Data Transfer Hub

部署手册

陈海云

戴晓斌

胡益恺

施乔

版本：v2.0.0

最后更新时间: 2021 年 06 月

亚马逊云科技

目　录

解决方案描述

Data Transfer Hub 是一个安全，可靠，可扩展和可追踪的数据传输解决方案，提供一致的用户体验，使您可以轻松地创建和管理不同数据类型，从不同的来源到 Amazon Web Service 云原生服务的传输任务。当此解决方案启动后，您可以在几分钟内开始数据传输任务。

使用场景

使用 Data Transfer Hub，您可以执行以下任何任务：

（1）在 Amazon S3 之间传输对象。

（2）将数据从其他云服务商的对象存储服务（包括阿里云 OSS，腾讯 COS，七牛 Kodo 以及兼容 Amazon S3 的云存储服务）传输到 Amazon S3。

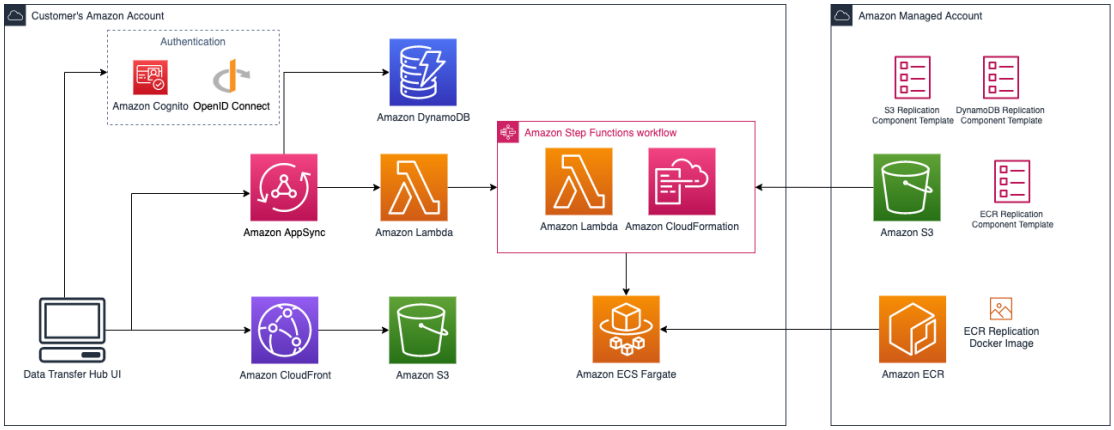
（3）在 Amazon ECR 之间传输容器镜像。

（4）将容器镜像从公共容器镜像仓库（例如 Docker Hub，Google gcr.io，Red Hat Quay.io）传输到 Amazon ECR。

系统架构

架构图

此方案的 CloudFormation 模板会自动部署和配置包含 Amazon AppSync, Amazon DynamoDB, Amazon ECS Fargate, Amazon Lambda, Amazon Step Functions 等服务的架构。该解决方案提供一个托管于 Amazon S3 的 Web 前端，通过 Amazon CloudFront 对外提供服务。Web 前端使用 Amazon Cognito User Pool 或 OpenID Connect（OIDC）服务提供商进行身份验证。



Web 前端提供一站式的创建和管理**传输**任务的用户体验。任何一种数据类型是 Data Transfer Hub 中的一个独立插件，通过 Amazon CloudFormation 打包，且托管于统一的 S3 桶中。当用户创建**传输**任务时，Lambda 将触发 CloudFormation 以开始创建**传输**任务所需要的资源。所有任务及任务的状态都存储在 DynamoDB 表中。

根据这种设计，所有**传输**插件都可以独立使用。开发人员可以轻松地将**传输**插件的 CloudFormation 贡献到 Data Transfer Hub **项目**中。

组件

Amazon Simple Storage Service (Amazon S3)

用于存放前端 UI 界面的部署代码及应用相关日志信息。

Amazon CloudFront

使用 CloudFront 进行前端界面访问加速。

Amazon AppSync

使用 AppSync GraphQL 作为接口的前端与后端数据的接口服务。

Amazon DynamoDB

使用 DynamoDB 存储**传输**任务以及对象**传输**的状态信息，保证在传输过程中如果某个文件**传输**失败，会进行重试。

Amazon Lambda

使用 Lambda 无服务器服务并行启动传输任务，加快数据迁移的效率以节省时间和服务运行成本。

Amazon Elastic Container Service

- 本解决方案使用 Amazon Fargate 运行 ECS 集群，前者是容器的无服务器计算服务。使用 Fargate，您无需预置和管理服务器，而且可以为每个应用程序指定资源并为其付费，并通过设计隔离应用程序来提高安全性。
- Amazon ECS 支持 Docker 并使您能够运行和管理 Docker 容器。 传输任务的容器镜像在 Amazon ECS 上部署和运行，无需进行任何配置更改。

Amazon Cognito User Pool

标准区域使用 Cognito User Pool 服务 提供用户身份验证。

Authing/Auth0/Okta

用户可以选择使用 OpenID Connect (OIDC) Provider 提供用户身份验证。例如，可以使用 Authing，Auth0，Okta 或者其他标准的 OIDC Provider。

部署说明（中国区域）

以下部署说明适用于由光环新网运营的 Amazon Web Service（北京）区域和由西云数据运营的 Amazon Web Service（宁夏）区域（或者该区域暂时不提供 Cognito User Pool 的服务的情形）。

部署前提

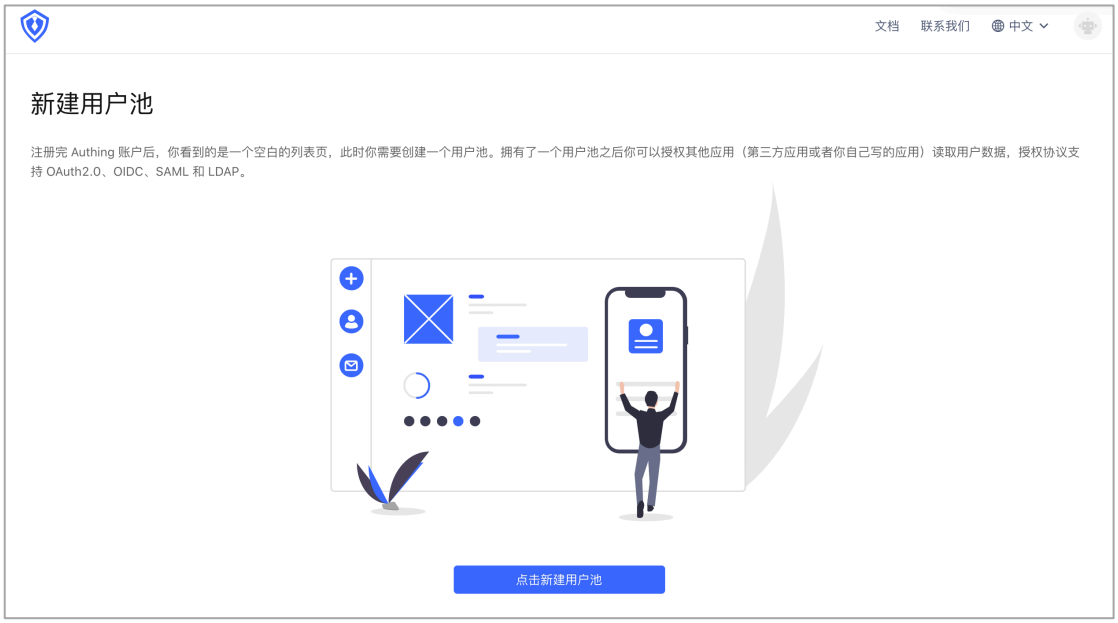
1. 准备已经通过 ICP 备案的域名。

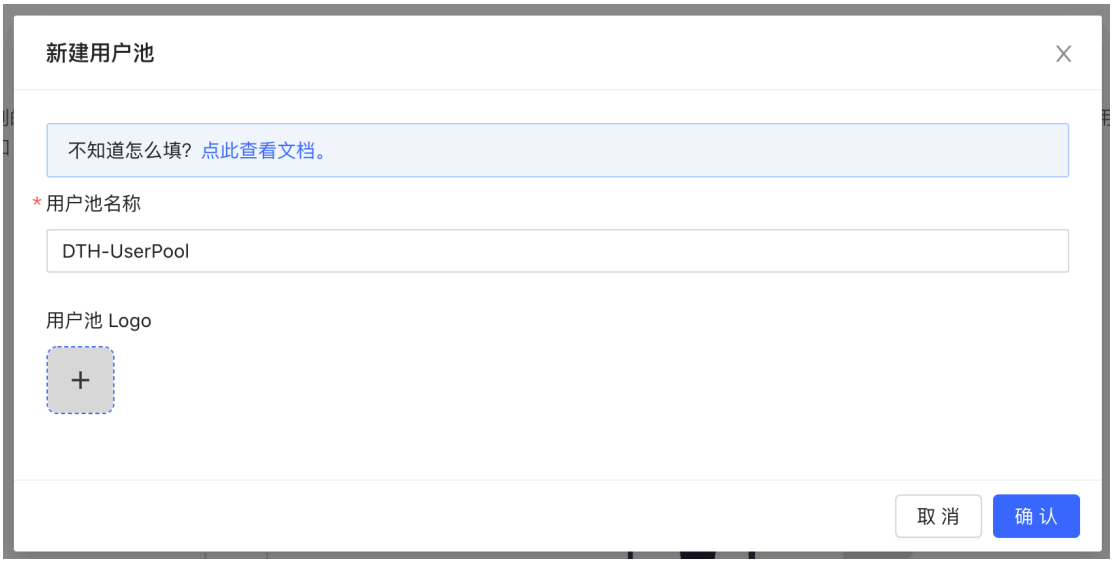
2. 配置 DNS 解析：CloudFormation 堆栈部署完成后，您需要配置 CNAME 解析将域名指向 CloudFront，并且等待解析生效后方可使用该解决方案。您可以通过查看 CloudFormation 的输出来获取 CloudFront 的 DNS 地址。

创建 OIDC 应用

在 Amazon Cognito User Pool 还没有上线的 Amazon Web Service 区域，您可以选择 OIDC 提供身份认证，以下使用 Amazon Web Service 合作伙伴 Authing 作为示例。如果您使用的是 Auth0 或者 Okta 作为 OpenIDConnector 服务提供商，请点击这里了解如何配置。

