Minimal AWS Cross Account Policy for Data Lake and Data Hub (Preview)

Date published: 2021-05-27
Date modified: 2021-08-06

Legal Notice

Contents

Revision history

2021-08-06 Updated the requirements for customizing the scope of iam:PassRole
2021-06-03 Updated link to the minimal policy
2021-06-02 Updated link to the minimal policy
2021-05-27 Initial draft

About this feature

If you are planning to only run Data Hub in your CDP environment, you don't need all the actions in the default cross account role provided on the create credential UI and in the provisioning credential documentation (See IAM policy definition). Using a significantly reduced cross account policy is possible as an alternative.

Note: This is a preview feature and the cross-account policy is subject to change without notification. Please check the github pages periodically to check for updates.

Reduced access policy definition

The IAM role used for the provisioning credential should use the updated reduced access policy instead of the policy provided in the UI and documentation. You should:

1. Copy this policy.
2. Update it as described in Customize the scope of iam:PassRole below.
3. You can optionally reduce the permissions further, as described in Further permissions reduction below.

Customize the scope of iam:PassRole

CDP utilizes the *iam:PassRole* as part of the cloud identity federation setup process via IDBroker on AWS. The default policy uses "*" as the resource scope for this action:

```
{
      "Effect": "Allow
       "Action": [
         "iam:PassRole"
       ],
       "Resource": "*"
}
```

However the scope can be reduced to just a few roles. To do this, you need to edit the following block and replace the "*" with the ARNs of the following roles used for the environment:

- IDBROKER_ROLE
- LOG_ROLE
- RANGER_AUDIT_ROLE

- DATALAKE_ADMIN_ROLE

You chose the name for these roles when you set up CDP as part of the Minimal Setup for Cloud Storage or AWS Quick Start.

After updating according to the Minimal Setup for Cloud Storage, the PassRole policy excerpt should like similar to the following:

```
{
      "Effect": "Allow
      "Action": [
        "iam:PassRole"
      ],
      "Resource": [
               "arn:aws:iam::$ACCOUNTID:role/IDBROKER_ROLE",
               "arn:aws:iam::$ACCOUNTID:role/LOG_ROLE",
               "arn:aws:iam::$ACCOUNTID:role/RANGER_AUDIT_ROLE",
               "arn:aws:iam::$ACCOUNTID:role/DATALAKE_ADMIN_ROLE"
          ]
}
```

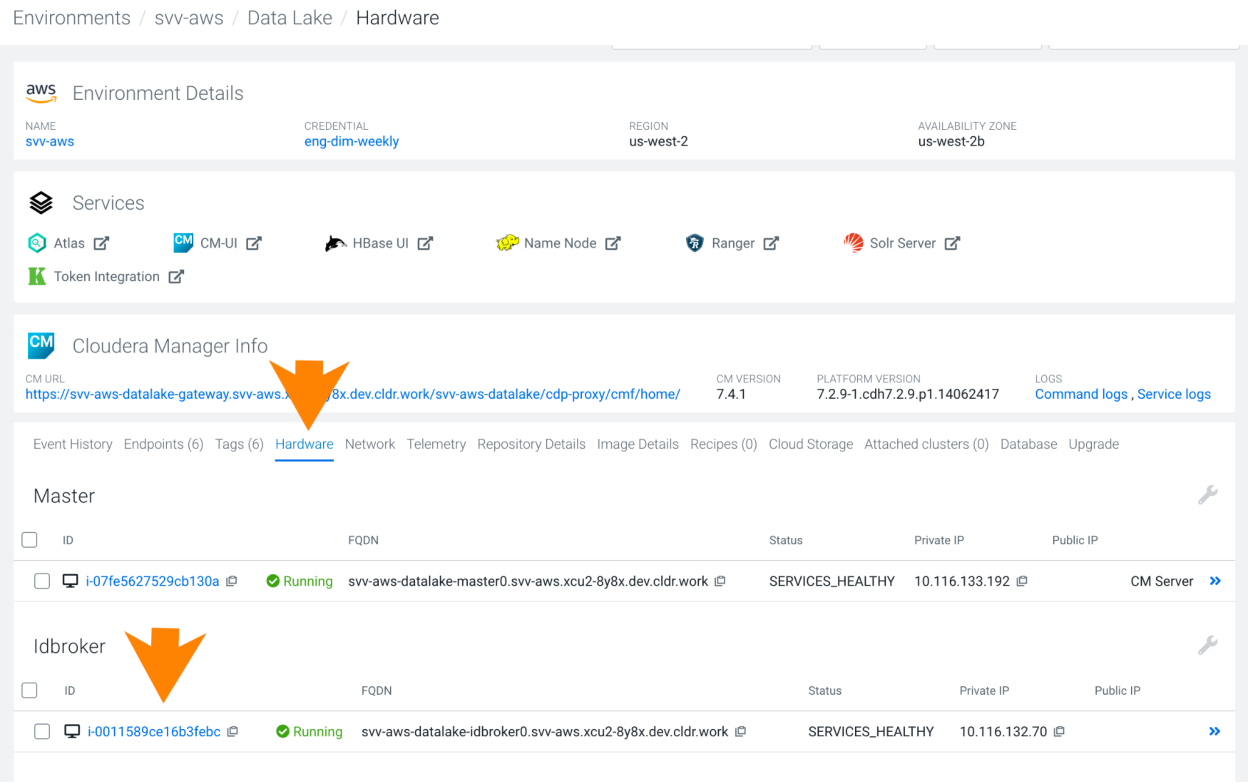
If you used the AWS Quick Start, the PassRole policy excerpt should look like this:

```
{
      "Effect": "Allow
      "Action": [
        "iam:PassRole"
      ],
      "Resource": [
               "arn:aws:iam::$ACCOUNTID:role/<prefix>-idbroker-role",
               "arn:aws:iam::$ACCOUNTID:role/<prefix>-log-role",
               "arn:aws:iam::$ACCOUNTID:role/<prefix>-ranger-audit-role",
               "arn:aws:iam::$ACCOUNTID:role/<prefix>-datalake-admin-role"
          ]
}
```

