API Reference

AWS Batch



API Version 2016-08-10

Copyright © 2026 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

AWS Batch: API Reference

Copyright © 2026 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**
Actions **2**
CancelJob 4
Request Syntax 4
URI Request Parameters 4
Request Body 4
Response Syntax 5
Response Elements 5
Errors 5
Examples 5
See Also 7
CancelJobs 8
Request Syntax 8
URI Request Parameters 9
Request Body 9
Response Syntax 9
Response Elements 10
Errors 10
Examples 11
See Also 12
CreateComputeEnvironment 13
Request Syntax 13
URI Request Parameters 16
Request Body 16
Response Syntax 20
Response Elements 20
Errors 21
Examples 21
See Also 25
CreateConsumableResource 26
Request Syntax 26
URI Request Parameters 26
Request Body 26
Response Syntax 27

Response Elements ... 27
Errors ... 28
See Also ... 28
CreateJobQueue ... 30
Request Syntax ... 30
URI Request Parameters ... 31
Request Body ... 31
Response Syntax ... 34
Response Elements ... 34
Errors ... 34
Examples ... 35
See Also ... 37
CreateQuotaShare ... 39
Request Syntax ... 39
URI Request Parameters ... 39
Request Body ... 39
Response Syntax ... 41
Response Elements ... 41
Errors ... 42
Examples ... 42
See Also ... 44
CreateSchedulingPolicy ... 45
Request Syntax ... 45
URI Request Parameters ... 45
Request Body ... 45
Response Syntax ... 46
Response Elements ... 47
Errors ... 47
Examples ... 48
See Also ... 50
CreateServiceEnvironment ... 51
Request Syntax ... 51
URI Request Parameters ... 51
Request Body ... 51
Response Syntax ... 53
Response Elements ... 53

Errors ... 53
Examples ... 54
See Also ... 55
DeleteComputeEnvironment ... 56
Request Syntax ... 56
URI Request Parameters ... 56
Request Body ... 56
Response Syntax ... 56
Response Elements ... 57
Errors ... 57
Examples ... 57
See Also ... 58
DeleteConsumableResource ... 60
Request Syntax ... 60
URI Request Parameters ... 60
Request Body ... 60
Response Syntax ... 60
Response Elements ... 60
Errors ... 61
See Also ... 61
DeleteJobQueue ... 62
Request Syntax ... 62
URI Request Parameters ... 62
Request Body ... 62
Response Syntax ... 62
Response Elements ... 63
Errors ... 63
Examples ... 63
See Also ... 64
DeleteQuotaShare ... 66
Request Syntax ... 66
URI Request Parameters ... 66
Request Body ... 66
Response Syntax ... 66
Response Elements ... 66
Errors ... 67

Examples ... 67
See Also ... 68
DeleteSchedulingPolicy ... 69
Request Syntax ... 69
URI Request Parameters ... 69
Request Body ... 69
Response Syntax ... 69
Response Elements ... 69
Errors ... 70
Examples ... 70
See Also ... 71
DeleteServiceEnvironment ... 73
Request Syntax ... 73
URI Request Parameters ... 73
Request Body ... 73
Response Syntax ... 73
Response Elements ... 73
Errors ... 74
Examples ... 74
See Also ... 75
DeregisterJobDefinition ... 76
Request Syntax ... 76
URI Request Parameters ... 76
Request Body ... 76
Response Syntax ... 76
Response Elements ... 76
Errors ... 77
Examples ... 77
See Also ... 78
DescribeComputeEnvironments ... 79
Request Syntax ... 79
URI Request Parameters ... 79
Request Body ... 79
Response Syntax ... 80
Response Elements ... 83
Errors ... 84

Examples ... 84
See Also ... 86
DescribeConsumableResource ... 87
Request Syntax ... 87
URI Request Parameters ... 87
Request Body ... 87
Response Syntax ... 87
Response Elements ... 88
Errors ... 89
See Also ... 89
DescribeJobDefinitions ... 91
Request Syntax ... 91
URI Request Parameters ... 91
Request Body ... 91
Response Syntax ... 93
Response Elements ... 111
Errors ... 112
Examples ... 112
See Also ... 114
DescribeJobQueues ... 115
Request Syntax ... 115
URI Request Parameters ... 115
Request Body ... 115
Response Syntax ... 116
Response Elements ... 117
Errors ... 118
Examples ... 118
See Also ... 119
DescribeJobs ... 121
Request Syntax ... 121
URI Request Parameters ... 121
Request Body ... 121
Response Syntax ... 121
Response Elements ... 143
Errors ... 143
Examples ... 143

See Also ... 146
DescribeQuotaShare ... 148
Request Syntax ... 148
URI Request Parameters ... 148
Request Body ... 148
Response Syntax ... 148
Response Elements ... 149
Errors ... 151
Examples ... 151
See Also ... 153
DescribeSchedulingPolicies ... 154
Request Syntax ... 154
URI Request Parameters ... 154
Request Body ... 154
Response Syntax ... 154
Response Elements ... 155
Errors ... 155
Examples ... 156
See Also ... 158
DescribeServiceEnvironments ... 160
Request Syntax ... 160
URI Request Parameters ... 160
Request Body ... 160
Response Syntax ... 161
Response Elements ... 162
Errors ... 162
Examples ... 163
See Also ... 164
DescribeServiceJob ... 165
Request Syntax ... 165
URI Request Parameters ... 165
Request Body ... 165
Response Syntax ... 165
Response Elements ... 167
Errors ... 171
Examples ... 171

See Also ... 173
GetJobQueueSnapshot ... 174
Request Syntax ... 174
URI Request Parameters ... 174
Request Body ... 174
Response Syntax ... 174
Response Elements ... 176
Errors ... 177
See Also ... 177
ListConsumableResources ... 178
Request Syntax ... 178
URI Request Parameters ... 178
Request Body ... 178
Response Syntax ... 179
Response Elements ... 180
Errors ... 180
See Also ... 181
ListJobs ... 182
Request Syntax ... 182
URI Request Parameters ... 182
Request Body ... 182
Response Syntax ... 186
Response Elements ... 187
Errors ... 188
Examples ... 188
See Also ... 190
ListJobsByConsumableResource ... 191
Request Syntax ... 191
URI Request Parameters ... 191
Request Body ... 191
Response Syntax ... 193
Response Elements ... 193
Errors ... 194
See Also ... 194
ListQuotaShares ... 196
Request Syntax ... 196

URI Request Parameters ... 196
Request Body ... 196
Response Syntax ... 197
Response Elements ... 198
Errors ... 198
Examples ... 199
See Also ... 200
ListSchedulingPolicies ... 202
Request Syntax ... 202
URI Request Parameters ... 202
Request Body ... 202
Response Syntax ... 203
Response Elements ... 203
Errors ... 204
Examples ... 204
See Also ... 205
ListServiceJobs ... 207
Request Syntax ... 207
URI Request Parameters ... 207
Request Body ... 207
Response Syntax ... 210
Response Elements ... 211
Errors ... 211
Examples ... 212
See Also ... 214
ListTagsForResource ... 215
Request Syntax ... 215
URI Request Parameters ... 215
Request Body ... 215
Response Syntax ... 215
Response Elements ... 216
Errors ... 216
Examples ... 216
See Also ... 218
RegisterJobDefinition ... 219
Request Syntax ... 219

URI Request Parameters ... 237
Request Body ... 237
Response Syntax ... 242
Response Elements ... 242
Errors ... 243
Examples ... 243
See Also ... 246
SubmitJob ... 248
Request Syntax ... 248
URI Request Parameters ... 254
Request Body ... 254
Response Syntax ... 259
Response Elements ... 259
Errors ... 260
Examples ... 260
See Also ... 261
SubmitServiceJob ... 263
Request Syntax ... 263
URI Request Parameters ... 263
Request Body ... 264
Response Syntax ... 266
Response Elements ... 266
Errors ... 267
Examples ... 267
See Also ... 269
TagResource ... 271
Request Syntax ... 271
URI Request Parameters ... 271
Request Body ... 271
Response Syntax ... 272
Response Elements ... 272
Errors ... 272
Examples ... 272
See Also ... 274
TerminateJob ... 275
Request Syntax ... 275

URI Request Parameters ... 275
Request Body ... 275
Response Syntax ... 276
Response Elements ... 276
Errors ... 276
Examples ... 276
See Also ... 277
TerminateJobs ... 279
Request Syntax ... 279
URI Request Parameters ... 279
Request Body ... 279
Response Syntax ... 280
Response Elements ... 280
Errors ... 281
Examples ... 281
See Also ... 283
TerminateServiceJob ... 284
Request Syntax ... 284
URI Request Parameters ... 284
Request Body ... 284
Response Syntax ... 284
Response Elements ... 285
Errors ... 285
Examples ... 285
See Also ... 286
TerminateServiceJobs ... 288
Request Syntax ... 288
URI Request Parameters ... 288
Request Body ... 288
Response Syntax ... 289
Response Elements ... 289
Errors ... 290
Examples ... 290
See Also ... 292
UntagResource ... 293
Request Syntax ... 293

URI Request Parameters ... 293
Request Body ... 293
Response Syntax ... 293
Response Elements ... 294
Errors ... 294
Examples ... 294
See Also ... 295
UpdateComputeEnvironment ... 296
Request Syntax ... 296
URI Request Parameters ... 298
Request Body ... 298
Response Syntax ... 301
Response Elements ... 302
Errors ... 302
Examples ... 302
See Also ... 304
UpdateConsumableResource ... 305
Request Syntax ... 305
URI Request Parameters ... 305
Request Body ... 305
Response Syntax ... 306
Response Elements ... 306
Errors ... 307
See Also ... 307
UpdateJobQueue ... 309
Request Syntax ... 309
URI Request Parameters ... 309
Request Body ... 310
Response Syntax ... 312
Response Elements ... 312
Errors ... 312
Examples ... 313
See Also ... 314
UpdateQuotaShare ... 315
Request Syntax ... 315
URI Request Parameters ... 315

Request Body ... 315
Response Syntax ... 316
Response Elements ... 317
Errors ... 317
Examples ... 317
See Also ... 319
UpdateSchedulingPolicy ... 320
Request Syntax ... 320
URI Request Parameters ... 320
Request Body ... 320
Response Syntax ... 321
Response Elements ... 321
Errors ... 321
Examples ... 322
See Also ... 323
UpdateServiceEnvironment ... 325
Request Syntax ... 325
URI Request Parameters ... 325
Request Body ... 325
Response Syntax ... 326
Response Elements ... 326
Errors ... 327
Examples ... 327
See Also ... 328
UpdateServiceJob ... 330
Request Syntax ... 330
URI Request Parameters ... 330
Request Body ... 330
Response Syntax ... 331
Response Elements ... 331
Errors ... 331
See Also ... 332
Data Types ... 333
ArrayProperties ... 339
Contents ... 339
See Also ... 339

ArrayPropertiesDetail 340
Contents 340
See Also 340
ArrayPropertiesSummary 342
Contents 342
See Also 342
AttemptContainerDetail 344
Contents 344
See Also 345
AttemptDetail 346
Contents 346
See Also 347
AttemptEcsTaskDetails 348
Contents 348
See Also 348
AttemptTaskContainerDetails 349
Contents 349
See Also 350
CancelJobsErrorDetail 351
Contents 351
See Also 351
CapacityLimit 353
Contents 353
See Also 353
CapacityReservationRequest 355
Contents 355
See Also 355
ComputeEnvironmentDetail 356
Contents 356
See Also 360
ComputeEnvironmentOrder 361
Contents 361
See Also 362
ComputeResource 363
Contents 363
See Also 375

ComputeResourceUpdate ... 376
Contents ... 376
See Also ... 389
ComputeScalingPolicy ... 390
Contents ... 390
See Also ... 390
ConsumableResourceProperties ... 392
Contents ... 392
See Also ... 392
ConsumableResourceRequirement ... 393
Contents ... 393
See Also ... 393
ConsumableResourceSummary ... 394
Contents ... 394
See Also ... 395
ContainerDetail ... 396
Contents ... 396
See Also ... 403
ContainerOverrides ... 404
Contents ... 404
See Also ... 406
ContainerProperties ... 407
Contents ... 407
See Also ... 414
ContainerSummary ... 415
Contents ... 415
See Also ... 415
Device ... 416
Contents ... 416
See Also ... 417
Ec2Configuration ... 418
Contents ... 418
See Also ... 422
EcsProperties ... 423
Contents ... 423
See Also ... 423

EcsPropertiesDetail ... 424
Contents ... 424
See Also ... 424
EcsPropertiesOverride ... 425
Contents ... 425
See Also ... 425
EcsSettings ... 426
Contents ... 426
See Also ... 426
EcsTaskDetails ... 428
Contents ... 428
See Also ... 431
EcsTaskProperties ... 432
Contents ... 432
See Also ... 435
EFSAuthorizationConfig ... 436
Contents ... 436
See Also ... 436
EFSVolumeConfiguration ... 438
Contents ... 438
See Also ... 439
EksAccessEntry ... 440
Contents ... 440
See Also ... 441
EksAttemptContainerDetail ... 442
Contents ... 442
See Also ... 442
EksAttemptDetail ... 444
Contents ... 444
See Also ... 445
EksConfiguration ... 447
Contents ... 447
See Also ... 448
EksConfigurationUpdate ... 449
Contents ... 449
See Also ... 449

EksContainer 450
Contents 450
See Also 452
EksContainerDetail 453
Contents 453
See Also 455
EksContainerEnvironmentVariable 456
Contents 456
See Also 456
EksContainerOverride 457
Contents 457
See Also 458
EksContainerResourceRequirements 459
Contents 459
See Also 461
EksContainerSecurityContext 462
Contents 462
See Also 463
EksContainerVolumeMount 464
Contents 464
See Also 464
EksEmptyDir 466
Contents 466
See Also 467
EksHostPath 468
Contents 468
See Also 468
EksMetadata 469
Contents 469
See Also 470
EksPersistentVolumeClaim 471
Contents 471
See Also 471
EksPodProperties 472
Contents 472
See Also 474

EksPodPropertiesDetail ... 475
Contents ... 475
See Also ... 477
EksPodPropertiesOverride ... 478
Contents ... 478
See Also ... 478
EksProperties ... 480
Contents ... 480
See Also ... 480
EksPropertiesDetail ... 481
Contents ... 481
See Also ... 481
EksPropertiesOverride ... 482
Contents ... 482
See Also ... 482
EksSecret ... 483
Contents ... 483
See Also ... 483
EksVolume ... 484
Contents ... 484
See Also ... 485
EphemeralStorage ... 486
Contents ... 486
See Also ... 486
EvaluateOnExit ... 487
Contents ... 487
See Also ... 488
FairshareCapacityUsage ... 489
Contents ... 489
See Also ... 489
FairshareCapacityUtilization ... 490
Contents ... 490
See Also ... 490
FairsharePolicy ... 491
Contents ... 491
See Also ... 492

FairshareUtilizationDetail ... 493
Contents ... 493
See Also ... 493
FargatePlatformConfiguration ... 494
Contents ... 494
See Also ... 494
FirelensConfiguration ... 495
Contents ... 495
See Also ... 495
FrontOfQueueDetail ... 497
Contents ... 497
See Also ... 497
FrontOfQueueJobSummary ... 498
Contents ... 498
See Also ... 498
FrontOfQuotaShareJobSummary ... 499
Contents ... 499
See Also ... 499
FrontOfQuotaSharesDetail ... 500
Contents ... 500
See Also ... 500
Host ... 501
Contents ... 501
See Also ... 501
ImagePullSecret ... 502
Contents ... 502
See Also ... 502
InfrastructureOptimization ... 503
Contents ... 503
See Also ... 503
InstanceLaunchTemplate ... 504
Contents ... 504
See Also ... 506
InstanceLaunchTemplateUpdate ... 507
Contents ... 507
See Also ... 508

InstanceRequirementsRequest ... 509
Contents ... 509
See Also ... 509
JobCapacityUsageSummary ... 510
Contents ... 510
See Also ... 510
JobDefinition ... 511
Contents ... 511
See Also ... 515
JobDependency ... 516
Contents ... 516
See Also ... 516
JobDetail ... 517
Contents ... 517
See Also ... 523
JobQueueDetail ... 524
Contents ... 524
See Also ... 527
JobStateTimeLimitAction ... 528
Contents ... 528
See Also ... 529
JobSummary ... 530
Contents ... 530
See Also ... 533
JobTimeout ... 534
Contents ... 534
See Also ... 534
KeyValuePair ... 535
Contents ... 535
See Also ... 535
KeyValuesPair ... 536
Contents ... 536
See Also ... 536
LatestServiceJobAttempt ... 537
Contents ... 537
See Also ... 537

LaunchTemplateSpecification 538
Contents 538
See Also 540
LaunchTemplateSpecificationOverride 541
Contents 541
See Also 543
LinuxParameters 544
Contents 544
See Also 547
ListJobsByConsumableResourceSummary 548
Contents 548
See Also 550
LogConfiguration 551
Contents 551
See Also 553
ManagedInstancesLocalStorageConfiguration 554
Contents 554
See Also 554
ManagedInstancesNetworkConfiguration 555
Contents 555
See Also 555
ManagedInstancesProvider 556
Contents 556
See Also 557
ManagedInstancesStorageConfiguration 558
Contents 558
See Also 558
MountPoint 559
Contents 559
See Also 559
NetworkConfiguration 561
Contents 561
See Also 561
NetworkInterface 562
Contents 562
See Also 562

NodeDetails ... 563
Contents ... 563
See Also ... 563
NodeOverrides ... 564
Contents ... 564
See Also ... 565
NodeProperties ... 566
Contents ... 566
See Also ... 566
NodePropertiesSummary ... 568
Contents ... 568
See Also ... 568
NodePropertyOverride ... 569
Contents ... 569
See Also ... 570
NodeRangeProperty ... 571
Contents ... 571
See Also ... 572
QueueSnapshotCapacityUsage ... 573
Contents ... 573
See Also ... 573
QueueSnapshotUtilizationDetail ... 574
Contents ... 574
See Also ... 574
QuotaShareCapacityLimit ... 576
Contents ... 576
See Also ... 576
QuotaShareCapacityUsage ... 577
Contents ... 577
See Also ... 577
QuotaShareCapacityUtilization ... 578
Contents ... 578
See Also ... 578
QuotaShareDetail ... 579
Contents ... 579
See Also ... 580

QuotaSharePolicy 581
Contents 581
See Also 581
QuotaSharePreemptionConfiguration 582
Contents 582
See Also 582
QuotaShareResourceSharingConfiguration 583
Contents 583
See Also 583
QuotaShareUtilizationDetail 585
Contents 585
See Also 585
RepositoryCredentials 586
Contents 586
See Also 586
ResourceRequirement 587
Contents 587
See Also 591
RetryStrategy 592
Contents 592
See Also 592
RuntimePlatform 593
Contents 593
See Also 594
S3FilesVolumeConfiguration 595
Contents 595
See Also 595
SchedulingPolicyDetail 597
Contents 597
See Also 598
SchedulingPolicyListingDetail 599
Contents 599
See Also 599
Secret 600
Contents 600
See Also 601

ServiceEnvironmentDetail ... 602
Contents ... 602
See Also ... 603
ServiceEnvironmentOrder ... 604
Contents ... 604
See Also ... 604
ServiceJobAttemptDetail ... 605
Contents ... 605
See Also ... 605
ServiceJobCapacityUsageDetail ... 607
Contents ... 607
See Also ... 607
ServiceJobCapacityUsageSummary ... 608
Contents ... 608
See Also ... 608
ServiceJobEvaluateOnExit ... 609
Contents ... 609
See Also ... 609
ServiceJobPreemptedAttempt ... 610
Contents ... 610
See Also ... 610
ServiceJobPreemptionConfiguration ... 612
Contents ... 612
See Also ... 612
ServiceJobPreemptionSummary ... 613
Contents ... 613
See Also ... 613
ServiceJobRetryStrategy ... 614
Contents ... 614
See Also ... 614
ServiceJobSummary ... 615
Contents ... 615
See Also ... 617
ServiceJobTimeout ... 619
Contents ... 619
See Also ... 619

ServiceResourceId ... 620
Contents ... 620
See Also ... 620
ShareAttributes ... 621
Contents ... 621
See Also ... 621
TaskContainerDependency ... 623
Contents ... 623
See Also ... 623
TaskContainerDetails ... 624
Contents ... 624
See Also ... 631
TaskContainerOverrides ... 632
Contents ... 632
See Also ... 633
TaskContainerProperties ... 634
Contents ... 634
See Also ... 640
TaskPropertiesOverride ... 642
Contents ... 642
See Also ... 642
TerminateJobsErrorDetail ... 643
Contents ... 643
See Also ... 643
TerminateServiceJobsErrorDetail ... 645
Contents ... 645
See Also ... 645
Tmpfs ... 647
Contents ... 647
See Also ... 648
Ulimit ... 649
Contents ... 649
See Also ... 649
UpdateManagedInstancesProviderConfiguration ... 651
Contents ... 651
See Also ... 652

UpdatePolicy 653
Contents 653
See Also 654
Volume 655
Contents 655
See Also 656
Common Parameters 657

Welcome

Using AWS Batch, you can run batch computing workloads on the AWS Cloud. Batch computing is a common means for developers, scientists, and engineers to access large amounts of compute resources. AWS Batch uses the advantages of the batch computing to remove the undifferentiated heavy lifting of configuring and managing required infrastructure. At the same time, it also adopts a familiar batch computing software approach. You can use AWS Batch to efficiently provision resources, and work toward eliminating capacity constraints, reducing your overall compute costs, and delivering results more quickly.

As a fully managed service, AWS Batch can run batch computing workloads of any scale. AWS Batch automatically provisions compute resources and optimizes workload distribution based on the quantity and scale of your specific workloads. With AWS Batch, there's no need to install or manage batch computing software. This means that you can focus on analyzing results and solving your specific problems instead.

This document was last published on October 21, 2026.

Actions

The following actions are supported:

- CancelJob
- CancelJobs
- CreateComputeEnvironment
- CreateConsumableResource
- CreateJobQueue
- CreateQuotaShare
- CreateSchedulingPolicy
- CreateServiceEnvironment
- DeleteComputeEnvironment
- DeleteConsumableResource
- DeleteJobQueue
- DeleteQuotaShare
- DeleteSchedulingPolicy
- DeleteServiceEnvironment
- DeregisterJobDefinition
- DescribeComputeEnvironments
- DescribeConsumableResource
- DescribeJobDefinitions
- DescribeJobQueues
- DescribeJobs
- DescribeQuotaShare
- DescribeSchedulingPolicies
- DescribeServiceEnvironments
- DescribeServiceJob
- GetJobQueueSnapshot
- ListConsumableResources
- ListJobs

- ListJobsByConsumableResource
- ListQuotaShares
- ListSchedulingPolicies
- ListServiceJobs
- ListTagsForResource
- RegisterJobDefinition
- SubmitJob
- SubmitServiceJob
- TagResource
- TerminateJob
- TerminateJobs
- TerminateServiceJob
- TerminateServiceJobs
- UntagResource
- UpdateComputeEnvironment
- UpdateConsumableResource
- UpdateJobQueue
- UpdateQuotaShare
- UpdateSchedulingPolicy
- UpdateServiceEnvironment
- UpdateServiceJob

CancelJob

Cancels a job in an AWS Batch job queue. Jobs that are in a SUBMITTED, PENDING, or RUNNABLE state are cancelled and the job status is updated to FAILED.

> **Note**
>
> A PENDING job is cancelled after all dependency jobs are completed. Therefore, it might take longer than expected to cancel a job in PENDING status.
> When you try to cancel an array parent job in PENDING, AWS Batch attempts to cancel all child jobs. The array parent job is cancelled when all child jobs are completed.

Jobs that progressed to the STARTING or RUNNING state aren't cancelled. However, the API operation still succeeds, even if no job is cancelled. These jobs must be terminated with the TerminateJob or TerminateJobs operation.

Request Syntax

```
POST /v1/canceljob HTTP/1.1
Content-type: application/json

{
   "jobId": "string",
   "reason": "string"
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

jobId

The AWS Batch job ID of the job to cancel.

Type: String

Required: Yes

reason

A message to attach to the job that explains the reason for cancelling it. This message is returned by future DescribeJobs operations on the job. It is also recorded in the AWS Batch activity logs.

This parameter has a limit of 1024 characters.

Type: String

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

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

Examples

In the following example or examples, the Authorization header contents (*[authorization-params]*) must be replaced with an AWS Signature Version 4 signature. For more information

about creating these signatures, see Signature Version 4 Signing Process in the *AWS General Reference*.

You only need to learn how to sign HTTP requests if you intend to manually create them. When you use the AWS Command Line Interface (AWS CLI) or one of the AWS SDKs to make requests to AWS, these tools automatically sign the requests for you with the access key that you specify when you configure the tools. When you use these tools, you don't need to learn how to sign requests yourself.

Example

This example cancels a job with the specified job ID.

Sample Request

```
POST /v1/canceljob HTTP/1.1
Host: batch.us-east-1.amazonaws.com
Accept-Encoding: identity
Content-Length: [content-length]
Authorization: [authorization-params]
X-Amz-Date: 20161130T001258Z
User-Agent: aws-cli/1.11.22 Python/2.7.12 Darwin/16.1.0 botocore/1.4.79

{
  "reason": "Cancelling job.",
  "jobId": "1d828f65-7a4d-42e8-996d-3b900ed59dc4"
}
```

Sample Response

```
HTTP/1.1 200 OK
Content-Type: application/json
Content-Length: 2
Connection: keep-alive
Date: Wed, 30 Nov 2016 00:12:59 GMT
x-amzn-RequestId: [request-id]
X-Amzn-Trace-Id: [trace-id]
X-Cache: Miss from cloudfront
Via: 1.1 bfdd5909914586f5bc4851846228c27f.cloudfront.net (CloudFront)
X-Amz-Cf-Id: whn1dX1uTx34Lvao7-7ZdkDXEbCZ_sjn3v3hHVFgbo1ORJtXyeggSw==

{}
```

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for Kotlin
- AWS SDK for PHP V3
- AWS SDK for Python (Boto3)
- AWS SDK for Ruby V3

CancelJobs

Cancels up to 50 jobs in an AWS Batch job queue. This is a bulk version of CancelJob. Jobs that are in a SUBMITTED, PENDING, or RUNNABLE state are cancelled and the job status is updated to FAILED.

> **Note**
>
> A PENDING job is cancelled after all dependency jobs are completed. Therefore, it might take longer than expected to cancel a job in PENDING status.
> When you try to cancel an array parent job in PENDING, AWS Batch attempts to cancel all child jobs. The array parent job is cancelled when all child jobs are completed.

Jobs that progressed to the STARTING or RUNNING state aren't cancelled. These jobs must be terminated with the TerminateJob or TerminateJobs operation.

AWS Batch reports the result for each job individually in the response. Jobs that were processed successfully are reported in the `successful` list. Jobs that encountered errors are reported in the `errors` list. The response returns an HTTP status code of `200` even when some jobs encountered errors, so check the `errors` list. Jobs that can't be found are treated as successfully processed.

> **Important**
>
> This operation requires `batch:CancelJob` permission for each job in the request. There is no separate `batch:CancelJobs` IAM action. If a caller's IAM policy grants `batch:CancelJob`, they can use both the singular CancelJob and bulk `CancelJobs` operations.

Request Syntax

```
POST /v1/canceljobs HTTP/1.1
Content-type: application/json

{
   "jobs": [ "string" ],
   "reason": "string"
```

```
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

jobs

An array of up to 50 AWS Batch job IDs of the jobs to cancel.

Type: Array of strings

Required: Yes

reason

A message to attach to the job that explains the reason for cancelling it. This message is returned by future DescribeJobs operations on the job. It is also recorded in the AWS Batch activity logs.

This parameter has a limit of 1024 characters.

Type: String

Required: Yes

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
   "errors": [
      {
         "code": "string",
         "job": "string",
```

```
            "message": "string"
        }
    ],
    "successful": [ "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

A list of `CancelJobsErrorDetail` items, one for each job that couldn't be cancelled. Each item includes the job ID along with a code and message that describe why the job wasn't cancelled.

Type: Array of CancelJobsErrorDetail objects

successful

A list of the job IDs whose cancellation request was accepted.

Type: Array of strings

Errors

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

Examples

In the following example or examples, the Authorization header contents (*[authorization-params]*) must be replaced with an AWS Signature Version 4 signature. For more information about creating these signatures, see Signature Version 4 Signing Process in the *AWS General Reference*.

You only need to learn how to sign HTTP requests if you intend to manually create them. When you use the AWS Command Line Interface (AWS CLI) or one of the AWS SDKs to make requests to AWS, these tools automatically sign the requests for you with the access key that you specify when you configure the tools. When you use these tools, you don't need to learn how to sign requests yourself.

Example

This example cancels the jobs with the specified job IDs.

Sample Request

```
POST /v1/canceljobs HTTP/1.1
Host: batch.us-east-1.amazonaws.com
Accept-Encoding: identity
Content-Length: [content-length]
Authorization: [authorization-params]
X-Amz-Date: 20161130T001258Z
User-Agent: aws-cli/1.11.22 Python/2.7.12 Darwin/16.1.0 botocore/1.4.79

{
  "jobs": [
    "1d828f65-7a4d-42e8-996d-3b900ed59dc4",
    "b3d0f26e-6d3a-4a5b-9c2e-7f4a1b2c3d4e"
  ],
  "reason": "Cancelling jobs."
}
```

Sample Response

```
HTTP/1.1 200 OK
Content-Type: application/json
Content-Length: [content-length]
Connection: keep-alive
```

```
Date: Wed, 30 Nov 2016 00:12:59 GMT
x-amzn-RequestId: [request-id]
X-Amzn-Trace-Id: [trace-id]
X-Cache: Miss from cloudfront
Via: 1.1 bfdd5909914586f5bc4851846228c27f.cloudfront.net (CloudFront)
X-Amz-Cf-Id: whn1dX1uTx34Lvao7-7ZdkDXEbCZ_sjn3v3hHVFgbo1ORJtXyeggSw==

{
  "successful": [
    "1d828f65-7a4d-42e8-996d-3b900ed59dc4"
  ],
  "errors": [
    {
      "job": "b3d0f26e-6d3a-4a5b-9c2e-7f4a1b2c3d4e",
      "code": "ServerException",
      "message": "Failed to read job state. Please retry this job."
    }
  ]
}
```

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for Kotlin
- AWS SDK for PHP V3
- AWS SDK for Python (Boto3)
- AWS SDK for Ruby V3

CreateComputeEnvironment

Creates an AWS Batch compute environment. You can create `MANAGED` or `UNMANAGED` compute environments. `MANAGED` compute environments can use Amazon EC2 or AWS Fargate resources. `UNMANAGED` compute environments can only use EC2 resources.

In a managed compute environment, AWS Batch manages the capacity and instance types of the compute resources within the environment. This is based on the compute resource specification that you define or the launch template that you specify when you create the compute environment. Either, you can choose to use EC2 On-Demand Instances and EC2 Spot Instances. Or, you can use Fargate and Fargate Spot capacity in your managed compute environment. You can optionally set a maximum price so that Spot Instances only launch when the Spot Instance price is less than a specified percentage of the On-Demand price.

In an unmanaged compute environment, you can manage your own EC2 compute resources and have flexibility with how you configure your compute resources. For example, you can use custom AMIs. However, you must verify that each of your AMIs meet the Amazon ECS container instance AMI specification. For more information, see container instance AMIs in the *Amazon Elastic Container Service Developer Guide*. After you created your unmanaged compute environment, you can use the DescribeComputeEnvironments operation to find the Amazon ECS cluster that's associated with it. Then, launch your container instances into that Amazon ECS cluster. For more information, see Launching an Amazon ECS container instance in the *Amazon Elastic Container Service Developer Guide*.

Note

AWS Batch doesn't automatically upgrade the AMIs in a compute environment after it's created. For more information on how to update a compute environment's AMI, see Updating compute environments in the *AWS Batch User Guide*.

Request Syntax

```
POST /v1/createcomputeenvironment HTTP/1.1
Content-type: application/json

{
   "computeEnvironmentName": "string",
   "computeResources": {
```

```
      "allocationStrategy": "string",
      "bidPercentage": number,
      "capacityTags": {
         "string" : "string"
      },
      "desiredvCpus": number,
      "ec2Configuration": [
         {
            "batchImageStatus": "string",
            "imageIdOverride": "string",
            "imageKubernetesVersion": "string",
            "imageType": "string"
         }
      ],
      "ec2KeyPair": "string",
      "imageId": "string",
      "instanceRole": "string",
      "instanceTypes": [ "string" ],
      "launchTemplate": {
         "launchTemplateId": "string",
         "launchTemplateName": "string",
         "overrides": [
            {
               "launchTemplateId": "string",
               "launchTemplateName": "string",
               "targetInstanceTypes": [ "string" ],
               "userdataType": "string",
               "version": "string"
            }
         ],
         "userdataType": "string",
         "version": "string"
      },
      "managedInstancesProvider": {
         "infrastructureOptimization": {
            "scaleInAfter": number
         },
         "infrastructureRoleArn": "string",
         "instanceLaunchTemplate": {
            "capacityOptionType": "string",
            "capacityReservations": {
               "reservationGroupArn": "string",
               "reservationPreference": "string"
            },
```

```
            "ec2InstanceProfileArn": "string",
            "fipsEnabled": boolean,
            "instanceMetadataTagsPropagation": boolean,
            "instanceRequirements": {
               "allowedInstanceTypes": [ "string" ]
            },
            "localStorageConfiguration": {
               "useLocalStorage": boolean
            },
            "monitoring": "string",
            "networkConfiguration": {
               "securityGroups": [ "string" ],
               "subnets": [ "string" ]
            },
            "storageConfiguration": {
               "storageSizeGiB": number
            }
         },
         "propagateTags": "string"
      },
      "maxvCpus": number,
      "minvCpus": number,
      "placementGroup": "string",
      "scalingPolicy": {
         "minScaleDownDelayMinutes": number
      },
      "securityGroupIds": [ "string" ],
      "spotIamFleetRole": "string",
      "subnets": [ "string" ],
      "tags": {
         "string" : "string"
      },
      "type": "string"
   },
   "context": "string",
   "ecsSettings": {
      "containerInsights": "string"
   },
   "eksConfiguration": {
      "accessEntry": {
         "desiredState": "string",
         "status": "string"
      },
      "eksClusterArn": "string",
```

```
      "kubernetesNamespace": "string"
   },
   "serviceRole": "string",
   "state": "string",
   "tags": {
      "string" : "string"
   },
   "type": "string",
   "unmanagedvCpus": number
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

computeEnvironmentName

The name for your compute environment. It can be up to 128 characters long. It can contain uppercase and lowercase letters, numbers, hyphens (-), and underscores (_).

Type: String

Required: Yes

computeResources

Details about the compute resources managed by the compute environment. This parameter is required for managed compute environments. For more information, see Compute Environments in the *AWS Batch User Guide*.

Type: ComputeResource object

Required: No

context

Reserved.

Type: String

Required: No

ecsSettings

The Amazon ECS settings for the compute environment. These settings control CloudWatch Container Insights collection for the compute environment.

Type: EcsSettings object

Required: No

eksConfiguration

The details for the Amazon EKS cluster that supports the compute environment.

> **Note**
>
> To create a compute environment that uses EKS resources, the caller must have permissions to call `eks:DescribeCluster`. Additional Amazon EKS permissions are required for AWS Batch to manage an access entry on the cluster; see Amazon EKS access entry authentication in the *AWS Batch User Guide*.

Type: EksConfiguration object

Required: No

serviceRole

The full Amazon Resource Name (ARN) of the IAM role that allows AWS Batch to make calls to other AWS services on your behalf. For more information, see AWS Batch service IAM role in the *AWS Batch User Guide*.

> **Important**
>
> If your account already created the AWS Batch service-linked role, that role is used by default for your compute environment unless you specify a different role here. If the AWS Batch service-linked role doesn't exist in your account, and no role is specified here, the service attempts to create the AWS Batch service-linked role in your account.
> This automatic service-linked role creation only applies to `MANAGED` compute environments. For `UNMANAGED` compute environments, you must explicitly specify a `serviceRole`.

If your specified role has a path other than `/`, then you must specify either the full role ARN (recommended) or prefix the role name with the path. For example, if a role with the name `bar` has a path of `/foo/`, specify `/foo/bar` as the role name. For more information, see Friendly names and paths in the *IAM User Guide*.

> **Note**
>
> Depending on how you created your AWS Batch service role, its ARN might contain the `service-role` path prefix. When you only specify the name of the service role, AWS Batch assumes that your ARN doesn't use the `service-role` path prefix. Because of this, we recommend that you specify the full ARN of your service role when you create compute environments.

Type: String

Required: No

state

The state of the compute environment. A compute environment must be created in the `ENABLED` state.

If the state is `ENABLED`, then the compute environment accepts jobs from a queue and can scale out automatically based on queues.

If the state is `ENABLED`, then the AWS Batch scheduler can attempt to place jobs from an associated job queue on the compute resources within the environment. If the compute environment is managed, then it can scale its instances out or in automatically, based on the job queue demand.

If the state is `DISABLED`, then the AWS Batch scheduler doesn't attempt to place jobs within the environment. Jobs in a `STARTING` or `RUNNING` state continue to progress normally. Managed compute environments in the `DISABLED` state don't scale out.

> **Note**
>
> Compute environments in a `DISABLED` state may continue to incur billing charges, for example, if they have running instances due to jobs that are still executing or a non-

zero `minvCpus` setting. To prevent additional charges, disable and delete the compute environment.

When an instance is idle, the instance scales down to the `minvCpus` value. However, the instance size doesn't change. For example, consider a `c5.8xlarge` instance with a `minvCpus` value of `4` and a `desiredvCpus` value of `36`. This instance doesn't scale down to a `c5.large` instance.

Type: String

Valid Values: `ENABLED | DISABLED`

Required: No

tags

The tags that you apply to the compute environment to help you categorize and organize your resources. Each tag consists of a key and an optional value. For more information, see Tagging AWS Resources in *AWS General Reference*.

These tags can be updated or removed using the TagResource and UntagResource API operations. These tags don't propagate to the underlying compute resources.

Type: String to string map

Map Entries: Maximum number of 50 items.

Key Length Constraints: Minimum length of 1. Maximum length of 128.

Value Length Constraints: Maximum length of 256.

Required: No

type

The type of the compute environment: `MANAGED` or `UNMANAGED`. For more information, see Compute Environments in the *AWS Batch User Guide*.

Type: String

Valid Values: `MANAGED | UNMANAGED`

Required: Yes

unmanagedvCpus

The maximum number of vCPUs for an unmanaged compute environment. This parameter is only used for fair-share scheduling to reserve vCPU capacity for new share identifiers. If this parameter isn't provided for a fair-share job queue, no vCPU capacity is reserved.

> **Note**
>
> This parameter is only supported when the `type` parameter is set to `UNMANAGED`.

Type: Integer

Required: No

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
   "computeEnvironmentArn": "string",
   "computeEnvironmentName": "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.

computeEnvironmentArn

The Amazon Resource Name (ARN) of the compute environment.

Type: String

computeEnvironmentName

The name of the compute environment. It can be up to 128 characters long. It can contain uppercase and lowercase letters, numbers, hyphens (-), and underscores (_).

Type: String

Errors

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

Examples

In the following example or examples, the Authorization header contents (*`[authorization-params]`*) must be replaced with an AWS Signature Version 4 signature. For more information about creating these signatures, see Signature Version 4 Signing Process in the *AWS General Reference*.

You only need to learn how to sign HTTP requests if you intend to manually create them. When you use the AWS Command Line Interface (AWS CLI) or one of the AWS SDKs to make requests to AWS, these tools automatically sign the requests for you with the access key that you specify when you configure the tools. When you use these tools, you don't need to learn how to sign requests yourself.

Example

This example creates a managed compute environment with specific C4 instance types that are launched on demand. The compute environment is called `C4OnDemand`.

Sample Request

```
POST /v1/createcomputeenvironment HTTP/1.1
Host: batch.us-east-1.amazonaws.com
```

```
Accept-Encoding: identity
Content-Length: [content-length]
Authorization: [authorization-params]
X-Amz-Date: 20161128T223128Z
User-Agent: aws-cli/1.11.21 Python/2.7.12 Darwin/16.1.0 botocore/1.4.78

{
  "computeEnvironmentName": "C4OnDemand",
  "state": "ENABLED",
  "type": "MANAGED",
  "computeResources": {
    "subnets": [
      "subnet-220c0e0a",
      "subnet-1a95556d",
      "subnet-978f6dce"
    ],
    "tags": {
      "Name": "Batch Instance - C4OnDemand",
      "Department": "Engineering"
    },
    "desiredvCpus": 48,
    "minvCpus": 0,
    "instanceTypes": [
      "c4.large",
      "c4.xlarge",
      "c4.2xlarge",
      "c4.4xlarge",
      "c4.8xlarge"
    ],
    "securityGroupIds": [
      "sg-cf5093b2"
    ],
    "instanceRole": "ecsInstanceRole",
    "maxvCpus": 128,
    "type": "EC2",
    "ec2KeyPair": "id_rsa"
  },
  "serviceRole": "arn:aws:iam::123456789012:role/AWSBatchServiceRole"
}
```

Sample Response

```
HTTP/1.1 200 OK
```

```
Date: Mon, 28 Nov 2016 22:31:28 GMT
Content-Type: application/json
Content-Length: [content-length]
Connection: keep-alive
x-amzn-RequestId: [request-id]
X-Amzn-Trace-Id: [trace-id]
X-Cache: Miss from cloudfront
Via: 1.1 7e587c722adb25336835ccb4e5814e4e.cloudfront.net (CloudFront)
X-Amz-Cf-Id: GwQRsxvmiuj1HYwbYq9MAEsQfJpN6BknGQlNX1jAd5qLQFXyHBwOUQ==

{
  "computeEnvironmentName": "C4OnDemand",
  "computeEnvironmentArn": "arn:aws:batch:us-east-1:123456789012:compute-environment/
C4OnDemand"
}
```

Example

This example creates a managed compute environment with the M4 instance type that's launched when the Spot Instance price less than or equal to 20% of the On-Demand price for the instance type. The compute environment is called `M4Spot`.

Sample Request

```
POST /v1/createcomputeenvironment HTTP/1.1
Host: batch.us-east-1.amazonaws.com
Accept-Encoding: identity
Content-Length: [content-length]
Authorization: [authorization-params]
X-Amz-Date: 20161128T223813Z
User-Agent: aws-cli/1.11.21 Python/2.7.12 Darwin/16.1.0 botocore/1.4.78

{
  "computeEnvironmentName": "M4Spot",
  "state": "ENABLED",
  "type": "MANAGED",
  "computeResources": {
    "subnets": [
      "subnet-220c0e0a",
      "subnet-1a95556d",
      "subnet-978f6dce"
    ],
    "type": "SPOT",
```

```
    "spotIamFleetRole": "arn:aws:iam::123456789012:role/aws-ec2-spot-fleet-role",
    "tags": {
      "Name": "Batch Instance - M4Spot",
      "Department": "Marketing"
    },
    "desiredvCpus": 4,
    "minvCpus": 0,
    "instanceTypes": [
      "m4"
    ],
    "securityGroupIds": [
      "sg-cf5093b2"
    ],
    "instanceRole": "ecsInstanceRole",
    "maxvCpus": 128,
    "bidPercentage": 20,
    "ec2KeyPair": "id_rsa"
  },
  "serviceRole": "arn:aws:iam::123456789012:role/AWSBatchServiceRole"
}
```

Sample Response

```
HTTP/1.1 200 OK
Date: Mon, 28 Nov 2016 22:38:16 GMT
Content-Type: application/json
Content-Length: [content-length]
Connection: keep-alive
x-amzn-RequestId: [request-id]
X-Amzn-Trace-Id: [trace-id]
X-Cache: Miss from cloudfront
Via: 1.1 8455edd9286a1292a39c993fdeccce65.cloudfront.net (CloudFront)
X-Amz-Cf-Id: 4mklLyUpygUko86fMNzPgA8_D64lSwPmG6iIKhAZkGpOp2e-3cKg_w==

{
  "computeEnvironmentName": "M4Spot",
  "computeEnvironmentArn": "arn:aws:batch:us-east-1:123456789012:compute-environment/
M4Spot"
}
```

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for Kotlin
- AWS SDK for PHP V3
- AWS SDK for Python (Boto3)
- AWS SDK for Ruby V3

CreateConsumableResource

Creates an AWS Batch consumable resource.

Request Syntax

```
POST /v1/createconsumableresource HTTP/1.1
Content-type: application/json

{
   "consumableResourceName": "string",
   "resourceType": "string",
   "tags": {
      "string" : "string"
   },
   "totalQuantity": number
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

consumableResourceName

The name of the consumable resource. Must be unique.

Type: String

Required: Yes

resourceType

Indicates whether the resource is available to be re-used after a job completes. Can be one of:

- `REPLENISHABLE` (default)
- `NON_REPLENISHABLE`

Type: String

Required: No

tags

The tags that you apply to the consumable resource to help you categorize and organize your resources. Each tag consists of a key and an optional value. For more information, see Tagging your AWS Batch resources.

Type: String to string map

Map Entries: Maximum number of 50 items.

Key Length Constraints: Minimum length of 1. Maximum length of 128.

Value Length Constraints: Maximum length of 256.

Required: No

totalQuantity

The total amount of the consumable resource that is available. Must be non-negative.

Type: Long

Required: No

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
   "consumableResourceArn": "string",
   "consumableResourceName": "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.

consumableResourceArn

The Amazon Resource Name (ARN) of the consumable resource.

Type: String

consumableResourceName

The name of the consumable resource.

Type: String

Errors

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for Kotlin

- AWS SDK for PHP V3
- AWS SDK for Python (Boto3)
- AWS SDK for Ruby V3

CreateJobQueue

Creates an AWS Batch job queue. When you create a job queue, you associate one or more compute environments to the queue and assign an order of preference for the compute environments.

You also set a priority to the job queue that determines the order that the AWS Batch scheduler places jobs onto its associated compute environments. For example, if a compute environment is associated with more than one job queue, the job queue with a higher priority is given preference for scheduling jobs to that compute environment.

Request Syntax

```
POST /v1/createjobqueue HTTP/1.1
Content-type: application/json

{
   "computeEnvironmentOrder": [
      {
         "computeEnvironment": "string",
         "order": number
      }
   ],
   "jobQueueName": "string",
   "jobQueueType": "string",
   "jobStateTimeLimitActions": [
      {
         "action": "string",
         "maxTimeSeconds": number,
         "reason": "string",
         "state": "string"
      }
   ],
   "priority": number,
   "schedulingPolicyArn": "string",
   "serviceEnvironmentOrder": [
      {
         "order": number,
         "serviceEnvironment": "string"
      }
   ],
   "state": "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.

computeEnvironmentOrder

The set of compute environments mapped to a job queue and their order relative to each other. The job scheduler uses this parameter to determine which compute environment runs a specific job. Compute environments must be in the `VALID` state before you can associate them with a job queue. You can associate up to three compute environments with a job queue. All of the compute environments must be either EC2 (`EC2` or `SPOT`) or Fargate (`FARGATE` or `FARGATE_SPOT`); EC2 and Fargate compute environments can't be mixed.

> **Note**
>
> All compute environments that are associated with a job queue must share the same architecture. AWS Batch doesn't support mixing compute environment architecture types in a single job queue.

Type: Array of ComputeEnvironmentOrder objects

Required: No

jobQueueName

The name of the job queue. It can be up to 128 letters long. It can contain uppercase and lowercase letters, numbers, hyphens (-), and underscores (_).

Type: String

Required: Yes

jobQueueType

The type of job queue. For service jobs that run on SageMaker Training, this value is `SAGEMAKER_TRAINING`. For regular container jobs, this value is `EKS`, `ECS`, or `ECS_FARGATE` depending on the compute environment.

Type: String

Valid Values: `EKS | ECS | ECS_FARGATE | SAGEMAKER_TRAINING | ECS_MANAGED_INSTANCES`

Required: No

jobStateTimeLimitActions

The set of actions that AWS Batch performs on jobs that remain at the head of the job queue in the specified state longer than specified times. AWS Batch will perform each action after `maxTimeSeconds` has passed. (**Note**: The minimum value for maxTimeSeconds is 600 (10 minutes) and its maximum value is 86,400 (24 hours).)

Type: Array of JobStateTimeLimitAction objects

Required: No

priority

The priority of the job queue. Job queues with a higher priority (or a higher integer value for the `priority` parameter) are evaluated first when associated with the same compute environment. Priority is determined in descending order. For example, a job queue with a priority value of `10` is given scheduling preference over a job queue with a priority value of `1`. All of the compute environments must be either EC2 (`EC2` or `SPOT`) or Fargate (`FARGATE` or `FARGATE_SPOT`); EC2 and Fargate compute environments can't be mixed.

Type: Integer

Required: Yes

schedulingPolicyArn

The Amazon Resource Name (ARN) of the fair-share scheduling policy. Job queues that don't have a fair-share scheduling policy are scheduled in a first-in, first-out (FIFO) model. After a job queue has a fair-share scheduling policy, it can be replaced but can't be removed.

The format is `aws:`*`Partition`*`:batch:`*`Region`*`:`*`Account`*`:scheduling-policy/`*`Name`* .

An example is `aws:aws:batch:us-west-2:123456789012:scheduling-policy/MySchedulingPolicy`.

A job queue without a fair-share scheduling policy is scheduled as a FIFO job queue and can't have a fair-share scheduling policy added. Jobs queues with a fair-share scheduling policy can have a maximum of 500 active share identifiers. When the limit has been reached, submissions of any jobs that add a new share identifier fail.

Type: String

Required: No

serviceEnvironmentOrder

A list of service environments that this job queue can use to allocate jobs. All serviceEnvironments must have the same type. A job queue can't have both a serviceEnvironmentOrder and a computeEnvironmentOrder field.

Type: Array of ServiceEnvironmentOrder objects

Required: No

state

The state of the job queue. If the job queue state is `ENABLED`, it is able to accept jobs. If the job queue state is `DISABLED`, new jobs can't be added to the queue, but jobs already in the queue can finish.

Type: String

Valid Values: `ENABLED | DISABLED`

Required: No

tags

The tags that you apply to the job queue to help you categorize and organize your resources. Each tag consists of a key and an optional value. For more information, see Tagging your AWS Batch resources in *AWS Batch User Guide*.

Type: String to string map

Map Entries: Maximum number of 50 items.

Key Length Constraints: Minimum length of 1. Maximum length of 128.

Value Length Constraints: Maximum length of 256.

Required: No

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
   "jobQueueArn": "string",
   "jobQueueName": "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.

jobQueueArn

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

Type: String

jobQueueName

The name of the job queue.

Type: String

Errors

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

Examples

In the following example or examples, the Authorization header contents (*[authorization-params]*) must be replaced with an AWS Signature Version 4 signature. For more information about creating these signatures, see Signature Version 4 Signing Process in the *AWS General Reference*.

You only need to learn how to sign HTTP requests if you intend to manually create them. When you use the AWS Command Line Interface (AWS CLI) or one of the AWS SDKs to make requests to AWS, these tools automatically sign the requests for you with the access key that you specify when you configure the tools. When you use these tools, you don't need to learn how to sign requests yourself.

Example

This example creates a job queue called `LowPriority` that uses the `M4Spot` compute environment.

Sample Request

```
POST /v1/createjobqueue HTTP/1.1
Host: batch.us-east-1.amazonaws.com
Accept-Encoding: identity
Content-Length: [content-length]
Authorization: [authorization-params]
X-Amz-Date: 20161128T234201Z
User-Agent: aws-cli/1.11.21 Python/2.7.12 Darwin/16.1.0 botocore/1.4.78

{
  "priority": 1,
  "state": "ENABLED",
  "computeEnvironmentOrder": [
    {
```

```
      "computeEnvironment": "M4Spot",
      "order": 1
    }
  ],
  "jobQueueName": "LowPriority"
}
```

Sample Response

```
HTTP/1.1 200 OK
Content-Type: application/json
Content-Length: [content-length]
Connection: keep-alive
Date: Mon, 28 Nov 2016 23:42:02 GMT
x-amzn-RequestId: [request-id]
X-Amzn-Trace-Id: [trace-id]
X-Cache: Miss from cloudfront
Via: 1.1 a44b4468444ef3ee67472bd5c5016098.cloudfront.net (CloudFront)
X-Amz-Cf-Id: bz9IuCM5FNkDfge5y-Zw7nFEjDdTHDYFwbEY2AKUqrt9l2XeKUcuyA==

{
  "jobQueueName": "LowPriority",
  "jobQueueArn": "arn:aws:batch:us-east-1:123456789012:job-queue/LowPriority"
}
```

Example

This example creates a job queue called `HighPriority` that uses the `C4OnDemand` compute environment with an order of 1 and the `M4Spot` compute environment with an order of 2.

Sample Request

```
POST /v1/createjobqueue HTTP/1.1
Host: batch.us-east-1.amazonaws.com
Accept-Encoding: identity
Content-Length: [content-length]
Authorization: [authorization-params]
X-Amz-Date: 20161128T234933Z
User-Agent: aws-cli/1.11.21 Python/2.7.12 Darwin/16.1.0 botocore/1.4.78

{
  "priority": 10,
```

```
  "state": "ENABLED",
  "computeEnvironmentOrder": [
    {
      "computeEnvironment": "C4OnDemand",
      "order": 1
    },
    {
      "computeEnvironment": "M4Spot",
      "order": 2
    }
  ],
  "jobQueueName": "HighPriority"
}
```

Sample Response

```
HTTP/1.1 200 OK
Date: Mon, 28 Nov 2016 23:49:34 GMT
Content-Type: application/json
Content-Length: [content-length]
Connection: keep-alive
x-amzn-RequestId: [request-id]
X-Amzn-Trace-Id: [trace-id]
X-Cache: Miss from cloudfront
Via: 1.1 e81bbcbc86832b655de5b9a19317ad01.cloudfront.net (CloudFront)
X-Amz-Cf-Id: 8NB20odDPMaKy9zHa6GPaGN_r562QsynDTRYPuhKwHSvQrMG70IHSQ==

{
  "jobQueueName": "HighPriority",
  "jobQueueArn": "arn:aws:batch:us-east-1:123456789012:job-queue/HighPriority"
}
```

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2

- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for Kotlin
- AWS SDK for PHP V3
- AWS SDK for Python (Boto3)
- AWS SDK for Ruby V3

CreateQuotaShare

Creates an AWS Batch quota share. Each quota share operates as a virtual queue with a configured compute capacity, resource sharing strategy, and borrow limits.

Request Syntax

```
POST /v1/createquotashare HTTP/1.1
Content-type: application/json

{
   "capacityLimits": [
      {
         "capacityUnit": "string",
         "maxCapacity": number
      }
   ],
   "jobQueue": "string",
   "preemptionConfiguration": {
      "inSharePreemption": "string"
   },
   "quotaShareName": "string",
   "resourceSharingConfiguration": {
      "borrowLimit": number,
      "strategy": "string"
   },
   "state": "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.

capacityLimits

A list that specifies the quantity and type of compute capacity allocated to the quota share.

Type: Array of QuotaShareCapacityLimit objects

Required: Yes

jobQueue

The AWS Batch job queue associated with the quota share. This can be the job queue name or ARN. A job queue must be in the `VALID` state before you can associate it with a quota share.

Type: String

Required: Yes

preemptionConfiguration

Specifies the preemption behavior for jobs in a quota share.

Type: QuotaSharePreemptionConfiguration object

Required: Yes

quotaShareName

The name of the quota share. It can be up to 128 characters long. It can contain uppercase and lowercase letters, numbers, hyphens (-), and underscores (_).

Type: String

Required: Yes

resourceSharingConfiguration

Specifies whether a quota share reserves, lends, or both lends and borrows idle compute capacity.

Type: QuotaShareResourceSharingConfiguration object

Required: Yes

state

The state of the quota share. If the quota share is `ENABLED`, it is able to accept jobs. If the quota share is `DISABLED`, new jobs won't be accepted but jobs already submitted can finish. The default state is `ENABLED`.

Type: String

Valid Values: `ENABLED | DISABLED`

Required: No

tags

The tags that you apply to the quota share to help you categorize and organize your resources. Each tag consists of a key and an optional value. For more information, see Tagging your AWS Batch resources in *AWS Batch User Guide*.

Type: String to string map

Map Entries: Maximum number of 50 items.

Key Length Constraints: Minimum length of 1. Maximum length of 128.

Value Length Constraints: Maximum length of 256.

Required: No

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
   "quotaShareArn": "string",
   "quotaShareName": "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.

quotaShareArn

The Amazon Resource Name (ARN) of the quota share.

Type: String

quotaShareName

The name of the quota share.

Type: String

Errors

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

Examples

In the following example or examples, the Authorization header contents (*`[authorization-params]`*) must be replaced with an AWS Signature Version 4 signature. For more information about creating these signatures, see Signature Version 4 Signing Process in the *AWS General Reference*.

You only need to learn how to sign HTTP requests if you intend to manually create them. When you use the AWS Command Line Interface (AWS CLI) or one of the AWS SDKs to make requests to AWS, these tools automatically sign the requests for you with the access key that you specify when you configure the tools. When you use these tools, you don't need to learn how to sign requests yourself.

Example

This example creates a quota share called `TeamAShare` associated with the `HighPriority` job queue, with capacity limits of 10 `ml.m5.large` and 5 `ml.m5.xlarge` instances, and a borrow limit of 200%.

Sample Request

```
POST /v1/createquotashare HTTP/1.1
Host: batch.us-east-1.amazonaws.com
Accept-Encoding: identity
Content-Length: [content-length]
Authorization: [authorization-params]
X-Amz-Date: 20250801T001258Z
User-Agent: aws-cli/2.27.33 Python/3.13.4 Darwin/24.3.0

{
  "quotaShareName": "TeamAShare",
  "jobQueue": "HighPriority",
  "capacityLimits": [
    {
      "maxCapacity": 10,
      "capacityUnit": "ml.m5.large"
    },
    {
      "maxCapacity": 5,
      "capacityUnit": "ml.m5.xlarge"
    }
  ],
  "resourceSharingConfiguration": {
    "strategy": "LEND_AND_BORROW",
    "borrowLimit": 200
  },
  "preemptionConfiguration": {
    "inSharePreemption": "ENABLED"
  },
  "state": "ENABLED",
  "tags": {
    "Department": "Engineering"
  }
}
```

Sample Response

```
HTTP/1.1 200 OK
Content-Type: application/json
Content-Length: [content-length]
Connection: keep-alive
Date: Fri, 01 Aug 2025 00:12:59 GMT
```

```
x-amzn-RequestId: [request-id]
X-Amzn-Trace-Id: [trace-id]
X-Cache: Miss from cloudfront
Via: 1.1 example7k9m3n8q4r2w5x1z6c4vexample.cloudfront.net (CloudFront)
X-Amz-Cf-Id: whn1dX1uTx34Lvao7-7ZdkDXEbCZ_sjn3v3hHVFgbo1ORJtXyexample

{
  "quotaShareName": "TeamAShare",
  "quotaShareArn": "arn:aws:batch:us-east-1:123456789012:job-queue/HighPriority/quota-
share/TeamAShare"
}
```

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for Kotlin
- AWS SDK for PHP V3
- AWS SDK for Python (Boto3)
- AWS SDK for Ruby V3

CreateSchedulingPolicy

Creates an AWS Batch scheduling policy.

Request Syntax

```
POST /v1/createschedulingpolicy HTTP/1.1
Content-type: application/json

{
   "fairsharePolicy": {
      "computeReservation": number,
      "shareDecaySeconds": number,
      "shareDistribution": [
         {
            "shareIdentifier": "string",
            "weightFactor": number
         }
      ]
   },
   "name": "string",
   "quotaSharePolicy": {
      "idleResourceAssignmentStrategy": "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.

fairsharePolicy

The fair-share scheduling policy details. Only one of fairsharePolicy or quotaSharePolicy can be set. Once set, this policy type cannot be removed or changed to a quotaSharePolicy.

Type: FairsharePolicy object

Required: No

name

The name of the fair-share scheduling policy. It can be up to 128 letters long. It can contain uppercase and lowercase letters, numbers, hyphens (-), and underscores (_).

Type: String

Required: Yes

quotaSharePolicy

The quota share scheduling policy details. Only one of fairsharePolicy or quotaSharePolicy can be set. Once set, this policy type cannot be removed or changed to a fairSharePolicy.

Type: QuotaSharePolicy object

Required: No

tags

The tags that you apply to the scheduling policy to help you categorize and organize your resources. Each tag consists of a key and an optional value. For more information, see Tagging AWS Resources in *AWS General Reference*.

These tags can be updated or removed using the TagResource and UntagResource API operations.

Type: String to string map

Map Entries: Maximum number of 50 items.

Key Length Constraints: Minimum length of 1. Maximum length of 128.

Value Length Constraints: Maximum length of 256.

Required: No

Response Syntax

```
HTTP/1.1 200
```

```
Content-type: application/json

{
   "arn": "string",
   "name": "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

The Amazon Resource Name (ARN) of the scheduling policy. The format is `aws:Partition:batch:Region:Account:scheduling-policy/Name` . For example, `aws:aws:batch:us-west-2:123456789012:scheduling-policy/MySchedulingPolicy`.

Type: String

name

The name of the scheduling policy.

Type: String

Errors

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

Examples

In the following example or examples, the Authorization header contents (*[authorization-params]*) must be replaced with an AWS Signature Version 4 signature. For more information about creating these signatures, see Signature Version 4 Signing Process in the *AWS General Reference*.

You only need to learn how to sign HTTP requests if you intend to manually create them. When you use the AWS Command Line Interface (AWS CLI) or one of the AWS SDKs to make requests to AWS, these tools automatically sign the requests for you with the access key that you specify when you configure the tools. When you use these tools, you don't need to learn how to sign requests yourself.

Example

This example creates a scheduling policy with the specified share identifiers and share identifier prefixes.

