Cloudera Public Cloud

AWS Onboarding Quickstart

Date published: 2019-08-22
Date modified:

https://docs.cloudera.com/

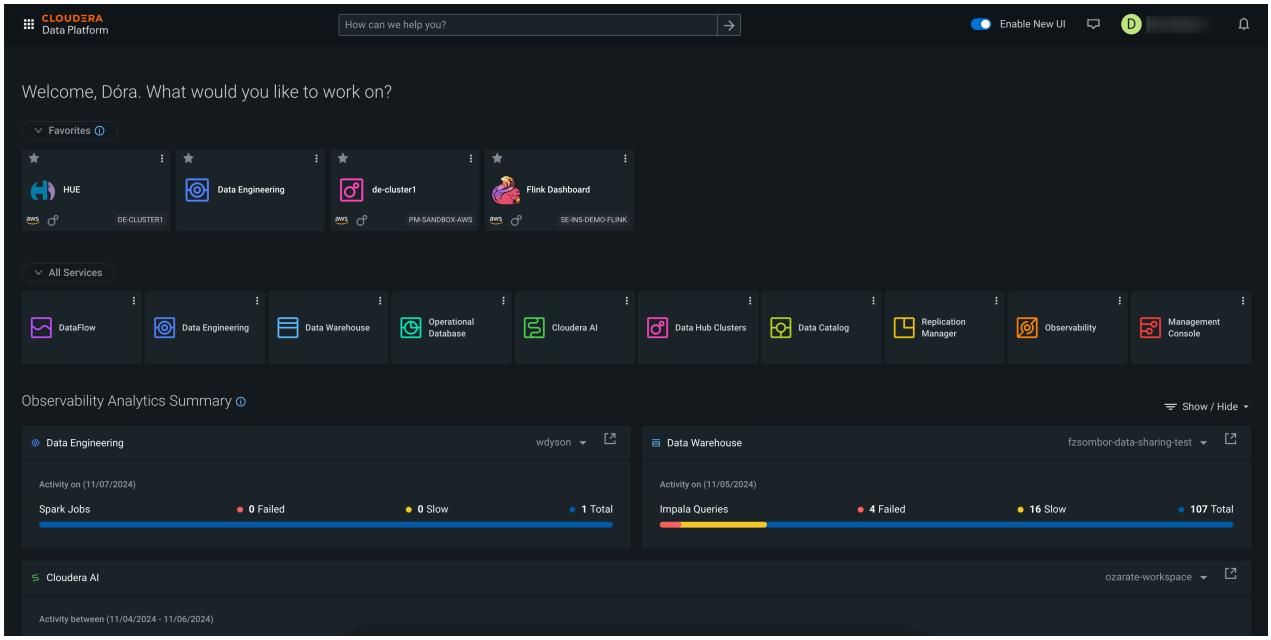
Legal Notice

Contents

AWS quickstart (Deprecated)

If you've reached the Cloudera landing page for the first time, you've come to the right place! In this quickstart, we'll show you step-by-step how to connect Cloudera to your AWS account, so that you can begin to provision clusters and workloads.

Warning: This quickstart has been deprecated and is no longer being maintained. For quickly setting up Cloudera on AWS, refer to Deploy Cloudera using Terraform.



To complete this quickstart, you'll need access to two things:

- The Cloudera console pictured above
- The AWS console

Note: This AWS onboarding quickstart is intended for simple Cloudera evaluation deployments only. It may not work for scenarios where AWS resources such as VPC, security group, storage accounts, and so on, are pre-created or AWS accounts have restrictions in place.

The steps that we will perform are:

Step 0: Verify the AWS prerequisites

Step 1: Create a provisioning credential

Step 2: Register an AWS environment in Cloudera

Verify AWS cloud platform prerequisites

Before getting started with the AWS onboarding quickstart, review and acknowledge the following:

- This AWS onboarding quickstart is intended for simple Cloudera evaluation deployments only. It may not work for scenarios where AWS resources such as VPC, security group, storage accounts, and so on, are pre-created or AWS accounts have restrictions in place.

