$
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
WorkspaceSettings
Service: Amazon SageMaker Service
The workspace settings for the SageMaker Canvas application.

Contents

S3ArtifactPath

The Amazon S3 bucket used to store artifacts generated by Canvas. Updating the Amazon S3 location impacts existing configuration settings, and Canvas users no longer have access to their artifacts. Canvas users must log out and log back in to apply the new location.

Type: String

Length Constraints: Maximum length of 1024.

Pattern: ^(https|s3)://([^/]+)/?(.*)$

Required: No

S3KmsKeyId

The AWS Key Management Service (KMS) encryption key ID that is used to encrypt artifacts generated by Canvas in the Amazon S3 bucket.

Type: String

Length Constraints: Maximum length of 2048.

Pattern: .*

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
Workteam
Service: Amazon SageMaker Service

Provides details about a labeling work team.

Contents

**Description**

A description of the work team.

Type: String


Pattern: .+

Required: Yes

**MemberDefinitions**

A list of MemberDefinition objects that contains objects that identify the workers that make up the work team.

Workforces can be created using Amazon Cognito or your own OIDC Identity Provider (IdP). For private workforces created using Amazon Cognito use CognitoMemberDefinition. For workforces created using your own OIDC identity provider (IdP) use OidcMemberDefinition.

Type: Array of MemberDefinition (p. 1647) objects

Array Members: Minimum number of 1 item. Maximum number of 10 items.

Required: Yes

**WorkteamArn**

The Amazon Resource Name (ARN) that identifies the work team.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-z\-]*:sagemaker:[a-z0-9\-]*:[0-9]{12}:workteam/.*

Required: Yes

**WorkteamName**

The name of the work team.

Type: String


Pattern: ^[a-zA-Z0-9](\-*[a-zA-Z0-9])\{0,62}\$

Required: Yes

**CreateDate**

The date and time that the work team was created (timestamp).

Type: Timestamp
Required: No

**LastUpdatedDate**

The date and time that the work team was last updated (timestamp).

Type: Timestamp

Required: No

**NotificationConfiguration**

Configures SNS notifications of available or expiring work items for work teams.

Type: NotificationConfiguration (p. 1775) object

Required: No

**ProductListingIds**

The Amazon Marketplace identifier for a vendor’s work team.

Type: Array of strings

Required: No

**SubDomain**

The URI of the labeling job’s user interface. Workers open this URI to start labeling your data objects.

Type: String

Required: No

**WorkforceArn**

The Amazon Resource Name (ARN) of the workforce.

Type: String

Length Constraints: Maximum length of 256.

Pattern: arn:aws[a-zA-Z\-]*:sagemaker:[a-zA-Z0-9\-]*:[0-9]{12}:workforce/.*

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3

Amazon SageMaker Runtime

The following data types are supported by Amazon SageMaker Runtime:

- PayloadPart (p. 2088)
• ResponseStream (p. 2089)
PayloadPart

Service: Amazon SageMaker Runtime

A wrapper for pieces of the payload that's returned in response to a streaming inference request. A streaming inference response consists of one or more payload parts.

Contents

Bytes

A blob that contains part of the response for your streaming inference request.

Type: Base64-encoded binary data object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ResponseStream

Service: Amazon SageMaker Runtime

A stream of payload parts. Each part contains a portion of the response for a streaming inference request.

Contents

InternalStreamFailure

The stream processing failed because of an unknown error, exception or failure. Try your request again.

Type: Exception
HTTP Status Code:
Required: No

ModelStreamError

An error occurred while streaming the response body. This error can have the following error codes:

ModelInvocationTimeExceeded

The model failed to finish sending the response within the timeout period allowed by Amazon SageMaker.

StreamBroken

The Transmission Control Protocol (TCP) connection between the client and the model was reset or closed.

Type: Exception
HTTP Status Code:
Required: No

PayloadPart

A wrapper for pieces of the payload that's returned in response to a streaming inference request. A streaming inference response consists of one or more payload parts.

Type: PayloadPart (p. 2088) object
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3

Amazon Sagemaker Edge Manager

The following data types are supported by Amazon Sagemaker Edge Manager:
- Checksum (p. 2091)
- Definition (p. 2092)
- DeploymentModel (p. 2093)
- DeploymentResult (p. 2095)
- EdgeDeployment (p. 2097)
- EdgeMetric (p. 2098)
- Model (p. 2099)
Checksum
Service: Amazon Sagemaker Edge Manager

Information about the checksum of a model deployed on a device.

Contents

Sum

The checksum of the model.
Type: String
Pattern: ^[a-z0-9](-*[a-z0-9])*$
Required: No

Type

The type of the checksum.
Type: String
Valid Values: SHA1
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
Definition
Service: Amazon Sagemaker Edge Manager
Information about deployment details based on the deployment type.

Contents

Checksum
The checksum information of the model.
Type: Checksum (p. 2091) object
Required: No

ModelHandle
The unique model handle.
Type: String
Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])*$
Required: No

S3Url
The absolute S3 location of the model.
Type: String
Length Constraints: Maximum length of 1024.
Pattern: ^s3://([^/]+)/?([^/]+)$
Required: No

State
The desired state of the model.
Type: String
Valid Values: DEPLOY | UNDEPLOY
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3

2092
DeploymentModel

Service: Amazon Sagemaker Edge Manager

Information about a model deployed on an edge device that is registered with SageMaker Edge Manager.

Contents

**DesiredState**

The desired state of the model.

- **Type:** String
- **Valid Values:** DEPLOY | UNDEPLOY
- **Required:** No

**ModelHandle**

The name by which the device refers to the model.

- **Type:** String
- **Length Constraints:** Minimum length of 1. Maximum length of 63.
- **Pattern:** ^[a-zA-Z0-9](-*[a-zA-Z0-9])*$
- **Required:** No

**ModelName**

The name of the model.

- **Type:** String
- **Length Constraints:** Minimum length of 4. Maximum length of 255.
- **Pattern:** ^[a-zA-Z0-9](-*[a-zA-Z0-9])*$
- **Required:** No

**ModelVersion**

The version of the model.

- **Type:** String
- **Length Constraints:** Minimum length of 1. Maximum length of 64.
- **Pattern:** [a-zA-Z0-9\ _\ .]+
- **Required:** No

**RollbackFailureReason**

Returns the error message if there is a rollback.

- **Type:** String
- **Required:** No

**State**

Returns the current state of the model.
DeploymentModel

Type: String
Valid Values: DEPLOY | UNDEPLOY
Required: No

**Status**
Returns the deployment status of the model.
Type: String
Valid Values: SUCCESS | FAIL
Required: No

**StatusReason**
Returns the error message for the deployment status result.
Type: String
Required: No

**See Also**
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
DeploymentResult

Service: Amazon Sagemaker Edge Manager

Information about the result of a deployment on an edge device that is registered with SageMaker Edge Manager.

Contents

DeploymentEndTime

The timestamp of when the deployment was ended, and the agent got the deployment results.