1. 注册 Authing 开发者账号，点击了解如何注册账号。
2. 账号创建成功后，登陆 Authing 控制台。
3. 在首次登录的时候，需要创建一个用户池

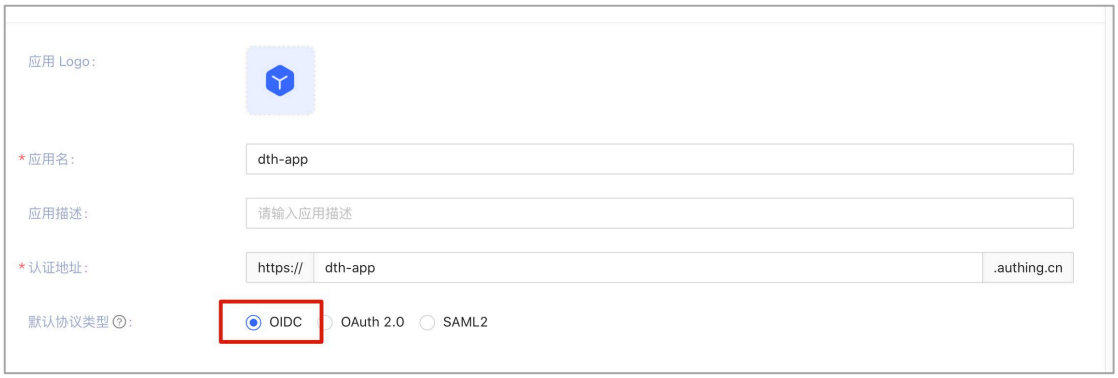




4. 创建 OIDC 应用

5. 设置协议类型为 OIDC

在基础设置中，确保默认协议类型为 OIDC，点击保存

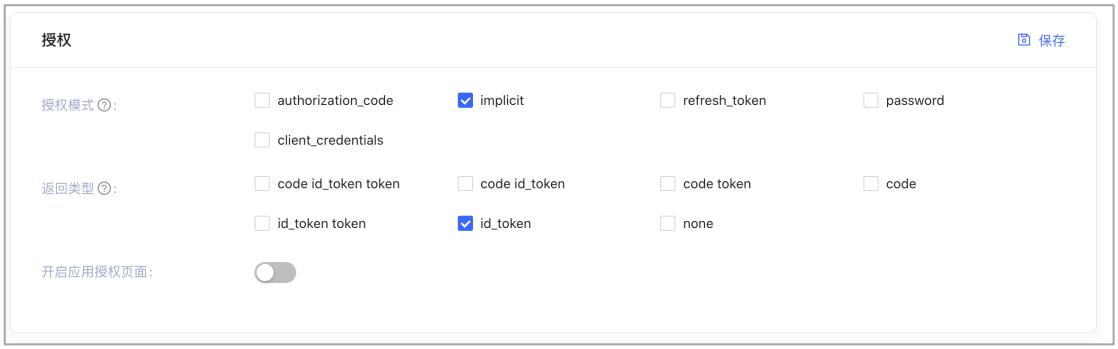


6. 配置回调链接

   在 URL 设置中，设置回调 URL 为 https://您部署 DataTransferHub 的域名/authentication/callback，点击保存。

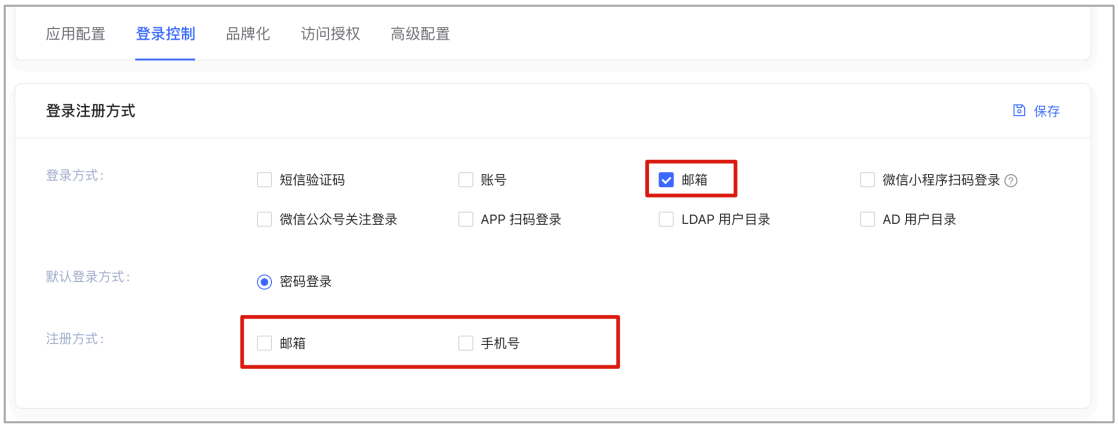
7. 授权配置

   在授权配置中，授权模式选择 implicit，返回类型选择 id_token，点击保存。



8. 修改登录控制配置

   点击登录控制标签页，在登录注册方式中，登录方式默认只允许邮箱登录，注册方式全部取消勾选（停用用户注册功能），点击保存。



9. 修改 id_token 认证签名算法

   在高级设置标签页中，安全性的 id_token 签名算法修改为 RS256，点击保存。

10. 创建 Data Transfer Hub 的登录用户

    在上述步骤中，我们禁用了用户注册功能，在此我们需要手动为 Data Transfer Hub 创建登录用户。

    依次点击左边菜单中的用户管理，用户列表进入到用户列表管理页面。

    点击创建用户按钮

在弹出的创建用户对话框中，选择邮件标签页，填写邮箱地址，密码，以及在选择用户池或应用选项中，选择我们刚创建的 Data Transfer Hub 的应用。点击确定。

注：用户会收到一封来自 Authing 的含有临时密码的邮件，用户首次登录的时候强制需要修改密码。

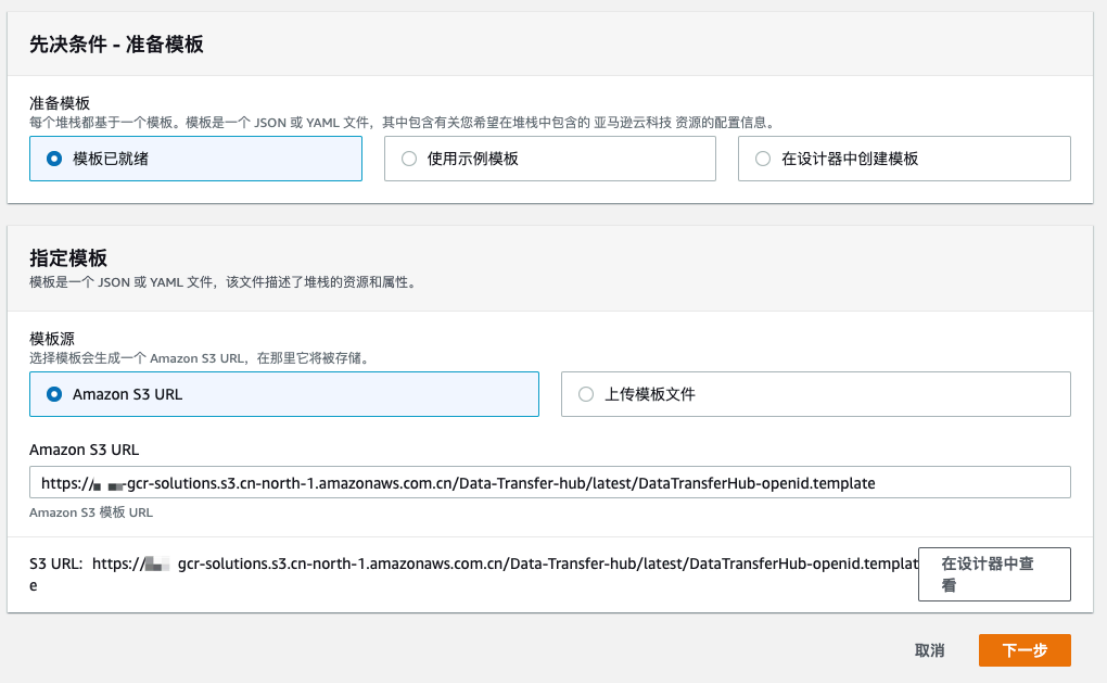
通过 Amazon CloudFormation 部署

1. 打开 Amazon Web Service 管理控制台（如果还没登录会先跳转到登录页面，登录后进入模板启动页面）。您可以使用控制台右上方的区域选择链接, 选择要部署的区域。然后单击下面的按钮以启动 Amazon CloudFormation 模板：