- Users running the AWS onboarding quickstart should have:
  - AWS Administrator permissions on the AWS account that you would like to use for Cloudera.
  - Rights to create AWS resources required by Cloudera. See list of AWS resources used by Cloudera.
  - Cloudera Admin role or Power User role in Cloudera subscription.
- This AWS onboarding quickstart uses a CloudFormation template that automatically creates the required resources such as buckets, IAM roles and policies, and so on.
- Cloudera on cloud relies on several AWS services that should be available and enabled in your region of choice. Verify if you have enough quota for each AWS service to set up Cloudera in your AWS account. See list of AWS resources used by Cloudera.

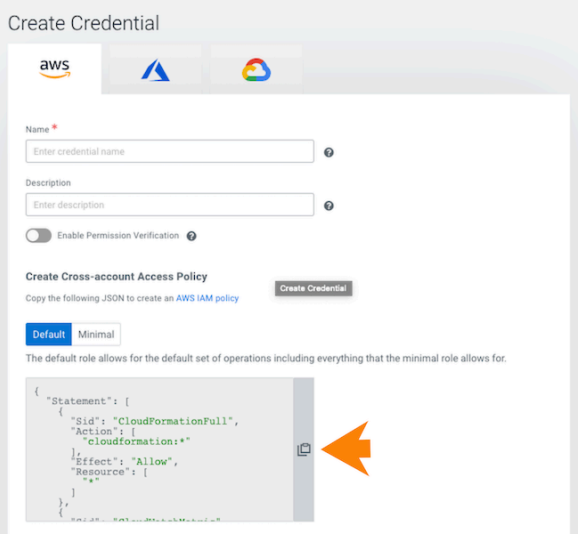
If you have more complex requirements than those listed here, contact Cloudera Sales Team to help you with the Cloudera onboarding.

Create a Cloudera credential

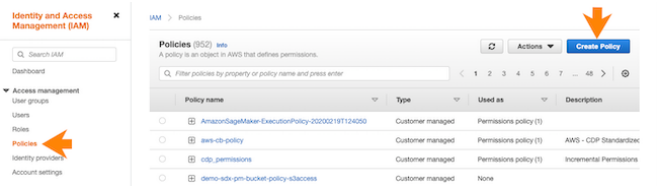
In the Cloudera console, the first step is to create a Cloudera credential. The Cloudera credential is the mechanism that allows Cloudera to create resources inside of your cloud account.

Procedure

1. Log in to the Cloudera web interface.
2. From the Cloudera home screen, click the Cloudera Management Console icon.
3. In the Cloudera Management Console, select Shared Resources > Credentials from the navigation pane.
4. Click Create Credential.
5. Click the Copy icon to the right of the **Create Cross-account Access Policy** text box.



6. In a second browser tab, open the **AWS Console** and navigate to Identity and Access ManagementPolicies. Click Create Policy.

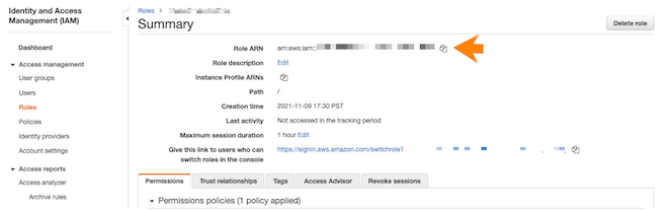


7. Click on the JSON tab and paste the access policy in the text box.

   You may get a warning related to using wildcards. You may ignore it and proceed to the next step.

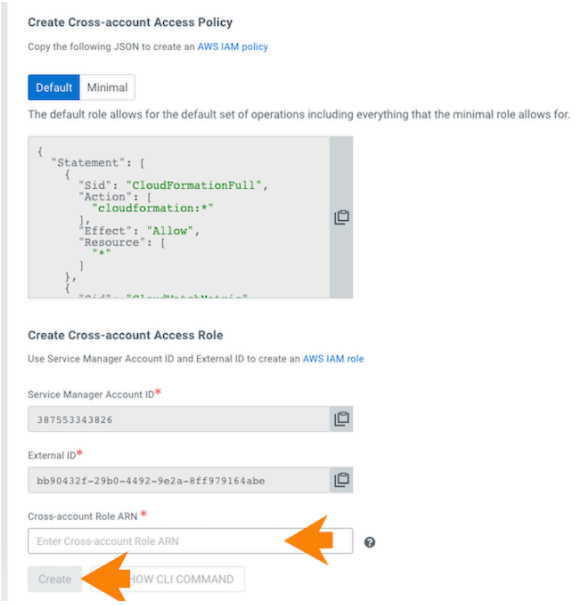
8. Click Next: Tags.
9. Click Review Policy.
10. Give the policy a unique name and a description.
11. Click Create Policy.

