



用户指南

# AWS Entity Resolution 数据匹配服务



# AWS Entity Resolution 数据匹配服务: 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 AWS Entity Resolution 数据匹配服务？ .....	1
你是首次 AWS Entity Resolution 数据匹配服务 使用吗？ .....	1
的特点 AWS Entity Resolution 数据匹配服务 .....	1
相关服务 .....	4
正在访问 AWS Entity Resolution 数据匹配服务 .....	4
的定价 AWS Entity Resolution 数据匹配服务 .....	5
设置 .....	6
报名参加 AWS .....	6
创建管理员用户 .....	6
为控制台用户创建 IAM 角色 .....	7
创建工作流程工作角色 .....	8
准备输入数据表 .....	15
准备第一方输入数据 .....	15
步骤 1：准备第一方数据表 .....	15
步骤 2：以支持的数据格式保存输入数据表 .....	17
步骤 3：将您的输入数据表上传到 Amazon S3 .....	17
步骤 4：创建 AWS Glue 表 .....	18
步骤 4：创建分区表 AWS Glue .....	19
准备第三方输入数据 .....	21
步骤 1：在上订阅提供商服务 AWS Data Exchange .....	22
步骤 2：准备第三方数据表 .....	23
步骤 3：以支持的数据格式保存输入数据表 .....	27
步骤 4：将您的输入数据表上传到 Amazon S3 .....	27
步骤 5：创建 AWS Glue 表 .....	28
架构映射 .....	30
创建架构映射 .....	30
克隆架构映射 .....	41
编辑架构映射 .....	42
删除架构映射 .....	43
ID 命名空间 .....	44
ID 命名空间来源 .....	44
创建 ID 命名空间源（基于规则） .....	45
创建 ID 命名空间源（提供者服务） .....	48
ID 命名空间目标 .....	50

创建 ID 命名空间目标 ( 基于规则的方法 ) .....	51
创建 ID 命名空间目标 ( 提供者服务方法 ) .....	53
编辑 ID 命名空间 .....	54
删除 ID 命名空间 .....	54
添加或更新 ID 命名空间的资源策略 .....	54
匹配工作流程 .....	56
匹配工作流程类型 .....	56
数据输出选项 .....	57
匹配工作流程结果 .....	58
创建基于规则的匹配工作流程 .....	58
高级规则类型 .....	60
简单规则类型 .....	76
创建基于机器学习的匹配工作流程 .....	85
创建基于提供商服务的匹配工作流程 .....	90
使用创建匹配的工作流程 LiveRamp .....	90
使用创建匹配的工作流程 TransUnion .....	97
使用 UID 2.0 创建匹配的工作流程 .....	102
编辑匹配的工作流程 .....	106
删除匹配的工作流程 .....	107
修改或生成匹配 ID .....	108
查找比赛编号 .....	111
从基于规则或 ML-based 匹配的工作流程中删除记录 .....	114
问题排查 .....	115
我在运行匹配的工作流程后收到了错误文件 .....	115
ID 映射工作流程 .....	117
一个人的身份映射工作流程 AWS 账户 .....	118
先决条件 .....	118
创建 ID 映射工作流程 ( 基于规则 ) .....	120
创建 ID 映射工作流程 ( 提供者服务 ) .....	125
跨两个 ID 映射工作流程 AWS 账户 .....	129
先决条件 .....	130
创建 ID 映射工作流程 ( 基于规则 ) .....	131
创建 ID 映射工作流程 ( 提供者服务 ) .....	135
运行 ID 映射工作流程 .....	140
运行自定义 ID 映射工作流程 .....	142
编辑 ID 映射工作流程 .....	144

删除 ID 映射工作流程 .....	145
为 ID 映射工作流程添加或更新资源策略 .....	145
提供商集成 .....	147
要求 .....	147
在上列出提供商服务 AWS Data Exchange .....	147
确定你的属性 .....	148
索取 AWS Entity Resolution 数据匹配服务 OpenAPI 规范 .....	149
使用 OpenAPI 规范 .....	149
批处理集成 .....	150
同步处理集成 .....	152
测试提供商集成 .....	153
安全性 .....	161
数据保护 .....	161
静态数据加密 AWS Entity Resolution 数据匹配服务 .....	162
密钥管理 .....	163
AWS PrivateLink .....	173
Identity and access management .....	175
受众 .....	175
使用身份进行身份验证 .....	176
使用策略管理访问 .....	177
如何 AWS Entity Resolution 数据匹配服务 与 IAM 配合使用 .....	178
基于身份的策略示例 .....	183
AWS 托管策略 .....	185
问题排查 .....	188
合规性验证 .....	189
AWS Entity Resolution 数据匹配服务 合规性最佳实践 .....	190
恢复能力 .....	190
监控 .....	191
CloudTrail 日志 .....	191
AWS Entity Resolution 数据匹配服务 信息在 CloudTrail .....	191
了解 AWS Entity Resolution 数据匹配服务 日志文件条目 .....	192
CloudWatch 日志 .....	192
设置日志传输 .....	193
禁用日志记录 ( 控制台 ) .....	200
读取日志 .....	200
AWS CloudFormation 资源 .....	203

AWS 实体解析和 CloudFormation 模板 .....	203
了解更多关于 CloudFormation .....	204
配额 .....	206
API 节流配额 .....	209
文档历史记录 .....	213
术语表 .....	219
Amazon 资源名称 ( ARN ) .....	219
属性类型 .....	219
自动处理 .....	219
AWS KMS key 进行筛选 .....	219
批量工作流程 .....	219
明文 .....	219
置信度 (ConfidenceLevel) .....	220
解密 .....	220
加密 .....	220
组名 .....	220
哈希 .....	220
哈希协议 (HashingProtocol) .....	220
ID 映射方法 .....	220
ID 映射工作流程 .....	221
ID 命名空间 .....	221
增量工作流程 .....	221
输入字段 .....	221
输入源 ARN (AR InputSource N) .....	222
基于机器学习的匹配 .....	222
手动处理 .....	222
Many-to-Many 匹配 .....	222
比赛 ID (matchID) .....	222
匹配键 (MatchKey) .....	223
匹配键名称 .....	223
匹配规则 (MatchRule) .....	223
Matching .....	223
匹配工作流程 .....	223
匹配的工作流程描述 .....	224
匹配工作流程名称 .....	224
匹配工作流程元数据 .....	224

标准化 (ApplyNormalization) .....	224
Name .....	225
电子邮件 .....	225
Phone .....	226
地址 .....	226
经过哈希处理 .....	229
Source_ID .....	229
标准化 (ApplyNormalization)- ML-based 仅限 .....	229
Name .....	230
电子邮件 .....	230
Phone .....	230
One-to-One 匹配 .....	230
Output .....	231
outputs3Path .....	231
OutputSourceConfig .....	231
基于提供商服务的匹配 .....	231
Rule-based 匹配 .....	232
传递匹配 .....	232
架构 .....	232
架构描述 .....	233
架构名称 .....	233
架构映射 .....	233
架构映射 ARN .....	233
唯一 ID .....	233
.....	CCXXXV

# 什么是 AWS Entity Resolution 数据匹配服务？

AWS Entity Resolution 数据匹配服务 是一项服务，可帮助您匹配、链接和增强存储在多个应用程序、渠道和数据存储中的相关记录。您可以开始使用灵活且可扩展的实体解析工作流程，并且可以连接到您现有的应用程序和数据服务提供商。

AWS Entity Resolution 数据匹配服务 提供高级匹配技术，例如基于规则的匹配、基于机器学习的匹配（机器学习匹配）和数据服务提供商主导的匹配。这些技术可以帮助您更准确地关联和增强客户信息、产品代码或业务数据代码的相关记录。

通过将最近发生的事件（例如广告点击、购物车放弃和购买）与来自数据服务提供商的匿名信号关联到一个唯一的实体 ID，您可以使用 AWS Entity Resolution 数据匹配服务 创建统一的客户互动视图。您还可以更好地追踪商店中使用不同代码（例如 SKU、UPC）的商品。您可以使用 AWS Entity Resolution 数据匹配服务 来控制匹配精度，更好地保护数据安全，同时最大限度地减少数据移动。

## 主题

- [你是首次 AWS Entity Resolution 数据匹配服务 使用吗？](#)
- [的特点 AWS Entity Resolution 数据匹配服务](#)
- [相关服务](#)
- [正在访问 AWS Entity Resolution 数据匹配服务](#)
- [的定价 AWS Entity Resolution 数据匹配服务](#)

## 你是首次 AWS Entity Resolution 数据匹配服务 使用吗？

如果您是首次使用 AWS Entity Resolution 数据匹配服务，我们建议您先阅读以下章节：

- [的特点 AWS Entity Resolution 数据匹配服务](#)
- [正在访问 AWS Entity Resolution 数据匹配服务](#)
- [设置 AWS Entity Resolution 数据匹配服务](#)

## 的特点 AWS Entity Resolution 数据匹配服务

AWS Entity Resolution 数据匹配服务 包括以下功能：

- 灵活且可定制的数据准备

AWS Entity Resolution 数据匹配服务 从中读取您的数据 AWS Glue ，用作匹配处理的输入。您最多可以指定 20 个数据输入。AWS Entity Resolution 数据匹配服务 将数据输入表的每一行作为记录处理，并使用唯一的实体作为主键。AWS Entity Resolution 数据匹配服务 可以对加密的数据集进行操作。首先定义[架构映射](#) AWS Entity Resolution 数据匹配服务 ，以了解要在[匹配工作流程](#)中使用哪些输入字段。您可以从现有的数据输入中引入自己的 AWS Glue 数据架构或蓝图。或者，您可以使用交互式用户界面或 JSON 编辑器构建自定义架构。默认情况下，AWS Entity Resolution 数据匹配服务 还会在匹配之前[对数据输入进行标准化](#)以改进匹配处理，例如删除特殊字符和多余空格，以及将文本格式化为小写。如果您的数据输入已经过标准化，则可以关闭标准化。我们还提供一个[GitHub 库](#)，您可以使用该库进一步自定义数据标准化过程以满足您的需求。

- 可配置的实体匹配工作流程

实体[匹配工作流程](#)是您设置的一系列步骤，用于说明 AWS Entity Resolution 数据匹配服务 如何匹配数据输入以及将合并数据输出写入何处。您可以设置一个或多个匹配工作流程来比较不同的数据输入，并使用不同的匹配技术，例如[基于规则的匹配](#)、[机器学习匹配](#)或[数据服务提供商主导的匹配](#)，无需实体解析或机器学习经验。您还可以查看现有匹配工作流程和指标的任务状态，例如资源编号、已处理的记录数和找到的匹配项数。

- Ready-to-use 基于规则的匹配

这种匹配方法在 o AWS 管理控制台 r AWS Command Line Interface (AWS CLI) 中包含一组 ready-to-use 规则。您可以使用这些规则根据您的输入字段查找相关记录。您还可以通过添加或删除每条规则的输入字段、删除规则、重新排列规则优先级以及创建新规则来自定义规则。您也可以重置规则，使其恢复到其原始配置。您的亚马逊简单存储服务 (Amazon S3) Simple Service 存储桶中输出的数据包含使用基于[规则](#)的匹配技术生成的匹配组 AWS Entity Resolution 数据匹配服务 。每个匹配组都有用于生成与之关联的匹配项的规则编号，以帮助您了解匹配项。例如，规则编号可以证明每个匹配组的精度，从而使规则一比规则二更精确。

- 预先配置的基于机器学习的匹配 ( 机器学习匹配 )

这种匹配技术包括预先配置的机器学习模型，用于在所有数据输入中查找匹配项，尤其是基于消费者的记录。该模型使用与姓名、电子邮件地址、电话号码、地址和出生日期数据类型关联的所有输入字段。该模型生成由相关记录组成的匹配组，每个组中都有一个[置信度分数](#)，解释了与其他比赛组相比的比赛质量。该模型会考虑缺失的输入字段，并一起分析整个记录以表示实体。您的 Amazon S3 存储桶中的数据输出包含使用机器学习匹配 AWS Entity Resolution 数据匹配服务 生成的匹配组。在这里，每个匹配组的相关置信度分数为 0.0—1.0，这表示匹配的精度。

- 将记录与数据服务提供商进行匹配

借助，AWS Entity Resolution 数据匹配服务 您可以与领先的数据服务供应商和许可数据集进行匹配、关联和增强记录，从而扩大您了解、接触和服务客户的能力。例如，您可以为数据附加属性以增强记录，也可以提高所使用的系统和平台的互操作性以实现业务目标。您只需点击几下即可使用此匹配的工作流程，无需构建和维护复杂的专有集成。您必须与这些数据服务提供商签订许可协议才能利用这种匹配技术。

- 手动批量处理和自动增量处理

您可以使用数据处理来帮助将您的数据输入或输入转换为合并的数据输出表，该表包含使用实体匹配工作流程配置生成的公共匹配 ID 的类似记录。使用 API 和/或 AWS 管理控制台 或 AWS CLI，您可以根据现有提取、转换和加载 (ETL) 数据管道按需运行 [手动批量处理](#)，该管道会重新处理所有新匹配项的数据，并更新现有匹配项。此外，对于基于规则的匹配方案，您可以启动 [自动增量处理](#)，这样，只要您的 Amazon S3 存储桶中有新数据，该服务就会读取这些新记录并将其与现有记录进行比较。这样可以使您的匹配项与 Amazon S3 数据的任何变化保持同步。

- 近乎实时的查询

通过 [AWS Entity Resolution 数据匹配服务 GetMatchId API 操作](#) 查找任何实体字段可帮助您同步检索现有的匹配 ID。您可以使用通过不同来源 AWS Entity Resolution 数据匹配服务 和渠道获取的个人身份信息 (PII) 属性致电。AWS Entity Resolution 数据匹配服务 对这些属性进行哈希处理以保护数据，并检索相应的匹配 ID 以关联和匹配客户。例如，您可以使用关联的姓名、电子邮件和邮寄地址进行网络注册。使用 AWS Entity Resolution 数据匹配服务 GetMatchId API 操作查找存储在 S3 存储桶中的匹配结果中是否已存在该客户或实体，以及与之关联的相应实体匹配 ID。获得实体匹配 ID 后，您可以在源应用程序（例如客户关系管理 (CRM) 或客户数据平台 (CDP) 系统）中找到与之相关的交易信息。

- 通过设计实现数据保护和区域化

AWS Entity Resolution 数据匹配服务 提供默认加密功能，可帮助您保护数据，并为输入到服务的每个数据提供加密密钥。例如，AWS Entity Resolution 数据匹配服务 允许您灵活地使用服务器端加密和哈希处理的数据来运行基于规则的匹配工作流程。AWS Entity Resolution 数据匹配服务 支持区域化，这意味着您的匹配工作流程以与您使用服务相同的位置 AWS 区域 来处理数据。在其他应用程序中使用已解析的数据之前，您还可以对 Amazon S3 中的数据输出进行加密和哈希处理。

- 多方转码

AWS Entity Resolution 数据匹配服务 帮助您定义数据源，并在想要使用数据协作的多方之间进行匹配配置，例如在 AWS Clean Rooms。

## 相关服务

以下内容与 AWS 服务 以下内容有关 AWS Entity Resolution 数据匹配服务：

- Amazon S3

将您导入的数据存储 AWS Entity Resolution 数据匹配服务 在 Amazon S3 中。

有关更多信息，请参阅[什么是 Amazon S3？](#) 在《Amazon 简单存储服务用户指南》中。

- AWS Glue

根据您在 Amazon S3 中的数据创建 AWS Glue 表以供在中使用 AWS Entity Resolution 数据匹配服务。

有关更多信息，请参阅[什么是 AWS Glue？](#) 在《AWS Glue 开发人员指南》中。

- AWS CloudTrail

AWS Entity Resolution 数据匹配服务 与 CloudTrail 日志配合使用可增强对 AWS 服务 活动的分析。

有关更多信息，请参阅[使用记录 AWS Entity Resolution 数据匹配服务 API 调用 AWS CloudTrail。](#)

- CloudFormation

在中创建以下资源 CloudFormation：AWS::EntityResolution::MatchingWorkflow, AWS::EntityResolution::SchemaMapping, AWS::EntityResolution::IdMappingWorkflow, AWS::EntityResolution::IdNamespace 和 AWS::EntityResolution::PolicyStatement

有关更多信息，请参阅[使用创建 AWS 实体解析资源 AWS CloudFormation。](#)

## 正在访问 AWS Entity Resolution 数据匹配服务

您可以 AWS Entity Resolution 数据匹配服务 通过以下选项进行访问：

- 直接通过 AWS Entity Resolution 数据匹配服务 控制台，网址为<https://console.aws.amazon.com/entityresolution/>。
- 通过 AWS Entity Resolution 数据匹配服务 API 以编程方式进行。有关更多信息，请参阅[AWS Entity Resolution 数据匹配服务 API 参考](#)。

- 如果您计划在 AWS Lambda Runtime 中调用 AWS Entity Resolution 数据匹配服务 API，请创建自己的部署包并添加所需版本的 AWS SDK 库。有关更多信息，请参阅《AWS Lambda 开发人员指南》中的以下示例：
  - [使用 .zip 或 JAR 文件存档部署 Java Lambda 函数](#)
  - [使用 Python Lambda 函数的 .zip 文件存档](#)

## 的定价 AWS Entity Resolution 数据匹配服务

有关定价信息，请参阅 [AWS Entity Resolution 数据匹配服务 定价](#)。

# 设置 AWS Entity Resolution 数据匹配服务

在 AWS Entity Resolution 数据匹配服务 首次使用之前，请注册 AWS 并创建管理员用户来创建角色。

## 报名参加 AWS

如果您已经有 AWS 账户，请跳过此步骤。

如果您没有 AWS 账户，请完成以下步骤来创建一个。

### 报名参加 AWS 账户

1. 打开<https://portal.aws.amazon.com/billing/注册>。
2. 按照屏幕上的说明操作。

在注册时，将接到电话或收到短信，要求使用电话键盘输入一个验证码。

当您注册时 AWS 账户，就会创建AWS 账户根用户一个。根用户有权访问该账户中的所有 AWS 服务和资源。作为最佳安全实践，请为用户分配管理访问权限，并且只使用根用户来执行[需要根用户访问权限的任务](#)。

## 创建管理员用户

要创建管理员用户，请选择以下选项之一。

选择一种方法来管理您的管理员	目标	方式	您也可以
在 IAM Identity Center 中	使用短期凭证访问 AWS。 这符合安全最佳实操。有关最佳实践的信息，请参阅《IAM 用户指	有关说明，请参阅《AWS IAM Identity Center 用户指南》中的 <a href="#">入门</a> 。	通过在《AWS Command Line Interface 用户指南》 <a href="#">AWS IAM Identity Center中配置 AWS CLI 要使用的来配置编程访问权限</a> 。

选择一种方法来管理您的管理员	目标	方式	您也可以
(推荐)	南》中的 <a href="#">IAM 中的安全最佳实践</a> 。		
在 IAM 中 (不推荐使用)	使用长期凭证访问 AWS。	按照《IAM 用户指南》中的 <a href="#">创建用于紧急访问的 IAM 用户</a> 中的说明进行操作。	按照《IAM 用户指南》中的 <a href="#">管理 IAM 用户的访问密钥</a> ，配置编程式访问。

## 为控制台用户创建 IAM 角色

如果您使用的是 AWS Entity Resolution 数据匹配服务 控制台，请完成以下步骤。

### 创建 IAM 角色

1. 使用您的管理员账户登录 IAM 控制台 (<https://console.aws.amazon.com/iam/>)。
2. 在 Access management (访问管理) 下，请选择 Roles (角色)。

您可以使用角色创建短期证书，建议使用此方法来提高安全性。您也可以选择用户来创建长期凭证。

3. 选择创建角色。
4. 在创建角色向导中，对于“可信实体类型”，选择AWS 账户。
5. 保持“此帐户”选项处于选中状态，然后选择“下一步”。
6. 对于添加权限，请选择创建策略。

将打开一个新选项卡。

- a. 选择 JSON 选项卡，然后根据授予控制台用户的权限添加策略。AWS Entity Resolution 数据匹配服务 根据常见用例提供以下托管策略：

- [AWS 托管策略：AWSEntityResolutionConsoleFullAccess](#)

- [AWS 托管策略：AWSEntityResolutionConsoleReadOnlyAccess](#)

- 选择下一步: 标签，添加标签（可选），然后选择下一步: 审核。
- 对于查看策略，输入名称和描述，然后查看摘要。
- 选择创建策略。

您已经为协作成员创建了策略。

- 返回原始选项卡，在“添加权限”下，输入您刚刚创建的策略的名称。（您可能需要重新加载页面。）
  - 选中您创建的策略名称旁边的复选框，然后选择下一步。
- 对于命名、查看和创建，输入角色名称和描述。
    - 查看选择受信任的实体，输入将担任该角色的一个或多个人员的 AWS 账户（如有必要）。
    - 在添加权限中查看权限，并在必要时进行编辑。
    - 查看标签，并在必要时添加标签。
    - 选择创建角色。

## 为创建工作流程工作角色 AWS Entity Resolution 数据匹配服务

AWS Entity Resolution 数据匹配服务 使用工作流程作业角色来运行工作流程。如果您具有必要的 IAM 权限，则可以使用控制台创建此角色。如果您没有CreateRole权限，请让您的管理员创建该角色。

### 为创建工作流程工作角色 AWS Entity Resolution 数据匹配服务

- 使用您的管理员账户登录 IAM 控制台。<https://console.aws.amazon.com/iam/>
- 在 Access management（访问管理）下，请选择 Roles（角色）。

您可以使用角色创建短期证书，建议使用此方法来提高安全性。您也可以选择用户来创建长期凭证。

- 选择创建角色。
- 在创建角色向导中，对于可信实体类型，选择自定义信任策略。
- 将以下自定义信任策略复制粘贴到 JSON 编辑器中。

JSON

```
{
```

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

6. 选择下一步。
7. 对于添加权限，请选择创建策略。

此时会出现一个新选项卡。

- a. 将以下策略复制并粘贴到 JSON 编辑器中。

#### Note

以下示例策略支持读取相应数据资源（如 Amazon S3 和）所需的权限 AWS Glue。但是，您可能需要修改此策略，具体取决于您设置数据源的方式。您可以在支持的 AWS 商业分区中使用来自任何区域的 AWS Glue 资源和底层 Amazon S3 资源，它们不必与位于同一区域 AWS Entity Resolution 数据匹配服务。AWS Glue 如果您的数据源未加密或解密，则无需授予 AWS KMS 权限。

## JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "s3:GetObject",
```

```
        "s3:ListBucket",
        "s3:GetBucketLocation"
    ],
    "Resource": [
        "arn:aws:s3:::{{input-buckets}}",
        "arn:aws:s3:::{{input-buckets}}/*"
    ],
    "Condition": {
        "StringEquals": {
            "s3:ResourceAccount": [
                "444455556666"
            ]
        }
    }
},
{
    "Effect": "Allow",
    "Action": [
        "s3:PutObject",
        "s3:ListBucket",
        "s3:GetBucketLocation"
    ],
    "Resource": [
        "arn:aws:s3:::{{output-bucket}}",
        "arn:aws:s3:::{{output-bucket}}/*"
    ],
    "Condition": {
        "StringEquals": {
            "s3:ResourceAccount": [
                "444455556666"
            ]
        }
    }
},
{
    "Effect": "Allow",
    "Action": [
        "glue:GetDatabase",
        "glue:GetTable",
        "glue:GetPartition",
        "glue:GetPartitions",
        "glue:GetSchema",
        "glue:GetSchemaVersion",
        "glue:BatchGetPartition"
```

```

    ],
    "Resource": [
      "arn:aws:glue:us-east-1:444455556666:database/{{input-
databases}}",
      "arn:aws:glue:us-east-1:444455556666:table/{{input-
database}}/{{input-tables}}",
      "arn:aws:glue:us-east-1:444455556666:catalog"
    ]
  }
]
}

```

将每个 *{{user input placeholder}}* 替换为您自己的信息。

*aws-region*

AWS 区域 你的资源。您可以使用 AWS Glue Amazon S3 和任何支持 AWS 区域 这些服务的商用 AWS KMS 资源。

*&ExampleAWSAccountNo1;*

你的 AWS 账户 身份证。

*input-buckets*

Amazon S3 存储桶，其中包含 AWS Entity Resolution ##### 将从 AWS Glue 何处读取的底层数据对象。

*output-buckets*

Amazon S3 存储桶 AWS Entity Resolution ##### 将在其中生成输出数据。

*input-databases*

AWS Glue AWS Entity Resolution ##### 将从中读取的数据库。

- b. ( 可选 ) 如果输入的 Amazon S3 存储桶是使用客户的 KMS 密钥加密的，请添加以下内容：

```

{
  "Effect": "Allow",
  "Action": [
    "kms:Decrypt"
  ],
  "Resource": [

```

```

        "arn:aws:kms:{{aws-
region}}:{{&ExampleAWSAccountNo1;}}:key/{{inputKeys}}"
    ]
}

```

将每个 `{{user input placeholder}}` 替换为您自己的信息。

`aws-region`

AWS 区域 你的资源。您可以使用 AWS Glue Amazon S3 和任何支持 AWS 区域 这些服务的商用 AWS KMS 资源。

`&ExampleAWSAccountNo1;`

你的 AWS 账户 身份证。

`inputKeys`

中的托管密钥 AWS Key Management Service。如果您的输入源已加密，则 AWS Entity Resolution ##### 必须使用您的密钥解密数据。

- c. ( 可选 ) 如果写入输出 Amazon S3 存储桶的数据需要加密，请添加以下内容：

```

{
  "Effect": "Allow",
  "Action": [
    "kms:GenerateDataKey",
    "kms:Encrypt"
  ],
  "Resource": [
    "arn:aws:kms:{{aws-
region}}:{{&ExampleAWSAccountNo1;}}:key/{{outputKeys}}"
  ]
}

```

将每个 `{{user input placeholder}}` 替换为您自己的信息。

`aws-region`

AWS 区域 你的资源。您可以使用 AWS Glue Amazon S3 和任何支持 AWS 区域 这些服务的商用 AWS KMS 资源。

`&ExampleAWSAccountNo1;`

你的 AWS 账户 身份证。

*outputKeys*

中的托管密钥 AWS Key Management Service。如果您需要对输出源进行加密，则AWS Entity Resolution ##### #必须使用您的密钥对输出数据进行加密。

- d. (可选) 如果您通过订阅了提供者服务 AWS Data Exchange，并且想要将现有角色用于基于提供者服务的工作流程，请添加以下内容：