Type: Timestamp

Required: No

DeploymentModels

Returns a list of models deployed on the agent.

Type: Array of DeploymentModel objects

Required: No

DeploymentName

The name and unique ID of the deployment.

Type: String


Pattern: ^[a-zA-Z0-9][-]*[a-zA-Z0-9]*$

Required: No

DeploymentStartTime

The timestamp of when the deployment was started on the agent.

Type: Timestamp

Required: No

DeploymentStatus

Returns the bucket error code.

Type: String


Pattern: ^[a-zA-Z0-9][-]*[a-zA-Z0-9]*$

Required: No

DeploymentStatusMessage

Returns the detailed error message.

Type: String

Required: No
See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
EdgeDeployment
Service: Amazon Sagemaker Edge Manager

Information about a deployment on an edge device that is registered with SageMaker Edge Manager.

Contents

Definitions

Returns a list of Definition objects.

Type: Array of Definition objects

Required: No

DeploymentName

The name and unique ID of the deployment.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9])*$

Required: No

FailureHandlingPolicy

Determines whether to rollback to previous configuration if deployment fails.

Type: String

Valid Values: ROLLBACK_ON_FAILURE | DO NOTHING

Required: No

Type

The type of the deployment.

Type: String

Valid Values: Model

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
EdgeMetric
Service: Amazon Sagemaker Edge Manager

Information required for edge device metrics.

Contents

Dimension

The dimension of metrics published.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 1000.

Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9\-])*$

Required: No

MetricName

Returns the name of the metric.

Type: String


Pattern: ^[a-zA-Z0-9](-*[a-zA-Z0-9-])*$

Required: No

Timestamp

Timestamp of when the metric was requested.

Type: Timestamp

Required: No

Value

Returns the value of the metric.

Type: Double

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
Model
Service: Amazon Sagemaker Edge Manager

Information about a model deployed on an edge device that is registered with SageMaker Edge Manager.

Contents

LatestInference
The timestamp of the last inference that was made.
Type: Timestamp
Required: No

LatestSampleTime
The timestamp of the last data sample taken.
Type: Timestamp
Required: No

ModelMetrics
Information required for model metrics.
Type: Array of EdgeMetric (p. 2098) objects
Required: No

ModelName
The name of the model.
Type: String
Pattern: ^[a-zA-Z0-9\-_\(\)*]([-\_\(\)*]*[a-zA-Z0-9\-_\(\)*])*$
Required: No

ModelVersion
The version of the model.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 64.
Pattern: [a-zA-Z0-9\-_\(\)\[\]\*\] \(\)\[\]\*\]+
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
Amazon SageMaker Feature Store Runtime

The following data types are supported by Amazon SageMaker Feature Store Runtime:

- BatchGetRecordError (p. 2101)
- BatchGetRecordIdentifier (p. 2103)
- BatchGetRecordResultDetail (p. 2105)
- FeatureValue (p. 2106)
- TtlDuration (p. 2107)
BatchGetRecordError
Service: Amazon SageMaker Feature Store Runtime

The error that has occurred when attempting to retrieve a batch of Records.

Contents

**ErrorCode**

The error code of an error that has occurred when attempting to retrieve a batch of Records. For more information on errors, see Errors.

Type: String

Length Constraints: Maximum length of 358400.

Pattern: .*

Required: Yes

**ErrorMessage**

The error message of an error that has occurred when attempting to retrieve a record in the batch.

Type: String

Length Constraints: Maximum length of 2048.

Required: Yes

**FeatureGroupName**

The name of the feature group that the record belongs to.

Type: String

Length Constraints: Maximum length of 358400.

Pattern: .*

Required: Yes

**RecordIdentifierValueAsString**

The value for the RecordIdentifier in string format of a Record from a FeatureGroup that is causing an error when attempting to be retrieved.

Type: String

Length Constraints: Maximum length of 358400.

Pattern: .*

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
• AWS SDK for Java V2
• AWS SDK for Ruby V3
**BatchGetRecordIdentifier**

Service: Amazon SageMaker Feature Store Runtime

The identifier that identifies the batch of Records you are retrieving in a batch.

**Contents**

**FeatureGroupName**

The name or Amazon Resource Name (ARN) of the FeatureGroup containing the records you are retrieving in a batch.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 150.

Pattern: `(arn:aws[\a-zA-Z0-9\-]+:sagemaker:[\a-zA-Z0-9\-]+:[0-9]{12}:feature-group/)?(\[\a-zA-Z0-9\-]+\[\-\_]*\[\a-zA-Z0-9\-]+\]{0,63})`

Required: Yes

**RecordIdentifiersValueAsString**

The value for a list of record identifiers in string format.

Type: Array of strings

Array Members: Minimum number of 1 item. Maximum number of 100 items.

Length Constraints: Maximum length of 358400.

Pattern: `.*`

Required: Yes

**FeatureNames**

List of names of Features to be retrieved. If not specified, the latest value for all the Features are returned.

Type: Array of strings

Array Members: Minimum number of 1 item.

Length Constraints: Minimum length of 1. Maximum length of 64.

Pattern: `^[\a-zA-Z0-9\-]+\[\-\_]*\[\a-zA-Z0-9\-]+\]{0,63}`

Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- **AWS SDK for C++**
- **AWS SDK for Go**
- **AWS SDK for Java V2**
- **AWS SDK for Ruby V3**
BatchGetRecordResultDetail
Service: Amazon SageMaker Feature Store Runtime

The output of records that have been retrieved in a batch.

Contents

**FeatureGroupName**

The FeatureGroupName containing Records you retrieved in a batch.

Type: String

Length Constraints: Maximum length of 358400.

Pattern: .*

Required: Yes

**Record**

The Record retrieved.

Type: Array of FeatureValue (p. 2106) objects

Array Members: Minimum number of 1 item.

Required: Yes

**RecordIdentifierValueAsString**

The value of the record identifier in string format.

Type: String

Length Constraints: Maximum length of 358400.

Pattern: .*

Required: Yes

**ExpiresAt**

The ExpiresAt ISO string of the requested record.

Type: String

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
FeatureValue
Service: Amazon SageMaker Feature Store Runtime

The value associated with a feature.

Contents

FeatureName

The name of a feature that a feature value corresponds to.

Type: String
Length Constraints: Minimum length of 1. Maximum length of 64.
Pattern: ^[a-zA-Z0-9]([-_]*[a-zA-Z0-9])\{0,63
Required: Yes

ValueAsString

The value in string format associated with a feature. Used when your CollectionType is None. Note that features types can be String, Integral, or Fractional. This value represents all three types as a string.

Type: String
Length Constraints: Maximum length of 358400.
Pattern: .*
Required: No

ValueAsStringList

The list of values in string format associated with a feature. Used when your CollectionType is a List, Set, or Vector. Note that features types can be String, Integral, or Fractional. These values represents all three types as a string.