Sample Request

```
POST /v1/createschedulingpolicy HTTP/1.1
Host: batch.us-east-1.amazonaws.com
Accept-Encoding: identity
User-Agent: aws-cli/1.20.21 Python/3.6.9 Linux/4.4.0-19041-Microsoft botocore/1.21.21
X-Amz-Date: 20210928T231724Z
X-Amz-Security-Token: [security-token]
Authorization: [authorization-params]
Content-Length: [content-length]

{
  "name": "ExampleFairSharePolicy",
  "fairsharePolicy": {
    "shareDecaySeconds": 3600,
    "computeReservation": 1,
    "shareDistribution": [
      {
        "shareIdentifier": "A1*",
        "weightFactor": 0.1
      },
      {
        "shareIdentifier": "A2",
        "weightFactor": 0.2
```

```
      },
      {
        "shareIdentifier": "B*",
        "weightFactor": 0.8
      },
      {
        "shareIdentifier": "C",
        "weightFactor": 1.2
      },
      {
        "shareIdentifier": "D*",
        "weightFactor": 1.5
      },
      {
        "shareIdentifier": "E",
        "weightFactor": 1.8
      }
    ]
  },
  "tags": {
    "Hot": "Dog",
    "Beef": "Brisket",
    "Pork": "Ribs",
    "Department": "Engineering"
  }
}
```

Sample Response

```
HTTP/1.1 200 OK
Date: Tue, 28 Sep 2021 23:17:50 GMT
Content-Type: application/json
Content-Length: [content-length]
x-amzn-RequestId: [request-id]
Access-Control-Allow-Origin: *
x-amz-apigw-id: [apigw-id]
Access-Control-Expose-Headers: X-amzn-errortype,X-amzn-requestid,X-amzn-errormessage,X-
amzn-trace-id,X-amz-apigw-id,date
X-Amzn-Trace-Id: [trace-id]
Connection: keep-alive

{
   "schedulingPolicies": [{
```

```
      "arn": "arn:aws:batch:us-east-1:123456789012:scheduling-policy/
ExampleFairSharePolicy"
   }]
}
```

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for Kotlin
- AWS SDK for PHP V3
- AWS SDK for Python (Boto3)
- AWS SDK for Ruby V3

CreateServiceEnvironment

Creates a service environment for running service jobs. Service environments define capacity limits for specific service types such as SageMaker Training jobs.

Request Syntax

```
POST /v1/createserviceenvironment HTTP/1.1
Content-type: application/json

{
   "capacityLimits": [
      {
         "capacityUnit": "string",
         "maxCapacity": number
      }
   ],
   "serviceEnvironmentName": "string",
   "serviceEnvironmentType": "string",
   "state": "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.

capacityLimits

The capacity limits for the service environment. The number of instances a job consumes is the total number of instances requested in the submit training job request resource configuration.

Type: Array of CapacityLimit objects

Required: Yes

serviceEnvironmentName

The name for the service environment. It can be up to 128 characters long and can contain letters, numbers, hyphens (-), and underscores (_).

Type: String

Required: Yes

serviceEnvironmentType

The type of service environment. For SageMaker Training jobs, specify `SAGEMAKER_TRAINING`.

Type: String

Valid Values: `SAGEMAKER_TRAINING`

Required: Yes

state

The state of the service environment. Valid values are `ENABLED` and `DISABLED`. The default value is `ENABLED`.

Type: String

Valid Values: `ENABLED | DISABLED`

Required: No

tags

The tags that you apply to the service environment to help you categorize and organize your resources. Each tag consists of a key and an optional value. For more information, see Tagging your AWS Batch resources.

Type: String to string map

Map Entries: Maximum number of 50 items.

Key Length Constraints: Minimum length of 1. Maximum length of 128.

Value Length Constraints: Maximum length of 256.

Required: No

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
   "serviceEnvironmentArn": "string",
   "serviceEnvironmentName": "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.

serviceEnvironmentArn

The Amazon Resource Name (ARN) of the service environment.

Type: String

serviceEnvironmentName

The name of the service environment.

Type: String

Errors

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

Examples

In the following example or examples, the Authorization header contents (*[authorization-params]*) must be replaced with an AWS Signature Version 4 signature. For more information about creating these signatures, see Signature Version 4 Signing Process in the *AWS General Reference*.

You only need to learn how to sign HTTP requests if you intend to manually create them. When you use the AWS Command Line Interface (AWS CLI) or one of the AWS SDKs to make requests to AWS, these tools automatically sign the requests for you with the access key that you specify when you configure the tools. When you use these tools, you don't need to learn how to sign requests yourself.

Example

This example creates a service environment for SageMaker Training jobs.

Sample Request

```
POST /v1/createserviceenvironment HTTP/1.1
Host: batch.us-east-1.amazonaws.com
Accept-Encoding: identity
Content-Length: [content-length]
Authorization: [authorization-params]
X-Amz-Date: 20250801T001258Z
User-Agent: aws-cli/2.27.33 Python/3.13.4 Darwin/24.3.0

{
  "serviceEnvironmentName": "SageMakerTrainingEnv",
  "serviceEnvironmentType": "SAGEMAKER_TRAINING",
  "capacityLimits": [
    {
      "maxCapacity": 50,
      "capacityUnit": "NUM_INSTANCES"
    }
  ]
}
```

Sample Response

```
HTTP/1.1 200 OK
```

```
Content-Type: application/json
Content-Length: [content-length]
Connection: keep-alive
Date: Fri, 01 Aug 2025 00:12:59 GMT
x-amzn-RequestId: [request-id]
X-Amzn-Trace-Id: [trace-id]
X-Cache: Miss from cloudfront
Via: 1.1 example7k9m3n8q4r2w5x1z6c4vexample.cloudfront.net (CloudFront)
X-Amz-Cf-Id: whn1dX1uTx34Lvao7-7ZdkDXEbCZ_sjn3v3hHVFgbo1ORJtXyexample

{
  "serviceEnvironmentName": "SageMakerTrainingEnv",
  "serviceEnvironmentArn": "arn:aws:batch:us-east-1:123456789012:service-environment/
SageMakerTrainingEnv"
}
```

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for Kotlin
- AWS SDK for PHP V3
- AWS SDK for Python (Boto3)
- AWS SDK for Ruby V3

DeleteComputeEnvironment

Deletes an AWS Batch compute environment.

Before you can delete a compute environment, you must set its state to `DISABLED` with the UpdateComputeEnvironment API operation and disassociate it from any job queues with the UpdateJobQueue API operation. Compute environments that use AWS Fargate resources must terminate all active jobs on that compute environment before deleting the compute environment. If this isn't done, the compute environment enters an invalid state.

Request Syntax

```
POST /v1/deletecomputeenvironment HTTP/1.1
Content-type: application/json

{
   "computeEnvironment": "string"
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

computeEnvironment

The name or Amazon Resource Name (ARN) of the compute environment to delete.

Type: String

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

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

Examples

In the following example or examples, the Authorization header contents (*`[authorization-params]`*) must be replaced with an AWS Signature Version 4 signature. For more information about creating these signatures, see Signature Version 4 Signing Process in the *AWS General Reference*.

You only need to learn how to sign HTTP requests if you intend to manually create them. When you use the AWS Command Line Interface (AWS CLI) or one of the AWS SDKs to make requests to AWS, these tools automatically sign the requests for you with the access key that you specify when you configure the tools. When you use these tools, you don't need to learn how to sign requests yourself.

Example

This example deletes the P3OnDemand compute environment.

Sample Request

```
POST /v1/deletecomputeenvironment HTTP/1.1
```

```
Host: batch.us-east-1.amazonaws.com
Accept-Encoding: identity
Content-Length: [content-length]
Authorization: [authorization-params]
X-Amz-Date: 20161128T202219Z
User-Agent: aws-cli/1.11.21 Python/2.7.12 Darwin/16.1.0 botocore/1.4.78

{
  "computeEnvironment": "P3OnDemand"
}
```

Sample Response

```
HTTP/1.1 200 OK
Content-Type: application/json
Content-Length: 2
Connection: keep-alive
Date: Mon, 28 Nov 2016 20:22:20 GMT
x-amzn-RequestId: [request-id]
X-Amzn-Trace-Id: [trace-id]
X-Cache: Miss from cloudfront
Via: 1.1 b63769e2d89c89274acd908e4bfcb9f4.cloudfront.net (CloudFront)
X-Amz-Cf-Id: mqHP9krdcbSbT0pivub4bJEM0_XCTTfENz0xPwwye-USu1CVGlj-nw==

{}
```

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for Kotlin
- AWS SDK for PHP V3

- AWS SDK for Python (Boto3)
- AWS SDK for Ruby V3

DeleteConsumableResource

Deletes the specified consumable resource.

Request Syntax

```
POST /v1/deleteconsumableresource HTTP/1.1
Content-type: application/json

{
   "consumableResource": "string"
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

consumableResource

The name or ARN of the consumable resource that will be deleted.

Type: String

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

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for Kotlin
- AWS SDK for PHP V3
- AWS SDK for Python (Boto3)
- AWS SDK for Ruby V3

DeleteJobQueue

Deletes the specified job queue. You must first disable submissions for a queue with the UpdateJobQueue operation. All jobs in the queue are eventually terminated when you delete a job queue.

It's not necessary to disassociate compute environments from a queue before submitting a `DeleteJobQueue` request.

Request Syntax

```
POST /v1/deletejobqueue HTTP/1.1
Content-type: application/json

{
   "jobQueue": "string"
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

jobQueue

The short name or full Amazon Resource Name (ARN) of the queue to delete.

Type: String

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

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

Examples

In the following example or examples, the Authorization header contents (*`[authorization-params]`*) must be replaced with an AWS Signature Version 4 signature. For more information about creating these signatures, see Signature Version 4 Signing Process in the *AWS General Reference*.

You only need to learn how to sign HTTP requests if you intend to manually create them. When you use the AWS Command Line Interface (AWS CLI) or one of the AWS SDKs to make requests to AWS, these tools automatically sign the requests for you with the access key that you specify when you configure the tools. When you use these tools, you don't need to learn how to sign requests yourself.

Example

This example deletes the GPGPU job queue.

Sample Request

```
POST /v1/deletejobqueue HTTP/1.1
```

```
Host: batch.us-east-1.amazonaws.com
Accept-Encoding: identity
Content-Length: [content-length]
Authorization: [authorization-params]
X-Amz-Date: 20161128T201857Z
User-Agent: aws-cli/1.11.21 Python/2.7.12 Darwin/16.1.0 botocore/1.4.78

{
  "jobQueue": "GPGPU"
}
```

Sample Response

```
HTTP/1.1 200 OK
Content-Type: application/json
Content-Length: 2
Connection: keep-alive
Date: Mon, 28 Nov 2016 20:18:57 GMT
x-amzn-RequestId: [request-id]
X-Amzn-Trace-Id: [trace-id]
X-Cache: Miss from cloudfront
Via: 1.1 56908f89e8d17ba579c0607313114955.cloudfront.net (CloudFront)
X-Amz-Cf-Id: UnpbX7PjdrV3N-Y79pD6eV3DfqYUXdEx3HAI9VYhUZ8h7yRBi5_ZVQ==

{}
```

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for Kotlin
- AWS SDK for PHP V3

- AWS SDK for Python (Boto3)
- AWS SDK for Ruby V3

DeleteQuotaShare

Deletes the specified quota share. You must first disable submissions for the share by updating the state to `DISABLED` using the UpdateQuotaShare operation. All jobs in the share are eventually terminated when you delete a quota share.

Request Syntax

```
POST /v1/deletequotashare HTTP/1.1
Content-type: application/json

{
   "quotaShareArn": "string"
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

quotaShareArn

The Amazon Resource Name (ARN) of the quota share.

Type: String

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

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

Examples

In the following example or examples, the Authorization header contents (*[authorization-params]*) must be replaced with an AWS Signature Version 4 signature. For more information about creating these signatures, see Signature Version 4 Signing Process in the *AWS General Reference*.

You only need to learn how to sign HTTP requests if you intend to manually create them. When you use the AWS Command Line Interface (AWS CLI) or one of the AWS SDKs to make requests to AWS, these tools automatically sign the requests for you with the access key that you specify when you configure the tools. When you use these tools, you don't need to learn how to sign requests yourself.

Example

This example deletes the specified quota share.

Sample Request

```
POST /v1/deletequotashare HTTP/1.1
Host: batch.us-east-1.amazonaws.com
Accept-Encoding: identity
Content-Length: [content-length]
Authorization: [authorization-params]
X-Amz-Date: 20250801T001258Z
```

```
User-Agent: aws-cli/2.27.33 Python/3.13.4 Darwin/24.3.0

{
  "quotaShareArn": "arn:aws:batch:us-east-1:123456789012:job-queue/HighPriority/quota-
share/TeamAShare"
}
```

Sample Response

```
HTTP/1.1 200 OK
Content-Type: application/json
Content-Length: 2
Connection: keep-alive
Date: Fri, 01 Aug 2025 00:12:59 GMT
x-amzn-RequestId: [request-id]
X-Amzn-Trace-Id: [trace-id]
X-Cache: Miss from cloudfront
Via: 1.1 example7k9m3n8q4r2w5x1z6c4vexample.cloudfront.net (CloudFront)
X-Amz-Cf-Id: whn1dX1uTx34Lvao7-7ZdkDXEbCZ_sjn3v3hHVFgbo1ORJtXyexample

{}
```

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for Kotlin
- AWS SDK for PHP V3
- AWS SDK for Python (Boto3)
- AWS SDK for Ruby V3

DeleteSchedulingPolicy

Deletes the specified scheduling policy.

You can't delete a scheduling policy that's used in any job queues.

Request Syntax

```
POST /v1/deleteschedulingpolicy 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

The Amazon Resource Name (ARN) of the scheduling policy to delete.

Type: String

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

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

Examples

In the following example or examples, the Authorization header contents (*[authorization-params]*) must be replaced with an AWS Signature Version 4 signature. For more information about creating these signatures, see Signature Version 4 Signing Process in the *AWS General Reference*.

You only need to learn how to sign HTTP requests if you intend to manually create them. When you use the AWS Command Line Interface (AWS CLI) or one of the AWS SDKs to make requests to AWS, these tools automatically sign the requests for you with the access key that you specify when you configure the tools. When you use these tools, you don't need to learn how to sign requests yourself.

Example

This example deletes a scheduling policy with the specified ARN.

Sample Request

```
POST /v1/deleteschedulingpolicy HTTP/1.1
Host: batch.us-east-1.amazonaws.com
Accept-Encoding: identity
User-Agent: aws-cli/1.20.21 Python/3.6.9 Linux/4.4.0-19041-Microsoft botocore/1.21.21
X-Amz-Date: 20210929T005824Z
X-Amz-Security-Token: [security-token]
```

```
Authorization: [authorization-params]
Content-Length: [content-length]

{
  "arn": "arn:aws:batch:us-east-1:123456789012:scheduling-policy/
ExampleFairSharePolicy2"
}
```

Sample Response

```
HTTP/1.1 200 OK
Date: Wed, 29 Sep 2021 00:58:26 GMT
Content-Type: application/json
Content-Length: 2
x-amzn-RequestId: [request-id]
Access-Control-Allow-Origin: *
x-amz-apigw-id: [apigw-id]
Access-Control-Expose-Headers: X-amzn-errortype,X-amzn-requestid,X-amzn-errormessage,X-
amzn-trace-id,X-amz-apigw-id,date
X-Amzn-Trace-Id: [trace-id]
Connection: keep-alive

{}
```

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for Kotlin
- AWS SDK for PHP V3
- AWS SDK for Python (Boto3)

- AWS SDK for Ruby V3

DeleteServiceEnvironment

Deletes a Service environment. Before you can delete a service environment, you must first set its state to `DISABLED` with the `UpdateServiceEnvironment` API operation and disassociate it from any job queues with the `UpdateJobQueue` API operation.

Request Syntax

```
POST /v1/deleteserviceenvironment HTTP/1.1
Content-type: application/json

{
   "serviceEnvironment": "string"
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

serviceEnvironment

The name or ARN of the service environment to delete.

Type: String

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

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

Examples

In the following example or examples, the Authorization header contents (*[authorization-params]*) must be replaced with an AWS Signature Version 4 signature. For more information about creating these signatures, see Signature Version 4 Signing Process in the *AWS General Reference*.

You only need to learn how to sign HTTP requests if you intend to manually create them. When you use the AWS Command Line Interface (AWS CLI) or one of the AWS SDKs to make requests to AWS, these tools automatically sign the requests for you with the access key that you specify when you configure the tools. When you use these tools, you don't need to learn how to sign requests yourself.

Example

This example the specified service environment.

Sample Request

```
POST /v1/canceljob HTTP/1.1
Host: batch.us-east-1.amazonaws.com
Accept-Encoding: identity
Content-Length: [content-length]
Authorization: [authorization-params]
X-Amz-Date: 20250801T001258Z
```

```
User-Agent: aws-cli/2.27.33 Python/3.13.4 Darwin/24.3.0

{
  "serviceEnvironment": "ExampleServiceEnvironment"
}
```

Sample Response

```
HTTP/1.1 200 OK
Content-Type: application/json
Content-Length: 2
Connection: keep-alive
Date: Fri, 1 Aug 2025 00:12:59 GMT
x-amzn-RequestId: [request-id]
X-Amzn-Trace-Id: [trace-id]
X-Cache: Miss from cloudfront
Via: 1.1 bfdd5909914586f5bc4851846228c27f.cloudfront.net (CloudFront)
X-Amz-Cf-Id: whn1dX1uTx34Lvao7-7ZdkDXEbCZ_sjn3v3hHVFgbo1ORJtXyeggSw==

{}
```

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for Kotlin
- AWS SDK for PHP V3
- AWS SDK for Python (Boto3)
- AWS SDK for Ruby V3

DeregisterJobDefinition

Deregisters an AWS Batch job definition. Job definitions are permanently deleted after 180 days.

Request Syntax

```
POST /v1/deregisterjobdefinition HTTP/1.1
Content-type: application/json

{
   "jobDefinition": "string"
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

jobDefinition

The name and revision (`name:revision`) or full Amazon Resource Name (ARN) of the job definition to deregister.

Type: String

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

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

Examples

In the following example or examples, the Authorization header contents (*`[authorization-params]`*) must be replaced with an AWS Signature Version 4 signature. For more information about creating these signatures, see Signature Version 4 Signing Process in the *AWS General Reference*.

You only need to learn how to sign HTTP requests if you intend to manually create them. When you use the AWS Command Line Interface (AWS CLI) or one of the AWS SDKs to make requests to AWS, these tools automatically sign the requests for you with the access key that you specify when you configure the tools. When you use these tools, you don't need to learn how to sign requests yourself.

Example

This example deregisters a job definition called `sleep10`.

Sample Request

```
POST /v1/deregisterjobdefinition HTTP/1.1
Host: batch.us-east-1.amazonaws.com
Accept-Encoding: identity
Content-Length: [content-length]
Authorization: [authorization-params]
X-Amz-Date: 20161128T215745Z
```

```
User-Agent: aws-cli/1.11.21 Python/2.7.12 Darwin/16.1.0 botocore/1.4.78

{
  "jobDefinition": "sleep10:1"
}
```

Sample Response

```
HTTP/1.1 200 OK
Content-Type: application/json
Content-Length: 2
Connection: keep-alive
Date: Mon, 28 Nov 2016 21:57:45 GMT
x-amzn-RequestId: [request-id]
X-Amzn-Trace-Id: [trace-id]
X-Cache: Miss from cloudfront
Via: 1.1 e892630891779ff1ccadccf205a776f3.cloudfront.net (CloudFront)
X-Amz-Cf-Id: wKAY_NOTbvY8PFcmo1aGja0xqGLxsTJgEtuc1KosPYAPYL8icYwvKw==

{}
```

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for Kotlin
- AWS SDK for PHP V3
- AWS SDK for Python (Boto3)
- AWS SDK for Ruby V3

DescribeComputeEnvironments

Describes one or more of your compute environments.

If you're using an unmanaged compute environment, you can use the `DescribeComputeEnvironment` operation to determine the `ecsClusterArn` that you launch your Amazon ECS container instances into.

Request Syntax

```
POST /v1/describecomputeenvironments HTTP/1.1
Content-type: application/json

{
   "computeEnvironments": [ "string" ],
   "maxResults": number,
   "nextToken": "string"
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

computeEnvironments

A list of up to 100 compute environment names or full Amazon Resource Name (ARN) entries.

Type: Array of strings

Required: No

maxResults

The maximum number of cluster results returned by `DescribeComputeEnvironments` in paginated output. When this parameter is used, `DescribeComputeEnvironments` only returns `maxResults` results in a single page along with a `nextToken` response element. The remaining results of the initial request can be seen by sending another `DescribeComputeEnvironments` request with the returned `nextToken` value. This value can

be between 1 and 100. If this parameter isn't used, then `DescribeComputeEnvironments` returns up to 100 results and a `nextToken` value if applicable.

Type: Integer

Required: No

nextToken

The `nextToken` value returned from a previous paginated `DescribeComputeEnvironments` request where `maxResults` was used and the results exceeded the value of that parameter. Pagination continues from the end of the previous results that returned the `nextToken` value. This value is `null` when there are no more results to return.

> **Note**
>
> Treat this token as an opaque identifier that's only used to retrieve the next items in a list and not for other programmatic purposes.

Type: String

Required: No

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
   "computeEnvironments": [
      {
         "computeEnvironmentArn": "string",
         "computeEnvironmentName": "string",
         "computeResources": {
            "allocationStrategy": "string",
            "bidPercentage": number,
            "capacityTags": {
               "string" : "string"
            },
            "desiredvCpus": number,
            "ec2Configuration": [
```

```
            {
               "batchImageStatus": "string",
               "imageIdOverride": "string",
               "imageKubernetesVersion": "string",
               "imageType": "string"
            }
         ],
         "ec2KeyPair": "string",
         "imageId": "string",
         "instanceRole": "string",
         "instanceTypes": [ "string" ],
         "launchTemplate": {
            "launchTemplateId": "string",
            "launchTemplateName": "string",
            "overrides": [
               {
                  "launchTemplateId": "string",
                  "launchTemplateName": "string",
                  "targetInstanceTypes": [ "string" ],
                  "userdataType": "string",
                  "version": "string"
               }
            ],
            "userdataType": "string",
            "version": "string"
         },
         "managedInstancesProvider": {
            "infrastructureOptimization": {
               "scaleInAfter": number
            },
            "infrastructureRoleArn": "string",
            "instanceLaunchTemplate": {
               "capacityOptionType": "string",
               "capacityReservations": {
                  "reservationGroupArn": "string",
                  "reservationPreference": "string"
               },
               "ec2InstanceProfileArn": "string",
               "fipsEnabled": boolean,
               "instanceMetadataTagsPropagation": boolean,
               "instanceRequirements": {
                  "allowedInstanceTypes": [ "string" ]
               },
               "localStorageConfiguration": {
```

```
                  "useLocalStorage": boolean
               },
               "monitoring": "string",
               "networkConfiguration": {
                  "securityGroups": [ "string" ],
                  "subnets": [ "string" ]
               },
               "storageConfiguration": {
                  "storageSizeGiB": number
               }
            },
            "propagateTags": "string"
         },
         "maxvCpus": number,
         "minvCpus": number,
         "placementGroup": "string",
         "scalingPolicy": {
            "minScaleDownDelayMinutes": number
         },
         "securityGroupIds": [ "string" ],
         "spotIamFleetRole": "string",
         "subnets": [ "string" ],
         "tags": {
            "string" : "string"
         },
         "type": "string"
      },
      "containerOrchestrationType": "string",
      "context": "string",
      "ecsClusterArn": "string",
      "ecsSettings": {
         "containerInsights": "string"
      },
      "eksConfiguration": {
         "accessEntry": {
            "desiredState": "string",
            "status": "string"
         },
         "eksClusterArn": "string",
         "kubernetesNamespace": "string"
      },
      "serviceRole": "string",
      "state": "string",
      "status": "string",
```

```
         "statusReason": "string",
         "tags": {
            "string" : "string"
         },
         "type": "string",
         "unmanagedvCpus": number,
         "updatePolicy": {
            "jobExecutionTimeoutMinutes": number,
            "terminateJobsOnUpdate": boolean
         },
         "uuid": "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.

computeEnvironments

The list of compute environments.

Type: Array of ComputeEnvironmentDetail objects

nextToken

The `nextToken` value to include in a future `DescribeComputeEnvironments` request. When the results of a `DescribeComputeEnvironments` request exceed `maxResults`, this value can be used to retrieve the next page of results. This value is `null` when there are no more results to return.

Type: String

Errors

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

Examples

In the following example or examples, the Authorization header contents (*[authorization-params]*) must be replaced with an AWS Signature Version 4 signature. For more information about creating these signatures, see Signature Version 4 Signing Process in the *AWS General Reference*.

You only need to learn how to sign HTTP requests if you intend to manually create them. When you use the AWS Command Line Interface (AWS CLI) or one of the AWS SDKs to make requests to AWS, these tools automatically sign the requests for you with the access key that you specify when you configure the tools. When you use these tools, you don't need to learn how to sign requests yourself.

Example

This example describes the P30nDemand compute environment.

Sample Request

```
POST /v1/describecomputeenvironments HTTP/1.1
Host: batch.us-east-1.amazonaws.com
Accept-Encoding: identity
Content-Length: [content-length]
Authorization: AUTHPARAMS
X-Amz-Date: 20161128T193355Z
```

```
User-Agent: aws-cli/1.11.21 Python/2.7.12 Darwin/16.1.0 botocore/1.4.78

{
  "computeEnvironments": [
    "P3OnDemand"
  ]
}
```

Sample Response

```
HTTP/1.1 200 OK
Content-Type: application/json
Content-Length: [content-length]
Connection: keep-alive
Date: Mon, 28 Nov 2016 19:33:56 GMT
x-amzn-RequestId: [request-id]
X-Amzn-Trace-Id: [trace-id]
X-Cache: Miss from cloudfront
Via: 1.1 56908f89e8d17ba579c0607313114955.cloudfront.net (CloudFront)
X-Amz-Cf-Id: FbgslaatWhj_yGhfkSCTPpPtjiuVuFOBns-kK5EsaasYQC5p2FnpiQ==

{
  "computeEnvironments": [{
    "computeEnvironmentName": "P3OnDemand",
    "computeEnvironmentArn": "arn:aws:batch:us-east-1:123456789012:compute-environment/
P3OnDemand",
    "ecsClusterArn": "arn:aws:ecs:us-east-1:123456789012:cluster/
P3OnDemand_Batch_2c06f29d-d1fe-3a49-879d-42394c86effc",
    "type": "MANAGED",
    "state": "ENABLED",
    "status": "VALID",
    "statusReason": "ComputeEnvironment Healthy",
    "computeResources": {
      "type": "EC2",
      "minvCpus": 0,
      "maxvCpus": 128,
      "desiredvCpus": 48,
      "instanceTypes": ["p3"],
      "subnets": ["subnet-220c0e0a", "subnet-1a95556d", "subnet-978f6dce"],
      "securityGroupIds": ["sg-cf5093b2"],
      "ec2KeyPair": "id_rsa",
      "instanceRole": "ecsInstanceRole",
      "tags": {
```

```
        "Name": "Batch Instance - P3OnDemand",
        "Department": "Management"
      }
    },
    "serviceRole": "arn:aws:iam::123456789012:role/AWSBatchServiceRole"
  }]
}
```

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for Kotlin
- AWS SDK for PHP V3
- AWS SDK for Python (Boto3)
- AWS SDK for Ruby V3

DescribeConsumableResource

Returns a description of the specified consumable resource.

Request Syntax

```
POST /v1/describeconsumableresource HTTP/1.1
Content-type: application/json

{
   "consumableResource": "string"
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

consumableResource

The name or ARN of the consumable resource whose description will be returned.

Type: String

Required: Yes

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
   "availableQuantity": number,
   "consumableResourceArn": "string",
   "consumableResourceName": "string",
   "createdAt": number,
   "inUseQuantity": number,
   "resourceType": "string",
```

```
   "tags": {
      "string" : "string"
   },
   "totalQuantity": 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.

availableQuantity

The amount of the consumable resource that is currently available to use.

Type: Long

consumableResourceArn

The Amazon Resource Name (ARN) of the consumable resource.

Type: String

consumableResourceName

The name of the consumable resource.

Type: String

createdAt

The Unix timestamp (in milliseconds) for when the consumable resource was created.

Type: Long

inUseQuantity

The amount of the consumable resource that is currently in use.

Type: Long

resourceType

Indicates whether the resource is available to be re-used after a job completes. Can be one of:

- REPLENISHABLE

- NON_REPLENISHABLE

Type: String

tags

The tags that you apply to the consumable resource to help you categorize and organize your resources. Each tag consists of a key and an optional value. For more information, see Tagging your AWS Batch resources.

Type: String to string map

Map Entries: Maximum number of 50 items.

Key Length Constraints: Minimum length of 1. Maximum length of 128.

Value Length Constraints: Maximum length of 256.

totalQuantity

The total amount of the consumable resource that is available.

Type: Long

Errors

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for Kotlin
- AWS SDK for PHP V3
- AWS SDK for Python (Boto3)
- AWS SDK for Ruby V3

DescribeJobDefinitions

Describes a list of job definitions. You can specify a `status` (such as `ACTIVE`) to only return job definitions that match that status.

Request Syntax

```
POST /v1/describejobdefinitions HTTP/1.1
Content-type: application/json

{
   "jobDefinitionName": "string",
   "jobDefinitions": [ "string" ],
   "maxResults": number,
   "nextToken": "string",
   "status": "string"
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

jobDefinitionName

The name of the job definition to describe.

Type: String

Required: No

jobDefinitions

A list of up to 100 job definitions. Each entry in the list can either be an ARN in the format `arn:aws:batch:${Region}:${Account}:job-definition/${JobDefinitionName}:${Revision}` or a short version using the form `${JobDefinitionName}:${Revision}`. This parameter can't be used with other parameters.

Type: Array of strings

Required: No

maxResults

The maximum number of results returned by `DescribeJobDefinitions` in paginated output. When this parameter is used, `DescribeJobDefinitions` only returns `maxResults` results in a single page and a `nextToken` response element. The remaining results of the initial request can be seen by sending another `DescribeJobDefinitions` request with the returned `nextToken` value. This value can be between 1 and 100. If this parameter isn't used, then `DescribeJobDefinitions` returns up to 100 results and a `nextToken` value if applicable.

Type: Integer

Required: No

nextToken

The `nextToken` value returned from a previous paginated `DescribeJobDefinitions` request where `maxResults` was used and the results exceeded the value of that parameter. Pagination continues from the end of the previous results that returned the `nextToken` value. This value is `null` when there are no more results to return.

> **Note**
>
> Treat this token as an opaque identifier that's only used to retrieve the next items in a list and not for other programmatic purposes.

Type: String

Required: No

status

The status used to filter job definitions.

Type: String

Required: No

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
   "jobDefinitions": [
      {
         "consumableResourceProperties": {
            "consumableResourceList": [
               {
                  "consumableResource": "string",
                  "quantity": number
               }
            ]
         },
         "containerOrchestrationType": "string",
         "containerProperties": {
            "command": [ "string" ],
            "enableExecuteCommand": boolean,
            "environment": [
               {
                  "name": "string",
                  "value": "string"
               }
            ],
            "ephemeralStorage": {
               "sizeInGiB": number
            },
            "executionRoleArn": "string",
            "fargatePlatformConfiguration": {
               "platformVersion": "string"
            },
            "image": "string",
            "instanceType": "string",
            "jobRoleArn": "string",
            "linuxParameters": {
               "devices": [
                  {
                     "containerPath": "string",
                     "hostPath": "string",
                     "permissions": [ "string" ]
                  }
```

```
      ],
      "initProcessEnabled": boolean,
      "maxSwap": number,
      "sharedMemorySize": number,
      "swappiness": number,
      "tmpfs": [
         {
            "containerPath": "string",
            "mountOptions": [ "string" ],
            "size": number
         }
      ]
   },
   "logConfiguration": {
      "logDriver": "string",
      "options": {
         "string" : "string"
      },
      "secretOptions": [
         {
            "name": "string",
            "valueFrom": "string"
         }
      ]
   },
   "memory": number,
   "mountPoints": [
      {
         "containerPath": "string",
         "readOnly": boolean,
         "sourceVolume": "string"
      }
   ],
   "networkConfiguration": {
      "assignPublicIp": "string"
   },
   "privileged": boolean,
   "readonlyRootFilesystem": boolean,
   "repositoryCredentials": {
      "credentialsParameter": "string"
   },
   "resourceRequirements": [
      {
         "type": "string",
```

```
            "value": "string"
         }
      ],
      "runtimePlatform": {
         "cpuArchitecture": "string",
         "operatingSystemFamily": "string"
      },
      "secrets": [
         {
            "name": "string",
            "valueFrom": "string"
         }
      ],
      "ulimits": [
         {
            "hardLimit": number,
            "name": "string",
            "softLimit": number
         }
      ],
      "user": "string",
      "vcpus": number,
      "volumes": [
         {
            "efsVolumeConfiguration": {
               "authorizationConfig": {
                  "accessPointId": "string",
                  "iam": "string"
               },
               "fileSystemId": "string",
               "rootDirectory": "string",
               "transitEncryption": "string",
               "transitEncryptionPort": number
            },
            "host": {
               "sourcePath": "string"
            },
            "name": "string",
            "s3filesVolumeConfiguration": {
               "accessPointArn": "string",
               "fileSystemArn": "string",
               "rootDirectory": "string",
               "transitEncryptionPort": number
            }
```

```
            }
         ]
      },
      "ecsProperties": {
         "taskProperties": [
            {
               "containers": [
                  {
                     "command": [ "string" ],
                     "dependsOn": [
                        {
                           "condition": "string",
                           "containerName": "string"
                        }
                     ],
                     "environment": [
                        {
                           "name": "string",
                           "value": "string"
                        }
                     ],
                     "essential": boolean,
                     "firelensConfiguration": {
                        "options": {
                           "string" : "string"
                        },
                        "type": "string"
                     },
                     "image": "string",
                     "linuxParameters": {
                        "devices": [
                           {
                              "containerPath": "string",
                              "hostPath": "string",
                              "permissions": [ "string" ]
                           }
                        ],
                        "initProcessEnabled": boolean,
                        "maxSwap": number,
                        "sharedMemorySize": number,
                        "swappiness": number,
                        "tmpfs": [
                           {
                              "containerPath": "string",
```

```
                "mountOptions": [ "string" ],
                "size": number
            }
        ]
    },
    "logConfiguration": {
        "logDriver": "string",
        "options": {
            "string" : "string"
        },
        "secretOptions": [
            {
                "name": "string",
                "valueFrom": "string"
            }
        ]
    },
    "mountPoints": [
        {
            "containerPath": "string",
            "readOnly": boolean,
            "sourceVolume": "string"
        }
    ],
    "name": "string",
    "privileged": boolean,
    "readonlyRootFilesystem": boolean,
    "repositoryCredentials": {
        "credentialsParameter": "string"
    },
    "resourceRequirements": [
        {
            "type": "string",
            "value": "string"
        }
    ],
    "secrets": [
        {
            "name": "string",
            "valueFrom": "string"
        }
    ],
    "startTimeout": number,
    "stopTimeout": number,
```

```
            "ulimits": [
               {
                  "hardLimit": number,
                  "name": "string",
                  "softLimit": number
               }
            ],
            "user": "string"
         }
      ],
      "enableExecuteCommand": boolean,
      "ephemeralStorage": {
         "sizeInGiB": number
      },
      "executionRoleArn": "string",
      "ipcMode": "string",
      "networkConfiguration": {
         "assignPublicIp": "string"
      },
      "networkMode": "string",
      "pidMode": "string",
      "platformVersion": "string",
      "runtimePlatform": {
         "cpuArchitecture": "string",
         "operatingSystemFamily": "string"
      },
      "taskRoleArn": "string",
      "volumes": [
         {
            "efsVolumeConfiguration": {
               "authorizationConfig": {
                  "accessPointId": "string",
                  "iam": "string"
               },
               "fileSystemId": "string",
               "rootDirectory": "string",
               "transitEncryption": "string",
               "transitEncryptionPort": number
            },
            "host": {
               "sourcePath": "string"
            },
            "name": "string",
            "s3filesVolumeConfiguration": {
```

```
                        "accessPointArn": "string",
                        "fileSystemArn": "string",
                        "rootDirectory": "string",
                        "transitEncryptionPort": number
                     }
                  }
               ]
            }
         ]
      },
      "eksProperties": {
         "podProperties": {
            "containers": [
               {
                  "args": [ "string" ],
                  "command": [ "string" ],
                  "env": [
                     {
                        "name": "string",
                        "value": "string"
                     }
                  ],
                  "image": "string",
                  "imagePullPolicy": "string",
                  "name": "string",
                  "resources": {
                     "limits": {
                        "string" : "string"
                     },
                     "requests": {
                        "string" : "string"
                     }
                  },
                  "securityContext": {
                     "allowPrivilegeEscalation": boolean,
                     "privileged": boolean,
                     "readOnlyRootFilesystem": boolean,
                     "runAsGroup": number,
                     "runAsNonRoot": boolean,
                     "runAsUser": number
                  },
                  "volumeMounts": [
                     {
                        "mountPath": "string",
```

```
                        "name": "string",
                        "readOnly": boolean,
                        "subPath": "string"
                     }
                  ]
               }
            ],
            "dnsPolicy": "string",
            "hostNetwork": boolean,
            "imagePullSecrets": [
               {
                  "name": "string"
               }
            ],
            "initContainers": [
               {
                  "args": [ "string" ],
                  "command": [ "string" ],
                  "env": [
                     {
                        "name": "string",
                        "value": "string"
                     }
                  ],
                  "image": "string",
                  "imagePullPolicy": "string",
                  "name": "string",
                  "resources": {
                     "limits": {
                        "string" : "string"
                     },
                     "requests": {
                        "string" : "string"
                     }
                  },
                  "securityContext": {
                     "allowPrivilegeEscalation": boolean,
                     "privileged": boolean,
                     "readOnlyRootFilesystem": boolean,
                     "runAsGroup": number,
                     "runAsNonRoot": boolean,
                     "runAsUser": number
                  },
                  "volumeMounts": [
```

```
                        {
                           "mountPath": "string",
                           "name": "string",
                           "readOnly": boolean,
                           "subPath": "string"
                        }
                     ]
                  }
               ],
               "metadata": {
                  "annotations": {
                     "string" : "string"
                  },
                  "labels": {
                     "string" : "string"
                  },
                  "namespace": "string"
               },
               "serviceAccountName": "string",
               "shareProcessNamespace": boolean,
               "volumes": [
                  {
                     "emptyDir": {
                        "medium": "string",
                        "sizeLimit": "string"
                     },
                     "hostPath": {
                        "path": "string"
                     },
                     "name": "string",
                     "persistentVolumeClaim": {
                        "claimName": "string",
                        "readOnly": boolean
                     },
                     "secret": {
                        "optional": boolean,
                        "secretName": "string"
                     }
                  }
               ]
            }
         },
         "jobDefinitionArn": "string",
         "jobDefinitionName": "string",
```

```
      "nodeProperties": {
         "mainNode": number,
         "nodeRangeProperties": [
            {
               "consumableResourceProperties": {
                  "consumableResourceList": [
                     {
                        "consumableResource": "string",
                        "quantity": number
                     }
                  ]
               },
               "container": {
                  "command": [ "string" ],
                  "enableExecuteCommand": boolean,
                  "environment": [
                     {
                        "name": "string",
                        "value": "string"
                     }
                  ],
                  "ephemeralStorage": {
                     "sizeInGiB": number
                  },
                  "executionRoleArn": "string",
                  "fargatePlatformConfiguration": {
                     "platformVersion": "string"
                  },
                  "image": "string",
                  "instanceType": "string",
                  "jobRoleArn": "string",
                  "linuxParameters": {
                     "devices": [
                        {
                           "containerPath": "string",
                           "hostPath": "string",
                           "permissions": [ "string" ]
                        }
                     ],
                     "initProcessEnabled": boolean,
                     "maxSwap": number,
                     "sharedMemorySize": number,
                     "swappiness": number,
                     "tmpfs": [
```

```
            {
               "containerPath": "string",
               "mountOptions": [ "string" ],
               "size": number
            }
         ]
      },
      "logConfiguration": {
         "logDriver": "string",
         "options": {
            "string" : "string"
         },
         "secretOptions": [
            {
               "name": "string",
               "valueFrom": "string"
            }
         ]
      },
      "memory": number,
      "mountPoints": [
         {
            "containerPath": "string",
            "readOnly": boolean,
            "sourceVolume": "string"
         }
      ],
      "networkConfiguration": {
         "assignPublicIp": "string"
      },
      "privileged": boolean,
      "readonlyRootFilesystem": boolean,
      "repositoryCredentials": {
         "credentialsParameter": "string"
      },
      "resourceRequirements": [
         {
            "type": "string",
            "value": "string"
         }
      ],
      "runtimePlatform": {
         "cpuArchitecture": "string",
         "operatingSystemFamily": "string"
```

```
            },
            "secrets": [
               {
                  "name": "string",
                  "valueFrom": "string"
               }
            ],
            "ulimits": [
               {
                  "hardLimit": number,
                  "name": "string",
                  "softLimit": number
               }
            ],
            "user": "string",
            "vcpus": number,
            "volumes": [
               {
                  "efsVolumeConfiguration": {
                     "authorizationConfig": {
                        "accessPointId": "string",
                        "iam": "string"
                     },
                     "fileSystemId": "string",
                     "rootDirectory": "string",
                     "transitEncryption": "string",
                     "transitEncryptionPort": number
                  },
                  "host": {
                     "sourcePath": "string"
                  },
                  "name": "string",
                  "s3filesVolumeConfiguration": {
                     "accessPointArn": "string",
                     "fileSystemArn": "string",
                     "rootDirectory": "string",
                     "transitEncryptionPort": number
                  }
               }
            ]
         },
         "ecsProperties": {
            "taskProperties": [
               {
```

```
"containers": [
   {
      "command": [ "string" ],
      "dependsOn": [
         {
            "condition": "string",
            "containerName": "string"
         }
      ],
      "environment": [
         {
            "name": "string",
            "value": "string"
         }
      ],
      "essential": boolean,
      "firelensConfiguration": {
         "options": {
            "string" : "string"
         },
         "type": "string"
      },
      "image": "string",
      "linuxParameters": {
         "devices": [
            {
               "containerPath": "string",
               "hostPath": "string",
               "permissions": [ "string" ]
            }
         ],
         "initProcessEnabled": boolean,
         "maxSwap": number,
         "sharedMemorySize": number,
         "swappiness": number,
         "tmpfs": [
            {
               "containerPath": "string",
               "mountOptions": [ "string" ],
               "size": number
            }
         ]
      },
      "logConfiguration": {
```

```
            "logDriver": "string",
            "options": {
               "string" : "string"
            },
            "secretOptions": [
               {
                  "name": "string",
                  "valueFrom": "string"
               }
            ]
         },
         "mountPoints": [
            {
               "containerPath": "string",
               "readOnly": boolean,
               "sourceVolume": "string"
            }
         ],
         "name": "string",
         "privileged": boolean,
         "readonlyRootFilesystem": boolean,
         "repositoryCredentials": {
            "credentialsParameter": "string"
         },
         "resourceRequirements": [
            {
               "type": "string",
               "value": "string"
            }
         ],
         "secrets": [
            {
               "name": "string",
               "valueFrom": "string"
            }
         ],
         "startTimeout": number,
         "stopTimeout": number,
         "ulimits": [
            {
               "hardLimit": number,
               "name": "string",
               "softLimit": number
            }
```

```
            ],
            "user": "string"
         }
      ],
      "enableExecuteCommand": boolean,
      "ephemeralStorage": {
         "sizeInGiB": number
      },
      "executionRoleArn": "string",
      "ipcMode": "string",
      "networkConfiguration": {
         "assignPublicIp": "string"
      },
      "networkMode": "string",
      "pidMode": "string",
      "platformVersion": "string",
      "runtimePlatform": {
         "cpuArchitecture": "string",
         "operatingSystemFamily": "string"
      },
      "taskRoleArn": "string",
      "volumes": [
         {
            "efsVolumeConfiguration": {
               "authorizationConfig": {
                  "accessPointId": "string",
                  "iam": "string"
               },
               "fileSystemId": "string",
               "rootDirectory": "string",
               "transitEncryption": "string",
               "transitEncryptionPort": number
            },
            "host": {
               "sourcePath": "string"
            },
            "name": "string",
            "s3filesVolumeConfiguration": {
               "accessPointArn": "string",
               "fileSystemArn": "string",
               "rootDirectory": "string",
               "transitEncryptionPort": number
            }
         }
```

```
                    ]
                 }
              ]
           },
           "eksProperties": {
              "podProperties": {
                 "containers": [
                    {
                       "args": [ "string" ],
                       "command": [ "string" ],
                       "env": [
                          {
                             "name": "string",
                             "value": "string"
                          }
                       ],
                       "image": "string",
                       "imagePullPolicy": "string",
                       "name": "string",
                       "resources": {
                          "limits": {
                             "string" : "string"
                          },
                          "requests": {
                             "string" : "string"
                          }
                       },
                       "securityContext": {
                          "allowPrivilegeEscalation": boolean,
                          "privileged": boolean,
                          "readOnlyRootFilesystem": boolean,
                          "runAsGroup": number,
                          "runAsNonRoot": boolean,
                          "runAsUser": number
                       },
                       "volumeMounts": [
                          {
                             "mountPath": "string",
                             "name": "string",
                             "readOnly": boolean,
                             "subPath": "string"
                          }
                       ]
                    }
```

```
      ],
      "dnsPolicy": "string",
      "hostNetwork": boolean,
      "imagePullSecrets": [
         {
            "name": "string"
         }
      ],
      "initContainers": [
         {
            "args": [ "string" ],
            "command": [ "string" ],
            "env": [
               {
                  "name": "string",
                  "value": "string"
               }
            ],
            "image": "string",
            "imagePullPolicy": "string",
            "name": "string",
            "resources": {
               "limits": {
                  "string" : "string"
               },
               "requests": {
                  "string" : "string"
               }
            },
            "securityContext": {
               "allowPrivilegeEscalation": boolean,
               "privileged": boolean,
               "readOnlyRootFilesystem": boolean,
               "runAsGroup": number,
               "runAsNonRoot": boolean,
               "runAsUser": number
            },
            "volumeMounts": [
               {
                  "mountPath": "string",
                  "name": "string",
                  "readOnly": boolean,
                  "subPath": "string"
               }
```

```
                        ]
                     }
                  ],
                  "metadata": {
                     "annotations": {
                        "string" : "string"
                     },
                     "labels": {
                        "string" : "string"
                     },
                     "namespace": "string"
                  },
                  "serviceAccountName": "string",
                  "shareProcessNamespace": boolean,
                  "volumes": [
                     {
                        "emptyDir": {
                           "medium": "string",
                           "sizeLimit": "string"
                        },
                        "hostPath": {
                           "path": "string"
                        },
                        "name": "string",
                        "persistentVolumeClaim": {
                           "claimName": "string",
                           "readOnly": boolean
                        },
                        "secret": {
                           "optional": boolean,
                           "secretName": "string"
                        }
                     }
                  ]
               }
            },
            "instanceTypes": [ "string" ],
            "targetNodes": "string"
         }
      ],
      "numNodes": number
   },
   "parameters": {
      "string" : "string"
```

```
         },
         "platformCapabilities": [ "string" ],
         "propagateTags": boolean,
         "retryStrategy": {
            "attempts": number,
            "evaluateOnExit": [
               {
                  "action": "string",
                  "onExitCode": "string",
                  "onReason": "string",
                  "onStatusReason": "string"
               }
            ]
         },
         "revision": number,
         "schedulingPriority": number,
         "status": "string",
         "tags": {
            "string" : "string"
         },
         "timeout": {
            "attemptDurationSeconds": number
         },
         "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.

jobDefinitions

The list of job definitions.

Type: Array of JobDefinition objects

nextToken

The `nextToken` value to include in a future `DescribeJobDefinitions` request. When the results of a `DescribeJobDefinitions` request exceed `maxResults`, this value can be used to retrieve the next page of results. This value is `null` when there are no more results to return.

Type: String

Errors

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

Examples

In the following example or examples, the Authorization header contents (*`[authorization-params]`*) must be replaced with an AWS Signature Version 4 signature. For more information about creating these signatures, see Signature Version 4 Signing Process in the *AWS General Reference*.

You only need to learn how to sign HTTP requests if you intend to manually create them. When you use the AWS Command Line Interface (AWS CLI) or one of the AWS SDKs to make requests to AWS, these tools automatically sign the requests for you with the access key that you specify when you configure the tools. When you use these tools, you don't need to learn how to sign requests yourself.

Example

This example describes all of your active job definitions.

Sample Request

```
POST /v1/describejobdefinitions HTTP/1.1
Host: batch.us-east-1.amazonaws.com
Accept-Encoding: identity
Content-Length: [content-length]
Authorization: [authorization-params]
X-Amz-Date: 20161128T221855Z
User-Agent: aws-cli/1.11.21 Python/2.7.12 Darwin/16.1.0 botocore/1.4.78

{
  "status": "ACTIVE"
}
```

Sample Response

```
HTTP/1.1 200 OK
Content-Type: application/json
Content-Length: [content-length]
Connection: keep-alive
Date: Mon, 28 Nov 2016 22:18:55 GMT
x-amzn-RequestId: [request-id]
X-Amzn-Trace-Id: [trace-id]
X-Cache: Miss from cloudfront
Via: 1.1 688936cc730f240888e6a59a81892a3d.cloudfront.net (CloudFront)
X-Amz-Cf-Id: hd-CAMqfaCJt-1oH7tBu9j5c-IhLQuMjFHFPck6F0MMt5CBea8mQBQ==

{
  "jobDefinitions": [{
    "jobDefinitionName": "sleep60",
    "jobDefinitionArn": "arn:aws:batch:us-east-1:123456789012:job-definition/
sleep60:1",
    "revision": 1,
    "status": "ACTIVE",
    "type": "container",
    "containerProperties": {
      "image": "busybox",
      "vcpus": 1,
      "memory": 128,
      "command": ["sleep", "60"],
      "volumes": [],
      "environment": [],
      "mountPoints": [],
```

```
      "ulimits": []
    }
  }]
}
```

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for Kotlin
- AWS SDK for PHP V3
- AWS SDK for Python (Boto3)
- AWS SDK for Ruby V3

DescribeJobQueues

Describes one or more of your job queues.

Request Syntax

```
POST /v1/describejobqueues HTTP/1.1
Content-type: application/json

{
   "jobQueues": [ "string" ],
   "maxResults": number,
   "nextToken": "string"
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

jobQueues

A list of up to 100 queue names or full queue Amazon Resource Name (ARN) entries.

Type: Array of strings

Required: No

maxResults

The maximum number of results returned by `DescribeJobQueues` in paginated output. When this parameter is used, `DescribeJobQueues` only returns `maxResults` results in a single page and a `nextToken` response element. The remaining results of the initial request can be seen by sending another `DescribeJobQueues` request with the returned `nextToken` value. This value can be between 1 and 100. If this parameter isn't used, then `DescribeJobQueues` returns up to 100 results and a `nextToken` value if applicable.

Type: Integer

Required: No

nextToken

The `nextToken` value returned from a previous paginated `DescribeJobQueues` request where `maxResults` was used and the results exceeded the value of that parameter. Pagination continues from the end of the previous results that returned the `nextToken` value. This value is `null` when there are no more results to return.

> **Note**
>
> Treat this token as an opaque identifier that's only used to retrieve the next items in a list and not for other programmatic purposes.

Type: String

Required: No

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
   "jobQueues": [
      {
         "computeEnvironmentOrder": [
            {
               "computeEnvironment": "string",
               "order": number
            }
         ],
         "jobQueueArn": "string",
         "jobQueueName": "string",
         "jobQueueType": "string",
         "jobStateTimeLimitActions": [
            {
               "action": "string",
               "maxTimeSeconds": number,
               "reason": "string",
```

```
                "state": "string"
            }
         ],
         "priority": number,
         "schedulingPolicyArn": "string",
         "serviceEnvironmentOrder": [
            {
               "order": number,
               "serviceEnvironment": "string"
            }
         ],
         "state": "string",
         "status": "string",
         "statusReason": "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.

jobQueues

The list of job queues.

Type: Array of JobQueueDetail objects

nextToken

The `nextToken` value to include in a future `DescribeJobQueues` request. When the results of a `DescribeJobQueues` request exceed `maxResults`, this value can be used to retrieve the next page of results. This value is `null` when there are no more results to return.

Type: String

Errors

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

Examples

In the following example or examples, the Authorization header contents (*`[authorization-params]`*) must be replaced with an AWS Signature Version 4 signature. For more information about creating these signatures, see Signature Version 4 Signing Process in the *AWS General Reference*.

You only need to learn how to sign HTTP requests if you intend to manually create them. When you use the AWS Command Line Interface (AWS CLI) or one of the AWS SDKs to make requests to AWS, these tools automatically sign the requests for you with the access key that you specify when you configure the tools. When you use these tools, you don't need to learn how to sign requests yourself.

Example

This example describes the `HighPriority` job queue.

Sample Request

```
POST /v1/describejobqueues HTTP/1.1
Host: batch.us-east-1.amazonaws.com
Accept-Encoding: identity
Content-Length: [content-length]
Authorization: [authorization-params]
```

```
X-Amz-Date: 20161128T194731Z
User-Agent: aws-cli/1.11.21 Python/2.7.12 Darwin/16.1.0 botocore/1.4.78

{
  "jobQueues": [
    "HighPriority"
  ]
}
```

Sample Response

```
HTTP/1.1 200 OK
Date: Mon, 28 Nov 2016 19:47:32 GMT
Content-Type: application/json
Content-Length: [content-length]
Connection: keep-alive
x-amzn-RequestId: [request-id]
X-Amzn-Trace-Id: [trace-id]
X-Cache: Miss from cloudfront
Via: 1.1 7bfcc2251021d9dc94a87ff179d69731.cloudfront.net (CloudFront)
X-Amz-Cf-Id: dwf7P2pnEYCxN1C3XdApqDtqzLfjpWAjbHvskd9oUqz4OUn9pvtx3Q==

{
  "jobQueues": [{
    "jobQueueName": "HighPriority",
    "jobQueueArn": "arn:aws:batch:us-east-1:123456789012:job-queue/HighPriority",
    "state": "ENABLED",
    "status": "VALID",
    "statusReason": "JobQueue Healthy",
    "priority": 10,
    "computeEnvironmentOrder": [{
      "order": 1,
      "computeEnvironment": "arn:aws:batch:us-east-1:123456789012:compute-environment/
C4OnDemand"
    }]
  }]
}
```

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for Kotlin
- AWS SDK for PHP V3
- AWS SDK for Python (Boto3)
- AWS SDK for Ruby V3

DescribeJobs

Describes a list of AWS Batch jobs.

