Cloudera Public Cloud

AWS Onboarding Quickstart

Date published: 2019-08-22 Date modified:



https://docs.cloudera.com/

Legal Notice

© Cloudera Inc. 2025. All rights reserved.

The documentation is and contains Cloudera proprietary information protected by copyright and other intellectual property rights. No license under copyright or any other intellectual property right is granted herein.

Unless otherwise noted, scripts and sample code are licensed under the Apache License, Version 2.0.

Copyright information for Cloudera software may be found within the documentation accompanying each component in a particular release.

Cloudera software includes software from various open source or other third party projects, and may be released under the Apache Software License 2.0 ("ASLv2"), the Affero General Public License version 3 (AGPLv3), or other license terms. Other software included may be released under the terms of alternative open source licenses. Please review the license and notice files accompanying the software for additional licensing information.

Please visit the Cloudera software product page for more information on Cloudera software. For more information on Cloudera support services, please visit either the Support or Sales page. Feel free to contact us directly to discuss your specific needs.

Cloudera reserves the right to change any products at any time, and without notice. Cloudera assumes no responsibility nor liability arising from the use of products, except as expressly agreed to in writing by Cloudera.

Cloudera, Cloudera Altus, HUE, Impala, Cloudera Impala, and other Cloudera marks are registered or unregistered trademarks in the United States and other countries. All other trademarks are the property of their respective owners.

Disclaimer: EXCEPT AS EXPRESSLY PROVIDED IN A WRITTEN AGREEMENT WITH CLOUDERA, CLOUDERA DOES NOT MAKE NOR GIVE ANY REPRESENTATION, WARRANTY, NOR COVENANT OF ANY KIND, WHETHER EXPRESS OR IMPLIED, IN CONNECTION WITH CLOUDERA TECHNOLOGY OR RELATED SUPPORT PROVIDED IN CONNECTION THEREWITH. CLOUDERA DOES NOT WARRANT THAT CLOUDERA PRODUCTS NOR SOFTWARE WILL OPERATE UNINTERRUPTED NOR THAT IT WILL BE FREE FROM DEFECTS NOR ERRORS, THAT IT WILL PROTECT YOUR DATA FROM LOSS, CORRUPTION NOR UNAVAILABILITY, NOR THAT IT WILL MEET ALL OF CUSTOMER'S BUSINESS REQUIREMENTS. WITHOUT LIMITING THE FOREGOING, AND TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, CLOUDERA EXPRESSLY DISCLAIMS ANY AND ALL IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY, QUALITY, NON-INFRINGEMENT, TITLE, AND FITNESS FOR A PARTICULAR PURPOSE AND ANY REPRESENTATION, WARRANTY, OR COVENANT BASED ON COURSE OF DEALING OR USAGE IN TRADE.

Contents

AWS quickstart (Deprecated)	4
Verify AWS prerequisites	4
Create a Cloudera credential	5
Register a Cloudera environment	7

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.

ELOUDERA Data Platform	How can we help you?		→		C Er	nable New UI 🛛 🖵	D D
Welcome, Dóra. What would you like	to work on?						
✓ Favorites ① ★ : ★	= * = *	:					
HUE Data Engineering	de-cluster1	Flink Dashboard					
V All Services							
: : : DataFlow Data Engineering	: : Data Warehouse Operational Database	: Cloudera Al	E Data Hub Clusters	E Data Catalog	E Replication Manager	Observability	: : Management Console
Observability Analytics Summary o							च Show / Hide ▼
Data Engineering			🗧 Data Warehouse				
Spark Jobs • 0 Failed	• 0 Slow	• 1 Total	Impala Queries	• 4	Failed	• 16 Slow	• 107 Total
Activity between (11/04/2024 - 11/06/2024)							

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.

reate Credential	
aws 🔥 🙆	
Name *	
Enter credential name	0
Description	
Enter description	Θ
Default Minimal	
The default role allows for the default set of operations	s including everything that the minimal role allows for.
{ "Statement": [
{ "Statement": [{ "Sid": "CloudFormationFull", "Action": [} }	
<pre>{ Statement": [</pre>	e 🔶

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

Identity and Access X Management (IAM)	IAM > Policies		•
Q, Search IAM Dashboard	Policies (952) into A policy is an object in AWS that defines perm Q. Filter policies by property or policy name	lissions.	C Actions ▼ Create Policy < 1 2 3 4 5 6 7 45 ♥
 Access management User groups 	Policy name	∞ Туре ч	♥ Used as ♥ Description
Users	AmazonSageMaker-ExecutionP	olicy-202002197124050 Customer managed	Permissions policy (1)
Policies	○	Customer managed	Permissions policy (1) AWS - CDP Standardize
Identity providers	O cdp_permissions	Customer managed	Permissions policy (1) Incremental Permission
Account settings	C 🗄 demo-sdx-pm-bucket-policy-st	3access Customer managed	None

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.

Default Mi	himal				
The default rol	e allows for the default se	et of operations inclu	ding everythin	g that the minin	nal role allows
{	nt": [[::oudFormation loudformation:*" ect": "Allow", cource": ["	Full",	Ľ		
) { {		2 - 9			
) }, Create Cross Use Service Manage	account Access Role ager Account ID and Externa Account ID*	" il ID to create an AWS Ia	AM role		
) (reate Cross Use Service Mar Service Manage 3875533438	account Access Role ager Account ID and Externa Account ID* 26	z _ ==	AM role		
) Create Cross Use Service Manage 3875533438 External ID*	account ID and Externi Account ID and Externi Account ID*	/ _ "	AM role		
] } Create Cross Use Service Mar Service Manage 3875533438 External ID* bb90432f-2	account ID and Externi Account ID and Externi Account ID* 26	/ _ = " I ID to create an AWS L	M role		
1 } Create Cross Use Service Manage 3875533438 External ID* bb90432f=2 Cross-account R	account Access Role ager Account ID and Externi Account ID* 26 20 20 20 20 20 20 20 20 20 20 20 20 20	/ _ = " II ID to create an AWS L 164 abe	LM role		

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.