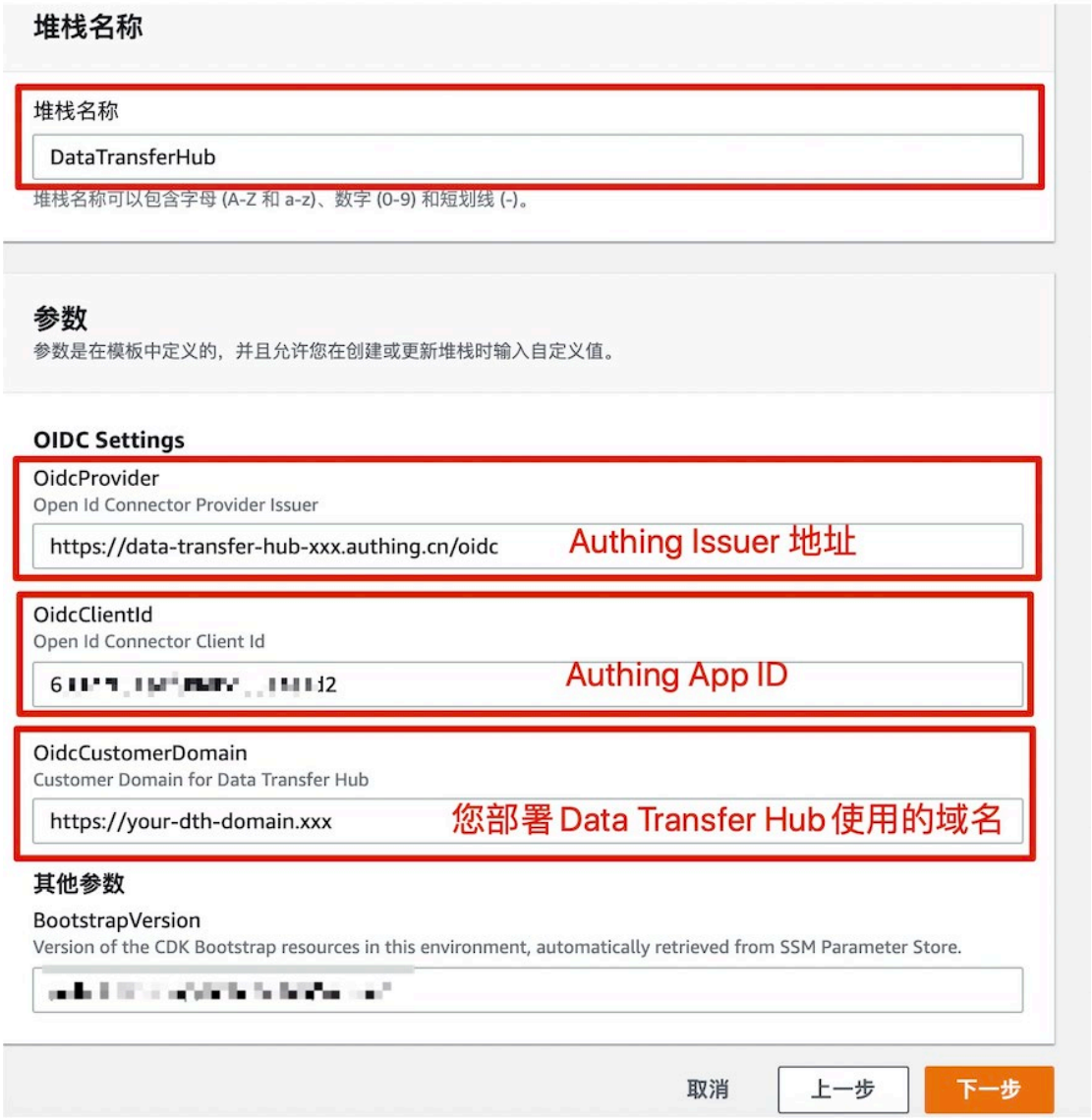
Launch Stack

2. 默认情况下，该模板将在 您登录控制台后后默认的区域启动。 若需在指定的 Amazon Web Service 区域中启动 该解决方案，请使用控制台导航栏中的区域选择器。

3. 在创建堆栈页面上，确认 Amazon S3 URL 文本框中显示正确的模板 URL，然后点击 【下一步】。

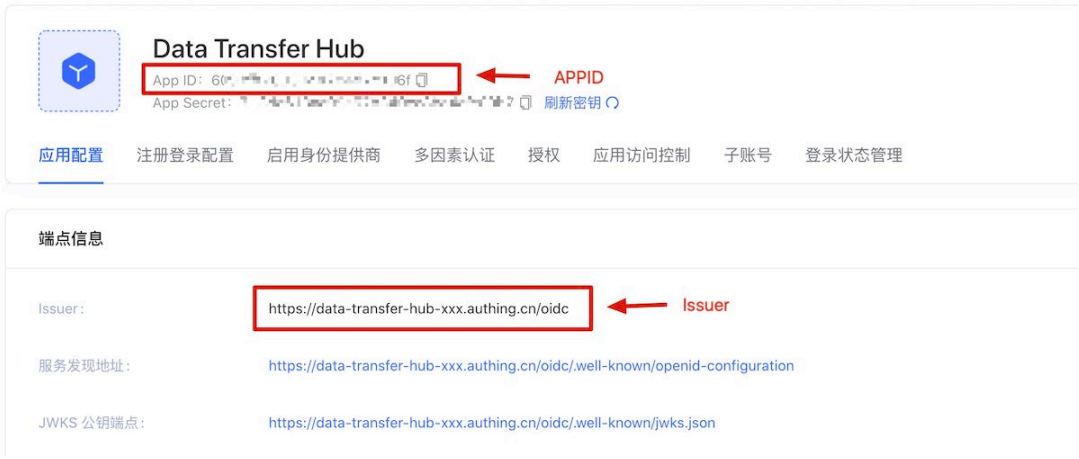


4. 在指定堆栈详细信息下，填写如下的参数。



参数名称	值	用途
OidcProvider	OpenID Connect Provider 应用的 Issuer 地址	服务端对 Token 及用户身份进行校验的服务地址（参考下图填写）
OidcClientId	OpenID Connect Provider 应用的 APPID	在 OIDC 中创建的应用的 APPID（参考下图填写）
OidcCustomerDomain	用户部署 DataTransfer 使用的域名	用户通过该域名访问 DataTransferHub，在中国区为已完成 ICP 备案的域名

打开在之前在 Authing.cn 中创建的 OIDC 应用配置信息，参考下图填写对应参数。



5. 参数填写完成后选择【下一步】。

6. 在配置堆栈选项页面上，保持默认值，点击 【下一步】。

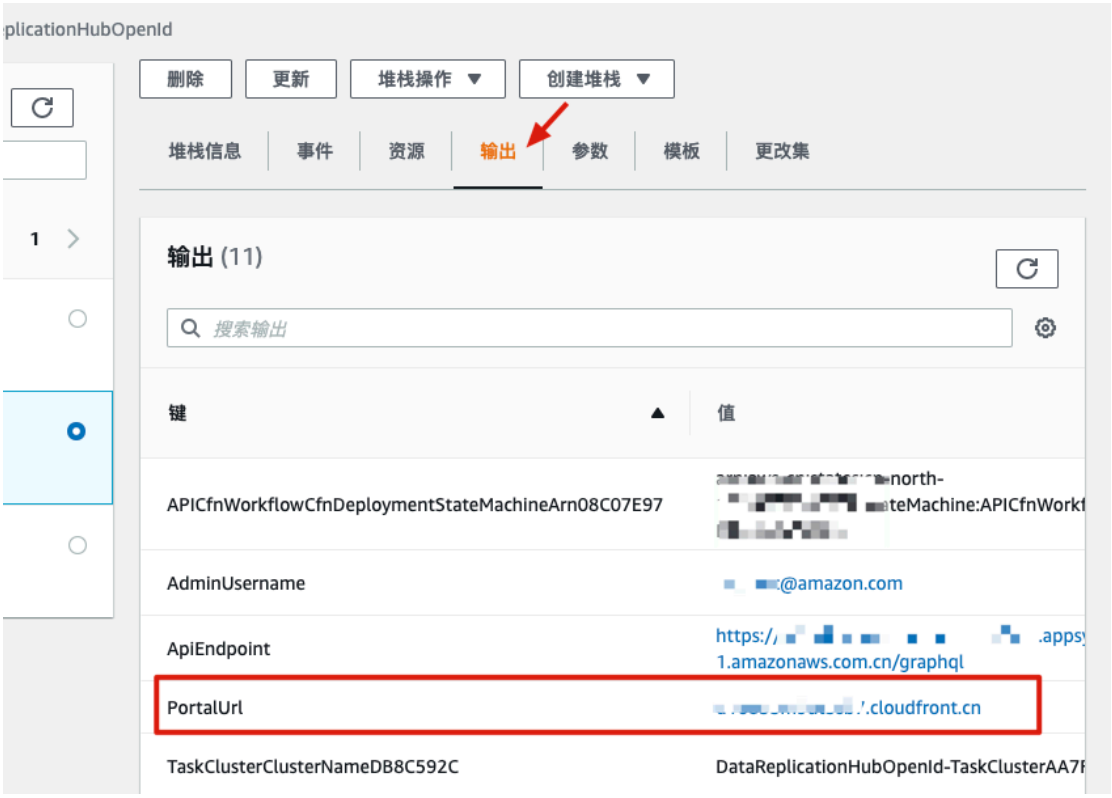
7. 在审核页面上，查看并确认设置。 确保选中确认模板将创建 Amazon Identity and Access Management（IAM）资源的框。

8. 选择【创建堆栈】以部署堆栈。

   您可以在 Amazon CloudFormation 控制台的状态列中查看堆栈的状态。 您应该在大约 15 分钟内看到状态为 CREATE_COMPLETE。

   堆栈创建成功后，可在 Amazon CloudFormation 中输出标签页中看到：

   • **PortalUrl** – 这个 URL 就是最终需要使用域名进行解析访问的 URL。请记住此 URL，在下一步配置域名解析的时候需要用到。



配置域名解析

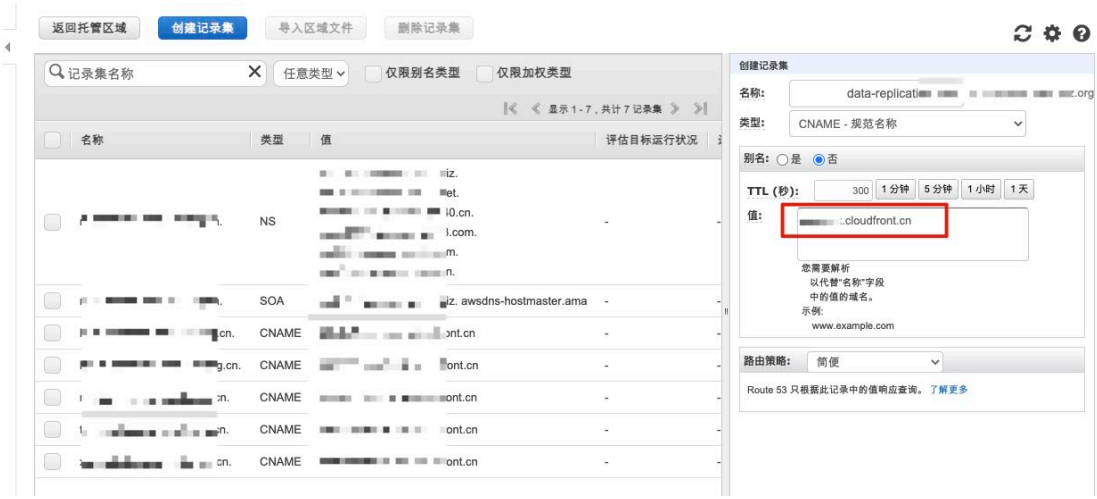
配置 DNS 解析，将 ICP 备案过的域名指向 CloudFront 默认域名，您可以使用自己的 DNS 解析商。以下为 Amazon Route53 上的配置示例：

1. 在 Route 53 中解析域名到 CloudFront

   打开 Amazon Web Service 控制台在 Route 53 中添加 CNAME 记录，以认证该域名为您所有并可以使用。首先，进入 Route 53 的管理页面，点击你的域名。如果没有创建托管区域，请参考文档。

   其中名称为 Data Transfer Hub 指定的访问域名，类型选择 CNAME-规范名称，值为上一步 CloudFormation 输出中的 PortalUrl。