Request Syntax

```
POST /v1/describejobs HTTP/1.1
Content-type: application/json

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

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

jobs

A list of up to 100 job IDs.

Type: Array of strings

Required: Yes

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
   "jobs": [
      {
         "arrayProperties": {
            "index": number,
            "size": number,
            "statusSummary": {
```

```
            "string" : number
         },
         "statusSummaryLastUpdatedAt": number
      },
      "attempts": [
         {
            "container": {
               "containerInstanceArn": "string",
               "exitCode": number,
               "logStreamName": "string",
               "networkInterfaces": [
                  {
                     "attachmentId": "string",
                     "ipv6Address": "string",
                     "privateIpv4Address": "string"
                  }
               ],
               "reason": "string",
               "taskArn": "string"
            },
            "startedAt": number,
            "statusReason": "string",
            "stoppedAt": number,
            "taskProperties": [
               {
                  "containerInstanceArn": "string",
                  "containers": [
                     {
                        "exitCode": number,
                        "logStreamName": "string",
                        "name": "string",
                        "networkInterfaces": [
                           {
                              "attachmentId": "string",
                              "ipv6Address": "string",
                              "privateIpv4Address": "string"
                           }
                        ],
                        "reason": "string"
                     }
                  ],
                  "taskArn": "string"
               }
            ]
```

```
            }
         ],
         "consumableResourceProperties": {
            "consumableResourceList": [
               {
                  "consumableResource": "string",
                  "quantity": number
               }
            ]
         },
         "container": {
            "command": [ "string" ],
            "containerInstanceArn": "string",
            "enableExecuteCommand": boolean,
            "environment": [
               {
                  "name": "string",
                  "value": "string"
               }
            ],
            "ephemeralStorage": {
               "sizeInGiB": number
            },
            "executionRoleArn": "string",
            "exitCode": number,
            "fargatePlatformConfiguration": {
               "platformVersion": "string"
            },
            "image": "string",
            "instanceType": "string",
            "jobRoleArn": "string",
            "linuxParameters": {
               "devices": [
                  {
                     "containerPath": "string",
                     "hostPath": "string",
                     "permissions": [ "string" ]
                  }
               ],
               "initProcessEnabled": boolean,
               "maxSwap": number,
               "sharedMemorySize": number,
               "swappiness": number,
               "tmpfs": [
```

```
            {
               "containerPath": "string",
               "mountOptions": [ "string" ],
               "size": number
            }
         ]
      },
      "logConfiguration": {
         "logDriver": "string",
         "options": {
            "string" : "string"
         },
         "secretOptions": [
            {
               "name": "string",
               "valueFrom": "string"
            }
         ]
      },
      "logStreamName": "string",
      "memory": number,
      "mountPoints": [
         {
            "containerPath": "string",
            "readOnly": boolean,
            "sourceVolume": "string"
         }
      ],
      "networkConfiguration": {
         "assignPublicIp": "string"
      },
      "networkInterfaces": [
         {
            "attachmentId": "string",
            "ipv6Address": "string",
            "privateIpv4Address": "string"
         }
      ],
      "privileged": boolean,
      "readonlyRootFilesystem": boolean,
      "reason": "string",
      "repositoryCredentials": {
         "credentialsParameter": "string"
      },
```

```
            "resourceRequirements": [
               {
                  "type": "string",
                  "value": "string"
               }
            ],
            "runtimePlatform": {
               "cpuArchitecture": "string",
               "operatingSystemFamily": "string"
            },
            "secrets": [
               {
                  "name": "string",
                  "valueFrom": "string"
               }
            ],
            "taskArn": "string",
            "ulimits": [
               {
                  "hardLimit": number,
                  "name": "string",
                  "softLimit": number
               }
            ],
            "user": "string",
            "vcpus": number,
            "volumes": [
               {
                  "efsVolumeConfiguration": {
                     "authorizationConfig": {
                        "accessPointId": "string",
                        "iam": "string"
                     },
                     "fileSystemId": "string",
                     "rootDirectory": "string",
                     "transitEncryption": "string",
                     "transitEncryptionPort": number
                  },
                  "host": {
                     "sourcePath": "string"
                  },
                  "name": "string",
                  "s3filesVolumeConfiguration": {
                     "accessPointArn": "string",
```

```
                    "fileSystemArn": "string",
                    "rootDirectory": "string",
                    "transitEncryptionPort": number
                }
            }
        ]
    },
    "createdAt": number,
    "dependsOn": [
        {
            "jobId": "string",
            "type": "string"
        }
    ],
    "ecsProperties": {
        "taskProperties": [
            {
                "containerInstanceArn": "string",
                "containers": [
                    {
                        "command": [ "string" ],
                        "dependsOn": [
                            {
                                "condition": "string",
                                "containerName": "string"
                            }
                        ],
                        "environment": [
                            {
                                "name": "string",
                                "value": "string"
                            }
                        ],
                        "essential": boolean,
                        "exitCode": number,
                        "firelensConfiguration": {
                            "options": {
                                "string" : "string"
                            },
                            "type": "string"
                        },
                        "image": "string",
                        "linuxParameters": {
                            "devices": [
```

```
            {
               "containerPath": "string",
               "hostPath": "string",
               "permissions": [ "string" ]
            }
         ],
         "initProcessEnabled": boolean,
         "maxSwap": number,
         "sharedMemorySize": number,
         "swappiness": number,
         "tmpfs": [
            {
               "containerPath": "string",
               "mountOptions": [ "string" ],
               "size": number
            }
         ]
      },
      "logConfiguration": {
         "logDriver": "string",
         "options": {
            "string" : "string"
         },
         "secretOptions": [
            {
               "name": "string",
               "valueFrom": "string"
            }
         ]
      },
      "logStreamName": "string",
      "mountPoints": [
         {
            "containerPath": "string",
            "readOnly": boolean,
            "sourceVolume": "string"
         }
      ],
      "name": "string",
      "networkInterfaces": [
         {
            "attachmentId": "string",
            "ipv6Address": "string",
            "privateIpv4Address": "string"
```

```
            }
         ],
         "privileged": boolean,
         "readonlyRootFilesystem": boolean,
         "reason": "string",
         "repositoryCredentials": {
            "credentialsParameter": "string"
         },
         "resourceRequirements": [
            {
               "type": "string",
               "value": "string"
            }
         ],
         "secrets": [
            {
               "name": "string",
               "valueFrom": "string"
            }
         ],
         "startTimeout": number,
         "stopTimeout": number,
         "ulimits": [
            {
               "hardLimit": number,
               "name": "string",
               "softLimit": number
            }
         ],
         "user": "string"
      }
   ],
   "enableExecuteCommand": boolean,
   "ephemeralStorage": {
      "sizeInGiB": number
   },
   "executionRoleArn": "string",
   "ipcMode": "string",
   "networkConfiguration": {
      "assignPublicIp": "string"
   },
   "networkMode": "string",
   "pidMode": "string",
   "platformVersion": "string",
```

```
            "runtimePlatform": {
               "cpuArchitecture": "string",
               "operatingSystemFamily": "string"
            },
            "taskArn": "string",
            "taskRoleArn": "string",
            "volumes": [
               {
                  "efsVolumeConfiguration": {
                     "authorizationConfig": {
                        "accessPointId": "string",
                        "iam": "string"
                     },
                     "fileSystemId": "string",
                     "rootDirectory": "string",
                     "transitEncryption": "string",
                     "transitEncryptionPort": number
                  },
                  "host": {
                     "sourcePath": "string"
                  },
                  "name": "string",
                  "s3filesVolumeConfiguration": {
                     "accessPointArn": "string",
                     "fileSystemArn": "string",
                     "rootDirectory": "string",
                     "transitEncryptionPort": number
                  }
               }
            ]
         }
      ]
   },
   "eksAttempts": [
      {
         "containers": [
            {
               "containerID": "string",
               "exitCode": number,
               "name": "string",
               "reason": "string"
            }
         ],
         "eksClusterArn": "string",
```

```
            "initContainers": [
               {
                  "containerID": "string",
                  "exitCode": number,
                  "name": "string",
                  "reason": "string"
               }
            ],
            "nodeName": "string",
            "podName": "string",
            "podNamespace": "string",
            "startedAt": number,
            "statusReason": "string",
            "stoppedAt": number
         }
      ],
      "eksProperties": {
         "podProperties": {
            "containers": [
               {
                  "args": [ "string" ],
                  "command": [ "string" ],
                  "env": [
                     {
                        "name": "string",
                        "value": "string"
                     }
                  ],
                  "exitCode": number,
                  "image": "string",
                  "imagePullPolicy": "string",
                  "name": "string",
                  "reason": "string",
                  "resources": {
                     "limits": {
                        "string" : "string"
                     },
                     "requests": {
                        "string" : "string"
                     }
                  },
                  "securityContext": {
                     "allowPrivilegeEscalation": boolean,
                     "privileged": boolean,
```

```
                        "readOnlyRootFilesystem": boolean,
                        "runAsGroup": number,
                        "runAsNonRoot": boolean,
                        "runAsUser": number
                     },
                     "volumeMounts": [
                        {
                           "mountPath": "string",
                           "name": "string",
                           "readOnly": boolean,
                           "subPath": "string"
                        }
                     ]
                  }
               ],
               "dnsPolicy": "string",
               "hostNetwork": boolean,
               "imagePullSecrets": [
                  {
                     "name": "string"
                  }
               ],
               "initContainers": [
                  {
                     "args": [ "string" ],
                     "command": [ "string" ],
                     "env": [
                        {
                           "name": "string",
                           "value": "string"
                        }
                     ],
                     "exitCode": number,
                     "image": "string",
                     "imagePullPolicy": "string",
                     "name": "string",
                     "reason": "string",
                     "resources": {
                        "limits": {
                           "string" : "string"
                        },
                        "requests": {
                           "string" : "string"
                        }
```

```
            },
            "securityContext": {
               "allowPrivilegeEscalation": boolean,
               "privileged": boolean,
               "readOnlyRootFilesystem": boolean,
               "runAsGroup": number,
               "runAsNonRoot": boolean,
               "runAsUser": number
            },
            "volumeMounts": [
               {
                  "mountPath": "string",
                  "name": "string",
                  "readOnly": boolean,
                  "subPath": "string"
               }
            ]
         }
      ],
      "metadata": {
         "annotations": {
            "string" : "string"
         },
         "labels": {
            "string" : "string"
         },
         "namespace": "string"
      },
      "nodeName": "string",
      "podName": "string",
      "serviceAccountName": "string",
      "shareProcessNamespace": boolean,
      "volumes": [
         {
            "emptyDir": {
               "medium": "string",
               "sizeLimit": "string"
            },
            "hostPath": {
               "path": "string"
            },
            "name": "string",
            "persistentVolumeClaim": {
               "claimName": "string",
```

```
                        "readOnly": boolean
                     },
                     "secret": {
                        "optional": boolean,
                        "secretName": "string"
                     }
                  }
               ]
            }
         },
         "isCancelled": boolean,
         "isTerminated": boolean,
         "jobArn": "string",
         "jobDefinition": "string",
         "jobId": "string",
         "jobName": "string",
         "jobQueue": "string",
         "nodeDetails": {
            "isMainNode": boolean,
            "nodeIndex": number
         },
         "nodeProperties": {
            "mainNode": number,
            "nodeRangeProperties": [
               {
                  "consumableResourceProperties": {
                     "consumableResourceList": [
                        {
                           "consumableResource": "string",
                           "quantity": number
                        }
                     ]
                  },
                  "container": {
                     "command": [ "string" ],
                     "enableExecuteCommand": boolean,
                     "environment": [
                        {
                           "name": "string",
                           "value": "string"
                        }
                     ],
                     "ephemeralStorage": {
                        "sizeInGiB": number
```

```
      },
      "executionRoleArn": "string",
      "fargatePlatformConfiguration": {
         "platformVersion": "string"
      },
      "image": "string",
      "instanceType": "string",
      "jobRoleArn": "string",
      "linuxParameters": {
         "devices": [
            {
               "containerPath": "string",
               "hostPath": "string",
               "permissions": [ "string" ]
            }
         ],
         "initProcessEnabled": boolean,
         "maxSwap": number,
         "sharedMemorySize": number,
         "swappiness": number,
         "tmpfs": [
            {
               "containerPath": "string",
               "mountOptions": [ "string" ],
               "size": number
            }
         ]
      },
      "logConfiguration": {
         "logDriver": "string",
         "options": {
            "string" : "string"
         },
         "secretOptions": [
            {
               "name": "string",
               "valueFrom": "string"
            }
         ]
      },
      "memory": number,
      "mountPoints": [
         {
            "containerPath": "string",
```

```
            "readOnly": boolean,
            "sourceVolume": "string"
         }
      ],
      "networkConfiguration": {
         "assignPublicIp": "string"
      },
      "privileged": boolean,
      "readonlyRootFilesystem": boolean,
      "repositoryCredentials": {
         "credentialsParameter": "string"
      },
      "resourceRequirements": [
         {
            "type": "string",
            "value": "string"
         }
      ],
      "runtimePlatform": {
         "cpuArchitecture": "string",
         "operatingSystemFamily": "string"
      },
      "secrets": [
         {
            "name": "string",
            "valueFrom": "string"
         }
      ],
      "ulimits": [
         {
            "hardLimit": number,
            "name": "string",
            "softLimit": number
         }
      ],
      "user": "string",
      "vcpus": number,
      "volumes": [
         {
            "efsVolumeConfiguration": {
               "authorizationConfig": {
                  "accessPointId": "string",
                  "iam": "string"
               },
```

```
                        "fileSystemId": "string",
                        "rootDirectory": "string",
                        "transitEncryption": "string",
                        "transitEncryptionPort": number
                     },
                     "host": {
                        "sourcePath": "string"
                     },
                     "name": "string",
                     "s3filesVolumeConfiguration": {
                        "accessPointArn": "string",
                        "fileSystemArn": "string",
                        "rootDirectory": "string",
                        "transitEncryptionPort": number
                     }
                  }
               ]
            },
            "ecsProperties": {
               "taskProperties": [
                  {
                     "containers": [
                        {
                           "command": [ "string" ],
                           "dependsOn": [
                              {
                                 "condition": "string",
                                 "containerName": "string"
                              }
                           ],
                           "environment": [
                              {
                                 "name": "string",
                                 "value": "string"
                              }
                           ],
                           "essential": boolean,
                           "firelensConfiguration": {
                              "options": {
                                 "string" : "string"
                              },
                              "type": "string"
                           },
                           "image": "string",
```

```
      "linuxParameters": {
         "devices": [
            {
               "containerPath": "string",
               "hostPath": "string",
               "permissions": [ "string" ]
            }
         ],
         "initProcessEnabled": boolean,
         "maxSwap": number,
         "sharedMemorySize": number,
         "swappiness": number,
         "tmpfs": [
            {
               "containerPath": "string",
               "mountOptions": [ "string" ],
               "size": number
            }
         ]
      },
      "logConfiguration": {
         "logDriver": "string",
         "options": {
            "string" : "string"
         },
         "secretOptions": [
            {
               "name": "string",
               "valueFrom": "string"
            }
         ]
      },
      "mountPoints": [
         {
            "containerPath": "string",
            "readOnly": boolean,
            "sourceVolume": "string"
         }
      ],
      "name": "string",
      "privileged": boolean,
      "readonlyRootFilesystem": boolean,
      "repositoryCredentials": {
         "credentialsParameter": "string"
```

```
            },
            "resourceRequirements": [
               {
                  "type": "string",
                  "value": "string"
               }
            ],
            "secrets": [
               {
                  "name": "string",
                  "valueFrom": "string"
               }
            ],
            "startTimeout": number,
            "stopTimeout": number,
            "ulimits": [
               {
                  "hardLimit": number,
                  "name": "string",
                  "softLimit": number
               }
            ],
            "user": "string"
         }
      ],
      "enableExecuteCommand": boolean,
      "ephemeralStorage": {
         "sizeInGiB": number
      },
      "executionRoleArn": "string",
      "ipcMode": "string",
      "networkConfiguration": {
         "assignPublicIp": "string"
      },
      "networkMode": "string",
      "pidMode": "string",
      "platformVersion": "string",
      "runtimePlatform": {
         "cpuArchitecture": "string",
         "operatingSystemFamily": "string"
      },
      "taskRoleArn": "string",
      "volumes": [
         {
```

```
                        "efsVolumeConfiguration": {
                           "authorizationConfig": {
                              "accessPointId": "string",
                              "iam": "string"
                           },
                           "fileSystemId": "string",
                           "rootDirectory": "string",
                           "transitEncryption": "string",
                           "transitEncryptionPort": number
                        },
                        "host": {
                           "sourcePath": "string"
                        },
                        "name": "string",
                        "s3filesVolumeConfiguration": {
                           "accessPointArn": "string",
                           "fileSystemArn": "string",
                           "rootDirectory": "string",
                           "transitEncryptionPort": number
                        }
                     }
                  ]
               }
            ]
         },
         "eksProperties": {
            "podProperties": {
               "containers": [
                  {
                     "args": [ "string" ],
                     "command": [ "string" ],
                     "env": [
                        {
                           "name": "string",
                           "value": "string"
                        }
                     ],
                     "image": "string",
                     "imagePullPolicy": "string",
                     "name": "string",
                     "resources": {
                        "limits": {
                           "string" : "string"
                        },
```

```
                        "requests": {
                           "string" : "string"
                        }
                     },
                     "securityContext": {
                        "allowPrivilegeEscalation": boolean,
                        "privileged": boolean,
                        "readOnlyRootFilesystem": boolean,
                        "runAsGroup": number,
                        "runAsNonRoot": boolean,
                        "runAsUser": number
                     },
                     "volumeMounts": [
                        {
                           "mountPath": "string",
                           "name": "string",
                           "readOnly": boolean,
                           "subPath": "string"
                        }
                     ]
                  }
               ],
               "dnsPolicy": "string",
               "hostNetwork": boolean,
               "imagePullSecrets": [
                  {
                     "name": "string"
                  }
               ],
               "initContainers": [
                  {
                     "args": [ "string" ],
                     "command": [ "string" ],
                     "env": [
                        {
                           "name": "string",
                           "value": "string"
                        }
                     ],
                     "image": "string",
                     "imagePullPolicy": "string",
                     "name": "string",
                     "resources": {
                        "limits": {
```

```
                        "string" : "string"
                     },
                     "requests": {
                        "string" : "string"
                     }
                  },
                  "securityContext": {
                     "allowPrivilegeEscalation": boolean,
                     "privileged": boolean,
                     "readOnlyRootFilesystem": boolean,
                     "runAsGroup": number,
                     "runAsNonRoot": boolean,
                     "runAsUser": number
                  },
                  "volumeMounts": [
                     {
                        "mountPath": "string",
                        "name": "string",
                        "readOnly": boolean,
                        "subPath": "string"
                     }
                  ]
               }
            ],
            "metadata": {
               "annotations": {
                  "string" : "string"
               },
               "labels": {
                  "string" : "string"
               },
               "namespace": "string"
            },
            "serviceAccountName": "string",
            "shareProcessNamespace": boolean,
            "volumes": [
               {
                  "emptyDir": {
                     "medium": "string",
                     "sizeLimit": "string"
                  },
                  "hostPath": {
                     "path": "string"
                  },
```

```
                        "name": "string",
                        "persistentVolumeClaim": {
                           "claimName": "string",
                           "readOnly": boolean
                        },
                        "secret": {
                           "optional": boolean,
                           "secretName": "string"
                        }
                     }
                  ]
               }
            },
            "instanceTypes": [ "string" ],
            "targetNodes": "string"
         }
      ],
      "numNodes": number
   },
   "parameters": {
      "string" : "string"
   },
   "platformCapabilities": [ "string" ],
   "propagateTags": boolean,
   "retryStrategy": {
      "attempts": number,
      "evaluateOnExit": [
         {
            "action": "string",
            "onExitCode": "string",
            "onReason": "string",
            "onStatusReason": "string"
         }
      ]
   },
   "schedulingPriority": number,
   "shareIdentifier": "string",
   "startedAt": number,
   "status": "string",
   "statusReason": "string",
   "stoppedAt": number,
   "tags": {
      "string" : "string"
   },
```

```
         "timeout": {
            "attemptDurationSeconds": 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.

jobs

The list of jobs.

Type: Array of JobDetail objects

Errors

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

Examples

In the following example or examples, the Authorization header contents (*[authorization-params]*) must be replaced with an AWS Signature Version 4 signature. For more information about creating these signatures, see Signature Version 4 Signing Process in the *AWS General Reference*.

You only need to learn how to sign HTTP requests if you intend to manually create them. When you use the AWS Command Line Interface (AWS CLI) or one of the AWS SDKs to make requests to AWS, these tools automatically sign the requests for you with the access key that you specify when you configure the tools. When you use these tools, you don't need to learn how to sign requests yourself.

Example

This example describes a job with the specified job ID.

Sample Request

```
POST /v1/describejobs HTTP/1.1
Host: batch.us-east-1.amazonaws.com
Accept-Encoding: identity
Content-Length: [content-length]
Authorization: [authorization-params]
X-Amz-Date: 20170327T151323Z
User-Agent: aws-cli/1.11.22 Python/2.7.12 Darwin/16.1.0 botocore/1.4.79

{
  "jobs": [
    "0668da57-1bcc-478b-bc14-5d4f1c1cef48"
  ]
}
```

Sample Response

```
HTTP/1.1 200 OK
Date: Mon, 27 Mar 2017 15:13:13 GMT
Content-Type: application/json
Content-Length: [content-length]
Connection: keep-alive
x-amzn-RequestId: [request-id]
X-Amzn-Trace-Id: [trace-id]
X-Cache: Miss from cloudfront
Via: 1.1 8a78b675adb2cce925860f2fe4383e71.cloudfront.net (CloudFront)
X-Amz-Cf-Id: TaW8k7yrDyXHgEU2udEEOAbliIY1iPmQr4LpN80ULdqyVGR6qP0q4Q==

{
  "jobs": [
    {
      "jobName": "EchoAttemptNumber",
```

```
      "jobId": "0668da57-1bcc-478b-bc14-5d4f1c1cef48",
      "jobQueue": "arn:aws:batch:us-east-1:123456789012:job-queue/HighPriority",
      "status": "FAILED",
      "attempts": [
        {
          "container": {
            "containerInstanceArn": "arn:aws:ecs:us-east-1:123456789012:container-
instance/90bfe527-119c-494a-b8fe-f5999c66d214",
            "taskArn": "arn:aws:ecs:us-east-1:123456789012:task/af37d830-6978-4a2b-
b796-e890e9b477b3",
            "exitCode": 1
          },
          "startedAt": 1490627002951,
          "stoppedAt": 1490627003065,
          "statusReason": "Essential container in task exited"
        },
        {
          "container": {
            "containerInstanceArn": "arn:aws:ecs:us-east-1:123456789012:container-
instance/90bfe527-119c-494a-b8fe-f5999c66d214",
            "taskArn": "arn:aws:ecs:us-east-1:123456789012:task/3dfd4d0e-
a177-4798-9c13-21b7148217bc",
            "exitCode": 2
          },
          "startedAt": 1490627019948,
          "stoppedAt": 1490627020059,
          "statusReason": "Essential container in task exited"
        },
        {
          "container": {
            "containerInstanceArn": "arn:aws:ecs:us-east-1:123456789012:container-
instance/90bfe527-119c-494a-b8fe-f5999c66d214",
            "taskArn": "arn:aws:ecs:us-
east-1:123456789012:task/22857040-182c-4af3-85f5-bb2c71edd282",
            "exitCode": 3
          },
          "startedAt": 1490627034798,
          "stoppedAt": 1490627034949,
          "statusReason": "Essential container in task exited"
        }
      ],
      "statusReason": "Essential container in task exited",
      "createdAt": 1490626709525,
      "retryStrategy": {
```

```
        "attempts": 3
      },
      "startedAt": 1490627034798,
      "stoppedAt": 1490627034949,
      "dependsOn": [],
      "jobDefinition": "arn:aws:batch:us-east-1:123456789012:job-definition/
EchoAttemptNumber:1",
      "parameters": {},
      "container": {
        "image": "amazonlinux",
        "vcpus": 1,
        "memory": 2,
        "command": [
          "/bin/bash",
          "-c",
          "exit $AWS_BATCH_JOB_ATTEMPT"
        ],
        "volumes": [],
        "environment": [],
        "mountPoints": [],
        "ulimits": [],
        "exitCode": 3,
        "containerInstanceArn": "arn:aws:ecs:us-east-1:123456789012:container-
instance/90bfe527-119c-494a-b8fe-f5999c66d214",
        "taskArn": "arn:aws:ecs:us-east-1:123456789012:task/22857040-182c-4af3-85f5-
bb2c71edd282"
      }
    }
  ]
}
```

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2

- AWS SDK for JavaScript V3
- AWS SDK for Kotlin
- AWS SDK for PHP V3
- AWS SDK for Python (Boto3)
- AWS SDK for Ruby V3

DescribeQuotaShare

Returns a description of the specified quota share.

Request Syntax

```
POST /v1/describequotashare HTTP/1.1
Content-type: application/json

{
   "quotaShareArn": "string"
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

quotaShareArn

The Amazon Resource Name (ARN) of the quota share.

Type: String

Required: Yes

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
   "capacityLimits": [
      {
         "capacityUnit": "string",
         "maxCapacity": number
      }
```

```
    ],
    "jobQueueArn": "string",
    "preemptionConfiguration": {
       "inSharePreemption": "string"
    },
    "quotaShareArn": "string",
    "quotaShareName": "string",
    "resourceSharingConfiguration": {
       "borrowLimit": number,
       "strategy": "string"
    },
    "state": "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.

capacityLimits

A list that specifies the quantity and type of compute capacity allocated to the quota share.

Type: Array of QuotaShareCapacityLimit objects

jobQueueArn

The ARN of the job queue associated with the quota share.

Type: String

preemptionConfiguration

Specifies the preemption behavior for jobs in a quota share.

Type: QuotaSharePreemptionConfiguration object

quotaShareArn

The Amazon Resource Name (ARN) of the quota share.

Type: String

quotaShareName

The name of the quota share.

Type: String

resourceSharingConfiguration

Specifies whether a quota share reserves, lends, or both lends and borrows idle compute capacity.

Type: QuotaShareResourceSharingConfiguration object

state

The state of the quota share.

Type: String

Valid Values: `ENABLED | DISABLED`

status

The current status of the quota share.

Type: String

Valid Values: `CREATING | VALID | INVALID | UPDATING | DELETING`

tags

The tags applied to the quota share.

Type: String to string map

Map Entries: Maximum number of 50 items.

Key Length Constraints: Minimum length of 1. Maximum length of 128.

Value Length Constraints: Maximum length of 256.

Errors

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

Examples

In the following example or examples, the Authorization header contents (*[authorization-params]*) must be replaced with an AWS Signature Version 4 signature. For more information about creating these signatures, see Signature Version 4 Signing Process in the *AWS General Reference*.

You only need to learn how to sign HTTP requests if you intend to manually create them. When you use the AWS Command Line Interface (AWS CLI) or one of the AWS SDKs to make requests to AWS, these tools automatically sign the requests for you with the access key that you specify when you configure the tools. When you use these tools, you don't need to learn how to sign requests yourself.

Example

This example describes the specified quota share.

Sample Request

```
POST /v1/describequotashare HTTP/1.1
Host: batch.us-east-1.amazonaws.com
Accept-Encoding: identity
Content-Length: [content-length]
Authorization: [authorization-params]
X-Amz-Date: 20250801T001258Z
```

```
User-Agent: aws-cli/2.27.33 Python/3.13.4 Darwin/24.3.0

{
  "quotaShareArn": "arn:aws:batch:us-east-1:123456789012:job-queue/HighPriority/quota-
share/TeamAShare"
}
```

Sample Response

```
HTTP/1.1 200 OK
Content-Type: application/json
Content-Length: [content-length]
Connection: keep-alive
Date: Fri, 01 Aug 2025 00:12:59 GMT
x-amzn-RequestId: [request-id]
X-Amzn-Trace-Id: [trace-id]
X-Cache: Miss from cloudfront
Via: 1.1 example7k9m3n8q4r2w5x1z6c4vexample.cloudfront.net (CloudFront)
X-Amz-Cf-Id: whn1dX1uTx34Lvao7-7ZdkDXEbCZ_sjn3v3hHVFgbo1ORJtXyexample

{
  "quotaShareName": "TeamAShare",
  "quotaShareArn": "arn:aws:batch:us-east-1:123456789012:job-queue/HighPriority/quota-
share/TeamAShare",
  "jobQueueArn": "arn:aws:batch:us-east-1:123456789012:job-queue/HighPriority",
  "capacityLimits": [
    {
      "maxCapacity": 10,
      "capacityUnit": "ml.m5.large"
    },
    {
      "maxCapacity": 5,
      "capacityUnit": "ml.m5.xlarge"
    }
  ],
  "resourceSharingConfiguration": {
    "strategy": "LEND_AND_BORROW",
    "borrowLimit": 200
  },
  "preemptionConfiguration": {
    "inSharePreemption": "ENABLED"
  },
  "state": "ENABLED",
```

```
  "status": "VALID",
  "tags": {
    "Department": "Engineering"
  }
}
```

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for Kotlin
- AWS SDK for PHP V3
- AWS SDK for Python (Boto3)
- AWS SDK for Ruby V3

DescribeSchedulingPolicies

Describes one or more of your scheduling policies.

Request Syntax

```
POST /v1/describeschedulingpolicies HTTP/1.1
Content-type: application/json

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

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

arns

A list of up to 100 scheduling policy Amazon Resource Name (ARN) entries.

Type: Array of strings

Required: Yes

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
   "schedulingPolicies": [
      {
         "arn": "string",
         "fairsharePolicy": {
```

```
            "computeReservation": number,
            "shareDecaySeconds": number,
            "shareDistribution": [
               {
                  "shareIdentifier": "string",
                  "weightFactor": number
               }
            ]
         },
         "name": "string",
         "quotaSharePolicy": {
            "idleResourceAssignmentStrategy": "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.

schedulingPolicies

The list of scheduling policies.

Type: Array of SchedulingPolicyDetail objects

Errors

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

Examples

In the following example or examples, the Authorization header contents (*[authorization-params]*) must be replaced with an AWS Signature Version 4 signature. For more information about creating these signatures, see Signature Version 4 Signing Process in the *AWS General Reference*.

You only need to learn how to sign HTTP requests if you intend to manually create them. When you use the AWS Command Line Interface (AWS CLI) or one of the AWS SDKs to make requests to AWS, these tools automatically sign the requests for you with the access key that you specify when you configure the tools. When you use these tools, you don't need to learn how to sign requests yourself.

Example

This example describes the specified scheduling policies.

Sample Request

```
POST /v1/describeschedulingpolicies HTTP/1.1
Host: batch.us-east-1.amazonaws.com
Accept-Encoding: identity
User-Agent: aws-cli/1.20.21 Python/3.6.9 Linux/4.4.0-19041-Microsoft botocore/1.21.21
X-Amz-Date: 20210929T002059Z
X-Amz-Security-Token: [security-token]
Authorization: [authorization-params]
Content-Length: [content-length]

{
  "arns": [
    "arn:aws:batch:us-east-1:123456789012:scheduling-policy/ExampleFairSharePolicy",
    "arn:aws:batch:us-east-1:123456789012:scheduling-policy/ExampleFairSharePolicy2"
  ]
}
```

Sample Response

```
HTTP/1.1 200 OK
Date: Wed, 29 Sep 2021 00:21:00 GMT
Content-Type: application/json
Content-Length: [content-length]
x-amzn-RequestId: [request-id]
Access-Control-Allow-Origin: *
x-amz-apigw-id: [apigw-id]
Access-Control-Expose-Headers: X-amzn-errortype,X-amzn-requestid,X-amzn-errormessage,X-
amzn-trace-id,X-amz-apigw-id,date
X-Amzn-Trace-Id: [trace-id]
Connection: keep-alive

{
  "schedulingPolicies" : [ {
    "name" : "ExampleFairSharePolicy2",
    "arn" : "arn:aws:batch:us-east-1:123456789012:scheduling-policy/
ExampleFairSharePolicy2",
    "fairsharePolicy" : {
      "shareDecaySeconds" : 3600,
      "computeReservation" : 1,
      "shareDistribution" : [ {
        "shareIdentifier" : "MostImportant",
        "weightFactor" : 1.0E-4
      }, {
        "shareIdentifier" : "LeastImportant",
        "weightFactor" : 999.9999
      } ]
    },
    "tags" : {
      "Department" : "Documentation",
      "Beef" : "Ribs",
      "Hot" : "Dogs",
      "Pork" : "Shoulder"
    }
  }, {
    "name" : "ExampleFairSharePolicy",
    "arn" : "arn:aws:batch:us-east-1:123456789012:scheduling-policy/
ExampleFairSharePolicy",
    "fairsharePolicy" : {
      "shareDecaySeconds" : 3600,
      "computeReservation" : 1,
      "shareDistribution" : [ {
```

```
        "shareIdentifier" : "A1*",
        "weightFactor" : 0.1
      }, {
        "shareIdentifier" : "A2",
        "weightFactor" : 0.2
      }, {
        "shareIdentifier" : "B*",
        "weightFactor" : 0.8
      }, {
        "shareIdentifier" : "C",
        "weightFactor" : 1.2
      }, {
        "shareIdentifier" : "D*",
        "weightFactor" : 1.5
      }, {
        "shareIdentifier" : "E",
        "weightFactor" : 1.8
      } ]
    },
    "tags" : {
      "Department" : "Engineering",
      "Beef" : "Brisket",
      "Hot" : "Dog",
      "Pork" : "Ribs"
    }
  } ]
}
```

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for Kotlin

- AWS SDK for PHP V3
- AWS SDK for Python (Boto3)
- AWS SDK for Ruby V3

DescribeServiceEnvironments

Describes one or more of your service environments.

Request Syntax

```
POST /v1/describeserviceenvironments HTTP/1.1
Content-type: application/json

{
   "maxResults": number,
   "nextToken": "string",
   "serviceEnvironments": [ "string" ]
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

maxResults

The maximum number of results returned by `DescribeServiceEnvironments` in paginated output. When this parameter is used, `DescribeServiceEnvironments` only returns `maxResults` results in a single page and a `nextToken` response element. The remaining results of the initial request can be seen by sending another `DescribeServiceEnvironments` request with the returned `nextToken` value. This value can be between 1 and 100. If this parameter isn't used, then `DescribeServiceEnvironments` returns up to 100 results and a `nextToken` value if applicable.

Type: Integer

Required: No

nextToken

The `nextToken` value returned from a previous paginated `DescribeServiceEnvironments` request where `maxResults` was used and the results exceeded the value of that parameter.

Pagination continues from the end of the previous results that returned the `nextToken` value. This value is `null` when there are no more results to return.

> **Note**
>
> Treat this token as an opaque identifier that's only used to retrieve the next items in a list and not for other programmatic purposes.

Type: String

Required: No

serviceEnvironments

An array of service environment names or ARN entries.

Type: Array of strings

Required: No

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
   "nextToken": "string",
   "serviceEnvironments": [
      {
         "capacityLimits": [
            {
               "capacityUnit": "string",
               "maxCapacity": number
            }
         ],
         "serviceEnvironmentArn": "string",
         "serviceEnvironmentName": "string",
         "serviceEnvironmentType": "string",
         "state": "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.

nextToken

The `nextToken` value to include in a future `DescribeServiceEnvironments` request. When the results of a `DescribeServiceEnvironments` request exceed `maxResults`, this value can be used to retrieve the next page of results. This value is `null` when there are no more results to return.

Type: String

serviceEnvironments

The list of service environments that match the request.

Type: Array of ServiceEnvironmentDetail objects

Errors

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

Examples

In the following example or examples, the Authorization header contents (*[authorization-params]*) must be replaced with an AWS Signature Version 4 signature. For more information about creating these signatures, see Signature Version 4 Signing Process in the *AWS General Reference*.

You only need to learn how to sign HTTP requests if you intend to manually create them. When you use the AWS Command Line Interface (AWS CLI) or one of the AWS SDKs to make requests to AWS, these tools automatically sign the requests for you with the access key that you specify when you configure the tools. When you use these tools, you don't need to learn how to sign requests yourself.

Example

This example describes the specified service environment.

Sample Request

```
POST /v1/describeserviceenvironments HTTP/1.1
Host: batch.us-east-1.amazonaws.com
Accept-Encoding: identity
Content-Length: [content-length]
Authorization: [authorization-params]
X-Amz-Date: 20250801T142030Z
User-Agent: aws-cli/2.27.33 Python/3.13.4 Darwin/24.3.0

{
  "serviceEnvironments": ["SageMakerTrainingEnv"]
}
```

Sample Response

```
HTTP/1.1 200 OK
Content-Type: application/json
Content-Length: [content-length]
Connection: keep-alive
Date: Mon, 01 Aug 2016 14:20:31 GMT
x-amzn-RequestId: [request-id]
X-Amzn-Trace-Id: [trace-id]
X-Cache: Miss from cloudfront
Via: 1.1 example2m8k7n9p6q3r4s5t8w1xexample.cloudfront.net (CloudFront)
```

```
X-Amz-Cf-Id: abc2def5ghi8jkl1mno4pqr7stu0vwx3yz6789abcdefghijklmnopqexample

{
  "serviceEnvironments": [
    {
      "serviceEnvironmentName": "SageMakerTrainingEnv",
      "serviceEnvironmentArn": "arn:aws:batch:us-east-1:123456789012:service-
environment/SageMakerTrainingEnv",
      "serviceEnvironmentType": "SAGEMAKER_TRAINING",
      "state": "ENABLED",
      "status": "VALID",
      "capacityLimits": [
        {
          "maxCapacity": 50,
          "capacityUnit": "NUM_INSTANCES"
        }
      ],
      "tags": {}
    }
  ]
}
```

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for Kotlin
- AWS SDK for PHP V3
- AWS SDK for Python (Boto3)
- AWS SDK for Ruby V3

DescribeServiceJob

The details of a service job.

Request Syntax

```
POST /v1/describeservicejob HTTP/1.1
Content-type: application/json

{
   "jobId": "string"
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

jobId

The job ID for the service job to describe.

Type: String

Required: Yes

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
   "attempts": [
      {
         "serviceResourceId": {
            "name": "string",
            "value": "string"
         },
```

```
            "startedAt": number,
            "statusReason": "string",
            "stoppedAt": number
         }
      ],
      "capacityUsage": [
         {
            "capacityUnit": "string",
            "quantity": number
         }
      ],
      "createdAt": number,
      "isTerminated": boolean,
      "jobArn": "string",
      "jobId": "string",
      "jobName": "string",
      "jobQueue": "string",
      "latestAttempt": {
         "serviceResourceId": {
            "name": "string",
            "value": "string"
         }
      },
      "preemptionConfiguration": {
         "preemptionRetriesBeforeTermination": number
      },
      "preemptionSummary": {
         "preemptedAttemptCount": number,
         "recentPreemptedAttempts": [
            {
               "serviceResourceId": {
                  "name": "string",
                  "value": "string"
               },
               "startedAt": number,
               "statusReason": "string",
               "stoppedAt": number
            }
         ]
      },
      "quotaShareName": "string",
      "retryStrategy": {
         "attempts": number,
         "evaluateOnExit": [
```

```
            {
               "action": "string",
               "onStatusReason": "string"
            }
         ]
      },
      "scheduledAt": number,
      "schedulingPriority": number,
      "serviceJobType": "string",
      "serviceRequestPayload": "string",
      "shareIdentifier": "string",
      "startedAt": number,
      "status": "string",
      "statusReason": "string",
      "stoppedAt": number,
      "tags": {
         "string" : "string"
      },
      "timeoutConfig": {
         "attemptDurationSeconds": 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.

attempts

A list of job attempts associated with the service job.

Type: Array of ServiceJobAttemptDetail objects

capacityUsage

The configured capacity for the service job, such as the number of instances. The number of instances should be the same value as the `serviceRequestPayload.InstanceCount` field.

Type: Array of ServiceJobCapacityUsageDetail objects

createdAt

The Unix timestamp (in milliseconds) for when the service job was created.

Type: Long

isTerminated

Indicates whether the service job has been terminated.

Type: Boolean

jobArn

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

Type: String

jobId

The job ID for the service job.

Type: String

jobName

The name of the service job.

Type: String

jobQueue

The ARN of the job queue that the service job is associated with.

Type: String

latestAttempt

The latest attempt associated with the service job.

Type: LatestServiceJobAttempt object

preemptionConfiguration

Specifies the service job behavior when preempted.

Type: ServiceJobPreemptionConfiguration object

preemptionSummary

Summarizes the preemptions of the service job. This field appears on a service job when it has been preempted.

Type: ServiceJobPreemptionSummary object

quotaShareName

The name of the quota share that the service job is associated with.

Type: String

retryStrategy

The retry strategy to use for failed service jobs that are submitted with this service job.

Type: ServiceJobRetryStrategy object

scheduledAt

The Unix timestamp (in milliseconds) for when the service job was scheduled. This represents when the service job was dispatched to SageMaker and the service job transitioned to the SCHEDULED state.

Type: Long

schedulingPriority

The scheduling priority of the service job.

Type: Integer

serviceJobType

The type of service job. For SageMaker Training jobs, this value is SAGEMAKER_TRAINING.

Type: String

Valid Values: SAGEMAKER_TRAINING

serviceRequestPayload

The request, in JSON, for the service that the SubmitServiceJob operation is queueing.

Type: String

shareIdentifier

The share identifier for the service job. This is used for fair-share scheduling.

Type: String

startedAt

The Unix timestamp (in milliseconds) for when the service job was started.

Type: Long

status

The current status of the service job.

Type: String

Valid Values: `SUBMITTED | PENDING | RUNNABLE | SCHEDULED | STARTING | RUNNING | SUCCEEDED | FAILED`

statusReason

A short, human-readable string to provide more details for the current status of the service job.

Type: String

stoppedAt

The Unix timestamp (in milliseconds) for when the service job stopped running.

Type: Long

tags

The tags that are associated with the service job. Each tag consists of a key and an optional value. For more information, see Tagging your AWS Batch resources.

Type: String to string map

Map Entries: Maximum number of 50 items.

Key Length Constraints: Minimum length of 1. Maximum length of 128.

Value Length Constraints: Maximum length of 256.

timeoutConfig

The timeout configuration for the service job.

Type: ServiceJobTimeout object

Errors

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

Examples

In the following example or examples, the Authorization header contents (*[authorization-params]*) must be replaced with an AWS Signature Version 4 signature. For more information about creating these signatures, see Signature Version 4 Signing Process in the *AWS General Reference*.

You only need to learn how to sign HTTP requests if you intend to manually create them. When you use the AWS Command Line Interface (AWS CLI) or one of the AWS SDKs to make requests to AWS, these tools automatically sign the requests for you with the access key that you specify when you configure the tools. When you use these tools, you don't need to learn how to sign requests yourself.

Example

This example describes the specified service job.

Sample Request

```
POST /v1/describeservicejob HTTP/1.1
Host: batch.us-east-1.amazonaws.com
Accept-Encoding: identity
Content-Length: [content-length]
Authorization: [authorization-params]
X-Amz-Date: 20250801T092145Z
```

```
User-Agent: aws-cli/2.27.33 Python/3.13.4 Darwin/24.3.0

{
  "jobId": "a4d6c728-8ee8-4c65-8e2a-9a5e8f4b7c3d"
}
```

Sample Response

```
HTTP/1.1 200 OK
Content-Type: application/json
Content-Length: [content-length]
Connection: keep-alive
Date: Fri, 01 Aug 2025 09:21:46 GMT
x-amzn-RequestId: [request-id]
X-Amzn-Trace-Id: [trace-id]
X-Cache: Miss from cloudfront
Via: 1.1 49j3k6m2n8p5q7r4s1t9w2f76dsexample.cloudfront.net (CloudFront)
X-Amz-Cf-Id: def3ghi6jkl9mno2pqr5stu8vwx2yz5012cdefghijklmnopqrexample

{
  "jobId": "a4d6c728-8ee8-4c65-8e2a-9a5e8f4b7c3d",
  "jobName": "sagemaker-training-job-example",
  "jobArn": "arn:aws:batch:us-east-1:123456789012:service-job/
a4d6c728-8ee8-4c65-8e2a-9a5e8f4b7c3d",
  "jobQueue": "arn:aws:batch:us-east-1:123456789012:job-queue/sagemaker-training-
queue",
  "status": "SUCCEEDED",
  "statusReason": "Job completed successfully",
  "serviceJobType": "SAGEMAKER_TRAINING",
  "schedulingPriority": 100,
  "isTerminated": false,
  "createdAt": 1722507600000,
  "startedAt": 1722507660000,
  "stoppedAt": 1722511260000,
  "serviceRequestPayload": "{\"TrainingJobName\": \"sagemaker-training-job-example
\", \"AlgorithmSpecification\": {\"TrainingImage\": \"123456789012.dkr.ecr.us-
east-1.amazonaws.com/pytorch-inference:1.8.0-cpu-py3\", \"TrainingInputMode
\": \"File\", \"ContainerEntrypoint\": [\"sleep\", \"1\"]}, \"RoleArn\":
\"arn:aws:iam::123456789012:role/SageMakerExecutionRole\", \"OutputDataConfig\":
 {\"S3OutputPath\": \"s3://example-bucket/model-output/\"}, \"ResourceConfig\":
 {\"InstanceType\": \"ml.m5.large\", \"InstanceCount\": 1, \"VolumeSizeInGB\": 1}}",
  "latestAttempt": {
   "serviceResourceId": {
```

```
      "name": "TrainingJobArn",
      "value": "arn:aws:sagemaker:us-east-1:123456789012:training-job/sagemaker-
training-job-example"
      }
  },
  "attempts": [
    {
      "serviceResourceId": {
      "name": "TrainingJobArn",
      "value": "arn:aws:sagemaker:us-east-1:123456789012:training-job/sagemaker-
training-job-example"
      },
      "startedAt": 1722507660000,
      "stoppedAt": 1722511260000,
      "statusReason": "Completed"
    }
  ],
  "tags": {}
}
```

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for Kotlin
- AWS SDK for PHP V3
- AWS SDK for Python (Boto3)
- AWS SDK for Ruby V3

GetJobQueueSnapshot

Provides a snapshot of job queue state, including ordering of RUNNABLE jobs, as well as capacity utilization for already dispatched jobs. The first 100 RUNNABLE jobs in the job queue are listed in order of dispatch. For job queues with an attached quota-share policy, the first RUNNABLE job in each quota share is also listed. Capacity utilization for the job queue is provided, as well as break downs by share for job queues with attached fair-share or quota-share scheduling policies.

Request Syntax

```
POST /v1/getjobqueuesnapshot HTTP/1.1
Content-type: application/json

{
   "jobQueue": "string"
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

jobQueue

The job queue's name or full queue Amazon Resource Name (ARN).

Type: String

Required: Yes

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
   "frontOfQueue": {
      "jobs": [
```

```
            {
               "earliestTimeAtPosition": number,
               "jobArn": "string"
            }
         ],
         "lastUpdatedAt": number
      },
      "frontOfQuotaShares": {
         "lastUpdatedAt": number,
         "quotaShares": {
            "string" : [
               {
                  "earliestTimeAtPosition": number,
                  "jobArn": "string"
               }
            ]
         }
      },
      "queueUtilization": {
         "fairshareUtilization": {
            "activeShareCount": number,
            "topCapacityUtilization": [
               {
                  "capacityUsage": [
                     {
                        "capacityUnit": "string",
                        "quantity": number
                     }
                  ],
                  "shareIdentifier": "string"
               }
            ]
         },
         "lastUpdatedAt": number,
         "quotaShareUtilization": {
            "topCapacityUtilization": [
               {
                  "capacityUsage": [
                     {
                        "capacityUnit": "string",
                        "quantity": number
                     }
                  ],
                  "quotaShareName": "string"
```

```
            }
         ]
      },
      "totalCapacityUsage": [
         {
            "capacityUnit": "string",
            "quantity": 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.

frontOfQueue

The list of the first 100 RUNNABLE jobs in each job queue. For first-in-first-out (FIFO) job queues, jobs are ordered based on their submission time. For job queues with an attached fair-share scheduling (FSS) or quota-share policy, jobs are ordered based on their job priority and share usage.

Type: FrontOfQueueDetail object

frontOfQuotaShares

The first RUNNABLE job in each quota share. Jobs are ordered based on their job priority and share usage.

Type: FrontOfQuotaSharesDetail object

queueUtilization

The job queue's capacity utilization, including total usage and breakdown per given share.

Type: QueueSnapshotUtilizationDetail object

Errors

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for Kotlin
- AWS SDK for PHP V3
- AWS SDK for Python (Boto3)
- AWS SDK for Ruby V3

ListConsumableResources

Returns a list of AWS Batch consumable resources.

Request Syntax

```
POST /v1/listconsumableresources HTTP/1.1
Content-type: application/json

{
   "filters": [
      {
         "name": "string",
         "values": [ "string" ]
      }
   ],
   "maxResults": number,
   "nextToken": "string"
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

filters

The filters to apply to the consumable resource list query. If used, only those consumable resources that match the filter are listed. Filter names and values can be:

- name: CONSUMABLE_RESOURCE_NAME

  values: case-insensitive matches for the consumable resource name. If a filter value ends with an asterisk (*), it matches any consumable resource name that begins with the string before the '*'.

Type: Array of KeyValuesPair objects

Required: No

maxResults

The maximum number of results returned by `ListConsumableResources` in paginated output. When this parameter is used, `ListConsumableResources` only returns `maxResults` results in a single page and a `nextToken` response element. The remaining results of the initial request can be seen by sending another `ListConsumableResources` request with the returned `nextToken` value. This value can be between 1 and 100. If this parameter isn't used, then `ListConsumableResources` returns up to 100 results and a `nextToken` value if applicable.

Type: Integer

Required: No

nextToken

The `nextToken` value returned from a previous paginated `ListConsumableResources` request where `maxResults` was used and the results exceeded the value of that parameter. Pagination continues from the end of the previous results that returned the `nextToken` value. This value is `null` when there are no more results to return.

> **Note**
>
> Treat this token as an opaque identifier that's only used to retrieve the next items in a list and not for other programmatic purposes.

Type: String

Required: No

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
   "consumableResources": [
      {
         "consumableResourceArn": "string",
```

```
            "consumableResourceName": "string",
            "inUseQuantity": number,
            "resourceType": "string",
            "totalQuantity": 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.

consumableResources

A list of consumable resources that match the request.

Type: Array of ConsumableResourceSummary objects

nextToken

The `nextToken` value to include in a future `ListConsumableResources` request. When the results of a `ListConsumableResources` request exceed `maxResults`, this value can be used to retrieve the next page of results. This value is `null` when there are no more results to return.

Type: String

Errors

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for Kotlin
- AWS SDK for PHP V3
- AWS SDK for Python (Boto3)
- AWS SDK for Ruby V3

ListJobs

Returns a list of AWS Batch jobs.

You must specify only one of the following items:

- A job queue ID to return a list of jobs in that job queue
- A multi-node parallel job ID to return a list of nodes for that job
- An array job ID to return a list of the children for that job

Request Syntax

```
POST /v1/listjobs HTTP/1.1
Content-type: application/json

{
   "arrayJobId": "string",
   "filters": [
      {
         "name": "string",
         "values": [ "string" ]
      }
   ],
   "jobQueue": "string",
   "jobStatus": "string",
   "maxResults": number,
   "multiNodeJobId": "string",
   "nextToken": "string"
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

arrayJobId

The job ID for an array job. Specifying an array job ID with this parameter lists all child jobs from within the specified array.

Type: String

Required: No

filters

The filter to apply to the query. Only one filter can be used at a time. When the filter is used, `jobStatus` is ignored with the exception that `SHARE_IDENTIFIER` and `jobStatus` can be used together. The filter doesn't apply to child jobs in an array or multi-node parallel (MNP) jobs. The results are sorted by the `createdAt` field, with the most recent jobs being first.

> **Note**
>
> The `SHARE_IDENTIFIER` filter and the `jobStatus` field can be used together to filter results.

JOB_NAME

The value of the filter is a case-insensitive match for the job name. If the value ends with an asterisk (*), the filter matches any job name that begins with the string before the '*'. This corresponds to the `jobName` value. For example, `test1` matches both `Test1` and `test1`, and `test1*` matches both `test1` and `Test10`. When the `JOB_NAME` filter is used, the results are grouped by the job name and version.

JOB_DEFINITION

The value for the filter is the name or Amazon Resource Name (ARN) of the job definition. This corresponds to the `jobDefinition` value. The value is case sensitive. When the value for the filter is the job definition name, the results include all the jobs that used any revision of that job definition name. If the value ends with an asterisk (*), the filter matches any job definition name that begins with the string before the '*'. For example, `jd1` matches only `jd1`, and `jd1*` matches both `jd1` and `jd1A`. The version of the job definition that's used doesn't affect the sort order. When the `JOB_DEFINITION` filter is used and the ARN is used (which is in the form `arn:${Partition}:batch:${Region}:${Account}:job-`

definition/${JobDefinitionName}:${Revision}), the results include jobs that used the specified revision of the job definition. Asterisk (*) isn't supported when the ARN is used.

BEFORE_CREATED_AT

The value for the filter is the time that's before the job was created. This corresponds to the createdAt value. The value is a string representation of the number of milliseconds since 00:00:00 UTC (midnight) on January 1, 1970.

AFTER_CREATED_AT

The value for the filter is the time that's after the job was created. This corresponds to the createdAt value. The value is a string representation of the number of milliseconds since 00:00:00 UTC (midnight) on January 1, 1970.

SHARE_IDENTIFIER

The value for the filter is the fairshare scheduling share identifier.

Type: Array of KeyValuesPair objects

Required: No

jobQueue

The name or full Amazon Resource Name (ARN) of the job queue used to list jobs.

Type: String

Required: No

jobStatus

The job status used to filter jobs in the specified queue. If the filters parameter is specified, the jobStatus parameter is ignored and jobs with any status are returned. The exception is the SHARE_IDENTIFIER filter and jobStatus can be used together. If you don't specify a status, only RUNNING jobs are returned.

> **Note**
>
> Array job parents are updated to PENDING when any child job is updated to RUNNABLE and remain in PENDING status while child jobs are running. To view these jobs, filter by PENDING status until all child jobs reach a terminal state.

Type: String

Valid Values: `SUBMITTED | PENDING | RUNNABLE | STARTING | RUNNING | SUCCEEDED | FAILED`

Required: No

maxResults

The maximum number of results returned by `ListJobs` in a paginated output. When this parameter is used, `ListJobs` returns up to `maxResults` results in a single page and a `nextToken` response element, if applicable. The remaining results of the initial request can be seen by sending another `ListJobs` request with the returned `nextToken` value.

The following outlines key parameters and limitations:

- The minimum value is 1.
- When `--job-status` is used, AWS Batch returns up to 1000 values.
- When `--filters` is used, AWS Batch returns up to 100 values.
- If neither parameter is used, then `ListJobs` returns up to 1000 results (jobs that are in the `RUNNING` status) and a `nextToken` value, if applicable.

Type: Integer

Required: No

multiNodeJobId

The job ID for a multi-node parallel job. Specifying a multi-node parallel job ID with this parameter lists all nodes that are associated with the specified job.

Type: String

Required: No

nextToken

The `nextToken` value returned from a previous paginated `ListJobs` request where `maxResults` was used and the results exceeded the value of that parameter. Pagination continues from the end of the previous results that returned the `nextToken` value. This value is `null` when there are no more results to return.

> **Note**
>
> Treat this token as an opaque identifier that's only used to retrieve the next items in a list and not for other programmatic purposes.

Type: String

Required: No

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
   "jobSummaryList": [
      {
         "arrayProperties": {
            "index": number,
            "size": number,
            "statusSummary": {
               "string" : number
            },
            "statusSummaryLastUpdatedAt": number
         },
         "capacityUsage": [
            {
               "capacityUnit": "string",
               "quantity": number
            }
         ],
         "container": {
            "exitCode": number,
            "reason": "string"
         },
         "createdAt": number,
         "isCancelled": boolean,
         "isTerminated": boolean,
         "jobArn": "string",
         "jobDefinition": "string",
```

```
            "jobId": "string",
            "jobName": "string",
            "nodeProperties": {
               "isMainNode": boolean,
               "nodeIndex": number,
               "numNodes": number
            },
            "scheduledAt": number,
            "shareIdentifier": "string",
            "startedAt": number,
            "status": "string",
            "statusReason": "string",
            "stoppedAt": 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.

jobSummaryList

A list of job summaries that match the request.

Type: Array of JobSummary objects

nextToken

The `nextToken` value to include in a future `ListJobs` request. When the results of a `ListJobs` request exceed `maxResults`, this value can be used to retrieve the next page of results. This value is `null` when there are no more results to return.

Type: String

Errors

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

Examples

In the following example or examples, the Authorization header contents (*[authorization-params]*) must be replaced with an AWS Signature Version 4 signature. For more information about creating these signatures, see Signature Version 4 Signing Process in the *AWS General Reference*.

You only need to learn how to sign HTTP requests if you intend to manually create them. When you use the AWS Command Line Interface (AWS CLI) or one of the AWS SDKs to make requests to AWS, these tools automatically sign the requests for you with the access key that you specify when you configure the tools. When you use these tools, you don't need to learn how to sign requests yourself.

Example

This example lists the running jobs in the `HighPriority` job queue.

Sample Request

```
POST /v1/listjobs HTTP/1.1
Host: batch.us-east-1.amazonaws.com
Accept-Encoding: identity
Content-Length: [content-length]
Authorization: [authorization-params]
X-Amz-Date: 20161129T201622Z
```

```
User-Agent: aws-cli/1.11.22 Python/2.7.12 Darwin/16.1.0 botocore/1.4.79

{
  "jobQueue": "HighPriority"
}
```

Sample Response

```
HTTP/1.1 200 OK
Content-Type: application/json
Content-Length: [content-length]
Connection: keep-alive
Date: Tue, 29 Nov 2016 20:16:22 GMT
x-amzn-RequestId: [request-id]
X-Amzn-Trace-Id: [trace-id]
X-Cache: Miss from cloudfront
Via: 1.1 7f3f42df8af148df1f9f1ee7175987ad.cloudfront.net (CloudFront)
X-Amz-Cf-Id: idKR5mD8f7Luom03P9DV1bFGXsq_SIFNy_nMrTCOqZrRc0nXgHqZfg==

{
  "jobSummaryList": [{
    "jobId": "e66ff5fd-a1ff-4640-b1a2-0b0a142f49bb",
    "jobName": "example"
  }]
}
```

Example

This example lists jobs in the `HighPriority` job queue that are in the `SUBMITTED` job status.

Sample Request

```
POST /v1/listjobs HTTP/1.1
Host: batch.us-east-1.amazonaws.com
Accept-Encoding: identity
Content-Length: [content-length]
Authorization: [authorization-params]
X-Amz-Date: 20161129T201642Z
User-Agent: aws-cli/1.11.22 Python/2.7.12 Darwin/16.1.0 botocore/1.4.79

{
  "jobQueue": "HighPriority",
  "jobStatus": "SUBMITTED"
```

```
}
```

Sample Response

```
HTTP/1.1 200 OK
Date: Tue, 29 Nov 2016 20:16:42 GMT
Content-Type: application/json
Content-Length: [content-length]
Connection: keep-alive
x-amzn-RequestId: [request-id]
X-Amzn-Trace-Id: [trace-id]
X-Cache: Miss from cloudfront
Via: 1.1 ebc28fb0ad14691ee5d6c1a49f41b878.cloudfront.net (CloudFront)
X-Amz-Cf-Id: Ngsjm0gBg2y4cDFG4uwpAmaKaT6Dejh7oGlVDmewQrUaeW_SPst_Bw==

{
  "jobSummaryList": [{
    "jobId": "68f0c163-fbd4-44e6-9fd1-25b14a434786",
    "jobName": "example"
  }]
}
```

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for Kotlin
- AWS SDK for PHP V3
- AWS SDK for Python (Boto3)
- AWS SDK for Ruby V3

ListJobsByConsumableResource

Returns a list of AWS Batch jobs that require a specific consumable resource.

Request Syntax

```
POST /v1/listjobsbyconsumableresource HTTP/1.1
Content-type: application/json

{
   "consumableResource": "string",
   "filters": [
      {
         "name": "string",
         "values": [ "string" ]
      }
   ],
   "maxResults": number,
   "nextToken": "string"
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

consumableResource

The name or ARN of the consumable resource.

Type: String

Required: Yes

filters

The filters to apply to the job list query. If used, only those jobs requiring the specified consumable resource (`consumableResource`) and that match the value of the filters are listed. The filter names and values can be:

- name: JOB_STATUS

  values: SUBMITTED | PENDING | RUNNABLE | STARTING | RUNNING | SUCCEEDED | FAILED

- name: JOB_NAME

  The values are case-insensitive matches for the job name. If a filter value ends with an asterisk (*), it matches any job name that begins with the string before the '*'.

Type: Array of KeyValuesPair objects

Required: No

maxResults

The maximum number of results returned by ListJobsByConsumableResource in paginated output. When this parameter is used, ListJobsByConsumableResource only returns maxResults results in a single page and a nextToken response element. The remaining results of the initial request can be seen by sending another ListJobsByConsumableResource request with the returned nextToken value. This value can be between 1 and 100. If this parameter isn't used, then ListJobsByConsumableResource returns up to 100 results and a nextToken value if applicable.

Type: Integer

Required: No

nextToken

The nextToken value returned from a previous paginated ListJobsByConsumableResource request where maxResults was used and the results exceeded the value of that parameter. Pagination continues from the end of the previous results that returned the nextToken value. This value is null when there are no more results to return.

> **Note**
>
> Treat this token as an opaque identifier that's only used to retrieve the next items in a list and not for other programmatic purposes.

Type: String

Required: No

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
   "jobs": [
      {
         "consumableResourceProperties": {
            "consumableResourceList": [
               {
                  "consumableResource": "string",
                  "quantity": number
               }
            ]
         },
         "createdAt": number,
         "jobArn": "string",
         "jobDefinitionArn": "string",
         "jobName": "string",
         "jobQueueArn": "string",
         "jobStatus": "string",
         "quantity": number,
         "shareIdentifier": "string",
         "startedAt": number,
         "statusReason": "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.

jobs

The list of jobs that require the specified consumable resources.

Type: Array of ListJobsByConsumableResourceSummary objects

nextToken

The `nextToken` value to include in a future `ListJobsByConsumableResource` request. When the results of a `ListJobsByConsumableResource` request exceed `maxResults`, this value can be used to retrieve the next page of results. This value is `null` when there are no more results to return.

Type: String

Errors

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for Kotlin

- AWS SDK for PHP V3
- AWS SDK for Python (Boto3)
- AWS SDK for Ruby V3

ListQuotaShares

Returns a list of AWS Batch quota shares associated with a job queue.

Request Syntax

```
POST /v1/listquotashares HTTP/1.1
Content-type: application/json

{
   "jobQueue": "string",
   "maxResults": number,
   "nextToken": "string"
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

jobQueue

The name or full Amazon Resource Name (ARN) of the job queue used to list quota shares.

Type: String

Required: Yes

maxResults

The maximum number of results returned by `ListQuotaShares` in paginated output. When this parameter is used, `ListQuotaShares` only returns `maxResults` results in a single page and a `nextToken` response element. You can see the remaining results of the initial request by sending another `ListQuotaShares` request with the returned `nextToken` value. This value can be between 1 and 100. If this parameter isn't used, `ListQuotaShares` returns up to 100 results and a `nextToken` value if applicable.

Type: Integer

Required: No

nextToken

The `nextToken` value that's returned from a previous paginated `ListQuotaShares` request where `maxResults` was used and the results exceeded the value of that parameter. Pagination continues from the end of the previous results that returned the `nextToken` value. This value is `null` when there are no more results to return.

> **Note**
>
> Treat this token as an opaque identifier that's only used to retrieve the next items in a list and not for other programmatic purposes.

Type: String

Required: No

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
   "nextToken": "string",
   "quotaShares": [
      {
         "capacityLimits": [
            {
               "capacityUnit": "string",
               "maxCapacity": number
            }
         ],
         "jobQueueArn": "string",
         "preemptionConfiguration": {
            "inSharePreemption": "string"
         },
         "quotaShareArn": "string",
         "quotaShareName": "string",
         "resourceSharingConfiguration": {
```

```
            "borrowLimit": number,
            "strategy": "string"
         },
         "state": "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

The `nextToken` value to include in a future `ListQuotaShares` request. When the results of a `ListQuotaShares` request exceed `maxResults`, this value can be used to retrieve the next page of results. This value is `null` when there are no more results to return.

Type: String

quotaShares

A list of quota shares that match the request.

Type: Array of QuotaShareDetail objects

Errors

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

Examples

In the following example or examples, the Authorization header contents (*[authorization-params]*) must be replaced with an AWS Signature Version 4 signature. For more information about creating these signatures, see Signature Version 4 Signing Process in the *AWS General Reference*.

You only need to learn how to sign HTTP requests if you intend to manually create them. When you use the AWS Command Line Interface (AWS CLI) or one of the AWS SDKs to make requests to AWS, these tools automatically sign the requests for you with the access key that you specify when you configure the tools. When you use these tools, you don't need to learn how to sign requests yourself.

Example

This example lists the quota shares associated with the `HighPriority` job queue.

Sample Request

```
POST /v1/listquotashares HTTP/1.1
Host: batch.us-east-1.amazonaws.com
Accept-Encoding: identity
Content-Length: [content-length]
Authorization: [authorization-params]
X-Amz-Date: 20250801T001258Z
User-Agent: aws-cli/2.27.33 Python/3.13.4 Darwin/24.3.0

{
  "jobQueue": "HighPriority"
}
```

Sample Response

```
HTTP/1.1 200 OK
Content-Type: application/json
Content-Length: [content-length]
Connection: keep-alive
Date: Fri, 01 Aug 2025 00:12:59 GMT
x-amzn-RequestId: [request-id]
```

```
X-Amzn-Trace-Id: [trace-id]
X-Cache: Miss from cloudfront
Via: 1.1 example7k9m3n8q4r2w5x1z6c4vexample.cloudfront.net (CloudFront)
X-Amz-Cf-Id: whn1dX1uTx34Lvao7-7ZdkDXEbCZ_sjn3v3hHVFgbo1ORJtXyexample

{
  "quotaShares": [
    {
      "quotaShareName": "TeamAShare",
      "quotaShareArn": "arn:aws:batch:us-east-1:123456789012:job-queue/HighPriority/
quota-share/TeamAShare",
      "jobQueueArn": "arn:aws:batch:us-east-1:123456789012:job-queue/HighPriority",
      "capacityLimits": [
        {
          "maxCapacity": 10,
          "capacityUnit": "ml.m5.large"
        },
        {
          "maxCapacity": 5,
          "capacityUnit": "ml.m5.xlarge"
        }
      ],
      "resourceSharingConfiguration": {
        "strategy": "LEND_AND_BORROW",
        "borrowLimit": 200
      },
      "preemptionConfiguration": {
        "inSharePreemption": "ENABLED"
      },
      "state": "ENABLED",
      "status": "VALID",
      "tags": {
        "Department": "Engineering"
      }
    },
    ...
  ]
}
```

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for Kotlin
- AWS SDK for PHP V3
- AWS SDK for Python (Boto3)
- AWS SDK for Ruby V3

ListSchedulingPolicies

Returns a list of AWS Batch scheduling policies.

Request Syntax

```
POST /v1/listschedulingpolicies HTTP/1.1
Content-type: application/json

{
   "maxResults": number,
   "nextToken": "string"
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

maxResults

The maximum number of results that's returned by `ListSchedulingPolicies` in paginated output. When this parameter is used, `ListSchedulingPolicies` only returns `maxResults` results in a single page and a `nextToken` response element. You can see the remaining results of the initial request by sending another `ListSchedulingPolicies` request with the returned `nextToken` value. This value can be between 1 and 100. If this parameter isn't used, `ListSchedulingPolicies` returns up to 100 results and a `nextToken` value if applicable.

Type: Integer

Required: No

nextToken

The `nextToken` value that's returned from a previous paginated `ListSchedulingPolicies` request where `maxResults` was used and the results exceeded the value of that parameter. Pagination continues from the end of the previous results that returned the `nextToken` value. This value is `null` when there are no more results to return.

> **Note**
>
> Treat this token as an opaque identifier that's only used to retrieve the next items in a list and not for other programmatic purposes.

Type: String

Required: No

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
   "nextToken": "string",
   "schedulingPolicies": [
      {
         "arn": "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

The `nextToken` value to include in a future `ListSchedulingPolicies` request. When the results of a `ListSchedulingPolicies` request exceed `maxResults`, this value can be used to retrieve the next page of results. This value is `null` when there are no more results to return.

Type: String

schedulingPolicies

A list of scheduling policies that match the request.

Type: Array of SchedulingPolicyListingDetail objects

Errors

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

Examples

In the following example or examples, the Authorization header contents (*`[authorization-params]`*) must be replaced with an AWS Signature Version 4 signature. For more information about creating these signatures, see Signature Version 4 Signing Process in the *AWS General Reference*.

You only need to learn how to sign HTTP requests if you intend to manually create them. When you use the AWS Command Line Interface (AWS CLI) or one of the AWS SDKs to make requests to AWS, these tools automatically sign the requests for you with the access key that you specify when you configure the tools. When you use these tools, you don't need to learn how to sign requests yourself.

Example

This example lists the scheduling policies.

Sample Request

```
POST /v1/listschedulingpolicies HTTP/1.1
Host: batch.us-east-1.amazonaws.com
Accept-Encoding: identity
```

```
User-Agent: aws-cli/1.20.21 Python/3.6.9 Linux/4.4.0-19041-Microsoft botocore/1.21.21
X-Amz-Date: 20210929T001942Z
X-Amz-Security-Token: [security-token]
Authorization: [authorization-params]
Content-Length: 0
```

Sample Response

```
HTTP/1.1 200 OK
Date: Wed, 29 Sep 2021 00:19:43 GMT
Content-Type: application/json
Content-Length: [content-length]
x-amzn-RequestId: [request-id]
Access-Control-Allow-Origin: *
x-amz-apigw-id: [apigw-id]
Access-Control-Expose-Headers: X-amzn-errortype,X-amzn-requestid,X-amzn-errormessage,X-
amzn-trace-id,X-amz-apigw-id,date
X-Amzn-Trace-Id: [trace-id]
Connection: keep-alive

{
    "schedulingPolicies": [{
        "arn": "arn:aws:batch:us-east-1:123456789012:scheduling-policy/
ExampleFairSharePolicy"
    }, {
        "arn": "arn:aws:batch:us-east-1:123456789012:scheduling-policy/
ExampleFairSharePolicy2"
    }]
}
```

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2

- AWS SDK for JavaScript V3
- AWS SDK for Kotlin
- AWS SDK for PHP V3
- AWS SDK for Python (Boto3)
- AWS SDK for Ruby V3

ListServiceJobs

Returns a list of service jobs for a specified job queue.

Request Syntax

```
POST /v1/listservicejobs HTTP/1.1
Content-type: application/json

{
   "filters": [
      {
         "name": "string",
         "values": [ "string" ]
      }
   ],
   "jobQueue": "string",
   "jobStatus": "string",
   "maxResults": number,
   "nextToken": "string"
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

filters

The filter to apply to the query. Only one filter can be used at a time. When the filter is used, `jobStatus` is ignored with the exception that `SHARE_IDENTIFIER` or `QUOTA_SHARE_NAME` and `jobStatus` can be used together. The results are sorted by the `createdAt` field, with the most recent jobs being first.