Type: Array of strings
Array Members: Minimum number of 0 items. Maximum number of 358400 items.
Length Constraints: Maximum length of 358400.
Pattern: .*
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
**TtlDuration**

Service: Amazon SageMaker Feature Store Runtime

Time to live duration, where the record is hard deleted after the expiration time is reached; ExpiresAt = EventTime + TtlDuration. For information on HardDelete, see the `DeleteRecord` API in the Amazon SageMaker API Reference guide.

**Contents**

**Unit**

TtlDuration time unit.

Type: String

Valid Values: Seconds | Minutes | Hours | Days | Weeks

Required: Yes

**Value**

TtlDuration time value.

Type: Integer

Valid Range: Minimum value of 1.

Required: Yes

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3

**Amazon SageMaker geospatial capabilities**

The following data types are supported by Amazon SageMaker geospatial capabilities:

- AreaOfInterest (p. 2110)
- AreaOfInterestGeometry (p. 2111)
- AssetValue (p. 2112)
- BandMathConfigInput (p. 2113)
- CloudMaskingConfigInput (p. 2114)
- CloudRemovalConfigInput (p. 2115)
- CustomIndicesInput (p. 2116)
- EarthObservationJobErrorDetails (p. 2117)
- EoCloudCoverInput (p. 2118)
- ExportErrorDetails (p. 2119)
• ExportErrorDetailsOutput (p. 2120)
• ExportS3DataInput (p. 2121)
• ExportVectorEnrichmentJobOutputConfig (p. 2122)
• Filter (p. 2123)
• Geometry (p. 2124)
• GeoMosaicConfigInput (p. 2125)
• InputConfigInput (p. 2126)
• InputConfigOutput (p. 2127)
• ItemSource (p. 2128)
• JobConfigInput (p. 2129)
• LandCoverSegmentationConfigInput (p. 2131)
• LandsatCloudCoverLandInput (p. 2132)
• ListEarthObservationJobOutputConfig (p. 2133)
• ListVectorEnrichmentJobOutputConfig (p. 2135)
• MapMatchingConfig (p. 2137)
• MultiPolygonGeometryInput (p. 2138)
• Operation (p. 2139)
• OutputBand (p. 2140)
• OutputConfigInput (p. 2141)
• OutputResolutionResamplingInput (p. 2142)
• OutputResolutionStackInput (p. 2143)
• PlatformInput (p. 2144)
• PolygonGeometryInput (p. 2145)
• Properties (p. 2146)
• Property (p. 2148)
• PropertyFilter (p. 2150)
• PropertyFilters (p. 2151)
• RasterDataCollectionMetadata (p. 2152)
• RasterDataCollectionQueryInput (p. 2154)
• RasterDataCollectionQueryOutput (p. 2155)
• RasterDataCollectionQueryWithBandFilterInput (p. 2157)
• ResamplingConfigInput (p. 2158)
• ReverseGeocodingConfig (p. 2159)
• StackConfigInput (p. 2160)
• TemporalStatisticsConfigInput (p. 2161)
• TimeRangeFilterInput (p. 2162)
• TimeRangeFilterOutput (p. 2163)
• UserDefined (p. 2164)
• VectorEnrichmentJobConfig (p. 2165)
• VectorEnrichmentJobDataSourceConfigInput (p. 2166)
• VectorEnrichmentJobErrorDetails (p. 2167)
• VectorEnrichmentJobExportErrorDetails (p. 2168)
• VectorEnrichmentJobInputConfig (p. 2169)
• VectorEnrichmentJobS3Data (p. 2170)
• ViewOffNadirInput (p. 2171)
• ViewSunAzimuthInput (p. 2172)
• ViewSunElevationInput (p. 2173)
• ZonalStatisticsConfigInput (p. 2174)
AreaOfInterest
Service: Amazon SageMaker geospatial capabilities
The geographic extent of the Earth Observation job.

Contents

Important
This data type is a UNION, so only one of the following members can be specified when used or returned.

AreaOfInterestGeometry
A GeoJSON object representing the geographic extent in the coordinate space.

Type: AreaOfInterestGeometry (p. 2111) object

Note: This object is a Union. Only one member of this object can be specified or returned.

Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
AreaOfInterestGeometry

Service: Amazon SageMaker geospatial capabilities

A GeoJSON object representing the geographic extent in the coordinate space.

Contents

Important
This data type is a UNION, so only one of the following members can be specified when used or returned.

MultiPolygonGeometry

The structure representing the MultiPolygon Geometry.

Type: MultiPolygonGeometryInput (p. 2138) object

Required: No

PolygonGeometry

The structure representing Polygon Geometry.

Type: PolygonGeometryInput (p. 2145) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
AssetValue
Service: Amazon SageMaker geospatial capabilities
The structure containing the asset properties.

Contents

Href
- Link to the asset object.
  - Type: String
  - Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
BandMathConfigInput
Service: Amazon SageMaker geospatial capabilities

Input structure for the BandMath operation type. Defines Predefined and CustomIndices to be computed using BandMath.

Contents

CustomIndices

CustomIndices that are computed.

Type: CustomIndicesInput (p. 2116) object

Required: No

PredefinedIndices

One or many of the supported predefined indices to compute. Allowed values: NDVI, EVI2, MSAVI, NDWI, NDMI, NDSI, and WDRVI.

Type: Array of strings

Array Members: Minimum number of 1 item.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
CloudMaskingConfigInput
Service: Amazon SageMaker geospatial capabilities
Input structure for CloudMasking operation type.

Contents

The members of this exception structure are context-dependent.

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
CloudRemovalConfigInput
Service: Amazon SageMaker geospatial capabilities
Input structure for Cloud Removal Operation type

Contents

AlgorithmName

The name of the algorithm used for cloud removal.
Type: String
Valid Values: INTERPOLATION
Required: No

InterpolationValue

The interpolation value you provide for cloud removal.
Type: String
Required: No

TargetBands

TargetBands to be returned in the output of CloudRemoval operation.
Type: Array of strings
Array Members: Minimum number of 1 item.
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
CustomIndicesInput
Service: Amazon SageMaker geospatial capabilities

Input object defining the custom BandMath indices to compute.

Contents

Operations

A list of BandMath indices to compute.

Type: Array of Operation (p. 2139) objects

Array Members: Minimum number of 1 item.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
EarthObservationJobErrorDetails
Service: Amazon SageMaker geospatial capabilities

The structure representing the errors in an EarthObservationJob.

Contents

Message

A detailed message describing the error in an Earth Observation job.

Type: String
Required: No

Type

The type of error in an Earth Observation job.

Type: String
Valid Values: CLIENT_ERROR | SERVER_ERROR
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
EoCloudCoverInput

Service: Amazon SageMaker geospatial capabilities

The structure representing the EoCloudCover filter.

Contents

LowerBound

Lower bound for EoCloudCover.

Type: Float

Required: Yes

UpperBound

Upper bound for EoCloudCover.