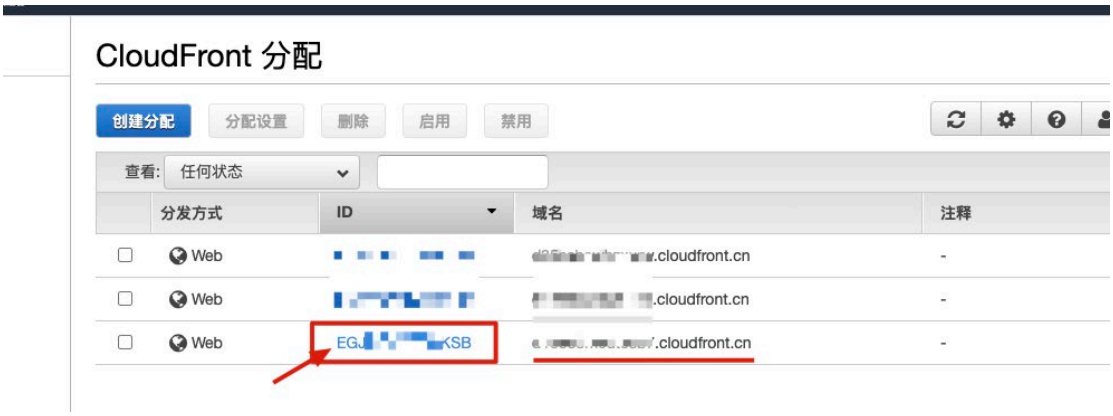
   点击【创建记录集】，在右侧的窗口【名称】输入框中输入记录的【名称】【类型】及【值】三个字段。



2. 点击【创建】

3. 在 CloudFront 中配置对应访问的域名

打开 CloudFront 控制台，在列表中找到刚才 PortalUrl 对应的分配 ID，点击 ID，（或勾选复选框，点击分配设置）



在常规选项卡中点击【编辑】

在备用域名设置中填写在 Route 53 中定义的域名

4. 点击【是的，请修改】等待修改完成。

到此为止，Data Transfer Hub 已经成功部署在了 Amazon Web Service 上。

部署说明（海外区域）

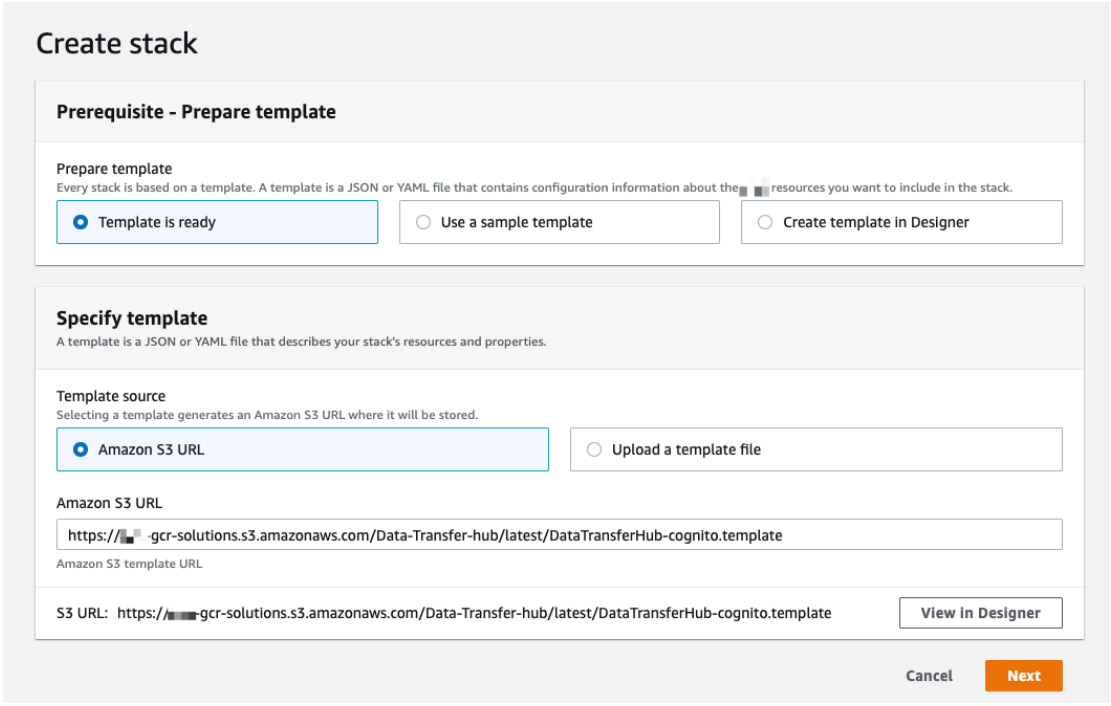
以下部署说明适用于除由光环新网运营的 Amazon Web Service（北京）区域和由西云数据运营的 Amazon Web Service（宁夏）区域以外的 Amazon Web Service 区域（并且该区域提供了 Cognito User Pool 服务，否则请参考中国站的部署文档，使用 OIDC 作为身份验证的方式部署）。

通过 Amazon CloudFormation 部署

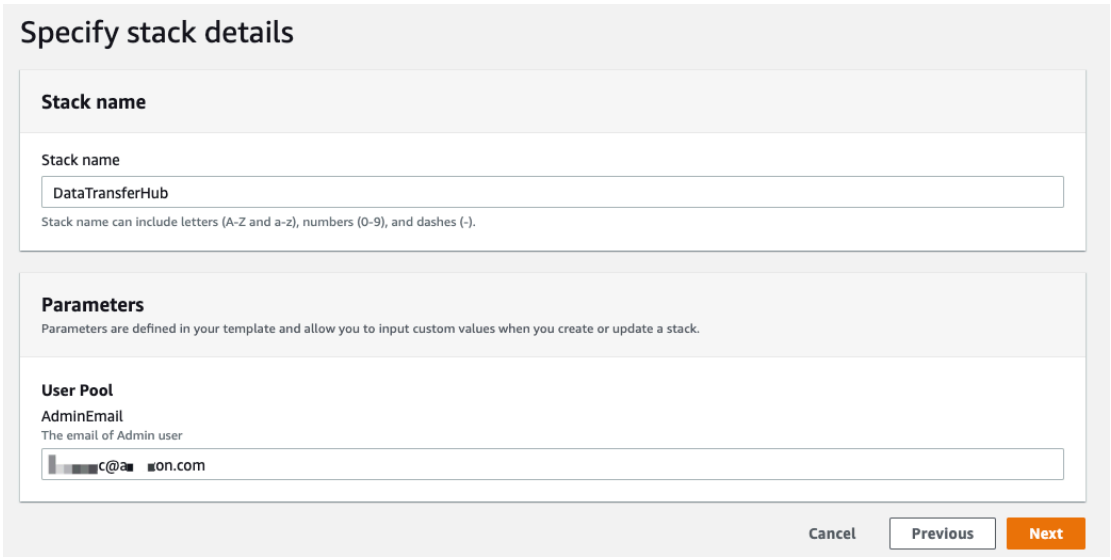
1. 首先，登录 Amazon Web Service 控制台，以下步骤针对于在 Amazon Web Service 标准区域中使用 Amazon Cognito User Pool 认证方式快速部署 Data Transfer Hub 解决方案。您可以使用以下链接快速启动一个 CloudFormation 堆栈来部署和管理整个方案：

2. 默认情况下，该模板将在 您登录控制台后后默认的区域启动。 若需在指定的 Amazon Web Service 区域中启动 该解决方案，请使用控制台导航栏中的区域选择器。
3. 在创建堆栈页面上，确认 Amazon S3 URL 文本框中显示正确的模板 URL，然后选择 【下一步】。