> **Note**
>
> The `SHARE_IDENTIFIER` or `QUOTA_SHARE_NAME` filter and the `jobStatus` field can be used together to filter results.

JOB_NAME

The value of the filter is a case-insensitive match for the job name. If the value ends with an asterisk (*), the filter matches any job name that begins with the string before the '*'. This corresponds to the `jobName` value. For example, `test1` matches both `Test1` and `test1`, and `test1*` matches both `test1` and `Test10`. When the `JOB_NAME` filter is used, the results are grouped by the job name and version.

BEFORE_CREATED_AT

The value for the filter is the time that's before the job was created. This corresponds to the `createdAt` value. The value is a string representation of the number of milliseconds since 00:00:00 UTC (midnight) on January 1, 1970.

AFTER_CREATED_AT

The value for the filter is the time that's after the job was created. This corresponds to the `createdAt` value. The value is a string representation of the number of milliseconds since 00:00:00 UTC (midnight) on January 1, 1970.

SHARE_IDENTIFIER

The value for the filter is the fairshare scheduling share identifier.

QUOTA_SHARE_NAME

The value for the filter is the quota management share name.

Type: Array of KeyValuesPair objects

Required: No

jobQueue

The name or ARN of the job queue with which to list service jobs.

Type: String

Required: No

jobStatus

The job status used to filter service jobs in the specified queue. If the `filters` parameter is specified, the `jobStatus` parameter is ignored and jobs with any status are returned. The exceptions are the `SHARE_IDENTIFIER` filter and `QUOTA_SHARE_NAME` filter, which can be used with `jobStatus`. If you don't specify a status, only `RUNNING` jobs are returned.

> **Note**
>
> The `SHARE_IDENTIFIER` filter or `QUOTA_SHARE_NAME` filter can be used with the `jobStatus` field to filter results.

Type: String

Valid Values: `SUBMITTED | PENDING | RUNNABLE | SCHEDULED | STARTING | RUNNING | SUCCEEDED | FAILED`

Required: No

maxResults

The maximum number of results returned by `ListServiceJobs` in paginated output. When this parameter is used, `ListServiceJobs` only returns `maxResults` results in a single page and a `nextToken` response element. The remaining results of the initial request can be seen by sending another `ListServiceJobs` request with the returned `nextToken` value. This value can be between 1 and 100. If this parameter isn't used, then `ListServiceJobs` returns up to 100 results and a `nextToken` value if applicable.

Type: Integer

Required: No

nextToken

The `nextToken` value returned from a previous paginated `ListServiceJobs` request where `maxResults` was used and the results exceeded the value of that parameter. Pagination continues from the end of the previous results that returned the `nextToken` value. This value is `null` when there are no more results to return.

> **Note**
>
> Treat this token as an opaque identifier that's only used to retrieve the next items in a list and not for other programmatic purposes.

Type: String

Required: No

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
   "jobSummaryList": [
      {
         "capacityUsage": [
            {
               "capacityUnit": "string",
               "quantity": number
            }
         ],
         "createdAt": number,
         "isTerminated": boolean,
         "jobArn": "string",
         "jobId": "string",
         "jobName": "string",
         "latestAttempt": {
            "serviceResourceId": {
               "name": "string",
               "value": "string"
            }
         },
         "quotaShareName": "string",
         "scheduledAt": number,
         "serviceJobType": "string",
         "shareIdentifier": "string",
         "startedAt": number,
         "status": "string",
```

```
            "statusReason": "string",
            "stoppedAt": 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.

jobSummaryList

A list of service job summaries.

Type: Array of ServiceJobSummary objects

nextToken

The `nextToken` value to include in a future `ListServiceJobs` request. When the results of a `ListServiceJobs` request exceed `maxResults`, this value can be used to retrieve the next page of results. This value is `null` when there are no more results to return.

Type: String

Errors

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

Examples

In the following example or examples, the Authorization header contents (*[authorization-params]*) must be replaced with an AWS Signature Version 4 signature. For more information about creating these signatures, see Signature Version 4 Signing Process in the *AWS General Reference*.

You only need to learn how to sign HTTP requests if you intend to manually create them. When you use the AWS Command Line Interface (AWS CLI) or one of the AWS SDKs to make requests to AWS, these tools automatically sign the requests for you with the access key that you specify when you configure the tools. When you use these tools, you don't need to learn how to sign requests yourself.

Example

This example lists all succeeded service jobs from the specified job queue.

Sample Request

```
POST /v1/listservicejobs HTTP/1.1
Host: batch.us-east-1.amazonaws.com
Accept-Encoding: identity
Content-Length: [content-length]
Authorization: [authorization-params]
X-Amz-Date: 20250801T103040Z
User-Agent: aws-cli/2.27.33 Python/3.13.4 Darwin/24.3.0

{
  "jobQueue": "sagemaker-training-queue",
  "jobStatus": "SUCCEEDED",
  "maxResults": 10
}
```

Sample Response

```
HTTP/1.1 200 OK
Content-Type: application/json
Content-Length: [content-length]
Connection: keep-alive
Date: Fri, 01 Aug 2025 10:30:41 GMT
x-amzn-RequestId: [request-id]
X-Amzn-Trace-Id: [trace-id]
```

```
X-Cache: Miss from cloudfront
Via: 1.1 fnhd4k2s8l3n6p9r2t5u8v1w4y7zexample.cloudfront.net (CloudFront)
X-Amz-Cf-Id: mno6pqr9stu2vwx5yz8901defghijklmnopqrstuvexample

{
  "jobSummaryList": [
    {
      "jobId": "a4d6c728-8ee8-4c65-8e2a-9a5e8f4b7c3d",
      "jobName": "sagemaker-training-job-example",
      "jobArn": "arn:aws:batch:us-east-1:123456789012:service-job/
a4d6c728-8ee8-4c65-8e2a-9a5e8f4b7c3d",
      "status": "SUCCEEDED",
      "serviceJobType": "SAGEMAKER_TRAINING",
      "createdAt": 1722507600000,
      "startedAt": 1722507660000,
      "stoppedAt": 1722511260000,
      "latestAttempt": {
      "serviceResourceId": {
            "name": "TrainingJobArn",
            "value": "arn:aws:sagemaker:us-east-1:123456789012:training-job/sagemaker-
training-job-example"
         }
      }
    },
    {
      "jobId": "b7e9f032-1aa2-4d78-9b3c-8e6f5a4d2c1b",
      "jobName": "image-classification-training",
      "jobArn": "arn:aws:batch:us-east-1:123456789012:service-job/
b7e9f032-1aa2-4d78-9b3c-8e6f5a4d2c1b",
      "status": "SUCCEEDED",
      "serviceJobType": "SAGEMAKER_TRAINING",
      "createdAt": 1722495000000,
      "startedAt": 1722495120000,
      "stoppedAt": 1722498720000,
      "latestAttempt": {
      "serviceResourceId": {
            "name": "TrainingJobArn",
            "value": "arn:aws:sagemaker:us-east-1:123456789012:training-job/image-
classification-training-example"
         }
      }
    }
  ]
```

```
}
```

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for Kotlin
- AWS SDK for PHP V3
- AWS SDK for Python (Boto3)
- AWS SDK for Ruby V3

ListTagsForResource

Lists the tags for an AWS Batch resource. AWS Batch resources that support tags are compute environments, jobs, job definitions, job queues, and scheduling policies. ARNs for child jobs of array and multi-node parallel (MNP) jobs aren't supported.

Request Syntax

```
GET /v1/tags/resourceArn HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

resourceArn

The Amazon Resource Name (ARN) that identifies the resource that tags are listed for. AWS Batch resources that support tags are compute environments, jobs, job definitions, job queues, and scheduling policies. ARNs for child jobs of array and multi-node parallel (MNP) jobs aren't supported.

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

The tags for the resource.

Type: String to string map

Map Entries: Maximum number of 50 items.

Key Length Constraints: Minimum length of 1. Maximum length of 128.

Value Length Constraints: Maximum length of 256.

Errors

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

Examples

In the following example or examples, the Authorization header contents (*`[authorization-params]`*) must be replaced with an AWS Signature Version 4 signature. For more information about creating these signatures, see Signature Version 4 Signing Process in the *AWS General Reference*.

You only need to learn how to sign HTTP requests if you intend to manually create them. When you use the AWS Command Line Interface (AWS CLI) or one of the AWS SDKs to make requests to AWS, these tools automatically sign the requests for you with the access key that you specify when you configure the tools. When you use these tools, you don't need to learn how to sign requests yourself.

Example

This example lists tags on the job definition with an ARN of "*arn:aws:batch:us-east-1:123456789012:job-definition/sleep30:1*".

Sample Request

```
GET /v1/tags/arn%3Aaws%3Abatch%3Aus-east-1%3A123456789012%3Ajob-definition
%2Fsleep30%3A1 HTTP/1.1
Host: batch.us-east-1.amazonaws.com
Accept-Encoding: identity
X-Amz-Date: 20200604T172419Z
X-Amz-Security-Token: [security-token]
Authorization: [authorization-params]
```

Sample Response

```
HTTP/1.1 200 OK
Date: Thu, 05 Jun 2020 17:24:20 GMT
Content-Type: application/json

x-amzn-RequestId: [request-id]
Access-Control-Allow-Origin: *
x-amz-apigw-id: [apigw-id]
X-Amzn-Trace-Id: [trace-id]
Connection: keep-alive

{
  "tags": {
    "Stage": "Alpha",
    "Department": "Engineering",
    "User": "JaneDoe"
  }
}
```

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for Kotlin
- AWS SDK for PHP V3
- AWS SDK for Python (Boto3)
- AWS SDK for Ruby V3

RegisterJobDefinition

Registers an AWS Batch job definition.

Request Syntax

```
POST /v1/registerjobdefinition HTTP/1.1
Content-type: application/json

{
   "consumableResourceProperties": {
      "consumableResourceList": [
         {
            "consumableResource": "string",
            "quantity": number
         }
      ]
   },
   "containerProperties": {
      "command": [ "string" ],
      "enableExecuteCommand": boolean,
      "environment": [
         {
            "name": "string",
            "value": "string"
         }
      ],
      "ephemeralStorage": {
         "sizeInGiB": number
      },
      "executionRoleArn": "string",
      "fargatePlatformConfiguration": {
         "platformVersion": "string"
      },
      "image": "string",
      "instanceType": "string",
      "jobRoleArn": "string",
      "linuxParameters": {
         "devices": [
            {
               "containerPath": "string",
               "hostPath": "string",
               "permissions": [ "string" ]
```

```
            }
         ],
         "initProcessEnabled": boolean,
         "maxSwap": number,
         "sharedMemorySize": number,
         "swappiness": number,
         "tmpfs": [
            {
               "containerPath": "string",
               "mountOptions": [ "string" ],
               "size": number
            }
         ]
      },
      "logConfiguration": {
         "logDriver": "string",
         "options": {
            "string" : "string"
         },
         "secretOptions": [
            {
               "name": "string",
               "valueFrom": "string"
            }
         ]
      },
      "memory": number,
      "mountPoints": [
         {
            "containerPath": "string",
            "readOnly": boolean,
            "sourceVolume": "string"
         }
      ],
      "networkConfiguration": {
         "assignPublicIp": "string"
      },
      "privileged": boolean,
      "readonlyRootFilesystem": boolean,
      "repositoryCredentials": {
         "credentialsParameter": "string"
      },
      "resourceRequirements": [
         {
```

```
            "type": "string",
            "value": "string"
         }
      ],
      "runtimePlatform": {
         "cpuArchitecture": "string",
         "operatingSystemFamily": "string"
      },
      "secrets": [
         {
            "name": "string",
            "valueFrom": "string"
         }
      ],
      "ulimits": [
         {
            "hardLimit": number,
            "name": "string",
            "softLimit": number
         }
      ],
      "user": "string",
      "vcpus": number,
      "volumes": [
         {
            "efsVolumeConfiguration": {
               "authorizationConfig": {
                  "accessPointId": "string",
                  "iam": "string"
               },
               "fileSystemId": "string",
               "rootDirectory": "string",
               "transitEncryption": "string",
               "transitEncryptionPort": number
            },
            "host": {
               "sourcePath": "string"
            },
            "name": "string",
            "s3filesVolumeConfiguration": {
               "accessPointArn": "string",
               "fileSystemArn": "string",
               "rootDirectory": "string",
               "transitEncryptionPort": number
```

```
            }
         }
      ]
   },
   "ecsProperties": {
      "taskProperties": [
         {
            "containers": [
               {
                  "command": [ "string" ],
                  "dependsOn": [
                     {
                        "condition": "string",
                        "containerName": "string"
                     }
                  ],
                  "environment": [
                     {
                        "name": "string",
                        "value": "string"
                     }
                  ],
                  "essential": boolean,
                  "firelensConfiguration": {
                     "options": {
                        "string" : "string"
                     },
                     "type": "string"
                  },
                  "image": "string",
                  "linuxParameters": {
                     "devices": [
                        {
                           "containerPath": "string",
                           "hostPath": "string",
                           "permissions": [ "string" ]
                        }
                     ],
                     "initProcessEnabled": boolean,
                     "maxSwap": number,
                     "sharedMemorySize": number,
                     "swappiness": number,
                     "tmpfs": [
                        {
```

```
            "containerPath": "string",
            "mountOptions": [ "string" ],
            "size": number
         }
      ]
   },
   "logConfiguration": {
      "logDriver": "string",
      "options": {
         "string" : "string"
      },
      "secretOptions": [
         {
            "name": "string",
            "valueFrom": "string"
         }
      ]
   },
   "mountPoints": [
      {
         "containerPath": "string",
         "readOnly": boolean,
         "sourceVolume": "string"
      }
   ],
   "name": "string",
   "privileged": boolean,
   "readonlyRootFilesystem": boolean,
   "repositoryCredentials": {
      "credentialsParameter": "string"
   },
   "resourceRequirements": [
      {
         "type": "string",
         "value": "string"
      }
   ],
   "secrets": [
      {
         "name": "string",
         "valueFrom": "string"
      }
   ],
   "startTimeout": number,
```

```
            "stopTimeout": number,
            "ulimits": [
               {
                  "hardLimit": number,
                  "name": "string",
                  "softLimit": number
               }
            ],
            "user": "string"
         }
      ],
      "enableExecuteCommand": boolean,
      "ephemeralStorage": {
         "sizeInGiB": number
      },
      "executionRoleArn": "string",
      "ipcMode": "string",
      "networkConfiguration": {
         "assignPublicIp": "string"
      },
      "networkMode": "string",
      "pidMode": "string",
      "platformVersion": "string",
      "runtimePlatform": {
         "cpuArchitecture": "string",
         "operatingSystemFamily": "string"
      },
      "taskRoleArn": "string",
      "volumes": [
         {
            "efsVolumeConfiguration": {
               "authorizationConfig": {
                  "accessPointId": "string",
                  "iam": "string"
               },
               "fileSystemId": "string",
               "rootDirectory": "string",
               "transitEncryption": "string",
               "transitEncryptionPort": number
            },
            "host": {
               "sourcePath": "string"
            },
            "name": "string",
```

```
                "s3filesVolumeConfiguration": {
                   "accessPointArn": "string",
                   "fileSystemArn": "string",
                   "rootDirectory": "string",
                   "transitEncryptionPort": number
                }
             }
          ]
       }
    ]
 },
 "eksProperties": {
    "podProperties": {
       "containers": [
          {
             "args": [ "string" ],
             "command": [ "string" ],
             "env": [
                {
                   "name": "string",
                   "value": "string"
                }
             ],
             "image": "string",
             "imagePullPolicy": "string",
             "name": "string",
             "resources": {
                "limits": {
                   "string" : "string"
                },
                "requests": {
                   "string" : "string"
                }
             },
             "securityContext": {
                "allowPrivilegeEscalation": boolean,
                "privileged": boolean,
                "readOnlyRootFilesystem": boolean,
                "runAsGroup": number,
                "runAsNonRoot": boolean,
                "runAsUser": number
             },
             "volumeMounts": [
                {
```

```
                    "mountPath": "string",
                    "name": "string",
                    "readOnly": boolean,
                    "subPath": "string"
                 }
              ]
           }
        ],
        "dnsPolicy": "string",
        "hostNetwork": boolean,
        "imagePullSecrets": [
           {
              "name": "string"
           }
        ],
        "initContainers": [
           {
              "args": [ "string" ],
              "command": [ "string" ],
              "env": [
                 {
                    "name": "string",
                    "value": "string"
                 }
              ],
              "image": "string",
              "imagePullPolicy": "string",
              "name": "string",
              "resources": {
                 "limits": {
                    "string" : "string"
                 },
                 "requests": {
                    "string" : "string"
                 }
              },
              "securityContext": {
                 "allowPrivilegeEscalation": boolean,
                 "privileged": boolean,
                 "readOnlyRootFilesystem": boolean,
                 "runAsGroup": number,
                 "runAsNonRoot": boolean,
                 "runAsUser": number
              },
```

```
            "volumeMounts": [
               {
                  "mountPath": "string",
                  "name": "string",
                  "readOnly": boolean,
                  "subPath": "string"
               }
            ]
         }
      ],
      "metadata": {
         "annotations": {
            "string" : "string"
         },
         "labels": {
            "string" : "string"
         },
         "namespace": "string"
      },
      "serviceAccountName": "string",
      "shareProcessNamespace": boolean,
      "volumes": [
         {
            "emptyDir": {
               "medium": "string",
               "sizeLimit": "string"
            },
            "hostPath": {
               "path": "string"
            },
            "name": "string",
            "persistentVolumeClaim": {
               "claimName": "string",
               "readOnly": boolean
            },
            "secret": {
               "optional": boolean,
               "secretName": "string"
            }
         }
      ]
   }
},
"jobDefinitionName": "string",
```

```
   "nodeProperties": {
      "mainNode": number,
      "nodeRangeProperties": [
         {
            "consumableResourceProperties": {
               "consumableResourceList": [
                  {
                     "consumableResource": "string",
                     "quantity": number
                  }
               ]
            },
            "container": {
               "command": [ "string" ],
               "enableExecuteCommand": boolean,
               "environment": [
                  {
                     "name": "string",
                     "value": "string"
                  }
               ],
               "ephemeralStorage": {
                  "sizeInGiB": number
               },
               "executionRoleArn": "string",
               "fargatePlatformConfiguration": {
                  "platformVersion": "string"
               },
               "image": "string",
               "instanceType": "string",
               "jobRoleArn": "string",
               "linuxParameters": {
                  "devices": [
                     {
                        "containerPath": "string",
                        "hostPath": "string",
                        "permissions": [ "string" ]
                     }
                  ],
                  "initProcessEnabled": boolean,
                  "maxSwap": number,
                  "sharedMemorySize": number,
                  "swappiness": number,
                  "tmpfs": [
```

```
            {
               "containerPath": "string",
               "mountOptions": [ "string" ],
               "size": number
            }
         ]
      },
      "logConfiguration": {
         "logDriver": "string",
         "options": {
            "string" : "string"
         },
         "secretOptions": [
            {
               "name": "string",
               "valueFrom": "string"
            }
         ]
      },
      "memory": number,
      "mountPoints": [
         {
            "containerPath": "string",
            "readOnly": boolean,
            "sourceVolume": "string"
         }
      ],
      "networkConfiguration": {
         "assignPublicIp": "string"
      },
      "privileged": boolean,
      "readonlyRootFilesystem": boolean,
      "repositoryCredentials": {
         "credentialsParameter": "string"
      },
      "resourceRequirements": [
         {
            "type": "string",
            "value": "string"
         }
      ],
      "runtimePlatform": {
         "cpuArchitecture": "string",
         "operatingSystemFamily": "string"
```

```
            },
            "secrets": [
               {
                  "name": "string",
                  "valueFrom": "string"
               }
            ],
            "ulimits": [
               {
                  "hardLimit": number,
                  "name": "string",
                  "softLimit": number
               }
            ],
            "user": "string",
            "vcpus": number,
            "volumes": [
               {
                  "efsVolumeConfiguration": {
                     "authorizationConfig": {
                        "accessPointId": "string",
                        "iam": "string"
                     },
                     "fileSystemId": "string",
                     "rootDirectory": "string",
                     "transitEncryption": "string",
                     "transitEncryptionPort": number
                  },
                  "host": {
                     "sourcePath": "string"
                  },
                  "name": "string",
                  "s3filesVolumeConfiguration": {
                     "accessPointArn": "string",
                     "fileSystemArn": "string",
                     "rootDirectory": "string",
                     "transitEncryptionPort": number
                  }
               }
            ]
         },
         "ecsProperties": {
            "taskProperties": [
               {
```

```
   "containers": [
      {
         "command": [ "string" ],
         "dependsOn": [
            {
               "condition": "string",
               "containerName": "string"
            }
         ],
         "environment": [
            {
               "name": "string",
               "value": "string"
            }
         ],
         "essential": boolean,
         "firelensConfiguration": {
            "options": {
               "string" : "string"
            },
            "type": "string"
         },
         "image": "string",
         "linuxParameters": {
            "devices": [
               {
                  "containerPath": "string",
                  "hostPath": "string",
                  "permissions": [ "string" ]
               }
            ],
            "initProcessEnabled": boolean,
            "maxSwap": number,
            "sharedMemorySize": number,
            "swappiness": number,
            "tmpfs": [
               {
                  "containerPath": "string",
                  "mountOptions": [ "string" ],
                  "size": number
               }
            ]
         },
         "logConfiguration": {
```

```
            "logDriver": "string",
            "options": {
               "string" : "string"
            },
            "secretOptions": [
               {
                  "name": "string",
                  "valueFrom": "string"
               }
            ]
         },
         "mountPoints": [
            {
               "containerPath": "string",
               "readOnly": boolean,
               "sourceVolume": "string"
            }
         ],
         "name": "string",
         "privileged": boolean,
         "readonlyRootFilesystem": boolean,
         "repositoryCredentials": {
            "credentialsParameter": "string"
         },
         "resourceRequirements": [
            {
               "type": "string",
               "value": "string"
            }
         ],
         "secrets": [
            {
               "name": "string",
               "valueFrom": "string"
            }
         ],
         "startTimeout": number,
         "stopTimeout": number,
         "ulimits": [
            {
               "hardLimit": number,
               "name": "string",
               "softLimit": number
            }
```

```
            ],
            "user": "string"
         }
      ],
      "enableExecuteCommand": boolean,
      "ephemeralStorage": {
         "sizeInGiB": number
      },
      "executionRoleArn": "string",
      "ipcMode": "string",
      "networkConfiguration": {
         "assignPublicIp": "string"
      },
      "networkMode": "string",
      "pidMode": "string",
      "platformVersion": "string",
      "runtimePlatform": {
         "cpuArchitecture": "string",
         "operatingSystemFamily": "string"
      },
      "taskRoleArn": "string",
      "volumes": [
         {
            "efsVolumeConfiguration": {
               "authorizationConfig": {
                  "accessPointId": "string",
                  "iam": "string"
               },
               "fileSystemId": "string",
               "rootDirectory": "string",
               "transitEncryption": "string",
               "transitEncryptionPort": number
            },
            "host": {
               "sourcePath": "string"
            },
            "name": "string",
            "s3filesVolumeConfiguration": {
               "accessPointArn": "string",
               "fileSystemArn": "string",
               "rootDirectory": "string",
               "transitEncryptionPort": number
            }
         }
```

```
                ]
             }
          ]
       },
       "eksProperties": {
          "podProperties": {
             "containers": [
                {
                   "args": [ "string" ],
                   "command": [ "string" ],
                   "env": [
                      {
                         "name": "string",
                         "value": "string"
                      }
                   ],
                   "image": "string",
                   "imagePullPolicy": "string",
                   "name": "string",
                   "resources": {
                      "limits": {
                         "string" : "string"
                      },
                      "requests": {
                         "string" : "string"
                      }
                   },
                   "securityContext": {
                      "allowPrivilegeEscalation": boolean,
                      "privileged": boolean,
                      "readOnlyRootFilesystem": boolean,
                      "runAsGroup": number,
                      "runAsNonRoot": boolean,
                      "runAsUser": number
                   },
                   "volumeMounts": [
                      {
                         "mountPath": "string",
                         "name": "string",
                         "readOnly": boolean,
                         "subPath": "string"
                      }
                   ]
                }
```

```
        ],
        "dnsPolicy": "string",
        "hostNetwork": boolean,
        "imagePullSecrets": [
           {
              "name": "string"
           }
        ],
        "initContainers": [
           {
              "args": [ "string" ],
              "command": [ "string" ],
              "env": [
                 {
                    "name": "string",
                    "value": "string"
                 }
              ],
              "image": "string",
              "imagePullPolicy": "string",
              "name": "string",
              "resources": {
                 "limits": {
                    "string" : "string"
                 },
                 "requests": {
                    "string" : "string"
                 }
              },
              "securityContext": {
                 "allowPrivilegeEscalation": boolean,
                 "privileged": boolean,
                 "readOnlyRootFilesystem": boolean,
                 "runAsGroup": number,
                 "runAsNonRoot": boolean,
                 "runAsUser": number
              },
              "volumeMounts": [
                 {
                    "mountPath": "string",
                    "name": "string",
                    "readOnly": boolean,
                    "subPath": "string"
                 }
```

```
                        ]
                     }
                  ],
                  "metadata": {
                     "annotations": {
                        "string" : "string"
                     },
                     "labels": {
                        "string" : "string"
                     },
                     "namespace": "string"
                  },
                  "serviceAccountName": "string",
                  "shareProcessNamespace": boolean,
                  "volumes": [
                     {
                        "emptyDir": {
                           "medium": "string",
                           "sizeLimit": "string"
                        },
                        "hostPath": {
                           "path": "string"
                        },
                        "name": "string",
                        "persistentVolumeClaim": {
                           "claimName": "string",
                           "readOnly": boolean
                        },
                        "secret": {
                           "optional": boolean,
                           "secretName": "string"
                        }
                     }
                  ]
               }
            },
            "instanceTypes": [ "string" ],
            "targetNodes": "string"
         }
      ],
      "numNodes": number
   },
   "parameters": {
      "string" : "string"
```

```
    },
    "platformCapabilities": [ "string" ],
    "propagateTags": boolean,
    "retryStrategy": {
       "attempts": number,
       "evaluateOnExit": [
          {
             "action": "string",
             "onExitCode": "string",
             "onReason": "string",
             "onStatusReason": "string"
          }
       ]
    },
    "schedulingPriority": number,
    "tags": {
       "string" : "string"
    },
    "timeout": {
       "attemptDurationSeconds": number
    },
    "type": "string"
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

consumableResourceProperties

Contains a list of consumable resources required by the job.

Type: ConsumableResourceProperties object

Required: No

containerProperties

An object with properties specific to Amazon ECS-based single-node container-based jobs. If the job definition's `type` parameter is `container`, then you must specify either

`containerProperties` or `nodeProperties`. This must not be specified for Amazon EKS-based job definitions.

> **Note**
>
> If the job runs on Fargate resources, then you must not specify `nodeProperties`. We recommend `ecsProperties`, which supports both single-container and multi-container jobs. `containerProperties` is also supported for single-container jobs. For more information, see Multi-container jobs in the *AWS Batch User Guide*.

Type: ContainerProperties object

Required: No

ecsProperties

An object with properties that are specific to Amazon ECS-based jobs. This must not be specified for Amazon EKS-based job definitions.

Type: EcsProperties object

Required: No

eksProperties

An object with properties that are specific to Amazon EKS-based jobs. This must not be specified for Amazon ECS based job definitions.

Type: EksProperties object

Required: No

jobDefinitionName

The name of the job definition to register. It can be up to 128 letters long. It can contain uppercase and lowercase letters, numbers, hyphens (-), and underscores (_).

Type: String

Required: Yes

nodeProperties

An object with properties specific to multi-node parallel jobs. If you specify node properties for a job, it becomes a multi-node parallel job. For more information, see Multi-node Parallel Jobs in the *AWS Batch User Guide*.

> **Note**
>
> If the job runs on Fargate resources, then you must not specify `nodeProperties`. We recommend `ecsProperties`, which supports both single-container and multi-container jobs. `containerProperties` is also supported for single-container jobs.

> **Note**
>
> If the job runs on Amazon EKS resources, then you must not specify `nodeProperties`.

Type: NodeProperties object

Required: No

parameters

Default parameter substitution placeholders to set in the job definition. Parameters are specified as a key-value pair mapping. Parameters in a `SubmitJob` request override any corresponding parameter defaults from the job definition.

Type: String to string map

Required: No

platformCapabilities

The platform capabilities required by the job definition. If no value is specified, it defaults to `EC2`. To run the job on Fargate resources, specify `FARGATE`. To run the job on Amazon ECS Managed Instances, specify `MANAGED_INSTANCES`.

Jobs with the `MANAGED_INSTANCES` platform capability must use `ecsProperties` (not `containerProperties`) and do not support multi-node parallel jobs.

> **Note**
>
> If the job runs on Amazon EKS resources, then you must not specify `platformCapabilities`.

Type: Array of strings

Valid Values: `EC2 | FARGATE | MANAGED_INSTANCES`

Required: No

propagateTags

Specifies whether to propagate the tags from the job or job definition to the corresponding Amazon ECS task. If no value is specified, the tags are not propagated. Tags can only be propagated to the tasks during task creation. For tags with the same name, job tags are given priority over job definitions tags. If the total number of combined tags from the job and job definition is over 50, the job is moved to the `FAILED` state.

> **Note**
>
> If the job runs on Amazon EKS resources, then you must not specify `propagateTags`.

Type: Boolean

Required: No

retryStrategy

The retry strategy to use for failed jobs that are submitted with this job definition. Any retry strategy that's specified during a SubmitJob operation overrides the retry strategy defined here. If a job is terminated due to a timeout, it isn't retried.

Type: RetryStrategy object

Required: No

schedulingPriority

The scheduling priority for jobs that are submitted with this job definition. This only affects jobs in job queues with a fair-share policy. Jobs with a higher scheduling priority are scheduled before jobs with a lower scheduling priority.

The minimum supported value is 0 and the maximum supported value is 9999.

Type: Integer

Required: No

tags

The tags that you apply to the job definition to help you categorize and organize your resources. Each tag consists of a key and an optional value. For more information, see Tagging AWS Resources in *AWS Batch User Guide*.

Type: String to string map

Map Entries: Maximum number of 50 items.

Key Length Constraints: Minimum length of 1. Maximum length of 128.

Value Length Constraints: Maximum length of 256.

Required: No

timeout

The timeout configuration for jobs that are submitted with this job definition, after which AWS Batch terminates your jobs if they have not finished. If a job is terminated due to a timeout, it isn't retried. The minimum value for the timeout is 60 seconds. Any timeout configuration that's specified during a SubmitJob operation overrides the timeout configuration defined here. For more information, see Job Timeouts in the *AWS Batch User Guide*.

Type: JobTimeout object

Required: No

type

The type of job definition. For more information about multi-node parallel jobs, see Creating a multi-node parallel job definition in the *AWS Batch User Guide*.

- If the value is `container`, then one of the following is required: `containerProperties`, `ecsProperties`, or `eksProperties`.
- If the value is `multinode`, then `nodeProperties` is required.

> **Note**
>
> If the job is run on Fargate resources, then `multinode` isn't supported.

Type: String

Valid Values: `container | multinode`

Required: Yes

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
   "jobDefinitionArn": "string",
   "jobDefinitionName": "string",
   "revision": 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.

jobDefinitionArn

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

Type: String

jobDefinitionName

The name of the job definition.

Type: String

revision

The revision of the job definition.

Type: Integer

Errors

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

Examples

In the following example or examples, the Authorization header contents (*`[authorization-params]`*) must be replaced with an AWS Signature Version 4 signature. For more information about creating these signatures, see Signature Version 4 Signing Process in the *AWS General Reference*.

You only need to learn how to sign HTTP requests if you intend to manually create them. When you use the AWS Command Line Interface (AWS CLI) or one of the AWS SDKs to make requests to AWS, these tools automatically sign the requests for you with the access key that you specify when you configure the tools. When you use these tools, you don't need to learn how to sign requests yourself.

Example

This example registers a job definition for a simple container job.

Sample Request

```
POST /v1/registerjobdefinition HTTP/1.1
Host: batch.us-east-1.amazonaws.com
Accept-Encoding: identity
Content-Length: [content-length]
Authorization: [authorization-params]
X-Amz-Date: 20201128T215526Z
User-Agent: aws-cli/1.11.21 Python/2.7.12 Darwin/16.1.0 botocore/1.4.78

{
  "containerProperties": {
    "image": "busybox",
    "command": [
      "sleep",
      "10"
    ],
    "resourceRequirements": [
      {
        "type": "MEMORY",
        "value": "128"
      },
      {
        "type": "VCPU",
        "value": "1"
      }
    ]
  },
  "type": "container",
  "jobDefinitionName": "sleep10"
}
```

Sample Response

```
HTTP/1.1 200 OK
Content-Type: application/json
Content-Length: [content-length]
Connection: keep-alive
Date: Mon, 28 Nov 2020 21:55:27 GMT
x-amzn-RequestId: [request-id]
X-Amzn-Trace-Id: [trace-id]
X-Cache: Miss from cloudfront
Via: 1.1 7a06af51e583997d8673ab89482dd45a.cloudfront.net (CloudFront)
```

```
X-Amz-Cf-Id: Y14HPNWWcKgm1U0wJPfLBzLDvrMSdyuHo4GMi0oQwI0ukruLpi0nFw==

{
  "jobDefinitionName": "sleep10",
  "jobDefinitionArn": "arn:aws:batch:us-east-1:123456789012:job-definition/sleep10:1",
  "revision": 1
}
```

Example

This example registers a job definition for a simple container job with retries.

Sample Request

```
POST /v1/registerjobdefinition HTTP/1.1
Host: batch.us-east-1.amazonaws.com
Accept-Encoding: identity
Content-Length: [content-length]
Authorization: [authorization-params]
X-Amz-Date: 20170327T145208Z
User-Agent: aws-cli/1.11.66 Python/2.7.10 Darwin/16.4.0 botocore/1.5.29

{
  "containerProperties": {
    "mountPoints": [],
    "image": "amazonlinux",
    "environment": [],
    "command": [
      "/bin/bash",
      "-c",
      "exit $AWS_BATCH_JOB_ATTEMPT"
    ],
    "volumes": [],
    "resourceRequirements": [
      {
        "type": "MEMORY",
        "value": "8"
      },
      {
        "type": "VCPU",
        "value": "1"
      }
    ],
```

```
    "ulimits": []
  },
  "retryStrategy": {
    "attempts": 3
  },
  "jobDefinitionName": "EchoAttemptNumber",
  "parameters": {},
  "type": "container"
}
```

Sample Response

```
HTTP/1.1 200 OK
Date: Mon, 27 Mar 2017 14:51:58 GMT
Content-Type: application/json
Content-Length: [content-length]
Connection: keep-alive
x-amzn-RequestId: [request-id]
X-Amzn-Trace-Id: [trace-id]
X-Cache: Miss from cloudfront
Via: 1.1 861b49a34b383ce3ac4ea8b7117b8953.cloudfront.net (CloudFront)
X-Amz-Cf-Id: l3zCxHtlIx4c1-RN2vkqIlpSbU9tsZNxfMSg6oIf700wg1BX0V5f_A==

{
  "jobDefinitionName": "EchoAttemptNumber",
  "jobDefinitionArn": "arn:aws:batch:us-east-1:123456789012:job-definition/
EchoAttemptNumber:1",
  "revision": 1
}
```

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2

- AWS SDK for JavaScript V3
- AWS SDK for Kotlin
- AWS SDK for PHP V3
- AWS SDK for Python (Boto3)
- AWS SDK for Ruby V3

SubmitJob

Submits an AWS Batch job from a job definition. Parameters that are specified during SubmitJob override parameters defined in the job definition. vCPU and memory requirements that are specified in the `resourceRequirements` objects in the job definition are the exception. They can't be overridden this way using the `memory` and `vcpus` parameters. Rather, you must specify updates to job definition parameters in a `resourceRequirements` object that's included in the `containerOverrides` parameter.

Note

Job queues with a scheduling policy are limited to 500 active share identifiers at a time.

Important

Jobs that run on Fargate resources can't be guaranteed to run for more than 14 days. This is because, after 14 days, Fargate resources might become unavailable and job might be terminated.

Request Syntax

```
POST /v1/submitjob HTTP/1.1
Content-type: application/json

{
   "arrayProperties": {
      "size": number
   },
   "consumableResourcePropertiesOverride": {
      "consumableResourceList": [
         {
            "consumableResource": "string",
            "quantity": number
         }
      ]
   },
   "containerOverrides": {
      "command": [ "string" ],
```

```
      "environment": [
         {
            "name": "string",
            "value": "string"
         }
      ],
      "instanceType": "string",
      "memory": number,
      "resourceRequirements": [
         {
            "type": "string",
            "value": "string"
         }
      ],
      "vcpus": number
   },
   "dependsOn": [
      {
         "jobId": "string",
         "type": "string"
      }
   ],
   "ecsPropertiesOverride": {
      "taskProperties": [
         {
            "containers": [
               {
                  "command": [ "string" ],
                  "environment": [
                     {
                        "name": "string",
                        "value": "string"
                     }
                  ],
                  "name": "string",
                  "resourceRequirements": [
                     {
                        "type": "string",
                        "value": "string"
                     }
                  ]
               }
            ]
         }
```

```
      ]
   },
   "eksPropertiesOverride": {
      "podProperties": {
         "containers": [
            {
               "args": [ "string" ],
               "command": [ "string" ],
               "env": [
                  {
                     "name": "string",
                     "value": "string"
                  }
               ],
               "image": "string",
               "name": "string",
               "resources": {
                  "limits": {
                     "string" : "string"
                  },
                  "requests": {
                     "string" : "string"
                  }
               }
            }
         ],
         "initContainers": [
            {
               "args": [ "string" ],
               "command": [ "string" ],
               "env": [
                  {
                     "name": "string",
                     "value": "string"
                  }
               ],
               "image": "string",
               "name": "string",
               "resources": {
                  "limits": {
                     "string" : "string"
                  },
                  "requests": {
                     "string" : "string"
```

```
                }
             }
          }
       ],
       "metadata": {
          "annotations": {
             "string" : "string"
          },
          "labels": {
             "string" : "string"
          },
          "namespace": "string"
       }
    }
 },
 "jobDefinition": "string",
 "jobName": "string",
 "jobQueue": "string",
 "nodeOverrides": {
    "nodePropertyOverrides": [
       {
          "consumableResourcePropertiesOverride": {
             "consumableResourceList": [
                {
                   "consumableResource": "string",
                   "quantity": number
                }
             ]
          },
          "containerOverrides": {
             "command": [ "string" ],
             "environment": [
                {
                   "name": "string",
                   "value": "string"
                }
             ],
             "instanceType": "string",
             "memory": number,
             "resourceRequirements": [
                {
                   "type": "string",
                   "value": "string"
                }
```

```
            ],
            "vcpus": number
         },
         "ecsPropertiesOverride": {
            "taskProperties": [
               {
                  "containers": [
                     {
                        "command": [ "string" ],
                        "environment": [
                           {
                              "name": "string",
                              "value": "string"
                           }
                        ],
                        "name": "string",
                        "resourceRequirements": [
                           {
                              "type": "string",
                              "value": "string"
                           }
                        ]
                     }
                  ]
               }
            ]
         },
         "eksPropertiesOverride": {
            "podProperties": {
               "containers": [
                  {
                     "args": [ "string" ],
                     "command": [ "string" ],
                     "env": [
                        {
                           "name": "string",
                           "value": "string"
                        }
                     ],
                     "image": "string",
                     "name": "string",
                     "resources": {
                        "limits": {
                           "string" : "string"
```

```
                        },
                        "requests": {
                           "string" : "string"
                        }
                     }
                  }
               ],
               "initContainers": [
                  {
                     "args": [ "string" ],
                     "command": [ "string" ],
                     "env": [
                        {
                           "name": "string",
                           "value": "string"
                        }
                     ],
                     "image": "string",
                     "name": "string",
                     "resources": {
                        "limits": {
                           "string" : "string"
                        },
                        "requests": {
                           "string" : "string"
                        }
                     }
                  }
               ],
               "metadata": {
                  "annotations": {
                     "string" : "string"
                  },
                  "labels": {
                     "string" : "string"
                  },
                  "namespace": "string"
               }
            }
         },
         "instanceTypes": [ "string" ],
         "targetNodes": "string"
      }
   ],
```

```
      "numNodes": number
   },
   "parameters": {
      "string" : "string"
   },
   "propagateTags": boolean,
   "retryStrategy": {
      "attempts": number,
      "evaluateOnExit": [
         {
            "action": "string",
            "onExitCode": "string",
            "onReason": "string",
            "onStatusReason": "string"
         }
      ]
   },
   "schedulingPriorityOverride": number,
   "shareIdentifier": "string",
   "tags": {
      "string" : "string"
   },
   "timeout": {
      "attemptDurationSeconds": number
   }
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

arrayProperties

The array properties for the submitted job, such as the size of the array. The array size can be between 2 and 10,000. If you specify array properties for a job, it becomes an array job. For more information, see Array Jobs in the *AWS Batch User Guide*.

Type: ArrayProperties object

Required: No

consumableResourcePropertiesOverride

An object that contains overrides for the consumable resources of a job.

Type: ConsumableResourceProperties object

Required: No

containerOverrides

An object with properties that override the defaults for the job definition that specify the name of a container in the specified job definition and the overrides it should receive. You can override the default command for a container, which is specified in the job definition or the Docker image, with a `command` override. You can also override existing environment variables on a container or add new environment variables to it with an `environment` override.

Type: ContainerOverrides object

Required: No

dependsOn

A list of dependencies for the job. A job can depend upon a maximum of 20 jobs. You can specify a `SEQUENTIAL` type dependency without specifying a job ID for array jobs so that each child array job completes sequentially, starting at index 0. You can also specify an `N_TO_N` type dependency with a job ID for array jobs. In that case, each index child of this job must wait for the corresponding index child of each dependency to complete before it can begin.

Type: Array of JobDependency objects

Required: No

ecsPropertiesOverride

An object, with properties that override defaults for the job definition, can only be specified for jobs that are run on Amazon ECS resources.

Type: EcsPropertiesOverride object

Required: No

eksPropertiesOverride

An object, with properties that override defaults for the job definition, can only be specified for jobs that are run on Amazon EKS resources.

Type: EksPropertiesOverride object

Required: No

jobDefinition

The job definition used by this job. This value can be one of `definition-name`, `definition-name:revision`, or the Amazon Resource Name (ARN) for the job definition, with or without the revision (`arn:aws:batch:`*`region`*`:`*`account`*`:job-definition/`*`definition-name`*`:`*`revision`* , or `arn:aws:batch:`*`region`*`:`*`account`*`:job-definition/`*`definition-name`*).

If the revision is not specified, then the latest active revision is used.

Type: String

Required: Yes

jobName

The name of the job. It can be up to 128 letters long. The first character must be alphanumeric, can contain uppercase and lowercase letters, numbers, hyphens (-), and underscores (_).

Type: String

Required: Yes

jobQueue

The job queue where the job is submitted. You can specify either the name or the Amazon Resource Name (ARN) of the queue.

Type: String

Required: Yes

nodeOverrides

A list of node overrides in JSON format that specify the node range to target and the container overrides for that node range.

> **Note**
>
> This parameter isn't applicable to jobs that are running on Fargate resources; use `containerOverrides` instead.

Type: NodeOverrides object

Required: No

parameters

Additional parameters passed to the job that replace parameter substitution placeholders that are set in the job definition. Parameters are specified as a key and value pair mapping. Parameters in a `SubmitJob` request override any corresponding parameter defaults from the job definition.

Type: String to string map

Required: No

propagateTags

Specifies whether to propagate the tags from the job or job definition to the corresponding Amazon ECS task. If no value is specified, the tags aren't propagated. Tags can only be propagated to the tasks during task creation. For tags with the same name, job tags are given priority over job definitions tags. If the total number of combined tags from the job and job definition is over 50, the job is moved to the `FAILED` state. When specified, this overrides the tag propagation setting in the job definition.

Type: Boolean

Required: No

retryStrategy

The retry strategy to use for failed jobs from this SubmitJob operation. When a retry strategy is specified here, it overrides the retry strategy defined in the job definition.

Type: RetryStrategy object

Required: No

schedulingPriorityOverride

The scheduling priority for the job. This only affects jobs in job queues with a fair-share policy. Jobs with a higher scheduling priority are scheduled before jobs with a lower scheduling priority. This overrides any scheduling priority in the job definition and works only within a single share identifier.

The minimum supported value is 0 and the maximum supported value is 9999.

Type: Integer

Required: No

shareIdentifier

The share identifier for the job. Don't specify this parameter if the job queue doesn't have a fair-share scheduling policy. If the job queue has a fair-share scheduling policy, then this parameter must be specified.

This string is limited to 255 alphanumeric characters, and can be followed by an asterisk (*).

Type: String

Required: No

tags

The tags that you apply to the job request to help you categorize and organize your resources. Each tag consists of a key and an optional value. For more information, see Tagging AWS Resources in *AWS General Reference*.

Type: String to string map

Map Entries: Maximum number of 50 items.

Key Length Constraints: Minimum length of 1. Maximum length of 128.

Value Length Constraints: Maximum length of 256.

Required: No

timeout

The timeout configuration for this SubmitJob operation. You can specify a timeout duration after which AWS Batch terminates your jobs if they haven't finished. If a job is terminated

due to a timeout, it isn't retried. The minimum value for the timeout is 60 seconds. This configuration overrides any timeout configuration specified in the job definition. For array jobs, child jobs have the same timeout configuration as the parent job. For more information, see Job Timeouts in the *Amazon Elastic Container Service Developer Guide*.

Type: JobTimeout object

Required: No

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
   "jobArn": "string",
   "jobId": "string",
   "jobName": "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

The Amazon Resource Name (ARN) for the job.

Type: String

jobId

The unique identifier for the job.

Type: String

jobName

The name of the job.

Type: String

Errors

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

Examples

In the following example or examples, the Authorization header contents (*[authorization-params]*) must be replaced with an AWS Signature Version 4 signature. For more information about creating these signatures, see Signature Version 4 Signing Process in the *AWS General Reference*.

You only need to learn how to sign HTTP requests if you intend to manually create them. When you use the AWS Command Line Interface (AWS CLI) or one of the AWS SDKs to make requests to AWS, these tools automatically sign the requests for you with the access key that you specify when you configure the tools. When you use these tools, you don't need to learn how to sign requests yourself.

Example

This example submits a simple container job called `example` to the `HighPriority` job queue.

Sample Request

```
POST /v1/submitjob HTTP/1.1
Host: batch.us-east-1.amazonaws.com
Accept-Encoding: identity
Content-Length: [content-length]
Authorization: [authorization-params]
X-Amz-Date: 20161129T201034Z
```

```
User-Agent: aws-cli/1.11.22 Python/2.7.12 Darwin/16.1.0 botocore/1.4.79

{
  "jobName": "example",
  "jobQueue": "HighPriority",
  "jobDefinition": "sleep60"
}
```

Sample Response

```
HTTP/1.1 200 OK
Content-Type: application/json
Content-Length: [content-length]
Connection: keep-alive
Date: Tue, 29 Nov 2016 20:10:34 GMT
x-amzn-RequestId: [request-id]
X-Amzn-Trace-Id: [trace-id]
X-Cache: Miss from cloudfront
Via: 1.1 6ba12aeff47e3e7677e084594bfce5e1.cloudfront.net (CloudFront)
X-Amz-Cf-Id: QoWCvISTvYBbaP2hAoxC8_TWTl2JN-kNASYCjCJ5HztN0e1OuzvpSA==

{
  "jobName": "example",
  "jobId": "876da822-4198-45f2-a252-6cea32512ea8"
}
```

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for Kotlin
- AWS SDK for PHP V3

- AWS SDK for Python (Boto3)
- AWS SDK for Ruby V3

SubmitServiceJob

Submits a service job to a specified job queue to run on SageMaker AI. A service job is a unit of work that you submit to AWS Batch for execution on SageMaker AI.

Request Syntax

```
POST /v1/submitservicejob HTTP/1.1
Content-type: application/json

{
   "clientToken": "string",
   "jobName": "string",
   "jobQueue": "string",
   "preemptionConfiguration": {
      "preemptionRetriesBeforeTermination": number
   },
   "quotaShareName": "string",
   "retryStrategy": {
      "attempts": number,
      "evaluateOnExit": [
         {
            "action": "string",
            "onStatusReason": "string"
         }
      ]
   },
   "schedulingPriority": number,
   "serviceJobType": "string",
   "serviceRequestPayload": "string",
   "shareIdentifier": "string",
   "tags": {
      "string" : "string"
   },
   "timeoutConfig": {
      "attemptDurationSeconds": number
   }
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

clientToken

A unique identifier for the request. This token is used to ensure idempotency of requests. If this parameter is specified and two submit requests with identical payloads and `clientTokens` are received, these requests are considered the same request and the second request is rejected.

Type: String

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

Required: No

jobName

The name of the service job. It can be up to 128 characters long. It can contain uppercase and lowercase letters, numbers, hyphens (-), and underscores (_).

Type: String

Required: Yes

jobQueue

The job queue into which the service job is submitted. You can specify either the name or the ARN of the queue. The job queue must have the type `SAGEMAKER_TRAINING`.

Type: String

Required: Yes

preemptionConfiguration

Specifies the service job behavior when preempted.

Type: ServiceJobPreemptionConfiguration object

Required: No

quotaShareName

The quota share for the service job. Don't specify this parameter if the job queue doesn't have a quota share scheduling policy. If the job queue has a quota share scheduling policy, then this parameter must be specified.

Type: String

Required: No

retryStrategy

The retry strategy to use for failed service jobs that are submitted with this service job request.

Type: ServiceJobRetryStrategy object

Required: No

schedulingPriority

The scheduling priority of the service job. Valid values are integers between 0 and 9999.

Type: Integer

Required: No

serviceJobType

The type of service job. For SageMaker Training jobs, specify `SAGEMAKER_TRAINING`.

Type: String

Valid Values: `SAGEMAKER_TRAINING`

Required: Yes

serviceRequestPayload

The request, in JSON, for the service that the SubmitServiceJob operation is queueing.

Type: String

Required: Yes

shareIdentifier

The share identifier for the service job. Don't specify this parameter if the job queue doesn't have a fair-share scheduling policy. If the job queue has a fair-share scheduling policy, then this parameter must be specified.

Type: String

Required: No

tags

The tags that you apply to the service job request. Each tag consists of a key and an optional value. For more information, see Tagging your AWS Batch resources.

Type: String to string map

Map Entries: Maximum number of 50 items.

Key Length Constraints: Minimum length of 1. Maximum length of 128.

Value Length Constraints: Maximum length of 256.

Required: No

timeoutConfig

The timeout configuration for the service job. If none is specified, AWS Batch defers to the default timeout of the underlying service handling the job.

Type: ServiceJobTimeout object

Required: No

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
   "jobArn": "string",
   "jobId": "string",
   "jobName": "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

The Amazon Resource Name (ARN) for the service job.

Type: String

jobId

The unique identifier for the service job.

Type: String

jobName

The name of the service job.

Type: String

Errors

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

Examples

In the following example or examples, the Authorization header contents (*[authorization-params]*) must be replaced with an AWS Signature Version 4 signature. For more information about creating these signatures, see Signature Version 4 Signing Process in the *AWS General Reference*.

You only need to learn how to sign HTTP requests if you intend to manually create them. When you use the AWS Command Line Interface (AWS CLI) or one of the AWS SDKs to make requests to

AWS, these tools automatically sign the requests for you with the access key that you specify when you configure the tools. When you use these tools, you don't need to learn how to sign requests yourself.

Example

This example submits a SageMaker training job to the specified job queue.

Sample Request

```
POST /v1/submitservicejob HTTP/1.1
Host: batch.us-east-1.amazonaws.com
Accept-Encoding: identity
Content-Length: [content-length]
Authorization: [authorization-params]
X-Amz-Date: 20250801T083015Z
User-Agent: aws-cli/2.27.33 Python/3.13.4 Darwin/24.3.0

{
    "jobName": "sagemaker-training-job-example",
    "jobQueue": "sagemaker-training-queue",
    "retryStrategy": {
        "attempts": 2,
        "evaluateOnExit": [
            {
                "action": "Retry",
                "onStatusReason": "Received status from SageMaker: AlgorithmError: *"
            },
            {
                "action": "EXIT",
                "onStatusReason": "*"
            }
        ]
    },
    "serviceJobType": "SAGEMAKER_TRAINING",
    "serviceRequestPayload": "{\"TrainingJobName\": \"sagemaker-training-job-example
\", \"AlgorithmSpecification\": {\"TrainingImage\": \"123456789012.dkr.ecr.us-
east-1.amazonaws.com/pytorch-inference:1.8.0-cpu-py3\", \"TrainingInputMode
\": \"File\", \"ContainerEntrypoint\":  [\"sleep\", \"1\"]}, \"RoleArn\":
\"arn:aws:iam::123456789012:role/SageMakerExecutionRole\", \"OutputDataConfig\":
 {\"S3OutputPath\": \"s3://example-bucket/model-output/\"}, \"ResourceConfig\":
 {\"InstanceType\": \"ml.m5.large\", \"InstanceCount\": 1, \"VolumeSizeInGB\": 1}}",
    "timeoutConfig": {
```

```
        "attemptDurationSeconds": 300
    },
    "tags": {
        "tag-name": "value-123"
    }
}
```

Sample Response

```
HTTP/1.1 200 OK
Content-Type: application/json
Content-Length: [content-length]
Connection: keep-alive
Date: Fri, 01 Aug 2025 08:30:16 GMT
x-amzn-RequestId: [request-id]
X-Amzn-Trace-Id: [trace-id]
X-Cache: Miss from cloudfront
Via: 1.1 hf65sd33h6j9k2l5m8n1p4q7r0sexample.cloudfront.net (CloudFront)
X-Amz-Cf-Id: jkl5mno8pqr1stu4vwx7yz0123ghijklmnopqrstuvwexample

{
  "jobName": "sagemaker-training-job-example",
  "jobId": "a4d6c728-8ee8-4c65-8e2a-9a5e8f4b7c3d",
  "jobArn": "arn:aws:batch:us-east-1:123456789012:service-job/
a4d6c728-8ee8-4c65-8e2a-9a5e8f4b7c3d"
}
```

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for Kotlin

- AWS SDK for PHP V3
- AWS SDK for Python (Boto3)
- AWS SDK for Ruby V3

TagResource

Associates the specified tags to a resource with the specified `resourceArn`. If existing tags on a resource aren't specified in the request parameters, they aren't changed. When a resource is deleted, the tags that are associated with that resource are deleted as well. AWS Batch resources that support tags are compute environments, jobs, job definitions, job queues, and scheduling policies. ARNs for child jobs of array and multi-node parallel (MNP) jobs aren't supported.

Request Syntax

```
POST /v1/tags/resourceArn HTTP/1.1
Content-type: application/json

{
   "tags": {
      "string" : "string"
   }
}
```

URI Request Parameters

The request uses the following URI parameters.

resourceArn

The Amazon Resource Name (ARN) of the resource that tags are added to. AWS Batch resources that support tags are compute environments, jobs, job definitions, job queues, and scheduling policies. ARNs for child jobs of array and multi-node parallel (MNP) jobs aren't supported.

Required: Yes

Request Body

The request accepts the following data in JSON format.

tags

The tags that you apply to the resource to help you categorize and organize your resources. Each tag consists of a key and an optional value. For more information, see Tagging AWS Resources in *AWS General Reference*.

Type: String to string map

Map Entries: Maximum number of 50 items.

Key Length Constraints: Minimum length of 1. Maximum length of 128.

Value Length Constraints: Maximum length of 256.

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

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

Examples

In the following example or examples, the Authorization header contents (*[authorization-params]*) must be replaced with an AWS Signature Version 4 signature. For more information

about creating these signatures, see Signature Version 4 Signing Process in the *AWS General Reference*.

You only need to learn how to sign HTTP requests if you intend to manually create them. When you use the AWS Command Line Interface (AWS CLI) or one of the AWS SDKs to make requests to AWS, these tools automatically sign the requests for you with the access key that you specify when you configure the tools. When you use these tools, you don't need to learn how to sign requests yourself.

Example

This example adds a tag to the job definition with an ARN of "*arn:aws:batch:us-east-1:123456789012:job-definition/sleep30:1*".

Sample Request

```
POST /v1/tags/arn%3Aaws%3Abatch%3Aus-east-1%3A123456789012%3Ajob-definition
%2Fsleep30%3A1 HTTP/1.1
Host: batch.us-east-1.amazonaws.com
Accept-Encoding: identity
X-Amz-Date: 20200604T172359Z
X-Amz-Security-Token: [security-token]
Authorization: [authorization-params]
Content-Length: [length-of-JSON]

{
  "tags": {
    "Stage": "alpha"
  }
}
```

Sample Response

```
HTTP/1.1 204 No Content
Date: Thu, 05 Jun 2020 17:24:04 GMT
Content-Type: application/json
Content-Length: 0
x-amzn-RequestId: [request-id]
Access-Control-Allow-Origin: *
x-amz-apigw-id: [apigw-id]
X-Amzn-Trace-Id: [trace-id]
```

```
Connection: keep-alive
```

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for Kotlin
- AWS SDK for PHP V3
- AWS SDK for Python (Boto3)
- AWS SDK for Ruby V3

TerminateJob

Terminates a job in a job queue. Jobs that are in the STARTING or RUNNING state are terminated, which causes them to transition to FAILED. Jobs that have not progressed to the STARTING state are cancelled.

Request Syntax

```
POST /v1/terminatejob HTTP/1.1
Content-type: application/json

{
   "jobId": "string",
   "reason": "string"
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

jobId

The AWS Batch job ID of the job to terminate.

Type: String

Required: Yes

reason

A message to attach to the job that explains the reason for terminating it. This message is returned by future DescribeJobs operations on the job. It is also recorded in the AWS Batch activity logs.

This parameter has a limit of 1024 characters.

Type: String

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

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

Examples

In the following example or examples, the Authorization header contents (*`[authorization-params]`*) must be replaced with an AWS Signature Version 4 signature. For more information about creating these signatures, see Signature Version 4 Signing Process in the *AWS General Reference*.

You only need to learn how to sign HTTP requests if you intend to manually create them. When you use the AWS Command Line Interface (AWS CLI) or one of the AWS SDKs to make requests to AWS, these tools automatically sign the requests for you with the access key that you specify when you configure the tools. When you use these tools, you don't need to learn how to sign requests yourself.

Example

This example terminates a job with the specified job ID.

Sample Request

```
POST /v1/terminatejob HTTP/1.1
Host: batch.us-east-1.amazonaws.com
Accept-Encoding: identity
Content-Length: [content-length]
Authorization: [authorization-params]
X-Amz-Date: 20161129T202905Z
User-Agent: aws-cli/1.11.22 Python/2.7.12 Darwin/16.1.0 botocore/1.4.79

{
  "reason": "Terminating job.",
  "jobId": "61e743ed-35e4-48da-b2de-5c8333821c84"
}
```

Sample Response

```
HTTP/1.1 200 OK
Content-Type: application/json
Content-Length: 2
Connection: keep-alive
Date: Tue, 29 Nov 2016 20:29:06 GMT
x-amzn-RequestId: [request-id]
X-Amzn-Trace-Id: [trace-id]
X-Cache: Miss from cloudfront
Via: 1.1 16d2657cebef5191828b055567b4efeb.cloudfront.net (CloudFront)
X-Amz-Cf-Id: 681NTs_bPulMwja2HekWMwngcUzx2a8w_oaG27W0L4Pjct7W1T-Fvw==

{}
```

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4

- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for Kotlin
- AWS SDK for PHP V3
- AWS SDK for Python (Boto3)
- AWS SDK for Ruby V3

TerminateJobs

Terminates up to 50 jobs in a job queue. This is a bulk version of TerminateJob. Jobs that are in the `STARTING` or `RUNNING` state are terminated, which causes them to transition to `FAILED`. Jobs that have not progressed to the `STARTING` state are cancelled.

AWS Batch reports the result for each job individually in the response. Jobs that were processed successfully are reported in the `successful` list. Jobs that encountered errors are reported in the `errors` list. The response returns an HTTP status code of `200` even when some jobs encountered errors, so check the `errors` list. Jobs that can't be found are treated as successfully processed.

> **Important**
>
> This operation requires `batch:TerminateJob` permission for each job in the request. There is no separate `batch:TerminateJobs` IAM action. If a caller's IAM policy grants `batch:TerminateJob`, they can use both the singular TerminateJob and bulk `TerminateJobs` operations.

Request Syntax

```
POST /v1/terminatejobs HTTP/1.1
Content-type: application/json

{
   "jobs": [ "string" ],
   "reason": "string"
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

jobs

An array of up to 50 AWS Batch job IDs of the jobs to terminate.

Type: Array of strings

Required: Yes

reason

A message to attach to the job that explains the reason for terminating it. This message is returned by future DescribeJobs operations on the job. It is also recorded in the AWS Batch activity logs.

This parameter has a limit of 1024 characters.

Type: String

Required: Yes

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
   "errors": [
      {
         "code": "string",
         "job": "string",
         "message": "string"
      }
   ],
   "successful": [ "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

A list of `TerminateJobsErrorDetail` items, one for each job that couldn't be terminated. Each item includes the job ID along with a code and message that describe why the job wasn't terminated.

Type: Array of TerminateJobsErrorDetail objects

successful

A list of the job IDs whose termination request was accepted.

Type: Array of strings

Errors

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

Examples

In the following example or examples, the Authorization header contents (*`[authorization-params]`*) must be replaced with an AWS Signature Version 4 signature. For more information about creating these signatures, see Signature Version 4 Signing Process in the *AWS General Reference*.

