aws

用户指南

亚马逊 DataZone

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

亚马逊 DataZone: 用户指南

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

Amazon 的商标和商业外观不得用于任何非 Amazon 的商品或服务，也不得以任何可能引起客户混淆、贬低或诋毁 Amazon 的方式使用。所有非 Amazon 拥有的其他商标均为各自所有者的财产，这些所有者可能附属于 Amazon、与 Amazon 有关联或由 Amazon 赞助，也可能不是如此。

Table of Contents

什么是亚马逊 DataZone？ 1
..... 1
Amazon 如何 DataZone 支持其他服务并与其他 AWS 服务集成？ 1
如何访问亚马逊 DataZone？ 2
亚马逊 SageMaker 以及何时使用亚马逊 SageMaker 与亚马逊 DataZone 3
术语和概念 4
亚马逊 DataZone 组件 4
什么是 Amazon DataZone 域名？ 5
Amazon 的 DataZone 项目和环境是什么？ 5
什么是亚马逊 DataZone 蓝图？ 7
Amazon DataZone 库存和发布工作流程是什么？ 8
创建项目库存资产 8
将项目库存资产发布到 Amazon DataZone 目录 9
Amazon DataZone 订阅和配送流程是什么？ 9
Amazon 的用户角色 DataZone 10
亚马逊 DataZone 术语 10
新增功能 17
2024 17
Amazon DataZone 推出针对订阅请求的元数据强制执行规则 17
亚马逊 DataZone 定制 AWS 服务蓝图现在 SageMaker 为亚马逊 DataZone 项目提供了全新的设置体验 17
Amazon DataZone 推出对定制 AWS 服务蓝图的 AWS CloudFormation 支持 17
亚马逊 DataZone 推出域名单位和授权政策 18
亚马逊 DataZone 推出数据产品 18
亚马逊 DataZone 推出精细访问控制功能 18
亚马逊 DataZone 推出数据血统功能 18
亚马逊 DataZone 推出定制 AWS 服务蓝图 19
数据来源创建流程的增强功能 19
亚马逊 DataZone 启动与亚马逊的整合 SageMaker 19
亚马逊 DataZone 推出与 L AWS ake Formation 混合访问模式的集成 20
亚马逊 DataZone 推出与 Glue 数据 AWS 质量的集成 20
Amazon 中描述的 AI 推荐正式发布版 DataZone 20
亚马逊 DataZone 推出亚马逊 Redshift 集成增强功能 20
AWS Amazon 的 Cloud Formation DataZone 21

直接将 IAM 委托人添加为 Amazon DataZone 项目的成员 21
对来自数据门户的自定义资产类型的支持 22
2023 22
删除域 22
混合模式 22
HIPAA 资格 22
Amazon 中描述的 AI 建议 DataZone （预览版） 22
DefaultDataLake 蓝图增强 23
支持的区域： 24
设置 25
注册参加 AWS 账户 25
配置使用该管理控制台所需的 IAM 权限 25
将必需和可选的策略附加到用户、组或角色，以便访问管理控制台 26
为 IAM 权限创建自定义策略，以启用管理服务控制台简化角色创建 26
创建自定义权限策略以管理与域关联的账户 28
（可选）为创建自定义策略 AWS 添加和删除 SSO 用户和 SSO 组对域的访问权限的 Identity Center 权限 31
（可选）将您的 IAM 委托人添加为密钥用户，使用来自的客户管理密钥创建您的域 AWS 千米 32
配置使用数据门户所需的 IAM 权限 32
将所需的策略附加到用户、组或角色以便访问数据门户 33
将所需的策略附加到用户、组或角色以便访问目录 34
如果您的域已使用 AWS KMS 中的客户自主管理型密钥进行加密，则将可选策略附加到用户、组或角色以访问数据门户或目录 35
正在设置 AWS 亚马逊 IAM 身份中心 DataZone 36
开始使用 38
包含示例 AWS Glue 数据的快速入门指南 38
第 1 步-创建 Amazon DataZone 域名和数据门户 39
步骤 2 – 创建发布项目 41
步骤 3 – 创建环境 41
步骤 4 – 创建数据以供发布 41
第 5 步-从 AWS Glue 收集元数据 42
步骤 6 – 整理和发布数据资产 42
步骤 7 – 创建用于数据分析的项目 43
步骤 8 – 创建用于数据分析的环境 43
步骤 9 – 搜索数据目录并订阅数据 43

步骤 10 – 批准订阅请求 44
步骤 11 – 在 Amazon Athena 中构建查询并分析数据 44
包含 Amazon Redshift 示例数据的快速入门指南 44
第 1 步-创建 Amazon DataZone 域名和数据门户 45
步骤 2 – 创建发布项目 47
步骤 3 – 创建环境 47
步骤 4 – 创建数据以供发布 48
步骤 5 – 从 Amazon Redshift 收集元数据 48
步骤 6 – 整理和发布数据资产 49
步骤 7 – 创建用于数据分析的项目 49
步骤 8 – 创建用于数据分析的环境 49
步骤 9 – 搜索数据目录并订阅数据 50
步骤 10 – 批准订阅请求 50
步骤 11 – 在 Amazon Redshift 中构建查询并分析数据 51
常见任务的示例脚本 51
创建 Amazon DataZone 域名和数据门户 51
创建发布项目 52
创建环境配置文件 52
创建环境 55
从 AWS Glue 收集元数据 56
整理和发布数据资产 58
搜索数据目录并订阅数据 62
在数据目录中搜索资产 62
其他有用的示例脚本 64
域和用户访问权限 66
创建域 66
编辑域 68
删除域 69
启用 Amazon 的 IAM 身份中心 DataZone 70
禁用 Amazon 的 IAM 身份中心 DataZone 71
在 Amazon DataZone 控制台中管理用户 72
管理 IAM 角色和用户 72
管理 SSO 用户 73
管理 SSO 组 74
在数据门户中管理用户权限 75
限制访问 Amazon DataZone 75

将亚马逊 DataZone 域名升级为亚马逊 SageMaker 统一域名 75
升级域之前的考虑事项 75
将您的亚马逊 DataZone 域名升级为亚马逊 SageMaker 统一域名 76
有关将 Amazon 域名升级为亚马逊 SageMaker 统一 DataZone 域名的常见问题 77
域单元和授权策略 79
创建域单元 80
编辑域单元 81
删除域单元 81
管理域单元所有者 82
向域单元中的用户和组分配授权策略 82
Amazon 域单元层次结构中的项目成员资格政策 DataZone 83
向域单元中的项目分配授权策略 89
在蓝图配置中分配授权策略 90
内置蓝图 92
在拥有 Amazon DataZone 域 AWS 名的账户中启用内置蓝图 92
将亚马逊 SageMaker 作为可信服务添加到拥有亚马逊 DataZone 域名的 AWS 账户中 97
自定义 AWS 服务蓝图 98
启用自定义 AWS 服务蓝图 98
使用自定义 AWS 服务蓝图创建环境 99
在自定义 AWS 服务环境中创建操作 100
将项目成员添加到自定义 AWS 服务环境 101
在 AWS 服务环境中配置数据源 101
在 AWS 服务环境中配置订阅目标 102
关联账户 103
请求与其他 AWS 账户关联 103
向账户提供对客户自主管理型 KMS 密钥的访问权限 104
接受来自 Amazon DataZone 域的账户关联请求并启用环境蓝图 104
在关联 AWS 账户中启用环境蓝图 105
在关联 AWS 账户中 SageMaker 将 Amazon 添加为可信服务 110
拒绝来自亚马逊 DataZone 域名的账户关联请求 110
在 Amazon 中移除关联账户 DataZone 110
数据目录 112
创建业务术语表 113
编辑业务术语表 114
删除业务术语表 114
创建术语表术语 115

编辑术语表术语 116
删除术语表术语 116
创建元数据表单 117
编辑元数据表单 118
删除元数据表单 118
创建元数据表单字段 119
编辑元数据表单字段 120
删除元数据表单字段 120
项目和环境 122
创建环境配置文件 123
编辑环境配置文件 125
删除环境配置文件 125
创建新环境 126
编辑环境 127
删除环境 127
创建新项目 128
编辑项目 128
将项目移动到其他域单元 129
删除项目 129
离开项目 131
向项目添加成员 131
从项目中移除成员 132
数据库存和发布 133
为亚马逊配置 Lake Formation 权限 DataZone 134
亚马逊 DataZone 与 AWS Lake Formation 混合模式 135
创建自定义资产类型 138
为创建并运行数据源 AWS Glue Data Catalog 142
为 Amazon Redshift 创建并运行数据来源 144
编辑数据来源 146
删除数据来源 147
将项目库存中的资产发布到目录 148
发布资产 148
管理库存并整理资产 149
将其他元数据表单附加到资产 150
在完成整理后将资产发布到目录 151
手动创建资产 151

从目录中取消发布资产 152
删除资产 152
手动启动数据来源运行 153
资产版本控制 153
Amazon 的数据质量 DataZone 154
启用数据质量 AWS Glue 资产 154
为自定义资产类型启用数据质量 155
在 Amazon 中使用机器学习和生成人工智能 DataZone 157
支持的区域： 158
使用 GenAI 的步骤 159
对自定义关系资产类型的支持 160
配额 160
Amazon 中的数据谱系 DataZone 160
Amazon 中的血统节点类型 DataZone 162
世系节点中的关键属性 162
可视化数据世系 163
Amazon 中的数据沿袭授权 DataZone 163
Amazon 中的数据沿袭示例体验 DataZone 164
在管理控制台中启用数据血统 164
以编程方式使用 Amazon DataZone 数据谱系 165
自动执行血统 AWS Glue 目录 165
从 Amazon Redshift 实现血统自动化 168
针对发布的元数据强制规则 168
将 Snowflake 作为数据源连接 169
先决条件 169
第 1 步：为亚马逊创建 Snowflake 角色 DataZone 169
第 2 步：将 Snowflake 凭据存储在 AWS Secrets Manager 171
第 3 步：查找您的项目环境 ID 171
步骤 4：创建 Snowflake 连接 172
验证连接 174
适用于 Spark 任务的雪花血统 AWS Glue 174
先决条件 174
步骤 1：在上升级 OpenLineage 核心客户端 AWS Glue 作业 174
第 2 步：在上启用世系事件生成 AWS Glue 作业 175
验证血统是否被捕获 175
数据产品 176

创建新的数据产品 176
发布数据产品 177
编辑数据产品 177
取消发布数据产品 178
删除数据产品 179
订阅数据产品 179
审查订阅请求并授权对数据产品的订阅 180
重新发布数据产品 180
数据发现、订阅和使用 182
在目录中搜索和查看资产 183
请求订阅资产 184
批准或拒绝订阅请求 185
自动审批订阅请求 186
撤销现有订阅 187
取消订阅请求 188
取消订阅资产 188
使用现有 IAM 角色完成亚马逊 DataZone 订阅 189
授予对托管 AWS Glue Data Catalog 资产的访问权限 192
授予对托管 Amazon Redshift 资产的访问权限 193
向非托管资产授予对已批准订阅的访问权限 194
查询 Amazon Athena 或 Amazon Redshift 中的数据 194
使用 Amazon Athena 查询数据 195
使用 Amazon Redshift 查询数据 197
针对订阅请求的元数据强制规则 198
通过 JDBC 连接使用外部分析应用程序分析订阅的数据 200
RedeemAccessToken API 参考 202
对数据的精细访问控制 205
创建行筛选条件 205
创建列筛选条件 206
删除行或列筛选条件 207
编辑行或列筛选条件 207
使用筛选条件授予访问权限 208
AWS Glue 桌子 208
Amazon Redshift 209
事件和通知 210
通过 Amazon DataZone 数据门户中的专用收件箱进行活动 210

通过 Amazon EventBridge 默认总线举办的活动 214
安全性 217
数据保护 217
数据加密 218
传输中加密 219
Inter-network 交通隐私 219
亚马逊的静态数据加密 DataZone 219
使用亚马逊接口 VPC 终端节点 DataZone 237
亚马逊授权 DataZone 238
亚马逊授权 DataZone 控制台 238
亚马逊授权 DataZone 门户 238
亚马逊 DataZone 配置文件和角色 239
控制访问权限 239
AWS 托管策略 240
亚马逊的 IAM 角色 DataZone 258
临时证书 268
主体权限 269
合规性验证 269
安全最佳实践 269
实施最低权限访问 269
使用 IAM 角色 270
实施 Server-Side 依赖资源中的加密 270
使用 CloudTrail 监控 API 调用 270
在亚马逊中使用 RAM DataZone 270
恢复能力 270
数据来源韧性 271
资产韧性 271
资产类型和元数据表单韧性 272
术语表韧性 272
全局搜索韧性 272
订阅韧性 272
环境韧性 272
环境蓝图韧性 273
项目韧性 273
RAM 韧性 273
用户配置文件管理韧性 273

域韧性 273
亚马逊的基础设施安全 DataZone 273
Cross-service 亚马逊的副手防范措施令人困惑 DataZone 273
亚马逊的配置和漏洞分析 DataZone 274
要添加到允许列表的域 274
监控 275
监控事件 275
CloudTrail 日志 275
CloudTrail 中的 Amazon DataZone 信息 276
问题排查 277
问题排查 AWS 亚马逊的 Lake Formation 权限 DataZone 277
对 Amazon DataZone 世系资产与上游数据集关联进行故障排除 279
SourceIdentifier 在血统节点上 279
Amazon 如何 DataZone 根据事件构建 sourceIdentifier？ OpenLineage 279
替代方法 285
排查资产世系节点缺少上游的问题 285
配额 289
亚马逊 DataZone 配额 7
亚马逊 DataZone API 速率限制 290
文档历史记录 295
..... cccxxi

什么是亚马逊 DataZone？

Amazon DataZone 是一项数据管理服务，可让您更快、更轻松地对存储在本地和第三方来源的数据进行分类 AWS、发现、共享和管理。借助 Amazon DataZone，监督组织数据资产的管理员可以使用精细的控制来管理和控制对数据的访问。这些控件有助于确保使用适当级别的权限和上下文进行访问。Amazon DataZone 让工程师、数据科学家、产品经理、分析师和业务用户可以轻松地在整个组织中共享和访问数据，这样他们就可以发现、使用和协作以获得数据驱动的见解。

亚马逊通过集成数据管理服务，包括亚马逊 Redshift、Amazon Athena、Amazon、 QuickSight Glue、Lak AWS e Formation、本地来源、第三方来源 AWS 等， DataZone 帮助您直接向最终用户交付数据并简化架构。

主题

- 我能用 Amazon 做 DataZone什么？
- Amazon 如何 DataZone 支持其他服务并与其他 AWS 服务集成？
- 如何访问亚马逊 DataZone？

我能用 Amazon 做 DataZone什么？

借助 Amazon DataZone，您可以执行以下操作：

- 跨组织边界管理数据访问。借助 Amazon DataZone，您可以根据贵组织的安全法规，帮助确保正确的用户出于正确的目的访问正确的数据，而不必依赖个人证书。您还可以提供数据资产使用情况的透明度，并通过受监管的工作流来批准数据订阅。还可以通过使用情况审计功能来监控项目之间的数据资产。
- 通过共享的数据和工具与数据工作人员建立联系，以获得业务见解。借助 Amazon DataZone，您可以通过跨团队的无缝协作以及提供对数据和分析工具的自助访问来提高业务团队的效率。您可以使用商业术语来搜索、共享和访问存储在 AWS本地或第三方提供商处的分类数据。此外，您还可以使用亚马逊 DataZone 企业术语表进一步了解您要使用的数据。
- 利用机器学习自动执行数据发现和编目操作。借助 Amazon DataZone，您可以减少手动将数据属性输入业务数据目录所花费的时间。数据目录所包含的更丰富的数据还将改善搜索体验。

Amazon 如何 DataZone 支持其他服务并与其他 AWS 服务集成？

Amazon DataZone 支持与其他 AWS 服务的三种集成：

- 生产者数据源-您可以根据存储在 AWS Glue 数据 DataZone 目录和 Amazon Redshift 表和视图中的数据将数据资产发布到亚马逊目录。您也可以手动将对象从亚马逊简单存储服务 (S3) Simple Storage Service 发布到亚马逊 DataZone 目录。
- 使用者工具 – 您可以使用 Amazon Athena 或 Amazon Redshift 查询编辑器来访问和分析数据资产。
- 访问控制和配送——亚马逊 DataZone 支持授予对 AWS Lake Formation 托管 AWS Glue 表格以及亚马逊 Redshift 表格和视图的访问权限。对于所有其他数据资产，亚马逊会向亚马逊 DataZone 发布与您的操作相关的标准事件（例如，批准订阅请求） EventBridge。您可以使用这些标准事件与其他 AWS 服务或第三方解决方案集成，以实现自定义集成。

如何访问亚马逊 DataZone？

您可以通过以下任何 DataZone 一种方式访问 Amazon：

- 亚马逊 SageMaker 管理控制台或亚马逊 DataZone 控制台

  您可以使用亚马逊 SageMaker 管理控制台或亚马逊 DataZone 管理控制台来访问和配置您的亚马逊 DataZone 域名、蓝图和用户。有关更多信息，请参阅 https://console.aws.amazon.com/datazone。此控制台还用于创建 Amazon DataZone 数据门户。
- 亚马逊 DataZone 数据门户

  Amazon DataZone 数据门户是一个基于浏览器的网络应用程序，您可以在其中以自助方式对数据进行分类、发现、管理、共享和分析。数据门户可以通过 AWS IAM Identity Center（ AWS SSO 的继任者）使用身份提供商提供的证书或您的 IAM 凭证对您进行身份验证。您可以通过访问 https://console.aws.amazon.com/datazone 上的亚马逊 DataZone 控制台来获取数据门户 URL。
- 亚马逊 DataZone HTTPS API

  您可以使用亚马逊 DataZone HTTPS API DataZone 以编程方式访问亚马逊，它允许您直接向服务发出 HTTPS 请求。有关更多信息，请参阅 Amazon DataZone API 参考。

亚马逊 SageMaker 以及何时使用亚马逊 SageMaker 与亚马逊 DataZone

Amazon C SageMaker atal og 基于亚马逊构建 DataZone，允许用户集中管理其数据资产。您可以对数据资产进行编目、搜索和发现数据或使用内置生成式人工智能功能创建元数据，也可以直接向 Amazon Q 开发者版提出自然语言问题来查找数据。用户可以在 Amazon Unified Studio 中使用具有精细访问控制的单一权限模型来 SageMaker 统一定义和强制执行访问策略。您可以创建业务词汇表、扩展元数据并构建可与大型团队共享的数据产品（具有精细访问控制）。您还可以查看数据质量分数并发现数据资产的数据血统。

您可以从亚马逊 SageMaker Unified Studio 访问亚马逊 SageMaker 目录。Unified Studio SageMaker 是亚马逊内部的开发体验，它汇集了 AWS 数据、分析、人工智能 (AI) 和机器学习 (ML) 服务。它提供了从单一界面构建、部署、执行和监控工作流的位置。这有助于推动团队间的协作和敏捷开发。

亚马逊 DataZone 术语和概念

Amazon DataZone 是一项数据管理服务，可让您更快、更轻松地对存储在本地和第三方来源的数据进行分类 AWS、发现、共享和管理。借助 Amazon DataZone，监督组织数据资产的管理员和数据管理员可以使用精细的控制来管理和控制对数据的访问。这些控件旨在确保使用适当级别的权限和上下文进行访问。Amazon DataZone 使工程师、数据科学家、产品经理、分析师和业务用户可以更轻松地访问整个组织的数据，以便他们可以发现、使用和协作以获得数据驱动的见解。

在开始使用 Amazon 时 DataZone，了解其关键概念、术语和组成部分非常重要。

主题

- 亚马逊 DataZone 组件
- 什么是 Amazon DataZone 域名？
- Amazon 的 DataZone 项目和环境是什么？
- 什么是亚马逊 DataZone 蓝图？
- Amazon DataZone 库存和发布工作流程是什么？
- Amazon DataZone 订阅和配送流程是什么？
- Amazon 的用户角色 DataZone
- 亚马逊 DataZone 术语

亚马逊 DataZone 组件

Amazon DataZone 包括以下四个主要组成部分：

- 业务数据目录 – 您可以使用此组件根据业务背景对整个组织内的数据进行编目，使组织内部的每个人员都能快速找到和理解数据。
- 发布和订阅工作流 - 您可以使用这些自动化工作流以自助方式保护创建者和使用者之间的数据，并确保组织中的每个人员都能访问适当的数据来实现既定目标。
- 项目和环境
  - Amazon DataZone 项目中包含基于业务用例的人员分组、资产（数据）和工具，用于简化对分析的访问。 AWS 项目提供了可供项目成员用来协作、交换数据和共享资产的区域。默认情况下，项目已配置为仅允许显式添加到项目中的人员访问项目中的数据和分析工具。项目管理根据项目策略生成的资产的所有权，以供数据使用者访问。

- 在 Amazon DataZone 项目中，环境是由零个或多个已配置的资源（例如 Amazon S3 存储桶、AWS Glue 数据库或 Amazon Athena 工作组）组成的集合，一组给定的 IAM 委托人（例如，具有贡献者权限的用户）可以对其进行操作。
- 数据门户（ AWS 管理控制台外）-这是一个基于浏览器的 Web 应用程序，不同的用户可以在其中以自助方式对数据进行编目、发现、管理、共享和分析。数据门户使用 IAM 凭证或您的身份提供商提供的现有凭证通过 AWS IAM Identity Center对用户进行身份验证。

什么是 Amazon DataZone 域名？

您可以使用 Amazon DataZone 域来组织您的资产、用户及其项目。通过将其他 AWS 账户与您的 Amazon DataZone 域名关联，您可以汇集您的数据源。之后，您可以使用元数据表单和术语表将这些数据来源中的资产发布到域的目录，从而提升元数据的完整性和质量。您也可以搜索和浏览这些资产，以查看域中已发布的数据。此外，您可以加入项目来与其他用户协作，订阅资产，并使用项目环境访问分析工具，包括 Amazon Athena 和 Amazon Redshift。无论是为企业创建单个 Amazon DataZone 域名，还是为不同的业务部门创建多个 Amazon 域名，Amazon DataZone 域名都能让您灵活地反映组织结构的数据和分析需求。 DataZone

Amazon 的 DataZone 项目和环境是什么？

Amazon 通过创建基于用例的团队、工具和数据分组， DataZone 使团队和分析用户能够在项目上进行协作。

- 在亚马逊中 DataZone，项目使一组用户能够就各种业务用例进行协作，这些用例涉及发布、发现、订阅和使用亚马逊 DataZone 目录中的数据。项目成员使用 Amazon DataZone 目录中的资产，并使用一个或多个分析工作流程生成新资产。项目支持数据门户中的以下活动：
  - 项目所有者可以添加具有所有者、贡献者、使用者、管理者和查看者权限的成员
  - 项目成员可以是 SSO 用户、SSO 组和 IAM 用户
  - 项目成员可以请求订阅数据目录中的资产

    提供项目订阅批准

	Create delete 项目	Create delete 项目概况	Create delete 环境概况	Create delete 环境	Add/delete 项目的成员	搜索和发现	Create delete 元数据 forms/glossaries	创建数据来源运行并摄取数据	发布数据并创建资产筛选器	请求订阅	Approve/eject 订阅请求	读取 Amazon Athena 和 Amazon Redshift 中的已订阅数据
所有者	由域单元成员管理	由域单元成员管理	由域单元成员管理	由域单元成员管理	支持	是	是	是	是	是	是	是
贡献者	由域单元成员管理	由域单元成员管理	由域单元成员管理	由域单元成员管理	否	是	是	是	是	是	是	是
使用者	由域单元成员管理	由域单元成员管理	由域单元成员管理	由域单元成员管理	否	是	否	否	否	是	否	是
查看者	由域单元成员管理	由域单元成员管理	由域单元成员管理	由域单元成员管理	否	是	否	否	否	否	否	是
管理者	由域单元成员管理	由域单元成员管理	由域单元成员管理	由域单元成员管理	否	是	是	是	是	否	是	是

- 在 Amazon DataZone 项目中，环境是由零个或多个已配置的资源（例如 Amazon S3、 AWS Glue 数据库或 Amazon Athena 工作组）组成的集合，其中有一组可以操作这些资源的 IAM 委托人。环境是通过环境配置文件创建的，这些配置文件是一组预先配置的资源和蓝图，它们提供用于创建环境的可重用模板。环境配置文件定义设置，例如部署环境的 AWS 账户 或区域。

什么是亚马逊 DataZone 蓝图？

用于创建环境的蓝图定义了环境所属项目的成员在处理亚马逊目录中的资产时可以使用哪些 AWS 工具和服务（例如， AWS Glue 或 Ama DataZone zon Redshift）。

在当前版本的 Amazon 中 DataZone，支持以下默认蓝图：

蓝图名称	说明	创建的资源
数据湖蓝图	使 Amazon DataZone 项目成员能够在环境中启动数据湖生成器和使用者服务。 作为消费者，它使亚马逊 DataZone 项目成员能够直接在Amazon Athena和其他Lake查询引擎中访问Lake Formation-managed 资产的 “只读” 副本。Formation-supported 作为制作者，它使亚马逊 DataZone 项目成员能够使用Amazon Athena创建新 LakeFormation-managed 表并将其发布到亚马逊目录中。DataZone	可让用户使用 Amazon Athena 创建和查询 Lake Formation 表。Amazon Athena 工作组 AWS Glue 、具有 “只读” Lake Formation 权限的数据库、“只读” 的 IAM 权限以及对由项目管理的 Amazon S3 的访问权限。 AWS Glue 具有 “创建” 和 “授予” Lake Formation 权限的数据库、“读取” 和 “写入” IAM 权限、带标签的 AWS Glue ETL（提取、转换和加载）。
数据仓库蓝图	作为消费者，该蓝图使亚马逊 DataZone 项目成员能够连接到自己的Amazon Redshift集群，以查询远程数据存储以及创建和存储新的数据集。	访问亚马逊 Redshift 查询编辑器，“读取” 亚马逊 DataZone 目录中订阅的数据源，能够在配置的 Amazon Redshift 集群中创建本地资产。访问 Amazon Redshift 查询编辑

蓝图名称	说明	创建的资源
	作为制作者，该蓝图使亚马逊 DataZone 项目成员能够连接到自己的 Amazon Redshift 集群，以查询远程数据存储、创建新数据集并将其发布到亚马逊 DataZone 目录。	器，“读取” 亚马逊 DataZone 目录中已订阅的数据源，能够从已配置的 Amazon Redshift 集群创建和发布资产。
亚马逊 Sagemaker 蓝图	该蓝图可帮助数据生产者和消费者无缝切换 SageMaker 到 Amazon，在机器学习 (ML) 项目上进行协作，同时对数据和机器学习资产实施访问管理。借助 Amazon DataZone 和 Amazon 之间新的内置集成 SageMaker，数据使用者和创建者可以简化基础设施设置中的机器学习管理，协作开展业务计划，并轻松管理数据和机器学习资产。	您可以创建一个可以在亚马逊中搜索、订阅和发布数据和机器学习资产的亚马逊 SageMaker 域名 DataZone。还可以按照配置订阅和发布 AWS Glue 数据库和湖泊形成。

Amazon DataZone 库存和发布工作流程是什么？

创建项目库存资产

要使用亚马逊对您的数据 DataZone 进行分类，您必须先将您的数据（资产）作为项目库存带到亚马逊 DataZone。为项目创建库存，从而仅允许该项目的成员发现资产。 search/browse 除非明确发布，否则并非所有域用户都可以使用项目清单资产。在当前版本的 Amazon 中 DataZone，您可以通过以下方式向项目库存添加资产：

- 通过数据门户或使用 Amazon DataZone API 创建和运行数据源。在当前版本的亚马逊中 DataZone，你可以为 AWS Glue 和 Amazon Redshift 创建和运行数据源。通过创建和运行 AWS Glue 或 Amazon Redshift 数据源，您可以在选定的项目清单中创建资产，并将其技术元数据从源数据库表或数据仓库中作为库存导入到亚马逊。 DataZone
- 使用 API，您可以根据可用的系统资产类型（AWS Glue、Amazon Redshift、Amazon S3 对象）或自定义资产类型创建资产。

  - 使用 Amazon DataZone API 在项目清单中创建自定义资产类型。自定义资产类型可以包括机器学习模型、控制面板、本地表等。
  - 使用 Amazon DataZone API 根据这些自定义资产类型创建资产。
- 使用 Amazon DataZone 数据门户手动为 S3 对象创建资产。

整理项目库存资产 - 创建项目库存后，数据所有者可以添加或更新业务名称（资产和架构）、描述（资产和架构）、自述文件、术语表术语（资产和架构）和元数据表单，从而使用所需的业务元数据来整理库存资产。您可以通过数据门户网站或使用 Amazon DataZone API 来执行此操作。每次编辑资产时，都会创建一个新的库存版本。

将项目库存资产发布到 Amazon DataZone 目录

使用 Amazon DataZone 对您的数据进行分类的下一步是让域名用户可以发现您项目的库存资产。您可以通过将库存资产发布到 Amazon DataZone 目录来做到这一点。只能将最新版本的库存资产发布到目录，并且仅最新发布的版本在发现目录中处于活动状态。如果库存资产在发布到亚马逊 DataZone 目录后进行了更新，则必须再次明确发布该库存资产，以使最新版本出现在发现目录中。在当前版本的 Amazon 中 DataZone，您可以通过以下方式将项目库存资产发布到亚马逊 DataZone 目录中：

- 通过数据门户或使用亚马逊 DataZone API 将您的项目库存资产手动发布到亚马逊 DataZone 目录。
- 在创建或编辑数据来源的过程中，启用可选的将您的 AWS Glue 资产发布到目录或将您的 Amazon Redshift 资产发布到目录设置，以便在计划或自动化的数据来源运行期间使用。启用此设置后，数据源运行会将资产添加到项目的库存中，然后还会将库存资产发布到 Amazon DataZone 目录。请注意，如果您直接发布，则资产可能不包含任何业务元数据，并且所有域用户都能直接发现资产。您可以通过数据门户或使用 Amazon DataZone API 在数据源上使用此设置。

Amazon DataZone 订阅和配送流程是什么？

将您的资产发布到亚马逊 DataZone 目录后，您的域用户就可以发现这些资产，请求和访问这些资产，并继续使用亚马逊 DataZone 来管理、共享和分析这些资产。

用户可通过代表项目订阅某个资产来请求访问该资产。创建订阅请求后，资产所有者会收到通知，可以查看订阅请求并决定批准还是拒绝该请求。如果数据所有者批准了订阅请求，则向订阅项目授予对该资产的访问权限。

订阅申请获得批准后，亚马逊将 DataZone 开始订阅配送工作流程，通过在 La AWS ke Formation 或 Amazon Redshift 中创建必要的授权，自动将资产添加到项目内的所有适用环境中。这使订阅项目成员能够在其环境中使用某个查询工具（Amazon Athena 或 Amazon Redshift 查询编辑器）来查询资产。

亚马逊 DataZone 只能针对托管资产（包括 AWS Glue 表格和 Amazon Redshift 表格和视图）触发此自动配送逻辑。对于所有其他资产类型（非托管资产），亚马逊 DataZone 无法自动触发配送，而是在 Amazon Eventbridge中发布事件，并在事件负载中包含所有必要的详细信息，以便您可以在亚马逊之外创建必要的补助金。 DataZone亚马逊 DataZone 还提供了 `updateSubscriptionStatus` API，允许您在亚马逊以外的地方完成订阅后更新订阅状态， DataZone 以便亚马逊 DataZone 可以通知项目成员他们可以开始使用资产。

Amazon 的用户角色 DataZone

以下是 Amazon DataZone 用户的主要角色：

- 负责将 Amazon 设置 DataZone 为其组织分析平台的域管理员。

  在亚马逊环境中 DataZone，域管理员 DataZone 在 AWS 账户中安装亚马逊，创建亚马逊 DataZone 域名，配置 AWS 账户关联和身份提供者与亚马逊 DataZone 域的关联。域管理员还使用其他 AWS 服务控制台，例如 AWS 组织和服务目录，来配置 Amazon DataZone。
- 作为 Amazon DataZone （资产发布者和订阅者）执行分析和机器学习任务的主要用户的数据用户。

  数据用户包括数据分析工作人员、数据科学家以及生产和使用数据资产的系统用户。在亚马逊环境中 DataZone，数据用户创建和加入项目和环境，使用预先配置的分析或机器学习工具订阅和使用数据资产，并将输出数据资产发布回亚马逊 DataZone 域名目录以与其他人共享。
- 构建自定义基础设施模板并将 Amazon DataZone 与内部目录或生产系统集成的系统开发人员。

  在 Amazon 环境中 DataZone，系统开发人员以环境提供者的身份构建环境蓝图（基础设施模板）或 Infrastructure-As-Code CI/CD 管道、用于跨环境推广数据资产的数据管道、用于与内部目录集成的目录同步和订阅赠款配送适配器，或者根据需要在 Amazon DataZone API 与内部用户界面或生产系统之间进行集成。
- 数据治理官员，他们拥有组织安全、隐私和其他合规政策的定义和风险，并确保其组织 DataZone 中对亚马逊的使用符合这些定义。

亚马逊 DataZone 术语

域：

Amazon DataZone 域名是将您的资产、用户及其项目连接在一起的组织实体。借助 Amazon DataZone 域名，您可以灵活地反映组织结构的数据和分析需求，无论是为企业创建单个 Amazon DataZone 域还是为不同的业务部门或团队创建多个数据区；域名。

域单元

域单元使您能够轻松地在特定的业务部门和团队下组织资产和其他域实体。要在组织各业务部门内部和各业务部门之间设置安全高效的数据共享，您可以在 Amazon 内创建域单元， DataZone 并允许每个业务部门内的选定用户登录并将其资产共享到目录中。域单位还可用于使资源所有者（例如 AWS 账户所有者）能够对其资源设置 Amazon DataZone 授权权限。域单元提供从账户所有者到域单元所有者的委托授权，他们可以代表账户所有者对环境配置文件（使用蓝图配置创建）设置授权权限。有关更多信息，请参阅 Amazon 中的域名单位和授权政策 DataZone。

授权策略

亚马逊 DataZone 授权策略是亚马逊内部的一组控制措施， DataZone 适用于项目、蓝图、环境、词汇表和元数据表单等实体。这些策略定义了谁可以在 Amazon DataZone 门户中创建这些实体并管理其生命周期。

在 Amazon DataZone 域单位内，您可以将以下授权策略分配给您的用户和群组，以授予他们特定的权限：

- 域单元创建策略
- 项目创建策略
- 项目成员资格策略
- 域单元所有权代入策略
- 项目所有权代入策略

有关更多信息，请参阅 为 Amazon DataZone 域单位内的用户和群组分配授权策略。

在 Amazon DataZone 域单位内，您可以将以下授权策略分配给您的项目，以授予其特定权限：

- 术语表创建策略
- 元数据表单创建策略
- 自定义资产类型创建策略

有关更多信息，请参阅 为 Amazon DataZone 域单位内的项目分配授权策略。

在特定的蓝图配置中，您可以将以下授权策略分配给项目和域单元所有者：

- 使用此蓝图创建环境配置文件-此策略可以分配给 Amazon DataZone 项目，并授权他们使用此蓝图创建环境配置文件。
- 授予使用此蓝图创建环境配置文件所需的权限 - 可将此策略分配给域单元所有者，以便授权他们允许项目使用此蓝图创建环境配置文件。

有关更多信息，请参阅 在 Amazon DataZone 蓝图配置中分配授权策略。

关联账户

将您的 AWS 账户与亚马逊 DataZone 域名关联后，您可以将这些 AWS 账户中的数据发布到亚马逊 DataZone 目录中，并创建亚马逊 DataZone 项目来处理多个 AWS 账户中的数据。账户关联请求只能在拥有 Amazon DataZone 域名的 AWS 账户中发起。只有被邀请账户的管理员用户才能接受 AWS 账户关联请求。 AWS 账户与亚马逊 DataZone 域名关联后，您可以将该账户中的数据源（例如 AWS Glue 目录和 Amazon Redshift）注册到该域名。关联还使 AWS 账户能够创建 Amazon DataZone 项目和环境。

AWS 账户 可以与一个或多个 Amazon DataZone 域名相关联。

数据来源

在 Amazon 中 DataZone，您可以使用数据源将来自源数据库或数据仓库的资产（数据）的技术元数据导入亚马逊 DataZone。在当前版本的亚马逊中 DataZone，你可以为 AWS Glue 和 Amazon Redshift 创建和运行数据源。通过创建数据源，您可以在亚马逊 DataZone 与数据源（AWS Glue Data Catalog 或 Amazon Redshift Warehoushift Warehoushift Warehouse）之间建立连接，从而使您能够读取技术元数据，包括表名称、列名和数据类型。通过创建数据源，您还可以启动初始数据源运行，在 Amazon 中创建新资产或更新现有资产 DataZone。在创建数据来源时或在成功创建数据来源后，您还可以选择为数据来源运行指定计划。

数据来源运行

在亚马逊中 DataZone，数据源运行是亚马逊 DataZone 执行的一项任务，目的是在项目清单中创建资产，也可以选择将项目库存资产发布到亚马逊 DataZone 目录。数据来源运行可以是自动化运行（在最初创建数据来源时启动）、计划运行或手动运行。数据选择标准使您能够微调要提取到项目清单或 Amazon 目录中的现有和未来数据集，以及这些库存或 DataZone目录资产的元数据更新频率。

订阅目标

在 Amazon 中 DataZone，订阅目标允许您访问在项目中订阅的数据。订阅目标指定了位置（例如，数据库或架构）和所需的权限（例如，IAM 角色），亚马逊 DataZone 可以使用这些权限与源数据建立连接并创建必要的授权，以便亚马逊 DataZone 项目的成员可以开始查询他们订阅的数据。

订阅请求

在亚马逊 DataZone，订阅请求是亚马逊 DataZone 项目必须遵循的流程才能获得对特定资产的访问权限。可以批准、拒绝、撤销或授予订阅请求。

资产

在 Amazon DataZone 中，资产是呈现单个物理数据对象（例如，表、仪表板、文件）或虚拟数据对象（例如视图）的实体。

Asset type

资产类型定义了资产在 Amazon DataZone 目录中的呈现方式。资产类型定义特定类型的资产的架构。在创建资产时，将根据资产类型（默认为最新版本）定义的架构来验证资产。当资产更新发生时，Amazon DataZone 会创建一个新的资产版本，并允许亚马逊 DataZone 用户对所有资产版本进行操作。

业务词汇表

在亚马逊中 DataZone，业务词汇表是可能与资产相关的商业术语的集合。业务术语表有助于确保组织内部在执行各种数据分析任务期间使用相同的术语和定义。

可以将业务术语表中的术语添加到资产和列中，以便在搜索过程中对这些属性进行分类或进一步识别这些属性。可以选择词汇表作为元数据表单中与资产关联的字段的值类型。当选择特定术语作为资产元数据表单字段的值时，用户可以搜索业务词汇表术语并找到关联的资产。

元数据表单类型

元数据表单类型是一种模板，用于定义在将资产创建为库存或在 Amazon DataZone 域中发布时收集和保存的元数据。元数据表单类型可以与数据资产关联。元数据表单类型可帮助域管理员定义该域所需的元数据表单，例如合规性信息、法规信息或分类。它使域管理员能够为其资产自定义其他元数据。亚马逊 DataZone 有系统元数据表单类型，例如资产通用详情表单类型、列业务元数据表单类型、胶水表格表单类型、胶水视图表单类型、redshift-view-form-table 表单类型、redshift-view-form-type、s3 对象集合表单类型、订阅条款表单类型和建议表单类型。

元数据表单

在亚马逊中 DataZone，元数据表单定义了在将资产创建为库存或在亚马逊 DataZone 域中发布时收集和保存的元数据。元数据表单定义是域管理员在目录域中创建的。元数据表单定义包含一个或多个字段定义，并支持布尔值、日期、小数、整数、字符串和业务术语表字段值数据类型。

域管理员通过将元数据表单添加到其域，来将元数据表单应用于其域中的资产。之后，资源发布者会在元数据表单中提供任何可选和必填的字段值。

Project

在亚马逊中 DataZone，项目允许一组用户就各种业务用例进行协作，这些用例涉及在项目清单中创建资产，从而使所有项目成员都能发现这些资产，然后发布、发现、订阅和使用亚马逊

DataZone 目录中的资产。项目成员使用 Amazon DataZone 目录中的资产，并使用一个或多个分析工作流程生成新资产。项目成员可以是所有者、贡献者、使用者、管理者和查看者。

	Create delete 项目	Create delete 项目概况	Create delete 环境概况	Create delete 环境	Add/delete 项目的成员	搜索和发现	Create delete 元数据 forms/glossarie	创建数据来源运行并摄取数据	发布数据并创建资产筛选器	请求订阅	Appror eject 订阅请求	读取 Amazon Athena 和 Amazon Redshift 中的已订阅数据
所有者	由域单元成员管理	由域单元成员管理	由域单元成员管理	由域单元成员管理	支持	是	是	是	是	是	是	是
贡献者	由域单元成员管理	由域单元成员管理	由域单元成员管理	由域单元成员管理	否	是	是	是	是	是	是	是
使用者	由域单元成员管理	由域单元成员管理	由域单元成员管理	由域单元成员管理	否	是	否	否	否	是	否	是
查看者	由域单元成员管理	由域单元成员管理	由域单元成员管理	由域单元成员管理	否	是	否	否	否	否	否	是
管理者	由域单元成员管理	由域单元成员管理	由域单元成员管理	由域单元成员管理	否	是	是	是	是	否	是	是

项目所有者可以将其他用户作为所有者或贡献者进行添加或删除，也可以修改或删除项目。可以使用策略定义针对贡献者的其他限制。当用户创建一个项目时，他们将成为该项目的第一个所有者。

环境

环境是一个集合，其中包含已配置的资源（例如，Amazon S3 存储桶、 AWS Glue 数据库或 Amazon Athena 工作组）和一组可操作这些资源的给定 IAM 主体（具有分配的贡献者权限）。每个环境还可具有用户主体，他们有权访问资源并能通过订阅和履行来访问数据。环境旨在将可操作链接存储到 AWS 服务、外部 IDE 和控制台。项目成员可以通过环境中配置的深度链接访问 Amazon Athena 控制台等服务。可以进一步将项目中的 SSO 用户和 IAM 用户范围缩小到 use/access 特定的环境。

环境配置文件

在 Amazon 中 DataZone，环境配置文件是您可以用来创建环境的模板。环境配置文件是使用蓝图创建的。

利用环境配置文件，域管理员可以用预先配置的参数封装蓝图，之后数据工作人员可以通过选择现有环境配置文件并指定新环境的名称来快速创建任意数量的新环境。这使数据工作人员能够高效地管理其项目和环境，同时确保他们符合域管理员实施的数据治理策略。

蓝图

用于创建环境的蓝图定义了环境所属项目的成员在处理亚马逊目录中的资产时可以使用哪些 AWS 工具和服务（例如， AWS Glue 或 Ama DataZone zon Redshift）。

在当前版本的 Amazon DataZone 中，支持以下默认蓝图：

- 数据湖蓝图
- 数据仓库蓝图
- Amazon Sagemaker 蓝图

用户配置文件

用户个人资料代表 Amazon DataZone 用户。Amazon DataZone 支持 IAM 角色和 SSO 身份，以便出于不同的目的与亚马逊 DataZone 管理控制台和数据门户进行交互。域管理员使用 IAM 角色在亚马逊 DataZone 管理控制台中执行与域相关的初始管理工作，包括创建新的 Amazon DataZone 域名、配置元数据表单类型和实施策略。数据工作者通过 Identity Center 使用他们的 SSO 企业身份登录亚马逊 DataZone 数据门户并访问他们拥有成员资格的项目。

组配置文件

群组资料代表一组 Amazon DataZone 用户。可以手动创建组，也可以将组映射到企业客户的 Active Directory 组。在 Amazon 中 DataZone，群组有两个用途。首先，一个小组可以映射到组

织结构图中的用户团队，从而在有新员工加入或离开团队时减少 Amazon DataZone 项目负责人的管理工作。其次，企业管理员使用 Active Directory 群组来管理和更新用户状态，因此亚马逊 DataZone 域管理员可以使用这些群组成员资格来实施亚马逊 DataZone 域名政策。

域管理员

在亚马逊中 DataZone，创建亚马逊 DataZone 域名的 IAM 委托人是该域的默认域管理员。Amazon 中的域管理员为域 DataZone 执行关键功能，包括创建域、分配其他域管理员、添加数据源和订阅目标、创建项目和环境以及分配项目所有者。

发布者

在亚马逊 DataZone，出版商将资产发布到亚马逊 DataZone 目录中，并且可以编辑他们发布的资产的元数据。如果获得此权限，出版商可以批准或拒绝其在 Amazon DataZone 目录中发布的资产的订阅请求。

订阅者

在亚马逊中 DataZone，订阅者是一个想要查找、访问和使用亚马逊 DataZone 目录中的资产的亚马逊 DataZone 项目。

AWS 账户 owner

在亚马逊中 DataZone， AWS 账户 所有者在其中创建角色、策略和权限 AWS 账户 ，使这些角色和权限 AWS 账户 能够与亚马逊 DataZone 域名关联。

Amazon 有哪些新内容 DataZone？

本节 DataZone 按发布日期介绍Amazon的新功能和改进。

主题

- 2024
- 2023

2024

Amazon DataZone 推出针对订阅请求的元数据强制执行规则

发布日期：2024 年 11 月 20 日

Amazon DataZone 针对订阅请求的新元数据强制执行规则使域单位所有者能够为数据使用者制定明确的元数据要求，简化访问请求并增强数据治理，从而加强数据治理。此功能使组织能够遵守其元数据标准、实施自定义工作流和提供一致、受管控的数据访问体验。有关更多信息，请参阅 针对订阅请求的元数据强制规则。

亚马逊 DataZone 定制 AWS 服务蓝图现在 SageMaker 为亚马逊 DataZone 项目提供了全新的设置体验

发布日期：2024 年 11 月 15 日

借助亚马逊 DataZone 定制 AWS 服务打印，您可以将现有的亚马逊 SageMaker 域名迁移到亚马逊 DataZone。借助此功能，管理员现在可以通过从 Amazon SageMaker 域中导入其现有授权用户、安全配置和策略来设置 Amazon DataZone 项目。有关更多信息，请参阅设置 SageMaker 资产（管理员指南）。

Amazon DataZone 推出对定制 AWS 服务蓝图的 AWS CloudFormation 支持

发布日期：2024 年 9 月 12 日

Amazon DataZone 增加了对定制 AWS 服务蓝图的 AWS CloudFormation 支持。这项新功能使您 AWS CloudFormation 能够使用在 Amazon 中自动创建环境 DataZone。借助自定义蓝图，管理员现在可以使用现有 IAM 角色将 Amazon 无缝集成 DataZone 到现有的数据管道中，将数据资产发布到

Amazon DataZone 目录，从而促进这些资产的受控共享，并增强对整个基础设施的治理。有关更多信息，请参阅 Amazon DataZone 资源类型参考。

亚马逊 DataZone 推出域名单位和授权政策

发布日期：2024 年 8 月 12 日

Amazon DataZone 推出了一组新的数据治理功能，称为域单元和授权策略，使客户能够创建业务部门/团队级别的组织并根据其业务需求管理策略。通过添加域单元，用户可以组织、创建、搜索和查找与业务部门或团队关联的数据资产和项目。通过授权策略，这些域单位用户可以设置访问策略，以便在 Amazon 中创建项目、术语表和使用计算资源。 DataZone有关更多信息，请参阅 Amazon 中的域名单位和授权政策 DataZone。

亚马逊 DataZone 推出数据产品

发布日期：2024 年 8 月 5 日

Amazon DataZone 推出了数据产品，可将数据资产分组为针对特定业务用例量身定制的定义明确、独立的软件包。例如，营销分析数据产品可以捆绑营销活动数据、管道数据和客户数据等各种数据资产。借助数据产品，客户可以简化发现和订阅流程，使它们与业务目标保持一致，并减少处理单个资产时的冗余。有关更多信息，请参阅 亚马逊 DataZone 数据产品。

亚马逊 DataZone 推出精细访问控制功能

发布日期：2024 年 7 月 2 日

亚马逊引入 DataZone 了精细的访问控制，使您可以精细控制亚马逊 DataZone业务数据目录中的数据资产，跨数据湖和数据仓库。利用此新功能，数据所有者现在可以仅允许访问行级和列级的特定数据记录，而不是授予对整个数据资产的访问权限。例如，如果您的数据列包含个人身份信息（PII）等敏感信息，则可以仅允许访问必要的列，从而在确保敏感信息受到保护的同时仍允许访问非敏感数据。同样，您可以控制行级访问权限，只允许用户查看与其角色或任务相关的记录。有关更多信息，请参阅对 Amazon 中数据的精细访问控制 DataZone。

亚马逊 DataZone 推出数据血统功能

发布日期：2024 年 6 月 27 日

Amazon DataZone 推出数据沿袭预览版，帮助客户可视化来自 OpenLineage支持系统的系统或 API 的世系事件，并跟踪数据从源头到消费的移动。使用与亚马逊 OpenLineage兼容 DataZone的功能 APIs，域管理员和数据制作者可以捕获和存储超出亚马逊可用范围的谱系事件 DataZone，包括

Amazon S3、G AWS lue和其他服务中的转换。此外，Amazon DataZone 版本与每个事件保持一致，使用户能够在任何时间点可视化血统或比较资产或任务历史的转换。此历史世系可让用户更深入地了解数据的演变过程，这对于故障排除、审计和验证数据资产的完整性至关重要。有关更多信息，请参阅 Amazon 中的数据谱系 DataZone。

亚马逊 DataZone 推出定制 AWS 服务蓝图

发布日期：2024 年 6 月 17 日

借助定制 AWS 服务蓝图，如果您拥有包括 IAM 角色、数据湖、数据网格、Amazon S3 存储桶和 Amazon Redshift 集群在内的现有 AWS 资源，则现在可以使用自己的自定义 IAM 角色指定对这些现有资源的权限，这样您的亚马逊 DataZone 用户就可以利用发布和订阅来共享和管理这些资源。借助定制 AWS 服务蓝图，Amazon DataZone 管理员可以使用自己的自定义角色配置 AWS 服务环境。他们可以为这些 AWS 服务环境配置操作链接，从而提供对其任何现有 AWS 资源的联合访问权限。他们还可以在这些自定义 AWS 服务环境中配置订阅目标和数据源。管理员可以在自己的 Amazon DataZone 域账户中或他们想要发布、订阅、发现或管理数据的任何关联账户中设置 AWS 服务环境。有关更多信息，请参阅 Amazon DataZone 定制 AWS 服务蓝图。

数据来源创建流程的增强功能

发布日期：2024 年 6 月 10 日

Amazon DataZone 对数据源创建流程进行了增强，以简化数据生产者的访问管理。通过这些更新，当数据创建者创建用于发布其 AWS Glue 和 Amazon Redshift 资产的数据源时，亚马逊会向项目成员 DataZone 授予只读权限。创建 AWS Glue 数据源时，Amazon DataZone 会自动向用于创建数据源的环境的 IAM 角色授予 “只读” 权限，允许访问相关 AWS Glue 数据库中的所有表。同样，对于亚马逊 Redshift 数据源，亚马逊 DataZone 授予对数据源中使用的亚马逊 Redshift 架构中所有表的 “只读” 访问权限。有关更多信息，请参阅为创建并运行 Amazon DataZone 数据源 AWS Glue Data Catalog和为亚马逊 Redshift 创建并运行亚马逊 DataZone 数据源。

亚马逊 DataZone 启动与亚马逊的整合 SageMaker

发布日期：2024 年 5 月 6 日

亚马逊 DataZone 推出与亚马逊的集成， SageMaker以帮助数据生产者和消费者无缝切换 SageMaker到亚马逊，在机器学习 (ML) 项目上进行协作，同时对数据和机器学习资产实施访问管理。借助 Amazon DataZone 和 Amazon 之间新的内置集成 SageMaker，数据使用者和创建者可以简化基础设施设置中的机器学习管理，协作开展业务计划，并轻松管理数据和机器学习资产。有关更多信息，请参阅亚马逊 DataZone 内置蓝图和Amazon 中的关联账户 DataZone。

亚马逊 DataZone 推出与 L AWS ake Formation 混合访问模式的集成

发布日期：2024 年 4 月 3 日

亚马逊推 DataZone 出了与 AWS Lake Formation混合访问模式的集成。这种集成使您能够轻松地通过亚马逊发布和共享您的 AWS Glue 表 DataZone，而无需先在 AWS Lake Formation 中注册它们。首先，管理员在 Amazon DataZone 控制台中启用`DefaultDataLake`蓝图下的数据位置注册设置。然后，当数据使用者订阅通过 IAM 权限管理的 AWS Glue 表时，亚马逊 DataZone 首先以混合模式注册该表的 Amazon S3 位置，然后通过 La AWS ke Formation 管理该表的权限，向数据使用者授予访问权限。这样可以确保使用新授予的 La AWS ke Formation 权限继续存在表上的 IAM 权限，而不会中断任何现有工作流程。有关更多信息，请参阅亚马逊 DataZone 与 AWS Lake Formation 混合模式。

亚马逊 DataZone 推出与 Glue 数据 AWS 质量的集成

发布日期：2024 年 4 月 3 日

亚马逊 DataZone 推出与 AWS Glue 数据质量的集成 APIs ，并提供集成来自第三方数据质量解决方案的数据质量指标的服务。新的集成使您能够将Glue AWS 数据质量分数自动发布到亚马逊 DataZone 业务数据目录中。Amazon DataZone APIs 可用于从第三方来源获取质量指标。发布后，数据使用者可以轻松搜索数据资产，查看精细的质量指标并识别失败的检查和规则，从而加快制定业务决策。有关更多信息，请参阅Amazon 的数据质量 DataZone。

Amazon 中描述的 AI 推荐正式发布版 DataZone

发布日期：2024 年 3 月 27 日

Amazon DataZone 宣布正式发布基于人工智能的新生成功能，通过丰富业务数据目录来改善数据发现、数据理解和数据使用。只需单击一下，数据创建者即可生成全面的业务数据描述和上下文，突出显示有影响力的列，并包含有关分析应用场景的建议。此次发布增加了 APIs 对数据生产者可用于以编程方式生成资产描述的支持。有关更多信息，请参阅 在 Amazon 中使用机器学习和生成人工智能 DataZone。

亚马逊 DataZone 推出亚马逊 Redshift 集成增强功能

发布日期：2024 年 3 月 21 日

亚马逊对其亚马逊Redshift集成 DataZone 进行了多项增强，简化了发布和订阅亚马逊Redshift表格和视图的过程。这些更新简化了数据创建者和使用者的体验，使他们能够使用 Amazon DataZone 管理员

提供的预配置凭证和连接参数快速创建数据仓库环境。此外，这些增强功能使管理员能够更好地控制谁可以使用其 AWS 账户和 Amazon Redshift 集群中的资源以及用于什么目的。

- 蓝图配置：启用 DefaultDataWarehouseBlueprint 蓝图后，可以将管理项目分配给已启用的蓝图，从而控制哪些项目可以使用您账户中的 DefaultDataWarehouseBlueprint 蓝图来创建环境配置文件。您还可以DefaultDataWarehouseBlueprint通过提供诸如集群、数据库和 AWS 密钥之类的参数来创建参数集。您也可以从 Amazon DataZone 控制台中创建 AWS 密钥。
- 环境配置文件：创建环境配置文件时，您可以选择提供自己的 Amazon Redshift 参数，也可以使用蓝图配置中的某个参数集。如果您选择使用在蓝图配置中创建的参数集，则 AWS 密钥只需要AmazonDataZoneDomainAmazonDataZoneProject标签（只有当您选择在环境配置文件中提供自己的参数集时，才需要标记）。在环境配置文件中，您可以指定已授权项目的列表。仅已授权项目能够使用此环境配置文件来创建数据仓库环境。还可以指定允许发布哪些已获数据授权的项目。目前，您可以选择下列选项之一：1) 从任何架构发布，2) 从默认环境架构发布，3) 不允许发布。
- 环境：数据创建者或使用者现在可以选择环境配置文件来创建环境，而无需提供自己的 Amazon Redshift 参数，包括 AWS 密钥、集群、工作组和数据库。这些参数将从环境配置文件移植到环境中。除了创建环境外，Amazon DataZone 现在还会为环境创建默认架构。项目成员对此架构具有读写访问权限，并且可以通过运行在创建环境期间创建的默认数据来源，轻松地将在此架构中创建的任何表发布到目录。用于创建环境的 Amazon Redshift 参数也可用于创建新的数据来源（而不是让数据创建者在创建数据来源期间提供自己的参数）。

AWS Amazon 的 Cloud Formation DataZone

发布日期：2024 年 1 月 18 日

现在，Amazon 的用户 DataZone 可以利用它 AWS CloudFormation 来有效地建模和管理一套亚马逊 DataZone 资源。此方法有助于一致地预置资源，并支持通过基础设施即代码来进行生命周期管理。利用自定义模板，您可以精确地定义所需的资源及其相互依赖项。有关更多信息，请参阅 Amazon DataZone 资源类型参考。

直接将 IAM 委托人添加为 Amazon DataZone 项目的成员

发布日期：2024 年 1 月 5 日

现在，您可以将 IAM 委托人添加为项目成员，即使这些 IAM 委托人尚未登录 Amazon DataZone（之前的要求）。在域管理员或 IT 管理员将 iam:GetUser 和 iam:GetRole 添加到域的域执行角色后，项目所有者只需提供 IAM 角色或 IAM 用户的 Amazon 资源名称（ARN）即可将 IAM 主体添加为成员。IAM 委托人仍然必须拥有访问 Amazon 所需的 IAM 权限 DataZone ，这些权限可以在 IAM 控制台中进行配置。有关更多信息，请参阅 向项目添加成员。

对来自数据门户的自定义资产类型的支持

发布日期：2024 年 1 月 5 日

对自定义资产的支持使Amazon DataZone 能够通过数据门户对非结构化数据（包括仪表板、查询和模型）的资产进行分类，从而使您可以更轻松地直接在数据门户中添加自定义资产以及之前提供的 API 支持。通过在 Amazon 中创建 DataZone、更新和发布自定义资产，您可以共享、查找、订阅任何类型的资产，并构建可管理这些资产的业务工作流程。有关更多信息，请参阅 在 Amazon 中创建自定义资产类型 DataZone。

2023

删除域

发布日期：2023 年 12 月 27 日

这是一项功能，可让您更轻松地删除域。现在，即使域不为空（如包含项目、环境、资产、数据来源等），也可以继续删除域。有关更多信息，请参阅 删除 Amazon DataZone 域名。

混合模式

发布日期：2023 年 12 月 22 日

亚马逊 DataZone 增加了对La AWS ke Formation混合模式的支持。有了这种支持，如果您将 AWS Glue表发布到亚马逊 DataZone ，其 AWS S3位置在混合模式下注册在Lake Formation中，则亚马逊 DataZone 会将此表视为托管资产，并且可以管理该表的订阅授权。在此功能发布之前，亚马逊 DataZone 会将此表视为非托管资产，也就是说，亚马逊 DataZone 将无法授予对该表的订阅。有关更多信息，请参阅 为亚马逊配置 Lake Formation 权限 DataZone。

HIPAA 资格

发布日期：2023 年 12 月 14 日

亚马逊 DataZone 现已符合《1996 年美国健康保险流通与责任法案》(HIPAA)。要查看符合 HIPAA 标准的 AWS 服务列表，请参阅 https://aws.amazon.com/compliance/hipaa-eligible-services-reference/。

Amazon 中描述的 AI 建议 DataZone （预览版）

发布日期：2023 年 11 月 28 日

AWS 宣布在 Amazon 中预览基于人工智能的新生成功能，该功能通过丰富业务数据目录 DataZone来改善数据发现、数据理解和数据使用。只需单击一下，数据创建者即可生成全面的业务数据描述和上下文，突出显示有影响力的列，并包含有关分析应用场景的建议。借助 Amazon 中描述的人工智能建议 DataZone，数据使用者可以识别分析所需的数据表和列，从而提高数据可发现性并减少与数据生产者的 back-and-forth通信。预览版适用于在以下 AWS 区域配置的 Amazon DataZone 域名：美国东部（弗吉尼亚北部）、美国西部（俄勒冈）。有关更多信息，请参阅 在 Amazon 中使用机器学习和生成人工智能 DataZone。

DefaultDataLake 蓝图增强

发布日期：2023 年 11 月 20 日

Amazon为 DefaultDataLake 蓝图添加 DataZone 了一项增强功能，使您可以更好地控制谁可以从您的 AWS 账户发布哪些数据。推出此功能时引入了两项关键更改。

- 在控制台中，启用 DefaultDataLake 蓝图后，您可以通过将管理项目分配给已启用的 DefaultDataLake 蓝图来控制哪些项目可以使用您账户中的蓝图来创建环境配置文件。
- 第二项更改是门户内的更改。如果您使用 DefaultDataLake 蓝图创建环境配置文件，则还可以选择允许使用该环境配置文件创建环境的授权项目。默认情况下，允许所有项目使用数据湖环境配置文件，但您可以将环境配置文件限制为特定项目，还可以控制可使用通过配置文件创建的环境发布哪些数据。

有关更多信息，请参阅 创建环境配置文件。

Amazon 支持的区域 DataZone

在当前版本中，以下 AWS 区域支持 DataZone Amazon：

- 美国东部（俄亥俄州）
- 美国东部（弗吉尼亚州北部）
- 美国西部（俄勒冈州）
- 亚太地区（香港）
- Asia Pacific (Mumbai)
- 亚太地区（新加坡）
- 亚太地区（悉尼）
- 亚太地区（马来西亚）
- 亚太地区（泰国）
- 亚太地区（东京）
- 亚太地区 (首尔)
- 加拿大（中部）
- 欧洲地区（法兰克福）
- 欧洲（苏黎世）
- 欧洲地区（爱尔兰）
- 欧洲地区（伦敦）
- 欧洲地区（巴黎）
- 欧洲地区（斯德哥尔摩）
- 南美洲（圣保罗）

设置亚马逊 DataZone

设置亚马逊 DataZone，你必须有一个 AWS 账户并为亚马逊设置所需的 IAM 政策和权限 DataZone。

设置亚马逊后 DataZone 权限，建议您完成中的步骤 入门 该部分将引导您完成创建亚马逊的过程 DataZone 域、获取数据门户 URL 和基本亚马逊 DataZone 数据生产者和数据使用者的工作流程。

主题

- 注册参加 AWS 账户
- 配置使用亚马逊所需的 IAM 权限 DataZone 管理控制台
- 配置使用亚马逊所需的 IAM 权限 DataZone 数据门户
- 设置 AWS 亚马逊 IAM 身份中心 DataZone

注册参加 AWS 账户

要开始使用 AWS，你需要一个 AWS 账户。有两种注册方式 AWS。如果您想完全控制账户配置和内容的存储位置，或者如果您有 HIPAA 或 FedRAMP 等受监管的工作负载，请使用 注册参加 AWS （高级）。如果您是新用户并想使用您已经拥有的登录名登录 AWS 并创建一个 AWS 选择中包含预先配置的默认值的环境 AWS 区域，使用 注册参加 AWS。

配置使用亚马逊所需的 IAM 权限 DataZone 管理控制台

为了访问和配置您的亚马逊 DataZone 域、蓝图和用户，以及创建亚马逊 DataZone 数据门户，你必须使用亚马逊 DataZone 管理控制台。

必须完成以下步骤才能配置所需的 and/or 任何想要使用亚马逊的用户、群组或角色的可选权限 DataZone 管理控制台。

用于设置 IAM 权限以使用管理控制台的过程

- 为亚马逊用户、群组或角色附加必需和可选政策 DataZone 控制台访问权限
- 为 IAM 权限创建自定义策略以启用亚马逊 DataZone 服务控制台简化了角色创建
- 为管理与亚马逊关联的账户的权限创建自定义政策 DataZone 域
- （可选）为创建自定义策略 AWS 身份中心权限添加和删除 SSO 用户和 SSO 组对亚马逊的访问权限 DataZone 域

- （可选）将您的 IAM 委托人添加为密钥用户以创建您的亚马逊 DataZone 带有来自的客户管理密钥的域 AWS Key Management Service（KMS）

为亚马逊用户、群组或角色附加必需和可选政策 DataZone 控制台访问权限

完成以下过程，将必需和可选的自定义策略附加到用户、组或角色。有关更多信息，请参阅 AWS 亚马逊的管理政策 DataZone。

1. 登录 AWS 管理控制台，然后在以下位置打开 IAM 控制台 https://console.aws.amazon.com/iam/。
2. 在导航窗格中，选择策略。
3. 选择要附加到用户、组或角色的以下策略。
   - 在策略列表中，选中旁边的复选框 AmazonDataZoneFullAccess。您可以使用 Filter 菜单和搜索框来筛选策略列表。有关更多信息，请参阅 AWS 管理策略：AmazonDataZoneFullAccess。
   - （可选）为 IAM 权限创建自定义策略以启用亚马逊 DataZone 服务控制台简化了角色创建。
   - （可选）为创建自定义策略 AWS 身份中心权限，可添加和删除 SSO 用户和 SSO 群组对您的亚马逊的访问权限 DataZone 域。
4. 选择 Actions（操作），然后选择 Attach（附加）。
5. 选择要将该策略附加到的用户、组或角色。您可以使用 Filter（筛选条件）菜单和搜索框来筛选委托人实体列表。选择用户、组或角色后，选择附加策略。

为 IAM 权限创建自定义策略以启用亚马逊 DataZone 服务控制台简化了角色创建

完成以下步骤以创建自定义内联策略，以获得启用 Amazon 所需的权限 DataZone 在中创建必要的角色 AWS 代表您管理控制台。

> **Note**
>
> 有关配置权限以允许创建服务角色的最佳实践信息，请参阅 https://docs.aws.amazon.com/IAM/latest/UserGuide/id_roles_create_for-service.html。

1. 登录 AWS 管理控制台，然后在以下位置打开 IAM 控制台 https://console.aws.amazon.com/iam/。

2. 在导航窗格中，选择用户或用户组。
3. 在列表中，请选择要在其中嵌入策略的用户或组的名称。
4. 选择 Permissions (权限) 选项卡，然后展开 Permissions policies (权限策略) 部分（如有必要）。
5. 选择添加权限，然后选择创建内联策略链接。
6. 在创建策略屏幕上的策略编辑器部分中，选择 JSON。

   使用以下 JSON 语句创建策略文档，然后选择下一步。

   JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "iam:CreatePolicy",
                "iam:CreateRole"
            ],
            "Resource": [
                "arn:aws:iam::*:policy/service-role/AmazonDataZone*",
                "arn:aws:iam::*:role/service-role/AmazonDataZone*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": "iam:AttachRolePolicy",
            "Resource": "arn:aws:iam::*:role/service-role/AmazonDataZone*",
            "Condition": {
                "ArnLike": {
                    "iam:PolicyARN": [
                        "arn:aws:iam::aws:policy/AmazonDataZone*",
                        "arn:aws:iam::*:policy/service-role/AmazonDataZone*"
                    ]
                }
            }
        }
    ]
}
```

7. 在查看策略屏幕上，输入此策略的名称。如果您对该策略感到满意，请选择 Create policy (创建策略)。确保屏幕顶部的红框中没有显示错误。更正报告的任何错误。

为管理与亚马逊关联的账户的权限创建自定义政策 DataZone 域

完成以下步骤以创建自定义内联策略，以便在关联策略中获得必要的权限 AWS 帐户列出、接受和拒绝域的资源共享，然后在关联账户中启用、配置和禁用环境蓝图。启用可选亚马逊 DataZone 在蓝图配置期间可以使用服务控制台简化的角色创建，您还必须 为 IAM 权限创建自定义策略以启用亚马逊 DataZone 服务控制台简化了角色创建。

> **Note**
>
> 有关配置权限以允许创建服务角色的最佳实践信息，请参阅 https://docs.aws.amazon.com/IAM/latest/UserGuide/id_roles_create_for-service.html。

1. 登录 AWS 管理控制台，然后在以下位置打开 IAM 控制台 https://console.aws.amazon.com/iam/。
2. 在导航窗格中，选择用户或用户组。
3. 在列表中，请选择要在其中嵌入策略的用户或组的名称。
4. 选择 Permissions (权限) 选项卡，然后展开 Permissions policies (权限策略) 部分（如有必要）。
5. 选择添加权限，然后选择创建内联策略链接。
6. 在创建策略屏幕上的策略编辑器部分中，选择 JSON。使用以下 JSON 语句创建策略文档，然后选择下一步。

   JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "datazone:ListEnvironmentBlueprintConfigurations",
                "datazone:PutEnvironmentBlueprintConfiguration",
```

```
                "datazone:GetDomain",
                "datazone:ListDomains",
                "datazone:GetEnvironmentBlueprintConfiguration",
                "datazone:ListEnvironmentBlueprints",
                "datazone:GetEnvironmentBlueprint",
                "datazone:ListAccountEnvironments",
                "datazone:DeleteEnvironmentBlueprintConfiguration"
            ],
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": "iam:PassRole",
            "Resource": [
                "arn:aws:iam::*:role/AmazonDataZone",
                "arn:aws:iam::*:role/service-role/AmazonDataZone*"
            ],
            "Condition": {
                "StringEquals": {
                    "iam:passedToService": "datazone.amazonaws.com"
                }
            }
        },
        {
            "Effect": "Allow",
            "Action": "iam:AttachRolePolicy",
            "Resource": "arn:aws:iam::*:role/service-role/AmazonDataZone*",
            "Condition": {
                "ArnLike": {
                    "iam:PolicyARN": [
                        "arn:aws:iam::aws:policy/AmazonDataZone*",
                        "arn:aws:iam::*:policy/service-role/AmazonDataZone*"
                    ]
                }
            }
        },
        {
            "Effect": "Allow",
            "Action": "iam:ListRoles",
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": [
```

```
                "iam:CreatePolicy",
                "iam:CreateRole"
            ],
            "Resource": [
                "arn:aws:iam::*:policy/service-role/AmazonDataZone*",
                "arn:aws:iam::*:role/service-role/AmazonDataZone*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "ram:AcceptResourceShareInvitation",
                "ram:RejectResourceShareInvitation",
                "ram:GetResourceShareInvitations"
            ],
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": [
                "s3:ListAllMyBuckets",
                "s3:ListBucket",
                "s3:GetBucketLocation"
            ],
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": "s3:CreateBucket",
            "Resource": "arn:aws:s3:::amazon-datazone*"
        }
    ]
}
```

7. 在查看策略屏幕上，输入此策略的名称。如果您对该策略感到满意，请选择 Create policy (创建策略)。确保屏幕顶部的红框中没有显示错误。更正报告的任何错误。