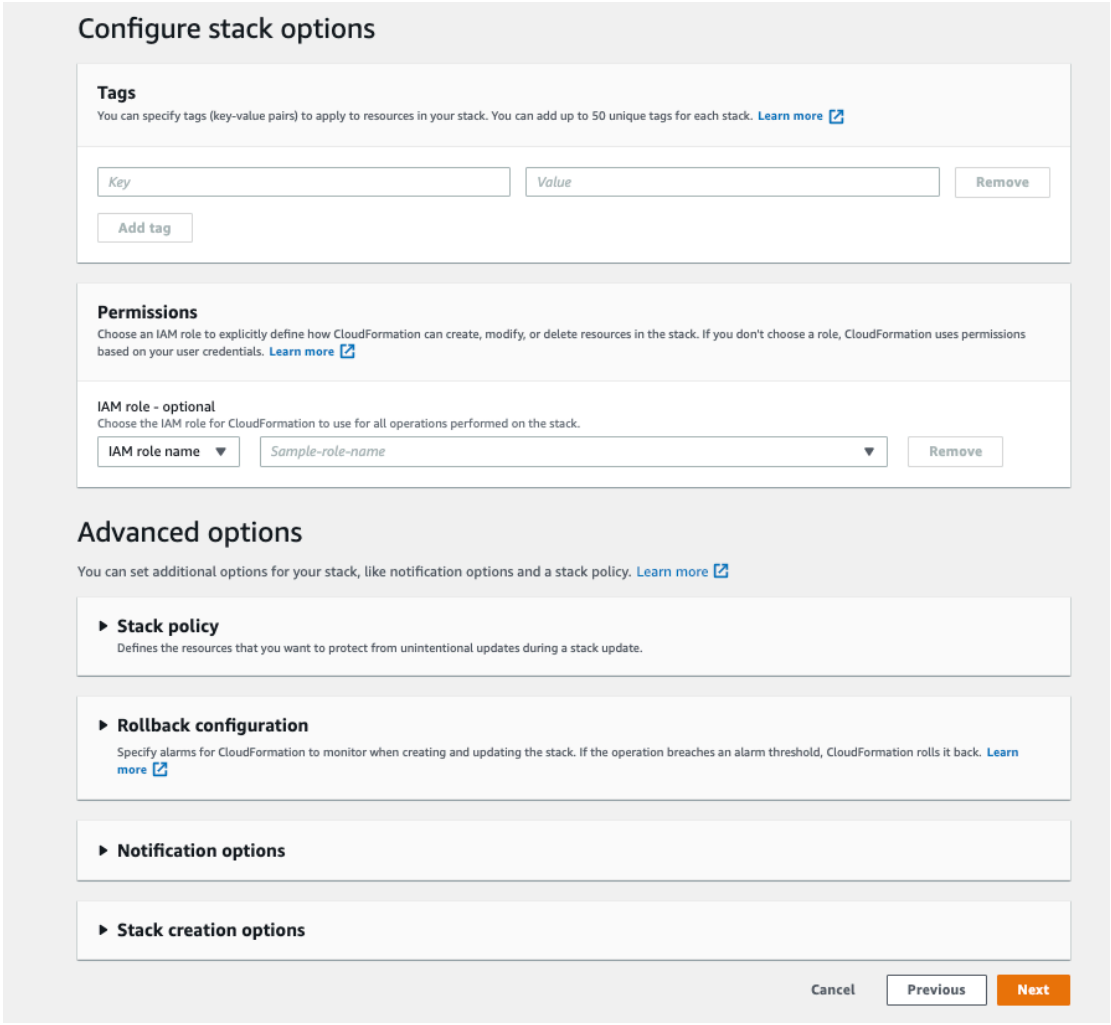
4. 在指定堆栈详细信息下，填写如下的参数。



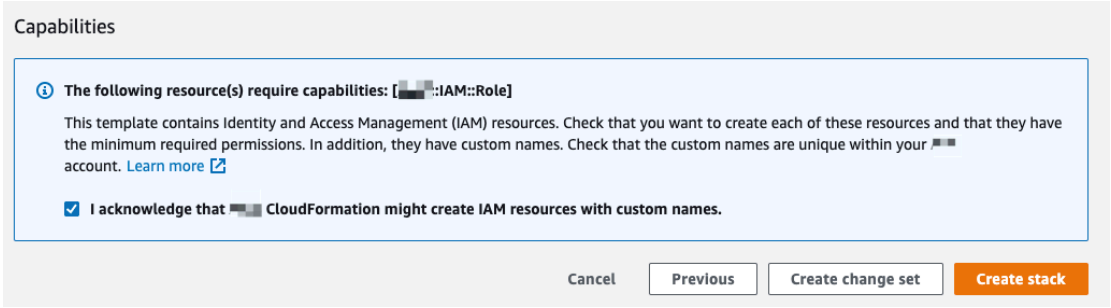
参数名称	值	用途
Admin Email	填写管理员 Email 地址	管理员邮箱，接收应用的临时密码

5. 参数填写完成后选择【下一步】。

6. 在配置堆栈选项页面上，保持默认值，点击 【下一步】。



7. 在审核页面上，查看并确认设置。 确保选中确认模板将创建 Amazon Identity and Access Management（IAM）资源的框。

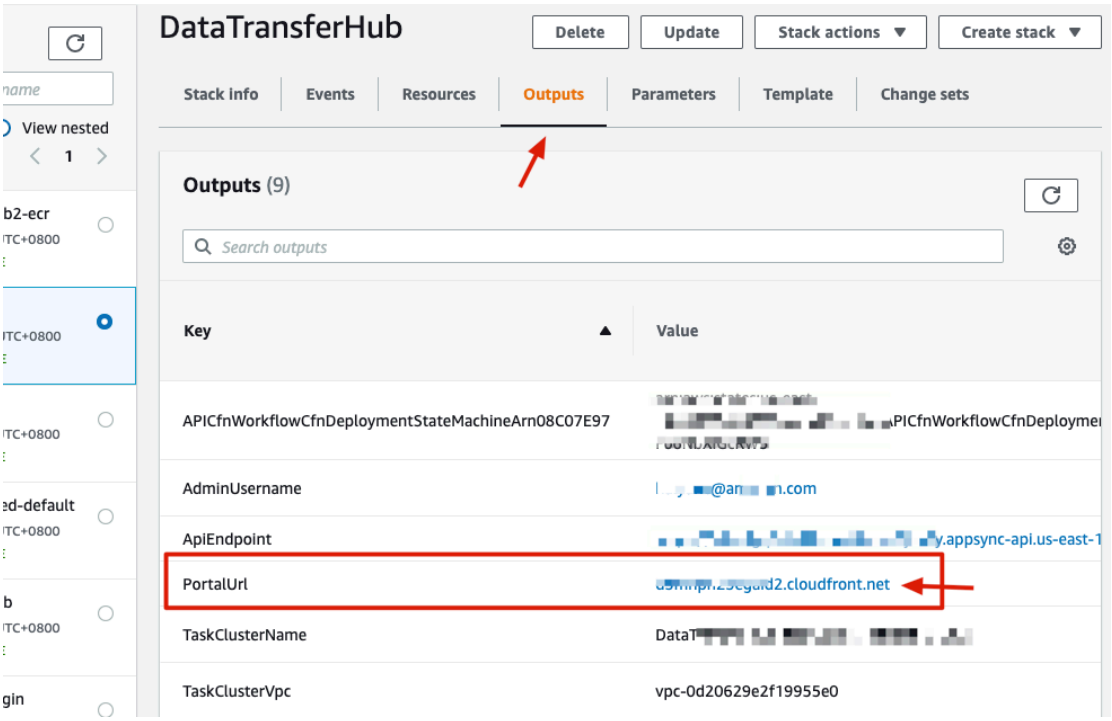


8. 选择【创建堆栈】以部署堆栈。

   您可以在 Amazon CloudFormation 控制台的状态列中查看堆栈的状态。 您应该在大约 15 分钟内看到状态为 CREATE_COMPLETE。

   堆栈创建成功后，可在 Amazon CloudFormation 中输出标签页中看到：

• PortalUrl – 这个 URL 就是最终应用部署的 Portal 的 URL 地址。



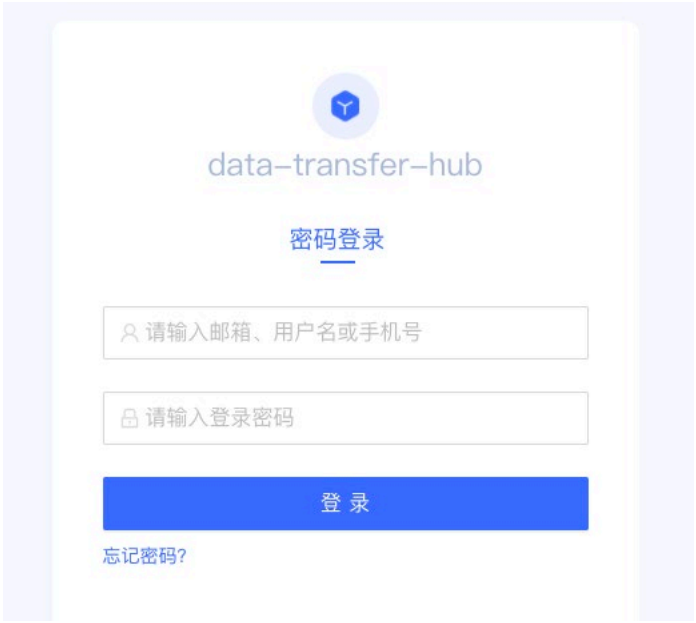
到此为止，Data Transfer Hub 已经成功部署在了 Amazon Web Service 上。

用户使用手册

登录系统

OpenId 认证方式

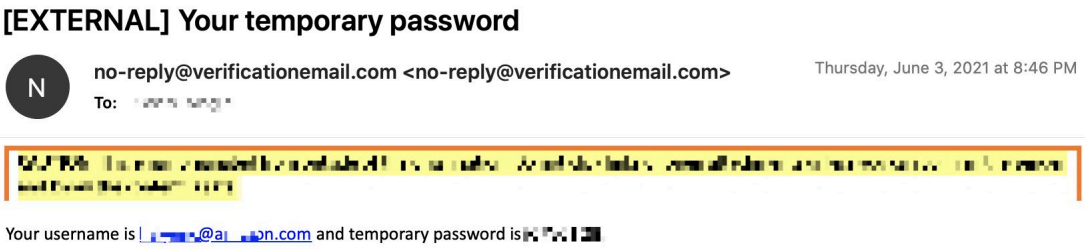
这里以 Authing 为例，在浏览器的地址栏输入 https://您的 Data Transfer Hub 域名。如果是首次登录系统，系统将会跳转到 Authing.cn 提供的登录界面，输入您在部署解决方案时创建的用户名和密码，点击【登录】，系统会跳转到 Data Transfer Hub 的首页。



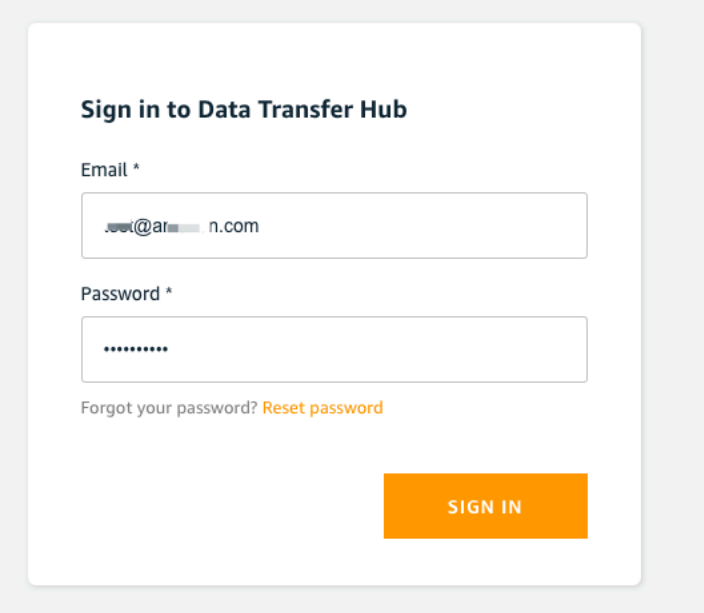
Cognito User Pool 认证方式

1. 在浏览器的地址栏输入 CloudFormation 输出中的 PortalUrl。系统会跳转到 Cognito 的登录界面，Email 输入管理员邮箱；在您部署成功后会收到一封 Your temporary password 的邮件，里面包含临时登录密码。