```
{
  "Effect": "Allow",
  "Sid": "DataExchangePermissions",
  "Action": "dataexchange:SendApiAsset",
  "Resource": [
    "arn:aws:dataexchange:{{aws-region}}::data-sets/{{datasetId}}/
revisions/{{revisionId}}/assets/{{assetId}}"
  ]
}
```

将每个 *{{user input placeholder}}* 替换为您自己的信息。

*aws-region*

授予提供者资源 AWS 区域 的地方。您可以在控制台的资产 ARN 中找到此值。AWS Data Exchange 例如：arn:aws:dataexchange:us-east-2::data-sets/111122223333/revisions/339ffc6444examplef3bc15cf0b2346b/assets/546468b8examplea37bfc73b8f79fefafa

*datasetId*

数据集的 ID，可在 AWS Data Exchange 控制台上找到。


*revisionId*

数据集的修订版，可在 AWS Data Exchange 控制台上找到。

*assetId*

资产的 ID，可在 AWS Data Exchange 控制台上找到。

8. 返回原始选项卡，在“添加权限”下，输入您刚刚创建的策略的名称。（您可能需要重新加载页面。）
9. 选中您创建的策略名称旁边的复选框，然后选择下一步。
10. 对于命名、查看和创建，输入角色名称和描述。

 Note

角色名称必须与授予成员的passRole权限模式相匹配，该成员可以传递权限workflow job role以创建匹配的工作流程。

例如，如果您使用的是AWSIdentityResolutionConsoleFullAccess托管策略，请记得在角色名称中加entityresolution入。

- a. 查看选择受信任的实体，并在必要时进行编辑。
- b. 在添加权限中查看权限，并在必要时进行编辑。
- c. 查看标签，并在必要时添加标签。
- d. 选择创建角色。

的工作流工作角色 AWS Entity Resolution 数据匹配服务 已创建。

## 准备输入数据表

在中 AWS Entity Resolution 数据匹配服务，您的每个输入数据表都包含源记录。这些记录包含消费者标识符，例如名字、姓氏、电子邮件地址或电话号码。这些源记录可以与您在相同或其他输入数据表中提供的其他源记录相匹配。每条记录都必须有一个唯一的记录 ID ([唯一 ID](#))，并且在其中创建架构映射时必须将其定义为主键 AWS Entity Resolution 数据匹配服务。

每个输入数据表都以由 Amazon S3 支持的 AWS Glue 表的形式提供。您可以使用已存在于 Amazon S3 中的第一方数据，也可以将来自其他第三方 SaaS 提供商的数据表导入 Amazon S3。将数据上传到 Amazon S3 后，您可以使用 AWS Glue 爬虫在中创建数据表。AWS Glue Data Catalog 然后，您可以使用该数据表作为输入 AWS Entity Resolution 数据匹配服务。

以下各节介绍如何准备第一方数据和第三方数据。

### 主题

- [准备第一方输入数据](#)
- [准备第三方输入数据](#)

## 准备第一方输入数据

[以下步骤描述了如何准备第一方数据，以用于基于规则的匹配工作流程、基于机器学习的匹配工作流程或身份映射工作流程。](#)

### 步骤 1：准备第一方数据表

每种匹配的工作流程类型都有一套不同的建议和指南，以帮助确保成功。

要准备第一方数据表，请查阅下表：

#### 第一方数据表指南

工作流程类型	必需
具有高级规则类型的基于规则的匹配工作流程	<ul style="list-style-type: none"> <li>• 需要一个<a href="#">唯一 ID</a>。</li> <li>• 唯一 ID 不能超过 38 个字符。</li> <li>• ( 可选 ) 一个 DELETE 列，用于指定在工作流程完成处理 AWS Entity Resolution 数据匹配服务 后要从哪些记录中删除。<i>false</i> 如果该列存在没有任何值，则默认值为。“删除”列设置为的记</li> </ul>

工作流程类型	必需
	<p>记录 <i>true</i> 将被删除。DELETE 列设置为 <i>false</i> 或为空的记录将由处理 AWS Entity Resolution 数据匹配服务。</p> <p>架构必须有一个 DELETE 列，其类型必须为 String no matchKey and groupName 。</p> <div data-bbox="574 464 1507 684" style="border: 1px solid #add8e6; border-radius: 10px; padding: 10px; margin: 10px 0;"> <p> <b>Note</b></p> <p>不支持查找匹配项 ID (GetMatchID )，因为手动处理节奏的高级规则类型不存储任何提取的数据。</p> </div> <p>在以下示例中，S1 将被摄取并 S2 删除。</p> <p>Example</p> <div data-bbox="574 905 1507 1066" style="border: 1px solid #add8e6; border-radius: 10px; padding: 10px; margin: 10px 0;"> <pre>sourceID, name, lastName, DELETE S1, name, lastname, false S2, name2, lastname2, true</pre> </div>
具有简单规则类型的基于规则的匹配工作流程	<ul style="list-style-type: none"> <li>• 需要一个 <a href="#">唯一 ID</a>。</li> <li>• 唯一 ID 不能超过 38 个字符。</li> </ul>
基于机器学习的匹配工作流程	<ul style="list-style-type: none"> <li>• 需要一个 <a href="#">唯一 ID</a>。</li> <li>• 该数据集包含以下类型之一： <ul style="list-style-type: none"> <li>• <b>Full Name</b></li> <li>• <b>Full Address</b></li> <li>• <b>Full phone</b></li> <li>• <b>Email address</b></li> <li>• <b>Date</b>— 匹配密钥名称为出生日期</li> </ul> </li> <li>• 所有列名均不使用以下保留名称：MatchId“”、MatchRule “、” RecordId、SourceId “、” 和 TargetId”。</li> </ul>

工作流程类型	必需
ID 映射工作流程	<ul style="list-style-type: none"> <li>• 需要一个<a href="#">唯一 ID</a>。</li> <li>• 唯一 ID 不能超过 257 个字符。</li> <li>• ( 可选 ) 一个 DELETE 列，用于指定在工作流程完成处理 AWS Entity Resolution 数据匹配服务 后要从哪些记录中删除。<i>false</i> 如果该列存在没有任何值，则默认值为。“删除”列设置为的记录 <i>true</i> 将被删除。DELETE 列设置为 <i>false</i> 或为空的记录将由处理 AWS Entity Resolution 数据匹配服务。</li> </ul> <p>架构必须有一个 DELETE 列，其类型必须为 String no matchKey and groupName 。</p> <p>在以下示例中，S1 将被摄取并 S2 删除。</p> <p>Example</p> <pre>sourceID, name, lastName, DELETE S1, name, lastname, false S2, name2, lastname2, true</pre>

## 步骤 2：以支持的数据格式保存输入数据表

如果您已经以支持的数据格式保存了第一方输入数据，则可以跳过此步骤。

要使用 AWS Entity Resolution 数据匹配服务，输入数据必须采用 AWS Entity Resolution 数据匹配服务支持的格式。

AWS Entity Resolution 数据匹配服务支持以下数据格式：

- 逗号分隔值 (CSV)
- Parquet

## 步骤 3：将您的输入数据表上传到 Amazon S3

如果您在 Amazon S3 中已经有了第一方数据表，则可以跳过此步骤。

**Note**

您可以将输入数据存储在支持 S3 的 AWS 商业分区中任何区域的 Amazon S3Resources 中。可以从其他地区访问这些数据，也可以在运行匹配工作流程 AWS 账户 时访问这些数据。

将您的输入数据表上传到 Amazon S3

1. 登录 AWS 管理控制台 并打开 Amazon S3 控制台，网址为<https://console.aws.amazon.com/s3/>。
2. 选择 Buckets，然后选择一个存储桶来存储您的数据表。
3. 选择上传，然后按照提示进行操作。
4. 选择对象选项卡，查看存储数据的前缀。记下文件夹的名称。

您可以选择要查看数据表的文件夹。

## 步骤 4：创建 AWS Glue 表

**Note**

如果您需要分区 AWS Glue 表，请跳至。[步骤 4：创建分区表 AWS Glue](#)

Amazon S3 中的输入数据必须编入目录 AWS Glue 并以 AWS Glue 表格形式表示。有关如何使用 Amazon S3 作为输入创建 AWS Glue 表的更多信息，请参阅[AWS Glue 开发者指南中的在 AWS Glue 控制台上使用爬虫](#)。

在此步骤中，您将在中设置一个爬虫 AWS Glue 来抓取 S3 存储桶中的所有文件并创建 AWS Glue 表。

**Note**

AWS Entity Resolution 数据匹配服务 目前不支持注册的 Amazon S3 地点 AWS Lake Formation。

## 创建 AWS Glue 表

1. 登录 AWS 管理控制台 并打开 AWS Glue 控制台，网址为 <https://console.aws.amazon.com/glue/>。
2. 从导航栏中，选择爬网程序。
3. 从列表中选择您的 S3 存储桶，然后选择创建抓取工具。
4. 在“设置 Crawler 属性”页上，输入 Crawler 名称（可选描述），然后选择“下一步”。
5. 继续浏览添加爬网程序页面，指定详细信息。
6. 在选择 IAM 角色页面上，选择选择现有 IAM 角色，然后选择下一步。

如果需要，您也可以选择创建 IAM 角色或让管理员创建 IAM 角色。

7. 对于为此爬网程序创建计划，请保留默认频率（按需运行），然后选择下一步。
8. 对于“配置 Crawler 的输出”，输入 AWS Glue 数据库，然后选择“下一步”。
9. 查看所有详细信息，然后选择“完成”。
10. 在爬网程序页面上，选中 S3 存储桶旁边的复选框，然后选择运行爬网程序。
11. 爬网程序运行完毕后，在 AWS Glue 导航栏上选择数据库，然后选择您的数据库名称。
12. 在数据库页面上，选择 {your database name} 中的表。
  - a. 查看 AWS Glue 数据库中的表。
  - b. 要查看表的架构，请选择一个特定的表。
  - c. 记下 AWS Glue 数据库名称和 AWS Glue 表名。

现在，您可以创建架构映射了。有关更多信息，请参阅 [创建架构映射](#)。

## 步骤 4：创建分区表 AWS Glue

### Note

中的 AWS Glue 分区功能 AWS Entity Resolution 数据匹配服务 仅在 ID 映射工作流程中受支持。此 AWS Glue 分区功能使您可以选择用于处理 AWS Entity Resolution 数据匹配服务的特定分区。

如果您不需要分区 AWS Glue 表，则可以跳过此步骤。

当您在数据结构中添加新文件夹（例如一个月以下的新日文件夹）时，分区 AWS Glue AWS Glue 表会自动反映表中的新分区。

在中创建分区 AWS Glue 表时 AWS Entity Resolution 数据匹配服务，可以指定要在 ID 映射工作流程中处理哪些分区。然后，每次运行 ID 映射工作流时，只处理这些分区中的数据，而不是处理整个 AWS Glue 表中的所有数据。此功能允许在中进行更精确、更高效、更具成本效益的数据处理 AWS Entity Resolution 数据匹配服务，从而让您在管理实体解析任务时拥有更大的控制权和灵活性。

您可以在 ID 映射工作流程中为源账户创建分区 AWS Glue 表。

您必须首先将 Amazon S3 中的输入数据编入目录，AWS Glue 并将其表示为 AWS Glue 表。有关如何使用 Amazon S3 作为输入创建 AWS Glue 表的更多信息，请参阅[AWS Glue 开发者指南中的在 AWS Glue 控制台上使用爬虫](#)。

在此步骤中，您将在中设置一个爬虫 AWS Glue 来抓取 S3 存储桶中的所有文件，然后创建分区 AWS Glue 表。

#### Note

AWS Entity Resolution 数据匹配服务 目前不支持注册的 Amazon S3 地点 AWS Lake Formation。

## 创建分区表 AWS Glue

1. 登录 AWS 管理控制台 并打开 AWS Glue 控制台，网址为<https://console.aws.amazon.com/glue/>。
2. 从导航栏中，选择爬网程序。
3. 从列表中选择您的 S3 存储桶，然后选择创建抓取工具。
4. 在“设置 Crawler 属性”页上，输入 Crawler 名称、可选描述，然后选择“下一步”。
5. 继续浏览添加爬网程序页面，指定详细信息。
6. 在选择 IAM 角色页面上，选择选择现有 IAM 角色，然后选择下一步。

如果需要，您也可以选择创建 IAM 角色或让管理员创建 IAM 角色。

7. 对于为此爬网程序创建计划，请保留默认频率（按需运行），然后选择下一步。
8. 对于“配置 Crawler 的输出”，输入 AWS Glue 数据库，然后选择“下一步”。
9. 查看所有详细信息，然后选择“完成”。
10. 在爬网程序页面上，选中 S3 存储桶旁边的复选框，然后选择运行爬网程序。

11. 爬网程序运行完毕后，在 AWS Glue 导航栏上选择数据库，然后选择您的数据库名称。
12. 在数据库页面的表下，选择要分区的表。
13. 在表格概述上，选择操作下拉列表，然后选择编辑表格。
  - a. 在表格属性下，选择添加。
  - b. 对于新密钥，请输入 `aerPushDownPredicateString`。
  - c. 对于新值，请输入 '`<PartitionKey>=<PartitionValue>`'。
  - d. 记下 AWS Glue 数据库名称和 AWS Glue 表名。

您现在已准备好执行以下操作：

- [创建架构映射](#)，然后为架构映射[创建 ID 映射工作流程 AWS 账户](#)。
- [创建 ID 命名空间源](#)，[创建 ID 命名空间目标](#)，然后[跨两个命名空间创建 ID 映射工作流程 AWS 账户](#)。

## 准备第三方输入数据

第三方数据服务提供的标识符可以与您的已知标识符相匹配。

AWS Entity Resolution 数据匹配服务 目前支持以下第三方数据提供商服务：

### 数据提供商服务

公司名	可用 AWS 区域	标识符
LiveRamp	美国东部（弗吉尼亚北部）（us-east-1）、美国东部（俄亥俄州）（us-east-2）和美国西部（俄勒冈）（us-west-2）	坡道 ID
TransUnion	美国东部（弗吉尼亚北部）（us-east-1）、美国东部（俄亥俄州）（us-east-2）和美国西部（俄勒冈）（us-west-2）	TransUnion 个人和家庭 IDs
统一身份证 2.0	美国东部（弗吉尼亚北部）（us-east-1）、美国东部（俄	未处理的 UID 2

公司名	可用 AWS 区域	标识符
	亥俄州 ) ( us-east-2 ) 和美国西部 ( 俄勒冈 ) ( us-west-2 )	

以下步骤介绍如何准备第三方数据，以使用基于[提供商服务的匹配工作流程](#)或基于[提供商服务的身份映射](#)工作流程。

## 主题

- [步骤 1：在上订阅提供商服务 AWS Data Exchange](#)
- [步骤 2：准备第三方数据表](#)
- [步骤 3：以支持的数据格式保存输入数据表](#)
- [步骤 4：将您的输入数据表上传到 Amazon S3](#)
- [步骤 5：创建 AWS Glue 表](#)

## 步骤 1：在上订阅提供商服务 AWS Data Exchange

如果您通过订阅了提供商服务 AWS Data Exchange，则可以使用以下提供商服务之一运行匹配的工作流程，将您的已知标识符与您的首选提供商进行匹配。您的数据将与您的首选提供商定义的一组输入相匹配。

### 要在上订阅提供商服务 AWS Data Exchange

1. 在上查看提供商列表 AWS Data Exchange。以下提供商列表可用：
  - LiveRamp
    - [LiveRamp 身份解析](#)
    - [LiveRamp 转码](#)
  - TransUnion
    - TruAudience 身份解析和充实
  - 统一身份证 2.0
    - [统一 ID 2.0 身份解析](#)
2. 根据您的报价类型，完成以下步骤之一。

- 私人报价 — 如果您与提供商存在关系，请按照《AWS Data Exchange 用户指南》中的“[私人产品和报价](#)”程序接受私人报价 AWS Data Exchange。
  - 自带订阅 — 如果您已经向提供商订阅了现有的数据，请按照AWS Data Exchange 用户指南中的[自带订阅 \(BYOS\) 优惠](#)程序接受自带订阅 (BYOS) 优惠。 AWS Data Exchange
3. 在上订阅提供者服务后 AWS Data Exchange，即可使用该提供商服务创建匹配的工作流程或 ID 映射工作流程。

有关如何访问包含以下内容的提供商产品的更多信息 APIs，请参阅AWS Data Exchange 用户指南中的[访问 API 产品](#)。

## 步骤 2：准备第三方数据表

每种第三方服务都有一套不同的建议和指南，以帮助确保成功的匹配工作流程。

要准备第三方数据表，请参阅下表：

### 数据提供商服务指南

提供者服务	需要唯一的身份证吗？	操作
LiveRamp	是	<p>确保进行以下配置：</p> <ul style="list-style-type: none"> <li>• <a href="#">唯一 ID</a> 可以是您自己的匿名标识符，也可以是行 ID。</li> <li>• 您的数据输入文件格式和标准化符合 LiveRamp 指导方针。</li> </ul> <p>有关匹配工作流程的输入文件格式指南的更多信息，请参阅 LiveRamp 文档中的<a href="#">通过 ADX 执行身份解析</a>。</p> <p>有关 ID 映射工作流程的输入文件格式指南的更多信息，请参阅文档中的<a href="#">通过 ADX 执行转码</a>。 LiveRamp</p>
TransUnion	是	<p>确保以下是输入视图中的 string 类型列：</p> <ul style="list-style-type: none"> <li>• <a href="#">唯一的 ID</a> 是必需的，可以是 CRM ID、联系人 ID、用户 ID 或任何唯一 ID。</li> </ul>

提供者服务	需要唯一的身份证吗？	操作
		<ul style="list-style-type: none"> <li>• <b>Name</b> <ul style="list-style-type: none"> <li>• <b>First Name</b>可以是小写或大写，支持昵称，但应排除标题和后缀。</li> <li>• <b>Last Name</b>可以是小写或大写，中间的首字母可以排除在外。</li> </ul> </li> <li>• <b>Address</b> <ul style="list-style-type: none"> <li>• <b>Street address1</b>和<b>Street address1</b>合并为<b>Full address</b>行（如果存在）。</li> <li>• <b>City</b>已与<b>Full address</b>分开。</li> <li>• <b>Zip</b>（或<b>zip plus4</b>），不含任何特殊字符，例如空格、连字符或空白。如果没有数据，则使用空值。</li> <li>• <b>State</b>指定为大写的 2 个字母的代码。</li> </ul> </li> <li>• <b>Phone</b> <ul style="list-style-type: none"> <li>• <b>Phone number</b>应为 10 位数字，不含任何特殊字符，例如空格或连字符。</li> </ul> </li> <li>• <b>Email addresses</b> 要么是纯文本，要么是 SHA256 经过哈希处理的小写字符串。</li> <li>• <b>Date of Birth</b>采用 y yy-yy-mm-dd 格式。</li> <li>• <b>Digital identifiers</b>（设备 IDs）可以包含 IDs 带连字符（长度为 36 个字符的原始设备 IDs/MAIDs/IFAs）和不带连字符（32 和 40 个字符长的哈希设备//）。IDs MAIDs IFAs <ul style="list-style-type: none"> <li>• <b>IPV4</b>是一个 32 位 IP 地址，用点分十进制记法表示。例如：192.0.2.1</li> <li>• <b>IPV6</b>是一个 128 位 IP 地址，用十六进制表示法，用冒号分隔。例如：2001:db8:0000:0000:0000:0000:0000:0001</li> </ul> </li> </ul>

提供者服务	需要唯一的身份证吗？	操作
		<ul style="list-style-type: none"><li>• <b>MAID</b> ( 移动广告 ID ) 是分配给移动设备用于广告目的的唯一字母数字字符串。女佣通常有 36 个字符。例如 : a1b2c3d4-5678-90ab-cdef-EXAMPLE11111</li></ul>

提供者服务	需要唯一的身份证吗？	操作
统一身份证 2.0	是	<p>确保进行以下配置：</p> <ul style="list-style-type: none"> <li>• <a href="#">唯一 ID</a> 不能是哈希。</li> <li>• 架构中使用了 <b>Phone number</b> 或 <b>Email addresses</b> ，而不是两者兼而有之。</li> <li>• UID2 支持 UID2 生成电子邮件和电话号码。但是，如果两个值都存在于架构映射中，则工作流会复制输出中的每条记录。一条记录使用电子邮件生 UID2 成，第二条记录使用电话号码。如果您的数据混合包含电子邮件和电话号码，并且您不希望在输出中出现这种重复的记录，那么最好的方法是为每个数据创建一个单独的工作流程，并使用不同的架构映射。在这种情况下，请执行两次步骤——为电子邮件创建一个工作流程，为电话号码创建一个单独的工作流程。</li> </ul> <div data-bbox="852 1056 1507 1801" style="border: 1px solid #add8e6; border-radius: 15px; padding: 15px; margin-top: 20px;"> <p> <b>Note</b></p> <p>无论是谁提出请求，特定的电子邮件或电话号码在任何特定时间都会产生相同的原始 UID2 价值。</p> <p>生盐 UID2s 是通过添加盐桶中的盐来制成的，这些盐桶大约每年轮换一次，这样生的盐也会 UID2 随之旋转。不同的盐桶在一年中的不同时间轮换。</p> <p>AWS Entity Resolution 数据匹配服务目前无法跟踪旋转盐桶和未加工盐桶的情况 UID2s，因此建议你每天重新生成未加工 UID2s 的盐桶。有关更多信息，请参阅<a href="#">增量更新 UID2s 应多久刷新一次</a>？在 UID 2.0 文档中。</p> </div>

## 步骤 3：以支持的数据格式保存输入数据表

如果您已经以支持的数据格式保存了第三方输入数据，则可以跳过此步骤。

要使用 AWS Entity Resolution 数据匹配服务，输入数据必须采用 AWS Entity Resolution 数据匹配服务支持的格式。

AWS Entity Resolution 数据匹配服务支持以下数据格式：

- 逗号分隔值 (CSV)

### Note

LiveRamp 仅支持 CSV 文件。

- Parquet

## 步骤 4：将您的输入数据表上传到 Amazon S3

如果您在 Amazon S3 中已有第三方数据表，则可以跳过此步骤。

### Note

您可以将输入数据存储在不支持 S3 的 AWS 商业分区中任何区域的 Amazon S3 资源中。可以从其他地区访问这些数据，也可以在运行匹配工作流程 AWS 账户时访问这些数据。

将您的输入数据表上传到 Amazon S3

1. 登录 AWS 管理控制台并打开 Amazon S3 控制台，网址为 <https://console.aws.amazon.com/s3/>。
2. 选择 Buckets，然后选择一个存储桶来存储您的数据表。
3. 选择上传，然后按照提示进行操作。
4. 选择对象选项卡，查看存储数据的前缀。记下文件夹的名称。

您可以选择要查看数据表的文件夹。

## 步骤 5：创建 AWS Glue 表

Amazon S3 中的输入数据必须编入目录 AWS Glue 并以 AWS Glue 表格形式表示。有关如何使用 Amazon S3 作为输入创建 AWS Glue 表的更多信息，请参阅[AWS Glue 开发者指南中的在 AWS Glue 控制台上使用爬虫](#)。

### Note

AWS Entity Resolution 数据匹配服务 不支持分区表。

在此步骤中，您将在中设置一个爬虫 AWS Glue 来抓取 S3 存储桶中的所有文件并创建 AWS Glue 表。

### Note

AWS Entity Resolution 数据匹配服务 目前不支持注册的 Amazon S3 地点 AWS Lake Formation。

### 创建 AWS Glue 表

1. 登录 AWS 管理控制台 并打开 AWS Glue 控制台，网址为<https://console.aws.amazon.com/glue/>。
2. 从导航栏中，选择爬网程序。
3. 从列表中选择您的 S3 存储桶，然后选择添加爬网程序。
4. 在添加爬网程序页面上，输入爬网程序名称，然后选择下一步。
5. 继续浏览添加爬网程序页面，指定详细信息。
6. 在选择 IAM 角色页面上，选择选择现有 IAM 角色，然后选择下一步。

如果需要，您也可以选择创建 IAM 角色或让管理员创建 IAM 角色。

7. 对于为此爬网程序创建计划，请保留默认频率（按需运行），然后选择下一步。
8. 对于“配置 Crawler 的输出”，输入 AWS Glue 数据库，然后选择“下一步”。
9. 检查所有详细信息，然后选择完成。
10. 在爬网程序页面上，选中 S3 存储桶旁边的复选框，然后选择运行爬网程序。
11. 爬网程序运行完毕后，在 AWS Glue 导航栏上选择数据库，然后选择您的数据库名称。

12. 在数据库页面上，选择 {your database name} 中的表。
  - a. 查看 AWS Glue 数据库中的表。
  - b. 要查看表的架构，请选择一个特定的表。
  - c. 记下 AWS Glue 数据库名称和 AWS Glue 表名。

现在，您可以创建架构映射了。有关更多信息，请参阅 [创建架构映射](#)。

## 使用架构映射定义输入数据

架构映射定义了您要解析的输入数据。它还提供有关输入数据的元数据，例如列的属性类型（输入字段）以及要匹配的列。

创建架构映射时，首先要定义输入字段和属性类型，然后定义匹配键和组相关数据。下图总结了如何创建架构映射。



### Define your data

Import columns from an AWS Glue table, build a custom schema, or use a JSON editor.



### Select input types

Assign a pre-defined input type for each input field to classify your data.



### Assign match keys

Define a match key for each input field to enable comparison for your matching workflow.



### Create data groups

Group related data that is separated into two or more input fields.

在创建架构映射之前，必须先设置 [AWS Entity Resolution 数据匹配服务](#) 和准备数据表。有关更多信息，请参阅[设置 AWS Entity Resolution 数据匹配服务](#)和[准备输入数据表](#)。

创建架构映射后，您可以执行以下操作之一：

- [创建匹配的工作流程](#)以查找不同数据输入之间的匹配项。
- [创建 ID 命名空间源](#)，您可以在 ID 映射工作流程中使用该源将数据转换为目标。
- AWS 账户使用您的架构@@ [映射作为源，在同一工作流程中创建 ID 映射](#) 映射工作流程。

### 主题


- [创建架构映射](#)
- [克隆架构映射](#)
- [编辑架构映射](#)
- [删除架构映射](#)

## 创建架构映射

此过程描述了使用[AWS Entity Resolution 数据匹配服务 控制台](#)创建架构映射的过程。

有三种方法可以创建架构映射：

- 使用“导入自 AWS Glue”选项导入现有输入数据-使用此创建方法使用引导流程从 AWS Glue 表中预填充的列开始定义输入字段。
- 使用“构建自定义架构”选项手动定义输入数据-使用此创建方法通过引导流程手动定义输入字段。
- 使用“使用 JSON 编辑器”选项手动创建-使用 JSON 编辑器手动创建、使用示例或导入现有输入数据。

 Note

此选项不可用“唯一 ID”和“输入”字段。

## Import from AWS Glue

通过从中导入现有输入数据来创建架构映射 AWS Glue

1. 登录 AWS 管理控制台 并打开 AWS Entity Resolution 数据匹配服务 控制台，网址为<https://console.aws.amazon.com/entityresolution/>。
2. 在左侧导航窗格的“数据准备”下，选择“架构映射”。
3. 在架构映射页面的右上角，选择创建架构映射。
4. 对于步骤 1：指定架构详细信息，请执行以下操作：
  - a. 在名称和创建方法中，输入架构映射名称和可选的描述。
  - b. 在“创建方法”中，选择“从中导入”AWS Glue。
  - c. 选择 AWS 区域。
  - d. 选择AWS Glue 数据库。
  - e. 选择AWS Glue 桌子。

要创建新表，请转到 AWS Glue 控制台<https://console.aws.amazon.com/glue/>。有关更多信息，请参阅《AWS Glue 用户指南》中的[AWS Glue 表格](#)。

- f. 对于唯一 ID，请指定明确引用每行数据的列。

### Example

例如，**Primary\_key**、**Row\_ID** 或 **Record\_ID**。

**Note**

“唯一 ID”列为必填字段。唯一 ID 必须是单个表中的唯一标识符。但是，在不同的表中，唯一 ID 可能有重复的值。如果未指定唯一 ID、在同一个源中不是唯一的，或者不同源的属性名称重叠，则在运行匹配的工作流程时会 AWS Entity Resolution 数据匹配服务 拒绝该记录。如果您在基于规则的匹配工作流程中使用此架构映射，则唯一 ID 不得超过 38 个字符。

- g. 对于输入字段，选择要用于匹配和可选传递的列。

您最多可以选择 34 列进行匹配和直通。

- i. 在“匹配”下，选择要用作匹配输入字段的列。

您最多可以选择 24 列进行匹配。

- ii. 如果要指定不用于匹配的列，请选择“添加直通列”。
- iii. （可选）在“直通”下，选择要包含为直通列的列。

**Note**

运行基于机器学习的匹配工作流程时，请勿在数据中使用以下任何保留名称作为列名：MatchId“”、“”、”、MatchRule SourceId “RecordId、” 和TargetId”。使用这些保留名称中的任何一个都将导致命名冲突和基于 ML 的匹配工作流程失败。

- h. （可选）如果要为资源启用标签，请选择添加新标签，然后输入密钥和值对。
- i. 选择下一步。

5. 对于“步骤 2：映射输入字段”，定义要用于匹配和可选传递的输入字段。

- a. 对于要匹配的输入字段，对于每个输入字段，

- 指定属性类型以对数据进行分类。
- 指定 Match 键名称以启用与匹配工作流程的输入字段比较。默认情况下，某些匹配键名称会自动与特定的属性类型相关联。
- 如果该输入字段的列值经过哈希处理，请选中“哈希”复选框；如果该值为明文，则将该复选框留空。

**Note**

如果您要创建架构映射以与基于 LiveRamp 提供者服务的匹配技术一起使用，则可以：

- 将提供商 ID 的属性类型指定为 LiveRamp ID。
- 将姓名字段的属性类型指定为多个字段（例如名字、姓氏）或一个字段。
- 将街道地址字段的属性类型指定为多个字段（例如街道地址 1、街道地址 2）或一个字段（完整地址）。

如果与地址匹配，则需要邮政编码（邮政编码）。

- 如果您在姓名中包含电子邮件（电子邮件地址）或电话（电话号码），则这些字段可以与街道地址匹配。

**Note**

如果您要创建架构映射以与基于 TransUnion 提供者服务的匹配技术一起使用，则可以指定以下任何属性类型：

- 全名、名字、姓氏
- 完整地址、街道地址 1、城市、州、国家、邮政编码
- Phone number（电话号码）
- 电子邮件地址
- 日期
- 数字标识符：IPV4IPV6、或 MA ID

**Note**

如果您要创建架构映射以用于基于机器学习的匹配工作流程，则您的数据集必须至少包含以下属性类型之一：

- 全名
- 完整地址

- 手机已满
- 电子邮件地址
- 带有匹配键名称的日期为出生日期

请勿将其中任何属性的属性类型指定为自定义字符串。

- b. (可选) 对于直通输入字段，添加不匹配的输入字段及其相应的哈希状态。

哈希状态表示该输入字段的列值是经过哈希处理还是明文。

- c. 选择下一步。

6. 对于“步骤 3：分组数据”，如果姓名、地址和电话号码输入字段已分成多个字段，则可以将它们分组。

此步骤将相关的输入字段连接成一个字段，这样您就可以在匹配的工作流程中将它们作为一个字段进行比较。

如果您没有任何数据映射到“姓名”、“地址”或“电话号码”输入字段，则此部分将为空。

如果您有更多类型的数据，也可以添加更多组。

- a. 如果要对名称输入数据进行分组，请执行以下操作：

在“全名”中，选择两个或更多要分组的输入字段。

群组名称和匹配键会自动与数据类型关联。

您可以更新群组名称，使用自定义匹配键的匹配键最多可包含 255 个字符，包括字母、数字、下划线 ( \_ ) 或连字符 ( - )。

选择“添加群组”以添加另一个群组。

**Note**

仅支持对全名进行标准化。

如果要标准化全名字类型，请将以下子类型分配给全名组：名字、中间名和姓氏。


- b. 如果要对地址输入数据进行分组，请执行以下操作：

对于完整地址，请选择两个或更多要分组的输入字段。

群组名称和匹配键。自动与数据类型关联。

您可以更新群组名称，使用自定义匹配键的匹配键最多可包含 255 个字符，包括字母、数字、下划线 ( \_ ) 或连字符 ( - )。

选择“添加群组”以添加另一个群组。

 Note

仅支持完整地址的标准化。

如果要标准化完整地址子类型，请将以下子类型分配给完整地址组：街道地址 1、街道地址 2：街道地址 3 名称、城市名称、州、国家/地区和邮政编码。


- c. 如果要对电话输入数据进行分组，请执行以下操作：

对于 Full phone，请选择两个或更多要分组的输入字段。

群组名称和匹配键。自动与数据类型关联。

您可以更新群组名称，使用自定义匹配键的匹配键最多可包含 255 个字符，包括字母、数字、下划线 ( \_ ) 或连字符 ( - )。

选择“添加群组”以添加另一个群组。

 Note

只有完整版手机支持标准化。

如果要标准化完整电话子类型，请将以下子类型分配给完整电话组：电话号码和电话国家/地区代码。

- d. 选择下一步。

7. 对于“步骤 4：查看并创建”，请执行以下操作：

- a. 查看您在之前的步骤中所做的选择，并在必要时进行编辑。
- b. 选择“创建架构映射”。

**Note**

将架构映射与工作流程关联后，您无法对其进行修改。如果要使用现有配置来创建新的架构映射，则可以克隆架构映射。

创建架构映射后，就可以[创建匹配的工作流程](#)或[创建 ID 命名空间](#)了。

## Build custom schema

使用“构建自定义架构”选项创建架构映射

1. 登录 AWS 管理控制台 并打开 AWS Entity Resolution 数据匹配服务 控制台，网址为<https://console.aws.amazon.com/entityresolution/>。
2. 在左侧导航窗格的“数据准备”下，选择“架构映射”。
3. 在架构映射页面的右上角，选择创建架构映射。
4. 对于步骤 1：指定架构详细信息，请执行以下操作：
  - a. 在名称和创建方法中，输入架构映射名称和可选的描述。
  - b. 在“创建方法”中，选择“生成自定义架构”。
  - c. 在“唯一 ID”中，输入唯一的 ID 以识别您的每一行数据。

### Example

例如，**Primary\_key**、**Row\_ID** 或 **Record\_ID**。


**Note**

“唯一 ID”列为必填字段。唯一 ID 必须是单个表中的唯一标识符。但是，在不同的表中，唯一 ID 可能有重复的值。如果未指定唯一 ID、在同一个源中不是唯一的，或者不同源的属性名称重叠，则在运行匹配的工作流程时会 AWS Entity Resolution 数据匹配服务 拒绝该记录。如果您在基于规则的匹配工作流程中使用此架构映射，则唯一 ID 不得超过 38 个字符。

- d. （可选）如果要为资源启用标签，请选择添加新标签，然后输入密钥和值对。
  - e. 选择下一步。
5. 对于“步骤 2：映射输入字段”，定义要用于匹配和可选传递的输入字段。


您最多可以为匹配和传递定义总共 34 列。

- a. 对于要匹配的输入字段，请输入输入字段。


 Note

运行基于机器学习的匹配工作流程时，请勿在数据中使用以下任何保留名称作为列名：MatchId“”、“”、”、MatchRule SourceId “RecordId、” 和TargetId”。使用这些保留名称中的任何一个都将导致命名冲突和基于 ML 的匹配工作流程失败。

- b. 选择属性类型对数据进行分类。

 Note

如果您要创建架构映射以用于[基于LiveRamp 提供商服务的匹配技术](#)，则可以将 ProviderID 属性类型指定为 ID。LiveRamp 如果要在输出中包含 PII 数据，则必须将属性类型指定为自定义字符串。

 Note

如果您要创建架构映射以与基于 TransUnion 提供者服务的匹配技术一起使用，则可以指定以下任何属性类型：

- 全名、名字、姓氏
- 完整地址、街道地址 1、城市、州、国家、邮政编码
- Phone number ( 电话号码 )
- 电子邮件地址
- 日期
- 数字标识符：IPV4IPV6、或 MA ID

**Note**

如果您要创建架构映射以用于[基于机器学习的匹配工作流程](#)，则您的数据集必须至少包含以下属性类型之一：

- 全名
- 完整地址
- 手机已满
- 电子邮件地址
- 带有匹配键名称的日期为出生日期

请勿将其中任何属性的属性类型指定为自定义字符串。

- c. 选择 Match 键名称以启用与匹配工作流程的输入字段比较。

默认情况下，某些匹配键名称会自动与特定的属性类型相关联。

- d. 如果该输入字段的列值经过哈希处理，请选中“哈希”复选框；如果该值为明文，则将该复选框留空。
- e. 选择“添加输入字段”以添加更多输入字段。

您最多可以添加 24 个输入字段进行匹配。

- f. ( 可选 ) 对于直通输入字段，添加不匹配的输入字段及其相应的哈希状态。
- g. 选择下一步。

6. 对于“步骤 3：分组数据”，如果姓名、地址、电话号码输入字段已分成多个字段，则可以将它们分组。

此步骤将相关的输入字段连接成一个字段，这样您就可以在匹配的工作流程中将它们作为一个字段进行比较。

如果您没有任何数据映射到“姓名”、“地址”、“电话号码”输入字段，则此部分将为空。

如果您有更多类型的数据，也可以添加更多组。


- a. 如果要对名称输入数据进行分组，请执行以下操作：

在“全名”中，选择两个或更多要分组的输入字段。

群组名称和匹配键会自动与数据类型关联。

您可以更新群组名称，使用自定义匹配键的匹配键最多可包含 255 个字符，包括字母、数字、下划线 (\_) 或连字符 (-)。

选择“添加群组”以添加另一个群组。

 Note

仅支持对全名进行标准化。

如果要标准化全名字类型，请将以下子类型分配给全名组：名字、中间名和姓氏。


- b. 如果要对地址输入数据进行分组，请执行以下操作：

对于完整地址，请选择两个或更多要分组的输入字段。

群组名称和匹配键。自动与数据类型关联。

您可以更新群组名称，使用自定义匹配键的匹配键最多可包含 255 个字符，包括字母、数字、下划线 (\_) 或连字符 (-)。

选择“添加群组”以添加另一个群组。

 Note

仅支持完整地址的标准化。

如果要标准化完整地址子类型，请将以下子类型分配给完整地址组：街道地址 1、街道地址 2：街道地址 3 名称、城市名称、州、国家/地区和邮政编码。

- c. 如果要对电话输入数据进行分组，请执行以下操作：

对于 Full phone，请选择两个或更多要分组的输入字段。

群组名称和匹配键。自动与数据类型关联。

您可以更新群组名称，使用自定义匹配键的匹配键最多可包含 255 个字符，包括字母、数字、下划线 (\_) 或连字符 (-)。

选择“添加群组”以添加另一个群组。

**Note**

只有完整版手机支持标准化。

如果要标准化完整电话子类型，请将以下子类型分配给完整电话组：电话号码和电话国家/地区代码。

- d. 选择下一步。
7. 对于“步骤 4：查看并创建”，请执行以下操作：
    - a. 查看您在之前的步骤中所做的选择，并在必要时进行编辑。
    - b. 选择“创建架构映射”。

**Note**

将架构映射与工作流程关联后，您无法对其进行修改。如果要使用现有配置来创建新的架构映射，则可以克隆架构映射。

创建架构映射后，就可以[创建匹配的工作流程](#)或[创建 ID 命名空间](#)了。

## Use JSON editor

### 使用 JSON 编辑器创建架构映射

1. 登录 AWS 管理控制台 并打开 AWS Entity Resolution 数据匹配服务 控制台，网址为<https://console.aws.amazon.com/entityresolution/>。
2. 在左侧导航窗格的“数据准备”下，选择“架构映射”。
3. 在架构映射页面的右上角，选择创建架构映射。
4. 对于步骤 1：指定架构详细信息，请执行以下操作：
  - a. 在名称和创建方法中，输入架构映射名称和可选的描述。
  - b. 在“创建方法”中，选择“使用 JSON 编辑器”。
  - c. （可选）如果要为资源启用标签，请选择添加新标签，然后输入密钥和值对。
  - d. 选择下一步。
5. 对于步骤 2：指定映射：
  - a. 开始在 JSON 编辑器中构建架构，或者根据您的目标选择以下选项之一：

您的目标	建议的选项
开始构建架构映射	插入示例 JSON，然后根据需要编辑信息。
使用现有的 JSON 文件	从文件导入

**Note**

仅以下类型支持标准化：NAME、ADDRESSPHONE、和EMAIL\_ADRESS。

如果要对子类型进行标准化处理，请将以下NAME子类型分配给 NAME group  
pName：、和 NAME\_FIRST NAME\_MIDDLE NAME\_LAST

如果要对子类型进行标准化处理，请将以下ADDRESS子类型分配给 ADDRESS

groupNameADDRESS\_STREET1：ADDRESS\_STREET2、、、、、ADDRESS\_STREET3和  
ADDRESS\_STATE ADDRESS\_COUNTRY ADDRESS\_POSTALCODE

如果要对子类型进行标准化处理，请将以下PHONE子类型分配给 groupName：  
和。PHONE PHONE\_NUMBER PHONE\_COUNTRYCODE

- b. 选择下一步。
6. 对于步骤 3：查看并创建：
    - a. 查看您在之前的步骤中所做的选择，并在必要时进行编辑。
    - b. 选择“创建架构映射”。

**Note**

将架构映射与工作流程关联后，您无法对其进行修改。如果要使用现有配置来创建新的架构映射，则可以克隆架构映射。

创建架构映射后，就可以[创建匹配的工作流程](#)或[创建 ID 命名空间](#)了。

## 克隆架构映射

如果要使用现有配置来创建新的架构映射，则可以克隆架构映射。

要克隆架构映射，请执行以下操作：

1. 登录 AWS 管理控制台 并打开 AWS Entity Resolution 数据匹配服务 控制台，网址为 <https://console.aws.amazon.com/entityresolution/>。
2. 在左侧导航窗格的“数据准备”下，选择“架构映射”。
3. 选择架构映射。
4. 选择克隆。
5. 在“指定架构详细信息”页面上，进行任何必要的更改，然后选择“下一步”。
6. 在“选择匹配技术”页面上，进行必要的更改，然后选择“下一步”。
7. 在地图输入字段页面上，进行任何必要的更改，然后选择下一步。
8. 在“分组数据”页面上，进行必要的更改，然后选择“下一步”。
9. 在“查看并保存”页面上，进行任何必要的更改，然后选择“克隆架构映射”。

## 编辑架构映射

在将架构映射关联到工作流程之前，您只能对其进行编辑。将架构映射与工作流程关联后，便无法对其进行编辑。如果要使用现有配置来创建新的架构映射，则可以克隆架构映射。

要编辑架构映射，请执行以下操作：

1. 登录 AWS 管理控制台 并打开 AWS Entity Resolution 数据匹配服务 控制台，网址为 <https://console.aws.amazon.com/entityresolution/>。
2. 在左侧导航窗格的“数据准备”下，选择“架构映射”。
3. 选择架构映射。
4. 选择编辑。
5. 在“指定架构详细信息”页面上，进行任何必要的更改，然后选择“下一步”。
6. 在“选择匹配技术”页面上，进行必要的更改，然后选择“下一步”。
7. 在地图输入字段页面上，进行任何必要的更改，然后选择下一步。
8. 在“分组数据”页面上，进行必要的更改，然后选择“下一步”。

### Note

只有全名、完整地址、完整电话和电子邮件地址支持标准化。  
如果要标准化全名子类型，请将以下子类型分配给全名组：名字、中间名和姓氏。

如果要标准化完整地址子类型，请将以下子类型分配给完整地址组：街道地址 1、街道地址 2：街道地址 3 名称、城市名称、州、国家/地区和邮政编码。  
如果要标准化完整电话子类型，请将以下子类型分配给完整电话组：电话号码和电话国家/地区代码。

9. 在“查看并保存”页面上，进行任何必要的更改，然后选择“编辑架构映射”。

## 删除架构映射

当架构映射与匹配的工作流程关联时，您无法将其删除。必须先从所有关联的匹配工作流中移除架构映射，然后才能将其删除。

要删除架构映射，请执行以下操作：

1. 登录 AWS 管理控制台 并打开 AWS Entity Resolution 数据匹配服务 控制台，网址为 <https://console.aws.amazon.com/entityresolution/>。
2. 在左侧导航窗格的“数据准备”下，选择“架构映射”。
3. 选择架构映射。
4. 选择删除。
5. 确认删除，然后选择删除。

# 使用 ID 命名空间定义输入数据

ID 命名空间是围绕输入数据表的包装。您可以使用 ID 命名空间来提供元数据，解释您的输入数据和匹配技术，以及如何在 [ID 映射工作流程](#) 中使用它们。

ID 命名空间有两种类型：源和目标。

- 源包含在 ID 映射工作 AWS Entity Resolution 数据匹配服务 流程中处理的源数据的配置。
- Target 包含所有源解析的目标数据配置。

您可以在 ID 映射工作流程中定义要 AWS 账户 在两个之间解析的输入数据。一个参与者创建 ID 命名空间源，另一个参与者创建 ID 命名空间目标。参与者创建源和目标后，您可以运行 ID 映射工作流程将数据从源转换为目标。

下图总结了如何创建用于 ID 映射工作流程的 ID 命名空间。



#### Prerequisite

An ID namespace that is a source requires a data input: [schema mapping](#) and an associated AWS Glue database. An ID namespace that is the target requires a target domain.



#### Create ID namespace

Provide the name and description, and then choose the type: source or target.



#### Configure your data

Select the configuration method and enter your source or target information.



#### Use in ID mapping workflows

Use your ID namespace as either a source or a target in an ID mapping workflow across two AWS accounts.

以下各节介绍如何创建 ID 命名空间源和 ID 命名空间目标。

## 主题

- [ID 命名空间来源](#)
- [ID 命名空间目标](#)
- [编辑 ID 命名空间](#)
- [删除 ID 命名空间](#)
- [添加或更新 ID 命名空间的资源策略](#)

## ID 命名空间来源

ID 命名空间源是 [ID 映射工作流程](#) 中的数据源。

在创建 ID 命名空间源之前，必须先创建架构映射或匹配的工作流程，具体取决于您的用例。有关更多信息，请参阅[创建架构映射](#)和[使用匹配流程匹配输入数据](#)。

创建 ID 命名空间源后，您可以在 ID 映射工作流程中将其与 ID 命名空间目标一起使用。有关更多信息，请参阅[使用 ID 映射工作流程映射输入数据](#)。

在 AWS Entity Resolution 数据匹配服务 控制台中创建 ID 命名空间源有两种方法：[基于规则的方法](#)或[提供者服务方法](#)。

## 主题

- [创建 ID 命名空间源 \( 基于规则 \)](#)
- [创建 ID 命名空间源 \( 提供者服务 \)](#)

## 创建 ID 命名空间源 ( 基于规则 )

本主题介绍使用基于规则的方法创建 ID 命名空间源的过程。此方法使用匹配规则在 ID 映射工作流程中将第一方数据从源转换为目标。

### Note

如果输入数据是源，则它必须具有架构映射和关联 AWS Glue 的数据库。

## 创建 ID 命名空间源 ( 基于规则 )

1. 登录 AWS 管理控制台 并打开 AWS Entity Resolution 数据匹配服务 控制台，网址为<https://console.aws.amazon.com/entityresolution/>。
2. 在左侧导航窗格的数据准备下，选择 ID 命名空间。
3. 在 ID 命名空间页面的右上角，选择创建 ID 命名空间。
4. 有关详细信息，请执行以下操作：
  - a. 在 ID 命名空间名称中，输入一个唯一的名称。
  - b. ( 可选 ) 在描述中，输入可选描述。
  - c. 对于 ID 命名空间类型，请选择来源。
5. 对于 ID 命名空间方法，请选择基于规则。
6. 对于数据输入，请选择要使用的输入类型，然后采取建议的操作。

输入类型	建议的操作
现有的架构映射	<ol style="list-style-type: none"> <li>1. 选择架构映射。</li> <li>2. 从下拉列表中选择AWS 区域、AWS Glue 数据库、AWS Glue 表和架构映射。</li> </ol> <p>您最多可以添加 19 个数据输入。</p> <div style="border: 1px solid #add8e6; border-radius: 10px; padding: 10px; margin-top: 10px;"> <p> <b>Note</b></p> <p>如果您的数据表有 DELETE 列，则架构映射的类型必须为 <code>String</code> 并且不能使用 <code>matchKey</code> 和 <code>groupName</code>。</p> </div>
现有的匹配工作流程	<ol style="list-style-type: none"> <li>1. 选择匹配工作流程。</li> <li>2. 选择与 ID 命名空间关联的账户：您的账户 AWS 账户或其他 AWS 账户。</li> <li>3. 根据账户类型，选择匹配工作流程名称或输入匹配工作流程 ARN。</li> </ol>

7. 对于规则参数，请执行以下操作。
- a. 根据您的目标选择以下选项之一，指定规则控件。

您的目标	建议的选项
允许来自源和目标的规则	没有偏好
选择来源、目标或两者是否可以在 ID 映射工作流程中提供规则	有限的规则

规则控件必须在源和目标之间兼容，才能在 ID 映射工作流程中使用。例如，如果源 ID 命名空间将规则限制于目标，但目标 ID 命名空间将规则限制于源，则会导致错误。

- b. 根据您的数据输入类型选择以下选项之一，指定匹配规则。

数据输入类型	推荐操作
架构映射	选择“添加其他规则”以添加匹配的规则。  您最多可以应用 25 个匹配规则来定义匹配标准。
匹配工作流程	选择“使用匹配工作流程中的规则”或“提供新规则”来定义您的匹配规则。

8. 对于比较和匹配参数，请执行以下操作。

- a. 根据您的目标选择以下选项之一，指定比较类型。

您的目标	建议的选项
在创建 ID 映射工作流程时，允许使用任何比较类型。	没有偏好
查找存储在多个输入字段中的数据的任意匹配组合，无论数据位于相同还是不同的输入字段中。	多个输入字段
当存储在多个输入字段中的相似数据不应匹配时，请限制在单个输入字段内进行比较。	单一输入字段

- b. 根据您的目标选择以下选项之一，指定记录匹配类型。

您的目标	建议的选项
在创建 ID 映射工作流程时，允许使用任何比较类型。	没有偏好
创建 ID 映射工作流程时，将记录匹配类型限制为：对于目标中的每条匹配记录，仅存储源中的一条匹配记录。	有限的记录匹配  and  一个源对一个目标

您的目标	建议的选项
创建 ID 映射工作流程时，将记录匹配类型限制为：对于目标中的每条匹配记录，存储源中的所有匹配记录。	有限的记录匹配  and  一个目标有多个来源

### Note

您必须为源 ID 和目标 ID 命名空间指定兼容限制。例如，如果源 ID 命名空间将规则限制于目标，但目标 ID 命名空间将规则限制于源，则会导致错误。

9. 通过从下拉列表中选择现有服务角色名称来指定服务访问权限。
10. ( 可选 ) 要为资源启用标签，请选择添加新标签，然后输入密钥和值对。
11. 选择创建 ID 命名空间。

ID 命名空间源已创建。现在，您可以[创建 ID 命名空间目标](#)了。

## 创建 ID 命名空间源 ( 提供者服务 )

本主题介绍使用提供者服务方法创建 ID 命名空间源的过程。此方法使用名为的提供者服务 LiveRamp。LiveRamp 在 ID 映射工作流程中，将第三方编码的数据从源转换为目标。


### Note

如果输入数据是源，则它必须具有架构映射和关联 AWS Glue 的数据库。

## 创建 ID 命名空间源 ( 提供者服务 )

1. 登录 AWS 管理控制台 并打开 AWS Entity Resolution 数据匹配服务 控制台，网址为<https://console.aws.amazon.com/entityresolution/>。
2. 在左侧导航窗格的数据准备下，选择 ID 命名空间。
3. 在 ID 命名空间页面的右上角，选择创建 ID 命名空间。
4. 有关详细信息，请执行以下操作：

- a. 在 ID 命名空间名称中，输入一个唯一的名称。
  - b. ( 可选 ) 在描述中，输入可选描述。
  - c. 对于 ID 命名空间类型，请选择来源。
5. 对于 ID 命名空间方法，请选择提供者服务。

 Note

AWS Entity Resolution 数据匹配服务 目前将 LiveRamp 提供者服务作为 ID 命名空间方法提供。如果您订阅了 LiveRamp，则状态将显示为“已订阅”。有关如何订阅的更多信息 LiveRamp，请参阅[步骤 1：在上订阅提供商服务 AWS Data Exchange](#)。

6. 对于数据输入，请从下拉列表中选择、AWS Glue 数据库、AWS Glue 表和架构映射。AWS 区域您最多可以添加 20 个数据输入。
7. 要指定服务访问权限，请选择一个选项并采取建议的操作。

Option	推荐操作
创建并使用新的服务角色	<ul style="list-style-type: none"> <li>• AWS Entity Resolution 数据匹配服务 使用此表所需的策略创建服务角色。</li> <li>• 默认的服务角色名称为entityresolution-id-mapping-workflow-<code>&lt;timestamp&gt;</code>。</li> <li>• 您必须拥有创建角色并附加策略的权限。</li> <li>• 如果您的输入数据已加密，请选择此数据由 KMS 密钥加密选项。然后，输入用于解密输入数据的密AWS KMS 钥。</li> </ul>
使用现有服务角色	<ol style="list-style-type: none"> <li>1. 从下拉列表中选择一个现有服务角色名称。</li> </ol> <p>如果您有列出角色的权限，则会显示角色列表。</p> <p>如果您没有列出角色的权限，可以输入要使用的角色的 Amazon 资源名称 (ARN)。</p>

Option	推荐操作
	<p>如果没有现有的服务角色，则使用现有服务角色选项不可用。</p> <p>2. 通过选择在 IAM 中查看外部链接来查看服务角色。</p> <p>默认情况下，AWS Entity Resolution 数据匹配服务不会尝试更新现有角色策略以添加必要的权限。</p>

8. (可选) 要为资源启用标签，请选择添加新标签，然后输入密钥和值对。
9. 选择创建 ID 命名空间。

ID 命名空间源已创建。现在，您可以[创建 ID 命名空间目标](#)了。

## ID 命名空间目标

ID 命名空间目标是 [ID 映射工作流程](#) 中数据的目标。所有来源都解析到目标。

在创建 ID 命名空间目标之前，必须先创建匹配的工作流程或订阅提供者服务 (LiveRamp)，具体取决于您的用例。有关更多信息，请参阅[使用匹配流程匹配输入数据](#)和[步骤 1：在上订阅提供商服务 AWS Data Exchange](#)。

创建 ID 命名空间目标后，您可以在 ID 映射工作流程中将其与 ID 命名空间源一起使用。有关更多信息，请参阅[使用 ID 映射工作流程映射输入数据](#)。

在 AWS Entity Resolution 数据匹配服务 控制台中创建 ID 命名空间目标有两种方法：[基于规则的方法](#)或[提供者服务方法](#)。

### 主题

- [创建 ID 命名空间目标 \(基于规则的方法\)](#)
- [创建 ID 命名空间目标 \(提供者服务方法\)](#)

## 创建 ID 命名空间目标 ( 基于规则的方法 )

本主题介绍使用基于规则的方法创建 ID 命名空间目标的过程。在 ID 映射工作流程中，此方法使用匹配规则将第一方数据从源转换为目标。

### 创建 ID 命名空间目标 ( 基于规则 )

1. 登录 AWS 管理控制台 并打开 AWS Entity Resolution 数据匹配服务 控制台，网址为<https://console.aws.amazon.com/entityresolution/>。
2. 在左侧导航窗格的数据准备下，选择 ID 命名空间。
3. 在 ID 命名空间页面的右上角，选择创建 ID 命名空间。
4. 有关详细信息，请执行以下操作：
  - a. 在 ID 命名空间名称中，输入一个唯一的名称。
  - b. ( 可选 ) 在描述中，输入可选描述。
  - c. 对于 ID 命名空间类型，请选择目标。
5. 对于 ID 命名空间方法，请选择基于规则。
6. 对于数据输入，在“匹配工作流程”下，执行以下操作。
  - a. 选择与 ID 命名空间关联的账户：您的账户 AWS 账户或其他 AWS 账户。
  - b. 根据账户类型，选择匹配工作流程名称或输入匹配工作流程 ARN。
7. 对于规则参数，请执行以下操作。
  - a. 根据您的目标选择以下选项之一，指定规则控件。

您的目标	建议的选项
允许来自源和目标的规则	没有偏好
选择来源、目标或两者是否可以在 ID 映射工作流程中提供规则	有限的规则

规则控件必须在源和目标之间兼容，才能在 ID 映射工作流程中使用。例如，如果源 ID 命名空间将规则限制于目标，但目标 ID 命名空间将规则限制于源，则会导致错误。

- b. 对于匹配规则，AWS Entity Resolution 数据匹配服务 自动添加匹配工作流程中的规则。
8. 对于比较和匹配参数，请执行以下操作。

- a. 根据您的目标选择以下选项之一，指定比较类型。

您的目标	建议的选项
在创建 ID 映射工作流程时，允许使用任何比较类型。	没有偏好
查找存储在多个输入字段中的数据的任意匹配组合，无论数据位于相同还是不同的输入字段中。	多个输入字段
当存储在多个输入字段中的相似数据不应匹配时，请限制在单个输入字段内进行比较。	单一输入字段

- b. 根据您的目标选择以下选项之一，指定记录匹配类型。

您的目标	建议的选项
在创建 ID 映射工作流程时，允许使用任何比较类型。	没有偏好
创建 ID 映射工作流程时，将记录匹配类型限制为：对于目标中的每条匹配记录，仅存储源中的一条匹配记录。	有限的记录匹配 and 一个源对一个目标
创建 ID 映射工作流程时，将记录匹配类型限制为：对于目标中的每条匹配记录，存储源中的所有匹配记录。	有限的记录匹配 and 一个目标有多个来源

**Note**

您必须为源 ID 和目标 ID 命名空间指定兼容限制。例如，如果源 ID 命名空间将规则限制于目标，但目标 ID 命名空间将规则限制于源，则会导致错误。

9. 通过从下拉列表中选择现有服务角色名称来指定服务访问权限。
10. ( 可选 ) 要为资源启用标签，请选择添加新标签，然后输入密钥和值对。
11. 选择创建 ID 命名空间。

ID 命名空间目标已创建。创建 ID 映射工作流程所需的 ID 命名空间 ( 源和目标 ) 后，就可以[创建 ID 映射](#)工作流程了。

## 创建 ID 命名空间目标 ( 提供者服务方法 )

本主题介绍使用提供者服务方法创建 ID 命名空间目标的过程。此方法使用名为的提供者服务 LiveRamp。LiveRamp 在 ID 映射工作流程中，将第三方编码的数据从源转换为目标。

### 创建 ID 命名空间目标 ( 提供者服务 )

1. 登录 AWS 管理控制台 并打开 AWS Entity Resolution 数据匹配服务 控制台，网址为<https://console.aws.amazon.com/entityresolution/>。
2. 在左侧导航窗格的数据准备下，选择 ID 命名空间。
3. 在 ID 命名空间页面的右上角，选择创建 ID 命名空间。
4. 有关详细信息，请执行以下操作：
  - a. 在 ID 命名空间名称中，输入一个唯一的名称。
  - b. ( 可选 ) 在描述中，输入可选描述。
  - c. 对于 ID 命名空间类型，请选择目标。
5. 对于 ID 命名空间方法，请选择提供者服务。

#### Note

AWS Entity Resolution 数据匹配服务 目前将 LiveRamp 提供者服务作为 ID 命名空间方法提供。

如果您订阅了 LiveRamp，则状态将显示为“已订阅”。

有关如何订阅的更多信息 LiveRamp，请参阅[步骤 1：在上订阅提供商服务 AWS Data Exchange](#)。

6. 在目标域中，输入 LiveRamp 提供的转码目标 LiveRamp 客户机域标识符。
7. ( 可选 ) 要为资源启用标签，请选择添加新标签，然后输入密钥和值对。
8. 选择创建 ID 命名空间。

ID 命名空间目标已创建。创建 ID 映射工作流程所需的 ID 命名空间（源和目标）后，即可开始[创建 ID 映射](#)工作流程。

## 编辑 ID 命名空间

在将 ID 命名空间关联到 ID 映射工作流程之前，您只能对其进行编辑。将 ID 命名空间与 ID 映射工作流程关联后，便无法对其进行编辑。

要编辑 ID 命名空间，请执行以下操作：

1. 登录 AWS 管理控制台 并打开 AWS Entity Resolution 数据匹配服务 控制台，网址为<https://console.aws.amazon.com/entityresolution/>。
2. 在左侧导航窗格的数据准备下，选择 ID 命名空间。
3. 选择 ID 命名空间。
4. 选择编辑。
5. 在编辑 ID 命名空间页面上，进行必要的更改，然后选择保存。

## 删除 ID 命名空间

当 ID 命名空间与 ID 映射工作流程关联时，您无法将其删除。必须先从所有关联的 ID 映射工作流中移除架构映射，然后才能将其删除。

要删除 ID 命名空间，请执行以下操作：

1. 登录 AWS 管理控制台 并打开 AWS Entity Resolution 数据匹配服务 控制台，网址为<https://console.aws.amazon.com/entityresolution/>。
2. 在左侧导航窗格的数据准备下，选择 ID 命名空间。
3. 选择 ID 命名空间。
4. 选择删除。
5. 确认删除，然后选择删除。

## 添加或更新 ID 命名空间的资源策略

资源策略允许 ID 映射资源的创建者访问您的 ID 命名空间资源。

## 添加或更新资源策略

1. 登录 AWS 管理控制台 并打开 AWS Entity Resolution 数据匹配服务 控制台，网址为<https://console.aws.amazon.com/entityresolution/>。
2. 在左侧导航窗格的“工作流程”下，选择 ID 命名空间。
3. 选择 ID 命名空间。
4. 在 ID 命名空间详细信息页面上，选择权限选项卡。
5. 在资源策略部分，选择编辑。
6. 在 JSON 编辑器中添加或更新策略。
7. 选择保存更改。

## 使用匹配流程匹配输入数据

匹配 workflow 是一种数据处理作业，它合并和比较来自不同输入源的数据，并根据不同的匹配技术确定哪些记录匹配。AWS Entity Resolution 数据匹配服务 从您指定的位置读取您的数据，查找记录之间的匹配项，并为每组 [匹配的数据分配一个匹配 ID](#)。

下图总结了如何创建匹配的工作流程。



### Complete prerequisite

Create a schema mapping to define your data.



### Choose your data input

Select the AWS Glue database and table that contains your data and the associated schema mapping.



### Set up matching techniques

Configure rule-based matching, use machine learning matching, or choose a provider service.



### Specify data output

Choose your data output fields and format to write to your S3 location.

## 主题

- [匹配 workflow 类型](#)
- [数据输出选项](#)
- [匹配 workflow 结果](#)
- [创建基于规则的匹配 workflow](#)
- [创建基于机器学习的匹配 workflow](#)
- [创建基于提供商服务的匹配 workflow](#)
- [编辑匹配的工作流程](#)
- [删除匹配的工作流程](#)
- [为基于规则的匹配 workflow 修改或生成匹配 ID](#)
- [为基于规则的匹配 workflow 查找匹配 ID](#)
- [从基于规则或 ML-based 匹配的工作流程中删除记录](#)
- [匹配 workflow 疑难解答](#)

## 匹配 workflow 类型

AWS Entity Resolution 数据匹配服务 支持三种类型的匹配 workflow：

## Rule-based 匹配

使用可配置的规则，根据指定字段的精确匹配或模糊匹配来识别匹配的记录。您可以定义匹配条件，例如拼写相似的匹配名称或格式不同的地址。

### 基于机器学习的匹配

使用机器学习模型识别相似的记录，即使数据存在差异、错误或缺失字段也是如此。这种方法可以检测比基于规则的匹配更复杂的匹配项。

### 基于提供商服务的匹配

在匹配之前，使用第三方数据提供商来丰富和验证您的数据。这种匹配类型与 Connect 客户档案输出不兼容。

## 数据输出选项

AWS Entity Resolution 数据匹配服务 可以将数据输出文件写入：

- 您指定的 Amazon S3 地点
- Connect 客户档案 ( 用于客户重复数据删除 )

### Important

导出到 Connect 客户档案与基于提供商的匹配不兼容。要导出到 Connect 客户档案，必须使用基于规则的匹配或基于机器学习的匹配。

如果需要 AWS Entity Resolution 数据匹配服务，您可以使用对输出数据进行哈希处理，从而帮助您保持对数据的控制。

下表显示了三种类型的匹配工作流及其支持的输出目标。

匹配类型	S3 输出	客户档案输出
<a href="#">基于规则</a>	是	是
<a href="#">基于机器学习</a>	是	是

匹配类型	S3 输出	客户档案输出
<a href="#">基于提供商服务</a>	是	否

## 匹配工作流程结果

创建并运行匹配的工作流程后，您可以在指定的 S3 位置或 Connect 客户档案中查看结果。匹配的工作流程会在数据编制索引后生成 ID。

匹配的工作流程可以有多个运行，结果（成功或错误）将写入名称为 jobId 的文件夹。

对于 S3 输出目标的每次运行：

- 数据输出既包含成功匹配的文件又包含错误文件
- 成功的结果将写入包含多个文件的文件夹 success
- 错误被写入包含多个字段的文件夹 error

每次运行 Connect 客户档案时，输出目的地：

- 经过重复数据删除的客户记录将直接发送到您的 Connect 客户实例
- 您可以在 AWS Entity Resolution 数据匹配服务 控制台中查看你最近的任务历史记录
- 重复数据删除流程中不包括 Connect Customer 中的现有配置文件

创建并运行匹配工作流程后，您可以使用[基于规则的匹配或机器学习 \(ML\) 匹配的输出作为基于提供商服务的匹配](#)的输入，或者反过来满足您的业务需求。

例如，为了节省提供商订阅成本，您可以先运行[基于规则的匹配](#)来查找数据上的匹配项。然后，您可以将不匹配记录的子集发送给[基于提供商服务的匹配](#)。请注意，如果您计划导出到客户档案，则应仅使用基于规则或基于机器学习的匹配。

有关故障排除的更多信息，请参阅[匹配工作流程疑难解答](#)。

## 创建基于规则的匹配工作流程

[Rule-based matching](#) 是一组分层的瀑布匹配规则 AWS Entity Resolution 数据匹配服务，由您根据输入的数据推荐，并且完全可以由您进行配置。基于规则的匹配工作流程使您可以比较明文数据或哈希数据，以根据您的自定义条件找到精确的匹配项。

在您的数据中 AWS Entity Resolution 数据匹配服务 发现两条或多条记录之间存在匹配项时，它会分配：

- 与[匹配数据集中的记录的匹配 ID](#)
- 生成[匹配项的匹配规则](#)。

在中创建基于规则的匹配工作流程时 AWS Entity Resolution 数据匹配服务，必须选择“简单”或“高级”规则类型。规则类型决定了您可以创建的规则条件的复杂性。创建工作流程后，您无法更改规则类型。

您可以使用下表来比较两种规则类型，并确定哪一种适合您的用例。

#### 规则类型对比图

使用案例	高级规则类型	简单规则类型
架构映射与输入类型一对一映射	是	否
将多个数据列映射到相同输入类型的架构映射	没有	是
支持精确匹配和模糊匹配	是	否（仅限精确匹配）
支持 AND、OR 和圆括号运算符	是	否（仅限 AND 运算符）
支持批处理工作流程	是	是
支持增量工作流程	是	是
支持实时工作流程	否	是
支持 ID 映射工作流程	没有	是

确定要使用哪种规则类型后，请使用以下主题创建基于规则的匹配工作流程，其中包括“高级”或“简单”规则类型。

#### 主题

- [使用高级规则类型创建基于规则的匹配工作流](#)
- [使用“简单”规则类型创建基于规则的匹配工作流程](#)

## 使用高级规则类型创建基于规则的匹配 workflow

### 先决条件

在创建基于规则的匹配 workflow 之前，您必须：

1. 创建架构映射。有关更多信息，请参阅 [创建架构映射](#)。
2. 如果使用 Connect 客户档案作为输出目标，请确保配置了相应的权限。

以下过程演示如何使用 AWS Entity Resolution 数据匹配服务 控制台或 API 创建具有高级规则类型的基于规则的匹配 workflow。CreateMatchingWorkflow

### Console

使用创建基于规则的匹配 workflow 高级 使用控制台的规则类型

1. 登录 AWS 管理控制台 并打开 AWS Entity Resolution 数据匹配服务 控制台，网址为 <https://console.aws.amazon.com/entityresolution/>。
2. 在左侧导航窗格的“workflow”下，选择“匹配”。
3. 在匹配 workflow 页面的右上角，选择创建匹配 workflow。
4. 对于“步骤 1：指定匹配的 workflow 详细信息”，请执行以下操作：
  - a. 输入匹配的 workflow 名称和可选的描述。
  - b. 对于数据输入 AWS 区域，选择 AWS Glue 数据库、AWS Glue 表，然后选择相应的架构映射。

您最多可以添加 19 个数据输入。

#### Note

要使用高级规则，您的架构映射必须满足以下要求：

1. 除非字段组合在一起，否则每个输入字段都必须映射到唯一的匹配键。
2. 如果将输入字段组合在一起，则它们可以共享相同的匹配键。

例如，以下架构映射对高级规则有效：