（可选）为创建自定义策略 AWS 身份中心权限添加和删除 SSO 用户和 SSO 组对亚马逊的访问权限 DataZone 域

完成以下步骤以创建自定义内联策略，以获得添加和删除 SSO 用户和 SSO 组对您的亚马逊的访问权限所需的权限 DataZone 域。

1. 登录 AWS 管理控制台，然后在以下位置打开 IAM 控制台 https://console.aws.amazon.com/iam/。
2. 在导航窗格中，选择用户或用户组。
3. 在列表中，请选择要在其中嵌入策略的用户或组的名称。
4. 选择 Permissions (权限) 选项卡，然后展开 Permissions policies (权限策略) 部分（如有必要）。
5. 选择添加权限和创建内联策略。
6. 在创建策略屏幕上的策略编辑器部分中，选择 JSON。

   使用以下 JSON 语句创建策略文档，然后选择下一步。

   JSON

```
{
  "Version":"2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "sso:GetManagedApplicationInstance",
        "sso:ListProfiles",
        "sso:AssociateProfile",
        "sso:DisassociateProfile",
        "sso:GetProfile"
      ],
      "Resource": "*"
    }
  ]
}
```

7. 在查看策略屏幕上，输入此策略的名称。如果您对该策略感到满意，请选择 Create policy (创建策略)。确保屏幕顶部的红框中没有显示错误。更正报告的任何错误。

（可选）将您的 IAM 委托人添加为密钥用户以创建您的亚马逊 DataZone 带有来自的客户管理密钥的域 AWS Key Management Service（KMS）

在您可以选择创建您的亚马逊之前 DataZone 使用来自的客户管理密钥 (CMK) 的域 AWS 密钥管理服务 (KMS)，完成以下程序，使您的 IAM 委托人成为您的 KMS 密钥的用户。

1. 登录 AWS 管理控制台，然后在以下位置打开 KMS 控制台 https://console.aws.amazon.com/kms/。
2. 要查看您账户中自己所创建和管理的密钥，请在导航窗格中选择 Customer managed keys (客户托管密钥)。
3. 在 KMS 密钥列表中，选择要检查的 KMS 密钥的别名或密钥 ID。
4. 添加或删除密钥用户，以及允许或禁止外部用户 AWS 账户要使用 KMS 密钥，请使用中的控件 关键用户 页面的部分。密钥用户可以在加密操作（如加密、解密、重新加密和生成数据密钥）中使用 KMS 密钥。

配置使用亚马逊所需的 IAM 权限 DataZone 数据门户

亚马逊 DataZone 数据门户（外部 AWS Management Console）是一个基于浏览器的 Web 应用程序，用户可以在其中以自助方式访问目录、发现、管理、共享和分析数据。数据门户通过 IAM 证书或身份提供商提供的现有证书对用户进行身份验证 AWS IAM 身份中心。

您必须完成以下程序，才能为想要使用亚马逊的任何用户、群组或角色配置所需的权限 DataZone 数据门户或目录：

用于配置 IAM 权限以使用数据门户的过程

- 将所需的策略附加到亚马逊的用户、群组或角色 DataZone 数据门户访问权限
- 将所需的策略附加到亚马逊的用户、群组或角色 DataZone 目录访问权限
- 为亚马逊的用户、群组或角色附加可选政策 DataZone 如果您的域是使用来自的客户管理密钥加密的，则可以访问数据门户或目录 AWS Key Management Service（KMS）

将所需的策略附加到亚马逊的用户、群组或角色 DataZone 数据门户访问权限

你可以访问亚马逊 DataZone 使用您的任一数据门户 AWS 凭据或您的单点登录 (SSO) 证书。按照以下部分中的说明设置使用您的访问数据门户所需的权限 AWS 证书。有关使用亚马逊的更多信息 DataZone 使用 SSO，请参阅 设置 AWS 亚马逊 IAM 身份中心 DataZone。

> **Note**
>
> 您的域中只有 IAM 委托人 AWS 帐户可以访问域的数据门户。来自其他人的 IAM 负责人 AWS 帐户无法访问域的数据门户。

完成以下过程以将所需的策略附加到用户、组或角色。有关更多信息，请参阅 AWS 亚马逊的管理政策 DataZone。

1. 登录 AWS 管理控制台，然后在以下位置打开 IAM 控制台 https://console.aws.amazon.com/iam/。
2. 在导航窗格中，选择用户、“用户组”或“角色”。
3. 在列表中，选择要在其中嵌入策略的用户、组或角色的名称。
4. 选择 Permissions (权限) 选项卡，然后展开 Permissions policies (权限策略) 部分（如有必要）。
5. 选择添加权限，然后选择创建内联策略链接。
6. 在创建策略屏幕上的策略编辑器部分中，选择 JSON。使用以下 JSON 语句创建策略文档，然后选择下一步。

   JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "datazone:GetIamPortalLoginUrl"
            ],
            "Resource": [
```

```
                "*"
            ]
        }
    ]
}
```

7. 在查看策略屏幕上，输入此策略的名称。如果您对该策略感到满意，请选择 Create policy (创建策略)。确保屏幕顶部的红框中没有显示错误。更正报告的任何错误。

将所需的策略附加到亚马逊的用户、群组或角色 DataZone 目录访问权限

> **Note**
>
> 您的域中只有 IAM 委托人 AWS 帐户可以访问该域的目录。来自其他人的 IAM 负责人 AWS 账户无法访问该域的目录。

您可以授予您的 IAM 身份访问您的亚马逊的权限 DataZone 通过 API 和 SDK 获取域名目录，步骤如下。如果你想让这些 IAM 身份也能访问亚马逊 DataZone 数据门户，然后另外按照上述步骤操作 将所需的策略附加到亚马逊的用户、群组或角色 DataZone 数据门户访问权限。有关更多信息，请参阅 AWS 亚马逊的管理政策 DataZone。

1. 登录 AWS 管理控制台，然后在以下位置打开 IAM 控制台 https://console.aws.amazon.com/iam/。
2. 在导航窗格中，选择策略。
3. 在策略列表中，选择旁边的单选按钮 AmazonDataZoneFullUserAccess 政策。您可以使用 Filter 菜单和搜索框来筛选策略列表。有关更多信息，请参阅 AWS 管理策略：AmazonDataZoneFullUserAccess。
4. 选择 Actions（操作），然后选择 Attach（附加）。
5. 通过选中每个主体旁边的复选框来选择要将策略附加到的用户、组或角色。您可以使用 Filter（筛选条件）菜单和搜索框来筛选委托人实体列表。选择用户、组或角色后，选择附加策略。

为亚马逊的用户、群组或角色附加可选政策 DataZone 如果您的域是使用来自的客户管理密钥加密的，则可以访问数据门户或目录 AWS Key Management Service（KMS）

如果你创建了自己的亚马逊 DataZone 使用您自己的 KMS 密钥进行数据加密的域，您还必须创建具有以下权限的内联策略并将其附加到您的 IAM 委托人，以便他们可以访问亚马逊 DataZone 数据门户或目录。

1. 登录 AWS 管理控制台，然后在以下位置打开 IAM 控制台 https://console.aws.amazon.com/iam/。
2. 在导航窗格中，选择用户、“用户组”或“角色”。
3. 在列表中，选择要在其中嵌入策略的用户、组或角色的名称。
4. 选择 Permissions (权限) 选项卡，然后展开 Permissions policies (权限策略) 部分（如有必要）。
5. 选择添加权限，然后选择创建内联策略链接。
6. 在创建策略屏幕上的策略编辑器部分中，选择 JSON。使用以下 JSON 语句创建策略文档，然后选择下一步。

   JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Sid": "Statement1",
            "Effect": "Allow",
            "Action": [
                "kms:Decrypt",
                "kms:DescribeKey",
                "kms:GenerateDataKey"
            ],
            "Resource": [
                "arn:aws:kms:us-
east-1:111122223333:key/1234abcd-12ab-34cd-56ef-1234567890ab"
            ]
        }
    ]
}
```

7. 在查看策略屏幕上，输入此策略的名称。如果您对该策略感到满意，请选择 Create policy (创建策略)。确保屏幕顶部的红框中没有显示错误。更正报告的任何错误。

设置 AWS 亚马逊 IAM 身份中心 DataZone

> **Note**
>
> AWS 必须在同一位置启用身份中心 AWS 区域即您的亚马逊 DataZone 域。目前， AWS 只能在单个身份中心中启用 AWS 地区。

你可以访问亚马逊 DataZone 使用您的单点登录 (SSO) 凭据或 AWS 证书。按照本节中的说明进行设置 AWS 亚马逊 IAM 身份中心 DataZone。有关使用亚马逊的更多信息 DataZone 和你的 AWS 证书，请参阅 配置使用亚马逊所需的 IAM 权限 DataZone 管理控制台。

如果你已经有了，你可以跳过本节中的步骤 AWS IAM 身份中心（继任者 AWS 单身 Sign-On) 启用和配置相同 AWS 您想要创建亚马逊的地区 DataZone 域。

完成以下步骤以启用 AWS IAM 身份中心（继任者 AWS 单身 Sign-On）。

1. 启用 AWS IAM 身份中心，您必须登录 AWS 使用您的证书进行管理控制台 AWS 组织管理账户。使用来自 AWS Organizations 成员账户的凭证登录时，您无法启用 IAM Identity Center。有关更多信息，请参阅 创建和管理组织 在 AWS 组织用户指南。
2. 打开 AWS IAM 身份中心（继任者 AWS 单身 Sign-On) 控制台 然后使用顶部导航栏中的区域选择器来选择 AWS 您要在其中创建亚马逊的地区 DataZone 域。
3. 请选择启用。
4. 选择身份来源。

   默认情况下，您将获得一个 IAM Identity Center 存储，以便快速轻松地管理用户。您也可以选择使用外部身份提供程序。我们这里使用默认的 IAM Identity Center 存储。

   有关更多信息，请参阅 Choose your identity source。
5. 在 IAM Identity Center 导航窗格中，选择组，然后选择创建组。输入组名称并选择创建。
6. 在 IAM Identity Center 导航窗格中，选择用户。
7. 在添加用户页面上，输入所需信息，然后选择向用户发送包含密码设置说明的电子邮件。用户将收到一封包含后续设置步骤的电子邮件。

8. 选择下一步：组，再选择所需的组，然后选择添加用户。用户将收到一封邀请他们使用 SSO 的电子邮件。在这封电子邮件中，他们需要选择接受邀请并设置密码。

在您创建亚马逊之后 DataZone 域，你可以启用 AWS 亚马逊身份中心 DataZone 并向您的 SSO 用户和 SSO 群组提供访问权限。有关更多信息，请参阅 启用 Amazon 的 IAM 身份中心 DataZone。

开始使用亚马逊 DataZone

本节中的信息可帮助您开始使用 Amazon DataZone。如果您不熟悉 Amazon DataZone，请先熟悉中介绍的概念和术语亚马逊 DataZone 术语和概念。

在开始执行这两个快速入门工作流中的任一工作流的步骤之前，您必须先完成本指南的设置部分中描述的过程。如果您使用的是全新的 AWS 账户，则必须配置使用亚马逊 DataZone 管理控制台所需的权限。如果您使用的 AWS 账户已有 AWS Glue 数据目录对象，则还必须为亚马逊配置 Lake Formation 权限 DataZone。

本入门部分将引导您完成以下 Amazon DataZone 快速入门工作流程：

主题

- 亚马逊 DataZone 快速入门 Glue AWS 数据
- 亚马逊使用亚马逊 DataZone Redshift 数据快速入门
- 使用示例脚本的 Amazon DataZone 快速入门

亚马逊 DataZone 快速入门 Glue AWS 数据

完成以下快速入门步骤，使用示例 AWS Glue 数据在 Amazon 中运行完整的数据生成器和数据 DataZone 使用者工作流程。

快速入门步骤

- 第 1 步-创建 Amazon DataZone 域名和数据门户
- 步骤 2 – 创建发布项目
- 步骤 3 – 创建环境
- 步骤 4 – 创建数据以供发布
- 第 5 步-从 AWS Glue 收集元数据
- 步骤 6 – 整理和发布数据资产
- 步骤 7 – 创建用于数据分析的项目
- 步骤 8 – 创建用于数据分析的环境
- 步骤 9 – 搜索数据目录并订阅数据
- 步骤 10 – 批准订阅请求

- 步骤 11 – 在 Amazon Athena 中构建查询并分析数据

第 1 步-创建 Amazon DataZone 域名和数据门户

本节介绍为此工作流程创建 Amazon DataZone 域和数据门户的步骤。

完成以下步骤以创建 Amazon DataZone 域名。有关 Amazon DataZone 域名的更多信息，请参阅亚马逊 DataZone 术语和概念。

1. 导航至 https://console.aws.amazon.com/datazone 上的亚马逊 DataZone 控制台，登录，然后选择创建域名。

> **Note**
>
> 如果您想在此工作流程中使用现有 Amazon DataZone 域名，请选择查看域名，然后选择要使用的域名，然后继续执行创建发布项目的第 2 步。

2. 在创建域页面上，提供以下字段的值：

   - 名称 – 指定您的域的名称。在此工作流中，您可以将此域命名为 Marketing。
   - 描述 – 指定可选的域描述。
   - 数据加密-默认情况下，您的数据使用为您 AWS 拥有和管理的密钥进行加密。在此应用场景中，您可以保留默认的数据加密设置。

     有关客户自主管理型密钥的更多信息，请参阅 亚马逊的静态数据加密 DataZone。如果您使用自己的 KMS 密钥进行数据加密，则必须在默认 AmazonDataZoneDomainExecutionRole 中包含以下语句。

     JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Sid": "Statement1",
            "Effect": "Allow",
            "Action": [
                "kms:Decrypt",
                "kms:DescribeKey",
                "kms:GenerateDataKey"
```

```
            ],
            "Resource": [
                "arn:aws:kms:us-
east-1:111122223333:key/1234abcd-12ab-34cd-56ef-1234567890ab"
            ]
        }
    ]
}
```

- 服务访问权限 – 默认情况下，将已选定的使用默认角色选项保持不变。

> **Note**
>
> 如果您在此工作流程中使用现有 Amazon DataZone 域名，则可以选择 “使用现有服务角色” 选项，然后从下拉菜单中选择现有角色。

- 在快速设置功能下，选择设置此账户以使用和发布数据。此选项启用内置的 Amazon 数据湖和数据仓库 DataZone 蓝图，并为该账户配置所需的权限、资源、默认项目以及默认数据湖和数据仓库环境配置文件。有关 Amazon DataZone 蓝图的更多信息，请参阅亚马逊 DataZone 术语和概念。
- 将权限详细信息下的其余字段保持不变。

> **Note**
>
> 如果您已有 Amazon DataZone 域名，则可以选择 “使用现有服务角色” 选项，然后从 Glue 管理访问角色、Redshift 管理访问角色和配置角色的下拉菜单中选择现有角色。

- 将标签下的字段保持不变。
- 选择创建域。

3. 成功创建一个域后，选择此域，然后在此域的摘要页面上记下此域的数据门户 URL。您可以使用此 URL 访问您的 Amazon DataZone 数据门户，以完成此工作流程中的其余步骤。您也可以通过选择打开数据门户来导航到数据门户。

> **Note**
>
> 在当前版本的 Amazon 中 DataZone，一旦创建了域，就无法修改为数据门户生成的 URL。

域创建过程可能需要几分钟的时间才能完成。等待域状态变为可用，然后再继续执行下一步。

步骤 2 – 创建发布项目

此部分介绍为此工作流创建发布项目所需的步骤。

1. 完成上述第 1 步并创建域名后，您将看到 “欢迎来到亚马逊 DataZone！ 窗口。在此窗口中，选择创建项目。
2. 例如，为该工作流程指定项目名称，您可以为其命名 SalesDataPublishingProject，然后将其余字段保持不变，然后选择创建。

步骤 3 – 创建环境

此部分介绍为此工作流创建环境所需的步骤。

1. 完成上述步骤 2 并创建项目后，您将看到您的项目已准备就绪，可以开始使用了！窗口。在此窗口中，选择创建环境。
2. 在创建环境页面上，指定以下内容，然后选择创建环境。
3. 为以下字段指定值：
   - 名称 – 指定环境的名称。在本演练中，您可以将它命名为 `Default data lake environment`。
   - 描述 – 指定环境的描述。
   - 环境配置文件-选择DataLakeProfile环境配置文件。这使您能够在此工作流程 DataZone 中使用亚马逊来处理亚马逊 S3、 AWS Glue Catalog 和 Amazon Athena 中的数据。
   - 在本演练中，将其余字段保持不变。
4. 选择创建环境。

步骤 4 – 创建数据以供发布

此部分介绍为此工作流创建数据以供发布所需的步骤。

1. 完成上述步骤 3 后，在 `SalesDataPublishingProject` 项目中，在右侧面板中的分析工具下，选择 Amazon Athena。这将打开 Athena 查询编辑器，使用项目的凭证进行身份验证。确保在 Amazon 环境下拉列表中选择了您的发布 DataZone 环境，并选择了查询编辑器中的`<environment_name>%_pub_db`数据库。

2. 在本演练中，您将使用 “按选择创建表” (CTAS) 查询脚本来创建要发布到 Amazon 的新表。 DataZone在查询编辑器中，执行此 CTAS 脚本来创建一个可发布并可供搜索和订阅的 `mkt_sls_table` 表。

```
CREATE TABLE mkt_sls_table AS
SELECT 146776932 AS ord_num, 23 AS sales_qty_sld, 23.4 AS wholesale_cost, 45.0 as
 lst_pr, 43.0 as sell_pr, 2.0 as disnt, 12 as ship_mode,13 as warehouse_id, 23 as
 item_id, 34 as ctlg_page, 232 as ship_cust_id, 4556 as bill_cust_id
UNION ALL SELECT 46776931, 24, 24.4, 46, 44, 1, 14, 15, 24, 35, 222, 4551
UNION ALL SELECT 46777394, 42, 43.4, 60, 50, 10, 30, 20, 27, 43, 241, 4565
UNION ALL SELECT 46777831, 33, 40.4, 51, 46, 15, 16, 26, 33, 40, 234, 4563
UNION ALL SELECT 46779160, 29, 26.4, 50, 61, 8, 31, 15, 36, 40, 242, 4562
UNION ALL SELECT 46778595, 43, 28.4, 49, 47, 7, 28, 22, 27, 43, 224, 4555
UNION ALL SELECT 46779482, 34, 33.4, 64, 44, 10, 17, 27, 43, 52, 222, 4556
UNION ALL SELECT 46779650, 39, 37.4, 51, 62, 13, 31, 25, 31, 52, 224, 4551
UNION ALL SELECT 46780524, 33, 40.4, 60, 53, 18, 32, 31, 31, 39, 232, 4563
UNION ALL SELECT 46780634, 39, 35.4, 46, 44, 16, 33, 19, 31, 52, 242, 4557
UNION ALL SELECT 46781887, 24, 30.4, 54, 62, 13, 18, 29, 24, 52, 223, 4561
```

   确保已在左侧的表和视图部分中成功创建 mkt_sls_table 表。现在，您有了可以发布到 Amazon DataZone 目录中的数据资产。

第 5 步-从 AWS Glue 收集元数据

本节介绍为该工作流程从 AWS Glue 收集元数据的步骤。

1. 完成上述步骤 4 后，在 Amazon DataZone 数据门户中，选择`SalesDataPublishingProject`项目，然后选择 “数据” 选项卡，然后在左侧面板中选择 “数据源”。
2. 选择在环境创建过程中创建的来源。
3. 选择操作下拉菜单旁边的运行，然后选择刷新按钮。数据源运行完成后，资产将添加到 Amazon DataZone 库存中。

步骤 6 – 整理和发布数据资产

此部分介绍在此工作流中整理和发布数据资产的步骤。

1. 完成上述步骤 5 后，在 Amazon DataZone 数据门户中，选择您在上一步中创建的SalesDataPublishingProject项目，选择 “数据” 选项卡，在左侧面板中选择 “库存数据”，然后找到mkt_sls_table表格。
2. 打开 mkt_sls_table 资产的详细信息页面以查看自动生成的企业名称。选择自动生成的元数据图标以查看自动生成的资产名称和列名称。您可以分别接受或拒绝每个名称，也可以选择全部接受以应用生成的名称。或者，您也可以将可用的元数据表单添加到资产中，并选择术语表术语来对数据进行分类。
3. 选择发布资产以发布 mkt_sls_table 资产。

步骤 7 – 创建用于数据分析的项目

此部分介绍创建用于数据分析的项目的步骤。这是此工作流包含的多个数据使用者步骤中的第一个步骤。

1. 完成上述步骤 6 后，在亚马逊 DataZone 数据门户中，从项目下拉菜单中选择创建项目。
2. 在创建项目页面上，指定项目名称，例如，您可以为此工作流程命名 MarketingDataAnalysisProject，然后将其余字段保持不变，然后选择创建。

步骤 8 – 创建用于数据分析的环境

此部分介绍创建用于数据分析的环境的步骤。

1. 完成上述步骤 7 后，在 Amazon DataZone 数据门户中，选择MarketingDataAnalysisProject项目，然后选择环境选项卡，然后选择创建环境。
2. 在创建环境页面上，指定以下内容，然后选择创建环境。
   - 名称 – 指定环境的名称。在本演练中，您可以将它命名为 Default data lake environment。
   - 描述 – 指定环境的描述。
   - 环境配置文件-选择内置DataLakeProfile环境配置文件。
   - 在本演练中，将其余字段保持不变。

步骤 9 – 搜索数据目录并订阅数据

此部分介绍搜索数据目录和订阅数据的步骤。

1. 完成上述步骤 8 后，在亚马逊 DataZone 数据门户中，选择亚马逊 DataZone图标，然后在亚马逊 DataZone 搜索字段中，使用数据门户搜索栏中的关键词（例如 “目录” 或 “销售”）搜索数据资产。

   如有必要，可应用筛选条件或排序，在找到产品销售数据资产后，可选择该资产以打开其详细信息页面。
2. 在目录销售数据资产的详细信息页面上，选择订阅。
3. 在 “订阅” 对话框中，从下拉列表中选择您的MarketingDataAnalysisProject消费者项目，然后指定订阅请求的原因，然后选择 “订阅”。

步骤 10 – 批准订阅请求

此部分介绍批准订阅请求的步骤。

1. 完成上述步骤 9 后，在 Amazon DataZone 数据门户中，选择用于发布资产的SalesDataPublishingProject项目。
2. 选择数据选项卡，再选择已发布的数据，然后选择传入的请求。
3. 现在，您可以看到需审批的新请求所在的行。选择查看请求。提供审批的原因，然后选择批准。

步骤 11 – 在 Amazon Athena 中构建查询并分析数据

现在，您已成功将资产发布到 Amazon DataZone 目录并订阅了该资产，您可以对其进行分析。

1. 在亚马逊 DataZone 数据门户中，选择您的MarketingDataAnalysisProject消费者项目，然后从右侧面板的 “分析工具” 下，选择 Amazon Athena 的 “查询数据” 链接。这将打开 Amazon Athena 查询编辑器，使用项目的凭证进行身份验证。从查询编辑器的 Amazon Environment 下拉列表中选择使用MarketingDataAnalysisProject者 DataZone 环境，然后<environment_name>%sub_db从数据库下拉列表中选择您的项目。
2. 现在，您可以对订阅的表运行查询。您可以从表和视图中选择表，然后选择预览以在编辑器屏幕上显示 select 语句。运行查询以查看结果。

亚马逊使用亚马逊 DataZone Redshift 数据快速入门

完成以下快速入门步骤，使用 Amazon Redshift 示例数据在亚马逊中运行完整的数据生成器和数据 DataZone 使用者工作流程。

快速入门步骤

- 第 1 步-创建 Amazon DataZone 域名和数据门户
- 步骤 2 – 创建发布项目
- 步骤 3 – 创建环境
- 步骤 4 – 创建数据以供发布
- 步骤 5 – 从 Amazon Redshift 收集元数据
- 步骤 6 – 整理和发布数据资产
- 步骤 7 – 创建用于数据分析的项目
- 步骤 8 – 创建用于数据分析的环境
- 步骤 9 – 搜索数据目录并订阅数据
- 步骤 10 – 批准订阅请求
- 步骤 11 – 在 Amazon Redshift 中构建查询并分析数据

第 1 步-创建 Amazon DataZone 域名和数据门户

完成以下步骤以创建 Amazon DataZone 域名。有关 Amazon DataZone 域名的更多信息，请参阅亚马逊 DataZone 术语和概念。

1. 导航至 https://console.aws.amazon.com/datazone 上的亚马逊 DataZone 控制台，登录，然后选择创建域名。

   > **Note**
   >
   > 如果您想在此工作流程中使用现有 Amazon DataZone 域名，请选择查看域名，然后选择要使用的域名，然后继续执行创建发布项目的第 2 步。

2. 在创建域页面上，提供以下字段的值：

   - 名称 – 指定您的域的名称。在此工作流中，您可以将此域命名为 `Marketing`。
   - 描述 – 指定可选的域描述。
   - 数据加密-默认情况下，您的数据使用为您 AWS 拥有和管理的密钥进行加密。在本演练中，您可以保留默认的数据加密设置。

     有关客户自主管理型密钥的更多信息，请参阅 亚马逊的静态数据加密 DataZone。如果您使用自己的 KMS 密钥进行数据加密，则必须在默认 AmazonDataZoneDomainExecutionRole 中包含以下语句。

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Sid": "Statement1",
            "Effect": "Allow",
            "Action": [
                "kms:Decrypt",
                "kms:DescribeKey",
                "kms:GenerateDataKey"
            ],
            "Resource": [
                "arn:aws:kms:us-
east-1:111122223333:key/1234abcd-12ab-34cd-56ef-1234567890ab"
            ]
        }
    ]
}
```

- 服务访问权限-选择 “使用自定义服务角色” 选项，然后AmazonDataZoneDomainExecutionRole从下拉菜单中选择。
- 在快速设置功能下，选择设置此账户以使用和发布数据。此选项启用内置的 Amazon 数据湖和数据仓库 DataZone 蓝图，并配置完成此工作流程中其余步骤所需的权限和资源。有关 Amazon DataZone 蓝图的更多信息，请参阅亚马逊 DataZone 术语和概念。
- 将权限详细信息和标签下的其余字段保持不变，然后选择创建域。

3. 成功创建一个域后，选择此域，然后在此域的摘要页面上记下此域的数据门户 URL。您可以使用此 URL 访问您的 Amazon DataZone 数据门户，以完成此工作流程中的其余步骤。

> **Note**
>
> 在当前版本的 Amazon 中 DataZone，一旦创建了域，就无法修改为数据门户生成的 URL。

域创建过程可能需要几分钟的时间才能完成。等待域状态变为可用，然后再继续执行下一步。

步骤 2 – 创建发布项目

以下部分介绍在此工作流中创建发布项目的步骤。

1. 完成步骤 1 后，使用数据门户 URL 导航至 Amazon DataZone 数据门户，然后使用单点登录 (SSO) 或 AWS IAM 凭证登录。
2. 选择 “创建项目”，指定项目名称，例如，为该工作流程指定项目名称 SalesDataPublishingProject，然后将其余字段保持不变，然后选择 “创建”。

步骤 3 – 创建环境

以下部分介绍在此工作流中创建环境的步骤。

1. 完成步骤 2 后，在 Amazon DataZone 数据门户中，选择您在上一步中创建的`SalesDataPublishingProject`项目，然后选择环境选项卡，然后选择创建环境。
2. 在创建环境页面上，指定以下内容，然后选择创建环境。
   - 名称 – 指定环境的名称。在本演练中，您可以将它命名为 `Default data warehouse environment`。
   - 描述 – 指定环境的描述。
   - 环境配置文件-选择DataWarehouseProfile环境配置文件。
   - 提供您的 Amazon Redshift 集群的名称、数据库名称以及存储数据的 Amazon Redshift 集群的密钥 ARN。

> **Note**
>
> 确保你在 Secrets Manager 中的 AWS 密钥包含以下标签（键/值）：
>
> - 对于 Amazon Redshift 集群 – datazone.rs.cluster：<cluster_name:database name>
>
> 对于 Amazon Redshift Serverless 工作组 – datazone.rs.workgroup：<workgroup_name:database_name>
> - AmazonDataZoneProject: <projectID>
> - AmazonDataZoneDomain: <domainID>
>
> 有关更多信息，请参阅在 S AWS ecrets Manager 中存储数据库凭据。
> 您在 S AWS ecrets Manager 中提供的数据库用户必须具有超级用户权限。

步骤 4 – 创建数据以供发布

以下部分介绍在此工作流中创建用于发布的数据的步骤。

1. 完成步骤 3 后，在亚马逊 DataZone 数据门户中，选择SalesDataPublishingProject项目，然后在右侧面板的 “分析工具” 下，选择 Amazon Redshift。这将打开 Amazon Redshift 查询编辑器，使用项目的凭证进行身份验证。
2. 在本演练中，您将使用 “按选择创建表” (CTAS) 查询脚本来创建要发布到 Amazon 的新表。 DataZone在查询编辑器中，执行此 CTAS 脚本来创建一个可发布并可供搜索和订阅的 mkt_sls_table 表。

```
CREATE TABLE mkt_sls_table AS
SELECT 146776932 AS ord_num, 23 AS sales_qty_sld, 23.4 AS wholesale_cost, 45.0 as
 lst_pr, 43.0 as sell_pr, 2.0 as disnt, 12 as ship_mode,13 as warehouse_id, 23 as
 item_id, 34 as ctlg_page, 232 as ship_cust_id, 4556 as bill_cust_id
UNION ALL SELECT 46776931, 24, 24.4, 46, 44, 1, 14, 15, 24, 35, 222, 4551
UNION ALL SELECT 46777394, 42, 43.4, 60, 50, 10, 30, 20, 27, 43, 241, 4565
UNION ALL SELECT 46777831, 33, 40.4, 51, 46, 15, 16, 26, 33, 40, 234, 4563
UNION ALL SELECT 46779160, 29, 26.4, 50, 61, 8, 31, 15, 36, 40, 242, 4562
UNION ALL SELECT 46778595, 43, 28.4, 49, 47, 7, 28, 22, 27, 43, 224, 4555
UNION ALL SELECT 46779482, 34, 33.4, 64, 44, 10, 17, 27, 43, 52, 222, 4556
UNION ALL SELECT 46779650, 39, 37.4, 51, 62, 13, 31, 25, 31, 52, 224, 4551
UNION ALL SELECT 46780524, 33, 40.4, 60, 53, 18, 32, 31, 31, 39, 232, 4563
UNION ALL SELECT 46780634, 39, 35.4, 46, 44, 16, 33, 19, 31, 52, 242, 4557
UNION ALL SELECT 46781887, 24, 30.4, 54, 62, 13, 18, 29, 24, 52, 223, 4561
```

确保已成功创建 mkt_sls_table 表。现在，您有了可以发布到 Amazon DataZone 目录中的数据资产。

步骤 5 – 从 Amazon Redshift 收集元数据

以下部分介绍从 Amazon Redshift 收集元数据的步骤。

1. 完成步骤 4 后，在 Amazon DataZone 数据门户中，选择SalesDataPublishingProject项目，然后选择 “数据” 选项卡，然后选择 “数据源”。
2. 选择在环境创建过程中创建的来源。

3. 选择操作下拉菜单旁边的运行，然后选择刷新按钮。数据源运行完成后，资产将添加到 Amazon DataZone 库存中。

步骤 6 – 整理和发布数据资产

以下部分介绍在此工作流中整理和发布数据资产的步骤。

1. 完成第 5 步后，在 Amazon DataZone 数据门户中，选择SalesDataPublishingProject项目，然后选择数据选项卡，选择库存数据，然后找到mkt_sls_table表。
2. 打开 mkt_sls_table 资产的详细信息页面以查看自动生成的企业名称。选择自动生成的元数据图标以查看自动生成的资产名称和列名称。您可以分别接受或拒绝每个名称，也可以选择全部接受以应用生成的名称。或者，您也可以将可用的元数据表单添加到资产中，并选择术语表术语来对数据进行分类。
3. 选择发布以发布 mkt_sls_table 资产。

步骤 7 – 创建用于数据分析的项目

以下部分介绍在此工作流中创建用于数据分析的项目的步骤。

1. 完成步骤 6 后，在 Amazon DataZone 数据门户中，选择创建项目。
2. 在 “创建项目” 页面中，指定项目名称，例如，为该工作流程命名 MarketingDataAnalysisProject，然后将其余字段保持不变，然后选择 “创建”。

步骤 8 – 创建用于数据分析的环境

以下部分介绍在此工作流中创建用于数据分析的环境的步骤。

1. 完成步骤 7 后，在 Amazon DataZone 数据门户中，选择您在上一步中创建的MarketingDataAnalysisProject项目，然后选择环境选项卡，然后选择添加环境。
2. 在创建环境页面上，指定以下内容，然后选择创建环境。
   - 名称 – 指定环境的名称。在本演练中，您可以将它命名为 Default data warehouse environment。
   - 描述 – 指定环境的描述。
   - 环境配置文件-选择DataWarehouseProfile环境配置文件。

- 提供您的 Amazon Redshift 集群的名称、数据库名称以及存储数据的 Amazon Redshift 集群的密钥 ARN。

> **Note**
>
> 确保你在 Secrets Manager 中的 AWS 密钥包含以下标签（键/值）：
>
> - 对于 Amazon Redshift 集群 – datazone.rs.cluster：<cluster_name:database name>
>
> 对于 Amazon Redshift Serverless 工作组 – datazone.rs.workgroup：<workgroup_name:database_name>
> - AmazonDataZoneProject: <projectID>
> - AmazonDataZoneDomain: <domainID>
>
> 有关更多信息，请参阅在 S AWS ecrets Manager 中存储数据库凭据。
> 您在 S AWS ecrets Manager 中提供的数据库用户必须具有超级用户权限。

- 在本演练中，将其余字段保持不变。

步骤 9 – 搜索数据目录并订阅数据

以下部分介绍搜索数据目录和订阅数据的步骤。

1. 完成步骤 8 后，在亚马逊 DataZone 数据门户中，使用数据门户搜索栏中的关键词（例如“目录”或“销售”）搜索数据资产。

   如有必要，可应用筛选条件或排序，在找到产品销售数据资产后，可选择该资产以打开其详细信息页面。
2. 在产品销售数据资产的详细信息页面上，选择订阅。
3. 在对话框中，从下拉列表中选择使用者项目，提供访问请求的原因，然后选择订阅。

步骤 10 – 批准订阅请求

以下部分介绍在此工作流中批准订阅请求的步骤。

1. 完成步骤 9 后，在 Amazon DataZone 数据门户中，选择用于发布资产的SalesDataPublishingProject项目。
2. 选择数据选项卡，再选择已发布的数据，然后选择传入的请求。

3. 选择查看请求链接，然后选择批准。

步骤 11 – 在 Amazon Redshift 中构建查询并分析数据

现在，您已成功将资产发布到 Amazon DataZone 目录并订阅了该资产，您可以对其进行分析。

1. 在亚马逊 DataZone 数据门户网站的右侧面板上，单击 Amazon Redshift 链接。这将打开 Amazon Redshift 查询编辑器，使用项目的凭证进行身份验证。
2. 现在，您可以对订阅的表运行查询（select 语句）。您可以单击表格（three-vertical-dots 选项），然后选择预览以在编辑器屏幕上显示选择语句。执行查询以查看结果。

使用示例脚本的 Amazon DataZone 快速入门

您可以 DataZone 通过管理门户网站或亚马逊 DataZone 数据门户网站访问亚马逊，也可以使用亚马逊 DataZone HTTPS API 以编程方式访问亚马逊，它允许您直接向服务发出 HTTPS 请求。本节包含调用 Amazon 的示例脚本 DataZone APIs ，您可以使用这些脚本来完成以下常见任务：

示例脚本

- 创建 Amazon DataZone 域名和数据门户
- 创建发布项目
- 创建环境配置文件
- 创建环境
- 从 AWS Glue 收集元数据
- 整理和发布数据资产
- 搜索数据目录并订阅数据
- 在数据目录中搜索资产
- 其他有用的示例脚本

创建 Amazon DataZone 域名和数据门户

您可以使用以下示例脚本创建 Amazon DataZone 域名。有关 Amazon DataZone 域名的更多信息，请参阅亚马逊 DataZone 术语和概念。

```
import sys
import boto3

// Initialize datazone client
region = 'us-east-1'
dzclient = boto3.client(service_name='datazone', region_name='us-east-1')

// Create DataZone domain
def create_domain(name):
    return dzclient.create_domain(
        name = name,
        description = "this is a description",
        domainExecutionRole = "arn:aws:iam::<account>:role/
AmazonDataZoneDomainExecutionRole",
    )
```

创建发布项目

您可以使用以下示例脚本在 Amazon 中创建发布项目 DataZone。

```
// Create Project
def create_project(domainId):
    return dzclient.create_project(
        domainIdentifier = domainId,
        name = "sample-project"
    )
```

创建环境配置文件

您可以使用以下示例脚本在 Amazon 中创建环境配置文件 DataZone。

调用 `CreateEnvironmentProfile` API 时将使用以下示例有效载荷：

```
Sample Payload
{
    "Content":{
        "project_name": "Admin_project",
        "domain_name": "Drug-Research-and-Development",
        "blueprint_account_region": [
```

```
            {
                "blueprint_name": "DefaultDataLake",
                "account_id": ["066535990535",
                "413878397724",
                "676266385322",
                "747721550195",
                "755347404384"
                ],
                "region": ["us-west-2", "us-east-1"]
            },
            {
                "blueprint_name": "DefaultDataWarehouse",
                "account_id": ["066535990535",
                "413878397724",
                "676266385322",
                "747721550195",
                "755347404384"
                ],
                "region":["us-west-2", "us-east-1"]
            }
        ]
    }
}
```

此示例脚本调用 CreateEnvironmentProfile API：

```
def create_environment_profile(domain_id, project_id, env_blueprints)
        try:
            response = dz.list_environment_blueprints(
                domainIdentifier=domain_id,
                managed=True
            )
            env_blueprints = response.get("items")
            env_blueprints_map = {}
            for i in env_blueprints:
                env_blueprints_map[i["name"]] = i['id']

            print("Environment Blueprint map", env_blueprints_map)
            for i in blueprint_account_region:
                print(i)
                for j in i["account_id"]:
```

```
                    for k in i["region"]:
                        print("The env blueprint name is", i['blueprint_name'])
                        dz.create_environment_profile(
                            description='This is a test environment profile created via
 lambda function',
                            domainIdentifier=domain_id,
                            awsAccountId=j,
                            awsAccountRegion=k,

 environmentBlueprintIdentifier=env_blueprints_map.get(i["blueprint_name"]),
                            name=i["blueprint_name"] + j + k + "_profile",
                            projectIdentifier=project_id
                        )
        except Exception as e:
            print("Failed to created Environment Profile")
            raise e
```

这是调用 CreateEnvironmentProfile API 后的示例输出有效载荷：

```
{
    "Content":{
        "project_name": "Admin_project",
        "domain_name": "Drug-Research-and-Development",
        "blueprint_account_region": [
            {
                "blueprint_name": "DefaultDataWarehouse",
                "account_id": ["111111111111"],
                "region":["us-west-2"],
                "user_parameters":[
                    {
                        "name": "dataAccessSecretsArn",
                        "value": ""
                    }
                ]
            }
        ]
    }
}
```

创建环境

您可以使用以下示例脚本在 Amazon 中创建环境 DataZone。

```
def create_environment(domain_id, project_id,blueprint_account_region ):
         try:
            #refer to get_domain_id and get_project_id for fetching ids using names.
            sts_client = boto3.client("sts")
            # Get the current account ID
            account_id = sts_client.get_caller_identity()["Account"]
            print("Fetching environment profile ids")
            env_profile_map = get_env_profile_map(domain_id, project_id)

            for i in blueprint_account_region:
                for j in i["account_id"]:
                    for k in i["region"]:
                        print(" env blueprint name", i['blueprint_name'])
                        profile_name = i["blueprint_name"] + j + k + "_profile"
                        env_name = i["blueprint_name"] + j + k + "_env"
                        description = f'This is environment is created for
 {profile_name}, Account {account_id} and region {i["region"]}'
                        try:
                            dz.create_environment(
                                description=description,
                                domainIdentifier=domain_id,

 environmentProfileIdentifier=env_profile_map.get(profile_name),
                                name=env_name,
                                projectIdentifier=project_id
                            )
                            print(f"Environment created - {env_name}")
                        except:
                            dz.create_environment(
                                description=description,
                                domainIdentifier=domain_id,

 environmentProfileIdentifier=env_profile_map.get(profile_name),
                                name=env_name,
                                projectIdentifier=project_id,
                                userParameters= i["user_parameters"]
                            )
                            print(f"Environment created - {env_name}")
```

```
        except Exception as e:
            print("Failed to created Environment")
            raise e
```

从 AWS Glue 收集元数据

您可以使用此示例脚本从 AWS Glue 收集元数据。此脚本按标准计划运行。可从示例脚本中检索参数并将它们设置为全局参数。使用标准函数获取项目、环境和域 ID。 AWS Glue 数据来源按标准时间创建和运行，可以在脚本的 cron 部分进行更新。

```
def crcreate_data_source(domain_id, project_id,data_source_name)
        print("Creating Data Source")
        data_source_creation = dz.create_data_source(
            # Define data source : Customize the data source to which you'd like to
 connect
            # define the name of the Data source to create, example: name
 ='TestGlueDataSource'
            name=data_source_name,
            # give a description for the datasource (optional), example:
 description='This is a dorra test for creation on DZ datasources'
            description=data_source_description,
            # insert the domain identifier corresponding to the domain to which the
 datasource will belong, example: domainIdentifier= 'dzd_6f3gst5jjmrrmv'
            domainIdentifier=domain_id,
            # give environment identifier , example: environmentIdentifier=
 '3weyt6hhn8qcvb'
            environmentIdentifier=environment_id,
            # give corresponding project identifier, example: projectIdentifier=
 '6tl4csoyrg16ef',
            projectIdentifier=project_id,
            enableSetting="ENABLED",
            # publishOnImport used to select whether assets are added to the inventory
 and/or discovery catalog .
            # publishOnImport = True : Assets will be added to project's inventory as
 well as published to the discovery catalog
            # publishOnImport = False : Assets will only be added to project's
 inventory.
            # You can later curate the metadata of the assets and choose subscription
 terms to publish them from the inventory to the discovery catalog.
            publishOnImport=False,
```

```
            # Automated business name generation : Use AI to automatically generate
metadata for assets as they are published or updated by this data source run.
            # Automatically generated metadata can be be approved, rejected, or edited
by data publishers.
            # Automatically generated metadata is badged with a small icon next to the
corresponding metadata field.
            recommendation={"enableBusinessNameGeneration": True},
            type="GLUE",
            configuration={
                "glueRunConfiguration": {
                    "dataAccessRole": "arn:aws:iam::"
                    + account_id
                    + ":role/service-role/AmazonDataZoneGlueAccess-"
                    + current_region
                    + "-"
                    + domain_id
                    + "",
                    "relationalFilterConfigurations": [
                        {
                            #
                            "databaseName": glue_database_name,
                            "filterExpressions": [
                                {"expression": "*", "type": "INCLUDE"},
                            ],
                            #    "schemaName": "TestSchemaName",
                        },
                    ],
                },
            },
            # Add metadata forms to the data source (OPTIONAL).
            # Metadata forms will be automatically applied to any assets that are
created by the data source.
            # assetFormsInput=[
            #     {
            #         "content": "string",
            #         "formName": "string",
            #         "typeIdentifier": "string",
            #         "typeRevision": "string",
            #     },
            # ],
            schedule={
                "schedule": "cron(5 20 * * ? *)",
                "timezone": "UTC",
            },
```

```
        )
        # This is a suggested syntax to return values
        #        return_values["data_source_creation"] = data_source_creation["items"]
        print("Data Source Created")


//This is the sample response payload after the CreateDataSource API is invoked:

{
    "Content":{
        "project_name": "Admin",
        "domain_name": "Drug-Research-and-Development",
        "env_name": "GlueEnvironment",
        "glue_database_name": "test",
        "data_source_name" : "test",
        "data_source_description" : "This is a test data source"
    }
}
```

整理和发布数据资产

您可以使用以下示例脚本在 Amazon DataZone 中整理和发布数据资产。

可使用以下脚本创建自定义表单类型：

```
def create_form_type(domainId, projectId):
    return dzclient.create_form_type(
        domainIdentifier = domainId,
        name = "customForm",
        model = {
            "smithy": "structure customForm { simple: String }"
        },
        owningProjectIdentifier = projectId,
        status = "ENABLED"
    )
```

可使用以下示例脚本创建自定义资产类型：

```
def create_custom_asset_type(domainId, projectId):
    return dzclient.create_asset_type(
        domainIdentifier = domainId,
        name = "userCustomAssetType",
        formsInput = {
            "Model": {
                "typeIdentifier": "customForm",
                "typeRevision": "1",
                "required": False
            }
        },
        owningProjectIdentifier = projectId,
    )
```

可使用以下示例脚本创建自定义资产：

```
def create_custom_asset(domainId, projectId):
    return dzclient.create_asset(
        domainIdentifier = domainId,
        name = 'custom asset',
        description = "custom asset",
        owningProjectIdentifier = projectId,
        typeIdentifier = "userCustomAssetType",
        formsInput = [
            {
                "formName": "UserCustomForm",
                "typeIdentifier": "customForm",
                "content": "{\"simple\":\"sample-catalogId\"}"
            }
        ]
    )
```

可使用以下示例脚本创建术语表：

```
def create_glossary(domainId, projectId):
    return dzclient.create_glossary(
        domainIdentifier = domainId,
        name = "test7",
        description = "this is a test glossary",
```

```
        owningProjectIdentifier = projectId
    )
```

可使用以下示例脚本创建术语表术语：

```
def create_glossary_term(domainId, glossaryId):
    return dzclient.create_glossary_term(
        domainIdentifier = domainId,
        name = "soccer",
        shortDescription = "this is a test glossary",
        glossaryIdentifier = glossaryId,
    )
```

可使用以下示例脚本通过系统定义的资产类型创建资产：

```
def create_asset(domainId, projectId):
    return dzclient.create_asset(
        domainIdentifier = domainId,
        name = 'sample asset name',
        description = "this is a glue table asset",
        owningProjectIdentifier = projectId,
        typeIdentifier = "amazon.datazone.GlueTableAssetType",
        formsInput = [
            {
                "formName": "GlueTableForm",
                "content": "{\"catalogId\":\"sample-catalogId\",\"columns\":
[{\"columnDescription\":\"sample-columnDescription\",\"columnName\":\"sample-
columnName\",\"dataType\":\"sample-dataType\",\"lakeFormationTags\":{\"sample-
key1\":\"sample-value1\",\"sample-key2\":\"sample-value2\"}}],\"compressionType\":
\"sample-compressionType\",\"lakeFormationDetails\":{\"lakeFormationManagedTable
\":false,\"lakeFormationTags\":{\"sample-key1\":\"sample-value1\",\"sample-key2\":
\"sample-value2\"}},\"primaryKeys\":[\"sample-Key1\",\"sample-Key2\"],\"region\":
\"us-east-1\",\"sortKeys\":[\"sample-sortKey1\"],\"sourceClassification\":\"sample-
sourceClassification\",\"sourceLocation\":\"sample-sourceLocation\",\"tableArn\":
\"sample-tableArn\",\"tableDescription\":\"sample-tableDescription\",\"tableName\":
\"sample-tableName\"}"
            }
        ]
    )
```

可使用以下示例脚本创建资产修订并附加术语表术语：

```
def create_asset_revision(domainId, assetId):
    return dzclient.create_asset_revision(
        domainIdentifier = domainId,
        identifier = assetId,
        name = 'glue table asset 7',
        description = "glue table asset description update",
        formsInput = [
            {
                "formName": "GlueTableForm",
                "content": "{\"catalogId\":\"sample-catalogId\",\"columns\":
[{\"columnDescription\":\"sample-columnDescription\",\"columnName\":\"sample-
columnName\",\"dataType\":\"sample-dataType\",\"lakeFormationTags\":{\"sample-
key1\":\"sample-value1\",\"sample-key2\":\"sample-value2\"}}],\"compressionType\":
\"sample-compressionType\",\"lakeFormationDetails\":{\"lakeFormationManagedTable
\":false,\"lakeFormationTags\":{\"sample-key1\":\"sample-value1\",\"sample-key2\":
\"sample-value2\"}},\"primaryKeys\":[\"sample-Key1\",\"sample-Key2\"],\"region\":
\"us-east-1\",\"sortKeys\":[\"sample-sortKey1\"],\"sourceClassification\":\"sample-
sourceClassification\",\"sourceLocation\":\"sample-sourceLocation\",\"tableArn\":
\"sample-tableArn\",\"tableDescription\":\"sample-tableDescription\",\"tableName\":
\"sample-tableName\"}"
            }
        ],
        glossaryTerms = ["<glossaryTermId:>"]
    )
```

可使用以下示例脚本发布资产：

```
def publish_asset(domainId, assetId):
    return dzclient.create_listing_change_set(
        domainIdentifier = domainId,
        entityIdentifier = assetId,
        entityType = "ASSET",
        action = "PUBLISH",
    )
```

搜索数据目录并订阅数据

可使用以下示例脚本搜索数据目录并订阅数据：

```
def search_asset(domainId, projectId, text):
    return dzclient.search(
        domainIdentifier = domainId,
        owningProjectIdentifier = projectId,
        searchScope = "ASSET",
        searchText = text,
    )
```

可使用以下示例脚本获取资产的列表 ID：

```
def search_listings(domainId, assetName, assetId):
    listings = dzclient.search_listings(
        domainIdentifier=domainId,
        searchText=assetName,
        additionalAttributes=["FORMS"]
    )

    assetListing = None
    for listing in listings['items']:
        if listing['assetListing']['entityId'] == assetId:
            assetListing = listing

    return listing['assetListing']['listingId']
```

可使用以下示例脚本通过列表 ID 创建订阅请求：

```
create_subscription_response = def create_subscription_request(domainId, projectId,
 listingId):
    return dzclient.create_subscription_request(
        subscribedPrincipals=[{
            "project": {
                "identifier": projectId
            }
```

```
        }],
        subscribedListings=[{
            "identifier": listingId
        }],
        requestReason="Give request reason here."
    )
```

使用create_subscription_response上述方法，使用以下示例脚本获取订阅，然后 accept/approve 进行订阅： subscription_request_id

```
subscription_request_id = create_subscription_response["id"]

def accept_subscription_request(domainId, subscriptionRequestId):
    return dzclient.accept_subscription_request(
        domainIdentifier=domainId,
        identifier=subscriptionRequestId
    )
```

在数据目录中搜索资产

您可以使用以下利用自由文本搜索的示例脚本在 Amazon DataZone 目录中查找您发布的数据资产（清单）。

- 以下示例在域中执行自由文本关键字搜索，并返回与提供的关键字“credit”匹配的所有列表：

```
aws datazone search-listings \
  --domain-identifier dzd_c1s7uxe71prrtz \
  --search-text "credit"
```

- 也可以组合多个关键字以进一步缩小搜索范围。例如，如果要查找所有已发布的数据资产（列表），其中包含与墨西哥的销量相关的数据，则可以使用两个关键字“Mexico”和“sales”来制定查询。

```
          aws datazone search-listings \
  --domain-identifier dzd_c1s7uxe71prrtz \
  --search-text "mexico sales"
```

也可以使用筛选条件搜索列表。 SearchListings API 中的filters参数允许您从网域中检索经过筛选的结果。此 API 支持多个默认筛选条件，您也可以组合两个或更多筛选条件并对它们执行 AND/OR 操作。筛选条件句采用两个参数：属性和值。默认支持的筛选条件属性为typeName、owningProjectId 和 glossaryTerms。

- 以下示例使用 assetType 筛选条件（其中列表类型为 Redshift 表）在给定域中搜索所有列表。

```
            aws datazone search-listings \
--domain-identifier dzd_c1s7uxe71prrtz \
--filters '{"or":[{"filter":
{"attribute":"typeName","value":"RedshiftTableAssetType"}} ]}'
```

- 您也可以使用 AND/OR 操作将多个过滤器组合在一起。在以下示例中，可以组合 typeName 和 project 筛选条件。

```
            aws datazone search-listings \
--domain-identifier dzd_c1s7uxe71prrtz \
--filters '{"or":[{"filter":
{"attribute":"typeName","value":"RedshiftTableAssetType"}},  {"filter":
{"attribute":"owningProjectId","value":"cwrrjch7f5kppj"}} ]}'
```

- 您甚至可以将自由文本搜索与过滤器结合使用来查找确切的结果，并按列表的 creation/last 更新时间对其进行进一步排序，如以下示例所示：

```
            aws datazone search-listings \
--domain-identifier dzd_c1s7uxe71prrtz \
--search-text "finance sales" \
--filters '{"or":[{"filter":{"attribute":"typeName","value":"GlueTableViewType"}} ]}'
 \
--sort '{"attribute": "UPDATED_AT", "order":"ASCENDING"}'
```

其他有用的示例脚本

在 Amazon 中处理数据时，您可以使用以下示例脚本来完成各种任务 DataZone。

使用以下示例脚本列出现有的 Amazon DataZone 域名：

```
def list_domains():
    datazone = boto3.client('datazone')
    response = datazone.list_domains(status='AVAILABLE')
    [print("%12s | %16s | %12s | %52s" % (item['id'], item['name'],
 item['managedAccountId'], item['portalUrl'])) for item in response['items']]
    return
```

使用以下示例脚本列出现有的 Amazon DataZone 项目：

```
def list_projects(domain_id):
    datazone = boto3.client('datazone')
    response = datazone.list_projects(domainIdentifier=domain_id)
    [print("%12s | %16s " % (item['id'], item['name'])) for item in response['items']]
    return
```

使用以下示例脚本列出现有的 Amazon DataZone 元数据表单：

```
def list_metadata_forms(domain_id):
    datazone = boto3.client('datazone')
    response = datazone.search_types(domainIdentifier=domain_id,
        managed=False,
        searchScope='FORM_TYPE')
    [print("%16s | %16s | %3s | %8s" % (item['formTypeItem']['name'],
 item['formTypeItem']['owningProjectId'],item['formTypeItem']['revision'],
 item['formTypeItem']['status'])) for item in response['items']]
    return
```

Amazon 中的域名和用户访问权限 DataZone

本节介绍如何在 Amazon 中创建和管理域名和用户访问权限 DataZone。

Amazon DataZone 域名是将您的资产、用户及其项目连接在一起的组织实体。借助 Amazon DataZone 域名，您可以灵活地反映组织结构的数据和分析需求，无论是为企业创建单个 Amazon DataZone 域还是为不同的业务部门或团队创建多个数据区；域名。

本节还介绍如何管理用户对亚马逊 DataZone 控制台和亚马逊门户的访问权限。 DataZone

有关更多信息，请参阅 亚马逊 DataZone 术语和概念。

主题

- 创建 Amazon DataZone 域名
- 编辑 Amazon DataZone 域名
- 删除 Amazon DataZone 域名
- 启用 Amazon 的 IAM 身份中心 DataZone
- 禁用 Amazon 的 IAM 身份中心 DataZone
- 在 Amazon DataZone 控制台中管理用户
- 在 Amazon DataZone 数据门户中管理用户权限
- 限制访问 Amazon DataZone
- 将亚马逊 DataZone 域名升级为亚马逊 SageMaker 统一域名

创建 Amazon DataZone 域名

> **Note**
>
> 如果您使用 DataZone 带 AWS 身份中心的 Amazon 来向 SSO 用户和群组提供访问权限，则当前您的亚马逊 DataZone 域必须与您的 AWS 身份中心实例位于同一 AWS 区域。

Amazon DataZone，域名是一个组织实体，用于将您的资产、用户及其项目联系在一起。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。

要创建 Amazon DataZone 域名，您必须在账户中扮演具有管理权限的 IAM 角色。 配置使用亚马逊所需的 IAM 权限 DataZone 管理控制台以获得创建域所需的最低权限。

Amazon 需要其他 IAM 角色 DataZone 才能代表具有默认配置的域用户执行操作。您可以提前创建这些 IAM 角色，也可以让 Amazon 为您 DataZone 创建它们。如果您希望 Amazon DataZone 在域名创建过程中为您创建这些 IAM 角色，那么要创建域，您必须担任具有角色创建权限的 IAM 角色。请参阅为 IAM 权限创建自定义策略以启用亚马逊 DataZone 服务控制台简化了角色创建。根据您的域名创建选择，Amazon DataZone 将为您创建最多四个新的 IAM 角色：AmazonDataZoneDomainExecutionRoleAmazonDataZoneGlueManageAccessRole、AmazonDataZone和AmazonDataZoneProvisioningRole。

完成以下步骤以创建 Amazon DataZone 域名。

1. 前往位于 https://console.aws.amazon.com/datazone 的亚马逊 DataZone 控制台，然后使用顶部导航栏中的区域选择器选择相应的区域。 AWS
2. 选择创建域，并提供以下字段的值：

   - 名称 – 指定域的友好名称。创建域后，便无法更改此名称。
   - 描述 – (可选) 指定域描述。
   - 数据加密-您的亚马逊 DataZone 域名、元数据和报告数据由 AWS 密钥管理服务 (KMS) 使用您的亚马逊特有的密钥进行加密 DataZone。使用此字段指定是要使用 AWS 自有密钥还是选择其他 AWS KMS 密钥。

     有关客户自主管理型密钥的更多信息，请参阅 亚马逊的静态数据加密 DataZone。如果您使用自己的 KMS 密钥进行数据加密，则必须在默认 AmazonDataZoneDomainExecutionRole 中包含以下语句。

     JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Sid": "Statement1",
            "Effect": "Allow",
            "Action": [
                "kms:Decrypt",
                "kms:DescribeKey",
                "kms:GenerateDataKey"
            ],
```