    Next, you create the required cross-account role.
12. In the AWS console, navigate back to Identity and Access Management.
13. Click RolesCreate Role.
14. Under **Select type of trusted entity**, select AWS Account > Another AWS account.
15. Return to the Cloudera Management Console and copy the contents of the Service Manager Account ID field on the **Credentials** page.
16. In the AWS console, paste the Service Manager Account ID into the Account ID field.
17. Return to the Cloudera Management Console and copy the contents of the External ID field on the **Credentials** page.
18. In the AWS console, check the "Require external ID" options box, and then paste the External ID copied from Cloudera into the External ID field.
19. Click Next: Permissions.
20. Under Permissions, select the checkbox next to the name of the policy that you created in Step 8.
21. Click Next: Tags.
22. Click Next: Review.
23. Give the role a unique name and description, then click Create Role.
24. Still in the role page of the AWS console, search for the role you just created, and click on it.
25. Copy the Role ARN at the top of the **Summary** page.



26. Return to the **Credentials** page in the Cloudera Management Console.
27. Give the Cloudera credential a name and description. The name can be any valid name.

28. Paste the Role ARN that you copied from the AWS console into the Cross-account Role ARN field, then click Create.



Now that you've created a cross-account role, proceed to creating a Cloudera environment.

Register a Cloudera environment

Before you register an environment, you'll want to create specific IAM roles and policies so that Cloudera can operate in a secure manner.

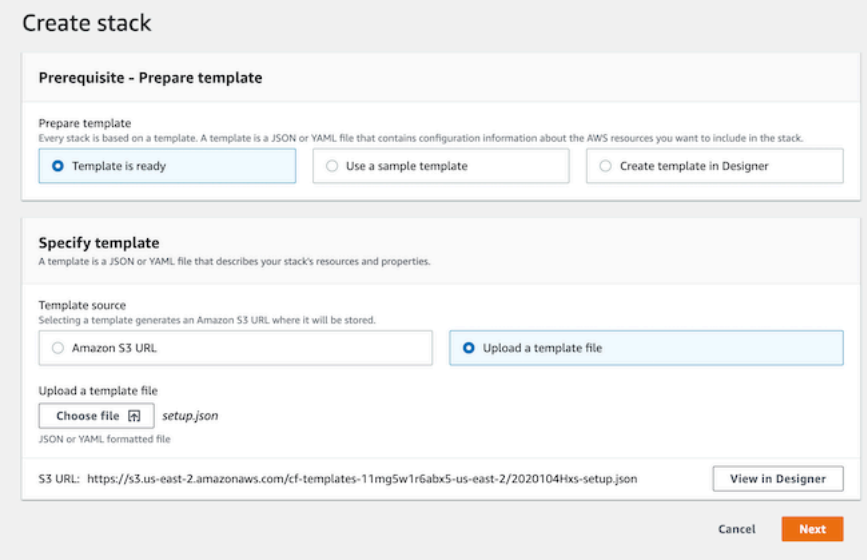
About this task

For background information, a description of what we're building and why can found here. For this quickstart, we'll use CloudFormation to set all of this up for you.