2. 邮件模版标题

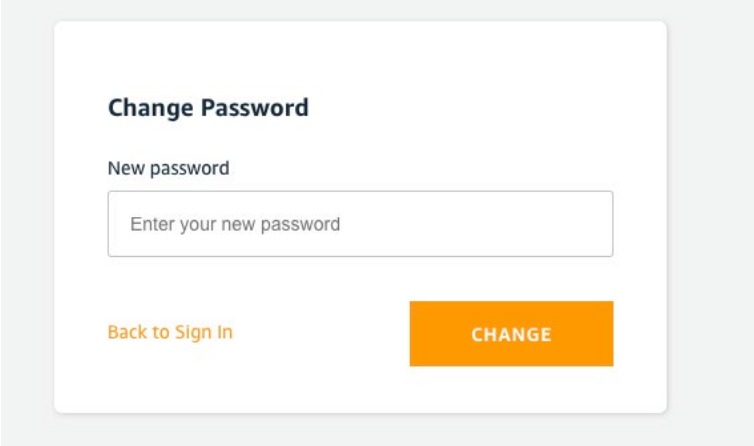


3. 登录界面，输入邮箱和密码

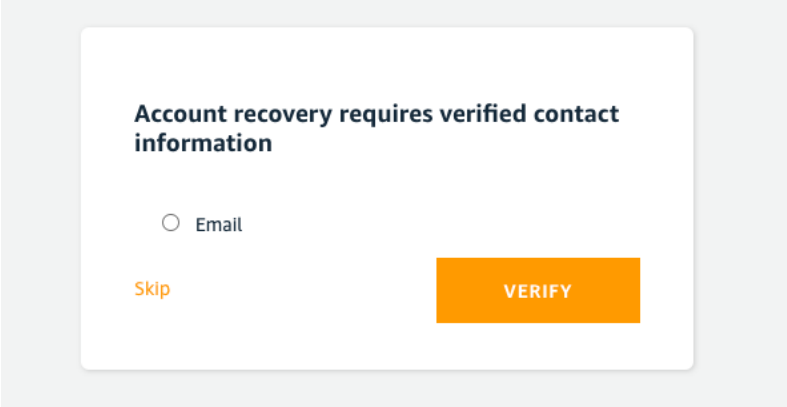


4. 点击【Sign In】

5. 首次登录时，系统强制要求您修改您的临时密码



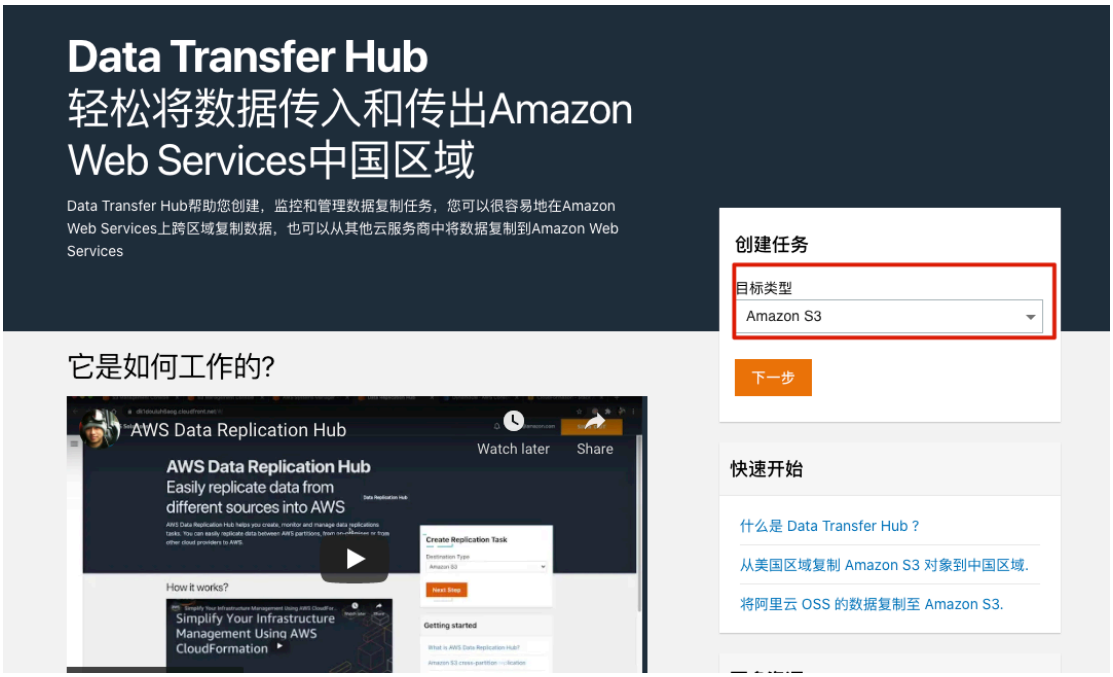
6. 验证您的邮箱地址



7. 验证完成后，系统将跳转到 Data Transfer Hub 的首页。

创建任务

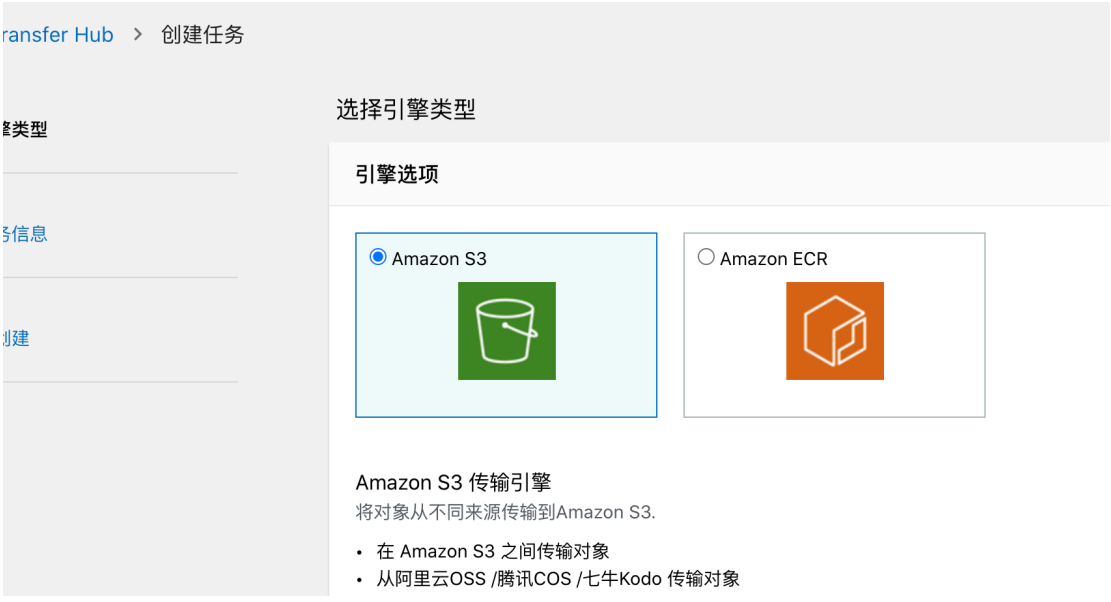
在用户登陆系统之后，就可以使用 Portal 来创建**传输**任务了。



在选择您所需要创建的目标类型，点击【下一步】

创建 Amazon S3 传输任务

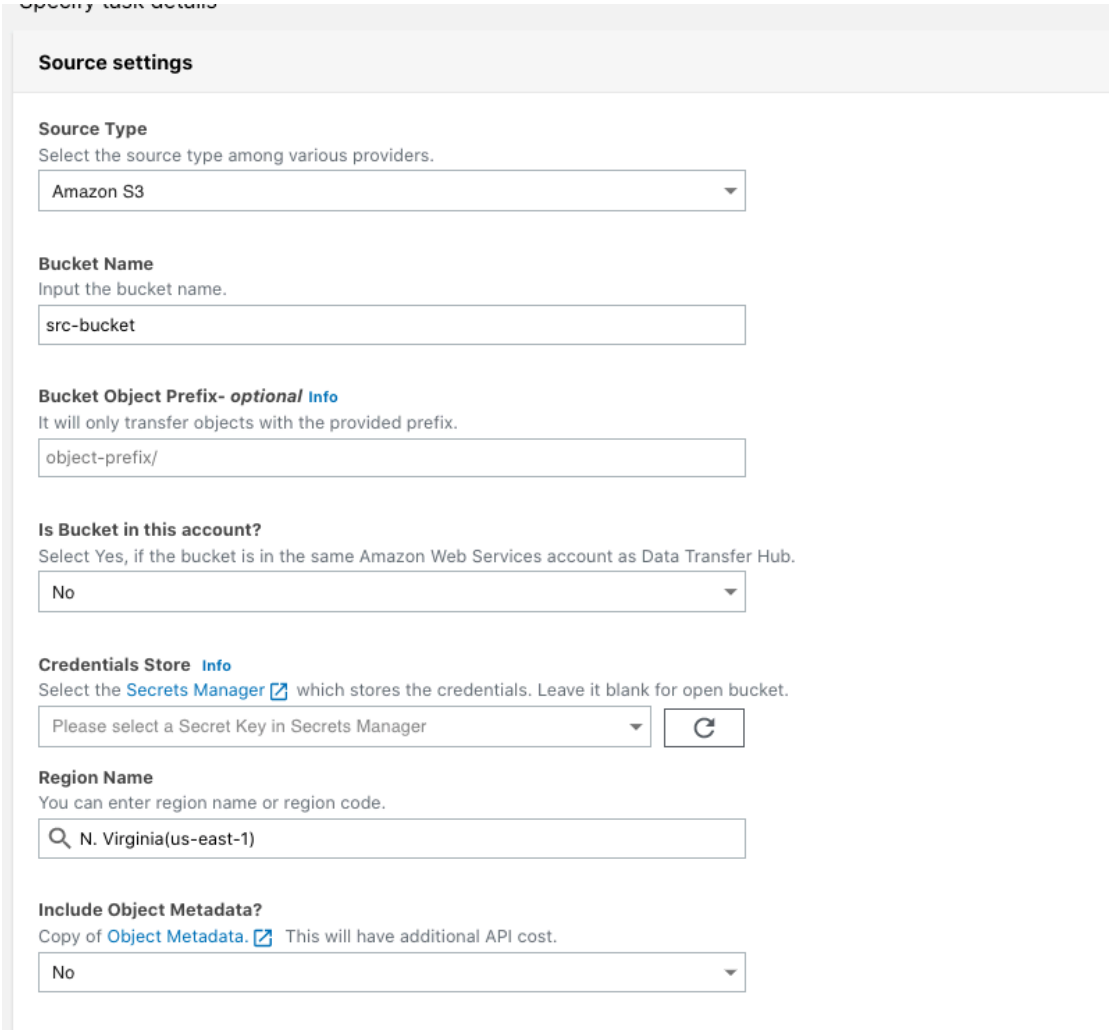
1. 选择引擎为 Amazon S3。点击【下一步】



2. 在填写**传输**任务信息

1）在数据源类型选项中选择所需**传输**的数据来源，下面以 Amazon S3 作为例子

2）在数据源**设**置和目**标**桶**设**置的表**单**中填写您需要**传输**的桶的**对应**的名称，前**缀**以及**选择**数据桶**对应**的区域信息。