```
firstName: { matchKey: 'name', groupName: 'name' }
```


```
lastName: { matchKey: 'name', groupName: 'name' }
```

在这种情况下，firstName和lastName字段组合在一起并共享相同的名称匹配密钥，这是允许的。

检查您的架构映射并更新它们以遵循此一对一匹配规则，除非字段已正确分组，否则才能使用高级规则。

3. 如果您的数据表有 DELETE 列，则架构映射的类型必须为，String并且不能使用matchKey和groupName。

- c. 默认情况下，“标准化数据”选项处于选中状态，以便在匹配之前对数据输入进行标准化。如果您不想对数据进行标准化处理，请取消选择“标准化数据”选项。

 Note

创建架构映射中仅支持以下场景的标准化：

- 如果将以下“名称”子类型分组：名字、中间名、姓氏。
- 如果将以下地址子类型分组：街道地址 1、街道地址 2、街道地址 3、城市、州、国家、邮政编码。
- 如果将以下电话子类型分组：电话号码、电话国家/地区代码。

- d. 要指定服务访问权限，请选择一个选项并采取建议的操作。

Option	推荐操作
创建并使用新的服务角色	<ul style="list-style-type: none"> <li>• AWS Entity Resolution 数据匹配服务使用此表所需的策略创建服务角色。</li> <li>• 默认服务角色名称为 <code>entityresolution-matching-workflow- &lt;timestamp&gt;</code> 。</li> <li>• 您必须拥有创建角色并附加策略的权限。</li> <li>• 如果您的输入数据已加密，则可以选择“此数据使用 KMS 密钥加密”选项，然后输入用于解密输入数据的密 AWS KMS 钥。</li> </ul>
使用现有服务角色	<ol style="list-style-type: none"> <li>1. 从下拉列表中选择一个现有服务角色名称。  如果您有列出角色的权限，则会显示角色列表。  如果您没有列出角色的权限，可以输入要使用的角色的 Amazon 资源名称 (ARN)。  如果没有现有的服务角色，则使用现有服务角色选项不可用。</li> <li>2. 通过选择在 IAM 中查看外部链接来查看服务角色。  默认情况下，AWS Entity Resolution 数据匹配服务不会尝试更新现有角色策略以添加必要的权限。</li> </ol>

e. (可选) 要为资源启用标签，请选择添加新标签，然后输入密钥和值对。

f. 选择下一步。

5. 对于步骤 2：选择匹配技术：

- a. 在“匹配方法”中，选择Rule-based匹配。
- b. 对于“规则类型”，选择“高级”。

The screenshot shows the 'Choose matching technique' step in the AWS Entity Resolution console. The 'Resolution type' section has 'Rule-based matching' selected. The 'Rule type' section has 'Advanced - new' selected. The 'Processing cadence' section has 'Manual' selected. The 'Matching rules (1)' section shows a single rule with the name 'Enter rule name' and the condition 'Exact(Name) AND Exact(Phone)'. The 'Add another rule' button is visible at the bottom of the rule list.

- c. 在“处理节奏”中，选择以下选项之一。
  - 选择“手动”，按需运行工作流以进行批量更新
  - 选择“自动”以在 S3 存储桶中有新数据后立即运行工作流程

#### Note

如果您选择“自动”，请确保您的 S3 存储桶已启用 Amazon EventBridge 通知。有关 EventBridge 使用 S3 控制台启用亚马逊的说明，请参阅 [Amazon S3 用户指南 EventBridge 中的启用亚马逊](#)。

- d. 对于匹配规则，输入规则名称，然后根据您的目标从下拉列表中选择相应的匹配函数和运算符来构建规则条件。

您最多可以创建 25 条规则。

**Note**

AWS Entity Resolution 数据匹配服务 还支持[传递匹配](#)，它可以处理所有规则级别的记录，以传递方式连接匹配组。传递匹配作为一项 API-only 功能提供。启用传递匹配后，不支持 EmptyValues=Ignore 修饰符。有关更多信息，请参阅[使用传递匹配](#)。

必须使用 AND 运算符将模糊匹配函数 ( Cosine、Levenshtein 或 Soundex ) 与精确匹配函数 ( Exact、ExactManyToMany ) 组合在一起。

根据您的目标，您可以使用下表来帮助决定要使用哪种类型的函数或运算符。

您的目标	推荐的函数或运算符	推荐的可选修饰符	优点
在准确的数据上匹配相同的字符串，但在空值上不匹配。	精确	EmptyValues=进程	
在准确的数据上匹配相同的字符串并忽略空值。	精确 ( <i>matchKey</i> )	EmptyValues=忽略	
在匹配键中匹配多条记录。适合灵活配对。限制：15 个匹配密钥	ExactMany ToMany( <i>matchKey</i> , <i>matchKey</i> , ...)	n/a	

您的目标	推荐的函数或运算符	推荐的可选修饰符	优点
测量数据的数字表示形式之间的相似性，但在空值上不匹配。适用于文本、数字或两者的混合。	余弦	EmptyValues=进程	简单、高效。  与加 TF-IDF 权结合使用时，可以很好地处理长文本。  非常适合基于单词的精确匹配。
测量数据的数值表示之间的相似性并忽略空值。	余弦 ( <i>matchKey, thresho</i> ,...)	EmptyValues=忽略	很好地处理错别字、拼写错误和移位。
计算将一个单词更改为另一个单词但与空值不匹配所需的最小更改次数。适用于拼写略有差异的文本。	莱文斯泰因	EmptyValues=进程	对各种 PII 类型有效。  适用于短字符串（例如，姓名或电话号码）。
计算将一个单词更改为另一个单词并忽略空值所需的最小更改次数。	Levenshtein ( <i>matchKey, ,... thres</i> )	EmptyValues=忽略	

您的目标	推荐的函数或运算符	推荐的可选修饰符	优点
根据文本字符串的相似程度比较和匹配这些字符串，但在空值上不匹配。适用于拼写或发音变化的文本。	Sounde	EmptyValues=进程	有效进行语音匹配，识别发音相似的单词。  速度快且计算成本低。
根据文本字符串的相似程度对其进行比较和匹配，并忽略空值。	Soundex () <i>matchKey</i>	EmptyValues=忽略	适合匹配发音相似但拼写不同的名字。
组合函数。	AND	n/a	
单独的函数。	或者	n/a	
对条件进行分组以创建嵌套条件。	(...)	n/a	

### Example与电话号码和电子邮件相匹配的规则条件

以下是匹配电话号码（电话匹配密钥）和电子邮件地址（电子邮件地址匹配密钥）记录的规则条件示例：

```
Exact(Phone,EmptyValues=Process) AND Levenshtein("Email address",2)
```

**Matching rules (1)**  
Define match criteria by creating a rule condition for each matching rule. Rearrange the priority to optimize results. You can create up to 25 rules.

**Rule name**  
 Remove ▼ ▲  
5 of 255 characters. Use alphanumeric, underscore (\_), or hyphen (-) characters.

**Rule condition - beta** [Info](#)  
Choose the appropriate matching functions and operators to build this rule condition.

1 Exact(Phone,EmptyValues=Process) AND Levenshtein("Email address",2)

Errors: 0 Line 1, Column 67

+ Add another rule Reset rules  
You can add up to 24 more rules.

Cancel Previous Next

电话匹配键使用精确匹配功能来匹配相同的字符串。电话匹配键使用 EmptyValues=Process 修饰符处理匹配中的空值。

电子邮件地址匹配键使用 Levenshtein 匹配功能使用默认的 Levenshtein 距离算法阈值 2 来匹配拼写错误的的数据。电子邮件匹配键不使用任何可选修饰符。

AND 运算符结合了精确匹配函数和 Levenshtein 匹配函数。

Example用于执行火柴键匹配 ExactManyToMany 的规则条件

以下是一个规则条件示例，该规则条件用于匹配三个地址字段（ HomeAddress匹配键、BillingAddress匹配键和ShippingAddress匹配键）上的记录，通过检查其中是否有相同值来查找潜在的匹配项。

ExactManyToMany运算符会评估指定地址字段的所有可能组合，以确定任意两个或多个地址之间的精确匹配项。例如，它将检测是否与或HomeAddress匹配ShippingAddress，BillingAddress或者所有三个地址是否完全匹配。

```
ExactManyToMany(HomeAddress, BillingAddress, ShippingAddress)
```

Example使用聚类的规则条件

在具有模糊条件的基于规则的高级匹配中，系统首先根据精确匹配将记录分组为聚类。形成这些初始聚类后，系统会应用模糊匹配过滤器来识别每个聚类中的其他匹配项。为了获得最佳性能，您应该根据自己的数据模式选择精确匹配条件来创建定义明确的初始群集。

以下是将多个精确匹配与模糊匹配要求相结合的规则条件示例。它使用AND运算符来检查三个字段 — FullName、出生日期 (DOB) 和 Address — 在记录之间是否完全匹配。使用 Levenshtein 的距离，它还允许在InternalID场上进行细微的变化。1Levenshtein 距离测量将一个字符串更改为另一个字符串所需的最小单字符编辑次数。距离为 1 表示它将匹配仅InternalIDs相差一个字符（例如单个错字、删除或插入）。这种条件组合有助于识别极有可能代表同一实体的记录，即使标识符之间存在细微差异。

```
Exact(FullName) AND Exact(DOB) AND Exact(Address) and
Levenshtein(InternalID, 1)
```

- e. 选择下一步。
6. 对于步骤 3：指定数据输出和格式：
    - a. 对于数据输出目标和格式，选择数据输出的 Amazon S3 位置，以及数据格式是标准化数据还是原始数据。
    - b. 对于加密，如果您选择自定义加密设置，请输入AWS KMS 密钥 ARN。
    - c. 查看系统生成的输出。
    - d. 对于数据输出，请决定要包含、隐藏或掩盖哪些字段，然后根据目标采取建议的操作。

您的目标	推荐操作
包括字段	将输出状态保持为“已包含”。
隐藏字段（从输出中排除）	选择“输出”字段，然后选择“隐藏”。
掩码字段	选择“输出”字段，然后选择“哈希输出”。
重置之前的设置	选择 重置。

- e. 选择下一步。
7. 对于步骤 4：查看并创建：
    - a. 查看您在之前的步骤中所做的选择，并在必要时进行编辑。
    - b. 选择创建并运行。

将出现一条消息，表示匹配的工作流程已创建且作业已启动。

8. 在匹配的工作流程详细信息页面的指标选项卡上，在“上次作业指标”下查看以下内容：

- 作业 ID。
- 匹配 workflow 作业的状态：已排队、进行中、已完成、失败
- workflow 作业的完成时间。
- 已处理的记录数。
- 未处理的记录数。
- 生成的唯一匹配 ID。
- 输入记录的数量。

您还可以查看任务历史记录下先前运行过的匹配 workflow 作业的作业指标。

9. 匹配的 workflow 任务完成（状态为已完成）后，您可以转到数据输出选项卡，然后选择您的 Amazon S3 位置以查看结果。
10. （仅限手动处理类型）如果您创建了与“手动”处理类型的 Rule-based 匹配 workflow，则可以在匹配的 workflow 详细信息页面上选择“运行 workflow”，随时运行匹配的 workflow。
11. （仅限自动处理类型）如果您的数据表有 DELETE 列，则：
  - 删除列 *true* 中设置为的记录。
  - 在“删除”列 *false* 中设置为的记录将提取到 S3 中。

有关更多信息，请参阅 [步骤 1：准备第一方数据表](#)。

## API

使用创建基于规则的匹配 workflow 高级 使用 API 的规则类型


### Note

默认情况下，workflow 使用标准（批处理）处理。要使用增量（自动）处理，必须对其进行明确配置。

1. 打开终端或命令提示符发出 API 请求。
2. 向以下端点创建 POST 请求：

```
/matchingworkflows
```

- 在请求标头中，将设置 Content-type 为 application/json。

 Note

有关支持的编程语言的完整列表，请参阅 [AWS Entity Resolution 数据匹配服务 API 参考](#)。

- 对于请求正文，请提供以下必需的 JSON 参数：

```
{
  "description": "string",
  "incrementalRunConfig": {
    "incrementalRunType": "string"
  },
  "inputSourceConfig": [
    {
      "applyNormalization": boolean,
      "inputSourceARN": "string",
      "schemaName": "string"
    }
  ],
  "outputSourceConfig": [
    {
      "applyNormalization": boolean,
      "KMSArn": "string",
      "output": [
        {
          "hashed": boolean,
          "name": "string"
        }
      ],
      "outputS3Path": "string"
    }
  ],
  "resolutionTechniques": {
    "providerProperties": {
      "intermediateSourceConfiguration": {
        "intermediateS3Path": "string"
      },
      "providerConfiguration": JSON value,
      "providerServiceArn": "string"
    },
    "resolutionType": "RULE_MATCHING",
  }
```

```

    "ruleBasedProperties": {
      "attributeMatchingModel": "string",
      "matchPurpose": "string",
      "rules": [
        {
          "matchingKeys": [ "string" ],
          "ruleName": "string"
        }
      ]
    },
    "ruleConditionProperties": {
      "rules": [
        {
          "condition": "string",
          "ruleName": "string"
        }
      ]
    }
  },
  "roleArn": "string",
  "tags": {
    "string" : "string"
  },
  "workflowName": "string"
}

```

其中：

- workflowName (必填) — 必须是唯一的，并且必须介于 1—255 个字符之间，匹配模式 [a-z A-Z \_0-9-]\*
- inputSourceConfig (必填) — 1—20 个输入源配置列表
- outputSourceConfig (必填) — 只有一个输出源配置
- resolutionTechniques (必填) — 设置为 “RULE\_MATCHING” 作为基于规则的匹配的分辨率类型
- roleArn (必填) — 用于执行工作流程的 IAM 角色 ARN
- ruleConditionProperties (必填) -规则条件列表和匹配规则的名称。

可选参数包括：

- description— 最多 255 个字符

- `incrementalRunConfig`— 增量运行类型配置
  - `tags`— 最多 200 个键值对
5. (可选) 要使用增量处理而不是默认的标准 (批处理) 处理, 请在请求正文中添加以下参数:

```
"incrementalRunConfig": {
  "incrementalRunType": "AUTOMATIC"
}
```

6. 发送请求。
7. 如果成功, 您将收到状态码 200 的响应和包含以下内容的 JSON 正文:

```
{
  "workflowArn": "string",
  "workflowName": "string",
  // Plus all configured workflow details
}
```

8. 如果呼叫失败, 您可能会收到以下错误之一:
- 400- `ConflictException` 如果工作流程名称已经存在
  - 400 — `ValidationException` 如果输入未通过验证
  - 402- `ExceedsLimitException` 如果超过账户限额
  - 403 — `AccessDeniedException` 如果你没有足够的访问权限
  - 429 — `ThrottlingException` 如果请求被限制
  - 500 — `InternalServerErrorException` 如果内部服务出现故障

## 使用传递匹配

默认情况下, AWS Entity Resolution 数据匹配服务 使用瀑布匹配方法, 将匹配更高规则级别的记录排除在后续规则之外。这意味着下一个规则只评估不匹配的记录。虽然这种方法适用于单源匹配, 但当有多个具有不同属性的数据源时, 它可能会导致问题。

使用瀑布法, 您可能需要将所有匹配的逻辑组合成一个过于宽松的规则, 以匹配不同来源的记录。这可能导致过度匹配, 即不匹配的记录被错误地分组在一起。

传递匹配通过处理所有规则级别的所有记录来解决这个问题。记录与规则匹配后, 其匹配 ID 即已固定, 但该记录仍可充当链接, 将后期规则中的不匹配记录与先前规则中的匹配组连接起来。此功能目前仅可通过 API 使用。

## 传递匹配的工作原理

传递匹配对当前规则中的每个匹配组使用以下匹配 ID 解析流程：

- 系统会检查组中是否有记录已有来自较早规则级别的匹配 ID。
- 如果存在匹配 ID，则整个群组将继承之前的匹配 ID。
- 如果存在多个候选项，则选择最早规则级别中最小的匹配 ID。
- 如果之前不存在匹配项，则会为该组分配一个新的匹配 ID。

## 启用传递匹配 (API)

### Important

传递匹配是一项 API-only 功能。您无法通过 AWS Entity Resolution 数据匹配服务 控制台启用传递匹配。

要使用传递匹配，必须使用 CreateMatchingWorkflow API 创建具有高级规则类型的新匹配工作流程。您不能在现有工作流程中添加传递匹配。

在请求正文 true 中加入 enableTransitiveMatching 设置为的 matchingConfig 参数。以下示例显示了启用传递匹配的完整 CreateMatchingWorkflow 请求正文：

```
{
  "workflowName": "my-transitive-workflow",
  "inputSourceConfig": [
    {
      "inputSourceARN": "arn:aws:glue:us-east-1:123456789012:table/my-database/my-table",
      "schemaName": "my-schema",
      "applyNormalization": true
    }
  ],
  "outputSourceConfig": [
    {
      "outputS3Path": "s3://my-bucket/output/",
      "output": [
        {
          "name": "name",
          "hashed": false
        }
      ]
    }
  ]
}
```

```
    }
  ]
}
],
"resolutionTechniques": {
  "resolutionType": "RULE_MATCHING",
  "ruleConditionProperties": {
    "rules": [
      {
        "ruleName": "Rule1",
        "condition": "Exact(Email) AND Exact(Phone)"
      },
      {
        "ruleName": "Rule2",
        "condition": "Exact(Name) AND Exact(Address)"
      }
    ]
  }
},
"matchingConfig": {
  "enableTransitiveMatching": true
},
"roleArn": "arn:aws:iam::123456789012:role/my-er-role"
}
```

### Important

该enableTransitiveMatching参数是不可变的。您只能在创建工作流程期间设置此参数，之后不能对其进行更改。

## 规则排序的最佳实践

使用传递匹配时，规则排序至关重要。遵循如下最佳实践：

- 从最具体（最高可信度）到最不具体，对规则进行排序。
- 对于诸如 SSN 或出生日期之类的唯一标识符属性，请将规则按相邻的对排列——一条规则包含该属性，下一条规则不包含该属性。这使系统能够正确处理缺少这些属性的记录。
- 请注意，规则的数量会影响工作流程延迟，因为记录是在所有规则级别上处理的。

## 限制

传递匹配有以下限制：

- 仅适用于高级匹配工作流程。不适用于简单规则类型或基于机器学习的匹配工作流程。
- 启用传递匹配时，不支持 `EmptyValues=Ignore` 修饰符。传递匹配包括对空值的内置处理，可防止过度匹配。
- 工作流程运行时间与定义的规则数量成比例增加，因为记录是在所有规则级别上处理的。
- 该 `enableTransitiveMatching` 设置只能在创建工作流程期间进行配置。要使用传递匹配，必须创建新的工作流程。
- 您有责任正确遵守订购规则。具有更多属性的规则必须先于具有较少属性的规则。不正确的排序可能会导致具有不同值的唯一字段（例如出生日期或 SSN）的记录被错误地组合在一起。

## 禁用传递匹配

该 `enableTransitiveMatching` 设置是不可变的。创建工作流程后，您无法更改此设置。

要禁用传递匹配，您必须创建一个不带 `enableTransitiveMatching` 参数的新匹配工作流程，或者在 `CreateMatchingWorkflow` API 请求 `false` 中将其设置为。没有 AWS Entity Resolution 数据匹配服务 控制台选项可以禁用传递匹配。

要将传递匹配工作流程替换为非传递性工作流程，请完成以下步骤：

1. 使用 `CreateMatchingWorkflow` API 创建新的匹配工作流程，不带 `matchingConfig` 参数或将参数 `enableTransitiveMatching` 设置为 `false`。
2. 运行新的工作流程以验证其是否产生了预期的结果。
3. 如果不再需要旧的传递匹配工作流程，请将其删除。有关更多信息，请参阅 [删除匹配的工作流程](#)。

## 性能注意事项

传递匹配工作流程比非传递性工作流程慢，因为记录是在所有规则级别上处理的，而不是在第一次匹配之后被排除在外。以下因素会影响性能：

## 规则计数

规则越多，处理时间越长。每增加一个规则级别都会添加合并迭代，因为系统必须在所有级别上传播匹配 ID。

## Batch 和增量处理

增量工作流程的开销高于大规模批处理工作流程。较大的基础记录存储需要更多的合并迭代才能在现有记录和新记录之间传播匹配 ID。

### 数据偏斜

分布偏斜的数据集（例如大型匹配组）会放大开销。当许多记录链接到同一个匹配组时，系统需要更多的传递链接操作来解析所有连接。

我们建议您在生产环境中使用具有代表性的数据量测试传递匹配工作流程。这可以帮助您了解特定数据和规则配置的性能特征。

## 使用“简单”规则类型创建基于规则的匹配工作流程

### 先决条件

在创建基于规则的匹配工作流程之前，您必须：

1. 创建架构映射。有关更多信息，请参阅 [创建架构映射](#)。
2. 如果使用 Connect 客户档案作为输出目标，请确保配置了相应的权限。

以下过程演示如何使用 AWS Entity Resolution 数据匹配服务 控制台或 API 创建具有简单规则类型的基于规则的匹配工作流程。CreateMatchingWorkflow


### Console

使用创建基于规则的匹配工作流程 简便 使用控制台的规则类型

1. 登录 AWS 管理控制台 并打开 AWS Entity Resolution 数据匹配服务 控制台，网址为<https://console.aws.amazon.com/entityresolution/>。
2. 在左侧导航窗格的“工作流程”下，选择“匹配”。
3. 在匹配工作流程页面的右上角，选择创建匹配工作流程。
4. 对于“步骤 1：指定匹配的工作流程详细信息”，请执行以下操作：
  - a. 输入匹配的工作流程名称和可选的描述。
  - b. 对于数据输入 AWS 区域，选择AWS Glue 数据库、AWS Glue 表，然后选择相应的架构映射。

您最多可以添加 19 个数据输入。

- c. 默认情况下，“标准化数据”选项处于选中状态，以便在匹配之前对数据输入进行标准化。如果您不想对数据进行标准化处理，请取消选择“标准化数据”选项。

 Note

创建架构映射中仅支持以下场景的标准化：

- 如果将以下“名称”子类型分组：名字、中间名、姓氏。
- 如果将以下地址子类型分组：街道地址 1、街道地址 2、街道地址 3、城市、州、国家、邮政编码。
- 如果将以下电话子类型分组：电话号码、电话国家/地区代码。

- d. 要指定服务访问权限，请选择一个选项并采取建议的操作。

Option	推荐操作
创建并使用新的服务角色	<ul style="list-style-type: none"> <li>• AWS Entity Resolution 数据匹配服务使用此表所需的策略创建服务角色。</li> <li>• 默认服务角色名称为 <code>entityresolution-matching-workflow- &lt;timestamp&gt;</code>。</li> <li>• 您必须拥有创建角色并附加策略的权限。</li> <li>• 如果您的输入数据已加密，则可以选择“此数据使用 KMS 密钥加密”选项，然后输入用于解密输入数据的密 AWS KMS 钥。</li> </ul>

Option	推荐操作
使用现有服务角色	<ol style="list-style-type: none"><li>1. 从下拉列表中选择一个现有服务角色名称。  如果您有列出角色的权限，则会显示角色列表。  如果您没有列出角色的权限，可以输入要使用的角色的 Amazon 资源名称 (ARN)。  如果没有现有的服务角色，则使用现有服务角色选项不可用。</li><li>2. 通过选择在 IAM 中查看外部链接来查看服务角色。  默认情况下，AWS Entity Resolution 数据匹配服务 不会尝试更新现有角色策略以添加必要的权限。</li></ol>

e. (可选) 要为资源启用标签，请选择添加新标签，然后输入密钥和值对。

f. 选择下一步。

5. 对于步骤 2：选择匹配技术：

a. 在“匹配方法”中，选择Rule-based匹配。

b. 对于规则类型，选择简单。

Step 1 Specify matching workflow details

Step 2 **Choose matching technique**

Step 3 Specify data output

Step 4 Review and create

### Choose matching technique Info

Specify how you want your data to be matched or choose a provider service.

#### Matching method

**Resolution type**

**Rule-based matching**  
Use customized rules to find exact matches.

**Machine learning-based matching**  
Use our machine learning model to help find a broader range of matches.

**Provider services**  
Use this option if you have a subscription to a preferred provider through AWS Data Exchange.

**Rule type Info**

The rule type determines whether you can create simple rule conditions or more complex rule conditions for your rule-based matching workflow. After creating the workflow, you can't change the rule type. [Learn more](#)

**Advanced - new**  
Suitable for fuzzy matching, exact matching, and schema mappings with data columns mapped one-to-one with input types. Real-time and ID mapping workflows not currently supported.

**Simple**  
Suitable for exact matching and schema mappings with multiple data columns mapped to the same input types. Supports real-time and ID mapping workflows.

**Processing cadence Info**

Determine how often to run your matching workflow job. The first job runs after you create the matching workflow. [See pricing](#)

**Manual**  
Your matching workflow job is run on demand. Useful for bulk processing.

**Automatic**  
Your matching workflow job is run automatically when you add or update your data inputs. Useful for incremental updates. This option is available only for rule-based matching. When using this option, matching rules can't be edited after creation.

**Index only for ID mapping - new**

**Turn on**  
By default, matching workflows generate IDs after the data is indexed. If you want to use the matching workflow as a source or a target in an ID mapping workflow, choose to only index the data and not generate IDs.

c. 在“处理节奏”中，选择以下选项之一。

- 选择“手动”，按需运行工作流以进行批量更新
- 选择“自动”以在 S3 存储桶中有新数据后立即运行工作流程

#### Note

如果您选择“自动”，请确保您的 S3 存储桶已启用 Amazon EventBridge 通知。有关 EventBridge 使用 S3 控制台启用亚马逊的说明，请参阅 [Amazon S3 用户指南 EventBridge 中的启用亚马逊](#)。

d. (可选) 对于仅适用于 ID 映射的索引，您可以选择启用仅索引数据而不生成 ID 的功能。

默认情况下，匹配的工作流程会在数据编制索引后生成 ID。

e. 在匹配规则中，输入规则名称，然后为该规则选择匹配密钥。

您最多可以创建 15 个规则，并且可以在规则中应用最多 15 个不同的匹配密钥来定义匹配条件。

**▼ Matching rules (1)**  
Apply up to 15 different match keys across your rules to define match criteria. Add or remove match keys, remove rules, create new rules, and rearrange the priority to optimize results. You can create up to 15 rules.

Rule name  
 Remove ▼ ▲  
 0 of 255 characters. Use alphanumeric, underscore (\_), or hyphen (-) characters.

Match keys  
 ▼  
 You can choose up to 15 more match keys.

+ Add another rule  
 You can add up to 14 more rules.

- f. 对于比较类型，请根据您的目标选择以下选项之一。

您的目标	建议的选项
在存储在多个输入字段中的数据中查找任意匹配项组合	多个输入字段
将比较限制为单个输入字段	单一输入字段

**▼ Comparison type**  
Choose how you want to compare similar data stored in different input fields when they are assigned the same match key.

Comparison type [Info](#)

**Multiple input fields**  
Find any combination of matches across data stored in multiple input fields, regardless of whether the data is in the same or different input field.

**Single input field**  
Limit comparison within a single input field, when similar data stored across multiple input fields should not be matched.

Cancel Previous Next

- g. 选择下一步。

6. 对于步骤 3：指定数据输出和格式：

- 对于数据输出目标和格式，选择数据输出的 Amazon S3 位置，以及数据格式是标准化数据还是原始数据。
- 对于加密，如果您选择自定义加密设置，请输入 AWS KMS 密钥 ARN。
- 查看系统生成的输出。

- d. 对于数据输出，请决定要包含、隐藏或掩盖哪些字段，然后根据目标采取建议的操作。

您的目标	推荐操作
包括字段	将输出状态保持为“已包含”。
隐藏字段（从输出中排除）	选择“输出”字段，然后选择“隐藏”。
掩码字段	选择“输出”字段，然后选择“哈希输出”。
重置之前的设置	选择 重置。

- e. 选择下一步。

7. 对于步骤 4：查看并创建：

- 查看您在之前的步骤中所做的选择，并在必要时进行编辑。
- 选择创建并运行。

将出现一条消息，表示匹配的工作流程已创建且作业已启动。

8. 在匹配的工作流程详细信息页面的指标选项卡上，在“上次作业指标”下查看以下内容：

- 作业 ID。
- 匹配工作流作业的状态：已排队、进行中、已完成、失败
- 工作流作业的完成时间。
- 已处理的记录数。
- 未处理的记录数。
- 生成的唯一匹配 ID。
- 输入记录的数量。

您还可以查看任务历史记录下先前运行过的匹配工作流作业的作业指标。

- 匹配的工作流程任务完成（状态为已完成）后，您可以转到数据输出选项卡，然后选择您的 Amazon S3 位置以查看结果。
- （仅限手动处理类型）如果您创建了与“手动”处理类型的 Rule-based 匹配工作流，则可以在匹配的工作流详细信息页面上选择“运行工作流”，随时运行匹配的工作流。

## API

使用创建基于规则的匹配工作流程 简便 使用 API 的规则类型

**Note**

默认情况下，工作流程使用标准（批处理）处理。要使用增量（自动）处理，必须对其进行明确配置。

1. 打开终端或命令提示符发出 API 请求。
2. 向以下端点创建 POST 请求：

```
/matchingworkflows
```

3. 在请求标头中，将设置 Content-type 为 application/json。

**Note**

有关支持的编程语言的完整列表，请参阅 [AWS Entity Resolution 数据匹配服务 API 参考](#)。

4. 对于请求正文，请提供以下必需的 JSON 参数：

```
{
  "description": "string",
  "incrementalRunConfig": {
    "incrementalRunType": "string"
  },
  "inputSourceConfig": [
    {
      "applyNormalization": boolean,
      "inputSourceARN": "string",
      "schemaName": "string"
    }
  ],
  "outputSourceConfig": [
    {
      "applyNormalization": boolean,
      "KMSArn": "string",
      "output": [
```

```
    {
      "hashed": boolean,
      "name": "string"
    }
  ],
  "outputS3Path": "string"
}
],
"resolutionTechniques": {
  "providerProperties": {
    "intermediateSourceConfiguration": {
      "intermediateS3Path": "string"
    },
    "providerConfiguration": JSON value,
    "providerServiceArn": "string"
  },
  "resolutionType": "RULE_MATCHING",
  "ruleBasedProperties": {
    "attributeMatchingModel": "string",
    "matchPurpose": "string",
    "rules": [
      {
        "matchingKeys": [ "string" ],
        "ruleName": "string"
      }
    ]
  },
  "ruleConditionProperties": {
    "rules": [
      {
        "condition": "string",
        "ruleName": "string"
      }
    ]
  }
},
"roleArn": "string",
"tags": {
  "string" : "string"
},
"workflowName": "string"
}
```

其中：

- `workflowName` (必填) — 必须是唯一的，并且必须介于 1—255 个字符之间，匹配模式 `[a-z A-Z _0-9-]*`
- `inputSourceConfig` (必填) — 1—20 个输入源配置列表
- `outputSourceConfig` (必填) — 只有一个输出源配置
- `resolutionTechniques` (必需) -设置为“规则匹配”以进行基于规则的匹配
- `roleArn` (必填) — 用于执行工作流程的 IAM 角色 ARN
- `ruleConditionProperties` (必填) -规则条件列表和匹配规则的名称。

可选参数包括：

- `description`— 最多 255 个字符
- `incrementalRunConfig`— 增量运行类型配置
- `tags`— 最多 200 个键值对

5. (可选) 要使用增量处理而不是默认的标准 (批处理) 处理，请在请求正文中添加以下参数：

```
"incrementalRunConfig": {
  "incrementalRunType": "AUTOMATIC"
}
```

6. 发送 请求。
7. 如果成功，您将收到状态码 200 的响应和包含以下内容的 JSON 正文：

```
{
  "workflowArn": "string",
  "workflowName": "string",
  // Plus all configured workflow details
}
```

8. 如果呼叫失败，您可能会收到以下错误之一：
  - 400- `ConflictException` 如果工作流程名称已经存在
  - 400 — `ValidationException` 如果输入未通过验证
  - 402- `ExceedsLimitException` 如果超过账户限额
  - 403 — `AccessDeniedException` 如果你没有足够的访问权限

- 429 — ThrottlingException 如果请求被限制
- 500 — InternalServerErrorException 如果内部服务出现故障

## 创建基于机器学习的匹配工作流程

[基于机器学习的匹配](#)是一个预设过程，它会尝试匹配您输入的所有数据的记录。基于机器学习的匹配工作流程使您能够使用机器学习模型比较明文数据以找到广泛的匹配项。

### Note

机器学习模型不支持哈希数据的比较。

在您的数据中 AWS Entity Resolution 数据匹配服务 发现两条或多条记录之间存在匹配项时，它会分配：

- 与[匹配数据集中的记录的匹配 ID](#)
- 匹配[置信度](#)百分比。

您可以使用 ML-based 匹配工作流程的输出作为数据服务提供商匹配的输入，反之亦然，以实现您的特定目标。例如，您可以先运行 ML-based 匹配，在自己的记录中查找跨数据源的匹配项。如果子集未匹配，则可以运行[基于提供商服务的匹配](#)来查找其他匹配项。

### 先决条件

在创建 ML-based 匹配的工作流程之前，您必须：

1. 创建架构映射。有关更多信息，请参阅 [创建架构映射](#)。
2. 如果使用 Connect 客户档案作为输出目标，请确保配置了相应的权限。

要创建 ML-based 匹配的工作流程，请执行以下操作：

1. 登录 AWS 管理控制台 并打开 AWS Entity Resolution 数据匹配服务 控制台，网址为<https://console.aws.amazon.com/entityresolution/>。
2. 在左侧导航窗格的“工作流程”下，选择“匹配”。
3. 在匹配工作流程页面的右上角，选择创建匹配工作流程。
4. 对于“步骤 1：指定匹配的工作流程详细信息”，请执行以下操作：

- a. 输入匹配的工作流程名称和可选的描述。
- b. 对于数据输入 AWS 区域，选择AWS Glue 数据库、AWS Glue 表，然后选择相应的架构映射。

您最多可以添加 20 个数据输入。

- c. 默认情况下，“标准化数据”选项处于选中状态，以便在匹配之前对数据输入进行标准化。如果您不想对数据进行标准化处理，请取消选择“标准化数据”选项。

基于机器学习的匹配仅对[Name](#)、[Phone](#)和进行标准化。[电子邮件](#)

- d. 要指定服务访问权限，请选择一个选项并采取建议的操作。

Option	推荐操作
创建并使用新的服务角色	<ul style="list-style-type: none"> <li>• AWS Entity Resolution 数据匹配服务 使用此表所需的策略创建服务角色。</li> <li>• 默认服务角色名称为 <code>entityresolution-matching-workflow- &lt;timestamp&gt;</code> 。</li> <li>• 您必须拥有创建角色并附加策略的权限。</li> <li>• 如果您的输入数据已加密，请选择“此数据由 KMS 密钥加密”选项。然后，输入用于解密输入数据的密AWS KMS 钥。</li> </ul>

Option	推荐操作
使用现有服务角色	<p>1. 从下拉列表中选择一个现有服务角色名称。</p> <p>如果您有列出角色的权限，则会显示角色列表。</p> <p>如果您没有列出角色的权限，可以输入要使用的角色的 Amazon 资源名称 (ARN)。</p> <p>如果没有现有的服务角色，则使用现有服务角色选项不可用。</p> <p>2. 通过选择在 IAM 中查看外部链接来查看服务角色。</p> <p>默认情况下，AWS Entity Resolution 数据匹配服务 不会尝试更新现有角色策略以添加必要的权限。</p>

e. ( 可选 ) 要为资源启用标签，请选择添加新标签，然后输入密钥和值对。

f. 选择下一步。

5. 对于步骤 2：选择匹配技术：

a. 对于匹配方法，选择基于机器学习的匹配。

AWS Entity Resolution > Matching workflows > Create matching workflow

Step 1  
[Specify matching workflow details](#)

Step 2  
**Choose matching technique**

Step 3  
Specify data output

Step 4  
Review and create

### Choose matching technique [Info](#)

Specify how you want your data to be matched or choose a provider service.

**Matching method**

Rule-based matching  
Use customized rules to find exact matches.

Machine learning-based matching  
Use our machine learning model to help find a broader range of matches.

Provider services  
Use this option if you have a subscription to a preferred provider through AWS Data Exchange.

**Machine learning-based matching [Info](#)**

Your data will be evaluated against a set of rules defining the criteria to find exact matches. This can help find matches across your data that may be incomplete or may not look exactly the same.

**Processing cadence [Info](#)**  
Determine how often to run your matching workflow job. The first job runs after you create the matching workflow. [See pricing](#)

Manual  
Your matching workflow job is run on demand. Useful for bulk processing.

Automatic  
Your matching workflow job is run automatically when you add or update your data inputs. Useful for incremental updates. This option is available only for rule-based matching.

**Using hashed data may limit matching functionality**  
Rule-based matching is recommended when comparing hashed data. The machine learning model is unable to compare hashed data. [Learn more](#)

- b. 对于处理节奏，选择了“手动”选项。

此选项允许您按需运行工作流程以进行批量更新。

**Note**

基于机器学习的匹配工作流程不支持自动（增量）处理。

- c. 选择下一步。
6. 对于步骤 3：指定数据输出和格式：
- 对于数据输出目标和格式，选择数据输出的 Amazon S3 位置，以及数据格式是标准化数据还是原始数据。
  - 对于加密，如果您选择自定义加密设置，请输入 AWS KMS 密钥 ARN。
  - 查看系统生成的输出。
  - 对于数据输出，请决定要包含、隐藏或掩盖哪些字段，然后根据目标采取建议的操作。

您的目标	建议的选项
包括字段	将输出状态保持为“已包含”。
隐藏字段 ( 从输出中排除 )	选择“输出”字段，然后选择“隐藏”。
掩码字段	选择“输出”字段，然后选择“哈希输出”。
重置之前的设置	选择 重置。

e. 选择下一步。

7. 对于步骤 4：查看并创建：

- a. 查看您在之前的步骤中所做的选择，并在必要时进行编辑。
- b. 选择创建并运行。

将出现一条消息，表示匹配的工作流程已创建且作业已启动。

8. 在匹配的工作流程详细信息页面的指标选项卡上，在“上次作业指标”下查看以下内容：

- 作业 ID。
- 匹配 workflow 作业的状态：已排队、进行中、已完成、失败
- 工作流作业的完成时间。
- 已处理的记录数。
- 未处理的记录数。
- 生成的唯一匹配 ID。
- 输入记录的数量。

您还可以查看任务历史记录下先前运行过的匹配 workflow 作业的作业指标。

9. 匹配的工作流程任务完成 ( 状态为已完成 ) 后，您可以转到数据输出选项卡，然后选择您的 Amazon S3 位置以查看结果。

10. ( 仅限手动处理类型 ) 如果您创建了手动处理类型的基于机器学习的匹配 workflow，则可以在匹配 workflow 详细信息页面上选择“运行 workflow”，随时运行匹配 workflow。

# 创建基于提供商服务的匹配工作流程

[基于提供商服务的匹配](#)使您能够将已知标识符与首选数据服务提供商进行匹配。

AWS Entity Resolution 数据匹配服务 目前支持以下数据提供商服务：

- LiveRamp
- TransUnion
- 统一身份证 2.0

有关支持的提供商服务的更多信息，请参阅[准备第三方输入数据](#)。

您可以对这些提供商使用公开订阅，AWS Data Exchange 也可以直接与数据提供商协商私人报价。有关创建新订阅或重复使用提供商服务的现有订阅的更多信息，请参阅[步骤 1：在上订阅提供商服务 AWS Data Exchange](#)。

以下各节介绍如何创建基于提供者的匹配工作流程。

## 主题

- [使用创建匹配的工作流程 LiveRamp](#)
- [使用创建匹配的工作流程 TransUnion](#)
- [使用 UID 2.0 创建匹配的工作流程](#)

## 使用创建匹配的工作流程 LiveRamp

该 LiveRamp 服务提供了一个名为 rampID 的标识符。RamPid是需求方平台中最常用的ID之一，用于为广告活动吸引受众。使用与匹配的工作流程 LiveRamp，您可以将经过哈希处理的电子邮件地址解析为 rampID。

### Note

AWS Entity Resolution 数据匹配服务 支持 PII-based rampID 赋值。

## 先决条件

在使用创建匹配的工作流程之前 LiveRamp，您必须：

1. 创建架构映射。有关更多信息，请参阅 [创建架构映射](#)。
2. 订阅该 LiveRamp 服务
3. 为 Amazon S3 数据暂存存储桶配置适当的权限，以便在其中临时写入匹配的工作流程输出

在使用创建 ID 映射工作流程之前 LiveRamp，请向 S3 数据暂存存储桶添加以下权限。

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "AWS": "arn:aws:iam::715724997226:root"
      },
      "Action": [
        "s3:PutObject",
        "s3:GetObject",
        "s3:GetObjectVersion",
        "s3:DeleteObject"
      ],
      "Resource": [
        "arn:aws:s3:::<staging-bucket>",
        "arn:aws:s3:::<staging-bucket>/*"
      ]
    },
    {
      "Effect": "Allow",
      "Principal": {
        "AWS": "arn:aws:iam::715724997226:root"
      },
      "Action": [
        "s3:ListBucket",
        "s3:GetBucketLocation",
        "s3:GetBucketPolicy",
        "s3:ListBucketVersions",
        "s3:GetBucketAcl"
      ],
      "Resource": [
```

```

        "arn:aws:s3:::<staging-bucket>",
        "arn:aws:s3:::<staging-bucket>/*"
    ]
}
]
}

```

将每个 *<user input placeholder>* 替换为您自己的信息。

*staging-bucket*

Amazon S3 存储桶，用于在运行基于提供商服务的工作流程时临时存储您的数据。

要创建匹配的工作流程，请执行 LiveRamp 以下操作：

1. 登录 AWS 管理控制台 并打开 AWS Entity Resolution 数据匹配服务 控制台，网址为 <https://console.aws.amazon.com/entityresolution/>。
2. 在左侧导航窗格的“工作流程”下，选择“匹配”。
3. 在匹配工作流程页面的右上角，选择创建匹配工作流程。
4. 对于“步骤 1：指定匹配的工作流程详细信息”，请执行以下操作：
  - a. 输入匹配的工作流程名称和可选的描述。
  - b. 对于数据输入 AWS 区域，选择 AWS Glue 数据库、AWS Glue 表，然后选择相应的架构映射。  
  
您最多可以添加 20 个数据输入。
  - c. 默认情况下，“标准化数据”选项处于选中状态，以便在匹配之前对数据输入进行标准化。

#### Note

创建架构映射中仅支持以下场景的标准化：

- 如果将以下“名称”子类型分组：名字、中间名、姓氏。
- 如果将以下地址子类型分组：街道地址 1、街道地址 2：街道地址 3 名称、城市名称、州、国家、邮政编码。
- 如果将以下电话子类型分组：电话号码、电话国家/地区代码。

如果您使用的是仅限电子邮件的解析流程，请取消选择“标准化数据”选项，因为只有经过哈希处理的电子邮件才用于输入数据。

- d. 要指定服务访问权限，请选择一个选项并采取建议的操作。

Option	推荐操作
创建并使用新的服务角色	<ul style="list-style-type: none"> <li>• AWS Entity Resolution 数据匹配服务 使用此表所需的策略创建服务角色。</li> <li>• 默认服务角色名称为 <code>entityresolution-matching-workflow-&lt;timestamp&gt;</code> 。</li> <li>• 您必须拥有创建角色并附加策略的权限。</li> <li>• 如果您的输入数据已加密，请选择“此数据由 KMS 密钥加密”选项。然后，输入用于解密输入数据的密AWS KMS 钥。</li> </ul>
使用现有服务角色	<ol style="list-style-type: none"> <li>1. 从下拉列表中选择一个现有服务角色名称。           <p>如果您有列出角色的权限，则会显示角色列表。</p> <p>如果您没有列出角色的权限，可以输入要使用的角色的 Amazon 资源名称 (ARN)。</p> <p>如果没有现有的服务角色，则使用现有服务角色选项不可用。</p> </li> <li>2. 通过选择在 IAM 中查看外部链接来查看服务角色。           <p>默认情况下，AWS Entity Resolution 数据匹配服务 不会尝试更新现有角色策略以添加必要的权限。</p> </li> </ol>

- e. ( 可选 ) 要为资源启用标签，请选择添加新标签，然后输入密钥和值对。

- f. 选择下一步。
5. 对于步骤 2：选择匹配技术：
    - a. 在“匹配方法”中，选择“提供者服务”。
    - b. 对于提供商服务，请选择LiveRamp。

**Note**

确保您的数据输入文件格式和标准化符合提供商服务的指南。

有关匹配工作流程的输入文件格式指南的更多信息，请参阅 LiveRamp 文档中的[通过 ADX 执行身份解析](#)。

- c. 对于LiveRamp 产品，请从下拉列表中选择产品。

**Matching method**

**Rule-based matching**  
Use customized rules to find exact matches.


**Machine learning-based matching**  
Use our machine learning model to help find a broader range of matches.

**Provider services**  
Use this option if you have a subscription to a preferred provider through AWS Data Exchange.

**Provider services** [Info](#)

You must have a provider agreement to use a provider service. Your data will be matched with a set of inputs defined by your preferred provider. Some information may be required and shared between you and your provider service.