Type: Float

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ExportErrorDetails
Service: Amazon SageMaker geospatial capabilities

The structure for returning the export error details in a GetEarthObservationJob.

Contents

ExportResults
   The structure for returning the export error details while exporting results of an Earth Observation job.
   Type: ExportErrorDetailsOutput (p. 2120) object
   Required: No

ExportSourceImages
   The structure for returning the export error details while exporting the source images of an Earth Observation job.
   Type: ExportErrorDetailsOutput (p. 2120) object
   Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ExportErrorDetailsOutput

Service: Amazon SageMaker geospatial capabilities

The structure representing the errors in an export EarthObservationJob operation.

Contents

Message

A detailed message describing the error in an export EarthObservationJob operation.

Type: String

Required: No

Type

The type of error in an export EarthObservationJob operation.

Type: String

Valid Values: CLIENT_ERROR | SERVER_ERROR

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ExportS3DataInput

Service: Amazon SageMaker geospatial capabilities

The structure containing the Amazon S3 path to export the Earth Observation job output.

Contents

S3Uri

The URL to the Amazon S3 data input.

Type: String

Pattern: ^s3://([^/]+)/?([^/]+)$

Required: Yes

KmsKeyId

The Key Management Service key ID for server-side encryption.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ExportVectorEnrichmentJobOutputConfig

Service: Amazon SageMaker geospatial capabilities

An object containing information about the output file.

Contents

S3Data

The input structure for Amazon S3 data; representing the Amazon S3 location of the input data objects.

Type: VectorEnrichmentJobS3Data (p. 2170) object

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
Filter

Service: Amazon SageMaker geospatial capabilities

The structure representing the filters supported by a RasterDataCollection.

Contents

Name

The name of the filter.

Type: String

Required: Yes

Type

The type of the filter being used.

Type: String

Required: Yes

Maximum

The maximum value of the filter.

Type: Float

Required: No

Minimum

The minimum value of the filter.

Type: Float

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
Geometry
Service: Amazon SageMaker geospatial capabilities

The structure representing a Geometry in terms of Type and Coordinates as per GeoJson spec.

Contents

Coordinates

The coordinates of the GeoJson Geometry.

Type: Array of arrays of arrays of doubles

Array Members: Minimum number of 1 item.

Array Members: Minimum number of 4 items.

Array Members: Fixed number of 2 items.

Required: Yes

Type

GeoJson Geometry types like Polygon and MultiPolygon.

Type: String

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
GeoMosaicConfigInput
Service: Amazon SageMaker geospatial capabilities

Input configuration information for the geomosaic.

Contents

AlgorithmName
The name of the algorithm being used for geomosaic.
Type: String
Valid Values: NEAR | BILINEAR | CUBIC | CUBICSPLINE | LANCZOS | AVERAGE | RMS | MODE | MAX | MIN | MED | Q1 | Q3 | SUM
Required: No

TargetBands
The target bands for geomosaic.
Type: Array of strings
Array Members: Minimum number of 1 item.
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
InputConfigInput
Service: Amazon SageMaker geospatial capabilities

Input configuration information.

Contents

PreviousEarthObservationJobArn
The Amazon Resource Name (ARN) of the previous Earth Observation job.

Type: String

Pattern: ^arn:aws[a-z-]{0,12}:sagemaker-geospatial:[a-z0-9-]{1,25}:[0-9]{12}:earth-observation-job/[a-z0-9]{12,}$

Required: No

RasterDataCollectionQuery
The structure representing the RasterDataCollection Query consisting of the Area of Interest, RasterDataCollectionArn, TimeRange and Property Filters.

Type: RasterDataCollectionQueryInput (p. 2154) object

Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
InputConfigOutput

Service: Amazon SageMaker geospatial capabilities

The InputConfig for an EarthObservationJob response.

Contents

PreviousEarthObservationJobArn

The Amazon Resource Name (ARN) of the previous Earth Observation job.

Type: String

Pattern: ^arn:aws[a-z-]{0,12}:sagemaker-geospatial:[a-z0-9-]{1,25}:[0-9]\{12\}:earth-observation-job/[a-z0-9]{12}\$

Required: No

RasterDataCollectionQuery

The structure representing the RasterDataCollection Query consisting of the Area of Interest, RasterDataCollectionArn, RasterDataCollectionName, TimeRange, and Property Filters.

Type: RasterDataCollectionQueryOutput (p. 2155) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ItemSource
Service: Amazon SageMaker geospatial capabilities
The structure representing the items in the response for SearchRasterDataCollection.

Contents

DateTime
The searchable date and time of the item, in UTC.
Type: Timestamp
Required: Yes

Geometry
The item Geometry in GeoJson format.
Type: Geometry (p. 2124) object
Required: Yes

Id
A unique Id for the source item.
Type: String
Required: Yes

Assets
This is a dictionary of Asset Objects data associated with the Item that can be downloaded or streamed, each with a unique key.
Type: String to AssetValue (p. 2112) object map
Required: No

Properties
This field contains additional properties of the item.
Type: Properties (p. 2146) object
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
JobConfigInput

Service: Amazon SageMaker geospatial capabilities

The input structure for the JobConfig in an EarthObservationJob.

Contents

**Important**

This data type is a UNION, so only one of the following members can be specified when used or returned.

- **BandMathConfig**
  
  An object containing information about the job configuration for BandMath.
  
  Type: `BandMathConfigInput (p. 2113)` object
  
  Required: No

- **CloudMaskingConfig**
  
  An object containing information about the job configuration for cloud masking.
  
  Type: `CloudMaskingConfigInput (p. 2114)` object
  
  Required: No

- **CloudRemovalConfig**
  
  An object containing information about the job configuration for cloud removal.
  
  Type: `CloudRemovalConfigInput (p. 2115)` object
  
  Required: No

- **GeoMosaicConfig**
  
  An object containing information about the job configuration for geomosaic.
  
  Type: `GeoMosaicConfigInput (p. 2125)` object
  
  Required: No

- **LandCoverSegmentationConfig**
  
  An object containing information about the job configuration for land cover segmentation.
  
  Type: `LandCoverSegmentationConfigInput (p. 2131)` object
  
  Required: No

- **ResamplingConfig**
  
  An object containing information about the job configuration for resampling.
  
  Type: `ResamplingConfigInput (p. 2158)` object
  
  Required: No

- **StackConfig**
  
  An object containing information about the job configuration for a Stacking Earth Observation job.
  
  Type: `StackConfigInput (p. 2160)` object
TemporalsStatisticsConfig
An object containing information about the job configuration for temporal statistics.
Type: TemporalsStatisticsConfigInput (p. 2161) object

ZonalStatisticsConfig
An object containing information about the job configuration for zonal statistics.
Type: ZonalStatisticsConfigInput (p. 2174) object

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
LandCoverSegmentationConfigInput

Service: Amazon SageMaker geospatial capabilities

The input structure for Land Cover Operation type.

Contents

The members of this exception structure are context-dependent.