If you have already created these roles and you don't know the name of these roles, you can obtain them using the following steps:

IDBROKER_ROLE

1. Log in to CDP web interface and navigate to the **Management Console** service > **Data Lakes**.
2. Scroll down to where you see **Event History** and select the **Hardware** tab.
3. Locate the EC2 instance for the **idbroker** node (if you are running Medium Duty Data Lake, you will see two IDBroker nodes, you can use either) :



4. Click on the instance id. You will be redirected to the EC2 console.
5. In the **Instance Details** section for this node, you can see the name of the IAM role being used by this instance. Click on this role to get to the details.
6. Copy the ARN of the instance profile. It follows the following naming convention: `arn:aws:iam::<12-digit-AWS-account-id>:instance-profile/<name-of-idbroker-role>`

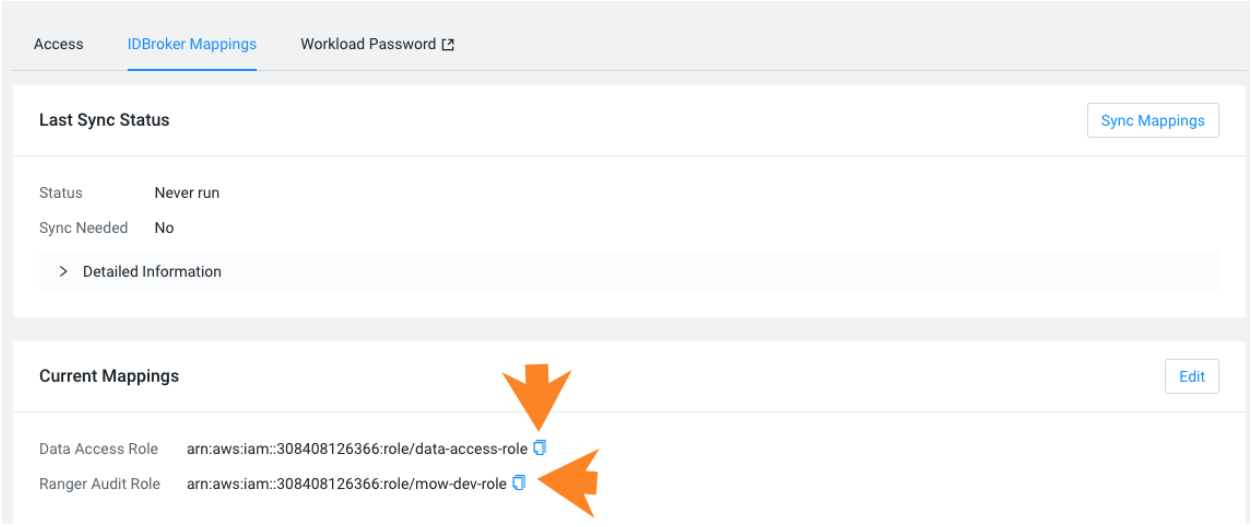
LOG_ROLE

1. Navigate to **Environments** .
2. Click on the existing environment
3. Click on the **Summary** tab.
4. Scroll to **Logs Storage and Audits** and you will see the **Instance Profile** entry.
5. Copy the ARN of the **Instance Profile**:

RANGER_AUDIT_ROLE and DATALAKE_ADMIN_ROLE

1. Navigate to **Environments**.
2. Click on the existing environment.
3. From the **Actions** menu, select **Manage Access**.

4. Click on the **IDBroker Mappings** tab.
5. Under **Current Mappings** you will see the mappings for the **Data Access Role** and **Ranger Audit Role**.
6. Copy the ARNs of the **Data Access Role** and **Ranger Audit Role**.



Further permissions reduction

You can reduce these policies even further by pre-creating some resources:

- If you set up your own VPC and subnets (using this guide), then the following actions can be removed from the policy:
  - `ec2:CreateVpc`
  - `ec2:CreateNatGateway`
  - `ec2:CreateRouteTable`
  - `ec2:CreateSubnet`
  - `ec2:CreateVpcEndpoint`
  - `ec2:CreateInternetGateway`
  - `ec2:DeleteSubnet`
  - `ec2:DeleteInternetGateway`
  - `ec2:AttachInternetGateway`
  - `ec2:DetachInternetGateway`
  - `ec2:DescribePrefixLists`
  - `ec2:AllocateAddress`
  - `ec2:AssociateRouteTable`
  - `ec2:CreateRoute`
  - `ec2:DeleteRouteTable`
  - `ec2:DeleteVpcEndpoints`
  - `ec2:DisassociateRouteTable`

    - `ec2:ReleaseAddress`
    - `ec2:DeleteRoute`
    - `ec2:DeleteNatGateway`
    - `ec2:DeleteVpc`
- If you use your own security groups (using this guide), then the following actions can be removed from the policy:
    - `ec2:CreateSecurityGroup`
    - `ec2:AuthorizeSecurityGroupIngress`
    - `ec2:RevokeSecurityGroupEgress`
    - `ec2:AuthorizeSecurityGroupEgress`
- If you use private IPs or set up CCM to communicate with the CDP Control Plane (as described here), then the following actions can be removed from the policy:
    - `ec2:AllocateAddress`
    - `ec2:ReleaseAddress`
- If you create your own DynamoDB table for S3Guard (described here), then the following actions can be removed from the policy:
    - `dynamodb:DeleteTable`