You only need to learn how to sign HTTP requests if you intend to manually create them. When you use the AWS Command Line Interface (AWS CLI) or one of the AWS SDKs to make requests to AWS, these tools automatically sign the requests for you with the access key that you specify when you configure the tools. When you use these tools, you don't need to learn how to sign requests yourself.

Example

This example terminates the jobs with the specified job IDs.

Sample Request

```
POST /v1/terminatejobs HTTP/1.1
Host: batch.us-east-1.amazonaws.com
Accept-Encoding: identity
Content-Length: [content-length]
Authorization: [authorization-params]
X-Amz-Date: 20161129T202905Z
User-Agent: aws-cli/1.11.22 Python/2.7.12 Darwin/16.1.0 botocore/1.4.79

{
  "jobs": [
    "61e743ed-35e4-48da-b2de-5c8333821c84",
    "b3d0f26e-6d3a-4a5b-9c2e-7f4a1b2c3d4e"
  ],
  "reason": "Terminating jobs."
}
```

Sample Response

```
HTTP/1.1 200 OK
Content-Type: application/json
Content-Length: [content-length]
Connection: keep-alive
Date: Tue, 29 Nov 2016 20:29:06 GMT
x-amzn-RequestId: [request-id]
X-Amzn-Trace-Id: [trace-id]
X-Cache: Miss from cloudfront
Via: 1.1 16d2657cebef5191828b055567b4efeb.cloudfront.net (CloudFront)
X-Amz-Cf-Id: 681NTs_bPulMwja2HekWMwngcUzx2a8w_oaG27W0L4Pjct7W1T-Fvw==

{
  "successful": [
    "61e743ed-35e4-48da-b2de-5c8333821c84"
  ],
  "errors": [
    {
      "job": "b3d0f26e-6d3a-4a5b-9c2e-7f4a1b2c3d4e",
      "code": "ServerException",
      "message": "Failed to read job state. Please retry this job."
    }
  ]
}
```

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for Kotlin
- AWS SDK for PHP V3
- AWS SDK for Python (Boto3)
- AWS SDK for Ruby V3

TerminateServiceJob

Terminates a service job in a job queue.

Request Syntax

```
POST /v1/terminateservicejob HTTP/1.1
Content-type: application/json

{
   "jobId": "string",
   "reason": "string"
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

jobId

The service job ID of the service job to terminate.

Type: String

Required: Yes

reason

A message to attach to the service job that explains the reason for terminating it. This message is returned by `DescribeServiceJob` operations on the service job.

Type: String

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

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

Examples

In the following example or examples, the Authorization header contents (*`[authorization-params]`*) must be replaced with an AWS Signature Version 4 signature. For more information about creating these signatures, see Signature Version 4 Signing Process in the *AWS General Reference*.

You only need to learn how to sign HTTP requests if you intend to manually create them. When you use the AWS Command Line Interface (AWS CLI) or one of the AWS SDKs to make requests to AWS, these tools automatically sign the requests for you with the access key that you specify when you configure the tools. When you use these tools, you don't need to learn how to sign requests yourself.

Example

This example terminates the specified service job with a reason.

Sample Request

```
POST /v1/terminateservicejob HTTP/1.1
```

```
Host: batch.us-east-1.amazonaws.com
Accept-Encoding: identity
Content-Length: [content-length]
Authorization: [authorization-params]
X-Amz-Date: 20250801T164055Z
User-Agent: aws-cli/2.27.33 Python/3.13.4 Darwin/24.3.0

{
  "jobId": "a4d6c728-8ee8-4c65-8e2a-9a5e8f4b7c3d",
  "reason": "Job terminated by user request"
}
```

Sample Response

```
HTTP/1.1 200 OK
Content-Type: application/json
Content-Length: 0
Connection: keep-alive
Date: Fri, 01 Aug 2025 16:40:56 GMT
x-amzn-RequestId: [request-id]
X-Amzn-Trace-Id: [trace-id]
X-Cache: Miss from cloudfront
Via: 1.1 254fde64p7r0s3t6u9v2w5x8y1zexample.cloudfront.net (CloudFront)
X-Amz-Cf-Id: pqr8stu1vwx4yz7012fghijklmnopqrstuvwxyzabexample
```

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for Kotlin
- AWS SDK for PHP V3
- AWS SDK for Python (Boto3)

- AWS SDK for Ruby V3

TerminateServiceJobs

Terminates up to 50 service jobs in a job queue. This is a bulk version of TerminateServiceJob.

AWS Batch reports the result for each service job individually in the response. Service jobs that were processed successfully are reported in the `successful` list. Service jobs that encountered errors are reported in the `errors` list. The response returns an HTTP status code of `200` even when some service jobs encountered errors, so check the `errors` list. Service jobs that can't be found are treated as successfully processed.

> **Important**
>
> This operation requires `batch:TerminateServiceJob` permission for each service job in the request. There is no separate `batch:TerminateServiceJobs` IAM action. If a caller's IAM policy grants `batch:TerminateServiceJob`, they can use both the singular `TerminateServiceJob` and bulk `TerminateServiceJobs` operations.

Request Syntax

```
POST /v1/terminateservicejobs HTTP/1.1
Content-type: application/json

{
   "jobs": [ "string" ],
   "reason": "string"
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

jobs

An array of up to 50 service job IDs of the service jobs to terminate.

Type: Array of strings

Required: Yes

reason

A message to attach to the service job that explains the reason for terminating it. This message is returned by `DescribeServiceJob` operations on the service job.

Type: String

Required: Yes

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
   "errors": [
      {
         "code": "string",
         "job": "string",
         "message": "string"
      }
   ],
   "successful": [ "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

A list of `TerminateServiceJobsErrorDetail` items, one for each service job that couldn't be terminated. Each item includes the service job ID along with a code and message that describe why the service job wasn't terminated.

Type: Array of TerminateServiceJobsErrorDetail objects

successful

A list of the service job IDs whose termination request was accepted.

Type: Array of strings

Errors

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

Examples

In the following example or examples, the Authorization header contents (*`[authorization-params]`*) must be replaced with an AWS Signature Version 4 signature. For more information about creating these signatures, see Signature Version 4 Signing Process in the *AWS General Reference*.

You only need to learn how to sign HTTP requests if you intend to manually create them. When you use the AWS Command Line Interface (AWS CLI) or one of the AWS SDKs to make requests to AWS, these tools automatically sign the requests for you with the access key that you specify when you configure the tools. When you use these tools, you don't need to learn how to sign requests yourself.

Example

This example terminates the specified service jobs with a reason.

Sample Request

```
POST /v1/terminateservicejobs HTTP/1.1
Host: batch.us-east-1.amazonaws.com
Accept-Encoding: identity
Content-Length: [content-length]
Authorization: [authorization-params]
X-Amz-Date: 20250801T164055Z
User-Agent: aws-cli/2.27.33 Python/3.13.4 Darwin/24.3.0

{
  "jobs": [
    "a4d6c728-8ee8-4c65-8e2a-9a5e8f4b7c3d",
    "b3d0f26e-6d3a-4a5b-9c2e-7f4a1b2c3d4e"
  ],
  "reason": "Job terminated by user request"
}
```

Sample Response

```
HTTP/1.1 200 OK
Content-Type: application/json
Content-Length: [content-length]
Connection: keep-alive
Date: Fri, 01 Aug 2025 16:40:56 GMT
x-amzn-RequestId: [request-id]
X-Amzn-Trace-Id: [trace-id]
X-Cache: Miss from cloudfront
Via: 1.1 254fde64p7r0s3t6u9v2w5x8y1zexample.cloudfront.net (CloudFront)
X-Amz-Cf-Id: pqr8stu1vwx4yz7012fghijklmnopqrstuvwxyzabexample

{
  "successful": [
    "a4d6c728-8ee8-4c65-8e2a-9a5e8f4b7c3d"
  ],
  "errors": [
    {
      "job": "b3d0f26e-6d3a-4a5b-9c2e-7f4a1b2c3d4e",
      "code": "ServerException",
      "message": "Failed to read job state. Please retry this job."
    }
  ]
}
```

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for Kotlin
- AWS SDK for PHP V3
- AWS SDK for Python (Boto3)
- AWS SDK for Ruby V3

UntagResource

Deletes specified tags from an AWS Batch resource.

Request Syntax

```
DELETE /v1/tags/resourceArn?tagKeys=tagKeys HTTP/1.1
```

URI Request Parameters

The request uses the following URI parameters.

resourceArn

The Amazon Resource Name (ARN) of the resource from which to delete tags. AWS Batch resources that support tags are compute environments, jobs, job definitions, job queues, and scheduling policies. ARNs for child jobs of array and multi-node parallel (MNP) jobs aren't supported.

Required: Yes

tagKeys

The keys of the tags to be removed.

Array Members: Minimum number of 1 item. Maximum number of 50 items.

Length Constraints: Minimum length of 1. Maximum length of 128.

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

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

Examples

In the following example or examples, the Authorization header contents (*`[authorization-params]`*) must be replaced with an AWS Signature Version 4 signature. For more information about creating these signatures, see Signature Version 4 Signing Process in the *AWS General Reference*.

You only need to learn how to sign HTTP requests if you intend to manually create them. When you use the AWS Command Line Interface (AWS CLI) or one of the AWS SDKs to make requests to AWS, these tools automatically sign the requests for you with the access key that you specify when you configure the tools. When you use these tools, you don't need to learn how to sign requests yourself.

Example

This example removes the "*Stage*" tag from the job definition with an ARN of "*arn:aws:batch:us-east-1:123456789012:job-definition/sleep30:1*".

Sample Request

```
DELETE /v1/tags/arn%3Aaws%3Abatch%3Aus-east-1%3A123456789012%3Ajob-definition
%2Fsleep30%3A1?tagKeys=Stage HTTP/1.1
```

```
Host: batch.us-east-1.amazonaws.com
Accept-Encoding: identity
X-Amz-Date: 20200604T174552Z
X-Amz-Security-Token: [security-token]
Authorization: [authorization-params]
Content-Length: 0
```

Sample Response

```
HTTP/1.1 204 No Content
Date: Thu, 05 Jun 2020 17:45:55 GMT
Content-Type: application/json
Content-Length: 0
x-amzn-RequestId: [request-id]
Access-Control-Allow-Origin: *
x-amz-apigw-id: [apigw-id]
X-Amzn-Trace-Id: [trace-id]
Connection: keep-alive
```

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for Kotlin
- AWS SDK for PHP V3
- AWS SDK for Python (Boto3)
- AWS SDK for Ruby V3

UpdateComputeEnvironment

Updates an AWS Batch compute environment.

Request Syntax

```
POST /v1/updatecomputeenvironment HTTP/1.1
Content-type: application/json

{
   "computeEnvironment": "string",
   "computeResources": {
      "allocationStrategy": "string",
      "bidPercentage": number,
      "capacityTags": {
         "string" : "string"
      },
      "desiredvCpus": number,
      "ec2Configuration": [
         {
            "batchImageStatus": "string",
            "imageIdOverride": "string",
            "imageKubernetesVersion": "string",
            "imageType": "string"
         }
      ],
      "ec2KeyPair": "string",
      "imageId": "string",
      "instanceRole": "string",
      "instanceTypes": [ "string" ],
      "launchTemplate": {
         "launchTemplateId": "string",
         "launchTemplateName": "string",
         "overrides": [
            {
               "launchTemplateId": "string",
               "launchTemplateName": "string",
               "targetInstanceTypes": [ "string" ],
               "userdataType": "string",
               "version": "string"
            }
         ],
         "userdataType": "string",
```

```
      "version": "string"
   },
   "managedInstancesProvider": {
      "infrastructureOptimization": {
         "scaleInAfter": number
      },
      "infrastructureRoleArn": "string",
      "instanceLaunchTemplate": {
         "capacityReservations": {
            "reservationGroupArn": "string",
            "reservationPreference": "string"
         },
         "ec2InstanceProfileArn": "string",
         "instanceMetadataTagsPropagation": boolean,
         "instanceRequirements": {
            "allowedInstanceTypes": [ "string" ]
         },
         "localStorageConfiguration": {
            "useLocalStorage": boolean
         },
         "monitoring": "string",
         "networkConfiguration": {
            "securityGroups": [ "string" ],
            "subnets": [ "string" ]
         },
         "storageConfiguration": {
            "storageSizeGiB": number
         }
      },
      "propagateTags": "string"
   },
   "maxvCpus": number,
   "minvCpus": number,
   "placementGroup": "string",
   "scalingPolicy": {
      "minScaleDownDelayMinutes": number
   },
   "securityGroupIds": [ "string" ],
   "subnets": [ "string" ],
   "tags": {
      "string" : "string"
   },
   "type": "string",
   "updateToLatestImageVersion": boolean
```

```
    },
    "context": "string",
    "ecsSettings": {
       "containerInsights": "string"
    },
    "eksConfiguration": {
       "accessEntry": {
          "desiredState": "string",
          "status": "string"
       }
    },
    "serviceRole": "string",
    "state": "string",
    "unmanagedvCpus": number,
    "updatePolicy": {
       "jobExecutionTimeoutMinutes": number,
       "terminateJobsOnUpdate": boolean
    }
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

computeEnvironment

The name or full Amazon Resource Name (ARN) of the compute environment to update.

Type: String

Required: Yes

computeResources

Details of the compute resources managed by the compute environment. Required for a managed compute environment. For more information, see Compute Environments in the *AWS Batch User Guide*.

Type: ComputeResourceUpdate object

Required: No

context

Reserved.

Type: String

Required: No

ecsSettings

The Amazon ECS settings for the compute environment. These settings control CloudWatch Container Insights collection for the compute environment.

Type: EcsSettings object

Required: No

eksConfiguration

Updates the Amazon EKS configuration for the compute environment. Only specify this parameter if the compute environment's `containerOrchestrationType` is EKS. Currently, the `accessEntry` setting is the only Amazon EKS configuration that you can change after the compute environment is created. For more information, see Amazon EKS access entry authentication in the *AWS Batch User Guide*.

Type: EksConfigurationUpdate object

Required: No

serviceRole

The full Amazon Resource Name (ARN) of the IAM role that allows AWS Batch to make calls to other AWS services on your behalf. For more information, see AWS Batch service IAM role in the *AWS Batch User Guide*.

> **Important**
>
> If the compute environment has a service-linked role, it can't be changed to use a regular IAM role. Likewise, if the compute environment has a regular IAM role, it can't be changed to use a service-linked role. To update the parameters for the compute environment that require an infrastructure update to change, the

AWSServiceRoleForBatch service-linked role must be used. For more information, see Updating compute environments in the *AWS Batch User Guide*.

If your specified role has a path other than `/`, then you must either specify the full role ARN (recommended) or prefix the role name with the path.

> **Note**
>
> Depending on how you created your AWS Batch service role, its ARN might contain the `service-role` path prefix. When you only specify the name of the service role, AWS Batch assumes that your ARN doesn't use the `service-role` path prefix. Because of this, we recommend that you specify the full ARN of your service role when you create compute environments.

Type: String

Required: No

state

The state of the compute environment. Compute environments in the `ENABLED` state can accept jobs from a queue and scale in or out automatically based on the workload demand of its associated queues.

If the state is `ENABLED`, then the AWS Batch scheduler can attempt to place jobs from an associated job queue on the compute resources within the environment. If the compute environment is managed, then it can scale its instances out or in automatically, based on the job queue demand.

If the state is `DISABLED`, then the AWS Batch scheduler doesn't attempt to place jobs within the environment. Jobs in a `STARTING` or `RUNNING` state continue to progress normally. Managed compute environments in the `DISABLED` state don't scale out.

> **Note**
>
> Compute environments in a `DISABLED` state may continue to incur billing charges, for example, if they have running instances due to jobs that are still executing or a non-

zero `minvCpus` setting. To prevent additional charges, disable and delete the compute environment.

When an instance is idle, the instance scales down to the `minvCpus` value. However, the instance size doesn't change. For example, consider a `c5.8xlarge` instance with a `minvCpus` value of `4` and a `desiredvCpus` value of `36`. This instance doesn't scale down to a `c5.large` instance.

Type: String

Valid Values: `ENABLED | DISABLED`

Required: No

unmanagedvCpus

The maximum number of vCPUs expected to be used for an unmanaged compute environment. Don't specify this parameter for a managed compute environment. This parameter is only used for fair-share scheduling to reserve vCPU capacity for new share identifiers. If this parameter isn't provided for a fair-share job queue, no vCPU capacity is reserved.

Type: Integer

Required: No

updatePolicy

Specifies the updated infrastructure update policy for the compute environment. For more information about infrastructure updates, see Updating compute environments in the *AWS Batch User Guide*.

Type: UpdatePolicy object

Required: No

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
```

```
    "computeEnvironmentArn": "string",
    "computeEnvironmentName": "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.

computeEnvironmentArn

The Amazon Resource Name (ARN) of the compute environment.

Type: String

computeEnvironmentName

The name of the compute environment. It can be up to 128 characters long. It can contain uppercase and lowercase letters, numbers, hyphens (-), and underscores (_).

Type: String

Errors

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

Examples

In the following example or examples, the Authorization header contents (*[authorization-params]*) must be replaced with an AWS Signature Version 4 signature. For more information

about creating these signatures, see Signature Version 4 Signing Process in the *AWS General Reference*.

You only need to learn how to sign HTTP requests if you intend to manually create them. When you use the AWS Command Line Interface (AWS CLI) or one of the AWS SDKs to make requests to AWS, these tools automatically sign the requests for you with the access key that you specify when you configure the tools. When you use these tools, you don't need to learn how to sign requests yourself.

Example

This example disables the P3OnDemand compute environment so it can be deleted.

Sample Request

```
POST /v1/updatecomputeenvironment HTTP/1.1
Host: batch.us-east-1.amazonaws.com
Accept-Encoding: identity
Content-Length: [content-length]
Authorization: AUTHPARAMS
X-Amz-Date: 20161128T194248Z
User-Agent: aws-cli/1.11.21 Python/2.7.12 Darwin/16.1.0 botocore/1.4.78

{
  "computeEnvironment": "P3OnDemand",
  "state": "DISABLED"
}
```

Sample Response

```
HTTP/1.1 200 OK
Content-Type: application/json
Content-Length: [content-length]
Connection: keep-alive
Date: Mon, 28 Nov 2016 19:42:49 GMT
x-amzn-RequestId: [request-id]
X-Amzn-Trace-Id: [trace-id]
X-Cache: Miss from cloudfront
Via: 1.1 7f3f42df8af148df1f9f1ee7175987ad.cloudfront.net (CloudFront)
X-Amz-Cf-Id: uxJn0P7cg_1RTxOs15FkCItWfmCeniKMZdXlFWaOfPfjqATHw3j-AA==

{
```

```
  "computeEnvironmentName": "P3OnDemand",
  "computeEnvironmentArn": "arn:aws:batch:us-east-1:123456789012:compute-environment/
P3OnDemand"
}
```

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for Kotlin
- AWS SDK for PHP V3
- AWS SDK for Python (Boto3)
- AWS SDK for Ruby V3

UpdateConsumableResource

Updates a consumable resource.

Request Syntax

```
POST /v1/updateconsumableresource HTTP/1.1
Content-type: application/json

{
   "clientToken": "string",
   "consumableResource": "string",
   "operation": "string",
   "quantity": number
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

clientToken

If this parameter is specified and two update requests with identical payloads and `clientTokens` are received, these requests are considered the same request. Both requests will succeed, but the update will only happen once. A `clientToken` is valid for 8 hours.

Type: String

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

Required: No

consumableResource

The name or ARN of the consumable resource to be updated.

Type: String

Required: Yes

operation

Indicates how the quantity of the consumable resource will be updated. Must be one of:

- `SET`

  Sets the quantity of the resource to the value specified by the `quantity` parameter.

- `ADD`

  Increases the quantity of the resource by the value specified by the `quantity` parameter.

- `REMOVE`

  Reduces the quantity of the resource by the value specified by the `quantity` parameter.

Type: String

Required: No

quantity

The change in the total quantity of the consumable resource. The `operation` parameter determines whether the value specified here will be the new total quantity, or the amount by which the total quantity will be increased or reduced. Must be a non-negative value.

Type: Long

Required: No

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
   "consumableResourceArn": "string",
   "consumableResourceName": "string",
   "totalQuantity": 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.

consumableResourceArn

The Amazon Resource Name (ARN) of the consumable resource.

Type: String

consumableResourceName

The name of the consumable resource to be updated.

Type: String

totalQuantity

The total amount of the consumable resource that is available.

Type: Long

Errors

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4

- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for Kotlin
- AWS SDK for PHP V3
- AWS SDK for Python (Boto3)
- AWS SDK for Ruby V3

UpdateJobQueue

Updates a job queue.

Request Syntax

```
POST /v1/updatejobqueue HTTP/1.1
Content-type: application/json

{
   "computeEnvironmentOrder": [
      {
         "computeEnvironment": "string",
         "order": number
      }
   ],
   "jobQueue": "string",
   "jobStateTimeLimitActions": [
      {
         "action": "string",
         "maxTimeSeconds": number,
         "reason": "string",
         "state": "string"
      }
   ],
   "priority": number,
   "schedulingPolicyArn": "string",
   "serviceEnvironmentOrder": [
      {
         "order": number,
         "serviceEnvironment": "string"
      }
   ],
   "state": "string"
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

computeEnvironmentOrder

Details the set of compute environments mapped to a job queue and their order relative to each other. This is one of the parameters used by the job scheduler to determine which compute environment runs a given job. Compute environments must be in the `VALID` state before you can associate them with a job queue. All of the compute environments must be either EC2 (`EC2` or `SPOT`) or Fargate (`FARGATE` or `FARGATE_SPOT`). EC2 and Fargate compute environments can't be mixed.

> **Note**
>
> All compute environments that are associated with a job queue must share the same architecture. AWS Batch doesn't support mixing compute environment architecture types in a single job queue.

Type: Array of ComputeEnvironmentOrder objects

Required: No

jobQueue

The name or the Amazon Resource Name (ARN) of the job queue.

Type: String

Required: Yes

jobStateTimeLimitActions

The set of actions that AWS Batch perform on jobs that remain at the head of the job queue in the specified state longer than specified times. AWS Batch will perform each action after `maxTimeSeconds` has passed. (**Note**: The minimum value for maxTimeSeconds is 600 (10 minutes) and its maximum value is 86,400 (24 hours).)

Type: Array of JobStateTimeLimitAction objects

Required: No

priority

The priority of the job queue. Job queues with a higher priority (or a higher integer value for the `priority` parameter) are evaluated first when associated with the same compute environment. Priority is determined in descending order. For example, a job queue with a priority value of `10` is given scheduling preference over a job queue with a priority value of 1. All of the compute environments must be either EC2 (`EC2` or `SPOT`) or Fargate (`FARGATE` or `FARGATE_SPOT`). EC2 and Fargate compute environments can't be mixed.

Type: Integer

Required: No

schedulingPolicyArn

Amazon Resource Name (ARN) of the fair-share scheduling policy. Once a job queue is created, the fair-share scheduling policy can be replaced but not removed. The format is `aws:`*`Partition`*`:batch:`*`Region`*`:`*`Account`*`:scheduling-policy/`*`Name`* . For example, `aws:aws:batch:us-west-2:123456789012:scheduling-policy/MySchedulingPolicy`.

Type: String

Required: No

serviceEnvironmentOrder

The order of the service environment associated with the job queue. Job queues with a higher priority are evaluated first when associated with the same service environment.

Type: Array of ServiceEnvironmentOrder objects

Required: No

state

Describes the queue's ability to accept new jobs. If the job queue state is `ENABLED`, it can accept jobs. If the job queue state is `DISABLED`, new jobs can't be added to the queue, but jobs already in the queue can finish.

Type: String

Valid Values: `ENABLED | DISABLED`

Required: No

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
   "jobQueueArn": "string",
   "jobQueueName": "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.

jobQueueArn

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

Type: String

jobQueueName

The name of the job queue.

Type: String

Errors

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

Examples

In the following example or examples, the Authorization header contents (*[authorization-params]*) must be replaced with an AWS Signature Version 4 signature. For more information about creating these signatures, see Signature Version 4 Signing Process in the *AWS General Reference*.

You only need to learn how to sign HTTP requests if you intend to manually create them. When you use the AWS Command Line Interface (AWS CLI) or one of the AWS SDKs to make requests to AWS, these tools automatically sign the requests for you with the access key that you specify when you configure the tools. When you use these tools, you don't need to learn how to sign requests yourself.

Example

This example disables a job queue so that it can be deleted.

Sample Request

```
POST /v1/updatejobqueue HTTP/1.1
Host: batch.us-east-1.amazonaws.com
Accept-Encoding: identity
Content-Length: [content-length]
Authorization: [authorization-params]
X-Amz-Date: 20161128T201802Z
User-Agent: aws-cli/1.11.21 Python/2.7.12 Darwin/16.1.0 botocore/1.4.78

{
  "state": "DISABLED",
  "jobQueue": "GPGPU"
}
```

Sample Response

```
HTTP/1.1 200 OK
Date: Mon, 28 Nov 2016 20:18:03 GMT
Content-Type: application/json
Content-Length: [content-length]
Connection: keep-alive
x-amzn-RequestId: [request-id]
X-Amzn-Trace-Id: [trace-id]
```

```
X-Cache: Miss from cloudfront
Via: 1.1 17de248e6d780f737234d37cc490dbe3.cloudfront.net (CloudFront)
X-Amz-Cf-Id: aVju0hE8eLpjSFl8Y3fOuxgOZXdigQlLcDMwO0plxnynw0dEsOsEgw==

{
  "jobQueueName": "GPGPU",
  "jobQueueArn": "arn:aws:batch:us-east-1:123456789012:job-queue/GPGPU"
}
```

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for Kotlin
- AWS SDK for PHP V3
- AWS SDK for Python (Boto3)
- AWS SDK for Ruby V3

UpdateQuotaShare

Updates a quota share.

Request Syntax

```
POST /v1/updatequotashare HTTP/1.1
Content-type: application/json

{
   "capacityLimits": [
      {
         "capacityUnit": "string",
         "maxCapacity": number
      }
   ],
   "preemptionConfiguration": {
      "inSharePreemption": "string"
   },
   "quotaShareArn": "string",
   "resourceSharingConfiguration": {
      "borrowLimit": number,
      "strategy": "string"
   },
   "state": "string"
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

capacityLimits

A list that specifies the quantity and type of compute capacity allocated to the quota share.

Type: Array of QuotaShareCapacityLimit objects

Required: No

preemptionConfiguration

Specifies the preemption behavior for jobs in a quota share.

Type: QuotaSharePreemptionConfiguration object

Required: No

quotaShareArn

The Amazon Resource Name (ARN) of the quota share to update.

Type: String

Required: Yes

resourceSharingConfiguration

Specifies whether a quota share reserves, lends, or both lends and borrows idle compute capacity.

Type: QuotaShareResourceSharingConfiguration object

Required: No

state

The state of the quota share. If the quota share is `ENABLED`, it is able to accept jobs. If the quota share is `DISABLED`, new jobs won't be accepted but jobs already submitted can finish.

Type: String

Valid Values: `ENABLED | DISABLED`

Required: No

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
   "quotaShareArn": "string",
```

```
   "quotaShareName": "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.

quotaShareArn

The Amazon Resource Name (ARN) of the quota share.

Type: String

quotaShareName

The name of the quota share.

Type: String

Errors

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

Examples

In the following example or examples, the Authorization header contents (*[authorization-params]*) must be replaced with an AWS Signature Version 4 signature. For more information

about creating these signatures, see Signature Version 4 Signing Process in the *AWS General Reference*.

You only need to learn how to sign HTTP requests if you intend to manually create them. When you use the AWS Command Line Interface (AWS CLI) or one of the AWS SDKs to make requests to AWS, these tools automatically sign the requests for you with the access key that you specify when you configure the tools. When you use these tools, you don't need to learn how to sign requests yourself.

Example

This example updates a quota share to use a reserve strategy and disable in-share preemption.

Sample Request

```
POST /v1/updatequotashare HTTP/1.1
Host: batch.us-east-1.amazonaws.com
Accept-Encoding: identity
Content-Length: [content-length]
Authorization: [authorization-params]
X-Amz-Date: 20250801T001258Z
User-Agent: aws-cli/2.27.33 Python/3.13.4 Darwin/24.3.0

{
  "quotaShareArn": "arn:aws:batch:us-east-1:123456789012:job-queue/HighPriority/quota-
share/TeamAShare",
  "resourceSharingConfiguration": {
    "strategy": "RESERVE"
  },
  "preemptionConfiguration": {
    "inSharePreemption": "DISABLED"
  }
}
```

Sample Response

```
HTTP/1.1 200 OK
Content-Type: application/json
Content-Length: [content-length]
Connection: keep-alive
Date: Fri, 01 Aug 2025 00:12:59 GMT
x-amzn-RequestId: [request-id]
```

```
X-Amzn-Trace-Id: [trace-id]
X-Cache: Miss from cloudfront
Via: 1.1 example7k9m3n8q4r2w5x1z6c4vexample.cloudfront.net (CloudFront)
X-Amz-Cf-Id: whn1dX1uTx34Lvao7-7ZdkDXEbCZ_sjn3v3hHVFgbo1ORJtXyexample

{
  "quotaShareName": "TeamAShare",
  "quotaShareArn": "arn:aws:batch:us-east-1:123456789012:job-queue/HighPriority/quota-
share/TeamAShare"
}
```

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for Kotlin
- AWS SDK for PHP V3
- AWS SDK for Python (Boto3)
- AWS SDK for Ruby V3

UpdateSchedulingPolicy

Updates a scheduling policy.

Request Syntax

```
POST /v1/updateschedulingpolicy HTTP/1.1
Content-type: application/json

{
   "arn": "string",
   "fairsharePolicy": {
      "computeReservation": number,
      "shareDecaySeconds": number,
      "shareDistribution": [
         {
            "shareIdentifier": "string",
            "weightFactor": number
         }
      ]
   },
   "quotaSharePolicy": {
      "idleResourceAssignmentStrategy": "string"
   }
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

arn

The Amazon Resource Name (ARN) of the scheduling policy to update.

Type: String

Required: Yes

fairsharePolicy

The fair-share policy scheduling details. Once set during creation, a fairsharePolicy cannot be removed or changed to a quotaSharePolicy.

Type: FairsharePolicy object

Required: No

quotaSharePolicy

The quota share scheduling policy details. Once set during creation, a quotaSharePolicy cannot be removed or changed to a fairsharePolicy.

Type: QuotaSharePolicy 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

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

Examples

In the following example or examples, the Authorization header contents (*[authorization-params]*) must be replaced with an AWS Signature Version 4 signature. For more information about creating these signatures, see Signature Version 4 Signing Process in the *AWS General Reference*.

You only need to learn how to sign HTTP requests if you intend to manually create them. When you use the AWS Command Line Interface (AWS CLI) or one of the AWS SDKs to make requests to AWS, these tools automatically sign the requests for you with the access key that you specify when you configure the tools. When you use these tools, you don't need to learn how to sign requests yourself.

Example

This example removes the "*Stage*" tag from the job definition with an ARN of "*arn:aws:batch:us-east-1:123456789012:job-definition/sleep30:1*".

Sample Request

```
POST /v1/updateschedulingpolicy HTTP/1.1
Host: batch.us-east-1.amazonaws.com
Accept-Encoding: identity
User-Agent: aws-cli/1.20.21 Python/3.6.9 Linux/4.4.0-19041-Microsoft botocore/1.21.21
X-Amz-Date: 20210929T022142Z
X-Amz-Security-Token: [security-token]
Authorization: [authorization-params]


{
  "arn": "arn:aws:batch:us-east-1:123456789012:scheduling-policy/
ExampleFairSharePolicy3",
  "fairsharePolicy": {
    "shareDecaySeconds": 3600,
    "computeReservation": 1,
    "shareDistribution": [
      {
        "shareIdentifier": "MostImportant",
        "weightFactor": 0.0001
      },{
        "shareIdentifier": "Normal",
        "weightFactor": 1.0
```

```
      },{
        "shareIdentifier": "LeastImportant",
        "weightFactor": 999.9999
      }
    ]
  }
}
```

Sample Response

```
HTTP/1.1 200 OK
Date: Wed, 29 Sep 2021 02:21:43 GMT
Content-Type: application/json
Content-Length: 2
x-amzn-RequestId: [request-id]
Access-Control-Allow-Origin: *
x-amz-apigw-id: [apigw-id]
Access-Control-Expose-Headers: X-amzn-errortype,X-amzn-requestid,X-amzn-errormessage,X-
amzn-trace-id,X-amz-apigw-id,date
X-Amzn-Trace-Id: [trace-id]
Connection: keep-alive

{}
```

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for Kotlin
- AWS SDK for PHP V3
- AWS SDK for Python (Boto3)

- AWS SDK for Ruby V3

UpdateServiceEnvironment

Updates a service environment. You can update the state of a service environment from ENABLED to DISABLED to prevent new service jobs from being placed in the service environment.

Request Syntax

```
POST /v1/updateserviceenvironment HTTP/1.1
Content-type: application/json

{
   "capacityLimits": [
      {
         "capacityUnit": "string",
         "maxCapacity": number
      }
   ],
   "serviceEnvironment": "string",
   "state": "string"
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

capacityLimits

The capacity limits for the service environment. This defines the maximum resources that can be used by service jobs in this environment.

Type: Array of CapacityLimit objects

Required: No

serviceEnvironment

The name or ARN of the service environment to update.

Type: String

Required: Yes

state

The state of the service environment.

Type: String

Valid Values: `ENABLED | DISABLED`

Required: No

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
   "serviceEnvironmentArn": "string",
   "serviceEnvironmentName": "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.

serviceEnvironmentArn

The Amazon Resource Name (ARN) of the service environment that was updated.

Type: String

serviceEnvironmentName

The name of the service environment that was updated.

Type: String

Errors

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

Examples

In the following example or examples, the Authorization header contents (*[authorization-params]*) must be replaced with an AWS Signature Version 4 signature. For more information about creating these signatures, see Signature Version 4 Signing Process in the *AWS General Reference*.

You only need to learn how to sign HTTP requests if you intend to manually create them. When you use the AWS Command Line Interface (AWS CLI) or one of the AWS SDKs to make requests to AWS, these tools automatically sign the requests for you with the access key that you specify when you configure the tools. When you use these tools, you don't need to learn how to sign requests yourself.

Example

This example updates a service environment to disable it.

Sample Request

```
POST /v1/updateserviceenvironment HTTP/1.1
Host: batch.us-east-1.amazonaws.com
Accept-Encoding: identity
Content-Length: [content-length]
Authorization: [authorization-params]
X-Amz-Date: 20250801T154520Z
```

```
User-Agent: aws-cli/2.27.33 Python/3.13.4 Darwin/24.3.0

{
  "serviceEnvironment": "SageMakerTrainingEnv",
  "state": "DISABLED"
}
```

Sample Response

```
HTTP/1.1 200 OK
Content-Type: application/json
Content-Length: [content-length]
Connection: keep-alive
Date: Fri, 01 Aug 2025 15:45:21 GMT
x-amzn-RequestId: [request-id]
X-Amzn-Trace-Id: [trace-id]
X-Cache: Miss from cloudfront
Via: 1.1 25g84de7k2m5n8p1q4r9s6t3w2xexample.cloudfront.net (CloudFront)
X-Amz-Cf-Id: ghi4jkl7mno0pqr3stu6vwx9yz2345fghijklmnopqrstuexample

{
  "serviceEnvironmentName": "SageMakerTrainingEnv",
  "serviceEnvironmentArn": "arn:aws:batch:us-east-1:123456789012:service-environment/
SageMakerTrainingEnv"
}
```

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for Kotlin
- AWS SDK for PHP V3

- AWS SDK for Python (Boto3)
- AWS SDK for Ruby V3

UpdateServiceJob

Updates the priority of a specified service job in an AWS Batch job queue.

Request Syntax

```
POST /v1/updateservicejob HTTP/1.1
Content-type: application/json

{
   "jobId": "string",
   "schedulingPriority": number
}
```

URI Request Parameters

The request does not use any URI parameters.

Request Body

The request accepts the following data in JSON format.

jobId

The AWS Batch job ID of the job to update.

Type: String

Required: Yes

schedulingPriority

The scheduling priority for the job. This only affects jobs in job queues with a quota-share or fair-share scheduling policy. Jobs with a higher scheduling priority are scheduled before jobs with a lower scheduling priority within a share.

The minimum supported value is 0 and the maximum supported value is 9999.