**LiveRamp**  
  
**/LiveRamp**

**TransUnion**  
  
**TransUnion** 

**Unified ID 2.0**  
  
**Unified ID 2.0**

**LiveRamp products**  
Choose from available products from LiveRamp.

Choose product ▲

Assignment Email

Assignment PII

Cancel Previous Next

**Note**

如果您选择赋值 PII，则在执行实体解析时必须至少提供一个非标识符列。例如，性别。

- d. 要进行LiveRamp 配置，请输入客户端 ID 管理器 ARN 和客户机密管理器 AR N。

**LiveRamp configuration**  
These are the required fields to use the LiveRamp service.

**Client ID manager ARN**  
Enter the Client ID manager ARN provided by LiveRamp.  
arn:aws:secretsmanager:us-east-1:██████████:secret:██████████  
83 of 2,048 characters.

**Client secret manager ARN**  
Enter the Client secret manager ARN provided by LiveRamp.  
arn:aws:secretsmanager:us-east-1:██████████:secret:██████████  
87 of 2,048 characters.

**Data staging** [Info](#)  
Choose the Amazon S3 location for temporarily storing your data while it processes. Your information will not be saved permanently.

**Amazon S3 location**  
s3://██████████  
View [Browse S3](#)

Cancel Previous **Next**

- e. 对于数据暂存，请选择 Amazon S3 位置，以便在处理数据时临时存储数据。

您必须拥有访问 Amazon S3 数据暂存位置的权限。有关更多信息，请参阅 [为创建工作流程工作角色 AWS Entity Resolution 数据匹配服务](#)。

- f. 选择下一步。

6. 对于步骤 3：指定数据输出：

- 对于数据输出目标和格式，选择数据输出的 Amazon S3 位置，以及数据格式是标准化数据还是原始数据。
- 对于加密，如果您选择自定义加密设置，请输入 AWS KMS 密钥 ARN。
- 查看 LiveRamp 生成的输出。

这是由生成的其他信息 LiveRamp。

- d. 对于数据输出，请决定要包含、隐藏或掩盖哪些字段，然后根据目标采取建议的操作。

**Note**

如果您已选择 LiveRamp，则由于 LiveRamp 隐私过滤器会删除个人身份信息 (PII)，某些字段将显示“不可用”的输出状态。

您的目标	建议的选项
包括字段	将输出状态保持为“已包含”。
隐藏字段（从输出中排除）	选择“输出”字段，然后选择“隐藏”。
掩码字段	选择“输出”字段，然后选择“哈希输出”。
重置之前的设置	选择 重置。

AWS Entity Resolution > ID mapping workflows > Create ID mapping workflow

Step 1 Specify ID mapping workflow details  
Step 2 Specify source and target  
Step 3 - optional Specify data output location  
Step 4 Review and create

### Specify data output location - optional Info

Choose your S3 location to write your data output.

**Data output destination Info**  
Choose the Amazon S3 location for the data output.

Amazon S3 location

**Encryption - optional Info**  
Your data is encrypted by default with a key that AWS owns and manages for you. To specify a different key, customize your encryption settings.

Customize encryption settings  
Specify an AWS KMS key to customize your encryption settings.

▼ **LiveRamp generated output (2)**  
Additional information generated by LiveRamp.

Output field	Description
RAMPID	LiveRamp's universal identifier that is tied to devices in the LiveRamp Identity Graph
TRANSCODED_IDENTIFIER	LiveRamp's universal identifier that is tied to devices in the LiveRamp Identity Graph

e. 选择下一步。

7. 对于步骤 4：查看并创建：

- 查看您在之前的步骤中所做的选择，并在必要时进行编辑。
- 选择创建并运行。

将出现一条消息，表示匹配的工作流程已创建且作业已启动。

8. 在匹配的工作流程详细信息页面的指标选项卡上，在“上次作业指标”下查看以下内容：

- 作业 ID。
- 匹配 workflow 作业的状态：已排队、进行中、已完成、失败

- workflow 作业的完成时间。
- 已处理的记录数。
- 未处理的记录数。
- 生成的唯一匹配 ID。
- 输入记录的数量。

您还可以查看任务历史记录下先前运行过的匹配 workflow 作业的作业指标。

9. 匹配的工作流程任务完成（状态为已完成）后，您可以转到数据输出选项卡，然后选择您的 Amazon S3 位置以查看结果。

## 使用创建匹配的工作流程 TransUnion

如果您订阅了该 TransUnion 服务，则可以通过使用 TransUnion 个人和家庭电子密钥以及 200 多个数据属性链接、匹配和增强存储在不同渠道上的客户相关记录来提高对客户理解。

该 TransUnion 服务提供被称为 TransUnion 个人和家庭身份证的标识符。TransUnion 提供已知标识符（例如姓名、地址、电话号码和电子邮件地址）的 ID 分配（也称为编码）。

### 先决条件

在使用创建匹配的工作流程之前 LiveRamp，您必须：

1. 创建架构映射。有关更多信息，请参阅 [创建架构映射](#)。
2. 订阅该 TransUnion 服务
3. 为 Amazon S3 数据暂存存储桶配置适当的权限，以便在其中临时写入匹配的工作流程输出

在使用创建匹配的工作流程之前 TransUnion，请向 S3 数据暂存存储桶添加以下权限。

### JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
```

```
        "AWS": "arn:aws:iam::381491956555:root"
    },
    "Action": [
        "s3:PutObject",
        "s3:GetObject",
        "s3:GetObjectVersion",
        "s3:DeleteObject"
    ],
    "Resource": [
        "arn:aws:s3:::<staging-bucket>",
        "arn:aws:s3:::<staging-bucket>/*"
    ]
},
{
    "Effect": "Allow",
    "Principal": {
        "AWS": "arn:aws:iam::381491956555:root"
    },
    "Action": [
        "s3:ListBucket",
        "s3:GetBucketLocation",
        "s3:GetBucketPolicy",
        "s3:ListBucketVersions",
        "s3:GetBucketAcl"
    ],
    "Resource": [
        "arn:aws:s3:::<staging-bucket>",
        "arn:aws:s3:::<staging-bucket>/*"
    ]
}
]
```

将每个 *<user input placeholder>* 替换为您自己的信息。

*staging-bucket*

Amazon S3 存储桶，用于在运行基于提供商服务的工作流程时临时存储您的数据。

要创建匹配的工作流程，请执行 TransUnion 以下操作：

1. 登录 AWS 管理控制台 并打开 AWS Entity Resolution 数据匹配服务 控制台，网址为 <https://console.aws.amazon.com/entityresolution/>。
2. 在左侧导航窗格的“工作流程”下，选择“匹配”。
3. 在匹配工作流程页面的右上角，选择创建匹配工作流程。
4. 对于“步骤 1：指定匹配的工作流程详细信息”，请执行以下操作：
  - a. 输入匹配的工作流程名称和可选的描述。
  - b. 对于数据输入 AWS 区域，选择 AWS Glue 数据库、AWS Glue 表，然后选择相应的架构映射。

您最多可以添加 20 个数据输入。

- c. 默认情况下，“标准化数据”选项处于选中状态，以便在匹配之前对数据输入进行标准化。如果您不想对数据进行标准化处理，请取消选择“标准化数据”选项。

#### Note

创建架构映射中仅支持以下场景的标准化：

- 如果将以下“名称”子类型分组：名字、中间名、姓氏。
- 如果将以下地址子类型分组：街道地址 1、街道地址 2：街道地址 3 名称、城市名称、州、国家、邮政编码。
- 如果将以下电话子类型分组：电话号码、电话国家/地区代码。

- d. 要指定服务访问权限，请选择一个选项并采取建议的操作。

Option	推荐操作
创建并使用新的服务角色	<ul style="list-style-type: none"> <li>• AWS Entity Resolution 数据匹配服务 使用此表所需的策略创建服务角色。</li> <li>• 默认服务角色名称为 <code>entityresolution-matching-workflow-&lt;timestamp&gt;</code>。</li> <li>• 您必须拥有创建角色并附加策略的权限。</li> </ul>

Option	推荐操作
使用现有服务角色	<ul style="list-style-type: none"> <li>• 如果您的输入数据已加密，请选择“此数据由 KMS 密钥加密”选项。然后，输入用于解密输入数据的密AWS KMS 钥。</li> </ul> <ol style="list-style-type: none"> <li>1. 从下拉列表中选择一个现有服务角色名称。  如果您有列出角色的权限，则会显示角色列表。  如果您没有列出角色的权限，可以输入要使用的角色的 Amazon 资源名称 (ARN)。  如果没有现有的服务角色，则使用现有服务角色选项不可用。</li> <li>2. 通过选择在 IAM 中查看外部链接来查看服务角色。  默认情况下，AWS Entity Resolution 数据匹配服务 不会尝试更新现有角色策略以添加必要的权限。</li> </ol>

e. ( 可选 ) 要为资源启用标签，请选择添加新标签，然后输入密钥和值对。

f. 选择下一步。

5. 对于步骤 2：选择匹配技术：

a. 在“匹配方法”中，选择“提供者服务”。

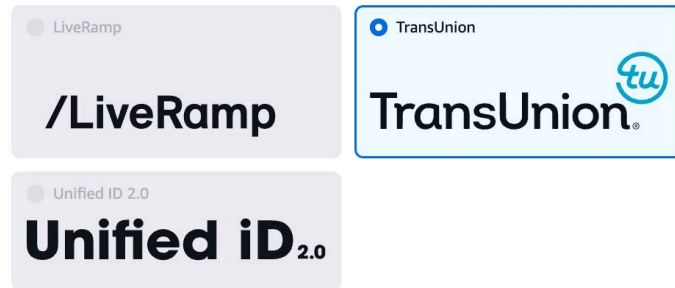
b. 对于提供商服务，请选择TransUnion。

#### Note

确保您的数据输入文件格式和标准化符合提供商服务的指南。

**Provider services** [Info](#)

You must have a provider agreement to use a provider service. Your data will be matched with a set of inputs defined by your preferred provider. Some information may be required and shared between you and your provider service.

**Access to TransUnion provider subscription**

✔ Subscribed

ⓘ To ensure a successful workflow run, your data input file format and normalization must be aligned with the provider service's guidelines. [Learn more](#)

- c. 对于数据暂存，请选择 Amazon S3 位置，以便在处理数据时临时存储数据。

您必须拥有访问 Amazon S3 数据暂存位置的权限。有关更多信息，请参阅 [the section called “创建工作流工作角色”](#)。

6. 选择下一步。

7. 对于步骤 3：指定数据输出：

- a. 对于数据输出目标和格式，选择数据输出的 Amazon S3 位置，以及数据格式是标准化数据还是原始数据。
- b. 对于加密，如果您选择自定义加密设置，请输入 AWS KMS 密钥 ARN。
- c. 查看 TransUnion 生成的输出。

这是由生成的其他信息 TransUnion。

- d. 对于数据输出，请决定要包含、隐藏或掩盖哪些字段，然后根据目标采取建议的操作。

您的目标	建议的选项
包括字段	将输出状态保持为“已包含”。
隐藏字段（从输出中排除）	选择“输出”字段，然后选择“隐藏”。
掩码字段	选择“输出”字段，然后选择“哈希输出”。
重置之前的设置	选择 重置。

- e. 对于系统生成的输出，请查看包含的所有字段。
  - f. 选择下一步。
8. 对于步骤 4：查看并创建：
- a. 查看您在之前的步骤中所做的选择，并在必要时进行编辑。
  - b. 选择创建并运行。

将出现一条消息，表示匹配的工作流程已创建且作业已启动。

9. 在匹配的工作流程详细信息页面的指标选项卡上，在“上次作业指标”下查看以下内容：
- 作业 ID。
  - 匹配工作流作业的状态：已排队、进行中、已完成、失败
  - 工作流作业的完成时间。
  - 已处理的记录数。
  - 未处理的记录数。
  - 生成的唯一匹配 ID。
  - 输入记录的数量。

您还可以查看任务历史记录下先前运行过的匹配工作流作业的作业指标。

10. 匹配的工作流程任务完成（状态为已完成）后，您可以转到数据输出选项卡，然后选择您的 Amazon S3 位置以查看结果。

## 使用 UID 2.0 创建匹配的工作流程

如果您订阅了 Unified ID 2.0 服务，则可以激活具有确定性身份的广告活动，并依靠与广告生态系统中许多 UID2-enabled 参与者的互操作性。有关更多信息，请参阅 [Unified ID 2.0 概述](#)。

Unified ID 2.0 服务提供原始的 UID 2，用于在 The Trade Desk 平台中制作广告活动。UID 2.0 是使用开源框架生成的。

在一个工作流程中，您可以使用 **Email Address** 或 **Phone number** 生成原始 UID2，但不能同时使用两者。如果两者都存在于架构映射中，则工作流程将选择 **Email Address** 然后 **Phone number** 将是直通字段。要同时支持这两者，请创建一个新的架构映射 **Phone number**，其中已映射但 **Email Address** 未映射。然后，使用这个新的架构映射创建第二个工作流程。

**Note**

未加工的 UID2 是通过添加盐桶中的盐来创建的，这些盐桶大约每年轮换一次，从而使原始 UID2 也随之旋转。因此，建议您每天刷新原始 UID2。有关更多信息，请参阅 <https://unifiedid.com/docs/getting-started/gs-faqs#how-often-should-uid2s-be-refreshed-for> **增量更新**。

**先决条件**

在使用 UID 2.0 创建匹配的工作流程之前，您必须：

1. 创建架构映射。有关更多信息，请参阅 [创建架构映射](#)。
2. 订阅 UID 2.0 服务

要使用 UID 2.0 创建匹配的工作流程，请执行以下操作：

1. 登录 AWS 管理控制台 并打开 AWS Entity Resolution 数据匹配服务 控制台，网址为 <https://console.aws.amazon.com/entityresolution/>。
2. 在左侧导航窗格的“工作流程”下，选择“匹配”。
3. 在匹配工作流程页面的右上角，选择创建匹配工作流程。
4. 对于“步骤 1：指定匹配的工作流程详细信息”，请执行以下操作：
  - a. 输入匹配的工作流程名称和可选的描述。
  - b. 对于数据输入 AWS 区域，选择 AWS Glue 数据库、AWS Glue 表，然后选择相应的架构映射。

您最多可以添加 20 个数据输入。

- c. 保持“标准化数据”选项处于选中状态，以便在匹配之前对数据输入 (**Email Address** 或 **Phone number**) 进行标准化。

有关 **Email Address** 标准化的更多信息，请参阅 UID 2.0 文档中的 [电子邮件地址标准化](#)。

有关 **Phone number** 标准化的更多信息，请参阅 UID 2.0 文档中的 [电话号码标准化](#)。

- d. 要指定服务访问权限，请选择一个选项并采取建议的操作。

Option	推荐操作
创建并使用新的服务角色	<ul style="list-style-type: none"> <li>• AWS Entity Resolution 数据匹配服务 使用此表所需的策略创建服务角色。</li> <li>• 默认服务角色名称为 <code>entityresolution-matching-workflow-&lt;timestamp&gt;</code> 。</li> <li>• 您必须拥有创建角色并附加策略的权限。</li> <li>• 如果您的输入数据已加密，请选择“此数据由 KMS 密钥加密”选项。然后，输入用于解密输入数据的密AWS KMS 钥。</li> </ul>
使用现有服务角色	<ol style="list-style-type: none"> <li>1. 从下拉列表中选择一个现有服务角色名称。 <p>如果您有列出角色的权限，则会显示角色列表。</p> <p>如果您没有列出角色的权限，可以输入要使用的角色的 Amazon 资源名称 (ARN)。</p> <p>如果没有现有的服务角色，则使用现有服务角色选项不可用。</p> </li> <li>2. 通过选择在 IAM 中查看外部链接来查看服务角色。 <p>默认情况下，AWS Entity Resolution 数据匹配服务 不会尝试更新现有角色策略以添加必要的权限。</p> </li> </ol>

e. ( 可选 ) 要为资源启用标签，请选择添加新标签，然后输入密钥和值对。

f. 选择下一步。

5. 对于步骤 2：选择匹配技术：

a. 在“匹配方法”中，选择“提供者服务”。

b. 对于提供商服务，请选择统一 ID 2.0。

The screenshot shows the 'Choose matching technique' step in the AWS Entity Resolution console. The breadcrumb trail is 'AWS Entity Resolution > Matching workflows > Create matching workflow'. The left sidebar shows four steps: Step 1 (Specify matching workflow details), Step 2 (Choose matching technique), Step 3 (Specify data output), and Step 4 (Review and create). The main content area is titled 'Choose matching technique' and includes a sub-header 'Matching method'. Three options are available: 'Rule-based matching' (unselected), 'Machine learning-based matching' (unselected), and 'Provider services' (selected). Below 'Provider services', there are two options: 'LiveRamp' and 'TransUnion' (both unselected), and 'Unified ID 2.0' (selected). A note states 'Access to Unified ID 2.0 provider subscription' with a 'Subscribed' status. At the bottom, there are 'Cancel', 'Previous', and 'Next' buttons.

c. 选择下一步。

6. 对于步骤 3：指定数据输出：

- 对于数据输出目标和格式，选择数据输出的 Amazon S3 位置，以及数据格式是标准化数据还是原始数据。
- 对于加密，如果您选择自定义加密设置，请输入 AWS KMS 密钥 ARN。
- 查看 Unified ID 2.0 生成的输出。

这是 UID 2.0 生成的所有其他信息的列表

- 对于数据输出，请决定要包含、隐藏或掩盖哪些字段，然后根据目标采取建议的操作。

您的目标	建议的选项
包括字段	将输出状态保持为“已包含”。
隐藏字段（从输出中排除）	选择“输出”字段，然后选择“隐藏”。

您的目标	建议的选项
掩码字段	选择“输出”字段，然后选择“哈希输出”。
重置之前的设置	选择 重置。

- e. 对于系统生成的输出，请查看包含的所有字段。
  - f. 选择下一步。
7. 对于步骤 4：查看并创建：
    - a. 查看您在之前的步骤中所做的选择，并在必要时进行编辑。
    - b. 选择创建并运行。

将出现一条消息，表示匹配的工作流程已创建且作业已启动。

8. 在匹配的工作流程详细信息页面的指标选项卡上，在“上次作业指标”下查看以下内容：
  - 作业 ID。
  - 匹配 workflow 作业的状态：已排队、进行中、已完成、失败
  - 工作流作业的完成时间。
  - 已处理的记录数。
  - 未处理的记录数。
  - 生成的唯一匹配 ID。
  - 输入记录的数量。

您还可以查看任务历史记录下先前运行过的匹配 workflow 作业的作业指标。

9. 匹配的工作流程任务完成（状态为已完成）后，您可以转到数据输出选项卡，然后选择您的 Amazon S3 位置以查看结果。

## 编辑匹配的工作流程

通过编辑匹配的工作流程，您可以使实体解析流程保持最新状态，并随着时间的推移响应组织不断变化的需求。您可能需要调整匹配标准、技术或数据输出，以提高实体解析过程的准确性和效率。如果您发现当前 workflow 的结果存在问题或错误，则对其进行编辑可以帮助您诊断和解决这些问题。

要编辑匹配的工作流程，请执行以下操作：

1. 登录 AWS 管理控制台 并打开 AWS Entity Resolution 数据匹配服务 控制台，网址为 <https://console.aws.amazon.com/entityresolution/>。
2. 在左侧导航窗格的“工作流程”下，选择“匹配”。
3. 选择匹配的工作流程。
4. 在匹配的工作流程详细信息页面的右上角，选择编辑工作流程。
5. 在“指定匹配的工作流程详细信息”页面上，进行任何必要的更改，然后选择“下一步”。
6. 在“选择匹配技术”页面上，进行必要的更改，然后选择“下一步”。

#### Important

您可以将处理节奏从“手动”更改为“自动”，但在将其更改为“自动”之后，则无法将其更改回“手动”。

如果“处理节奏”已设置为“自动”，则无法将其更改为“手动”。

7. 在“指定数据输出”页面上，进行必要的更改，然后选择“下一步”。
8. 在“查看并保存”页面上，进行必要的更改，然后选择“保存”。

## 删除匹配的工作流程

如果不再使用匹配的工作流程或已过时，删除它可以帮助您的工作空间保持井井有条和整洁。如果你开发了一个新的、经过改进的工作流程来取代旧的工作流程，那么删除旧的工作流程可以帮助确保你只使用最新的流程。

要删除匹配的工作流程，请执行以下操作：

1. 登录 AWS 管理控制台 并打开 AWS Entity Resolution 数据匹配服务 控制台，网址为 <https://console.aws.amazon.com/entityresolution/>。
2. 在左侧导航窗格的“工作流程”下，选择“匹配”。
3. 选择匹配的工作流程。
4. 在匹配的工作流程详细信息页面的右上角，选择删除。
5. 确认删除，然后选择删除。

## 为基于规则的匹配工作流程修改或生成匹配 ID

匹配 ID 是在运行匹配工作流程后，由每个匹配的记录集生成 AWS Entity Resolution 数据匹配服务 并应用于这些记录集的标识符。这是输出中包含的匹配工作流程元数据的一部分。

当您需要更新现有客户的记录或向数据集中添加新客户时，可以使用 AWS Entity Resolution 数据匹配服务 控制台或 GenerateMatchID API。修改现有的匹配 ID 有助于在更新客户信息时保持一致性，而在向系统中添加以前未识别的客户时，则需要生成新的匹配 ID。

### Note

无论您使用控制台还是 API，都需要支付额外费用。您选择的处理类型会影响操作的准确性和响应时间。

### Important

如果您在任务进行期间撤消了对 S3 存储桶的 AWS Entity Resolution 数据匹配服务 权限，则仍 AWS Entity Resolution 数据匹配服务 会处理向 S3 输出结果并收取费用，但无法将结果传送到您的存储桶。为避免出现此问题，请确保在开始任务之前 AWS Entity Resolution 数据匹配服务 具有写入您的 S3 存储桶的正确权限。如果在处理过程中权限被撤销，则在您恢复正确的存储桶权限后，会在任务完成后最长 30 天内 AWS Entity Resolution 数据匹配服务 尝试重新交付结果。

以下步骤将指导您完成查找或生成匹配 ID、选择处理类型以及查看结果的过程。

### Console

#### 使用控制台修改或生成匹配 ID

1. 登录 AWS 管理控制台 并打开 AWS Entity Resolution 数据匹配服务 控制台，网址为 <https://console.aws.amazon.com/entityresolution/>。
2. 在左侧导航窗格的“工作流程”下，选择“匹配”。
3. 选择已处理的基于规则的匹配工作流（Job 状态为“已完成”）。
4. 在匹配工作流程详细信息页面上，选择匹配 ID 选项卡。
5. 选择“修改”或“生成匹配 ID”。

**Note**

“修改或生成匹配 ID” 选项仅适用于使用自动处理节奏的匹配工作流程。如果您选择了手动处理节奏，则此选项将显示为非活动状态。要使用此选项，请编辑您的工作流程以使用自动处理节奏。有关编辑工作流程的更多信息，请参阅[编辑匹配的工作流程](#)。

## 6. 从下拉列表中选择表。AWS Glue

如果工作流程中只有一个 AWS Glue 表，则默认情况下会将其选中。

## 7. 选择处理类型。

- 一致 — 您可以查找现有的匹配 ID 或立即生成并保存新的匹配 ID。此选项具有最高的精度和较慢的响应时间。
- 背景 (如 API EVENTUAL 所示) — 您可以查找现有的匹配 ID 或立即生成新的匹配 ID。更新的记录保存在后台。此选项具有快速的初始响应，以后将在 S3 中提供完整的结果。
- 快速生成 ID (如 API EVENTUAL\_NO\_LOOKUP 所示) — 您可以创建新的匹配 ID，而无需查找现有的匹配 ID。更新的记录保存在后台。此选项的响应速度最快。建议仅用于唯一记录。

## 8. 对于记录属性，

- a. 输入唯一 ID 的值。
- b. 根据工作流程中配置的规则，为每个匹配键输入一个值，该值将与现有记录相匹配。

## 9. 选择“查找匹配 ID”并保存记录。

将显示一条成功消息，说明已找到匹配 ID 或生成了新的匹配 ID 并保存了记录。

## 10. 在成功消息中查看相应的匹配 ID 和保存到匹配工作流程的关联规则。

## 11. (可选) 要复制匹配 ID，请选择“复制”。

## API

## 使用 API 修改或生成匹配 ID

**Note**

要成功调用此 API，您必须先使用该 API 成功运行基于规则的匹配工作流程。[StartMatchingJob](#)

有关支持的编程语言的完整列表，请参阅 [GenerateMatchID](#) 的“另请参阅”部分。

1. 打开终端或命令提示符发出 API 请求。
2. 向以下端点创建 POST 请求：

```
/matchingworkflows/workflowName/generateMatches
```

3. 在请求标头中，将设置 Content-type 为 application/json。
4. 在请求 URI 中，指定您的 workflowName。

必 workflowName 须：

- 长度介于 1 到 255 个字符之间
  - 匹配模式 [a-z A-Z \_0-9-] \*
5. 对于请求正文，请提供以下 JSON：

```
{
  "processingType": "string",
  "records": [
    {
      "inputSourceARN": "string",
      "recordAttributeMap": {
        "string": "string"
      },
      "uniqueId": "string"
    }
  ]
}
```

其中：

- processingType ( 可选 ) -默认为CONSISTENT。选择以下值之一：
    - CONSISTENT-以较慢的响应时间获得最高的精度
    - EVENTUAL-通过后台处理实现更快的初始响应
    - EVENTUAL\_NO\_LOOKUP-当已知记录是唯一时，可实现最快的响应
  - records ( 必填 ) -仅包含一个记录对象的数组
6. 发送 请求。

如果成功，您将收到状态码 200 的响应和包含以下内容的 JSON 正文：

```
{
  "failedRecords": [
    {
      "errorMessage": "string",
      "inputSourceARN": "string",
      "uniqueId": "string"
    }
  ],
  "matchGroups": [
    {
      "matchId": "string",
      "matchRule": "string",
      "records": [
        {
          "inputSourceARN": "string",
          "recordId": "string"
        }
      ]
    }
  ]
}
```

如果呼叫失败，您可能会收到以下错误之一：

- 403- AccessDeniedException 如果你没有足够的访问权限
- 404- ResourceNotFoundException 如果找不到资源
- 429- ThrottlingException 如果请求被限制
- 400- ValidationException 如果输入未通过验证
- 500- InternalServerErrorException 如果内部服务出现故障

## 为基于规则的匹配工作流程查找匹配 ID

完成基于规则的匹配工作流程后，您可以检索每条已处理记录的匹配 ID 和关联规则。此信息可帮助您了解记录是如何匹配的，以及应用了哪些规则。以下过程演示如何使用 AWS Entity Resolution 数据匹配服务 控制台或 GetMatchID API 访问这些数据。

## Console

### 使用控制台查找比赛 ID

1. 登录 AWS 管理控制台 并打开 AWS Entity Resolution 数据匹配服务 控制台，网址为 <https://console.aws.amazon.com/entityresolution/>。
2. 在左侧导航窗格的“工作流程”下，选择“匹配”。
3. 选择已处理的基于规则的匹配工作流（Job 状态为“已完成”）。
4. 在匹配工作流详细信息页面上，选择匹配 ID 选项卡。
5. 选择“查找匹配 ID”。

#### Note

“查找匹配 ID”选项仅适用于使用自动处理节奏的匹配工作流。如果您选择了手动处理节奏，则此选项将显示为非活动状态。要使用此选项，请编辑您的工作流以使用自动处理节奏。有关编辑工作流的更多信息，请参阅[编辑匹配的工作流程](#)。

6. 请执行以下操作之一：

如果...	操作...
只有一个架构映射与此工作流相关联。	查看默认情况下选择的架构映射。
有多个架构映射与此工作流相关联。	从下拉列表中选择架构映射。

7. 在“记录属性”中，输入现有匹配键的值，以查找每条现有记录。

#### Tip

输入尽可能多的值以帮助找到匹配 ID。

8. 默认情况下，“标准化数据”选项处于选中状态，以便在匹配之前对数据输入进行标准化。如果您不想对数据进行标准化处理，请取消选择“标准化数据”选项。
9. 如果要查看匹配规则，请展开查看匹配规则。
10. 选择查找。

将显示一条成功消息，说明已找到匹配 ID。

11. 查看相应的匹配 ID 和找到的关联规则。

## API

### 使用 API 查找匹配 ID

#### Note

要成功调用此 API，您必须先使用该 API 成功运行基于规则的匹配工作流程。StartMatchingJob

有关支持的编程语言的完整列表，请参阅 [GetMatchID API](#) 的“另请参阅”部分。

1. 打开终端或命令提示符发出 API 请求。
2. 向以下端点创建 POST 请求：

```
/matchingworkflows/workflowName/matches
```

3. 在请求标头中，将设置 Content-type 为 application/json。
4. 在请求 URI 中，指定您的 workflowName。

必 workflowName 须：

- 长度介于 1 到 255 个字符之间
- 匹配模式 [a-z A-Z \_0-9-]\*

5. 对于请求正文，请提供以下 JSON：

```
{
  "applyNormalization": boolean,
  "record": {
    "string" : "string"
  }
}
```

其中：

applyNormalization ( 可选 ) - 设置为 true 以标准化架构中定义的属性

record ( 必填 ) - 要获取匹配 ID 的记录

6. 发送 请求。

如果成功，您将收到状态码 200 的响应和包含以下内容的 JSON 正文：

```
{
  "matchId": "string",
  "matchRule": "string"
}
```

matchId 是这组匹配记录的唯一标识符，它 matchRule 表示该记录与哪条规则匹配。

如果呼叫失败，您可能会收到以下错误之一：

- 403- AccessDeniedException 如果你没有足够的访问权限
- 404- ResourceNotFoundException 如果找不到资源
- 429- ThrottlingException 如果请求被限制
- 400- ValidationException 如果输入未通过验证
- 500- InternalServerErrorException 如果内部服务出现故障

## 从基于规则或 ML-based 匹配的工作流程中删除记录

如果您需要遵守数据管理法规，则可以从基于规则的工作流程或 ML-based 匹配的工作流程中删除记录。

从基于规则或 ML-based 匹配的工作流程中删除记录

1. 登录 AWS 管理控制台 并打开 AWS Entity Resolution 数据匹配服务 控制台，网址为 <https://console.aws.amazon.com/entityresolution/>。
2. 在左侧导航窗格的“工作流程”下，选择“匹配”。
3. 选择基于规则的工作流程或 ML-based 匹配的工作流程。
4. 在匹配的工作流程详细信息页面上，从“操作”下拉列表中选择“删除唯一 ID”。
5. 在“唯一 ID”部分输入要删除的唯一 ID。

您最多可以输入 10 个唯一 ID。

6. 指定要从中删除唯一 ID 的输入源。

如果工作流程只有一个输入源，则默认情况下会列出输入源。

如果您只指定一个输入源，则其他输入源中的唯一 ID 不会受到影响。

## 7. 选择删除唯一 ID。

# 匹配工作流程疑难解答

使用以下信息来帮助您诊断和修复运行匹配工作流程时可能遇到的常见问题。

## 我在运行匹配的工作流程后收到了错误文件

### 常见原因

匹配的工作流程可以有多个运行，结果（成功或错误）将写入名称为jobId的文件夹。

匹配工作流程的成功结果将写入包含多个文件的文件夹success，每个文件都包含成功记录的子集。

匹配工作流程的错误将写入包含多个文件夹error，每个文件夹都包含错误记录的子集。

创建错误文件的原因如下：

- [唯一 ID](#) 是：
  - null
  - 一行数据中缺失
  - 数据表中的一条记录中缺失
  - 在数据表的另一行数据中重复
  - 未指定
  - 在同一个来源中不是唯一的
  - 在多个来源中不是唯一的
  - 跨源重叠
  - 超过 38 个字符（仅限基于规则的匹配工作流程）
- [架构映射](#) 中的一个字段包含一个保留名称：
  - EmailAddress
  - InputSourceARN
  - MatchRule
  - matchID

- HashingProtocol
- ConfidenceLevel
- 来源

#### Note

如果错误文件中的记录是由于前面列出的原因而创建的，则需要向您收费，因为这会产生服务的处理成本。如果错误文件中的记录是由于内部服务器错误造成的，则无需向您收费。

## 解决方案

### 要解决此问题

1. 检查[唯一 ID](#) 是否有效。

如果[唯一 ID](#) 无效，请更新数据表中的唯一 ID，保存新的数据表，创建新的架构映射，然后再次运行匹配的工作流程。

2. 检查[架构映射](#)中的一个字段是否包含保留名称。

如果其中一个字段包含保留名称，请使用新名称创建新的架构映射，然后再次运行匹配的工作流程。

## 使用 ID 映射工作流程映射输入数据

ID 映射工作流程是一种数据处理作业，它根据指定的 ID 映射方法将数据从输入数据源映射到输入数据目标。它会生成一个 ID 映射表。

ID 映射工作流程需要输入数据源和输入数据目标。您的数据输入源和目标取决于您要执行的 ID 映射类型。执行 ID 映射的方法有两种：基于规则的服务或提供商的服务：

- 基于规则 ID 映射 - 您可以使用匹配规则将第一方数据从源转换为目标。
- 提供商服务 ID 映射-您可以使用 LiveRamp 提供者服务将第三方数据从源转换为目标。

### Note

中的提供商服务 ID 映射工作流程当前 AWS Entity Resolution 数据匹配服务 已与集成 LiveRamp。如果您订阅了该 LiveRamp 服务，则可以使用创建 ID 映射工作流程 LiveRamp 来执行转码。通过 LiveRamp 转码，您可以将一组源 Ramp IDs 转换为任何目标目标 rampID。通过使用 RamPid 作为代表客户的代币，您可以避免直接与广告平台共享客户数据。

有关更多信息，请参阅 LiveRamp 文档网站上的[通过 ADX 执行翻译](#)。

在以下任一方案中，您可以在两个数据集之间执行 ID 映射：

- 在你自己的内心深处 AWS 账户
- 跨越两个不同的 AWS 账户

下图总结了如何设置 ID 映射工作流程。



#### Complete prerequisite

Create a [schema mapping](#) for ID mapping in your AWS account or an [ID namespace](#) for ID mapping across AWS accounts to define your data.



#### Specify ID mapping details

Provide details for your ID mapping workflow and choose an ID mapping method.



#### Specify source and target

Use a schema mapping or ID namespace to describe your input data depending on your ID mapping type.



#### Specify data output location - *optional*

Choose your S3 location to write your data output.

## 主题

- [一个人的身份映射工作流程 AWS 账户](#)
- [跨两个 ID 映射工作流程 AWS 账户](#)
- [运行 ID 映射工作流程](#)
- [运行自定义 ID 映射工作流程](#)
- [编辑 ID 映射工作流程](#)
- [删除 ID 映射工作流程](#)
- [为 ID 映射工作流程添加或更新资源策略](#)

## 一个人的身份映射工作流程 AWS 账户

一个人的 ID 映射工作流程 AWS 账户使您可以自己在两个数据集之间执行 ID 映射 AWS 账户。

在自己创建 ID 映射工作流程之前 AWS 账户，必须先完成[先决条件](#)。

创建并运行 ID 映射工作流程后，您可以查看输出（ID 映射表）并将其用于分析。

以下主题将引导您完成一系列步骤，以便在同一工作流程中创建 ID 映射工作流程 AWS 账户。

### 主题

- [先决条件](#)
- [创建 ID 映射工作流程（基于规则）](#)
- [创建 ID 映射工作流程（提供者服务）](#)

## 先决条件

在 AWS 账户使用基于规则或提供者服务 ID 映射方法为其创建 ID 映射工作流之前，必须先执行以下操作：

- 完成“[设置](#)”中的任务 AWS Entity Resolution 数据匹配服务。
- 根据您使用的输入数据类型，完成中的[准备输入数据表](#)任务。
- [创建架构映射](#)或[创建匹配的工作流程](#)。
- （仅限提供者服务 ID 映射）在使用创建 ID 映射工作流程之前 LiveRamp，必须选择要临时写入身份映射工作流程输出的亚马逊简单存储服务 (Amazon S3) Service 数据暂存存储桶。

如果您使用 LiveRamp 提供者服务来翻译第三方数据，请添加以下权限策略，该策略允许您访问数据暂存存储桶。

## JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "AWS": "arn:aws:iam::715724997226:root"
      },
      "Action": [
        "s3:PutObject",
        "s3:GetObject",
        "s3:GetObjectVersion",
        "s3:DeleteObject"
      ],
      "Resource": [
        "arn:aws:s3:::<staging-bucket>",
        "arn:aws:s3:::<staging-bucket>/*"
      ]
    },
    {
      "Effect": "Allow",
      "Principal": {
        "AWS": "arn:aws:iam::715724997226:root"
      },
      "Action": [
        "s3:ListBucket",
        "s3:GetBucketLocation",
        "s3:GetBucketPolicy",
        "s3:ListBucketVersions",
        "s3:GetBucketAcl"
      ],
      "Resource": [
        "arn:aws:s3:::<staging-bucket>",
        "arn:aws:s3:::<staging-bucket>/*"
      ]
    }
  ]
}
```

在前面的权限策略中，将每项 `<user input placeholder>` 策略替换为您自己的信息。

*staging-bucket*

Amazon S3 存储桶，用于在运行基于提供商服务的工作流程时临时存储您的数据。

## 创建 ID 映射工作流程（基于规则）

本主题介绍为使用匹配规则将第一方数据从源转换为目标 AWS 账户的 ID 映射工作流程创建 ID 映射工作流程的过程。

为一个人创建基于规则的 ID 映射工作流程 AWS 账户

1. 登录 AWS 管理控制台 并打开 AWS Entity Resolution 数据匹配服务 控制台，网址为 <https://console.aws.amazon.com/entityresolution/>。
2. 在左侧导航窗格的“工作流程”下，选择 ID 映射。
3. 在 ID 映射工作流程页面的右上角，选择创建 ID 映射工作流程。
4. 对于步骤 1：指定 ID 映射工作流程详细信息，请执行以下操作。
  - a. 输入 ID 映射工作流程名称和可选描述。
  - b. 对于 ID 映射方法，请选择基于规则。
  - c. （可选）要仅处理工作流程中新的、已更新或已删除的记录，请选择“启用增量处理”。

### ID mapping method Info

Choose the ID mapping method you want to use.