```
            "Resource": [
                "arn:aws:kms:us-
east-1:111122223333:key/1234abcd-12ab-34cd-56ef-1234567890ab"
            ]
        }
    ]
}
```

- 服务访问权限-选择是让 Amazon DomainExecutionRole为您 DataZone 创建和使用新的 IAM 角色，还是选择现有的 IAM 角色。
- 快速设置-（可选）勾选此复选框，让 Amazon 为您的账户 DataZone 设置数据消耗和发布功能，从而更快地开始使用。亚马逊 DataZone 将创建三个 IAM 角色用于配置、接收和管理对 GI AWS ue 和 Amazon Redshift 资源的访问权限，创建一个新的 Amazon S3 存储桶，创建管理 DataZone 亚马逊项目，以及为数据湖和数据仓库默认蓝图创建环境配置文件。
- 标签-（可选）为域指定 AWS 标签（键和值对）。
- 成功创建域名后，您的浏览器应刷新以显示您的新 Amazon DataZone 域名的详情页面。

编辑 Amazon DataZone 域名

在 Amazon 中 DataZone，域名是一个组织实体，用于将您的资产、用户及其项目联系在一起。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。

创建 Amazon DataZone 域后，您可以稍后编辑该域名以：更改描述、启用 IAM Identity Center 以及添加、编辑或删除标签密钥及其值。要编辑 Amazon DataZone 域名，您必须在账户中扮演具有管理权限的 IAM 角色。 配置使用亚马逊所需的 IAM 权限 DataZone 管理控制台以获得编辑域所需的最低权限。

要编辑域，请完成以下步骤：

1. 登录 AWS 管理控制台并在 https://console.aws.amazon.com/data zon DataZone e 上打开亚马逊控制台。
2. 选择查看域，然后从列表中选择域名。该名称是一个超链接。
3. 在域的详细信息页面上，选择编辑。
4. - 编辑描述。
   - 设置 IAM Identity Center 设置。在设置 AWS 亚马逊 IAM 身份中心 DataZone中详细了解这些设置。
   - 添加、编辑或移除标签键及其值。
5. 完成编辑操作后，选择更新域。

删除 Amazon DataZone 域名

在 Amazon 中 DataZone，域名是一个组织实体，用于将您的资产、用户及其项目联系在一起。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。

删除域是最终操作。删除操作将不可撤销地删除每个 Amazon DataZone 实体，包括数据源、项目、环境、资产、术语表和元数据表单。删除不会删除亚马逊 DataZone 可能帮助您创建的非亚马逊 DataZone AWS 资源，例如 IAM 角色、S3 存储桶、G AWS lue 数据库以及通过或 LakeFormation Redshift 授予的订阅授权。如果您不再需要这些资源，请在相应的 AWS 服务中将其删除。

为防止他人恶意删除域名，删除域名需要亚马逊的 IAM 管理权限 DataZone，您可以使用 IAM 进行配置。为防止他人意外删除域名，删除域名需要输入确认词（在 Amazon DataZone 控制台中）。

要删除域，请完成以下步骤：

1. 登录 AWS 管理控制台并在 https://console.aws.amazon.com/data zon DataZone e 上打开亚马逊控制台。
2. 选择查看域，然后从列表中选择域名。该名称是一个超链接。
3. 选择删除并查看信息性警告。
4. 键入请求的文本以确认您理解这些警告。选择删除。

> **Important**
>
> 删除域是一项不可撤销的操作，您或 AWS无法取消此操作。

> **Note**
>
> 当您或您的域用户在项目中创建环境时，Amazon DataZone 会在您的域名或关联账户中创建 AWS 资源，为您和您的域用户提供功能。以下是 Amazon DataZone 可能为您域中的项目创建的 AWS 资源列表以及默认名称。删除域名并不会删除您 AWS 账户中的任何此类 AWS 资源。
>
> - IAM 角色：datazone_usr_<environmentId>。
> - Glue 数据库：（1）<environmentName>_pub_db-*、（2）<environmentName>_sub_db-*。如果已经存在同名数据库，Amazon DataZone 将添加环境 ID。
> - Athena 工作组：<environmentName>-*。如果已经存在同名工作组，Amazon DataZone 将添加环境 ID。

- CloudWatch 日志组：datazone_ <environmentId>

启用 Amazon 的 IAM 身份中心 DataZone

> **Note**
>
> 要完成此过程，您必须在与您的 Amazon DataZone 域相同的 AWS 区域启用 IA AWS M 身份中心。

您可以使用 AWS IAM Identity Center 为 SSO 用户和群组提供访问您的 Amazon DataZone 数据门户的权限。完成后设置 AWS 亚马逊 IAM 身份中心 DataZone，您可以允许您的 SSO 用户和群组访问您的 Amazon DataZone 域名数据门户。

要启用 AWS IAM Identity Center 与您的 Amazon DataZone 域名一起使用，您必须在账户中扮演具有管理权限的 IAM 角色。 配置使用亚马逊所需的 IAM 权限 DataZone 管理控制台并为 IAM 权限创建自定义策略以启用亚马逊 DataZone 服务控制台简化了角色创建获得启用 IAM 身份中心以便在 Amazon 上使用所需的最低权限 DataZone。

完成以下步骤以启用 Amazon AWS 的 IAM 身份中心 DataZone。

1. 登录 AWS 管理控制台并在 https://console.aws.amazon.com/data z DataZone one 上打开控制台。
2. 选择查看域，然后从列表中选择域名。该名称是一个超链接。
3. 在域的详细信息页面上，选择编辑。
   - 选中启用 IAM Identity Center 中的用户对应的复选框。
   - 选择是连接到 IAM Identity Center 的组织实例还是连接到 IAM Identity Center 的账户实例。
   - 在两种用户分配模式之间进行选择。在使用选定项更新域后，以后将无法更改域。
     - 通过隐式用户分配，任何添加到您的 IAM 身份中心目录的用户都可以访问您的 Amazon DataZone 域。
     - 使用显式用户分配，您将从 IAM Identity Center 目录中添加特定用户或群组，为他们提供访问您的 Amazon DataZone 域的权限。稍后，您将在 Amazon DataZone 控制台中添加和删除这些用户和群组。
4. 在对选定项感到满意后，选择更新域。

禁用 Amazon 的 IAM 身份中心 DataZone

禁用 AWS Amazon DataZone 域名的 IAM 身份中心将取消所有 SSO 用户的访问权限。

> **Note**
>
> 禁用 IAM Identity Center 将不会停止对 SSO 用户计费。要停止对 SSO 用户计费，您必须在您的域中停用他们。计费将一直持续到用户被停用当月的月底。要停用用户，请参阅在 Amazon DataZone 控制台中管理用户。

您可以使用 AWS IAM Identity Center 为 SSO 用户和群组提供访问您的 Amazon DataZone 数据门户的权限。如果您已启用适用于 Amazon AWS 的 IAM Identity Center DataZone，则可以稍后禁用所有用户的访问权限。

要禁用用于您的 Amazon DataZone 域的 IA AWS M 身份中心，您必须在账户中扮演具有管理权限的 IAM 角色。 配置使用亚马逊所需的 IAM 权限 DataZone 管理控制台并为 IAM 权限创建自定义策略以启用亚马逊 DataZone 服务控制台简化了角色创建获得禁用 IAM 身份中心在 Amazon 上使用所需的最低权限 DataZone。

完成以下步骤以禁用 Amazon AWS 的 IAM 身份中心 DataZone。

1. 登录 AWS 管理控制台并在 https://console.aws.amazon.com/data z DataZone one 上打开控制台。
2. 选择查看域，然后从列表中选择域名。该名称是一个超链接。
3. 复制域的 Amazon 资源名称（ARN），其开头为 arn:aws:datazone:<regionName>:<accountId>:domain/<domainName>。
4. 打开 IAM 身份中心控制台，网址为https://console.aws.amazon.com/singlesignon/。
5. 选择应用程序。
6. 选择您要禁用 AWS IAM Identity Center 的域，这将取消所有 SSO 用户对该域数据门户的访问权限。您可以使用筛选条件菜单和搜索框来筛选应用程序列表。
7. 从操作菜单上，选择禁用。
8. SSO 用户将无法访问 Amazon DataZone 域名。
9. 要为亚马逊 DataZone 域重新启用 AWS IAM 身份中心，请选择要为其重新启用 AWS IAM 身份中心的域，然后从 “操作” 菜单中选择 “启用”。

在 Amazon DataZone 控制台中管理用户

您的用户可以使用其 AWS 凭证或单点登录 (SSO) 凭证访问亚马逊 DataZone 数据门户。要在亚马逊 DataZone控制台中管理亚马逊 DataZone 域的用户，您必须在该账户中扮演具有亚马逊 DataZone 管理控制台权限的 IAM 角色。 配置使用亚马逊所需的 IAM 权限 DataZone 管理控制台以获得在 Amazon DataZone 控制台中管理用户所需的最低权限。

主题

- 管理 IAM 角色和用户
- 管理 SSO 用户
- 管理 SSO 组

管理 IAM 角色和用户

IAM 角色和用户使用 AWS 身份和访问管理 (IAM) 创建，并通过策略附加的权限访问您的 DataZone 亚马逊域名。有关更多信息，请参阅 配置使用亚马逊所需的 IAM 权限 DataZone 数据门户。在当前版本的 Amazon 中 DataZone，来自亚马逊 DataZone 域名所有者账户的管理员可以为自己账户中的用户或关联账户中的用户创建 IAM 用户个人资料。亚马逊 DataZone 域名所有者账户的管理员也可以将现有用户的状态设置为 “已分配” 或 “未分配”（如已分配或未分配以使用亚马逊 DataZone），或者激活或停用任何现有用户。

1. 登录 AWS 管理控制台并在 https://console.aws.amazon.com/data z DataZone one 上打开控制台。
2. 选择查看域，然后从列表中选择域名。该名称是一个超链接。
3. 在域详细信息页面上，选择用户管理。
4. 要在 Amazon DataZone 域名所有者账户或关联账户中添加用户 IAM 用户，请选择添加，然后选择添加 IAM 用户。
5. 在添加用户页面上，选择当前账户或关联的账户，使用查找和添加用户或角色字段以查找要添加的用户，然后选择添加用户。
6. 要查看现有 IAM 用户的状态，请在用户管理页面上的用户类型下拉菜单中选择 IAM 用户。

    - 名称列显示 IAM 用户或角色的 ARN。
    - 状态列显示域中的 IAM 用户或角色的当前状态。
        - 已指定 IAM 用户已被分配使用亚马逊 DataZone。
        - “未分配” 表示已取消指定 IAM 用户使用亚马逊。 DataZone

- 已激活意味着 IAM 用户或角色已调用 API、发出命令（通过命令行界面）或访问了您的域名的 Amazon DataZone 门户。
- 停用意味着 IAM 用户或角色无法再使用 Amazon DataZone 数据门户。要限制以编程方式进行访问，请参阅限制访问 Amazon DataZone。

7. 要停用当前已激活的 IAM 用户或角色，请选中该用户旁边的框，然后从操作菜单中选择停用。这将导致用户无法再使用亚马逊 DataZone 数据门户。要限制以编程方式进行访问，请参阅限制访问 Amazon DataZone。
8. 要激活当前已停用的 IAM 用户或角色，请选中该用户旁边的框，然后从操作菜单中选择激活。如果 IAM 用户或角色拥有`datazone:GetUserPortalLoginUrl`权限，则该用户将获得对 Amazon DataZone 数据门户的访问权限。

管理 SSO 用户

可以通过身份提供商创建或同步 SSO 用户。有关更多信息，请参阅设置 AWS 亚马逊 IAM 身份中心 DataZone和启用 Amazon 的 IAM 身份中心 DataZone以启用和配置适用于 Amazon 的 AWS IAM 身份中心 DataZone。您可以查看分配给域的 SSO 用户的列表、添加 SSO 用户和移除 SSO 用户。

1. 登录 AWS 管理控制台并在 https://console.aws.amazon.com/data z DataZone one 上打开控制台。
2. 选择查看域，然后从列表中选择域名。该名称是一个超链接。
3. 在域详细信息页面上，向下滚动并选择用户管理。
4. 对于用户类型，请选择 SSO 用户以查看之前已通过数据门户进行身份验证的 SSO 用户的当前列表。使用隐式用户分配时，之前未通过数据门户进行身份验证的 SSO 用户不会被列出。

   - 名称列显示 SSO 用户名。
   - 状态列显示域中的 SSO 用户的当前状态。
     - “已分配”表示已将 SSO 用户显式分配给域。因此，用户可以访问亚马逊 DataZone。仅当域的身份提供商模式设置为显式分配时，才使用此状态。
     - 已激活表示 SSO 用户已访问该域名的 Amazon DataZone 门户。将自动进行激活。
     - “已停用”表示 SSO 用户被阻止访问域的数据门户。
     - “已移除”表示 SSO 用户之前已被分配到域，但该用户在进行访问前已被移除。
5. 通过选择添加和添加用户来添加 SSO 用户。如果域名设置为隐式用户分配，则此选项不可用，这意味着身份池中的所有用户都可以访问该Amazon DataZone 域。

   - 在添加用户页面上，搜索要添加的用户的别名。搜索框下方将显示包含潜在匹配项的列表。

- 选择要添加的用户。他们的别名将以木条形式显示在搜索框下方。
- 如果您对要添加的用户列表感到满意，请选择添加用户。
- 用户被分配到状态为 “已分配” 的 Amazon DataZone 域。
- 当用户首次访问域的数据门户时，状态将自动更改为已激活。

6. 通过以下方式移除已分配 SSO 用户：选择该用户并从操作菜单中选择取消分配。因此，用户将失去对 Amazon DataZone 域的访问权限。用户的状态将显示为未分配。如果将域设置为隐式用户分配，则此选项不可用。
7. 通过以下方式停用已激活 SSO 用户：选择该用户并从操作菜单中选择停用。因此，用户对 Amazon DataZone 数据门户的访问权限将丢失并被阻止。用户的状态将显示为已停用。
8. 通过以下方式激活已停用 SSO 用户：选择该用户并从操作菜单中选择激活。因此，用户将重新获得对 Amazon DataZone 数据门户的访问权限。用户的状态将显示为已激活。

管理 SSO 组

SSO 组是在 AWS IAM 身份中心中创建的，或者与您的身份提供商同步。有关更多信息，请参阅设置 AWS 亚马逊 IAM 身份中心 DataZone和启用 Amazon 的 IAM 身份中心 DataZone以启用和配置适用于 Amazon 的 AWS IAM 身份中心 DataZone。您可以查看分配给域的 SSO 组的列表、添加 SSO 组和移除 SSO 组。

1. 登录 AWS 管理控制台并在 https://console.aws.amazon.com/data z DataZone one 上打开控制台。
2. 选择查看域，然后从列表中选择域名。该名称是一个超链接。
3. 在域详细信息页面上，向下滚动并选择用户管理。
4. 对于用户类型，选择 SSO 组以查看当前的 SSO 组列表。

   - 名称列显示 SSO 组名。
   - 状态列显示域中的 SSO 组的当前状态。
     - 已分配表示已将 SSO 组显式分配给域。因此，组中的所有用户都可以访问域的数据门户（除非用户已被停用）。
     - 未分配表示已从域中移除 SSO 组。组中的用户无法通过其在组中的成员资格访问域的数据门户。
5. 通过选择添加和添加组来添加 SSO 组。如果域名设置为隐式用户分配，则此选项不可用，这意味着无论群组成员资格如何，身份池中的所有用户都可以访问Amazon DataZone 域。

- 在添加组页面上，搜索要添加的组的别名。搜索框下方将显示包含潜在匹配项的列表。
- 选择要添加的组。他们的别名将以木条形式显示在搜索框下方。
- 如果您对要添加的组列表感到满意，请选择添加组。
- 这些群组被分配到状态为 “已分配” 的 Amazon DataZone 域。
- 当组成员首次访问域的数据门户时，状态将自动更改为已激活。

6. 通过以下方式移除已分配 SSO 组：选择该组并从操作菜单中选择取消分配。因此，该群组将无法访问 Amazon DataZone 域名。组的状态将显示为未分配。 DataZone 通过该群组的成员资格获得Amazon访问权限的用户将失去访问权限。如果将域设置为隐式用户分配，则此选项不可用。

在 Amazon DataZone 数据门户中管理用户权限

您可以使用 Amazon DataZone 管理门户为 IAM 用户和角色、SSO 用户和群组以及 SAML 用户配置身份验证。亚马逊 DataZone 会为每位使用亚马逊 DataZone的用户分配一个用户档案。

使用项目、创建实体等操作的用户配置文件权限是通过域单元和策略授权管理的。具体项目中通过指定项目成员资格（所有者、贡献者、查看者）来决定操作授权。

限制访问 Amazon DataZone

限制对 Amazon 的编程访问权限 DataZone-对于进行编程 API 调用的 IAM 用户或角色，可以通过 IAM 策略限制访问权限。如果您想撤销任何已经为角色颁发的短期凭证，则可以对角色或服务控制策略使用 IAM 撤销会话机制。

限制登录访问亚马逊 DataZone 数据门户-为了限制对亚马逊 DataZone 数据门户的登录访问权限，对于 IAM 用户或角色，IAM 策略可以限制对`datazone:GetUserPortalLoginUrl`操作的访问权限。对于 SSO 用户和群组，请将亚马逊 DataZone 用户资料状态设置为 “已停用”，从而限制对亚马逊 DataZone 数据门户的访问。如果您的域名配置为隐式分配，并且该用户以前未使用过 Amazon DataZone，则需要将该用户从身份提供商中移除。

将亚马逊 DataZone 域名升级为亚马逊 SageMaker 统一域名

升级域之前的考虑事项

在将您的亚马逊 DataZone 域名升级到亚马逊 SageMaker 统一域名之前，请查看以下重要注意事项，确保升级过程顺利进行。

- 升级过程只能通过 AWS 管理控制台进行。目前，不提供用于升级域的 API 支持。您可以从 Amazon DataZone 域名的域名详情页面初始化升级流程。
- 升级过程需要配置以下角色（您可以选择现有角色或让 Amazon SageMaker Unified Studio 代表您创建角色）：
  - 域名执行角色-对于亚马逊 DataZone 域名，您可以使用亚马逊要求的角色 DataZone 对域中的数据进行分类、发现、管理、共享和分析。AmazonDataZoneDomainExecutionRole对于 Amazon SageMaker 统一域名，您必须使用现有域名或创建新AmazonSageMakerDomainExecution角色。
  - 域名服务角色-Amazon DataZone 不需要域名服务角色。对于 Amazon SageMaker 统一域名，您必须使用现有域名或创建新AmazonSageMakerDomainService角色。这是由 Amazon SageMaker Unified Studio 执行的域级操作的服务角色。
- 根域所有权考虑事项：
  - 在升级过程中， users/groups 可以选择将 IAM 用户或 SSO 分配为根域所有者。
  - 如果根域单位仅将 IAM 角色分配为所有者，则建议您添加 IAM 用户或 SSO user/group 作为所有者。有关更多信息，请参阅《Amazon DataZone 管理员指南》中的用户管理。
  - 重要提示：IAM 角色无法登录亚马逊 SageMaker 统一工作室。
- 关联账户和 Res AWS ource Access Manager (AWS RAM) 的更改：
  - 关联账户使用 AWS RAM 中的资源共享来允许根域账户执行 API 操作。
  - 升级过程会更改由 Amazon 创建和管理的 AWS RAM 共享的底层托管权限 DataZone。受影响的托管权限为 `AWSRAMPermissionsAmazonDatazoneDomainExtendedServiceAccess` 和 `AWSRAMPermissionsAmazonDatazoneDomainExtendedServiceWithPortalAccess`。
- Amazon Q 订阅变更 – 升级后的域将使 Amazon Q 订阅默认为免费套餐。域升级完成后，域管理员可以更改此设置。
- 升级后，域的 `domainVersion` 属性将从 `V1` 更改为 `V2`。

将您的亚马逊 DataZone 域名升级为亚马逊 SageMaker 统一域名

您可以完成以下步骤将您的亚马逊 DataZone 域名升级为亚马逊 SageMaker 统一域名。

1. 前往位于 https://console.aws.amazon.com/datazone 的亚马逊 DataZone 控制台，然后使用顶部导航栏中的区域选择器选择相应的区域。 AWS
2. 选择您要升级的 Amazon DataZone 域名，然后导航到其详情页面。
3. 在域名的详情页面上，选择位于将您的域名升级到 Amazon SageMaker Unified Studio 通知中的开始使用按钮。

4. 在将您的域名升级到 Amazon SageMaker Unified Studio 页面上，选择开始。
5. 接下来，如果您要升级的 Amazon 域没有类型为 IAM 用户、SSO 用户/组的所有者，请指定 DataZone 域执行角色和域服务角色以及根域单位所有者。然后，选择升级域。

有关将 Amazon 域名升级为亚马逊 SageMaker 统一 DataZone 域名的常见问题

- 升级后，哪些属性和配置会随域一起转移？

  在亚马逊 DataZone 域上配置的所有属性都将转移到升级后的亚马逊 SageMaker 统一域中。这包括数据加密属性、身份验证应用程序属性等。

- 我是否需要再次为用户设置单点登录（SSO）访问权限？

  不是。 您与该域关联的 IAM Identity Center SSO 应用程序将延续到升级后的 Amazon SageMaker 统一域中。此外，分配给该域的任何 IAM 用户或角色都将在升级后的 Amazon SageMaker 统一域中可用。

- 升级后我还能使用 Amazon DataZone 门户吗？

  可以。升级后，亚马逊 DataZone 门户网站和亚马逊 SageMaker 统一工作室都将可供最终用户进行互动。在域管理员通过亚马逊 SageMaker 管理控制台停用亚马逊门户网站之前，这两个 DataZone 门户网站都将保持打开状态。

- 我能否在 Amazon SageMaker Unified Studio 中看到在亚马逊 DataZone 门户网站上创建的项目和其他实体？

  可以。通过亚马逊 DataZone 门户网站创建的大多数实体（项目、元数据表单、词汇表、域单元）都将显示在 Amazon Uni SageMaker fied Studio 中。项目将保留与资产、资产订阅、成员等相关的所有资产、元数据表单和词汇表。这些项目需要从 AWS Athena 或 Amazon Redshift 查询编辑器中查询数据。元数据表单和词汇表将显示在 Amazon SageMaker Unified Studio 中，可以从亚马逊 SageMaker 对其进行编辑，也可以分配给通过亚马逊创建的项目中的资产。 SageMaker亚马逊的环境和环境配置文件 DataZone 不会显示在 Amazon SageMaker Unified Studio 中，这些实体已被亚马逊 SageMaker 项目概况所取代。在 Amazon SageMaker Unified Studio 中创建的项目将无法通过亚马逊 DataZone 门户网站查看。

- 升级到 Amazon SageMaker 统一域名后，域标识符和项目标识符会怎样？

  升级后，包括域和项目在内的所有实体标识符都将保持不变。

- 我的 AWS CloudFormation (CFN) 堆栈能否继续适用于新升级的 Amazon SageMaker 统一域名？

Amazon SageMaker Unified Studio 的使用方式与亚马逊 APIs 相同 DataZone。但是，CFN 模板中的逻辑需要进行某些修改。例如，来自亚马逊的域名与亚马逊 DataZone SageMaker统一域名的区别在于名为 domainVersion 的属性（值 V1 | V2）。

- 回滚升级后会出现什么情况？
  - 回滚升级会将域版本从 V2 更改为 V1。Amazon SageMaker Unified Studio 将无法再访问。该域名的控制台视图将返回到 Amazon DataZone 视图。回滚之前创建的资源只要不与从 Amazon Unified Studio 创建的项目绑定，就将保持不变，只有在不存在从 Amazon SageMaker Uni SageMaker fied Studio 内部创建的项目时，才允许回滚。
  - 回滚后，诸如 AWS Q 订阅之类的设置也将保留。
  - 如果 VPCs 是为供Amazon使用而创建的 SageMaker，则这些内容将在回滚后继续存在。SageMaker 服务创建的 VPC 将带有标签：名称 = SageMakerUnifiedStudio VPC
  - RAM 资源共享下的托管权限不会被回滚。托管权限是亚马逊 DataZone和亚马逊 U SageMaker nified Studio 的超集。
  - 已回滚的域可以再次升级为 Amazon SageMaker 统一域名。

Amazon 中的域名单位和授权政策 DataZone

使用域单元可轻松地在特定的业务部门和团队下组织资产和其他域实体。要在组织各业务部门内部和各业务部门之间设置安全高效的数据共享，请在 Amazon 中创建域单元， DataZone 并允许每个业务部门内的选定用户登录并将其资产共享到目录中。企业中任何地方的用户都可以轻松搜索这些业务部门下的资产，并请求对这些资产的访问权限。

域单位还可用于使资源所有者（例如 AWS 账户所有者）能够对其资源设置 Amazon DataZone 授权权限。域单元提供从账户所有者到域单元所有者的委托授权，他们可以代表账户所有者对环境配置文件（使用蓝图配置创建）设置授权权限。这可让您根据他们所属的业务部门来限制谁可以创建和使用哪些环境配置文件。Amazon DataZone 授权权限还可用于强制执行元数据标准，并仅允许选定的项目创建元数据表单和词汇表。这有助于维护一致的高质量元数据。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。

在 Amazon DataZone 域单位内，您可以将以下授权策略分配给您的用户和群组，以授予他们特定的权限：

- 域单元创建策略
- 项目创建策略
- 项目成员资格策略
- 域单元所有权代入策略
- 项目所有权代入策略

有关更多信息，请参阅 为 Amazon DataZone 域单位内的用户和群组分配授权策略。

在 Amazon DataZone 域单位内，您可以将以下授权策略分配给您的项目，以授予其特定权限：

- 术语表创建策略
- 元数据表单创建策略
- 自定义资产类型创建策略

有关更多信息，请参阅 为 Amazon DataZone 域单位内的项目分配授权策略。

在 Amazon 中使用授权机制的另一种方法 DataZone 是将授权策略应用于亚马逊 DataZone 蓝图配置中的项目和域单元所有者。

Amazon DataZone 蓝图配置是一个实体，它封装了创建和配置发布和订阅用户工作流程中使用的资源所需的信息。此信息包括 AWS 账号和区域、 CloudFormation 模板、账户级别参数（例如 VPCs 和子网），还可以包含数据库连接信息和凭证。为了控制成本并提高安全性，数据平台用户需要能够控制谁可以使用这些蓝图并创建环境。

在特定的蓝图配置中，您可以将以下授权策略分配给项目和域单元所有者：

- 使用此蓝图创建环境配置文件-此策略可以分配给 Amazon DataZone 项目，并授权他们使用此蓝图创建环境配置文件。
- 授予使用此蓝图创建环境配置文件所需的权限 - 可将此策略分配给域单元所有者，以便授权他们允许项目使用此蓝图创建环境配置文件。

有关更多信息，请参阅 在 Amazon DataZone 蓝图配置中分配授权策略。

主题

- 在 Amazon 中创建域名单位 DataZone
- 在 Amazon 中编辑域名单位 DataZone
- 在 Amazon 中删除域名单位 DataZone
- 在 Amazon 中管理域名单位所有者 DataZone
- 为 Amazon DataZone 域单位内的用户和群组分配授权策略
- 为 Amazon DataZone 域单位内的项目分配授权策略
- 在 Amazon DataZone 蓝图配置中分配授权策略

在 Amazon 中创建域名单位 DataZone

在 Amazon 中 DataZone，域名单位使您能够在特定的业务部门和团队下组织您的资产和其他域名实体。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。

创建域单元

1. 使用 DataZone 数据门户 URL 导航至 Amazon 数据门户，然后使用您的 SSO 或 AWS 凭证登录。如果您是亚马逊 DataZone 管理员，则可以通过访问创建亚马逊 DataZone 域名的 AWS 账户中的 https://console.aws.amazon.com/datazon e 上的亚马逊 DataZone 控制台来获取数据门户 URL。
2. 选择查看域，然后选择要在其中创建域单元的域。

3. 在域详细信息页面上，导航到域单元选项卡。
4. 选择创建域单元。
5. 指定以下项，然后选择创建域单元：

   - 在域单元详细信息下，对于名称，指定域单元名称。
   - 在域单元详细信息下，对于描述，指定域单元描述。
   - 域单元父级 – 选择要在其下添加新域单元的父域单元。
   - 域单元所有者 – 指定可以编辑此域单元的域单元所有者。

在 Amazon 中编辑域名单位 DataZone

在 Amazon 中 DataZone，域名单位使您能够在特定的业务部门和团队下组织您的资产和其他域名实体。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。

编辑域单元

1. 使用 DataZone 数据门户 URL 导航至 Amazon 数据门户，然后使用您的 SSO 或 AWS 凭证登录。如果您是亚马逊 DataZone 管理员，则可以通过访问创建亚马逊 DataZone 域名的 AWS 账户中的 https://console.aws.amazon.com/datazon e 上的亚马逊 DataZone 控制台来获取数据门户 URL。
2. 选择查看域，然后选择要在其中编辑域单元的域。
3. 在域详细信息页面上，导航到域单元选项卡，然后选择要编辑的域单元。
4. 展开操作并选择编辑域单元。
5. 对域单元名称和描述进行更改，然后选择保存更改。

在 Amazon 中删除域名单位 DataZone

在 Amazon 中 DataZone，域名单位使您能够在特定的业务部门和团队下组织您的资产和其他域名实体。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。

编辑域单元

1. 使用 DataZone 数据门户 URL 导航至 Amazon 数据门户，然后使用您的 SSO 或 AWS 凭证登录。如果您是亚马逊 DataZone 管理员，则可以通过访问创建亚马逊 DataZone 域名的 AWS 账户中的 https://console.aws.amazon.com/datazon e 上的亚马逊 DataZone 控制台来获取数据门户 URL。

2. 选择查看域，然后选择要在其中删除域单元的域。
3. 在域详细信息页面上，导航到域单元选项卡，然后选择要删除的域单元。
4. 展开“操作”并选择删除域单元。
5. 在删除域单元弹出窗口中，通过选择删除域单元来确认删除。

在 Amazon 中管理域名单位所有者 DataZone

在 Amazon 中 DataZone，域名单位使您能够在特定的业务部门和团队下组织您的资产和其他域名实体。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。

要通过 Amazon DataZone 管理控制台向顶级域单元添加所有者，请完成以下步骤。

1. 前往位于 https://console.aws.amazon.com/datazone 的亚马逊 DataZone 控制台，然后使用您的账户凭证登录。
2. 选择 “查看域名”，然后选择要向其中添加 DataZone 域名单位所有者的 Amazon 域名。
3. 在域详细信息页面上，导航到域根所有者表。
4. 选择添加，然后指定要设为域单元所有者的用户。选择添加根域所有者。

要通过 Amazon DataZone 数据门户添加域名单位所有者，请完成以下步骤：

1. 使用 DataZone 数据门户 URL 导航至 Amazon 数据门户，然后使用您的 SSO 或 AWS 凭证登录。如果您是亚马逊 DataZone 管理员，则可以通过访问创建亚马逊 DataZone 域名的 AWS 账户中的 https://console.aws.amazon.com/datazon e 上的亚马逊 DataZone 控制台来获取数据门户 URL。
2. 选择查看域，然后选择要在其中添加域单元所有者的域和域单元。
3. 在域单元详细信息页面上，选择所有者选项卡，然后选择添加所有者。
4. 在添加域单元所有者弹出窗口中，指定要设为域单元所有者的用户，然后选择添加所有者。

为 Amazon DataZone 域单位内的用户和群组分配授权策略

在 Amazon 中 DataZone，域名单位使您能够在特定的业务部门和团队下组织您的资产和其他域名实体。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。

在 Amazon DataZone 域单元中，您可以将以下授权策略分配给您的用户和群组，以授予他们在该域单位内的各种授权权限：

- 域单元创建策略
- 项目创建策略
- 项目成员资格策略
- 域单元所有权代入策略
- 项目所有权代入策略

要向域单元中的用户和组分配授权策略，请完成以下过程：

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。如果您是亚马逊 DataZone 管理员，则可以通过 https://console.aws.amazon.com/datazon e 导航到亚马逊 DataZone 控制台，使用域名创建 AWS 账户 地登录，然后选择打开数据门户。
2. 选择查看域，然后选择要为其分配授权策略的域和域单元。
3. 在域单元详细信息页面上，选择要分配给的授权策略， users/groups 然后选择添加用户。
4. 在添加用户弹出窗口中，执行下列操作之一：
   - 选择选定的用户和组，指定要向其分配所选授权策略的用户和组，然后选择添加用户。
   - 选择所有用户，然后选择添加用户。
   - 选择所有组，然后选择添加用户。
5. 您也可以为选定用户启用或禁用所选授权策略的级联权限。为此，请选择要为其启用级联权限的用户，再展开操作，然后选择将级联权限设置为 true。选定用户将在此域单元下的所有子域单元中拥有该策略授予的权限。也可以选择要为其禁用级联权限的用户，再展开操作，然后选择将级联权限设置为 false。

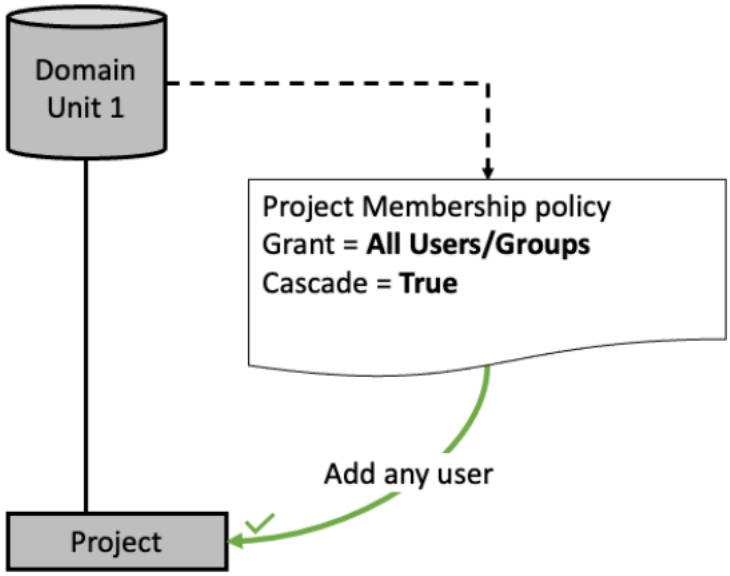
Amazon 域单元层次结构中的项目成员资格政策 DataZone

项目成员资格策略定义有资格作为成员被添加到域单元中的项目的人员或组。本主题描述了策略对分层结构中的单个和多个域单元产生的影响的场景。

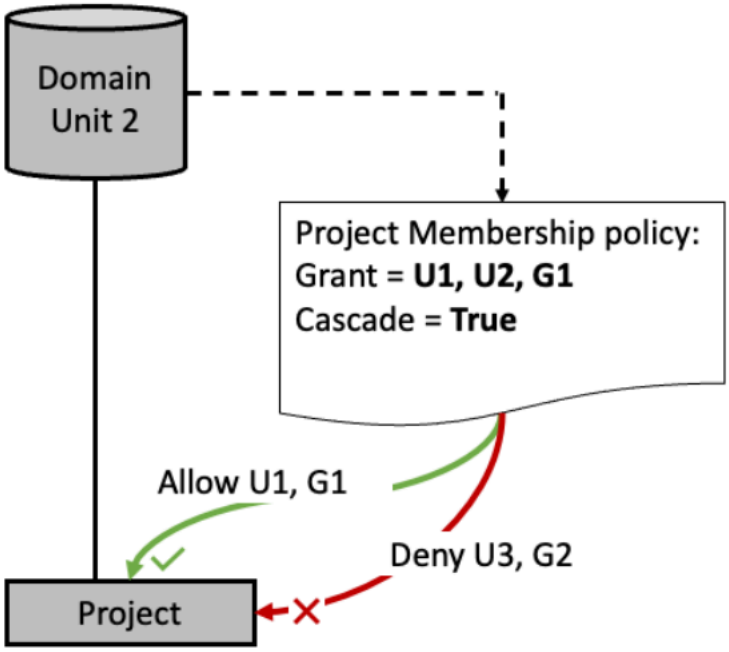
请务必注意本主题中使用的几个概念：

- 成员资格池 – 通过项目成员资格策略向其授予访问权限的主体（用户或组），被视为在项目成员资格池中。例如，如果将域单元 DU1 策略授予用户 U1 和 U2 以及单点登录 (SSO) 组 G1，则其项目成员资格池 DU1 将包括 {U1、U2、G1}。
- 级联 – 能够将授权向下传递给通过域单元层次结构连接的所有子域单元。
- 授权 – 用户或组执行操作所需的权限。

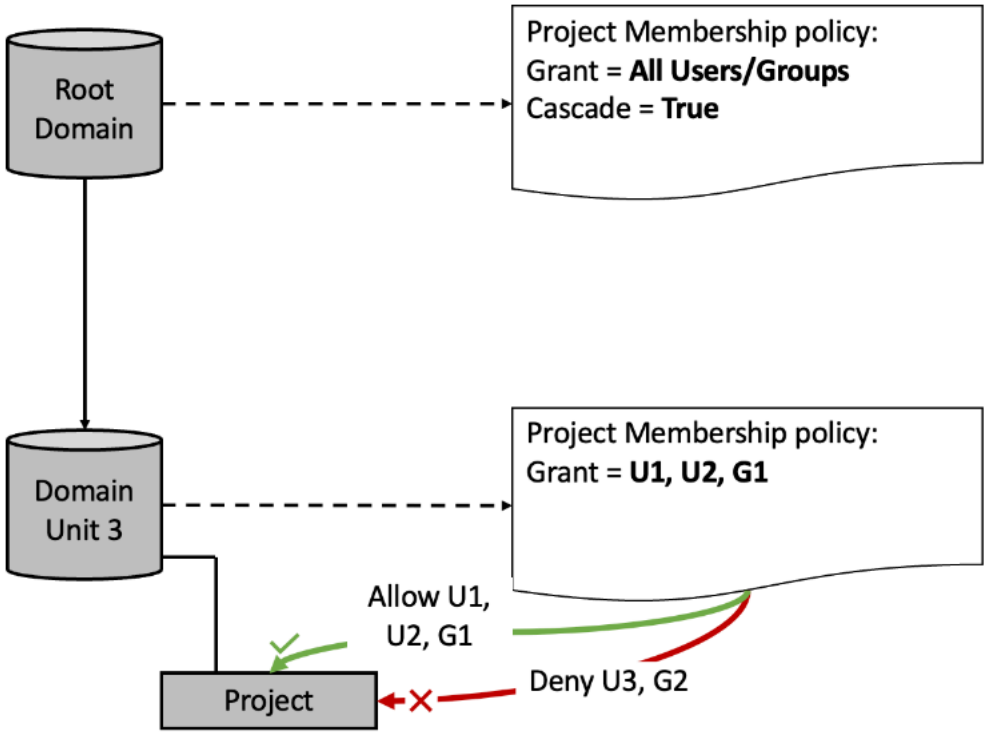
场景 1 – 任何用户或组均可添加到域单元 1 下的项目中，因为成员资格池包含 {All Users/Groups}。



场景 2 – 用户 {U1, G1} 可添加到域单元 2 下的项目中，因为他们在域单元 2 下的成员资格池中。用户 {U3, G2} 无法添加到任何项目中，因为他们不在成员资格池中。

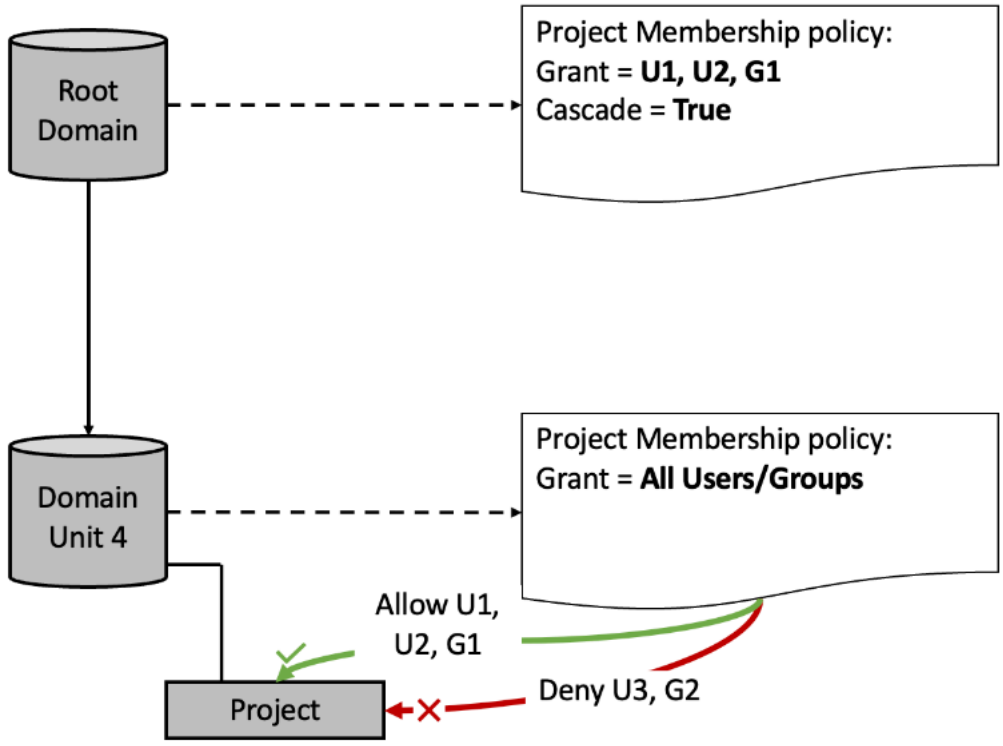


场景 3 – 成员资格池的交集：如果存在具有不同的域单元层次结构级别的成员资格池，则只能将位于所有成员资格池中的用户和组添加到项目中。



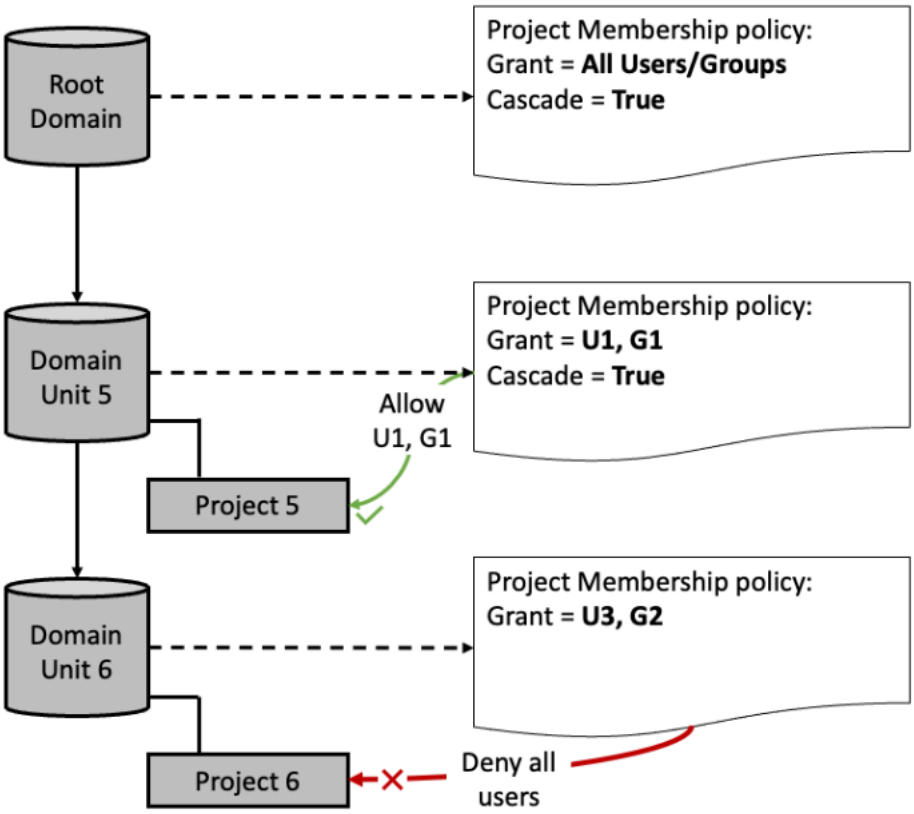
- 两个成员资格池中的用户交集为 {U1, U2, G1}。
- 用户 {U1, U2, G1} 可添加到域单元 3 下的项目中。
- 即使所有用户和所有组位于根域单元级别的成员资格池中，也无法将用户 {U3, G2} 添加到域单元 3 下的项目中。

场景 4 – 成员资格池的交集：如果存在具有不同的域单元层次结构级别的成员资格池，则只能将位于所有成员资格池中的用户和组添加到项目中。

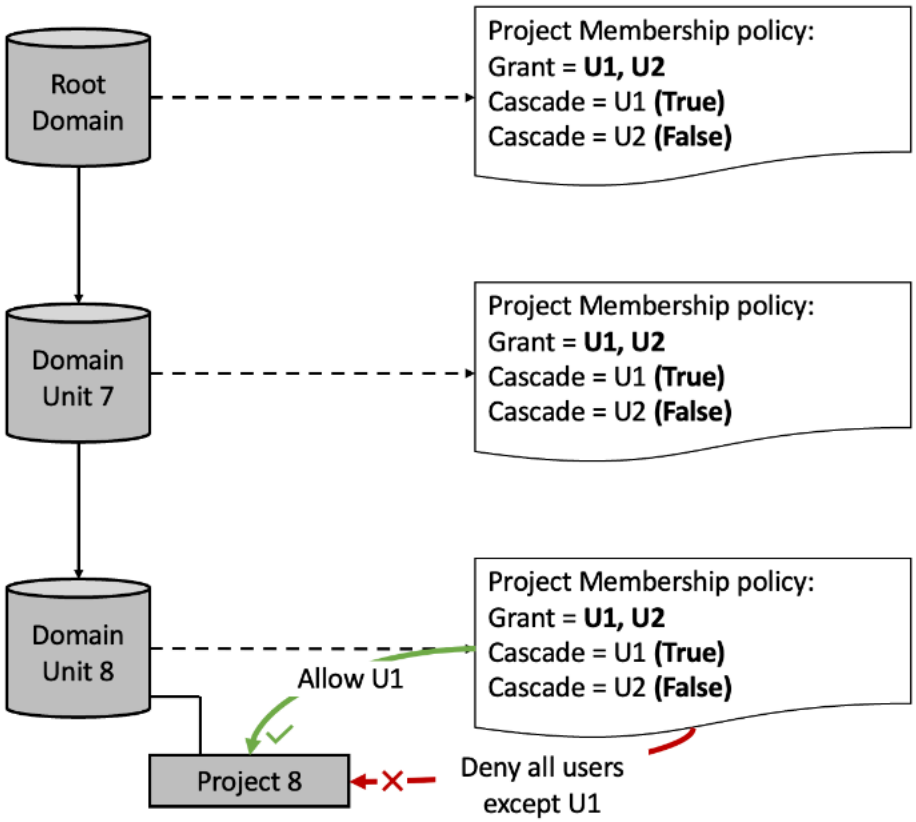


- 两个成员资格池中的用户交集为 {U1, U2, G1}。
- 域单元 4 的成员资格池为 {All Users / Groups}，但成员资格池不能扩展到根域 {U1, U2, G1} 的成员资格池之外。
- 即使所有用户和所有组位于域单元 4 的成员资格池中，也无法将用户 {U3, G2} 添加到域单元 4 下的项目中。

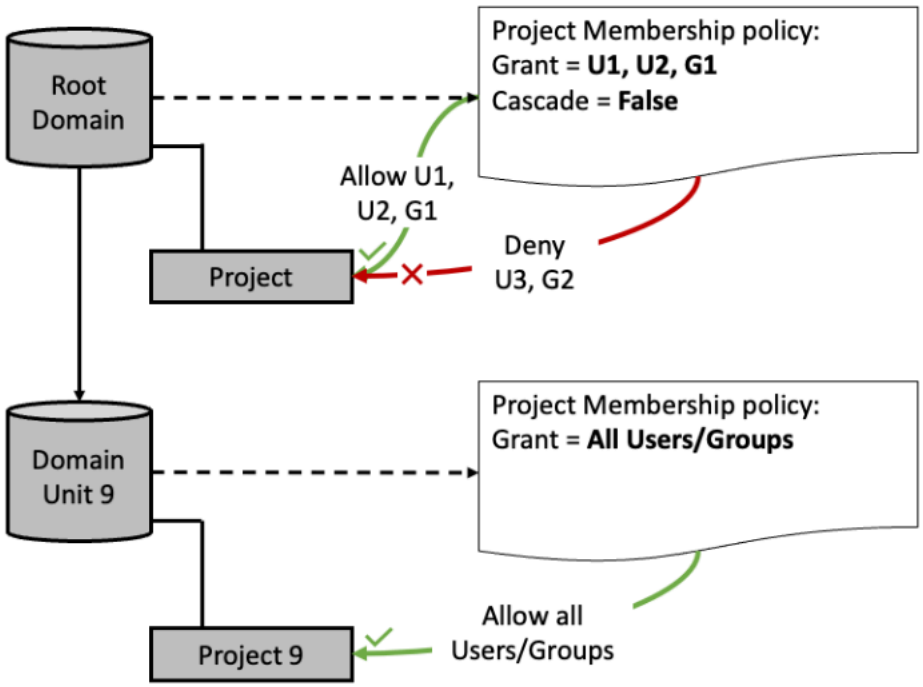
场景 5 – 用户 {U1, G1} 可添加到项目 5 中，因为他们在根域和域单元 5 之间的成员资格池交集中。由于三个成员池的交叉点为空，因此 user/group 无法向 Project 6 中添加。



场景 6 – 所有三个成员资格池的交集意味着仅用户 {U1} 可添加到项目 8 中。域单元 8 的交集是 {U1}、{U1}、{U1, U2} – 仅 {U1} 同时位于三个池中。



场景 7 – 用户 {U1, U2, G1} 可添加到根域的项目中，因为他们在根域的成员资格池中。任何用户或组都可添加到域单元 9 下的项目中，因为成员资格池包含 {All Users/Groups}，并且其上方的根域中的级联已设置为 false。



为 Amazon DataZone 域单位内的项目分配授权策略

在 Amazon 中 DataZone，域名单位使您能够在特定的业务部门和团队下组织您的资产和其他域名实体。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。

在 Amazon DataZone 域单元中，您可以将以下授权策略分配给您的项目，以向这些实体授予该域单位内的各种授权权限：

- 术语表创建策略
- 元数据表单创建策略
- 自定义资产类型创建策略

要向域单元中的项目分配授权策略，请完成以下过程：

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。如果您是亚马逊 DataZone 管理员，则可以通过 https://console.aws.amazon.com/datazon e 导航到亚马逊 DataZone 控制台，使用域名创建 AWS 账户 地登录，然后选择打开数据门户。
2. 选择管理域，然后选择要为其分配授权策略的域和域单元。
3. 在域单元详细信息页面上，选择要分配的授权策略，然后单击它进行配置。

在 Amazon DataZone 蓝图配置中分配授权策略

在 Amazon 中使用授权机制的另一种方法 DataZone 是将授权策略应用于亚马逊 DataZone 蓝图配置中的项目和域单元所有者。

Amazon DataZone 蓝图配置是一个实体，它封装了创建和配置发布和订阅用户工作流程中使用的资源所需的信息。此信息包括 AWS 账号和区域、CFN 模板、账户级别参数（例如 VPCs 和子网），还可以包含数据库连接信息和凭证。为了控制成本并提高安全性，数据平台用户需要能够控制谁可以使用这些蓝图并创建环境。

在特定的蓝图配置中，您可以将以下授权策略分配给项目和域单元所有者：

- 使用此蓝图创建环境配置文件-此策略可以分配给 Amazon DataZone 项目，并授权他们使用此蓝图创建环境配置文件。
- 授予使用此蓝图创建环境配置文件所需的权限 - 可将此策略分配给域单元所有者，以便授权他们允许项目使用此蓝图创建环境配置文件。

通过 Amazon DataZone 数据门户将使用此蓝图授权策略创建环境配置文件分配给蓝图配置中的项目

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。如果您是亚马逊 DataZone 管理员，则可以通过 https://console.aws.amazon.com/datazon e 导航到亚马逊 DataZone 控制台，使用域名创建 AWS 账户 地登录，然后选择打开数据门户。
2. 在数据门户中，选择具有要使用的已启用蓝图的域，然后导航到蓝图配置选项卡。
3. 在蓝图配置选项卡中，选择要使用的已启用蓝图，再在此蓝图的详细信息页面中，导航到授权策略选项卡，然后选择使用此蓝图创建环境配置文件授权策略。
4. 在使用此蓝图创建环境配置文件授权策略详细信息页面中，展开操作并选择添加项目。
5. 在添加项目弹出窗口中，可以执行下列操作之一：

    - 选择域单元中的所有项目选项，然后搜索并指定包含要授权使用此蓝图创建环境配置文件的项目的域单元，然后选择添加项目。
    - 选择域单元中的选定项目选项，搜索并指定包含要向其分配此策略的项目的域单元，再搜索并选择要向其分配此策略的项目，然后选择添加项目。

通过 Amazon DataZone 管理控制台，通过蓝图配置向域单元所有者分配使用此蓝图授权策略创建环境配置文件的授予权限

1. 前往位于 https://console.aws.amazon.com/datazone 的亚马逊 DataZone 控制台，然后使用您的账户凭证登录。
2. 在 Amazon DataZone 控制台中，选择要使用的已启用蓝图的域，然后导航到 Blueprin ts 选项卡。
3. 在蓝图选项卡中，选择要使用的已启用蓝图，然后在蓝图的详细信息页面中，导航到委派权限选项卡。
4. 在委派权限选项卡中，搜索并选择要为其所有者分配授予使用此蓝图创建环境配置文件的权限策略的域单元，然后选择添加委派权限。

亚马逊 DataZone 内置蓝图

用于创建环境的蓝图定义了环境所属项目的成员在处理 Amazon DataZone 目录中的资产时可以使用的工具和服务。在当前版本的 Amazon 中 DataZone，有以下内置蓝图：

- 数据湖蓝图
- 数据仓库蓝图
- 亚马逊 SageMaker 蓝图

您可以按照以下过程中的步骤在 Amazon DataZone 中启用默认蓝图：

- 在拥有 Amazon DataZone 域 AWS 名的账户中启用内置蓝图
- 将亚马逊 SageMaker 作为可信服务添加到拥有亚马逊 DataZone 域名的 AWS 账户中

在拥有 Amazon DataZone 域 AWS 名的账户中启用内置蓝图

用于创建环境的蓝图定义了环境所属项目的成员在处理 Amazon DataZone 目录中的资产时可以使用的工具和服务。

在当前版本的 Amazon 中 DataZone，有几个内置蓝图：数据湖蓝图、数据仓库蓝图和亚马逊 SageMaker 蓝图。

- 数据湖蓝图包含启动和配置一组服务（AWS Glue、 AWS Lake Formation、Amazon Athena）以发布和使用亚马逊目录中的数据湖资产的定义。 DataZone
- 数据仓库蓝图包含启动和配置一组服务（Amazon Redshift）的定义，以发布和使用亚马逊目录中的亚马逊 Redshift 资产。 DataZone
- 亚马逊 SageMaker 蓝图包含启动和配置一组服务（Amazon SageMaker Studio）以发布和使用亚马逊 DataZone 目录中的亚马逊 SageMaker 资产的定义。

有关更多信息，请参阅 亚马逊 DataZone 术语和概念。

创建 Amazon DataZone 域时，您可以选择在域创建过程中自动启用默认数据湖和默认数据仓库内置蓝图的快速设置。快速设置功能还使用这些内置蓝图为您创建默认环境配置文件和默认环境。

如果您在创建亚马逊 DataZone 域名时未选择快速设置，则可以使用以下步骤在存放此亚马逊 DataZone 域名的 AWS 账户中启用可用的内置蓝图。您必须先启用这些内置蓝图，之后才能使用它们在此域中创建环境配置文件和环境。

要通过亚马逊 DataZone 管理控制台在亚马逊 DataZone 域中启用内置蓝图，您必须在账户中扮演具有管理权限的 IAM 角色。 配置使用亚马逊所需的 IAM 权限 DataZone 管理控制台以获得最低权限。

在 Amazon DataZone 域中启用内置蓝图

1. 前往位于 https://console.aws.amazon.com/datazone 的亚马逊 DataZone 控制台，然后使用您的账户凭证登录。
2. 选择查看域，然后选择要在其中启用一个或多个内置蓝图的域。
3. 在域详细信息页面上，导航到蓝图选项卡。
4. 从蓝图列表中选择、DefaultDataLake或 Amazon SageMaker 蓝图。DefaultDataWarehouse
5. 在所选蓝图的详细信息页面上，选择在此账户中启用。
6. 在权限和资源页面上，指定以下角色：

   - 如果您要启用DefaultDataLake蓝图，请为 Glue 管理访问权限角色指定一个新的或现有的服务角色，该角色 DataZone 授予亚马逊收录和管理对 G AWS lue 和 La AWS ke Formation 中表的访问权限的授权。
   - 如果您要启用DefaultDataWarehouse蓝图，请为 Redshift 管理访问权限角色指定一个新的或现有的服务角色，该角色 DataZone 授权亚马逊获取和管理对 Amazon Redshift 中的数据共享、表和视图的访问权限。
   - 如果您要启用亚马逊 SageMaker蓝图，请为SageMaker 管理访问角色指定一个新的或现有的服务角色，以授予亚马逊向目录发布亚马逊 SageMaker 数据的 DataZone权限。它还授予亚马逊授予访问 DataZone 权限或撤销对亚马逊在目录中 SageMaker 发布的资产的访问权限的权限。

     > **Important**
     >
     > 在您启用亚马逊 SageMaker蓝图时，亚马逊 DataZone 会检查当前账户和地区中是否 DataZone 存在以下 Amazon 的 IAM 角色。如果这些角色不存在，Amazon DataZone 会自动创建它们。
     >
     > - AmazonDataZoneGlueAccess-<region>-<domainId>
     > - AmazonDataZoneRedshiftAccess-<region>-<domainId>

   - 对于配置角色，请指定一个新的或现有的服务角色，该角色 DataZone 授予 Amazon 在环境账户和区域 AWS CloudFormation 中使用创建和配置环境资源的授权。

- 如果您要为 SageMaker-Glu e 数据源的 Amazon S3 存储桶启用亚马逊 SageMaker蓝图，请指定 AWS 账户中所有 SageMaker 环境都要使用的 Amazon S3 存储桶。您指定的存储桶前缀必须为以下项之一：
  - amazon-datazone*
  - datazone-sagemaker*
  - sagemaker-datazone*
  - DataZone-Sagemaker*
  - Sagemaker-* DataZone
  - DataZone-SageMaker*
  - SageMaker-DataZone*

7. 选择启用蓝图。

启用所选蓝图后，可以控制哪些项目可以使用您账户中的蓝图来创建环境配置文件。您可以通过将管理项目分配给蓝图的配置来做到这一点。

> **Important**
>
> 默认情况下，没有为环境蓝图指定管理项目，这意味着任何 Amazon DataZone 用户都可以为环境蓝图创建配置文件。因此，强烈建议您始终为环境蓝图指定管理项目以确保加强治理。

在已启用的蓝图上指定管理项目

1. 前往位于 https://console.aws.amazon.com/datazone 的亚马逊 DataZone 控制台，然后使用您的账户凭证登录。
2. 选择查看域，然后选择要在其中为所选蓝图添加管理项目的域。
3. 选择蓝图选项卡，然后选择要处理的蓝图。
4. 默认情况下，域内的所有项目都可以使用账户中的 DefaultDataLake 或或 DefaultDataWareshouse Amazon SageMaker 蓝图来创建环境配置文件。但是，您可以通过将管理项目分配给蓝图来施加限制。要添加管理项目，请选择选择管理项目，然后从下拉菜单中选择要添加为管理项目的项目，然后选择选择管理项目。

在 AWS 账户中启用 DefaultDataWarehouse 蓝图后，您可以向蓝图配置中添加参数集。参数集是一组键和值，是亚马逊 DataZone 与您的 Amazon Redshift 集群建立连接所必需的，用于创建数据仓库环境。这些参数包括您的 Amazon Redshift 集群的名称、数据库以及保存集群凭证的 AWS 密钥。

向 DefaultDataWarehouse 蓝图添加参数集

1. 前往位于 https://console.aws.amazon.com/datazone 的亚马逊 DataZone 控制台，然后使用您的账户凭证登录。
2. 选择查看域，然后选择包含要在其中添加参数集的域。
3. 选择蓝图选项卡，然后选择 DefaultDataWareshouse 蓝图以打开蓝图详细信息页面。
4. 在蓝图详细信息页面上的参数集选项卡下，选择创建参数集。
   - 提供参数集的名称。
   - （可选）提供参数集的描述。
   - 选择一个区域
   - 选择 Amazon Redshift 集群或 Amazon Redshift Serverless。
   - 选择保存所选 Amazon Redshift 集群或 Amazon Redshift 无服务器工作组凭证的 AWS 秘密 ARN。必须使用 `AmazonDataZoneDomain : [Domain_ID]` 标签标记 AWS 密钥才能在参数集中使用该密钥。
     - 如果您没有现有 AWS 密钥，也可以通过选择 “创建新密钥” 来创建新 AWS 密钥。这将打开一个对话框，可在其中提供密钥的名称、用户名和密码。选择 “创建新 AWS 密钥” 后，Amazon 将在 Secr AWS ets Manager 服务中 DataZone创建一个新密钥，并确保该密钥使用您尝试创建参数集的域进行标记。
   - 如果您在上述步骤中选择了 Amazon Redshift 集群，现在请从下拉列表中选择一个集群。如果您在上述步骤中选择了 Amazon Redshift 工作组，现在请从下拉列表中选择一个工作组。
   - 输入所选 Amazon Redshift 集群或 Amazon Redshift Serverless 工作组中的数据库名称。
   - 选择创建参数集。

> **Note**
>
> 您最多只能向 DefaultDataWarehouse蓝图添加 10 个参数集。

在您的 AWS 账户中启用 Amazon SageMaker 蓝图后，您可以向蓝图配置中添加参数集。参数集是一组键和值，是亚马逊与您的亚马逊 DataZone SageMaker 建立连接所必需的，用于创建 sagemaker 环境。

向 Amazon SageMaker 蓝图添加参数集

1. 前往位于 https://console.aws.amazon.com/datazone 的亚马逊 DataZone 控制台，然后使用您的账户凭证登录。
2. 选择查看域，然后选择包含要在其中添加参数集的已启用蓝图的域。
3. 选择蓝图选项卡，然后选择亚马逊 SageMaker 蓝图以打开蓝图的详细信息页面。
4. 在蓝图详细信息页面上的参数集选项卡下，选择创建参数集，然后指定以下项：

    - 提供参数集的名称。
    - （可选）提供参数集的描述。
    - 指定 Amazon SageMaker 域名身份验证类型。您可以选择 IAM 或 IAM Identity Center（SSO）。
    - 指定 AWS 区域。
    - 为数据加密指定 AWS KMS 密钥。您可以选择现有密钥对或创建新密钥。
    - 在环境参数下，指定以下项：
        - VPC ID-您用于亚马逊 SageMaker 环境的 VPC 的 ID。您可以指定现有 VPC，也可以创建新 VPC。
        - 子网-一个或多个子网 IDs 代表您的 VPC 内特定资源的 IP 地址范围。
        - 网络访问 – 选择仅限 VPC 或仅限公共互联网。
        - 安全组 – 配置 VPC 和子网时使用的安全组。
    - 在“数据来源参数”下，选择下列项之一：
        - AWS 仅限 Glue
        - AWS Glue + Amazon Redshift Serverless。如果您选择此选项，请指定以下项：
            - 指定保存所选 Amazon Redshift 集群凭证的 AWS 秘密 ARN。必须使用 `AmazonDataZoneDomain : [Domain_ID]` 标签标记 AWS 密钥才能在参数集中使用该密钥。

              如果您没有现有 AWS 密钥，也可以通过选择 “创建新密钥” 来创建新 AWS 密钥。这将打开一个对话框，可在其中提供密钥的名称、用户名和密码。选择 “创建新 AWS 密钥” 后，Amazon 将在 Secr AWS ets Manager 服务中 DataZone 创建一个新密钥，并确保该密钥使用您尝试创建参数集的域进行标记。

- 指定要在创建环境时使用的 Amazon Redshift 工作组。
- 指定要在创建环境时使用的数据库（在所选工作组中）的名称。

- AWS 仅限 Glue + 亚马逊 Redshift 集群
  - 指定保存所选 Amazon Redshift 集群凭证的 AWS 秘密 ARN。必须使用 `AmazonDataZoneDomain : [Domain_ID]` 标签标记 AWS 密钥才能在参数集中使用该密钥。

    如果您没有现有 AWS 密钥，也可以通过选择 “创建新密钥” 来创建新 AWS 密钥。这将打开一个对话框，可在其中提供密钥的名称、用户名和密码。选择 “创建新 AWS 密钥” 后，Amazon 将在 Secr AWS ets Manager 服务中 DataZone 创建一个新密钥，并确保该密钥使用您尝试创建参数集的域进行标记。
  - 指定要在创建环境时使用的 Amazon Redshift 集群。
  - 指定要在创建环境时使用的数据库（在所选集群中）的名称。

5. 选择创建参数集。

将亚马逊 SageMaker 作为可信服务添加到拥有亚马逊 DataZone 域名的 AWS 账户中

如果您启用了亚马逊 SageMaker 蓝图，则还必须将其添加 SageMaker 为亚马逊内部的可信服务之一 DataZone。为此，请完成以下过程：

1. 前往位于 https://console.aws.amazon.com/datazone 的亚马逊 DataZone 控制台，然后使用您的账户凭证登录。
2. 选择查看域，然后选择包含已启用 SageMaker 蓝图的域。
3. 选择可信服务，然后选择亚马逊 SageMaker，然后选择启用。

Amazon DataZone 定制 AWS 服务蓝图

在亚马逊中 DataZone，自定义 AWS 服务蓝图允许您通过 DataZone 将亚马逊配置为使用您自己已在组织中设置的现有 AWS 身份和访问管理 (IAM) 角色和 AWS 服务，从而优化资源使用和成本。

用于创建 Amazon DataZone 环境的蓝图定义了该环境所属项目的成员在处理 Amazon DataZone 目录中的资产时可以使用哪些工具和服务。在当前版本的 Amazon 中 DataZone，有以下内置蓝图：

- 数据湖蓝图
- 数据仓库蓝图
- 亚马逊 SageMaker 蓝图

借助 Amazon DataZone 定制 AWS 服务蓝图，您可以创建针对您当前在组织中使用的任何 AWS 服务进行定制的环境和项目。借助自定义蓝图，您可以将 Amazon 纳入现有的数据管道 DataZone 中，方法是将其配置为使用现有的 IAM 角色来增强对基础设施设置的监管，并就业务计划进行协作。

> **Important**
>
> 借助亚马逊 DataZone 定制 AWS 服务打印，您可以将现有的亚马逊 SageMaker 域名迁移到亚马逊 DataZone。借助此功能，管理员现在可以通过从 Amazon SageMaker 域中导入其现有授权用户、安全配置和策略来设置 Amazon DataZone 项目。有关更多信息，请参阅设置 SageMaker 资产（管理员指南）。

主题

- 启用自定义 AWS 服务蓝图
- 使用自定义 AWS 服务蓝图创建环境
- 在自定义 AWS 服务环境中创建操作
- 将项目成员添加到自定义 AWS 服务环境
- 在 AWS 服务环境中配置数据源
- 在 AWS 服务环境中配置订阅目标

启用自定义 AWS 服务蓝图

完成以下步骤以在您的域中启用自定义 AWS 服务蓝图。

1. 登录 AWS 管理控制台并在 https://console.aws.amazon.com/data zone 上打开亚马逊 DataZone 管理控制台。
2. 选择 “查看域”，然后选择要在其中启用自定义 AWS 服务蓝图的域。
3. 选择蓝图选项卡，再从可用蓝图列表中选择 AWS 服务蓝图，然后选择启用。

使用自定义 AWS 服务蓝图创建环境

完成以下过程，使用自定义 AWS 服务蓝图创建环境。

1. 登录 AWS 管理控制台并在 https://console.aws.amazon.com/data zone 上打开亚马逊 DataZone 管理控制台。
2. 选择 “查看域”，然后选择启用自定义 AWS 服务蓝图的域。
3. 选择蓝图选项卡，再选择已启用的 AWS 服务蓝图，然后选择创建环境。
4. 在创建环境页面上，指定以下内容，然后选择创建环境：

   - 名称 – 指定环境的名称。
   - 描述 – 指定环境的描述。
   - 项目 – 为环境指定新的或现有的所属项目。项目使一群用户能够发现、发布、订阅和使用 Amazon 中的资产 DataZone。该环境将可供指定项目的所有成员使用。所有环境都归其用户有权访问环境的项目所有。
   - 环境角色-指定一个现有 IAM 角色，该角色将授予亚马逊在此环境中 DataZone 访问您的现有 AWS 服务和资源（例如 Amazon S3 和 AWS Glue）的权限。

   > **Note**
   >
   > Amazon DataZone 不会为您配置此角色。您必须拥有一个现有 IAM 角色，该角色具有您想要在此环境中启用的现有 AWS 服务和资源的权限。
   > 确保该 IAM 角色具有所需的最低权限，换句话说，缩小范围以仅提供对要在此环境中启用的 AWS 服务和资源的访问权限。
   > 您可以使用 AWS 策略生成器来构建符合您要求的策略，并将其附加到您要使用的自定义 IAM 角色。
   > 确保该角色以 `AmazonDataZone` 开头以便遵循约定。虽然这不是强制性要求，但建议您这样做。如果 IAM 管理员使用的是 `AmazonDataZoneFullAccess` 策略，则必须遵循此约定，因为存在传递角色检查验证。
   > 在创建自定义角色时，请确保它在信任策略中信任 `datazone.amazonaws.com`：

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Principal": {
                "Service": [
                    "datazone.amazonaws.com"
                ]
            },
            "Action": [
                "sts:AssumeRole",
                "sts:TagSession"
            ]
        }
    ]
}
```

- AWS region-指定要在其中创建此环境的 AWS 区域。

在自定义 AWS 服务环境中创建操作

完成以下步骤以在自定义 AWS 服务环境中创建操作。通过在自定义 AWS 服务环境中创建操作，您可以将指向 Amazon DataZone 数据门户的深度链接添加到该环境中可用的分析工具。

1. 登录 AWS 管理控制台并在 https://console.aws.amazon.com/data zone 上打开亚马逊 DataZone 管理控制台。
2. 选择 “查看域”，然后选择启用自定义 AWS 服务蓝图的域。
3. 选择蓝图选项卡，再选择已启用的 AWS 服务蓝图，然后选择要在其中添加操作的 AWS 服务环境。
4. 在 AWS 控制台链接页面上，从 “热门链接” 或 “自定义 AWS 链接” 部分中选择 AWS 链接（操作），以启用通过 DataZone 亚马逊数据门户指向您的 Amazon S3 存储桶、Amazon Athena AWS 工作组、Glue 任务或该环境中 AWS 任何其他自定义控制台资源的深度链接。

5. 如果您使用此环境的摘要部分中的数据门户链接在数据门户中导航到此环境，则可以在分析工具部分下看到您添加的深度链接。

将项目成员添加到自定义 AWS 服务环境

完成以下步骤，将项目成员添加到 AWS 服务环境。

1. 登录 AWS 管理控制台并在 https://console.aws.amazon.com/data zone 上打开亚马逊 DataZone 管理控制台。
2. 选择 “项目” 选项卡，然后在 AWS 服务环境中选择要向其中添加成员的项目。
3. 选择添加，再在添加成员页面上，查找并添加 IAM 用户、SSO 用户或 SSO 组中的成员。将分配的项目角色指定为所有者、贡献者、使用者、管理者或查看者。在查找并添加成员后，选择添加成员。

在 AWS 服务环境中配置数据源

完成以下步骤，在 AWS 服务环境中配置数据源。

1. 登录 AWS 管理控制台并在 https://console.aws.amazon.com/data zone 上打开亚马逊 DataZone 管理控制台。
2. 选择蓝图选项卡，然后选择自定义 AWS 服务蓝图。
3. 在 “已创建的环境” 下，选择要在其中配置数据源的 AWS 服务环境。
4. 选择数据来源选项卡，再选择添加，指定以下内容，然后选择添加。

    - 名称 – 数据来源名称。
    - 资源 ——选择 AWS Glue 或 Amazon Redshift。
        - 对于 AWS Glue，请指定资源数据库。
        - 对于 Amazon Redshift，选择集群或无服务器，然后指定 Redshift 凭证，包括新的或现有的 AWS 密钥、创建环境时要使用的集群或无服务器工作组、创建环境时要使用的数据库以及指定数据库中的架构。
    - 权限-指定一个管理访问角色，该角色将授权亚马逊 DataZone 提取和管理对 La AWS ke Formation 中表的访问权限（适用于 G AWS lue），或者授予亚马逊采集和管理对亚马逊 DataZone Redshift 中表的访问权限的授权。
    - 用于数据消费-在亚马逊中 DataZone，项目成员可以通过订阅目标 DataZone使用数据，亚马逊使用订阅目标来访问您在项目中订阅的数据。指定是否也将此数据来源添加为订阅目标。

在 AWS 服务环境中配置订阅目标

完成以下过程可在 AWS 服务环境中配置订阅目标。

1. 登录 AWS 管理控制台并在 https://console.aws.amazon.com/data zone 上打开亚马逊 DataZone 管理控制台。
2. 选择蓝图选项卡，然后选择 AWS 服务蓝图。
3. 在 “已创建的环境” 下，选择要在其中配置订阅目标的 AWS 服务环境。
4. 选择订阅目标选项卡，再选择添加，指定以下内容，然后选择添加。

   - 名称 – 订阅目标名称。
   - 资源 ——选择 AWS Glue 或 Amazon Redshift。
     - 对于 AWS Glue，请指定资源数据库。
     - 对于 Amazon Redshift，选择集群或无服务器，然后指定 Redshift 凭证，包括新的或现有的 AWS 密钥、创建环境时要使用的集群或无服务器工作组、创建环境时要使用的数据库以及指定数据库中的架构。
   - 权限-指定一个管理访问角色，该角色将授权亚马逊 DataZone 提取和管理对 La AWS ke Formation 中表的访问权限（适用于 G AWS lue），或者授予亚马逊采集和管理对亚马逊 DataZone Redshift 中表的访问权限的授权。
   - 用于数据消费-在 Amazon 中 DataZone，您可以通过允许提取元数据的数据源将数据发布到数据目录中。指定是否也将此订阅目标添加为数据来源。

Amazon 中的关联账户 DataZone

将您的 AWS 账户与您的 Amazon DataZone 域名关联后，域用户就可以发布和使用这些 AWS 账户中的数据。需执行三个步骤来设置账户关联。

- 首先，通过请求关联将域名与所需 AWS 账户共享。如果账户与域名 AWS 账户不同，亚马逊将 DataZone 使用 AWS 资源访问管理器 (RAM)。 AWS 账户关联只能由 Amazon DataZone 域名发起。
- 第二步，让账户所有者接受关联请求。
- 第三步，让账户所有者启用所需的环境蓝图。通过启用蓝图，账户所有者为网域中的用户提供在其账户中创建和访问资源（例如 AWS Glue 数据库和 Amazon Redshift 集群）所需的 IAM 角色和资源配置。

完成以下步骤，将账户与 Amazon 关联 DataZone：

- 步骤 1 – 请求与其他 AWS 账户关联
- 步骤 2 – 接受来自 Amazon DataZone 域的账户关联请求并启用环境蓝图
- 步骤 3 – 在关联 AWS 账户中启用环境蓝图

请求与其他 AWS 账户关联

> **Note**
>
> 通过向其他 AWS 账户发送关联请求，您就是在使用 Resource Acc AWS ess Manager (RAM) 与其他 AWS 账户共享您的域。请务必检查您输入的账户 ID 的准确性。

要在亚马逊 DataZone 控制台中请求与其他 AWS 账户关联亚马逊 DataZone 域名，您必须在该账户中扮演具有管理权限的 IAM 角色。 配置使用亚马逊所需的 IAM 权限 DataZone 管理控制台以获得申请账户关联所需的最低权限。

完成以下步骤以请求与其他 AWS 账户关联。

1. 登录 AWS 管理控制台并在 https://console.aws.amazon.com/data zone 上打开亚马逊 DataZone 管理控制台。
2. 选择查看域，然后从列表中选择域名。该名称是一个超链接。

3. 向下滚动到关联账户选项卡，然后选择请求关联。
4. 输入 IDs 您要申请关联的账户。如果您对账户列表感到满意 IDs，请选择请求关联。
5. 在“RAM 策略”下，指定账户关联的 RAM 策略。您可以选择`AWSRAMPermissionDataZonePortalReadWrite`哪个账户将允许关联账户执行亚马逊 DataZone APIs 并访问数据门户，也可以选择`AWSRAMPermissionDataZoneDefault`，这将仅允许关联账户执行亚马逊 DataZone APIs ，不提供数据门户访问权限。 DataZone 然后，亚马逊代表您的账户在 Resource Access Manager 中创建 AWS 资源共享，并将输入的账户 ID 作为委托人。
6. 您必须通知其他 AWS 账户的所有者接受您的请求。邀请将在七（7）天后过期。

向账户提供对客户自主管理型 KMS 密钥的访问权限

Amazon DataZone 域及其元数据是加密的，要么是（默认情况下）使用由您持有的密钥进行加密 AWS，或者（可选）使用您在域创建期间拥有并提供的 AWS 密钥管理服务 (KMS) 中的客户管理密钥。如果您的域是通过客户自主管理型密钥加密的，请按照以下过程操作，向关联账户授予对 KMS 密钥的使用权限。

1. 登录 AWS 管理控制台并打开 KMS 控制台，网址为https://console.aws.amazon.com/kms/。
2. 要查看您账户中自己所创建和管理的密钥，请在导航窗格中选择 Customer managed keys (客户托管密钥)。
3. 要查看您账户中自己所创建和管理的密钥，请在导航窗格中选择 Customer managed keys (客户托管密钥)。
4. 在 KMS 密钥列表中，选择要检查的 KMS 密钥的别名或密钥 ID。
5. 要允许或禁止外部 AWS 账户使用 KMS 密钥，请使用页面其他 AWS 账户部分中的控件。这些账户中的 IAM 主体（自身具有适当的 KMS 权限）可以在加密操作（如加密、解密、重新加密和生成数据密钥）中使用 KMS 密钥。

接受来自 Amazon DataZone 域的账户关联请求并启用环境蓝图

要在亚马逊 DataZone 管理控制台中接受与亚马逊 DataZone 域的关联，您必须在账户中扮演具有管理权限的 IAM 角色。 配置使用亚马逊所需的 IAM 权限 DataZone 管理控制台以获得最低权限。

完成以下操作以接受与 Amazon DataZone 域名的关联。

1. 登录 AWS 管理控制台并在 https://console.aws.amazon.com/data zone 上打开亚马逊 DataZone 管理控制台。

2. 选择查看请求，然后从列表中选择邀请域。邀请状态应为已请求。选择审核请求。
3. 选择是否启用默认的数据湖 and/or 数据仓库环境蓝图，方法是选中两者都选中或其中一个复选框。您可以稍后执行此操作。
   - 利用数据湖环境蓝图，域用户能够创建和管理 AWS Glue、Amazon S3 和 Amazon Athena 资源，以便从数据湖发布和使用。
   - 利用数据仓库环境蓝图，域用户能够创建和管理 Amazon Redshift 资源，以便从数据仓库发布和使用。
4. 如果您选择一个或两个默认环境蓝图，请配置以下权限和资源。
   - 管理访问权限 IAM 角色向亚马逊提供权限，使域用户 DataZone能够提取和管理对表（例如 G AWS lue 和 Amazon Redshift）的访问权限。您可以选择让 Amazon DataZone 创建和使用新的 IAM 角色，也可以从现有 IAM 角色列表中进行选择。
   - 配置 IAM 角色向 A mazon 提供权限 DataZone，使域用户能够创建和配置环境资源，例如 AWS Glue 数据库。您可以选择让 Amazon DataZone 创建和使用新的 IAM 角色，也可以从现有 IAM 角色列表中进行选择。
   - 用于数据湖的 Amazon S3 存储桶是域用户存储数据湖数据时亚马逊 DataZone 将使用的存储桶或路径。您可以使用亚马逊选择的默认存储桶， DataZone 也可以通过输入其路径字符串来选择自己的现有 Amazon S3 路径。如果您选择自己的 Amazon S3 路径，则需要更新 IAM 策略以向亚马逊 DataZone 提供使用该路径的权限。
5. 如果您对配置感到满意，请选择接受并配置关联。

在关联 AWS 账户中启用环境蓝图

要在 Amazon DataZone 管理控制台中启用环境蓝图，您必须在账户中扮演具有管理权限的 IAM 角色。 配置使用亚马逊所需的 IAM 权限 DataZone 管理控制台以获得最低权限。

完成以下操作可在关联域中启用蓝图。

1. 登录 AWS 管理控制台并在 https://console.aws.amazon.com/data zone 上打开亚马逊 DataZone 管理控制台。
2. 打开左侧导航面板，然后选择关联域。
3. 选择要为其启用环境蓝图的域。
4. 从蓝图列表中，选择DefaultDataLake或DefaultDataWarehouse SageMaker、Amazon 或定制 AWS 服务蓝图。

> **Note**
>
> 如果您启用自定义 AWS 服务蓝图，则无需指定管理访问角色。自定义 AWS 服务 bluerpint 的权限和授权机制是在您使用此蓝图创建环境时处理的。有关更多信息，请参阅使用自定义 AWS 服务蓝图创建环境。

5. 在所选蓝图的详细信息页面上，选择在此账户中启用。
6. 在权限和资源页面上，指定以下角色：

- 如果您要启用DefaultDataLake蓝图，请为 Glue 管理访问权限角色指定一个新的或现有的服务角色，该角色 DataZone 授予亚马逊收录和管理对 G AWS lue 和 La AWS ke Formation 中表的访问权限的授权。
- 如果您要启用DefaultDataWarehouse蓝图，请为 Redshift 管理访问权限角色指定一个新的或现有的服务角色，该角色 DataZone 授权亚马逊获取和管理对 Amazon Redshift 中的数据共享、表和视图的访问权限。
- 如果您要启用亚马逊 SageMaker蓝图，请为SageMaker 管理访问角色指定一个新的或现有的服务角色，以授予亚马逊向目录发布亚马逊 SageMaker 数据的 DataZone 权限。它还授予亚马逊授予访问 DataZone 权限或撤销对亚马逊在目录中 SageMaker 发布的资产的访问权限的权限。

> **Important**
>
> 在您启用亚马逊 SageMaker蓝图时，亚马逊 DataZone 会检查当前账户和地区中是否 DataZone 存在以下 Amazon 的 IAM 角色。如果这些角色不存在，Amazon DataZone 会自动创建它们。
>
> - AmazonDataZoneGlueAccess-<region>-<domainId>
> - AmazonDataZoneRedshiftAccess-<region>-<domainId>

- 对于配置角色，请指定一个新的或现有的服务角色，该角色 DataZone 授予 Amazon 在环境账户和区域 AWS CloudFormation 中使用创建和配置环境资源的授权。
- 如果您要为 SageMaker-Glu e 数据源的 Amazon S3 存储桶启用亚马逊 SageMaker蓝图，请指定 AWS 账户中所有 SageMaker 环境都要使用的 Amazon S3 存储桶。您指定的存储桶前缀必须为以下项之一：
  - amazon-datazone*
  - datazone-sagemaker*
  - sagemaker-datazone*

- DataZone-Sagemaker*
- Sagemaker-* DataZone
- DataZone-SageMaker*
- SageMaker-DataZone*