Type: Integer

Required: Yes

Response Syntax

```
HTTP/1.1 200
Content-type: application/json

{
   "jobArn": "string",
   "jobId": "string",
   "jobName": "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

The Amazon Resource Name (ARN) for the job.

Type: String

jobId

The unique identifier for the job.

Type: String

jobName

The name of the job.

Type: String

Errors

ClientException

These errors are usually caused by a client action. One example cause is using an action or resource on behalf of a user that doesn't have permissions to use the action or resource. Another cause is specifying an identifier that's not valid.

HTTP Status Code: 400

ServerException

These errors are usually caused by a server issue.

HTTP Status Code: 500

See Also

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

- AWS Command Line Interface V2
- AWS SDK for .NET V4
- AWS SDK for C++
- AWS SDK for Go v2
- AWS SDK for Java V2
- AWS SDK for JavaScript V3
- AWS SDK for Kotlin
- AWS SDK for PHP V3
- AWS SDK for Python (Boto3)
- AWS SDK for Ruby V3

Data Types

The AWS Batch API contains several data types that various actions use. This section describes each data type in detail.

> **Note**
>
> The order of each element in a data type structure is not guaranteed. Applications should not assume a particular order.

The following data types are supported:

- ArrayProperties
- ArrayPropertiesDetail
- ArrayPropertiesSummary
- AttemptContainerDetail
- AttemptDetail
- AttemptEcsTaskDetails
- AttemptTaskContainerDetails
- CancelJobsErrorDetail
- CapacityLimit
- CapacityReservationRequest
- ComputeEnvironmentDetail
- ComputeEnvironmentOrder
- ComputeResource
- ComputeResourceUpdate
- ComputeScalingPolicy
- ConsumableResourceProperties
- ConsumableResourceRequirement
- ConsumableResourceSummary
- ContainerDetail
- ContainerOverrides

- ContainerProperties
- ContainerSummary
- Device
- Ec2Configuration
- EcsProperties
- EcsPropertiesDetail
- EcsPropertiesOverride
- EcsSettings
- EcsTaskDetails
- EcsTaskProperties
- EFSAuthorizationConfig
- EFSVolumeConfiguration
- EksAccessEntry
- EksAttemptContainerDetail
- EksAttemptDetail
- EksConfiguration
- EksConfigurationUpdate
- EksContainer
- EksContainerDetail
- EksContainerEnvironmentVariable
- EksContainerOverride
- EksContainerResourceRequirements
- EksContainerSecurityContext
- EksContainerVolumeMount
- EksEmptyDir
- EksHostPath
- EksMetadata
- EksPersistentVolumeClaim
- EksPodProperties
- EksPodPropertiesDetail

- EksPodPropertiesOverride
- EksProperties
- EksPropertiesDetail
- EksPropertiesOverride
- EksSecret
- EksVolume
- EphemeralStorage
- EvaluateOnExit
- FairshareCapacityUsage
- FairshareCapacityUtilization
- FairsharePolicy
- FairshareUtilizationDetail
- FargatePlatformConfiguration
- FirelensConfiguration
- FrontOfQueueDetail
- FrontOfQueueJobSummary
- FrontOfQuotaShareJobSummary
- FrontOfQuotaSharesDetail
- Host
- ImagePullSecret
- InfrastructureOptimization
- InstanceLaunchTemplate
- InstanceLaunchTemplateUpdate
- InstanceRequirementsRequest
- JobCapacityUsageSummary
- JobDefinition
- JobDependency
- JobDetail
- JobQueueDetail
- JobStateTimeLimitAction

- JobSummary
- JobTimeout
- KeyValuePair
- KeyValuesPair
- LatestServiceJobAttempt
- LaunchTemplateSpecification
- LaunchTemplateSpecificationOverride
- LinuxParameters
- ListJobsByConsumableResourceSummary
- LogConfiguration
- ManagedInstancesLocalStorageConfiguration
- ManagedInstancesNetworkConfiguration
- ManagedInstancesProvider
- ManagedInstancesStorageConfiguration
- MountPoint
- NetworkConfiguration
- NetworkInterface
- NodeDetails
- NodeOverrides
- NodeProperties
- NodePropertiesSummary
- NodePropertyOverride
- NodeRangeProperty
- QueueSnapshotCapacityUsage
- QueueSnapshotUtilizationDetail
- QuotaShareCapacityLimit
- QuotaShareCapacityUsage
- QuotaShareCapacityUtilization
- QuotaShareDetail
- QuotaSharePolicy

- QuotaSharePreemptionConfiguration
- QuotaShareResourceSharingConfiguration
- QuotaShareUtilizationDetail
- RepositoryCredentials
- ResourceRequirement
- RetryStrategy
- RuntimePlatform
- S3FilesVolumeConfiguration
- SchedulingPolicyDetail
- SchedulingPolicyListingDetail
- Secret
- ServiceEnvironmentDetail
- ServiceEnvironmentOrder
- ServiceJobAttemptDetail
- ServiceJobCapacityUsageDetail
- ServiceJobCapacityUsageSummary
- ServiceJobEvaluateOnExit
- ServiceJobPreemptedAttempt
- ServiceJobPreemptionConfiguration
- ServiceJobPreemptionSummary
- ServiceJobRetryStrategy
- ServiceJobSummary
- ServiceJobTimeout
- ServiceResourceId
- ShareAttributes
- TaskContainerDependency
- TaskContainerDetails
- TaskContainerOverrides
- TaskContainerProperties
- TaskPropertiesOverride

- TerminateJobsErrorDetail
- TerminateServiceJobsErrorDetail
- Tmpfs
- Ulimit
- UpdateManagedInstancesProviderConfiguration
- UpdatePolicy
- Volume

ArrayProperties

An object that represents an AWS Batch array job.

Contents

size

The size of the array 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 Java V2
- AWS SDK for Ruby V3

ArrayPropertiesDetail

An object that represents the array properties of a job.

Contents

index

The job index within the array that's associated with this job. This parameter is returned for array job children.

Type: Integer

Required: No

size

The size of the array job. This parameter is returned for parent array jobs.

Type: Integer

Required: No

statusSummary

A summary of the number of array job children in each available job status. This parameter is returned for parent array jobs.

Type: String to integer map

Required: No

statusSummaryLastUpdatedAt

The Unix timestamp (in milliseconds) for when the `statusSummary` was last updated.

Type: Long

Required: No

See Also

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

- AWS SDK for C++
- AWS SDK for Java V2
- AWS SDK for Ruby V3

ArrayPropertiesSummary

An object that represents the array properties of a job.

Contents

index

The job index within the array that's associated with this job. This parameter is returned for children of array jobs.

Type: Integer

Required: No

size

The size of the array job. This parameter is returned for parent array jobs.

Type: Integer

Required: No

statusSummary

A summary of the number of array job children in each available job status. This parameter is returned for parent array jobs.

Type: String to integer map

Required: No

statusSummaryLastUpdatedAt

The Unix timestamp (in milliseconds) for when the `statusSummary` was last updated.

Type: Long

Required: No

See Also

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

- AWS SDK for C++
- AWS SDK for Java V2
- AWS SDK for Ruby V3

AttemptContainerDetail

An object that represents the details of a container that's part of a job attempt.

Contents

containerInstanceArn

The Amazon Resource Name (ARN) of the Amazon ECS container instance that hosts the job attempt.

Type: String

Required: No

exitCode

The exit code for the job attempt. A non-zero exit code is considered failed.

Type: Integer

Required: No

logStreamName

The name of the CloudWatch Logs log stream that's associated with the container. The log group for AWS Batch jobs is `/aws/batch/job`. Each container attempt receives a log stream name when they reach the `RUNNING` status.

Type: String

Required: No

networkInterfaces

The network interfaces that are associated with the job attempt.

Type: Array of NetworkInterface objects

Required: No

reason

A short (255 max characters) human-readable string to provide additional details for a running or stopped container.

Type: String

Required: No

taskArn

The Amazon Resource Name (ARN) of the Amazon ECS task that's associated with the job attempt. Each container attempt receives a task ARN when they reach the `STARTING` status.

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 Java V2
- AWS SDK for Ruby V3

AttemptDetail

An object that represents a job attempt.

Contents

container

The details for the container in this job attempt.

Type: AttemptContainerDetail object

Required: No

startedAt

The Unix timestamp (in milliseconds) for when the attempt was started (when the attempt transitioned from the `STARTING` state to the `RUNNING` state).

Type: Long

Required: No

statusReason

A short, human-readable string to provide additional details for the current status of the job attempt.

Type: String

Required: No

stoppedAt

The Unix timestamp (in milliseconds) for when the attempt was stopped (when the attempt transitioned from the `RUNNING` state to a terminal state, such as `SUCCEEDED` or `FAILED`).

Type: Long

Required: No

taskProperties

The properties for a task definition that describes the container and volume definitions of an Amazon ECS task.

Type: Array of AttemptEcsTaskDetails 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 Java V2
- AWS SDK for Ruby V3

AttemptEcsTaskDetails

An object that represents the details of a task.

Contents

containerInstanceArn

The Amazon Resource Name (ARN) of the container instance that hosts the task.

Type: String

Required: No

containers

A list of containers that are included in the `taskProperties` list.

Type: Array of AttemptTaskContainerDetails objects

Required: No

taskArn

The ARN of the Amazon ECS task.

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 Java V2
- AWS SDK for Ruby V3

AttemptTaskContainerDetails

An object that represents the details of a container that's part of a job attempt.

Contents

exitCode

The exit code for the container’s attempt. A non-zero exit code is considered failed.

Type: Integer

Required: No

logStreamName

The name of the Amazon CloudWatch Logs log stream that's associated with the container. The log group for AWS Batch jobs is `/aws/batch/job`. Each container attempt receives a log stream name when they reach the `RUNNING` status.

Type: String

Required: No

name

The name of a container.

Type: String

Required: No

networkInterfaces

The network interfaces that are associated with the job attempt.

Type: Array of NetworkInterface objects

Required: No

reason

A short (255 max characters) string that's easy to understand and provides additional details for a running or stopped container.

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 Java V2
- AWS SDK for Ruby V3

CancelJobsErrorDetail

An object that contains the details of a job that couldn't be cancelled by a `CancelJobs` operation.

Contents

code

An error code that identifies the reason the job couldn't be cancelled. Valid values are:

- `ValidationException` – A job identifier in the request is malformed or isn't valid.
- `ClientException` – The request failed because of a client error.
- `ThrottlingException` – The request was throttled. Retry the request.
- `ServerException` – An internal error occurred. Retry the request.
- `AccessDenied` – The caller isn't authorized to perform the action on the specified job.

Type: String

Required: Yes

job

The AWS Batch job ID of the job that couldn't be cancelled.

Type: String

Required: Yes

message

A message that describes the reason the job couldn't be cancelled.

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 Java V2
- AWS SDK for Ruby V3

CapacityLimit

Defines the type and maximum quantity of resources that can be allocated to service jobs in a service environment.

Contents

capacityUnit

The unit of measure for the capacity limit, which defines how `maxCapacity` is interpreted. For `SAGEMAKER_TRAINING` jobs in a quota management enabled service environment, specify the instance type (for example, `ml.m5.large`). Otherwise, use `NUM_INSTANCES`.

Type: String

Required: No

maxCapacity

The maximum capacity available for the service environment. For a quota management enabled service environment, this value represents the maximum quantity of a particular resource type (specified by `capacityUnit`) that can be allocated to service jobs. For other service environments, this value represents the maximum quantity of all resources that can be allocated to service jobs.

For example, if `maxCapacity=50` and `capacityUnit=NUM_INSTANCES`, you can run up to 50 instances concurrently. If you run 5 SageMaker Training jobs that each use 10 instances, a subsequent job requiring 10 instances waits in the queue until capacity is available. In a quota management enabled service environment with `capacityUnit=ml.m5.large`, only `ml.m5.large` instances count against this limit, and jobs requiring other instance types wait until a matching capacity limit is configured.

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 Java V2
- AWS SDK for Ruby V3

CapacityReservationRequest

The capacity reservation configuration for Amazon ECS Managed Instances. Use this to target On-Demand Capacity Reservations or Reserved Instances.

Contents

reservationGroupArn

The Amazon Resource Name (ARN) of the capacity reservation group to target.

Type: String

Required: No

reservationPreference

The capacity reservation preference. Valid values:

- `RESERVATIONS_ONLY` — Use only capacity reservations.
- `RESERVATIONS_FIRST` — Prefer capacity reservations but fall back to On-Demand if unavailable.
- `RESERVATIONS_EXCLUDED` — Do not use capacity reservations.

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 Java V2
- AWS SDK for Ruby V3

ComputeEnvironmentDetail

An object that represents an AWS Batch compute environment.

Contents

computeEnvironmentArn

The Amazon Resource Name (ARN) of the compute environment.

Type: String

Required: Yes

computeEnvironmentName

The name of the compute environment. It can be up to 128 characters long. It can contain uppercase and lowercase letters, numbers, hyphens (-), and underscores (_).

Type: String

Required: Yes

computeResources

The compute resources defined for the compute environment. For more information, see Compute environments in the *AWS Batch User Guide*.

Type: ComputeResource object

Required: No

containerOrchestrationType

The orchestration type of the compute environment. The valid values are ECS (default) or EKS.

Type: String

Valid Values: ECS | EKS

Required: No

context

Reserved.

Type: String

Required: No

ecsClusterArn

The Amazon Resource Name (ARN) of the underlying Amazon ECS cluster that the compute environment uses.

Type: String

Required: No

ecsSettings

The Amazon ECS settings for the compute environment. These settings control CloudWatch Container Insights collection.

Type: EcsSettings object

Required: No

eksConfiguration

The configuration for the Amazon EKS cluster that supports the AWS Batch compute environment. Only specify this parameter if the `containerOrchestrationType` is `EKS`.

Type: EksConfiguration object

Required: No

serviceRole

The service role that's associated with the compute environment that allows AWS Batch to make calls to AWS API operations on your behalf. For more information, see Batch service IAM role in the *AWS Batch User Guide*.

Type: String

Required: No

state

The state of the compute environment. The valid values are `ENABLED` or `DISABLED`.

If the state is `ENABLED`, then the AWS Batch scheduler can attempt to place jobs from an associated job queue on the compute resources within the environment. If the compute

environment is managed, then it can scale its instances out or in automatically based on the job queue demand.

If the state is `DISABLED`, then the AWS Batch scheduler doesn't attempt to place jobs within the environment. Jobs in a `STARTING` or `RUNNING` state continue to progress normally. Managed compute environments in the `DISABLED` state don't scale out.

> **Note**
>
> Compute environments in a `DISABLED` state may continue to incur billing charges, for example, if they have running instances due to jobs that are still executing or a non-zero `minvCpus` setting. To prevent additional charges, disable and delete the compute environment.

When an instance is idle, the instance scales down to the `minvCpus` value. However, the instance size doesn't change. For example, consider a `c5.8xlarge` instance with a `minvCpus` value of `4` and a `desiredvCpus` value of `36`. This instance doesn't scale down to a `c5.large` instance.

Type: String

Valid Values: `ENABLED | DISABLED`

Required: No

status

The current status of the compute environment (for example, `CREATING` or `VALID`).

Type: String

Valid Values: `CREATING | UPDATING | DELETING | DELETED | VALID | INVALID`

Required: No

statusReason

A short, human-readable string to provide additional details for the current status of the compute environment.

Type: String

Required: No

tags

The tags applied to the compute environment.

Type: String to string map

Map Entries: Maximum number of 50 items.

Key Length Constraints: Minimum length of 1. Maximum length of 128.

Value Length Constraints: Maximum length of 256.

Required: No

type

The type of the compute environment: `MANAGED` or `UNMANAGED`. For more information, see Compute environments in the *AWS Batch User Guide*.

Type: String

Valid Values: `MANAGED | UNMANAGED`

Required: No

unmanagedvCpus

The maximum number of VCPUs expected to be used for an unmanaged compute environment.

Type: Integer

Required: No

updatePolicy

Specifies the infrastructure update policy for the compute environment. For more information about infrastructure updates, see Updating compute environments in the *AWS Batch User Guide*.

Type: UpdatePolicy object

Required: No

uuid

Unique identifier for the compute environment.

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 Java V2
- AWS SDK for Ruby V3

ComputeEnvironmentOrder

The order that compute environments are tried in for job placement within a queue. Compute environments are tried in ascending order. For example, if two compute environments are associated with a job queue, the compute environment with a lower order integer value is tried for job placement first. Compute environments must be in the `VALID` state before you can associate them with a job queue. All of the compute environments must be either EC2 (`EC2` or `SPOT`) or Fargate (`FARGATE` or `FARGATE_SPOT`); Amazon EC2 and Fargate compute environments can't be mixed.

> **Note**
>
> All compute environments that are associated with a job queue must share the same architecture. AWS Batch doesn't support mixing compute environment architecture types in a single job queue.

Contents

computeEnvironment

The Amazon Resource Name (ARN) of the compute environment.

Type: String

Required: Yes

order

The order of the compute environment. Compute environments are tried in ascending order. For example, if two compute environments are associated with a job queue, the compute environment with a lower `order` integer value is tried for job placement first.

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 Java V2
- AWS SDK for Ruby V3

ComputeResource

An object that represents an AWS Batch compute resource. For more information, see Compute environments in the *AWS Batch User Guide*.

Contents

maxvCpus

The maximum number of vCPUs that a compute environment can support.

> **Note**
>
> With any allocation strategy except `BEST_FIT` using On-Demand (`EC2`) compute resources, AWS Batch might need to exceed `maxvCpus` to meet your capacity requirements. In this event, AWS Batch never exceeds `maxvCpus` by more than a single instance.

Type: Integer

Required: Yes

type

The type of compute environment: `EC2`, `SPOT`, `FARGATE`, `FARGATE_SPOT`, or `ECS_MANAGED_INSTANCES`. For more information, see Compute environments in the *AWS Batch User Guide*.

If you choose `SPOT`, you must also specify an Amazon EC2 Spot Fleet role with the `spotIamFleetRole` parameter. For more information, see Amazon EC2 spot fleet role in the *AWS Batch User Guide*.

If you choose `ECS_MANAGED_INSTANCES`, you must also specify a `managedInstancesProvider` configuration. To use Spot capacity, set `capacityOptionType` to `SPOT` in the `managedInstancesProvider.instanceLaunchTemplate` configuration. For more information, see Amazon ECS Managed Instances compute environments in the *AWS Batch User Guide*.

> **Note**
>
> Multi-node parallel jobs aren't supported on Spot Instances or Amazon ECS Managed Instances.

Type: String

Valid Values: `EC2` | `SPOT` | `FARGATE` | `FARGATE_SPOT` | `ECS_MANAGED_INSTANCES`

Required: Yes

allocationStrategy

The allocation strategy to use for the compute resource if not enough instances of the best fitting instance type can be allocated. This might be because of availability of the instance type in the Region or Amazon EC2 service limits. For more information, see Allocation strategies in the *AWS Batch User Guide*.

> **Note**
>
> This parameter isn't applicable to jobs that are running on Fargate resources. Don't specify it.

> **Note**
>
> This parameter is required for Amazon EKS compute environments. For Amazon ECS compute environments, if this parameter isn't specified, the `BEST_FIT` allocation strategy is used by default.

BEST_FIT (default)

AWS Batch selects an instance type that best fits the needs of the jobs with a preference for the lowest-cost instance type. If additional instances of the selected instance type aren't available, AWS Batch waits for the additional instances to be available. If there aren't enough instances available or the user is reaching Amazon EC2 service limits, additional jobs aren't run until the currently running jobs are completed. This allocation strategy keeps costs

lower but can limit scaling. If you're using Spot Fleets with BEST_FIT, the Spot Fleet IAM Role must be specified. Compute resources that use a BEST_FIT allocation strategy don't support infrastructure updates and can't update some parameters. For more information, see Updating compute environments in the *AWS Batch User Guide*.

BEST_FIT_PROGRESSIVE

AWS Batch selects additional instance types that are large enough to meet the requirements of the jobs in the queue. Its preference is for instance types with lower cost vCPUs. If additional instances of the previously selected instance types aren't available, AWS Batch selects new instance types.

BEST_FIT_PROGRESSIVE_ORDERED

> **Important**
>
> This is an advanced allocation strategy only for customers who want to control which instance types are preferred during scaling.
> Placing large instance types at the top of the list may result in **over-provisioning** for small jobs. Placing small instance types at the top may cause the compute environment to reach Amazon EC2 instance count limits before reaching `maxvCpus`.

AWS Batch selects instance types in the order they appear in the `instanceTypes` list. When an instance family is specified, sizes within that family are expanded using BEST_FIT_PROGRESSIVE logic—preferring sizes that best fit the jobs, with larger sizes as fallback. Instance types that cannot meet the resource requirements of the jobs are skipped. This strategy is only available for On-Demand Instance (EC2) compute resources.

If an instance family and an explicit instance type from that family both appear in `instanceTypes`, the explicit type takes its listed position and is excluded from the family expansion. For example, in `["m7a.4xlarge", "m7a", "m6a"]`, `m7a.4xlarge` is always placed first and is excluded from the `m7a` family expansion.

SPOT_CAPACITY_OPTIMIZED

AWS Batch selects one or more instance types that are large enough to meet the requirements of the jobs in the queue. Its preference is for instance types that are less likely to be interrupted. This allocation strategy is only available for Spot Instance compute resources.

SPOT_PRICE_CAPACITY_OPTIMIZED

The price and capacity optimized allocation strategy looks at both price and capacity to select the Spot Instance pools that are the least likely to be interrupted and have the lowest possible price. This allocation strategy is only available for Spot Instance compute resources.

SPOT_CAPACITY_OPTIMIZED_PRIORITIZED

> **Important**
>
> This is an advanced allocation strategy for customers who want to influence instance type selection during scaling. This strategy optimizes for **capacity first**, and honors instance type priorities on a best-effort basis (priorities are honored when they do not significantly reduce available Spot capacity).
> Placing large instance types at the top of the list may result in **over-provisioning** for small jobs. Placing small instance types at the top may cause the compute environment to reach Amazon EC2 instance count limits before reaching `maxvCpus`.

AWS Batch selects instance types in the order they appear in the `instanceTypes` list, but **optimizes for capacity first**. The customer-defined priority is honored on a best-effort basis. When Spot Instance capacity pools are similarly available, priority order is respected. When capacity is constrained, AWS Batch selects from the most available pools regardless of priority to minimize the likelihood of Spot Instance interruptions. This strategy is only available for Spot Instance compute resources.

With any allocation strategy except `BEST_FIT` using On-Demand (`EC2`) compute resources, AWS Batch might need to exceed `maxvCpus` to meet your capacity requirements. In this event, AWS Batch never exceeds `maxvCpus` by more than a single instance.

Type: String

Valid Values: `BEST_FIT | BEST_FIT_PROGRESSIVE | BEST_FIT_PROGRESSIVE_ORDERED | SPOT_CAPACITY_OPTIMIZED | SPOT_PRICE_CAPACITY_OPTIMIZED | SPOT_CAPACITY_OPTIMIZED_PRIORITIZED`

Required: No

bidPercentage

The maximum percentage that a Spot Instance price can be when compared with the On-Demand price for that instance type before instances are launched. For example, if your maximum percentage is 20%, then the Spot price must be less than 20% of the current On-Demand price for that Amazon EC2 instance. You always pay the lowest (market) price and never more than your maximum percentage. If you leave this field empty, the default value is 100% of the On-Demand price. For most use cases, we recommend leaving this field empty.

> **Note**
>
> This parameter isn't applicable to jobs that are running on Fargate resources. Don't specify it.

Type: Integer

Required: No

capacityTags

The tags to apply to the Amazon ECS capacity provider and Amazon EC2 instances launched by the compute environment. These tags are separate from the compute environment resource tags (the top-level `tags` parameter). Use `capacityTags` for cost allocation and organization of the underlying infrastructure resources.

This parameter is only valid for `ECS_MANAGED_INSTANCES` compute environments. You must have the `batch:SetCapacityTags` permission on the compute environment resource to use this parameter.

Type: String to string map

Map Entries: Maximum number of 50 items.

Key Length Constraints: Minimum length of 1. Maximum length of 128.

Value Length Constraints: Maximum length of 256.

Required: No

desiredvCpus

The desired number of vCPUS in the compute environment. AWS Batch modifies this value between the minimum and maximum values based on job queue demand.

> **Note**
>
> This parameter isn't applicable to jobs that are running on Fargate resources. Don't specify it.

Type: Integer

Required: No

ec2Configuration

Provides information that's used to select Amazon Machine Images (AMIs) for Amazon EC2 instances in the compute environment. If `Ec2Configuration` isn't specified, the default is `ECS_AL2023` for EC2 (ECS) compute environments and `EKS_AL2023` for EKS compute environments.

One or two values can be provided.

> **Note**
>
> This parameter isn't applicable to jobs that are running on Fargate resources. Don't specify it.

Type: Array of Ec2Configuration objects

Required: No

ec2KeyPair

The Amazon EC2 key pair that's used for instances launched in the compute environment. You can use this key pair to log in to your instances with SSH.

> **Note**
>
> This parameter isn't applicable to jobs that are running on Fargate resources. Don't specify it.

Type: String

Required: No

imageId

This member has been deprecated.

The Amazon Machine Image (AMI) ID used for instances launched in the compute environment. This parameter is overridden by the `imageIdOverride` member of the `Ec2Configuration` structure.

> **Note**
>
> This parameter isn't applicable to jobs that are running on Fargate resources. Don't specify it.

> **Note**
>
> The AMI that you choose for a compute environment must match the architecture of the instance types that you intend to use for that compute environment. For example, if your compute environment uses A1 instance types, the compute resource AMI that you choose must support ARM instances. Amazon ECS vends both x86 and ARM versions of the Amazon ECS-optimized Amazon Linux 2023 AMI. For more information, see Amazon ECS-optimized Amazon Linux 2023 AMI in the *Amazon Elastic Container Service Developer Guide*.

Type: String

Required: No

instanceRole

The Amazon ECS instance profile applied to Amazon EC2 instances in a compute environment. This parameter is required for Amazon EC2 instances types. You can specify the short name or full Amazon Resource Name (ARN) of an instance profile. For example, *ecsInstanceRole* or `arn:aws:iam::<aws_account_id>:instance-profile/ecsInstanceRole` . For more information, see Amazon ECS instance role in the *AWS Batch User Guide*.

> **Note**
>
> This parameter isn't applicable to jobs that are running on Fargate resources. Don't specify it.

Type: String

Required: No

instanceTypes

The instances types that can be launched. You can specify instance families to launch any instance type within those families (for example, `c5` or `p3`), or you can specify specific sizes within a family (such as `c5.8xlarge`).

AWS Batch can select the instance type for you if you choose one of the following:

- `default_x86_64` to choose x86 based instance types (from the `m6i`, `c6i`, `r6i`, and `c7i` instance families) that matches the resource demands of the job queue.
- `default_arm64` to choose ARM based instance types (from the `m6g`, `c6g`, `r6g`, and `c7g` instance families) that matches the resource demands of the job queue.
- `optimal` Semantically equivalent to `default_x86_64`, see Optimal instance type configuration to receive automatic instance family updates for details.

> **Note**
>
> Instance family availability varies by AWS Region. For example, some AWS Regions may not have any fourth generation instance families but have fifth and sixth generation instance families.
> When using `default_x86_64` or `default_arm64` instance bundles, AWS Batch selects instance families based on a balance of cost-effectiveness and performance. While

newer generation instances often provide better price-performance, AWS Batch may choose an earlier generation instance family if it provides the optimal combination of availability, cost, and performance for your workload. For example, in an AWS Region where both c6i and c7i instances are available, AWS Batch might select c6i instances if they offer better cost-effectiveness for your specific job requirements. For more information on AWS Batch instance types and AWS Region availability, see Instance type compute table in the *AWS Batch User Guide*.
AWS Batch periodically updates your instances in default bundles to newer, more cost-effective options. Updates happen automatically without requiring any action from you. Your workloads continue running during updates with no interruption

Note

This parameter isn't applicable to jobs that are running on Fargate resources. Don't specify it.

Note

When you create a compute environment, the instance types that you select for the compute environment must share the same architecture. For example, you can't mix x86 and ARM instances in the same compute environment.

Type: Array of strings

Required: No

launchTemplate

The launch template to use for your compute resources. Any other compute resource parameters that you specify in a CreateComputeEnvironment API operation override the same parameters in the launch template. You must specify either the launch template ID or launch template name in the request, but not both. For more information, see Launch template support in the *AWS Batch User Guide*.

> **Note**
>
> This parameter isn't applicable to jobs that are running on Fargate resources. Don't specify it.

Type: LaunchTemplateSpecification object

Required: No

managedInstancesProvider

The configuration for the Amazon ECS Managed Instances capacity provider. This parameter is required when `computeResources.type` is `ECS_MANAGED_INSTANCES` and must not be specified for other compute environment types.

For more information, see Amazon ECS Managed Instances compute environments in the *AWS Batch User Guide*.

Type: ManagedInstancesProvider object

Required: No

minvCpus

The minimum number of vCPUs that a compute environment should maintain (even if the compute environment is `DISABLED`).

> **Note**
>
> This parameter isn't applicable to jobs that are running on Fargate resources. Don't specify it.

Type: Integer

Required: No

placementGroup

The Amazon EC2 placement group to associate with your compute resources. If you intend to submit multi-node parallel jobs to your compute environment, you should consider creating a

cluster placement group and associate it with your compute resources. This keeps your multi-node parallel job on a logical grouping of instances within a single Availability Zone with high network flow potential. For more information, see Placement groups in the *Amazon EC2 User Guide for Linux Instances*.

> **Note**
>
> This parameter isn't applicable to jobs that are running on Fargate resources. Don't specify it.

Type: String

Required: No

scalingPolicy

The scaling policy configuration for the compute environment.

> **Note**
>
> This parameter isn't applicable to jobs that are running on Fargate resources. Don't specify it.

Type: ComputeScalingPolicy object

Required: No

securityGroupIds

The Amazon EC2 security groups that are associated with instances launched in the compute environment. One or more security groups must be specified, either in `securityGroupIds` or using a launch template referenced in `launchTemplate`. This parameter is required for jobs that are running on Fargate resources and must contain at least one security group. Fargate doesn't support launch templates. If security groups are specified using both `securityGroupIds` and `launchTemplate`, the values in `securityGroupIds` are used.

Type: Array of strings

Required: No

spotIamFleetRole

The Amazon Resource Name (ARN) of the Amazon EC2 Spot Fleet IAM role applied to a SPOT compute environment. This role is required if the allocation strategy set to BEST_FIT or if the allocation strategy isn't specified. For more information, see Amazon EC2 spot fleet role in the *AWS Batch User Guide*.

> **Note**
>
> This parameter isn't applicable to jobs that are running on Fargate resources. Don't specify it.

> **Important**
>
> To tag your Spot Instances on creation, the Spot Fleet IAM role specified here must use the newer **AmazonEC2SpotFleetTaggingRole** managed policy. The previously recommended **AmazonEC2SpotFleetRole** managed policy doesn't have the required permissions to tag Spot Instances. For more information, see Spot instances not tagged on creation in the *AWS Batch User Guide*.

Type: String

Required: No

subnets

The VPC subnets where the compute resources are launched. These subnets must be within the same VPC. Fargate compute resources can contain up to 16 subnets. For more information, see VPCs and subnets in the *Amazon VPC User Guide*. This parameter is required for compute environments using EC2, SPOT, FARGATE, or FARGATE_SPOT compute resources.

> **Note**
>
> AWS Batch on Amazon EC2 and AWS Batch on Amazon EKS support Local Zones. For more information, see Local Zones in the *Amazon EC2 User Guide for Linux Instances*, Amazon EKS and AWS Local Zones in the *Amazon EKS User Guide* and Amazon ECS clusters in Local Zones, Wavelength Zones, and AWS Outposts in the *Amazon ECS Developer Guide*.

AWS Batch on Fargate doesn't currently support Local Zones.

Type: Array of strings

Required: No

tags

Key-value pair tags to be applied to Amazon EC2 resources that are launched in the compute environment. For AWS Batch, these take the form of `"String1": "String2"`, where `String1` is the tag key and `String2` is the tag value (for example, `{ "Name": "Batch Instance - C4OnDemand" }`). This is helpful for recognizing your AWS Batch instances in the Amazon EC2 console. Updating these tags requires an infrastructure update to the compute environment. For more information, see Updating compute environments in the *AWS Batch User Guide*. These tags aren't seen when using the AWS Batch `ListTagsForResource` API operation.

Note

This parameter isn't applicable to jobs that are running on Fargate resources. Don't specify it.

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 Java V2
- AWS SDK for Ruby V3

ComputeResourceUpdate

An object that represents the attributes of a compute environment that can be updated. For more information, see Updating compute environments in the *AWS Batch User Guide*.

Contents

allocationStrategy

The allocation strategy to use for the compute resource if there's not enough instances of the best fitting instance type that can be allocated. This might be because of availability of the instance type in the Region or Amazon EC2 service limits. For more information, see Allocation strategies in the *AWS Batch User Guide*.

When updating a compute environment, changing the allocation strategy requires an infrastructure update of the compute environment. For more information, see Updating compute environments in the *AWS Batch User Guide*. `BEST_FIT` isn't supported when updating a compute environment.

> **Note**
>
> This parameter isn't applicable to jobs that are running on Fargate resources. Don't specify it.

BEST_FIT_PROGRESSIVE

AWS Batch selects additional instance types that are large enough to meet the requirements of the jobs in the queue. Its preference is for instance types with lower cost vCPUs. If additional instances of the previously selected instance types aren't available, AWS Batch selects new instance types.

BEST_FIT_PROGRESSIVE_ORDERED

> **Important**
>
> This is an advanced allocation strategy only for customers who want to control which instance types are preferred during scaling.

Placing large instance types at the top of the list may result in **over-provisioning** for small jobs. Placing small instance types at the top may cause the compute environment to reach Amazon EC2 instance count limits before reaching `maxvCpus`.

AWS Batch selects instance types in the order they appear in the `instanceTypes` list. When an instance family is specified, sizes within that family are expanded using `BEST_FIT_PROGRESSIVE` logic—preferring sizes that best fit the jobs, with larger sizes as fallback. Instance types that cannot meet the resource requirements of the jobs are skipped. This strategy is only available for On-Demand Instance (`EC2`) compute resources.

If an instance family and an explicit instance type from that family both appear in `instanceTypes`, the explicit type takes its listed position and is excluded from the family expansion. For example, in `["m7a.4xlarge", "m7a", "m6a"]`, `m7a.4xlarge` is always placed first and is excluded from the `m7a` family expansion.

SPOT_CAPACITY_OPTIMIZED

AWS Batch selects one or more instance types that are large enough to meet the requirements of the jobs in the queue. Its preference is for instance types that are less likely to be interrupted. This allocation strategy is only available for Spot Instance compute resources.

SPOT_PRICE_CAPACITY_OPTIMIZED

The price and capacity optimized allocation strategy looks at both price and capacity to select the Spot Instance pools that are the least likely to be interrupted and have the lowest possible price. This allocation strategy is only available for Spot Instance compute resources.

SPOT_CAPACITY_OPTIMIZED_PRIORITIZED

Important

This is an advanced allocation strategy for customers who want to influence instance type selection during scaling. This strategy optimizes for **capacity first**, and honors instance type priorities on a best-effort basis (priorities are honored when they do not significantly reduce available Spot capacity).
Placing large instance types at the top of the list may result in **over-provisioning** for small jobs. Placing small instance types at the top may cause the compute environment to reach Amazon EC2 instance count limits before reaching `maxvCpus`.

AWS Batch selects instance types in the order they appear in the `instanceTypes` list, but **optimizes for capacity first**. The customer-defined priority is honored on a best-effort basis. When Spot Instance capacity pools are similarly available, priority order is respected. When capacity is constrained, AWS Batch selects from the most available pools regardless of priority to minimize the likelihood of Spot Instance interruptions. This strategy is only available for Spot Instance compute resources.

With any allocation strategy except `BEST_FIT` using On-Demand (`EC2`) compute resources, AWS Batch might need to exceed `maxvCpus` to meet your capacity requirements. In this event, AWS Batch never exceeds `maxvCpus` by more than a single instance.

Type: String

Valid Values: `BEST_FIT_PROGRESSIVE | BEST_FIT_PROGRESSIVE_ORDERED | SPOT_CAPACITY_OPTIMIZED | SPOT_PRICE_CAPACITY_OPTIMIZED | SPOT_CAPACITY_OPTIMIZED_PRIORITIZED`

Required: No

bidPercentage

The maximum percentage that a Spot Instance price can be when compared with the On-Demand price for that instance type before instances are launched. For example, if your maximum percentage is 20%, the Spot price must be less than 20% of the current On-Demand price for that Amazon EC2 instance. You always pay the lowest (market) price and never more than your maximum percentage. For most use cases, we recommend leaving this field empty.

When updating a compute environment, changing the bid percentage requires an infrastructure update of the compute environment. For more information, see Updating compute environments in the *AWS Batch User Guide*.

> **Note**
>
> This parameter isn't applicable to jobs that are running on Fargate resources. Don't specify it.

Type: Integer

Required: No

capacityTags

The updated tags to apply to the Amazon ECS capacity provider and Amazon EC2 instances. This parameter is only valid for `ECS_MANAGED_INSTANCES` compute environments. You must have the `batch:SetCapacityTags` permission on the compute environment resource to use this parameter.

Type: String to string map

Map Entries: Maximum number of 50 items.

Key Length Constraints: Minimum length of 1. Maximum length of 128.

Value Length Constraints: Maximum length of 256.

Required: No

desiredvCpus

The desired number of vCPUS in the compute environment. AWS Batch modifies this value between the minimum and maximum values based on job queue demand.

> **Note**
>
> This parameter isn't applicable to jobs that are running on Fargate resources. Don't specify it.

> **Note**
>
> AWS Batch doesn't support changing the desired number of vCPUs of an existing compute environment. Don't specify this parameter for compute environments using Amazon EKS clusters.

> **Note**
>
> When you update the `desiredvCpus` setting, the value must be between the `minvCpus` and `maxvCpus` values.

Additionally, the updated `desiredvCpus` value must be greater than or equal to the current `desiredvCpus` value. For more information, see Troubleshooting AWS Batch in the *AWS Batch User Guide*.

Type: Integer

Required: No

ec2Configuration

Provides information used to select Amazon Machine Images (AMIs) for Amazon EC2 instances in the compute environment. If `Ec2Configuration` isn't specified, the default is `ECS_AL2023` for EC2 (ECS) compute environments and `EKS_AL2023` for EKS compute environments.

When updating a compute environment, changing this setting requires an infrastructure update of the compute environment. For more information, see Updating compute environments in the *AWS Batch User Guide*. To remove the Amazon EC2 configuration and any custom AMI ID specified in `imageIdOverride`, set this value to an empty string.

One or two values can be provided.

Note

This parameter isn't applicable to jobs that are running on Fargate resources. Don't specify it.

Type: Array of Ec2Configuration objects

Required: No

ec2KeyPair

The Amazon EC2 key pair that's used for instances launched in the compute environment. You can use this key pair to log in to your instances with SSH. To remove the Amazon EC2 key pair, set this value to an empty string.

When updating a compute environment, changing the Amazon EC2 key pair requires an infrastructure update of the compute environment. For more information, see Updating compute environments in the *AWS Batch User Guide*.

> **Note**
>
> This parameter isn't applicable to jobs that are running on Fargate resources. Don't specify it.

Type: String

Required: No

imageId

The Amazon Machine Image (AMI) ID used for instances launched in the compute environment. This parameter is overridden by the `imageIdOverride` member of the `Ec2Configuration` structure. To remove the custom AMI ID and use the default AMI ID, set this value to an empty string.

When updating a compute environment, changing the AMI ID requires an infrastructure update of the compute environment. For more information, see Updating compute environments in the *AWS Batch User Guide*.

> **Note**
>
> This parameter isn't applicable to jobs that are running on Fargate resources. Don't specify it.

> **Note**
>
> The AMI that you choose for a compute environment must match the architecture of the instance types that you intend to use for that compute environment. For example, if your compute environment uses A1 instance types, the compute resource AMI that you choose must support ARM instances. Amazon ECS vends both x86 and ARM versions of the Amazon ECS-optimized Amazon Linux 2023 AMI. For more information, see Amazon ECS-optimized Amazon Linux 2023 AMI in the *Amazon Elastic Container Service Developer Guide*.

Type: String

Required: No

instanceRole

The Amazon ECS instance profile applied to Amazon EC2 instances in a compute environment. Required for Amazon EC2 instances. You can specify the short name or full Amazon Resource Name (ARN) of an instance profile. For example, *ecsInstanceRole* or `arn:aws:iam::<aws_account_id>:instance-profile/ecsInstanceRole` . For more information, see Amazon ECS instance role in the *AWS Batch User Guide*.

When updating a compute environment, changing this setting requires an infrastructure update of the compute environment. For more information, see Updating compute environments in the *AWS Batch User Guide*.

> **Note**
>
> This parameter isn't applicable to jobs that are running on Fargate resources. Don't specify it.

Type: String

Required: No

instanceTypes

The instances types that can be launched. You can specify instance families to launch any instance type within those families (for example, `c5` or `p3`), or you can specify specific sizes within a family (such as `c5.8xlarge`).

AWS Batch can select the instance type for you if you choose one of the following:

- `optimal` to select instance types (from the `c4`, `m4`, `r4`, `c5`, `m5`, and `r5` instance families) that match the demand of your job queues.
- `default_x86_64` to choose x86 based instance types (from the `m6i`, `c6i`, `r6i`, and `c7i` instance families) that matches the resource demands of the job queue.
- `default_arm64` to choose x86 based instance types (from the `m6g`, `c6g`, `r6g`, and `c7g` instance families) that matches the resource demands of the job queue.

Note

Starting on 11/01/2025 the behavior of `optimal` is going to be changed to match `default_x86_64`. During the change your instance families could be updated to a newer generation. You do not need to perform any actions for the upgrade to happen. For more information about change, see Optimal instance type configuration to receive automatic instance family updates.

Note

Instance family availability varies by AWS Region. For example, some AWS Regions may not have any fourth generation instance families but have fifth and sixth generation instance families.
When using `default_x86_64` or `default_arm64` instance bundles, AWS Batch selects instance families based on a balance of cost-effectiveness and performance. While newer generation instances often provide better price-performance, AWS Batch may choose an earlier generation instance family if it provides the optimal combination of availability, cost, and performance for your workload. For example, in an AWS Region where both c6i and c7i instances are available, AWS Batch might select c6i instances if they offer better cost-effectiveness for your specific job requirements. For more information on AWS Batch instance types and AWS Region availability, see Instance type compute table in the *AWS Batch User Guide*.
AWS Batch periodically updates your instances in default bundles to newer, more cost-effective options. Updates happen automatically without requiring any action from you. Your workloads continue running during updates with no interruption

Note

This parameter isn't applicable to jobs that are running on Fargate resources. Don't specify it.

> **Note**
>
> When you create a compute environment, the instance types that you select for the compute environment must share the same architecture. For example, you can't mix x86 and ARM instances in the same compute environment.

Type: Array of strings

Required: No

launchTemplate

The updated launch template to use for your compute resources. You must specify either the launch template ID or launch template name in the request, but not both. For more information, see Launch template support in the *AWS Batch User Guide*. To remove the custom launch template and use the default launch template, set `launchTemplateId` or `launchTemplateName` member of the launch template specification to an empty string. Removing the launch template from a compute environment will not remove the AMI specified in the launch template. In order to update the AMI specified in a launch template, the `updateToLatestImageVersion` parameter must be set to `true`.

When updating a compute environment, changing the launch template requires an infrastructure update of the compute environment. For more information, see Updating compute environments in the *AWS Batch User Guide*.

> **Note**
>
> This parameter isn't applicable to jobs that are running on Fargate resources. Don't specify it.

Type: LaunchTemplateSpecification object

Required: No

managedInstancesProvider

The updated configuration for the Amazon ECS Managed Instances capacity provider. This parameter is only valid when the compute environment type is `ECS_MANAGED_INSTANCES`. You cannot change `capacityOptionType` or `fipsEnabled` on update.

Type: UpdateManagedInstancesProviderConfiguration object

Required: No

maxvCpus

The maximum number of Amazon EC2 vCPUs that an environment can reach.

> **Note**
>
> With any allocation strategy except `BEST_FIT` using On-Demand (`EC2`) compute resources, AWS Batch might need to exceed `maxvCpus` to meet your capacity requirements. In this event, AWS Batch never exceeds `maxvCpus` by more than a single instance.

Type: Integer

Required: No

minvCpus

The minimum number of vCPUs that an environment should maintain (even if the compute environment is `DISABLED`).

> **Note**
>
> This parameter isn't applicable to jobs that are running on Fargate resources. Don't specify it.

Type: Integer

Required: No

placementGroup

The Amazon EC2 placement group to associate with your compute resources. If you intend to submit multi-node parallel jobs to your compute environment, you should consider creating a cluster placement group and associate it with your compute resources. This keeps your multi-node parallel job on a logical grouping of instances within a single Availability Zone with high network flow potential. For more information, see Placement groups in the *Amazon EC2 User Guide for Linux Instances*.

When updating a compute environment, changing the placement group requires an infrastructure update of the compute environment. For more information, see Updating compute environments in the *AWS Batch User Guide*.

> **Note**
>
> This parameter isn't applicable to jobs that are running on Fargate resources. Don't specify it.

Type: String

Required: No

scalingPolicy

The scaling policy configuration for the compute environment.

> **Note**
>
> This parameter isn't applicable to jobs that are running on Fargate resources. Don't specify it.

Type: ComputeScalingPolicy object

Required: No

securityGroupIds

The Amazon EC2 security groups that are associated with instances launched in the compute environment. This parameter is required for Fargate compute resources, where it can contain up to 5 security groups. For Fargate compute resources, providing an empty list is handled as if this parameter wasn't specified and no change is made. For Amazon EC2 compute resources, providing an empty list removes the security groups from the compute resource.

When updating a compute environment, changing the Amazon EC2 security groups requires an infrastructure update of the compute environment. For more information, see Updating compute environments in the *AWS Batch User Guide*.

Type: Array of strings

Required: No

subnets

The VPC subnets where the compute resources are launched. Fargate compute resources can contain up to 16 subnets. For Fargate compute resources, providing an empty list will be handled as if this parameter wasn't specified and no change is made. For Amazon EC2 compute resources, providing an empty list removes the VPC subnets from the compute resource. For more information, see VPCs and subnets in the *Amazon VPC User Guide*.

When updating a compute environment, changing the VPC subnets requires an infrastructure update of the compute environment. For more information, see Updating compute environments in the *AWS Batch User Guide*.

> **Note**
>
> AWS Batch on Amazon EC2 and AWS Batch on Amazon EKS support Local Zones. For more information, see Local Zones in the *Amazon EC2 User Guide for Linux Instances*, Amazon EKS and AWS Local Zones in the *Amazon EKS User Guide* and Amazon ECS clusters in Local Zones, Wavelength Zones, and AWS Outposts in the *Amazon ECS Developer Guide*.
> AWS Batch on Fargate doesn't currently support Local Zones.

Type: Array of strings

Required: No

tags

Key-value pair tags to be applied to Amazon EC2 resources that are launched in the compute environment. For AWS Batch, these take the form of `"String1": "String2"`, where `String1` is the tag key and `String2` is the tag value (for example, `{ "Name": "Batch Instance - C4OnDemand" }`). This is helpful for recognizing your Batch instances in the Amazon EC2 console. These tags aren't seen when using the AWS Batch `ListTagsForResource` API operation.

When updating a compute environment, changing this setting requires an infrastructure update of the compute environment. For more information, see Updating compute environments in the *AWS Batch User Guide*.

> **Note**
>
> This parameter isn't applicable to jobs that are running on Fargate resources. Don't specify it.

Type: String to string map

Required: No

type

The type of compute environment: `EC2`, `SPOT`, `FARGATE`, `FARGATE_SPOT`, or `ECS_MANAGED_INSTANCES`. For more information, see Compute environments in the *AWS Batch User Guide*.

If you choose `SPOT`, you must also specify an Amazon EC2 Spot Fleet role with the `spotIamFleetRole` parameter. For more information, see Amazon EC2 spot fleet role in the *AWS Batch User Guide*.

When updating a compute environment, changing the type of a compute environment requires an infrastructure update of the compute environment. For more information, see Updating compute environments in the *AWS Batch User Guide*.

You cannot change the type to or from `ECS_MANAGED_INSTANCES`.

Type: String

Valid Values: `EC2` | `SPOT` | `FARGATE` | `FARGATE_SPOT` | `ECS_MANAGED_INSTANCES`

Required: No

updateToLatestImageVersion

Specifies whether the AMI ID is updated to the latest one that's supported by AWS Batch when the compute environment has an infrastructure update. The default value is `false`.

> **Note**
>
> An AMI ID can either be specified in the `imageId` or `imageIdOverride` parameters or be determined by the launch template that's specified in the `launchTemplate` parameter. If an AMI ID is specified any of these ways, this parameter is ignored.

For more information about to update AMI IDs during an infrastructure update, see Updating the AMI ID in the *AWS Batch User Guide*.

When updating a compute environment, changing this setting requires an infrastructure update of the compute environment. For more information, see Updating compute environments in the *AWS Batch User Guide*.

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 Java V2
- AWS SDK for Ruby V3

ComputeScalingPolicy

An object that represents a scaling policy for a compute environment.

Contents

minScaleDownDelayMinutes

The minimum time (in minutes) that AWS Batch keeps instances running in the compute environment after their jobs complete. For each instance, the delay period begins when the last job finishes. If no new jobs are placed on the instance during this delay, AWS Batch terminates the instance once the delay expires.

Valid Range: Minimum value of 20. Maximum value of 10080. Use 0 to unset and disable the scale down delay.

> **Note**
>
> Idle instances retained during the scale-down delay period are billable at standard EC2 pricing.

> **Note**
>
> The scale down delay does not apply to:
>
> - Instances being replaced during infrastructure updates
> - Newly launched instances that have not yet run any jobs
> - Spot instances reclaimed due to interruption

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 Java V2
- AWS SDK for Ruby V3

ConsumableResourceProperties

Contains a list of consumable resources required by a job.

Contents

consumableResourceList

The list of consumable resources required by a job.

Type: Array of ConsumableResourceRequirement 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 Java V2
- AWS SDK for Ruby V3

ConsumableResourceRequirement

Information about a consumable resource required to run a job.

Contents

consumableResource

The name or ARN of the consumable resource.

Type: String

Required: No

quantity

The quantity of the consumable resource that is needed.

Type: Long

Required: No

See Also

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

- AWS SDK for C++
- AWS SDK for Java V2
- AWS SDK for Ruby V3

ConsumableResourceSummary

Current information about a consumable resource.

Contents

consumableResourceArn

The Amazon Resource Name (ARN) of the consumable resource.

Type: String

Required: Yes

consumableResourceName

The name of the consumable resource.

Type: String

Required: Yes

inUseQuantity

The amount of the consumable resource that is currently in use.

Type: Long

Required: No

resourceType

Indicates whether the resource is available to be re-used after a job completes. Can be one of:

- `REPLENISHABLE`
- `NON_REPLENISHABLE`

Type: String

Required: No

totalQuantity

The total amount of the consumable resource that is available.

Type: Long

Required: No

See Also

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

- AWS SDK for C++
- AWS SDK for Java V2
- AWS SDK for Ruby V3

ContainerDetail

An object that represents the details of a container that's part of a job.

Contents

command

The command that's passed to the container.

Type: Array of strings

Required: No

containerInstanceArn

The Amazon Resource Name (ARN) of the container instance that the container is running on.

Type: String

Required: No

enableExecuteCommand

Determines whether execute command functionality is turned on for this task. If `true`, execute command functionality is turned on all the containers in the task.

Type: Boolean

Required: No

environment

The environment variables to pass to a container.

> **Note**
>
> Environment variables cannot start with "`AWS_BATCH`". This naming convention is reserved for variables that AWS Batch sets.

Type: Array of KeyValuePair objects

Required: No

ephemeralStorage

The amount of ephemeral storage allocated for the task. This parameter is used to expand the total amount of ephemeral storage available, beyond the default amount, for tasks hosted on AWS Fargate.

Type: EphemeralStorage object

Required: No

executionRoleArn

The Amazon Resource Name (ARN) of the execution role that AWS Batch can assume. For more information, see Batch execution IAM role in the *AWS Batch User Guide*.

Type: String

Required: No

exitCode

The exit code returned upon completion.

Type: Integer

Required: No

fargatePlatformConfiguration

The platform configuration for jobs that are running on Fargate resources. Jobs that are running on Amazon EC2 resources must not specify this parameter.

Type: FargatePlatformConfiguration object

Required: No

image

The image used to start the container.

Type: String

Required: No

instanceType

The instance type of the underlying host infrastructure of a multi-node parallel job.

> **Note**
>
> This parameter isn't applicable to jobs that are running on Fargate resources.

Type: String

Required: No

jobRoleArn

The Amazon Resource Name (ARN) that's associated with the job when run.

Type: String

Required: No

linuxParameters

Linux-specific modifications that are applied to the container, such as details for device mappings.

Type: LinuxParameters object

Required: No

logConfiguration

The log configuration specification for the container.

This parameter maps to `LogConfig` in the Create a container section of the Docker Remote API and the `--log-driver` option to docker run. By default, containers use the same logging driver that the Docker daemon uses. However, the container might use a different logging driver than the Docker daemon by specifying a log driver with this parameter in the container definition. To use a different logging driver for a container, the log system must be configured properly on the container instance. Or, alternatively, it must be configured on a different log server for remote logging options. For more information on the options for different supported log drivers, see Configure logging drivers in the Docker documentation.

> **Note**
>
> AWS Batch currently supports a subset of the logging drivers available to the Docker daemon (shown in the LogConfiguration data type). Additional log drivers might be available in future releases of the Amazon ECS container agent.

This parameter requires version 1.18 of the Docker Remote API or greater on your container instance. To check the Docker Remote API version on your container instance, log in to your container instance and run the following command: `sudo docker version | grep "Server API version"`

> **Note**
>
> The Amazon ECS container agent running on a container instance must register the logging drivers available on that instance with the `ECS_AVAILABLE_LOGGING_DRIVERS` environment variable before containers placed on that instance can use these log configuration options. For more information, see Amazon ECS container agent configuration in the *Amazon Elastic Container Service Developer Guide*.

Type: LogConfiguration object

Required: No

logStreamName

The name of the Amazon CloudWatch Logs log stream that's associated with the container. The log group for AWS Batch jobs is `/aws/batch/job`. Each container attempt receives a log stream name when they reach the `RUNNING` status.

Type: String

Required: No

memory

For jobs running on Amazon EC2 resources that didn't specify memory requirements using `resourceRequirements`, the number of MiB of memory reserved for the job. For other jobs, including all run on Fargate resources, see `resourceRequirements`.

Type: Integer

Required: No

mountPoints

The mount points for data volumes in your container.

Type: Array of MountPoint objects

Required: No

networkConfiguration

The network configuration for jobs that are running on Fargate resources. Jobs that are running on Amazon EC2 resources must not specify this parameter.

Type: NetworkConfiguration object

Required: No

networkInterfaces

The network interfaces that are associated with the job.

Type: Array of NetworkInterface objects

Required: No

privileged

When this parameter is true, the container is given elevated permissions on the host container instance (similar to the `root` user). The default value is `false`.

> **Note**
>
> This parameter isn't applicable to jobs that are running on Fargate resources and shouldn't be provided, or specified as `false`.

Type: Boolean

Required: No

readonlyRootFilesystem

When this parameter is true, the container is given read-only access to its root file system. This parameter maps to `ReadonlyRootfs` in the Create a container section of the Docker Remote API and the `--read-only` option to `docker run`.

Type: Boolean

Required: No

reason

A short (255 max characters) human-readable string to provide additional details for a running or stopped container.

Type: String

Required: No

repositoryCredentials

The private repository authentication credentials to use.

Type: RepositoryCredentials object

Required: No

resourceRequirements

The type and amount of resources to assign to a container. The supported resources include `GPU`, `MEMORY`, and `VCPU`.

Type: Array of ResourceRequirement objects

Required: No

runtimePlatform

An object that represents the compute environment architecture for AWS Batch jobs on Fargate.

Type: RuntimePlatform object

Required: No

secrets

The secrets to pass to the container. For more information, see Specifying sensitive data in the *AWS Batch User Guide*.

Type: Array of Secret objects

Required: No

taskArn

The Amazon Resource Name (ARN) of the Amazon ECS task that's associated with the container job. Each container attempt receives a task ARN when they reach the `STARTING` status.

Type: String

Required: No

ulimits

A list of `ulimit` values to set in the container. This parameter maps to `Ulimits` in the Create a container section of the Docker Remote API and the `--ulimit` option to docker run.

> **Note**
>
> This parameter isn't applicable to jobs that are running on Fargate resources.

Type: Array of Ulimit objects

Required: No

user

The user name to use inside the container. This parameter maps to `User` in the Create a container section of the Docker Remote API and the `--user` option to docker run.

Type: String

Required: No

vcpus

The number of vCPUs reserved for the container. For jobs that run on Amazon EC2 resources, you can specify the vCPU requirement for the job using `resourceRequirements`, but you can't specify the vCPU requirements in both the `vcpus` and `resourceRequirements` object. This parameter maps to `CpuShares` in the Create a container section of the Docker Remote API and the `--cpu-shares` option to docker run. Each vCPU is equivalent to 1,024 CPU shares. You

must specify at least one vCPU. This is required but can be specified in several places. It must be specified for each node at least once.

> **Note**
>
> This parameter isn't applicable to jobs that run on Fargate resources. For jobs that run on Fargate resources, you must specify the vCPU requirement for the job using `resourceRequirements`.

Type: Integer

Required: No

volumes

A list of volumes that are associated with the job.

Type: Array of Volume 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 Java V2
- AWS SDK for Ruby V3

ContainerOverrides

The overrides that should be sent to a container.

For information about using AWS Batch overrides when you connect event sources to targets, see BatchContainerOverrides.

Contents

command

The command to send to the container that overrides the default command from the Docker image or the job definition.

> **Note**
>
> This parameter can't contain an empty string.

Type: Array of strings

Required: No

environment

The environment variables to send to the container. You can add new environment variables, which are added to the container at launch, or you can override the existing environment variables from the Docker image or the job definition.

> **Note**
>
> Environment variables cannot start with "AWS_BATCH". This naming convention is reserved for variables that AWS Batch sets.

Type: Array of KeyValuePair objects

Required: No

instanceType

The instance type to use for a multi-node parallel job.

> **Note**
>
> This parameter isn't applicable to single-node container jobs or jobs that run on Fargate resources, and shouldn't be provided.

Type: String

Required: No

memory

This parameter is deprecated, use `resourceRequirements` to override the memory requirements specified in the job definition. It's not supported for jobs running on Fargate resources. For jobs that run on Amazon EC2 resources, it overrides the `memory` parameter set in the job definition, but doesn't override any memory requirement that's specified in the `resourceRequirements` structure in the job definition. To override memory requirements that are specified in the `resourceRequirements` structure in the job definition, `resourceRequirements` must be specified in the `SubmitJob` request, with `type` set to `MEMORY` and `value` set to the new value. For more information, see Can't override job definition resource requirements in the *AWS Batch User Guide*.

Type: Integer

Required: No

resourceRequirements

The type and amount of resources to assign to a container. This overrides the settings in the job definition. The supported resources include `GPU`, `MEMORY`, and `VCPU`.

Type: Array of ResourceRequirement objects

Required: No

vcpus

This parameter is deprecated, use `resourceRequirements` to override the `vcpus` parameter that's set in the job definition. It's not supported for jobs running on Fargate resources. For jobs that run on Amazon EC2 resources, it overrides the `vcpus` parameter set in the job definition, but doesn't override any vCPU requirement specified in the `resourceRequirements` structure in the job definition. To override vCPU requirements that are specified in the

`resourceRequirements` structure in the job definition, `resourceRequirements` must be specified in the `SubmitJob` request, with `type` set to `VCPU` and `value` set to the new value. For more information, see Can't override job definition resource requirements in the *AWS Batch 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 Java V2
- AWS SDK for Ruby V3

ContainerProperties

Container properties are used for Amazon ECS based job definitions. These properties to describe the container that's launched as part of a job.

Contents

command

The command that's passed to the container. This parameter maps to `Cmd` in the Create a container section of the Docker Remote API and the `COMMAND` parameter to docker run. For more information, see https://docs.docker.com/engine/reference/builder/#cmd.

Type: Array of strings

Required: No

enableExecuteCommand

Determines whether execute command functionality is turned on for this task. If `true`, execute command functionality is turned on all the containers in the task.

Type: Boolean

Required: No

environment

The environment variables to pass to a container. This parameter maps to `Env` in the Create a container section of the Docker Remote API and the `--env` option to docker run.

> **Important**
>
> We don't recommend using plaintext environment variables for sensitive information, such as credential data.

> **Note**
>
> Environment variables cannot start with "`AWS_BATCH`". This naming convention is reserved for variables that AWS Batch sets.

Type: Array of KeyValuePair objects

Required: No

ephemeralStorage

The amount of ephemeral storage to allocate for the task. This parameter is used to expand the total amount of ephemeral storage available, beyond the default amount, for tasks hosted on AWS Fargate.

Type: EphemeralStorage object

Required: No

executionRoleArn

The Amazon Resource Name (ARN) of the execution role that AWS Batch can assume. For jobs that run on Fargate resources, you must provide an execution role. For more information, see AWS Batch execution IAM role in the *AWS Batch User Guide*.

Type: String

Required: No

fargatePlatformConfiguration

The platform configuration for jobs that are running on Fargate resources. Jobs that are running on Amazon EC2 resources must not specify this parameter.

Type: FargatePlatformConfiguration object

Required: No

image

Required. The image used to start a container. This string is passed directly to the Docker daemon. Images in the Docker Hub registry are available by default. Other repositories are specified with *`repository-url/image:tag`* . It can be 255 characters long. It can contain uppercase and lowercase letters, numbers, hyphens (-), underscores (_), colons (:), periods (.), forward slashes (/), and number signs (#). This parameter maps to `Image` in the Create a container section of the Docker Remote API and the `IMAGE` parameter of docker run.

> **Note**
>
> Docker image architecture must match the processor architecture of the compute resources that they're scheduled on. For example, ARM-based Docker images can only run on ARM-based compute resources.

- Images in Amazon ECR Public repositories use the full `registry/repository[:tag]` or `registry/repository[@digest]` naming conventions. For example, `public.ecr.aws/`*`registry_alias/my-web-app:latest`* `.`
- Images in Amazon ECR repositories use the full registry and repository URI (for example, `123456789012.dkr.ecr.<region-name>.amazonaws.com/<repository-name>`).
- Images in official repositories on Docker Hub use a single name (for example, `ubuntu` or `mongo`).
- Images in other repositories on Docker Hub are qualified with an organization name (for example, `amazon/amazon-ecs-agent`).
- Images in other online repositories are qualified further by a domain name (for example, `quay.io/assemblyline/ubuntu`).

Type: String

Required: No

instanceType

The instance type to use for a multi-node parallel job. All node groups in a multi-node parallel job must use the same instance type.

> **Note**
>
> This parameter isn't applicable to single-node container jobs or jobs that run on Fargate resources, and shouldn't be provided.

Type: String

Required: No

jobRoleArn

The Amazon Resource Name (ARN) of the IAM role that the container can assume for AWS permissions. For more information, see IAM roles for tasks in the *Amazon Elastic Container Service Developer Guide*.

Type: String

Required: No

linuxParameters

Linux-specific modifications that are applied to the container, such as details for device mappings.

Type: LinuxParameters object

Required: No

logConfiguration

The log configuration specification for the container.

This parameter maps to `LogConfig` in the Create a container section of the Docker Remote API and the `--log-driver` option to docker run. By default, containers use the same logging driver that the Docker daemon uses. However the container might use a different logging driver than the Docker daemon by specifying a log driver with this parameter in the container definition. To use a different logging driver for a container, the log system must be configured properly on the container instance (or on a different log server for remote logging options). For more information on the options for different supported log drivers, see Configure logging drivers in the Docker documentation.

> **Note**
>
> AWS Batch currently supports a subset of the logging drivers available to the Docker daemon (shown in the LogConfiguration data type).

This parameter requires version 1.18 of the Docker Remote API or greater on your container instance. To check the Docker Remote API version on your container instance, log in to your container instance and run the following command: `sudo docker version | grep "Server API version"`

> **Note**
>
> The Amazon ECS container agent running on a container instance must register the logging drivers available on that instance with the `ECS_AVAILABLE_LOGGING_DRIVERS` environment variable before containers placed on that instance can use these log configuration options. For more information, see Amazon ECS container agent configuration in the *Amazon Elastic Container Service Developer Guide*.

Type: LogConfiguration object

Required: No

memory

This parameter is deprecated, use `resourceRequirements` to specify the memory requirements for the job definition. It's not supported for jobs running on Fargate resources. For jobs that run on Amazon EC2 resources, it specifies the memory hard limit (in MiB) for a container. If your container attempts to exceed the specified number, it's terminated. You must specify at least 4 MiB of memory for a job using this parameter. The memory hard limit can be specified in several places. It must be specified for each node at least once.

Type: Integer

Required: No

mountPoints

The mount points for data volumes in your container. This parameter maps to `Volumes` in the Create a container section of the Docker Remote API and the `--volume` option to docker run.

Type: Array of MountPoint objects

Required: No

networkConfiguration

The network configuration for jobs that are running on Fargate resources. Jobs that are running on Amazon EC2 resources must not specify this parameter.

Type: NetworkConfiguration object

Required: No

privileged

When this parameter is true, the container is given elevated permissions on the host container instance (similar to the `root` user). This parameter maps to `Privileged` in the Create a container section of the Docker Remote API and the `--privileged` option to docker run. The default value is false.

> **Note**
>
> This parameter isn't applicable to jobs that are running on Fargate resources and shouldn't be provided, or specified as false.

Type: Boolean

Required: No

readonlyRootFilesystem

When this parameter is true, the container is given read-only access to its root file system. This parameter maps to `ReadonlyRootfs` in the Create a container section of the Docker Remote API and the `--read-only` option to `docker run`.

Type: Boolean

Required: No

repositoryCredentials

The private repository authentication credentials to use.

Type: RepositoryCredentials object

Required: No

resourceRequirements

The type and amount of resources to assign to a container. The supported resources include `GPU`, `MEMORY`, and `VCPU`.

Type: Array of ResourceRequirement objects

Required: No

runtimePlatform

An object that represents the compute environment architecture for AWS Batch jobs on Fargate.

Type: RuntimePlatform object

Required: No

secrets

The secrets for the container. For more information, see Specifying sensitive data in the *AWS Batch User Guide*.

Type: Array of Secret objects

Required: No

ulimits

A list of `ulimits` to set in the container. This parameter maps to `Ulimits` in the Create a container section of the Docker Remote API and the `--ulimit` option to docker run.

> **Note**
>
> This parameter isn't applicable to jobs that are running on Fargate resources and shouldn't be provided.

Type: Array of Ulimit objects

Required: No

user

The user name to use inside the container. This parameter maps to `User` in the Create a container section of the Docker Remote API and the `--user` option to docker run.

Type: String

Required: No

vcpus

This parameter is deprecated, use `resourceRequirements` to specify the vCPU requirements for the job definition. It's not supported for jobs running on Fargate resources. For jobs running on Amazon EC2 resources, it specifies the number of vCPUs reserved for the job.

Each vCPU is equivalent to 1,024 CPU shares. This parameter maps to `CpuShares` in the Create a container section of the Docker Remote API and the `--cpu-shares` option to docker run. The number of vCPUs must be specified but can be specified in several places. You must specify it at least once for each node.

Type: Integer

Required: No

volumes

A list of data volumes used in a job.

Type: Array of Volume 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 Java V2
- AWS SDK for Ruby V3

ContainerSummary

An object that represents summary details of a container within a job.

Contents

exitCode

The exit code to return upon completion.

Type: Integer

Required: No

reason

A short (255 max characters) human-readable string to provide additional details for a running or stopped container.

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 Java V2
- AWS SDK for Ruby V3

Device

An object that represents a container instance host device.

> **Note**
>
> This object isn't applicable to jobs that are running on Fargate resources and shouldn't be provided.

Contents

hostPath

The path for the device on the host container instance.

Type: String

Required: Yes

containerPath

The path inside the container that's used to expose the host device. By default, the `hostPath` value is used.

Type: String

Required: No

permissions

The explicit permissions to provide to the container for the device. By default, the container has permissions for `read`, `write`, and `mknod` for the device.

Type: Array of strings

Valid Values: `READ | WRITE | MKNOD`

Required: No

See Also

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

- AWS SDK for C++
- AWS SDK for Java V2
- AWS SDK for Ruby V3

Ec2Configuration

Provides information used to select Amazon Machine Images (AMIs) for instances in the compute environment. If `Ec2Configuration` isn't specified, the default is `ECS_AL2023` (Amazon ECS-optimized Amazon Linux 2023) for EC2 (ECS) compute environments and `EKS_AL2023` (Amazon EKS-optimized Amazon Linux 2023 AMI) for EKS compute environments.

Note

This object isn't applicable to jobs that are running on Fargate resources.

Contents

imageType

The image type to match with the instance type to select an AMI. The supported values are different for ECS and EKS resources.

ECS

If the `imageIdOverride` parameter isn't specified, then a recent Amazon ECS-optimized Amazon Linux 2023 AMI (`ECS_AL2023`) is used. If a new image type is specified in an update, but neither an `imageId` nor a `imageIdOverride` parameter is specified, then the latest Amazon ECS optimized AMI for that image type that's supported by AWS Batch is used.

Important

AWS is ending support for Amazon ECS Amazon Linux 2-optimized and accelerated AMIs on June 30, 2026. On January 12, 2026, AWS Batch changed the default AMI for new Amazon ECS compute environments from Amazon Linux 2 to Amazon Linux 2023. Effective June 30, 2026, AWS Batch will block creation of new Amazon ECS compute environments using Batch-provided Amazon Linux 2 AMIs. We strongly recommend migrating your existing AWS Batch Amazon ECS compute environments to Amazon Linux 2023 prior to June 30, 2026. For more information on upgrading from AL2 to AL2023, see How to migrate from ECS AL2 to ECS AL2023 in the *AWS Batch User Guide*.

ECS_AL2

Amazon Linux 2: Used for non-GPU instance families.

ECS_AL2_NVIDIA

Amazon Linux 2 (GPU): Used for GPU instance families (for example P4 and G4) and non AWS Graviton-based instance types.

ECS_AL2023

Amazon Linux 2023: Default for all non-GPU instance families.

> **Note**
>
> Amazon Linux 2023 does not support A1 instances.

ECS_AL2023_NVIDIA

Amazon Linux 2023 (GPU): Default for all GPU instance families and can be used for all non AWS Graviton-based instance types.

> **Note**
>
> ECS_AL2023_NVIDIA doesn't support p3 and g3 instance types.

EKS

If the `imageIdOverride` parameter isn't specified, then a recent Amazon EKS-optimized Amazon Linux 2023 AMI (EKS_AL2023) is used. If a new image type is specified in an update, but neither an `imageId` nor a `imageIdOverride` parameter is specified, then the latest Amazon EKS optimized AMI for that image type that AWS Batch supports is used.

> **Important**
>
> Amazon Linux 2023 AMIs are the default on AWS Batch for Amazon EKS.
> AWS ended support for Amazon EKS AL2-optimized and AL2-accelerated AMIs on November 26, 2025. AWS Batch Amazon EKS compute environments using Amazon Linux 2 will no longer receive software updates, security patches, or bug fixes from

AWS. We recommend migrating to Amazon Linux 2023. For more information on upgrading from AL2 to AL2023, see How to upgrade from EKS AL2 to EKS AL2023 in the *AWS Batch User Guide*.

EKS_AL2

Amazon Linux 2: Used for non-GPU instance families.

EKS_AL2_NVIDIA

Amazon Linux 2 (accelerated): Used for GPU instance families (for example, P4 and G4) and can be used for all non AWS Graviton-based instance types.

EKS_AL2023

Amazon Linux 2023: Default for non-GPU instance families.

Note

Amazon Linux 2023 does not support A1 instances.

EKS_AL2023_NVIDIA

Amazon Linux 2023 (accelerated): Default for GPU instance families and can be used for all non AWS Graviton-based instance types.

Type: String

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

Required: Yes

batchImageStatus

The status of the Batch-provided default AMIs associated with the `imageType`.

The field only appears after the compute environment has begun scaling instances using the `imageType`. The field is not present when an image is specified in `ComputeResources.imageId` (deprecated), the default launch template, or `Ec2Configuration.imageIdOverride`. The field is also not present when the compute