Procedure

1. Download the CloudFormation provided template here.

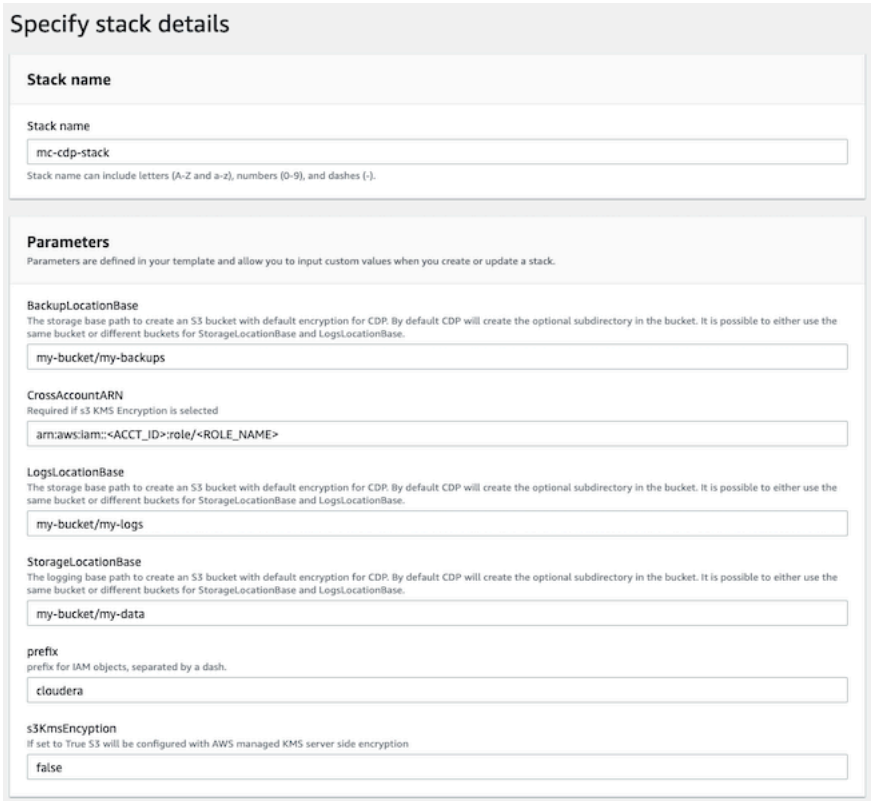
2. In the AWS console, deploy the CloudFormation template:
   a) In **AWS Services**, search for CloudFormation.
   b) Click Create Stack and select With new resources (standard).
   c) Select Template is ready and then Upload a template file.



   d) Click Choose file and select the CloudFormation template that you downloaded.
   e) Click Next.
   f) Under Stack name, enter a stack name. The name can be any valid name.
   g) Under **Parameters**, complete the following fields:

   - BackupLocationBase: Choose an unused bucket name and path for the FreeIPA backups. Cloudera will be creating the bucket for you. The same bucket can be used for BackupLocationBase, LogsLocationBase, and StorageLocationBase. By default this is set to my-bucket/my-backups.
   - CrossAccountARN: Do not change the default value. This parameter is only required when encryption is enabled, but since in this quickstart we do not enable encryption, you should leave this value as is.
   - LogsLocationBase: Choose an unused bucket name and path for the logs. Cloudera will be creating the bucket for you. The same bucket can be used for BackupLocationBase, LogsLocationBase, and StorageLocationBase. By default this is set to my-bucket/my-logs.
   - StorageLocationBase: Choose an unused bucket name and path for the data. Cloudera will be creating the bucket for you. The same bucket can be used for BackupLocationBase, LogsLocationBase, and StorageLocationBase. By default this is set to my-bucket/my-data.
   - Prefix: A short prefix of your choosing, which will be added to the names of the IAM resources Cloudera will be creating. We chose "cloudera" as an example.
   - s3KmsEncryption: Encryption will be disabled for the created bucket. You don't need to change this value.

   For example:

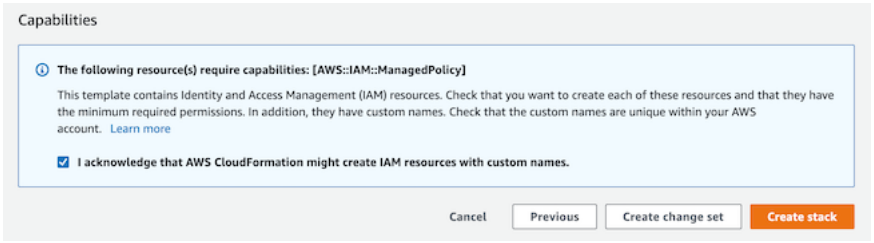


Make a note of the BackupLocationBase, LogsLocationBase, StorageLocationBase, and Prefix that you define. You will need them later.

h) Click Next.

i) At the **Configure Stack Options** page, click Next.

j) At the bottom of the **Review** page, under Capabilities, click the checkbox next to I acknowledge that AWS Cloudformation might create IAM resources with custom names, as that is exactly what we will be doing.