7. 选择启用蓝图。

启用所选蓝图后，可以控制哪些项目可以使用您账户中的蓝图来创建环境配置文件。您可以通过将管理项目分配给蓝图的配置来做到这一点。

指定在已启用 DefaultDataLake 或 DefaultDataWarehouse 蓝图上管理项目

1. 前往位于 https://console.aws.amazon.com/datazone 的亚马逊 DataZone 控制台，然后使用您的账户凭证登录。
2. 打开左侧导航面板并选择关联域，然后选择要在其中添加管理项目的域。
3. 选择蓝图选项卡，然后选择 DefaultDataLake 或 DefaultDataWareshouse 蓝图。
4. 默认情况下，域内的所有项目都可以使用账户中的 DefaultDataLake 或 DefaultDataWareshouse 蓝图来创建环境配置文件。但是，您可以通过将管理项目分配给蓝图来施加限制。要添加管理项目，请选择选择管理项目，然后从下拉菜单中选择要添加为管理项目的项目，然后选择选择管理项目。

在 AWS 账户中启用 DefaultDataWarehouse 蓝图后，您可以向蓝图配置中添加参数集。参数集是一组键和值，是亚马逊 DataZone 与您的 Amazon Redshift 集群建立连接所必需的，用于创建数据仓库环境。这些参数包括您的 Amazon Redshift 集群的名称、数据库以及保存集群凭证的 AWS 密钥。

> **Important**
>
> 默认情况下，没有为环境蓝图指定管理项目，这意味着任何 Amazon DataZone 用户都可以为环境蓝图创建配置文件。因此，强烈建议您始终为环境蓝图指定管理项目以确保加强治理。

向 DefaultDataWarehouse 蓝图添加参数集

1. 前往位于 https://console.aws.amazon.com/datazone 的亚马逊 DataZone 控制台，然后使用您的账户凭证登录。
2. 打开左侧导航面板并选择关联域，然后选择要在其中添加参数集的域。

3. 选择蓝图选项卡，然后选择 DefaultDataWareshouse 蓝图以打开蓝图详细信息页面。
4. 在蓝图详细信息页面上的参数集选项卡下，选择创建参数集。
   - 提供参数集的名称。
   - （可选）提供参数集的描述。
   - 选择一个区域
   - 选择 Amazon Redshift 集群或 Amazon Redshift Serverless。
   - 选择保存所选 Amazon Redshift 集群或 Amazon Redshift 无服务器工作组凭证的 AWS 秘密 ARN。 AWS 密钥必须用`AmazonDataZoneDomain : [Domain_ID]`标签进行标记，才有资格在参数集中使用。
     - 如果您没有现有 AWS 密钥，也可以通过选择 “创建新密钥” 来创建新 AWS 密钥。这将打开一个对话框，可在其中提供密钥的名称、用户名和密码。选择 “创建新 AWS 密钥” 后，Amazon 将在 Secr AWS ets Manager 服务中 DataZone 创建一个新密钥，并确保该密钥使用您尝试创建参数集的域进行标记。
   - 选择 Amazon Redshift 集群或 Amazon Redshift Serverless 工作组。
   - 输入所选 Amazon Redshift 集群或 Amazon Redshift Serverless 工作组中的数据库名称。
   - 选择创建参数集。

> **Note**
>
> 您最多只能向 DefaultDataWarehouse 蓝图添加 10 个参数集。

在您的 AWS 账户中启用 Amazon SageMaker 蓝图后，您可以向蓝图配置中添加参数集。参数集是一组键和值，是亚马逊与您的亚马逊 DataZone SageMaker 建立连接所必需的，用于创建 sagemaker 环境。

向 Amazon SageMaker 蓝图添加参数集

1. 前往位于 https://console.aws.amazon.com/datazone 的亚马逊 DataZone 控制台，然后使用您的账户凭证登录。
2. 选择查看域，然后选择包含要在其中添加参数集的已启用蓝图的域。
3. 选择蓝图选项卡，然后选择 Amazon SageMaker 蓝图以打开蓝图的详细信息页面。
4. 在蓝图详细信息页面上的参数集选项卡下，选择创建参数集，然后指定以下项：

- 提供参数集的名称。
- （可选）提供参数集的描述。
- 指定 Amazon SageMaker 域名身份验证类型。您可以选择 IAM 或 IAM Identity Center（SSO）。
- 指定 AWS 区域。
- 为数据加密指定 AWS KMS 密钥。您可以选择现有密钥对或创建新密钥。
- 在环境参数下，指定以下项：
  - VPC ID-您用于亚马逊 SageMaker环境的 VPC 的 ID。您可以指定现有 VPC，也可以创建新 VPC。
  - 子网-一个或多个子网 IDs 代表您的 VPC 内特定资源的 IP 地址范围。
  - 网络访问 – 选择仅限 VPC 或仅限公共互联网。
  - 安全组 – 配置 VPC 和子网时使用的安全组。
- 在“数据来源参数”下，选择下列项之一：
  - AWS 仅限 Glue
  - AWS Glue + Amazon Redshift Serverless。如果您选择此选项，请指定以下项：
    - 指定保存所选 Amazon Redshift 集群凭证的 AWS 秘密 ARN。 AWS 密钥必须用`AmazonDataZoneDomain : [Domain_ID]`标签进行标记，才有资格在参数集中使用。

      如果您没有现有 AWS 密钥，也可以通过选择 “创建新密钥” 来创建新 AWS 密钥。这将打开一个对话框，可在其中提供密钥的名称、用户名和密码。选择 “创建新 AWS 密钥” 后，Amazon 将在 Secr AWS ets Manager 服务中 DataZone 创建一个新密钥，并确保该密钥使用您尝试创建参数集的域进行标记。
    - 指定要在创建环境时使用的 Amazon Redshift 工作组。
    - 指定要在创建环境时使用的数据库（在所选工作组中）的名称。
  - AWS 仅限 Glue + 亚马逊 Redshift 集群
    - 指定保存所选 Amazon Redshift 集群凭证的 AWS 秘密 ARN。 AWS 密钥必须用`AmazonDataZoneDomain : [Domain_ID]`标签进行标记，才有资格在参数集中使用。

      如果您没有现有 AWS 密钥，也可以通过选择 “创建新密钥” 来创建新 AWS 密钥。这将打开一个对话框，可在其中提供密钥的名称、用户名和密码。选择 “创建新 AWS 密钥”

后，Amazon 将在 Secr AWS ets Manager 服务中 DataZone 创建一个新密钥，并确保该密钥使用您尝试创建参数集的域进行标记。

- 指定要在创建环境时使用的 Amazon Redshift 集群。
- 指定要在创建环境时使用的数据库（在所选集群中）的名称。

5. 选择创建参数集。

在关联 AWS 账户中 SageMaker 将 Amazon 添加为可信服务

如果您启用了亚马逊 SageMaker 蓝图，则还必须将其添加 SageMaker 为亚马逊内部的可信服务之一 DataZone。为此，请完成以下过程：

1. 前往位于 https://console.aws.amazon.com/datazone 的亚马逊 DataZone 控制台，然后使用您的账户凭证登录。
2. 选择查看域，然后选择包含已启用 SageMaker 蓝图的域。
3. 选择“可信服务”，然后选择 Amazon SageMaker，然后选择“启用”。

拒绝来自亚马逊 DataZone 域名的账户关联请求

要在亚马逊 DataZone 管理控制台中拒绝来自亚马逊 DataZone域的关联请求，您必须在该账户中扮演具有管理权限的 IAM 角色。 配置使用亚马逊所需的 IAM 权限 DataZone 管理控制台以获得最低权限。

完成以下操作，拒绝来自亚马逊 DataZone 域的关联请求。

1. 登录 AWS 管理控制台并在 https://console.aws.amazon.com/data zone 上打开亚马逊 DataZone 管理控制台。
2. 选择查看请求，然后从列表中选择邀请域。邀请状态应为已请求。选择拒绝关联。通过选择拒绝关联来确认您的选择。

在 Amazon 中移除关联账户 DataZone

要在 Amazon DataZone 管理控制台中删除关联 AWS 账户，您必须在该账户中扮演具有管理权限的 IAM 角色。 配置使用亚马逊所需的 IAM 权限 DataZone 管理控制台以获得最低权限。

完成以下过程可从域中删除关联账户。

1. 登录 AWS 管理控制台并在 https://console.aws.amazon.com/data zone 上打开亚马逊 DataZone 管理控制台。
2. 选择查看域，然后从列表中选择域名。该名称是一个超链接。
3. 向下滚动至关联账户选项卡。为要删除的 AWS 账户选择账户 ID。
4. 选择取消关联。在字段中输入“disassociate”并选择取消关联来确认您的选择。
5. 现已从域中删除此账户，并且域的用户无法使用此账户发布和使用数据。

亚马逊 DataZone 数据目录

您可以使用 Amazon B DataZone usiness 数据目录根据业务背景对整个组织的数据进行分类，从而使组织中的每个人都能快速查找和理解数据。

要使用亚马逊对您的数据 DataZone 进行分类，您必须先将您的数据（资产）作为项目库存带到亚马逊 DataZone。为项目创建库存，从而仅允许该项目的成员发现资产。 search/browse 除非明确发布，否则并非所有域用户都可以使用项目清单资产。

创建项目库存后，数据所有者可以添加或更新业务名称（资产和架构）、描述（资产和架构）、自述文件、术语表术语（资产和架构）和元数据表单，从而使用所需的业务元数据来整理库存资产。

使用 Amazon DataZone 对您的数据进行分类的下一步是让域名用户可以发现您项目的库存资产。您可以通过将库存资产发布到 Amazon DataZone 目录来做到这一点。只能将最新版本的库存资产发布到目录，并且仅最新发布的版本在发现目录中处于活动状态。如果库存资产在发布到亚马逊 DataZone 目录后进行了更新，则必须再次明确发布该库存资产，以使最新版本出现在发现目录中。

有关更多信息，请参阅 亚马逊 DataZone 术语和概念。

主题

- 在 Amazon 中创建业务词汇表 DataZone
- 在 Amazon 中编辑企业词汇表 DataZone
- 在 Amazon 中删除企业词汇表 DataZone
- 在 Amazon 的词汇表中创建术语 DataZone
- 在 Amazon 的词汇表中编辑术语 DataZone
- 删除 Amazon 词汇表中的术语 DataZone
- 在 Amazon 中创建元数据表单 DataZone
- 在 Amazon 中编辑元数据表单 DataZone
- 在 Amazon 中删除元数据表单 DataZone
- 在 Amazon 的元数据表单中创建字段 DataZone
- 在 Amazon 中编辑元数据表单中的字段 DataZone
- 在 Amazon 中删除元数据表单中的字段 DataZone

在 Amazon 中创建业务词汇表 DataZone

在亚马逊中 DataZone，业务词汇表是可能与资产（数据）相关的商业术语（单词）的集合。它为业务用户提供了相应的词汇表以及商业术语及其定义的列表，从而确保在分析数据时，整个组织内使用相同的定义。业务术语表是在目录域中创建的，可应用于资产和列以帮助理解该资产或列的关键特征。可以应用一个或多个术语表术语。业务术语表可以是术语的平面列表，业务术语表中的任何术语都可与其他术语的子列表相关联。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。要在您的 Amazon DataZone 域中创建、编辑或删除词汇表，您必须是拥有该域名的相应权限的拥有项目的成员。

要创建术语表，请完成以下步骤：

1. 使用 DataZone 数据门户 URL 导航至 Amazon 数据门户，然后使用您的 SSO 或 AWS 凭证登录。如果您是亚马逊 DataZone 管理员，则可以通过访问创建亚马逊 DataZone 域名的 AWS 账户中的 https://console.aws.amazon.com /datazone 上的亚马逊 DataZone 控制台来获取数据门户 URL。
2. 在顶部导航栏中，浏览找到搜索旁边的目录菜单。
3. 在 Amazon DataZone 数据门户中，选择 “词汇表”，然后选择 “创建词汇表”。
4. 指定术语表的名称、描述和所有者，然后选择创建词汇表。
5. 通过选择已启用开关来启用新术语表。
6. 在术语表的详细信息页面上，您可以选择创建自述文件来添加有关此术语表的一些其他信息。

要禁用或启用业务术语表，请完成以下步骤：

1. 使用 DataZone 数据门户 URL 导航至 Amazon 数据门户，然后使用您的 SSO 或 AWS 凭证登录。如果您是亚马逊 DataZone 管理员，则可以通过访问创建亚马逊 DataZone 域名的 AWS 账户中的 https://console.aws.amazon.com /datazone 上的亚马逊 DataZone 控制台来获取数据门户 URL。
2. 在顶部导航栏中，浏览找到搜索旁边的目录菜单。
3. 在 Amazon DataZone 数据门户中，选择术语表，然后找到要禁用/启用的业务词汇表。
4. 在术语表详细信息页面上，找到启用/禁用开关，然后使用它来启用或禁用所选术语表。

> **Note**
>
> 禁用某个术语表也会禁用该术语表中的所有术语。

在 Amazon 中编辑企业词汇表 DataZone

在亚马逊中 DataZone，业务词汇表是可能与资产（数据）相关的商业术语（单词）的集合。它为业务用户提供了相应的词汇表以及商业术语及其定义的列表，从而确保在分析数据时，整个组织内使用相同的定义。业务术语表是在目录域中创建的，可应用于资产和列以帮助理解该资产或列的关键特征。可以应用一个或多个术语表术语。业务术语表可以是术语的平面列表，业务术语表中的任何术语都可与其他术语的子列表相关联。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。要编辑您的 Amazon DataZone 域中的词汇表，您必须是拥有该域名的相应权限的拥有项目的成员。

要编辑业务术语表，请完成以下步骤：

1. 使用 DataZone 数据门户 URL 导航至 Amazon 数据门户，然后使用您的 SSO 或 AWS 凭证登录。如果您是亚马逊 DataZone 管理员，则可以通过访问创建亚马逊 DataZone 域名的 AWS 账户中的 https://console.aws.amazon.com /datazone 上的亚马逊 DataZone 控制台来获取数据门户 URL。
2. 在顶部导航栏中，浏览找到搜索旁边的目录菜单。
3. 在 Amazon DataZone 数据门户中，选择 “术语表”，然后找到要编辑的业务词汇表。
4. 在术语表详细信息页面上，展开操作，然后选择编辑以编辑术语表。
5. 对名称和描述进行更新，然后选择保存。

在 Amazon 中删除企业词汇表 DataZone

在亚马逊中 DataZone，业务词汇表是可能与资产（数据）相关的商业术语（单词）的集合。它为业务用户提供了相应的词汇表以及商业术语及其定义的列表，从而确保在分析数据时，整个组织内使用相同的定义。业务术语表是在目录域中创建的，可应用于资产和列以帮助理解该资产或列的关键特征。可以应用一个或多个术语表术语。业务术语表可以是术语的平面列表，业务术语表中的任何术语都可与其他术语的子列表相关联。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。要删除您的 Amazon DataZone 域中的词汇表，您必须是拥有该域名的相应权限的拥有项目的成员。

要删除业务术语表，请完成以下步骤：

1. 使用 DataZone 数据门户 URL 导航至 Amazon 数据门户，然后使用您的 SSO 或 AWS 凭证登录。如果您是亚马逊 DataZone 管理员，则可以通过访问创建亚马逊 DataZone 域名的 AWS 账户中的 https://console.aws.amazon.com /datazone 上的亚马逊 DataZone 控制台来获取数据门户 URL。
2. 在顶部导航栏中，浏览找到搜索旁边的目录菜单。

3. 在 Amazon DataZone 数据门户中，选择 “术语表”，然后找到要删除的业务词汇表。
4. 在术语表详细信息页面上，展开操作，然后选择删除以删除术语表。

> **Note**
>
> 您必须先删除术语表中的所有现有术语，然后才能删除术语表。

5. 通过选择删除来确认删除术语表。

在 Amazon 的词汇表中创建术语 DataZone

在亚马逊中 DataZone，业务词汇表是可能与资产（数据）相关的商业术语的集合。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。要创建、编辑或删除您的 Amazon DataZone 域名词汇表中的术语，您必须是拥有该域名的相应权限的拥有项目的成员。

在 Amazon 中 DataZone，商业词汇表术语可以有近似的描述。要设置特定术语的上下文，可以指定术语之间的关系。在定义术语的关系后，该关系会自动添加到相关术语的定义中。Amazon 中提供的词汇表术语关系 DataZone 包括以下内容：

- 是一种类型 – 表示当前术语是一类已确定的术语。表示已确定的术语是当前术语的父术语。
- 有类型 – 表示当前术语是指示的一个或多个特定术语的通用术语。此关系可以表示通用术语的子术语。

要创建新术语，请完成以下步骤：

1. 使用 DataZone 数据门户 URL 导航至 Amazon 数据门户，然后使用您的 SSO 或 AWS 凭证登录。如果您是亚马逊 DataZone 管理员，则可以通过访问创建亚马逊 DataZone 域名的 AWS 账户中的 https://console.aws.amazon.com /datazone 上的亚马逊 DataZone 控制台来获取数据门户 URL。
2. 在顶部导航栏中，浏览找到搜索旁边的目录菜单。
3. 在 Amazon DataZone 数据门户中，选择 “词汇表”，然后选择要在其中创建新术语的词汇表。
4. 指定术语的名称、描述和所有者，然后选择创建术语。
5. 通过选择已启用开关来启用新术语。
6. 要添加自述文件，请导航到术语详细信息页面，然后可以选择创建自述文件来添加有关此术语表的一些其他信息。

7. 要添加关系，请导航到术语详细信息页面，选择术语关系部分，然后选择添加术语表术语。在对话框中，选择要关联的关系和术语，然后选择关闭将术语添加到相应的关系类型中。此关系还将添加到已关联的所有术语中。

在 Amazon 的词汇表中编辑术语 DataZone

在亚马逊中 DataZone，业务词汇表是可能与资产（数据）相关的商业术语的集合。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。要创建、编辑或删除您的 Amazon DataZone 域名词汇表中的术语，您必须是拥有该域名的相应权限的拥有项目的成员。

在 Amazon 中 DataZone，商业词汇表术语可以有近似的描述。要设置特定术语的上下文，可以指定术语之间的关系。在定义术语的关系后，该关系会自动添加到相关术语的定义中。Amazon 中提供的词汇表术语关系 DataZone 包括以下内容：

- 是一种类型 – 表示当前术语是一类已确定的术语。表示已确定的术语是当前术语的父术语。
- 有类型 – 表示当前术语是指示的一个或多个特定术语的通用术语。此关系可以表示通用术语的子术语。

要编辑术语表术语，请完成以下步骤：

1. 使用 DataZone 数据门户 URL 导航至 Amazon 数据门户，然后使用您的 SSO 或 AWS 凭证登录。如果您是亚马逊 DataZone 管理员，则可以通过访问创建亚马逊 DataZone 域名的 AWS 账户中的 https://console.aws.amazon.com /datazone 上的亚马逊 DataZone 控制台来获取数据门户 URL。
2. 在顶部导航栏中，浏览找到搜索旁边的目录菜单。
3. 在 Amazon DataZone 数据门户中，选择 “词汇表”，找到包含您要编辑的术语的词汇表，然后选择该术语。
4. 在术语详细信息页面上，展开操作，然后选择编辑以编辑术语。
5. 对名称和描述进行更新，然后选择保存。

删除 Amazon 词汇表中的术语 DataZone

在亚马逊中 DataZone，业务词汇表是可能与资产（数据）相关的商业术语的集合。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。要创建、编辑或删除您的 Amazon DataZone 域名词汇表中的术语，您必须是拥有该域名的相应权限的拥有项目的成员。

在 Amazon 中 DataZone，商业词汇表术语可以有近似的描述。要设置特定术语的上下文，可以指定术语之间的关系。在定义术语的关系后，该关系会自动添加到相关术语的定义中。Amazon 中提供的词汇表术语关系 DataZone 包括以下内容：

- 是一种类型 – 表示当前术语是一类已确定的术语。表示已确定的术语是当前术语的父术语。
- 有类型 – 表示当前术语是指示的一个或多个特定术语的通用术语。此关系可以表示通用术语的子术语。

要删除术语表术语，请完成以下步骤：

1. 使用 DataZone 数据门户 URL 导航至 Amazon 数据门户，然后使用您的 SSO 或 AWS 凭证登录。如果您是亚马逊 DataZone 管理员，则可以通过访问创建亚马逊 DataZone 域名的 AWS 账户中的 https://console.aws.amazon.com /datazone 上的亚马逊 DataZone 控制台来获取数据门户 URL。
2. 在顶部导航栏中，浏览找到搜索旁边的目录菜单。
3. 在亚马逊 DataZone 数据门户中，选择词汇表，找到包含您要删除的术语的词汇表，然后选择该术语。
4. 在术语表详细信息页面上，展开操作，然后选择删除以删除术语。
5. 通过选择删除来确认删除术语。

在 Amazon 中创建元数据表单 DataZone

在 Amazon 中 DataZone，元数据表单是一种简单的表单，用于为目录中的资产元数据补充额外的业务背景。它充当面向数据所有者的可扩展机制，用于为资产补充信息，从而帮助数据用户搜索和查找相关数据。元数据表单还可以作为一种机制，确保发布到 Amazon DataZone 目录的所有资产保持一致。

元数据表单定义包含一个或多个字段定义，并支持布尔值、日期、小数、整数、字符串和业务术语表字段值数据类型。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。要在您的 Amazon DataZone 域中创建、编辑或删除元数据表单，您必须是拥有相应证书的拥有项目的成员。

要创建元数据表单，请完成以下步骤：

1. 使用 DataZone 数据门户 URL 导航至 Amazon 数据门户，然后使用您的 SSO 或 AWS 凭证登录。如果您是亚马逊 DataZone 管理员，则可以通过访问创建亚马逊 DataZone 域名的 AWS 账户中的 https://console.aws.amazon.com /datazone 上的亚马逊 DataZone 控制台来获取数据门户 URL。

2. 在顶部导航栏中，浏览找到搜索旁边的目录菜单。
3. 在 Amazon DataZone 数据门户中，选择元数据表单，然后选择创建表单。
4. 指定元数据表单的名称、描述和所有者，然后选择创建表单。

在 Amazon 中编辑元数据表单 DataZone

在 Amazon 中 DataZone，元数据表单是一种简单的表单，用于为目录中的资产元数据补充额外的业务背景。它充当面向数据所有者的可扩展机制，用于为资产补充信息，从而帮助数据用户搜索和查找相关数据。元数据表单还可以作为一种机制，确保发布到 Amazon DataZone 目录的所有资产保持一致。

元数据表单定义包含一个或多个字段定义，并支持布尔值、日期、小数、整数、字符串和业务术语表字段值数据类型。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。要在您的 Amazon DataZone 域中创建、编辑或删除元数据表单，您必须是拥有相应证书的拥有项目的成员。

要编辑元数据表单，请完成以下步骤：

1. 使用 DataZone 数据门户 URL 导航至 Amazon 数据门户，然后使用您的 SSO 或 AWS 凭证登录。如果您是亚马逊 DataZone 管理员，则可以通过访问创建亚马逊 DataZone 域名的 AWS 账户中的 https://console.aws.amazon.com /datazone 上的亚马逊 DataZone 控制台来获取数据门户 URL。
2. 在顶部导航栏中，浏览找到搜索旁边的目录菜单。
3. 在 Amazon DataZone 数据门户中，选择元数据表单，然后找到要编辑的元数据表单。
4. 在元数据表单的详细信息页面上，展开操作，然后选择编辑。
5. 对名称、描述和所有者字段进行更新，然后选择更新表单。

在 Amazon 中删除元数据表单 DataZone

在 Amazon 中 DataZone，元数据表单是一种简单的表单，用于为目录中的资产元数据补充额外的业务背景。它充当面向数据所有者的可扩展机制，用于为资产补充信息，从而帮助数据用户搜索和查找相关数据。元数据表单还可以作为一种机制，确保发布到 Amazon DataZone 目录的所有资产保持一致。

元数据表单定义包含一个或多个字段定义，并支持布尔值、日期、小数、整数、字符串和业务术语表字段值数据类型。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。要在您的 Amazon DataZone 域中创建、编辑或删除元数据表单，您必须是拥有相应证书的拥有项目的成员。

要删除元数据表单，请完成以下步骤：

> **Note**
>
> 您必须先从应用元数据表单的所有资产类型或资产中移除该表单，然后才能删除该表单。

1. 使用 DataZone 数据门户 URL 导航至 Amazon 数据门户，然后使用您的 SSO 或 AWS 凭证登录。如果您是亚马逊 DataZone 管理员，则可以通过访问创建亚马逊 DataZone 域名的 AWS 账户中的 https://console.aws.amazon.com /datazone 上的亚马逊 DataZone 控制台来获取数据门户 URL。
2. 在顶部导航栏中，浏览找到搜索旁边的目录菜单。
3. 在 Amazon DataZone 数据门户中，选择元数据表单，然后找到要删除的元数据表单。
4. 如果要删除的元数据表单已启用，请通过选择已启用开关来禁用该元数据表单。
5. 在元数据表单的详细信息页面上，展开操作，然后选择删除。
6. 通过选择删除来确认删除。

在 Amazon 的元数据表单中创建字段 DataZone

在 Amazon 中 DataZone，元数据表单是一种简单的表单，用于为目录中的资产元数据补充额外的业务背景。它充当面向数据所有者的可扩展机制，用于为资产补充信息，从而帮助数据用户搜索和查找相关数据。元数据表单还可以作为一种机制，确保发布到 Amazon DataZone 目录的所有资产保持一致。

元数据表单定义包含一个或多个字段定义，并支持布尔值、日期、小数、整数、字符串和业务术语表字段值数据类型。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。要在您的 Amazon DataZone 域中创建、编辑或删除元数据表单中的字段，您必须是拥有相应证书的拥有项目的成员。

要创建元数据表单字段，请完成以下步骤：

1. 使用 DataZone 数据门户 URL 导航至 Amazon 数据门户，然后使用您的 SSO 或 AWS 凭证登录。如果您是亚马逊 DataZone 管理员，则可以通过访问创建亚马逊 DataZone 域名的 AWS 账户中的 https://console.aws.amazon.com /datazone 上的亚马逊 DataZone 控制台来获取数据门户 URL。
2. 在顶部导航栏中，浏览找到搜索旁边的目录菜单。
3. 在 Amazon DataZone 数据门户中，选择元数据表单，然后选择要在其中创建字段的元数据表单。
4. 在表单的详细信息页面上，选择创建字段。
5. 指定字段名称、描述、类型以及是否为必填字段，然后选择创建字段。

在 Amazon 中编辑元数据表单中的字段 DataZone

在 Amazon 中 DataZone，元数据表单是一种简单的表单，用于为目录中的资产元数据补充额外的业务背景。它充当面向数据所有者的可扩展机制，用于为资产补充信息，从而帮助数据用户搜索和查找相关数据。元数据表单还可以作为一种机制，确保发布到 Amazon DataZone 目录的所有资产保持一致。

元数据表单定义包含一个或多个字段定义，并支持布尔值、日期、小数、整数、字符串和业务术语表字段值数据类型。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。要在您的 Amazon DataZone 域中创建、编辑或删除元数据表单中的字段，您必须是拥有相应证书的拥有项目的成员。

要编辑元数据表单字段，请完成以下步骤：

1. 使用 DataZone 数据门户 URL 导航至 Amazon 数据门户，然后使用您的 SSO 或 AWS 凭证登录。如果您是亚马逊 DataZone 管理员，则可以通过访问创建亚马逊 DataZone 域名的 AWS 账户中的 https://console.aws.amazon.com /datazone 上的亚马逊 DataZone 控制台来获取数据门户 URL。
2. 在顶部导航栏中，浏览找到搜索旁边的目录菜单。
3. 在 Amazon DataZone 数据门户中，选择元数据表单，然后选择要在其中编辑字段的元数据表单。
4. 在表单的详细信息页面上，选择要编辑的字段，展开操作，然后选择编辑。
5. 对字段名称、描述、类型以及是否为必填字段进行更新，然后选择更新字段。

在 Amazon 中删除元数据表单中的字段 DataZone

在 Amazon 中 DataZone，元数据表单是一种简单的表单，用于为目录中的资产元数据补充额外的业务背景。它充当面向数据所有者的可扩展机制，用于为资产补充信息，从而帮助数据用户搜索和查找相关数据。元数据表单还可以作为一种机制，确保发布到 Amazon DataZone 目录的所有资产保持一致。

元数据表单定义包含一个或多个字段定义，并支持布尔值、日期、小数、整数、字符串和业务术语表字段值数据类型。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。要在您的 Amazon DataZone 域中创建、编辑或删除元数据表单中的字段，您必须是拥有相应证书的拥有项目的成员。

要删除元数据表单字段，请完成以下步骤：

1. 使用 DataZone 数据门户 URL 导航至 Amazon 数据门户，然后使用您的 SSO 或 AWS 凭证登录。如果您是亚马逊 DataZone 管理员，则可以通过访问创建亚马逊 DataZone 域名的 AWS 账户中的 https://console.aws.amazon.com /datazone 上的亚马逊 DataZone 控制台来获取数据门户 URL。

2. 在顶部导航栏中，浏览找到搜索旁边的目录菜单。
3. 在 Amazon DataZone 数据门户中，选择元数据表单，然后选择要删除字段的元数据表单。
4. 在表单的详细信息页面上，选择要删除的字段，展开操作，然后选择删除。
5. 通过选择删除来确认删除。

Amazon DataZone 项目和环境

在亚马逊中 DataZone，项目使一组用户能够就各种业务用例进行协作，这些用例涉及发布、发现、订阅和使用亚马逊 DataZone 目录中的数据资产。每个 Amazon DataZone 项目都应用了一组访问控制，因此只有获得授权的个人、团体和角色才能访问该项目和该项目订阅的数据资产，并且只能使用由项目权限定义的工具。项目充当身份主体，接收对底层资源的访问授权，从而使 Amazon DataZone 能够在组织的基础设施内运营，而无需依赖个人用户的证书。

在 Amazon 中 DataZone，环境是已配置资源的集合（例如，Amazon S3 存储桶、 AWS Glue 数据库或 Amazon Athena 工作组），具有一组给定的 IAM 委托人（具有分配的贡献者权限），他们可以对这些资源进行操作。每个环境还可具有用户主体，他们有权访问资源并能通过订阅和履行来访问数据。环境旨在存储指向 AWS 服务、外部 IDE 和控制台的可操作链接。项目成员可以通过环境中配置的深度链接访问 Amazon Athena 控制台等服务。可以进一步将项目中的 SSO 用户和 IAM 用户范围缩小到 use/access 特定的环境。

在 Amazon 中 DataZone，您可以使用名为环境配置文件的模板来创建环境。反过来，环境配置文件是通过使用内置和自定义 AWS 服务蓝图创建的。利用环境配置文件，域管理员可以用预先配置的参数封装蓝图，之后数据工作人员可以通过选择现有环境配置文件并指定新环境的名称来快速创建任意数量的新环境。这使数据工作人员能够高效地管理其项目和环境，同时确保他们符合域管理员实施的数据治理策略。

有关更多信息，请参阅 亚马逊 DataZone 术语和概念。

主题

- 创建环境配置文件
- 编辑环境配置文件
- 删除环境配置文件
- 创建新环境
- 编辑环境
- 删除环境
- 创建新项目
- 编辑项目
- 将项目移动到其他域单元
- 删除项目
- 离开项目

- 向项目添加成员
- 从项目中移除成员

创建环境配置文件

在 Amazon 中 DataZone，环境配置文件是您可以用来创建环境的模板。环境配置文件的目的是通过在配置文件中嵌入放置信息（例如 AWS 账户和区域）来简化环境创建。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。要在亚马逊 DataZone域中创建环境配置文件，您必须属于亚马逊 DataZone 项目。所有环境配置文件归项目所有，任何项目中的所有授权用户均可使用这些配置文件来创建新环境。

创建环境配置文件

1. 使用 DataZone 数据门户 URL 导航至 Amazon 数据门户，然后使用您的 SSO 或 AWS 凭证登录。如果您是亚马逊 DataZone 管理员，则可以通过访问创建亚马逊 DataZone 域名的 AWS 账户https://console.aws.amazon.com/datazone中的亚马逊 DataZone 控制台来获取数据门户 URL。
2. 在数据门户中，选择浏览项目，然后选择要在其中创建环境配置文件的项目。
3. 导航到项目中的环境选项卡，然后选择创建环境配置文件。
4. 配置以下字段：

   - 名称 – 环境配置文件的名称。
   - 描述 –（可选）环境配置文件的描述。
   - 所有者项目 – 此字段中默认选择将在其中创建配置文件的项目。
   - 蓝图 – 为其创建此配置文件的蓝图。您可以选择一个默认的 Amazon DataZone 蓝图（数据湖或数据仓库）。

     如果已指定数据仓库蓝图，请执行以下操作：

     - 提供一个参数集。要选择现有参数集，请选择选择参数集选项。如果要输入自己的参数，请选择输入自己的。
     - 如果您选择现有参数，请执行以下操作：
       - 从下拉列表中选择一个 AWS 账户。
       - 从下拉列表中选择一个参数集。
     - 如果您选择输入自己的参数，请执行以下操作：
       - 通过从下拉列表中选择 “ AWS 账户和区域” 来提供 AWS 参数。

- 提供 Redshift 数据仓库参数：
  - 选择 Amazon Redshift 集群或 Amazon Redshift Serverless
  - 输入保存所选 Amazon Redshift 集群或 Amazon Redshift Serverless 工作组凭证的 AWS 秘密 ARN。必须使用在其中创建环境配置文件的域 ID 和项目 ID 来标记 AWS 密钥。
    - `AmazonDataZoneDomain: [Domain_ID]`
    - `AmazonDataZoneProject: [Project_ID]`
  - 输入 Amazon Redshift 集群或 Amazon Redshift Serverless 工作组的名称。
  - 输入所选 Amazon Redshift 集群或 Amazon Redshift Serverless 工作组中的数据库名称。
- 在授权项目部分中，指定可以使用环境配置文件来创建环境的项目。默认情况下，域中的所有项目都可以使用账户中的环境配置文件来创建环境。要保留此默认设置，请选择所有项目。不过，您可以通过将授权项目分配给环境来限制这一点。为此，请选择仅限授权项目，然后指定可以使用此项目配置文件来创建环境的项目。
- 在发布部分中，选择下列选项之一：
  - 从任何架构发布：如果选择此选项，则使用此环境配置文件创建的环境可用于从上面提供的 Redshift 参数中选定的数据库中的任何架构进行发布。使用此环境配置文件创建的环境的用户也可以提供自己的 Amazon Redshift 参数，以便在环境配置文件中选择的 AWS 账户和区域内的任何架构中进行发布。
  - 仅从默认环境架构发布：如果您选择此选项，则使用此选项创建的环境只能用于从 Amazon DataZone 为该环境创建的默认架构进行发布。使用此环境配置文件创建的环境的用户无法提供自己的 Amazon Redshift 参数。
  - 不允许发布：如果选择此选项，则使用此环境配置文件创建的环境只能用于订阅和使用数据。环境根本无法用于发布任何数据。

如果已指定数据湖蓝图，请执行以下操作：

- 在AWS 账户参数部分，指定 AWS 账号和将在其中创建潜在环境的 AWS 账户区域。
- 在授权项目部分中，指定可以结合使用此环境配置文件与内置数据湖环境配置文件来创建环境的项目。默认情况下，域中的所有项目都可以使用账户中的数据湖蓝图来创建环境配置文件。要保留此默认设置，请选择所有项目。但是，您可以通过将项目分配给蓝图来施加限制。为此，请选择仅限授权项目，然后指定可以使用此项目配置文件来创建环境的项目。
- 在 “数据库” 部分中，选择 “任何数据库” 以允许从创建环境的 AWS 帐户和区域内的任何数据库进行发布，或者选择 “仅默认数据库” 以仅允许从使用环境创建的默认发布数据库进行发布。

5. 选择创建环境配置文件。

编辑环境配置文件

在 Amazon 中 DataZone，环境配置文件是您可以用来创建环境的模板。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。要编辑亚马逊 DataZone 域中的现有环境配置文件，您必须属于亚马逊 DataZone 项目。

编辑环境配置文件

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。
   如果您是亚马逊 DataZone 管理员，则可以导航到亚马逊 DataZone 控制台，https://console.aws.amazon.com/datazone并使用域名创建 AWS 账户 地登录，然后选择打开数据门户。
2. 在数据门户中，选择浏览项目，然后选择要在其中编辑环境配置文件的项目。
3. 导航到项目中的环境选项卡，选择环境配置文件，然后选择要编辑的环境配置文件。

   如果您编辑的是数据仓库环境配置文件，则只能编辑现有环境配置文件的名称和描述。

   如果您编辑的是数据湖环境配置文件，则可以编辑配置文件的名称和描述、有权使用此配置文件来创建环境的项目以及数据库。要编辑这些设置，请执行以下操作：

   - 在授权项目部分中，指定可以结合使用此环境配置文件与内置数据湖环境配置文件来创建环境的项目。默认情况下，域中的所有项目都可以使用账户中的数据湖蓝图来创建环境配置文件。要保留此默认设置，请选择所有项目。但是，您可以通过将项目分配给蓝图来施加限制。为此，请选择仅限授权项目，然后指定可以使用此项目配置文件来创建环境的项目。
   - 在 “数据库” 部分中，选择 “任何数据库” 以允许从创建环境的 AWS 帐户和区域内的任何数据库进行发布，或者选择 “仅默认数据库” 以仅允许从使用环境创建的默认发布数据库进行发布。

   完成编辑后，选择编辑环境配置文件。

删除环境配置文件

在 Amazon 中 DataZone，环境配置文件是您可以用来创建环境的模板。环境配置文件的目的是通过在配置文件中嵌入放置信息（例如 AWS 账户和区域）来简化环境创建。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。要删除亚马逊 DataZone域中的环境配置文件，您必须属于亚马逊 DataZone 项目。

> **Note**
>
> 删除一个环境配置文件后，您将无法再使用此配置文件来创建任何环境。

删除环境配置文件

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。如果您是亚马逊 DataZone 管理员，则可以导航到亚马逊 DataZone 控制台，https://console.aws.amazon.com/datazone并使用域名创建 AWS 账户 地登录，然后选择打开数据门户。
2. 在数据门户中，选择浏览项目，然后选择要在其中删除环境配置文件的项目。
3. 导航到项目中的环境选项卡，选择环境配置文件，然后选择要删除的环境配置文件。
4. 选择要删除的环境配置文件，然后选择操作和删除，并确认删除。

创建新环境

在 Amazon DataZone 项目中，环境是已配置资源的集合（例如，Amazon S3 存储桶、 AWS Glue 数据库或 Amazon Athena 工作组），具有一组给定的 IAM 委托人（环境用户角色），分配了可以对这些资源进行操作的所有者或贡献者权限。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。

任何具有访问数据门户所需权限的亚马逊 DataZone 用户都可以在项目中创建亚马逊 DataZone 环境。

要创建新环境，请完成以下步骤。

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。如果您是亚马逊 DataZone 管理员，则可以导航到亚马逊 DataZone 控制台，https://console.aws.amazon.com/datazone并使用域名创建 AWS 账户 地登录，然后选择打开数据门户。
2. 选择浏览所有项目，然后选择要在其中创建新环境的项目。
3. 选择创建环境，指定以下字段的值，然后选择创建环境：
   - 名称 – 环境名称
   - 描述 – 环境的描述
   - 环境配置文件 – 选择现有的环境配置文件或创建新的环境配置文件。环境配置文件是一个可用于创建环境的模板。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。

     选择环境配置文件后，在参数部分下，为该环境配置文件中的字段指定值。

编辑环境

在 Amazon DataZone 项目中，环境是已配置资源的集合（例如，Amazon S3 存储桶、 AWS Glue 数据库或 Amazon Athena 工作组），其中的一组给定的 IAM 委托人（具有分配的贡献者权限）可以对这些资源进行操作。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。

任何具有访问数据门户所需权限的亚马逊 DataZone 用户都可以在项目中编辑亚马逊 DataZone 环境。

要编辑现有环境，请完成以下步骤。

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。
   如果您是亚马逊 DataZone 管理员，则可以导航到亚马逊 DataZone 控制台，https://console.aws.amazon.com/datazone并使用域名创建 AWS 账户 地登录，然后选择打开数据门户。
2. 从顶部导航窗格中选择浏览项目，然后选择包含要编辑的环境的项目。
3. 找到并选择该环境以打开其详细信息页面。然后展开操作，并选择编辑环境。
4. 对环境的名称和描述进行更改，然后选择保存更改。

删除环境

在 Amazon DataZone 项目中，环境是已配置资源的集合（例如，Amazon S3 存储桶、 AWS Glue 数据库或 Amazon Athena 工作组），其中的一组给定的 IAM 委托人（具有分配的贡献者权限）可以对这些资源进行操作。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。

任何具有访问数据门户所需权限的亚马逊 DataZone 用户都可以在项目中删除亚马逊 DataZone 环境。

要删除现有环境，请完成以下步骤。

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。
   如果您是亚马逊 DataZone 管理员，则可以导航到亚马逊 DataZone 控制台，https://console.aws.amazon.com/datazone并使用域名创建 AWS 账户 地登录，然后选择打开数据门户。
2. 从顶部导航窗格中选择浏览项目，然后选择包含要删除的环境的项目。
3. 找到并选择该环境以打开其详细信息页面，然后展开操作并选择删除环境。
4. 在删除环境弹出窗口中，通过在字段中键入 `Delete` 来确认删除，然后选择删除环境。

   仅在删除所有依赖某个环境的实体后，才能成功删除该环境。要删除环境，您必须先删除其所有关联的数据来源和订阅目标。

创建新项目

在亚马逊中 DataZone，项目使一组用户能够就各种业务用例进行协作，这些用例涉及发布、发现、订阅和使用亚马逊 DataZone 目录中的数据资产。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。

任何具有访问数据门户所需权限的亚马逊 DataZone 用户都可以创建亚马逊 DataZone 项目。

要创建新项目，请完成以下步骤。

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。
   如果您是亚马逊 DataZone 管理员，则可以导航到亚马逊 DataZone 控制台，https://console.aws.amazon.com/datazone并使用域名创建 AWS 账户 地登录，然后选择打开数据门户。
2. 在 Amazon DataZone 数据门户中，选择创建项目。
3. 指定以下字段的值，然后选择创建项目：
   - 名称 – 项目名称。
   - 描述 – 项目的描述。
   - 域单元 – 要在其下创建此项目的域单元。

编辑项目

在亚马逊中 DataZone，项目使一组用户能够就各种业务用例进行协作，这些用例涉及发布、发现、订阅和使用亚马逊 DataZone 目录中的数据资产。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。要编辑 Amazon DataZone 项目，您必须是该项目的所有者或包含该项目的域名的域管理员。

要编辑现有项目，请完成以下步骤。

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。
   如果您是亚马逊 DataZone 管理员，则可以导航到亚马逊 DataZone 控制台，https://console.aws.amazon.com/datazone并使用域名创建 AWS 账户 地登录，然后选择打开数据门户。
2. 选择浏览项目。
3. 选择要编辑的项目。如果项目列表中未显示此项目，则可以通过在查找项目字段中指定项目名称来搜索此项目。
4. 展开操作并选择编辑项目。
5. 对项目名称和描述进行更新，然后选择保存。

将项目移动到其他域单元

在亚马逊中 DataZone，项目使一组用户能够就各种业务用例进行协作，这些用例涉及发布、发现、订阅和使用亚马逊 DataZone 目录中的数据资产。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。

要将 Amazon DataZone 项目移至其他域单位，您必须满足以下要求：

- 您必须拥有在要将项目移动到的域单元内创建项目的策略授权。
- 项目的所有成员都必须拥有您要将项目移动到的域单元中的项目成员资格。
- 您必须是要将项目移动到的域单元的所有者。
- 您必须是项目的所有者。

要将现有项目移动到其他域单元，请完成以下步骤。

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。如果您是亚马逊 DataZone 管理员，则可以导航到亚马逊 DataZone 控制台，https://console.aws.amazon.com/datazone并使用域名创建 AWS 账户 地登录，然后选择打开数据门户。
2. 选择浏览项目。
3. 选择要移动的项目。如果项目列表中未显示此项目，则可以通过在查找项目字段中指定项目名称来搜索此项目。
4. 展开操作并选择移动项目。
5. 指定要在其下移动此项目的域单元，然后选择移动。

删除项目

在亚马逊 DataZone，项目使一组用户能够就各种业务用例进行协作，这些 and/or 用例涉及发布、发现、订阅、使用亚马逊 DataZone 目录中的数据资产。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。

删除项目是最终操作。删除操作将不可撤销地删除项目的内容，包括数据来源、环境、资产、术语表和元数据表单。亚马逊 DataZone 撤销 DataZone 了亚马逊通过Lake Formation和Amazon Redshift对托管资产发放的补助。删除项目不会删除亚马逊 DataZone 可能帮助您创建的非亚马逊 DataZone AWS 资源。如果您不再需要这些 AWS 资源，请在相应的 AWS 服务和帐户中将其删除。

要删除 Amazon DataZone 项目，您必须是该项目的所有者。

要删除现有项目，请完成以下步骤。

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。IAM 委托人可以导航到 Amazon DataZone 控制台，使用域名创建 AWS 账户 地登录，然后选择打开数据门户。https://console.aws.amazon.com/datazone
2. 从顶部导航窗格中选择浏览项目。
3. 选择要删除的项目。如果项目列表中未显示此项目，则可以通过在查找项目字段中指定项目名称来搜索此项目。
4. 展开操作并选择删除项目。

   查看有关删除项目可能造成的影响的信息性警告。
5. 如果您接受警告，请键入确认文本，然后选择删除。

> **Important**
>
> 删除项目是一项不可撤销的操作，您或 AWS无法取消此操作。

> **Note**
>
> 当您或您的域用户在项目中创建环境时，Amazon DataZone 会在您的域名或关联账户中创建 AWS 资源，为您和您的域用户提供功能。以下是 Amazon DataZone 可能为项目创建的 AWS 资源列表以及默认名称。删除项目不会删除您 AWS 账户中的任何此类 AWS 资源。
>
> - IAM 角色：datazone_usr_<environmentId>。
> - Glue 数据库：（1）<environmentName>_pub_db-*、（2）<environmentName>_sub_db-*。如果已经存在同名数据库，Amazon DataZone 将添加环境 ID。
> - Athena 工作组：<environmentName>-*。如果已经存在同名工作组，Amazon DataZone 将添加环境 ID。
> - CloudWatch 日志组：datazone_ <environmentId>

离开项目

在亚马逊中 DataZone，项目使一组用户能够就各种业务用例进行协作，这些用例涉及发布、发现、订阅和使用亚马逊 DataZone 目录中的数据资产。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。

要离开现有项目，请完成以下步骤。

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。如果您是亚马逊 DataZone 管理员，则可以导航到亚马逊 DataZone 控制台，https://console.aws.amazon.com/datazone并使用域名创建 AWS 账户 地登录，然后选择打开数据门户。
2. 从顶部导航窗格中选择选择项目，然后选择该项目。
3. 选择要离开的项目。如果项目列表中未显示此项目，则可以通过在查找项目字段中指定项目名称来搜索此项目。
4. 展开操作并选择离开项目。

向项目添加成员

在亚马逊中 DataZone，项目使一组用户能够就各种业务用例进行协作，这些用例涉及发布、发现、订阅和使用亚马逊 DataZone 目录中的数据资产。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。

要向项目添加成员，您必须是项目所有者。您可以将 SSO 组、SSO 用户或 IAM 主体（角色或用户）添加为项目成员。

要向现有项目添加成员，请完成以下步骤。

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。如果您是亚马逊 DataZone 管理员，则可以导航到亚马逊 DataZone 控制台，https://console.aws.amazon.com/datazone并使用域名创建 AWS 账户 地登录，然后选择打开数据门户。
2. 从顶部导航窗格中选择选择项目，然后选择该项目。
3. 选择要将成员添加到的项目。如果项目列表中未显示此项目，则可以通过在查找项目字段中指定项目名称来搜索此项目。
4. 在项目的详细信息页面上，选择成员选项卡，然后选择所有成员节点。
5. 在“项目成员”选项卡中，选择添加成员。
6. 在向项目添加成员弹出窗口中，指定要添加的用户以及他们在项目中的角色（所有者、贡献者、使用者、管理者或查看者），然后选择添加成员。

> **Important**
>
> 您只能将这些用户添加为项目成员，这些成员通过为该项目所在的域单元配置的项目成员资格授权策略授权成为该项目的成员。有关更多信息，请参阅 为 Amazon DataZone 域单位内的用户和群组分配授权策略。

> **Note**
>
> 如果 IAM 委托人的域中已有 Amazon DataZone 用户资料，则可以将该委托人添加为项目成员。当亚马逊通过门户、API 或 CLI 成功与 IAM 委托人交互时，它 DataZone 会自动为该域名创建用户档案。您无法为 IAM 主体创建用户配置文件。如果 IAM 委托人在域中没有现有 Amazon DataZone 用户资料，则要将 IAM 委托人添加为项目成员，请管理员在 IAM 控制台AmazonDataZoneDomainExecutionRole中为您的域添加以下两个 IAM 权限：`iam:GetUser`和`iam:GetRole`。另外，IAM 主体必须拥有相应的 IAM 权限才能在域中执行某些操作。

从项目中移除成员

在亚马逊中 DataZone，项目使一组用户能够就各种业务用例进行协作，这些用例涉及发布、发现、订阅和使用亚马逊 DataZone 目录中的数据资产。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。您必须是项目所有者才能从项目中移除成员。

要从现有项目中移除成员，请完成以下步骤。

1. 使用 DataZone 数据门户 URL 导航至 Amazon 数据门户，然后使用您的 SSO 或 AWS 凭证登录。如果您是亚马逊 DataZone 管理员，则可以通过访问创建亚马逊 DataZone 域名的 AWS 账户https://console.aws.amazon.com/datazone中的亚马逊 DataZone 控制台来获取数据门户 URL。
2. 从顶部导航窗格中选择选择项目，然后选择该项目。
3. 选择要从中移除成员的项目。如果项目列表中未显示此项目，则可以通过在查找项目字段中指定项目名称来搜索此项目。
4. 在项目的详细信息页面上，选择成员选项卡，然后选择所有成员节点。
5. 在“项目成员”选项卡中，选择要从项目中移除的成员，然后选择移除。
6. 在移除成员弹出窗口中，通过选择移除成员来确认移除。

数据库存和在 Amazon 上发布 DataZone

本节介绍您要执行的任务和程序，以便在亚马逊上创建数据清单， DataZone 并在亚马逊上发布数据 DataZone。

要使用亚马逊对您的数据 DataZone 进行分类，您必须先将您的数据（资产）作为项目库存带到亚马逊 DataZone。为特定项目创建库存，从而仅允许该项目的成员发现资产。 search/browse 除非明确发布，否则并非所有域用户都可以使用项目清单资产。创建项目库存后，数据所有者可以添加或更新业务名称（资产和架构）、描述（资产和架构）、自述文件、术语表术语（资产和架构）和元数据表单，从而使用所需的业务元数据来整理库存资产。

使用 Amazon DataZone 对您的数据进行分类的下一步是让域名用户可以发现您项目的库存资产。您可以通过将库存资产发布到 Amazon DataZone 目录来做到这一点。只能将最新版本的库存资产发布到目录，并且仅最新发布的版本在发现目录中处于活动状态。如果库存资产在发布到亚马逊 DataZone 目录后进行了更新，则必须再次明确发布该库存资产，以使最新版本出现在发现目录中。

有关更多信息，请参阅 亚马逊 DataZone 术语和概念。

主题

- 为亚马逊配置 Lake Formation 权限 DataZone
- 在 Amazon 中创建自定义资产类型 DataZone
- 为创建并运行 Amazon DataZone 数据源 AWS Glue Data Catalog
- 为亚马逊 Redshift 创建并运行亚马逊 DataZone 数据源
- 在 Amazon 中编辑数据源 DataZone
- 在 Amazon 中删除数据源 DataZone
- 将项目库存中的资产发布到 Amazon DataZone 目录
- 在 Amazon 中管理库存和整理资产 DataZone
- 在 Amazon 中手动创建资产 DataZone
- 从 Amazon DataZone 目录中取消发布资产
- 删除亚马逊 DataZone 资产
- 手动启动在 Amazon 中运行的数据源 DataZone
- Amazon 中的资产修订 DataZone
- Amazon 的数据质量 DataZone

- 在 Amazon 中使用机器学习和生成人工智能 DataZone
- Amazon 中的数据谱系 DataZone
- 针对发布的元数据强制规则
- 在亚马逊中将 Snowflake 作为数据源进行连接 DataZone
- 启用 Snowflake 血统 AWS Glue Spark 作业

为亚马逊配置 Lake Formation 权限 DataZone

当您使用内置的数据湖蓝图 (DefaultDataLake) 创建环境时，将在该环境的创建过程中在 Amazon DataZone 中添加一个 AWS Glue 数据库。如果要从此 AWS Glue 数据库发布资产，则无需其他权限。

但是，如果您想发布资产并订阅存在于亚马逊 DataZone 环境之外的 AWS Glue 数据库中的资产，则必须明确向亚马逊 DataZone 提供访问此外部 AWS Glue 数据库中表的权限。为此，你必须在 AWS Lake Formation 中完成以下设置，并将必要的 Lake Formation 权限附加到AmazonDataZoneGlueAccess-<领域>-<域名>。

- 使用 Lake Formation 权限模式或混合访问模式在 AWS Lake Formation 中为您的数据湖配置 Amazon S3 位置。有关更多信息，请参阅 https://docs.aws.amazon.com/lake-formation/latest/dg/register-data-lake.html。
- 从亚马逊 DataZone 处理`IAMAllowedPrincipals`权限的 Amazon Lake Formation 表中移除权限。有关更多信息，请参阅 https://docs.aws.amazon.com/lake-formation/latest/dg/upgrade-glue-lake-formation-background.html。
- 将以下 AWS Lake Formation 权限附加到AmazonDataZoneGlueAccess-<领域>-<域名>：
  - 表所在的数据库的 `Describe` 和 `Describe grantable` 权限
  - `Describe`、`SelectDescribe Grantable`、、以上数据库中您 DataZone 要代表您管理访问`Select Grantable`权限的所有表的权限。

> **Note**
>
> 亚马逊 DataZone 支持 AWS Lake Formation 混合模式。Lake For AWS mation 混合模式使您可以开始通过 Lake Formation 管理您的 Glue 数据库和表的权限，同时继续保留对这些表和数据库的任何现有 IAM 权限。有关更多信息，请参阅 亚马逊 DataZone 与 AWS Lake Formation 混合模式。

有关更多信息，请参阅 问题排查 AWS 亚马逊的 Lake Formation 权限 DataZone。

亚马逊 DataZone 与 AWS Lake Formation 混合模式

亚马逊 DataZone 已与 La AWS ke Formation 混合模式集成。这种集成使您能够轻松地通过亚马逊发布和共享您的 AWS Glue 表， DataZone而无需先在 AWS Lake Formation 中注册它们。混合模式允许您开始通过 AWS Lake For AWS mation 管理您的 Glue 表的权限，同时继续保持对这些表的任何现有 IAM 权限。

首先，您可以在 Amazon DataZone 管理控制台中启用DefaultDataLake蓝图下的数据位置注册设置。

启用与集成 AWS Lake Formation 混合模式

1. 导航至 Amazon DataZone 控制台 https://console.aws.amazon.com/datazone，并使用您的账户凭证登录。
2. 选择 “查看域”，然后选择要在其中启用与 AWS Lake Formation 混合模式集成的域。
3. 在域详细信息页面上，导航到蓝图选项卡。
4. 从蓝图列表中选择DefaultDataLake蓝图。
5. 确保 DefaultDataLake 蓝图已启用。如果未启用此蓝图，请按照在拥有 Amazon DataZone 域 AWS 名的账户中启用内置蓝图中的步骤操作，在您的 AWS 账户中启用它。
6. 在 DefaultDataLake 详细信息页面上，打开配置选项卡，然后选择页面右上角的编辑按钮。
7. 选中数据位置注册下方的框以启用数据位置注册。
8. 对于数据位置管理角色，您可以创建新 IAM 角色或选择现有 IAM 角色。亚马逊 DataZone 使用此角色通过 Lake Formation 混合 read/write访问模式管理对为数据湖选择的 Amazon S3 存储桶的访问权限。 AWS 有关更多信息，请参阅 AmazonDataZoneS3Manage-<领域>-<域名>。
9. 或者，如果您不希望亚马逊在混合模式下自动注册某些 Amazon S3 地点 DataZone ，则可以选择将其排除在外。为此，请完成以下步骤：

   - 选择切换按钮以排除指定的 Amazon S3 位置。
   - 提供要排除的 Amazon S3 存储桶的 URI。
   - 要添加其他存储桶，请选择添加 S3 位置。

> **Note**
>
> Amazon DataZone 仅允许排除 S3 根位置。系统将自动从注册中排除根 S3 位置路径内的任何 S3 位置。

- 选择保存更改。

在 AWS 账户中启用数据位置注册设置后，当数据使用者订阅通过 IAM 权限管理的 AWS Glue 表时，亚马逊 DataZone 将首先以混合模式注册该表的 Amazon S3 位置，然后通过 La AWS ke Formation 管理表的权限，向数据使用者授予访问权限。这样可以确保使用新授予的 La AWS ke Formation 权限继续存在表上的 IAM 权限，而不会中断任何现有工作流程。

启用时如何处理加密的 Amazon S3 位置 AWS 亚马逊中的 Lake Formation 混合模式集成 DataZone

如果您使用的是使用客户托 AWS 管或托管 KMS 密钥加密的 Amazon S3 位置，则该AmazonDataZoneS3Manage角色必须具有使用 KMS 密钥加密和解密数据的权限，或者 KMS 密钥策略必须向该角色授予对密钥的权限。

如果您的 Amazon S3 位置使用 AWS 托管密钥加密，请向该AmazonDataZoneDataLocationManagement角色添加以下内联策略：

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "kms:Encrypt",
                "kms:Decrypt",
                "kms:ReEncrypt*",
                "kms:GenerateDataKey*",
                "kms:DescribeKey"
            ],
            "Resource": "arn:aws:kms:us-
east-1:111122223333:key/1234abcd-12ab-34cd-56ef-1234567890ab"
        }
    ]
}
```

如果您的 Amazon S3 位置已使用客户自主管理型密钥进行加密，请执行以下操作：

1. 打开 AWS KMS 控制台，https://console.aws.amazon.com/kms并以 AWS 身份和访问管理 (IAM) 管理用户或可以修改用于加密位置的 KMS 密钥策略的用户身份登录。
2. 在导航窗格中，选择客户自主管理型密钥，然后选择所需的 KMS 密钥的名称。
3. 在 KMS 密钥详细信息页面上，选择密钥策略选项卡，然后执行以下任一操作将您的自定义角色或 Lake Formation 服务相关角色添加为 KMS 密钥用户：

   - 如果显示默认视图（包括“密钥管理员”、“密钥删除”、“密钥用户”和“其他 AWS 账户”部分），请在“密钥用户”部分下添加AmazonDataZoneDataLocationManagement角色。
   - 如果显示密钥策略 (JSON)，请编辑策略以向“允许使用密钥”对象添加AmazonDataZoneDataLocationManagement角色，如以下示例所示

```
...
        {
            "Sid": "Allow use of the key",
            "Effect": "Allow",
            "Principal": {
                "AWS": [
                    "arn:aws:iam::111122223333:role/service-role/
AmazonDataZoneDataLocationManage-<region>-<domain-id>"
                ]
            },
            "Action": [
                "kms:Encrypt",
                "kms:Decrypt",
                "kms:ReEncrypt*",
                "kms:GenerateDataKey*",
                "kms:DescribeKey"
            ],
            "Resource": "*"
        },
        ...
```

> **Note**
>
> 如果 KMS 密钥或 Amazon S3 位置与数据目录不在同一个 AWS 账户中，请按照跨 AWS 账户注册加密的 Amazon S3 位置中的说明进行操作。

在 Amazon 中创建自定义资产类型 DataZone

在 Amazon 中 DataZone，资产代表特定类型的数据资源，例如数据库表、控制面板或机器学习模型。为了在描述目录资产时保持一致性和标准化，Amazon DataZone 域必须有一组资产类型来定义资产在目录中的表示方式。资产类型定义特定类型的资产的架构。资产类型具有一组必填和可选的可命名元数据表单类型（例如 GovForm 或 GovernanceFormType）。Amazon 中的资产类型 DataZone 是版本化的。在创建资产时，将根据资产类型（通常是最新版本）定义的架构来验证资产，如果指定的结构无效，则将无法创建资产。

系统资产类型-Ama DataZone zon 预置服务拥有的系统资产类型（包括 GlueTableAssetType GlueViewAssetType RedshiftTableAssetType RedshiftViewAssetType、、和 S3ObjectCollectionAssetType）和系统表单类型（包括 DataSourceReferenceFormType AssetCommonDetailsFormType、和 SubscriptionTermsFormType）。无法编辑系统资产类型。

自定义资产类型 – 要创建自定义资产类型，首先创建要在表单类型中使用的所需的元数据表单类型和术语表。之后，您可以通过指定名称、描述和关联的元数据表单（必需或可选）来创建自定义资产类型。

对于具有结构化数据的资产类型，要表示数据门户中的列架构，您可以使用 `RelationalTableFormType` 向列添加技术元数据（包括列名、描述和数据类型），并使用 `ColumnBusinessMetadataForm` 添加列的企业描述，包括企业名称、术语表术语和自定义键值对。

要通过数据门户创建自定义资产类型，请完成以下步骤：

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。
   如果您是亚马逊 DataZone 管理员，则可以导航到亚马逊 DataZone 控制台，https://console.aws.amazon.com/datazone并使用域名创建 AWS 账户 地登录，然后选择打开数据门户。
2. 从顶部导航窗格中选择选择项目，然后选择要在其中创建自定义资产类型的项目。
3. 导航到项目的数据选项卡。
4. 从左侧导航窗格中选择资产类型，然后选择创建资产类型。
5. 指定以下内容，然后选择创建。
   - 名称 – 自定义资产类型的名称
   - 描述 – 自定义资产类型的描述。
   - 选择“添加元数据表单”以将元数据表单添加到此自定义资产类型。
6. 创建自定义资产类型后，您可以使用它创建资产。

要通过 API 创建自定义资产类型，请完成以下步骤：

1. 通过调用 CreateFormType API 操作来创建元数据表单类型。

   以下是 Amazon 的 SageMaker 示例：

```
m_model = "

structure SageMakerModelFormType {
   @required
   @amazon.datazone#searchable
   modelName: String

   @required
   modelArn: String

   @required
   creationTime: String
}
"

CreateFormType(
    domainIdentifier="my-dz-domain",
    owningProjectIdentifier="d4bywm0cja1dbb",
    name="SageMakerModelFormType",
    model=m_model
    status="ENABLED"
    )
```

2. 接下来，您可以通过调用 CreateAssetType API 操作来创建资产类型。您只能使用可用的系统表单类型（SubscriptionTermsFormType在以下示例中）或自定义表单类型通过 Amazon DataZone API 创建资产类型。对于系统表单类型，类型名称必须以 amazon.datazone 开头。

```
CreateAssetType(
    domainIdentifier="my-dz-domain",
    owningProjectIdentifier="d4bywm0cja1dbb",
    name="SageMakerModelAssetType",
    formsInput={
        "SageMakerModelForm": {
            "typeIdentifier": "SageMakerModelFormType",
```

```
            "typeRevision": 7,
            "required": True,
        },
        "SubscriptionTerms": {
            "typeIdentifier": "amazon.datazone.SubscriptionTermsFormType",
            "typeRevision": 1,
            "required": False,
        },
    },
)
```

以下是为结构化数据创建资产类型的示例：

```
CreateAssetType(
    domainIdentifier="my-dz-domain",
    owningProjectIdentifier="d4bywm0cja1dbb",
    name="OnPremMySQLAssetType",
    formsInput={
        "OnpremMySQLForm": {
            "typeIdentifier": "OnpremMySQLFormType",
            "typeRevision": 5,
            "required": True,
        },
        "RelationalTableForm": {
            "typeIdentifier": "amazon.datazone.RelationalTableFormType",
            "typeRevision": 1,
            "required": True,
        },
        "ColumnBusinessMetadataForm": {
            "typeIdentifier": "amazon.datazone.ColumnBusinessMetadataFormType",
            "typeRevision": 1,
            "required": False,
        },
        "SubscriptionTerms": {
            "typeIdentifier": "amazon.datazone.SubscriptionTermsFormType",
            "typeRevision": 1,
            "required": False,
        },
    },
)
```

3. 现在，您可以使用上述步骤中创建的自定义资产类型来创建资产。

```
CreateAsset(
   domainIdentifier="my-dz-domain",
   owningProjectIdentifier="d4bywm0cja1dbb",
   typeIdentifier="SageMakerModelAssetType",
   name="MyModelAsset",
   glossaryTerms="xxx",
   formsInput=[{
        "formName": "SageMakerModelForm",
        "typeIdentifier": "SageMakerModelFormType",
        "content": "{\n \"ModelName\" : \"sample-ModelName\",\n \"ModelArn\" :
 \"999999911111\",\n \"CreationTime\" : \"2025-01-01 18:00:00.000\"}"
        }
        ]
)
```

在此示例中，您创建的是结构化数据资产：

```
CreateAsset(
   domainIdentifier="my-dz-domain",
   owningProjectIdentifier="d4bywm0cja1dbb",
   typeIdentifier="OnPremMySQLAssetType",
   name="MyModelAsset",
   glossaryTerms="xxx",
   formsInput=[{
        "formName": "RelationalTableForm",
        "typeIdentifier": "amazon.datazone.RelationalTableFormType",
        "content": ".."
        },
        {
        "formName": "OnpremMySQLForm",
        "typeIdentifier": "OnpremMySQLFormType",
        "content": ".."
        },
        {
        "formName": "mySQLTableForm",
        "typeIdentifier": "MySQLTableFormType",
        "typeRevision": "1",
        "content": ".."
```