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
LandsatCloudCoverLandInput

Service: Amazon SageMaker geospatial capabilities

The structure representing Land Cloud Cover property for Landsat data collection.

Contents

**LowerBound**

The minimum value for Land Cloud Cover property filter. This will filter items having Land Cloud Cover greater than or equal to this value.

Type: Float

Required: Yes

**UpperBound**

The maximum value for Land Cloud Cover property filter. This will filter items having Land Cloud Cover less than or equal to this value.

Type: Float

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ListEarthObservationJobOutputConfig

Service: Amazon SageMaker geospatial capabilities

An object containing information about the output file.

Contents

Arn
The Amazon Resource Name (ARN) of the list of the Earth Observation jobs.
Type: String
Required: Yes

CreationTime
The creation time.
Type: Timestamp
Required: Yes

DurationInSeconds
The duration of the session, in seconds.
Type: Integer
Required: Yes

Name
The names of the Earth Observation jobs in the list.
Type: String
Required: Yes

OperationType
The operation type for an Earth Observation job.
Type: String
Required: Yes

Status
The status of the list of the Earth Observation jobs.
Type: String
Valid Values: INITIALIZING | IN_PROGRESS | STOPPING | COMPLETED | STOPPED | FAILED | DELETING | DELETED
Required: Yes

Tags
Each tag consists of a key and a value.
Type: String to string map
Required: No
See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ListVectorEnrichmentJobOutputConfig

Service: Amazon SageMaker geospatial capabilities

An object containing information about the output file.

Contents

Arn

The Amazon Resource Name (ARN) of the list of the Vector Enrichment jobs.

Type: String

Pattern: ^arn:aws[a-z-][0,12]:sagemaker-geospatial:[a-z0-9-]{1,12}:vector-enrichment-job/[a-z0-9]{12},$

Required: Yes

CreationTime

The creation time.

Type: Timestamp

Required: Yes

DurationInSeconds

The duration of the session, in seconds.

Type: Integer

Required: Yes

Name

The names of the Vector Enrichment jobs in the list.

Type: String

Required: Yes

Status

The status of the Vector Enrichment jobs list.

Type: String

Valid Values: INITIALIZING | IN_PROGRESS | STOPPING | STOPPED | COMPLETED | FAILED | DELETING | DELETED

Required: Yes

Type

The type of the list of Vector Enrichment jobs.

Type: String

Valid Values: REVERSE_GEOCODING | MAP_MATCHING

Required: Yes
Tags

Each tag consists of a key and a value.

Type: String to string map

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
MapMatchingConfig

Service: Amazon SageMaker geospatial capabilities

The input structure for Map Matching operation type.

Contents

IdAttributeName

The field name for the data that describes the identifier representing a collection of GPS points belonging to an individual trace.

Type: String

Required: Yes

TimestampAttributeName

The name of the timestamp attribute.

Type: String

Required: Yes

XAttributeName

The name of the X-attribute

Type: String

Required: Yes

YAttributeName

The name of the Y-attribute

Type: String

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
MultiPolygonGeometryInput
Service: Amazon SageMaker geospatial capabilities

The structure representing Polygon Geometry based on the GeoJson spec.

Contents

Coordinates

The coordinates of the multipolygon geometry.
Type: Array of arrays of arrays of arrays of doubles
Array Members: Minimum number of 1 item.
Array Members: Minimum number of 4 items.
Array Members: Fixed number of 2 items.
Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
Operation
Service: Amazon SageMaker geospatial capabilities

Represents an arithmetic operation to compute spectral index.

Contents

Equation
Textual representation of the math operation; Equation used to compute the spectral index.
Type: String
Required: Yes

Name
The name of the operation.
Type: String
Required: Yes

OutputType
The type of the operation.
Type: String
Valid Values: INT32 | FLOAT32 | INT16 | FLOAT64 | UINT16
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:
- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
OutputBand
Service: Amazon SageMaker geospatial capabilities
A single EarthObservationJob output band.

Contents

BandName
The name of the band.
Type: String
Required: Yes
OutputDataType
The datatype of the output band.
Type: String
Valid Values: INT32 | FLOAT32 | INT16 | FLOAT64 | UINT16
Required: Yes

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
OutputConfigInput

Service: Amazon SageMaker geospatial capabilities

The response structure for an OutputConfig returned by an ExportEarthObservationJob.

Contents

S3Data

Path to Amazon S3 storage location for the output configuration file.

Type: ExportS3DataInput (p. 2121) object

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
OutputResolutionResamplingInput

Service: Amazon SageMaker geospatial capabilities

OutputResolution Configuration indicating the target resolution for the output of Resampling operation.

Contents

UserDefined

User Defined Resolution for the output of Resampling operation defined by value and unit.

Type: UserDefined (p. 2164) object

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
OutputResolutionStackInput

Service: Amazon SageMaker geospatial capabilities

The input structure representing Output Resolution for Stacking Operation.

Contents

**Predefined**

A string value representing Predefined Output Resolution for a stacking operation. Allowed values are HIGHEST, LOWEST, and AVERAGE.

Type: String

Valid Values: HIGHEST | LOWEST | AVERAGE

Required: No

**UserDefined**

The structure representing User Output Resolution for a Stacking operation defined as a value and unit.

Type: [UserDefined](p. 2164) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
PlatformInput
Service: Amazon SageMaker geospatial capabilities

The input structure for specifying Platform. Platform refers to the unique name of the specific platform the instrument is attached to. For satellites it is the name of the satellite, eg. landsat-8 (Landsat-8), sentinel-2a.

Contents

Value
The value of the platform.
Type: String
Required: Yes

ComparisonOperator
The ComparisonOperator to use with PlatformInput.
Type: String
Valid Values: EQUALS | NOT_EQUALS | STARTS_WITH
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
**PolygonGeometryInput**
Service: Amazon SageMaker geospatial capabilities

The structure representing Polygon Geometry based on the GeoJson spec.

**Contents**

**Coordinates**

Coordinates representing a Polygon based on the GeoJson spec.

Type: Array of arrays of arrays of doubles

Array Members: Minimum number of 1 item.

Array Members: Minimum number of 4 items.

Array Members: Fixed number of 2 items.

Required: Yes

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
Properties
Service: Amazon SageMaker geospatial capabilities
Properties associated with the Item.

Contents

**EoCloudCover**

Estimate of cloud cover.

Type: Float

Required: No

**LandsatCloudCoverLand**

Land cloud cover for Landsat Data Collection.

Type: Float

Required: No

**Platform**

Platform property. Platform refers to the unique name of the specific platform the instrument is attached to. For satellites it is the name of the satellite, eg. landsat-8 (Landsat-8), sentinel-2a.

Type: String

Required: No

**ViewOffNadir**

The angle from the sensor between nadir (straight down) and the scene center. Measured in degrees (0-90).

Type: Float

Required: No

**ViewSunAzimuth**

The sun azimuth angle. From the scene center point on the ground, this is the angle between truth north and the sun. Measured clockwise in degrees (0-360).