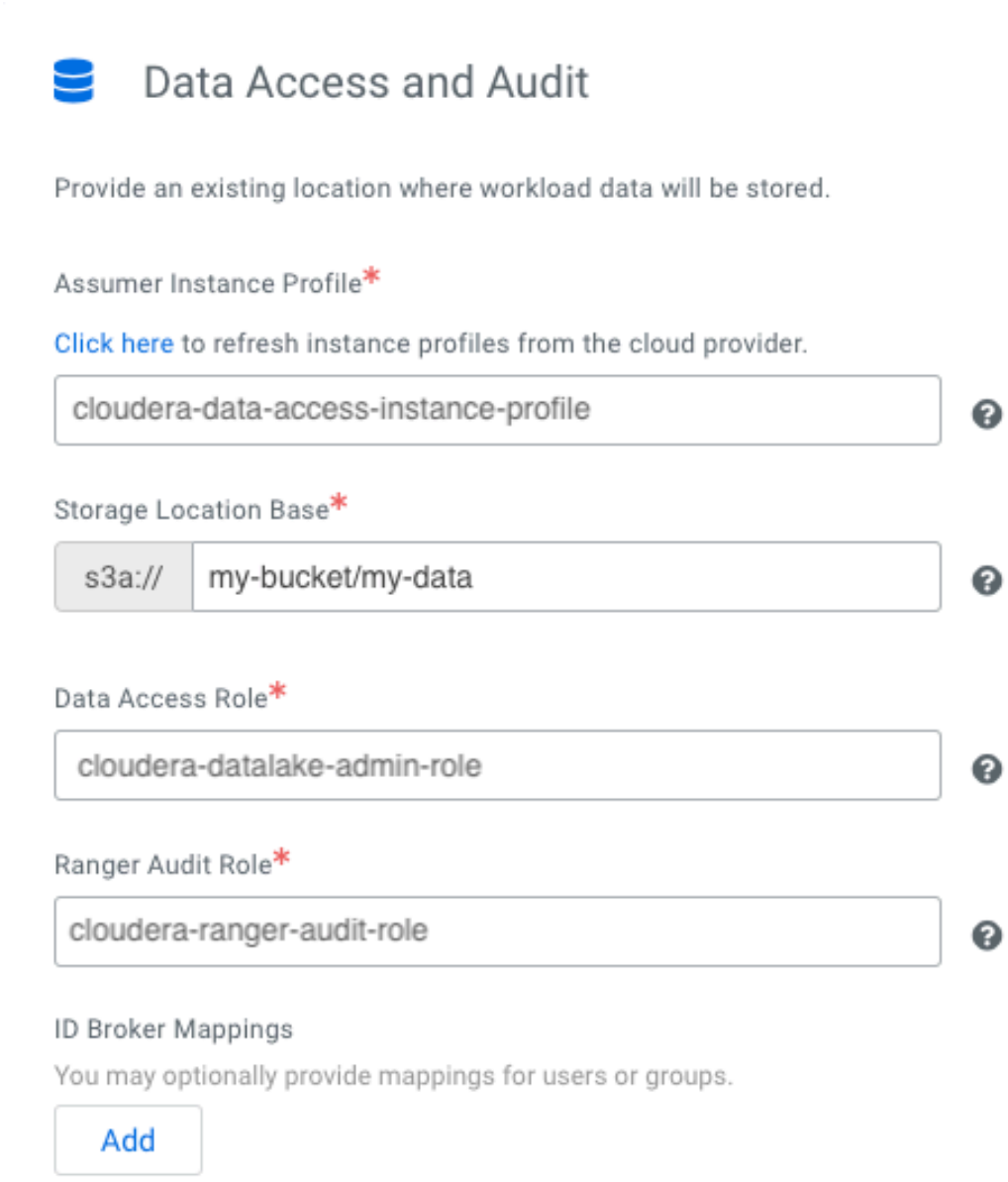


k) Click Submit.