```
        },
        {
        "formName": "AssetCommonDetailsForm",
        "typeIdentifier": "amazon.datazone.AssetCommonDetailsFormType",
        "content": "..."
        },
        .....
        ]
)
```

为创建并运行 Amazon DataZone 数据源 AWS Glue Data Catalog

在 Amazon 中 DataZone，您可以创建 AWS Glue Data Catalog 数据源，以便从中导入数据库表的技术元数据 AWS Glue。要为添加数据源 AWS Glue Data Catalog，源数据库必须已存在于 AWS Glue。

创建和运行 AWS Glue 数据源时，会将源 AWS Glue 数据库中的资产添加到您的 Amazon DataZone 项目的库存中。您可以按设定的时间表或按需运行 AWS Glue 数据源，以创建或更新资产的技术元数据。在数据源运行期间，您可以选择将您的资产发布到 Amazon DataZone 目录，从而让所有域用户都能发现这些资产。也可以在编辑项目库存资产的企业元数据后发布这些资产。域用户可以搜索和发现已发布的资产，并请求订阅这些资产。

要添加 AWS Glue 数据来源

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。
   如果您是亚马逊 DataZone 管理员，则可以导航到亚马逊 DataZone 控制台，https://console.aws.amazon.com/datazone并使用域名创建 AWS 账户 地登录，然后选择打开数据门户。
2. 从顶部导航窗格中选择选择项目，然后选择要将数据来源添加到的项目。
3. 导航到项目的数据选项卡。
4. 从左侧导航窗格中选择数据来源，然后选择创建数据来源。
5. 配置以下字段：
   - 名称 – 数据来源名称。
   - 描述 – 数据来源描述。
6. 在数据来源类型下，选择 AWS Glue。
7. 在 “选择环境” 下，指定要在其中发布 AWS Glue 表的环境。

8. 在数据选择下，提供一个 AWS Glue 数据库并输入您的表选择标准。例如，如果您选择包括并输入 `*corporate`，则数据库将包括所有以 `corporate` 一词结尾的源表。

   您可以从下拉列表中选择一个 AWS Glue 数据库，也可以键入数据库名称。下拉列表包括两个数据库：环境的发布数据库和订阅数据库。如果要从并非由环境创建的数据库引入资产，您必须键入数据库的名称，而不是从下拉列表中选择数据库。

   可以为单个数据库中的表添加多个包含和排除规则。也可以使用添加另一个数据库按钮来添加多个数据库。

9. 在数据质量下，可以选择对此数据来源启用数据质量自动监测功能。如果您这样做，亚马逊会将您现有的 AWS Glue 数据质量输出 DataZone导入您的亚马逊 DataZone目录中。默认情况下，亚马逊会从 AWS Glue DataZone 导入现有 100 份没有有效期的最新 100 份质量报告。

   Amazon 的数据质量指标 DataZone 可帮助您了解数据源的完整性和准确性。亚马逊从 AWS G DataZone lue 中提取这些数据质量指标，以便在某个时间点（例如在搜索业务数据目录期间）提供背景信息。数据用户可以查看其订阅的资产的数据质量指标随时间变化的情况。数据创建者可以按计划摄取 AWS Glue 数据质量分数。亚马逊 DataZone 企业数据目录还可以通过数据质量 API 显示来自第三方系统的数据质量指标。有关更多信息，请参阅 Amazon 的数据质量 DataZone。

10. 选择下一步。
11. 对于发布设置，选择是否可以在企业数据目录中立即发现资产。如果您仅将资产添加到库存中，则可以稍后选择订阅条款并将资产发布到企业数据目录。
12. 对于自动生成企业名称，请选择是否在从来源导入资产时自动为其生成元数据。
13. （可选）对于元数据表单，添加表单以定义在资产导入 Amazon 时收集和保存的元数据 DataZone。有关更多信息，请参阅 the section called “创建元数据表单”。
14. 在运行偏好中，选择何时运行数据来源。

    - 按时间表运行 – 指定数据来源的运行日期和时间。
    - 按需运行 – 可以手动启动数据来源运行。

15. 选择下一步。
16. 检查您的数据来源配置，然后选择创建。

> **Note**
>
> 创建 AWS Glue 数据源时，亚马逊 DataZone 会为环境的 IAM 角色创建 Lake Formation “只读” 权限，该角色用于创建数据源，以访问数据源中使用的 G AWS lue 数据库中的所

有表。您可在环境的详细信息页面上，在数据来源下监控这些授权的状态。在向发布环境的 IAM 角色授予访问权限时，亚马逊会向 AWS Glue 数据库 DataZone 添加以下 AWS 标签：`DataZoneDiscoverable_${domainId}: true`
对于在 Amazon 当前版本之前创建的环境 DataZone，项目成员将无法在 Amazon Athena 中查看已授权的表。

为亚马逊 Redshift 创建并运行亚马逊 DataZone 数据源

在亚马逊中 DataZone，您可以创建亚马逊 Redshift 数据源，以便从亚马逊 Redshift 数据仓库中导入数据库表和视图的技术元数据。要为亚马逊 Redshift 添加亚马逊 DataZone 数据源，源数据仓库必须已经存在于亚马逊 Redshift 中。

创建和运行 Amazon Redshift 数据源时，您可以将源亚马逊 Redshift 数据仓库中的资产添加到您的 DataZone 亚马逊项目的库存中。您可以按设定的时间表或按需运行 Amazon Redshift 数据来源，以创建或更新资产的技术元数据。在数据源运行期间，您可以选择将项目库存资产发布到 Amazon DataZone 目录，从而使所有域用户都能发现这些资产。也可以在编辑库存资产的企业元数据后发布这些资产。域用户可以搜索和发现已发布的资产，并请求订阅这些资产。

添加 Amazon Redshift 数据来源

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。
如果您是亚马逊 DataZone 管理员，则可以导航到亚马逊 DataZone 控制台，https://console.aws.amazon.com/datazone并使用域名创建 AWS 账户 地登录，然后选择打开数据门户。
2. 从顶部导航窗格中选择选择项目，然后选择要将数据来源添加到的项目。
3. 导航到项目的数据选项卡。
4. 从左侧导航窗格中选择数据来源，然后选择创建数据来源。
5. 配置以下字段：
   - 名称 – 数据来源名称。
   - 描述 – 数据来源描述。
6. 在数据来源类型下，选择 Amazon Redshift。
7. 在选择环境下，指定要在其中发布 Amazon Redshift 表的环境。
8. 根据您选择的环境，亚马逊 DataZone 将自动直接从环境中应用Amazon Redshift凭证和其他参数，或者允许您选择自己的凭证和其他参数。

- 如果您选择的环境仅允许通过环境的默认 Amazon Redshift 架构进行发布，则亚马逊 DataZone 将自动应用亚马逊 Redshift 凭证和其他参数，包括亚马逊 Redshift 集群或工作组名称 AWS 、密钥、数据库名称和架构名称。您无法编辑这些自动填充的参数。
- 如果您选择的环境不允许发布任何数据，则将无法继续创建数据来源。
- 如果您选择的环境允许从任何架构发布数据，则可以选择使用该环境中的证书和其他 Amazon Redshift 参数，也可以输入自己的凭证和参数。 credentials/parameters

9. 如果您选择使用自己的凭证来创建数据来源，请提供以下详细信息：

   - 在提供 Amazon Redshift 凭证下，选择是使用预置的 Amazon Redshift 集群还是 Amazon Redshift Serverless 工作区作为数据来源。
   - 根据您在上述步骤中的选择，从下拉菜单中选择您的 Amazon Redshift 集群或工作空间，然后在 Secr AWS ets Manager 中选择用于身份验证的密钥。可以选择现有密钥或创建新密钥。
   - 为了使现有密钥显示在下拉列表中，请确保你在 Secrets Manager 中的 AWS 密钥包含以下标签 (key/value)：
     - AmazonDataZoneProject: <projectID>
     - AmazonDataZoneDomain: <domainID>

     如果您选择创建新密钥，系统会自动使用上面引用的标签来标记密钥，无需执行任何额外步骤。有关更多信息，请参阅中存储数据库凭据 AWS Secrets Manager。

     为创建数据源而提供的 AWS 密钥中的 Amazon Redshift 用户必须拥有要发布的表的SELECT权限。如果您希望 Amazon DataZone 同时代表您管理订阅（访问权限），则 AWS 密钥中的数据库用户还必须具有以下权限：
     - `CREATE DATASHARE`
     - `ALTER DATASHARE`
     - `DROP DATASHARE`

10. 在数据选择下，提供一个 Amazon Redshift 数据库、架构，并输入您的表或视图选择标准。例如，如果您选择包括并输入 `*corporate`，则资产将包括所有以 `corporate` 一词结尾的源表。

    可以为单个数据库中的表添加多个包含规则。也可以使用添加另一个数据库按钮来添加多个数据库。

11. 选择下一步。

12. 对于发布设置，选择是否可以在数据目录中立即发现资产。如果您仅将资产添加到库存中，则可以稍后选择订阅条款并将资产发布到企业数据目录。

13. 对于自动生成企业名称，请选择是否在从来源发布和更新资产时自动为其生成元数据。
14. （可选）对于元数据表单，添加表单以定义在资产导入 Amazon 时收集和保存的元数据 DataZone。有关更多信息，请参阅 the section called “创建元数据表单”。
15. 在运行偏好中，选择何时运行数据来源。
    - 按时间表运行 – 指定数据来源的运行日期和时间。
    - 按需运行 – 可以手动启动数据来源运行。
16. 选择下一步。
17. 检查您的数据来源配置，然后选择创建。

> **Note**
>
> 创建 Amazon Redshift 数据源时，亚马逊会 DataZone 授予对用于创建数据源的环境的 “只读” 访问权限，以访问数据源中使用的 Amazon Redshift 架构中的所有表。您可在环境的详细信息页面上，在数据来源下监控这些授权的状态。
> 使用不同于创建环境的 Amazon Redshift 集群或无服务器工作组时，必须确保将以下 AWS 标签添加到集群或工作组。必须执行此操作才能使环境用户能够在 Amazon Redshift 查询编辑器 V2 中查看授权的数据库：`DataZoneDiscoverable_${domainId}: true`
> 对于在亚马逊当前版本之前创建的环境 DataZone，项目成员将无法在 Amazon Redshift 中查看已授权的表。

在 Amazon 中编辑数据源 DataZone

创建 Amazon DataZone 数据源后，您可以随时对其进行修改以更改源详细信息或数据选择标准。如果您不再需要某个数据来源，可以将其删除。

要完成这些步骤，您必须附加AmazonDataZoneFullAccess AWS 托管策略。有关更多信息，请参阅 the section called “AWS 托管策略”。

您可以编辑 Amazon DataZone 数据源以修改其数据选择设置，包括添加、删除或更改表选择标准。还可以添加和删除数据库。您无法更改数据来源类型或在其中发布数据来源的环境。

编辑数据来源

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。如果您是亚马逊 DataZone 管理员，则可以导航到亚马逊 DataZone 控制台，https://console.aws.amazon.com/datazone并使用域名创建 AWS 账户 地登录，然后选择打开数据门户。
2. 从顶部导航窗格中选择选择项目，然后选择数据来源所属的项目。
3. 导航到项目的数据选项卡。
4. 从左侧导航窗格中选择数据来源，然后选择要修改的数据来源。
5. 导航到数据来源定义选项卡，然后选择编辑。
6. 对数据来源定义进行更改。您可以更新数据来源详细信息并更改数据选择标准。
7. 完成更改后，选择保存。

在 Amazon 中删除数据源 DataZone

创建 Amazon DataZone 数据源后，您可以随时对其进行修改以更改源详细信息或数据选择标准。

要完成这些步骤，您必须附加AmazonDataZoneFullAccess AWS 托管策略。有关更多信息，请参阅 the section called “AWS 托管策略”。

当您不再需要 Amazon DataZone 数据源时，可以将其永久删除。删除数据来源后，仍可在目录中使用该数据来源中的所有资产，并且用户仍可以订阅它们。但是，资产将停止接收来自该来源的更新。建议您先将依赖资产移至其他数据来源，然后再删除该数据来源。

> **Note**
>
> 您必须先删除数据来源中的所有履行，之后才能将其删除。有关更多信息，请参阅 数据发现、订阅和使用。

删除数据来源

1. 在项目的数据选项卡上，从左侧导航窗格中选择数据来源。
2. 选择要删除的数据来源。
3. 依次选择操作和删除数据来源，然后确认删除。

将项目库存中的资产发布到 Amazon DataZone 目录

您可以将项目清单中的亚马逊 DataZone 资产及其元数据发布到亚马逊 DataZone 目录中。只能将资产的最新版本发布到目录。

将资产发布到目录时，请注意以下几点：

- 要将资产发布到目录，您必须是该项目的所有者或贡献者。
- 对于亚马逊 Redshift 资产，请确保与发布商和订阅者集群关联的亚马逊 Redshift 集群满足亚马逊 Redshift 数据共享的所有要求，以便亚马逊 DataZone 能够管理 Redshift 表和视图的访问权限。请参阅 Amazon Redshift 的数据共享概念。
- 亚马逊 DataZone 仅支持对从和亚马逊 Redshift 发布 AWS Glue Data Catalog 的资产进行访问管理。对于所有其他资产，例如 Amazon S3 对象，Amazon DataZone 不管理经批准的订阅者的访问权限。如果您订阅了这些非托管资产，则会收到以下消息：

  ```
  Subscription approval does not provide access to data. Subscription
  grants on this asset are not managed by Amazon DataZone. For more
  information or help, reach out to your administrator.
  ```

在 Amazon 上发布资产 DataZone

如果您在创建数据来源时未选择使资产能够立即在数据目录中被发现，请执行以下步骤以便稍后发布资产。

发布资产

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。
   如果您是亚马逊 DataZone管理员，则可以导航到亚马逊 DataZone 控制台，https://console.aws.amazon.com/datazone并使用域名创建 AWS 账户 地登录，然后选择打开数据门户。
2. 从顶部导航窗格中选择选择项目，然后选择资产所属的项目。
3. 导航到项目的数据选项卡。
4. 从左侧导航窗格中选择库存数据，然后选择要发布的资产。

> **Note**
>
> 默认情况下，所有资产都需要订阅批准，这意味着数据所有者必须批准针对资产的所有订阅请求。如果您想在发布资产前更改此设置，请打开“资产详情”并选择订阅批准旁边的编辑。稍后可通过修改并重新发布资产来更改此设置。

5. 选择发布资产。这会将资产直接发布到目录。

   如果您对资产进行了更改，例如修改其批准要求，则可以选择Re-publish将更新发布到目录。

在 Amazon 中管理库存和整理资产 DataZone

要使用亚马逊对您的数据 DataZone 进行分类，您必须先将您的数据（资产）作为项目库存带到亚马逊 DataZone。为特定项目创建库存，从而仅允许该项目的成员发现资产。

在项目库存中创建资产后，可以整理其元数据。例如，您可以编辑资产的名称、描述或自述文件。每次编辑资产时都会创建资源的新版本。可以使用资产详情页面上的“历史记录”选项卡来查看所有资产版本。

可以编辑自述文件部分，并为资产添加丰富描述。自述文件部分支持 markdown，这可让您根据需要设置描述的格式，并向使用者描述有关资产的关键信息。

可以通过填写可用表单在资产级别添加术语表术语。

要整理架构，您可以查看列，添加企业名称和描述，并在列级别添加术语表术语。

如果在创建数据来源时启用了自动元数据生成，则可以逐个或一次性接受/拒绝资产和列的企业名称。

也可以编辑订阅条款以指定资产是否需要获得批准。

借 DataZone 助 Amazon 中的元数据表单，您可以通过添加自定义属性（例如，销售区域、销售年度和销售季度）来扩展数据资产的元数据模型。附加到某个资产类型的元数据表单将应用于从该资产类型创建的所有资产。您还可以在数据来源运行过程中或创建数据来源后，向单个资产添加其他元数据表单。有关创建新表单的信息，请参阅the section called “创建元数据表单”。

要更新资产的元数据，您必须是该资产所属项目的所有者或贡献者。

更新资产的元数据

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。如果您是亚马逊 DataZone 管理员，则可以导航到亚马逊 DataZone 控制台，https://console.aws.amazon.com/datazone并使用域名创建 AWS 账户 地登录，然后选择打开数据门户。
2. 从顶部导航窗格中选择选择项目，然后选择包含要更新其元数据的资产的项目。
3. 导航到项目的数据选项卡。
4. 从左侧导航窗格中选择库存数据，然后选择要更新其元数据的资产的名称。
5. 在资产详情页面上的元数据表单下，选择编辑并根据需要编辑现有表单。您还可以为资产附加其他元数据表单。有关更多信息，请参阅 the section called “将其他元数据表单附加到资产”。
6. 更新完后，选择保存表单。

   当您保存表单时，Amazon DataZone 会生成该资产的新库存版本。要将更新的版本发布到目录中，请选择Re-publish 资产。

将其他元数据表单附加到资产

默认情况下，附加到某个域的元数据表单将附加到已发布到该域的所有资产。数据发布者可以将其他元数据表单与单个资产关联，从而提供更多上下文信息。

将其他元数据表单附加到资产

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。如果您是亚马逊 DataZone管理员，则可以导航到亚马逊 DataZone 控制台，https://console.aws.amazon.com/datazone并使用域名创建 AWS 账户 地登录，然后选择打开数据门户。
2. 从顶部导航窗格中选择选择项目，然后选择包含要将元数据添加到的资产的项目。
3. 导航到项目的数据选项卡。
4. 从左侧导航窗格中选择库存数据，然后选择要将元数据添加到的资产的名称。
5. 在资产详情页面上的元数据表单下，选择添加表单。
6. 选择要添加到资产的表单，然后选择添加表单。
7. 为每个元数据字段输入值，然后选择保存表单。

   当您保存表单时，Amazon DataZone 会生成该资产的新库存版本。要将更新的版本发布到目录中，请选择Re-publish 资产。

在 Amazon 中进行整理后，将资产发布到目录中 DataZone

一旦对资产管理感到满意，数据所有者就可以将资产版本发布到 Amazon DataZone 目录中，从而让所有域名用户都能发现该版本。资产显示库存版本和已发布的版本。在发现目录中，仅显示最新的已发布版本。如果元数据在发布后进行了更新，则新的库存版本将用于发布到目录。

在 Amazon 中手动创建资产 DataZone

在 Amazon DataZone 中，资产是呈现单个物理数据对象（例如表、控制面板、文件）或虚拟数据对象（例如视图）的实体。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。手动发布资产是一次性操作。您未指定资产的运行时间表，因此资产在其来源发生更改时不会自动更新。

要通过项目手动创建资产，您必须是项目的所有者或贡献者。

手动创建资产

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。
   如果您是亚马逊 DataZone 管理员，则可以导航到亚马逊 DataZone 控制台，https://console.aws.amazon.com/datazone并使用域名创建 AWS 账户 地登录，然后选择打开数据门户。
2. 从顶部导航窗格中选择选择项目，然后选择要为其创建资产的项目。
3. 导航到项目的数据选项卡。
4. 从左侧导航窗格中选择数据来源，然后选择创建数据资产。
5. 对于资产详情，请配置以下设置：
   - 资产类型 – 资产的类型。
   - 名称 – 资产的名称。
   - 描述 – 资产的描述。
6. 对于 S3 位置，输入源 S3 存储桶的 Amazon 资源名称（ARN）。

   （可选）输入 S3 接入点。有关更多信息，请参阅使用 Amazon S3 接入点管理数据访问。
7. 对于发布设置，选择是否可以在目录中立即发现资产。如果您仅将资产添加到库存中，则可以稍后选择订阅条款以将资产发布到目录。
8. 选择创建。

   在创建资产后，将资产作为活跃资产直接发布到目录，或将资产存储在库存中直到您决定发布它为止。

从 Amazon DataZone 目录中取消发布资产

当您从目录中取消发布某项 Amazon DataZone 资产时，该资产将不再出现在全球搜索结果中。新用户将无法在目录中找到或订阅资产清单，而所有现有订阅将保持不变。

要取消发布某个资产，您必须是该资产所属项目的所有者或贡献者：

取消发布资产

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。
   如果您是亚马逊 DataZone 管理员，则可以导航到亚马逊 DataZone 控制台，https://console.aws.amazon.com/datazone并使用域名创建 AWS 账户 地登录，然后选择打开数据门户。
2. 从顶部导航窗格中选择选择项目，然后选择资产所属的项目。
3. 导航到项目的数据选项卡。
4. 从左侧导航窗格中选择已发布的数据。
5. 从已发布的资产列表中找到该资产，然后选择取消发布。

   这将从目录中删除资产。可以随时通过选择发布来重新发布资产。

删除亚马逊 DataZone 资产

如果您不再需要 Amazon 中的某项资产 DataZone，则可以将其永久删除。从目录中删除资产的过程与从目录中取消发布资产的过程不同。您可以从目录中删除某个资产及其相关清单，使其不在显示在任何搜索结果中。要删除资产清单，您必须先撤销其所有订阅。

要删除某个资产，您必须是该资产所属项目的所有者或贡献者：

> **Note**
>
> 要删除资产清单，您必须先撤销该资产的所有现有订阅。您无法删除具有现有订阅用户的资产清单。

删除资产

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。
   如果您是亚马逊 DataZone 管理员，则可以导航到亚马逊 DataZone 控制台，https://console.aws.amazon.com/datazone并使用域名创建 AWS 账户 地登录，然后选择打开数据门户。

2. 从顶部导航窗格中选择选择项目，然后选择包含要删除的资产的项目。
3. 导航到项目的数据选项卡。
4. 从左侧导航窗格中选择已发布的数据，然后找到并选择要删除的资产。这将打开资产详情页面。
5. 依次选择操作和删除，然后确认删除。

   在删除资产后，便无法再查看资产，并且用户无法订阅资产。

手动启动在 Amazon 中运行的数据源 DataZone

当您运行数据源时，Amazon 会从源中 DataZone 提取所有新的或修改过的元数据，并更新库存中的关联资产。向 Amazon 添加数据源时 DataZone，您需要指定该源的运行首选项，该首选项定义了数据源是按计划运行还是按需运行。如果来源按需运行，则必须手动启动数据来源运行。

即使来源按时间表运行，也可以随时手动运行来源。向资产添加业务元数据后，您可以选择资产并将其发布到 Amazon DataZone 目录，以便所有域名用户都能发现这些资产。仅已发布的资产可供其他域用户搜索。

手动运行数据来源

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。
   如果您是亚马逊 DataZone 管理员，则可以导航到亚马逊 DataZone 控制台，https://console.aws.amazon.com/datazone并使用域名创建 AWS 账户 地登录，然后选择打开数据门户。
2. 从顶部导航窗格中选择选择项目，然后选择数据来源所属的项目。
3. 导航到项目的数据选项卡。
4. 从左侧导航窗格中选择数据来源，然后找到并选择要运行的数据来源。这将打开数据来源详细信息页面。
5. 选择按需运行。

   当 Amazon 使用源中的最新数据 DataZone 更新资产元数据时，数据源状态将更改`Running`为。可以在数据来源运行选项卡上监控运行的状态。

Amazon 中的资产修订 DataZone

当您编辑资产的业务或技术元数据时，Amazon 会 DataZone 增加该资产的修订量。这些编辑包括修改资产名称、描述、词汇表、列名、元数据表单和元数据表单字段值。可以通过手动编辑、数据来源作业

运行或 API 操作执行这些更改。每当您对资产进行编辑时，Amazon 都会 DataZone 自动生成新的资产修订。

在更新资产并生成新修订后，您必须将新修订发布到目录以使其更新并可供订阅用户使用。有关更多信息，请参阅 the section called “将项目库存中的资产发布到目录”。只能将资产的最新版本发布到目录。

查看资产的过去修订

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。如果您是亚马逊 DataZone 管理员，则可以导航到亚马逊 DataZone 控制台，https://console.aws.amazon.com/datazone并使用域名创建 AWS 账户 地登录，然后选择打开数据门户。
2. 从顶部导航窗格中选择选择项目，然后选择包含资产的项目。
3. 导航到项目的数据选项卡，然后找到并选择资产。这将打开资产详情页面。
4. 导航到历史记录选项卡，该选项卡显示资产的过去修订的列表。

Amazon 的数据质量 DataZone

Amazon 中的数据质量指标 DataZone 可帮助您了解不同的质量指标，例如数据源的完整性、及时性和准确性。Amazon DataZone 与 AWS Glue 数据质量集成，并提供 API 来集成来自第三方数据质量解决方案的数据质量指标。数据用户可以查看其订阅的资产的数据质量指标随时间变化的情况。要编写和运行数据质量规则，您可以使用自己选择的数据质量工具，例如 AWS Glue 数据质量。借助 Amazon 中的数据质量指标 DataZone，数据使用者可以可视化资产和列的数据质量分数，从而帮助建立对他们用于决策的数据的信任。

Pre-requisites 以及 IAM 角色的更改

如果您使用的是Amazon DataZone 的 AWS 托管策略，则无需执行其他配置步骤，并且这些托管策略会自动更新以支持数据质量。如果您对角色使用自己的策略来授予 Amazon DataZone 与支持的服务互操作所需的权限，则必须更新附加到这些角色的策略，以支持读取中的 AWS Glue 数据质量信息，AWS 管理策略：AmazonDataZoneGlueManageAccessRolePolicy并启用对AWS 管理策略：AmazonDataZoneDomainExecutionRolePolicy和中的时间序列 API 的支持。AWS 管理策略：AmazonDataZoneFullUserAccess

启用数据质量 AWS Glue 资产

亚马逊从 AWS G DataZone lue 中提取数据质量指标是为了提供某一时间点的背景信息，例如在搜索业务数据目录期间。数据用户可以查看其订阅的资产的数据质量指标随时间变化的情况。数据生成者可以按计划获 AWS 取 Glue 数据质量分数。亚马逊 DataZone 企业数据目录还可以通过数据质量 API

显示来自第三方系统的数据质量指标。有关更多信息，请参阅 AWS Glue 数据质量和数据目录的 AWS Glue 数据质量入门。

您可以通过以下方式为您的 Amazon DataZone 资产启用数据质量指标：

- 在创建新的 AWS Glue 数据源或编辑现有 Glue 数据源时，使用数据门户或 Amazon DataZone API 通过亚马逊 DataZone 数据门户启用 AWS Glue 数据源的数据质量。

  有关通过门户为数据来源启用数据质量的更多信息，请参阅为创建并运行 Amazon DataZone 数据源 AWS Glue Data Catalog。

  > **Note**
  >
  > 可以使用数据门户仅为 AWS Glue 库存资产启用数据质量。在此版本的 Amazon 中，不支持通过数据门户为 Amazon Redshift 或自定义类型资产 DataZone 启用数据质量。

  还可以使用 API 为新的或现有的数据来源启用数据质量。为此，您可以调用CreateDataSource或UpdateDataSource并将`autoImportDataQualityResult`参数设置为“True”。

  启用数据质量后，您可以按需或按时间表运行数据来源。每次运行最多可以为每个资产引入 100 个指标。在将数据来源用于数据质量时，无需手动创建表单或添加指标。在发布资产后，对数据质量表单所做的更新（每条历史记录规则最多 30 个数据点）将反映在面向使用者的清单中。随后，向资产添加的每一个新指标都会自动添加到清单中。无需重新发布资产即可向使用者提供最新的分数。

为自定义资产类型启用数据质量

您可以使用 Amazon DataZone API 为您的任何自定义类型资产启用数据质量。有关更多信息，请参阅下列内容：

- PostTimeSeriesDataPoints
- ListTimeSeriesDataPoints
- GetTimeSeriesDataPoint
- DeleteTimeSeriesDataPoints

以下步骤提供了使用 API 或 CLI 在亚马逊中为您的资产导入第三方指标的示例 DataZone：

1. 按如下方式调用 PostTimeSeriesDataPoints API：

```
aws datazone post-time-series-data-points  \
--cli-input-json file://createTimeSeriesPayload.json \
```

具有以下有效载荷：

```
"domainId": "dzd_5oo7xzoqltu8mf",
    "entityId": "4wyh64k2n8czaf",
    "entityType": "ASSET",
    "form": {
        "content": "{\n  \"evaluations\" : [ {\n    \"types\" : [ \"MaximumLength
\" ],\n    \"description\" : \"ColumnLength \\\"ShippingCountry\\\" <= 6\",\n
   \"details\" : { },\n    \"applicableFields\" : [ \"ShippingCountry\" ],\n
   \"status\" : \"PASS\"\n  }, {\n    \"types\" : [ \"MaximumLength\" ],\n
 \"description\" : \"ColumnLength \\\"ShippingState\\\" <= 2\",\n    \"details
\" : { },\n    \"applicableFields\" : [ \"ShippingState\" ],\n    \"status\" :
 \"PASS\"\n  }, {\n    \"types\" : [ \"MaximumLength\" ],\n    \"description
\" : \"ColumnLength \\\"ShippingCity\\\" <= 8\",\n    \"details\" : { },\n
 \"applicableFields\" : [ \"ShippingCity\" ],\n    \"status\" : \"PASS\"\n  },
 {\n    \"types\" : [ \"Completeness\" ],\n    \"description\" : \"Completeness \
\\"ShippingStreet\\\" >= 0.59\",\n    \"details\" : { },\n    \"applicableFields
\" : [ \"ShippingStreet\" ],\n    \"status\" : \"PASS\"\n  }, {\n    \"types\" :
 [ \"MaximumLength\" ],\n    \"description\" : \"ColumnLength \\\"ShippingStreet\\
\" <= 101\",\n    \"details\" : { },\n    \"applicableFields\" : [ \"ShippingStreet
\" ],\n    \"status\" : \"PASS\"\n  }, {\n    \"types\" : [ \"MaximumLength\" ],\n
    \"description\" : \"ColumnLength \\\"BillingCountry\\\" <= 6\",\n    \"details
\" : { },\n    \"applicableFields\" : [ \"BillingCountry\" ],\n    \"status\" :
 \"PASS\"\n  }, {\n    \"types\" : [ \"Completeness\" ],\n    \"description\" :
 \"Completeness \\\"biLlingcountry\\\" >= 0.5\",\n    \"details\" : {\n
 \"EVALUATION_MESSAGE\" : \"Value: 0.26666666666666666 does not meet the constraint
 requirement!\"\n    },\n    \"applicableFields\" : [ \"biLlingcountry\" ],\n
    \"status\" : \"FAIL\"\n  }, {\n    \"types\" : [ \"Completeness\" ],\n
 \"description\" : \"Completeness \\\"Billingstreet\\\" >= 0.5\",\n    \"details
\" : { },\n    \"applicableFields\" : [ \"Billingstreet\" ],\n    \"status\" :
 \"PASS\"\n  } ],\n  \"passingPercentage\" : 88.0,\n  \"evaluationsCount\" : 8\n}",
        "formName": "shortschemaruleset",
        "id": "athp9dyw75gzhj",
        "timestamp": 1.71700477757E9,
        "typeIdentifier": "amazon.datazone.DataQualityResultFormType",
```

```
        "typeRevision": "8"
    },
    "formName": "shortschemaruleset"
}
```

您可以通过调用以下 GetFormType 操作来获取此有效载荷：

```
aws datazone get-form-type --domain-identifier <your_domain_id> --form-type-
identifier amazon.datazone.DataQualityResultFormType --region <domain_region> --
output text --query 'model.smithy'
```

2. 按如下方式调用 DeleteTimeSeriesDataPoints API：

```
aws datazone delete-time-series-data-points\
--domain-identifier dzd_bqqlk3nz21zp2f \
--entity-identifier dzd_bqqlk3nz21zp2f \
--entity-type ASSET \
--form-name rulesET1 \
```

在 Amazon 中使用机器学习和生成人工智能 DataZone

> **Note**
>
> 由 Amazon Bedrock 提供支持： AWS 实现自动滥用检测。由于亚马逊中关于描述功能的人工智能建议 DataZone 是建立在 Amazon Bedrock 之上的，因此用户继承了 Amazon Bedrock 中实施的控制措施，以强制执行安全、安保和负责任地使用人工智能。

在当前版本的 Amazon 中 DataZone，您可以使用 AI 的名称和描述推荐功能来自动发现和编目数据。Amazon 对生成式 AI 的支持可为资产和列 DataZone 创建企业名称和描述。可以使用这些名称和描述为数据添加业务上下文并推荐数据集的分析，这有助于优化数据发现结果。

在 Amazon Bedrock 的大型语言模型的支持下，Amazon 中针对数据资产名称和描述的人工智能建议可 DataZone 帮助您确保您的数据易于理解且易于发现。人工智能建议还提供针对数据集的最相关的分

析应用程序。通过减少手动文档任务并建议适当的数据用法，自动生成的名称和描述可以帮助您提高数据的可信度，最大限度地减少对有用数据的忽视情况，从而加快做出明智的决策。

支持的区域：

在当前的 Amazon DataZone 版本中，以下区域支持 AI 姓名和描述推荐功能：

- 美国东部（弗吉尼亚州北部）
- 美国西部（俄勒冈）
- 亚太地区（东京）
- 欧洲地区（法兰克福）
- 亚太地区（悉尼）
- 加拿大（中部）
- 欧洲地区（伦敦）
- 南美洲（圣保罗）
- 欧洲地区（爱尔兰）
- 亚太地区（新加坡）
- 美国东部（俄亥俄州）
- 亚太地区（首尔）

Amazon DataZone 支持在以下地区生成企业描述。

- 亚太地区（孟买）
- 欧洲地区（巴黎）

Amazon DataZone 支持在以下地区生成企业名称。

- 欧洲地区（斯德哥尔摩）

> **Bedrock 跨区域推理**
>
> 亚马逊 DataZone 利用 Amazon Bedrock 的跨区域推理终端节点为美国东部（俄亥俄州）地区提供建议。所有其他地区都使用区域内端点。

使用 GenAI 的步骤

以下过程介绍如何在 Amazon 中为姓名和描述生成 AI 推荐 DataZone：

- 导航到亚马逊 DataZone 数据门户 URL，然后使用单点登录 (SSO) 或您的 AWS 凭证登录。如果您是亚马逊 DataZone管理员，请导航到亚马逊 DataZone 控制台，https://console.aws.amazon.com/datazone并使用域名创建 AWS 账户 地登录，然后选择打开数据门户。
- 在顶部导航窗格中，选择选择项目，然后选择包含要为其生成人工智能描述建议的资产的项目。

生成业务描述和摘要

- 导航到项目的数据选项卡。
- 在左侧导航窗格中，选择库存数据，然后选择要为其生成人工智能描述建议的资产的名称。
- 在资产详细信息页面上的业务元数据选项卡中，选择生成描述。

生成业务名称

- 导航到项目的数据选项卡。
- 在左侧导航窗格中，选择数据来源，然后选择要为其启用业务名称生成操作的数据来源。
- 转到详细信息选项卡并启用自动生成业务名称配置。
- BusinessNames 也可以在创建资产时通过启用 API 负载中的 PredictionConfiguration 下的业务 NameGeneration 标志，以编程方式生成。CreateAsset

Accepting/Rejecting 预测

- 在生成描述后，您可以编辑、接受或拒绝该描述。
- 每个自动生成的数据资产元数据描述的旁边都会显示绿色图标。在业务元数据选项卡中，您可以选择自动生成的摘要旁边的绿色图标，然后选择编辑、接受或拒绝来处理生成的描述。
- 也可以选择全部接受或全部拒绝选项（在选择业务元数据选项卡时，这两个选项会显示在页面顶部），从而对所有自动生成的描述执行选定操作。
- 或者，您可以选择架构选项卡，然后通过以下方式逐个处理自动生成的描述：一次选择一个列描述的绿色图标，并选择接受或拒绝。
- 在架构选项卡中，也可以选择全部接受或全部拒绝，从而对所有自动生成的描述执行选定操作。

要将资产与生成的描述一起发布到目录，请选择发布资产，然后在发布资产弹出窗口中再次选择发布资产来确认此操作。

> **Note**
>
> 如果您未接受或拒绝为某个资产生成的描述，并随后发布该资产，则此未经审核的自动生成的元数据将不会包含在发布的数据资产中。

对自定义关系资产类型的支持

亚马逊 DataZone 支持自定义资产类型的 GenAI 功能。以前，只有托管的 AWS Glue 和 Amazon Redshift 资产类型支持此功能。

要启用此功能，请创建自己的资产类型定义，并将 `RelationalTableFormType` 作为其中一个表单附加到定义中。Amazon DataZone 会自动检测此类表单的存在，并为这些资产启用 GenAI 功能。生成公司名称（通过 CreateAsset API 中的 PredictionConfiguration）和 BusinessDescription（通过生成描述按钮，点击资产详情页面）的总体体验保持不变。

有关创建自定义资产类型的更多信息，请参阅在 Amazon 中创建自定义资产类型 DataZone。

配额

Amazon DataZone 支持不同的企业名称生成和企业描述生成配额。您可以联系 AWS 支持团队以增加这些配额。

- BusinessDescriptionGeneration: 10K invocations/month
- BusinessNameGeneration: 50K invocations/month

Amazon 中的数据谱系 DataZone

Amazon DataZone 中的数据沿袭 OpenLineage-compatible 功能可以帮助您从 OpenLineage-enabled 系统或通过 API 捕获和可视化世系事件，以跟踪数据来源、跟踪转换和查看跨组织的数据消耗情况。它为您提供了数据资产的总体视图，以便查看资产的来源及其连接链。世系数据包括有关亚马逊 DataZone业务数据目录内活动的信息，包括有关编目资产、这些资产的订阅者以及使用 API 以编程方式捕获的业务数据目录之外发生的活动的信息。

主题

- Amazon 中的血统节点类型 DataZone
- 世系节点中的关键属性
- 可视化数据世系
- Amazon 中的数据沿袭授权 DataZone
- Amazon 中的数据沿袭示例体验 DataZone
- 在管理控制台中启用数据血统
- 以编程方式使用 Amazon DataZone 数据谱系
- 自动执行血统 AWS Glue 目录
- 从 Amazon Redshift 实现血统自动化

可以将血统设置为在添加到亚马逊时自动从 AWS Glue 和 Amazon Redshift 数据库中捕获。DataZone此外，可以将在 AWS Glue（v5.0 及更高版本）控制台中运行的 Spark ETL 作业配置为向亚马逊域发送世系事件。 DataZone

在 Amazon 中 DataZone，域管理员可以在设置数据湖和数据仓库内置蓝图的同时配置世系，从而确保使用这些资源创建的所有数据源运行都启用自动世系捕获。

使用亚马逊 DataZone的 OpenLineage-compatible API，域管理员和数据创建者可以捕获和存储超出亚马逊可用范围的世系事件 DataZone，包括 Amazon S3、Glue 和其他 AWS 服务中的转换。这为数据使用者提供了全面视图，帮助他们自信地了解资产来源，同时数据创建者可以通过了解资产的使用情况来评估资产更改产生的影响。此外，Amazon DataZone 版本与每个事件保持一致，使用户能够在任何时间点可视化血统或比较资产或任务历史的转换。此历史世系可让用户更深入地了解数据的演变过程，这对于故障排除、审计和确保数据资产的完整性至关重要。

借助数据沿袭，您可以在 Amazon DataZone 中完成以下任务：

- 了解数据的来源：了解数据源自何处可让您清楚地了解数据的源、依赖关系和转换，从而增强对数据的信任。此透明度有助于自信地做出数据驱动型决策。
- 了解数据管道更改产生的影响：在对数据管道进行更改时，可以使用世系功能来标识所有将受影响的下游使用者。这有助于确保在不中断关键数据流的情况下进行更改。
- 确定数据质量问题的根本原因：如果在下游报告中检测到数据质量问题，则可以使用世系（尤其是列级世系）来追溯数据（在列级别），以将问题追溯到其源。这可帮助数据工程师识别和修复问题。
- 改善数据治理和合规性：可使用列级世系来演示对数据治理和隐私法规的遵从性。例如，可使用列级世系来显示敏感数据（例如 PII）的存储位置以及下游活动中处理敏感数据的方式。

Amazon 中的血统节点类型 DataZone

在 Amazon 中 DataZone，数据谱系信息显示在代表表和视图的节点中。根据项目（例如，在数据门户左上角选择的项目）的上下文，创建者可以同时查看库存资产和已发布的资产，而使用者只能查看已发布的资产。首次在资产详细信息页面中打开世系选项卡时，已编目的数据集节点是通过世系图的世系节点向上游或下游导航的起点。

以下是 Amazon DataZone 支持的数据血统节点类型：

- 数据集节点 – 此节点类型包括有关特定数据资产的数据世系信息。
  - 包含亚马逊 DataZone 目录中发布的 AWS Glue 或 Amazon Redshift 资产相关信息的数据集节点是自动生成的，节点中包含相应的 G AWS lue 或 Amazon Redshift 图标。
  - 包含未在 Amazon DataZone 目录中发布的资产信息的数据集节点由域管理员（制作者）手动创建，并由节点内的默认自定义资产图标表示。
- 作业（运行）节点 – 此节点类型显示作业的详细信息，包括特定作业的最新运行和运行详细信息。此节点还捕获作业的多次运行，并且可在节点详细信息的历史记录选项卡中查看。您可以通过选择节点图标来查看节点详细信息。

世系节点中的关键属性

世系节点中的 `sourceIdentifier` 属性表示数据集上发生的事件。谱系节点`sourceIdentifier`的就是数据集的标识符（table/view 等）。它用于在世系节点上强制实施唯一性。例如，不能有两个具有同一 `sourceIdentifier` 的世系节点。以下是不同类型的节点的 `sourceIdentifier` 值的示例：

- 对于具有相应数据集类型的数据集节点：
  - 资产：amazon.datazone.asset/<assetId>
  - 清单（已发布的资产）：amazon.datazone.listing/<listingId>
  - AWS <region><account-id><database>Glue table：arn: aws: glue:: table//<table-name>
  - <database><schema>Amazon Redshift table/view：arn: aws: < redshift/redshift-serverless>: <region><account-id>: <table-type (等) >/< >//table/viewclusterIdentifier/workgroupName<table-name>
  - 对于使用 open-lineage 运行事件导入的任何其他类型的数据集节点，将<namespace><name>使用该 input/output数据集`sourceIdentifier`的/。
- 对于作业：
  - 对于使用 open-lineage 运行事件导入的作业节点，将 <jobs_namespace>.<job_name> 用作 sourceIdentifier。

- 对于作业运行：
  - 对于使用 open-lineage 运行事件导入的作业运行节点，将 <jobs_namespace>.<job_name>/<run_id> 用作 sourceIdentifier。

对于使用 `createAsset` API 创建的资产，必须使用 `createAssetRevision` API 更新 `sourceIdentifier` 以便能够将资产映射到上游资源。

可视化数据世系

Amazon DataZone 的资产详情页面以图形方式呈现数据谱系，便于直观呈现上游或下游的数据关系。资产详细信息页面提供以下功能来浏览图表：

- Column-level lineage：如果在数据集节点中可用，则扩展列级谱系。如果源列信息可用，这将自动显示与上游或下游数据集节点的关系。
- 列搜索：当列数的默认显示为 10 时。如果超过 10 个列，则将激活分页以导航到其余列。要快速查看特定列，可以在仅列出已搜索列的数据集节点上进行搜索。
- 仅查看数据集节点：如果要切换为仅查看数据集世系节点并筛选出作业节点，您可以选择图表查看器左上角的“打开视图控件”图标，然后切换仅显示数据集节点选项。这将从图表中删除所有作业节点，并让您仅浏览数据集节点。请注意，在启用“仅查看数据集节点”时，图表无法向上游或下游展开。
- 详细信息窗格：每个世系节点都具有捕获到的详细信息，并且会在选中时显示。
  - 数据集节点具有一个详细信息窗格，其中显示针对给定时间戳为该节点捕获的所有详细信息。每个数据集节点具有 3 个选项卡，即：“世系信息”、“架构”和“历史记录”选项卡。“历史记录”选项卡列出为该节点捕获的世系事件的不同版本。从 API 捕获的所有详细信息都使用元数据表单或 JSON 查看器显示。
  - 作业节点具有一个详细信息窗格，其中显示作业详细信息与“作业信息”和“历史记录”选项卡。详细信息窗格还捕获在作业运行过程中捕获到的查询或表达式。“历史记录”选项卡列出为该节点捕获的作业运行的不同版本。从 API 捕获的所有详细信息都使用元数据表单或 JSON 查看器显示。
- 版本选项卡：Amazon DataZone 数据谱系中的所有世系节点都有版本控制。对于每个数据集节点或作业节点，版本都将作为历史记录捕获，这使您能够在各个版本之间导航以确定随时间推移发生的变化。每个版本都会在世系页面中打开一个新的选项卡以帮助进行比较。

Amazon 中的数据沿袭授权 DataZone

写入权限-要将世系数据发布到 Amazon DataZone，您必须拥有一个 IAM 角色，其权限策略包括对 `PostLineageEvent` API 的`ALLOW`操作。此 IAM 授权在 API Gateway 层进行。

读取权限-有两个操作：GetLineageNode和ListLineageNodeHistory包含在AmazonDataZoneDomainExecutionRolePolicy托管策略中，因此 Amazon DataZone 域中的每个用户都可以调用这些操作来遍历数据谱系图。

Amazon 中的数据沿袭示例体验 DataZone

您可以使用数据沿袭示例体验来浏览和了解 Amazon 中的数据谱系 DataZone，包括在数据谱系图中遍历上游或下游、探索版本和列级谱系。

完成以下步骤，在 Amazon 中试用示例数据谱系体验： DataZone

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。如果您是亚马逊 DataZone管理员，则可以导航到亚马逊 DataZone 控制台，https://console.aws.amazon.com/datazone并使用域名创建 AWS 账户 地登录，然后选择打开数据门户。
2. 选择任何可用的数据资产以打开资产的详细信息页面。
3. 在资产的详细信息页面上，选择血统选项卡，将鼠标悬停在信息图标上，然后选择尝试示例血统。
4. 在数据世系弹出窗口中，选择开始引导式数据世系旅程。

   此时，将显示一个全屏选项卡，其中提供了世系信息的所有空间。示例数据世系图表最初在上游和下游两端显示一个深度为 1 的基本节点。您可以将图表扩展到上游或下游。您也可以选择列信息，并查看世系如何流经节点。

在管理控制台中启用数据血统

您可以在配置默认数据湖和默认数据仓库蓝图的过程中启用数据血统。

完成以下过程为默认数据湖蓝图启用数据血统。

1. 导航至 Amazon DataZone 控制台 https://console.aws.amazon.com/datazone，并使用您的账户凭证登录。
2. 选择 “查看域”，然后选择要为 DefaultDataLake 蓝图启用数据沿袭的域。
3. 在域详细信息页面上，导航到蓝图选项卡。
4. 在 DefaultDataLake 蓝图的详细信息页面上，选择区域选项卡。
5. 在为 DefaultDataLake 蓝图添加区域的过程中，您可以启用数据沿袭。因此，如果已经添加了一个区域，但其中的数据血统功能未启用（导入数据血统列中显示了否），则必须先删除该区域。要启用数据血统，请选择添加区域，然后选择要添加的区域，并确保在添加区域弹出窗口中选中启用导入数据血统复选框。

要为 DefaultDataWarehouse 蓝图启用数据沿袭，请完成以下步骤。

1. 导航至 Amazon DataZone 控制台 https://console.aws.amazon.com/datazone，并使用您的账户凭证登录。
2. 选择 “查看域”，然后选择要为 DefaultDataWarehouse 蓝图启用数据沿袭的域。
3. 在域详细信息页面上，导航到蓝图选项卡。
4. 在 DefaultDataWarehouse 蓝图的详细信息页面上，选择参数集选项卡。
5. 在为 DefaultDataWarehouse 蓝图添加参数集的过程中，您可以启用数据沿袭。为此，请选择创建参数集。
6. 在创建参数集页面上，指定以下内容，然后选择创建参数集。
   - 参数集的名称。
   - 参数集的描述。
   - AWS 您要在其中创建环境的区域。
   - 指定亚马逊 DataZone 是使用这些参数来建立与您的 Amazon Redshift 集群还是无服务器工作组的连接。
   - 指定密 AWS 钥。
   - 指定要在创建环境时使用的集群或无服务器工作组。
   - 指定要在创建环境时使用的数据库（在指定的集群或工作组中）的名称。
   - 在导入数据血统下，选中启用导入数据血统。

以编程方式使用 Amazon DataZone 数据谱系

要在 Amazon 中使用数据血统功能 DataZone，您可以调用以下 API：

- GetLineageNode
- ListLineageNodeHistory
- PostLineageEvent

自动执行血统 AWS Glue 目录

当 AWS Glue 数据库和表被添加到 Amazon DataZone 目录时，将使用数据源运行自动提取这些表的世系。对于此来源，有几种方法可以实现血统自动化：

- 蓝图配置 – 设置蓝图的管理员可以将蓝图配置为自动捕获血统。通过此配置，管理员能够定义血统捕获的关键数据来源，而不需要依赖数据生成者对数据进行编目。有关更多信息，请参阅 在管理控制台中启用数据血统。
- 数据源配置-数据生成者在为 AWS Glue 数据库配置数据源运行时，会看到一个视图以及数据质量，用于告知该数据源的自动数据沿袭。
  - 血统设置可以在数据来源定义选项卡中查看。数据生成者无法编辑此值。
  - Data Source 运行中的世系集合从表元数据中获取信息以建立世系。 AWS Glue crawler 支持不同类型的来源，在数据源运行中捕获血统的来源包括 Amazon S3、DynamoDB、Catalog、Delta Lake、Iceberg 表和存储在 Amazon S3 中的 Hudi 表。目前不支持 JDBC 和 DocumentDB 或 MongoDB 作为来源。
  - 限制 – 如果表的数量大于 100 个，血统运行将在 100 个表之后失败。确保 Gl AWS ue 爬网程序未配置为在一次运行中引入超过 100 张表。
- AWS Glue (v5.0) 配置——在 AWS Glue Studio 中运行 AWS Glue 作业时，可以为任务配置数据沿袭，将世系事件直接发送到亚马逊网域。 DataZone

  1. 导航到 AWS Glue 控制台https://console.aws.amazon.com/gluestudio，然后使用您的账户凭据登录。
  2. 选择 ETL 作业，然后创建新作业或单击任何现有作业。
  3. 转到作业详细信息（包括 ETL 流程作业）选项卡，然后向下滚动到“生成血统事件”部分。
  4. 选中该复选框可启用发送世系事件，该复选框会展开显示输入字段以输入 Amazon DataZone 域名 ID。
- AWS Glue (V5.0) 笔记本配置——在笔记本中，你可以通过添加 %%configure 魔法来自动收集 Spark 执行任务。此配置会将事件发送到 Amazon DataZone 域。

```
%%configure --name project.spark -f
{
    "--
conf":"spark.extraListeners=io.openlineage.spark.agent.OpenLineageSparkListener
 --conf spark.openlineage.transport.type=amazon_datazone_api --
conf spark.openlineage.transport.domainId={DOMAIN_ID}  --conf
 spark.glue.accountId={ACCOUNT_ID} --conf
 spark.openlineage.facets.custom_environment_variables=[AWS_DEFAULT_REGION;GLUE_VERSION;GLUE_
 --conf spark.glue.JOB_NAME={JOB_NAME}"
}
```

下面是参数详细信息：

- `spark.extraListeners=io.openlineage.spark.agent.OpenLineageSparkListener`-OpenLineageSparkListener 将在 Spark 的监听器总线中创建并注册
- `spark.openlineage.transport.type=amazon_datazone_api`-这是一项 OpenLineage 规范，用于告诉 OpenLineage插件使用 DataZone API 传输向的 API 发送世系事件。 DataZone PostLineageEvent 有关更多信息，请参阅 https://openlineage.io/docs/integrations/spark/configuration/spark_conf。
- `spark.openlineage.transport.domainId={DOMAIN_ID}` – 此参数用于建立一个域，API 传输将向该域提交血统事件。
- `spark.openlineage.facets.custom_environment_variables [AWS_DEFAULT_REGION;GLUE_VERSION;GLUE_COMMAND_CRITERIA;GLUE_PYTHON_VERSION;]`-Glue 交互式会话填充的以下环境变量（AWS_DEFAULT_REGION GLUE_VERSION、GLUE_COMMAND_CRITERIA 和 GLUE_PYTHON_VERSION）将被添加到 LineageEvent
- `spark.glue.accountId=<ACCOUNT_ID>` – 元数据所在的 Glue Data Catalog 的账户 ID。此账户 ID 用于在血统事件中构建 Glue ARN。
- `spark.glue.JOB_NAME` – 血统事件的作业名称。Notebook 中的作业名称可以设置为 `spark.glue.JOB_NAME: ${projectId}.${pathToNotebook}`。

- 设置参数以配置 DataZone 从 AWS Glue 与亚马逊的通信

  参数键：--conf

  参数值：

```
spark.extraListeners=io.openlineage.spark.agent.OpenLineageSparkListener
--conf spark.openlineage.transport.type=amazon_datazone_api
--conf spark.openlineage.transport.domainId=<DOMAIN_ID>
--conf
 spark.openlineage.facets.custom_environment_variables=[AWS_DEFAULT_REGION;GLUE_VERSION;GLUE_
--conf spark.glue.accountId=<ACCOUNT_ID> (replace <DOMAIN_ID> and <ACCOUNT_ID> with
 the right values)
```

  对于 Notebook，请添加以下附加参数：

```
--conf spark.glue.JobName=<SessionId> --conf spark.glue.JobRunId=<SessionId or NONE?>
replace <SessionId> and <SessionId> with the right values
```

从 Amazon Redshift 实现血统自动化

通过管理员设置的数据仓库蓝图配置，从 Amazon Redshift 服务中捕获世系，亚马逊会自动捕获世系。 DataZone世系运行会捕获为给定数据库执行的查询，并生成要存储在 Amazon 中的世系事件，DataZone 以便数据生成者或使用者在访问特定资产时进行可视化。

可以使用以下配置实现血统自动化：

- 蓝图配置 – 设置蓝图的管理员可以将蓝图配置为自动捕获血统。通过此配置，管理员能够定义血统捕获的关键数据来源，而不需要依赖数据生成者对数据进行编目。要进行设置，请转到在管理控制台中启用数据血统。
- 数据来源配置：数据生成者在为 Amazon Redshift 数据库配置数据来源运行时会看到该数据来源的自动数据血统设置。

  血统设置可以在数据来源定义选项卡中查看。数据生成者无法编辑此值。

针对发布的元数据强制规则

在 Amazon DataZone 上发布的元数据执行规则使域单位所有者能够为数据制作者制定明确的元数据要求，简化访问请求并增强数据治理，从而加强数据治理。

目前有 Amazon 的所有 AWS 商业区域都支持 DataZone 该功能。

域单位所有者可以完成以下程序，在 Amazon 中配置元数据强制执行 DataZone：

1. 使用 DataZone 数据门户 URL 导航至 Amazon 数据门户，然后使用您的 SSO 或 AWS 凭证登录。如果您是亚马逊 DataZone 管理员，则可以通过访问创建亚马逊 DataZone 域名的 AWS 账户 https://console.aws.amazon.com/datazone中的亚马逊 DataZone 控制台来获取数据门户 URL。
2. 选择域，导航到域单元选项卡，然后选择要使用的域单元。
3. 选择规则选项卡，然后选择添加。
4. 在创建必需的元数据表单规则页面上，执行以下操作，然后选择添加规则：

   - 为规则指定名称。

- 在操作下，选择数据资产和产品发布。
- 在必填表单下，选择添加元数据表单，在域/域单元内选择要添加到此规则的元数据表单，然后选择添加。最多可为每个规则添加 5 个元数据。
- 在范围下，指定要与这些表单关联的数据实体。您可以选择数据产品 and/or 数据资产。
- 在数据资产类型下，指定该规则是适用于所有资产类型还是选定的资产类型。
- 在 “项目” 下，指定所需的表单是与所有项目发布的数据产品 and/or 资产相关联，还是仅与该域单元中的选定项目相关联。此外，如果您希望子域单元继承此要求，请选中将规则级联到子域单元。

在亚马逊中将 Snowflake 作为数据源进行连接 DataZone

亚马逊 DataZone 支持 Snowflake 作为第三方数据源。将 Snowflake 连接到亚马逊 DataZone 项目后，该目录可以索引您的 Snowflake 数据库和架构，从 Snowflake 的查询历史记录中捕获列级谱系，并允许您通过联合来查询 Snowflake 表。本页将引导您使用 CLI AWS 创建 Snowflake 连接。

先决条件

在开始之前，请确认以下内容已准备就绪：

- 一个 Amazon DataZone 域名和一个你想要将 Snowflake 数据编入目录的项目。请参阅Amazon 中的域名和用户访问权限 DataZone和Amazon DataZone 项目和环境。
- 一个 Snowflake 帐户，具有在要编目的数据库、架构和仓库中创建角色和授予权限的权限。
- 使用与您 AWS Secrets Manager 的 Amazon DataZone 域名相同的 AWS 账户进行访问。
- 如果您打算从中查询 Snowflake 数据，则可以使用 Amazon S3 存储桶进行查询泄露。

您将按以下顺序执行以下步骤：创建 Snowflake 角色，将 Snowflake 凭证存储在中， AWS Secrets Manager找到您的亚马逊 DataZone 项目环境 ID，然后致电。`aws datazone create-connection`

第 1 步：为亚马逊创建 Snowflake 角色 DataZone

亚马逊使用您创建并向其授予特定权限的角色 DataZone 连接到 Snowflake。角色授权分为四类：

- 元数据同步 — Amazon DataZone 读取您的表和视图的结构以填充目录。
- 血统同步 — 亚马逊 DataZone 读取 Snowflake 的 QUERY_HISTORY 和 ACCESS_HISTORY 来推导列级血统。

- 查询权限 — 通过 Amazon DataZone 发出的查询（包括来自）的查询是必需的。
- 仓库使用情况-角色执行的任何查询都需要使用。

在 Snowflake 中运行以下语句，将占位符替换为你的值：

```
CREATE ROLE <role_name>;

-- Metadata: database and schema access
GRANT USAGE ON DATABASE <database_name> TO ROLE <role_name>;
GRANT USAGE ON SCHEMA <database_name>.<schema_name> TO ROLE <role_name>;

-- Metadata: table and view structure (no data access)
GRANT REFERENCES ON ALL TABLES IN SCHEMA <database_name>.<schema_name> TO ROLE
 <role_name>;
GRANT REFERENCES ON ALL VIEWS IN SCHEMA <database_name>.<schema_name> TO ROLE
 <role_name>;
GRANT REFERENCES ON FUTURE TABLES IN SCHEMA <database_name>.<schema_name> TO ROLE
 <role_name>;
GRANT REFERENCES ON FUTURE VIEWS IN SCHEMA <database_name>.<schema_name> TO ROLE
 <role_name>;

-- Lineage: access to QUERY_HISTORY and ACCESS_HISTORY
GRANT IMPORTED PRIVILEGES ON DATABASE SNOWFLAKE TO ROLE <role_name>;

-- Query access
GRANT SELECT ON ALL TABLES IN SCHEMA <database_name>.<schema_name> TO ROLE <role_name>;

-- Optional: include tables created after the role is granted
GRANT SELECT ON FUTURE TABLES IN SCHEMA <database_name>.<schema_name> TO ROLE
 <role_name>;

-- Warehouse access
GRANT USAGE ON WAREHOUSE <warehouse_name> TO ROLE <role_name>;

-- Assign the role to the Snowflake user whose credentials DataZone will use
GRANT ROLE <role_name> TO USER <snowflake_username>;
```

Important

GRANT IMPORTED PRIVILEGES ON DATABASE SNOWFLAKE是血统所必需的。如果没有它，元数据同步就会成功，但不会捕获任何世系记录。

第 2 步：将 Snowflake 凭据存储在 AWS Secrets Manager

在 AWS Secrets Manager与您的 Amazon DataZone 域名相同的 AWS 账户中创建密钥。Amazon DataZone 支持两种身份验证方法：用户名和密码或密钥对。

要进行用户名和密码身份验证，请使用以下 JSON 正文：

```
{
  "username": "<snowflake_username>",
  "password": "<snowflake_password>"
}
```

要进行密钥对身份验证，请使用以下 JSON 正文：

```
{
  "sfUser": "<snowflake_username>",
  "pem_private_key": "-----BEGIN PRIVATE KEY-----\n...\n-----END PRIVATE KEY-----"
}
```

给秘密加标签，这样 Amazon DataZone 就可以发现它。这两个标签都是必需的：

标签键	值
AmazonDataZoneDomain	您的亚马逊 DataZone 域名
AmazonDataZoneProject	您的亚马逊 DataZone 项目 ID

请注意密钥的 ARN。您将在步骤 4 `create-connection` 中将其传递给。

第 3 步：查找您的项目环境 ID

`create-connection`API 调用需要连接所在的项目环境的 ID。你可以从项目用户角色 ARN 中推导出来。

项目用户角色 ARN 遵循以下模式：

```
arn:aws:iam::<account_id>:role/datazone_usr_role_<project_id>_<environment_id>
```

最后一个下划线之后的最后一段是环境 ID。例如，在：

```
arn:aws:iam::123456789012:role/datazone_usr_role_a1b2c3dfadsf_58cjkfdsjfd
```

58cjkfdsjfd 是环境 ID。这是要传递的值，如步骤 4 --environment-identifier 所示。环境 ID 对应于您项目的工具环境。

步骤 4：创建 Snowflake 连接

运行以下 AWS CLI 命令，用您的值替换占位符：

```
aws datazone create-connection \
    --domain-identifier <domain_id> \
    --environment-identifier <environment_id> \
    --name "<connection_name>" \
    --description "<optional_description>" \
    --props '{
      "snowflakeProperties": {
        "connectivityProperties": {
          "connectionProperties": {
            "HOST": "<account>.snowflakecomputing.com",
            "PORT": "443",
            "DATABASE": "<database_name>",
            "WAREHOUSE": "<warehouse_name>",
            "SCHEMA": "<schema_name>",
            "ROLE_ARN": "<project_user_role_arn>",
            "CATALOG_CASING_FILTER": "UPPERCASE_ONLY"
          },
          "athenaProperties": {
            "spill_bucket": "<project_s3_bucket_name>",
            "spill_prefix": "<domain_id>/<project_id>/dev/sys/athena"
          },
          "authenticationConfiguration": {
            "authenticationType": "BASIC",
            "secretArn": "<secret_arn_from_step_2>"
          }
        },
        "snowflakeRole": "<role_name_from_step_1>",
        "identityMapping": {
          "usernameAttribute": "EMAIL"
        }
      }
```