**Rule-based - new**  
Use matching rules to translate first-party data from a source to a target in ID mapping.

**Provider services**  
Use a provider service to translate third party-encoded data from a source to a target in ID mapping.

**Enable incremental processing**  
AWS Entity Resolution will process only new, updated, or deleted records in either the Source or Target ID namespace, rather than recreating the entire ID mapping table.

AWS Entity Resolution 数据匹配服务 仅处理源 ID 或目标 ID 命名空间中的新记录、已更新记录或已删除记录，而不是重新创建整个 ID 映射表。

当您选择增量处理并且您的数据表具有 DELETE 列时，会根据 DELETE 列的值以不同的方式 AWS Entity Resolution 数据匹配服务 处理记录。

- 在 DELETE 列 true 中标记为的记录将从 ID 映射表中删除。
- false 在“删除”列中标记为的记录将提取到 Amazon S3 中。

如果未选中此选项，则会在 ID 映射表上 AWS Entity Resolution 数据匹配服务 运行默认的批处理 ID 映射 workflow。

- d. (可选) 要为资源启用标签，请选择添加新标签，然后输入密钥和值对。
  - e. 选择下一步。
5. 对于“步骤 2：指定源和目标”，请执行以下操作。
- a. 对于 Source，选择适用于您的场景，然后采取建议的操作。

场景	推荐操作
在 ID 映射工作流程中使用您自己的 AWS Glue 数据库、AWS Glue 表和架构映射。	<ol style="list-style-type: none"> <li>1. 选择架构映射。</li> <li>2. 选择 AWS Glue 数据库 AWS 区域、AWS Glue 表，然后选择相应的架构映射。</li> </ol> <p>您最多可以添加 19 个数据输入。</p>
使用现有的匹配工作流程，该工作流程指向要在 ID 映射工作流程中使用的记录数据。	<ol style="list-style-type: none"> <li>1. 选择“匹配工作流程”。</li> <li>2. 从下拉列表中选择现有的匹配工作流程。</li> </ol>

- b. 对于 Target，从下拉列表中选择一个现有的匹配工作流程。
- c. 对于规则参数，请执行以下操作。
  - i. 根据您的源类型选择以下选项之一，指定规则控件。

Source type (源类型)	推荐操作
匹配工作流程	通过选择来源、目标或两者是否可以在 ID 映射 workflow 中提供规则来指定规则控件。

Source type ( 源类型 )	推荐操作
	<p>规则控件必须在源和目标之间兼容，才能在 ID 映射工作流程中使用。</p> <p>例如，如果源 ID 命名空间将规则限制于目标，但目标 ID 命名空间将规则限制于源，则会导致错误。</p>
架构映射	跳过此步骤。

- ii. 对于比较和匹配参数，比较类型自动设置为多个输入字段。

这是因为两个参与者之前都选择了此选项。

- d. 根据您的目标选择以下选项之一，指定“记录”匹配类型。

您的目标	建议的选项
创建 ID 映射工作流程时，将记录匹配类型限制为：对于目标中的每条匹配记录，仅存储源中的一条匹配记录。	一个源对一个目标
创建 ID 映射工作流程时，将记录匹配类型限制为：对于目标中的每条匹配记录，存储源中的所有匹配记录。	同一个目标有多个来源

**Note**

您必须为源 ID 和目标 ID 命名空间指定兼容限制。

- e. 要指定服务访问权限，请选择一个选项并采取建议的操作。

**Service access**

AWS Entity Resolution requires permissions to read your data input from AWS Glue and write to S3 on your behalf. [View policy document](#)

**Choose a method to authorize AWS Entity Resolution**

- Create and use a new service role  
Automatically create the role and add the necessary permissions policy.
- Use an existing service role

**Service role name**


51 of 64 characters. Use alphanumeric and '+,=,@-\_' characters. Don't include spaces. Name must be unique across all roles in the account.

- This data is encrypted with a KMS key  
Specify the associated KMS key to enable AWS Entity Resolution to access each of your data inputs.

Option	推荐操作
创建并使用新的服务角色	<ul style="list-style-type: none"> <li>• AWS Entity Resolution 数据匹配服务 使用此表所需的策略创建服务角色。</li> <li>• 默认服务角色名称为 <code>entityresolution-id-mapping-workflow- -&lt;timestamp&gt;</code> 。</li> <li>• 您必须拥有创建角色并附加策略的权限。</li> <li>• 如果您的输入数据已加密，请选择“此数据由 KMS 密钥加密”选项。然后，输入用于解密输入数据的密AWS KMS 钥。</li> </ul>

Option	推荐操作
使用现有服务角色	<p>1. 从下拉列表中选择一个现有服务角色名称。</p> <p>如果您有列出角色的权限，则会显示角色列表。</p> <p>如果您没有列出角色的权限，可以输入要使用的角色的 Amazon 资源名称 (ARN)。</p> <p>如果没有现有的服务角色，则使用现有服务角色选项不可用。</p> <p>2. 通过选择在 IAM 中查看外部链接来查看服务角色。</p> <p>默认情况下，AWS Entity Resolution 数据匹配服务不会尝试更新现有角色策略以添加必要的权限。</p>

6. 选择下一步。
7. 对于步骤 3：指定数据输出位置（可选），请执行以下操作。
  - a. 对于数据输出目标，请执行以下操作：
    - i. 选择数据输出的 Amazon S3 位置。
    - ii. 对于加密，如果您选择自定义加密设置，请输入 AWS KMS 密钥 ARN 或选择创建 AWS KMS 密钥。
  - b. 选择下一步。
8. 对于“步骤 4：查看并创建”，请执行以下操作。
  - a. 查看您在之前的步骤中所做的选择，并在必要时对其进行编辑。
  - b. 选择创建。

将出现一条消息，表明身份映射工作流程已创建。

创建 ID 映射工作流程后，就可以[运行身份映射工作流程](#)了。

## 创建 ID 映射工作流程 ( 提供者服务 )

本主题介绍 AWS 账户 使用名为的提供者服务为一个人创建 ID 映射工作流的过程 LiveRamp。LiveRamp 使用维护或派生的 Ramp IDs 将一组源 Ram IDs p 转换为另一组。

为一个人创建基于提供者服务的身份映射工作流程 AWS 账户

1. 登录 AWS 管理控制台 并打开 AWS Entity Resolution 数据匹配服务 控制台，网址为<https://console.aws.amazon.com/entityresolution/>。
2. 在左侧导航窗格的“工作流程”下，选择 ID 映射。
3. 在 ID 映射工作流程页面的右上角，选择创建 ID 映射工作流程。
4. 对于步骤 1：指定 ID 映射工作流程详细信息，请执行以下操作。
  - a. 输入 ID 映射工作流程名称和可选描述。
  - b. 对于 ID 映射方法，请选择提供商服务。

AWS Entity Resolution 数据匹配服务 目前提供 LiveRamp 提供者服务作为 ID 映射方法。如果您订阅了 LiveRamp，则状态将显示为“已订阅”。有关如何订阅的更多信息 LiveRamp，请参阅[步骤 1：在上订阅提供商服务 AWS Data Exchange](#)。



### ID mapping method Info

## /LiveRamp

Currently we are only offering LiveRamp service as an ID mapping method.

#### Access to LiveRamp provider subscription

 **Subscribed**

 To ensure a successful workflow run, your data input file format and normalization must be aligned with the provider service's guidelines. [Learn more](#) 

### Note

确保您的数据输入文件格式符合提供商服务的指南。有关输入文件格式指南 LiveRamp 的更多信息，请参阅 LiveRamp 文档网站上的[通过 ADX 执行翻译](#)。

- c. 要进行LiveRamp 配置，请输入以下 LiveRamp 提供的值：
- 客户 ID 管理器 ARN
  - 客户密钥管理器 ARN

**LiveRamp configuration** [Info](#)

**Client ID manager ARN**  
Enter the Client ID manager ARN provided by LiveRamp.

Enter ARN

0 of 2,048 characters.

**Client secret manager ARN**  
Enter the Client secret manager ARN provided by LiveRamp.

Enter ARN

0 of 2,048 characters.

- d. ( 可选 ) 要为资源启用标签，请选择添加新标签，然后输入密钥和值对。
- e. 选择下一步。
5. 对于“步骤 2：指定源和目标”，请执行以下操作。
- a. 对于 Source，选择适用于您的场景，然后采取建议的操作。

场景	推荐操作
在 ID 映射工作流程中使用您自己的 AWS Glue 数据库、AWS Glue 表和架构映射。	<ol style="list-style-type: none"> <li>1. 选择架构映射。</li> <li>2. 选择AWS Glue 数据库AWS 区域、AWS Glue 表，然后选择相应的架构映射。</li> </ol> <p style="text-align: center;">您最多可以添加 19 个数据输入。</p>
使用现有的匹配工作流程，该工作流程指向要在 ID 映射工作流程中使用的记录数据。	<ol style="list-style-type: none"> <li>1. 选择“匹配工作流程”。</li> <li>2. 从下拉列表中选择现有的匹配工作流程。</li> </ol>

- b. 对于 Target，请根据您选择的 ID 映射方法执行以下操作之一。

ID 映射方法	推荐操作
基于规则	从下拉列表中选择现有的匹配工作流程。
提供商服务	输入目标域中 LiveRamp 提供的转码目标 LiveRamp 客户端域标识符。 <div data-bbox="889 464 1507 594" style="border: 1px solid #ccc; padding: 5px; margin-top: 10px;"> <p><b>Target</b> <small>Info</small> Enter the LiveRamp client domain identifier targeted for transcoding provided by LiveRamp.</p> <p><b>Target domain</b>  <input type="text" value="Enter target domain"/>  <small>0 of 4 characters.</small></p> </div>

- c. 对于数据暂存，请选择您要临时写入 ID 映射工作流程输出的 Amazon S3 位置。

**Data staging** Info  
Choose the Amazon S3 location for temporarily storing your data while it processes. Your information will not be saved permanently.

**Amazon S3 location**

View 
Browse S3

- d. 要指定服务访问权限，请选择一个选项并采取建议的操作。

**Service access**

AWS Entity Resolution requires permissions to read your data input from AWS Glue and write to S3 on your behalf. [View policy document](#)

**Choose a method to authorize AWS Entity Resolution**

Create and use a new service role  
Automatically create the role and add the necessary permissions policy.

Use an existing service role

**Service role name**

entityresolution-id-mapping-workflow-20240117121045

51 of 64 characters. Use alphanumeric and '+,=, @, -, \_' characters. Don't include spaces. Name must be unique across all roles in the account.

This data is encrypted with a KMS key  
Specify the associated KMS key to enable AWS Entity Resolution to access each of your data inputs.

Option	推荐操作
创建并使用新的服务角色	<ul style="list-style-type: none"> <li>• AWS Entity Resolution 数据匹配服务 使用此表所需的策略创建服务角色。</li> <li>• 默认服务角色名称为 <code>entityresolution-id-mapping-workflow- -&lt;timestamp&gt;</code> 。</li> <li>• 您必须拥有创建角色并附加策略的权限。</li> <li>• 如果您的输入数据已加密，请选择“此数据由 KMS 密钥加密”选项。然后，输入用于解密输入数据的密AWS KMS 钥。</li> </ul>
使用现有服务角色	<ol style="list-style-type: none"> <li>1. 从下拉列表中选择一个现有服务角色名称。 <p>如果您有列出角色的权限，则会显示角色列表。</p> <p>如果您没有列出角色的权限，可以输入要使用的角色的 Amazon 资源名称 (ARN)。</p> <p>如果没有现有的服务角色，则使用现有服务角色选项不可用。</p> </li> <li>2. 通过选择在 IAM 中查看外部链接来查看服务角色。 <p>默认情况下，AWS Entity Resolution 数据匹配服务 不会尝试更新现有角色策略以添加必要的权限。</p> </li> </ol>

6. 选择下一步。

7. 对于步骤 3：指定数据输出位置（可选），请执行以下操作。

a. 对于数据输出目标，请执行以下操作：

i. 选择数据输出的 Amazon S3 位置。

- ii. 对于加密，如果您选择自定义加密设置，请输入AWS KMS 密钥 ARN 或选择创建 AWS KMS 密钥。
- b. 查看LiveRamp 生成的输出。
- c. 选择下一步。

AWS Entity Resolution > ID mapping workflows > Create ID mapping workflow

Step 1 Specify ID mapping workflow details

Step 2 Specify source and target

Step 3 - optional Specify data output location

Step 4 Review and create

### Specify data output location - optional Info

Choose your S3 location to write your data output.

**Data output destination Info**  
Choose the Amazon S3 location for the data output.

**Amazon S3 location**

Q s3://bucket/prefix View 🔗 Browse S3

**Encryption - optional Info**  
Your data is encrypted by default with a key that AWS owns and manages for you. To specify a different key, customize your encryption settings.

Customize encryption settings  
Specify an AWS KMS key to customize your encryption settings.

▼ **LiveRamp generated output (2)**  
Additional information generated by LiveRamp.

Output field	Description
RAMPID	LiveRamp's universal identifier that is tied to devices in the LiveRamp Identity Graph
TRANSCODED_IDENTIFIER	LiveRamp's universal identifier that is tied to devices in the LiveRamp Identity Graph

Cancel Previous Next

8. 对于“步骤 4：查看并创建”，请执行以下操作。
  - a. 查看您在之前的步骤中所做的选择，并在必要时对其进行编辑。
  - b. 选择创建。

将出现一条消息，表明身份映射工作流程已创建。

9. 创建 ID 映射工作流程后，就可以[运行身份映射工作流程](#)了。

## 跨两个 ID 映射工作流程 AWS 账户

跨两个数据集的 ID 映射工作流 AWS 账户使您能够在两个数据集之间执行 ID 映射 AWS 账户。这通常是在你自己 AWS 账户 和另一个人之间完成 AWS 账户的。

例如，发布者可以使用自己的目标 ID 命名空间（在自己的命名空间中 AWS 账户）和广告商的来源 ID 命名空间（在另一个 AWS 账户命名空间中）创建 ID 映射工作流程。

在创建跨两个的 ID 映射工作流程之前 AWS 账户，必须先完成[先决条件](#)。

创建 ID 映射工作流程后，您可以查看输出（ID 映射表）并将其用于分析。

以下主题将引导您完成一系列步骤，以跨两个步骤创建 ID 映射工作流程 AWS 账户：

主题

- [先决条件](#)
- [创建 ID 映射工作流程（基于规则）](#)
- [创建 ID 映射工作流程（提供者服务）](#)

## 先决条件

在跨两个 ID 映射工作流程之前 AWS 账户，必须先执行以下操作：

- 完成[设置 AWS Entity Resolution 数据匹配服务](#)中所述的任务。
- [创建 ID 命名空间源](#)。
- [创建 ID 命名空间目标](#)。
- 如果您正在使用其他来源的 ID 命名空间源，请获取 ID 命名空间 ARN。AWS 账户
- （仅限提供商服务）在两个存储桶之间创建 ID 映射工作流程 AWS 账户 LiveRamp 需要获得访问权限 S3 存储桶和 AWS Key Management Service (AWS KMS) 客户托管密钥。

在使用创建跨两个 AWS 账户的 ID 映射工作流程之前 LiveRamp，请添加以下权限策略，该策略 LiveRamp 允许访问 S3 存储桶和客户托管密钥。

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [{
    "Effect": "Allow",
    "Principal": {
      "AWS": "arn:aws:iam::715724997226:root"
    },
    "Action": [
      "kms:Decrypt"
    ],
    "Resource": "arn:aws:kms:us-east-1:111122223333:key/key-id",
    "Condition": {
      "StringEquals": {
```

```

        "kms:ViaService": "s3.us-east-1.amazonaws.com"
      }
    }
  ]
}

```

在前面的权限策略中，将每项 *<user input placeholder>* 策略替换为您自己的信息。

*<KMSKeyARN>*

AWS KMS 客户托管密钥的 ARN。

## 创建 ID 映射工作流程（基于规则）

完成[先决条件](#)后，您可以创建一个或多个 ID 映射工作流程，以使用匹配规则将第一方数据从源转换为目标。

在两者之间创建基于规则的 ID 映射工作流程 AWS 账户

1. 登录 AWS 管理控制台 并打开 AWS Entity Resolution 数据匹配服务 控制台，网址为<https://console.aws.amazon.com/entityresolution/>。
2. 在左侧导航窗格的“工作流程”下，选择 ID 映射。
3. 在 ID 映射工作流程页面的右上角，选择创建 ID 映射工作流程。
4. 对于步骤 1：指定 ID 映射工作流程详细信息，请执行以下操作。
  - a. 输入 ID 映射工作流程名称和可选描述。
  - b. 对于 ID 映射方法，请选择基于规则。
  - c. （可选）要仅处理工作流中新的、已更新或已删除的记录，请选择“启用增量处理”。

**ID mapping method** Info

Choose the ID mapping method you want to use.

**Rule-based - new**

Use matching rules to translate first-party data from a source to a target in ID mapping.

**Provider services**

Use a provider service to translate third party-encoded data from a source to a target in ID mapping.

**Enable incremental processing**

AWS Entity Resolution will process only new, updated, or deleted records in either the Source or Target ID namespace, rather than recreating the entire ID mapping table.

AWS Entity Resolution 数据匹配服务 仅处理源 ID 或目标 ID 命名空间中的新记录、已更新记录或已删除记录，而不是重新创建整个 ID 映射表。

当您选择增量处理并且您的数据表具有 DELETE 列时，会根据 DELETE 列的值以不同的方式 AWS Entity Resolution 数据匹配服务 处理记录。

- 在 DELETE 列 true 中标记为的记录将从 ID 映射表中删除。
- false 在“删除”列中标记为的记录将提取到 Amazon S3 中。

如果未选中此选项，则会在 ID 映射表上 AWS Entity Resolution 数据匹配服务 运行默认的批处理 ID 映射 workflow。

- d. (可选) 要为资源启用标签，请选择添加新标签，然后输入密钥和值对。
  - e. 选择下一步。
5. 对于“步骤 2：指定源和目标”，请执行以下操作。
- a. 打开“高级选项”。
  - b. 对于“来源”，选择“匹配 workflow”，然后从下拉列表中选择现有的匹配 workflow。
  - c. 对于 Target，选择匹配 workflow，然后从下拉列表中选择现有的匹配 workflow。
  - d. 对于规则参数，通过选择源还是目标可以在 ID 映射 workflow 中提供规则来指定规则控件。


规则控件必须在源和目标之间兼容，才能在 ID 映射 workflow 中使用。例如，如果源 ID 命名空间将规则限制于目标，但目标 ID 命名空间将规则限制于源，则会导致错误。

- e. 对于比较和匹配参数，请执行以下操作。
  - i. 通过根据您的目标选择一个选项来指定比较类型。

您的目标	建议的选项
查找存储在多个输入字段中的数据的任意匹配组合，无论数据位于相同还是不同的输入字段中。	多个输入字段
当存储在多个输入字段中的相似数据不应匹配时，请限制在单个输入字段内进行比较。	单个输入字段

- ii. 通过根据您的目标选择一个选项来指定“记录”匹配类型。

您的目标	建议的选项
创建 ID 映射工作流程时，将记录匹配类型限制为：对于目标中的每条匹配记录，仅存储源中的一条匹配记录。	一个源对一个目标
创建 ID 映射工作流程时，将记录匹配类型限制为：对于目标中的每条匹配记录，存储源中的所有匹配记录。	同一个目标有多个来源

 Note

您必须为源 ID 和目标 ID 命名空间指定兼容限制。

- f. 要指定服务访问权限，请选择一个选项并采取建议的操作。

### Service access

AWS Entity Resolution requires permissions to read your data input from AWS Glue and write to S3 on your behalf. [View policy document](#)

#### Choose a method to authorize AWS Entity Resolution

- Create and use a new service role  
Automatically create the role and add the necessary permissions policy.
- Use an existing service role

#### Service role name

entityresolution-id-mapping-workflow-20240117121045

51 of 64 characters. Use alphanumeric and '+=, @-\_' characters. Don't include spaces. Name must be unique across all roles in the account.

- This data is encrypted with a KMS key  
Specify the associated KMS key to enable AWS Entity Resolution to access each of your data inputs.

Option	推荐操作
创建并使用新的服务角色	<ul style="list-style-type: none"> <li>• AWS Entity Resolution 数据匹配服务 使用此表所需的策略创建服务角色。</li> <li>• 默认服务角色名称为 <code>entityresolution-id-mapping-workflow- -&lt;timestamp&gt;</code> 。</li> <li>• 您必须拥有创建角色并附加策略的权限。</li> <li>• 如果您的输入数据已加密，请选择“此数据由 KMS 密钥加密”选项。然后，输入用于解密输入数据的密AWS KMS 钥。</li> </ul>
使用现有服务角色	<ol style="list-style-type: none"> <li>1. 从下拉列表中选择一个现有服务角色名称。 <p>如果您有列出角色的权限，则会显示角色列表。</p> <p>如果您没有列出角色的权限，可以输入要使用的角色的 Amazon 资源名称 (ARN)。</p> <p>如果没有现有的服务角色，则使用现有服务角色选项不可用。</p> </li> <li>2. 通过选择在 IAM 中查看外部链接来查看服务角色。 <p>默认情况下，AWS Entity Resolution 数据匹配服务 不会尝试更新现有角色策略以添加必要的权限。</p> </li> </ol>

6. 选择下一步。

7. 对于步骤 3：指定数据输出位置（可选），请执行以下操作。

a. 对于数据输出目标，请执行以下操作。

i. 选择数据输出的 Amazon S3 位置。

- ii. 对于加密，如果您选择自定义加密设置，请输入AWS KMS 密钥 ARN 或选择创建 AWS KMS 密钥。
  - b. 查看LiveRamp 生成的输出。
  - c. 选择下一步。
8. 对于“步骤 4：查看并创建”，请执行以下操作。
- a. 查看您在之前的步骤中所做的选择，并在必要时对其进行编辑。
  - b. 选择创建。

将出现一条消息，表明身份映射工作流程已创建。

创建 ID 映射工作流程后，就可以[运行身份映射工作流程](#)了。

## 创建 ID 映射工作流程（提供者服务）

完成[先决条件](#)后，您可以使用 LiveRamp 提供者服务创建一个或多个 ID 映射工作流程。LiveRamp 使用维护或派生的 Ramp IDs 将一组源 Ram IDs p 转换为另一组。

### 使用提供者服务创建 ID 映射工作流程

1. 登录 AWS 管理控制台 并打开 AWS Entity Resolution 数据匹配服务 控制台，网址为<https://console.aws.amazon.com/entityresolution/>。
2. 在左侧导航窗格的“工作流程”下，选择 ID 映射。
3. 在 ID 映射工作流程页面的右上角，选择创建 ID 映射工作流程。
4. 对于步骤 1：指定 ID 映射工作流程详细信息，请执行以下操作。
  - a. 输入 ID 映射工作流程名称和可选描述。
  - b. 对于 ID 映射方法，请选择提供商服务。

AWS Entity Resolution 数据匹配服务 目前提供 LiveRamp 提供者服务作为 ID 映射方法。如果您订阅了 LiveRamp，则状态将显示为“已订阅”。有关如何订阅的更多信息 LiveRamp，请[参阅步骤 1：在上订阅提供商服务 AWS Data Exchange](#)。

**ID mapping method** [Info](#)

# /LiveRamp

Currently we are only offering LiveRamp service as an ID mapping method.

**Access to LiveRamp provider subscription**

✔ Subscribed

ⓘ To ensure a successful workflow run, your data input file format and normalization must be aligned with the provider service's guidelines. [Learn more](#) [↗](#)

**Note**

确保您的数据输入文件格式符合提供商服务的指南。有关输入文件格式指南 LiveRamp 的更多信息，请参阅 LiveRamp 文档网站上的[通过 ADX 执行翻译](#)。

c. 要进行 LiveRamp 配置，请输入以下 LiveRamp 提供的值：

- 客户 ID 管理器 ARN
- 客户密钥管理器 ARN

**LiveRamp configuration** [Info](#)**Client ID manager ARN**

Enter the Client ID manager ARN provided by LiveRamp.

0 of 2,048 characters.

**Client secret manager ARN**

Enter the Client secret manager ARN provided by LiveRamp.

0 of 2,048 characters.

d. (可选) 要为资源启用标签，请选择添加新标签，然后输入密钥和值对。

e. 选择下一步。

5. 对于“步骤 2：指定源和目标”，请执行以下操作。

- 打开“高级选项”。
- 对于来源，选择 ID 命名空间。

AWS Entity Resolution > ID mapping workflows > Create ID mapping workflow

Step 1  
Specify ID mapping workflow details

Step 2  
**Specify source and target**

Step 3 - optional  
Specify data output location

Step 4  
Review and create

## Specify source and target Info

Use a schema mapping or ID namespace to describe your input data depending on your ID mapping type.

**Advanced options**  
Use advanced options if you are creating an ID mapping across AWS accounts and have created ID namespace resources to manage AWS account permissions.

### Source Info

The source of the data in an ID mapping workflow.

**Schema mapping**  
Use AWS Glue database, AWS Glue table, and schema mapping for ID mapping on your own AWS account.

**ID namespace**  
Use an ID namespace to describe your source data for ID mapping across two AWS accounts.

### ID namespace Info

Choose an AWS account associated with the ID namespace source. [Create ID namespace](#)

Your AWS account  
 Another AWS account

**Your ID namespaces**

Select ID namespace ▼

- c. 对于 ID 命名空间，请确定 ID 命名空间所在的位置，然后采取建议的操作。

ID 命名空间的位置	推荐操作
你自己的 AWS 账户	<ol style="list-style-type: none"> <li>1. 选择你的 AWS 账户。</li> <li>2. 从“您的 ID 命名空间”下拉列表中选择 ID 命名空间。</li> </ol>
别人的 AWS 账户	<ol style="list-style-type: none"> <li>1. 选择另一个 AWS 账户。</li> <li>2. 输入 ID 命名空间 ARN。</li> </ol>

- d. 对于“目标”，选择 ID 命名空间。

**Target** [Info](#)  
Select how you want to provide the domain to which you want to translate your data using ID mapping.

**Domain**  
Provide a specific target domain to which you want to translate the data to

**ID namespace**  
Use an ID namespace to describe your target configuration for ID mapping across two AWS accounts.

**ID namespace** [Info](#)  
Choose an AWS account associated with the ID namespace source. [Create ID namespace](#)

Your AWS account  
 Another AWS account

**Your ID namespaces**

- e. 要指定服务访问权限，请选择一个选项并采取建议的操作。

**Service access**  
AWS Entity Resolution requires permissions to read your data input from AWS Glue and write to S3 on your behalf. [View policy document](#)

**Choose a method to authorize AWS Entity Resolution**

Create and use a new service role  
Automatically create the role and add the necessary permissions policy.

Use an existing service role

**Service role name**

51 of 64 characters. Use alphanumeric and '+=, @-\_' characters. Don't include spaces. Name must be unique across all roles in the account.

This data is encrypted with a KMS key  
Specify the associated KMS key to enable AWS Entity Resolution to access each of your data inputs.

Option	推荐操作
创建并使用新的服务角色	<ul style="list-style-type: none"> <li>• AWS Entity Resolution 数据匹配服务 使用此表所需的策略创建服务角色。</li> <li>• 默认服务角色名称为 <code>entityresolution-id-mapping-workflow- -&lt;timestamp&gt;</code> 。</li> <li>• 您必须拥有创建角色并附加策略的权限。</li> <li>• 如果您的输入数据已加密，请选择“此数据由 KMS 密钥加密”选项。然后，输入用于解密输入数据的密AWS KMS 钥。</li> </ul>
使用现有服务角色	<ol style="list-style-type: none"> <li>1. 从下拉列表中选择一个现有服务角色名称。 <ul style="list-style-type: none"> <li>如果您有列出角色的权限，则会显示角色列表。</li> <li>如果您没有列出角色的权限，可以输入要使用的角色的 Amazon 资源名称 (ARN)。</li> <li>如果没有现有的服务角色，则使用现有服务角色选项不可用。</li> </ul> </li> <li>2. 通过选择在 IAM 中查看外部链接来查看服务角色。 <ul style="list-style-type: none"> <li>默认情况下，AWS Entity Resolution 数据匹配服务 不会尝试更新现有角色策略以添加必要的权限。</li> </ul> </li> </ol>

6. 选择下一步。

7. 对于步骤 3：指定数据输出位置（可选），请执行以下操作。

a. 对于数据输出目标，请执行以下操作。

i. 选择数据输出的 Amazon S3 位置。

- ii. 对于加密，如果您选择自定义加密设置，请输入AWS KMS 密钥 ARN 或选择创建 AWS KMS 密钥。
- b. 查看LiveRamp 生成的输出。
- c. 选择下一步。

AWS Entity Resolution > ID mapping workflows > Create ID mapping workflow

Step 1 Specify ID mapping workflow details

Step 2 Specify source and target

Step 3 - optional Specify data output location

Step 4 Review and create

### Specify data output location - optional Info

Choose your S3 location to write your data output.

**Data output destination** Info  
Choose the Amazon S3 location for the data output.

**Amazon S3 location**

Q s3://bucket/prefix View Browse S3

**Encryption - optional** Info  
Your data is encrypted by default with a key that AWS owns and manages for you. To specify a different key, customize your encryption settings.

Customize encryption settings  
Specify an AWS KMS key to customize your encryption settings.

▼ **LiveRamp generated output (2)**  
Additional information generated by LiveRamp.

Output field	Description
RAMPID	LiveRamp's universal identifier that is tied to devices in the LiveRamp Identity Graph
TRANSCODED_IDENTIFIER	LiveRamp's universal identifier that is tied to devices in the LiveRamp Identity Graph

Cancel Previous Next

8. 对于“步骤 4：查看并创建”，请执行以下操作。
  - a. 查看您在之前的步骤中所做的选择，并在必要时对其进行编辑。
  - b. 选择创建。

将出现一条消息，表明身份映射工作流程已创建。

创建 ID 映射工作流程后，就可以[运行身份映射工作流程](#)了。

## 运行 ID 映射工作流程

在为其中一个人创建 ID 映射工作流程 AWS 账户或在两个之间创建 ID 映射工作流程后 AWS 账户，您可以运行 ID 映射工作流程。ID 映射工作流程输出一个 CSV 文件。

## 运行 ID 映射工作流程

1. 登录 AWS 管理控制台 并打开 AWS Entity Resolution 数据匹配服务 控制台，网址为<https://console.aws.amazon.com/entityresolution/>。
2. 在左侧导航窗格的“工作流程”下，选择 ID 映射。
3. 选择 ID 映射工作流程。
4. 在 ID 映射工作流程详细信息页面的右上角，选择运行。
5. 在匹配的工作流程详细信息页面的指标选项卡上，在“上次作业指标”下查看以下内容：
  - Job ID
  - 匹配工作流作业的状态：已排队、进行中、已完成、失败
  - 运行类型
  - 工作流程作业的开始时间
  - 工作流作业的完成时间
  - 工作流程作业的持续时间
  - 输出目的地
  - 这些区域有：AWS KMS key
  - 服务角色
  - 输入记录的数量
  - 唯一记录的数量
  - 加载的新唯一记录的数量
  - 映射的记录数
  - 已删除的映射记录的数量
  - 新映射记录的数量
  - 映射的源记录数量
  - 新映射的源记录的数量
  - 已删除的映射源记录的数量
  - 映射的目标记录数
  - 新映射的目标记录的数量
  - 已删除的映射目标记录的数量
  - 已处理的删除记录数
  - 已处理的记录数

- 未处理的记录数

在 Job History 下，您还可以查看之前运行的 ID 映射工作流程作业的作业指标。

6. 身份映射工作流程任务完成（状态为已完成）后，选择数据输出，然后选择您的 Amazon S3 位置以查看结果。

获取 CSV 文件后，您可以通过 RAMPID 加入 TRANSCODED\_ID。

## 运行自定义 ID 映射工作流程

### Note

此过程适用于 [单个工作流 AWS 账户](#) 或 [跨越两个 AWS 账户且启用增量处理的工作流](#)。

运行 ID 映射工作流程时，您可以为输出数据指定不同于最初配置的 Amazon S3 位置。您还可以通过选择以下三种运行类型之一来选择如何处理数据：Batch（处理所有数据）、增量（仅处理新的或更改的数据）或仅删除（仅处理删除请求）。

### 使用新的输出目标运行 ID 映射工作流

1. 登录 AWS 管理控制台 并打开 AWS Entity Resolution 数据匹配服务 控制台，网址为 <https://console.aws.amazon.com/entityresolution/>。
2. 在左侧导航窗格的“工作流程”下，选择 ID 映射。
3. 选择要运行的 ID 映射工作流程。
4. 在 ID 映射工作流详细信息页面上，选择运行工作流程，然后选择使用新的输出目标运行。
5. 对于数据输出目标，请配置以下内容。
  - a. 对于“运行类型”，选择以下选项之一。
    - Batch — 处理整个 ID 映射表。

建议在初始设置、定期进行完全刷新或源和目标 ID 命名空间发生重大更改时使用。
    - 增量-仅处理源 ID 或目标 ID 命名空间中的新记录、已更新记录或已删除记录。

建议用于频繁更新、每日运行或实时数据同步。
    - 仅删除-仅处理从 Target ID 命名空间中删除的记录。

建议用于快速同步移除操作。

- b. 选择数据输出的 Amazon S3 位置。
- c. 对于加密，请执行以下任一操作：
  - 保留默认加密设置
  - 选择“自定义加密设置”，然后输入AWS KMS 密钥 ARN 或选择创建 AWS KMS 密钥。

6. 要指定服务访问权限，请选择一个选项并采取建议的操作。

Option	推荐操作
创建并使用新的服务角色	<ul style="list-style-type: none"> <li>• AWS Entity Resolution 数据匹配服务 使用此表所需的策略创建服务角色。</li> <li>• 默认服务角色名称为 <code>entityresolution-id-mapping-workflow-&lt;timestamp&gt;</code>。</li> <li>• 您必须拥有创建角色并附加策略的权限。</li> <li>• 如果您的输入数据已加密，请选择“此数据由 KMS 密钥加密”选项。然后，输入用于解密输入数据的密AWS KMS 钥。</li> </ul>
使用现有服务角色	<ol style="list-style-type: none"> <li>1. 从下拉列表中选择一个现有服务角色名称。  如果您有列出角色的权限，则会显示角色列表。  如果您没有列出角色的权限，可以输入要使用的角色的 Amazon 资源名称 (ARN)。  如果没有现有的服务角色，则使用现有服务角色选项不可用。</li> <li>2. 通过选择在 IAM 中查看外部链接来查看服务角色。</li> </ol>

Option	推荐操作
	默认情况下，AWS Entity Resolution 数据匹配服务不会尝试更新现有角色策略以添加必要的权限。

7. 选择运行。
8. 在匹配的工作流程详细信息页面的指标选项卡上，在“上次作业指标”下查看以下内容：
  - Job ID
  - workflows 作业的完成时间
  - 匹配 workflows 作业的状态：已排队、进行中、已完成、失败
  - 已处理的记录数
  - 未处理的记录数
  - 输入记录的数量
  - IDs 生成的唯一匹配项的数量。
  - 新映射记录的数量。
  - 新映射的目标记录的数量。
  - 新映射的源记录的数量。
  - 已删除的新映射源记录的数量。
  - 已删除的新映射目标记录的数量。
  - 已删除的新映射记录的数量。

在 Job History 下，您还可以查看之前运行的 ID 映射 workflows 作业的作业指标。

9. 身份映射 workflows 任务完成（状态为已完成）后，选择数据输出，然后选择您的 Amazon S3 位置以查看结果。

获取 CSV 文件后，您可以通过 RAMPID 加入 TRANSCODED\_ID。

## 编辑 ID 映射 workflows

编辑 ID 映射 workflows 允许您保持实体解析能力，up-to-date 并与随着时间的推移不断变化的业务需求保持一致。您可能需要调整映射规则、技术和参数，可以优化 workflows 以提供更准确、更可靠的 ID 匹配结果。您可能还想添加新的数据源，扩展要映射的类型，或者 IDs 在 workflows 中加入其他匹配条

件。如果您发现 ID 映射结果存在问题或错误，则使用工作流程进行编辑可以帮助您诊断和解决这些问题。

要编辑 ID 映射工作流程，请执行以下操作：

1. 登录 AWS 管理控制台 并打开 AWS Entity Resolution 数据匹配服务 控制台，网址为 <https://console.aws.amazon.com/entityresolution/>。
2. 在左侧导航窗格的“工作流程”下，选择 ID 映射。
3. 选择 ID 映射工作流程。
4. 在 ID 映射工作流程详细信息页面的右上角，选择编辑。
5. 在“指定 ID 映射工作流程详细信息”页面上，进行任何必要的更改，然后选择“下一步”。
6. 在“指定数据输出”页面上，进行必要的更改，然后选择“下一步”。
7. 在“查看并保存”页面上，进行必要的更改，然后选择“保存”。

## 删除 ID 映射工作流程

如果您不再使用 ID 映射工作流程，则将其删除可以帮助简化工作流程管理。此外，删除用于类似目的的冗余或效率较低的 ID 映射工作流程可以帮助您整合流程。

要删除 ID 映射工作流程，请执行以下操作：

1. 登录 AWS 管理控制台 并打开 AWS Entity Resolution 数据匹配服务 控制台，网址为 <https://console.aws.amazon.com/entityresolution/>。
2. 在左侧导航窗格的“工作流程”下，选择 ID 映射。
3. 选择 ID 映射工作流程。
4. 在 ID 映射工作流程详细信息页面的右上角，选择删除。
5. 确认删除，然后选择删除。

## 为 ID 映射工作流程添加或更新资源策略

资源策略允许 ID 映射资源的创建者访问您的 ID 映射工作流程资源。

添加或更新资源策略

1. 登录 AWS 管理控制台 并打开 AWS Entity Resolution 数据匹配服务 控制台，网址为 <https://console.aws.amazon.com/entityresolution/>。

2. 在左侧导航窗格的“工作流程”下，选择 ID 映射。
3. 选择 ID 映射工作流程。
4. 在 ID 映射工作流程详细信息页面上，选择权限选项卡。
5. 在资源策略部分中，选择编辑。
6. 在 JSON 编辑器中添加或更新策略。
7. 选择保存更改。

# 以提供商 AWS Entity Resolution 数据匹配服务 身份与之集成

AWS Entity Resolution 数据匹配服务 第三方提供商集成可帮助客户保护消费者隐私并遵守数据主权法。第三方提供商（例如 LiveRamp 和 TransUnion）将消费者标识符转换为广告 IDs，例如 Ramp IDs 和 Fabrick IDs。这些广告标识符通常用于广告和营销工具，以防止将消费者数据导出到非 AWS 托管系统。本节为提供商提供了与之集成的指南，AWS Entity Resolution 数据匹配服务 以将消费者标识符编码或转码 IDs 为广告，以便在[基于提供商服务的](#)匹配工作流程中使用。

有关当前与集成的提供商服务的更多信息 AWS Entity Resolution 数据匹配服务，请参阅[创建基于提供商服务的匹配工作流程](#)。

## 主题

- [要求](#)
- [使用 AWS Entity Resolution 数据匹配服务 OpenAPI 规范](#)
- [测试提供商集成](#)

## 要求

在作为提供者服务与集成之前 AWS Entity Resolution 数据匹配服务，请完成以下要求。

## 主题

- [在上列出提供商服务 AWS Data Exchange](#)
- [确定你的属性](#)
- [索取 AWS Entity Resolution 数据匹配服务 OpenAPI 规范](#)

## 在上列出提供商服务 AWS Data Exchange

作为第三方提供商，您必须在 [AWS Data Exchange \(ADX\)](#) 产品目录中发布您的产品。您的产品在 AWS Data Exchange 产品目录上列出后，订阅者可以通过公开或私有报价订阅您的产品。

## 要在上列出提供者服务 AWS Data Exchange

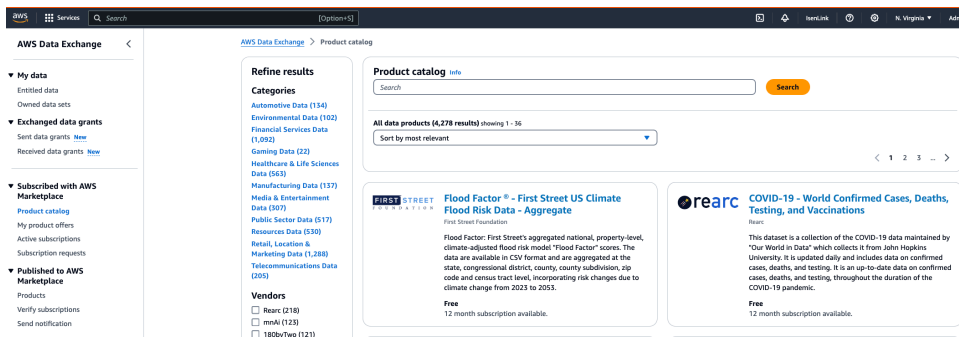
1. 如果您是新的数据产品提供商 AWS Data Exchange，请完成《AWS Data Exchange 用户指南》中标题为“[作为提供者入门](#)”一节中的步骤。

2. 按照AWS Data Exchange 用户指南中标题为“[如何发布包含 APIs 的产品](#)”一节中的步骤创建 REST API 数据集并发布包含 APIs 的新产品。AWS Data Exchange 您可以使用 AWS Data Exchange 控制台或使用来完成该过程 AWS Command Line Interface。

如果您已将产品知名度设置为公开，则所有订阅者都可以公开报价。

如果您已将产品可见性设置为私有，请根据您的用例完成AWS Data Exchange 用户指南中标题为“[创建自定义报价](#)”部分中的步骤。

下图显示了产品目录中可用产品的示例。AWS Data Exchange



3. 产品在 AWS Data Exchange 产品目录上发布后，订阅者可以通过以下方式订阅该产品。

- 订阅公共产品。
- 使用由[提供商服务发布的私有报价](#)（自定义报价）。
- 使用[自带订阅 \(BYOS\)](#) 优惠。

有关更多信息，请参阅[订阅和访问AWS Data Exchange 用户指南 APIs中包含的产品](#)。

## 确定你的属性

输入数据的@@ 属性是工作流程中要解析的实体的类型定义。属性的一些示例包括FirstName、LastNameEmail、或Custom String。

当你确定自己的属性时，你应该记下任何要求或指导方针。

### Example示例

以下是识别提供者属性的验证示例。

- “FirstName或LastName” 属性是必需的。

- 如果该Email属性存在，则必须对其进行哈希处理。

作为提供商，在继续操作之前，您必须识别提供商服务产品中的属性，然后<aws-entity-resolution-bd###@amazon.com> 将这些属性传达给 AWS Entity Resolution 数据匹配服务 业务开发团队进行进一步验证。

## 索取 AWS Entity Resolution 数据匹配服务 OpenAPI 规范

AWS Entity Resolution 数据匹配服务 有一个 OpenAPI 规范，作为提供者，您可以将其用作包含集成所 APIs 涉及内容的握手。有关更多信息，请参阅 [使用 AWS Entity Resolution 数据匹配服务 OpenAPI 规范](#)。

要申请 OpenAPI 定义，请通过 <aws-entity-resolution-bd@amazon> .com 与 AWS Entity Resolution 数据匹配服务 业务开发团队联系。

## 使用 AWS Entity Resolution 数据匹配服务 OpenAPI 规范

OpenAPI 规范定义了与之相关的所有协议。AWS Entity Resolution 数据匹配服务此规范是实现集成所必需的。

OpenAPI 定义包含以下 API 操作：

- POST AssignIdentities
- POST CreateJob
- GET GetJob
- POST StartJob
- POST MapIdentities
- GET Schema

要索取 OpenAPI 规范，请通过 <aws-entity-resolution-bd@amazon> .com 与 AWS Entity Resolution 数据匹配服务 业务开发团队联系。

OpenAPI 规范支持两种类型的集成，用于编码和转码消费者标识符、批处理和同步处理。获取 OpenAPI 规范后，针对您的用例实现处理集成类型。

主题

- [批处理集成](#)
- [同步处理集成](#)

## 批处理集成

批处理集成遵循异步设计模式。启动工作流程后 AWS Data Exchange，它会通过提供商集成端点提交作业，然后该工作流通过定期轮询作业状态来等待任务完成。对于可能需要更长时间且提供商吞吐量较低的任务运行，则更适合使用此解决方案。提供者将以 Amazon S3 链接的形式获取数据集位置，他们可以自行处理该链接，并将结果写入预先确定的输出 S3 位置。

使用三个 API 定义启用批处理集成。AWS Entity Resolution 数据匹配服务 将按以下顺序调用可用的提供 AWS Data Exchange 者端点：

1. POST CreateJob：此 API 操作将任务信息提交给提供商进行处理。这些信息与任务类型有关；编码或转码、S3 位置、客户提供的架构以及所需的任何其他作业属性。

此 API 返回 JobId，Job 的状态将为以下状态之

—：PENDINGREADY、IN\_PROGRESS、COMPLETE、或 FAILED。

### 编码请求示例