3. Still in the AWS console, create an SSH key in the region of your choice. If there is already an SSH key in your preferred region that you'd like to use, you can skip these steps.
   a) In **AWS Services**, search for EC2.
   b) In the top right corner, verify that you are in your preferred region.
   c) On the left hand navigation bar, choose Key Pairs.
   d) On the top right of the screen, select Create Key Pair.
   e) Provide a name. The name can be any valid name.
   f) Choose RSA type, and then choose the pem format.
   g) Click Create key pair.
4. Return to the Cloudera Management Console and navigate to Environments Register Environments .
5. Provide an environment name and description. The name can be any valid name.
6. Choose Amazon as the cloud provider.

7. Under **Amazon Web Services Credential**, choose the credential that you created earlier.
8. Click Next.
9. Under **Data Lake Settings**, give your new data lake a name. The name can be any valid name. Choose the latest data lake version.
10. Under **Data Access and Audit**:
    - Choose prefix-data-access-instance-profile>
    - For Storage Location Base, choose the StorageLocationBase from the cloud formation template.
    - For Data Access Role, choose prefix-datalake-admin-role.
    - For Ranger Audit Role, choose prefix-ranger-audit-role, where "prefix" is the prefix you defined in the **Parameters** section of the stack details in AWS.

    For example:



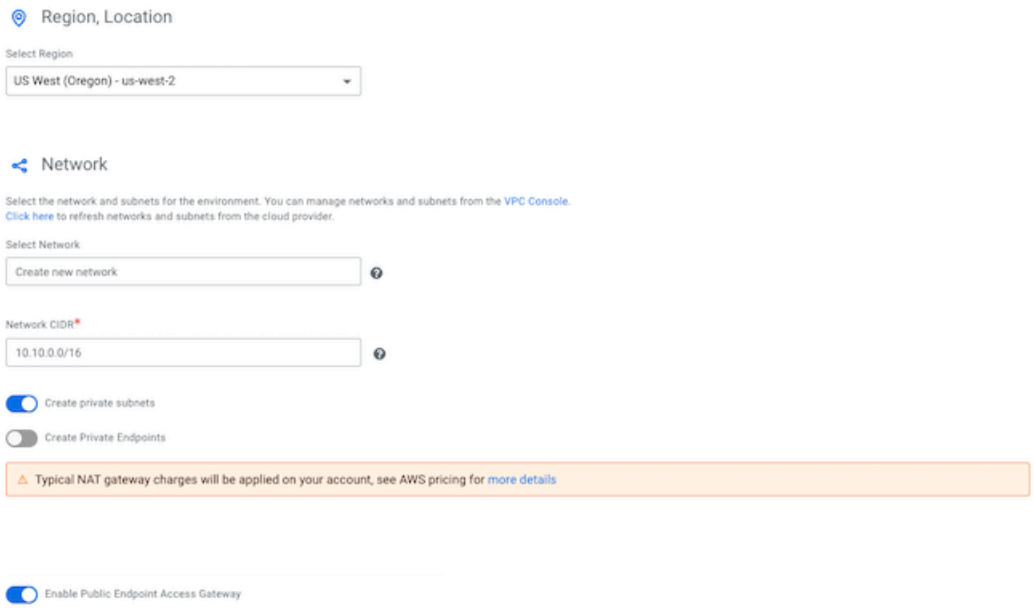
11. For Data Lake **Scale**, choose Light Duty.
12. Click Next.
13. Under Select Region, choose your desired region. This should be the same region you created an SSH key in previously.

14. Under **Select Network**, choose Create New Network.
15. Create private subnets should be enabled by default. If it isn't, enable it.