environment has a launch template override. For more information on image selection, see AMI selection order.

> **Note**
>
> This field is read-only and only appears in the DescribeComputeEnvironments response.

- `LATEST` – Using the most recent AMI supported
- `UPDATE_AVAILABLE` – An updated AMI is available
  - If a compute environment has multiple AMIs for the `imageType` and any one AMI has `UPDATE_AVAILABLE`, the status shows `UPDATE_AVAILABLE`.
  - For compute environments that use `BEST_FIT` as their allocation strategy, you can perform a blue/green update to update the AMI.
  - For all other compute environments, you can perform an AMI version update to update the AMI to the latest version.

Type: String

Required: No

imageIdOverride

The AMI ID used for instances launched in the compute environment that match the image type. This setting overrides the `imageId` set in the `computeResource` object.

> **Note**
>
> The AMI that you choose for a compute environment must match the architecture of the instance types that you intend to use for that compute environment. For example, if your compute environment uses A1 instance types, the compute resource AMI that you choose must support ARM instances. Amazon ECS vends both x86 and ARM versions of the Amazon ECS-optimized Amazon Linux 2023 AMI. For more information, see Amazon ECS-optimized Amazon Linux 2023 AMI in the *Amazon Elastic Container Service Developer Guide*.

Type: String

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

Required: No

imageKubernetesVersion

The Kubernetes version for the compute environment. If you don't specify a value, the latest version that AWS Batch supports is used.

Type: String

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 Java V2
- AWS SDK for Ruby V3

EcsProperties

An object that contains the properties for the Amazon ECS resources of a job.

Contents

taskProperties

An object that contains the properties for the Amazon ECS task definition of a job.

> **Note**
>
> This object is currently limited to one task element. However, the task element can run up to 10 containers.

Type: Array of EcsTaskProperties objects

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 Java V2
- AWS SDK for Ruby V3

EcsPropertiesDetail

An object that contains the details for the Amazon ECS resources of a job.

Contents

taskProperties

The properties for the Amazon ECS task definition of a job.

Type: Array of EcsTaskDetails 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 Java V2
- AWS SDK for Ruby V3

EcsPropertiesOverride

An object that contains overrides for the Amazon ECS task definition of a job.

Contents

taskProperties

The overrides for the Amazon ECS task definition of a job.

> **Note**
>
> This object is currently limited to one element.

Type: Array of TaskPropertiesOverride 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 Java V2
- AWS SDK for Ruby V3

EcsSettings

The Amazon ECS settings for a compute environment, including the CloudWatch Container Insights mode. Use this structure with `CreateComputeEnvironment` and `UpdateComputeEnvironment`.

Contents

containerInsights

Specifies the CloudWatch Container Insights mode for the compute environment. Valid values are:

ENABLED

Turns on standard Container Insights, which collects CPU, memory, disk, and network utilization metrics for the compute environment.

ENHANCED

Turns on enhanced Container Insights, which collects the standard metrics along with additional per-task observability metrics.

DISABLED

Turns off Container Insights for the compute environment.

If you don't specify a value, the default is `DISABLED`. For more information, see Container Insights in the *AWS Batch User Guide*.

Type: String

Valid Values: `ENABLED | ENHANCED | 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 Java V2

- AWS SDK for Ruby V3

EcsTaskDetails

The details of a task definition that describes the container and volume definitions of an Amazon ECS task.

Contents

containerInstanceArn

The Amazon Resource Name (ARN) of the container instance that hosts the task.

Type: String

Required: No

containers

A list of containers that are included in the `taskProperties` list.

Type: Array of TaskContainerDetails objects

Required: No

enableExecuteCommand

Determines whether execute command functionality is turned on for this task. If `true`, execute command functionality is turned on all the containers in the task.

Type: Boolean

Required: No

ephemeralStorage

The amount of ephemeral storage allocated for the task.

Type: EphemeralStorage object

Required: No

executionRoleArn

The Amazon Resource Name (ARN) of the execution role that AWS Batch can assume. For more information, see Batch execution IAM role in the *AWS Batch User Guide*.

Type: String

Required: No

ipcMode

The IPC resource namespace to use for the containers in the task. The valid values are `host`, `task`, or none. For more information see `ipcMode` in EcsTaskProperties.

Type: String

Required: No

networkConfiguration

The network configuration for jobs that are running on Fargate resources. Jobs that are running on Amazon EC2 resources must not specify this parameter.

Type: NetworkConfiguration object

Required: No

networkMode

The network mode configured for the task. This field is populated for jobs running on Amazon ECS Managed Instances (`MANAGED_INSTANCES` platform capability) and always returns `host`.

Type: String

Required: No

pidMode

The process namespace to use for the containers in the task. The valid values are `host`, or `task`. For more information see `pidMode` in EcsTaskProperties.

Type: String

Required: No

platformVersion

The Fargate platform version where the jobs are running.

Type: String

Required: No

runtimePlatform

An object that represents the compute environment architecture for AWS Batch jobs on Fargate or Amazon ECS Managed Instances. Contains the operating system family and CPU architecture of the task.

Type: RuntimePlatform object

Required: No

taskArn

The ARN of the Amazon ECS task.

Type: String

Required: No

taskRoleArn

The Amazon Resource Name (ARN) of the IAM role that the container can assume for AWS permissions. For more information, see IAM roles for tasks in the *Amazon Elastic Container Service Developer Guide*.

> **Note**
>
> This is object is comparable to ContainerProperties:jobRoleArn.

Type: String

Required: No

volumes

A list of data volumes used in a job.

Type: Array of Volume 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 Java V2
- AWS SDK for Ruby V3

EcsTaskProperties

The properties for a task definition that describes the container and volume definitions of an Amazon ECS task. You can specify which Docker images to use, the required resources, and other configurations related to launching the task definition through an Amazon ECS service or task.

Contents

containers

This object is a list of containers.

Type: Array of TaskContainerProperties objects

Required: Yes

enableExecuteCommand

Determines whether execute command functionality is turned on for this task. If `true`, execute command functionality is turned on all the containers in the task.

Type: Boolean

Required: No

ephemeralStorage

The amount of ephemeral storage to allocate for the task. This parameter is used to expand the total amount of ephemeral storage available, beyond the default amount, for tasks hosted on AWS Fargate.

Type: EphemeralStorage object

Required: No

executionRoleArn

The Amazon Resource Name (ARN) of the execution role that AWS Batch can assume. For jobs that run on Fargate resources, you must provide an execution role. For more information, see AWS Batch execution IAM role in the *AWS Batch User Guide*.

Type: String

Required: No

ipcMode

The IPC resource namespace to use for the containers in the task. The valid values are `host`, `task`, or none.

If `host` is specified, all containers within the tasks that specified the `host` IPC mode on the same container instance share the same IPC resources with the host Amazon EC2 instance.

If `task` is specified, all containers within the specified `task` share the same IPC resources.

If none is specified, the IPC resources within the containers of a task are private, and are not shared with other containers in a task or on the container instance.

If no value is specified, then the IPC resource namespace sharing depends on the Docker daemon setting on the container instance. For more information, see IPC settings in the Docker run reference.

> **Note**
>
> This parameter is not supported for jobs that run on Fargate resources.

Type: String

Required: No

networkConfiguration

The network configuration for jobs that are running on Fargate resources. Jobs that are running on Amazon EC2 resources or Amazon ECS Managed Instances must not specify this parameter.

Type: NetworkConfiguration object

Required: No

networkMode

The network mode to use for the task. Valid values: `host`. When not specified, the default is `host`.

With `host` mode, the container shares the host instance's network stack directly. When running tasks that use the `host` network mode, do not run containers using the root user (UID 0). Running as root grants unrestricted access to host resources and increases the attack surface.

This parameter only applies to jobs running on Amazon ECS Managed Instances (`MANAGED_INSTANCES` platform capability). It cannot be specified for Fargate or Amazon EC2 platform job definitions.

Type: String

Required: No

pidMode

The process namespace to use for the containers in the task. The valid values are `host` or `task`. For example, monitoring sidecars might need `pidMode` to access information about other containers running in the same task.

If `host` is specified, all containers within the tasks that specified the `host` PID mode on the same container instance share the process namespace with the host Amazon EC2 instance.

If `task` is specified, all containers within the specified task share the same process namespace.

If no value is specified, the default is a private namespace for each container. For more information, see PID settings in the Docker run reference.

Type: String

Required: No

platformVersion

The Fargate platform version where the jobs are running. A platform version is specified only for jobs that are running on Fargate resources. If one isn't specified, the `LATEST` platform version is used by default. This uses a recent, approved version of the Fargate platform for compute resources. For more information, see AWS Fargate platform versions in the *Amazon Elastic Container Service Developer Guide*.

Type: String

Required: No

runtimePlatform

An object that represents the compute environment architecture for AWS Batch jobs on Fargate or Amazon ECS Managed Instances. Use this to specify the operating system family (`operatingSystemFamily`) and CPU architecture (`cpuArchitecture`).

For Amazon ECS Managed Instances, the valid value for `operatingSystemFamily` is `LINUX` (default). The valid values for `cpuArchitecture` are `X86_64` and `ARM64`.

Type: RuntimePlatform object

Required: No

taskRoleArn

The Amazon Resource Name (ARN) that's associated with the Amazon ECS task.

> **Note**
>
> This is object is comparable to ContainerProperties:jobRoleArn.

Type: String

Required: No

volumes

A list of volumes that are associated with the job.

Type: Array of Volume 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 Java V2
- AWS SDK for Ruby V3

EFSAuthorizationConfig

The authorization configuration details for the Amazon EFS file system.

Contents

accessPointId

The Amazon EFS access point ID to use. If an access point is specified, the root directory value specified in the `EFSVolumeConfiguration` must either be omitted or set to `/` which enforces the path set on the EFS access point. If an access point is used, transit encryption must be enabled in the `EFSVolumeConfiguration`. For more information, see Working with Amazon EFS access points in the *Amazon Elastic File System User Guide*.

Type: String

Required: No

iam

Whether or not to use the AWS Batch job IAM role defined in a job definition when mounting the Amazon EFS file system. If enabled, transit encryption must be enabled in the `EFSVolumeConfiguration`. If this parameter is omitted, the default value of `DISABLED` is used. For more information, see Using Amazon EFS access points in the *AWS Batch User Guide*. EFS IAM authorization requires that `TransitEncryption` be `ENABLED` and that a `JobRoleArn` is specified.

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 Java V2

- AWS SDK for Ruby V3

EFSVolumeConfiguration

This is used when you're using an Amazon Elastic File System file system for job storage. For more information, see Amazon EFS Volumes in the *AWS Batch User Guide*.

Contents

fileSystemId

The Amazon EFS file system ID to use.

Type: String

Required: Yes

authorizationConfig

The authorization configuration details for the Amazon EFS file system.

Type: EFSAuthorizationConfig object

Required: No

rootDirectory

The directory within the Amazon EFS file system to mount as the root directory inside the host. If this parameter is omitted, the root of the Amazon EFS volume is used instead. Specifying `/` has the same effect as omitting this parameter. The maximum length is 4,096 characters.

> **Important**
>
> If an EFS access point is specified in the `authorizationConfig`, the root directory parameter must either be omitted or set to `/`, which enforces the path set on the Amazon EFS access point.

Type: String

Required: No

transitEncryption

Determines whether to enable encryption for Amazon EFS data in transit between the Amazon ECS host and the Amazon EFS server. Transit encryption must be enabled if Amazon EFS IAM

authorization is used. If this parameter is omitted, the default value of `DISABLED` is used. For more information, see Encrypting data in transit in the *Amazon Elastic File System User Guide*.

Type: String

Valid Values: `ENABLED | DISABLED`

Required: No

transitEncryptionPort

The port to use when sending encrypted data between the Amazon ECS host and the Amazon EFS server. If you don't specify a transit encryption port, it uses the port selection strategy that the Amazon EFS mount helper uses. The value must be between 0 and 65,535. For more information, see EFS mount helper in the *Amazon Elastic File System 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 Java V2
- AWS SDK for Ruby V3

EksAccessEntry

Configures whether AWS Batch manages an Amazon EKS access entry on the cluster for the compute environment. For information on how the fields interact with the cluster's `authenticationMode` and with other compute environments that share the cluster, see Amazon EKS access entry authentication in the *AWS Batch User Guide*.

Note

Setting `desiredState=ENABLED` on a single compute environment does not guarantee that AWS Batch creates an access entry, and setting `desiredState=DISABLED` on a single compute environment does not guarantee that AWS Batch deletes one. AWS Batch compares the `desiredState` across all compute environments that target the same cluster. The AWS Batch-managed access entry is created only when all compute environments have `desiredState=ENABLED`, and deleted only when all have `desiredState=DISABLED`. If you have multiple compute environments on the same cluster, set `desiredState` consistently across all of them to avoid uncertainty. For more information, see Reconciling desiredState across compute environments in the *AWS Batch User Guide*.

Contents

desiredState

The desired access entry state for the compute environment. Valid values:

ENABLED

AWS Batch manages an access entry on the cluster for the compute environment.

DISABLED

AWS Batch deletes the AWS Batch-managed access entry for the cluster. This value is rejected if the cluster's `authenticationMode` is `API`, because such a cluster doesn't support the `aws-auth` ConfigMap.

INHERIT_FROM_CLUSTER

AWS Batch defers to the cluster's current access entry `status`. On a cluster whose authentication mode is `API`, AWS Batch creates and manages an access entry. On a cluster

whose authentication mode is `API_AND_CONFIG_MAP` or `CONFIG_MAP`, AWS Batch neither adds nor removes an access entry.

Type: String

Valid Values: `ENABLED | DISABLED | INHERIT_FROM_CLUSTER`

Required: Yes

status

The observed state of the access entry on the cluster. `ACTIVE` means that an access entry for the compute environment exists on the cluster and takes precedence over the `aws-auth` ConfigMap. `INACTIVE` means that no AWS Batch-managed access entry is present. This is a read-only field returned by `DescribeComputeEnvironments`.

Type: String

Valid Values: `ACTIVE | INACTIVE`

Required: No

See Also

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

- AWS SDK for C++
- AWS SDK for Java V2
- AWS SDK for Ruby V3

EksAttemptContainerDetail

An object that represents the details for an attempt for a job attempt that an Amazon EKS container runs.

Contents

containerID

The ID for the container.

Type: String

Required: No

exitCode

The exit code returned for the job attempt. A non-zero exit code is considered failed.

Type: Integer

Required: No

name

The name of a container.

Type: String

Required: No

reason

A short (255 max characters) human-readable string to provide additional details for a running or stopped container.

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 Java V2
- AWS SDK for Ruby V3

EksAttemptDetail

An object that represents the details of a job attempt for a job attempt by an Amazon EKS container.

Contents

containers

The details for the final status of the containers for this job attempt.

Type: Array of EksAttemptContainerDetail objects

Required: No

eksClusterArn

The Amazon Resource Name (ARN) of the Amazon EKS cluster.

Type: String

Required: No

initContainers

The details for the init containers.

Type: Array of EksAttemptContainerDetail objects

Required: No

nodeName

The name of the node for this job attempt.

Type: String

Required: No

podName

The name of the pod for this job attempt.

Type: String

Required: No

podNamespace

The namespace of the Amazon EKS cluster that the pod exists in.

Type: String

Required: No

startedAt

The Unix timestamp (in milliseconds) for when the attempt was started (when the attempt transitioned from the `STARTING` state to the `RUNNING` state).

Type: Long

Required: No

statusReason

A short, human-readable string to provide additional details for the current status of the job attempt.

Type: String

Required: No

stoppedAt

The Unix timestamp (in milliseconds) for when the attempt was stopped. This happens when the attempt transitioned from the `RUNNING` state to a terminal state, such as `SUCCEEDED` or `FAILED`.

Type: Long

Required: No

See Also

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

- AWS SDK for C++
- AWS SDK for Java V2

- AWS SDK for Ruby V3

EksConfiguration

Configuration for the Amazon EKS cluster that supports the AWS Batch compute environment. The cluster must exist before the compute environment can be created.

Contents

eksClusterArn

The Amazon Resource Name (ARN) of the Amazon EKS cluster. An example is `arn:aws:eks:us-east-1:123456789012:cluster/ClusterForBatch` .

Type: String

Required: Yes

kubernetesNamespace

The namespace of the Amazon EKS cluster. AWS Batch manages pods in this namespace. The value can't left empty or null. It must be fewer than 64 characters long, can't be set to `default`, can't start with "`kube-`," and must match this regular expression: `^[a-z0-9]([-a-z0-9]*[a-z0-9])?$`. For more information, see Namespaces in the Kubernetes documentation.

Type: String

Required: Yes

accessEntry

The AWS Batch-managed Amazon EKS access entry for the compute environment. Set `desiredState` to declare whether AWS Batch manages an access entry on the cluster. In a `DescribeComputeEnvironments` response, `desiredState` is the value that AWS Batch recorded for the compute environment and `status` is the observed state of the access entry on the cluster. To change the access entry on an existing compute environment, use `EksConfigurationUpdate.accessEntry`.

Whether the entry is provisioned on the cluster depends on the cluster's `authenticationMode` and the `desiredState` recorded for each AWS Batch compute environment targeting the cluster. For more information, see Amazon EKS access entry authentication in the *AWS Batch User Guide*.

If you don't specify this field, AWS Batch doesn't record a `desiredState` for the compute environment and `DescribeComputeEnvironments` doesn't return one. For the purpose of provisioning the access entry, AWS Batch behaves as it does for `INHERIT_FROM_CLUSTER`.

Type: EksAccessEntry 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 Java V2
- AWS SDK for Ruby V3

EksConfigurationUpdate

An object that represents the attributes of an AWS Batch compute environment's Amazon EKS configuration that can be updated. Currently `accessEntry` is the only attribute that you can change after the compute environment is created. For more information, see Amazon EKS access entry authentication in the *AWS Batch User Guide*.

Contents

accessEntry

The updated access entry configuration for the compute environment. Set `desiredState` to declare whether AWS Batch will manage an access entry on the cluster. For the accepted values, see `EksAccessEntry`.

Type: EksAccessEntry 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 Java V2
- AWS SDK for Ruby V3

EksContainer

EKS container properties are used in job definitions for Amazon EKS based job definitions to describe the properties for a container node in the pod that's launched as part of a job. This can't be specified for Amazon ECS based job definitions.

Contents

image

The Docker image used to start the container.

Type: String

Required: Yes

args

An array of arguments to the entrypoint. If this isn't specified, the `CMD` of the container image is used. This corresponds to the `args` member in the Entrypoint portion of the Pod in Kubernetes. Environment variable references are expanded using the container's environment.

If the referenced environment variable doesn't exist, the reference in the command isn't changed. For example, if the reference is to "`$(NAME1)`" and the `NAME1` environment variable doesn't exist, the command string will remain "`$(NAME1)`." `$$` is replaced with `$`, and the resulting string isn't expanded. For example, `$$(VAR_NAME)` is passed as `$(VAR_NAME)` whether or not the `VAR_NAME` environment variable exists. For more information, see Dockerfile reference: CMD and Define a command and arguments for a pod in the *Kubernetes documentation*.

Type: Array of strings

Required: No

command

The entrypoint for the container. This isn't run within a shell. If this isn't specified, the `ENTRYPOINT` of the container image is used. Environment variable references are expanded using the container's environment.

If the referenced environment variable doesn't exist, the reference in the command isn't changed. For example, if the reference is to "`$(NAME1)`" and the `NAME1` environment variable

doesn't exist, the command string will remain "$(NAME1)." $$ is replaced with $ and the resulting string isn't expanded. For example, $$(VAR_NAME) will be passed as $(VAR_NAME) whether or not the VAR_NAME environment variable exists. The entrypoint can't be updated. For more information, see ENTRYPOINT in the *Dockerfile reference* and Define a command and arguments for a container and Entrypoint in the *Kubernetes documentation*.

Type: Array of strings

Required: No

env

The environment variables to pass to a container.

> **Note**
>
> Environment variables cannot start with "AWS_BATCH". This naming convention is reserved for variables that AWS Batch sets.

Type: Array of EksContainerEnvironmentVariable objects

Required: No

imagePullPolicy

The image pull policy for the container. Supported values are Always, IfNotPresent, and Never. This parameter defaults to IfNotPresent. However, if the :latest tag is specified, it defaults to Always. For more information, see Updating images in the *Kubernetes documentation*.

Type: String

Required: No

name

The name of the container. If the name isn't specified, the default name "Default" is used. Each container in a pod must have a unique name.

Type: String

Required: No

resources

The type and amount of resources to assign to a container. The supported resources include `memory`, `cpu`, and `nvidia.com/gpu`. For more information, see Resource management for pods and containers in the *Kubernetes documentation*.

Type: EksContainerResourceRequirements object

Required: No

securityContext

The security context for a job. For more information, see Configure a security context for a pod or container in the *Kubernetes documentation*.

Type: EksContainerSecurityContext object

Required: No

volumeMounts

The volume mounts for the container. AWS Batch supports `emptyDir`, `hostPath`, and `secret` volume types. For more information about volumes and volume mounts in Kubernetes, see Volumes in the *Kubernetes documentation*.

Type: Array of EksContainerVolumeMount 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 Java V2
- AWS SDK for Ruby V3

EksContainerDetail

The details for container properties that are returned by `DescribeJobs` for jobs that use Amazon EKS.

Contents

args

An array of arguments to the entrypoint. If this isn't specified, the `CMD` of the container image is used. This corresponds to the `args` member in the Entrypoint portion of the Pod in Kubernetes. Environment variable references are expanded using the container's environment.

If the referenced environment variable doesn't exist, the reference in the command isn't changed. For example, if the reference is to `"$(NAME1)"` and the `NAME1` environment variable doesn't exist, the command string will remain `"$(NAME1)"`. `$$` is replaced with `$` and the resulting string isn't expanded. For example, `$$(VAR_NAME)` is passed as `$(VAR_NAME)` whether or not the `VAR_NAME` environment variable exists. For more information, see Dockerfile reference: CMD and Define a command and arguments for a pod in the *Kubernetes documentation*.

Type: Array of strings

Required: No

command

The entrypoint for the container. For more information, see Entrypoint in the *Kubernetes documentation*.

Type: Array of strings

Required: No

env

The environment variables to pass to a container.

> **Note**
>
> Environment variables cannot start with `"AWS_BATCH"`. This naming convention is reserved for variables that AWS Batch sets.

Type: Array of EksContainerEnvironmentVariable objects

Required: No

exitCode

The exit code returned for the job attempt. A non-zero exit code is considered failed.

Type: Integer

Required: No

image

The Docker image used to start the container.

Type: String

Required: No

imagePullPolicy

The image pull policy for the container. Supported values are `Always`, `IfNotPresent`, and `Never`. This parameter defaults to `Always` if the `:latest` tag is specified, `IfNotPresent` otherwise. For more information, see Updating images in the *Kubernetes documentation*.

Type: String

Required: No

name

The name of the container. If the name isn't specified, the default name "`Default`" is used. Each container in a pod must have a unique name.

Type: String

Required: No

reason

A short human-readable string to provide additional details for a running or stopped container. It can be up to 255 characters long.

Type: String

Required: No

resources

The type and amount of resources to assign to a container. The supported resources include `memory`, `cpu`, and `nvidia.com/gpu`. For more information, see Resource management for pods and containers in the *Kubernetes documentation*.

Type: EksContainerResourceRequirements object

Required: No

securityContext

The security context for a job. For more information, see Configure a security context for a pod or container in the *Kubernetes documentation*.

Type: EksContainerSecurityContext object

Required: No

volumeMounts

The volume mounts for the container. AWS Batch supports `emptyDir`, `hostPath`, and `secret` volume types. For more information about volumes and volume mounts in Kubernetes, see Volumes in the *Kubernetes documentation*.

Type: Array of EksContainerVolumeMount 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 Java V2
- AWS SDK for Ruby V3

EksContainerEnvironmentVariable

An environment variable.

Contents

name

The name of the environment variable.

Type: String

Required: Yes

value

The value of the environment variable.

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 Java V2
- AWS SDK for Ruby V3

EksContainerOverride

Object representing any Kubernetes overrides to a job definition that's used in a SubmitJob API operation.

Contents

args

The arguments to the entrypoint to send to the container that overrides the default arguments from the Docker image or the job definition. For more information, see Dockerfile reference: CMD and Define a command an arguments for a pod in the *Kubernetes documentation*.

Type: Array of strings

Required: No

command

The command to send to the container that overrides the default command from the Docker image or the job definition.

Type: Array of strings

Required: No

env

The environment variables to send to the container. You can add new environment variables, which are added to the container at launch. Or, you can override the existing environment variables from the Docker image or the job definition.

> **Note**
>
> Environment variables cannot start with "`AWS_BATCH`". This naming convention is reserved for variables that AWS Batch sets.

Type: Array of EksContainerEnvironmentVariable objects

Required: No

image

The override of the Docker image that's used to start the container.

Type: String

Required: No

name

A pointer to the container that you want to override. The name must match a unique container name that you wish to override.

Type: String

Required: No

resources

The type and amount of resources to assign to a container. These override the settings in the job definition. The supported resources include `memory`, `cpu`, and `nvidia.com/gpu`. For more information, see Resource management for pods and containers in the *Kubernetes documentation*.

Type: EksContainerResourceRequirements 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 Java V2
- AWS SDK for Ruby V3

EksContainerResourceRequirements

The type and amount of resources to assign to a container. The supported resources include `memory`, `cpu`, and `nvidia.com/gpu`. For more information, see Resource management for pods and containers in the *Kubernetes documentation*.

Contents

limits

The type and quantity of the resources to reserve for the container. The values vary based on the `name` that's specified. Resources can be requested using either the `limits` or the `requests` objects.

memory

The memory hard limit (in MiB) for the container, using whole integers, with a "Mi" suffix. If your container attempts to exceed the memory specified, the container is terminated. You must specify at least 4 MiB of memory for a job. `memory` can be specified in `limits`, `requests`, or both. If `memory` is specified in both places, then the value that's specified in `limits` must be equal to the value that's specified in `requests`.

> **Note**
>
> To maximize your resource utilization, provide your jobs with as much memory as possible for the specific instance type that you are using. To learn how, see Memory management in the *AWS Batch User Guide*.

cpu

The number of CPUs that's reserved for the container. Values must be an even multiple of `0.25`. `cpu` can be specified in `limits`, `requests`, or both. If `cpu` is specified in both places, then the value that's specified in `limits` must be at least as large as the value that's specified in `requests`.

nvidia.com/gpu

The number of GPUs that's reserved for the container. Values must be a whole integer. `memory` can be specified in `limits`, `requests`, or both. If `memory` is specified in both

places, then the value that's specified in `limits` must be equal to the value that's specified in `requests`.

Type: String to string map

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

Required: No

requests

The type and quantity of the resources to request for the container. The values vary based on the `name` that's specified. Resources can be requested by using either the `limits` or the `requests` objects.

memory

The memory hard limit (in MiB) for the container, using whole integers, with a "Mi" suffix. If your container attempts to exceed the memory specified, the container is terminated. You must specify at least 4 MiB of memory for a job. `memory` can be specified in `limits`, `requests`, or both. If `memory` is specified in both, then the value that's specified in `limits` must be equal to the value that's specified in `requests`.

> **Note**
>
> If you're trying to maximize your resource utilization by providing your jobs as much memory as possible for a particular instance type, see Memory management in the *AWS Batch User Guide*.

cpu

The number of CPUs that are reserved for the container. Values must be an even multiple of `0.25`. `cpu` can be specified in `limits`, `requests`, or both. If `cpu` is specified in both, then the value that's specified in `limits` must be at least as large as the value that's specified in `requests`.

nvidia.com/gpu

The number of GPUs that are reserved for the container. Values must be a whole integer. `nvidia.com/gpu` can be specified in `limits`, `requests`, or both. If `nvidia.com/gpu` is

specified in both, then the value that's specified in `limits` must be equal to the value that's specified in `requests`.

Type: String to string map

Value 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 Java V2
- AWS SDK for Ruby V3

EksContainerSecurityContext

The security context for a job. For more information, see Configure a security context for a pod or container in the *Kubernetes documentation*.

Contents

allowPrivilegeEscalation

Whether or not a container or a Kubernetes pod is allowed to gain more privileges than its parent process. The default value is `false`.

Type: Boolean

Required: No

privileged

When this parameter is `true`, the container is given elevated permissions on the host container instance. The level of permissions are similar to the `root` user permissions. The default value is `false`. This parameter maps to `privileged` policy in the Privileged pod security policies in the *Kubernetes documentation*.

Type: Boolean

Required: No

readOnlyRootFilesystem

When this parameter is `true`, the container is given read-only access to its root file system. The default value is `false`. This parameter maps to `ReadOnlyRootFilesystem` policy in the Volumes and file systems pod security policies in the *Kubernetes documentation*.

Type: Boolean

Required: No

runAsGroup

When this parameter is specified, the container is run as the specified group ID (`gid`). If this parameter isn't specified, the default is the group that's specified in the image metadata. This parameter maps to `RunAsGroup` and `MustRunAs` policy in the Users and groups pod security policies in the *Kubernetes documentation*.

Type: Long

Required: No

runAsNonRoot

When this parameter is specified, the container is run as a user with a `uid` other than 0. If this parameter isn't specified, so such rule is enforced. This parameter maps to `RunAsUser` and `MustRunAsNonRoot` policy in the Users and groups pod security policies in the *Kubernetes documentation*.

Type: Boolean

Required: No

runAsUser

When this parameter is specified, the container is run as the specified user ID (`uid`). If this parameter isn't specified, the default is the user that's specified in the image metadata. This parameter maps to `RunAsUser` and `MustRanAs` policy in the Users and groups pod security policies in the *Kubernetes documentation*.

Type: Long

Required: No

See Also

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

- AWS SDK for C++
- AWS SDK for Java V2
- AWS SDK for Ruby V3

EksContainerVolumeMount

The volume mounts for a container for an Amazon EKS job. For more information about volumes and volume mounts in Kubernetes, see Volumes in the *Kubernetes documentation*.

Contents

mountPath

The path on the container where the volume is mounted.

Type: String

Required: No

name

The name the volume mount. This must match the name of one of the volumes in the pod.

Type: String

Required: No

readOnly

If this value is `true`, the container has read-only access to the volume. Otherwise, the container can write to the volume. The default value is `false`.

Type: Boolean

Required: No

subPath

A sub-path inside the referenced volume instead of its root.

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 Java V2
- AWS SDK for Ruby V3

EksEmptyDir

Specifies the configuration of a Kubernetes `emptyDir` volume. An `emptyDir` volume is first created when a pod is assigned to a node. It exists as long as that pod is running on that node. The `emptyDir` volume is initially empty. All containers in the pod can read and write the files in the `emptyDir` volume. However, the `emptyDir` volume can be mounted at the same or different paths in each container. When a pod is removed from a node for any reason, the data in the `emptyDir` is deleted permanently. For more information, see emptyDir in the *Kubernetes documentation*.

Contents

medium

The medium to store the volume. The default value is an empty string, which uses the storage of the node.

""

(Default) Use the disk storage of the node.

"Memory"

Use the `tmpfs` volume that's backed by the RAM of the node. Contents of the volume are lost when the node reboots, and any storage on the volume counts against the container's memory limit.

Type: String

Required: No

sizeLimit

The maximum size of the volume. By default, there's no maximum size defined.

Type: String

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 Java V2
- AWS SDK for Ruby V3

EksHostPath

Specifies the configuration of a Kubernetes `hostPath` volume. A `hostPath` volume mounts an existing file or directory from the host node's filesystem into your pod. For more information, see hostPath in the *Kubernetes documentation*.

Contents

path

The path of the file or directory on the host to mount into containers on the pod.

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 Java V2
- AWS SDK for Ruby V3

EksMetadata

Describes and uniquely identifies Kubernetes resources. For example, the compute environment that a pod runs in or the `jobID` for a job running in the pod. For more information, see Understanding Kubernetes Objects in the *Kubernetes documentation*.

Contents

annotations

Key-value pairs used to attach arbitrary, non-identifying metadata to Kubernetes objects. Valid annotation keys have two segments: an optional prefix and a name, separated by a slash (/).

- The prefix is optional and must be 253 characters or less. If specified, the prefix must be a DNS subdomain− a series of DNS labels separated by dots (.), and it must end with a slash (/).
- The name segment is required and must be 63 characters or less. It can include alphanumeric characters ([a-z0-9A-Z]), dashes (-), underscores (_), and dots (.), but must begin and end with an alphanumeric character.

> **Note**
>
> Annotation values must be 255 characters or less.

Annotations can be added or modified at any time. Each resource can have multiple annotations.

Type: String to string map

Required: No

labels

Key-value pairs used to identify, sort, and organize cube resources. Can contain up to 63 uppercase letters, lowercase letters, numbers, hyphens (-), and underscores (_). Labels can be added or modified at any time. Each resource can have multiple labels, but each key must be unique for a given object.

Type: String to string map

Required: No

namespace

The namespace of the Amazon EKS cluster. In Kubernetes, namespaces provide a mechanism for isolating groups of resources within a single cluster. Names of resources need to be unique within a namespace, but not across namespaces. AWS Batch places Batch Job pods in this namespace. If this field is provided, the value can't be empty or null. It must meet the following requirements:

- 1-63 characters long
- Can't be set to default
- Can't start with `kube`
- Must match the following regular expression: `^[a-z0-9]([-a-z0-9]*[a-z0-9])?$`

For more information, see Namespaces in the *Kubernetes documentation*. This namespace can be different from the `kubernetesNamespace` set in the compute environment's `EksConfiguration`, but must have identical role-based access control (RBAC) roles as the compute environment's `kubernetesNamespace`. For multi-node parallel jobs, the same value must be provided across all the node ranges.

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 Java V2
- AWS SDK for Ruby V3

EksPersistentVolumeClaim

A `persistentVolumeClaim` volume is used to mount a PersistentVolume into a Pod. PersistentVolumeClaims are a way for users to "claim" durable storage without knowing the details of the particular cloud environment. See the information about PersistentVolumes in the *Kubernetes documentation*.

Contents

claimName

The name of the `persistentVolumeClaim` bounded to a `persistentVolume`. For more information, see Persistent Volume Claims in the *Kubernetes documentation*.

Type: String

Required: Yes

readOnly

An optional boolean value indicating if the mount is read only. Default is false. For more information, see Read Only Mounts in the *Kubernetes documentation*.

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 Java V2
- AWS SDK for Ruby V3

EksPodProperties

The properties for the pod.

Contents

containers

The properties of the container that's used on the Amazon EKS pod.

> **Note**
>
> This object is limited to 10 elements.

Type: Array of EksContainer objects

Required: No

dnsPolicy

The DNS policy for the pod. The default value is `ClusterFirst`. If the `hostNetwork` parameter is not specified, the default is `ClusterFirstWithHostNet`. `ClusterFirst` indicates that any DNS query that does not match the configured cluster domain suffix is forwarded to the upstream nameserver inherited from the node. For more information, see Pod's DNS policy in the *Kubernetes documentation*.

Valid values: `Default` | `ClusterFirst` | `ClusterFirstWithHostNet`

Type: String

Required: No

hostNetwork

Indicates if the pod uses the hosts' network IP address. The default value is `true`. Setting this to `false` enables the Kubernetes pod networking model. Most AWS Batch workloads are egress-only and don't require the overhead of IP allocation for each pod for incoming connections. For more information, see Host namespaces and Pod networking in the *Kubernetes documentation*.

Type: Boolean

Required: No

imagePullSecrets

References a Kubernetes secret resource. It holds a list of secrets. These secrets help to gain access to pull an images from a private registry.

`ImagePullSecret$name` is required when this object is used.

Type: Array of ImagePullSecret objects

Required: No

initContainers

These containers run before application containers, always runs to completion, and must complete successfully before the next container starts. These containers are registered with the Amazon EKS Connector agent and persists the registration information in the Kubernetes backend data store. For more information, see Init Containers in the *Kubernetes documentation*.

> **Note**
>
> This object is limited to 10 elements.

Type: Array of EksContainer objects

Required: No

metadata

Metadata about the Kubernetes pod. For more information, see Understanding Kubernetes Objects in the *Kubernetes documentation*.

Type: EksMetadata object

Required: No

serviceAccountName

The name of the service account that's used to run the pod. For more information, see Kubernetes service accounts and Configure a Kubernetes service account to assume an IAM role in the *Amazon EKS User Guide* and Configure service accounts for pods in the *Kubernetes documentation*.

Type: String

Required: No

shareProcessNamespace

Indicates if the processes in a container are shared, or visible, to other containers in the same pod. For more information, see Share Process Namespace between Containers in a Pod.

Type: Boolean

Required: No

volumes

Specifies the volumes for a job definition that uses Amazon EKS resources.

Type: Array of EksVolume 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 Java V2
- AWS SDK for Ruby V3

EksPodPropertiesDetail

The details for the pod.

Contents

containers

The properties of the container that's used on the Amazon EKS pod.

Type: Array of EksContainerDetail objects

Required: No

dnsPolicy

The DNS policy for the pod. The default value is `ClusterFirst`. If the `hostNetwork` parameter is not specified, the default is `ClusterFirstWithHostNet`. `ClusterFirst` indicates that any DNS query that does not match the configured cluster domain suffix is forwarded to the upstream nameserver inherited from the node. If no value was specified for `dnsPolicy` in the RegisterJobDefinition API operation, then no value will be returned for `dnsPolicy` by either of DescribeJobDefinitions or DescribeJobs API operations. The pod spec setting will contain either `ClusterFirst` or `ClusterFirstWithHostNet`, depending on the value of the `hostNetwork` parameter. For more information, see Pod's DNS policy in the *Kubernetes documentation*.

Valid values: `Default` | `ClusterFirst` | `ClusterFirstWithHostNet`

Type: String

Required: No

hostNetwork

Indicates if the pod uses the hosts' network IP address. The default value is `true`. Setting this to `false` enables the Kubernetes pod networking model. Most AWS Batch workloads are egress-only and don't require the overhead of IP allocation for each pod for incoming connections. For more information, see Host namespaces and Pod networking in the *Kubernetes documentation*.

Type: Boolean

Required: No

imagePullSecrets

Displays the reference pointer to the Kubernetes secret resource. These secrets help to gain access to pull an images from a private registry.

Type: Array of ImagePullSecret objects

Required: No

initContainers

The container registered with the Amazon EKS Connector agent and persists the registration information in the Kubernetes backend data store.

Type: Array of EksContainerDetail objects

Required: No

metadata

Describes and uniquely identifies Kubernetes resources. For example, the compute environment that a pod runs in or the `jobID` for a job running in the pod. For more information, see Understanding Kubernetes Objects in the *Kubernetes documentation*.

Type: EksMetadata object

Required: No

nodeName

The name of the node for this job.

Type: String

Required: No

podName

The name of the pod for this job.

Type: String

Required: No

serviceAccountName

The name of the service account that's used to run the pod. For more information, see Kubernetes service accounts and Configure a Kubernetes service account to assume an IAM

role in the *Amazon EKS User Guide* and Configure service accounts for pods in the *Kubernetes documentation*.

Type: String

Required: No

shareProcessNamespace

Indicates if the processes in a container are shared, or visible, to other containers in the same pod. For more information, see Share Process Namespace between Containers in a Pod.

Type: Boolean

Required: No

volumes

Specifies the volumes for a job definition using Amazon EKS resources.

Type: Array of EksVolume 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 Java V2
- AWS SDK for Ruby V3

EksPodPropertiesOverride

An object that contains overrides for the Kubernetes pod properties of a job.

Contents

containers

The overrides for the container that's used on the Amazon EKS pod.

Type: Array of EksContainerOverride objects

Required: No

initContainers

The overrides for the `initContainers` defined in the Amazon EKS pod. These containers run before application containers, always run to completion, and must complete successfully before the next container starts. These containers are registered with the Amazon EKS Connector agent and persists the registration information in the Kubernetes backend data store. For more information, see Init Containers in the *Kubernetes documentation*.

Type: Array of EksContainerOverride objects

Required: No

metadata

Metadata about the overrides for the container that's used on the Amazon EKS pod.

Type: EksMetadata 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 Java V2

- AWS SDK for Ruby V3

EksProperties

An object that contains the properties for the Kubernetes resources of a job.

Contents

podProperties

The properties for the Kubernetes pod resources of a job.

Type: EksPodProperties 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 Java V2
- AWS SDK for Ruby V3

EksPropertiesDetail

An object that contains the details for the Kubernetes resources of a job.

Contents

podProperties

The properties for the Kubernetes pod resources of a job.

Type: EksPodPropertiesDetail 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 Java V2
- AWS SDK for Ruby V3

EksPropertiesOverride

An object that contains overrides for the Kubernetes resources of a job.

Contents

podProperties

The overrides for the Kubernetes pod resources of a job.

Type: EksPodPropertiesOverride 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 Java V2
- AWS SDK for Ruby V3

EksSecret

Specifies the configuration of a Kubernetes `secret` volume. For more information, see secret in the *Kubernetes documentation*.

Contents

secretName

The name of the secret. The name must be allowed as a DNS subdomain name. For more information, see DNS subdomain names in the *Kubernetes documentation*.

Type: String

Required: Yes

optional

Specifies whether the secret or the secret's keys must be defined.

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 Java V2
- AWS SDK for Ruby V3

EksVolume

Specifies an Amazon EKS volume for a job definition.

Contents

name

The name of the volume. The name must be allowed as a DNS subdomain name. For more information, see DNS subdomain names in the *Kubernetes documentation*.

Type: String

Required: Yes

emptyDir

Specifies the configuration of a Kubernetes `emptyDir` volume. For more information, see emptyDir in the *Kubernetes documentation*.

Type: EksEmptyDir object

Required: No

hostPath

Specifies the configuration of a Kubernetes `hostPath` volume. For more information, see hostPath in the *Kubernetes documentation*.

Type: EksHostPath object

Required: No

persistentVolumeClaim

Specifies the configuration of a Kubernetes `persistentVolumeClaim` bounded to a `persistentVolume`. For more information, see Persistent Volume Claims in the *Kubernetes documentation*.

Type: EksPersistentVolumeClaim object

Required: No

secret

Specifies the configuration of a Kubernetes `secret` volume. For more information, see secret in the *Kubernetes documentation*.

Type: EksSecret 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 Java V2
- AWS SDK for Ruby V3

EphemeralStorage

The amount of ephemeral storage to allocate for the task. This parameter is used to expand the total amount of ephemeral storage available, beyond the default amount, for tasks hosted on AWS Fargate.

Contents

sizeInGiB

The total amount, in GiB, of ephemeral storage to set for the task. The minimum supported value is `21` GiB and the maximum supported value is `200` GiB.

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 Java V2
- AWS SDK for Ruby V3

EvaluateOnExit

Specifies an array of up to 5 conditions to be met, and an action to take (RETRY or EXIT) if all conditions are met. If none of the `EvaluateOnExit` conditions in a `RetryStrategy` match, then the job is retried.

Contents

action

Specifies the action to take if all of the specified conditions (on`StatusReason`, on`Reason`, and on`ExitCode`) are met. The values aren't case sensitive.

Type: String

Valid Values: `RETRY | EXIT`

Required: Yes

onExitCode

Contains a glob pattern to match against the decimal representation of the `ExitCode` returned for a job. The pattern can be up to 512 characters long. It can contain only numbers, and can end with an asterisk (*) so that only the start of the string needs to be an exact match.

The string can contain up to 512 characters.

Type: String

Required: No

onReason

Contains a glob pattern to match against the `Reason` returned for a job. The pattern can contain up to 512 characters. It can contain letters, numbers, periods (.), colons (:), and white space (including spaces and tabs). It can optionally end with an asterisk (*) so that only the start of the string needs to be an exact match.

Type: String

Required: No

onStatusReason

Contains a glob pattern to match against the `StatusReason` returned for a job. The pattern can contain up to 512 characters. It can contain letters, numbers, periods (.), colons (:), and white spaces (including spaces or tabs). It can optionally end with an asterisk (*) so that only the start of the string needs to be an exact match.

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 Java V2
- AWS SDK for Ruby V3

FairshareCapacityUsage

The capacity usage for a fairshare scheduling job queue.

Contents

capacityUnit

The unit of measure for the capacity usage. For compute jobs, this is `VCPU` for Amazon EC2 and `cpu` for Amazon EKS. For service jobs, this is the instance type.

Type: String

Required: No

quantity

The quantity of capacity being used, measured in the units specified by `capacityUnit`.

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 Java V2
- AWS SDK for Ruby V3

FairshareCapacityUtilization

The capacity utilization for a specific share in a fairshare scheduling job queue, including the share identifier and its current usage.

Contents

capacityUsage

The capacity usage information for this share, including the unit of measure and quantity being used. This is VCPU for Amazon EC2 and cpu for Amazon EKS.

Type: Array of FairshareCapacityUsage objects

Required: No

shareIdentifier

The share identifier for the fairshare scheduling job queue.

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 Java V2
- AWS SDK for Ruby V3

FairsharePolicy

The fair-share scheduling policy details.

Contents

computeReservation

A value used to reserve some of the available maximum vCPU for share identifiers that aren't already used.

The reserved ratio is `(`*`computeReservation`*`/100)^`*`ActiveFairShares`* where *`ActiveFairShares`* is the number of active share identifiers.

For example, a `computeReservation` value of 50 indicates that AWS Batch reserves 50% of the maximum available vCPU if there's only one share identifier. It reserves 25% if there are two share identifiers. It reserves 12.5% if there are three share identifiers. A `computeReservation` value of 25 indicates that AWS Batch should reserve 25% of the maximum available vCPU if there's only one share identifier, 6.25% if there are two fair share identifiers, and 1.56% if there are three share identifiers.

The minimum value is 0 and the maximum value is 99.

Type: Integer

Required: No

shareDecaySeconds

The amount of time (in seconds) to use to calculate a fair-share percentage for each share identifier in use. A value of zero (0) indicates the default minimum time window (600 seconds). The maximum supported value is 604800 (1 week).

The decay allows for more recently run jobs to have more weight than jobs that ran earlier. Consider adjusting this number if you have jobs that (on average) run longer than ten minutes, or a large difference in job count or job run times between share identifiers, and the allocation of resources doesn't meet your needs.

Type: Integer

Required: No

shareDistribution

An array of `SharedIdentifier` objects that contain the weights for the share identifiers for the fair-share policy. Share identifiers that aren't included have a default weight of `1.0`.

Type: Array of ShareAttributes 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 Java V2
- AWS SDK for Ruby V3

FairshareUtilizationDetail

The fairshare utilization for a job queue, including the number of active shares and top capacity utilization.

Contents

activeShareCount

The total number of active shares in the fairshare scheduling job queue that are currently utilizing capacity.

Type: Long

Required: No

topCapacityUtilization

A list of the top 20 shares with the highest capacity utilization, ordered by usage amount.

Type: Array of FairshareCapacityUtilization 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 Java V2
- AWS SDK for Ruby V3

FargatePlatformConfiguration

The platform configuration for jobs that are running on Fargate resources. Jobs that run on Amazon EC2 resources must not specify this parameter.

Contents

platformVersion

The AWS Fargate platform version where the jobs are running. A platform version is specified only for jobs that are running on Fargate resources. If one isn't specified, the LATEST platform version is used by default. This uses a recent, approved version of the AWS Fargate platform for compute resources. For more information, see AWS Fargate platform versions in the *Amazon Elastic Container Service Developer Guide*.

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 Java V2
- AWS SDK for Ruby V3

FirelensConfiguration

The FireLens configuration for the container. This is used to specify and configure a log router for container logs. For more information, see Custom log routing in the *Amazon Elastic Container Service Developer Guide*.

Contents

type

The log router to use. The valid values are `fluentd` or `fluentbit`.

Type: String

Valid Values: `fluentd | fluentbit`

Required: Yes

options

The options to use when configuring the log router. This field is optional and can be used to specify a custom configuration file or to add additional metadata, such as the task, task definition, cluster, and container instance details to the log event. If specified, the syntax to use is `"options":{"enable-ecs-log-metadata":"true|false","config-file-type:"s3|file","config-file-value":"arn:aws:s3:::mybucket/fluent.conf|filepath"}`. For more information, see Creating a task definition that uses a FireLens configuration in the *Amazon Elastic Container Service Developer Guide*.

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 Java V2
- AWS SDK for Ruby V3

FrontOfQueueDetail

Contains a list of the first 100 RUNNABLE jobs associated to a single job queue.

Contents

jobs

The Amazon Resource Names (ARNs) of the first 100 RUNNABLE jobs in a named job queue. For first-in-first-out (FIFO) job queues, jobs are ordered based on their submission time. For fair-share scheduling (FSS) job queues, jobs are ordered based on their job priority and share usage.

Type: Array of FrontOfQueueJobSummary objects

Required: No

lastUpdatedAt

The Unix timestamp (in milliseconds) for when each of the first 100 RUNNABLE jobs were last updated.

Type: Long

Required: No

See Also

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

- AWS SDK for C++
- AWS SDK for Java V2
- AWS SDK for Ruby V3

FrontOfQueueJobSummary

An object that represents summary details for the first 100 `RUNNABLE` jobs in a job queue.

Contents

earliestTimeAtPosition

The Unix timestamp (in milliseconds) for when the job transitioned to its current position in the job queue.

Type: Long

Required: No

jobArn

The ARN for a job in a named job queue.

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 Java V2
- AWS SDK for Ruby V3

FrontOfQuotaShareJobSummary

An object that represents summary details for the first RUNNABLE job in a quota share.

Contents

earliestTimeAtPosition

The Unix timestamp (in milliseconds) for when the job transitioned to its current position in the quota share.

Type: Long

Required: No

jobArn

The ARN for a job in a named quota share.

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 Java V2
- AWS SDK for Ruby V3

FrontOfQuotaSharesDetail

An object that represents the details of the first RUNNABLE job in each named quota share associated with a single job queue.

Contents

lastUpdatedAt

The Unix timestamp (in milliseconds) for when the first RUNNABLE job per quota share were all last updated.

Type: Long

Required: No

quotaShares

Contains a list of the first RUNNABLE job in each named quota share.

Type: String to array of FrontOfQuotaShareJobSummary objects 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 Java V2
- AWS SDK for Ruby V3

Host

Determine whether your data volume persists on the host container instance and where it's stored. If this parameter is empty, then the Docker daemon assigns a host path for your data volume. However, the data isn't guaranteed to persist after the containers that are associated with it stop running.

Contents

sourcePath

The path on the host container instance that's presented to the container. If this parameter is empty, then the Docker daemon has assigned a host path for you. If this parameter contains a file location, then the data volume persists at the specified location on the host container instance until you delete it manually. If the source path location doesn't exist on the host container instance, the Docker daemon creates it. If the location does exist, the contents of the source path folder are exported.

> **Note**
>
> This parameter isn't applicable to jobs that run on Fargate resources. Don't provide this for these jobs.

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 Java V2
- AWS SDK for Ruby V3

ImagePullSecret

References a Kubernetes secret resource. This name of the secret must start and end with an alphanumeric character, is required to be lowercase, can include periods (.) and hyphens (-), and can't contain more than 253 characters.

Contents

name

Provides a unique identifier for the `ImagePullSecret`. This object is required when `EksPodProperties$imagePullSecrets` is used.

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 Java V2
- AWS SDK for Ruby V3

InfrastructureOptimization

The infrastructure optimization configuration for an Amazon ECS Managed Instances capacity provider. Specifies the idle-instance scale-in behavior.

Contents

scaleInAfter

The number of seconds an instance can remain idle before it is terminated. Valid values are `-1` or `0` to `3600`. Use `-1` as a special value to disable scale-in (instances are never terminated for being idle). If not specified, a default value applies.

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 Java V2
- AWS SDK for Ruby V3

InstanceLaunchTemplate

The instance launch configuration for an Amazon ECS Managed Instances capacity provider. Specifies the instance profile, networking, instance selection constraints, capacity pricing model, storage, and monitoring settings.

Contents

ec2InstanceProfileArn

The Amazon Resource Name (ARN) of the Amazon EC2 instance profile for the managed instances. The instance profile must use the `AmazonECSInstanceRolePolicyForManagedInstances` managed policy with a trust policy for `ec2.amazonaws.com`.

Type: String

Required: Yes

networkConfiguration

The network configuration for the managed instances. Specifies the VPC subnets and security groups where instances are launched.

Type: ManagedInstancesNetworkConfiguration object

Required: Yes

capacityOptionType

The capacity pricing model for the managed instances. Valid values:

- `ON_DEMAND` (default) — On-Demand pricing.
- `SPOT` — Spot Instances, which can provide significant cost savings for fault-tolerant workloads.

Type: String

Required: No

capacityReservations

The capacity reservation configuration for the managed instances. Use this to target On-Demand Capacity Reservations or Reserved Instances for predictable capacity and cost optimization.

Type: CapacityReservationRequest object

Required: No

fipsEnabled

Specifies whether FIPS 140-2 validated cryptographic modules are enabled on the managed instances. Not available in all Regions.

Type: Boolean

Required: No

instanceMetadataTagsPropagation

Specifies whether instance tags are accessible from the instance metadata service (IMDS). If not specified, instance tags are not accessible from IMDS.

Type: Boolean

Required: No

instanceRequirements

The instance type requirements for the capacity provider. Use this to constrain which Amazon EC2 instance types Amazon ECS can launch. If not specified, all available instance types are eligible.

Type: InstanceRequirementsRequest object

Required: No

localStorageConfiguration

The local storage configuration for the managed instances. If not specified, instance store volumes are not available to containers.

Type: ManagedInstancesLocalStorageConfiguration object

Required: No

monitoring

The level of CloudWatch monitoring for the managed instances. Valid values are `BASIC` and `DETAILED`.

Type: String

Required: No

storageConfiguration

The storage configuration for the managed instances. Configures the root EBS volume size. If not specified, the service uses the default EBS volume size for the instance type.

Type: ManagedInstancesStorageConfiguration 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 Java V2
- AWS SDK for Ruby V3

InstanceLaunchTemplateUpdate

The instance launch configuration for updating an Amazon ECS Managed Instances capacity provider. You cannot change `capacityOptionType` or `fipsEnabled` after the compute environment is created.

Contents

capacityReservations

The updated capacity reservation configuration.

Type: CapacityReservationRequest object

Required: No

ec2InstanceProfileArn

The updated Amazon Resource Name (ARN) of the Amazon EC2 instance profile for the managed instances.

Type: String

Required: No

instanceMetadataTagsPropagation

Specifies whether instance tags are accessible from the instance metadata service (IMDS).

Type: Boolean

Required: No

instanceRequirements

The updated instance type requirements for the capacity provider.

Type: InstanceRequirementsRequest object

Required: No

localStorageConfiguration

The updated local storage configuration.

Type: ManagedInstancesLocalStorageConfiguration object

Required: No

monitoring

The updated monitoring level. Valid values are `BASIC` and `DETAILED`.

Type: String

Required: No

networkConfiguration

The updated network configuration for the managed instances.

Type: ManagedInstancesNetworkConfiguration object

Required: No

storageConfiguration

The updated storage configuration for the managed instances.

Type: ManagedInstancesStorageConfiguration 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 Java V2
- AWS SDK for Ruby V3

InstanceRequirementsRequest

The instance type requirements for the Amazon ECS Managed Instances capacity provider. Use this to specify which Amazon EC2 instance types or instance families Amazon ECS can launch.

Contents

allowedInstanceTypes

A list of specific instance types or instance families that Amazon ECS can launch (for example, `m5.large` or `g5`). When specified, only these instance types are used.

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 Java V2
- AWS SDK for Ruby V3

JobCapacityUsageSummary

The capacity usage for a job, including the unit of measure and quantity of resources being used.

Contents

capacityUnit

The unit of measure for the capacity usage. This is `VCPU` for Amazon EC2 and `cpu` for Amazon EKS.

Type: String

Required: No

quantity

The quantity of capacity being used by the job, measured in the units specified by `capacityUnit`.

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 Java V2
- AWS SDK for Ruby V3

JobDefinition

An object that represents an AWS Batch job definition.

Contents

jobDefinitionArn

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

Type: String

Required: Yes

jobDefinitionName

The name of the job definition.

Type: String

Required: Yes

revision

The revision of the job definition.

Type: Integer

Required: Yes

type

The type of job definition. It's either `container` or `multinode`. If the job is run on Fargate resources, then `multinode` isn't supported. For more information about multi-node parallel jobs, see Creating a multi-node parallel job definition in the *AWS Batch User Guide*.

Type: String

Required: Yes

consumableResourceProperties

Contains a list of consumable resources required by the job.

Type: ConsumableResourceProperties object

Required: No

containerOrchestrationType

The orchestration type of the compute environment. The valid values are ECS (default) or EKS.

Type: String

Valid Values: ECS | EKS

Required: No

containerProperties

An object with properties specific to Amazon ECS-based jobs. When `containerProperties` is used in the job definition, it can't be used in addition to `eksProperties`, `ecsProperties`, or `nodeProperties`.

Type: ContainerProperties object

Required: No

ecsProperties

An object that contains the properties for the Amazon ECS resources of a job.When `ecsProperties` is used in the job definition, it can't be used in addition to `containerProperties`, `eksProperties`, or `nodeProperties`.

Type: EcsProperties object

Required: No

eksProperties

An object with properties that are specific to Amazon EKS-based jobs. When `eksProperties` is used in the job definition, it can't be used in addition to `containerProperties`, `ecsProperties`, or `nodeProperties`.

Type: EksProperties object

Required: No

nodeProperties

An object with properties that are specific to multi-node parallel jobs. When `nodeProperties` is used in the job definition, it can't be used in addition to `containerProperties`, `ecsProperties`, or `eksProperties`.

> **Note**
>
> If the job runs on Fargate resources, don't specify `nodeProperties`. We recommend `ecsProperties`, which supports both single-container and multi-container jobs. `containerProperties` is also supported for single-container jobs.

Type: NodeProperties object

Required: No

parameters

Default parameters or parameter substitution placeholders that are set in the job definition. Parameters are specified as a key-value pair mapping. Parameters in a `SubmitJob` request override any corresponding parameter defaults from the job definition. For more information about specifying parameters, see Job definition parameters in the *AWS Batch User Guide*.

Type: String to string map

Required: No

platformCapabilities

The platform capabilities required by the job definition. If no value is specified, it defaults to `EC2`. Jobs run on Fargate resources specify `FARGATE`. Jobs run on Amazon ECS Managed Instances specify `MANAGED_INSTANCES`.

Type: Array of strings

Valid Values: `EC2 | FARGATE | MANAGED_INSTANCES`

Required: No

propagateTags

Specifies whether to propagate the tags from the job or job definition to the corresponding Amazon ECS task. If no value is specified, the tags aren't propagated. Tags can only be propagated to the tasks when the tasks are created. For tags with the same name, job tags are given priority over job definitions tags. If the total number of combined tags from the job and job definition is over 50, the job is moved to the `FAILED` state.

Type: Boolean

Required: No

retryStrategy

The retry strategy to use for failed jobs that are submitted with this job definition.

Type: RetryStrategy object

Required: No

schedulingPriority

The scheduling priority of the job definition. This only affects jobs in job queues with a fair-share policy. Jobs with a higher scheduling priority are scheduled before jobs with a lower scheduling priority.

Type: Integer

Required: No

status

The status of the job definition.

Type: String

Required: No

tags

The tags that are applied to the job definition.

Type: String to string map

Map Entries: Maximum number of 50 items.

Key Length Constraints: Minimum length of 1. Maximum length of 128.

Value Length Constraints: Maximum length of 256.

Required: No

timeout

The timeout time for jobs that are submitted with this job definition. After the amount of time you specify passes, AWS Batch terminates your jobs if they aren't finished.

Type: JobTimeout 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 Java V2
- AWS SDK for Ruby V3

JobDependency

An object that represents an AWS Batch job dependency.

Contents

jobId

The job ID of the AWS Batch job that's associated with this dependency.

Type: String

Required: No

type

The type of the job dependency.

Type: String

Valid Values: `N_TO_N | SEQUENTIAL`

Required: No

See Also

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

- AWS SDK for C++
- AWS SDK for Java V2
- AWS SDK for Ruby V3

JobDetail

An object that represents an AWS Batch job.

Contents

jobDefinition

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

Type: String

Required: Yes

jobId

The job ID.

Type: String

Required: Yes

jobName

The job name.

Type: String

Required: Yes

jobQueue

The Amazon Resource Name (ARN) of the job queue that the job is associated with.

Type: String

Required: Yes

startedAt

The Unix timestamp (in milliseconds) for when the job was started. More specifically, it's when the job transitioned from the `STARTING` state to the `RUNNING` state.

Type: Long

Required: Yes

status

The current status for the job.

> **Note**
>
> If your jobs don't progress to `STARTING`, see Jobs stuck in RUNNABLE status in the troubleshooting section of the *AWS Batch User Guide*.

Type: String

Valid Values: `SUBMITTED | PENDING | RUNNABLE | STARTING | RUNNING | SUCCEEDED | FAILED`

Required: Yes

arrayProperties

The array properties of the job, if it's an array job.

Type: ArrayPropertiesDetail object

Required: No

attempts

A list of job attempts that are associated with this job.

Type: Array of AttemptDetail objects

Required: No

consumableResourceProperties

Contains a list of consumable resources required by the job.

Type: ConsumableResourceProperties object

Required: No

container

An object that represents the details for the container that's associated with the job. If the details are for a multiple-container job, this object will be empty.

Type: ContainerDetail object

Required: No

createdAt

The Unix timestamp (in milliseconds) for when the job was created. For non-array jobs and parent array jobs, this is when the job entered the `SUBMITTED` state. This is specifically at the time SubmitJob was called. For array child jobs, this is when the child job was spawned by its parent and entered the `PENDING` state.

Type: Long

Required: No

dependsOn

A list of job IDs that this job depends on.

Type: Array of JobDependency objects

Required: No

ecsProperties

An object with properties that are specific to Amazon ECS-based jobs.

Type: EcsPropertiesDetail object

Required: No

eksAttempts

A list of job attempts that are associated with this job.

Type: Array of EksAttemptDetail objects

Required: No

eksProperties

An object with various properties that are specific to Amazon EKS based jobs.

Type: EksPropertiesDetail object

Required: No

isCancelled

Indicates whether the job is cancelled.

Type: Boolean

Required: No

isTerminated

Indicates whether the job is terminated.

Type: Boolean

Required: No

jobArn

The Amazon Resource Name (ARN) of the job.

Type: String

Required: No

nodeDetails

An object that represents the details of a node that's associated with a multi-node parallel job.

Type: NodeDetails object

Required: No

nodeProperties

An object that represents the node properties of a multi-node parallel job.

> **Note**
>
> This isn't applicable to jobs that are running on Fargate resources.

Type: NodeProperties object

Required: No

parameters

Additional parameters that are passed to the job that replace parameter substitution placeholders or override any corresponding parameter defaults from the job definition.

Type: String to string map

Required: No

platformCapabilities

The platform capabilities required by the job definition. If no value is specified, it defaults to `EC2`. Jobs run on Fargate resources specify `FARGATE`. Jobs run on Amazon ECS Managed Instances specify `MANAGED_INSTANCES`.

Type: Array of strings

Valid Values: `EC2 | FARGATE | MANAGED_INSTANCES`

Required: No

propagateTags

Specifies whether to propagate the tags from the job or job definition to the corresponding Amazon ECS task. If no value is specified, the tags aren't propagated. Tags can only be propagated to the tasks when the tasks are created. For tags with the same name, job tags are given priority over job definitions tags. If the total number of combined tags from the job and job definition is over 50, the job is moved to the `FAILED` state.

Type: Boolean

Required: No

retryStrategy

The retry strategy to use for this job if an attempt fails.

Type: RetryStrategy object

Required: No

schedulingPriority

The scheduling policy of the job definition. This only affects jobs in job queues with a fair-share policy. Jobs with a higher scheduling priority are scheduled before jobs with a lower scheduling priority.

Type: Integer

Required: No

shareIdentifier

The share identifier for the job.

Type: String

Required: No

statusReason

A short, human-readable string to provide more details for the current status of the job.

- `CAPACITY:INSUFFICIENT_INSTANCE_CAPACITY` - All compute environments have insufficient capacity to service the job.
- `MISCONFIGURATION:COMPUTE_ENVIRONMENT_MAX_RESOURCE` - All compute environments have a `maxVcpu` setting that is smaller than the job requirements.
- `MISCONFIGURATION:JOB_RESOURCE_REQUIREMENT` - All compute environments have no connected instances that meet the job requirements.
- `MISCONFIGURATION:SERVICE_ROLE_PERMISSIONS` - All compute environments have problems with the service role permissions.

Type: String

Required: No

stoppedAt

The Unix timestamp (in milliseconds) for when the job was stopped. More specifically, it's when the job transitioned from the `RUNNING` state to a terminal state, such as `SUCCEEDED` or `FAILED`.

Type: Long

Required: No

tags

The tags that are applied to the job.

Type: String to string map

Map Entries: Maximum number of 50 items.

Key Length Constraints: Minimum length of 1. Maximum length of 128.

Value Length Constraints: Maximum length of 256.

Required: No

timeout

The timeout configuration for the job.

Type: JobTimeout 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 Java V2
- AWS SDK for Ruby V3

JobQueueDetail

An object that represents the details for an AWS Batch job queue.

Contents

computeEnvironmentOrder

The compute environments that are attached to the job queue and the order that job placement is preferred. Compute environments are selected for job placement in ascending order.

Type: Array of ComputeEnvironmentOrder objects

Required: Yes

jobQueueArn

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

Type: String

Required: Yes

jobQueueName

The job queue name.

Type: String

Required: Yes

priority

The priority of the job queue. Job queue priority determines the order that job queues are evaluated when multiple queues dispatch jobs within a shared compute environment. A higher value for `priority` indicates a higher priority. Queues are evaluated in cycles, in descending order by priority. For example, a job queue with a priority value of `10` is evaluated before a queue with a priority value of 1. All of the compute environments must be either Amazon EC2 (`EC2` or `SPOT`) or Fargate (`FARGATE` or `FARGATE_SPOT`). Amazon EC2 and Fargate compute environments can't be mixed.

> **Note**
>
> Job queue priority doesn't guarantee that a particular job executes before a job in a lower priority queue. Jobs added to higher priority queues during the queue evaluation cycle might not be evaluated until the next cycle. A job is dispatched from a queue only if resources are available when the queue is evaluated. If there are insufficient resources available at that time, the cycle proceeds to the next queue. This means that jobs added to higher priority queues might have to wait for jobs in multiple lower priority queues to complete before they are dispatched. You can use job dependencies to control the order for jobs from queues with different priorities. For more information, see Job Dependencies in the *AWS Batch User Guide*.

Type: Integer

Required: Yes

state

Describes the ability of the queue to accept new jobs. If the job queue state is `ENABLED`, it can accept jobs. If the job queue state is `DISABLED`, new jobs can't be added to the queue, but jobs already in the queue can finish.

Type: String

Valid Values: `ENABLED | DISABLED`

Required: Yes

jobQueueType

The type of job queue. For service jobs that run on SageMaker Training, this value is `SAGEMAKER_TRAINING`. For regular container jobs, this value is `EKS`, `ECS`, or `ECS_FARGATE` depending on the compute environment.

Type: String

Valid Values: `EKS | ECS | ECS_FARGATE | SAGEMAKER_TRAINING | ECS_MANAGED_INSTANCES`

Required: No

jobStateTimeLimitActions

The set of actions that AWS Batch perform on jobs that remain at the head of the job queue in the specified state longer than specified times. AWS Batch will perform each action after `maxTimeSeconds` has passed.

Type: Array of JobStateTimeLimitAction objects

Required: No

schedulingPolicyArn

The Amazon Resource Name (ARN) of the scheduling policy. The format is `aws:Partition:batch:Region:Account:scheduling-policy/Name` . For example, `aws:aws:batch:us-west-2:123456789012:scheduling-policy/MySchedulingPolicy`.

Type: String

Required: No

serviceEnvironmentOrder

The order of the service environment associated with the job queue. Job queues with a higher priority are evaluated first when associated with the same service environment.

Type: Array of ServiceEnvironmentOrder objects

Required: No

status

The status of the job queue (for example, `CREATING` or `VALID`).

Type: String

Valid Values: `CREATING | UPDATING | DELETING | DELETED | VALID | INVALID`

Required: No

statusReason

A short, human-readable string to provide additional details for the current status of the job queue.

Type: String

Required: No

tags

The tags that are applied to the job queue. For more information, see Tagging your AWS Batch resources in *AWS Batch User Guide*.

Type: String to string map

Map Entries: Maximum number of 50 items.

Key Length Constraints: Minimum length of 1. Maximum length of 128.

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 Java V2
- AWS SDK for Ruby V3

JobStateTimeLimitAction

Specifies an action that AWS Batch will take after the job has remained at the head of the queue in the specified state for longer than the specified time.

Contents

action

The action to take when a job is at the head of the job queue in the specified state for the specified period of time. For job queues connected to a `ECS`, `FARGATE` or `EKS` compute environment, the only supported value is `CANCEL`, which will cancel the job. For job queues connected to a `SAGEMAKER_TRAINING` service environment, the only supported value is `TERMINATE`, which will terminate the job.

Type: String

Valid Values: `CANCEL | TERMINATE`

Required: Yes

maxTimeSeconds

The approximate amount of time, in seconds, that must pass with the job in the specified state before the action is taken. The minimum value is 600 (10 minutes) and the maximum value is 86,400 (24 hours).

Type: Integer

Required: Yes

reason

The reason to log for the action being taken.

Type: String

Required: Yes

state

The state of the job needed to trigger the action. The only supported value is `RUNNABLE`.

Type: String

Valid Values: RUNNABLE

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 Java V2
- AWS SDK for Ruby V3

JobSummary

An object that represents summary details of a job.

Contents

jobId

The job ID.

Type: String

Required: Yes

jobName

The job name.

Type: String

Required: Yes

arrayProperties

The array properties of the job, if it's an array job.

Type: ArrayPropertiesSummary object

Required: No

capacityUsage

The configured capacity usage information for this job, including the unit of measure and quantity of resources.

Type: Array of JobCapacityUsageSummary objects

Required: No

container

An object that represents the details of the container that's associated with the job.

Type: ContainerSummary object

Required: No

createdAt

The Unix timestamp (in milliseconds) for when the job was created. For non-array jobs and parent array jobs, this is when the job entered the `SUBMITTED` state (at the time SubmitJob was called). For array child jobs, this is when the child job was spawned by its parent and entered the `PENDING` state.

Type: Long

Required: No

isCancelled

Indicates whether a cancellation request has been accepted for the job. This field is only present when the value is `true`.

Type: Boolean

Required: No

isTerminated

Indicates whether a termination request has been accepted for the job. This field is only present when the value is `true`.

Type: Boolean

Required: No

jobArn

The Amazon Resource Name (ARN) of the job.

Type: String

Required: No

jobDefinition

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

Type: String

Required: No

nodeProperties

The node properties for a single node in a job summary list.

> **Note**
>
> This isn't applicable to jobs that are running on Fargate resources.

Type: NodePropertiesSummary object