```
    }'
```

连接属性参考

属性	说明
HOST	你的 Snowflake 账号网址。
PORT	Snowflake 端口。使用 443。
DATABASE	要编目的 Snowflake 数据库。
WAREHOUSE	亚马逊 DataZone 用来运行查询的 Snowflake 仓库。
SCHEMA	要编目的 Snowflake 架构。
ROLE_ARN	项目用户角色 ARN（参见步骤 3）。
CATALOG_CASING_FILTER	控制已编目标识符的大小写标准化。有效值为 UPPERCASE_ONLY 和 LOWERCASE_ONLY 。如果省略，则默认为UPPERCASE_ONLY 。
spill_bucket	查询泄漏的 S3 存储桶名称。从中查询 Snowflake 时为必填项。
spill_prefix	S3 泄漏前缀，格式为<domain_id>/<project_id>/dev/sys/athena 。
authenticationType	BASIC用于 username/password 或密钥对身份验证。
secretArn	在步骤 2 中创建的密钥的 ARN。
snowflakeRole	步骤 1 中的 Snowflake 角色名称。
identityMapping.usernameAttribute	设置为 EMAIL。

验证连接

`aws datazone get-connection`使用您的域名 ID 和返回的连接标识符进行呼叫`create-connection`。首次运行元数据同步后，已编目的 Snowflake 表将出现在该项目的亚马逊 DataZone 目录中。

启用 Snowflake 血统 AWS Glue Spark 作业

亚马逊 DataZone 可以从读取 Snowflake 数据源或写入 Snowflake 数据源的 AWS Glue Spark 作业中捕获列级谱系。当您在 AWS Glue 作业上启用世系时， AWS Glue Spark 运行时会在任务运行时发出世系事件，Amazon SageMaker Unified Studio 会在世系视图中呈现这些事件。

AWS Glue的 Spark 运行时已经包含了它用来发送事件的 OpenLineage Spark 和 Amazon DataZone 集成。唯一的额外步骤是将工作中的 OpenLineage 核心客户端 (`openlineage-java`) 升级到版本`1.49.0`。

先决条件

在 AWS Glue Spark 作业上启用世系之前，请确认以下内容：

- 已经配置了具有 Snowflake 连接的 Amazon DataZone 域名和项目。请参阅在亚马逊中将 Snowflake 作为数据源进行连接 DataZone。
- 从连接的 Snowflake 源读取或写入的 AWS Glue Spark 作业。
- 修改 AWS Glue 作业配置的权限。
- AWS Glue 任务可以读取的 Amazon S3 存储桶。

以下步骤描述了如何在 AWS Glue 工作中升级 OpenLineage 核心客户端并启用世系事件生成。

步骤 1：在上升级 OpenLineage 核心客户端 AWS Glue 作业

AWS Glue Spark 运行时将 OpenLineage Spark 集成和亚马逊 DataZone集成捆绑在一起。要使其与 Amazon 兼容 DataZone，请将运行中的`openlineage-java`客户端升级到版本`1.49.0`。

1. `openlineage-java-1.49.0.jar`从 Maven Central 下载。请使用以下任一命令：
   - Maven 存储库页面。选择要下载的 jar。
   - 直接的 Maven 网址。
2. 将 JAR 上传到您的 AWS Glue 任务可以读取的 Amazon S3 存储桶。

3. 在 AWS Glue 控制台中，打开任务并选择 Job 详情。在 “高级属性” 下，找到 “依赖 JAR” 路径并添加已上传 JAR 的 S3 URI。
4. 仍然在 Job 详细信息中，添加以下 job 参数：

键	值
`--user-jars-first`	`true`

这会让 AWS Glue Spark 运行时加载你上传的`openlineage-java`客户端，而不是与 AWS Glue之捆绑的版本。

第 2 步：在上启用世系事件生成 AWS Glue 作业

1. 在 AWS Glue 控制台中，打开任务并选择 Job 详情 → 基本属性。
2. 选择 “生成世系事件”。
3. 在显示的字段中输入您的 Amazon DataZone 域名 ID。使用与创建 Snowflake 连接时相同的域 ID。
4. 保存作业。

验证血统是否被捕获

运行作 AWS Glue 业。任务完成后，在 Amazon SageMaker Unified Studio 中打开项目中目标表的世系视图。任务读取的 Snowflake 源的血统记录在图表中显示为上游节点。

亚马逊 DataZone 数据产品

Amazon DataZone 使数据生成者能够将数据资产分组为定义明确、自成一体的包，称为数据产品，专为特定业务用例量身定制。使用有凝聚力的、与业务协调一致的数据产品可以增强发布和订阅流程。数据使用者可将互连数据资产作为一个单元来进行搜索和查找，从而轻松识别这些资产。这种方法可以减少查找所有相关信息所需的时间和工作量，并降低丢失重要数据的风险。此外，数据产品还实施了统一的访问模型，通过单个请求即可轻松访问数据。这使得不再需要多个权限，从而更快地开始分析数据。此外，通过将资产编目为数据产品，数据创建者可在数据产品级别而不是单独启用元数据和访问控制管理，从而减少管理开销。另外，通过提供这些专用的已分组资产以供使用，可提高访问治理和数据利用的效率，确保资产与业务目标保持一致，且易于访问以用于预期用途。数据治理团队可以监控这些数据产品的使用率，并提供有用的见解来提升数据素养成熟度。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。

主题

- 在 Amazon 上创建新的数据产品 DataZone
- 在 Amazon 上发布数据产品 DataZone
- 在 Amazon 中编辑数据产品 DataZone
- 在 Amazon 上取消发布数据产品 DataZone
- 在 Amazon 中删除数据产品 DataZone
- 在 Amazon 上订阅数据产品 DataZone
- 在 Amazon 中查看订阅申请并授予对数据产品的订阅 DataZone
- 在 Amazon 上重新发布数据产品 DataZone

在 Amazon 上创建新的数据产品 DataZone

Amazon DataZone 使数据生成者能够将数据资产分组为定义明确、自成一体的包，称为数据产品，专为特定业务用例量身定制。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。

要创建数据产品，您必须是项目的所有者或贡献者。

要创建新的数据产品，请完成以下步骤。

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。如果您是亚马逊 DataZone 管理员，则可以通过 https://console.aws.amazon.com/datazon e 导航到亚马逊 DataZone 控制台，使用域名创建 AWS 账户 地登录，然后选择打开数据门户。

2. 在 Amazon DataZone 数据门户中，选择您要在其中创建数据产品的项目。
3. 选择数据选项卡，再选择库存数据，然后选择创建新的数据产品。
4. 在创建新的数据产品页面中，指定数据产品的名称和描述，然后选择选择资产向数据产品添加各种资产。在选择资产弹出窗口中，选择要添加到此数据产品的资产，然后选择选择。要完成数据产品的创建，请选择创建。

在 Amazon 上发布数据产品 DataZone

Amazon DataZone 使数据生成者能够将数据资产分组为定义明确、自成一体的包，称为数据产品，专为特定业务用例量身定制。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。

要发布数据产品，您必须是项目的所有者或贡献者。针对发布功能的元数据强制规则可以进行配置，以便为数据生成者制定明确的元数据要求，从而限制数据产品的发布时间。

要发布数据产品，请完成以下步骤。

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。如果您是亚马逊 DataZone 管理员，则可以通过 https://console.aws.amazon.com/datazon e 导航到亚马逊 DataZone 控制台，使用域名创建 AWS 账户 地登录，然后选择打开数据门户。
2. 在 Amazon DataZone 数据门户中，选择要发布的数据产品所在的项目。
3. 选择数据选项卡，再选择库存数据，然后选择数据产品筛选条件。这将显示所有未发布的现有数据产品。
4. 选择要发布的数据产品，然后选择发布。通过选择发布数据产品来确认发布此数据产品。

> **Note**
>
> 此数据产品中的任何未发布的数据资产都将被发布，但只能通过此数据产品获得这些资产。

在 Amazon 中编辑数据产品 DataZone

Amazon DataZone 使数据生成者能够将数据资产分组为定义明确、自成一体的包，称为数据产品，专为特定业务用例量身定制。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。

要编辑某个数据产品，您必须是该数据所属项目的所有者或贡献者。

要编辑数据产品，请完成以下步骤。

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。如果您是亚马逊 DataZone 管理员，则可以通过 https://console.aws.amazon.com/datazon e 导航到亚马逊 DataZone 控制台，使用域名创建 AWS 账户 地登录，然后选择打开数据门户。
2. 在 Amazon DataZone 数据门户中，选择要发布的数据产品所在的项目。
3. 选择数据选项卡，再选择库存数据或已发布的数据，然后选择数据产品筛选条件。
4. 选择要编辑的数据产品。在编辑数据产品的过程中，您可以执行以下操作：
   - 选择创建自述文件以添加自述文件，这将帮助用户更好地理解此页面内容。
   - 选择添加术语以添加术语表术语。在窗口中选择术语表术语，然后选择添加术语。
   - 选择添加元数据表单，再在添加元数据表单窗口中选择您的表单，然后选择添加。
   - 展开操作，选择编辑，对数据产品的名称和描述进行编辑，然后选择更新。

在 Amazon 上取消发布数据产品 DataZone

Amazon DataZone 使数据生成者能够将数据资产分组为定义明确、自成一体的包，称为数据产品，专为特定业务用例量身定制。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。

要取消发布某个数据产品，您必须是该数据所属项目的所有者或贡献者。

要取消发布数据产品，请完成以下步骤。

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。如果您是亚马逊 DataZone 管理员，则可以通过 https://console.aws.amazon.com/datazon e 导航到亚马逊 DataZone 控制台，使用域名创建 AWS 账户 地登录，然后选择打开数据门户。
2. 在 Amazon DataZone 数据门户中，选择要取消发布的数据产品所在的项目。
3. 选择数据选项卡，再选择库存数据或已发布的数据，然后选择数据产品筛选条件。这将显示所有现有的数据产品。
4. 选择要取消发布的数据产品，然后展开操作并选择取消发布。通过选择取消发布来确认取消发布此数据产品。

> **Note**
>
> 取消发布数据产品会产生以下影响：
>
> - 此数据产品将不再可供查看或订阅。

- 任何只能通过此数据产品获取的数据资产将不再可用。
- 此数据产品的所有有效订阅将保留。
- 任何单独发布的数据资产将不受影响。

在 Amazon 中删除数据产品 DataZone

Amazon DataZone 使数据生成者能够将数据资产分组为定义明确、自成一体的包，称为数据产品，专为特定业务用例量身定制。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。

要删除某个数据产品，您必须是该数据所属项目的所有者或贡献者。

要删除数据产品，请完成以下步骤。

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。如果您是亚马逊 DataZone 管理员，则可以通过 https://console.aws.amazon.com/datazon e 导航到亚马逊 DataZone 控制台，使用域名创建 AWS 账户 地登录，然后选择打开数据门户。
2. 在 Amazon DataZone 数据门户中，选择要删除的数据产品所在的项目。
3. 选择数据选项卡，再选择库存数据或已发布的数据，然后选择数据产品筛选条件。这将显示所有现有的数据产品。
4. 选择要删除的数据产品，然后展开操作并选择删除。通过在文本字段中键入 `delete`，然后选择删除来确认删除此数据产品。

> **Note**
>
> 删除数据产品会产生以下影响：
>
> - 此数据产品将不再可供发布、查看或订阅。
> - 任何只能通过此数据产品获取的数据资产将不再显示在数据目录中。不会从库存资产中删除数据产品。

在 Amazon 上订阅数据产品 DataZone

Amazon DataZone 使数据生成者能够将数据资产分组为定义明确、自成一体的包，称为数据产品，专为特定业务用例量身定制。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。

任何具有访问数据门户所需权限的亚马逊 DataZone 用户都可以订阅亚马逊 DataZone 数据产品。

要订阅或取消订阅数据产品，请完成以下步骤。

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。如果您是亚马逊 DataZone 管理员，则可以通过 https://console.aws.amazon.com/datazon e 导航到亚马逊 DataZone 控制台，使用域名创建 AWS 账户 地登录，然后选择打开数据门户。
2. 选择浏览目录以查找要订阅的数据产品，然后选择该数据产品。
3. 在数据产品的详细信息页面上，选择订阅。
4. 指定订阅的项目和原因，然后选择订阅。

在 Amazon 中查看订阅申请并授予对数据产品的订阅 DataZone

Amazon DataZone 使数据生成者能够将数据资产分组为定义明确、自成一体的包，称为数据产品，专为特定业务用例量身定制。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。

数据产品的所有者项目可以审查并授予对亚马逊 DataZone 数据产品的订阅。

要审查订阅请求并授权对数据产品的订阅，请完成以下步骤：

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。如果您是亚马逊 DataZone 管理员，则可以通过 https://console.aws.amazon.com/datazon e 导航到亚马逊 DataZone 控制台，使用域名创建 AWS 账户 地登录，然后选择打开数据门户。
2. 选择拥有要审查其传入的订阅请求的数据产品的项目。
3. 选择数据选项卡，然后选择传入的请求。
4. 选择要审查的请求，再在订阅请求窗口中选择批准或拒绝，然后键入指定备注。

在 Amazon 上重新发布数据产品 DataZone

Amazon DataZone 使数据生成者能够将数据资产分组为定义明确、自成一体的包，称为数据产品，专为特定业务用例量身定制。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。

要重新发布数据产品，您必须是项目的所有者或贡献者。针对发布功能的元数据强制规则可以进行配置，以便为数据生成者制定明确的元数据要求，从而限制数据产品的重新发布时间。

要重新发布数据产品，请完成以下步骤。

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。如果您是亚马逊 DataZone 管理员，则可以通过 https://console.aws.amazon.com/datazon e 导航到亚马逊 DataZone 控制台，使用域名创建 AWS 账户 地登录，然后选择打开数据门户。
2. 在 Amazon DataZone 数据门户中，选择要重新发布的数据产品所在的项目。
3. 选择数据选项卡，再选择已发布的数据，然后选择数据产品筛选条件。
4. 选择要重新发布的数据产品，然后选择资产选项卡。
5. 在资产选项卡上，执行下列操作之一：

   - 移除数据产品中的某个现有资产，方法是选择该资产，然后展开操作图标并选择移除资产。通过在移除资产弹出窗口中选择移除来确认移除资产。重新发布资产后，将从该数据产品的所有订阅用户中移除此资产。
   - 向数据产品添加新资产，方法是选择“添加”按钮，然后选择要添加到数据产品中的一个或多个资产。
6. 在数据产品的详细信息页面上，选择重新发布。通过在重新发布数据产品弹出窗口中选择重新发布来确认此操作。

> **Note**
>
> 重新发布此数据产品将对所有订阅用户产生以下影响：
>
> - 如果已从数据产品中移除资产，则订阅用户将无法再访问这些资产。
> - 如果已将资产添加到数据产品中，则订阅用户将获得对这些资产的访问权限。
> - 数据资产的新发布版本将可用。

Amazon DataZone 数据发现、订阅和使用

在 Amazon 中 DataZone，一旦资产发布到某个域名，订阅者就可以发现该资产并请求订阅该资产。在订阅过程中，订阅用户首先需搜索并浏览目录以找到所需的资产。在亚马逊 DataZone门户网站上，他们选择通过提交包含申请理由和理由的订阅请求来订阅资产。资产的所有者会审查请求。他们可以批准或拒绝请求。

授权订阅后，履行流程将开始以便订阅用户访问资产。有两种主要的资产访问控制和配送模式：亚马逊资产的访问控制和配送模式，以及非亚马逊 DataZone-managed 管理的资产的访问控制和配送模式 DataZone。

- 托管资产 — 亚马逊 DataZone 可以管理托管资产的配送和权限，例如 AWS Glue 表格和 Amazon Redshift 表格和视图。
- 非托管资产 — 亚马逊向亚马逊 DataZone 发布与您的操作相关的标准事件（例如，批准订阅请求）。 EventBridge您可以使用这些标准事件与其他 AWS 服务或第三方解决方案集成，以实现自定义集成。

主题

- 在 Amazon DataZone 目录中搜索和查看资产
- 申请订阅 Amazon 中的资产 DataZone
- 在 Amazon 上批准或拒绝订阅申请 DataZone
- 撤销 Amazon 中的现有订阅 DataZone
- 在 Amazon 上取消订阅请求 DataZone
- 取消订阅 Amazon 中的资产 DataZone
- 使用现有 IAM 角色完成亚马逊 DataZone 订阅
- 向托管授予访问权限 AWS Glue Data Catalog Amazon 中的资产 DataZone
- 授予访问亚马逊中托管的亚马逊 Redshift 资产的权限 DataZone
- 向经批准的亚马逊非托管资产的订阅授予访问权限 DataZone
- 在亚马逊 Athena 或亚马逊的 Amazon Redshift 中查询数据 DataZone
- 针对订阅请求的元数据强制规则
- 通过 JDBC 连接使用外部分析应用程序分析 Amazon DataZone 订阅的数据

在 Amazon DataZone 目录中搜索和查看资产

Amazon DataZone 提供了一种简化的数据搜索方式。任何有权访问数据门户的亚马逊 DataZone 用户都可以在亚马逊 DataZone 目录中搜索资产，并查看资产名称和分配给他们的元数据。您可以通过查看其详细信息页面来进一步了解资产。

> **Note**
>
> 要查看某个资产包含的实际数据，您必须先订阅该资产，使您的订阅请求获得批准并且您获得访问权限。

Amazon DataZone （新域和现有域名）中的搜索包括基于关键字和语义匹配的结果。搜索算法会对关键字匹配项进行优先排序，然后将语义匹配项附加到它们后面。

语义搜索功能使不同角色和职能的用户能够更有效地发现、访问和利用其组织的数据资产，从而改善决策、协作和整体数据驱动型能力。在语义搜索中，除了简单的关键字匹配项结果外，关键字输入还会生成基于同义词和含义的搜索结果。例如，在语义搜索中，如果您输入“flower”作为搜索输入，搜索结果中会返回名称中包含“rose”一词的数据资产。如果您输入“movie”作为搜索输入，搜索结果中会返回名称中包含“film”一词的数据资产。如果您输入“football”作为搜索输入，搜索结果中会返回名称中包含“soccer”一词的数据资产。

利用关键字搜索，您可以在搜索订阅的资产时输入各种关键字。例如，如果您有一个名为 `Catalog Sales Data` 的资产，当您输入以下任一关键字时，该资产就会在搜索结果中返回：`catalog_sales`、`Catalog Sales`、`CatalogSales` 和 `catalogsales`。

Amazon DataZone 还通过为列名和表名等技术标识符启用精确匹配和部分匹配功能来增强搜索体验。借助这项新功能，您可以用双引号（“”）将关键字括起来，从而确保搜索结果与技术名称完全匹配或部分匹配。此功能基于关键字和语义搜索功能而构建，使您能够通过概念和相关术语来发现资产。通过为技术标识符增加一层精确度，此增强功能使您能够管理具有复杂技术命名约定的大型数据目录。

在搜索数据时，您可能需要找到特定的技术资产来支持您的使用案例。借助技术标识符搜索功能，您可以准确地检索资产，从而节省时间和简化发现过程。例如，“customer_id”之类的查询会返回具有确切标识符的列或表，而“sales_”之类的部分查询可以识别相关资产，例如 sales_summary 和 sales_data_2024。这种增强功能可确保数据使用者能够高效地找到所需资产，从而提高工作效率。

在目录中搜索资产

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。如果您是亚马逊 DataZone 管理员，则可以导航到亚马逊 DataZone 控制台，https://console.aws.amazon.com/datazone并使用域名创建 AWS 账户 地登录，然后选择打开数据门户。
2. 您可以在数据门户主页上的搜索栏中键入要查找的资产的名称。
3. 要浏览命名空间，请选择页面右上角的目录以打开目录。目录提供了一种多维搜索体验，可让您通过搜索数据所有者和术语表术语等条件来查找资产。
4. 在某个搜索框中输入您的搜索词。运行搜索后，您可以应用各种筛选条件来缩小结果范围。筛选条件包括资产类型、来源账户和资产 AWS 区域 所属账户。
5. 要查看有关特定资产的详细信息，请选择该资产以打开其详细信息页面。详细信息页面包括以下信息：

   - 资产名称、数据来源（AWS Glue、Amazon Redshift 或 Amazon S3）、类型（表、视图或 S3 对象）、列数和大小。
   - 资产的描述。
   - 当前发布的资产修订、所有者、是否需要审批订阅、命名空间和更新历史记录。
   - 概述选项卡，包括术语表术语和元数据表单。
   - 架构选项卡，显示资产的架构，包括业务和技术列名称、数据类型以及列的业务描述。“架构”选项卡仅对表和视图可见（对 Amazon S3 对象不可见）。
   - 订阅选项卡，包含域的订阅用户列表。
   - 历史记录选项卡，包含资产的过去修订的列表。

申请订阅 Amazon 中的资产 DataZone

亚马逊 DataZone 允许您查找、访问和使用亚马逊 DataZone目录中的资产。在目录中查找要访问的某个资产时，您需要订阅该资产，这将创建订阅请求。之后，审批者会批准或拒绝您的请求。

您必须是某个项目的成员才能请求订阅该项目中的资产。

订阅资产

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。如果您是亚马逊 DataZone 管理员，则可以导航到亚马逊 DataZone 控制台，https://console.aws.amazon.com/datazone并使用域名创建 AWS 账户 地登录，然后选择打开数据门户。
2. 使用搜索栏搜索并选择要订阅的资产，然后选择订阅。

3. 在订阅弹出窗口中，提供以下信息：
   - 要订阅资产的项目。
   - 简短的订阅请求理由。
4. 选择订阅。

   在发布者批准您的请求后，您将在数据门户中收到通知。

要查看订阅请求的状态，请找到并选择订阅资产的项目。导航至项目的数据选项卡，然后从左侧导航窗格中选择请求的数据。此页面列出了项目已请求访问的资产。您可以按请求状态来筛选此列表。

在 Amazon 上批准或拒绝订阅申请 DataZone

亚马逊 DataZone 允许您查找、访问和使用亚马逊 DataZone目录中的资产。在目录中查找要访问的某个资产时，您必须订阅该资产，这将创建订阅请求。之后，审批者会批准或拒绝您的请求。

您必须是拥有项目（发布了资产的项目）的成员，才能批准或拒绝订阅请求。

批准或拒绝订阅请求

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。
   如果您是亚马逊 DataZone 管理员，则可以导航到亚马逊 DataZone 控制台，https://console.aws.amazon.com/datazone并使用域名创建 AWS 账户 地登录，然后选择打开数据门户。
2. 在数据门户中，选择浏览项目列表，然后选择包含带订阅请求的资产的项目。
3. 导航至数据选项卡，然后从左侧导航窗格中选择传入的请求。
4. 找到请求并选择查看请求。您可以按待处理进行筛选，以仅查看仍处于开放状态的请求。
5. 审查订阅请求和访问理由，并决定是批准还是拒绝该请求。
6. 要批准请求，请在以下两个选项之间进行选择：
   - 完全访问：如果您选择使用“完全访问”选项批准订阅，则订阅用户将有权访问您的数据资产中的所有行和列。
   - 使用行或列筛选条件进行批准：要限制对特定的数据行和数据列的访问，您可以选择该选项以使用行和列筛选条件进行批准。有关更多信息，请参阅 对 Amazon 中数据的精细访问控制 DataZone。
     - 选择选择筛选条件，然后从下拉列表中选择要应用于订阅的一个或多个可用筛选条件。

- 要创建新的筛选条件，您可以选择“创建新筛选条件”选项，这将打开一个新页面，可在其中创建新的行或列筛选条件。有关更多信息，请参阅在 Amazon 中创建列筛选条件 DataZone和在 Amazon 中创建行筛选条件 DataZone。

7. （可选）输入响应来说明您接受或拒绝请求的原因。
8. 选择批准或拒绝。

作为项目所有者，您可以随时撤销订阅请求。有关更多信息，请参阅 the section called “撤销现有订阅”。

要查看所有订阅请求，请参阅事件和通知。

> **Note**
>
> 亚马逊 DataZone 支持对 Glue 表、亚马逊 Redshif AWS t 表和亚马逊 Redshift 视图进行精细访问控制。

自动审批订阅请求

默认情况下，对已发布资产的订阅请求需要由数据所有者手动审批。但是，Amazon DataZone 支持两种可以自动批准订阅请求的方案：

- 在资源发布期间禁用审批 – 发布数据资产时，您可以选择不要求订阅审批。在这种情况下，对该资产的所有传入订阅请求都会自动获得批准。要了解如何禁用对资产的审批，请参阅将项目库存中的资产发布到 Amazon DataZone 目录。
- 请求者是发布资产的项目的所有者或贡献者 – 如果请求者已获得手动审批订阅请求的授权，则订阅请求也会自动获得批准。具体而言，就是他们既是发布资产的项目的成员，又是请求访问权限的项目的成员。

  要获得自动审批资格，需满足以下条件：

  - 请求者必须在最初发布资产的项目中被列为所有者或贡献者。
  - 请求者还必须在提出订阅请求的项目中被列为所有者或贡献者。

  这将确保仅当请求者在两个项目（共享资产的项目和请求访问权限的项目）中都具有可见性和权限时，才会触发自动审批。如果请求者同时满足这两个条件，系统会自动批准该请求。

撤销 Amazon 中的现有订阅 DataZone

亚马逊 DataZone 允许您查找、访问和使用亚马逊 DataZone目录中的资产。在目录中查找要访问的某个资产时，您需要订阅该资产，这将创建订阅请求。之后，审批者会批准或拒绝您的请求。您在批准某个订阅后可能需要撤销该订阅，原因是批准有误或订阅用户不再需要访问该资产。

您必须是拥有项目（发布了资产的项目）的成员才能撤销订阅。

撤销订阅

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。
   如果您是亚马逊 DataZone 管理员，则可以导航到亚马逊 DataZone 控制台，https://console.aws.amazon.com/datazone并使用域名创建 AWS 账户 地登录，然后选择打开数据门户。
2. 从顶部导航窗格中选择选择项目，然后选择包含要撤销的订阅的项目。
3. 导航至数据选项卡，然后从左侧导航窗格中选择传入的请求。
4. 找到要撤销的订阅，然后选择查看订阅。
5. （可选）启用该复选框可允许订阅用户将资产保留在项目的订阅目标中。订阅目标是对一组资源的引用，其中可在环境中使用订阅的数据。

   如果您稍后需要从订阅目标撤销对资产的访问权限，则必须在 AWS Lake Formation中执行此操作。
6. 选择撤销订阅。

撤销某个订阅后，您将无法重新批准该订阅。仅在订阅用户必须再次订阅该资产后，您才能批准该资产。

> **Note**
>
> 撤销订阅只会影响特定用户对资产的访问权限，即您要撤销其订阅的订阅用户。资产本身保持不变，用户（订阅用户）也保持不变。该用户只有在提交另一个订阅请求并获得批准后才能访问该资产。

在 Amazon 上取消订阅请求 DataZone

亚马逊 DataZone 允许您查找、访问和使用亚马逊 DataZone目录中的资产。在目录中查找要访问的某个资产时，您需要订阅该资产，这将创建订阅请求。之后，审批者会批准或拒绝您的请求。您可能需要取消待处理的订阅请求，原因是您错误地提交了该请求，或您不再需要对资产进行读访问。

要取消订阅请求，您必须是项目所有者或贡献者。

取消订阅请求

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。如果您是亚马逊 DataZone 管理员，则可以导航到亚马逊 DataZone 控制台，https://console.aws.amazon.com/datazone并使用域名创建 AWS 账户 地登录，然后选择打开数据门户。
2. 从顶部导航窗格中选择选择项目，然后选择包含订阅请求的项目。
3. 导航至项目的数据选项卡，然后从左侧导航窗格中选择请求的数据。此页面列出了项目已请求访问的资产。
4. 按已请求进行筛选，以仅查看仍处于开放状态的请求。找到请求并选择查看请求。
5. 审查订阅请求并选择取消请求。

如果您想重新订阅该资产（或其他资产），请参阅the section called “请求订阅资产”。

> **Note**
>
> 当不再需要对资产的“读取”访问权限时，可以取消待处理的订阅请求。已被取消待处理的订阅请求的资产和用户不受此操作的影响。

取消订阅 Amazon 中的资产 DataZone

亚马逊 DataZone 允许您查找、访问和使用亚马逊 DataZone目录中的资产。在目录中查找要访问的某个资产时，您需要订阅该资产，这将创建订阅请求。之后，审批者会批准或拒绝您的请求。您可能需要取消订阅资产，原因是您错误地订阅了资产并且订阅请求已获批，或您不再需要对资产进行读访问。

您必须是项目成员才能取消订阅某个资产。

取消订阅资产

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。如果您是亚马逊 DataZone 管理员，则可以导航到亚马逊 DataZone 控制台，https://console.aws.amazon.com/datazone并使用域名创建 AWS 账户 地登录，然后选择打开数据门户。
2. 从顶部导航窗格中选择选择项目，然后选择包含要取消订阅的资产的项目。
3. 导航至项目的数据选项卡，然后从左侧导航窗格中选择请求的数据。此页面列出了项目已请求访问的资产。
4. 按已批准进行筛选，以仅查看已批准的请求。找到请求并选择查看订阅。
5. 审查订阅并选择取消订阅。

如果您想重新订阅该资产（或其他资产），请参阅the section called “请求订阅资产”。

> **Note**
>
> 当用户不再需要对资产的访问权限时，他们可以选择取消订阅选项。资产将保持不变，此操作不会导致任何资源被删除。

使用现有 IAM 角色完成亚马逊 DataZone 订阅

在当前版本中，Amazon DataZone 支持您使用现有的 IAM 角色来访问数据。为此，您可以在 Amazon DataZone 环境中创建用于完成订阅的订阅目标。要在其中一个关联 AWS 账户中为环境创建订阅目标，可以使用以下步骤：

第 1 步：确保您的 Amazon DataZone 域名使用的是 RAM 策略的版本 2 或更高版本

1. 在 AWS RAM 控制台中导航到 “我共享：资源共享” 页面。
2. 由于 AWS RAM 资源共享存在于特定 AWS 区域中，因此请从控制台右上角的下拉列表中选择相应的 AWS 区域。
3. 选择与您的 Amazon DataZone 域名对应的资源共享，然后选择修改。您可以使用域名的名称或 ID 来识别 Amazon DataZone 域的 RAM 共享，因为创建的 RAM 共享名为:`DataZone-<domain-name>-<domain-id>`。
4. 选择下一步以继续执行下一步骤，以便检查 RAM 策略的版本并进行修改。
5. 确保 RAM 策略的版本为版本 2 或更高版本。如果不是这样，请使用下拉列表选择版本 2 或更高版本。

6. 选择跳至步骤 4: 审核和更新。
7. 选择更新资源共享。

步骤 2：从关联账户创建订阅目标

- 在当前版本中，Amazon 仅 DataZone 支持使用 API 创建订阅目标。以下是一些有效负载示例，您可以用来创建订阅目标，以满足您的 AWS Glue 表格和 Amazon Redshift 表或视图的订阅。有关更多信息，请参阅 CreateSubscriptionTarget。

  AWS Glue 的订阅目标示例

```
{
        "domainIdentifier": "<DOMAIN_ID>",
        "environmentIdentifier": "<ENVIRONMENT_ID>",
        "name": "<SUBSCRIPTION_TARGET_NAME>",
        "type": "GlueSubscriptionTargetType",
        "authorizedPrincipals" : ["IAM_ROLE_ARN"],
        "subscriptionTargetConfig" : [{"content": "{\"databaseName\":
 \"<DATABASE_NAME>\"}", "formName": "GlueSubscriptionTargetConfigForm"}],
        "manageAccessRole": "<GLUE_DATA_ACCESS_ROLE_IN_ASSOCIATED_ACCOUNT_ARN>",
        "applicableAssetTypes" : ["GlueTableAssetType"],
        "provider": "Amazon DataZone"
}
```

  Amazon Redshift 的订阅目标示例：

```
{
        "domainIdentifier": "<DOMAIN_ID>",
        "environmentIdentifier": "<ENVIRONMENT_ID>",
        "name": "<SUBSCRIPTION_TARGET_NAME>",
        "type": "RedshiftSubscriptionTargetType",
        "authorizedPrincipals" : ["REDSHIFT_DATABASE_ROLE_NAME"],
        "subscriptionTargetConfig" : [{"content": "{\"databaseName\":
 \"<DATABASE_NAME>\", \"secretManagerArn\": \"<SECRET_MANAGER_ARN>
\",\"clusterIdentifier\": \"<CLUSTER_IDENTIFIER>\"}", "formName":
 "RedshiftSubscriptionTargetConfigForm"}],
        "manageAccessRole":
 "<REDSHIFT_DATA_ACCESS_ROLE_IN_ASSOCIATED_ACCOUNT_ARN>",
```

```
        "applicableAssetTypes" : ["RedshiftViewAssetType",
 "RedshiftTableAssetType"],
        "provider": "Amazon DataZone"
}
```

Important

- 您在上述 API 调用中使用的 environmentIdentifier 应存在于您从中发出 API 调用的同一关联账户中。否则，API 调用将失败。
- 您在 “AuthorizedPrincipals” 中使用的 IAM 角色 ARN 是在订阅资产添加到订阅目标后，亚马逊 DataZone 将向其授予访问权限的角色。这些授权主体必须与在其中创建订阅目标的环境属于同一账户。
- 供应商字段的值必须为 “Amazon DataZone” DataZone ，亚马逊才能完成订阅配送。
- 订阅中提供的数据库名称TargetConfig 应该已经存在于创建目标的账户中。亚马逊 DataZone 不会创建此数据库。此外，请确保管理访问角色具有此数据库的 CREATE TABLE 权限。
- 此外，请确保作为授权委托人提供的角色（ AWS Glue 的 IAM 角色和 Amazon Redshift 的数据库角色）已存在于环境账户中。对于 Amazon Redshift 订阅目标，在连接到集群时需要对所代入的角色进行额外更新。此角色必须为该角色附加 RedshiftDbRoles 标签。此标签的值可以是逗号分隔的列表。该值应是创建订阅目标时作为授权主体提供的数据库角色。
- `CreateSubscriptionTarget`、`UpdateSubscriptionTarget`，并在`DeleteSubscriptionTarget` AWS 账户层面获得授权。调用者必须与拥有订阅目标的环境属于同一个 AWS 账户，但不必是拥有订阅目标的项目的所有者或贡献者。因此，关联账户中持有相应 IAM 操作的任何委托人都可以创建、修改或删除任何项目的订阅目标`authorizedPrincipals`，包括决定亚马逊在完成订阅时 DataZone 授予访问权限的委托人。这些操作不能`Resource: "*"`在下面的 IAM 身份策略中限定，因此应仅授予受信任的管理委托人。

步骤 3：订阅新表并履行新目标的订阅

- 创建订阅目标后，您可以订阅新表，Amazon DataZone 会将其实现上述目标。

向托管授予访问权限 AWS Glue Data Catalog Amazon 中的资产 DataZone

在 Amazon 中 DataZone，订阅请求以及已批准或授予的资产读取权限订阅均由资产所有者管理。

> **Note**
>
> 不支持使用该 AWS Lake Formation LF-TBAC 方法对 AWS Glue Data Catalog 资产进行访问管理。
> 不支持跨区域共享中的 AWS Glue Data Catalog 资产。

托管 AWS Glue Data Catalog 资产的订阅请求获得批准后，Amazon DataZone 会自动将这些资产添加到项目中的所有现有数据湖环境中。 DataZone 然后，Amazon 会代表您通过授予并管理对已批准 AWS Glue Data Catalog 表格的访问权限 AWS Lake Formation。对于订阅者项目，授予的资产将 AWS Glue Data Catalog 作为您账户中的资源显示在中。之后，您可以使用 Amazon Athena 查询表。

> **Note**
>
> 如果在已订阅的 AWS Glue Data Catalog 资产自动添加到现有数据湖环境后向项目中添加了新的数据湖环境，则必须手动将这些订阅的 AWS Glue Data Catalog 资产添加到这个新的数据湖环境中。为此，您可以在 Amazon DataZone 数据门户中项目概述页面的 “数据” 选项卡中选择 “添加授权” 选项。

为了 DataZone 使亚马逊能够授予对 G AWS lue 数据目录表的访问权限，必须满足以下条件。

- AWS Glue 表必须是 Lake， Formation-managed 因为亚马逊通过管理 Lake Formation 权限来 DataZone 授予访问权限。
- 用于发布 AWS Glue 数据目录表的数据湖环境的管理访问角色必须具有以下 Lake Formation 权限：
  - `DESCRIBE`以及对包含已发布表的 AWS Glue 数据库的`DESCRIBE GRANTABLE`权限。
  - Lake Formation 中对已发布的表的 `DESCRIBE`、`SELECT`、`DESCRIBE GRANTABLE`、`SELECT GRANTABLE` 权限。

有关更多信息，请参阅《AWS Lake Formation Developer Guide》中的 Granting and revoking permissions on catalog resources。

授予访问亚马逊中托管的亚马逊 Redshift 资产的权限 DataZone

当对 Amazon Redshift 表或视图的订阅获得批准后，Amazon DataZone 可以自动将订阅的资产添加到项目内的所有数据仓库环境中，这样项目成员就可以在其环境中使用 Amazon Redshift 查询编辑器链接查询数据。在幕后 DataZone，Amazon在来源和订阅目标之间创建了必要的赠款和数据共享。

根据源数据库（发布者）和目标数据库（订阅用户）所在的位置，授予访问权限的过程会有所不同。

- 同一个集群，同一个数据库-如果数据必须在同一个数据库中共享，Amazon 会直接 DataZone 授予对源表的权限。
- 同一个集群，不同的数据库-如果数据必须在同一个集群中的两个数据库之间共享，Amazon DataZone 将在目标数据库中创建一个视图并授予对已创建视图的权限。
- 同一账户不同的集群-Amazon DataZone 在源集群和目标集群之间创建数据共享，并在共享表的顶部创建视图。授予对视图的权限。
- Cross-account -与上面相同，但需要额外的步骤才能在生产者集群端授权跨账户数据共享，还需要另外一个步骤来关联消费者集群端的数据共享。

> **Note**
>
> 如果在自动将订阅的 Amazon Redshift 资产添加到现有数据仓库环境后，向项目添加了一个新的数据仓库环境，则必须手动将这些订阅的 Amazon Redshift 资产添加到这个新的数据仓库环境中。为此，您可以在 Amazon DataZone 数据门户中项目概述页面的 “数据” 选项卡中选择 “添加授权” 选项。

确保您的发布和订阅 Amazon Redshift 集群满足 Amazon Redshift 数据共享的所有要求。有关更多信息，请参阅《Amazon Redshift 开发人员指南》。

> **Note**
>
> 亚马逊 DataZone 支持自动授予对亚马逊 Redshift 集群和亚马逊 Redshift 无服务器资产的订阅。
>
> Cross-Region 不支持使用 Amazon Redshift 进行数据共享。

向经批准的亚马逊非托管资产的订阅授予访问权限 DataZone

在 Amazon 中 DataZone，订阅请求以及已批准或授予的资产读取权限订阅均由资产所有者管理。

Amazon DataZone 允许用户在业务数据目录中发布任何类型的资产。对于其中一些资产，Amazon DataZone 可以自动管理访问授权。这些资产称为托管资产，包括 Lake Formation-managed AWS Glue 数据目录表以及 Amazon Redshift 表和视图。Amazon DataZone 无法自动授予订阅权限的所有其他资产都称为非托管资产。

Amazon 为您 DataZone 提供了管理非托管资产访问权限的途径。当企业数据目录中某项资产的订阅获得数据所有者的批准后，亚马逊会在您的账户中 EventBridge 在亚马逊上 DataZone 发布一个事件，并在有效载荷中发布所有必要的信息，使您能够在来源和目标之间创建访问授权。在收到此事件后，您可以触发一个自定义处理程序，从而使用事件中的信息来创建必要的授予或权限。授予访问权限后，您可以在 Amazon 上报告和更新订阅状态， DataZone 这样它就可以通知订阅该资产的用户他们可以开始使用该资产。有关更多信息，请参阅 Amazon DataZone 事件和通知。

在亚马逊 Athena 或亚马逊的 Amazon Redshift 中查询数据 DataZone

在亚马逊中 DataZone，一旦订阅者可以访问目录中的资产，他们就可以使用 Amazon Athena 或 Amazon Redshift 查询编辑器 v2 使用该资产（查询和分析）。您必须是项目所有者或贡献者才能完成此任务。根据项目中启用的蓝图，亚马逊在数据门户项目页面的右侧窗格中 DataZone 提供指向 and/or 亚马逊 Athena Amazon Redshift 查询编辑器 v2 的链接。

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。
   如果您是亚马逊 DataZone 管理员，则可以导航到亚马逊 DataZone 控制台，https://console.aws.amazon.com/datazone并使用域名创建 AWS 账户 地登录，然后选择打开数据门户。
2. 在 Amazon DataZone 数据门户中，选择 “浏览项目列表”，然后找到并选择要分析的数据所在的项目。
3. 如果在此项目上启用了数据湖蓝图，则项目主页的右侧面板中将显示指向 Amazon Athena 的链接。

   如果在此项目上启用了数据仓库蓝图，则项目主页的右侧面板中将显示指向查询编辑器的链接。

> **Note**
>
> 蓝图是在用于创建项目的环境配置文件中定义的。

主题

- 使用 Amazon Athena 查询数据
- 使用 Amazon Redshift 查询数据

使用 Amazon Athena 查询数据

选择 Amazon Athena 链接，使用项目的身份验证凭证在浏览器的新标签页中打开 Amazon Athena 查询编辑器。在查询编辑器中，系统会自动选择您正在处理的 Amazon DataZone 项目作为当前工作组。

在 Amazon Athena 查询编辑器中，编写并运行您的查询。一些常见任务包括：

- 查询和分析订阅的资产
- 创建新表
- 从外部 S3 存储桶中的查询结果（CTAS）创建表

查询和分析订阅的资产

如果 Amazon 未自动授予您项目订阅的资产的访问权限 DataZone，则必须授权您访问基础数据。有关如何授予对这些资产的访问权限的更多信息，请参阅向经批准的亚马逊非托管资产的订阅授予访问权限 DataZone。

如果亚马逊自动授予您项目订阅的资产的访问权限 DataZone，则可以对表运行 SQL 查询并在 Amazon Athena 中查看结果。有关在 Amazon Athena 中使用 SQL 的更多信息，请参阅 SQL reference for Athena。

如果在项目主页的右侧面板中选择 Amazon Athena 链接后导航到 Amazon Athena 查询编辑器，则 Amazon Athena 查询编辑器的右上角会显示项目下拉列表，并且会自动选择您的项目上下文。

您会在数据库下拉列表中看到以下数据库：

- 发布数据库（*{environmentname}*_pub_db）。该数据库的目的是为您提供一个环境，让您可以在项目背景下生成新数据，然后将这些数据发布到 Amazon DataZone 目录中。项目所有者和贡献者具有此数据库的读写访问权限。项目查看者仅具有此数据库的读访问权限。
- 订阅数据库（*{environmentname}*_sub_db）。该数据库的目的是与您共享您作为项目成员在 Amazon DataZone 目录中订阅的数据，并使您能够查询这些数据。

创建新表

如果您已连接到一个外部 Amazon S3 存储桶，则可以使用 Amazon Athena 查询和分析该存储桶中的资产。在这种情况下，亚马逊 DataZone 无权直接授予对外部 Amazon S3 存储桶中基础数据的访问权限，并且在项目外部创建的外部 Amazon S3 数据不会在 Lake Formation 中自动管理，也无法由亚马逊管理 DataZone。另一种方法是使用 Amazon Athena 中的 CREATE TABLE 语句将数据从外部 Amazon S3 存储桶复制到项目的 Amazon S3 存储桶中的新表中。在 Amazon Athena 中运行 CREATE TABLE 查询时，会将表注册到 AWS Glue Data Catalog。

要在 Amazon S3 中指定数据的路径，请使用 LOCATION 属性，如以下示例中所示：

```
CREATE EXTERNAL TABLE 'test_table'(
...
)
ROW FORMAT ...
STORED AS INPUTFORMAT ...
OUTPUTFORMAT ...
LOCATION 's3://bucketname/folder/'
```

有关更多信息，请参阅 Amazon S3 中的表位置。

从外部 S3 存储桶中的查询结果（CTAS）创建表

在订阅资产时，只能对底层数据进行只读访问。您可以使用 Amazon Athena 创建表的副本。在 Amazon Athena 中，A CREATE TABLE AS SELECT (CTAS) 查询根据另一个查询中的 SELECT 语句的结果创建新表。有关 CTAS 语法的信息，请参阅 CREATE TABLE AS。

以下示例通过复制表的所有列来创建表：

```
CREATE TABLE new_table AS
SELECT *
FROM old_table;
```

在同一个示例的下列变化中，您的 SELECT 语句还包括 WHERE 子句。在这种情况下，查询将只从表中选择满足 WHERE 子句的行：

```
CREATE TABLE new_table AS
SELECT *
FROM old_table WHERE condition;
```

以下示例创建运行在其他表的一组列上的新查询：

```
CREATE TABLE new_table AS
SELECT column_1, column_2, ... column_n
FROM old_table;
```

同一个示例的此变化从多个表的特定列创建新表：

```
CREATE TABLE new_table AS
SELECT column_1, column_2, ... column_n
FROM old_table_1, old_table_2, ... old_table_n;
```

这些新创建的表现在已成为您项目 AWS Glue 数据库的一部分，通过将数据作为资产发布到亚马逊目录中，可以让其他人发现并与其他亚马逊 DataZone DataZone 项目共享。

使用 Amazon Redshift 查询数据

在 Amazon DataZone 数据门户中，打开使用数据仓库蓝图的环境。在环境页面上的右侧面板中，选择 Amazon Redshift 链接。这将打开一个确认对话框，其中包含必要的详细信息，可帮助您在 Amazon Redshift 查询编辑器 v2.0 中与环境的 Amazon Redshift 集群或 Amazon Redshift Serverless 工作组建立连接。在确定建立连接所需的详细信息后，选择打开 Amazon Redshift 按钮。这将使用亚马逊环境的临时凭证在浏览器的新选项卡中打开 Amazon Redshift 查询编辑器 v2.0。 DataZone

在查询编辑器中，根据您的环境使用的是 Amazon Redshift Serverless 工作组还是 Amazon Redshift 集群，执行以下步骤。

对于 Amazon Redshift Serverless 工作组

1. 在查询编辑器中，识别您的亚马逊 DataZone 环境的 Amazon Redshift Serverless 工作组，右键单击该工作组，然后选择创建连接。
2. 选择联合用户以进行身份验证。

3. 提供 Amazon DataZone 环境数据库的名称。
4. 选择创建连接。

对于 Amazon Redshift 集群：

1. 在查询编辑器中，识别您的亚马逊 DataZone 环境的 Amazon Redshift 集群，右键单击它并选择创建连接。
2. 选择使用您的 IAM 身份的临时凭证以进行身份验证。
3. 如果上述身份验证方法不可用，请通过选择左下角的齿轮按钮打开账户设置，然后选择使用 IAM 凭证进行身份验证并保存。这是一项一次性设置。
4. 提供用于创建连接的 Amazon DataZone 环境数据库的名称。
5. 选择创建连接。

现在，您可以开始查询为亚马逊环境配置的 Amazon Redshift 集群或 Amazon Redshift 无服务器工作组中的表和视图。 DataZone

您已订阅的任何 Amazon Redshift 表或视图都会链接到为该环境配置的 Amazon Redshift 集群或 Amazon Redshift Serverless 工作组。您可以订阅表和视图，也可以发布您在环境的集群或数据库中创建的任何新表和视图。

例如，我们来看看以下场景：一个环境链接到一个名为 `redshift-cluster-1` 的 Amazon Redshift 集群以及该集群中名为 `dev` 的数据库。使用 Amazon DataZone 数据门户，您可以查询已添加到您的环境中的表和视图。在数据门户的右侧窗格中的`Analytics tools`部分下，您可以选择此环境的 Amazon Redshift 链接，这将打开查询编辑器。之后，您可以右键单击 `redshift-cluster-1` 集群，并使用使用您的 IAM 身份的临时凭证创建连接。建立连接后，您可以在 dev 数据库下看到您的环境有权访问的所有表和视图。

针对订阅请求的元数据强制规则

Amazon DataZone 中的订阅请求元数据强制规则功能使域单位所有者能够为数据使用者制定明确的元数据要求，简化访问请求并增强数据治理，从而加强数据治理。此功能使组织能够遵守其元数据标准、实施自定义工作流和提供一致、受管控的数据访问体验。

目前有 Amazon 的所有 AWS 商业区域都支持 DataZone 该功能。

域单位所有者可以完成以下程序，在 Amazon 中配置元数据强制执行 DataZone：

1. 使用 DataZone 数据门户 URL 导航至 Amazon 数据门户，然后使用您的 SSO 或 AWS 凭证登录。如果您是亚马逊 DataZone 管理员，则可以通过访问创建亚马逊 DataZone 域名的 AWS 账户 https://console.aws.amazon.com/datazone中的亚马逊 DataZone 控制台来获取数据门户 URL。
2. 选择域，导航到域单元选项卡，然后选择要使用的域单元。
3. 选择规则选项卡，然后选择添加。
4. 在创建必需的元数据表单规则页面上，执行以下操作，然后选择添加规则：
   - 为规则指定名称。
   - 在操作下，选择订阅请求。
   - 在必填表单下，选择添加元数据表单，在域/域单元内选择要添加到此规则的元数据表单，然后选择添加。最多可为每个规则添加 5 个元数据。
   - 在范围下，指定要与这些表单关联的数据实体。您可以选择数据产品 and/or 数据资产。
   - 在数据资产类型下，指定该规则是适用于所有资产类型还是选定的资产类型。
   - 在 “项目” 下，指定所需的表单是与所有项目发布的数据产品 and/or 资产相关联，还是仅与该域单元中的选定项目相关联。此外，如果您希望子域单元继承此要求，请选中将规则级联到子域单元。

配置元数据强制后，数据使用者可以完成以下过程来请求访问权限：

1. 使用 DataZone 数据门户 URL 导航至 Amazon 数据门户，然后使用您的 SSO 或 AWS 凭证登录。如果您是亚马逊 DataZone 管理员，则可以通过访问创建亚马逊 DataZone 域名的 AWS 账户 https://console.aws.amazon.com/datazone中的亚马逊 DataZone 控制台来获取数据门户 URL。
2. 使用搜索栏搜索并选择要订阅的资产，然后选择订阅。
3. 在订阅弹出窗口中，提供以下信息：
   - 要订阅资产的项目。
   - 简短的订阅请求理由。
   - 填写必填元数据 – 指定域单元指定的必填元数据字段。如果必填字段不完整，系统将突出显示这些字段，并且在问题解决之前会禁止提交。输入所有必填字段后，选择应用。
4. 选择请求以提交订阅请求。提交后，将在中生成一个事件 EventBridge，该事件可根据需要在 Amazon DataZone 以外的自定义工作流程中使用。在发布者批准您的请求后，您将在数据门户中收到通知。

数据生成者可以完成以下过程来审批订阅请求：

批准或拒绝订阅请求

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。如果您是亚马逊 DataZone 管理员，则可以导航到亚马逊 DataZone 控制台，https://console.aws.amazon.com/datazone并使用域名创建 AWS 账户 地登录，然后选择打开数据门户。
2. 在数据门户中，选择浏览项目列表，然后选择包含带订阅请求的资产的项目。
3. 导航至数据选项卡，然后从左侧导航窗格中选择传入的请求。
4. 找到请求并选择查看请求。您可以按待处理进行筛选，以仅查看仍处于开放状态的请求。
5. 审查订阅请求和访问理由，并决定是批准还是拒绝该请求。

   数据生成者可以查看提供的元数据（包括文档链接和账户 ID），以在授予访问权限前确认请求是否符合合规性和工作流要求。
6. 要批准请求，请在以下两个选项之间进行选择：

   - 完全访问：如果您选择使用“完全访问”选项批准订阅，则订阅用户将有权访问您的数据资产中的所有行和列。
   - 使用行或列筛选条件进行批准：要限制对特定的数据行和数据列的访问，您可以选择该选项以使用行和列筛选条件进行批准。有关更多信息，请参阅 对 Amazon 中数据的精细访问控制 DataZone。
     - 选择选择筛选条件，然后从下拉列表中选择要应用于订阅的一个或多个可用筛选条件。
     - 要创建新的筛选条件，您可以选择“创建新筛选条件”选项，这将打开一个新页面，可在其中创建新的行或列筛选条件。有关更多信息，请参阅在 Amazon 中创建列筛选条件 DataZone和在 Amazon 中创建行筛选条件 DataZone。
7. （可选）输入响应来说明您接受或拒绝请求的原因。
8. 选择审批。

通过 JDBC 连接使用外部分析应用程序分析 Amazon DataZone 订阅的数据

亚马逊 DataZone 使数据使用者能够在单个项目中轻松查找和订阅来自多个来源的数据，并使用亚马逊 Athena、Amazon Redshift 查询编辑器和亚马逊分析这些数据。 SageMaker

亚马逊 DataZone 还支持通过 Athena JDBC 驱动程序进行身份验证，该驱动程序使用户能够使用流行的外部 SQL 和分析工具（例如 SQL Workbench、dBeaver、Tableau、Domino、Power BI 等）查询

他们订阅的 DataZone 亚马逊数据。用户可以通过 SSO 或 IAM 使用其公司凭证进行身份验证，然后开始分析他们在 Ama DataZone zon 项目中订阅的数据。

亚马逊对 A DataZone thena JDBC 驱动程序的支持具有以下好处：

- 更多的查询和可视化工具选择-数据 DataZone 使用者可以使用支持 JDBC 连接的各种分析工具中的首选工具连接到 Amazon。这使他们能够继续使用自己熟悉的软件，而无需学习新的数据使用工具。
- 编程访问 – 通过服务器或自定义应用程序与访问受管控的数据建立 JDBC 连接，使数据使用者能够执行自动化且更复杂的数据操作。

您可以使用您的 JDBC 网址将您的外部分析工具连接到您的 Amazon DataZone 订阅的数据。要获取 JDBC URL，请执行以下过程：

> **Important**
>
> 在当前版本中，亚马逊 DataZone 支持使用亚马逊 Athena JDBC 驱动程序进行身份验证。要完成此过程，请确保已为所选分析应用程序下载并安装了最新的 Athena JDBC 驱动程序。

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。如果您是亚马逊 DataZone 管理员，则可以导航到亚马逊 DataZone 控制台，https://console.aws.amazon.com/datazone并使用域名创建 AWS 账户 地登录，然后选择打开数据门户。
2. 在 Amazon DataZone 数据门户中，选择 “浏览项目列表”，然后找到并选择要分析的数据所在的项目。
3. 在项目主页的右侧面板中，选择使用 JDBC 进行连接。
4. 在 JDBC 参数弹出窗口中，选择您的身份验证方法（SSO 凭证或 IAM 凭证），然后复制 JDBC URL 的字符串或独立参数。然后，您可以使用它连接到您的外部分析应用程序。

当您 DataZone 使用 JBDC 查询或参数将外部分析应用程序连接到 Amazon 时，即`RedeemAccessToken`调用 API。`RedeemAccessToken` API 会将 Identity Center 访问令牌交换为用于调用 `GetEnvironmentCredentials` API 的 `AmazonDataZoneDomainExecutionRole` 凭证。

有关使用 IAM 证书连接到 Athena 中的亚马逊 DataZone-governed 数据的身份验证机制的更多信息，DataZone 请参阅 IAM 证书提供商。有关允许使用 IAM 身份中心在 Athena 中连接到 Amazon DataZone-governed 数据的身份验证机制的更多信息，DataZone 请参阅 Idc 证书提供商。

RedeemAccessToken API 参考

请求语法

```
POST /sso/redeem-token HTTP/1.1
Content-type: application/json

{
   "domainId": "string",
   "accessToken": "string"
}
```

请求参数

请求使用了以下参数。

DomainId

亚马逊 DataZone 域名的 ID。

模式：^dzd [-_] [a-z A-Z0-9 _-] {1,36} $

是否必需：是

accessToken

Identity Center 访问令牌。

类型：字符串。

是否必需：是

响应语法

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

{
   "credentials": AwsCredentials
```

```
}
```

响应元素

凭证

用于调用 GetEnvironmentCredentials API 的 AmazonDataZoneDomainExecutionRole 凭证。

类型：AwsCredentials 对象数组。此数据类型包含以下属性：

- 访问权限KeyId： AccessKeyId
- 秘密AccessKey： SecretAccessKey
- 会话令牌： SessionToken
- expiration：Timestamp

accessToken

Identity Center 访问令牌。

类型：字符串。

是否必需：是

错误

AccessDeniedException

您没有足够的访问权限，无法执行该操作。

HTTP 状态代码：403

ResourceNotFoundException

找不到指定的资源。

HTTP 状态代码：404

ValidationException

输入未能满足 AWS 服务指定的约束。

HTTP 状态代码：400

InternalServerException

由于未知错误、异常或故障，请求失败。

HTTP 状态代码：500

对 Amazon 中数据的精细访问控制 DataZone

在当前版本的 Amazon 中 DataZone，支持对您的数据进行精细的访问控制，使您能够对敏感数据进行精细的访问控制。您可以控制哪个项目可以访问发布到亚马逊 DataZone 业务数据目录的数据资产中的特定数据记录。Amazon DataZone 支持行和列筛选器来实现精细的访问控制。

行筛选条件使您能够根据您定义的条件限制对特定行的访问。例如，如果您的表包含两个区域（美洲和欧洲）的数据，并且您想确保欧洲的员工只能访问与其区域相关的数据，则可以创建一个行筛选条件，其中包含区域为欧洲（例如，区域 =“欧洲”）的行。这样一来，欧洲的员工便无法访问美洲的数据。

利用列筛选条件，您可以限制对数据资产中特定列的访问。例如，如果您的表包含个人身份信息（PII）等敏感信息，则可以创建一个列筛选条件来排除 PII 列。这可确保订阅用户只能访问非敏感数据。

要利用精细的访问控制，您可以为亚马逊上的 GI AWS ue 和 Amazon Redshift 资产创建行和列筛选器。 DataZone在收到访问数据资产的订阅请求后，您可以通过应用相应的行和列筛选条件来批准该请求。Amazon DataZone 确保订阅者只能访问您在批准订阅时应用的筛选条件所允许的行和列。

主题

- 在 Amazon 中创建行筛选条件 DataZone
- 在 Amazon 中创建列筛选条件 DataZone
- 删除 Amazon 中的行或列筛选条件 DataZone
- 在 Amazon 中编辑行或列筛选条件 DataZone
- 在 Amazon 中使用筛选条件授予访问权限 DataZone

在 Amazon 中创建行筛选条件 DataZone

Amazon DataZone 允许您创建行筛选器，以便在批准订阅时使用，以确保订阅者只能访问行筛选器中定义的数据行。要创建行筛选条件，请执行以下步骤：

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。如果您是亚马逊 DataZone 管理员，则可以通过 https://console.aws.amazon.com/datazon e 导航到亚马逊 DataZone 控制台，使用域名创建 AWS 账户 地登录，然后选择打开数据门户。
2. 从顶部导航窗格中选择选择项目，然后选择资产所属的项目。
3. 导航到项目的数据选项卡。

4. 从左侧导航窗格中选择已发布的数据，然后选择要为其创建行筛选条件的资产。如果您在亚马逊中的数据资产类型 DataZone 为 AWS Glue 表、亚马逊 Redshift 表或亚马逊 Redshift 视图，则可以添加行筛选器。
5. 在资产详细信息页面上，转至资产筛选条件选项卡，然后选择添加资产筛选条件。
6. 配置以下字段：
   - 名称 – 筛选条件的名称
   - 描述 – 筛选条件的描述
7. 在“筛选条件类型”下，选择行筛选条件。
8. 在“行筛选条件表达式”下，为行筛选条件提供一个或多个表达式。
   - 从下拉列表的列中选择列。
   - 从运算符下拉列表中选择运算符。
   - 在值字段中输入值。
9. 要向筛选条件表达式添加其他条件，请选择添加条件。
10. 在行筛选条件表达式中使用多个条件时，选择 And 或 Or 以链接这些条件。
11. 选择创建筛选条件。

有关如何向订阅应用行筛选条件的信息，请参阅在 Amazon 上批准或拒绝订阅申请 DataZone。

在 Amazon 中创建列筛选条件 DataZone

Amazon DataZone 允许您创建列筛选条件，以便在批准订阅时使用，以确保订阅者只能访问列筛选器中定义的数据列。要创建列筛选条件，请执行以下步骤：

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。如果您是亚马逊 DataZone 管理员，则可以通过 https://console.aws.amazon.com/datazon e 导航到亚马逊 DataZone 控制台，使用域名创建 AWS 账户 地登录，然后选择打开数据门户。
2. 从顶部导航窗格中选择选择项目，然后选择资产所属的项目。
3. 导航到项目的数据选项卡。
4. 从左侧导航窗格中选择已发布的数据，然后选择要为其创建列筛选条件的资产。如果您在亚马逊中的数据资产类型为 AWS G DataZone lue 表、亚马逊 Redshift 表或亚马逊 Redshift 视图，则可以添加列筛选器。
5. 在资产详细信息页面上，转至资产筛选条件选项卡，然后选择添加资产筛选条件。

6. 配置以下字段：
   - 名称 – 筛选条件的名称
   - 描述 – 筛选条件的描述
7. 在“筛选条件类型”下，选择列筛选条件。
8. 使用复选框再次选择要包含在筛选条件中的列，即数据资产中的列。
9. 选择“创建筛选条件”

有关如何向订阅应用列筛选条件的信息，请参阅在 Amazon 上批准或拒绝订阅申请 DataZone。

删除 Amazon 中的行或列筛选条件 DataZone

要删除行或列筛选条件，请执行以下步骤：

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。如果您是亚马逊 DataZone 管理员，则可以通过 https://console.aws.amazon.com/datazon e 导航到亚马逊 DataZone 控制台，使用域名创建 AWS 账户 地登录，然后选择打开数据门户。
2. 导航到项目的数据选项卡。
3. 从左侧导航窗格中选择已发布的数据或库存数据，然后选择要为其删除行或列筛选条件的资产。
4. 在资产详细信息页面上，转到资产筛选条件选项卡，然后打开要删除的筛选条件。
5. 依次选择操作和删除，然后确认删除。

> **Note**
>
> 仅在活跃订阅中未使用某个筛选条件时，才能删除该筛选条件。

在 Amazon 中编辑行或列筛选条件 DataZone

要编辑行或列筛选条件，请执行以下步骤：

1. 导航至 Amazon DataZone 数据门户 URL，然后使用单点登录 (SSO) 或凭证登录 AWS 。如果您是亚马逊 DataZone 管理员，则可以通过 https://console.aws.amazon.com/datazon e 导航到亚马逊 DataZone 控制台，使用域名创建 AWS 账户 地登录，然后选择打开数据门户。

2. 导航到项目的数据选项卡。
3. 从左侧导航窗格中选择已发布的数据或库存数据，然后选择要为其编辑行或列筛选条件的资产。
4. 在资产详细信息页面上，转到资产筛选条件选项卡，然后打开要编辑的筛选条件。
5. 您可以编辑以下字段：
   - 描述 – 筛选条件的描述
6. 如果您编辑的是行筛选条件，则可以更新行筛选条件表达式。
7. 如果您编辑的是列筛选条件，则可以在筛选条件中添加或删除选定的列。
8. 完成更改后，选择编辑资产筛选条件。

> **Note**
>
> 如果您编辑活跃订阅中使用的筛选条件，Amazon DataZone 将自动更新授予订阅者项目的权限。这意味着订阅用户将只能访问更新后的筛选条件中定义的行或列，从而确保一致地实施您的数据访问策略。

在 Amazon 中使用筛选条件授予访问权限 DataZone

亚马逊 DataZone 通过将定义的行和列筛选条件转换为对 La AWS ke Formation 和 Amazon Redshift 的适当授权，实现了精细的访问控制。以下是亚马逊如何为 AWS Glue 表和 Amazon Redshift DataZone 实现这些筛选条件的说明。

AWS Glue 桌子

当对带有行 and/or 列筛选器的 AWS Glue 表的订阅获得批准后，Amazon 会通过在 AWS Lake Formation 中创建带有数据单元筛选器的授权，从而 DataZone 实现订阅，确保订阅者项目的成员只能根据应用于订阅的筛选条件访问允许他们访问的行和列。

亚马逊 DataZone 首先将亚马逊 DataZone 中应用的行和列筛选器转换为 AWS Lake Formation 数据单元格筛选器。如果使用多个行和列筛选条件，Ama DataZone zon 会合并所有列和所有行筛选条件，以计算行和列级别的有效权限。 DataZone 然后，亚马逊使用有效的行和列权限创建了一个 AWS Lake Formation 数据单元格筛选器。

创建数据单元筛选器后，Amazon 将使用此数据单元筛选器在 La AWS ke Formation 中创建只读（SELECT）权限，从而与订阅者项目 DataZone 共享订阅表。

Amazon Redshift

当订阅 table/view 带有行 and/or 列筛选器的 Amazon Redshift 获得批准后，Amazon 会通过在 Amazon Redshift 中创建限定范围的延迟绑定视图来 DataZone 实现订阅，确保订阅者项目的成员只能根据应用于订阅的行和列筛选条件访问允许他们访问的行和列。

亚马逊 DataZone 首先将应用于亚马逊订阅的行和列筛选条件转换为 Amazon DataZone Redshift 后期绑定视图。如果使用多个行和列筛选条件，A DataZone mazon 会合并所有列和所有行筛选条件，以计算行和列级别的有效权限。 DataZone 然后，Amazon 使用有效的行和列权限创建后期绑定视图。

创建后期绑定视图后，亚马逊将通过在 Amazon Redshift 中创建只读（选择）权限，与订阅者项目的成员 DataZone 共享此视图。

Amazon DataZone 事件和通知

Amazon DataZone 会随时向您通报数据门户中的重要活动，例如订阅请求、更新、评论和系统事件。亚马逊通过在数据门户的专用收件箱中或通过亚马逊 EventBridge 默认总线传送消息来向您 DataZone 提供这些信息。

通过 Amazon DataZone 数据门户中的专用收件箱进行活动

Amazon在数据门户中 DataZone 提供了一个专用的收件箱，您可以在其中查看消息并对其采取行动。最新消息还将显示在主页、项目页面和目录页面上。例如，如果用户请求访问某个数据资产，则该资产的发布项目的所有者和贡献者会在数据门户中看到此请求，在执行操作后，与此请求相关的订阅项目的项目成员将在数据门户中看到通知。有两种类型的消息：

- 任务 – 这些消息告知收件人需要在某个位置执行操作。它们具有一个可选的状态字段，您可使用此字段进行跟踪。
- 事件 – 这些消息仅供参考，并且没有指定的状态。事件提供最近更新的审计跟踪记录。

在 Amazon 中 DataZone，会为以下事件类型生成消息：

事件类别	事件名称	事件描述	事件类型
订阅	已创建订阅请求	创建订阅请求时生成此事件	Task
订阅	已接受订阅请求	接受订阅请求时生成此事件	事件
订阅	已拒绝订阅请求	拒绝订阅请求时生成此事件	事件
订阅	已删除订阅请求	删除订阅请求时生成此事件	事件
Project	已成功创建项目	成功创建项目时生成此事件	事件

事件类别	事件名称	事件描述	事件类型
项目成员资格	已成功添加项目成员	向项目中添加新成员时生成此事件	事件
项目成员资格	已成功移除项目成员	从项目中移除成员时生成此事件	事件
项目成员资格	已成功更改项目成员角色	更改成员在项目中的角色时生成此事件	事件
环境	已开始环境部署	启动环境部署时生成此事件	事件
环境	已完成环境部署	成功完成环境部署时生成此事件	事件
环境	环境部署已失败	环境部署失败时生成此事件	事件
环境	环境部署自定义工作流已启动	启动具有自定义工作流的环境时生成此事件	事件
数据资产	已将资产添加到库存	将新的数据资产添加到库存（即在草稿状态下添加到目录）时生成此事件	事件
数据资产	已发布资产	在发布新的数据资产（即可供订阅）时生成此事件	事件
数据资产	已更改资产架构	自上一个摄取作业后，当资产架构发生更改时生成此事件	事件
订阅	已创建订阅	当有人请求订阅数据资产时生成此事件	Task

事件类别	事件名称	事件描述	事件类型
订阅	已批准订阅	当发布项目所有者或贡献者批准订阅时生成此事件	事件
订阅	已拒绝订阅	当发布项目所有者或贡献者拒绝订阅时生成此事件	事件
订阅	已删除订阅	当订阅用户取消订阅时生成此事件	事件
订阅	已请求订阅授权	当有人请求对资产的访问权限时生成此事件	事件
订阅	已完成订阅授权	当发布项目所有者或贡献者授权订阅资产时生成此事件	事件
订阅	订阅授权失败	在订阅授权失败时生成此事件	事件
订阅	已请求撤销订阅授权	当发布项目所有者或贡献者撤销订阅授权时生成此事件	事件
订阅	已完成订阅授权撤销	在完成订阅授权撤销时生成此事件	事件
订阅	订阅授权撤销失败	在订阅授权撤销失败时生成此事件	事件
自动生成企业名称	已成功生成企业名称	在成功完成自动企业名称生产作业时生成此事件	事件

事件类别	事件名称	事件描述	事件类型
自动生成企业名称	企业名称生成失败	在自动企业名称生产作业失败时生成此事件	事件
数据来源运行	已创建数据来源	在创建新的数据来源时生成此事件	事件
数据来源运行	已更新数据来源	在更新现有数据来源时生成此事件	事件
数据来源运行	已触发数据来源运行	在启动数据来源运行时生成此事件	事件
数据来源运行	数据来源运行成功	在数据来源运行成功时生成此事件	事件
数据来源运行	数据来源运行失败	在数据来源运行失败时生成此事件	事件

要查看数据门户收件箱中的任务，请完成以下步骤：

1. 使用 DataZone 数据门户 URL 导航至 Amazon 数据门户，然后使用您的 SSO 或 AWS 凭证登录。如果您是亚马逊 DataZone 管理员，则可以通过访问创建亚马逊 DataZone 域名的 AWS 账户中的 https://console.aws.amazon.com/datazon e 上的亚马逊 DataZone 控制台来获取数据门户 URL。
2. 在数据门户中，要查看包含一组最新任务的弹出窗口，请选择搜索栏旁边的铃铛图标。
3. 选择“查看全部”以查看所有任务。可以通过选择“事件”选项卡来更改视图并查看所有事件。
4. 可以按事件主题、活动或非活动状态或日期范围来筛选搜索内容。
5. 选择任一任务以导航到可响应该任务的位置。

要查看数据门户收件箱中的事件，请完成以下步骤：

1. 使用 DataZone 数据门户 URL 导航至 Amazon 数据门户，然后使用您的 SSO 或 AWS 凭证登录。如果您是亚马逊 DataZone 管理员，则可以通过使用创建亚马逊 DataZone 根域名的 AWS

账户访问亚马逊 DataZone 控制台 https://console.aws.amazon.com/datazon e 来获取数据门户 URL。

2. 在数据门户中，要查看包含一组最新事件的弹出窗口，请选择搜索栏旁边的铃铛图标。
3. 选择“查看全部”以查看所有事件。可以通过选择“任务”选项卡来更改视图并查看所有任务。
4. 按事件主题或日期范围来筛选搜索内容。
5. 选择任一事件以导航到可查看该事件详细信息的位置。

通过 Amazon EventBridge 默认总线举办的活动

除了将消息发送到数据门户中的专用收件箱外， DataZone 还可以使用托管亚马逊 DataZone 根域名的同一 AWS 账户将这些消息发送到您的亚马逊 EventBridge 默认事件总线。这将实现事件驱动型自动化，例如履行订阅或与其他工具的自定义集成。您可以创建与传入的亚马逊 EventBridge 事件相匹配的规则，并将它们发送到亚马逊 EventBridge 目标进行处理。一条规则可以将一个事件发送到多个目标，然后这些目标将可并行运行。

以下是示例事件：


```
{
  "version": "0",
  "id": "bd3d6239-2877-f464-0572-b1d76760e085",
  "detail-type": "Subscription Request Created",
  "source": "aws.datazone",
  "account": "111111111111",
  "time": "2023-11-13T17:57:00Z",
  "region": "us-east-1",
  "resources": [],
  "detail": {
    "version": "655",
    "metadata": {
      "domain": "dzd_bc8e1ez8r2a6xz",
      "user": "44f864b8-50a1-70cc-736f-c1f763934ab7",
      "id": "5jbc0lie0sr99j",
      "version": "1",
      "typeName": "SubscriptionRequestEntityType",
      "owningProjectId": "6oy92hwk937pgn",
      "awsAccountId": "111111111111",
      "clientToken": "e781b7b5-78c5-4608-961e-3792a6c3ff0d"
    },
    "data": {
```

```
      "autoApproved": true,
      "requesterId": "44f864b8-50a1-70cc-736f-c1f763934ab7",
      "status": "PENDING",
      "subscribedListings": [
        {
          "id": "ayzstznnx4dxyf",
          "ownerProjectId": "5a3se66qm88947",
          "version": "12"
        }
      ],
      "subscribedPrincipals": [
        {
          "id": "6oy92hwk937pgn",
          "type": "PROJECT"
        }
      ]
    }
  }
}
```

Ama DataZone zon 支持的详情类型的完整列表包括：

- 已创建订阅请求
- 已接受订阅请求
- 已拒绝订阅请求
- 已删除订阅请求
- 已请求订阅授权
- 已完成订阅授权
- 订阅授权失败
- 已请求撤销订阅授权
- 已完成订阅授权撤销
- 订阅授权撤销失败
- 已将资产添加到库存
- 已将资产添加到目录
- 已更改资产架构
- 数据来源状态更改

- 已创建数据来源
- 已更新数据来源
- 已触发数据来源运行
- 数据来源运行成功
- 数据来源运行失败
- 域创建成功
- 域创建失败
- 域删除成功
- 域删除失败
- 已开始环境部署
- 已完成环境部署
- 环境部署已失败
- 环境生成已开始
- 环境删除已完成
- 环境删除失败
- 已成功创建项目
- 已成功添加项目成员
- 已成功移除项目成员
- 已成功更改项目成员角色
- 环境部署客户工作流已启动
- 已成功生成企业名称
- 企业名称生成失败

有关更多信息，请参阅 Amazon EventBridge。

亚马逊的安全 DataZone

云安全位于 AWS 是最高优先级。作为一个 AWS 客户，您将受益于专为满足最安全敏感组织的要求而构建的数据中心和网络架构。

安全是双方的共同责任 AWS 还有你。责任共担模型将此描述为云的安全性与云中的安全性：

- 云的安全 — AWS 负责保护正在运行的基础架构 AWS 中的服务 AWS Cloud。 AWS 还为您提供可以安全使用的服务。 Third-party 审计师定期测试和验证我们的安全有效性，这是其中的一部分 AWS 合规计划。了解适用于亚马逊的合规计划 DataZone，参见 AWS 合规计划范围内的服务。
- 云端安全 — 您的责任由 AWS 您使用的服务。您还需要对其他因素负责，包括您的数据的敏感性、您公司的要求以及适用的法律法规。

本文档可帮助您了解在使用亚马逊时如何应用分担责任模式 DataZone。以下主题向您展示如何配置亚马逊 DataZone 以实现您的安全与合规性目标。你还将学习如何使用其他 AWS 帮助您监控和保护您的亚马逊的服务 DataZone 资源。

主题

- 亚马逊的数据保护 DataZone
- 亚马逊授权 DataZone
- 控制对亚马逊的访问权限 DataZone 使用 IAM 的资源
- 亚马逊合规性验证 DataZone
- 亚马逊安全最佳实践 DataZone
- 亚马逊的弹性 DataZone
- 亚马逊的基础设施安全 DataZone
- Cross-service 亚马逊的副手防范措施令人困惑 DataZone
- 亚马逊的配置和漏洞分析 DataZone
- 要添加到允许列表的域

亚马逊的数据保护 DataZone

这个 AWS 责任共担模式 适用于亚马逊的数据保护 DataZone。如本模型所述， AWS 负责保护运行所有网络的全球基础设施 AWS Cloud。您负责维护对托管在此基础结构上的内容的控制。您还负责您

所使用的 AWS 服务 的安全配置和管理任务。有关数据隐私的更多信息，请参阅 数据隐私常见问题解答。 有关欧洲数据保护的信息，请参阅通用数据保护条例（GDPR）中心。

出于数据保护的目的，我们建议您保护 AWS 账户 凭据并使用以下方式设置个人用户 AWS IAM Identity Center 要么 AWS Identity and Access Management （我是）。这样，每个用户只获得履行其工作职责所需的权限。还建议您通过以下方式保护数据：

- 对每个账户使用多重身份验证（MFA）。
- 使用 SSL/TLS 与之沟通 AWS 资源。我们要求使用 TLS 1.2，并建议使用 TLS 1.3。
- 使用设置 API 和用户活动日志 AWS CloudTrail。有关使用的信息 CloudTrail 有待捕捉的踪迹 AWS 活动，请参阅 与... 合作 CloudTrail 小径 在 AWS CloudTrail 用户指南。
- 使用 AWS 加密解决方案，以及其中的所有默认安全控件 AWS 服务。
- 使用高级托管安全服务（如 Amazon Macie），它有助于发现和保护存储在 Amazon S3 中的敏感数据。
- 如果您在访问时需要经过 FIPS 140-3 验证的加密模块 AWS 通过命令行接口或 API，使用 FIPS 端点。有关可用的 FIPS 端点的更多信息，请参阅《美国联邦信息处理标准（FIPS）第 140-3 版》https://aws.amazon.com/compliance/fips/。

强烈建议您切勿将机密信息或敏感信息（如客户的电子邮件地址）放入标签或自由格式文本字段（如名称字段）。这包括当你与亚马逊合作时 DataZone 或者其他 AWS 服务 使用控制台、API、 AWS CLI，或 AWS 软件开发工具包。在用于名称的标签或自由格式文本字段中输入的任何数据都可能会用于计费或诊断日志。如果您向外部服务器提供 URL，强烈建议您不要在网址中包含凭证信息来验证对该服务器的请求。

数据加密

授予权限时，由您决定谁获得哪个亚马逊的权限 DataZone 资源。您可以对这些资源启用希望允许的特定操作。因此，您应仅授予执行任务所需的权限。实施最低权限访问对于减小安全风险以及可能由错误或恶意意图造成的影响至关重要。

静态加密

亚马逊 DataZone 默认情况下，使用加密您的所有数据 AWS 密钥管理服务 (AWS KMS) 把那个关键 AWS 为您拥有和管理。您也可以加密存储在亚马逊中的数据 DataZone 使用您管理的密钥进行编录 AWS 千米。

当您在亚马逊创建域名时 DataZone，您可以通过选中旁边的复选框来提供加密设置 自定义加密设置（高级）下 数据加密，并提供 KMS 密钥。

传输中加密

亚马逊 DataZone 使用传输层安全 (TLS) 和客户端加密进行传输加密。与亚马逊的沟通 DataZone 始终通过 HTTPS 完成，因此您的数据在传输过程中始终处于加密状态。

Inter-network 交通隐私

为了保护账户之间的连接，亚马逊 DataZone 使用服务角色和 IAM 角色安全地连接到客户账户并代表客户执行操作。

主题

- 亚马逊的静态数据加密 DataZone
- 使用亚马逊接口 VPC 终端节点 DataZone

亚马逊的静态数据加密 DataZone

默认情况下，静态数据加密有助于降低保护敏感数据的操作开销和复杂性。同时，它还支持构建符合严格加密合规性和监管要求的安全应用程序。

亚马逊 DataZone 使用默认值 AWS-拥有的密钥可自动加密您的静态数据。您无法查看、管理或审核的使用情况 AWS 拥有的密钥。有关更多信息，请参阅 AWS 拥有的密钥。

虽然您无法禁用此加密层或选择其他加密类型，但可以在创建亚马逊时选择客户管理的密钥 DataZone 域。亚马逊 DataZone 支持使用您可以创建、拥有和管理的对称客户管理密钥。由于您能够完全控制加密，因此可执行以下任务：

- 建立和维护密钥策略
- 创建和维护 IAM 策略和授权
- 启用和禁用密钥策略
- 轮换密钥加密材料
- 添加标签
- 创建密钥别名
- 计划密钥删除

要使用自己的密钥，请在创建亚马逊时选择客户管理的密钥 DataZone 域。

有关更多信息，请参阅客户自主管理型密钥。

> **Note**
>
> 亚马逊 DataZone 使用自动启用静态加密 AWS 拥有免费保护客户数据的密钥。AWS 使用客户管理的密钥会收取 KMS 费用。有关定价的更多信息，请参阅 AWS Key Management Service 定价。

亚马逊怎么样 DataZone 将补助金用于 AWS KMS

亚马逊 DataZone 需要两个 补助 使用您的客户管理密钥。当你创建亚马逊时 DataZone 使用客户管理密钥加密的域名，亚马逊 DataZone 通过发送来代表您创建补助金 CreateGrant 的请求 AWS 千米。补助金 AWS KMS 用于向亚马逊提供 DataZone 访问您账户中的 KMS 密钥。亚马逊 DataZone 创建以下授权，使用您的客户管理密钥进行以下内部操作：

一项用于为以下操作加密静态数据的授权：

- 发送 DescribeKey 的请求 AWS KMS 用于验证在创建亚马逊时输入的对称客户管理的 KMS 密钥 ID DataZone 域名有效。
- 发送 GenerateDataKey 到 AWS KMS 生成由您的客户管理密钥加密的数据密钥。
- 发送 解密 请求启用亚马逊 DataZone 解密存储的数据。
- RetireGrant 在删除域名时取消授权。

一笔用于搜索、发现和 出口 你的数据：

- DescribeKey -提供允许亚马逊使用的客户管理密钥详细信息 DataZone 来验证密钥。
- 解密 -允许亚马逊 DataZone 解密存储的数据。

您可以随时撤消对指向客户自主管理型密钥的授权的访问权限。如果你这样做，亚马逊 DataZone 将无法访问由客户管理的密钥加密的任何数据，这会影响依赖该数据的操作。

创建客户托管密钥

您可以使用创建对称的客户管理密钥 AWS 管理控制台，或 AWS KMS API。

要创建对称的客户管理密钥，请按照以下步骤操作 创建对称的客户管理密钥 在 AWS 密钥管理服务开发人员指南。

密钥策略 – 密钥策略控制对客户自主管理型密钥的访问。每个客户托管式密钥必须只有一个密钥策略，其中包含确定谁可以使用密钥以及如何使用密钥的声明。创建客户托管式密钥时，可以指定密钥策略。有关更多信息，请参阅 管理对客户管理密钥的访问权限 在 AWS 密钥管理服务开发人员指南。

在亚马逊上使用您的客户管理密钥 DataZone 资源，密钥策略中必须允许以下 API 操作：

- kms:CreateGrant — 向客户管理的密钥添加授权。授予对指定 KMS 密钥的控制访问权限，这允许访问 拨款业务 亚马逊 DataZone 要求。有关以下内容的更多信息 使用补助金，请参阅 AWS 密钥管理服务开发人员指南。
- kms:DescribeKey — 提供客户管理的密钥详细信息以允许亚马逊 DataZone 来验证密钥。
- kms:GenerateDataKey — 返回唯一的对称数据密钥以供在外部使用 AWS 千米。
- kms:Decrypt – 解密已通过 KMS 密钥加密的加密文字。