Type: Float

Required: No

**ViewSunElevation**

The sun elevation angle. The angle from the tangent of the scene center point to the sun. Measured from the horizon in degrees (-90-90). Negative values indicate the sun is below the horizon, e.g. sun elevation of -10° means the data was captured during nautical twilight.

Type: Float

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:
Properties

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
Property
Service: Amazon SageMaker geospatial capabilities

Represents a single searchable property to search on.

Contents

**Important**
This data type is a UNION, so only one of the following members can be specified when used or returned.

**EoCloudCover**
The structure representing EoCloudCover property filter containing a lower bound and upper bound.

Type: `EoCloudCoverInput (p. 2118)` object

Required: No

**LandsatCloudCoverLand**
The structure representing Land Cloud Cover property filter for Landsat collection containing a lower bound and upper bound.

Type: `LandsatCloudCoverLandInput (p. 2132)` object

Required: No

**Platform**
The structure representing Platform property filter consisting of value and comparison operator.

Type: `PlatformInput (p. 2144)` object

Required: No

**ViewOffNadir**
The structure representing ViewOffNadir property filter containing a lower bound and upper bound.

Type: `ViewOffNadirInput (p. 2171)` object

Required: No

**ViewSunAzimuth**
The structure representing ViewSunAzimuth property filter containing a lower bound and upper bound.

Type: `ViewSunAzimuthInput (p. 2172)` object

Required: No

**ViewSunElevation**
The structure representing ViewSunElevation property filter containing a lower bound and upper bound.

Type: `ViewSunElevationInput (p. 2173)` object

Required: No
See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
PropertyFilter

Service: Amazon SageMaker geospatial capabilities

The structure representing a single PropertyFilter.

Contents

Property

Represents a single property to match with when searching a raster data collection.

Type: Property (p. 2148) object

Note: This object is a Union. Only one member of this object can be specified or returned.

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
PropertyFilters
Service: Amazon SageMaker geospatial capabilities
A list of PropertyFilter objects.

Contents

LogicalOperator
The Logical Operator used to combine the Property Filters.
Type: String
Valid Values: AND
Required: No

Properties
A list of Property Filters.
Type: Array of PropertyFilter objects
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
RasterDataCollectionMetadata

Service: Amazon SageMaker geospatial capabilities

Response object containing details for a specific RasterDataCollection.

Contents

Arn

The Amazon Resource Name (ARN) of the raster data collection.

Type: String

Pattern: ^arn:aws[a-z-]{0,12}:sagemaker-geospatial:[a-z0-9-]{1,25}:[0-9]{12}:raster-data-collection/(public|premium|user)/[a-z0-9]{12,}$

Required: Yes

Description

A description of the raster data collection.

Type: String

Required: Yes

Name

The name of the raster data collection.

Type: String

Required: Yes

SupportedFilters

The list of filters supported by the raster data collection.

Type: Array of Filter objects

Required: Yes

Type

The type of raster data collection.

Type: String

Valid Values: PUBLIC | PREMIUM | USER

Required: Yes

DescriptionPageUrl

The description URL of the raster data collection.

Type: String

Required: No

Tags

Each tag consists of a key and a value.

Type: String to string map
See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
RasterDataCollectionQueryInput
Service: Amazon SageMaker geospatial capabilities

The input structure for Raster Data Collection Query containing the Area of Interest, TimeRange Filters, and Property Filters.

Contents

RasterDataCollectionArn

The Amazon Resource Name (ARN) of the raster data collection.
Type: String
Pattern: ^arn:aws[a-z-]{0,12}:sagemaker-geospatial:[a-z0-9-]{1,25}:[0-9]{12}:raster-data-collection/(public|premium|user)/[a-z0-9]{12}$
Required: Yes

TimeRangeFilter

The TimeRange Filter used in the RasterDataCollection Query.
Type: TimeRangeFilterInput (p. 2162) object
Required: Yes

AreaOfInterest

The area of interest being queried for the raster data collection.
Type: AreaOfInterest (p. 2110) object

Note: This object is a Union. Only one member of this object can be specified or returned.
Required: No

PropertyFilters

The list of Property filters used in the Raster Data Collection Query.
Type: PropertyFilters (p. 2151) object
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
RasterDataCollectionQueryOutput

Service: Amazon SageMaker geospatial capabilities

The output structure contains the Raster Data Collection Query input along with some additional metadata.

Contents

RasterDataCollectionArn

The ARN of the Raster Data Collection against which the search is done.

Type: String

Pattern: ^arn:aws[a-z-]{0,12}:sagemaker-geospatial:[a-z0-9-]{1,25}:[0-9]{12}:raster-data-collection/(public|premium|user)/[a-z0-9]{12,}$

Required: Yes

RasterDataCollectionName

The name of the raster data collection.

Type: String

Required: Yes

TimeRangeFilter

The TimeRange filter used in the search.

Type: TimeRangeFilterOutput (p. 2163) object

Required: Yes

AreaOfInterest

The Area of Interest used in the search.

Type: AreaOfInterest (p. 2110) object

Note: This object is a Union. Only one member of this object can be specified or returned.

Required: No

PropertyFilters

Property filters used in the search.

Type: PropertyFilters (p. 2151) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
• AWS SDK for Ruby V3
**RasterDataCollectionQueryWithBandFilterInput**

Service: Amazon SageMaker geospatial capabilities

This is a RasterDataCollectionQueryInput containing AreaOfInterest, Time Range filter and Property filters.

**Contents**

**TimeRangeFilter**

- The TimeRange Filter used in the search query.
- Type: *TimeRangeFilterInput (p. 2162)* object
- Required: Yes

**AreaOfInterest**

- The Area of interest to be used in the search query.
- Type: *AreaOfInterest (p. 2110)* object
- **Note:** This object is a Union. Only one member of this object can be specified or returned.
- Required: No

**BandFilter**

- The list of Bands to be displayed in the result for each item.
- Type: Array of strings
- Array Members: Minimum number of 1 item.
- Required: No

**PropertyFilters**

- The Property Filters used in the search query.
- Type: *PropertyFilters (p. 2151)* object
- Required: No

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
ResamplingConfigInput
Service: Amazon SageMaker geospatial capabilities

The structure representing input for resampling operation.

Contents

OutputResolution
The structure representing output resolution (in target georeferenced units) of the result of resampling operation.

Type: OutputResolutionResamplingInput (p. 2142) object

Required: Yes

AlgorithmName
The name of the algorithm used for resampling.

Type: String

Valid Values: NEAR | BILINEAR | CUBIC | CUBICSPLINE | LANCZOS | AVERAGE | RMS | MODE | MAX | MIN | MED | Q1 | Q3 | SUM

Required: No

TargetBands
Bands used in the operation. If no target bands are specified, it uses all bands available in the input.

Type: Array of strings

Array Members: Minimum number of 1 item.

Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ReverseGeocodingConfig
Service: Amazon SageMaker geospatial capabilities

The input structure for Reverse Geocoding operation type.