    Note:

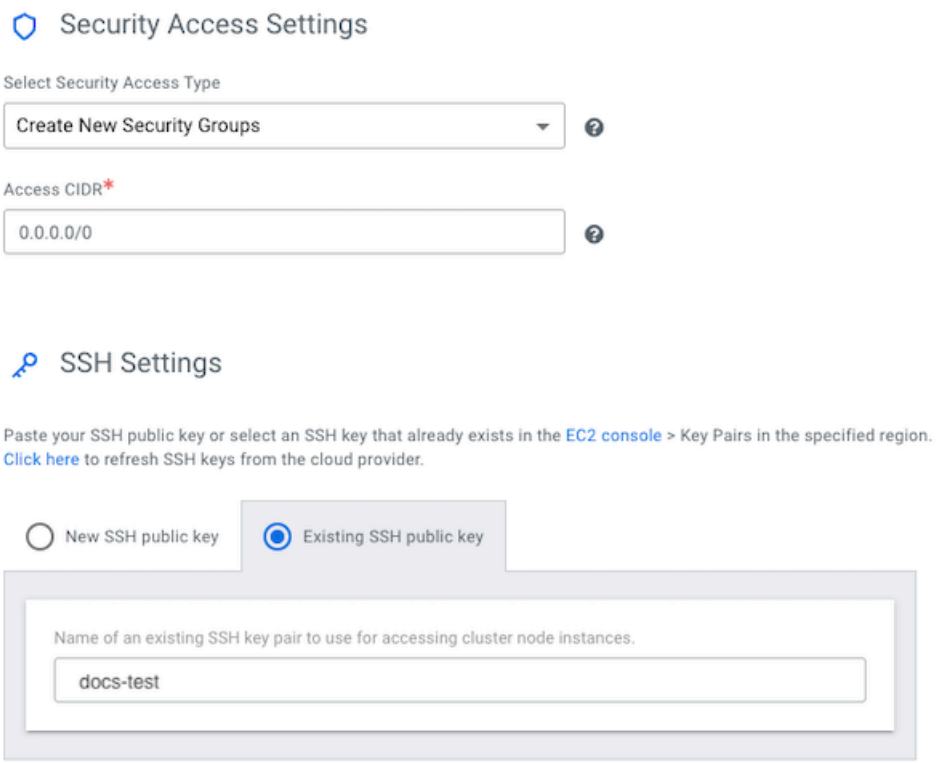
    By enabling private subnets you will not have SSH access to cluster nodes unless you have access to the VPC.
16. Click the toggle button to enable Enable Public Endpoint Access Gateway.

    For example:



17. Under **Security Access Settings**, choose Create New Security Groups.

18. Under **SSH Settings**, choose the SSH key you created earlier.
    For example:



19. Optionally, under **Add Tags**, provide any tags that you'd like the resources to be tagged with in your AWS account.
20. Click Next.

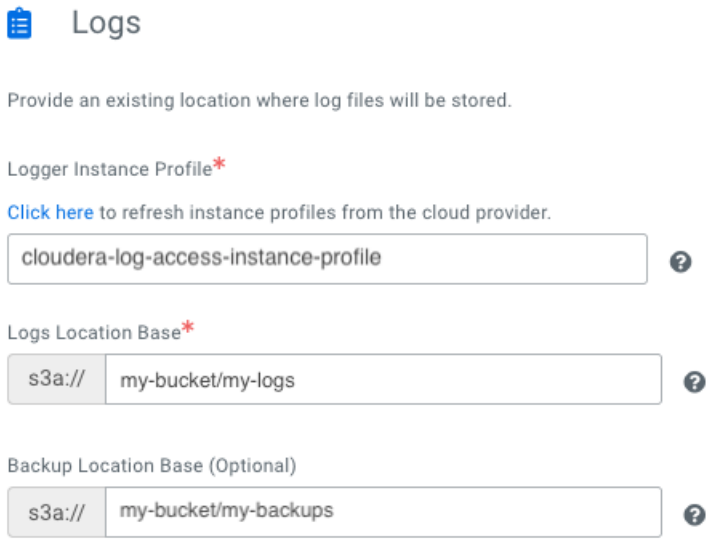
21. Under **Logs**:

a) Choose the Instance Profile titled prefix-log-access-instance-profile, where "prefix" is the prefix you defined in the **Parameters** section of the stack details in AWS.

b) For Logs Location Base, choose the LogsLocationBase from the CloudFormation template.

c) For Backup Location Base, choose the BackupLocationBase from the CloudFormation template.

For example, using the parameters we defined earlier:



22. Click Register Environment.