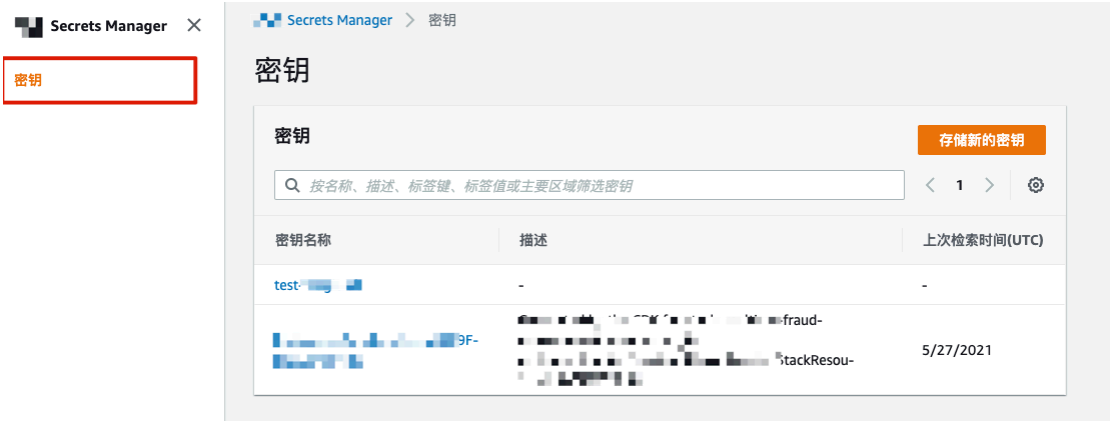


3）如果数据源桶是与 Data Transfer Hub 部署的账户一致，请选择 Yes(如果源数据桶与 Data Transfer Hub 部署的账户一致，则可以启用 S3 Event 触发数据传输)。

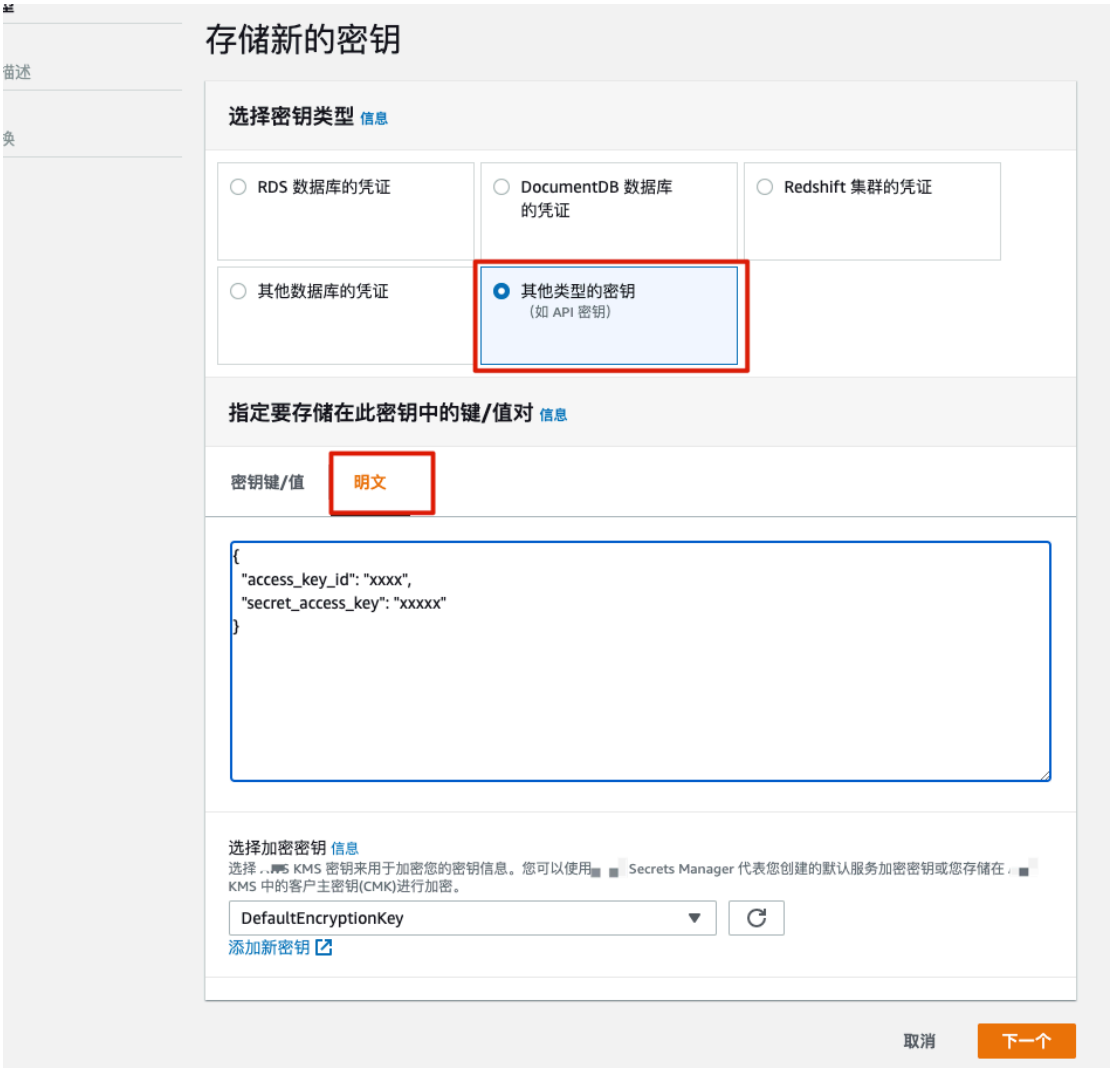
这里以二者不在同一个账户为例。启用传输数据任务需要选择一个用户凭据，接下来的步骤将指导如何添加凭据。

4）在 Secrets Manager 中添加凭据。

在控制台中打开 Secrets Manager，在左边菜单中点击【密钥】



点击【创建新的密钥】,选择密钥类型为“其他类型的密钥”。



在指定要存储在此密钥的键/值对的“明文“输入框中按照图中格式填写对应的 AK/SK 信息。

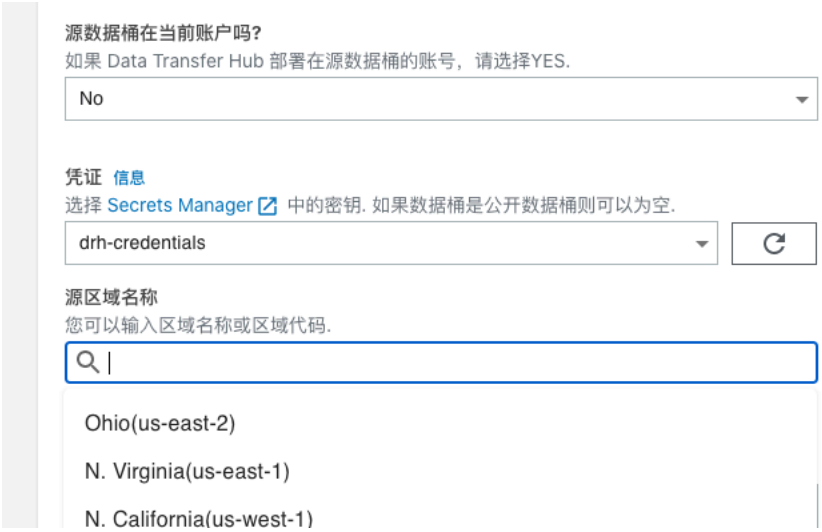
5）点击【下一个】,填写密钥名称和描述(可选)

6）点击【下一个】，在配置自动轮换中，选择“禁用自动轮换”。

7) 点击【下一个】，保持默认值后点击【存储】，完成密钥的创建。

8) 回到 Data Transfer Hub 创建任务的界面中，点击凭证选择框右边的刷新按钮，下拉选项中会显示刚才创建好的凭证(密钥)。选择该凭证(密钥)。

8）在区域选择器中，选择源数据桶所在的区域。



9）在目标桶设置中，填写目标桶的名称，选择目标桶区域。

10) 在高级选项中，可以填写目标桶的前缀(可选)，需要存储的 S3 的存储类型以及目标对象的 ACL。

11）在引擎**设置选项**中，可以根据实际情况可以调整 Graviton 2 的相关参数

12) 在引擎**设**置的高**级选项**中，可以修改**查**找器及**线**程的相关参数，如果没有特别需要，保持默**认值**即可，建**议**在 Amazon Web Services Solutions Architect 的指**导**下**调**整此**设**置。

13）在更多选项中，根据实际情况填写描述（可选），同时填写通知邮箱（如果**传输**任务中**传输**的过程中出现错误情况，则会发邮件通知到这个邮箱地址）。

14）点击【下一步】

3. 在预览任务信息

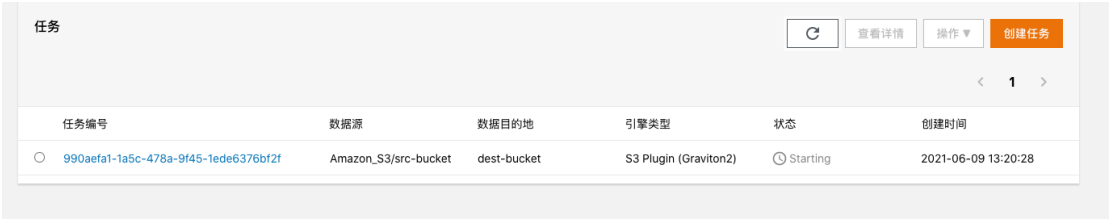
步骤二: 任务详细信息

任务参数 (23)

参数	值
数据源类型	Amazon_S3
源数据端点URL	
源数据桶名称	src-bucket
源数据桶对象前缀	
启用S3事件	No
源数据区域名称	cn-northwest-1
源数据桶是否在当前账户	false
源数据凭证中的密钥名称名称	drh-credentials
目标数据桶名称	dest-bucket
目标数据桶对象前缀	
目标S3存储类型	STANDARD
目标区域名称	us-east-1
目标数据桶是否在当前账户	true
目标区域的密钥名称	
是否包含元数据	false
目标对象ACL	bucket-owner-full-control
最大容量	20
最小容量	1
所需容量	1
查找器深度	0
查找器数量	1
工作线程数	4
通知邮箱	[redacted]