Required: No

scheduledAt

The Unix timestamp (in milliseconds) for when the job was scheduled for execution. For more information on job statues, see Service job status in the *AWS Batch User Guide*.

Type: Long

Required: No

shareIdentifier

The share identifier for the fairshare scheduling queue that this job is associated with.

Type: String

Required: No

startedAt

The Unix timestamp for when the job was started. More specifically, it's when the job transitioned from the `STARTING` state to the `RUNNING` state.

Type: Long

Required: No

status

The current status for the job.

Type: String

Valid Values: `SUBMITTED | PENDING | RUNNABLE | STARTING | RUNNING | SUCCEEDED | FAILED`

Required: No

statusReason

A short, human-readable string to provide more details for the current status of the job.

Type: String

Required: No

stoppedAt

The Unix timestamp for when the job was stopped. More specifically, it's when the job transitioned from the `RUNNING` state to a terminal state, such as `SUCCEEDED` or `FAILED`.

Type: Long

Required: No

See Also

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

- AWS SDK for C++
- AWS SDK for Java V2
- AWS SDK for Ruby V3

JobTimeout

An object that represents a job timeout configuration.

Contents

attemptDurationSeconds

The job timeout time (in seconds) that's measured from the job attempt's `startedAt` timestamp. After this time passes, AWS Batch terminates your jobs if they aren't finished. The minimum value for the timeout is 60 seconds.

For array jobs, the timeout applies to the child jobs, not to the parent array job.

For multi-node parallel (MNP) jobs, the timeout applies to the whole job, not to the individual nodes.

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 Java V2
- AWS SDK for Ruby V3

KeyValuePair

A key-value pair object.

Contents

name

The name of the key-value pair. For environment variables, this is the name of the environment variable.

Type: String

Required: No

value

The value of the key-value pair. For environment variables, this is the value of the environment variable.

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 Java V2
- AWS SDK for Ruby V3

KeyValuesPair

A filter name and value pair that's used to return a more specific list of results from a `ListJobs` or `ListJobsByConsumableResource` API operation.

The filters supported are documented in the `ListJobs` API operation.

Contents

name

The name of the filter. Filter names are case sensitive.

Type: String

Required: No

values

The filter values.

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 Java V2
- AWS SDK for Ruby V3

LatestServiceJobAttempt

Information about the latest attempt of a service job. A Service job can transition from SCHEDULED back to RUNNABLE state when they encounter capacity constraints.

Contents

serviceResourceId

The service resource identifier associated with the service job attempt.

Type: ServiceResourceId 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 Java V2
- AWS SDK for Ruby V3

LaunchTemplateSpecification

An object that represents a launch template that's associated with a compute resource. You must specify either the launch template ID or launch template name in the request, but not both.

If security groups are specified using both the `securityGroupIds` parameter of `CreateComputeEnvironment` and the launch template, the values in the `securityGroupIds` parameter of `CreateComputeEnvironment` will be used.

> **Note**
>
> This object isn't applicable to jobs that are running on Fargate resources.

Contents

launchTemplateId

The ID of the launch template.

Type: String

Required: No

launchTemplateName

The name of the launch template.

Type: String

Required: No

overrides

A launch template to use in place of the default launch template. You must specify either the launch template ID or launch template name in the request, but not both.

You can specify up to ten (10) launch template overrides that are associated to unique instance types or families for each compute environment.

> **Note**
>
> To unset all override templates for a compute environment, you can pass an empty array to the UpdateComputeEnvironment.overrides parameter, or not include the `overrides` parameter when submitting the `UpdateComputeEnvironment` API operation.

Type: Array of LaunchTemplateSpecificationOverride objects

Required: No

userdataType

The EKS node initialization process to use. You only need to specify this value if you are using a custom AMI. The default value is `EKS_BOOTSTRAP_SH`. If *imageType* is a custom AMI based on EKS_AL2023 or EKS_AL2023_NVIDIA then you must choose `EKS_NODEADM`.

Type: String

Valid Values: `EKS_BOOTSTRAP_SH | EKS_NODEADM`

Required: No

version

The version number of the launch template, `$Default`, or `$Latest`.

If the value is `$Default`, the default version of the launch template is used. If the value is `$Latest`, the latest version of the launch template is used.

> **Important**
>
> If the AMI ID that's used in a compute environment is from the launch template, the AMI isn't changed when the compute environment is updated. It's only changed if the `updateToLatestImageVersion` parameter for the compute environment is set to `true`. During an infrastructure update, if either `$Default` or `$Latest` is specified, AWS Batch re-evaluates the launch template version, and it might use a different version of the launch template. This is the case even if the launch template isn't specified in the update. When updating a compute environment, changing the launch template

requires an infrastructure update of the compute environment. For more information, see Updating compute environments in the *AWS Batch User Guide*.

Default: `$Default`

Latest: `$Latest`

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 Java V2
- AWS SDK for Ruby V3

LaunchTemplateSpecificationOverride

An object that represents a launch template to use in place of the default launch template. You must specify either the launch template ID or launch template name in the request, but not both.

If security groups are specified using both the `securityGroupIds` parameter of `CreateComputeEnvironment` and the launch template, the values in the `securityGroupIds` parameter of `CreateComputeEnvironment` will be used.

You can define up to ten (10) overrides for each compute environment.

> **Note**
>
> This object isn't applicable to jobs that are running on Fargate resources.

> **Note**
>
> To unset all override templates for a compute environment, you can pass an empty array to the UpdateComputeEnvironment.overrides parameter, or not include the `overrides` parameter when submitting the `UpdateComputeEnvironment` API operation.

Contents

launchTemplateId

The ID of the launch template.

Note: If you specify the `launchTemplateId` you can't specify the `launchTemplateName` as well.

Type: String

Required: No

launchTemplateName

The name of the launch template.

Note: If you specify the `launchTemplateName` you can't specify the `launchTemplateId` as well.

Type: String

Required: No

targetInstanceTypes

The instance type or family that this override launch template should be applied to.

This parameter is required when defining a launch template override.

Information included in this parameter must meet the following requirements:

- Must be a valid Amazon EC2 instance type or family.
- The following AWS Batch `InstanceTypes` are not allowed: `optimal`, `default_x86_64`, and `default_arm64`.
- `targetInstanceTypes` can target only instance types and families that are included within the `ComputeResource.instanceTypes` set. `targetInstanceTypes` doesn't need to include all of the instances from the `instanceType` set, but at least a subset. For example, if `ComputeResource.instanceTypes` includes `[m5, g5]`, `targetInstanceTypes` can include `[m5.2xlarge]` and `[m5.large]` but not `[c5.large]`.
- `targetInstanceTypes` included within the same launch template override or across launch template overrides can't overlap for the same compute environment. For example, you can't define one launch template override to target an instance family and another define an instance type within this same family.

Type: Array of strings

Required: No

userdataType

The EKS node initialization process to use. You only need to specify this value if you are using a custom AMI. The default value is EKS_BOOTSTRAP_SH. If *imageType* is a custom AMI based on EKS_AL2023 or EKS_AL2023_NVIDIA then you must choose EKS_NODEADM.

Type: String

Valid Values: `EKS_BOOTSTRAP_SH | EKS_NODEADM`

Required: No

version

The version number of the launch template, `$Default`, or `$Latest`.

If the value is `$Default`, the default version of the launch template is used. If the value is `$Latest`, the latest version of the launch template is used.

> **Important**
>
> If the AMI ID that's used in a compute environment is from the launch template, the AMI isn't changed when the compute environment is updated. It's only changed if the `updateToLatestImageVersion` parameter for the compute environment is set to `true`. During an infrastructure update, if either `$Default` or `$Latest` is specified, AWS Batch re-evaluates the launch template version, and it might use a different version of the launch template. This is the case even if the launch template isn't specified in the update. When updating a compute environment, changing the launch template requires an infrastructure update of the compute environment. For more information, see Updating compute environments in the *AWS Batch User Guide*.

Default: `$Default`

Latest: `$Latest`

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 Java V2
- AWS SDK for Ruby V3

LinuxParameters

Linux-specific modifications that are applied to the container, such as details for device mappings.

Contents

devices

Any of the host devices to expose to the container. This parameter maps to `Devices` in the Create a container section of the Docker Remote API and the `--device` option to docker run.

> **Note**
>
> This parameter isn't applicable to jobs that are running on Fargate resources. Don't provide it for these jobs.

Type: Array of Device objects

Required: No

initProcessEnabled

If true, run an `init` process inside the container that forwards signals and reaps processes. This parameter maps to the `--init` option to docker run. This parameter requires version 1.25 of the Docker Remote API or greater on your container instance. To check the Docker Remote API version on your container instance, log in to your container instance and run the following command: `sudo docker version | grep "Server API version"`

Type: Boolean

Required: No

maxSwap

The total amount of swap memory (in MiB) a container can use. This parameter is translated to the `--memory-swap` option to docker run where the value is the sum of the container memory plus the `maxSwap` value. For more information, see `--memory-swap` details in the Docker documentation.

If a `maxSwap` value of `0` is specified, the container doesn't use swap. Accepted values are `0` or any positive integer. If the `maxSwap` parameter is omitted, the container doesn't use the swap

configuration for the container instance on which it runs. A `maxSwap` value must be set for the `swappiness` parameter to be used.

> **Note**
>
> This parameter isn't applicable to jobs that are running on Fargate resources. Don't provide it for these jobs.

Type: Integer

Required: No

sharedMemorySize

The value for the size (in MiB) of the `/dev/shm` volume. This parameter maps to the `--shm-size` option to docker run.

> **Note**
>
> This parameter isn't applicable to jobs that are running on Fargate resources. Don't provide it for these jobs.

Type: Integer

Required: No

swappiness

You can use this parameter to tune a container's memory swappiness behavior. A `swappiness` value of `0` causes swapping to not occur unless absolutely necessary. A `swappiness` value of `100` causes pages to be swapped aggressively. Valid values are whole numbers between `0` and `100`. If the `swappiness` parameter isn't specified, a default value of `60` is used. If a value isn't specified for `maxSwap`, then this parameter is ignored. If `maxSwap` is set to 0, the container doesn't use swap. This parameter maps to the `--memory-swappiness` option to docker run.

Consider the following when you use a per-container swap configuration.

- Swap space must be enabled and allocated on the container instance for the containers to use.

> **Note**
>
> By default, the Amazon ECS optimized AMIs don't have swap enabled. You must enable swap on the instance to use this feature. For more information, see Instance store swap volumes in the *Amazon EC2 User Guide for Linux Instances* or How do I allocate memory to work as swap space in an Amazon EC2 instance by using a swap file?

- The swap space parameters are only supported for job definitions using EC2 resources.
- If the `maxSwap` and `swappiness` parameters are omitted from a job definition, each container has a default `swappiness` value of 60. Moreover, the total swap usage is limited to two times the memory reservation of the container.

> **Note**
>
> This parameter isn't applicable to jobs that are running on Fargate resources. Don't provide it for these jobs.

Type: Integer

Required: No

tmpfs

The container path, mount options, and size (in MiB) of the `tmpfs` mount. This parameter maps to the `--tmpfs` option to docker run.

> **Note**
>
> This parameter isn't applicable to jobs that are running on Fargate resources. Don't provide this parameter for this resource type.

Type: Array of Tmpfs 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 Java V2
- AWS SDK for Ruby V3

ListJobsByConsumableResourceSummary

Current information about a consumable resource required by a job.

Contents

consumableResourceProperties

Contains a list of consumable resources required by the job.

Type: ConsumableResourceProperties object

Required: Yes

createdAt

The Unix timestamp (in milliseconds) for when the consumable resource was created.

Type: Long

Required: Yes

jobArn

The Amazon Resource Name (ARN) of the job.

Type: String

Required: Yes

jobName

The name of the job.

Type: String

Required: Yes

jobQueueArn

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

Type: String

Required: Yes

jobStatus

The status of the job. Can be one of:

- SUBMITTED
- PENDING
- RUNNABLE
- STARTING
- RUNNING
- SUCCEEDED
- FAILED

Type: String

Required: Yes

quantity

The total amount of the consumable resource that is available.

Type: Long

Required: Yes

jobDefinitionArn

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

Type: String

Required: No

shareIdentifier

The fair-share scheduling identifier for the job.

Type: String

Required: No

startedAt

The Unix timestamp for when the job was started. More specifically, it's when the job transitioned from the STARTING state to the RUNNING state.

Type: Long

Required: No

statusReason

A short, human-readable string to provide more details for the current status of the job.

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 Java V2
- AWS SDK for Ruby V3

LogConfiguration

Log configuration options to send to a custom log driver for the container.

Contents

logDriver

The log driver to use for the container. The valid values that are listed for this parameter are log drivers that the Amazon ECS container agent can communicate with by default.

The supported log drivers are `awsfirelens`, `awslogs`, `fluentd`, `gelf`, `json-file`, `journald`, `logentries`, `syslog`, and `splunk`.

> **Note**
>
> Jobs that are running on Fargate resources are restricted to the `awslogs` and `splunk` log drivers.

awsfirelens

Specifies the firelens logging driver. For more information on configuring Firelens, see Send Amazon ECS logs to an AWS service or AWS Partner in the *Amazon Elastic Container Service Developer Guide*.

awslogs

Specifies the Amazon CloudWatch Logs logging driver. For more information, see Using the awslogs log driver in the *AWS Batch User Guide* and Amazon CloudWatch Logs logging driver in the Docker documentation.

fluentd

Specifies the Fluentd logging driver. For more information including usage and options, see Fluentd logging driver in the *Docker documentation*.

gelf

Specifies the Graylog Extended Format (GELF) logging driver. For more information including usage and options, see Graylog Extended Format logging driver in the *Docker documentation*.

journald

Specifies the journald logging driver. For more information including usage and options, see Journald logging driver in the *Docker documentation*.

json-file

Specifies the JSON file logging driver. For more information including usage and options, see JSON File logging driver in the *Docker documentation*.

splunk

Specifies the Splunk logging driver. For more information including usage and options, see Splunk logging driver in the *Docker documentation*.

syslog

Specifies the syslog logging driver. For more information including usage and options, see Syslog logging driver in the *Docker documentation*.

Note

If you have a custom driver that's not listed earlier that you want to work with the Amazon ECS container agent, you can fork the Amazon ECS container agent project that's available on GitHub and customize it to work with that driver. We encourage you to submit pull requests for changes that you want to have included. However, Amazon Web Services doesn't currently support running modified copies of this software.

This parameter requires version 1.18 of the Docker Remote API or greater on your container instance. To check the Docker Remote API version on your container instance, log in to your container instance and run the following command: `sudo docker version | grep "Server API version"`

Type: String

Valid Values: `json-file | syslog | journald | gelf | fluentd | awslogs | splunk | awsfirelens`

Required: Yes

options

The configuration options to send to the log driver. This parameter requires version 1.19 of the Docker Remote API or greater on your container instance. To check the Docker Remote API version on your container instance, log in to your container instance and run the following command: `sudo docker version | grep "Server API version"`

Type: String to string map

Required: No

secretOptions

The secrets to pass to the log configuration. For more information, see Specifying sensitive data in the *AWS Batch User Guide*.

Type: Array of Secret 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 Java V2
- AWS SDK for Ruby V3

ManagedInstancesLocalStorageConfiguration

The local storage configuration for Amazon ECS Managed Instances.

Contents

useLocalStorage

Specifies whether instance store volumes (local NVMe SSDs) are available to containers. When enabled, containers can use the instance store for high-performance temporary storage.

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 Java V2
- AWS SDK for Ruby V3

ManagedInstancesNetworkConfiguration

The network configuration for Amazon ECS Managed Instances. Specifies the VPC subnets and security groups where instances are launched.

Contents

securityGroups

The VPC security groups to associate with the managed instances.

Type: Array of strings

Required: Yes

subnets

The VPC subnets where managed instances are launched. If your subnets don't provide public IP addresses, they must have a NAT gateway for outbound internet access.

Type: Array of strings

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 Java V2
- AWS SDK for Ruby V3

ManagedInstancesProvider

The configuration for an Amazon ECS Managed Instances capacity provider. This object is required when creating a compute environment with `computeResources.type` set to `ECS_MANAGED_INSTANCES`.

Contents

infrastructureRoleArn

The Amazon Resource Name (ARN) of the IAM role that Amazon ECS assumes to manage Amazon EC2 instances on your behalf. This role must have a trust policy for `ecs.amazonaws.com`. You must have the `iam:PassRole` permission for this role with the condition `iam:PassedToService: ecs.amazonaws.com`.

Type: String

Required: Yes

instanceLaunchTemplate

The instance launch configuration for the Amazon ECS Managed Instances capacity provider. Contains networking, instance profile, instance requirements, capacity type, storage, and monitoring configuration.

Type: InstanceLaunchTemplate object

Required: Yes

infrastructureOptimization

The infrastructure optimization configuration for the capacity provider. Specifies the idle-instance scale-in behavior.

Type: InfrastructureOptimization object

Required: No

propagateTags

Specifies whether tags on the capacity provider are propagated to the Amazon EC2 instances it launches. Valid values:

- `CAPACITY_PROVIDER` — Propagates tags to instances.

- NONE (default) — Does not propagate tags to instances.

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 Java V2
- AWS SDK for Ruby V3

ManagedInstancesStorageConfiguration

The storage configuration for Amazon ECS Managed Instances.

Contents

storageSizeGiB

The size of the root EBS volume in GiB for the managed instances.

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 Java V2
- AWS SDK for Ruby V3

MountPoint

Details for a Docker volume mount point that's used in a job's container properties. This parameter maps to `Volumes` in the Create a container section of the *Docker Remote API* and the `--volume` option to docker run.

Contents

containerPath

The path on the container where the host volume is mounted.

Type: String

Required: No

readOnly

If this value is `true`, the container has read-only access to the volume. Otherwise, the container can write to the volume. The default value is `false`.

Type: Boolean

Required: No

sourceVolume

The name of the volume to mount.

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 Java V2
- AWS SDK for Ruby V3

NetworkConfiguration

The network configuration for jobs that are running on Fargate resources. Jobs that are running on Amazon EC2 resources must not specify this parameter.

Contents

assignPublicIp

Indicates whether the job has a public IP address. For a job that's running on Fargate resources in a private subnet to send outbound traffic to the internet (for example, to pull container images), the private subnet requires a NAT gateway be attached to route requests to the internet. For more information, see Amazon ECS task networking in the *Amazon Elastic Container Service Developer Guide*. The default value is "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 Java V2
- AWS SDK for Ruby V3

NetworkInterface

An object that represents the elastic network interface for a multi-node parallel job node.

Contents

attachmentId

The attachment ID for the network interface.

Type: String

Required: No

ipv6Address

The private IPv6 address for the network interface.

Type: String

Required: No

privateIpv4Address

The private IPv4 address for the network interface.

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 Java V2
- AWS SDK for Ruby V3

NodeDetails

An object that represents the details of a multi-node parallel job node.

Contents

isMainNode

Specifies whether the current node is the main node for a multi-node parallel job.

Type: Boolean

Required: No

nodeIndex

The node index for the node. Node index numbering starts at zero. This index is also available on the node with the `AWS_BATCH_JOB_NODE_INDEX` environment variable.

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 Java V2
- AWS SDK for Ruby V3

NodeOverrides

An object that represents any node overrides to a job definition that's used in a SubmitJob API operation.

> **Note**
>
> This parameter isn't applicable to jobs that are running on Fargate resources. Don't provide it for these jobs. Rather, use `containerOverrides` instead.

Contents

nodePropertyOverrides

The node property overrides for the job.

Type: Array of NodePropertyOverride objects

Required: No

numNodes

The number of nodes to use with a multi-node parallel job. This value overrides the number of nodes that are specified in the job definition. To use this override, you must meet the following conditions:

- There must be at least one node range in your job definition that has an open upper boundary, such as `:` or `n:`.
- The lower boundary of the node range that's specified in the job definition must be fewer than the number of nodes specified in the override.
- The main node index that's specified in the job definition must be fewer than the number of nodes specified in the override.

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 Java V2
- AWS SDK for Ruby V3

NodeProperties

An object that represents the node properties of a multi-node parallel job.

> **Note**
>
> Node properties can't be specified for Amazon EKS based job definitions.

Contents

mainNode

Specifies the node index for the main node of a multi-node parallel job. This node index value must be fewer than the number of nodes.

Type: Integer

Required: Yes

nodeRangeProperties

A list of node ranges and their properties that are associated with a multi-node parallel job.

Type: Array of NodeRangeProperty objects

Required: Yes

numNodes

The number of nodes that are associated with a multi-node parallel job.

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 Java V2
- AWS SDK for Ruby V3

NodePropertiesSummary

An object that represents the properties of a node that's associated with a multi-node parallel job.

Contents

isMainNode

Specifies whether the current node is the main node for a multi-node parallel job.

Type: Boolean

Required: No

nodeIndex

The node index for the node. Node index numbering begins at zero. This index is also available on the node with the `AWS_BATCH_JOB_NODE_INDEX` environment variable.

Type: Integer

Required: No

numNodes

The number of nodes that are associated with a multi-node parallel 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 Java V2
- AWS SDK for Ruby V3

NodePropertyOverride

The object that represents any node overrides to a job definition that's used in a SubmitJob API operation.

Contents

targetNodes

The range of nodes, using node index values, that's used to override. A range of `0:3` indicates nodes with index values of `0` through `3`. If the starting range value is omitted (`:n`), then `0` is used to start the range. If the ending range value is omitted (`n:`), then the highest possible node index is used to end the range.

Type: String

Required: Yes

consumableResourcePropertiesOverride

An object that contains overrides for the consumable resources of a job.

Type: ConsumableResourceProperties object

Required: No

containerOverrides

The overrides that are sent to a node range.

Type: ContainerOverrides object

Required: No

ecsPropertiesOverride

An object that contains the properties that you want to replace for the existing Amazon ECS resources of a job.

Type: EcsPropertiesOverride object

Required: No

eksPropertiesOverride

An object that contains the properties that you want to replace for the existing Amazon EKS resources of a job.

Type: EksPropertiesOverride object

Required: No

instanceTypes

An object that contains the instance types that you want to replace for the existing resources of a job.

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 Java V2
- AWS SDK for Ruby V3

NodeRangeProperty

This is an object that represents the properties of the node range for a multi-node parallel job.

Contents

targetNodes

The range of nodes, using node index values. A range of `0:3` indicates nodes with index values of `0` through `3`. If the starting range value is omitted (`:n`), then `0` is used to start the range. If the ending range value is omitted (`n:`), then the highest possible node index is used to end the range. Your accumulative node ranges must account for all nodes (`0:n`). You can nest node ranges (for example, `0:10` and `4:5`). In this case, the `4:5` range properties override the `0:10` properties.

Type: String

Required: Yes

consumableResourceProperties

Contains a list of consumable resources required by a job.

Type: ConsumableResourceProperties object

Required: No

container

The container details for the node range.

Type: ContainerProperties object

Required: No

ecsProperties

This is an object that represents the properties of the node range for a multi-node parallel job.

Type: EcsProperties object

Required: No

eksProperties

This is an object that represents the properties of the node range for a multi-node parallel job.

Type: EksProperties object

Required: No

instanceTypes

The instance types of the underlying host infrastructure of a multi-node parallel job.

> **Note**
>
> This parameter isn't applicable to jobs that are running on Fargate resources.
> In addition, this list object is currently limited to one element.

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 Java V2
- AWS SDK for Ruby V3

QueueSnapshotCapacityUsage

The configured capacity usage for a job queue snapshot, including the unit of measure and quantity of resources being used.

Contents

capacityUnit

The unit of measure for the capacity usage. For compute jobs, this is `VCPU` for Amazon EC2 and `cpu` for Amazon EKS. For service jobs, this is the instance type.

Type: String

Required: No

quantity

The quantity of capacity being used in the queue snapshot, measured in the units specified by `capacityUnit`.

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 Java V2
- AWS SDK for Ruby V3

QueueSnapshotUtilizationDetail

The job queue utilization at a specific point in time, including total capacity usage, and quota share or fairshare utilization breakdown depending on the job queue scheduling policy.

Contents

fairshareUtilization

The utilization information for a fairshare scheduling job queues, including active share count and top capacity utilization by share.

Type: FairshareUtilizationDetail object

Required: No

lastUpdatedAt

The Unix timestamp (in milliseconds) for when the queue utilization information was last updated.

Type: Long

Required: No

quotaShareUtilization

The utilization information for a job queue with a quota share scheduling policy.

Type: QuotaShareUtilizationDetail object

Required: No

totalCapacityUsage

The total capacity usage for the entire job queue.

Type: Array of QueueSnapshotCapacityUsage 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 Java V2
- AWS SDK for Ruby V3

QuotaShareCapacityLimit

Defines the capacity limit for a quota share, or the type and maximum quantity of a particular resource that can be allocated to jobs in the quota share without borrowing.

Contents

capacityUnit

The unit of compute capacity for the capacityLimit. For example, `ml.m5.large`.

Type: String

Required: Yes

maxCapacity

The maximum capacity available for the quota share. This value represents the maximum quantity of a resource that can be allocated to jobs in the quota share without borrowing.

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 Java V2
- AWS SDK for Ruby V3

QuotaShareCapacityUsage

The capacity usage for a quota share, including units of compute capacity and quantity of resources being used.

Contents

capacityUnit

The unit of compute capacity for the capacity usage.

Type: String

Required: No

quantity

The quantity of capacity being used.

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 Java V2
- AWS SDK for Ruby V3

QuotaShareCapacityUtilization

The capacity utilization for a specific quota share, including the quota share name and its current usage.

Contents

capacityUsage

The capacity usage information for this quota share, including the units of compute capacity and quantity being used.

Type: Array of QuotaShareCapacityUsage objects

Required: No

quotaShareName

The name of the quota share.

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 Java V2
- AWS SDK for Ruby V3

QuotaShareDetail

Detailed information about a quota share, including its configuration, state, and capacity limits.

Contents

capacityLimits

A list that specifies the quantity and type of compute capacity allocated to the quota share.

Type: Array of QuotaShareCapacityLimit objects

Required: No

jobQueueArn

The Amazon Resource Name (ARN) of the job queue associated with the quota share.

Type: String

Required: No

preemptionConfiguration

Specifies the preemption behavior for jobs in a quota share.

Type: QuotaSharePreemptionConfiguration object

Required: No

quotaShareArn

The Amazon Resource Name (ARN) of the quota share.

Type: String

Required: No

quotaShareName

The name of the quota share.

Type: String

Required: No

resourceSharingConfiguration

Specifies whether a quota share reserves, lends, or both lends and borrows idle compute capacity.

Type: QuotaShareResourceSharingConfiguration object

Required: No

state

The state of the quota share.

Type: String

Valid Values: `ENABLED | DISABLED`

Required: No

status

The current status of the quota share.

Type: String

Valid Values: `CREATING | VALID | INVALID | UPDATING | 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 Java V2
- AWS SDK for Ruby V3

QuotaSharePolicy

The quota share scheduling policy details for a job queue.

Contents

idleResourceAssignmentStrategy

The strategy that determines how idle resources are assigned to quota shares that are borrowing capacity. Currently, only `FIFO` is supported.

Type: String

Valid Values: `FIFO`

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 Java V2
- AWS SDK for Ruby V3

QuotaSharePreemptionConfiguration

Specifies the preemption behavior for jobs in a quota share.

Contents

inSharePreemption

Specifies whether jobs within a quota share can be preempted by another, higher priority job in the same quota share.

Type: String

Valid Values: `ENABLED | DISABLED`

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 Java V2
- AWS SDK for Ruby V3

QuotaShareResourceSharingConfiguration

Specifies whether a quota share reserves, lends, or both lends and borrows idle compute capacity.

Contents

strategy

The resource sharing strategy for the quota share. The `RESERVE` strategy allows a quota share to reserve idle capacity for itself. `LEND` configures the share to lend its idle capacity to another share in need of capacity. The `LEND_AND_BORROW` strategy configures the share to borrow idle capacity from an underutilized share, as well as lend to another share.

Type: String

Valid Values: `RESERVE | LEND | LEND_AND_BORROW`

Required: Yes

borrowLimit

The maximum percentage of additional capacity that the quota share can borrow from other shares. `borrowLimit` can only be applied to quota shares with a strategy of `LEND_AND_BORROW`. This value is expressed as a percentage of the quota share's configured CapacityLimits.

The `borrowLimit` is applied uniformly across all capacity units. For example, if the `borrowLimit` is 200, the quota share can borrow up to 200% of its configured `maxCapacity` for each capacity unit. The default `borrowLimit` is -1, which indicates unlimited borrowing.

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 Java V2
- AWS SDK for Ruby V3

QuotaShareUtilizationDetail

An object that represents the capacity utilization details of all quota shares associated with a single job queue.

Contents

topCapacityUtilization

A list of the top capacity utilizations across quota shares associated with a job queue.

Type: Array of QuotaShareCapacityUtilization 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 Java V2
- AWS SDK for Ruby V3

RepositoryCredentials

The repository credentials for private registry authentication.

Contents

credentialsParameter

The Amazon Resource Name (ARN) of the secret containing the private repository credentials.

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 Java V2
- AWS SDK for Ruby V3

ResourceRequirement

The type and amount of a resource to assign to a container. The supported resources include `GPU`, `MEMORY`, and `VCPU`.

Contents

type

The type of resource to assign to a container. The supported resources include `GPU`, `MEMORY`, and `VCPU`.

Type: String

Valid Values: `GPU` | `VCPU` | `MEMORY`

Required: Yes

value

The quantity of the specified resource to reserve for the container. The values vary based on the `type` specified.

type="GPU"

The number of physical GPUs to reserve for the container. Make sure that the number of GPUs reserved for all containers in a job doesn't exceed the number of available GPUs on the compute resource that the job is launched on.

> **Note**
>
> GPUs aren't available for jobs that are running on Fargate resources.

type="MEMORY"

The memory hard limit (in MiB) present to the container. This parameter is supported for jobs that are running on Amazon EC2 resources. If your container attempts to exceed the memory specified, the container is terminated. This parameter maps to `Memory` in the Create a container section of the Docker Remote API and the `--memory` option to

docker run. You must specify at least 4 MiB of memory for a job. This is required but can be specified in several places for multi-node parallel (MNP) jobs. It must be specified for each node at least once. This parameter maps to `Memory` in the Create a container section of the Docker Remote API and the `--memory` option to docker run.

> **Note**
>
> If you're trying to maximize your resource utilization by providing your jobs as much memory as possible for a particular instance type, see Memory management in the *AWS Batch User Guide*.

For jobs that are running on Fargate resources, then `value` is the hard limit (in MiB), and must match one of the supported values and the `VCPU` values must be one of the values supported for that memory value.

value = 512

VCPU = 0.25

value = 1024

VCPU = 0.25 or 0.5

value = 2048

VCPU = 0.25, 0.5, or 1

value = 3072

VCPU = 0.5, or 1

value = 4096

VCPU = 0.5, 1, or 2

value = 5120, 6144, or 7168

VCPU = 1 or 2

value = 8192

VCPU = 1, 2, or 4

value = 9216, 10240, 11264, 12288, 13312, 14336, or 15360

VCPU = 2 or 4

value = 16384

VCPU = 2, 4, or 8

value = 17408, 18432, 19456, 21504, 22528, 23552, 25600, 26624, 27648, 29696, or 30720

VCPU = 4

value = 20480, 24576, or 28672

VCPU = 4 or 8

value = 36864, 45056, or 53248

VCPU = 8

value = 61440

VCPU = 8 or 32

value = 32768, 40960, 49152, or 57344

VCPU = 8 or 16

value = 65536, 73728, 81920, 90112, 98304, 106496, or 114688

VCPU = 16

value = 122880

VCPU = 16 or 32

value = 249856

VCPU = 32

type="VCPU"

The number of vCPUs reserved for the container. This parameter maps to `CpuShares` in the Create a container section of the Docker Remote API and the `--cpu-shares` option to docker run. Each vCPU is equivalent to 1,024 CPU shares. For Amazon EC2 resources, you must specify at least one vCPU. This is required but can be specified in several places; it must be specified for each node at least once.

The default for the Fargate On-Demand vCPU resource count quota is 6 vCPUs. For more information about Fargate quotas, see AWS Fargate quotas in the *AWS General Reference*.

For jobs that are running on Fargate resources, then `value` must match one of the supported values and the `MEMORY` values must be one of the values supported for that `VCPU` value. The supported values are 0.25, 0.5, 1, 2, 4, 8, 16, and 32.

value = 0.25

`MEMORY` = 512, 1024, or 2048

value = 0.5

`MEMORY` = 1024, 2048, 3072, or 4096

value = 1

`MEMORY` = 2048, 3072, 4096, 5120, 6144, 7168, or 8192

value = 2

`MEMORY` = 4096, 5120, 6144, 7168, 8192, 9216, 10240, 11264, 12288, 13312, 14336, 15360, or 16384

value = 4

`MEMORY` = 8192, 9216, 10240, 11264, 12288, 13312, 14336, 15360, 16384, 17408, 18432, 19456, 20480, 21504, 22528, 23552, 24576, 25600, 26624, 27648, 28672, 29696, or 30720

value = 8

`MEMORY` = 16384, 20480, 24576, 28672, 32768, 36864, 40960, 45056, 49152, 53248, 57344, or 61440

value = 16

`MEMORY` = 32768, 40960, 49152, 57344, 65536, 73728, 81920, 90112, 98304, 106496, 114688, or 122880

value = 32

`MEMORY` = 61440, 122880, or 249856

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 Java V2
- AWS SDK for Ruby V3

RetryStrategy

The retry strategy that's associated with a job. For more information, see Automated job retries in the *AWS Batch User Guide*.

Contents

attempts

The number of times to move a job to the RUNNABLE status. You can specify between 1 and 10 attempts. If the value of `attempts` is greater than one, the job is retried on failure the same number of attempts as the value.

Type: Integer

Required: No

evaluateOnExit

Array of up to 5 objects that specify the conditions where jobs are retried or failed. If this parameter is specified, then the `attempts` parameter must also be specified. If none of the listed conditions match, then the job is retried.

Type: Array of EvaluateOnExit 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 Java V2
- AWS SDK for Ruby V3

RuntimePlatform

An object that represents the compute environment architecture for AWS Batch jobs on Fargate.

Contents

cpuArchitecture

The vCPU architecture. The default value is `X86_64`. Valid values are `X86_64` and `ARM64`.

> **Note**
>
> This parameter must be set to `X86_64` for Windows containers.

> **Note**
>
> Fargate Spot is not supported on Windows-based containers on Fargate. A job queue will be blocked if a Windows job is submitted to a job queue with only Fargate Spot compute environments. However, you can attach both `FARGATE` and `FARGATE_SPOT` compute environments to the same job queue.

Type: String

Required: No

operatingSystemFamily

The operating system for the compute environment. Valid values are: `LINUX` (default), `WINDOWS_SERVER_2019_CORE`, `WINDOWS_SERVER_2019_FULL`, `WINDOWS_SERVER_2022_CORE`, and `WINDOWS_SERVER_2022_FULL`.

> **Note**
>
> The following parameters can't be set for Windows containers: `linuxParameters`, `privileged`, `user`, `ulimits`, `readonlyRootFilesystem`, and `efsVolumeConfiguration`.

> **Note**
>
> The AWS Batch Scheduler checks the compute environments that are attached to the job queue before registering a task definition with Fargate. In this scenario, the job queue is where the job is submitted. If the job requires a Windows container and the first compute environment is `LINUX`, the compute environment is skipped and the next compute environment is checked until a Windows-based compute environment is found.

> **Note**
>
> Fargate Spot is not supported on Windows-based containers on Fargate. A job queue will be blocked if a Windows job is submitted to a job queue with only Fargate Spot compute environments. However, you can attach both `FARGATE` and `FARGATE_SPOT` compute environments to the same job queue.

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 Java V2
- AWS SDK for Ruby V3

S3FilesVolumeConfiguration

This is used when you're using an S3Files file system for job storage.

Contents

fileSystemArn

The Amazon Resource Name (ARN) of the S3Files file system to use.

Type: String

Required: Yes

accessPointArn

The Amazon Resource Name (ARN) of the S3Files access point to use.

Type: String

Required: No

rootDirectory

The directory within the S3Files file system to mount as the root directory.

Type: String

Required: No

transitEncryptionPort

The port to use when sending encrypted data between the Amazon ECS host and the S3Files file system server.

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 Java V2
- AWS SDK for Ruby V3

SchedulingPolicyDetail

An object that represents a scheduling policy.

Contents

arn

The Amazon Resource Name (ARN) of the scheduling policy. An example is `arn:aws:batch:us-east-1:123456789012:scheduling-policy/HighPriority` .

Type: String

Required: Yes

name

The name of the fair-share scheduling policy.

Type: String

Required: Yes

fairsharePolicy

The fair-share scheduling policy details.

Type: FairsharePolicy object

Required: No

quotaSharePolicy

The quota share scheduling policy details.

Type: QuotaSharePolicy object

Required: No

tags

The tags that you apply to the fair-share scheduling policy to categorize and organize your resources. Each tag consists of a key and an optional value. For more information, see Tagging AWS resources in *AWS General Reference*.

Type: String to string map

Map Entries: Maximum number of 50 items.

Key Length Constraints: Minimum length of 1. Maximum length of 128.

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 Java V2
- AWS SDK for Ruby V3

SchedulingPolicyListingDetail

An object that contains the details of a scheduling policy that's returned in a `ListSchedulingPolicy` action.

Contents

arn

Amazon Resource Name (ARN) of the scheduling policy.

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 Java V2
- AWS SDK for Ruby V3

Secret

An object that represents the secret to expose to your container. Secrets can be exposed to a container in the following ways:

- To inject sensitive data into your containers as environment variables, use the `secrets` container definition parameter.
- To reference sensitive information in the log configuration of a container, use the `secretOptions` container definition parameter.

For more information, see Specifying sensitive data in the *AWS Batch User Guide*.

Contents

name

The name of the secret.

Type: String

Required: Yes

valueFrom

The secret to expose to the container. The supported values are either the full Amazon Resource Name (ARN) of the AWS Secrets Manager secret or the full ARN of the parameter in the AWS Systems Manager Parameter Store.

> **Note**
>
> If the AWS Systems Manager Parameter Store parameter exists in the same Region as the job you're launching, then you can use either the full Amazon Resource Name (ARN) or name of the parameter. If the parameter exists in a different Region, then the full ARN must be specified.

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 Java V2
- AWS SDK for Ruby V3

ServiceEnvironmentDetail

Detailed information about a service environment, including its configuration, state, and capacity limits.

Contents

capacityLimits

The capacity limits for the service environment. This defines the maximum resources that can be used by service jobs in this environment.

Type: Array of CapacityLimit objects

Required: Yes

serviceEnvironmentArn

The Amazon Resource Name (ARN) of the service environment.

Type: String

Required: Yes

serviceEnvironmentName

The name of the service environment.

Type: String

Required: Yes

serviceEnvironmentType

The type of service environment. For SageMaker Training jobs, this value is `SAGEMAKER_TRAINING`.

Type: String

Valid Values: `SAGEMAKER_TRAINING`

Required: Yes

state

The state of the service environment. Valid values are `ENABLED` and `DISABLED`.

Type: String

Valid Values: `ENABLED | DISABLED`

Required: No

status

The current status of the service environment.

Type: String

Valid Values: `CREATING | UPDATING | DELETING | DELETED | VALID | INVALID`

Required: No

tags

The tags associated with the service environment. Each tag consists of a key and an optional value. For more information, see Tagging your AWS Batch resources.

Type: String to string map

Map Entries: Maximum number of 50 items.

Key Length Constraints: Minimum length of 1. Maximum length of 128.

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 Java V2
- AWS SDK for Ruby V3

ServiceEnvironmentOrder

Specifies the order of a service environment for a job queue. This determines the priority order when multiple service environments are associated with the same job queue.

Contents

order

The order of the service environment. Job queues with a higher priority are evaluated first when associated with the same service environment.

Type: Integer

Required: Yes

serviceEnvironment

The name or ARN of the service environment.

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 Java V2
- AWS SDK for Ruby V3

ServiceJobAttemptDetail

Detailed information about an attempt to run a service job.

Contents

serviceResourceId

The service resource identifier associated with the service job attempt.

Type: ServiceResourceId object

Required: No

startedAt

The Unix timestamp (in milliseconds) for when the service job attempt was started.

Type: Long

Required: No

statusReason

A string that provides additional details for the current status of the service job attempt.

Type: String

Required: No

stoppedAt

The Unix timestamp (in milliseconds) for when the service job attempt stopped running.

Type: Long

Required: No

See Also

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

- AWS SDK for C++

- AWS SDK for Java V2
- AWS SDK for Ruby V3

ServiceJobCapacityUsageDetail

The capacity usage for a service job, including the unit of measure and quantity of resources being consumed.

Contents

capacityUnit

The unit of measure for the service job capacity usage. For service jobs, this is the instance type.

Type: String

Required: No

quantity

The quantity of capacity being used by the service job, measured in the units specified by `capacityUnit`.

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 Java V2
- AWS SDK for Ruby V3

ServiceJobCapacityUsageSummary

The capacity usage for a service job, including the unit of measure and quantity of resources being used.

Contents

capacityUnit

The unit of measure for the service job capacity usage. For service jobs, this is the instance type.

Type: String

Required: No

quantity

The quantity of capacity being used by the service job, measured in the units specified by `capacityUnit`.

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 Java V2
- AWS SDK for Ruby V3

ServiceJobEvaluateOnExit

Specifies conditions for when to exit or retry a service job based on the exit status or status reason.

Contents

action

The action to take if the service job exits with the specified condition. Valid values are `RETRY` and `EXIT`.

Type: String

Valid Values: `RETRY | EXIT`

Required: No

onStatusReason

Contains a glob pattern to match against the StatusReason returned for a job. The pattern can contain up to 512 characters and can contain all printable characters. It can optionally end with an asterisk (*) so that only the start of the string needs to be an exact match.

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 Java V2
- AWS SDK for Ruby V3

ServiceJobPreemptedAttempt

Detailed information about a preempted attempt of a service job.

Contents

serviceResourceId

The service resource identifier associated with the service job attempt.

Type: ServiceResourceId object

Required: No

startedAt

The Unix timestamp (in milliseconds) for when the service job attempt was started.

Type: Long

Required: No

statusReason

A string that provides additional details for the current status of the service job attempt.

Type: String

Required: No

stoppedAt

The Unix timestamp (in milliseconds) for when the service job attempt stopped running.

Type: Long

Required: No

See Also

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

- AWS SDK for C++

- AWS SDK for Java V2
- AWS SDK for Ruby V3

ServiceJobPreemptionConfiguration

Specifies the service job behavior when preempted.

Contents

preemptionRetriesBeforeTermination

The number of times a service job can be retried after it is preempted. A job will be terminated when preemption retries have been exhausted. If this field is unset, preempted jobs will be requeued an unlimited number of times.

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 Java V2
- AWS SDK for Ruby V3

ServiceJobPreemptionSummary

Summarizes the preemptions of the service job. This field appears on a service job when it has been preempted.

Contents

preemptedAttemptCount

The total number of times the service job has been preempted.

Type: Integer

Required: No

recentPreemptedAttempts

A list of the most recent preemption attempts for the service job.

Type: Array of ServiceJobPreemptedAttempt 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 Java V2
- AWS SDK for Ruby V3

ServiceJobRetryStrategy

The retry strategy for service jobs. This defines how many times to retry a failed service job and under what conditions. For more information, see Service job retry strategies in the *AWS Batch User Guide*.

Contents

attempts

The number of times to move a service job to RUNNABLE status. You can specify between 1 and 10 attempts.

Type: Integer

Required: Yes

evaluateOnExit

Array of `ServiceJobEvaluateOnExit` objects that specify conditions under which the service job should be retried or failed.

Type: Array of ServiceJobEvaluateOnExit 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 Java V2
- AWS SDK for Ruby V3

ServiceJobSummary

Summary information about a service job.

Contents

jobId

The job ID for the service job.

Type: String

Required: Yes

jobName

The name of the service job.

Type: String

Required: Yes

serviceJobType

The type of service job. For SageMaker Training jobs, this value is `SAGEMAKER_TRAINING`.

Type: String

Valid Values: `SAGEMAKER_TRAINING`

Required: Yes

capacityUsage

The capacity usage information for this service job, including the unit of measure and quantity of resources being used.

Type: Array of ServiceJobCapacityUsageSummary objects

Required: No

createdAt

The Unix timestamp (in milliseconds) for when the service job was created.

Type: Long

Required: No

isTerminated

Indicates whether a termination request has been accepted for the service job. This field is only present when the value is `true`.

Type: Boolean

Required: No

jobArn

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

Type: String

Required: No

latestAttempt

Information about the latest attempt for the service job.

Type: LatestServiceJobAttempt object

Required: No

quotaShareName

The quota share for the service job.

Type: String

Required: No

scheduledAt

The Unix timestamp (in milliseconds) for when the service job was scheduled for execution.

Type: Long

Required: No

shareIdentifier

The share identifier for the job.

Type: String

Required: No

startedAt

The Unix timestamp (in milliseconds) for when the service job was started.

Type: Long

Required: No

status

The current status of the service job.

Type: String

Valid Values: `SUBMITTED | PENDING | RUNNABLE | SCHEDULED | STARTING | RUNNING | SUCCEEDED | FAILED`

Required: No

statusReason

A short string to provide more details on the current status of the service job.

Type: String

Required: No

stoppedAt

The Unix timestamp (in milliseconds) for when the service job stopped running.

Type: Long

Required: No

See Also

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

- AWS SDK for C++
- AWS SDK for Java V2

- AWS SDK for Ruby V3

ServiceJobTimeout

The timeout configuration for service jobs.

Contents

attemptDurationSeconds

The maximum duration in seconds that a service job attempt can run. After this time is reached, AWS Batch terminates the service job attempt.

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 Java V2
- AWS SDK for Ruby V3

ServiceResourceId

The AWS Batch unique identifier.

Contents

name

The name of the resource identifier.

Type: String

Valid Values: `TrainingJobArn`

Required: Yes

value

The value of the resource identifier.

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 Java V2
- AWS SDK for Ruby V3

ShareAttributes

Specifies the weights for the share identifiers for the fair-share policy. Share identifiers that aren't included have a default weight of `1.0`.

Contents

shareIdentifier

A share identifier or share identifier prefix. If the string ends with an asterisk (*), this entry specifies the weight factor to use for share identifiers that start with that prefix. The list of share identifiers in a fair-share policy can't overlap. For example, you can't have one that specifies a `shareIdentifier` of `UserA*` and another that specifies a `shareIdentifier` of `UserA1`.

There can be no more than 500 share identifiers active in a job queue.

The string is limited to 255 alphanumeric characters, and can be followed by an asterisk (*).

Type: String

Required: Yes

weightFactor

The weight factor for the share identifier. The default value is 1.0. A lower value has a higher priority for compute resources. For example, jobs that use a share identifier with a weight factor of 0.125 (1/8) get 8 times the compute resources of jobs that use a share identifier with a weight factor of 1.

The smallest supported value is 0.0001, and the largest supported value is 999.9999.

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 Java V2
- AWS SDK for Ruby V3

TaskContainerDependency

A list of containers that this task depends on.

Contents

condition

The dependency condition of the container. The following are the available conditions and their behavior:

- `START` - This condition emulates the behavior of links and volumes today. It validates that a dependent container is started before permitting other containers to start.
- `COMPLETE` - This condition validates that a dependent container runs to completion (exits) before permitting other containers to start. This can be useful for nonessential containers that run a script and then exit. This condition can't be set on an essential container.
- `SUCCESS` - This condition is the same as `COMPLETE`, but it also requires that the container exits with a zero status. This condition can't be set on an essential container.

Type: String

Required: No

containerName

A unique identifier for the container.

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 Java V2
- AWS SDK for Ruby V3

TaskContainerDetails

The details for the container in this task attempt.

Contents

command

The command that's passed to the container. This parameter maps to `Cmd` in the Create a container section of the Docker Remote API and the `COMMAND` parameter to docker run. For more information, see https://docs.docker.com/engine/reference/builder/#cmd.

Type: Array of strings

Required: No

dependsOn

A list of containers that this container depends on.

Type: Array of TaskContainerDependency objects

Required: No

environment

The environment variables to pass to a container. This parameter maps to `Env` in the Create a container section of the Docker Remote API and the `--env` option to docker run.

> **Important**
>
> We don't recommend using plaintext environment variables for sensitive information, such as credential data.

Type: Array of KeyValuePair objects

Required: No

essential

If the essential parameter of a container is marked as `true`, and that container fails or stops for any reason, all other containers that are part of the task are stopped. If the `essential`

parameter of a container is marked as false, its failure doesn't affect the rest of the containers in a task. If this parameter is omitted, a container is assumed to be essential.

All jobs must have at least one essential container. If you have an application that's composed of multiple containers, group containers that are used for a common purpose into components, and separate the different components into multiple task definitions. For more information, see Application Architecture in the *Amazon Elastic Container Service Developer Guide*.

Type: Boolean

Required: No

exitCode

The exit code returned upon completion.

Type: Integer

Required: No

firelensConfiguration

The FireLens configuration for the container. This is used to specify and configure a log router for container logs. For more information, see Custom log routing in the *Amazon Elastic Container Service Developer Guide*.

Type: FirelensConfiguration object

Required: No

image

The image used to start a container. This string is passed directly to the Docker daemon. By default, images in the Docker Hub registry are available. Other repositories are specified with either `repository-url/image:tag` or `repository-url/image@digest`. Up to 255 letters (uppercase and lowercase), numbers, hyphens, underscores, colons, periods, forward slashes, and number signs are allowed. This parameter maps to `Image` in the Create a container section of the Docker Remote API and the `IMAGE` parameter of the *docker run*.

Type: String

Required: No

linuxParameters

Linux-specific modifications that are applied to the container, such as Linux kernel capabilities. For more information, see KernelCapabilities.

> **Note**
>
> This parameter is not supported for Windows containers.

Type: LinuxParameters object

Required: No

logConfiguration

The log configuration specification for the container.

This parameter maps to `LogConfig` in the Create a container section of the Docker Remote API and the `--log-driver` option to docker run.

By default, containers use the same logging driver that the Docker daemon uses. However the container can use a different logging driver than the Docker daemon by specifying a log driver with this parameter in the container definition. To use a different logging driver for a container, the log system must be configured properly on the container instance (or on a different log server for remote logging options). For more information about the options for different supported log drivers, see Configure logging drivers in the *Docker documentation*.

> **Note**
>
> Amazon ECS currently supports a subset of the logging drivers available to the Docker daemon (shown in the `LogConfiguration` data type). Additional log drivers may be available in future releases of the Amazon ECS container agent.

This parameter requires version 1.18 of the Docker Remote API or greater on your container instance. To check the Docker Remote API version on your container instance, log in to your container instance and run the following command: sudo docker version `--format '{{.Server.APIVersion}}'`

> **Note**
>
> The Amazon ECS container agent running on a container instance must register the logging drivers available on that instance with the `ECS_AVAILABLE_LOGGING_DRIVERS` environment variable before containers placed on that instance can use these log configuration options. For more information, see Amazon ECS container agent configuration in the *Amazon Elastic Container Service Developer Guide*.

Type: LogConfiguration object

Required: No

logStreamName

The name of the CloudWatch Logs log stream that's associated with the container. The log group for AWS Batch jobs is /aws/batch/job. Each container attempt receives a log stream name when they reach the `RUNNING` status.

Type: String

Required: No

mountPoints

The mount points for data volumes in your container.

This parameter maps to `Volumes` in the Create a container section of the Docker Remote API and the --volume option to docker run.

Windows containers can mount whole directories on the same drive as `$env:ProgramData`. Windows containers can't mount directories on a different drive, and mount point can't be across drives.

Type: Array of MountPoint objects

Required: No

name

The name of a container.

Type: String

Required: No

networkInterfaces

The network interfaces that are associated with the job.

Type: Array of NetworkInterface objects

Required: No

privileged

When this parameter is `true`, the container is given elevated privileges on the host container instance (similar to the `root` user). This parameter maps to `Privileged` in the Create a container section of the Docker Remote API and the `--privileged` option to docker run.

> **Note**
>
> This parameter is not supported for Windows containers or tasks run on Fargate.

Type: Boolean

Required: No

readonlyRootFilesystem

When this parameter is true, the container is given read-only access to its root file system. This parameter maps to `ReadonlyRootfs` in the Create a container section of the Docker Remote API and the `--read-only` option to docker run.

> **Note**
>
> This parameter is not supported for Windows containers.

Type: Boolean

Required: No

reason

A short (255 max characters) human-readable string to provide additional details for a running or stopped container.

Type: String

Required: No

repositoryCredentials

The private repository authentication credentials to use.

Type: RepositoryCredentials object

Required: No

resourceRequirements

The type and amount of a resource to assign to a container. The only supported resource is a GPU.

Type: Array of ResourceRequirement objects

Required: No

secrets

The secrets to pass to the container. For more information, see Specifying Sensitive Data in the Amazon Elastic Container Service Developer Guide.

Type: Array of Secret objects

Required: No

startTimeout

Time duration (in seconds) to wait before giving up on resolving dependencies for a container. The minimum value is 2 seconds and the maximum value for Fargate is 120 seconds.

Type: Integer

Required: No

stopTimeout

Time duration (in seconds) to wait before the container is forcefully killed if it doesn't exit normally on its own. The minimum value is 2 seconds and the maximum value for Fargate is 120 seconds. If the parameter is not specified, the default value of 30 seconds is used. For tasks that use the EC2 launch type, if the `stopTimeout` parameter isn't specified, the value set for the Amazon ECS container agent configuration variable `ECS_CONTAINER_STOP_TIMEOUT` is

used. If neither the `stopTimeout` parameter nor the `ECS_CONTAINER_STOP_TIMEOUT` agent configuration variable are set, then the default value of 30 seconds is used.

Type: Integer

Required: No

ulimits

A list of `ulimits` to set in the container. If a `ulimit` value is specified in a task definition, it overrides the default values set by Docker. This parameter maps to `Ulimits` in the Create a container section of the Docker Remote API and the `--ulimit` option to docker run.

Amazon ECS tasks hosted on Fargate use the default resource limit values set by the operating system with the exception of the nofile resource limit parameter which Fargate overrides. The `nofile` resource limit sets a restriction on the number of open files that a container can use. The default `nofile` soft limit is `1024` and the default hard limit is `65535`.

This parameter requires version 1.18 of the Docker Remote API or greater on your container instance. To check the Docker Remote API version on your container instance, log in to your container instance and run the following command: sudo docker version `--format '{{.Server.APIVersion}}'`

> **Note**
>
> This parameter is not supported for Windows containers.

Type: Array of Ulimit objects

Required: No

user

The user to use inside the container. This parameter maps to User in the Create a container section of the Docker Remote API and the --user option to docker run.

> **Note**
>
> When running tasks using the `host` network mode, don't run containers using the `root user (UID 0)`. We recommend using a non-root user for better security.

You can specify the `user` using the following formats. If specifying a UID or GID, you must specify it as a positive integer.

- `user`
- `user:group`
- `uid`
- `uid:gid`
- `user:gi`
- `uid:group`
-

> **Note**
>
> This parameter is not supported for Windows containers.

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 Java V2
- AWS SDK for Ruby V3

TaskContainerOverrides

The overrides that should be sent to a container.

For information about using AWS Batch overrides when you connect event sources to targets, see BatchContainerOverrides.

Contents

command

The command to send to the container that overrides the default command from the Docker image or the job definition.

> **Note**
>
> This parameter can't contain an empty string.

Type: Array of strings

Required: No

environment

The environment variables to send to the container. You can add new environment variables, which are added to the container at launch, or you can override the existing environment variables from the Docker image or the job definition.

> **Note**
>
> Environment variables cannot start with `AWS_BATCH`. This naming convention is reserved for variables that AWS Batch sets.

Type: Array of KeyValuePair objects

Required: No

name

A pointer to the container that you want to override. The container's name provides a unique identifier for the container being used.

Type: String

Required: No

resourceRequirements

The type and amount of resources to assign to a container. This overrides the settings in the job definition. The supported resources include `GPU`, `MEMORY`, and `VCPU`.

Type: Array of ResourceRequirement 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 Java V2
- AWS SDK for Ruby V3

TaskContainerProperties

Container properties are used for Amazon ECS-based job definitions. These properties to describe the container that's launched as part of a job.

Contents

image

The image used to start a container. This string is passed directly to the Docker daemon. By default, images in the Docker Hub registry are available. Other repositories are specified with either `repository-url/image:tag` or `repository-url/image@digest`. Up to 255 letters (uppercase and lowercase), numbers, hyphens, underscores, colons, periods, forward slashes, and number signs are allowed. This parameter maps to `Image` in the Create a container section of the Docker Remote API and the `IMAGE` parameter of the *docker run*.

Type: String

Required: Yes

command

The command that's passed to the container. This parameter maps to `Cmd` in the Create a container section of the Docker Remote API and the `COMMAND` parameter to docker run. For more information, see Dockerfile reference: CMD.

Type: Array of strings

Required: No

dependsOn

A list of containers that this container depends on.

Type: Array of TaskContainerDependency objects

Required: No

environment

The environment variables to pass to a container. This parameter maps to Env in the Create a container section of the Docker Remote API and the `--env` parameter to docker run.

> **Important**
>
> We don't recommend using plaintext environment variables for sensitive information, such as credential data.

> **Note**
>
> Environment variables cannot start with `AWS_BATCH`. This naming convention is reserved for variables that AWS Batch sets.

Type: Array of KeyValuePair objects

Required: No

essential

If the essential parameter of a container is marked as `true`, and that container fails or stops for any reason, all other containers that are part of the task are stopped. If the `essential` parameter of a container is marked as false, its failure doesn't affect the rest of the containers in a task. If this parameter is omitted, a container is assumed to be essential.

All jobs must have at least one essential container. If you have an application that's composed of multiple containers, group containers that are used for a common purpose into components, and separate the different components into multiple task definitions. For more information, see Application Architecture in the *Amazon Elastic Container Service Developer Guide*.

Type: Boolean

Required: No

firelensConfiguration

The FireLens configuration for the container. This is used to specify and configure a log router for container logs. For more information, see Custom log routing in the *Amazon Elastic Container Service Developer Guide*.

Type: FirelensConfiguration object

Required: No

linuxParameters

Linux-specific modifications that are applied to the container, such as Linux kernel capabilities. For more information, see KernelCapabilities.

Type: LinuxParameters object

Required: No

logConfiguration

The log configuration specification for the container.

This parameter maps to `LogConfig` in the Create a container section of the Docker Remote API and the `--log-driver` option to docker run.

By default, containers use the same logging driver that the Docker daemon uses. However the container can use a different logging driver than the Docker daemon by specifying a log driver with this parameter in the container definition. To use a different logging driver for a container, the log system must be configured properly on the container instance (or on a different log server for remote logging options). For more information about the options for different supported log drivers, see Configure logging drivers in the *Docker documentation*.

> **Note**
>
> Amazon ECS currently supports a subset of the logging drivers available to the Docker daemon (shown in the `LogConfiguration` data type). Additional log drivers may be available in future releases of the Amazon ECS container agent.

This parameter requires version 1.18 of the Docker Remote API or greater on your container instance. To check the Docker Remote API version on your container instance, log in to your container instance and run the following command: sudo docker version `--format '{{.Server.APIVersion}}'`

> **Note**
>
> The Amazon ECS container agent running on a container instance must register the logging drivers available on that instance with the `ECS_AVAILABLE_LOGGING_DRIVERS` environment variable before containers placed

on that instance can use these log configuration options. For more information, see Amazon ECS container agent configuration in the *Amazon Elastic Container Service Developer Guide*.

Type: LogConfiguration object

Required: No

mountPoints

The mount points for data volumes in your container.

This parameter maps to `Volumes` in the Create a container section of the Docker Remote API and the --volume option to docker run.

Windows containers can mount whole directories on the same drive as `$env:ProgramData`. Windows containers can't mount directories on a different drive, and mount point can't be across drives.

Type: Array of MountPoint objects

Required: No

name

The name of a container. The name can be used as a unique identifier to target your `dependsOn` and `Overrides` objects.

Type: String

Required: No

privileged

When this parameter is `true`, the container is given elevated privileges on the host container instance (similar to the `root` user). This parameter maps to `Privileged` in the Create a container section of the Docker Remote API and the `--privileged` option to docker run.

> **Note**
>
> This parameter is not supported for Windows containers or tasks run on Fargate.

Type: Boolean

Required: No

readonlyRootFilesystem

When this parameter is true, the container is given read-only access to its root file system. This parameter maps to `ReadonlyRootfs` in the Create a container section of the Docker Remote API and the `--read-only` option to docker run.

> **Note**
>
> This parameter is not supported for Windows containers.

Type: Boolean

Required: No

repositoryCredentials

The private repository authentication credentials to use.

Type: RepositoryCredentials object

Required: No

resourceRequirements

The type and amount of a resource to assign to a container. The supported resources include `GPU`, `MEMORY`, and `VCPU`.

Type: Array of ResourceRequirement objects

Required: No

secrets

The secrets to pass to the container. For more information, see Specifying Sensitive Data in the Amazon Elastic Container Service Developer Guide.

Type: Array of Secret objects

Required: No

startTimeout

Time duration (in seconds) to wait before giving up on resolving dependencies for a container. The minimum value is 2 seconds and the maximum value for Fargate is 120 seconds.

Type: Integer

Required: No

stopTimeout

Time duration (in seconds) to wait before the container is forcefully killed if it doesn't exit normally on its own. The minimum value is 2 seconds and the maximum value for Fargate is 120 seconds. If the parameter is not specified, the default value of 30 seconds is used. For tasks that use the EC2 launch type, if the `stopTimeout` parameter isn't specified, the value set for the Amazon ECS container agent configuration variable `ECS_CONTAINER_STOP_TIMEOUT` is used. If neither the `stopTimeout` parameter nor the `ECS_CONTAINER_STOP_TIMEOUT` agent configuration variable are set, then the default value of 30 seconds is used.

Type: Integer

Required: No

ulimits

A list of `ulimits` to set in the container. If a `ulimit` value is specified in a task definition, it overrides the default values set by Docker. This parameter maps to `Ulimits` in the Create a container section of the Docker Remote API and the `--ulimit` option to docker run.

Amazon ECS tasks hosted on Fargate use the default resource limit values set by the operating system with the exception of the nofile resource limit parameter which Fargate overrides. The `nofile` resource limit sets a restriction on the number of open files that a container can use. The default `nofile` soft limit is `1024` and the default hard limit is `65535`.

This parameter requires version 1.18 of the Docker Remote API or greater on your container instance. To check the Docker Remote API version on your container instance, log in to your container instance and run the following command: sudo docker version `--format '{{.Server.APIVersion}}'`

> **Note**
>
> This parameter is not supported for Windows containers.

Type: Array of Ulimit objects

Required: No

user

The user to use inside the container. This parameter maps to User in the Create a container section of the Docker Remote API and the --user option to docker run.

> **Note**
>
> When running tasks using the `host` network mode, don't run containers using the `root user (UID 0)`. We recommend using a non-root user for better security.

You can specify the `user` using the following formats. If specifying a UID or GID, you must specify it as a positive integer.

- `user`
- `user:group`
- `uid`
- `uid:gid`
- `user:gi`
- `uid:group`

> **Note**
>
> This parameter is not supported for Windows containers.

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 Java V2
- AWS SDK for Ruby V3

TaskPropertiesOverride

An object that contains overrides for the task definition of a job.

Contents

containers

The overrides for the container definition of a job.

Type: Array of TaskContainerOverrides 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 Java V2
- AWS SDK for Ruby V3

TerminateJobsErrorDetail

An object that contains the details of a job that couldn't be terminated by a `TerminateJobs` operation.

Contents

code

An error code that identifies the reason the job couldn't be terminated. Valid values are:

- `ValidationException` – A job identifier in the request is malformed or isn't valid.
- `ClientException` – The request failed because of a client error.
- `ThrottlingException` – The request was throttled. Retry the request.
- `ServerException` – An internal error occurred. Retry the request.
- `AccessDenied` – The caller isn't authorized to perform the action on the specified job.

Type: String

Required: Yes

job

The AWS Batch job ID of the job that couldn't be terminated.

Type: String

Required: Yes

message

A message that describes the reason the job couldn't be terminated.

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 Java V2
- AWS SDK for Ruby V3

TerminateServiceJobsErrorDetail

An object that contains the details of a service job that couldn't be terminated by a `TerminateServiceJobs` operation.

Contents

code

An error code that identifies the reason the service job couldn't be terminated. Valid values are:

- `ValidationException` – A service job identifier in the request is malformed or isn't valid.
- `ClientException` – The request failed because of a client error.
- `ThrottlingException` – The request was throttled. Retry the request.
- `ServerException` – An internal error occurred. Retry the request.
- `AccessDenied` – The caller isn't authorized to perform the action on the specified service job.

Type: String

Required: Yes

job

The service job ID of the service job that couldn't be terminated.

Type: String

Required: Yes

message

A message that describes the reason the service job couldn't be terminated.

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 Java V2
- AWS SDK for Ruby V3

Tmpfs

The container path, mount options, and size of the `tmpfs` mount.

> **Note**
>
> This object isn't applicable to jobs that are running on Fargate resources.

Contents

containerPath

The absolute file path in the container where the `tmpfs` volume is mounted.

Type: String

Required: Yes

size

The size (in MiB) of the `tmpfs` volume.

Type: Integer

Required: Yes

mountOptions

The list of `tmpfs` volume mount options.

Valid values: `"defaults" | "ro" | "rw" | "suid" | "nosuid" | "dev" | "nodev" | "exec" | "noexec" | "sync" | "async" | "dirsync" | "remount" | "mand" | "nomand" | "atime" | "noatime" | "diratime" | "nodiratime" | "bind" | "rbind" | "unbindable" | "runbindable" | "private" | "rprivate" | "shared" | "rshared" | "slave" | "rslave" | "relatime" | "norelatime" | "strictatime" | "nostrictatime" | "mode" | "uid" | "gid" | "nr_inodes" | "nr_blocks" | "mpol"`

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 Java V2
- AWS SDK for Ruby V3

Ulimit

The `ulimit` settings to pass to the container. For more information, see Ulimit.

> **Note**
>
> This object isn't applicable to jobs that are running on Fargate resources.

Contents

hardLimit

The hard limit for the `ulimit` type.

Type: Integer

Required: Yes

name

The `type` of the `ulimit`. Valid values are: `core` | `cpu` | `data` | `fsize` | `locks` | `memlock` | `msgqueue` | `nice` | `nofile` | `nproc` | `rss` | `rtprio` | `rttime` | `sigpending` | `stack`.

Type: String

Required: Yes

softLimit

The soft limit for the `ulimit` type.

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 Java V2
- AWS SDK for Ruby V3

UpdateManagedInstancesProviderConfiguration

The configuration for updating an Amazon ECS Managed Instances capacity provider. Used in `UpdateComputeEnvironment` requests. The `capacityOptionType` and `fipsEnabled` fields cannot be changed on update.

Contents

infrastructureOptimization

The updated infrastructure optimization configuration.

Type: InfrastructureOptimization object

Required: No

infrastructureRoleArn

The updated Amazon Resource Name (ARN) of the IAM role that Amazon ECS assumes to manage Amazon EC2 instances on your behalf.

Type: String

Required: No

instanceLaunchTemplate

The updated instance launch configuration for the Amazon ECS Managed Instances capacity provider.

Type: InstanceLaunchTemplateUpdate object

Required: No

propagateTags

Specifies whether tags on the capacity provider are propagated to the Amazon EC2 instances it launches. Valid values:

- `CAPACITY_PROVIDER` — Propagates tags to instances.
- `NONE` — Does not propagate tags to instances.

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 Java V2
- AWS SDK for Ruby V3

UpdatePolicy

Specifies the infrastructure update policy for the Amazon EC2 compute environment. For more information about infrastructure updates, see Updating compute environments in the *AWS Batch User Guide*.

> **Important**
>
> The `updatePolicy` property applies only to Amazon EC2 managed compute environments. AWS Batch doesn't support `updatePolicy` for Fargate compute environments. If you specify it for a Fargate compute environment, AWS Batch ignores it.

Contents

jobExecutionTimeoutMinutes

Specifies the job timeout (in minutes) when the compute environment infrastructure is updated. The default value is 30. The maximum value is 7200.

> **Note**
>
> Increasing `jobExecutionTimeoutMinutes` during infrastructure updates delays the replacement of instances with new instances that include updates such as security patches, but provides more time for jobs to execute. Consider the security implications of this tradeoff when setting timeout values.

Type: Long

Valid Range: Minimum value of 1. Maximum value of 7200.

Required: No

terminateJobsOnUpdate

Specifies whether jobs are automatically terminated when the compute environment infrastructure is updated. The default value is `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 Java V2
- AWS SDK for Ruby V3

Volume

A data volume that's used in a job's container properties.

Contents

efsVolumeConfiguration

This parameter is specified when you're using an Amazon Elastic File System file system for job storage. Jobs that are running on Fargate resources must specify a `platformVersion` of at least `1.4.0`.

Type: EFSVolumeConfiguration object

Required: No

host

The contents of the `host` parameter determine whether your data volume persists on the host container instance and where it's stored. If the host parameter is empty, then the Docker daemon assigns a host path for your data volume. However, the data isn't guaranteed to persist after the containers that are associated with it stop running.

> **Note**
>
> This parameter isn't applicable to jobs that are running on Fargate resources and shouldn't be provided.

Type: Host object

Required: No

name

The name of the volume. It can be up to 255 characters long. It can contain uppercase and lowercase letters, numbers, hyphens (-), and underscores (_). This name is referenced in the `sourceVolume` parameter of container definition `mountPoints`.

Type: String

Required: No

s3filesVolumeConfiguration

This parameter is specified when you're using an S3Files file system for job storage.

Type: S3FilesVolumeConfiguration 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 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*.

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.

For more information, see Create a signed AWS API request 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

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 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 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