Contents

XAttributeName
The field name for the data that describes x-axis coordinate, eg. longitude of a point.
Type: String
Required: Yes

YAttributeName
The field name for the data that describes y-axis coordinate, eg. latitude of a point.
Type: String
Required: Yes

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
StackConfigInput
Service: Amazon SageMaker geospatial capabilities

The input structure for Stacking Operation.

Contents

OutputResolution
The structure representing output resolution (in target georeferenced units) of the result of stacking operation.

Type: OutputResolutionStackInput (p. 2143) object

Required: No

TargetBands
A list of bands to be stacked in the specified order. When the parameter is not provided, all the available bands in the data collection are stacked in the alphabetical order of their asset names.

Type: Array of strings

Array Members: Minimum number of 1 item.

Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
TemporalStatisticsConfigInput

Service: Amazon SageMaker geospatial capabilities

The structure representing the configuration for Temporal Statistics operation.

Contents

Statistics

The list of the statistics method options.

Type: Array of strings

Array Members: Minimum number of 1 item.

Valid Values: MEAN | MEDIAN | STANDARD_DEVIATION

Required: Yes

GroupBy

The input for the temporal statistics grouping by time frequency option.

Type: String

Valid Values: ALL | YEARLY

Required: No

TargetBands

The list of target band names for the temporal statistic to calculate.

Type: Array of strings

Array Members: Minimum number of 1 item.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
TimeRangeFilterInput
Service: Amazon SageMaker geospatial capabilities

The input for the time-range filter.

Contents

EndTime

The end time for the time-range filter.

Type: Timestamp

Required: Yes

StartTime

The start time for the time-range filter.

Type: Timestamp

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
TimeRangeFilterOutput
Service: Amazon SageMaker geospatial capabilities

The output structure of the time range filter.

Contents

EndTime
The ending time for the time range filter.
Type: Timestamp
Required: Yes

StartTime
The starting time for the time range filter.
Type: Timestamp
Required: Yes

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
UserDefined
Service: Amazon SageMaker geospatial capabilities
The output resolution (in target georeferenced units) of the result of the operation

Contents

Unit
The units for output resolution of the result.
Type: String
Valid Values: METERS
Required: Yes

Value
The value for output resolution of the result.
Type: Float
Required: Yes

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
VectorEnrichmentJobConfig

Service: Amazon SageMaker geospatial capabilities

It contains configs such as ReverseGeocodingConfig and MapMatchingConfig.

Contents

**Important**
This data type is a UNION, so only one of the following members can be specified when used or returned.

**MapMatchingConfig**

The input structure for Map Matching operation type.

Type: [MapMatchingConfig](p. 2137) object

Required: No

**ReverseGeocodingConfig**

The input structure for Reverse Geocoding operation type.

Type: [ReverseGeocodingConfig](p. 2159) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
VectorEnrichmentJobDataSourceConfigInput

Service: Amazon SageMaker geospatial capabilities

The input structure for the data source that represents the storage type of the input data objects.

Contents

**Important**

This data type is a UNION, so only one of the following members can be specified when used or returned.

**S3Data**

The input structure for the Amazon S3 data that represents the Amazon S3 location of the input data objects.

Type: [VectorEnrichmentJobS3Data](p. 2170) object

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
VectorEnrichmentJobErrorDetails

Service: Amazon SageMaker geospatial capabilities

VectorEnrichmentJob error details in response from GetVectorEnrichmentJob.

Contents

ErrorMessage

A message that you define and then is processed and rendered by the Vector Enrichment job when
the error occurs.

Type: String
Required: No

ErrorType

The type of error generated during the Vector Enrichment job.

Type: String
Valid Values: CLIENT_ERROR | SERVER_ERROR
Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
VectorEnrichmentJobExportErrorDetails

Service: Amazon SageMaker geospatial capabilities

VectorEnrichmentJob export error details in response from GetVectorEnrichmentJob.

Contents

**Message**

The message providing details about the errors generated during the Vector Enrichment job.

Type: String

Required: No

**Type**

The output error details for an Export operation on a Vector Enrichment job.

Type: String

Valid Values: CLIENT_ERROR | SERVER_ERROR

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
VectorEnrichmentJobInputConfig
Service: Amazon SageMaker geospatial capabilities
The input structure for the InputConfig in a VectorEnrichmentJob.

Contents

DataSourceConfig
The input structure for the data source that represents the storage type of the input data objects.
Type: VectorEnrichmentJobDataSourceConfigInput (p. 2166) object

Note: This object is a Union. Only one member of this object can be specified or returned.
Required: Yes

DocumentType
The input structure that defines the data source file type.
Type: String
Valid Values: CSV
Required: Yes

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
VectorEnrichmentJobS3Data

Service: Amazon SageMaker geospatial capabilities

The Amazon S3 data for the Vector Enrichment job.

Contents

S3Uri

The URL to the Amazon S3 data for the Vector Enrichment job.

Type: String

Pattern: ^s3://([^/]+)/(.*)$

Required: Yes

KmsKeyId

The Key Management Service key ID for server-side encryption.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
ViewOffNadirInput
Service: Amazon SageMaker geospatial capabilities

The input structure for specifying ViewOffNadir property filter. ViewOffNadir refers to the angle from the sensor between nadir (straight down) and the scene center. Measured in degrees (0-90).

Contents

LowerBound

The minimum value for ViewOffNadir property filter. This filters items having ViewOffNadir greater than or equal to this value.

Type: Float

Required: Yes

UpperBound

The maximum value for ViewOffNadir property filter. This filters items having ViewOffNadir lesser than or equal to this value.

Type: Float

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
ViewSunAzimuthInput

Service: Amazon SageMaker geospatial capabilities

The input structure for specifying ViewSunAzimuth property filter. ViewSunAzimuth refers to the Sun azimuth angle. From the scene center point on the ground, this is the angle between truth north and the sun. Measured clockwise in degrees (0-360).

Contents

LowerBound

The minimum value for ViewSunAzimuth property filter. This filters items having ViewSunAzimuth greater than or equal to this value.

Type: Float

Required: Yes

UpperBound

The maximum value for ViewSunAzimuth property filter. This filters items having ViewSunAzimuth lesser than or equal to this value.

Type: Float

Required: Yes

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
**ViewSunElevationInput**

Service: Amazon SageMaker geospatial capabilities

The input structure for specifying ViewSunElevation angle property filter.

**Contents**

**LowerBound**

The lower bound to view the sun elevation.

Type: Float

Required: Yes

**UpperBound**

The upper bound to view the sun elevation.

Type: Float

Required: Yes

**See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java V2](#)
- [AWS SDK for Ruby V3](#)
ZonalStatisticsConfigInput

Service: Amazon SageMaker geospatial capabilities

The structure representing input configuration of ZonalStatistics operation.

Contents

**Statistics**

List of zonal statistics to compute.