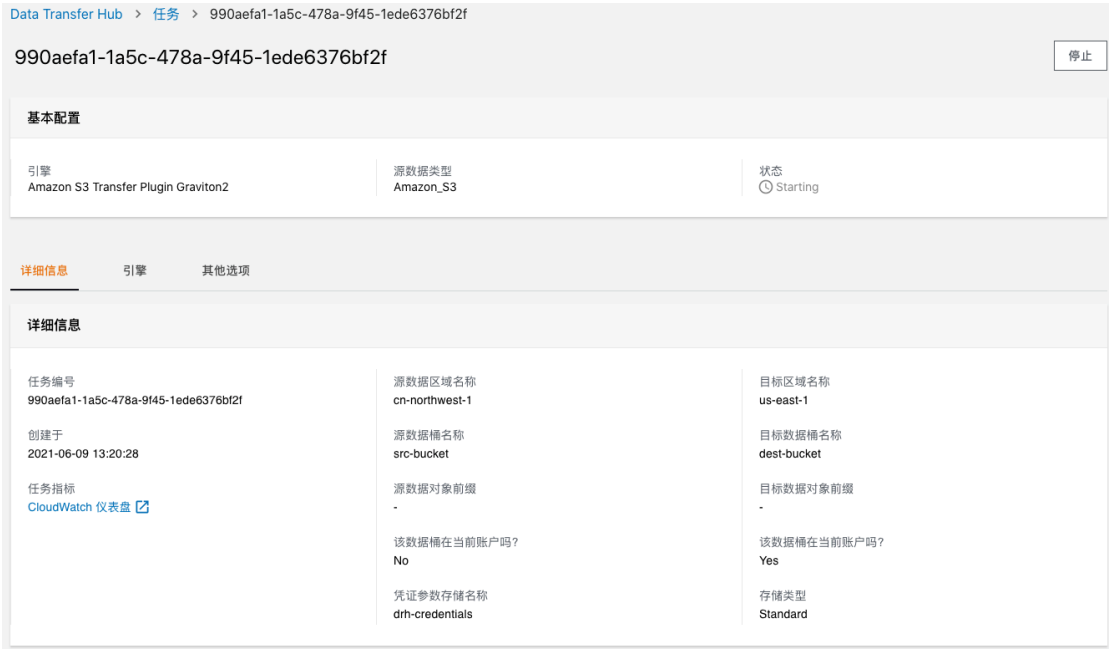
4. 点击【创建任务】

任务列表中会新增一条刚才创建的任务记录，至此，Amazon S3 的**传输**任务创建完成。



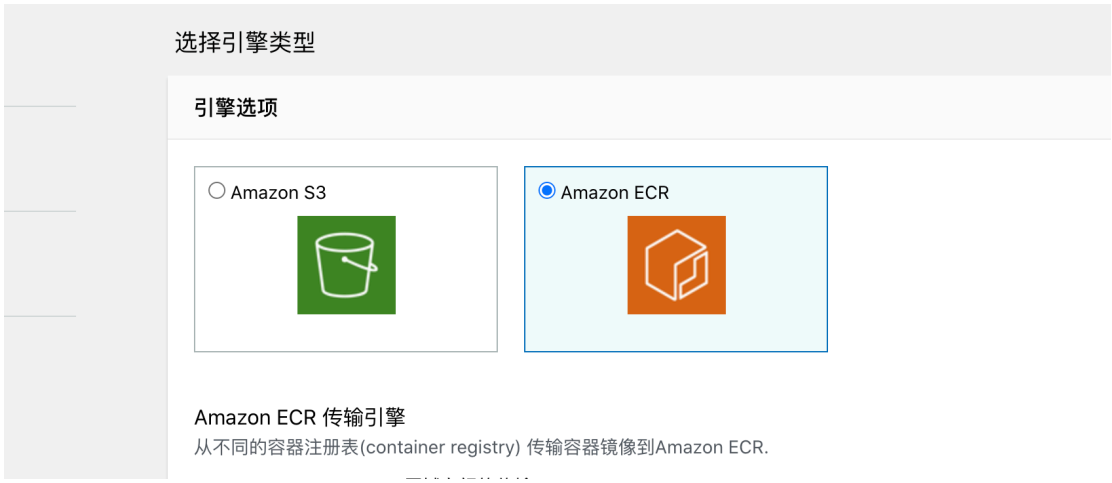
5. 查看任务详情

用户可以点击任务编号或者选择任务点击【查看详情】来查看任务详细信息。



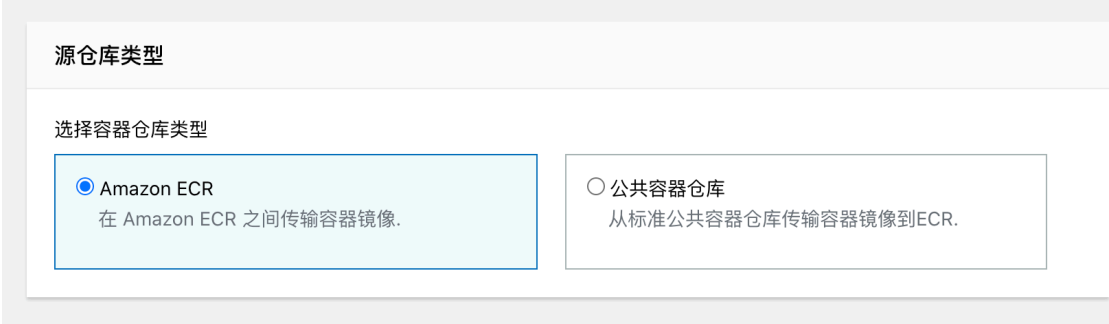
创建 Amazon ECR 传输任务

1. 选择引擎为 Amazon ECR。点击【下一步】



2. 填写任务信息

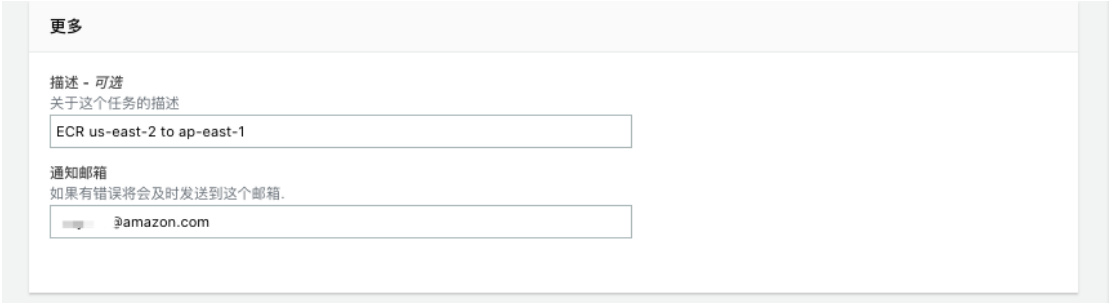
1）在数据源类型选项中选择所需**传输**的数据来源，下面以**传输** ECR 作为例子



2）在源仓库设置中填写需要**传输**的源地址信息，以及需要**传输**的容器**镜**像等信息。

3）在目标仓库设置中，填写需要**传输**的目标 Amazon ECR 的信息。

4）在更多选项中，根据实际情况填写描述，同时填写通知邮箱（如果传输任务中传输的过程中出现错误情况，则会发邮件通知到这个邮箱地址）。



5）点击【下一步】

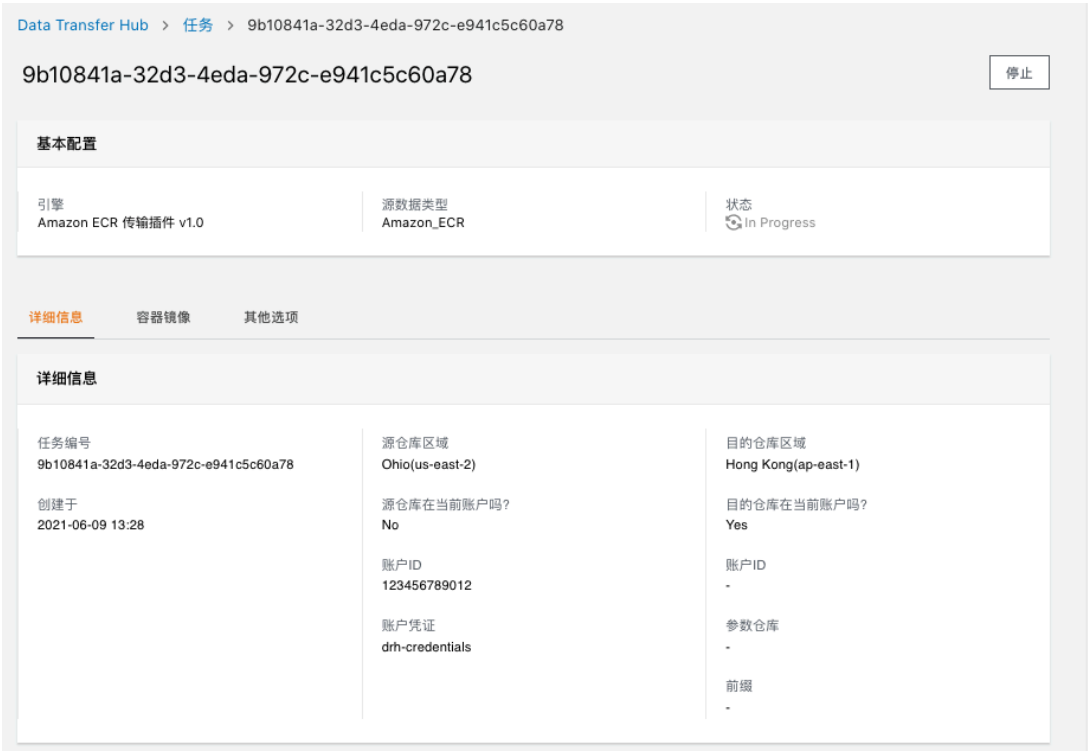
3. 在预览任务信息

4. 点击【创建任务】

任务列表中会新增一条刚才创建的任务记录，至此，Amazon ECR 的传输任务创建完成。

5. 查看任务详情

用户可以点击任务编号或者选择任务然后点击【查看详情】来查看任务的详细信息。



小结

综上所述，使用 Data Transfer Hub 解决方案，结合对用户友好的操作界面，可以方便快捷高效地创建您所需的传输任务，减少您对于复杂数据传输系统的构建和维护工作。