```
POST /jobs
{
  "actionType": "ID_ASSIGNMENT",
  "s3SourceLocation": "string",
  "s3TargetLocation": "string",
  "jobProperties": {
    "assignmentJobProperties": {
      "fieldMappings": [
        {
          "name": "string",
          "type": "NAME"
        }
      ]
    }
  },
  "customerSpecifiedJobProperties": {
    "property1": "string",
    "property2": "string"
  },
  "outputSourceConfiguration": {
```

```
"KMSArn": "string"
}
```

### 示例响应

```
{
  "jobId": "string",
  "status": "PENDING"
}
```

2. POST StartJob : 此 API 让提供者知道要根据JobId提供的内容开始作业。这允许提供者执行CreateJob直到之前StartJob所需的任何验证。

此 API 返回 a JobId、f Status or the Job statusMessage、和statusCode。

### 编码请求示例

```
POST/jobs/{jobId}
{
  "customerSpecifiedJobProperties": {
    "property1": "string",
    "property2": "string"
  }
}
```

### 示例响应

```
{
  "jobId": "string",
  "status": "PENDING",
  "statusMessage": "string",
  "statusCode": 200
}
```

3. GET GetJob : 此 API 会通知任务 AWS Entity Resolution 数据匹配服务 是否已完成或任何其他状态。

此 API 返回 a JobId、f Status or the Job statusMessage、和statusCode。

### 编码请求示例

```
GET /jobs/{jobId}
```

## 示例响应

```
{
  "jobId": "string",
  "status": "PENDING",
  "statusMessage": "string",
  "statusCode": 200
}
```

AWS Entity Resolution 数据匹配服务 OpenAPI 规范中提供了 APIs 这些内容的完整定义。

## 同步处理集成

对于具有近乎实时的响应时间、实时响应时间、更高的吞吐量和更高的 TPS 的提供商来说，同步处理解决方案更受青睐。此 AWS Entity Resolution 数据匹配服务 工作流程对数据集进行分区，并行发出多个 API 请求。然后，AWS Entity Resolution 数据匹配服务 工作流程负责将结果写入所需的输出位置。

此过程是使用其中一个 API 定义启用的。AWS Entity Resolution 数据匹配服务 调用提供者端点，该端点可通过 AWS Data Exchange 以下方式获得：

POST AssignIdentities：此 API 使用与该记录recordFields关联的source\_id标识符向提供商发送数据。

此 API 返回assignedRecords.

## 编码请求示例

```
POST /assignment
{
  "sourceRecords": [
    {
      "sourceId": "string",
      "recordFields": [
        {
          "name": "string",
          "type": "NAME",

```

```
        "value": "string"
      }
    ]
  }
]
```

## 示例响应

```
{
  "assignedRecords": [
    {
      "sourceRecord": {
        "sourceId": "string",
        "recordFields": [
          {
            "name": "string",
            "type": "NAME",
            "value": "string"
          }
        ]
      },
      "identity": any
    }
  ]
}
```

AWS Entity Resolution 数据匹配服务 OpenAPI 规范中提供了 APIs 这些内容的完整定义。

根据提供商选择的方式，AWS Entity Resolution 数据匹配服务 将为该提供者创建用于启动编码或转码的配置。此外，客户还可以使用 APIs 提供的配置来使用这些配置 AWS Entity Resolution 数据匹配服务。

此配置可使用 Amazon 资源名称 (ARN) 进行访问，该名称源自提供商服务的托管位置和提供商服务的类型。AWS Data Exchange AWS Entity Resolution 数据匹配服务 将此 ARN 称为 `providerServiceARN`

## 测试提供商集成

虽然 AWS Entity Resolution 数据匹配服务 托管数据匹配服务，但提供商集成是 end-to-end 匹配工作流程的重要第三方组件。已经为提供商定义 AWS Entity Resolution 数据匹配服务 了几项测试，可以在集

成失败时增加安全保障。这种方法为提供商提供了根据这些 end-to-end 测试案例监控其服务运行状况的机会。

提供商可以使用 AWS Entity Resolution 数据匹配服务 软件开发套件 (SDK) 使用自己的 end-to-end 测试账户和自己的数据来运行这些测试用例。如果提供商提出任何问题，请 AWS Entity Resolution 数据匹配服务 使用首选的上报路径将问题上报。此外，提供者需要对测试结果实施自己的监控。提供者需要与 AWS 账户 IDs 共享用于运行这些测试的内容 AWS Entity Resolution 数据匹配服务。

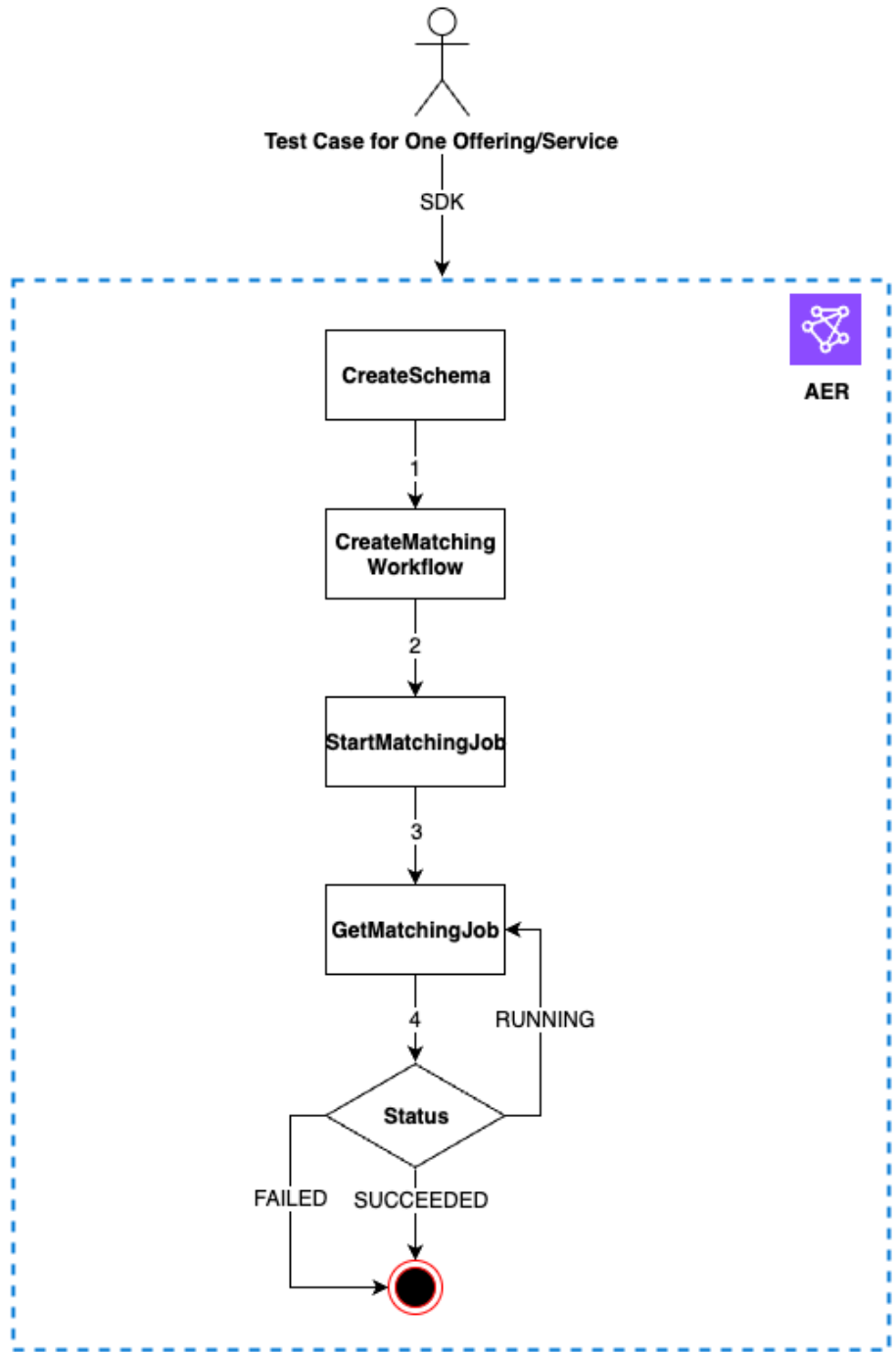
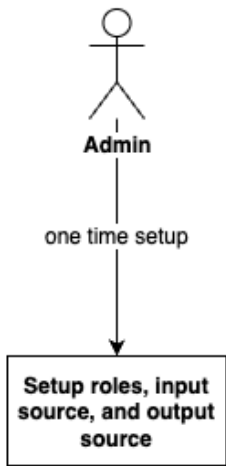
成功运行意味着提供商可以设置自己的数据，通过使用自己的服务，任务状态将返回 C AWS Entity Resolution 数据匹配服务 completed 而不会出现任何错误。这可以通过使用 APIs 提供的以编程方式完成。AWS Entity Resolution 数据匹配服务

例如，提供商可以根据其服务设置其 S3 存储桶、输入源、角色、架构和工作流程。这些设置完成后，提供商可以每天运行一次这些工作流程，其中包含200条记录，以测试其服务。在这种方法中，提供商使用他们选择的 SDK，并对通过 end-to-end 测试帐号提供的服务 AWS Data Exchange 进行测试。供应商应针对其每项产品或服务进行这些测试。

#### Note

提供商需要提供 AWS Entity Resolution 数据匹配服务 他们accountId)用来运行这些工作流程进行测试的 AWS 账户 ID。此外，提供商需要监控这些测试并确保它们通过，这意味着提供商需要在失败时启用通知并相应地解决问题。

下图显示了一个典型 end-to-end 的工作流程测试用例。



### 测试提供商集成

1. ( 一次性设置 ) AWS Entity Resolution 数据匹配服务 按照中的步骤为设置资源 [设置 AWS Entity Resolution 数据匹配服务](#)。

完成一次性设置过程后，应准备好角色、数据和数据源。现在，您已准备好使用 AWS Entity Resolution 数据匹配服务 控制台或测试提供商集成 APIs。

## 2. 使用 AWS Entity Resolution 数据匹配服务 APIs 或控制台测试提供商集成。

### API

要测试提供商集成，请使用 AWS Entity Resolution 数据匹配服务 APIs

1. 使用 [CreateSchemaMapping API](#) 创建架构映射。有关支持的编程语言的完整列表，请参阅 [CreateSchemaMapping API](#) 的“[另请参阅](#)”部分。

架构映射是告诉您 AWS Entity Resolution 数据匹配服务 如何解释数据以进行匹配的过程。您可以定义输入数据表的架构，您希望 AWS Entity Resolution 将其读入匹配的工作流程。

创建架构映射时，必须为 AWS Entity Resolution 读取的每行输入数据指定并分配一个[唯一标识符](#)。例如：Primary\_key、Row\_ID、Record\_ID。

Example为包含id和的数据源创建架构映射 email

以下是包含id和的数据源的架构映射示例email：

```
[
  {
    "fieldName": "id",
    "type": "UNIQUE_ID"
  },
  {
    "fieldName": "email",
    "type": "EMAIL_ADDRESS"
  }
]
```

Example为包含id和email使用 Java SDK 的数据源创建架构映射

以下是包含id并email使用 Java SDK 的数据源的架构映射示例：

```
EntityResolutionClient.createSchemaMapping(
    CreateSchemaMappingRequest.builder()
        .schemaName(<schema-name>)
        .mappedInputFields([
```

```

SchemaInputAttribute.builder().fieldName("id").type("UNIQUE_ID").build(),
SchemaInputAttribute.builder().fieldName("email").type("EMAIL_ADDRESS").build()
    ])
    .build()
)

```

2. 使用 [CreateMatchingWorkflow API](#) 创建匹配的工作流程。有关支持的编程语言的完整列表，请参阅 [CreateMatchingWorkflow API](#) 的“另请参阅”部分。

Example使用 Java 开发工具包创建匹配的工作流程

以下是使用 Java SDK 的匹配工作流程的示例：

```

EntityResolutionClient.createMatchingWorkflow(
    CreateMatchingWorkflowRequest.builder()
        .workflowName(<workflow-name>)
        .inputSourceConfig(
            InputSource.builder().inputSourceARN(<glue-inputsource-from-
            step1>).schemaName(<schema-name-from-step2>).build()
        )
        .outputSourceConfig(OutputSource.builder().outputS3Path(<output-s3-
        path>).output(<output-1>, <output-2>, <output-3>).build())
        .resolutionTechniques(ResolutionTechniques.builder()
            .resolutionType(PROVIDER)
            .providerProperties(ProviderProperties.builder()
                .providerServiceArn(<provider-arn>)
                .providerConfiguration(<configuration-
                depending-on-service>)
            .intermediateSourceConfiguration(<intermedaite-s3-path>)
            .build())
    )
)

```

```
.build()
                                .roleArn(<role-from-step1>)
                                .build()
)
```

设置匹配的工作流程后，您可以运行工作流程。

3. 使用 [StartMatchingJob API](#) 运行匹配的工作流程。要运行匹配的工作流程，您必须使用 `CreateMatchingWorkflow` 端点创建了匹配的工作流程。

有关支持的编程语言的完整列表，请参阅 [StartMatchingJob API](#) 的“另请参阅”部分。

Example 使用 Java SDK 运行匹配的工作流程

以下是使用 Java SDK 运行匹配工作流程的示例：

```
EntityResolutionClient.startMatchingJob(StartMatchingJobRequest.builder()
    .workflowName(<name-of-workflow-from-step3>)
    .build()
)
```

4. 使用 [GetMatchingJob API](#) 监控工作流程的状态。

此 API 返回与任务关联的状态、指标和错误（如果有）。

Example 使用 Java SDK 监控匹配的工作流程

以下是使用 Java SDK 监控匹配工作流程作业的示例：

```
EntityResolutionClient.getMatchingJob(GetMatchingJobRequest.builder()
    .workflowName(<name-of-workflow-from-step3>)
    .jobId(jobId-from-startMatchingJob)
    .build()
)
```

如果工作流程成功完成，则 end-to-end 测试已完成。

## Console

使用 AWS Entity Resolution 数据匹配服务 控制台测试提供商集成

1. 按照中的步骤创建架构映射[创建架构映射](#)。

架构映射是告诉您 AWS Entity Resolution 数据匹配服务 如何解释数据以进行匹配的过程。您可以定义要 AWS Entity Resolution 数据匹配服务 读入匹配工作流程的输入数据表的架构。

创建架构映射时，必须为 AWS Entity Resolution 数据匹配服务 读取的每一行输入数据指定并分配一个[唯一标识符](#)。例如：Primary\_key、Row\_ID、Record\_ID。

Example包含id和的数据源的架构映射 email

以下是包含id和的数据源的架构映射示例email：

```
[
  {
    "fieldName": "id",
    "type": "UNIQUE_ID"
  },
  {
    "fieldName": "email",
    "type": "EMAIL_ADDRESS"
  }
]
```

2. 按照中的步骤创建并运行匹配的工作流程[创建基于提供商服务的匹配工作流程](#)。

创建匹配工作流是您设置的过程，用于指定要匹配的输入数据以及如何执行匹配。在基于提供商的工作流程中，如果账户通过订阅了提供商服务 AWS Data Exchange，则可以将您的已知标识符与您的首选提供商进行匹配。根据您的用于执行端到端测试的提供商和服务，您可以相应地配置匹配的工作流程。

AWS Entity Resolution 数据匹配服务 控制台将创建和运行的操作组合在一个按钮中。选择“创建并运行”后，将显示一条消息，表明匹配的工作流程已创建且作业已启动。

3. 在“匹配工作流程”页面上监控工作流程的状态。

如果工作流成功完成（Job 状态为“已完成”），则 end-to-end测试已完成。

在匹配工作流程详细信息页面的指标选项卡上，您可以在“上次任务指标”下查看以下内容：

- Job ID。
- 匹配 workflow 作业的状态：已排队、进行中、已完成、失败
- workflow 作业的完成时间。
- 已处理的记录数。
- 未处理的记录数。
- IDs 生成的唯一匹配项。
- 输入记录的数量。

您还可以查看任务历史记录下先前运行过的匹配 workflow 作业的作业指标。

# 安全性 AWS Entity Resolution 数据匹配服务

云安全 AWS 是重中之重。作为 AWS 客户，您可以受益于专为满足大多数安全敏感型组织的要求而构建的数据中心和网络架构。

安全是双方共同承担 AWS 的责任。[责任共担模式](#)将其描述为云的安全性和云中的安全性：

- 云安全 — AWS 负责保护在云 AWS 服务 中运行的基础架构 AWS Cloud。AWS 还为您提供可以安全使用的服务。作为[AWS 合规计划](#)的一部分，第三方审计师定期测试和验证我们安全的有效性。要了解适用的合规计划 AWS Entity Resolution 数据匹配服务，请参阅[按合规计划划分的范围内的AWSAWS 服务按合规计划](#)。
- 云端安全 — 您的责任由您 AWS 服务 使用的内容决定。您还需要对其他因素负责，包括您的数据的敏感性、您的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时如何应用分担责任模型 AWS Entity Resolution 数据匹配服务。以下主题向您介绍如何进行配置 AWS Entity Resolution 数据匹配服务 以满足您的安全和合规性目标。您还将学习如何使用其他 AWS 服务 方法来监控和保护您的 AWS Entity Resolution 数据匹配服务 资源。

## 主题

- [中的数据保护 AWS Entity Resolution 数据匹配服务](#)
- [的身份和访问管理 AWS Entity Resolution 数据匹配服务](#)
- [的合规性验证 AWS Entity Resolution 数据匹配服务](#)
- [韧性在 AWS Entity Resolution 数据匹配服务](#)

## 中的数据保护 AWS Entity Resolution 数据匹配服务

AWS [责任共担模式](#)适用于保护 AWS Entity Resolution 数据匹配服务中的数据。如本模型所述 AWS ，负责保护运行所有内容的全球基础架构 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AWS](#)条款。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](#)。

出于数据保护目的，我们建议您保护 AWS 账户 凭证并使用 AWS IAM Identity Center 或 AWS Identity and Access Management (IAM) 设置个人用户。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 )。

- 用于 SSL/TLS 与 AWS 资源通信。我们要求使用 TLS 1.2，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 CloudTrail 跟踪捕获 AWS 活动的信息，请参阅《AWS CloudTrail 用户指南》中的[使用跟 CloudTrail 跟踪](#)。
- 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（例如 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 AWS 通过命令行界面或 API 进行访问时需要经过 FIPS 140-3 验证的加密模块，请使用 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（如您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）。这包括您使用控制台、API AWS Entity Resolution 数据匹配服务 或 SDK 或以其他 AWS 服务 方式使用控制台 AWS CLI、API 或 AWS SDK 的情况。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。

## 静态数据加密 AWS Entity Resolution 数据匹配服务

AWS Entity Resolution 数据匹配服务 默认提供加密，以使用 AWS 自有的加密密钥保护敏感的静态客户数据。

AWS 拥有的密钥 — 默认 AWS Entity Resolution 数据匹配服务 使用这些密钥自动加密个人身份数据。您无法查看、管理或使用 AWS 自有密钥，也无法审核其使用情况。但是，您无需采取任何措施来保护加密数据的密钥。有关更多信息，请参阅《AWS Key Management Service 开发人员指南》中的[AWS 拥有的密钥](#)。

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，您可以使用它来构建符合严格加密合规性和监管要求的安全应用程序。

或者，您也可以在创建匹配的工作流程资源时提供客户托管的 KMS 密钥进行加密。

客户托管密钥- AWS Entity Resolution 数据匹配服务 支持使用由您创建、拥有和管理的对称客户托管 KMS 密钥来加密您的敏感数据。由于您可以完全控制这层加密，因此可以执行以下任务：

- 制定和维护关键策略
- 建立和维护 IAM 策略和授权
- 启用和禁用密钥策略

- 轮换加密材料
- 添加 标签
- 创建密钥别名
- 计划删除密钥

有关更多信息，请参阅《AWS Key Management Service 开发人员指南》中的[客户托管密钥](#)。

有关更多信息 AWS KMS，请参阅[什么是 AWS Key Management Service ?](#)

## 密钥管理

### 操作方法 AWS Entity Resolution 数据匹配服务 将补助金用于 AWS KMS

AWS Entity Resolution 数据匹配服务 需要获得[授权](#)才能使用您的客户托管密钥。当您创建使用客户托管密钥加密的匹配工作流程时，AWS Entity Resolution 数据匹配服务 会通过向发送[CreateGrant](#)请求来代表您创建授权 AWS KMS。中的授权 AWS KMS 用于授予对客户账户中的 KMS 密钥的 AWS Entity Resolution 数据匹配服务 访问权限。AWS Entity Resolution 数据匹配服务 需要获得授权才能使用您的客户托管密钥进行以下内部操作：

- 向发送[GenerateDataKey](#)请求 AWS KMS 以生成由您的客户托管密钥加密的数据密钥。
- 将 [Decrypt](#) 请求发送 AWS KMS 到以解密加密的数据密钥，以便它们可用于加密您的数据。

您可以随时撤销授予访问权限，或删除服务对客户托管密钥的访问权限。如果这样做，将 AWS Entity Resolution 数据匹配服务 无法访问由客户托管密钥加密的任何数据，这会影响到依赖该数据的操作。例如，如果您通过授权删除了对密钥的服务访问权限，并尝试为使用客户密钥加密的匹配工作流程启动任务，则该操作将返回AccessDeniedException错误。

### 创建客户托管的密钥

您可以使用或 AWS KMS API 创建对称的客户托管密钥。AWS 管理控制台

#### 创建对称的客户托管密钥

AWS Entity Resolution 数据匹配服务 支持使用[对称加密 KMS 密钥进行加密](#)。按照《AWS Key Management Service 开发人员指南》中[创建对称的客户托管密钥](#)的步骤进行操作。

### 关键政策声明

密钥策略控制对客户托管密钥的访问。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《AWS Key Management Service 开发人员指南》中的[管理客户托管密钥的访问权限](#)。

要将客户托管密钥 AWS Entity Resolution 数据匹配服务 用于您的资源，必须在密钥策略中允许以下 API 操作：

- [kms:DescribeKey](#)— 提供密钥 ARN、创建日期（以及删除日期，如果适用）、密钥状态以及密钥材料的来源和到期日期（如果有）等信息。它包括可帮助您区分不同类型的 KMS 密钥的字段（例如 KeySpec）。它还显示密钥的使用情况（加密、签名或生成和验证 MAC）以及 KMS 密钥支持的算法。AWS Entity Resolution 数据匹配服务 验证是 SYMMETRIC\_DEFAULT 和 KeySpecKeyUsage 是 ENCRYPT\_DECRYPT。
- [kms:CreateGrant](#)：向客户托管密钥添加授权。授予对指定 KMS 密钥的控制访问权限，从而允许对[授予操作](#) AWS Entity Resolution 数据匹配服务 所需的访问权限。有关[使用授权](#)的更多信息，请参阅《AWS Key Management Service 开发人员指南》。

这 AWS Entity Resolution 数据匹配服务 允许执行以下操作：

- 调用 GenerateDataKey 生成加密的数据密钥并将其存储，因为数据密钥不会立即用于加密。
- 调用 Decrypt 使用存储的加密数据密钥访问加密数据。
- 设置停用主体，以允许服务 RetireGrant。

以下是您可以为其添加的策略声明示例 AWS Entity Resolution 数据匹配服务：

```
{
  "Sid" : "Allow access to principals authorized to use AWS Entity Resolution",
  "Effect" : "Allow",
  "Principal" : {
    "AWS" : "*"
  },
  "Action" : ["kms:DescribeKey","kms:CreateGrant"],
  "Resource" : "*",
  "Condition" : {
    "StringEquals" : {
      "kms:ViaService" : "entityresolution.region.amazonaws.com",
      "kms:CallerAccount" : "111122223333"
    }
  }
}
```

```
}
```

## 用户的权限

将 KMS 密钥配置为加密的默认密钥时，默认 KMS 密钥策略允许任何有权访问所需 KMS 操作的用户使用此 KMS 密钥来加密或解密资源。要使用客户托管的 KMS 密钥加密，您必须向用户授予调用以下操作的权限：

- kms:CreateGrant
- kms:Decrypt
- kms:DescribeKey
- kms:GenerateDataKey

在[CreateMatchingWorkflow](#)请求期间，AWS Entity Resolution 数据匹配服务 将代表您 AWS KMS 向发送[DescribeKey](#)和[CreateGrant](#)请求。这将要求使用客户托管的 KMS 密钥CreateMatchingWorkflow提出请求的 IAM 实体拥有 KMS 密钥策略的kms:DescribeKey权限。

在 an [CreateIdMappingWorkflow](#) [StartIdMappingJob](#)请求期间，AWS Entity Resolution 数据匹配服务 将 AWS KMS 代表您向发送[DescribeKey](#)和[CreateGrant](#)请求。这将要求使用客户托管的 KMS 密钥进行CreateIdMappingWorkflow和StartIdMappingJob请求的 IAM 实体拥有 KMS 密钥策略的kms:DescribeKey权限。提供商将能够访问客户托管的密钥来解密 AWS Entity Resolution 数据匹配服务 Amazon S3 存储桶中的数据。

以下是您可以添加的政策声明示例，供提供商解密 AWS Entity Resolution 数据匹配服务 Amazon S3 存储桶中的数据：

## JSON

```
{
  "Version": "2012-10-17",
  "Statement": [{
    "Effect": "Allow",
    "Principal": {
      "AWS": "arn:aws:iam::715724997226:root"
    },
    "Action": [
      "kms:Decrypt"
    ]
  }
]
```

```

    ],
    "Resource": "arn:aws:kms:us-east-1:111122223333:key/key-id",
    "Condition": {
      "StringEquals": {
        "kms:ViaService": "s3.us-east-1.amazonaws.com"
      }
    }
  }
}
}
}

```

将每个 *<user input placeholder>* 替换为您自己的信息。

*<KMSKeyARN>*

AWS KMS 亚马逊资源名称。

同样，调用 [StartMatchingJobAPI](#) 的 IAM 实体必须拥有匹配工作流程中提供的客户托管 KMS 密钥 `kms:Decrypt` 和 `kms:GenerateDataKey` 权限。

有关在 [策略中指定权限](#) 的更多信息，请参阅《AWS Key Management Service 开发人员指南》。

有关 [密钥访问疑难解答](#) 的更多信息，请参阅《AWS Key Management Service 开发人员指南》。

## 为指定客户管理的密钥 AWS Entity Resolution 数据匹配服务

您可以指定客户托管密钥作为以下资源的第二层加密：

[匹配工作流程](#)-创建匹配的工作流程资源时，您可以通过输入 KMSARN 来指定数据密钥，该密钥 AWS Entity Resolution 数据匹配服务 用于加密资源存储的可识别个人数据。

`kmSarn` — 输入密钥 ARN，这是客户托管 [密钥的密钥标识符](#)。AWS KMS

如果您要在两个资源之间创建或运行 ID 映射工作流程，则可以将客户托管密钥指定为以下资源的第二层加密 AWS 账户：

[ID 映射工作流程](#) 或 [启动 ID 映射工作流程](#)-创建 ID 映射工作流资源或启动 ID 映射工作流程作业时，您可以通过输入 `kmSarn` 来指定数据密钥，该密钥 AWS Entity Resolution 数据匹配服务 用于加密资源存储的可识别个人数据。

`kmSarn` — 输入密钥 ARN，这是客户托管 [密钥的密钥标识符](#)。AWS KMS

## 监控您的加密密钥 AWS Entity Resolution 数据匹配服务 服务

当您在 AWS Entity Resolution 数据匹配服务 服务资源中使用 AWS KMS 客户托管密钥时，您可以使用 [AWS CloudTrail](#) 或 [Amazon CloudWatch Logs](#) 来跟踪 AWS Entity Resolution 数据匹配服务 发送到的请求 AWS KMS。

以下示例是 CreateGrant、GenerateDataKeyDecrypt、和监控 DescribeKey 为访问由 AWS Entity Resolution 数据匹配服务 您的客户托管密钥加密的数据而调用的 AWS KMS 操作 AWS CloudTrail 的事件：

### 主题

- [CreateGrant](#)
- [DescribeKey](#)
- [GenerateDataKey](#)
- [Decrypt](#)

### CreateGrant

当您使用 AWS KMS 客户托管密钥加密匹配的工作流程资源时，AWS Entity Resolution 数据匹配服务 会代表您发送 CreateGrant 请求以访问您中的 KMS 密钥 AWS 账户。AWS Entity Resolution 数据匹配服务 创建的授权特定于与 AWS KMS 客户托管密钥关联的资源。此外，在您删除资源时，AWS Entity Resolution 数据匹配服务 使用 RetireGrant 操作来移除授权。

以下示例事件记录了 CreateGrant 操作：

```
{
  "eventVersion": "1.08",
  "userIdentity": {
    "type": "AssumedRole",
    "principalId": "AROAIQDTESTANDEXAMPLE:Sampleuser01",
    "arn": "arn:aws:sts::111122223333:assumed-role/Admin/Sampleuser01",
    "accountId": "111122223333",
    "accessKeyId": "AKIAIOSFODNN7EXAMPLE3",
    "sessionContext": {
      "sessionIssuer": {
        "type": "Role",
        "principalId": "AROAIQDTESTANDEXAMPLE:Sampleuser01",
        "arn": "arn:aws:sts::111122223333:assumed-role/Admin/Sampleuser01",
        "accountId": "111122223333",
```

```

        "userName": "Admin"
    },
    "webIdFederationData": {},
    "attributes": {
        "mfaAuthenticated": "false",
        "creationDate": "2021-04-22T17:02:00Z"
    }
},
"invokedBy": "entityresolution.amazonaws.com"
},
"eventTime": "2021-04-22T17:07:02Z",
"eventSource": "kms.amazonaws.com",
"eventName": "CreateGrant",
"awsRegion": "us-west-2",
"sourceIPAddress": "172.12.34.56",
"userAgent": "ExampleDesktop/1.0 (V1; OS)",
"requestParameters": {
    "retiringPrincipal": "entityresolution.region.amazonaws.com",
    "operations": [
        "GenerateDataKey",
        "Decrypt",
    ],
    "keyId": "arn:aws:kms:us-
west-2:111122223333:key/1234abcd-12ab-34cd-56ef-123456SAMPLE",
    "granteePrincipal": "entityresolution.region.amazonaws.com"
},
"responseElements": {
    "grantId":
"0ab0ac0d0b000f00ea00cc0a0e00fc00bce000c000f0000000c0bc0a0000aaafSAMPLE",
    "keyId": "arn:aws:kms:us-
west-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-
west-2:111122223333:key/1234abcd-12ab-34cd-56ef-123456SAMPLE"
    }
],
"eventType": "AwsApiCall",

```

```
"managementEvent": true,  
"eventCategory": "Management",  
"recipientAccountId": "111122223333"  
}
```

## DescribeKey

AWS Entity Resolution 数据匹配服务 使用该DescribeKey操作来验证与您的匹配资源关联的 AWS KMS 客户托管密钥是否存在于账户和区域中。

以下示例事件记录了 DescribeKey 操作：

```
{  
  "eventVersion": "1.08",  
  "userIdentity": {  
    "type": "AssumedRole",  
    "principalId": "AROAIQDTESTANDEXAMPLE:Sampleuser01",  
    "arn": "arn:aws:sts::111122223333:assumed-role/Admin/Sampleuser01",  
    "accountId": "111122223333",  
    "accessKeyId": "AKIAIOSFODNN7EXAMPLE3",  
    "sessionContext": {  
      "sessionIssuer": {  
        "type": "Role",  
        "principalId": "AROAIQDTESTANDEXAMPLE:Sampleuser01",  
        "arn": "arn:aws:sts::111122223333:assumed-role/Admin/Sampleuser01",  
        "accountId": "111122223333",  
        "userName": "Admin"  
      },  
      "webIdFederationData": {},  
      "attributes": {  
        "mfaAuthenticated": "false",  
        "creationDate": "2021-04-22T17:02:00Z"  
      }  
    },  
    "invokedBy": "entityresolution.amazonaws.com"  
  },  
  "eventTime": "2021-04-22T17:07:02Z",  
  "eventSource": "kms.amazonaws.com",  
  "eventName": "DescribeKey",  
  "awsRegion": "us-west-2",  
  "sourceIPAddress": "172.12.34.56",  
  "userAgent": "ExampleDesktop/1.0 (V1; OS)",  
  "requestParameters": {
```

```

    "keyId": "00dd0db0-0000-0000-ac00-b0c000SAMPLE"
  },
  "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-
west-2:111122223333:key/1234abcd-12ab-34cd-56ef-123456SAMPLE"
    }
  ],
  "eventType": "AwsApiCall",
  "managementEvent": true,
  "eventCategory": "Management",
  "recipientAccountId": "111122223333"
}

```

## GenerateDataKey

当您为匹配的工作流程资源启用 AWS KMS 客户托管密钥时，AWS Entity Resolution 数据匹配服务会通过亚马逊简单存储服务 (Amazon S3) Simple Service 向 AWS KMS 其发送GenerateDataKey请求，指定 AWS KMS 该资源的客户托管密钥。

以下示例事件记录了 GenerateDataKey 操作：

```

{
  "eventVersion": "1.08",
  "userIdentity": {
    "type": "AWSService",
    "invokedBy": "s3.amazonaws.com"
  },
  "eventTime": "2021-04-22T17:07:02Z",
  "eventSource": "kms.amazonaws.com",
  "eventName": "GenerateDataKey",
  "awsRegion": "us-west-2",
  "sourceIPAddress": "172.12.34.56",
  "userAgent": "ExampleDesktop/1.0 (V1; OS)",
  "requestParameters": {
    "keySpec": "AES_256",

```

```

    "keyId": "arn:aws:kms:us-
west-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-
west-2:111122223333:key/1234abcd-12ab-34cd-56ef-123456SAMPLE"
    }
  ],
  "eventType": "AwsApiCall",
  "managementEvent": true,
  "eventCategory": "Management",
  "recipientAccountId": "111122223333",
  "sharedEventID": "57f5dbee-16da-413e-979f-2c4c6663475e"
}

```

## Decrypt

当您为匹配的工作流程资源启用 AWS KMS 客户托管密钥时，AWS Entity Resolution 数据匹配服务会通过亚马逊简单存储服务 (Amazon S3) Simple Service 向 AWS KMS 其发送 Decrypt 请求，指定 AWS KMS 该资源的客户托管密钥。

以下示例事件记录了 Decrypt 操作：

```

{
  "eventVersion": "1.08",
  "userIdentity": {
    "type": "AWSService",
    "invokedBy": "s3.amazonaws.com"
  },
  "eventTime": "2021-04-22T17:10:51Z",
  "eventSource": "kms.amazonaws.com",
  "eventName": "Decrypt",
  "awsRegion": "us-west-2",
  "sourceIPAddress": "172.12.34.56",
  "userAgent": "ExampleDesktop/1.0 (V1; OS)",
  "requestParameters": {

```

```
    "keyId": "arn:aws:kms:us-west-2:111122223333:key/1234abcd-12ab-34cd-56ef-123456SAMPLE",
    "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-west-2:111122223333:key/1234abcd-12ab-34cd-56ef-123456SAMPLE"
    }
  ],
  "eventType": "AwsApiCall",
  "managementEvent": true,
  "eventCategory": "Management",
  "recipientAccountId": "111122223333",
  "sharedEventID": "dc129381-1d94-49bd-b522-f56a3482d088"
}
```

## 注意事项

AWS Entity Resolution 数据匹配服务 不支持使用新的客户托管 KMS 密钥更新匹配的工作流程。在这种情况下，您可以使用客户托管的 KMS 密钥创建新的工作流程。

## 了解详情

以下资源提供有关静态数据加密的更多信息。

有关 [AWS Key Management Service 基本概念](#)的更多信息，请参阅AWS Key Management Service 开发人员指南。

有关 [AWS Key Management Service 安全最佳实践](#)的更多信息，请参阅AWS Key Management Service 开发人员指南。

## AWS Entity Resolution 数据匹配服务 使用接口端点进行访问 (AWS PrivateLink)

您可以使用 AWS PrivateLink 在您的 VPC 和之间创建私有连接 AWS Entity Resolution 数据匹配服务。您可以像在 VPC 中 AWS Entity Resolution 数据匹配服务 一样进行访问，无需使用互联网网关、NAT 设备、VPN 连接或 Direct Connect 连接。VPC 中的实例不需要公有 IP 地址即可访问 AWS Entity Resolution 数据匹配服务。

您可以通过创建由 AWS PrivateLink 提供支持的接口端点来建立此私有连接。我们将在您为接口端点启用的每个子网中创建一个端点网络接口。这些是请求者托管的网络接口，用作发往 AWS Entity Resolution 数据匹配服务的流量的入口点。

有关更多信息，请参阅 AWS PrivateLink 指南 AWS PrivateLink 中的 [AWS 服务 通过访问](#)。

### 的注意事项 AWS Entity Resolution 数据匹配服务

在为设置接口端点之前 AWS Entity Resolution 数据匹配服务，请查看 AWS PrivateLink 指南中的 [注意事项](#)。

AWS Entity Resolution 数据匹配服务 支持通过接口端点调用其所有 API 操作。

支持 VPC 终端节点策略 AWS Entity Resolution 数据匹配服务。默认情况下，允许通过接口端点对 AWS Entity Resolution 数据匹配服务 进行完全访问。或者，您可以将安全组与端点网络接口关联，以控制通过接口端点流向 AWS Entity Resolution 数据匹配服务 的流量。

### 为创建接口终端节点 AWS Entity Resolution 数据匹配服务

您可以创建用于 AWS Entity Resolution 数据匹配服务 使用 Amazon VPC 控制台或 AWS Command Line Interface (AWS CLI) 的接口终端节点。有关更多信息，请参阅《AWS PrivateLink 指南》中的 [创建接口端点](#)。

AWS Entity Resolution 数据匹配服务 使用以下服务名称创建接口终端节点：

```
com.amazonaws.region.entityresolution
```

AWS Entity Resolution 数据匹配服务 还支持符合 FIPS ( 联邦信息处理标准 ) 的端点。要使用 FIPS 终端节点，请使用以下服务名称：

```
com.amazonaws.region.entityresolution-fips
```

如果为接口端点启用私有 DNS，则可使用其默认区域 DNS 名称向 AWS Entity Resolution 数据匹配服务发出 API 请求。例如 `entityresolution.us-east-1.amazonaws.com`。

## 为 VPC 端点创建端点策略

端点策略是一种 IAM 资源，您可以将其附加到接口端点。默认终端节点策略允许 AWS Entity Resolution 数据匹配服务通过接口终端节点进行完全访问。要控制允许 AWS Entity Resolution 数据匹配服务从您的 VPC 访问权限，请将自定义终端节点策略附加到接口终端节点。

端点策略指定以下信息：

- 可执行操作的主体（AWS 账户、IAM 用户和 IAM 角色）。
- 可执行的操作。
- 可对其执行操作的资源。

有关更多信息，请参阅《AWS PrivateLink 指南》中的[使用端点策略控制对服务的访问权限](#)。

示例：用于 AWS Entity Resolution 数据匹配服务操作的 VPC 终端节点策略

以下是自定义端点策略的示例。当您将此策略附加到接口终端节点时，它会授予所有委托人对所有资源 AWS Entity Resolution 数据匹配服务执行所列操作的访问权限。

```
{
  "Statement": [
    {
      "Principal": "*",
      "Effect": "Allow",
      "Action": [
        "entityresolution:CreateMatchingWorkflow",
        "entityresolution:StartMatchingJob",
        "entityresolution:GetMatchingJob"
      ],
      "Resource": "*"
    }
  ]
}
```

# 的身份和访问管理 AWS Entity Resolution 数据匹配服务

AWS Identity and Access Management (IAM) AWS 服务 可帮助管理员安全地控制对 AWS 资源的访问权限。IAM 管理员控制谁可以进行身份验证（登录）和授权（拥有权限）使用 AWS Entity Resolution 数据匹配服务 资源。您可以使用 IAM AWS 服务 ，无需支付额外费用。

## Note

AWS Entity Resolution 数据匹配服务 支持跨账户政策。有关更多信息，请参阅《IAM 用户指南》中的 [IAM 中的跨账户资源访问](#)。

## 主题

- [受众](#)
- [使用身份进行身份验证](#)
- [使用策略管理访问](#)
- [如何 AWS Entity Resolution 数据匹配服务 与 IAM 配合使用](#)
- [基于身份的策略示例 AWS Entity Resolution 数据匹配服务](#)
- [AWS 的托管策略 AWS Entity Resolution 数据匹配服务](#)
- [对 AWS Entity Resolution 数据匹配服务 身份和访问进行故障排除](#)

## 受众

您的使用方式 AWS Identity and Access Management (IAM) 因您的角色而异：

- 服务用户：如果您无法访问功能，请从管理员处请求权限（请参阅[对 AWS Entity Resolution 数据匹配服务 身份和访问进行故障排除](#)）
- 服务管理员：确定用户访问权限并提交权限请求（请参阅[如何 AWS Entity Resolution 数据匹配服务 与 IAM 配合使用](#)）
- IAM 管理员：编写用于管理访问权限的策略（请参阅[基于身份的策略示例 AWS Entity Resolution 数据匹配服务](#)）

## 使用身份进行身份验证

身份验证是您 AWS 使用身份凭证登录的方式。您必须以 IAM 用户身份进行身份验证 AWS 账户根用户，或者通过担任 IAM 角色进行身份验证。

您可以使用来自身份源的证书 AWS IAM Identity Center（例如（IAM Identity Center）、单点登录身份验证或 Google/Facebook 证书，以联合身份登录。有关登录的更多信息，请参阅《AWS 登录 用户指南》中的[如何登录您的 AWS 账户](#)。

对于编程访问，AWS 提供 SDK 和 CLI 来对请求进行加密签名。有关更多信息，请参阅《IAM 用户指南》中的[适用于 API 请求的 AWS 签名版本 4](#)。

### AWS 账户 root 用户

创建时 AWS 账户，首先会有一个名为 AWS 账户 root 用户的登录身份，该身份可以完全访问所有资源 AWS 服务和资源。我们强烈建议不要使用根用户进行日常任务。有关需要根用户凭证的任务，请参阅《IAM 用户指南》中的[需要根用户凭证的任务](#)。

### 联合身份

作为最佳实践，要求人类用户使用与身份提供商的联合身份验证才能 AWS 服务 使用临时证书进行访问。

联合身份是指来自您的企业目录、Web 身份提供商的用户 Directory Service，或者 AWS 服务 使用来自身份源的凭据进行访问的用户。联合身份代入可提供临时凭证的角色。

要集中管理访问权限，建议使用。AWS IAM Identity Center 有关更多信息，请参阅《AWS IAM Identity Center 用户指南》中的[什么是 IAM Identity Center？](#)。

### IAM 用户和群组

[IAM 用户](#)是对某个人员或应用程序具有特定权限的一个身份。建议使用临时凭证，而非具有长期凭证的 IAM 用户。有关更多信息，请参阅 IAM 用户指南中的[要求人类用户使用身份提供商的联合身份验证才能 AWS 使用临时证书进行访问](#)。

[IAM 组](#)指定一组 IAM 用户，便于更轻松地对大量用户进行权限管理。有关更多信息，请参阅《IAM 用户指南》中的[IAM 用户使用案例](#)。

## IAM 角色

[IAM 角色](#)是具有特定权限的身份，可提供临时凭证。您可以通过[从用户切换到 IAM 角色 \(控制台\)](#)或调用 AWS CLI 或 AWS API 操作来代入角色。有关更多信息，请参阅《IAM 用户指南》中的[担任角色的方法](#)。

IAM 角色对于联合用户访问、临时 IAM 用户权限、跨账户访问、跨服务访问以及在 Amazon EC2 上运行的应用程序非常有用。有关更多信息，请参阅《IAM 用户指南》中的[IAM 中的跨账户资源访问](#)。

## 使用策略管理访问

您可以 AWS 通过创建策略并将其附加到 AWS 身份或资源来控制中的访问权限。策略定义了与身份或资源关联时的权限。AWS 在委托人提出请求时评估这些政策。大多数策略都以 JSON 文档的 AWS 形式存储在中。有关 JSON 策略文档的更多信息，请参阅《IAM 用户指南》中的[JSON 策略概述](#)。

管理员使用策略，通过定义哪个主体可以在什么条件下对哪些资源执行哪些操作来指定谁有权访问什么。

默认情况下，用户和角色没有权限。IAM 管理员创建 IAM 策略并将其添加到角色中，然后用户可以担任这些角色。IAM 策略定义权限，与执行操作所用的方法无关。

## 基于身份的策略

基于身份的策略是您附加到身份（用户、组或角色）的 JSON 权限策略文档。这些策略控制身份可以执行什么操作、对哪些资源执行以及在什么条件下执行。要了解如何创建基于身份的策略，请参阅《IAM 用户指南》中的[使用客户管理型策略定义自定义 IAM 权限](#)。

基于身份的策略可以是内联策略（直接嵌入到单个身份中）或托管策略（附加到多个身份的独立策略）。要了解如何在托管策略和内联策略之间进行选择，请参阅《IAM 用户指南》中的[在托管策略与内联策略之间进行选择](#)。

## 基于资源的策略

基于资源的策略是附加到资源的 JSON 策略文档。示例包括 IAM 角色信任策略和 Amazon S3 存储桶策略。在支持基于资源的策略的服务中，服务管理员可以使用它们来控制对特定资源的访问。您必须在基于资源的策略中[指定主体](#)。

基于资源的策略是位于该服务中的内联策略。您不能在基于资源的策略中使用 IAM 中的 AWS 托管策略。

## 其他策略类型

AWS 支持其他策略类型，这些策略类型可以设置更常见的策略类型授予的最大权限：

- 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。有关更多信息，请参阅《IAM 用户指南》中的 [IAM 实体的权限边界](#)。
- 服务控制策略 (SCPs)-在中指定组织或组织单位的最大权限 AWS Organizations。有关更多信息，请参阅《AWS Organizations 用户指南》中的 [服务控制策略](#)。
- 资源控制策略 (RCPs)-设置账户中资源的最大可用权限。有关更多信息，请参阅《AWS Organizations 用户指南》中的 [资源控制策略 \(RCPs\)](#)。
- 会话策略 – 在为角色或联合用户创建临时会话时，作为参数传递的高级策略。有关更多信息，请参阅《IAM 用户指南》中的 [会话策略](#)。

## 多个策略类型

当多个类型的策略应用于一个请求时，生成的权限更加复杂和难以理解。要了解在涉及多种策略类型时如何 AWS 确定是否允许请求，请参阅 IAM 用户指南中的 [策略评估逻辑](#)。

## 如何 AWS Entity Resolution 数据匹配服务 与 IAM 配合使用

在使用 IAM 管理访问权限之前 AWS Entity Resolution 数据匹配服务，请先了解有哪些 IAM 功能可供使用 AWS Entity Resolution 数据匹配服务。

您可以搭配使用的 IAM 功能 AWS Entity Resolution 数据匹配服务

IAM 功能	AWS Entity Resolution 数据匹配服务 支持
<a href="#">基于身份的策略</a>	是
<a href="#">基于资源的策略</a>	是
<a href="#">策略操作</a>	是
<a href="#">策略资源</a>	是
<a href="#">策略条件键</a>	是
<a href="#">ACLs</a>	否

IAM 功能	AWS Entity Resolution 数据匹配服务 支持
<a href="#">ABAC (策略中的标签)</a>	部分
<a href="#">临时凭证</a>	是
<a href="#">转发访问会话 (FAS)</a>	是
<a href="#">服务角色</a>	是
<a href="#">服务关联角色</a>	否

要全面了解 AWS Entity Resolution 数据匹配服务 以及其他 AWS 服务如何与大多数 IAM 功能配合使用，请参阅 IAM 用户指南中的与 IAM [配合使用的AWS 服务](#)。

## 基于身份的策略 AWS Entity Resolution 数据匹配服务

支持基于身份的策略：是

基于身份的策略是可附加到身份（如 IAM 用户、用户组或角色）的 JSON 权限策略文档。这些策略控制用户和角色可在何种条件下对哪些资源执行哪些操作。要了解如何创建基于身份的策略，请参阅《IAM 用户指南》中的[使用客户管理型策略定义自定义 IAM 权限](#)。

通过使用 IAM 基于身份的策略，您可以指定允许或拒绝的操作和资源以及允许或拒绝操作的条件。要了解可在 JSON 策略中使用的所有元素，请参阅《IAM 用户指南》中的[IAM JSON 策略元素引用](#)。

### 基于身份的策略示例 AWS Entity Resolution 数据匹配服务

要查看 AWS Entity Resolution 数据匹配服务 基于身份的策略的示例，请参阅。[基于身份的策略示例 AWS Entity Resolution 数据匹配服务](#)

## 内部基于资源的政策 AWS Entity Resolution 数据匹配服务

支持基于资源的策略：是

基于资源的策略是附加到资源的 JSON 策略文档。基于资源的策略的示例包括 IAM 角色信任策略和 Amazon S3 存储桶策略。在支持基于资源的策略的服务中，服务管理员可以使用它们来控制对特定资源的访问。对于在其中附加策略的资源，策略定义指定主体可以对该资源执行哪些操作以及在什么条件下执行。您必须在基于资源的策略中[指定主体](#)。委托人可以包括账户、用户、角色、联合用户或 AWS 服务。

要启用跨账户访问，您可以将整个账户或其他账户中的 IAM 实体指定为基于资源的策略中的主体。有关更多信息，请参阅《IAM 用户指南》中的 [IAM 中的跨账户资源访问](#)。

## 的政策行动 AWS Entity Resolution 数据匹配服务

支持策略操作：是

管理员可以使用 AWS JSON 策略来指定谁有权访问什么。也就是说，哪个主体可以对什么资源执行操作，以及在什么条件下执行。

JSON 策略的 Action 元素描述可用于在策略中允许或拒绝访问的操作。在策略中包含操作以授予执行关联操作的权限。

要查看 AWS Entity Resolution 数据匹配服务 操作列表，请参阅《服务授权参考》AWS Entity Resolution 数据匹配服务中 [定义的操作](#)。

正在执行的策略操作在操作前 AWS Entity Resolution 数据匹配服务 使用以下前缀：