Prerequisite - Prepare template		
Prepare template Every stack is based on a template. A template is :	JSON or YAML file that contains configuration informatio	n about the AWS resources you want to include in the stack.
• Template is ready	Use a sample template	Create template in Designer
Specify template A template is a JSON or YAML file that describes y	our stack's resources and properties.	
Specify template A template is a JSON or YAML file that describes y	our stack's resources and properties.	
Specify template A template is a JSON or YAML file that describes y Template source Selecting a template generates an Amazon S3 UR	our stack's resources and properties.	
Specify template template is a JSON or YAML file that describes y remplate source electing a template generates an Amazon S3 URI Amazon S3 URL	our stack's resources and properties. . where it will be stored.	template file
Specify template A template is a JSON or YAML file that describes y Template source electing a template generates an Amazon S3 URL Amazon S3 URL Juload a template file	our stack's resources and properties. L where it will be stored.	template file
Specify template A template is a JSON or YAML file that describes y Template source Selecting a template generates an Amazon S3 URI Amazon S3 URL Upload a template file Choose file Setup json	our stack's resources and properties.	template file
Specify template A template is a JSON or YAML file that describes y Femplate source Selecting a template generates an Amazon S3 URL Amazon S3 URL Upload a template file Choose file SSON or YAML formatted file	our stack's resources and properties. where it will be stored. O Uptoad 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:

Stack name	
Stack name	
mc-cdp-stack	
Stack name can include let	ters (A-Z and a-z), numbers (0-9), and dashes (-).
Parameters	
Parameters are defined in y	our template and allow you to input custom values when you create or update a stack.
BackupLocationBase The storage base path to co same bucket or different be	eate an S3 bucket with default encryption for CDP. By default CDP will create the optional subdirectory in the bucket. It is possible to either use the ckets for StorageLocationBase and LogsLocationBase.
my-bucket/my-backu	ps
CrossAccountARN Required if s3 KMS Encrypt arn:aws:lam:: <acct_i< th=""><th>ion is selected D>:role/<role_name></role_name></th></acct_i<>	ion is selected D>:role/ <role_name></role_name>
LogsLocationBase The storage base path to cr same bucket or different be	eate an 53 bucket with default encryption for CDP. By default CDP will create the optional subdirectory in the bucket. It is possible to either use the ckets for StorageLocationBase and LogsLocationBase.
my-bucket/my-logs	
StorageLocationBase The logging base path to co same bucket or different be	eate an 53 bucket with default encryption for CDP. By default CDP will create the optional subdirectory in the bucket. It is possible to either use the ckets for StorageLocationBase and LogsLocationBase.
my-bucket/my-data	
prefix prefix for IAM objects, sepa	rated by a dash.
cloudera	
s3KmsEncyption	Roured with AWS managed KMS server side encryption
and the second s	

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.

Capabilities	
Interpretation of the second secon	Policy]
This template contains Identity and Access Management (IAM) resourd the minimum required permissions. In addition, they have custom nan account. Learn more	res. Check that you want to create each of these resources and that they have nes. Check that the custom names are unique within your AWS
I acknowledge that AWS CloudFormation might create IAM reso	urces with custom names.
	Cancel Previous Create change set Create stack

- 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.

0

0

ଢ

- 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:



Provide an existing location where workload data will be stored.

Assumer Instance Profile*

Click here to refresh instance profiles from the cloud provider.

cloudera-data-access-instance-profile

Storage Location Base*



Data Access Role*

cloudera-datalake-admin-role

Ranger Audit Role*

cloudera-ranger-audit-role

ID Broker Mappings

You may optionally provide mappings for users or groups.

Add

- **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.

٢	7	_		1	
I	τ		Ы	L	
I	1	5	N	L	
L				r	
ι	_	_	-		

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.

or example:	
Region, Location	
elect Region	
US West (Oregon) - us-west-2 👻	
Network ielect the network and subnets for the environment. You can manage a first have a formed and unbeen from the alound annufactor.	etworks and subnets from the VPC Console.
lick here to refresh networks and subnets from the cloud provider.	
Create new network	0
letwork CIDP*	
10.10.0.0/16	•
Create private subnets	
Create Private Endpoints	
△ Typical NAT gateway charges will be applied on your according to the second seco	unt, see AWS pricing for more details

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

18. Under SSH Settings, choose the SSH key you created earlier.

For example:

Create New Security Groups	- 0	
Access CIDR*		
0.0.0.0/0	Ø	
Aste your SSH public key or select an SSH key that a Slick here to refresh SSH keys from the cloud provide New SSH public key Existing SSH Name of an existing SSH key pair to use for acc	Iready exists in the EC2 console > Key Pairs in the spo r. ublic key ssing cluster node instances.	cified region.

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:

	~	
Provide an	existing location where log files will be stored.	
Logger Ins	tance Profile*	
Click here	to refresh instance profiles from the cloud provider.	
cloudera	a-log-access-instance-profile	0
	tion Base*	
Logs Loca		
s3a://	my-bucket/my-logs	
s3a:// Backup Lo	my-bucket/my-logs cation Base (Optional)	

22. Click Register Environment.