以下是您可以为亚马逊添加的政策声明示例 DataZone:

```
"Statement": [
    {
      "Sid": "Enable IAM User Permissions for DescribeKey",
      "Effect": "Allow",
      "Principal": {
        "AWS": "arn:aws:iam::111122223333:root"
      },
      "Action": "kms:DescribeKey",
      "Resource": "arn:aws:kms:us-east-1:111122223333:key/key_ID"
    },
    {
      "Sid": "Allow access to principals authorized to manage Amazon DataZone",
      "Effect": "Allow",
      "Principal": {
        "AWS": "arn:aws:iam::111122223333:root"
      },
      "Action": [
        "kms:Decrypt",
        "kms:GenerateDataKey"
      ],
      "Resource": "arn:aws:kms:us-east-1:111122223333:key/key_ID",
      "Condition": {
```

```
        "ForAnyValue:StringEquals": {
          "kms:EncryptionContextKeys": "aws:datazone:domainId"
        }
      }
    },
    {
      "Sid": "Allow creating grants when creating an Amazon DataZone for all principals
 in the account that are authorized to manage Amazon DataZone",
      "Effect": "Allow",
      "Principal": {
        "AWS": "arn:aws:iam::111122223333:root"
      },
      "Action": "kms:CreateGrant",
      "Resource": "arn:aws:kms:us-east-1:111122223333:key/key_ID",
      "Condition": {
        "StringLike": {
          "kms:CallerAccount": "111122223333",
          "kms:ViaService": "datazone.us-east-1.amazonaws.com"
        },
        "Bool": {
          "kms:GrantIsForAWSResource": "true"
        },
        "ForAnyValue:StringEquals": {
          "kms:EncryptionContextKeys": "aws:datazone:domainId"
        }
      }
    }
]
```

> **Note**
>
> 亚马逊 DataZone 数据门户被授予通过域执行角色主体访问您的客户管理密钥的权限。

有关以下内容的更多信息 在策略中指定权限，请参阅 AWS 密钥管理服务开发人员指南。

有关以下内容的更多信息 密钥访问疑难解答，请参阅 AWS 密钥管理服务开发人员指南。

为亚马逊指定客户管理密钥 DataZone

在域创建过程中，您可以将客户自主管理型密钥指定为第二层加密。

亚马逊 DataZone 加密上下文

加密上下文是一组可选的键值对，包含有关数据的其他上下文信息。

AWS KMS 使用加密上下文作为 其他经过身份验证的数据 来支持 经过验证的加密。当您在加密数据的请求中包含加密上下文时， AWS KMS 将加密上下文绑定到加密的数据。要解密数据，您必须在请求中包含相同的加密上下文。

亚马逊 DataZone 使用以下加密上下文：

```
"encryptionContextSubset": {
    "aws:datazone:domainId": "{dzd_samleid}"
}
```

使用加密上下文进行监控 -当您使用对称的客户管理密钥加密亚马逊时 DataZone，您还可以在审计记录和日志中使用加密上下文来确定如何使用客户管理的密钥。加密上下文也出现在由生成的日志中 AWS CloudTrail 或者亚马逊 CloudWatch 日志。

使用加密上下文控制对客户自主管理型密钥的访问 – 您可以使用密钥策略和 IAM 策略中的加密上下文作为条件来控制对您的对称客户自主管理型密钥的访问。您也可以在授予中使用加密上下文约束。

亚马逊 DataZone 在授权中使用加密上下文约束来控制对您账户或地区的客户管理密钥的访问权限。授权约束要求授权允许的操作使用指定的加密上下文。

以下是密钥策略声明示例，用于授予对特定加密上下文的客户托管密钥的访问权限。

```
{
     "Sid": "Enable DescribeKey",
     "Effect": "Allow",
     "Principal": {
       "AWS": "arn:aws:iam::111122223333:role/ExampleRole"
     },
     "Action": "kms:DescribeKey",
     "Resource": "arn:aws:kms:region:111122223333:key/key_ID"
   },
   {
     "Sid": "Allow access to principal to manage an Amazon DataZone domain with the
given domain id",
     "Effect": "Allow",
     "Principal": {
```

```
        "AWS": "arn:aws:iam::111122223333:role/ExampleRole"
      },
      "Action": [
        "kms:Decrypt",
        "kms:GenerateDataKey"
      ],
      "Resource": "arn:aws:kms:region:111122223333:key/key_ID",
      "Condition": {
        "StringEquals": {
          "kms:EncryptionContext:aws:datazone:domainId": "dzd_sampleid"
        }
      }
    },
    {
      "Sid": "Allow creating grants when creating an Amazon DataZone domain to
principal",
      "Effect": "Allow",
      "Principal": {
        "AWS": "arn:aws:iam::111122223333:role/ExampleRole"
      },
      "Action": "kms:CreateGrant",
      "Resource": "arn:aws:kms:region:111122223333:key/key_ID",
      "Condition": {
        "StringLike": {
          "kms:CallerAccount": "111122223333",
          "kms:ViaService": "datazone.region.amazonaws.com"
        },
        "Bool": {
          "kms:GrantIsForAWSResource": "true"
        },
        "ForAnyValue:StringEquals": {
          "kms:EncryptionContextKeys": "aws:datazone:domainId"
        }
      }
    }
```

监控您的亚马逊加密密钥 DataZone

当你使用 AWS 您的亚马逊的 KMS 客户管理密钥 DataZone 资源，你可以使用 AWS CloudTrail 追踪亚马逊的请求 DataZone 发送到 AWS 千米。以下示例是 AWS CloudTrail 的活动 `CreateGrant`，`GenerateDataKey`，`Decrypt`，以及 `RetireGrant` 监控亚马逊调用的 KMS 操作 DataZone 访问由您的客户管理密钥加密的数据。

CreateGrant

当你使用 AWS 用于加密您的亚马逊的 KMS 客户管理密钥 DataZone 域名，亚马逊 DataZone 发送 `CreateGrant` 代表您请求访问您的 KMS 密钥 AWS 账户。赞成亚马逊 DataZone 创建的内容特定于与关联的资源 AWS KMS 客户管理密钥。此外，亚马逊 DataZone 使用 `RetireGrant` 删除域名时移除授权的操作。

以下示例事件记录了 `CreateGrant` 操作：

```
{
    "eventVersion": "1.11",
    "userIdentity": {
        "type": "AssumedRole",
        "principalId": "AROAIGDTESTANDEXAMPLE:Sampleuser01",
        "arn": "arn:aws:sts::111122223333:assumed-role/Example/Sampleuser01",
        "accountId": "111122223333",
        "accessKeyId": "AKIAIOSFODNN7EXAMPLE3",
        "sessionContext": {
            "sessionIssuer": {
                "type": "Role",
                "principalId": "AROAIGDTESTANDEXAMPLE",
                "arn": "arn:aws:iam::111122223333:role/Example",
                "accountId": "111122223333",
                "userName": "Example"
            },
            "attributes": {
                "creationDate": "2024-04-22T17:02:00Z",
                "mfaAuthenticated": "false"
            }
        },
        "invokedBy": "datazone.amazonaws.com"
    },
    "eventTime": "2024-04-22T17:02:00Z",
    "eventSource": "kms.amazonaws.com",
    "eventName": "CreateGrant",
    "awsRegion": "us-east-2",
    "sourceIPAddress": "datazone.amazonaws.com",
    "userAgent": "datazone.amazonaws.com",
    "requestParameters": {
        "retiringPrincipal": "datazone.us-east-2.amazonaws.com",
        "operations": [
            "GenerateDataKey",
```

```
            "RetireGrant",
            "DescribeKey",
            "Decrypt"
        ],
        "granteePrincipal": "datazone.us-east-2.amazonaws.com",
        "constraints": {
            "encryptionContextSubset": {
                "aws:datazone:domainId": "dzd_sampleid"
            }
        },
        "keyId": "arn:aws:kms:us-
east-2:111122223333:key/1234abcd-12ab-34cd-56ef-123456SAMPLE"
    },
    "responseElements": {
        "grantId":
 "0ab0ac0d0b000f00ea00cc0a0e00fc00bce000c000f0000000c0bc0a0000aaafSAMPLE",
        "keyId": "arn:aws:kms:us-
east-2:111122223333:key/1234abcd-12ab-34cd-56ef-123456SAMPLE"
    },
    "requestID": "ff000af-00eb-00ce-0e00-ea000fb0fba0SAMPLE",
    "eventID": "ff000af-00eb-00ce-0e00-ea000fb0fba0SAMPLE",
    "readOnly": false,
    "resources": [
        {
            "accountId": "111122223333",
            "type": "AWS::KMS::Key",
            "ARN": "arn:aws:kms:us-
east-2:111122223333:key/1234abcd-12ab-34cd-56ef-123456SAMPLE"
        }
    ],
    "eventType": "AwsApiCall",
    "managementEvent": true,
    "recipientAccountId": "111122223333",
    "eventCategory": "Management",
    "sessionCredentialFromConsole": "true"
}
```

```
{
    "eventVersion": "1.11",
    "userIdentity": {
        "type": "AssumedRole",
```

```
        "principalId": "AROAIGDTESTANDEXAMPLE:Sampleuser01",
        "arn": "arn:aws:sts::111122223333:assumed-role/Example/Sampleuser01",
        "accountId": "111122223333",
        "accessKeyId": "AKIAIOSFODNN7EXAMPLE3",
        "sessionContext": {
            "sessionIssuer": {
                "type": "Role",
                "principalId": "AROAIGDTESTANDEXAMPLE",
                "arn": "arn:aws:iam::111122223333:role/Example",
                "accountId": "111122223333",
                "userName": "Example"
            },
            "attributes": {
                "creationDate": "2024-04-22T17:10:00Z",
                "mfaAuthenticated": "false"
            }
        },
        "invokedBy": "datazone.amazonaws.com"
    },
    "eventTime": "2024-04-22T17:49:00Z",
    "eventSource": "kms.amazonaws.com",
    "eventName": "CreateGrant",
    "awsRegion": "us-east-2",
    "sourceIPAddress": "datazone.amazonaws.com",
    "userAgent": "datazone.amazonaws.com",
    "requestParameters": {
        "retiringPrincipal": "datazone.us-east-2.amazonaws.com",
        "operations": [
            "DescribeKey",
            "Decrypt"
        ],
        "granteePrincipal": "datazone.us-east-2.amazonaws.com",
        "constraints": {
            "encryptionContextSubset": {
                "aws:datazone:domainId": "dzd_sampleid"
            }
        },
        "keyId": "arn:aws:kms:us-
east-2:111122223333:key/1234abcd-12ab-34cd-56ef-123456SAMPLE"
    },
    "responseElements": {
        "grantId":
 "0ab0ac0d0b000f00ea00cc0a0e00fc00bce000c000f0000000c0bc0a0000aaafSAMPLE",
```

```
        "keyId": "arn:aws:kms:us-
east-2:111122223333:key/1234abcd-12ab-34cd-56ef-123456SAMPLE"
    },
    "requestID": "ff000af-00eb-00ce-0e00-ea000fb0fba0SAMPLE",
    "eventID": "ff000af-00eb-00ce-0e00-ea000fb0fba0SAMPLE",
    "readOnly": false,
    "resources": [
        {
            "accountId": "111122223333",
            "type": "AWS::KMS::Key",
            "ARN": "arn:aws:kms:us-
east-2:111122223333:key/1234abcd-12ab-34cd-56ef-123456SAMPLE"
        }
    ],
    "eventType": "AwsApiCall",
    "managementEvent": true,
    "recipientAccountId": "111122223333",
    "eventCategory": "Management",
    "sessionCredentialFromConsole": "true"
}
```

GenerateDataKey

当你启用 AWS 您的亚马逊的 KMS 客户管理密钥 DataZone 域名，亚马逊 DataZone 生成数据密钥。它发送一个 GenerateDataKey 请求 AWS 指定 KMS AWS 该域的 KMS 客户管理密钥。

以下示例事件记录了 GenerateDataKey 操作：

```
{
    "eventVersion": "1.11",
    "userIdentity": {
        "type": "AssumedRole",
        "principalId": "AROAIGDTESTANDEXAMPLE:AmazonSageMakerDomainExecution",
        "arn": "arn:aws:sts::111122223333:assumed-role/
AmazonSageMakerDomainExecution/AmazonSageMakerDomainExecution",
        "accountId": "111122223333",
        "accessKeyId": "AKIAIOSFODNN7EXAMPLE3",
        "sessionContext": {
            "sessionIssuer": {
                "type": "Role",
                "principalId": "AROAIGDTESTANDEXAMPLE",
```

```
                "arn": "arn:aws:iam::111122223333:role/service-role/
AmazonSageMakerDomainExecution",
                "accountId": "111122223333",
                "userName": "AmazonSageMakerDomainExecution"
            },
            "attributes": {
                "creationDate": "2024-04-22T19:50:39Z",
                "mfaAuthenticated": "false"
            }
        },
        "invokedBy": "datazone.amazonaws.com"
    },
    "eventTime": "2024-04-22T19:50:40Z",
    "eventSource": "kms.amazonaws.com",
    "eventName": "GenerateDataKey",
    "awsRegion": "us-east-2",
    "sourceIPAddress": "datazone.amazonaws.com",
    "userAgent": "datazone.amazonaws.com",
    "requestParameters": {
        "keySpec": "AES_256",
        "encryptionContext": {
            "aws:datazone:domainId": "dzd_sampleid",
            "V": "2024-04-22T17:49:12.98177136Z|cacf3df7-7b99-49f6-ae14-sample",
            "version": "0",
            "N": "dzd_sampleid|arn:aws:kms:us-
east-2:111122223333:key/1234abcd-12ab-34cd-56ef-123456SAMPLE",
            "*aws-kms-table*": "awsdatazoneroaring-data-store-datakeys-prod-us-
east-2"
        },
        "keyId": "arn:aws:kms:us-
east-2:111122223333:key/1234abcd-12ab-34cd-56ef-123456SAMPLE"
    },
    "responseElements": null,
    "requestID": "ff000af-00eb-00ce-0e00-ea000fb0fba0SAMPLE",
    "eventID": "ff000af-00eb-00ce-0e00-ea000fb0fba0SAMPLE",
    "readOnly": true,
    "resources": [
        {
            "accountId": "111122223333",
            "type": "AWS::KMS::Key",
            "ARN": "arn:aws:kms:us-
east-2:111122223333:key/1234abcd-12ab-34cd-56ef-123456SAMPLE"
        }
    ],
```

```
    "eventType": "AwsApiCall",
    "managementEvent": true,
    "recipientAccountId": "111122223333",
    "eventCategory": "Management"
}
```

```
{
    "eventVersion": "1.11",
    "userIdentity": {
        "type": "AWSService",
        "invokedBy": "AWS Internal"
    },
    "eventTime": "2024-04-22T19:50:40Z",
    "eventSource": "kms.amazonaws.com",
    "eventName": "GenerateDataKey",
    "awsRegion": "us-east-2",
    "sourceIPAddress": "AWS Internal",
    "userAgent": "AWS Internal",
    "requestParameters": {
        "encryptionContext": {
            "aws:datazone:domainId": "dzd_sampleid",
            "aws:s3:arn": "arn:aws:s3:::amazon-datazone-us-
east-2-422ceee9465430bdb354d1c9efsample"
        },
        "keyId": "arn:aws:kms:us-
east-2:111122223333:key/1234abcd-12ab-34cd-56ef-123456SAMPLE",
        "keySpec": "AES_256"
    },
    "responseElements": null,
    "requestID": "ff000af-00eb-00ce-0e00-ea000fb0fba0SAMPLE",
    "eventID": "ff000af-00eb-00ce-0e00-ea000fb0fba0SAMPLE",
    "readOnly": true,
    "resources": [
        {
            "accountId": "111122223333",
            "type": "AWS::KMS::Key",
            "ARN": "arn:aws:kms:us-
east-2:111122223333:key/1234abcd-12ab-34cd-56ef-123456SAMPLE"
        }
    ],
    "eventType": "AwsApiCall",
```

```
    "managementEvent": true,
    "recipientAccountId": "111122223333",
    "sharedEventID": "ff000af-00eb-00ce-0e00-ea000fb0fba0SAMPLE",
    "eventCategory": "Management"
}
```

Decrypt

当您访问加密的亚马逊时 DataZone 域名，亚马逊 DataZone 打电话给 `Decrypt` 使用存储的加密数据密钥访问加密数据的操作。

以下示例事件记录了 `Decrypt` 操作：

```
{
    "eventVersion": "1.11",
    "userIdentity": {
        "type": "AssumedRole",
        "principalId": "AROAIGDTESTANDEXAMPLE:AmazonSageMakerDomainExecution",
        "arn": "arn:aws:sts::111122223333:assumed-role/
AmazonSageMakerDomainExecution/AmazonSageMakerDomainExecution",
        "accountId": "111122223333",
        "accessKeyId": "AKIAIOSFODNN7EXAMPLE3",
        "sessionContext": {
            "sessionIssuer": {
                "type": "Role",
                "principalId": "AROAIGDTESTANDEXAMPLE",
                "arn": "arn:aws:iam::111122223333:role/service-role/
AmazonSageMakerDomainExecution",
                "accountId": "111122223333",
                "userName": "AmazonSageMakerDomainExecution"
            },
            "attributes": {
                "creationDate": "2024-04-22T19:50:39Z",
                "mfaAuthenticated": "false"
            }
        },
        "invokedBy": "datazone.amazonaws.com"
    },
    "eventTime": "2024-04-22T19:51:54Z",
    "eventSource": "kms.amazonaws.com",
    "eventName": "Decrypt",
    "awsRegion": "us-east-2",
```

```
    "sourceIPAddress": "datazone.amazonaws.com",
    "userAgent": "datazone.amazonaws.com",
    "requestParameters": {
        "encryptionAlgorithm": "SYMMETRIC_DEFAULT",
        "encryptionContext": {
            "aws:datazone:domainId": "dzd_sampleid",
            "V": "2024-04-22T17:49:12.98177136Z|cacf3df7-7b99-49f6-ae14-sample",
            "version": "0",
            "N": "dzd_sampleid|arn:aws:kms:us-
east-2:111122223333:key/1234abcd-12ab-34cd-56ef-123456SAMPLE",
            "*aws-kms-table*": "awsdatazoneroaring-data-store-datakeys-prod-us-
east-2"
        }
    },
    "responseElements": null,
    "requestID": "ff000af-00eb-00ce-0e00-ea000fb0fba0SAMPLE",
    "eventID": "ff000af-00eb-00ce-0e00-ea000fb0fba0SAMPLE",
    "readOnly": true,
    "resources": [
        {
            "accountId": "111122223333",
            "type": "AWS::KMS::Key",
            "ARN": "arn:aws:kms:us-
east-2:111122223333:key/1234abcd-12ab-34cd-56ef-123456SAMPLE"
        }
    ],
    "eventType": "AwsApiCall",
    "managementEvent": true,
    "recipientAccountId": "111122223333",
    "eventCategory": "Management"
}
```

```
{
    "eventVersion": "1.11",
    "userIdentity": {
        "type": "AWSService",
        "invokedBy": "datazone.amazonaws.com"
    },
    "eventTime": "2024-04-22T19:51:54Z",
    "eventSource": "kms.amazonaws.com",
    "eventName": "Decrypt",
```

```
    "awsRegion": "us-east-2",
    "sourceIPAddress": "datazone.amazonaws.com",
    "userAgent": "datazone.amazonaws.com",
    "requestParameters": {
        "encryptionContext": {
            "aws:datazone:domainId": "dzd_sampleid",
            "V": "2024-04-22T17:49:12.98177136Z|cacf3df7-7b99-49f6-ae14-sample",
            "version": "0",
            "N": "dzd_sampleid|arn:aws:kms:us-
east-2:111122223333:key/1234abcd-12ab-34cd-56ef-123456SAMPLE",
            "*aws-kms-table*": "awsdatazoneroaring-data-store-datakeys-prod-us-
east-2"
        },
        "encryptionAlgorithm": "SYMMETRIC_DEFAULT"
    },
    "responseElements": null,
    "requestID": "ff000af-00eb-00ce-0e00-ea000fb0fba0SAMPLE",
    "eventID": "ff000af-00eb-00ce-0e00-ea000fb0fba0SAMPLE",
    "readOnly": true,
    "resources": [
        {
            "accountId": "111122223333",
            "type": "AWS::KMS::Key",
            "ARN": "arn:aws:kms:us-
east-2:111122223333:key/1234abcd-12ab-34cd-56ef-123456SAMPLE"
        }
    ],
    "eventType": "AwsApiCall",
    "managementEvent": true,
    "recipientAccountId": "111122223333",
    "sharedEventID": "ff000af-00eb-00ce-0e00-ea000fb0fba0SAMPLE",
    "eventCategory": "Management"
}
```



```
{
    "eventVersion": "1.11",
    "userIdentity": {
        "type": "AWSService",
        "invokedBy": "AWS Internal"
    },
    "eventTime": "2024-04-22T19:51:54Z",
```

```
    "eventSource": "kms.amazonaws.com",
    "eventName": "Decrypt",
    "awsRegion": "us-east-2",
    "sourceIPAddress": "AWS Internal",
    "userAgent": "AWS Internal",
    "requestParameters": {
        "encryptionAlgorithm": "SYMMETRIC_DEFAULT",
        "encryptionContext": {
            "aws:datazone:domainId": "dzd_sampleid",
            "aws:s3:arn": "arn:aws:s3:::amazon-datazone-us-
east-2-422ceee9465430bdb354d1c9efsample"
        }
    },
    "responseElements": null,
    "requestID": "ff000af-00eb-00ce-0e00-ea000fb0fba0SAMPLE",
    "eventID": "ff000af-00eb-00ce-0e00-ea000fb0fba0SAMPLE",
    "readOnly": true,
    "resources": [
        {
            "accountId": "111122223333",
            "type": "AWS::KMS::Key",
            "ARN": "arn:aws:kms:us-
east-2:111122223333:key/1234abcd-12ab-34cd-56ef-123456SAMPLE"
        }
    ],
    "eventType": "AwsApiCall",
    "managementEvent": true,
    "recipientAccountId": "111122223333",
    "sharedEventID": "ff000af-00eb-00ce-0e00-ea000fb0fba0SAMPLE",
    "eventCategory": "Management"
}
```

RetireGrant

以下示例事件记录了 RetireGrant 操作：

```
{
    "eventVersion": "1.11",
    "userIdentity": {
        "type": "AWSService",
        "invokedBy": "datazone.amazonaws.com"
    },
```

```
    "eventTime": "2025-04-29T22:18:50Z",
    "eventSource": "kms.amazonaws.com",
    "eventName": "RetireGrant",
    "awsRegion": "us-east-2",
    "sourceIPAddress": "datazone.amazonaws.com",
    "userAgent": "datazone.amazonaws.com",
    "requestParameters": null,
    "responseElements": {
        "keyId": "arn:aws:kms:us-
east-2:111122223333:key/1234abcd-12ab-34cd-56ef-123456SAMPLE"
    },
    "additionalEventData": {
        "grantId":
 "0ab0ac0d0b000f00ea00cc0a0e00fc00bce000c000f0000000c0bc0a0000aaafSAMPLE"
    },
    "requestID": "294308c0-7617-4727-b5c9-34eaf75aa8e3",
    "eventID": "273708f7-5fbb-3a90-b04d-2b3138bf0ec9",
    "readOnly": false,
    "resources": [
        {
            "accountId": "111122223333",
            "type": "AWS::KMS::Key",
            "ARN": "arn:aws:kms:us-
east-2:111122223333:key/1234abcd-12ab-34cd-56ef-123456SAMPLE"
        }
    ],
    "eventType": "AwsApiCall",
    "managementEvent": true,
    "recipientAccountId": "111122223333",
    "sharedEventID": "b46377d7-b3c3-4bfd-a257-722bd3f3411d",
    "eventCategory": "Management"
}
```


创建涉及加密的数据湖环境 AWS 胶水目录

在高级用例中，当你使用 AWS Glue 目录已加密，您必须授予访问亚马逊的权限 DataZone 使用您的客户管理的 KMS 密钥的服务。您可以通过更新自定义 KMS 策略并在密钥中添加标签来完成此操作。授予对亚马逊的访问权限 DataZone 用于处理加密数据的服务 AWS Glue 目录，完成以下操作：

- 将以下策略添加到您的自定义 KMS 密钥。有关更多信息，请参阅更改密钥策略。

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Sid": "Allow datazone environment roles to decrypt using the key",
            "Effect": "Allow",
            "Principal": {
                "AWS": "*"
            },
            "Action": [
                "kms:Decrypt"
            ],
            "Resource": "*",
            "Condition": {
                "StringEquals": {
                    "kms:EncryptionContext:glue_catalog_id":
 "<GLUE_CATALOG_ID>"
                },
                "ArnLike": {
                    "aws:PrincipalArn": [
                        "arn:aws:iam::111122223333:role/*datazone_usr*",
                        "arn:aws:iam::444455556666:role/*datazone_usr*"
                    ]
                }
            }
        },
        {
            "Sid": "Allow datazone environment roles to describe the key",
            "Effect": "Allow",
            "Principal": {
                "AWS": "*"
            },
            "Action": [
                "kms:DescribeKey"
            ],
            "Resource": "*",
            "Condition": {
                "ArnLike": {
                    "aws:PrincipalArn": [
                    "arn:aws:iam::111122223333:role/*datazone_usr*",
                    "arn:aws:iam::444455556666:role/*datazone_usr*"
```

```
                    ]
                }
            }
        }
    ]
}
```

> **Important**
>
> - 您必须使用要在其中创建环境的账户 ID 修改策略中的 "aws:PrincipalArn" ARN。您要在其中创建环境的每个账户都必须在策略中列为 "aws:PrincipalArn"。
> - 您还必须<GLUE_CATALOG_ID>替换为有效的 AWS 您的账户 ID AWS Glue 目录已找到。
> - 请注意，此政策向所有亚马逊授予使用密钥的权限 DataZone 指定账户中的环境用户角色。如果您只想允许特定的环境用户角色使用密钥，则必须指定整个环境用户角色名称，例如，arn:aws:iam::<ENVIRONMENT_ACCOUNT_ID>:role/datazone_usr_<ENVIRONMENT_ID>（其中 <ENVIRONMENT_ID> 是环境的 ID，而非通配符格式）。

- 将以下标签添加到您的自定义 KMS 密钥。有关更多信息，请参阅使用标签控制对 KMS 密钥的访问。

```
key: AmazonDataZoneEnvironment
value: all
```

使用亚马逊接口 VPC 终端节点 DataZone

如果您使用亚马逊虚拟私有云 (Amazon VPC) 来托管您的 AWS 资源，您可以在您的亚马逊 VPC 和亚马逊之间建立连接 DataZone。您可以与亚马逊使用此连接 DataZone 无需通过公共互联网。

亚马逊 VPC 允许您启动 AWS 自定义虚拟网络中的资源。可以使用 VPC 控制您的网络设置，例如 IP 地址范围、子网、路由表和网络网关。有关 VPC 的更多信息，请参阅《Amazon VPC 用户指南》https://docs.aws.amazon.com/vpc/latest/userguide/what-is-amazon-vpc.html。

将您的亚马逊 VPC 连接到亚马逊 DataZone，您必须先定义接口 VPC 终端节点，这样您就可以将您的 VPC 连接到其他 AWS 服务。该端点提供了可靠且可扩展的连接，无需互联网网关、网络地址转换

(NAT) 实例或 VPN 连接。有关如何创建 VPC 终端节点的更多信息和详细步骤，请参阅 接口 VPC 端点 (AWS PrivateLink) 在亚马逊 VPC 用户指南中。

> **Important**
>
> 在 VPC 中，终端节点策略是一种基于资源的策略，您可以将其附加到 VPC 终端节点以控制哪个 AWS 委托人可以使用该端点来访问 AWS 服务。
> 亚马逊当前发布的版本 DataZone 支持使用终端节点策略在您的亚马逊 VPC 和亚马逊之间建立和使用连接 DataZone 端点。

亚马逊授权 DataZone

亚马逊 DataZone的界面由内部的管理控制台组成 AWS 以及控制台外 Web 应用程序（数据门户）。

亚马逊 DataZone 管理控制台可用于 AWS 顶级资源 API 的管理员，包括创建和管理域， AWS 这些域的账户关联以及您想要委托亚马逊访问管理的数据源 DataZone。你可以使用亚马逊 DataZone 管理控制台，用于管理向亚马逊委托访问管理控制所需的所有 IAM 角色和配置 DataZone 为他们明确配置的服务 AWS 账户。亚马逊 DataZone 数据门户是第一方 AWS 适用于 SSO 用户的 Identity Center 应用程序。如果启用，则获得授权的 IAM 主体也可以使用控制台对数据门户进行联合身份验证，而不是使用 SSO 身份。

亚马逊 DataZone的数据门户设计主要供以下人员使用 AWS IAM 身份 Center-authenticated 用户可以管理数据访问权限并执行数据发布、发现、订阅和分析任务。

亚马逊授权 DataZone 控制台

亚马逊 DataZone 控制台授权模型使用 IAM 授权。此控制台主要供管理员用来进行设置。亚马逊 DataZone 使用域管理员的概念 AWS 账号和会员 AWS 账户，所有这些账户都使用控制台在尊重的同时建立信任关系 AWS 组织边界。

亚马逊授权 DataZone 门户

亚马逊 DataZone 数据门户授权模型是一种分层 ACL，具有静态角色原型（配置文件），其中包括管理员和查看者。例如，用户可以拥有管理员或用户的配置文件。在域级别，他们可以将域用户指定为数据所有者。在项目级别，用户既可以是所有者，也可以是贡献者。可以将这些配置文件配置为两种类型之一：用户和组。之后，这些配置文件会与域和项目关联，并且这些权限的状态将存储在关联表中。

在这种授权模式下，亚马逊 DataZone 允许用户管理用户和组权限。用户管理项目成员资格、请求项目的成员资格和批准成员资格。用户可发布数据、订阅数据和审批订阅。

当用户的数据门户客户端请求 Amazon 的 IAM 会话凭证时，用户在特定项目中执行数据分析 DataZone 根据用户在特定项目环境中的有效配置文件生成。此会话的范围限定在用户的权限和特定项目的资源。之后，用户将进入 Athena 或 Redshift 来查询相关数据，并且所有底层 IAM 工作都将完全抽象化。

亚马逊 DataZone 配置文件和角色

在对用户进行身份验证后，经过身份验证的上下文将映射到用户配置文件 ID。此用户配置文件可以具有多个不同的关联（项目所有者、域管理员等），并且可用于对用户进行授权。每个关联（例如，项目所有者、域管理员等）都具有基于上下文的某些活动的权限。例如，具有域管理员关联的用户可以创建其他域，为域分配其他域管理员以及在域中创建项目模板。项目所有者可以为其项目添加或删除项目成员，以及将资产发布到域。

控制对亚马逊的访问权限 DataZone 使用 IAM 的资源

你需要 AWS Identity and Access Management (IAM) 完成以下与安全相关的任务：

- 在 AWS 账户下创建用户和组。
- 为 AWS 账户下的每个用户分配唯一安全凭证。
- 使用以下命令控制每个用户执行任务的权限 AWS 资源。
- 允许用户进入另一个 AWS 账户 分享你的 AWS 资源。
- 为你创建角色 AWS 账户 并定义可以假设他们的用户或服务。
- 使用企业的现有身份授予使用以下方式执行任务的权限 AWS 资源

有关 IAM 的更多信息，请参阅以下文档：

- AWS Identity and Access Management （我是）
- IAM 入门
- IAM 用户指南

以下部分描述了设置亚马逊所需的政策和权限 DataZone 及其组件，例如域（包括域）、关联账户、项目和数据源。有关更多信息，请参阅 亚马逊 DataZone 术语和概念。

内容

- AWS 亚马逊的管理政策 DataZone

- 亚马逊的 IAM 角色 DataZone
- 临时证书
- 主体权限

AWS 亚马逊的管理政策 DataZone

一个 AWS 托管策略是一个独立的策略，由创建和管理 AWS。 AWS 托管策略旨在为许多常见用例提供权限，以便您可以开始为用户、群组和角色分配权限。

请记住 AWS 托管策略可能不会为您的特定用例授予最低权限权限，因为它们适用于所有人 AWS 可供客户使用。我们建议通过定义特定于使用案例的客户管理型策略来进一步减少权限。

您无法更改中定义的权限 AWS 托管策略。如果 AWS 更新中定义的权限 AWS 托管策略，此更新会影响该策略所关联的所有主体身份（用户、组和角色）。 AWS 最有可能更新 AWS 新建时的托管策略 AWS 服务 已启动或新的 API 操作可用于现有服务。

有关更多信息，请参阅《IAM 用户指南》中的 AWS 托管式策略。

内容

- AWS 管理策略：AmazonDataZoneFullAccess
- AWS 管理策略： AmazonDataZoneFullUserAccess
- AWS 管理策略：AmazonDataZoneEnvironmentRolePermissionsBoundary
- AWS 管理策略：AmazonDataZoneRedshiftGlueProvisioningPolicy
- AWS 管理策略：AmazonDataZoneGlueManageAccessRolePolicy
- AWS 管理策略：AmazonDataZoneRedshiftManageAccessRolePolicy
- AWS 管理策略：AmazonDataZoneDomainExecutionRolePolicy
- AWS 管理策略：AmazonDataZoneSageMakerProvisioningRolePolicy
- AWS 管理策略：AmazonDataZoneSageMakerManageAccessRolePolicy
- AWS 管理策略：AmazonDataZoneSageMakerEnvironmentRolePermissionsBoundary
- 亚马逊 DataZone 更新到 AWS 托管策略

AWS 管理策略：AmazonDataZoneFullAccess

您可以将 `AmazonDataZoneFullAccess` 策略附加到 IAM 身份。

本政策提供对亚马逊的完全访问权限 DataZone 通过 AWS 管理控制台。此政策还具有以下权限：AWS 用于加密 SSM 参数的 KMS。必须使用 KMS 密钥进行标记 EnableKeyForAmazonDataZone 允许解密 SSM 参数。

权限详细信息

该策略包含以下权限：

- datazone — 授予委托人对亚马逊的完全访问权限 DataZone 通过 AWS 管理控制台。
- kms – 允许主体列出别名、描述密钥和解密密钥。
- s3 — 允许委托人选择现有或创建新的 S3 存储桶来存储 Amazon DataZone 数据。
- ram — 允许委托人共享亚马逊 DataZone 跨域名 AWS 账户。
- iam – 允许主体列出和传递角色并获取策略。
- sso – 允许主体获取已启用 AWS IAM Identity Center 的区域。
- secretsmanager – 允许主体创建、标记和列出带特定前缀的密钥。
- aoss — 允许负责人创建和检索信息 OpenSearch 无服务器安全策略。
- bedrock – 允许主体创建、列出和检索推理配置文件和基础模型的信息。
- codeconnections – 允许主体删除、检索信息、列出连接和管理连接的标签。
- codewhisperer — 允许校长列出 CodeWhisperer 配置文件。
- ssm – 允许主体放置、删除和检索参数信息。
- redshift – 允许主体描述集群和列出无服务器工作组
- glue – 允许主体获取数据库。

要查看此策略的权限，请参阅《AWS 托管式策略参考》中的 AmazonDataZoneFullAccess。

策略注意事项和限制

AmazonDataZoneFullAccess 策略未涵盖某些功能。

- 如果你创建了亚马逊 DataZone 用你自己的域名 AWS KMS 密钥，你必须有权限 kms:CreateGrant 为了成功创建域名，以及 kms:GenerateDataKey， kms:Decrypt 让该密钥调用其他亚马逊 DataZone API，例如 listDataSources 和 createDataSource。此外，您还必须在该密钥的资源策略中拥有 kms:CreateGrant、kms:Decrypt、kms:GenerateDataKey 和 kms:DescribeKey 的权限。

  如果您使用默认的服务拥有的 KMS 密钥，则无需达到此要求。

有关更多信息，请参阅 AWS Key Management Service。

- 如果你想使用 创造 和 更新 亚马逊内部的角色功能 DataZone 控制台，您必须拥有管理员权限或拥有创建 IAM 角色所需的 IAM 权限以及 create/update 政策。所需的权限包括 `iam:CreateRole`、`iam:CreatePolicy`、`iam:CreatePolicyVersion`、`iam:DeletePolicyVersi` 和 `iam:AttachRolePolicy` 权限。
- 如果您在亚马逊创建新域名 DataZone 和 AWS IAM Identity Center 用户登录已激活，或者如果您为亚马逊的现有域名激活了登录 DataZone，您必须拥有以下权限：
  - 组织：DescribeOrganization
  - 组织：ListDelegatedAdministrators
  - sso:CreateInstance
  - sso:ListInstances
  - sso:GetSharedSsoConfiguration
  - sso:PutApplicationGrant
  - sso:PutApplicationAssignmentConfiguration
  - sso:PutApplicationAuthenticationMethod
  - sso:PutApplicationAccessScope
  - sso:CreateApplication
  - sso:DeleteApplication
  - sso:CreateApplicationAssignment
  - sso:DeleteApplicationAssignment
  - sso 目录：CreateUser
  - sso 目录：SearchUsers
  - sso:ListApplications
- 为了接受 AWS 亚马逊账户关联请求 DataZone，你必须有 `ram:AcceptResourceShareInvitation` 许可。
- 如果你想为创建所需的资源 SageMaker Unified Studio 网络设置，您必须拥有以下权限并连接 AmazonVpcFullAccess 政策：
  - 我是：PassRole
  - 云层形成：CreateStack

AWS 管理策略： AmazonDataZoneFullUserAccess

该政策授予对亚马逊的完全访问权限 DataZone，但它不允许管理域、用户或关联帐户。

权限详细信息

要查看此策略的权限，请参阅《AWS 托管式策略参考》中的 AmazonDataZoneFullUserAccess。

AWS 管理策略：AmazonDataZoneEnvironmentRolePermissionsBoundary

> **Note**
>
> 此策略是权限边界。权限边界设置基于身份的策略可以授予 IAM 实体的最大权限。您不应使用和附上亚马逊 DataZone 权限边界策略由您自己决定。亚马逊 DataZone 权限边界政策只能附加到亚马逊 DataZone 托管角色。有关权限边界的更多信息，请参阅《IAM 用户指南》中的 IAM 实体的权限边界。

当你通过亚马逊创建环境时 DataZone 数据门户，亚马逊 DataZone 将此权限边界应用于 在创建环境期间生成的 IAM 角色。权限边界限制了亚马逊角色的范围 DataZone 创建以及您添加的任何角色。

亚马逊 DataZone 使用 `AmazonDataZoneEnvironmentRolePermissionsBoundary` 托管策略限制其附加到的预配置 IAM 委托人。校长可以采取这样的形式 用户角色 那个亚马逊 DataZone 可以代表交互式企业用户或分析服务假设 (AWS Glue，例如），然后执行操作来处理数据，例如从 Amazon S3 读取和写入或运行 AWS Glue 爬网程序。

这个 `AmazonDataZoneEnvironmentRolePermissionsBoundary` 政策授予亚马逊读写权限 DataZone 转到诸如此类的服务 AWS Glue，亚马逊 S3， AWS Lake Formation、亚马逊 Redshift 和亚马逊 Athena。该策略还为使用这些服务所需的某些基础设施资源（例如网络接口和）提供读写权限。 AWS KMS 钥匙。

亚马逊 DataZone 适用 `AmazonDataZoneEnvironmentRolePermissionsBoundary` AWS 托管策略作为所有亚马逊的权限边界 DataZone 环境角色（所有者和贡献者）。该权限边界限制这些角色只能访问环境所需的资源和操作。

要查看此策略的权限，请参阅《AWS 托管式策略参考》中的 AmazonDataZoneEnvironmentRolePermissionsBoundary。

AWS 管理策略：AmazonDataZoneRedshiftGlueProvisioningPolicy

这个 AmazonDataZoneRedshiftGlueProvisioningPolicy 政策补助亚马逊 DataZone 与进行互操作所需的权限 AWS Glue 和亚马逊 Redshift。

要查看此策略的权限，请参阅《AWS 托管式策略参考》中的 AmazonDataZoneRedshiftGlueProvisioningPolicy。

AWS 管理策略：AmazonDataZoneGlueManageAccessRolePolicy

该政策赋予亚马逊 DataZone 发布权限 AWS 将数据粘贴到目录中。它还给了亚马逊 DataZone 授予访问权限或撤消访问权限的权限 AWS 将已发布的资源粘贴到目录中。

要查看此策略的权限，请参阅《AWS 托管式策略参考》中的 AmazonDataZoneGlueManageAccessRolePolicy。

AWS 管理策略：AmazonDataZoneRedshiftManageAccessRolePolicy

该政策赋予亚马逊 DataZone 向目录发布亚马逊 Redshift 数据的权限。它还给了亚马逊 DataZone 授予对目录中亚马逊 Redshift 或 Amazon Redshift Serverless Serverless 已发布资产的访问权限或撤消访问权限

要查看此策略的权限，请参阅《AWS 托管式策略参考》中的 AmazonDataZoneRedshiftManageAccessRolePolicy。

AWS 管理策略：AmazonDataZoneDomainExecutionRolePolicy

这是亚马逊的默认政策 DataZone `DomainExecutionRole` 服务角色。此角色由亚马逊使用 DataZone 对亚马逊中的数据进行分类、发现、管理、共享和分析 DataZone 域。此角色提供对所有亚马逊的访问权限 DataZone 使用数据门户所需的 API，以及支持在亚马逊中使用关联账户的 RAM 权限 DataZone 域。

你可以附上 AmazonDataZoneDomainExecutionRolePolicy 对你的政策 `AmazonDataZoneDomainExecutionRole`。

要查看此策略的权限，请参阅《AWS 托管式策略参考》中的 AmazonDataZoneDomainExecutionRolePolicy。

AWS 管理策略：AmazonDataZoneSageMakerProvisioningRolePolicy

这个 AmazonDataZoneSageMakerProvisioningRolePolicy 政策补助亚马逊 DataZone 与亚马逊互操作所需的权限 SageMaker。

要查看此策略的权限，请参阅《AWS 托管式策略参考》中的 AmazonDataZoneSageMakerProvisioningRolePolicy。

AWS 管理策略：AmazonDataZoneSageMakerManageAccessRolePolicy

该政策赋予亚马逊 DataZone 发布亚马逊的权限 SageMaker 将资产添加到目录中。它还给了亚马逊 DataZone 授予访问权限或撤销对亚马逊的访问权限的权限 SageMaker 目录中已发布的资产。

此策略包括以下权限：

- cloudtrail — 检索有关的信息 CloudTrail 小径。
- cloudwatch — 检索当前 CloudWatch 警报。
- 日志 — 检索指标过滤器 CloudWatch 日志。
- sns – 检索 SNS 主题的订阅列表。
- config — 检索有关配置记录器、资源和的信息 AWS 配置规则。还允许服务相关角色创建和删除 AWS 配置规则，并根据规则运行评估。
- iam – 获取和生成账户凭证报告。
- organizations – 检索组织的账户和组织单位（OU）信息。
- securityhub – 检索有关如何配置 Security Hub 服务、标准和控件的信息。
- tag – 检索有关资源标签的信息。

要查看此策略的权限，请参阅《AWS 托管式策略参考》中的 AmazonDataZoneSageMakerManageAccessRolePolicy。

AWS 管理策略：AmazonDataZoneSageMakerEnvironmentRolePermissionsBoundary

> **Note**
>
> 此策略是权限边界。权限边界设置基于身份的策略可以授予 IAM 实体的最大权限。您不应使用和附上亚马逊 DataZone 权限边界策略由您自己决定。亚马逊 DataZone 权限边界政策只能附加到亚马逊 DataZone 托管角色。有关权限边界的更多信息，请参阅《IAM 用户指南》中的 IAM 实体的权限边界。

当你创建亚马逊时 SageMaker 通过亚马逊获得环境 DataZone 数据门户，亚马逊 DataZone 将此权限边界应用于环境创建期间生成的 IAM 角色。权限边界限制了亚马逊角色的范围 DataZone 创建以及您添加的任何角色。

亚马逊 DataZone 使用 AmazonDataZoneSageMakerEnvironmentRolePermissionsBoundary 托管策略限制其附加到的预配置 IAM 委托人。委托人可能采用亚马逊的用户角色的形式 DataZone 可以代表交互式企业用户或分析服务假设 (AWS SageMaker，例如），然后执行操作来处理数据，例如从 Amazon S3 或 Amazon Redshift 读取和写入数据，或者运行 AWS 胶水爬虫。

这个 AmazonDataZoneSageMakerEnvironmentRolePermissionsBoundary 政策授予亚马逊读写权限 DataZone 到亚马逊等服务 SageMaker， AWS Glue，亚马逊 S3， AWS Lake Formation、亚马逊 Redshift 和亚马逊雅典娜。该政策还为使用这些服务所需的某些基础设施资源提供读写权限，例如网络接口、Amazon ECR 存储库和 AWS KMS 密钥。它还允许访问亚马逊 SageMaker 像亚马逊这样的应用程序 SageMaker 画布。

亚马逊 DataZone 适用 AmazonDataZoneSageMakerEnvironmentRolePermissionsBoundary 托管策略作为所有亚马逊的权限边界 DataZone 环境角色（所有者和贡献者）。该权限边界限制这些角色只能访问环境所需的资源和操作。

要查看此策略的权限，请参阅《AWS 托管式策略参考》中的 AmazonDataZoneSageMakerEnvironmentRolePermissionsBoundary。

亚马逊 DataZone 更新到 AWS 托管策略

查看有关更新的详细信息 AWS 亚马逊的管理政策 DataZone 自从该服务开始跟踪这些更改以来。如需有关此页面变更的自动提醒，请在亚马逊上订阅 RSS 提要 DataZone 文件历史 页面。

更改	描述	日期
AmazonDataZoneRedshiftGlueProvisioningPolicy 政策更新	此更新添加了 cloudformation:UntagResource 许可。亚马逊 DataZone 使用此权限从中移除标签 AWS CloudFormation 它在环境配置期间管理的堆栈。	2026 年 9 月 10 日
AmazonDataZoneSageMakerEnvironmentRolePermissionsBoundary 政策更新	的政策更新 AmazonDataZoneSageMakerEnvironmentRolePermissionsBoundary。此更新为添加了 Deny 语句 sagemaker	2026 年 7 月 15 日

更改	描述	日期
	:DeleteNotebookInstanceLifecycleConfig 行动。	
AmazonDataZoneSageMakerEnvironmentRolePermissionsBoundary -政策更新	的政策更新 AmazonDataZoneSageMakerEnvironmentRolePermissionsBoundary。为添加了 Deny 语句 sagemaker:CreateNotebookInstanceLifecycleConfig 行动。	2026 年 6 月 26 日
AmazonDataZoneSageMakerEnvironmentRolePermissionsBoundary -政策更新	的政策更新 AmazonDataZoneSageMakerEnvironmentRolePermissionsBoundary。为添加了 Deny 语句 sagemaker:UpdateNotebookInstanceLifecycleConfig 限制此高权限操作的操作。	2026年3月11日
AmazonDataZoneDomainExecutionRolePolicy -政策更新	的政策更新 AmazonDataZoneDomainExecutionRolePolicy -添加权限 QueryGraph 支持基于图形的实体搜索功能的操作。	2026 年 2 月 25 日

更改	描述	日期
AmazonDataZoneGlue ManageAccessRolePolicy -政策更新	的政策更新 AmazonDataZoneGlueManageAccessRolePolicy -向... 添加权限 `GetConnection` 支持对基于连接的数据源进行数据沿袭捕获的操作 AWS 胶水。	2025 年 7 月 30 日
AmazonDataZoneFullAccess -政策更新	的政策更新 AmazonDataZoneFullAccess -概括范围 SecretsManager `create` 和 `tag` 格式为的新域的权限 `dzd-` 而不是 `dzd_...`。	2025 年 7 月 23 日
AmazonDataZoneFullAccess -政策更新	的政策更新 AmazonDataZoneFullAccess -允许主机连接或更新 AWS 中的托管权限 AWS RAM 资源共享。	2025 年 5 月 22 日
AmazonDataZoneGlue ManageAccessRolePolicy -政策更新	的政策更新 AmazonDataZoneGlueManageAccessRolePolicy -亚马逊 DataZone 项目用户角色用作联合表的数据传输角色。此更新为 `iam:PassRole` 语句增加了 `datazone_usr_role*` ，使项目用户角色能够用于此目的。	2025 年 5 月 21 日

更改	描述	日期
AmazonDataZoneSageMakerProvisioningRolePolicy - 政策更新	的政策更新 AmazonDataZoneSageMakerProvisioningRolePolicy -添加对的支持 `glue:GetConnection` 行动。	2025 年 1 月 2 日
AmazonDataZoneSageMakerEnvironmentRolePermissionsBoundary -政策更新	的政策更新 AmazonDataZoneSageMakerEnvironmentRolePermissionsBoundary -此更改增加了 `sagemaker:AddTags` 到权限边界启用亚马逊 DataZone 成功致电 `CreateUserProfile` 带有必要的标签。	2024 年 12 月 3 日
AmazonDataZoneSageMakerAccess，以及 AmazonDataZoneGlueManageAccessRolePolicy -政策更新	的政策更新 AmazonDataZoneFullAccess，AmazonDataZoneSageMakerAccess，以及 AmazonDataZoneGlueManageAccessRolePolicy -启用对亚马逊的支持 SageMaker 统一工作室体验。	2024 年 12 月 3 日
AmazonDataZoneDomainExecutionRolePolicy 和 AmazonDataZoneFullUserAccess -政策更新	的政策更新 AmazonDataZoneDomainExecutionRolePolicy 和 AmazonDataZoneFullUserAccess -启用对订阅请求的元数据执行规则的支持。	2024 年 11 月 19 日

更改	描述	日期
AmazonDataZoneRedshiftGlueProvisioningPolicy -政策更新	的政策更新 AmazonDataZoneRedshiftGlueProvisioningPolicy -到 “添加” `iam:DeletePolicyVersion` 允许用户删除使用创建的策略的策略版本 `datazone*` 。这有助于解除对需要更新环境用户角色策略的用户的屏蔽。	2024 年 10 月 22 日
AmazonDataZoneDomainExecutionRolePolicy 和 AmazonDataZoneFullUserAccess -政策更新	的政策更新 AmazonDataZoneDomainExecutionRolePolicy 和 AmazonDataZoneFullUserAccess-启用对用于创建和管理亚马逊的新 API 的支持 DataZone 域单元和数据产品。	2024 年 7 月 31 日
AmazonDataZoneGlueManageAccessRolePolicy -政策更新	政策更新至 AmazonDataZoneGlueManageAccessRolePolicy -亚马逊 DataZone 正在添加用于细粒度访问控制功能的 IAM 权限，以缩小在 Lake Formation 中授予的权限范围。	2024 年 7 月 2 日
AmazonDataZoneExecutionRolePolicy 和 AmazonDataZoneFullUserAccess -政策更新	政策更新至 AmazonDataZoneExecutionRolePolicy 和 AmazonDataZoneFullUserAccess 启用对数据沿袭和细粒度访问控制 API 的支持。	2024 年 6 月 27 日

更改	描述	日期
AmazonDataZoneGlueManageAccessRolePolicy -政策更新	政策更新至 AmazonDataZoneGlueManageAccessRolePolicy 这增加了亚马逊自助订阅功能所需的 IAM 权限 DataZone 以缩小湖泊形成中授予的权限范围。使用自行订阅功能时，只能向标记的资源授予 Lake Formation 权限。	2024 年 6 月 14 日
AmazonDataZoneDomainExecutionRolePolicy -政策更新	政策更新至 AmazonDataZoneDomainExecutionRolePolicy 这为亚马逊增加了新的 API DataZone 这使用户能够为他们的亚马逊配置操作 DataZone 环境。	2024 年 6 月 14 日
AmazonDataZoneFullAccess -政策更新	政策更新至 AmazonDataZoneFullAccess 这使亚马逊成为可能 DataZone 管理控制台使用域和项目标签代表用户创建密钥。还包括 `ram:ListResourceSharePermissions` 操作以允许从域所有者账户进行管理，以便查看关联账户的账户关联状态。	2024 年 6 月 14 日

更改	描述	日期
AmazonDataZoneSageMakerEnvironmentRolePermissionsBoundary -新的权限边界	新的权限边界称为 AmazonDataZoneSageMakerEnvironmentRolePermissionsBoundary 。当你创建亚马逊时 SageMaker 通过亚马逊获得环境 DataZone 数据门户，亚马逊 DataZone 将此权限边界应用于环境创建期间生成的 IAM 角色。权限边界限制了亚马逊角色的范围 DataZone 创建以及您添加的任何角色。	2024 年 4 月 30 日
AmazonDataZoneSageMakerAccess -新政策	新政策被称为 AmazonDataZoneSageMakerAccess 给亚马逊 DataZone 发布亚马逊的权限 SageMaker 将资产添加到目录中。它还给了亚马逊 DataZone 授予访问权限或撤销对亚马逊的访问权限的权限 SageMaker 目录中已发布的资产。	2024 年 4 月 30 日
AmazonDataZoneFullAccess -政策更新	的更新 AmazonDataZoneFullAccess 增加访问权限的策略 `DescribeSecurityGroups` 采取措施提高账户管理员在控制台中配置蓝图的易用性，以及 `GetPolicy` 帮助检索有关指定托管策略的信息的操作。	2024 年 4 月 30 日

更改	描述	日期
AmazonDataZoneSageMakerProvisioningRolePolicy - 新政策	新政策被称为 AmazonDataZoneSageMakerProvisioningRolePolicy 补助亚马逊 DataZone 与亚马逊互操作所需的权限 SageMaker。	2024 年 4 月 30 日
AmazonDataZoneS3Manage-<region>-<domainId>-新角色	新角色叫做 AmazonDataZoneS3Manage-<region>-<domainId> 这是在亚马逊时使用的 DataZone 呼叫 AWS Lake Formation 将注册亚马逊简单存储服务（亚马逊 S3）地点。 AWS Lake Formation 在访问该位置的数据时会扮演此角色。	2024 年 4 月 1 日
AmazonDataZoneGlueManageAccessRolePolicy -政策更新	更新了 AmazonDataZoneGlueManageAccessRolePolicy 启用对允许亚马逊的权限的支持 DataZone 启用数据的发布和访问权限。	2024 年 4 月 1 日
AmazonDataZoneDomainExecutionRolePolicy 和 AmazonDataZoneFullUserAccess -政策更新	更新了 AmazonDataZoneDomainExecutionRolePolicy 和 AmazonDataZoneFullUserAccess 启用对的支持 `CancelMetadataGenerationRun` API。	2024 年 3 月 29 日

更改	描述	日期
AmazonDataZoneFullAccess - 政策更新	更新了 `AmazonDataZoneFullAccess` 使用户能够在亚马逊中选择自己的密钥、集群、VPC 和子网 DataZone 管理控制台，而不是在文本框中键入它们。	2024 年 3 月 13 日
AmazonDataZoneDomainExecutionRolePolicy -政策更新	更新了 AmazonDataZoneDomainExecutionRolePolicy 启用对的支持 `ListEnvironmentBlueprintConfigurationSummaries` 通过识别在哪个账户和区域启用了哪些蓝图来创建环境配置文件所需的 API。	2024 年 2 月 1 日
AmazonDataZoneGlueManageAccessRolePolicy -政策更新	更新了 AmazonDataZoneGlueManageAccessRolePolicy 启用对的支持 AWS 湖泊形成混合模式。	2023 年 12 月 14 日
AmazonDataZoneFullUserAccess 和 AmazonDataZoneDomainExecutionRolePolicy -政策更新	更新了 AmazonDataZoneFullUserAccess 还有 AmazonDataZoneDomainExecutionRolePolicy 支持发电的政策 AI-powered 亚马逊的数据描述功能 DataZone。	2023 年 11 月 28 日

更改	描述	日期
AmazonDataZoneEnvironmentRolePermissionsBoundary -政策更新	亚马逊 DataZone 更新了 AmazonDataZoneEnvironmentRolePermissionsBoundary 包含额外内容的托管策略 `athena:GetQueryResultsStream` 权限范围缩小为 `ResourceTag` 条件。	2023 年 11 月 17 日
AmazonDataZoneRedshiftManageAccessRolePolicy -政策更新	亚马逊 DataZone 更新了 AmazonDataZoneRedshiftManageAccessRolePolicy 通过移除对组织 ID 的检查 `redshift:AssociateDataShareConsumer` 行动。这使您能够在 AWS 组织之间共享资源。	2023 年 11 月 16 日
AmazonDataZoneFullUserAccess -政策更新	亚马逊 DataZone 更新了 AmazonDataZoneFullUserAccess 授予亚马逊完全访问权限的政策 DataZone，但它不允许管理域、用户或关联帐户。	2023 年 10 月 2 日
AmazonDataZonePortalFullAccessPolicy -政策已过时	亚马逊 DataZone 已弃用 AmazonDataZonePortalFullAccessPolicy。	2023 年 9 月 29 日
AmazonDataZonePreviewConsoleFullAccess -政策已过时	亚马逊 DataZone 已弃用 AmazonDataZonePreviewConsoleFullAccess。	2023 年 9 月 29 日

更改	描述	日期
AmazonDataZoneDomainExecutionRolePolicy -新政策	亚马逊 DataZone 添加了一个名为的新政策 AmazonDataZoneDomainExecutionRolePolicy。 这是亚马逊的默认政策 DataZone `AmazonDataZoneDomainExecutionRole` 服务角色。此角色由亚马逊使用 DataZone 对亚马逊中的数据进行分类、发现、管理、共享和分析 DataZone 域。 您可以将 `AmazonDataZoneDomainExecutionRolePolicy` 策略附加到 `AmazonDataZoneDomainExecutionRole` 。	2023 年 9 月 25 日
AmazonDataZoneCrossAccountAdmin -新政策	亚马逊 DataZone 添加了一个名为的新政策 AmazonDataZoneCrossAccountAdmin 这使用户能够与亚马逊合作 DataZone 及其关联账户。	2023 年 9 月 19 日
AmazonDataZoneFullUserAccess -新政策	亚马逊 DataZone 添加了一个名为的新政策 AmazonDataZoneFullUserAccess 这授予了对亚马逊的完全访问权限 DataZone，但它不允许管理域、用户或关联帐户。	2023 年 9 月 12 日

更改	描述	日期
AmazonDataZoneRedshiftManageAccessRolePolicy - 新政策	亚马逊 DataZone 添加了一个名为的新政策 AmazonDataZoneRedshiftManageAccessRolePolicy 它授予允许亚马逊的权限 DataZone 启用数据的发布和访问权限。	2023 年 9 月 12 日
AmazonDataZoneGlueManageAccessRolePolicy -新政策	亚马逊 DataZone 添加了一个名为的新政策 AmazonDataZoneGlueManageAccessRolePolicy 这给了亚马逊 DataZone 发布权限 AWS 将数据粘贴到目录中。它还给了亚马逊 DataZone 授予访问权限或撤消访问权限的权限 AWS 将已发布的资源粘贴到目录中。	2023 年 9 月 12 日
AmazonDataZoneRedshiftGlueProvisioningPolicy -新政策	亚马逊 DataZone 添加了一个名为的新政策 AmazonDataZoneRedshiftGlueProvisioningPolicy 这给了亚马逊 DataZone 与支持的数据源进行互操作所需的权限。	2023 年 9 月 12 日
AmazonDataZoneEnvironmentRolePermissionsBoundary -新政策	亚马逊 DataZone 添加了一个名为的新政策 AmazonDataZoneEnvironmentRolePermissionsBoundary 这限制了它所连接的预置IAM委托人。	2023 年 9 月 12 日

更改	描述	日期
AmazonDataZoneFullAccess - 新政策	亚马逊 DataZone 添加了一个名为的新政策 AmazonDataZoneFullAccess 这提供了对亚马逊的完全访问权限 DataZone 通过 AWS 管理控制台。	2023 年 9 月 12 日
托管式策略更新	的更新 AmazonDataZonePreviewConsoleFullAccess 包含额外内容的托管策略 `iam:GetPolicy` 权限。	2023 年 6 月 13 日
亚马逊 DataZone 已开始跟踪更改	亚马逊 DataZone 开始跟踪其更改 AWS 托管策略。	2023 年 3 月 20 日

亚马逊的 IAM 角色 DataZone

主题

- AmazonDataZoneProvisioningRole-<域AccountId>
- AmazonDataZoneDomainExecutionRole
- AmazonDataZoneGlueAccess-<领域>-<域名>
- AmazonDataZoneRedshiftAccess-<领域>-<域名>
- AmazonDataZoneS3Manage-<领域>-<域名>
- AmazonDataZoneSageMakerManageAccessRole-<领域>-<域名>
- AmazonDataZoneSageMakerProvisioningRolePolicyRole-<域AccountId>

AmazonDataZoneProvisioningRole-<域AccountId>

`AmazonDataZoneProvisioningRole-<domainAccountId>` 已附加 `AmazonDataZoneRedshiftGlueProvisioningPolicy`。此角色授予亚马逊 DataZone 与进行互操作所需的权限 AWS Glue 和亚马逊 Redshift。

默认 AmazonDataZoneProvisioningRole-<domainAccountId> 已附加以下信任策略：

JSON

```
{
  "Version":"2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "Service": "datazone.amazonaws.com"
      },
      "Action": "sts:AssumeRole",
      "Condition": {
        "StringEquals": {
            "aws:SourceAccount": "{{domain_account}}"
        }
      }
    }
  ]
}
```

AmazonDataZoneDomainExecutionRole

这个 AmazonDataZoneDomainExecutionRole 有 AWS 托管策略 AmazonDataZoneDomainExecutionRolePolicy 附上。亚马逊 DataZone 代表你创建这个角色。对于数据门户中的某些操作，亚马逊 DataZone 在创建该角色的账户中代入此角色，并检查该角色是否有权执行操作。

这个 AmazonDataZoneDomainExecutionRole 中的角色是必需的 AWS 账户 托管您的亚马逊 DataZone 域。此角色是在您创建亚马逊时自动为您创建的 DataZone 域。

默认 AmazonDataZoneDomainExecutionRole 角色具有以下信任策略。

JSON

```
{
```

```
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Principal": {
                "Service": "datazone.amazonaws.com"
            },
            "Action": [
                "sts:AssumeRole",
                "sts:TagSession"
            ],
            "Condition": {
                "StringEquals": {
                    "aws:SourceAccount": "{{source_account_id}}"
                },
                "ForAllValues:StringLike": {
                    "aws:TagKeys": [
                        "datazone*"
                    ]
                }
            }
        }
    ]
}
```

AmazonDataZoneGlueAccess-<领域>-<域名>

AmazonDataZoneGlueAccess-<region>-<domainId> 角色已附加 AmazonDataZoneGlueManageAccessRolePolicy。此角色授予亚马逊 DataZone 发布权限 AWS 将数据粘贴到目录中。它还给了亚马逊 DataZone 授予访问权限或撤消访问权限的权限 AWS 将已发布的资源粘贴到目录中。

默认 AmazonDataZoneGlueAccess-<region>-<domainId> 角色已附加以下信任策略：

JSON

```
{
  "Version":"2012-10-17",
  "Statement": [
```

```
    {
      "Effect": "Allow",
      "Principal": {
        "Service": "datazone.amazonaws.com"
      },
      "Action": "sts:AssumeRole",
      "Condition": {
            "StringEquals": {
            "aws:SourceAccount": "111122223333"
            },
            "ArnEquals": {
            "aws:SourceArn": "arn:aws:datazone:us-east-1:111122223333:domain/
dzd-12345"
            }
        }
    }
  ]
}
```

AmazonDataZoneRedshiftAccess-<领域>-<域名>

AmazonDataZoneRedshiftAccess-<region>-<domainId> 角色已附加 AmazonDataZoneRedshiftManageAccessRolePolicy。此角色授予亚马逊 DataZone 向目录发布亚马逊 Redshift 数据的权限。它还给了亚马逊 DataZone 授予对目录中亚马逊 Redshift 或 Amazon Redshift Serverless Serverless 已发布资产的访问权限或撤消访问权限

默认 AmazonDataZoneRedshiftAccess-<region>-<domainId> 角色已附加以下内联权限策略：

JSON

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Sid": "RedshiftSecretStatement",
         "Effect":"Allow",
         "Action":"secretsmanager:GetSecretValue",
```

```
         "Resource":"*",
         "Condition":{
            "StringEquals":{
               "secretsmanager:ResourceTag/AmazonDataZoneDomain":"{{domainId}}"
            }
         }
      }
   ]
}
```

默认 AmazonDataZoneRedshiftManageAccessRole<timestamp> 已附加以下信任策略：

JSON

```
{
  "Version":"2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "Service": "datazone.amazonaws.com"
      },
      "Action": "sts:AssumeRole",
      "Condition": {
            "StringEquals": {
            "aws:SourceAccount": "111122223333"
            },
            "ArnEquals": {
            "aws:SourceArn": "arn:aws:datazone:us-east-1:111122223333:domain/
dzd-12345"
            }
        }
    }
  ]
}
```

AmazonDataZoneS3Manage-<领域>-<域名>

这个 AmazonDataZoneS3Manage-<region>-<domainId>在亚马逊时使用 DataZone 呼叫 AWS Lake Formation 将注册亚马逊简单存储服务（亚马逊 S3）地点。 AWS Lake Formation 在访问该位置的数据时会扮演此角色。有关更多信息，请参阅用于注册位置的角色的要求。

此角色已附加以下内联权限策略。

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Sid": "LakeFormationDataAccessPermissionsForS3",
            "Effect": "Allow",
            "Action": [
                "s3:PutObject",
                "s3:GetObject",
                "s3:DeleteObject"
            ],
            "Resource": "*",
            "Condition": {
                "StringEquals": {
                    "aws:ResourceAccount": "{{accountId}}"
                }
            }
        },
        {
            "Sid": "LakeFormationDataAccessPermissionsForS3ListBucket",
            "Effect": "Allow",
            "Action": [
                "s3:ListBucket"
            ],
            "Resource": "*",
            "Condition": {
                "StringEquals": {
                    "aws:ResourceAccount": "{{accountId}}"
                }
            }
        },
```

```
        {
            "Sid": "LakeFormationDataAccessPermissionsForS3ListAllMyBuckets",
            "Effect": "Allow",
            "Action": [
                "s3:ListAllMyBuckets"
            ],
            "Resource": "arn:aws:s3:::*",
            "Condition": {
                "StringEquals": {
                    "aws:ResourceAccount": "{{accountId}}"
                }
            }
        },
        {
            "Sid": "LakeFormationExplicitDenyPermissionsForS3",
            "Effect": "Deny",
            "Action": [
                "s3:PutObject",
                "s3:GetObject",
                "s3:DeleteObject"
            ],
            "Resource": [
                "arn:aws:s3:::[[BucketNames]]/*"
            ],
            "Condition": {
                "StringEquals": {
                    "aws:ResourceAccount": "{{accountId}}"
                }
            }
        },
        {
            "Sid": "LakeFormationExplicitDenyPermissionsForS3ListBucket",
            "Effect": "Deny",
            "Action": [
                "s3:ListBucket"
            ],
            "Resource": [
                "arn:aws:s3:::[[BucketNames]]"
            ],
            "Condition": {
                "StringEquals": {
                    "aws:ResourceAccount": "{{accountId}}"
                }
            }
```

```
        }
    ]
}
```

这个 AmazonDataZoneS3Manage-<region>-<domainId>附有以下信任政策：

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Sid": "TrustLakeFormationForDataLocationRegistration",
            "Effect": "Allow",
            "Principal": {
                "Service": "lakeformation.amazonaws.com"
            },
            "Action": "sts:AssumeRole",
            "Condition": {
                "StringEquals": {
                    "aws:SourceAccount": "{{source_account_id}}"
                }
            }
        }
    ]
}
```

AmazonDataZoneSageMakerManageAccessRole-<领域>-<域名>

AmazonDataZoneSageMakerManageAccessRole 角色已附加 AmazonDataZoneSageMakerAccess、AmazonDataZoneRedshiftManageAccessRolePolicy 和 AmazonDataZoneGlueManageAccessRolePolicy。此角色授予亚马逊 DataZone 发布和管理数据湖、数据仓库和亚马逊 Sagemaker 资产订阅的权限。

AmazonDataZoneSageMakerManageAccessRole 角色已附加以下内联策略：

JSON

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Sid": "RedshiftSecretStatement",
         "Effect":"Allow",
         "Action":"secretsmanager:GetSecretValue",
         "Resource":"*",
         "Condition":{
            "StringEquals":{
               "secretsmanager:ResourceTag/AmazonDataZoneDomain":"{{domainId}}"
            }
         }
      }
   ]
}
```

AmazonDataZoneSageMakerManageAccessRole 角色已附加以下信任策略：

JSON

```
{
  "Version":"2012-10-17",
  "Statement": [
    {
      "Sid": "DatazoneTrustPolicyStatement",
      "Effect": "Allow",
      "Principal": {
        "Service":  ["datazone.amazonaws.com",
                    "sagemaker.amazonaws.com"]
      },
      "Action": "sts:AssumeRole",
      "Condition": {
            "StringEquals": {
            "aws:SourceAccount": "111122223333"
            },
```

```
            "ArnEquals": {
            "aws:SourceArn": "arn:aws:datazone:us-east-1:111122223333:domain/
dzd-12345"
            }
        }
    }
]
}
```

AmazonDataZoneSageMakerProvisioningRolePolicyRole-<域AccountId>

AmazonDataZoneSageMakerProvisioningRolePolicyRole 角色已附加 AmazonDataZoneSageMakerProvisioningRolePolicy 和 AmazonDataZoneRedshiftGlueProvisioningPolicy。此角色授予亚马逊 DataZone 与进行互操作所需的权限 AWS Glue、亚马逊 Redshift 和亚马逊 Sagemaker。

AmazonDataZoneSageMakerProvisioningRolePolicyRole 角色已附加以下内联策略：

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Sid": "SageMakerStudioTagOnCreate",
            "Effect": "Allow",
            "Action": [
                "sagemaker:AddTags"
            ],
            "Resource": "arn:aws:sagemaker:*:111122223333:*/*",
            "Condition": {
                "Null": {
                    "sagemaker:TaggingAction": "false"
                }
            }
        }
    ]
}
```

AmazonDataZoneSageMakerProvisioningRolePolicyRole 角色已附加以下信任策略：

JSON

```
{
  "Version":"2012-10-17",
  "Statement": [
    {
      "Sid": "DataZoneTrustPolicyStatement",
      "Effect": "Allow",
      "Principal": {
        "Service": "datazone.amazonaws.com"
      },
      "Action": "sts:AssumeRole",
      "Condition": {
        "StringEquals": {
            "aws:SourceAccount": "{{domain_account}}"
        }
      }
    }
  ]
}
```

临时证书

一些 AWS 当你使用临时证书登录时，服务不起作用。欲了解更多信息，包括 AWS 服务使用临时证书，请参阅 AWS 与 IAM 配合使用的服务 在 IAM 用户指南。

如果您登录，则使用的是临时证书 AWS 管理控制台 使用除用户名和密码之外的任何方法。例如，当你访问 AWS 使用贵公司的单点登录 (SSO) 链接，该过程会自动创建临时证书。当您以用户身份登录控制台，然后切换角色时，您还会自动创建临时凭证。有关切换角色的更多信息，请参阅《IAM 用户指南》中的切换到角色（控制台）。