```
entityresolution
```

要在单个语句中指定多项操作，请使用逗号将它们隔开。

```
"Action": [  
  "entityresolution:action1",  
  "entityresolution:action2"  
]
```

要查看 AWS Entity Resolution 数据匹配服务 基于身份的策略的示例，请参阅 [基于身份的策略示例](#) [AWS Entity Resolution 数据匹配服务](#)

## 的政策资源 AWS Entity Resolution 数据匹配服务

支持策略资源：是

管理员可以使用 AWS JSON 策略来指定谁有权访问什么。也就是说，哪个主体可以对什么资源执行操作，以及在什么条件下执行。

Resource JSON 策略元素指定要向其应用操作的一个或多个对象。作为最佳实践，请使用其 [Amazon 资源名称 \( ARN \)](#) 指定资源。对于不支持资源级权限的操作，请使用通配符 (\*) 指示语句应用于所有资源。

```
"Resource": "*"
```

要查看 AWS Entity Resolution 数据匹配服务 资源类型及其列表 ARNs，请参阅《服务授权参考》AWS Entity Resolution 数据匹配服务中的“[由定义的资源](#)”。要了解您可以在哪些操作中指定每个资源的 ARN，请参阅 [AWS Entity Resolution 数据匹配服务定义的操作](#)。

要查看 AWS Entity Resolution 数据匹配服务 基于身份的策略的示例，请参阅。[基于身份的策略示例](#)  
[AWS Entity Resolution 数据匹配服务](#)

## 的策略条件密钥 AWS Entity Resolution 数据匹配服务

支持特定于服务的策略条件键：是

管理员可以使用 AWS JSON 策略来指定谁有权访问什么。也就是说，哪个主体可以对什么资源执行操作，以及在什么条件下执行。

Condition 元素根据定义的条件指定语句何时执行。您可以创建使用[条件运算符](#)（例如，等于或小于）的条件表达式，以使策略中的条件与请求中的值相匹配。要查看所有 AWS 全局条件键，请参阅 IAM 用户指南中的[AWS 全局条件上下文密钥](#)。

要查看 AWS Entity Resolution 数据匹配服务 条件键列表，请参阅《服务授权参考》AWS Entity Resolution 数据匹配服务中的[条件密钥](#)。要了解您可以使用条件键的操作和资源，请参阅[操作定义者](#)  
[AWS Entity Resolution 数据匹配服务](#)。

要查看 AWS Entity Resolution 数据匹配服务 基于身份的策略的示例，请参阅。[基于身份的策略示例](#)  
[AWS Entity Resolution 数据匹配服务](#)

## ACLs in AWS Entity Resolution 数据匹配服务

支持 ACLs：否

访问控制列表 (ACLs) 控制哪些委托人（账户成员、用户或角色）有权访问资源。ACLs 与基于资源的策略类似，尽管它们不使用 JSON 策略文档格式。

## ABAC with AWS Entity Resolution 数据匹配服务

支持 ABAC（策略中的标签）：部分支持

基于属性的访问权限控制 (ABAC) 是一种授权策略，该策略基于称为标签的属性来定义权限。您可以将标签附加到 IAM 实体和 AWS 资源，然后设计 ABAC 策略以允许在委托人的标签与资源上的标签匹配时进行操作。

要基于标签控制访问，您需要使用 `aws:ResourceTag/key-name`、`aws:RequestTag/key-name` 或 `aws:TagKeys` 条件键在策略的 [条件元素](#) 中提供标签信息。

如果某个服务对于每种资源类型都支持所有这三个条件键，则对于该服务，该值为是。如果某个服务仅对于部分资源类型支持所有这三个条件键，则该值为部分。

有关 ABAC 的更多信息，请参阅《IAM 用户指南》中的 [使用 ABAC 授权定义权限](#)。要查看设置 ABAC 步骤的教程，请参阅《IAM 用户指南》中的 [使用基于属性的访问权限控制 \(ABAC\)](#)。

## 将临时凭证与配合使用 AWS Entity Resolution 数据匹配服务

支持临时凭证：是

临时证书提供对 AWS 资源的短期访问权限，并且是在您使用联合身份或切换角色时自动创建的。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 用户指南》中的 [IAM 中的临时安全凭证](#) 和 [使用 IAM 的 AWS 服务](#)

## 转发访问会话 AWS Entity Resolution 数据匹配服务

支持转发访问会话 (FAS)：是

转发访问会话 (FAS) 使用调用主体的权限 AWS 服务，再加上 AWS 服务 向下游服务发出请求的请求。有关发出 FAS 请求时的策略详情，请参阅 [转发访问会话](#)。

## 的服务角色 AWS Entity Resolution 数据匹配服务

支持服务角色：是

服务角色是由一项服务担任、代表您执行操作的 [IAM 角色](#)。IAM 管理员可以在 IAM 中创建、修改和删除服务角色。有关更多信息，请参阅《IAM 用户指南》中的 [创建向 AWS 服务委派权限的角色](#)。

### Warning

更改服务角色的权限可能会中断 AWS Entity Resolution 数据匹配服务 功能。只有在 AWS Entity Resolution 数据匹配服务 提供操作指导时才编辑服务角色。

## 的服务相关角色 AWS Entity Resolution 数据匹配服务

支持服务相关角色：否

服务相关角色是一种与服务相关联的 AWS 服务角色。服务可以代入代表您执行操作的角色。服务相关角色出现在您的 AWS 账户中，并且归服务所有。IAM 管理员可以查看但不能编辑服务关联角色的权限。

有关创建或管理服务相关角色的详细信息，请参阅[能够与 IAM 搭配使用的 AWS 服务](#)。在表中查找服务相关角色列中包含 Yes 的表。选择是链接以查看该服务的服务相关角色文档。

## 基于身份的策略示例 AWS Entity Resolution 数据匹配服务

默认情况下，用户和角色没有创建或修改 AWS Entity Resolution 数据匹配服务资源的权限。要授予用户对所需资源执行操作的权限，IAM 管理员可以创建 IAM 策略。

要了解如何使用这些示例 JSON 策略文档创建基于 IAM 身份的策略，请参阅《IAM 用户指南》中的[创建 IAM 策略 \(控制台\)](#)。

有关由 AWS Entity Resolution 数据匹配服务定义的操作和资源类型（包括每种资源类型的格式）的详细信息，请参阅《服务授权参考》AWS Entity Resolution 数据匹配服务中的[操作、资源和条件密钥](#)。  
ARNs

### 主题

- [策略最佳实践](#)
- [使用控制 AWS Entity Resolution 数据匹配服务台](#)
- [允许用户查看他们自己的权限](#)

## 策略最佳实践

基于身份的策略决定了某人是否可以在您的账户中创建、访问或删除 AWS Entity Resolution 数据匹配服务资源。这些操作可能会使 AWS 账户产生成本。创建或编辑基于身份的策略时，请遵循以下指南和建议：

- 开始使用 AWS 托管策略并转向最低权限权限 — 要开始向用户和工作负载授予权限，请使用为许多常见用例授予权限的 AWS 托管策略。它们在你的版本中可用 AWS 账户。我们建议您通过定义针对您的用例的 AWS 客户托管策略来进一步减少权限。有关更多信息，请参阅《IAM 用户指南》中的[AWS 托管策略](#)或[工作职能的 AWS 托管策略](#)。
- 应用最低权限：在使用 IAM 策略设置权限时，请仅授予执行任务所需的权限。为此，您可以定义在特定条件下可以对特定资源执行的操作，也称为最低权限许可。有关使用 IAM 应用权限的更多信息，请参阅《IAM 用户指南》中的[IAM 中的策略和权限](#)。

- 使用 IAM 策略中的条件进一步限制访问权限：您可以向策略添加条件来限制对操作和资源的访问。例如，您可以编写策略条件来指定必须使用 SSL 发送所有请求。如果服务操作是通过特定 AWS 服务的（例如）使用的，则也可以使用条件来授予对服务操作的访问权限 CloudFormation。有关更多信息，请参阅《IAM 用户指南》中的 [IAM JSON 策略元素：条件](#)。
- 使用 IAM Access Analyzer 验证您的 IAM 策略，以确保权限的安全性和功能性：IAM Access Analyzer 会验证新策略和现有策略，以确保策略符合 IAM 策略语言（JSON）和 IAM 最佳实践。IAM Access Analyzer 提供 100 多项策略检查和可操作的建议，以帮助您制定安全且功能性强的策略。有关更多信息，请参阅《IAM 用户指南》中的 [使用 IAM Access Analyzer 验证策略](#)。
- 需要多重身份验证 (MFA)-如果 AWS 账户您的场景需要 IAM 用户或根用户，请启用 MFA 以提高安全性。若要在调用 API 操作时需要 MFA，请将 MFA 条件添加到您的策略中。有关更多信息，请参阅《IAM 用户指南》中的 [使用 MFA 保护 API 访问](#)。

有关 IAM 中的最佳实践的更多信息，请参阅《IAM 用户指南》中的 [IAM 中的安全最佳实践](#)。

## 使用控制 AWS Entity Resolution 数据匹配服务 台

要访问 AWS Entity Resolution 数据匹配服务 控制台，您必须拥有一组最低权限。这些权限必须允许您列出和查看有关您的 AWS Entity Resolution 数据匹配服务 资源的详细信息 AWS 账户。如果创建比必需的最低权限更为严格的基于身份的策略，对于附加了该策略的实体（用户或角色），控制台将无法按预期正常运行。

对于仅调用 AWS CLI 或 AWS API 的用户，您无需为其设置最低控制台权限。相反，只允许访问与其尝试执行的 API 操作相匹配的操作。

为确保用户和角色仍然可以使用 AWS Entity Resolution 数据匹配服务 控制台，还需要将 AWS Entity Resolution 数据匹配服务 *ConsoleAccess* 或 *ReadOnly* AWS 托管策略附加到实体。有关更多信息，请参阅《IAM 用户指南》中的 [为用户添加权限](#)。

## 允许用户查看他们自己的权限

该示例说明了您如何创建策略，以允许 IAM 用户查看附加到其用户身份的内联和托管式策略。此策略包括在控制台上或使用 AWS CLI 或 AWS API 以编程方式完成此操作的权限。

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "ViewOwnUserInfo",
      "Effect": "Allow",
```

```
    "Action": [
      "iam:GetUserPolicy",
      "iam:ListGroupsForUser",
      "iam:ListAttachedUserPolicies",
      "iam:ListUserPolicies",
      "iam:GetUser"
    ],
    "Resource": ["arn:aws:iam::*:user/${aws:username}"]
  },
  {
    "Sid": "NavigateInConsole",
    "Effect": "Allow",
    "Action": [
      "iam:GetGroupPolicy",
      "iam:GetPolicyVersion",
      "iam:GetPolicy",
      "iam:ListAttachedGroupPolicies",
      "iam:ListGroupPolicies",
      "iam:ListPolicyVersions",
      "iam:ListPolicies",
      "iam:ListUsers"
    ],
    "Resource": "*"
  }
]
```

## AWS 的托管策略 AWS Entity Resolution 数据匹配服务

AWS 托管策略是由创建和管理的独立策略 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、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，因为它们可供所有 AWS 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 AWS 托管策略中定义的权限。如果 AWS 更新 AWS 托管策略中定义的权限，则更新会影响该策略所关联的所有委托人身份（用户、组和角色）。AWS 最有可能在启动新的 API 或现有服务可以使用新 AWS 服务的 API 操作时更新 AWS 托管策略。

有关更多信息，请参阅《IAM 用户指南》中的[AWS 托管式策略](#)。

## AWS 托管策略：AWSEntityResolutionConsoleFullAccess

您可以将 AWSEntityResolutionConsoleFullAccess 策略附加到 IAM 身份。

此策略授予对 AWS Entity Resolution 数据匹配服务 端点和资源的完全访问权限。

该策略还允许对相关内容 ( AWS 服务 例如 S3、Tagging AWS Glue AWS KMS EventBridge、Amazon 等 ) 进行某些读取权限，AWS Data Exchange 以便控制台可以显示选择并使用所选选项来执行实体解析操作。此外，此策略还授予对 Connect 客户档案 API 的访问权限，以实现自动匹配结果处理的集成。某些资源的范围已缩小到包含服务名称entityresolution。

由于 AWS Entity Resolution 数据匹配服务 依赖传递的角色对相关 AWS 资源执行操作，因此该策略还授予选择和传递所需角色的权限。

### 权限详细信息

该策略包含以下权限。

- EntityResolutionAccess— 允许委托人完全访问 AWS Entity Resolution 数据匹配服务 端点和资源。
- GlueSourcesConsoleDisplay— 授予将 AWS Glue 表列为数据源选项和导入数据源的表架构的访问权限，以获得用户体验。
- S3BucketsConsoleDisplay— 授予将所有 S3 存储桶列为数据源选项的权限。
- S3SourcesConsoleDisplay— 授予将 S3 存储桶显示为数据源选项的权限。
- TaggingConsoleDisplay— 授予读取标记键和值的权限。
- KMSConsoleDisplay— 授予描述密钥和列出别名的访问权限 AWS Key Management Service ，以解密和加密数据源。
- ListRolesToPickForPassing— 授予列出所有角色的权限，以使用户可以选择要传递的角色。
- PassRoleToEntityResolutionService— 授予将缩小范围的角色传递给 AWS Entity Resolution 数据匹配服务 服务的访问权限。
- ManageEventBridgeRules— 授予创建、更新和删除用于获取 S3 通知的 Amazon EventBridge 规则的权限。
- ADXReadAccess— 授予访问权限 AWS Data Exchange 以验证客户是否有权利或订阅。
- CustomerProfilesIntegrationAccess— 授予对 Amazon Connect 客户档案和 Connect 客户档案 API 的访问权限，以实现与 Connect 客户档案之间的 AWS Entity Resolution 数据匹配服务 集成，从而自动处理匹配结果。

要查看此策略的权限，请参阅《AWS 托管策略参考》中的 [AWSEntityResolutionConsoleFullAccess](#)。

## AWS 托管策略：AWSEntityResolutionConsoleReadOnlyAccess

您可以将 AWSEntityResolutionConsoleReadOnlyAccess 附加到 IAM 实体。

此策略授予对 AWS Entity Resolution 数据匹配服务 端点和资源的只读访问权限。

### 权限详细信息

该策略包含以下权限。

- EntityResolutionRead— 允许委托人对 AWS Entity Resolution 数据匹配服务 端点和资源进行只读访问。

要查看此策略的权限，请参阅《AWS 托管策略参考》中的 [AWSEntityResolutionConsoleReadOnlyAccess](#)。

## AWS Entity Resolution 数据匹配服务 更新到 AWS 托管策略

查看 AWS Entity Resolution 数据匹配服务 自该服务开始跟踪这些更改以来 AWS 托管策略更新的详细信息。要获得有关此页面变更的自动提醒，请订阅“AWS Entity Resolution 数据匹配服务 文档历史记录”页面上的 RSS feed。

更改	描述	日期
AWSEntityResolutionConsoleFullAccess 对现有策略的更新	- 添加CustomerProfilesIntegrationAccess 后可与 Amazon Connect 客户档案集成，以便自动处理匹配结果。	2025 年 12 月 15 日
AWSEntityResolutionConsoleFullAccess 对现有策略的更新	- 添加ADXReadAccess ManageEventBridgeRules 并启用匹配工作流程中的提供者服务选项。	2023 年 10 月 16 日

更改	描述	日期
AWS Entity Resolution 数据匹配服务 已开始跟踪更改	AWS Entity Resolution 数据匹配服务 开始跟踪其 AWS 托管策略的更改。	2023 年 8 月 18 日

## 对 AWS Entity Resolution 数据匹配服务 身份和访问进行故障排除

使用以下信息来帮助您诊断和修复在使用 AWS Entity Resolution 数据匹配服务 和 IAM 时可能遇到的常见问题。

### 主题

- [我无权在以下位置执行操作 AWS Entity Resolution 数据匹配服务](#)
- [我无权执行 iam : PassRole](#)
- [我想允许我以外的人 AWS 账户 访问我的 AWS Entity Resolution 数据匹配服务 资源](#)

### 我无权在以下位置执行操作 AWS Entity Resolution 数据匹配服务

如果 AWS 管理控制台 告诉您您无权执行某项操作，则必须联系管理员寻求帮助。管理员是指提供用户名和密码的人员。

当 mateojackson IAM 用户尝试使用控制台查看有关虚构 *my-example-widget* 资源的详细信息，但不拥有虚构 `entityresolution:GetWidget` 权限时，会发生以下示例错误。

```
User: arn:aws:iam::123456789012:user/mateojackson is not authorized to perform:
entityresolution:GetWidget on resource: my-example-widget
```

在这种情况下，Mateo 请求他的管理员更新其策略，以允许他使用 `entityresolution:GetWidget` 操作访问 *my-example-widget* 资源。

### 我无权执行 iam : PassRole

如果您收到一个错误，表明您无权执行 `iam:PassRole` 操作，则必须更新策略以允许您将角色传递给。AWS Entity Resolution 数据匹配服务

有些 AWS 服务 允许您将现有角色传递给该服务，而不是创建新的服务角色或服务相关角色。为此，您必须具有将角色传递到服务的权限。

当名为 marymajor 的 IAM 用户尝试使用控制台在 AWS Entity Resolution 数据匹配服务中执行操作时，会发生以下示例错误。但是，服务必须具有服务角色所授予的权限才可执行此操作。Mary 不具有将角色传递到服务的权限。

```
User: arn:aws:iam::123456789012:user/marymajor is not authorized to perform:
iam:PassRole
```

在这种情况下，必须更新 Mary 的策略以允许她执行 iam:PassRole 操作。

如果您需要帮助，请联系您的 AWS 管理员。您的管理员是提供登录凭证的人。

## 我想允许我以外的人 AWS 账户 访问我的 AWS Entity Resolution 数据匹配服务 资源

您可以创建一个角色，以便其他账户中的用户或您组织外的人员可以使用该角色来访问您的资源。您可以指定谁值得信赖，可以代入角色。对于支持基于资源的策略或访问控制列表 (ACLs) 的服务，您可以使用这些策略向人们授予访问您的资源的权限。

要了解更多信息，请参阅以下内容：

- 要了解是否 AWS Entity Resolution 数据匹配服务 支持这些功能，请参阅[如何 AWS Entity Resolution 数据匹配服务 与 IAM 配合使用](#)。
- 要了解如何提供对您拥有的资源的访问权限 AWS 账户，请参阅[IAM 用户指南中的向您拥有 AWS 账户 的另一个 IAM 用户提供访问权限](#)。
- 要了解如何向第三方提供对您的资源的访问[权限 AWS 账户](#)，请参阅[IAM 用户指南中的向第三方提供访问权限。AWS 账户](#)。
- 要了解如何通过身份联合验证提供访问权限，请参阅《IAM 用户指南》中的[为经过外部身份验证的用户（身份联合验证）提供访问权限](#)。
- 要了解使用角色和基于资源的策略进行跨账户访问之间的差别，请参阅《IAM 用户指南》中的[IAM 中的跨账户资源访问](#)。

## 的合规性验证 AWS Entity Resolution 数据匹配服务

要了解是否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](#)”，然后选择您感兴趣的合规计划。AWS 服务 有关一般信息，请参阅[AWS 合规计划AWS](#)。

您可以使用下载第三方审计报告 AWS Artifact。有关更多信息，请参阅中的“[下载报告](#)”中的“[AWS Artifact](#)”。

您在使用 AWS 服务 时的合规责任取决于您的数据的敏感性、贵公司的合规目标以及适用的法律和法规。有关您在使用时的合规责任的更多信息 AWS 服务，请参阅[AWS 安全文档](#)。

## AWS Entity Resolution 数据匹配服务 合规性最佳实践

本节提供使用时的最佳实践和合规性建议 AWS Entity Resolution 数据匹配服务。

### 支付卡行业数据安全标准 ( PCI DSS )

AWS Entity Resolution 数据匹配服务 支持商家或服务提供商处理、存储和传输信用卡数据，并且已被验证符合支付卡行业 (PCI) 数据安全标准 (DSS)。有关 PCI DSS 的更多信息，包括如何申请 PCI Compliance Package 的副本，请参阅 AWS [PCI DSS](#) 第 1 级。

### 系统和组织控制 ( SOC )

AWS Entity Resolution 数据匹配服务 符合系统和组织控制 (SOC) 措施，包括 SOC 1、SOC 2 和 SOC 3。SOC 报告是独立的第三方检查报告，用于展示如何 AWS 实现关键合规控制和目标。这些审计可确保采取适当的安全措施和程序，防止出现风险，影响客户和公司数据的安全性、机密性和可用性。这些第三方审计的结果可在 [AWS SOC 合规网站上](#) 查看，您可以在该网站上查看已发布的报告，以获取有关支持 AWS 运营和合规的控制措施的更多信息。

## 韧性在 AWS Entity Resolution 数据匹配服务

AWS 全球基础设施是围绕 AWS 区域 可用区构建的。AWS 区域 提供多个物理隔离和隔离的可用区，这些可用区通过低延迟、高吞吐量和高度冗余的网络连接。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结构相比，可用区具有更高的可用性、容错性和可扩展性。

有关 AWS 区域 和可用区的更多信息，请参阅[AWS 全球基础设施](#)。

除了 AWS 全球基础架构外，还 AWS Entity Resolution 数据匹配服务 提供多项功能来帮助支持您的数据弹性和备份需求。

## 监控 AWS Entity Resolution 数据匹配服务

监控是维护和其他 AWS 解决方案的可靠性、可用性和性能的重要组成部分。AWS Entity Resolution 数据匹配服务 AWS 提供以下监控工具 AWS Entity Resolution 数据匹配服务，供您监视、报告问题并在适当时自动采取措施：

- AWS CloudTrail捕获由您或代表您发起的 API 调用和相关事件，AWS 账户 并将日志文件传输到您指定的 Amazon S3 存储桶。您可以识别哪些用户和账户拨打了电话 AWS、来源 IP 讨论了从哪个拨打的呼叫以及呼叫发生的时间。有关更多信息，请参阅 [AWS CloudTrail 《用户指南》](#)。
- Amazon CloudWatch Logs 允许您检查、存储和访问来自 Amazon EC2 实例和其他来源的日志。CloudTrail CloudWatch 日志可以检查日志文件中的信息，并告诉您何时达到某些阈值。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 日志用户指南](#)。

### 主题

- [使用记录 AWS Entity Resolution 数据匹配服务 API 调用 AWS CloudTrail](#)
- [使用 Amazon CloudWatch 日志监控和记录工作流程](#)

## 使用记录 AWS Entity Resolution 数据匹配服务 API 调用 AWS CloudTrail

AWS Entity Resolution 数据匹配服务 与 AWS CloudTrail 一项服务集成，该服务提供用户、角色或 AWS 服务在中执行的操作的记录 AWS Entity Resolution 数据匹配服务。CloudTrail 将所有 API 调用捕获 AWS Entity Resolution 数据匹配服务 为事件。捕获的调用包括来自 AWS Entity Resolution 数据匹配服务 控制台的调用和对 AWS Entity Resolution 数据匹配服务 API 操作的代码调用。如果您创建了跟踪，则可以允许将 CloudTrail 事件持续传输到 Amazon S3 存储桶，包括的事件 AWS Entity Resolution 数据匹配服务。如果您未配置跟踪，您仍然可以在 CloudTrail 控制台的“事件历史记录”中查看最新的事件。使用收集的信息 CloudTrail，您可以确定向哪个请求发出 AWS Entity Resolution 数据匹配服务、发出请求的 IP 地址、谁发出了请求、何时发出请求以及其他详细信息。

要了解更多信息 CloudTrail，请参阅[AWS CloudTrail 用户指南](#)。

## AWS Entity Resolution 数据匹配服务 信息在 CloudTrail

CloudTrail 在您创建账户 AWS 账户 时已在您的账户上启用。当活动发生在中时 AWS Entity Resolution 数据匹配服务，该活动会与其他 AWS 服务 CloudTrail 事件一起记录在事件历史记录中。

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[使用事件历史记录查看 CloudTrail 事件](#)。

要持续记录您的 AWS 账户事件（包括的事件）AWS Entity Resolution 数据匹配服务，请创建跟踪。跟踪允许 CloudTrail 将日志文件传输到 Amazon S3 存储桶。默认情况下，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。跟踪记录 AWS 分区中所有区域的事件，并将日志文件传送到您指定的 Amazon S3 存储桶。此外，您可以配置其他 AWS 服务，以进一步分析和处理 CloudTrail 日志中收集的事件数据。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](#)
- [配置 Amazon SNS 通知 CloudTrail](#)
- [接收来自多个地区的 CloudTrail 日志文件和接收来自多个账户的 CloudTrail 日志文件](#)

所有 AWS Entity Resolution 数据匹配服务 操作均由《API 参考》记录 CloudTrail 并记录在《[AWS Entity Resolution 数据匹配服务 API 参考](#)》中。

每个事件或日志条目都包含有关生成请求的人员信息。身份信息有助于您确定以下内容：

- 请求是使用根证书还是 AWS Identity and Access Management (IAM) 用户凭证发出。
- 请求是使用角色还是联合用户的临时安全凭证发出的。
- 请求是否由其他 AWS 服务发出。

有关更多信息，请参阅 [CloudTrail userIdentity 元素](#)。

## 了解 AWS Entity Resolution 数据匹配服务 日志文件条目

跟踪是一种配置，允许将事件作为日志文件传输到您指定的 Amazon S3 存储桶。CloudTrail 日志文件包含一个或多个日志条目。事件代表来自任何来源的单个请求，包括有关请求的操作、操作的日期和时间、请求参数等的信息。CloudTrail 日志文件不是公共 API 调用的有序堆栈跟踪，因此它们不会按任何特定的顺序出现。

## 使用 Amazon CloudWatch 日志监控和记录工作流程

AWS Entity Resolution 数据匹配服务 提供全面的日志记录功能，可帮助您检查和分析匹配和 ID 映射工作流程。通过与 Amazon CloudWatch Logs 集成，您可以捕获有关工作流程执行的详细信息，

包括事件类型、时间戳、处理统计数据 and 错误计数。您可以选择将这些日志传送到日 CloudWatch 志、Amazon S3 或 Amazon Data Firehose 目的地。通过分析这些日志，您可以评估服务性能，解决问题，深入了解您的客户群，并更好地了解您的 AWS Entity Resolution 数据匹配服务 使用情况和账单。虽然默认情况下日志记录处于禁用状态，但您可以通过控制台或 API 为新的和现有的工作流程启用日志记录。

启用 AWS Entity Resolution 数据匹配服务 工作流程日志记录功能时，将收取标准的 Amazon CloudWatch 售货费，包括与日志提取、存储和分析相关的费用；有关详细的定价信息，请访问[CloudWatch 定价](#)页面。。

## 主题

- [设置日志传输](#)
- [禁用日志记录 \(控制台\)](#)
- [读取日志](#)

## 设置日志传输

本节将说明使用 AWS Entity Resolution 数据匹配服务 日志记录所需的必要权限，以及如何使用控制台和启用日志传输 APIs。

## 主题

- [Permissions](#)
- [为新工作流程启用日志记录 \(控制台\)](#)
- [为新工作流程启用日志记录 \(API\)](#)
- [为现有工作流程启用日志记录 \(控制台\)](#)

## Permissions

AWS Entity Resolution 数据匹配服务 使用 CloudWatch 已发送的日志来传送工作流程日志。要传送工作流程日志，您需要拥有对所指定日志记录目标的权限。

要查看每个日志目标所需的权限，请在《Amazon L CloudWatch ogs 用户指南》中从以下 AWS 服务中进行选择。

- [Amazon CloudWatch 日志](#)
- [Amazon Simple Storage Service \( Amazon S3 \)](#)

- [Amazon Data Firehose](#)

要在中创建、查看或更改日志配置 AWS Entity Resolution 数据匹配服务，您必须具有所需的权限。您的 IAM 角色必须包含以下最低权限才能在 AWS Entity Resolution 数据匹配服务 控制台中管理工作流程日志。

JSON

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "AllowLogDeliveryActionsConsoleCWL",
      "Effect": "Allow",
      "Action": [
        "logs:DescribeLogGroups"
      ],
      "Resource": [
        "arn:aws:logs:us-east-1:111122223333:log-group:*"
      ]
    },
    {
      "Sid": "AllowLogDeliveryActionsConsoleS3",
      "Effect": "Allow",
      "Action": [
        "s3:ListAllMyBuckets",
        "s3:ListBucket",
        "s3:GetBucketLocation"
      ],
      "Resource": [
        "arn:aws:s3:::*"
      ]
    },
    {
      "Sid": "AllowLogDeliveryActionsConsoleFH",
      "Effect": "Allow",
      "Action": [
        "firehose:ListDeliveryStreams",
        "firehose:DescribeDeliveryStream"
      ],
      "Resource": [
        "*"
      ]
    }
  ]
}
```

```
    ]
  }
]
}
```

有关管理工作流程日志的权限的更多信息，请参阅 Amazon [Log CloudWatch s 用户指南中的启用 AWS 服务](#) 日志记录。

## 为新工作流程启用日志记录 ( 控制台 )

设置日志记录目标的权限后，您可以使用控制台为新工作流程启用日志记录。AWS Entity Resolution 数据匹配服务

### 为新工作流程启用日志记录 ( 控制台 )

1. 在家中打开 <https://console.aws.amazon.com/entityresolution/> 主 AWS Entity Resolution 数据匹配服务机。
2. 在“工作流程”下，选择“匹配工作流”或“ID 映射工作流”。
3. 按照步骤创建以下工作流程之一：
  - [基于规则的匹配工作流程](#)
  - [基于机器学习的匹配工作流程](#)
  - [基于提供商服务的匹配工作流程](#)
  - [一个账户的 ID 映射工作流程](#)
  - [跨两个账户的 ID 映射工作流程](#)
4. 对于步骤 1 指定匹配的工作流程详细信息，对于日志传送- EntityResolution 工作流程日志，选择添加。
  - 选择以下日志记录目标之一。
    - 到 Amazon CloudWatch 日志
    - 前往亚马逊 S3
    - 前往亚马逊 Data Firehose

#### Tip

如果您选择 Amazon S3 或 Firehose，则可以将日志传送到跨账户或活期账户。

要启用跨账户交付，两者都 AWS 账户 必须具有所需的权限。有关更多信息，请参阅 Amazon L CloudWatch logs 用户指南中的[跨账户交付示例](#)。

5. 对于目标日志组，将自动创建以“/aws/vendedlogs/”为前缀的日志组。如果您使用的是其他日志组，请在设置日志传输之前对其进行操作。有关更多信息，请参阅 Amazon 日志用户指南中的使用日志组和 CloudWatch 日志[流](#)。
6. 如需更多设置-可选，请选择以下选项：
  - a. 在“字段选择”中，选择要包含在每条日志记录中的日志字段。
  - b. ( CloudWatch 日志 ) 对于输出格式，选择日志的输出格式。
  - c. 对于字段分隔符，选择如何分隔每个日志字段。
  - d. (Amazon S3) 对于后缀，请指定用于对数据进行分区的路径。
  - e. (Amazon S3) 对于兼容 Hive，如果要使用兼容 Hive 的 S3 路径，请选择“启用”。
7. 要创建另一个日志目标，请选择添加并重复步骤 4-6。
8. 完成其余步骤以设置和运行工作流程。
9. 工作流程作业完成后，在您指定的日志传输目标中检查工作流日志。

## 为新工作流程启用日志记录 (API)

设置登录目标权限后，您可以 AWS Entity Resolution 数据匹配服务 使用 Amazon Logs 为新工作流程启用 CloudWatch 日志功能 APIs。

### 为新工作流程启用日志记录 (API)

1. 在 AWS Entity Resolution 数据匹配服务 控制台中创建工作流程后，获取该工作流程的 Amazon 资源名称 (ARN)。

您可以从 AWS Entity Resolution 数据匹配服务 控制台的工作流程页面找到 ARN，也可以调用 GetMatchingWorkflow 或 GetIdMappingWorkflow API 操作。

工作流程 ARN 遵循以下格式：

```
arn:(aws|aws-us-gov|aws-cn):entityresolution:[a-z]{2}-[a-z]{1,10}-[0-9]:[0-9]{12}:(matchingworkflow/[a-zA-Z_0-9-]{1,255})
```

ID 映射 ARN 遵循以下格式：

```
arn:(aws|aws-us-gov|aws-cn):entityresolution:[a-z]{2}-[a-z]{1,10}-[0-9]:[0-9]{12}:(idmappingworkflow/[a-zA-Z_0-9-]{1,255})
```

有关更多信息，请参阅《AWS Entity Resolution 数据匹配服务 API 参考》中的 [GetMatchingWorkflow](#) 或 [GetIdMappingWorkflow](#)。

2. 使用 CloudWatch Logs PutDeliverySource API 操作作为工作流程日志创建传送源。

有关更多信息，请参阅 Amazon CloudWatch 日志 API 参考 [PutDeliverySource](#) 中的。

- a. 通过 resourceArn.
- b. 对于 logType，收集的日志类型为 WORKFLOW\_LOGS：

### Example

#### PutDeliverySourceAPI 操作示例

```
{
  "logType": "WORKFLOW_LOGS",
  "name": "my-delivery-source",
  "resourceArn": "arn:aws:entityresolution:region:accoungId:matchingworkflow/
XXXWorkflow"
}
```

3. 使用 PutDeliveryDestination API 操作配置日志的存储位置。

您可以选择 CloudWatch Logs、Amazon S3 或 Firehose 作为目标。您必须为日志的存储位置指定其中一个目标选项的 ARN。

有关更多信息，请参阅 Amazon CloudWatch 日志 API 参考 [PutDeliveryDestination](#) 中的。

### Example

#### PutDeliveryDestinationAPI 操作示例

```
{
  "delivery-destination-configuration": {
    "destinationResourceArn": "arn:aws:logs:region:accountId:log-group:my-log-
group"
  },
  "name": "my-delivery-destination",
}
```

```
"outputFormat": "json",
}
}
```

#### Note

如果您要跨账户传送日志，则必须使用 `PutDeliveryDestinationPolicyAPI` 为目标账户分配 AWS Identity and Access Management (IAM) 策略。IAM 策略支持将日志从一个账户传输到另一个账户。

4. 使用 `CreateDelivery` API 操作将传送源链接到您在前面步骤中创建的目标。此 API 操作会将传输源与最终目标关联起来。

有关更多信息，请参阅 Amazon CloudWatch 日志 API 参考 [PutDeliveryDestination](#) 中的。

#### Example

#### CreateDeliveryAPI 操作示例

```
{
  "delivery-destination-arn": "arn:aws:logs:region:accountId:log-group:my-log-
group",
  "delivery-source-name": "my-delivery-source",
  "tags": {
    "string" : "string"
  }
}
```

5. 运行 工作流。
6. 工作流作业完成后，在您指定的日志传输目标中检查工作流日志。

## 为现有工作流程启用日志记录（控制台）

设置日志目标权限后，您可以使用控制台上的“日志交付”选项卡为现有工作流程启用日志记录。AWS Entity Resolution 数据匹配服务

使用“日志交付”选项卡（控制台）为现有工作流程启用日志记录

1. 在家中打开 <https://console.aws.amazon.com/entityresolution/> 主 AWS Entity Resolution 数据匹配服务机。

- 在“工作流程”下，选择“匹配工作流程”或“ID 映射工作流程”，然后选择您的现有工作流程。
- 在“日志传送”选项卡的“日志传送”下，选择“添加”，然后选择以下日志记录目标之一。

- 到 Amazon CloudWatch 日志
- 到 Amazon S3
  - 跨账户
  - 在当前账户中
- 前往亚马逊 Data Firehose
  - 跨账户
  - 在当前账户中

 Tip

如果您选择 Amazon S3 或 Firehose，则可以将日志传送到跨账户或活期账户。要启用跨账户交付，两者都 AWS 账户 必须具有所需的权限。有关更多信息，请参阅 Amazon L CloudWatch logs 用户指南中的[跨账户交付示例](#)。

- 在模式中，根据您选择的日志传送类型，执行以下操作。
  - 查看日志类型：WORKFLOW\_LOGS。

无法更改日志类型。
  - ( CloudWatch 日志 ) 对于目标日志组，将自动创建以“/aws/vendedlogs/”为前缀的日志组。如果您使用的是其他日志组，请在设置日志传输之前对其进行操作。有关更多信息，请参阅 Amazon 日志用户指南中的使用日志组和 CloudWatch 日志[流](#)。

( 往来账户中的 Amazon S3 ) 对于目标 S3 存储桶，选择一个存储桶或输入 ARN。  
( 亚马逊 S3 交叉账户 ) 对于配送目的地 ARN，输入配送目的地 ARN。  
( 当前账户中的 Firehose ) 对于目标传送流，请输入在其他账户中创建的传送目标资源的 ARN。  
( Firehose 交叉账户 ) 对于配送目的地 ARN，请输入配送目的地 ARN。
- 如需更多设置-可选，请选择以下选项：

- 在“字段选择”中，选择要包含在每条日志记录中的日志字段。

- b. ( CloudWatch 日志 ) 对于输出格式，选择日志的输出格式。
  - c. 对于字段分隔符，选择如何分隔每个日志字段。
  - d. (Amazon S3) 对于后缀，请指定用于对数据进行分区的路径。
  - e. (Amazon S3) 对于兼容 Hive，如果要使用兼容 Hive 的 S3 路径，请选择“启用”。
6. 选择添加。
  7. 在 workflows 页面上，选择运行。
  8. 工作流作业完成后，在您指定的日志传输目标中检查工作流日志。

## 禁用日志记录 ( 控制台 )

您可以随时在控制台中禁用 AWS Entity Resolution 数据匹配服务 工作流的日志记录。

### 禁用工作流日志 ( 控制台 )

1. 在家中打开<https://console.aws.amazon.com/entityresolution/>主 AWS Entity Resolution 数据匹配服务机。
2. 在“工作流”下，选择“匹配工作流”或“ID 映射工作流”，然后选择您的工作流。
3. 在“日志传送”选项卡的“日志传送”下，选择目的地，然后选择“删除”。
4. 查看您的更改，然后导航到下一步以保存更改。

## 读取日志

阅读 Amazon CloudWatch 日志可帮助您保持高效 AWS Entity Resolution 数据匹配服务 的工作流。日志可以详细了解您的工作流执行情况，包括处理的记录数量和遇到的任何错误等重要指标，从而帮助您确保数据处理顺利进行。此外，日志还可通过时间戳和事件类型实时跟踪工作流进度，使您能够快速识别数据处理管道中的瓶颈或问题。全面的错误跟踪和记录计数信息可以准确显示成功处理了多少记录以及是否还有未处理的记录，从而帮助您保持数据的质量和完整性。

如果您使用 CloudWatch 日志作为目标，则可以使用 CloudWatch Logs Insights 来读取工作流日志。通常 CloudWatch 会收取日志费用。有关更多信息，请参阅 Amazon Logs 用户指南中的使用 CloudWatch 日志见解分析 CloudWatch 日志[数据](#)。

**Note**

工作流程日志可能需要几分钟才能显示在您的目标中。如果您没有看到日志，请等待几分钟并刷新页面。

工作流日志由一系列格式化的日志记录组成，其中每条日志记录代表一个工作流程。日志中字段的顺序可能会变化。