Type: Array of strings

Array Members: Minimum number of 1 item.

Valid Values: MEAN | MEDIAN | STANDARD_DEVIATION | MAX | MIN | SUM

Required: Yes

**ZoneS3Path**

The Amazon S3 path pointing to the GeoJSON containing the polygonal zones.

Type: String

Pattern: ^s3://([^/]+)/?(.*)$

Required: Yes

**TargetBands**

Bands used in the operation. If no target bands are specified, it uses all bands available input.

Type: Array of strings

Array Members: Minimum number of 1 item.

Required: No

**ZoneS3PathKmsKeyId**

The Amazon Resource Name (ARN) or an ID of a AWS Key Management Service (AWS KMS) key that Amazon SageMaker uses to decrypt your output artifacts with Amazon S3 server-side encryption. The SageMaker execution role must have kms:GenerateDataKey permission.

The KmsKeyId can be any of the following formats:

- // KMS Key ID
  
  "1234abcd-12ab-34cd-56ef-1234567890ab"

- // Amazon Resource Name (ARN) of a KMS Key
  
  "arn:aws:kms:<region>:<account>:key/<key-id-12ab-34cd-56ef-1234567890ab>"

For more information about key identifiers, see [Key identifiers (KeyID)] in the AWS Key Management Service (AWS KMS) documentation.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 2048.

Required: No
See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3

Amazon SageMaker Metrics Service

The following data types are supported by Amazon SageMaker Metrics Service:

- BatchPutMetricsError (p. 2176)
- RawMetricData (p. 2177)
BatchPutMetricsError
Service: Amazon SageMaker Metrics Service

An error that occurred when putting the metric data.

Contents

Code

The error code of an error that occurred when attempting to put metrics.

- METRIC_LIMIT_EXCEEDED: The maximum amount of metrics per resource is exceeded.
- INTERNAL_ERROR: An internal error occurred.
- VALIDATION_ERROR: The metric data failed validation.
- CONFLICT_ERROR: Multiple requests attempted to modify the same data simultaneously.

Type: String

Valid Values: METRIC_LIMIT_EXCEEDED | INTERNAL_ERROR | VALIDATION_ERROR | CONFLICT_ERROR

Required: No

MetricIndex

An index that corresponds to the metric in the request.

Type: Integer

Required: No

See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
RawMetricData
Service: Amazon SageMaker Metrics Service
The raw metric data to associate with the resource.

Contents

MetricName
The name of the metric.
Type: String
Length Constraints: Minimum length of 1. Maximum length of 255.
Pattern: .+
Required: Yes

Timestamp
The time that the metric was recorded.
Type: Timestamp
Required: Yes

Value
The metric value.
Type: Double
Required: Yes

Step
The metric step (epoch).
Type: Integer
Valid Range: Minimum value of 0.
Required: No

See Also
For more information about using this API in one of the language-specific AWS SDKs, see the following:

- AWS SDK for C++
- AWS SDK for Go
- AWS SDK for Java V2
- AWS SDK for Ruby V3
Common Parameters

The following list contains the parameters that all actions use for signing Signature Version 4 requests with a query string. Any action-specific parameters are listed in the topic for that action. For more information about Signature Version 4, see Signing AWS API requests in the IAM User Guide.

Action
The action to be performed.
Type: string
Required: Yes

Version
The API version that the request is written for, expressed in the format YYYY-MM-DD.
Type: string
Required: Yes

X-Amz-Algorithm
The hash algorithm that you used to create the request signature.
Condition: Specify this parameter when you include authentication information in a query string instead of in the HTTP authorization header.
Type: string
Valid Values: AWS4-HMAC-SHA256
Required: Conditional

X-Amz-Credential
The credential scope value, which is a string that includes your access key, the date, the region you are targeting, the service you are requesting, and a termination string ("aws4_request"). The value is expressed in the following format: access_key/YYYYMMDD/region/service/aws4_request.
Condition: Specify this parameter when you include authentication information in a query string instead of in the HTTP authorization header.
Type: string
Required: Conditional

X-Amz-Date
The date that is used to create the signature. The format must be ISO 8601 basic format (YYYYMMDD'T'HHMMSS'Z'). For example, the following date time is a valid X-Amz-Date value: 20120325T120000Z.
Condition: X-Amz-Date is optional for all requests; it can be used to override the date used for signing requests. If the Date header is specified in the ISO 8601 basic format, X-Amz-Date is not required. When X-Amz-Date is used, it always overrides the value of the Date header. For more information, see Elements of an AWS API request signature in the IAM User Guide.
Type: string
Required: Conditional

**X-Amz-Security-Token**

The temporary security token that was obtained through a call to AWS Security Token Service (AWS STS). For a list of services that support temporary security credentials from AWS STS, see [AWS services that work with IAM](https://docs.aws.amazon.com/IAM/latest/UserGuide/id我不想做你的老师.html) in the *IAM User Guide*.

Condition: If you're using temporary security credentials from AWS STS, you must include the security token.

Type: string
Required: Conditional

**X-Amz-Signature**

Specifies the hex-encoded signature that was calculated from the string to sign and the derived signing key.

Condition: Specify this parameter when you include authentication information in a query string instead of in the HTTP authorization header.

Type: string
Required: Conditional

**X-Amz-SignedHeaders**

Specifies all the HTTP headers that were included as part of the canonical request. For more information about specifying signed headers, see [Create a signed AWS API request](https://docs.aws.amazon.com/IAM/latest/UserGuide/id我不想做你的老师.html) in the *IAM User Guide*.

Condition: Specify this parameter when you include authentication information in a query string instead of in the HTTP authorization header.

Type: string
Required: Conditional
Common Errors

This section lists the errors common to the API actions of all AWS services. For errors specific to an API action for this service, see the topic for that API action.

AccessDeniedException
You do not have sufficient access to perform this action.
HTTP Status Code: 400

IncompleteSignature
The request signature does not conform to AWS standards.
HTTP Status Code: 400

InternalFailure
The request processing has failed because of an unknown error, exception or failure.
HTTP Status Code: 500

InvalidAction
The action or operation requested is invalid. Verify that the action is typed correctly.
HTTP Status Code: 400

InvalidClientTokenId
The X.509 certificate or AWS access key ID provided does not exist in our records.
HTTP Status Code: 403

NotAuthorized
You do not have permission to perform this action.
HTTP Status Code: 400

OptInRequired
The AWS access key ID needs a subscription for the service.
HTTP Status Code: 403

RequestExpired
The request reached the service more than 15 minutes after the date stamp on the request or more than 15 minutes after the request expiration date (such as for pre-signed URLs), or the date stamp on the request is more than 15 minutes in the future.
HTTP Status Code: 400

ServiceUnavailable
The request has failed due to a temporary failure of the server.
HTTP Status Code: 503

ThrottlingException
The request was denied due to request throttling.
HTTP Status Code: 400

ValidationError

The input fails to satisfy the constraints specified by an AWS service.

HTTP Status Code: 400