您可以使用手动创建临时证书 AWS CLI 要么 AWS API。然后，您可以使用这些临时证书来访问 AWS。 AWS 建议您动态生成临时证书，而不是使用长期访问密钥。有关更多信息，请参阅 IAM 中的临时安全凭证。

亚马逊 DataZone 门户临时证书

当你登录亚马逊时 DataZone 门户，您将收到的临时证书 AmazonDataZoneDomainExecutionRole。当你使用时 AmazonDataZoneDomainExecutionRole，这些凭证在使用时会自动刷新。如果一段时间未使用，它们会自动过期。

主体权限

当您使用 IAM 用户或角色在中执行操作时 AWS，你被视为校长。策略向主体授予权限。使用某些服务时，您可能会执行一个操作，此操作然后在不同服务中触发另一个操作。在这种情况下，您必须具有执行这两个操作的权限。要查看某项操作是否需要在策略中添加其他相关操作，请参阅 的操作、资源和条件键 AWS 文档要点 在 服务授权参考。

亚马逊合规性验证 DataZone

我们的新 AWS 注册体验不是为受监管的工作负载设计的。如果你正在使用我们的新品 AWS 注册体验，但你想使用 AWS 对于受监管的工作负载，您可以 报名参加 AWS （高级） 要么 激活高级功能 为了你的 AWS 环境。

要了解是否 AWS 服务 属于特定合规计划的范围，请参阅 AWS 服务 按合规计划划分的范围 并选择您感兴趣的合规计划。有关一般信息，请参见 AWS 合规计划。

您可以使用以下方式下载第三方审计报告 AWS Artifact。有关更多信息，请参阅 在中下载报告 AWS Artifact。

您在使用时的合规责任 AWS 服务 由您的数据的敏感度、贵公司的合规目标以及适用的法律法规决定。有关您在使用时的合规责任的更多信息 AWS 服务，参见 AWS 安全文档。

亚马逊安全最佳实践 DataZone

亚马逊 DataZone 提供了许多安全功能，供您在制定和实施自己的安全策略时考虑。以下最佳实践是一般指导原则，并不代表完整安全解决方案。由于这些最佳实践可能不适合您的环境或不满足您的环境要求，因此将其视为有用的考虑因素而不是惯例。

实施最低权限访问

授予权限时，由您决定谁获得哪个亚马逊的权限 DataZone 资源。您可以对这些资源启用希望允许的特定操作。因此，您应仅授予执行任务所需的权限。实施最低权限访问对于减小安全风险以及可能由错误或恶意意图造成的影响至关重要。

有关更多信息，请参阅AWS 亚马逊的管理政策 DataZone和服务控制策略（SCP）。

使用 IAM 角色

生产者和客户端应用程序必须具有有效的凭证才能访问亚马逊 DataZone 资源。你不应该存储 AWS 证书直接存储在客户端应用程序或 Amazon S3 存储桶中。这些是不会自动轮换的长期凭证，如果它们受到损害，可能会对业务产生重大影响。

相反，您应该使用 IAM 角色来管理您的生产者和客户端应用程序访问亚马逊的临时证书 DataZone 资源。在使用角色时，您不必使用长期凭证（如用户名和密码或访问密钥）来访问其他资源。

有关更多信息，请参阅 IAM 用户指南中的以下主题：

- IAM 角色
- 针对角色的常见情形：用户、应用程序和服务

实施 Server-Side 依赖资源中的加密

静态数据和传输中的数据可以在亚马逊中加密 DataZone。

使用 CloudTrail 监控 API 调用

亚马逊 DataZone 已集成到 AWS CloudTrail，一项提供用户、角色或某人所采取操作记录的服务 AWS 亚马逊的服务 DataZone。

使用收集的信息 CloudTrail，您可以确定向亚马逊提出的请求 DataZone、提出请求的 IP 地址、谁提出请求、何时提出请求以及其他详细信息。

在亚马逊中使用 RAM DataZone

关联你的 AWS 在亚马逊开设的账户 DataZone 域名使域用户能够发布和使用来自这些域的数据 AWS 账户。亚马逊 DataZone 使用 AWS 资源访问管理器 (RAM)，用于管理跨账户访问权限。有关更多信息，请参阅Amazon 中的关联账户 DataZone和 AWS RAM 中的安全性。

亚马逊的弹性 DataZone

这个 AWS 全球基础设施是围绕它建立的 AWS 区域 和可用区。 AWS 区域 提供多个物理上分隔和隔离的可用区，这些可用区通过低延迟、高吞吐量和高度冗余的网络相连。利用可用区，您可以设计和操

作在可用区之间无中断地自动实现失效转移的应用程序和数据库。与传统的单个或多个数据中心基础设施相比，可用区具有更高的可用性、容错能力和可扩展性。

有关以下内容的更多信息 AWS 区域 和可用区，请参阅 AWS 全球基础设施。

除了 AWS 全球基础设施，亚马逊 DataZone 提供多种功能，可帮助支持您的数据弹性和备份需求。

主题

- 数据来源韧性
- 资产韧性
- 资产类型和元数据表单韧性
- 术语表韧性
- 全局搜索韧性
- 订阅韧性
- 环境韧性
- 环境蓝图韧性
- 项目韧性
- RAM 韧性
- 用户配置文件管理韧性
- 域韧性

数据来源韧性

在亚马逊期间 DataZone 可用性事件， `DataSource` 作业将定期重试，最长可达 24 小时。如果作业因配置错误而失败，则将发出一个 `DataSourceRunFailed` 事件。如果亚马逊 DataZone 域名使用 KMS 密钥进行配置，并且 AmazonDataZoneDomainExecutionRole 在作业运行期间失去对该密钥的访问权限，运行将在 `INACCESSIBLE` 州。在恢复 KMS 访问权限后，应手动更新作业以触发过渡回到可用状态。

资产韧性

在亚马逊 DataZone，资产是版本控制的。如果资产的某个版本需要回滚，则可以使用上一个稳定版本的内容来创建新版本。可以发布资产版本。除非发布新版本，否则无法编辑资产的已发布版本。可以订阅已发布的资产（又名列入清单的资产）。为防止对某项资源进行新订阅，可以将其取消发布。 Un-

publishing 资产对现有订阅没有影响。删除某个资产将删除该资产的所有取消发布版本。必须单独删除资产的已发布版本。只能在没有订阅的情况下删除资产的已发布版本。

资产类型和元数据表单韧性

在亚马逊 DataZone、资产类型和元数据表单类型是版本控制的。如果一个资产类型正在由某个资产使用，则无法删除该资产类型。如果一个元数据表单类型正在由资产类型或资产使用，则无法删除该元数据表单类型。如果您不想将特定的元数据表单类型用于策划，您可以将其禁用，这不会影响它已附加到的资产。

术语表韧性

在亚马逊 DataZone，如果正在使用词汇表和词汇表，则不能将其删除。如果您不想将特定的术语表和术语表术语用于策划，您可以将其禁用，这不会影响它已附加到的资产。

全局搜索韧性

在亚马逊 DataZone，可以通过全球搜索发现已发布的资产（又名清单）。可以通过取消发布某个资产来回滚该资产的发布。取消发布资产不会影响现有订阅。可以将已发布资产回滚到特定版本，方式是重新发布该特定版本。这将不会影响现有订阅。

订阅韧性

在亚马逊 DataZone，订阅补助金的履行将在失败之前尝试两次退休。如果失败，则必须手动将其删除以进行重试。如果亚马逊 DataZone 无法撤消订阅权限，删除订阅可能会失败。应该解决潜在的错误，或者 `retainPermissions` 标志可以用于 `DeleteSubscriptionGrant` 强制从亚马逊删除补助金的 API 操作 DataZone 无需撤消权限。

如果亚马逊 DataZone 域名使用 KMS 密钥进行配置，并且 `AmazonDataZoneDomainExecutionRole` 在此期间无法访问此密钥 `SubscriptionGrant` 工作流程，赠款已标记 `INACCESSIBLE`。在恢复 KMS 访问权限后，必须删除并重新创建 `INACCESSIBLE` 授权。

环境韧性

如果亚马逊 DataZone 域名使用 KMS 密钥进行配置，并且 `AmazonDataZoneDomainExecutionRole` 在环境工作流程中失去对该密钥的访问权限，环境将被标记 `INACCESSIBLE`。在恢复 KMS 访问权限后，必须删除并重新创建 `INACCESSIBLE` 环境。环境创建将在失败前尝试两次停用。如果失败，则必须手动将其删除以进行重试。如果环境工作流失败，则环境将进入失败状态。此时，只能删除并重新创建环境。

环境蓝图韧性

在亚马逊 DataZone，如果有任何底层环境配置文件，则无法删除环境蓝图。

项目韧性

在亚马逊 DataZone，如果有任何包含的环境，则无法删除项目。

RAM 韧性

有关 RAM 弹性信息，请参阅 https://docs.aws.amazon.com/ram/latest/userguide/security-disaster-recovery-resiliency.html。

用户配置文件管理韧性

有关用户配置文件韧性信息，请参阅 AWS Identity Center。

域韧性

在亚马逊 DataZone，如果域包含项目或数据源，则无法将其删除。

亚马逊的基础设施安全 DataZone

作为一项托管服务，亚马逊 DataZone 受以下保护 AWS 全球网络安全。有关信息 AWS 安全服务及其方法 AWS 保护基础架构，请参阅 AWS 云安全。设计你的 AWS 使用基础设施安全最佳实践的环境，请参阅 基础设施保护 在 安全支柱 AWS 架构良好的框架。

你用 AWS 发布了访问亚马逊的 API 调用 DataZone 通过网络。客户端必须支持以下内容：

- 传输层安全性协议（TLS）。我们要求使用 TLS 1.2，并建议使用 TLS 1.3。
- 具有完美向前保密 (PFS) 的密码套件，例如 DHE（临时保密） Diffie-Hellman) 或 ECDHE（椭圆曲线短暂的 Diffie-Hellman）。大多数现代系统（如 Java 7 及更高版本）都支持这些模式。

Cross-service 亚马逊的副手防范措施令人困惑 DataZone

混淆代理问题是一个安全性问题，即不具有某操作执行权限的实体可能会迫使具有更高权限的实体执行该操作。在 AWS，跨服务模仿可能会导致副手混乱的问题。 Cross-service 当一项服务（呼叫服务）调用另一项服务（被叫服务）时，可能会发生模拟。可以操纵调用服务以使用其权限对另一个客户的资

源进行操作，否则该服务不应有访问权限。为了防止这种情况， AWS 提供的工具可帮助您使用已获准访问您账户中资源的服务主体保护所有服务的数据。

我们建议使用 aws：SourceAccount 资源策略中的全局条件上下文密钥，用于限制亚马逊的权限 DataZone 为资源提供另一项服务。使用 aws：SourceAccount 如果你想允许该账户中的任何资源与跨服务使用相关联。

亚马逊的配置和漏洞分析 DataZone

AWS 处理基本的安全任务，例如客户机操作系统 (OS) 和数据库修补、防火墙配置和灾难恢复。这些流程已通过相应第三方审核和认证。有关更多信息，请参阅 AWS 责任共担模式。

要添加到允许列表的域

对于亚马逊 DataZone 访问亚马逊的数据门户 DataZone 服务，必须将以下域添加到数据门户尝试访问该服务的网络上的允许列表中。

- `*.api.aws`
- `*.on.aws`

监控 Amazon DataZone

监控是保持 Amazon DataZone 和您的其他 AWS 解决方案的可靠性、可用性和性能的重要方面。AWS 提供以下监控工具来监控 Amazon DataZone，在出现问题时进行报告，并在适当的时候采取自动措施：

- Amazon CloudWatch 可实时监控您的 AWS 资源以及您在 AWS 上运行的应用程序。您可以收集和跟踪指标，创建自定义的控制平面，以及设置警报以在指定的指标达到您指定的阈值时通知您或采取措施。例如，您可以使用 CloudWatch 跟踪 Amazon EC2 实例的 CPU 使用率或其他指标并且在需要时自动启动新实例。有关更多信息，请参阅《Amazon CloudWatch 用户指南》https://docs.aws.amazon.com/AmazonCloudWatch/latest/monitoring/。
- Amazon CloudWatch Logs 使您能够监控、存储和访问来自 Amazon EC2 实例、CloudTrail 和其他来源的日志文件。CloudWatch Logs 可以监控日志文件中的信息，并在达到特定阈值时通知您。您还可以在高持久性存储中检索您的日志数据。有关更多信息，请参阅 Amazon CloudWatch Logs 用户指南。
- 您可以使用 Amazon EventBridge 自动执行您的 AWS 服务并自动响应系统事件，例如应用程序可用性问题或资源更改。AWS 服务中的事件将近乎实时传输到 EventBridge。您可以编写简单的规则来指示您关注的事件，并指示要在事件匹配规则时执行的自动化操作。有关更多信息，请参阅 Amazon EventBridge 用户指南。
- AWS CloudTrail 捕获由您的 AWS 账户或代表该账户发出的 API 调用和相关事件，并将日志文件传输到您指定的 Amazon S3 桶。您可以标识哪些用户和账户调用了 AWS、发出调用的源 IP 地址以及调用的发生时间。有关更多信息，请参阅 《AWS CloudTrail 用户指南》。

监控 Amazon EventBridge 中的 Amazon DataZone 事件

您可以在 EventBridge 中监控 Amazon DataZone 事件，这将从您自己的应用程序、软件即服务（SaaS）应用程序和 AWS 服务传输实时数据流。EventBridge 将该数据路由到 AWS Lambda 和 Amazon Simple Notification Service 等目标。这些事件与 Amazon CloudWatch Events 中出现的事件相同，可提供近乎实时的系统事件流，这些事件描述 AWS 资源的更改。

有关更多信息，请参阅 通过 Amazon EventBridge 默认总线举办的活动。

使用 AWS CloudTrail 记录 Amazon DataZone API 调用

Amazon DataZone 与 AWS CloudTrail 集成，后者是在 Amazon DataZone 中提供用户、角色或 AWS 服务所执行操作的记录的服务。CloudTrail 将 Amazon DataZone 的所有 API 调用捕获为事件。捕获的

调用包括通过 Amazon DataZone 控制台的调用以及对 Amazon DataZone API 操作的代码调用。如果您创建跟踪，则可以使 CloudTrail 事件持续传送到 Amazon S3 存储桶（包括 Amazon DataZone 的事件）。如果您不配置跟踪，则仍可在 CloudTrail 控制台中的事件历史记录中查看最新事件。使用通过 CloudTrail 收集的信息，您可以确定向 Amazon DataZone 发出了什么请求、发出请求的 IP 地址、何人发出的请求、请求的发出时间以及其他详细信息。

要了解有关 CloudTrail 的更多信息，请参阅《AWS CloudTrail 用户指南》https://docs.aws.amazon.com/awscloudtrail/latest/userguide/cloudtrail-user-guide.html。

CloudTrail 中的 Amazon DataZone 信息

在您创建 AWS 账户时，将在该账户上启用 CloudTrail。当 Amazon DataZone 管理控制台中发生活动时，该活动将记录在 CloudTrail 事件中，并与其他 AWS 服务事件一同保存在事件历史记录中。您可以在 AWS 账户 中查看、搜索和下载最新事件。有关更多信息，请参阅使用 CloudTrail 事件历史记录查看事件。

要持续记录 AWS 账户中的事件（包括 Amazon DataZone 的事件），请创建跟踪。通过跟踪记录，CloudTrail 可将日志文件传送至 Amazon S3 存储桶。预设情况下，在控制台中创建跟踪记录时，此跟踪记录应用于所有AWS 区域。此跟踪记录在 AWS 分区中记录所有区域中的事件，并将日志文件传送至您指定的 Amazon S3 存储桶。此外，您可以配置其他 AWS 服务，进一步分析在 CloudTrail 日志中收集的事件数据并采取行动。有关更多信息，请参阅下列内容：

- 创建跟踪记录概述
- CloudTrail 支持的服务和集成
- Configuring Amazon SNS notifications for CloudTrail
- 从多个区域接收 CloudTrail 日志文件和从多个账户接收 CloudTrail 日志文件

所有 Amazon DataZone 操作均由 CloudTrail 记录。

对亚马逊进行故障排除 DataZone

如果您在与 Amazon 合作时遇到访问被拒绝问题或类似困难， DataZone请参阅本节的主题。

问题排查 AWS 亚马逊的 Lake Formation 权限 DataZone

此部分包含您在为亚马逊配置 Lake Formation 权限 DataZone时可能遇到的问题的问题排查说明。

数据门户中的错误消息	解决方案
无法代入数据访问角色。	当 Amazon DataZone 无法假设您在账户DefaultDataLakeBlueprint中启用时使用的AmazonDataZoneGlueDataAccessRole，则会显示此错误。要解决此问题，请访问您的数据资产所在账户中的 AWS IAM 控制台，并确保与 Amazon DataZone 服务委托人AmazonDataZoneGlueDataAccessRole有正确的信任关系。有关更多信息，请参阅 AmazonDataZoneGlueAccess-<领域>-<域名>。
数据访问角色没有必要的权限，无法读取您尝试订阅的资产的元数据。	当 Amazon DataZone 成功担任该AmazonDataZoneGlueDataAccessRole角色但该角色没有必要的权限时，就会显示此错误。要修复此问题，请转到您的数据资产所在账户中的 AWS IAM 控制台，并确保该角色已AmazonDataZoneGlueManageAccessRolePolicy附加该数据。有关更多信息，请参阅 AmazonDataZoneGlueAccess-<领域>-<域名>。
资产是一项资源链接。Amazon DataZone 不支持订阅资源链接。	当您尝试发布到亚马逊的资产是指向 AWS GI DataZone ue 表的资源链接时，就会显示此错误。
资产不由 La AWS ke Formation 管理。	此错误表示未对您要发布的资产强制执行 AWS Lake Formation 权限。此错误会在以下情况下出现。

数据门户中的错误消息	解决方案
	• 资产的 Amazon S3 位置未注册到 AWS Lake Formation。要解决此问题，请使用表格所在的账户登录您的 L AWS ake Formation 控制台，然后在 L AWS ake Formation 模式或混合模式下注册 Amazon S3 位置。有关更多信息，请参阅 Registering an Amazon S3 location（注册 Amazon S3 位置）。在几种情况下，需要进行进一步的修改。其中包括加密的 AmazonS3 存储桶或跨账户 S3 存储桶和 Glue 目录设置 AWS 。在这种情况下，可能需要修改 KMS and/or S3 设置。有关更多信息，请参阅 Registering an encrypted Amazon S3 location。 • Amazon S3 位置已在 AWS Lake Format IAMAllowedPrincipalion 模式下注册，但已添加到表的权限中。要解决此问题，您可以IAMAllowedPrincipal从表的权限中移除，也可以在混合模式下注册 S3 位置。有关更多信息，请参阅关于升级为 Lake Formation 权限模型。如果您的 S3 位置已加密或 S3 位置与您的 AWS Glue 表位于不同的账户中，请按照 Registering an encrypted Amazon S3 location 中的说明进行操作。

数据门户中的错误消息	解决方案
数据访问角色没有必要的 Lake Formation 权限，无法授予对该资产的访问权限。	此错误AmazonDataZoneGlueDataAccessRole表明，您用于在账户DefaultDataLakeBlueprint中启用的，不具备亚马逊 DataZone 管理已发布资产权限所需的权限。您可以通过添加AmazonDataZoneGlueDataAccessRole作为 AWS Lake Formati AmazonDataZoneGlueDataAccessRoleon 管理员或向要发布的资产授予以下权限来解决问题。 • 对资产所在的数据库的“描述”和“描述可授予”权限 • 描述、选择、描述可授予权限、选择对数据库中所有资产的可授予权限，这些资产是您希望 Ama DataZone zon 代表您管理的访问权限。

对 Amazon DataZone 世系资产与上游数据集关联进行故障排除

本节包含有关您可能遇到的 Amazon DataZone 世系问题的疑难解答说明。对于某些 AWS Glue 和 Amazon Redshift-related 开放血统运行事件，您可能会看到资产谱系未链接到上游数据集。本主题说明了场景和几种缓解问题的方法。有关世系的更多信息，请参阅 Amazon 中的数据谱系 DataZone。

SourceIdentifier 在血统节点上

世系节点中的 `sourceIdentifier` 属性表示数据集上发生的事件。有关更多信息，请参阅世系节点中的关键属性。

世系节点表示在相应的数据集或作业上发生的所有事件。血统节点包含一个 “sourceDentifier” 属性，该属性包含相应的标识符。 dataset/job由于我们支持 open-lineage 事件，因此默认情况下，`sourceIdentifier` 值将以数据集、作业和作业运行的“命名空间”和“名称”的组合形式填充。

对于诸如 AWS Glue Amazon Redshift 之类的 AWS 资源，`sourceIdentifier`将是表 AWS Glue ARN 和 Redshift 表 ARN，亚马逊 DataZone 将从中构建运行事件以及其他详细信息，如下所示：

> **Note**
>
> 在中 AWS，ARN 包含每个资源的账户 ID、区域、数据库和表等信息。

- OpenLineage 这些数据集的事件包含数据库和表名。
- 区域是在运行的“environment-properties”分面捕获的。如果区域不存在，系统将使用调用方凭证中的区域。
- AccountId 取自来电者凭证。

SourceIdentifier 关于其中的资产 DataZone

`AssetCommonDetailForm` 具有一个名为“sourceIdentifier”的属性，该属性表示资产代表的数据集的标识符。对于要链接到上游数据集的资产世系节点，需要将该属性与数据集节点的 `sourceIdentifier` 所对应的值一起填充。如果资源是由数据源导入的，则工作流程会自动填充`sourceIdentifier`为表 AWS Glue ARN/Redshift 表 ARN，而通过 `CreateAsset` API 创建的其他资产（包括自定义资产）则应由调用者填充该值。

Amazon 如何 DataZone 根据事件构建 sourceIdentifier？ OpenLineage

对于 AWS Glue 和 Redshift 资产，`sourceIdentifier`是由 Glue 和 Redshift ARN 构造的。以下是亚马逊的 DataZone 构造方式：

AWS Glue 进行筛选

目标是构造一个 OpenLineage 事件，其中输出血统节点`sourceIdentifier`为：

```
arn:aws:glue:us-east-1:123456789012:table/testlfdb/testlftb-1
```

要确定运行是否使用来自的数据 AWS Glue，请查看分`environment-properties`面中是否存在某些关键字。具体而言，如果存在这些指定字段中的任一字段，则系统会假定 `RunEvent` 源自 AWS Glue。

- GLUE_VERSION
- GLUE_COMMAND_CRITERIA
- GLUE_PYTHON_VERSION

```
"run": {
   "runId":"4e3da9e8-6228-4679-b0a2-fa916119fthr",
   "facets":{
      "environment-properties":{
         "_producer":"https://github.com/OpenLineage/OpenLineage/tree/1.9.1/
integration/spark",
         "_schemaURL":"https://openlineage.io/spec/2-0-2/OpenLineage.json#/$defs/
RunFacet",
         "environment-properties":{
            "GLUE_VERSION":"3.0",
            "GLUE_COMMAND_CRITERIA":"glueetl",
            "GLUE_PYTHON_VERSION":"3"
         }
      }
   }
```

对于 AWS Glue 运行，您可以使用分symlinks面中的名称来获取数据库和表名，这些名称可用于构造 ARN。

需要确保名称为 databaseName.tableName：

```
"symlinks": {
   "_producer":"https://github.com/OpenLineage/OpenLineage/tree/1.9.1/integration/
spark",
   "_schemaURL":"https://openlineage.io/spec/facets/1-0-0/SymlinksDatasetFacet.json#/
$defs/SymlinksDatasetFacet",
   "identifiers":[
      {
         "namespace":"s3://object-path",
         "name":"testlfdb.testlftb-1",
         "type":"TABLE"
      }
   ]
}
```

示例 COMPLETE 事件：

```
{
   "eventTime":"2024-07-01T12:00:00.000000Z",
   "producer":"https://github.com/OpenLineage/OpenLineage/tree/1.9.1/integration/glue",
   "schemaURL":"https://openlineage.io/spec/2-0-2/OpenLineage.json#/$defs/RunEvent",
   "eventType":"COMPLETE",
```

```
   "run": {
      "runId":"4e3da9e8-6228-4679-b0a2-fa916119fthr",
      "facets":{
         "environment-properties":{
            "_producer":"https://github.com/OpenLineage/OpenLineage/tree/1.9.1/
integration/spark",
            "_schemaURL":"https://openlineage.io/spec/2-0-2/OpenLineage.json#/$defs/
RunFacet",
            "environment-properties":{
               "GLUE_VERSION":"3.0",
               "GLUE_COMMAND_CRITERIA":"glueetl",
               "GLUE_PYTHON_VERSION":"3"
            }
         }
      }
   },
   "job":{
      "namespace":"namespace",
      "name":"job_name",
      "facets":{
         "jobType":{
            "_producer":"https://github.com/OpenLineage/OpenLineage/tree/1.9.1/
integration/glue",
            "_schemaURL":"https://openlineage.io/spec/facets/2-0-2/
JobTypeJobFacet.json#/$defs/JobTypeJobFacet",
            "processingType":"BATCH",
            "integration":"glue",
            "jobType":"JOB"
         }
      }
   },
   "inputs":[
      {
         "namespace":"namespace",
         "name":"input_name"
      }
   ],
   "outputs":[
      {
         "namespace":"namespace.output",
         "name":"output_name",
         "facets":{
            "symlinks":{
```

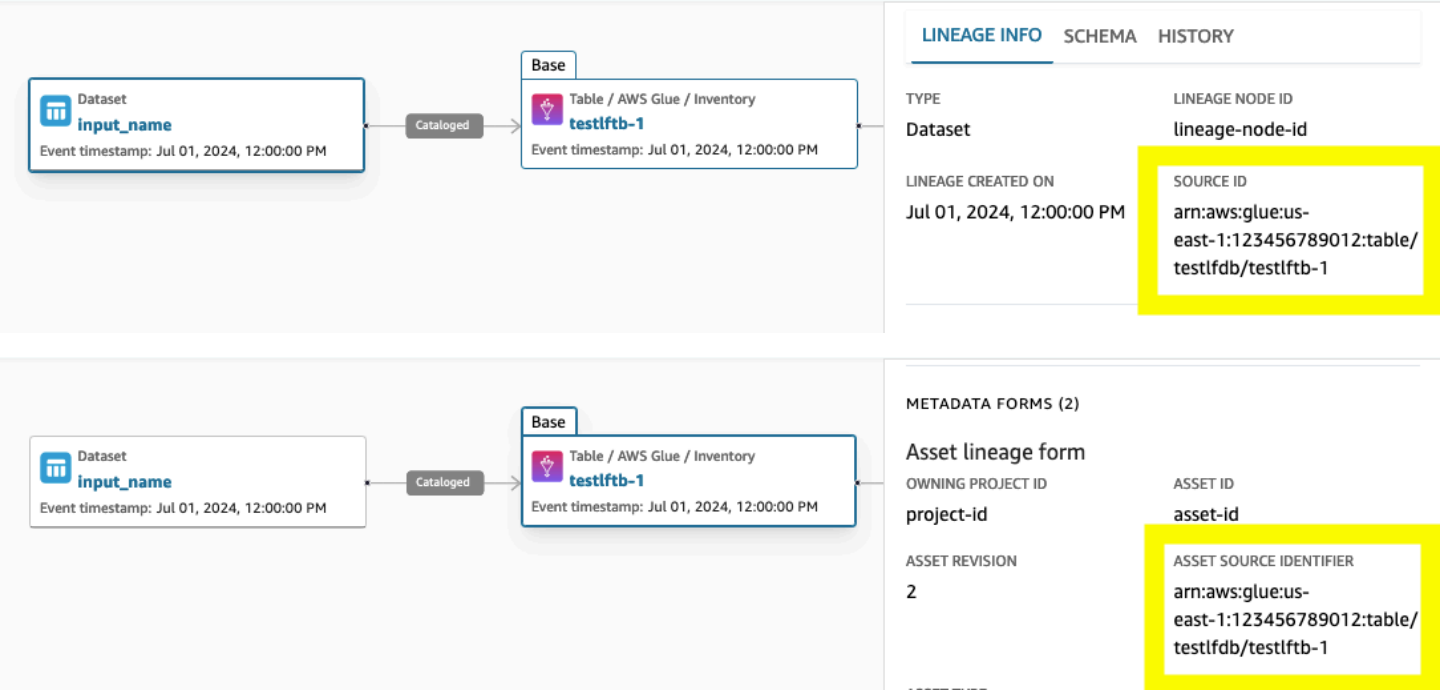
```
                "_producer":"https://github.com/OpenLineage/OpenLineage/tree/1.9.1/
integration/spark",
                "_schemaURL":"https://openlineage.io/spec/facets/1-0-0/
SymlinksDatasetFacet.json#/$defs/SymlinksDatasetFacet",
                "identifiers":[
                   {
                      "namespace":"s3://object-path",
                      "name":"testlfdb.testlftb-1",
                      "type":"TABLE"
                   }
                ]
             }
          }
       }
    ]
}
```

根据提交的 OpenLineage 事件，输出世系节点的 sourceIdentifier 将为：

```
arn:aws:glue:us-east-1:123456789012:table/testlfdb/testlftb-1
```

输出世系节点将连接到资产的世系节点，其中资产的 sourceIdentifier 为：

```
arn:aws:glue:us-east-1:123456789012:table/testlfdb/testlftb-1
```



Amazon Redshift ARN

目标是构造一个 OpenLineage 事件，其中输出血统节点`sourceIdentifier`为：

```
arn:aws:redshift:us-east-1:123456789012:table/workgroup-20240715/tpcds_data/public/
dws_tpcds_7
```

系统根据命名空间来确定输入或输出是否存储在 Redshift 中。具体而言，如果命名空间以 redshift:// 开头或包含字符串 `redshift-serverless.amazonaws.com` 或 `redshift.amazonaws.com`，则它是 Redshift 资源。

```
"outputs": [
    {
        "namespace":"redshift://workgroup-20240715.123456789012.us-
east-1.redshift.amazonaws.com:5439",
        "name":"tpcds_data.public.dws_tpcds_7"
    }
]
```

请注意，命名空间需采用以下格式：

```
provider://{cluster_identifier}.{region_name}:{port}
```

对于 `redshift-serverless`：

```
"outputs": [
    {
        "namespace":"redshift://workgroup-20240715.123456789012.us-east-1.redshift-
serverless.amazonaws.com:5439",
        "name":"tpcds_data.public.dws_tpcds_7"
    }
]
```

这将产生以下 `sourceIdentifier`

```
arn:aws:redshift-serverless:us-east-1:123456789012:table/workgroup-20240715/tpcds_data/
public/dws_tpcds_7
```

根据提交 OpenLineage 的事件，`sourceIdentifier`要映射到下游（即事件的输出）血统节点为：

```
arn:aws:redshift-serverless:us-e:us-east-1:123456789012:table/workgroup-20240715/
tpcds_data/public/dws_tpcds_7
```

该映射可帮助您可视化目录中资产的世系。

替代方法

当上述条件均不满足时，系统使用命名空间/名称来构造 sourceIdentifier：

```
"inputs": [
  {
     "namespace":"arn:aws:redshift:us-east-1:123456789012:table",
     "name":"workgroup-20240715/tpcds_data/public/dws_tpcds_7"
  }
],
"outputs": [
  {
     "namespace":"arn:aws:glue:us-east-1:123456789012:table",
     "name":"testlfdb/testlftb-1"
  }
]
```

排查资产世系节点缺少上游的问题

如果您看不到资产世系节点的上游，则可以执行以下操作来排查它未与数据集关联的原因：

1. 在提供 domainId 和 assetId 的同时调用 GetAsset：

   ```
   aws datazone get-asset --domain-identifier <domain-id> --identifier <asset-id>
   ```

   响应如下所示：

   ```
   {
       .....
       "formsOutput": [
           .....
           {
               "content": "{\"sourceIdentifier\":\"arn:aws:glue:eu-
   west-1:123456789012:table/testlfdb/testlftb-1\"}",
               "formName": "AssetCommonDetailsForm",
               "typeName": "amazon.datazone.AssetCommonDetailsFormType",
   ```

```
            "typeRevision": "6"
        },
        .....
    ],
    "id": "<asset-id>",
    ....
}
```

2. 调用 GetLineageNode 以获取数据集世系节点的 sourceIdentifier。由于无法直接获取相应数据集节点的世系节点，因此您可以先对作业运行调用 GetLineageNode：

```
aws datazone get-lineage-node --domain-identifier <domain-id> --identifier
 <job_namespace>.<job_name>/<run_id>

if you are using the getting started scripts, job name and run ID are printed in the
 console
and namespace is "default". Otherwise you can get these values from run event
 content.
```

示例响应看起来与以下内容类似：

```
{
    .....
    "downstreamNodes": [
        {
            "eventTimestamp": "2024-07-24T18:08:55+08:00",
            "id": "afymge5k4v0euf"
        }
    ],
    "formsOutput": [
        <some forms corresponding to run and job>
    ],
    "id": "<system generated node-id for run>",
    "sourceIdentifier": "default.redshift.create/2f41298b-1ee7-3302-
a14b-09addffa7580",
    "typeName": "amazon.datazone.JobRunLineageNodeType",
    ....
    "upstreamNodes": [
        {
            "eventTimestamp": "2024-07-24T18:08:55+08:00",
            "id": "6wf2z27c8hghev"
        },
        {
```

```
            "eventTimestamp": "2024-07-24T18:08:55+08:00",
            "id": "4tjbcsnre6banb"
        }
    ]
}
```

3. 通过传入 downstream/upstream 节点标识符（您认为该标识符应链接到资产节点）GetLineageNode再次调用，因为这些标识符对应于数据集：

   使用上面的示例响应的示例命令：

```
aws datazone get-lineage-node --domain-identifier <domain-id> --identifier
 afymge5k4v0euf
```

   这将返回与数据集对应的世系节点详细信息：afymge5k4v0euf

```
{
    .....
    "domainId": "dzd_cklzc5s2jcr7on",
    "downstreamNodes": [],
    "eventTimestamp": "2024-07-24T18:08:55+08:00",
    "formsOutput": [
        .....
    ],
    "id": "afymge5k4v0euf",
    "sourceIdentifier": "arn:aws:redshift:us-east-1:123456789012:table/
workgroup-20240715/tpcds_data/public/dws_tpcds_7",
    "typeName": "amazon.datazone.DatasetLineageNodeType",
    "typeRevision": "1",
    ....
    "upstreamNodes": [
        ...
    ]
}
```

4. 比较此数据集节点的 sourceIdentifier 和来自 GetAsset 的响应。如果它们未关联，则它们不匹配，因此在世系 UI 中将不可见。

Non-matching 情景和缓解措施

以下是它们不匹配的常见场景以及可能的缓解措施：

根本原因：这些表存在于与 Amazon DataZone 域名账户不同的账户中。

缓解措施：您可以从关联账户调用 PostLineageEvent 操作。由于将从调用方凭证中选取用于构造 ARN 的 accountId，因此在运行入门脚本或调用 PostLineageEvent 时，您可以从包含表的账户代入角色。这样做将有助于正确构造 ARN 并与资产节点关联。

根本原因：Redshift 的 ARN table/views 包含 Redshift/Redshift-serverless 基于运行事件中相应数据集信息的命名空间和名称属性。 OpenLineage

缓解措施：由于没有确定的方法来获知给定名称是属于集群还是工作组，因此我们将使用以下启发法：

- 如果与数据集对应的“名称”包含“redshift-serverless.amazonaws.com”，则我们使用 redshift-serverless 作为 ARN 的一部分，否则默认为“redshift”。
- 上述情况表示工作组名称的别名不起作用。

根本原因：自定义资产的上游数据集未正常关联。

缓解措施：请务必通过调用与数据集节点的 sourceIdentifier 匹配的 CreateAsset/CreateAssetRevision（对于自定义节点，为 <namespace>/<name>）来填充资产的 sourceIdentifier。

Amazon 的配额 DataZone

您的 AWS 账户对每项 AWS 服务都有默认配额（以前称为限制）。除非另有说明，否则，每个配额是区域特定的。

Amazon DataZone 有以下配额和限制。

亚马逊 DataZone 配额

资源	说明	值
数据资产类型	在一个 DataZone 域中可以创建的最大数据资产类型数量	1000
数据资产	可以在一个 Amazon DataZone 域中创建的最大数据资产数量	100 万
词汇表	可在域中创建的业务术语表的最大数量	1000
业务术语表术语	可在域中创建的业务术语表术语的最大总数	10000
域中的环境	一个 Amazon DataZone 域中的最大环境数量	1000
每项资产的资产筛选条件数量	每个 Amazon DataZone 资产的最大资产筛选器数量	100
每个订阅的筛选条件数量	每次 DataZone 订阅 Amazon 的最大筛选器数量	5
域中的域单元数	Amazon DataZone 域名中域名单位的最大数量	500
域单元中的层次结构级别	域单元的层次结构级别的最大数量	5

资源	说明	值
每个域单元的每策略授权数	每个域单元的每个策略的最大授权数	20
数据产品	一个 DataZone 域中可以创建的最大数据产品数量	500,000
数据来源运行	每个数据来源每天的最大数据来源运行数。	25

亚马逊 DataZone API 速率限制

下表描述了 Amazon DataZone API 的速率限制。这些是每个 AWS 账户在每个区域的限制。

亚马逊 DataZone API 速率限制

API	API 速率限制
CreateGlossary	每秒 5 个事务 (TPS)
UpdateGlossary	20 TPS
GetGlossary	20 TPS
DeleteGlossary	20 TPS
UpdateGlossaryTerm	20 TPS
DeleteGlossaryTerm	20 TPS
CreateAsset	20 TPS
ListAssetRevisions	20 TPS
CreateAssetRevision	20 TPS
DeleteAsset	20 TPS
CreateDataProduct	20 TPS

API	API 速率限制
ListDataProductRevisions	20 TPS
CreateDataProductRevision	20 TPS
DeleteDataProduct	20 TPS
CreateAssetType	20 TPS
DeleteAssetType	20 TPS
CreateFormType	20 TPS
DeleteFormType	20 TPS
搜索	20 TPS
SearchTypes	20 TPS
AcceptPredictions	20 TPS
RejectPredictions	20 TPS
AcceptSubscriptionRequest	3 TPS
CancelSubscription	3 TPS
CreateSubscriptionGrant	3 TPS
CreateSubscriptionRequest	3 TPS
GetSubscriptionEligibility	30 TPS
DeleteSubscriptionGrant	3 TPS
DeleteSubscriptionRequest	3 TPS
DeleteSubscriptionTarget	3 TPS
GetSubscription	8 TPS

API	API 速率限制
GetSubscriptionGrant	8 TPS
GetSubscriptionRequestDetails	8 TPS
ListSubscriptionGrants	8 TPS
ListSubscriptionRequests	8 TPS
ListSubscriptions	8 TPS
ListSubscriptionTargets	8 TPS
RejectSubscriptionRequest	3 TPS
RevokeSubscription	3 TPS
UpdateSubscriptionRequest	3 TPS
UpdateSubscriptionTarget	3 TPS
CreateProjectProfile	3 TPS
UpdateProjectProfile	3 TPS
CreateDomain	8 TPS
UpdateDomain	8 TPS
CreateProject	3 TPS
UpdateProject	3 TPS
DeleteProject	3 TPS
ListProjects	8 TPS
CreateProjectMembership	3 TPS
ListProjectMemberships	8 TPS

API	API 速率限制
DeleteProjectMembership	3 TPS
CreateEnvironment	3 TPS
DeleteEnvironment	3 TPS
UpdateEnvironment	3 TPS
ListEnvironments	8 TPS
GetEnvironment	8 TPS
GetEnvironmentCredentials	8 TPS
CreateEnvironmentProfile	8 TPS
ListEnvironmentProfiles	8 TPS
ListEnvironmentBlueprints	8 TPS
PutEnvironmentBlueprintConfiguration	10 TPS
StartMetadataGenerationRun	10 TPS
CancelMetadataGenerationRun	20 TPS
CreateDomainUnit	20 TPS
AddPolicyGrant	20 TPS
AddEntityOwner	20 TPS
CreateRule	20 TPS
UpdateRule	20 TPS
CreateDataSource	20 TPS
UpdateDataSource	20 TPS

API	API 速率限制
DeleteDataSource	20 TPS
ListDataSources	20 TPS
SearchListings	16 TPS
StartDataSourceRun	20 TPS
UpdateDataSourceRunActivities	20 TPS
PostLineageEvent	20 TPS
CreateConnection	20 TPS
UpdateConnection	20 TPS
GetConnection	20 TPS
ListConnections	20 TPS
DeleteConnection	20 TPS
CreateListingChangeSet	20 TPS

Amazon DataZone 用户指南的文档历史记录

下表描述了 Amazon 发布的文档 DataZone。

变更	说明	日期
订阅目标 API 的授权模型	澄清了`CreateSubscriptionTarget` `UpdateSubscriptionTarget` 、和`DeleteSubscriptionTarget` 是在 AWS 账户级别授权的，不需要项目成员资格。有关更多信息，请参阅使用现有 IAM 角色配送 Amazon DataZone 订阅。	2026年10月7日
AmazonDataZoneRedshiftGlueProvisioningPolicy 政策更新	的政策更新AmazonDataZoneRedshiftGlueProvisioningPolicy。此更新添加了`cloudformation:UntagResource` 权限。在环境配置期间，Amazon DataZone 使用此权限从其管理的 AWS CloudFormation 堆栈中移除标签。有关更多信息，请参阅 Amazon 对 AWS 托管策略的 DataZone 更新。	2026年9月10日
AmazonDataZoneSageMakerEnvironmentRolePermissionsBoundary 政策更新	的政策更新AmazonDataZoneSageMakerEnvironmentRolePermissionsBoundary。此更新为`sagemaker:DeleteNotebookInstanceLifecycleConfig` 操作添加了“拒绝”语句。有关更多信息，	2026 年 7 月 15 日

	请参阅 Amazon 对 AWS 托管策略的 DataZone 更新。	
AmazonDataZoneSageMakerEnvironmentRolePermissionsBoundary -政策更新	的政策更新AmazonDataZoneSageMakerEnvironmentRolePermissionsBoundary。为`sagemaker:CreateNotebookInstanceLifecycleConfig` 操作添加了“拒绝”语句。有关更多信息，请参阅 Amazon 对 AWS 托管策略的 DataZone 更新。	2026 年 6 月 26 日
AmazonDataZoneSageMakerEnvironmentRolePermissionsBoundary -政策更新	的政策更新AmazonDataZoneSageMakerEnvironmentRolePermissionsBoundary。为该`sagemaker:UpdateNotebookInstanceLifecycleConfig` 操作添加了“拒绝”语句，以限制此高权限操作。有关更多信息，请参阅 Amazon 对 AWS 托管策略的 DataZone 更新。	2026年3月11日
AmazonDataZoneDomainExecutionRolePolicy -政策更新	策略更新 AmazonDataZoneDomainExecutionRolePolicy-为`QueryGraph` 操作添加权限以支持基于图表的实体搜索功能。有关更多信息，请参阅 Amazon 对 AWS 托管策略的 DataZone 更新。	2026 年 2 月 25 日

AmazonDataZoneGlueManageAccessRolePolicy -政策更新	政策更新 AmazonDataZoneGlueManageAccessRolePolicy-为`GetConnection` 操作添加权限，以支持对基于连接的 AWS Glue 数据源进行数据沿袭捕获。有关更多信息，请参阅 Amazon 对 AWS 托管策略的 DataZone 更新。	2025 年 7 月 30 日
AmazonDataZoneFullAccess -政策更新	的政策更新 AmazonDataZoneFullAccess-概括了新域名的范围 SecretsManager `create`和`tag`权限，其格式将`dzd-`改为。有关更多信息，请参阅 Amazon 对 AWS 托管策略的 DataZone 更新。`dzd_..`	2025 年 7 月 23 日
AmazonDataZoneFullAccess -政策更新	策略更新 AmazonDataZoneFullAccess-允许控制台附加或更新 AWS RAM 资源共享中的 AWS 托管权限。有关更多信息，请参阅 Amazon 对 AWS 托管策略的 DataZone 更新。	2025 年 5 月 22 日

AmazonDataZoneGlueManageAccessRolePolicy -政策更新	政策更新 AmazonDataZoneGlueManageAccessRolePolicy-Amazon DataZone 项目用户角色用作联合表的数据传输角色。此更新为 `iam:PassRole` 语句增加了 `datazone_usr_role*` ，使项目用户角色能够用于此目的。有关更多信息，请参阅 Amazon 对 AWS 托管策略的 DataZone 更新。	2025 年 5 月 21 日
AmazonDataZoneSageMakerProvisioningRolePolicy -政策更新	政策更新 AmazonDataZoneSageMakerProvisioningRolePolicy-增加了对该`glue:GetConnection` 操作的支持。有关更多信息，请参阅 Amazon 对 AWS 托管策略的 DataZone 更新。	2025 年 1 月 2 日
AmazonDataZoneSageMakerEnvironmentRolePermissionsBoundary -政策更新	的政策更新 AmazonDataZoneSageMakerEnvironmentRolePermissionsBoundary——此更改增加了权限边界，使亚马逊 DataZone 能够`CreateUserProfile` 使用必要的标签成功调用。`sagemaker:AddTags` 有关更多信息，请参阅 Amazon 对 AWS 托管策略的 DataZone 更新。	2024 年 12 月 3 日

AmazonDataZoneSageMakerAccess，以及 AmazonDataZoneGlueManageAccessRolePolicy -政策更新	对、和 AmazonDataZoneGlueManageAccessRolePolicy-的政策进行了更新 AmazonDataZoneFullAccessAmazonDataZoneSageMakerAccess，以支持亚马逊 SageMaker Unified Studio 体验。有关更多信息，请参阅 Amazon 对 AWS 托管策略的 DataZone 更新。	2024 年 12 月 3 日
AmazonDataZoneDomainExecutionRolePolicy 以及 AmazonDataZoneFullUserAccess -政策更新	策略更新，旨在支持针对订阅请求的元数据强制规则。有关更多信息，请参阅 Amazon 对 AWS 托管策略的 DataZone 更新。	2024 年 11 月 20 日
Amazon DataZone 推出针对订阅请求的元数据强制执行规则	Amazon DataZone 针对订阅请求的新元数据强制执行规则使域单位所有者能够为数据使用者制定明确的元数据要求，简化访问请求并增强数据治理，从而加强数据治理。此功能使组织能够遵守其元数据标准、实施自定义工作流和提供一致、受管控的数据访问体验。有关更多信息，请参阅针对订阅请求的元数据强制规则。	2024 年 11 月 20 日
AmazonDataZoneRedshiftGlueProvisioningPolicy -政策更新	添加 `iam:DeletePolicyVersion` 以允许用户删除使用 `datazone*` 创建的策略的策略版本。这有助于解除对需要更新环境用户角色策略的用户的屏蔽。有关更多信息，请参阅 Amazon 对 AWS 托管策略的 DataZone 更新。	2024 年 10 月 22 日

AWS CloudFormation 支持自定义 AWS 服务蓝图	Amazon DataZone 增加了对定制 AWS 服务蓝图的 AWS CloudFormation 支持。这项新功能使您 AWS CloudFormation 能够使用在 Amazon 中自动创建环境 DataZone。借助自定义蓝图，管理员现在可以使用现有 IAM 角色将 Amazon 无缝集成 DataZone 到现有的数据管道中，将数据资产发布到 Amazon DataZone 目录，从而促进这些资产的受控共享，并增强对整个基础设施的治理。有关更多信息，请参阅 Amazon DataZone 资源类型参考。	2024 年 9 月 12 日
域单元	Amazon DataZone 推出了一组新的数据治理功能，称为域单元和授权策略，使客户能够根据其业务需求创建业务 unit/team 级组织并管理策略。通过添加域单元，用户可以组织、创建、搜索和查找与业务部门或团队关联的数据资产和项目。通过授权策略，这些域单位用户可以设置访问策略，以便在 Amazon 中创建项目、术语表和使用计算资源。DataZone	2024 年 8 月 5 日

数据产品	Amazon DataZone 推出了数据产品，可将数据资产分组为针对特定业务用例量身定制的定义明确、独立的软件包。例如，营销分析数据产品可以捆绑营销活动数据、管道数据和客户数据等各种数据资产。借助数据产品，客户可以简化发现和订阅流程，使它们与业务目标保持一致，并减少处理单个资产时的冗余。	2024 年 8 月 5 日
AmazonDataZoneDomainExecutionRolePolicy 以及 AmazonDataZoneFullUserAccess -政策更新	对AmazonDataZoneDomainExecutionRolePolicy和的政策进行了更新，AmazonDataZoneFullUserAccess以支持用于创建和管理 Amazon DataZone 域单元和数据产品的新 API。有关更多信息，请参阅 Amazon 对 AWS 托管策略的 DataZone 更新。	2024 年 8 月 5 日

Fine-grained 访问控制	亚马逊引入 DataZone 了精细的访问控制，使您可以精细控制亚马逊 DataZone业务数据目录中的数据资产，跨数据湖和数据仓库。利用此新功能，数据所有者现在可以仅允许访问行级和列级的特定数据记录，而不是授予对整个数据资产的访问权限。例如，如果您的数据列包含个人身份信息（PII）等敏感信息，则可以仅允许访问必要的列，从而在确保敏感信息受到保护的同时仍允许访问非敏感数据。同样，您可以控制行级访问权限，只允许用户查看与其角色或任务相关的记录。	2024 年 7 月 2 日
AmazonDataZoneGlueManageAccessRolePolicy -政策更新	政策更新 AmazonDataZoneGlueManageAccessRolePolicy——亚马逊 DataZone 正在添加用于细粒度访问控制功能的 IAM 权限，以缩小在 Lake Formation 中授予的权限范围。有关更多信息，请参阅 Amazon 对 AWS 托管策略的 DataZone 更新。	2024 年 7 月 2 日

数据世系	Amazon DataZone 推出数据沿袭预览版，帮助客户可视化来自 OpenLineage-enabled 系统或通过 API 的世系事件，并跟踪数据从源头到消费的移动。使用亚马逊 DataZone 的 OpenLineage-compatible API，域管理员和数据创建者可以捕获和存储超出亚马逊可用范围的世系事件 DataZone，包括 Amazon S3、Glue 和其他 AWS 服务中的转换。此外，Amazon DataZone 版本与每个事件保持一致，使用户能够在任何时间点可视化血统或比较资产或任务历史的转换。此历史世系可让用户更深入地了解数据的演变过程，这对于故障排除、审计和验证数据资产的完整性至关重要。	2024 年 6 月 27 日
AmazonDataZoneExecutionRolePolicy 以及 AmazonDataZoneFullUserAccess -政策更新	对AmazonDataZoneExecutionRolePolicy和的策略进行了更新 AmazonDataZoneFullUserAccess，以启用对数据沿袭和细粒度访问控制 API 的支持。有关更多信息，请参阅 Amazon 对 AWS 托管策略的 DataZone 更新。	2024 年 6 月 27 日

自定义 AWS 服务蓝图	借助定制 AWS 服务蓝图，如果您拥有包括 IAM 角色、数据湖、数据网格、Amazon S3 存储桶和 Amazon Redshift 集群在内的现有 AWS 资源，您现在可以使用自己的自定义 IAM 角色指定对这些现有资源的权限，这样您的亚马逊 DataZone 用户就可以利用发布和订阅来共享和管理这些资源。借助定制 AWS 服务蓝图，Amazon DataZone 管理员可以使用自己的自定义角色配置 AWS 服务环境。他们可以为这些 AWS 服务环境配置操作链接，从而提供对其任何现有 AWS 资源的联合访问权限。他们还可以在这些自定义 AWS 服务环境中配置订阅目标和数据源。管理员可以在自己的 Amazon DataZone 域账户中或他们想要发布、订阅、发现或管理数据的任何关联账户中设置 AWS 服务环境。	2024 年 6 月 17 日
AmazonDataZoneGlueManageAccessRolePolicy -政策更新	的AmazonDataZoneGlueManageAccessRolePolicy 政策更新增加了亚马逊自助订阅功能所需的 IAM 权限，DataZone 以缩小湖形成时授予的权限范围。使用自行订阅功能时，只能向标记的资源授予 Lake Formation 权限。有关更多信息，请参阅 Amazon 对 AWS 托管策略的 DataZone 更新。	2024 年 6 月 14 日

AmazonDataZoneFullAccess - 政策更新	的政策更新AmazonDataZoneFullAccess 使得 Amazon DataZone 管理控制台能够代表用户使用域和项目标签创建密钥。还包括 `ram:ListResourceSharePermissions` 操作以允许从域所有者账户进行管理，以便查看关联账户的账户关联状态。有关更多信息，请参阅 Amazon 对 AWS 托管策略的 DataZone 更新。	2024 年 6 月 14 日
AmazonDataZoneDomainExecutionRolePolicy -政策更新	的政策更新为AmazonDataZoneDomainExecutionRolePolicy 亚马逊添加了新的 API DataZone ，使用户能够为其亚马逊 DataZone 环境配置操作。有关更多信息，请参阅 Amazon 对 AWS 托管策略的 DataZone 更新。	2024 年 6 月 14 日

数据来源创建增强功能	Amazon DataZone 对数据源创建流程进行了增强，以简化数据生产者的访问管理。通过这些更新，当数据创建者创建用于发布其 AWS Glue 和 Amazon Redshift 资产的数据源时，亚马逊会向项目成员 DataZone 授予只读权限。创建 AWS Glue 数据源时，Amazon DataZone 会自动向用于创建数据源的环境的 IAM 角色授予 “只读” 权限，允许访问相关 AWS Glue 数据库中的所有表。同样，对于亚马逊 Redshift 数据源，亚马逊 DataZone 授予对数据源中使用的亚马逊 Redshift 架构中所有表的 “只读” 访问权限。	2024 年 6 月 10 日
与亚马逊集成 SageMaker	亚马逊 DataZone 推出与亚马逊的集成， SageMaker以帮助数据生产者和消费者无缝切换 SageMaker 到亚马逊，在机器学习 (ML) 项目上进行协作，同时对数据和机器学习资产实施访问管理。借助 Amazon DataZone 和 Amazon 之间新的内置集成 SageMaker，数据使用者和创建者可以简化基础设施设置中的机器学习管理，协作开展业务计划，并轻松管理数据和机器学习资产。	2024 年 5 月 6 日

AmazonDataZoneSageMakerProvisioningRolePolicy - 新政策	名为的新政策AmazonDataZoneSageMakerProvisioningRolePolicy授 DataZone 予亚马逊与亚马逊 SageMaker互操作所需的权限。有关更多信息，请参阅 Amazon 对 AWS 托管策略的 DataZone 更新。	2024 年 4 月 30 日
AmazonDataZoneSageMakerEnvironmentRolePermissionsBoundary -新的权限边界	新的权限边界已调用 AmazonDataZoneSageMakerEnvironmentRolePermissionsBoundary 。当您通过亚马逊 DataZone 数据门户创建亚马逊 SageMaker 环境时，亚马逊会 DataZone将此权限边界应用于在创建环境期间生成的 IAM 角色。权限边界限制了 Amazon DataZone 创建的角色和您添加的任何角色的范围。有关更多信息，请参阅 Amazon 对 AWS 托管策略的 DataZone 更新。	2024 年 4 月 30 日
AmazonDataZoneSageMakerAccess -新政策	名为的新政策AmazonDataZoneSageMakerAccess 授 DataZone 予亚马逊授予用户访问亚马逊 SageMaker 环境中各种资源所需的权限。有关更多信息，请参阅 Amazon 对 AWS 托管策略的 DataZone 更新。	2024 年 4 月 30 日

AmazonDataZoneFullAccess - 政策更新	对AmazonDataZoneFullAccess策略的更新，增加了 `DescribeSecurityGroups` 操作访问权限，以提高账户管理员的可用性，在控制台中配置蓝图和`GetPolicy` 操作以帮助检索有关指定托管策略的信息。有关更多信息，请参阅 Amazon 对 AWS 托管策略的 DataZone 更新。	2024 年 4 月 30 日
Lake Formation 混合访问模式	亚马逊推 DataZone 出了与 AWS Lake Formation混合访问模式的集成。这种集成使您能够轻松地通过亚马逊发布和共享您的 AWS Glue 表 DataZone，而无需先在 AWS Lake Formation 中注册它们。首先，管理员在 Amazon DataZone 控制台中启用`DefaultDataLake` 蓝图下的数据位置注册设置。然后，当数据使用者订阅通过 IAM 权限管理的 AWS Glue 表时，亚马逊 DataZone首先以混合模式注册该表的 Amazon S3 位置，然后通过 La AWS ke Formation 管理该表的权限，向数据使用者授予访问权限。这样可以确保使用新授予的 La AWS ke Formation 权限继续存在表上的 IAM 权限，而不会中断任何现有工作流程。有关更多信息，请参阅亚马逊与 AWS Lake F DataZone ormation 混合模式的集成。	2024 年 4 月 3 日

数据质量	亚马逊 DataZone 推出与 AWS Glue 数据质量的集成，并提供 API 来集成来自第三方数据质量解决方案的数据质量指标。新的集成使您能够将Glue AWS 数据质量分数自动发布到亚马逊 DataZone业务数据目录中。Amazon DataZone API 可用于从第三方来源获取质量指标。发布后，数据使用者可以轻松搜索数据资产，查看精细的质量指标并识别失败的检查和规则，从而加快制定业务决策。有关更多信息，请参阅 Amazon 中的数据质量 DataZone。	2024 年 4 月 3 日
AmazonDataZoneS3Manage---新角色 <region><domainId>	名为 AmazonDataZoneS3Manage- <region>的新角色在 <domainId>亚马逊 DataZone 调用 AWS Lake Formation 注册亚马逊简单存储服务 (Amazon S3) 位置时使用。AWS Lake Formation 在访问该位置的数据时扮演这个角色。有关更多信息，请参阅 Amazon 对 AWS 托管策略的 DataZone 更新。	2024 年 4 月 1 日
AmazonDataZoneGlueManageAccessRolePolicy -政策更新	更新了AmazonDataZoneGlueManageAccessRolePolicy以启用对允许 Amazon DataZone 启用发布和数据访问权限的权限的支持。有关更多信息，请参阅 Amazon 对 AWS 托管策略的 DataZone 更新。	2024 年 4 月 1 日

AmazonDataZoneDomainExecutionRolePolicy 以及 AmazonDataZoneFullUserAccess -政策更新	更新了AmazonDataZoneDomainExecutionRolePolicy和AmazonDataZoneFullUserAccess以启用对 `CancelMetadataGenerationRun` API 的支持。有关更多信息，请参阅 Amazon 对 AWS 托管策略的 DataZone 更新。	2024 年 3 月 29 日
AmazonDataZoneFullAccess -政策更新	Amazon DataZone 宣布正式发布新的生成 AI-based功能，通过丰富业务数据目录来改善数据发现、数据理解和数据使用。只需单击一下，数据创建者即可生成全面的业务数据描述和上下文，突出显示有影响力的列，并包含有关分析应用场景的建议。此次发布增加了对 API 的支持，数据创建者可以使用这些 API 以编程方式生成资产描述。	2024 年 3 月 27 日
AmazonDataZoneFullAccess -政策更新	亚马逊对其亚马逊Redshift集成 DataZone 进行了多项增强，简化了发布和订阅亚马逊Redshift表格和视图的过程。这些更新简化了数据创建者和使用者的体验，使他们能够使用 Amazon DataZone 管理员提供的预配置凭证和连接参数快速创建数据仓库环境。此外，这些增强功能使管理员能够更好地控制谁可以使用其 AWS 账户和 Amazon Redshift 集群中的资源以及用于什么目的。	2024 年 3 月 21 日

AmazonDataZoneFullAccess - 政策更新	更新了，使用户AmazonDataZoneFullAccess 能够在 Amazon DataZone 管理控制台中选择自己的密钥、集群、vpc 和子网，而不必在文本框中键入它们。有关更多信息，请参阅 Amazon 对 AWS 托管策略的 DataZone 更新。	2024 年 3 月 13 日
AmazonDataZoneDomainExecutionRolePolicy -政策更新	通过确定哪些蓝图在哪个账户和区域启用，更新了以启用对创建环境配置文件所需的 ListEnvironmentBlueprintConfigurationSummaries API 的支持。AmazonDataZoneDomainExecutionRolePolicy有关更多信息，请参阅 Amazon 对 AWS 托管策略的 DataZone 更新。	2024 年 2 月 1 日
增强了对 CloudFormation 的使用	现在，Amazon 的用户 DataZone 可以利用它 AWS CloudFormation 来有效地建模和管理一套亚马逊 DataZone 资源。此方法有助于一致地预置资源，并支持通过基础设施即代码来进行生命周期管理。利用自定义模板，您可以精确地定义所需的资源及其相互依赖项。有关更多信息，请参阅 Amazon DataZone 资源类型参考。	2024 年 1 月 18 日

自定义资产	对自定义资产的支持使Amazon DataZone 能够通过数据门户对非结构化数据（包括仪表板、查询和模型）的资产进行分类，从而使您可以更轻松地直接在数据门户中添加自定义资产以及之前提供的 API 支持。通过在 Amazon 中创建 DataZone、更新和发布自定义资产，您可以共享、查找、订阅任何类型的资产，并构建可管理这些资产的业务工作流程。有关更多信息，请参阅创建自定义资产类型。	2024 年 1 月 5 日
将 IAM 主体添加为项目成员	现在，您可以将 IAM 委托人添加为项目成员，即使这些 IAM 委托人尚未登录 Amazon DataZone （之前的要求）。在域管理员或 IT 管理员将 `iam:GetUser` 和 `iam:GetRole` 添加到域的域执行角色后，项目所有者只需提供 IAM 角色或 IAM 用户的 Amazon 资源名称（ARN）即可将 IAM 主体添加为成员。IAM 委托人仍然必须拥有访问 Amazon 所需的 IAM 权限 DataZone ，这些权限可以在 IAM 控制台中进行配置。有关更多信息，请参阅向项目添加成员。	2024 年 1 月 5 日

删除域	删除域是一项功能，可让您更轻松地删除域。现在，即使域不为空（如包含项目、环境、资产、数据来源等），也可以继续删除域。有关更多信息，请参阅删除 Amazon DataZone 域名。	2023 年 12 月 27 日
Lake Formation 混合模式	亚马逊 DataZone 增加了对 La AWS ke Formation混合模式的支持。有了这种支持，如果您将 AWS Glue表发布到亚马逊 DataZone ，其 AWS S3位置在混合模式下注册在 Lake Formation中，则亚马逊 DataZone 会将此表视为托管资产，并且可以管理该表的订阅授权。在此功能发布之前，亚马逊 DataZone 会将此表视为非托管资产，也就是说，亚马逊 DataZone 将无法授予对该表的订阅。有关更多信息，请参阅为亚马逊配置 Lake Formation 权限 DataZone。	2023 年 12 月 22 日
HIPAA 合规性	亚马逊现在 DataZone 是 U.S。符合 1996 年《健康保险流通与责任法》(HIPAA)。要查看符合 HIPAA 标准的 AWS 服务列表，请参阅。https://aws.amazon.com/compliance/hipaa-eligible-services-reference/	2023 年 12 月 14 日

AmazonDataZoneGlueManageAccessRolePolicy -政策更新	更新了AmazonDataZoneGlueManageAccessRolePolicy以启用对 AWS Lake Formation 混合模式的支持。有关更多信息，请参阅 Amazon 对 AWS 托管策略的 DataZone 更新。	2023 年 12 月 14 日
AmazonDataZoneFullUserAccess 以及 AmazonDataZoneDomainExecutionRolePolicy -政策更新	亚马逊 DataZone 更新了AmazonDataZoneFullUserAccess和AmazonDataZoneDomainExecutionRolePolicy政策，以支持亚马逊中的生成 AI-powered 数据描述功能 DataZone。有关更多信息，请参阅 Amazon 对 AWS 托管策略的 DataZone 更新。	2023 年 11 月 28 日

人工智能建议	AWS 宣布在 Amazon DataZone 中预览一项新的生成 AI-based 功能，该功能可通过丰富业务数据目录来改善数据发现、数据理解和数据使用。只需单击一下，数据创建者即可生成全面的业务数据描述和上下文，突出显示有影响力的列，并包含有关分析应用场景的建议。借助 Amazon 中描述的人工智能建议 DataZone，数据使用者可以识别分析所需的数据表和列，从而提高数据可发现性并减少与数据生成者的来回通信。预览版适用于在以下 AWS 区域配置的 Amazon DataZone 域名：美国东部（弗吉尼亚北部）、美国西部（俄勒冈）。有关更多信息，请参阅使用机器学习和生成式人工智能。	2023 年 11 月 28 日
DefaultDataLake 蓝图	Amazon为 DefaultDataLake 蓝图添加 DataZone 了一项增强功能，使您可以更好地控制谁可以从您的 AWS 账户发布哪些数据。推出此功能时引入了两项关键更改。	2023 年 11 月 20 日

AmazonDataZoneEnvironmentRolePermissionsBoundary -政策更新	Amazon 对AmazonDataZoneEnvironmentRolePermissionsBoundary托管策略 DataZone 进行了更新，其中包括根据ResourceTag 条件限定的额外athena:GetQueryResultsStream 权限。有关更多信息，请参阅 Amazon 对 AWS 托管策略的 DataZone 更新。	2023 年 11 月 17 日
AmazonDataZoneRedshiftManageAccessRolePolicy -政策更新	亚马逊 DataZone 更新了AmazonDataZoneRedshiftManageAccessRolePolicy政策，取消了对该redshift:AssociateDataShareConsumer 操作的组织编号的检查。这使您能够在 AWS 组织之间共享资源。有关更多信息，请参阅 Amazon 对 AWS 托管策略的 DataZone 更新。	2023 年 11 月 16 日
用户指南的 GA 版本	《亚马逊 DataZone 用户指南》正式上市 (GA) 版本。	2023 年 10 月 15 日
AmazonDataZoneFullUserAccess -政策更新	亚马逊 DataZone 更新了授予亚马逊完全访问权限的AmazonDataZoneFullUserAccess政策 DataZone，但不允许管理域名、用户或关联账户。有关更多信息，请参阅亚马逊对 AWS 托管政策的 DataZone 更新。	2023 年 10 月 2 日

AmazonDataZonePreviewConsoleFullAccess -政策已弃用	亚马逊 DataZone 已弃用AmazonDataZonePreviewConsoleFullAccess。有关更多信息，请参阅亚马逊对 AWS 托管 DataZone 策略的更新。	2023 年 9 月 29 日
AmazonDataZonePortalFullAccessPolicy -政策已弃用	亚马逊 DataZone 已弃用AmazonDataZonePortalFullAccessPolicy。有关更多信息，请参阅亚马逊对 AWS 托管 DataZone 策略的更新。	2023 年 9 月 29 日
AmazonDataZoneDomainExecutionRolePolicy -新政策	亚马逊 DataZone 添加了一项名为 “” 的新政策AmazonDataZoneDomainExecutionRolePolicy。这是 Amazon DataZone `AmazonDataZoneDomainExecutionRole` 服务角色的默认策略。亚马逊使用此角色 DataZone 对亚马逊 DataZone 域中的数据进行分类、发现、管理、共享和分析。您可以将 `AmazonDataZoneDomainExecutionRolePolicy` 策略附加到 `AmazonDataZoneDomainExecutionRole` 。有关更多信息，请参阅 Amazon 对 AWS 托管策略的 DataZone 更新。	2023 年 9 月 25 日

AmazonDataZoneCrossAccountAdmin -新政策	亚马逊 DataZone 添加了一项名为的新政策 AmazonDataZoneCrossAccountAdmin，允许用户使用亚马逊 DataZone 及其关联账户。有关更多信息，请参阅 Amazon 对 AWS 托管策略的 DataZone 更新。	2023 年 9 月 19 日
AmazonDataZoneRedshiftManageAccessRolePolicy -新政策	亚马逊 DataZone 添加了一项名为的新政策 AmazonDataZoneRedshiftManageAccessRolePolicy，该政策授予 DataZone 允许亚马逊启用数据发布和访问权限的权限。有关更多信息，请参阅 Amazon 对 AWS 托管策略的 DataZone 更新。	2023 年 9 月 12 日
AmazonDataZoneRedshiftGlueProvisioningPolicy -新政策	亚马逊 DataZone 添加了一项名为的新政策 AmazonDataZoneRedshiftGlueProvisioningPolicy，该政策向亚马逊 DataZone授予与支持的数据源进行互操作所需的权限。有关更多信息，请参阅 Amazon 对 AWS 托管策略的 DataZone 更新。	2023 年 9 月 12 日

AmazonDataZoneGlueManageAccessRolePolicy -新政策	亚马逊 DataZone 添加了一项名为 “AmazonDataZoneGlueManageAccessRolePolicy授予亚马逊向目录发布 AWS Glue 数据的 DataZone权限” 的新政策。它还授予亚马逊授予访问DataZone权限或撤销对目录中已发布的 AWS Glue 资产的访问权限的权限。有关更多信息，请参阅 Amazon 对 AWS 托管策略的 DataZone 更新。	2023 年 9 月 12 日
AmazonDataZoneFullUserAccess -新政策	亚马逊 DataZone 添加了一项名为的新政策 AmazonDataZoneFullUserAccess，该政策允许 DataZone 通过数据门户网站访问亚马逊。有关更多信息，请参阅 Amazon 对 AWS 托管策略的 DataZone 更新。	2023 年 9 月 12 日
AmazonDataZoneFullAccess -新政策	亚马逊 DataZone 添加了一项名为的新政策 AmazonDataZoneFullAccess， DataZone 通过 AWS 管理控制台提供对亚马逊的完全访问权限。有关更多信息，请参阅 Amazon 对 AWS 托管策略的 DataZone 更新。	2023 年 9 月 12 日

AmazonDataZoneEnvironmentRolePermissionsBoundary -新政策	Amazon DataZone 添加了一项名为的新政策 AmazonDataZoneEnvironmentRolePermissionsBoundary，该政策限制了其所关联的预配置 IAM 委托人。有关更多信息，请参阅 Amazon 对 AWS 托管策略的 DataZone 更新。	2023 年 9 月 12 日
托管式策略更新	对 AmazonDataZonePreviewConsoleFullAccess 托管策略的更新。有关更多信息，请参阅 Amazon 对 AWS 托管策略的 DataZone 更新。	2023 年 6 月 13 日
托管式策略更新	对 AmazonDataZoneProjectDeploymentPermissionsBoundary 托管策略的更新。有关更多信息，请参阅 Amazon 对 AWS 托管策略的 DataZone 更新。	2023 年 4 月 3 日
???	亚马逊 DataZone （预览版）用户指南的初始版本。	2023 年 3 月 29 日

本文属于机器翻译版本。若本译文内容与英语原文存在差异，则一律以英文原文为准。