```
{
  "resource_arn": "arn:aws:ses:us-east-1:1234567890:mailmanager-ingress-point/inp-xxxxx",
  "event_type": "JOB_START",
  "event_timestamp": 1728562395042,
  "job_id": "b01eea4678d4423a4b43eeada003f6",
  "workflow_name": "TestWorkflow",
  "workflow_start_time": "2025-03-11 10:19:56",
  "data_processing_progression": "Matching Job Starts ...",
  "total_records_processed": 1500,
  "total_records_unprocessed": 0,
  "incremental_records_processed": 0,
  "error_message": "sample error that caused workflow failure"
}
```

以下列表按顺序描述了日志记录字段：

**resource\_arn**

Amazon 资源名称 (ARN)，用于唯一标识工作流程中使用的 AWS 资源。

**event\_type**

工作流程执行期间发生的事件类型。AWS Entity Resolution 数据匹配服务 目前支持：

JOB\_START

DATA\_PROCESSING\_STEP\_START

DATA\_PROCESSING\_STEP\_END

JOB\_SUCCESS

JOB\_FAILURE

## event\_timestamp

指示工作流程中事件发生时间的 Unix 时间戳。

## job\_id

分配给特定工作流程任务执行的唯一标识符。

## workflow\_name

为正在执行的工作流程指定的名称。

## workflow\_start\_time

开始执行工作流程的日期和时间。

## data\_processing\_progression

对数据处理工作流程当前阶段的描述。示例："Matching Job Starts"、"Loading Step Starts"、"ID\_Mapping Job Ends Successfully"。

## total\_records\_processed

在工作流程中成功处理的记录总数。

## total\_records\_unprocessed

在工作流程执行期间未处理的记录数。

## incremental\_records\_processed

在增量工作流程更新中处理的新记录数。

## error\_message

工作流程失败的根本原因。

# 使用创建 AWS 实体解析资源 AWS CloudFormation

AWS Entity Resolution 与 AWS CloudFormation 一项服务集成，可帮助您对 AWS 资源进行建模和设置，从而减少创建和管理资源和基础设施所花费的时间。您可以创建一个描述所需的所有 AWS 资源（例如 `AWS::EntityResolution::MatchingWorkflow`，`AWS::EntityResolution::SchemaMapping`，`AWS::EntityResolution::IdMappingWorkflow`，`AWS::EntityResolution::IdNamespace` 和 `AWS::EntityResolution::PolicyStatement`）的模板，然后为您 CloudFormation 预置和配置这些资源。

使用时 CloudFormation，您可以重复使用模板来一致且重复地设置 AWS 实体解析资源。只需描述一次您的资源，然后在多个 AWS 账户 区域中一遍又一遍地配置相同的资源。

## AWS 实体解析和 CloudFormation 模板

要为 AWS 实体解析和相关服务预置和配置资源，您必须了解 [CloudFormation 模板](#)。模板是 JSON 或 YAML 格式的文本文件。这些模板描述了您要在 CloudFormation 堆栈中配置的资源。如果你不熟悉 JSON 或 YAML，可以使用 D CloudFormation esigner 来帮助你开始使用 CloudFormation 模板。有关更多信息，请参阅《AWS CloudFormation 用户指南》中的 [什么是 CloudFormation Designer？](#)。

AWS 实体解析支持创建 `AWS::EntityResolution::MatchingWorkflow`，`AWS::EntityResolution::SchemaMapping`，`AWS::EntityResolution::IdMappingWorkflow`，`AWS::EntityResolution::IdNamespace` 和 `AWS::EntityResolution::PolicyStatement` 输入 CloudFormation。有关更多信息，包括和的 JSON 和 YAML 模板示例 `AWS::EntityResolution::PolicyStatement`，请参阅 AWS CloudFormation 用户指南中的 [AWS 实体解析资源类型参考](#)。 `AWS::EntityResolution::MatchingWorkflow`，`AWS::EntityResolution::SchemaMapping`，`AWS::EntityResolution::IdMappingWorkflow`，`AWS::EntityResolution::IdNamespace`

可用模板如下：

- 匹配工作流程

创建一个 `MatchingWorkflow` 对象，用于存储要运行的数据处理作业的配置。

有关更多信息，请参阅以下主题：

《CloudFormation 用户指南》中的 [AWS::EntityResolution::MatchingWorkflow](#)

《AWS Entity Resolution 数据匹配服务 API 参考》中的 [CreateMatchingWorkflow](#)

- 架构映射

创建架构映射，用于定义输入客户记录表的架构。

有关更多信息，请参阅以下主题：

《CloudFormation 用户指南》中的 [AWS::EntityResolution::SchemaMapping](#)

《AWS Entity Resolution 数据匹配服务 API 参考》中的 [CreateSchemaMapping](#)

- 身份映射工作流程

创建一个 IdMappingWorkflow 对象，用于存储要运行的数据处理作业的配置。

有关更多信息，请参阅以下主题：

《CloudFormation 用户指南》中的 [AWS::EntityResolution::IdMappingWorkflow](#)

《AWS Entity Resolution 数据匹配服务 API 参考》中的 [CreateIdMappingWorkflow](#)

- ID 命名空间

创建一个 IdNamespace 对象，该对象存储解释数据集及其使用方法的元数据。

有关更多信息，请参阅以下主题：

《CloudFormation 用户指南》中的 [AWS::EntityResolution::IdNamespace](#)

《AWS Entity Resolution 数据匹配服务 API 参考》中的 [CreateIdNamespace](#)

- PolicyStatement

创建一个 PolicyStatement 对象。

有关更多信息，请参阅以下主题：

《CloudFormation 用户指南》中的 [AWS::EntityResolution::PolicyStatement](#)

《AWS Entity Resolution 数据匹配服务 API 参考》中的 [AddPolicyStatement](#)

## 了解更多关于 CloudFormation

要了解更多关于 CloudFormation，请参阅以下资源：

- [AWS CloudFormation](#)

了解更多关于 CloudFormation

- [AWS CloudFormation 用户指南](#)
- [CloudFormation API 参考](#)
- [AWS CloudFormation 命令行界面用户指南](#)

## 的配额 AWS Entity Resolution 数据匹配服务

您的每个配额 AWS 账户 都有默认配额，以前称为限制 AWS 服务。除非另有说明，否则每个配额都是特定于区域的。您可以申请增加某些配额，但无法增加其他配额。

要查看的配额 AWS Entity Resolution 数据匹配服务，请打开 [Service Quotas 控制台](#)。在导航窗格中，选择 AWS 服务，然后选择 AWS Entity Resolution 数据匹配服务。

要请求提高配额，请参阅《Service Quotas 用户指南》中的[请求提高配额](#)。如果配额在服务配额中尚不可用，请使用[提高限制表格](#)。

您的 AWS 账户 配额与以下有关 AWS Entity Resolution 数据匹配服务。

Name	默认值	可调整	说明
并发 ID 映射作业数	每个受支持的区域：1 个	否	在当前 AWS 区域中可以同时处理的 ID 映射工作流程的最大数量。
并发匹配作业	每个受支持的区域：1 个	否	当前 AWS 区域中可以同时处理的匹配工作流程的最大数量。
ID 映射工作流程数	每个受支持的区域：10 个	<u>是</u>	您可在当前 AWS 区域的此账户中创建的最大 ID 映射工作流程数量。
ID 命名空间	每个受支持的区域：10 个	<u>是</u>	您可在当前 AWS 区域的此账户中创建的最大 ID 命名空间数量。
匹配流程数	每个受支持的区域：10 个	<u>是</u>	您可在当前 AWS 区域的此账户中创建的最大匹配工作流程数量。
GenerateMatchId API 请求的速率	每个受支持的区域：10 个	<u>是</u>	每秒 GenerateMatchId API 请求的最大数量

Name	默认值	可调整	说明
GetMatchId API 请求的速率	每个受支持的区域：50 个	<a href="#"><u>是</u></a>	每秒 GetMatchId API 请求的最大数量。
每个基于机器学习的匹配工作流程的记录数	每个受支持的区域：1.5 亿个	<a href="#"><u>是</u></a>	在 af-south-1、ap-northeast-2、eu-west-2 的 AWS 区域中，此账户中基于机器学习的匹配工作流程可以处理的最大记录数。
每个基于机器学习的匹配工作流程的记录数	每个受支持的区域：6 亿个	<a href="#"><u>是</u></a>	在 ap-northeast-1、ap-southeast-1、ap-southeast-2、ca-central-1、eu-central-1、eu-central-1、eu-west-1、eu-west-1、eu-west-1、eu-west-1、us-east-1、us-east-2、us-west-2。AWS
每个提供商 ID 映射工作流程的记录数	每个受支持的区域：1.5 亿个	<a href="#"><u>是</u></a>	在 af-south-1、ap-northeast-2、eu-west-2 的 AWS 区域中，此账户中可以处理的最大提供商 ID 映射记录数。

Name	默认值	可调整	说明
每个提供商 ID 映射工作流程的记录数	每个受支持的区域：2.5 亿个	<a href="#"><u>是</u></a>	在 ap-northeast-1、ap-southeast-1、ap-southeast-2、ca-central-1、eu-central-1、eu-central-1、eu-central-1、eu-west-1、us-east-1、us-east-1 等 AWS 区域中可以处理的最大提供商 ID 映射记录数 us-east-2、us-west-2。
每个基于提供商服务的匹配流程的记录数	每个受支持的区域：1 亿个	<a href="#"><u>是</u></a>	当前 AWS 区域中该账户中基于提供商服务的匹配工作流程可以处理的最大记录数。
每个基于规则的 ID 映射工作流程的记录数	每个受支持的区域：10 亿个	<a href="#"><u>是</u></a>	在 ap-northeast-1、ap-southeast-1、ap-southeast-2、ca-central-1、eu-central-1、eu-central-1、eu-central-1、eu-west-1、us- eaa 的 AWS 区域中，可以处理的最大记录数 us-1、us-east-2、us-west-2。
每个基于规则的 ID 映射工作流程的记录数	每个受支持的区域：1.5 亿个	<a href="#"><u>是</u></a>	在 af-south-1、ap-northeast-2、eu-west-2 等 AWS 区域中，此账户中可以处理的最大记录数，用于基于规则的 ID 映射。

Name	默认值	可调整	说明
每个基于规则的匹配工作流程的记录数	每个受支持的区域：1 亿个	<a href="#">是</a>	当前 AWS 区域中该账户中基于规则的匹配工作流程可以处理的最大记录数。
架构映射数	每个受支持的区域：50 个	<a href="#">是</a>	您可以在当前 AWS 区域的此账户中创建的最大架构映射数。

## API 节流配额

资源	速率限制	说明
CreateMatchingWorkflow 请求速率	5 TPS	每秒 CreateMatchingWorkflow API 调用的最大次数。
DeleteMatchingWorkflow 请求速率	5 TPS	每秒 DeleteMatchingWorkflow API 调用的最大次数。
GetMatchingWorkflow 请求速率	5 TPS	每秒 GetMatchingWorkflow API 调用的最大次数。
ListMatchingWorkflows 请求速率	5 TPS	每秒 ListMatchingWorkflows API 调用的最大次数。
UpdateMatchingWorkflow 请求速率	5 TPS	每秒 UpdateMatchingWorkflow API 调用的最大次数。
CreateSchemaMapping 请求速率	5 TPS	每秒 CreateSchemaMapping API 调用的最大次数。
DeleteSchemaMapping 请求速率	5 TPS	每秒 DeleteSchemaMapping API 调用的最大次数。

资源	速率限制	说明
GetSchemaMapping 请求速率	5 TPS	每秒 GetSchemaMapping API 调用的最大次数。
ListSchemaMappings 请求速率	5 TPS	每秒 ListSchemaMappings API 调用的最大次数。
UpdateSchemaMapping 请求速率	5 TPS	每秒 UpdateSchemaMapping API 调用的最大次数。
GetPartnerComponent 请求速率	5 TPS	每秒 GetPartnerComponent API 调用的最大次数。
ListPartnerComponents 请求速率	5 TPS	每秒 ListPartnerComponents API 调用的最大次数。
TagResource 请求速率	5 TPS	每秒 TagResource API 调用的最大次数。
UntagResource 请求速率	5 TPS	每秒 UntagResource API 调用的最大次数。
ListTagsForResource 请求速率	5 TPS	每秒 ListTagsForResource API 调用的最大次数。
CreateIdMappingWorkflow 请求速率	5 TPS	每秒 CreateIdMappingWorkflow API 调用的最大次数。
DeleteIdMappingWorkflow 请求速率	5 TPS	每秒 DeleteIdMappingWorkflow API 调用的最大次数。
GetIdMappingWorkflow 请求速率	5 TPS	每秒 GetIdMappingWorkflow API 调用的最大次数。
ListIdMappingWorkflow 请求速率	5 TPS	每秒 ListIdMappingWorkflow API 调用的最大次数。

资源	速率限制	说明
UpdateIdMappingWorkflow 请求速率	5 TPS	每秒 UpdateIdMappingWorkflow API 调用的最大次数。
ListProviderServices 请求速率	5 TPS	每秒 ListProviderServices API 调用的最大次数。
GetProviderService 请求速率	5 TPS	每秒 GetProviderService API 调用的最大次数。
CreateIdNamespace 请求速率	5 TPS	每秒 CreateIdNamespace API 调用的最大次数。
DeleteIdNamespace 请求速率	5 TPS	每秒 DeleteIdNamespace API 调用的最大次数。
GetIdNamespace 请求速率	5 TPS	每秒 GetIdNamespace API 调用的最大次数。
ListIdNamespaces 请求速率	5 TPS	每秒 ListIdNamespaces API 调用的最大次数。
UpdateIdNamespace 请求速率	5 TPS	每秒 UpdateIdNamespace API 调用的最大次数。
AddPolicyStatement 请求速率	5 TPS	每秒 AddPolicyStatement API 调用的最大次数。
DeletePolicyStatement 请求速率	5 TPS	每秒 DeletePolicyStatement API 调用的最大次数。
GetPolicy 请求速率	5 TPS	每秒 GetPolicy API 调用的最大次数。
PutPolicy 请求速率	5 TPS	每秒 PutPolicy API 调用的最大次数。

资源	速率限制	说明
GetMatchingJob 请求速率	10 TPS	每秒 GetMatchingJob API 调用的最大次数。
ListMatchingJobs 请求速率	5 TPS	每秒 ListMatchingJobs API 调用的最大次数。
StartMatchingJob 请求速率	5 TPS	每秒 StartMatchingJob API 调用的最大次数。
GetMatchId 请求速率	50 TPS	每秒 GetMatchId API 调用的最大次数。
GetIdMappingJob 请求速率	10 TPS	每秒 GetIdMappingJob API 调用的最大次数。
ListIdMappingJobs 请求速率	5 TPS	每秒 ListIdMappingJobs API 调用的最大次数。
StartIdMappingJob 请求速率	5 TPS	每秒 StartIdMappingJob API 调用的最大次数。
BatchDeleteUniqueId 请求速率	5 TPS	每秒 BatchDeleteUniqueId API 调用的最大次数。

# 的文档历史记录 AWS Entity Resolution 数据匹配服务 用户指南

下表描述了文档版本 AWS Entity Resolution 数据匹配服务。

如需有关此文档的更新通知，您可以订阅 RSS 源。要订阅 RSS 更新，您必须为当前使用的浏览器启用 RSS 插件。

变更	说明	日期
<a href="#">Support 支持传递匹配</a>	AWS Entity Resolution 数据匹配服务 现在支持使用高级规则类型的基于规则的匹配工作流的传递匹配。传递匹配处理所有规则级别的记录，允许匹配的记录充当链接，将不匹配的记录与现有匹配组连接起来。	2026年4月27日
<a href="#">更新现有策略</a>	在 AWSEntityResolutionConsoleFullAccess 托管策略添加了以下新权限：CustomerProfilesIntegrationAccess 。	2025 年 12 月 15 日
<a href="#">支持 Connect 客户客户档案</a>	添加了在使用基于规则或基于机器学习的匹配工作流程时，将经过重复数据删除的客户记录直接导出到 Connect 客户档案的功能。	2025 年 12 月 15 日
<a href="#">对 FIPS 的支持</a>	AWS Entity Resolution 数据匹配服务 现在通过以下方式支持符合联邦信息处理标准 (FIPS) 140-2 的端点。AWS PrivateLink	2025 年 10 月 21 日

<a href="#">身份映射工作流程-更新</a>	现在，客户可以在基于规则的 ID 映射工作流程中使用增量处理来更高效地处理大型数据集。客户还可以从 ID 映射工作流程中删除记录，以帮助遵守数据管理法规。	2025 年 9 月 22 日
<a href="#">对跨区域的 Support</a>	现在，客户可以使用其他名称空间中的数据 AWS 区域 作为 ID 命名空间、映射工作流程或 ID 映射工作流程的输入。	2025 年 9 月 8 日
<a href="#">Support 支持增强的规则条件和增量删除</a>	客户现在可以将规则条件与布尔运算符和新的匹配函数（例如）一起使用 ExactMany ToMany，从而通过精确匹配和模糊匹配的组合来实现更精确的匹配标准。此外，客户可以使用 Amazon S3 文件在高级匹配工作流程中逐步删除记录。	2025 年 7 月 30 日
<a href="#">比赛编号处理说明</a>	添加了以下说明：“修改或生成匹配 ID”和“查找匹配 ID”选项需要在匹配工作流程中自动处理节奏。	2025 年 7 月 17 日
<a href="#">生成新的匹配 ID</a>	使用基于规则的匹配工作流程时，客户现在可以查找和修改现有的匹配 ID 或生成新的匹配 ID。	2025 年 6 月 2 日
<a href="#">基于提供商服务的匹配工作流程 — 更新</a>	现在，在使用基于 TransUnion 提供商服务的匹配工作流程时，客户可以使用 IPV4、IPV6 和 MAID 等数字标识符。	2025 年 4 月 21 日

<a href="#">Amazon CloudWatch 日志</a>	AWS Entity Resolution 数据匹配服务 现在支持 CloudWatch 日志集成，允许您启用详细的工作流程日志记录，以捕获任务执行指标、时间和处理统计数据，这些统计数据可以传输到 CloudWatch 日志、Amazon S3 或 Amazon Data Firehose 目标。	2025 年 4 月 14 日
<a href="#">身份映射工作流程-更新</a>	现在，客户可以在使用 ID 映射工作流程时设置 AWS Glue 分区。	2025 年 3 月 25 日
<a href="#">配额-更新</a>	Documentation-only 更新。 Rule-based 匹配工作流程可以处理多达 1 亿条记录，而基于机器学习的匹配工作流程可以处理多达 2.5 亿条记录。需要更高限额的客户请联系服务团队。	2025 年 2 月 7 日
<a href="#">架构映射-更新</a>	Documentation-only 更新以说明全名、完整地址和完整电话属性类型支持标准化。	2025 年 1 月 17 日
<a href="#">提供商集成</a>	Documentation-only 更新。 客户可以学习如何作为提供商服务与集成 AWS Entity Resolution 数据匹配服务。	2024 年 8 月 8 日
<a href="#">身份映射工作流程-更新</a>	现在，客户可以使用匹配规则在 ID 映射工作流程中转换第一方数据。	2024 年 7 月 23 日

<a href="#">匹配工作流程-更新</a>	客户现在可以从基于规则的工作流程或 ML-based 匹配的工作流程中删除记录，以帮助遵守数据管理法规。	2024 年 4 月 8 日
<a href="#">身份映射工作流程-更新</a>	客户现在可以跨多个使用身份映射工作流程 AWS 账户。	2024 年 4 月 2 日
<a href="#">CloudFormation 资源-新的和更新的资源</a>	AWS Entity Resolution 数据匹配服务 添加了以下资源：AWS::EntityResolution::IdNamespace AWS::EntityResolution::PolicyStatement 并更新了以下资源：AWS::EntityResolution::IdMappingWorkflow 。	2024 年 4 月 2 日
<a href="#">查找比赛编号</a>	客户现在可以为已处理的基于规则的工作流程找到相应的 Match ID 和关联规则。	2024 年 3 月 25 日
<a href="#">匹配工作流程-更新</a>	AWS Entity Resolution 数据匹配服务 现在支持在基于 LiveRamp 提供商服务的 PII-based 匹配工作流程中分配 RAMPID。	2024 年 2 月 12 日
<a href="#">AWS PrivateLink</a>	AWS Entity Resolution 数据匹配服务 现在支持额外的数据安全性 AWS PrivateLink ，可帮助客户私密访问托管在上的服务 AWS。	2023 年 10 月 20 日

<a href="#">CloudFormation 资源-新的和更新资源</a>	AWS Entity Resolution 数据匹配服务 添加了以下资源 : AWS::EntityResolution:IdMappingWorkflow 并更新了以下资源 : AWS::EntityResolution::MatchingWorkflow 和AWS::EntityResolution::Schemamapping 。	2023 年 10 月 19 日
<a href="#">更新现有策略</a>	AWSEntityResolutionConsoleFullAccess 托管策略中添加了以下新权限 : ADXReadAccess 和ManageEventBridgeRules 。	2023 年 10 月 16 日
<a href="#">架构映射-更新</a>	客户现在可以编辑和更新现有的数据架构。	2023 年 10 月 16 日
<a href="#">匹配工作流程-更新</a>	客户现在可以选择首选的数据提供商服务来帮助匹配和关联他们的数据。	2023 年 10 月 16 日
<a href="#">身份映射工作流程</a>	客户可以使用此新工作流程来指定 ID 映射详细信息、选择所需的 ID 映射方法以及指定数据输入和输出字段。	2023 年 10 月 16 日
<a href="#">CloudFormation 整合</a>	AWS Entity Resolution 数据匹配服务 现在与集成 CloudFormation。	2023 年 8 月 24 日
<a href="#">AWS 托管策略更新-新策略</a>	AWS Entity Resolution 数据匹配服务 添加了两个新的托管策略。	2023 年 8 月 18 日

[初始版本](#)

《AWS Entity Resolution 数据匹配服务 用户指南》的初始版本  
2023 年 7 月 26 日

# AWS Entity Resolution 数据匹配服务 术语表

## Amazon 资源名称 ( ARN )

AWS 资源的唯一标识符。当您需要在所有资源中明确指定资源时，例如在 AWS Entity Resolution 数据匹配服务 策略 AWS Entity Resolution 数据匹配服务、Amazon Relational Database Service (Amazon RDS) 标签和 API 调用中，则需要使用 ARN。

## 属性类型

输入字段的属性类型。[创建架构映射](#)时，您可以从预先配置的值列表中选择属性类型，例如姓名、地址、电话号码或电子邮件地址。属性类型告诉您呈现的是 AWS Entity Resolution 数据匹配服务 哪种数据，从而可以对其进行正确分类和标准化。

## 自动处理

匹配 workflow 作业的处理节奏选项，当您的数据输入发生变化时，它可以自动运行。

此选项仅适用于[基于规则的匹配](#)。

默认情况下，匹配 workflow 作业的处理节奏设置为“[手动](#)”，这样便可以按需运行。您可以将自动处理设置为在数据输入发生变化时自动运行匹配的工作流程作业。这样可以使您的匹配工作流程输出保持最新。

## AWS KMS key 进行筛选

这是您用于静态加密的 AWS KMS Amazon 资源名称 (ARN)。如果未提供，则系统将使用 AWS Entity Resolution 数据匹配服务 托管 KMS 密钥。

## 批量工作流程

按计划间隔运行的进程，用于匹配和解析整个数据集中的数据。中的批处理 workflow 最 AWS Entity Resolution 数据匹配服务 适合用于初始设置、定期完全刷新以及源数据集和目标数据集均发生重大变化的场景。

## 明文

未受加密保护的数据。

## 置信度 (ConfidenceLevel)

对于 ML 匹配，这是 ML 识别匹配的记录集 AWS Entity Resolution 数据匹配服务 时所应用的置信水平。这是将包含在输出中的[匹配工作流程元数据](#)的一部分。

## 解密

将加密数据转换回其原始形式的过程。只有获得密钥，才能进行解密。

## 加密

使用称为密钥的机密值将数据编码成看似随机的形式的过程。如果无法访问密钥，就无法确定原始明文。

## 组名

组名引用整个输入字段组，可以帮助您将已解析的数据分组在一起以进行匹配。

例如，如果有三个输入字段：`first_name`、`middle_name`、`last_name`、和 `suffix`，则可以通过在组名中输入匹配和输出来将它们分组在一起。`full_name`

## 哈希

哈希意味着应用一种加密算法，该算法会生成不可逆且唯一的固定大小的字符串，称为哈希。AWS Entity Resolution 数据匹配服务 使用安全哈希算法 256 位 (SHA256) 哈希协议，并将输出 32 字节的字符串。在中 AWS Entity Resolution 数据匹配服务，您可以选择是否对输出中的数据值进行哈希处理。

## 哈希协议 (HashingProtocol)

AWS Entity Resolution 数据匹配服务 使用安全哈希算法 256 位 (SHA256) 哈希协议，并将输出 32 字节的字符串。这是将包含在输出中的[匹配工作流程元数据](#)的一部分。

## ID 映射方法

您希望如何执行 ID 映射。

有两种 ID 映射方法：

- Rule-based — 在 ID 映射工作流程中，使用匹配规则将第一方数据从源转换为目标的方法。
- 提供者服务-在 ID 映射工作流程中，使用提供者服务将第三方编码的数据从源转换为目标的方法。

AWS Entity Resolution 数据匹配服务 目前支持 LiveRamp 作为基于提供商服务的身份映射方法。您必须订阅 LiveRamp 直通 AWS Data Exchange 才能使用此方法。有关更多信息，请参阅 [步骤 1：在上订阅提供商服务 AWS Data Exchange](#)。

## ID 映射工作流程

一种数据处理作业，它根据指定的 ID 映射方法将输入数据源中的数据映射到输入数据目标。它会生成一个 ID 映射表。此工作流程要求您指定 [ID 映射方法](#) 以及要从源转换为目标的输入数据。

您可以将 ID 映射工作流程设置为自己运行 AWS 账户 或跨两个运行 AWS 账户。

## ID 命名空间

中的一种资源 AWS Entity Resolution 数据匹配服务，其中包含解释多个数据集 AWS 账户 以及如何在 [ID 映射工作流程](#) 中使用这些数据集的元数据。

ID 命名空间有两种类型：SOURCE 和 TARGET。SOURCE 包含将在 ID 映射工作流程中处理的源数据的配置。TARGET 包含所有源都将解析为的目标数据的配置。要定义要跨两个集合解析的输入数据 AWS 账户，请创建一个 ID 命名空间源和一个 ID 命名空间目标，以将数据从一个集合 (SOURCE) 转换为另一个集合 (TARGET)。

在您和其他成员创建 ID 命名空间并运行 ID 映射工作流程后，您可以加入协作，在 AWS Clean Rooms ID 映射表上运行多表联接，并分析数据。

有关更多信息，请参阅 [AWS Clean Rooms 《用户指南》](#)。

## 增量工作流程

仅匹配和解析自上次运行以来的新记录或更新的记录，而不是处理整个数据集的流程。中的增量工作流程最 AWS Entity Resolution 数据匹配服务 适合频繁更新，以便在数据集只有一小部分发生变化时保持数据的新鲜度。

## 输入字段

输入字段对应于 AWS Glue 输入数据表中的列名。

## 输入源 ARN (AR InputSource N)

为 AWS Glue 表输入生成的亚马逊资源名称 (ARN)。这是将包含在输出中的[匹配工作流程元数据](#)的一部分。

## 基于机器学习的匹配

基于机器学习的匹配 (机器学习匹配) 可在您的数据中查找可能不完整或可能看起来不完全相同的匹配项。机器学习匹配是一个预设过程，它将尝试匹配您输入的所有数据的记录。机器学习匹配返回每个[匹配数据集的匹配 ID](#) 和[置信度](#)。

## 手动处理

匹配 workflow 作业的处理节奏选项，使其能够按需运行。

此选项是默认设置的，可用于[基于规则的匹配和基于机器学习的匹配](#)。

## Many-to-Many 匹配

Many-to-many matching 会比较相似数据的多个实例。已分配相同匹配键的输入字段中的值将相互匹配，无论它们位于同一个输入字段还是不同的输入字段中。

例如，您可能有多个电话号码输入字段，例如mobile\_phone和home\_phone，它们具有相同的匹配键“Phone”。使用多对多匹配将输入字段中的数据与mobile\_phone输入字段中的数据以及mobile\_phone输入字段中的数据进行比较。home\_phone

匹配规则使用 (或) 运算评估具有相同匹配键的多个输入字段中的数据，一对多匹配比较多个输入字段中的值。这意味着，如果两条记录之间有任何组合mobile\_phone或home\_phone匹配，“电话”匹配键将返回匹配项。对于匹配键“Phone”来查找匹配项，Record One mobile\_phone = Record Two mobile\_phone或Record One mobile\_phone = Record Two home\_phone或Record One home\_phone = Record Two home\_phone或Record One home\_phone = Record Two mobile\_phone。

## 比赛 ID (matchID)

对于基于规则的匹配和 ML 匹配，这是由每个匹配的记录集生成 AWS Entity Resolution 数据匹配服务并应用于的 ID。这是将包含在输出中的[匹配工作流程元数据](#)的一部分。

## 匹配键 (MatchKey)

Match key 指示将 AWS Entity Resolution 数据匹配服务 哪些输入字段视为相似数据，哪些输入字段应视为不同的数据。这有助于 AWS Entity Resolution 数据匹配服务 自动配置基于规则的匹配规则，并比较存储在不同输入字段中的相似数据。

如果您想将数据中的mobile\_phone输入字段和输入字段等多种类型的电话号码信息进行比较，则可以为其提供匹配键“Phone”。home\_phone然后，可以将基于规则的匹配配置为在所有输入字段中使用“或”语句与“电话”匹配键比较数据（参见“[One-to-One 匹配](#)工作流程”部分中的[Many-to-Many 匹配](#)和匹配定义）。

如果您希望基于规则的匹配完全分开考虑不同类型的电话号码信息，则可以创建更具体的匹配键，例如“Mobile\_Phone”和“Home\_Phone”。然后，在设置匹配工作流程时，您可以指定在基于规则的匹配中如何使用每个电话匹配键。

如果没有 MatchKey 为特定的输入字段指定“否”，则该字段不能用于匹配，但可以执行匹配工作流程，并且可以根据需要输出。

## 匹配键名称

分配给匹配键的名称。

## 匹配规则 (MatchRule)

对于基于规则的匹配，这是生成匹配记录集所应用的规则编号。这是将包含在输出中的[匹配工作流程元数据](#)的一部分。

## Matching

组合和比较来自不同输入字段、表或数据库的数据，并根据满足某些匹配条件（例如，通过匹配规则或模型）确定其中哪个相似或“匹配”的过程。

## 匹配工作流程

您为指定要匹配的输入数据而设置的过程以及如何执行匹配。

## 匹配的工作流程描述

您可以选择输入的匹配工作流程的可选描述。如果您创建多个工作流程，描述可以帮助您区分匹配的工作流程。

## 匹配工作流程名称

您指定的匹配工作流程的名称。

### Note

匹配的工作流程名称必须是唯一的。它们的名称不能相同，否则将返回错误。

## 匹配工作流程元数据

在匹配的工作流作业 AWS Entity Resolution 数据匹配服务 期间生成和输出的信息。输出时需要此信息。

## 标准化 (ApplyNormalization)

选择是否按照架构中的定义对输入数据进行标准化。标准化通过删除多余的空格和特殊字符并标准化为小写格式来标准化数据。

例如，如果输入字段的属性类型为 [Full phone](#)，并且输入表中的值格式为 (123) 456-7890，则 AWS Entity Resolution 数据匹配服务 会将值标准化为 1234567890。

### Note

仅支持[姓名](#)、[地址](#)、[电话](#)和[电子邮件](#)的群组类型标准化。

以下各节描述了我们的标准标准化规则。

要具体进行 ML-based 匹配，请参阅[标准化 \(ApplyNormalization\)- ML-based 仅限](#)。

主题

- [Name](#)

- [电子邮件](#)
- [Phone](#)
- [地址](#)
- [经过哈希处理](#)
- [Source\\_ID](#)

## Name

### Note

只有名称组类型支持标准化。

名称组类型在控制台中显示为全名，在 API NAME 中也显示为全名。

如果要规范化名称组类型的子类型，请执行以下操作：

- 在控制台中，为全名组分配以下子类型：名字、中间名和姓氏。
- 在 [CreateSchemaMapping](#) API 中，为 NAME 组名分配以下类型：NAME\_FIRSTNAME\_MIDDLE、和。NAME\_LAST

- TRIM = 修剪前导和尾随的空格
- 小写 = 小写所有字母字符
- CONVER@@@ T\_ACCENT = 将带有重音符号的字母转换为普通字母
- REMOVE\_ALL\_NON\_ALPHA = 移除所有非字母字符 [a-z] A-Z

## 电子邮件

### Note

电子邮件群组类型支持标准化。

电子邮件群组类型在控制台中显示为电子邮件地址，在 API EMAIL\_ADDRESS 中也显示为电子邮件地址。

- TRIM = 修剪前导和尾随的空格
- 小写 = 小写所有字母字符

- CONVER@@ T\_ACCENT = 将带有重音符号的字母转换为普通字母
- EMAIL\_ADDRESS\_UTIL\_NORM = 删除用户名中的所有点 (.)，删除用户名中加号 (+) 之后的任何内容，并标准化常见的域名变体
- REMOVE\_ALL\_NON\_EMAIL\_CHARS = 移除所有非字母数字字符 [a-z] 和 [.-@-] A-Z0-9

## Phone

### Note

仅电话组类型支持标准化。

电话组类型在控制台中显示为“完整电话”，在 API PHONE 中显示为“完整电话”。

如果要标准化电话组类型的子类型，请执行以下操作：

- 在控制台中，将以下子类型分配给完整电话组：电话号码和电话国家/地区代码。
  - 在 [CreateSchemaMapping](#) API 中，为 PHONE 组名分配以下类型：**PHONE\_NUMBER** 和 **PHONE\_COUNTRYCODE**
- TRIM = 修剪前导和尾随的空格
  - REMOVE\_ALL\_NON\_NUMERIC = 移除所有非数字字符 [0- 9]
  - REMOVE\_ALL\_LEADING\_ZEROES = 移除所有前导零
  - ENSURE\_PREFIX\_WITH\_MAP，“phonePrefixMap” = 检查每个电话号码并尝试将其与电话中的模式进行匹配。PrefixMap 如果找到匹配项，该规则将添加或修改电话号码的前缀，以确保其符合地图中指定的标准格式。

## 地址

### Note

仅地址组类型支持标准化。

地址组类型在控制台中显示为完整地址，在 API ADDRESS 中也显示为完整地址。

如果要对地址组类型的子类型进行标准化，请执行以下操作：

- 在控制台中，将以下子类型分配给完整地址组：街道地址 1、街道地址 2：街道地址 3 名称、城市名称、州、国家/地区和邮政编码 t

- 在 [CreateSchemaMapping](#) API 中，为 ADDRESS 组名分配以下类型：ADDRESS\_STREET1、ADDRESS\_STREET2、ADDRESS\_STREET3、ADDRESS\_CITYADDRESS 和 ADDRESS\_POSTALCODE

- TRIM = 修剪前导和尾随的空格
- 小写 = 小写所有字母字符
- CONVER@@ T\_ACCENT = 将带有重音符号的字母转换为普通字母
- REMOVE\_ALL\_NON\_ALPHA = 移除所有非字母字符 [a-z] A-Z
- 使用 ADDRESS\_RENAME\_WORD\_MAP 重命名单词 = 用 ADDRESS\_RENAME\_WORD\_MAP 中的单词替换地址 [字符串中的单词](#)
- 使用 ADDRESS\_RENAME\_DELIMATILER\_MAP 重命名分隔符 = 将地址字符串中的分隔符替换为 ADDRESS\_RENAME\_DELIMER\_MAP [中的字符串](#)
- 使用 ADDRESS\_RENAME\_DIRECTION\_MAP 重命名方向 = 将地址字符串中的分隔符替换为 ADDRESS\_RENAME\_DIRECTION\_M [AP 中的字符串](#)
- [使用 ADDRESS\\_RENAME\\_NUMBER\\_MAP 重命名数字 = 用 ADDRESS\\_RENAME\\_NUMBER\\_MAP 中的字符串替换地址字符串中的数字](#)
- 使用 ADDRESS\_RENAME\_SPECIAL\_CHAR\_MAP 重命名\_特殊\_CHAR\_CHAR\_MAP = 将地址字符串中的特殊字符替换为 ADDRESS\_RENAME\_SPECIAL\_CHAR\_ [MAP 中的字符串](#)

## 地址\_重命名\_WORD\_MAP

这些是标准化地址字符串时将重命名的单词。

```
"avenue": "ave",
"bouled": "blvd",
"circle": "cir",
"circles": "cirs",
"court": "ct",
"centre": "ctr",
"center": "ctr",
"drive": "dr",
"freeway": "fwy",
"frwy": "fwy",
"highway": "hwy",
"lane": "ln",
```

```
"parks": "park",
"parkways": "pkwy",
"pky": "pkwy",
"pkway": "pkwy",
"pkwys": "pkwy",
"parkway": "pkwy",
"parkwy": "pkwy",
"place": "pl",
"plaza": "plz",
"plza": "plz",
"road": "rd",
"square": "sq",
"squ": "sq",
"sqr": "sq",
"street": "st",
"str": "st",
"str.": "strasse"
```

## 地址\_重命名\_分隔符\_地图

这些是标准化地址字符串时将重命名的分隔符。

```
",": " ",
".": " ",
"[": " ",
]": " ",
"/": " ",
"_": " ",
"#": " number "
```

## 地址\_重命名\_方向\_地图

这些是在标准化地址字符串时将重命名的方向标识符。

```
"east": "e",
"north": "n",
"south": "s",
"west": "w",
"northeast": "ne",
"northwest": "nw",
"southeast": "se",
"southwest": "sw"
```

## 地址\_重命名\_数字\_地图

这些是在标准化地址字符串时将重命名的数字字符串。

```
"número": "number",  
"numero": "number",  
"no": "number",  
"núm": "number",  
"num": "number"
```

## 地址\_重命名\_SPECIAL\_CHAR\_MAP

这些是在标准化地址字符串时将被重命名的特殊字符串。

```
"ß": "ss",  
"ä": "ae",  
"ö": "oe",  
"ü": "ue",  
"ø": "o",  
"æ": "ae"
```

## 经过哈希处理

- TRIM = 修剪前导和尾随的空格

## Source\_ID

- TRIM = 修剪前导和尾随的空格

## 标准化 (ApplyNormalization)- ML-based 仅限

选择是否按照架构中的定义对输入数据进行标准化。标准化通过删除多余的空格和特殊字符并标准化为小写格式来标准化数据。

例如，如果输入字段的属性类型为NAME，并且输入表中的值格式为Johns Smith，则 AWS Entity Resolution 数据匹配服务 会将值标准化为john smith。

以下各节描述了[基于机器学习的匹配](#)工作流程的标准化规则。

## 主题

- [Name](#)
- [电子邮件](#)
- [Phone](#)

## Name

- TRIM = 修剪前导和尾随的空格
- 小写 = 小写所有字母字符

## 电子邮件

- 小写 = 小写所有字母字符
- 仅用 @ 符号替换 (at) ( 区分大小写 )
- 删除值中任意位置的所有空格
- 移除第一个之外的所有内容 ( "<>" 如果存在 )

## Phone

- TRIM = 修剪前导和尾随的空格
- REMOVE\_ALL\_NON\_NUMERIC = 移除所有非数字字符 [0- 9]
- REMOVE\_ALL\_LEADING\_ZEROES = 移除所有前导零
- ENSURE\_PREFIX\_WITH\_MAP , “phonePrefixMap” = 检查每个电话号码并尝试将其与电话中的模式进行匹配。PrefixMap如果找到匹配项，该规则将添加或修改电话号码的前缀，以确保其符合地图中指定的标准格式。

## One-to-One 匹配

One-to-one matching 比较相似数据的单个实例。具有相同匹配键和相同输入字段中的值的输入字段将相互匹配。

例如，您可能有多个电话号码输入字段，例如mobile\_phone和home\_phone，它们具有相同的匹配键“Phone”。使用一对一匹配将输入字段中的数据与mobile\_phone输入字段中的数

据进行比较，并将mobile\_phone输入字段中的数据与home\_phone输入字段中的数据进行比较。home\_phonemobile\_phone输入字段中的数据不会与home\_phone输入字段中的数据进行比较。

匹配规则使用 ( 或 ) 运算评估具有相同匹配键的多个输入字段中的数据，一对多匹配比较单个输入字段中的值。这意味着，如果mobile\_phone或在两条记录之间匹home\_phone配，“电话”匹配键将返回匹配项。要使用匹配键“电话”来查找匹配项，Record One mobile\_phone = Record Two mobile\_phone或者Record One home\_phone = Record Two home\_phone。

匹配规则通过 ( 和 ) 运算评估具有不同匹配键的输入字段中的数据。如果您希望基于规则的匹配完全分开考虑不同类型的电话号码信息，则可以创建更具体的匹配键，例如“mobile\_phone”和“home\_phone”。如果您想在规则中同时使用两个匹配键来查找匹配项，请使用 AN Record One mobile\_phone = Record Two mobile\_phone D Record One home\_phone = Record Two home\_phone。

## Output

对象列表，每个OutputAttribute对象都有“名称”和“哈希”字段。这些对象中的每一个都表示要包含在AWS Glue 输出表中的一列，以及您是否要对该列中的值进行哈希处理。

## outputs3Path

AWS Entity Resolution 数据匹配服务 将输出表写入的 S3 目标。

## OutputSourceConfig

对象列表，每个 OutputSource 对象都有字段 outputs3Path、ApplyNormalization和 Output t。

## 基于提供商服务的匹配

基于提供商服务的匹配过程旨在将您的记录与首选数据服务提供商和许可数据集进行匹配、关联和增强。您必须通过 AWS Data Exchange 提供商服务订阅才能使用此匹配技术。

AWS Entity Resolution 数据匹配服务 目前已与以下数据服务提供商集成：

- LiveRamp
- TransUnion

- UID 2.0

## Rule-based 匹配

Rule-based 匹配是旨在查找精确匹配项的过程。Rule-based matching 是一组分层的瀑布匹配规则 AWS Entity Resolution 数据匹配服务，由您建议，基于您输入的数据，并且完全可以由您配置。规则条件中提供的所有匹配键都必须完全匹配，才能将比较的数据声明为匹配项并输出关联的元数据。Rule-based matching 会为每个[匹配的数据集返回一个匹配 ID](#) 和一个规则编号。

我们建议定义能够唯一标识实体的规则。对规则进行排序，先找到更精确的匹配项。

例如，假设你有两个规则，规则 1 和规则 2。

这些规则具有以下匹配密钥：

- 规则 1 包括全名和地址
- 规则 2 包括全名、地址和电话

由于规则 1 首先运行，因此规则 2 不会找到任何匹配项，因为规则 1 本来可以找到所有匹配项。

要查找按电话区分的匹配项，请重新排列规则，如下所示：

- 规则 2 包括全名、地址和电话
- 规则 1 包括全名和地址

## 传递匹配

对于使用高级规则类型的[基于规则的匹配工作流程](#)，[传递匹配](#)是一项可选功能。默认情况下，AWS Entity Resolution 数据匹配服务 使用瀑布匹配方法，即在更高规则级别匹配的记录将排除在后续规则之外。启用传递匹配后，将在所有规则级别上处理所有记录。记录的[匹配 ID](#) 在第一次匹配时是固定的，但该记录继续充当链接，将后来的规则中的不匹配记录与先前规则中的匹配组连接起来。

有关更多信息，请参阅 [使用传递匹配](#)。

## 架构

该术语用于定义一组数据的组织和连接方式的结构或布局。

## 架构描述

您可以选择输入的架构的可选描述。如果您创建了多个架构映射，则描述可以帮助您区分架构映射。

## 架构名称

架构的名称。

### Note

架构名称必须是唯一的。它们的名称不能相同，否则将返回错误。

## 架构映射

中的架构映射 [AWS Entity Resolution 数据匹配服务](#) 是告诉您 [AWS Entity Resolution 数据匹配服务](#) 如何解释数据以进行匹配的过程。您可以定义要 [AWS Entity Resolution 数据匹配服务](#) 读入匹配工作流程的输入数据表的架构。

## 架构映射 ARN

为 [架构映射](#) 生成的亚马逊资源名称 (ARN)。

## 唯一 ID

您指定的唯一标识符，必须将其分配给 [AWS Entity Resolution 数据匹配服务](#) 读取的每一行输入数据。

### Example

例如，**Primary\_key**、**Row\_ID** 或 **Record\_ID**。

“唯一 ID” 列为必填字段。

唯一 ID 必须是单个表中的唯一标识符。

唯一 ID 必须满足以下模式：`[a-zA-Z0-9_-]`

在不同的表中，唯一 ID 可以有重复的值。

## [匹配工作流程](#)的最大唯一 ID 长度为 38

唯一标识的最大长度为 257 个字符 [ID 映射工作流程](#)

运行[匹配工作流程](#)时，如果唯一 ID 满足以下条件，则该记录将被拒绝：

- 未指定
- 在同一个表中不是唯一的
- 不同源的属性名称存在重叠
- 超过 38 个字符（仅限基于规则的匹配工作流程）

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。