

Azure Onboarding Quickstart

Date published: 2019-08-22

Date modified:

CLOUDERA

Legal Notice

© Cloudera Inc. 2024. 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

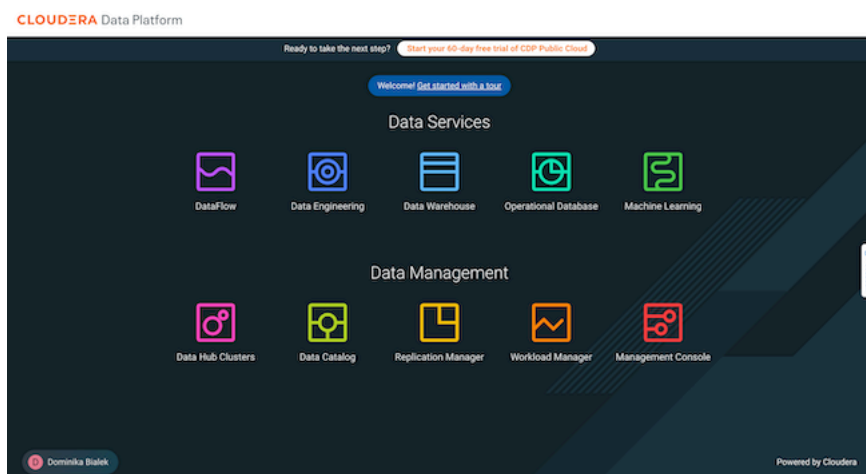
- Azure quickstart (Deprecated).....4**
 - Verify Azure prerequisites..... 5
 - Create an Azure AD app.....6
 - Deploy the template.....7
 - Assign roles.....8
 - Create or locate an SSH Key..... 10
 - Create a CDP credential..... 10
 - Register a CDP environment.....10

Azure quickstart (Deprecated)

If you've reached the CDP 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 CDP to your Azure subscription, 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 CDP on Azure, refer to [Deploy CDP using Terraform](#).



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

- The CDP console pictured above
- The Azure console
- Azure Cloud shell



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

In addition to this documentation, you can refer to the following video:



The steps that we will perform are:

Step 0: Verify the Azure prerequisites

Step 1: Create an Azure AD app

Step 2: Deploy the Azure quickstart template

Step 3: Assign roles

Step 4: Create or locate an SSH key

Step 5: Create a CDP credential

Step 6: Register a CDP environment

Verify Azure cloud platform prerequisites

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

- This Azure onboarding quickstart is intended for simple CDP evaluation deployments only. It may not work for scenarios where Azure resources such as VNet, security group, storage accounts, and so on, are pre-created or Azure accounts have restrictions in place.
- User running the Azure onboarding quickstart should have:
 - Owner permissions on the Azure subscription that you would like to use for CDP.
 - Rights to create Azure resources required by CDP. See list of [Azure resources used by CDP](#).
 - Rights to create an Azure AD application (service principal) and assign Contributor role at subscription level.
 - CDP Admin role or Power User role in CDP subscription.
- This Azure onboarding quickstart uses an Azure ARM template that automatically creates the required resources such as storage accounts, containers, managed identities, resource groups, and so on.

- CDP Public Cloud relies on several Azure services that should be available and enabled in your region of choice. Verify if you have enough quota for each Azure service to set up CDP in your Azure account. See list of [Azure resources used by CDP](#).

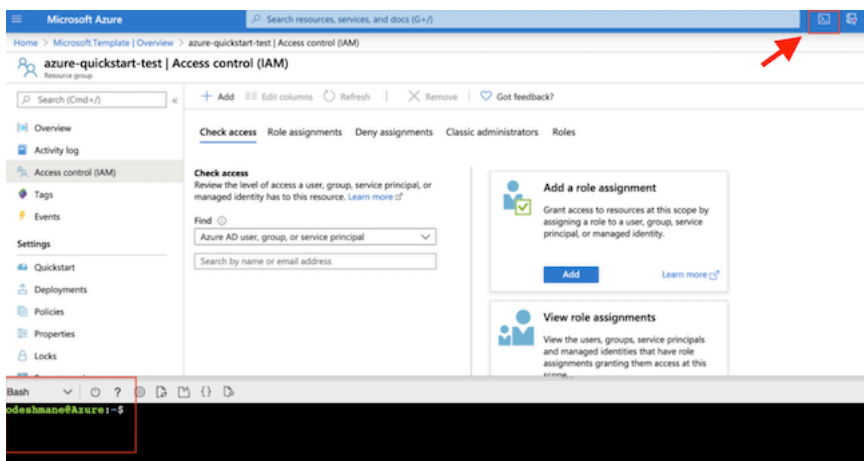
If you have more complex requirements than those listed here, contact Cloudera Sales Team to help you with CDP onboarding.

Create an Azure AD app

In the Azure portal, create an application in your Azure Active Directory tenant. This steps allows you to use the native Cloud Shell terminal and not have to set up Azure CLI.

Procedure

1. Log in to the Azure portal and launch Cloud Shell.



2. When prompted, select Bash to use Bash shell.
3. Run the following command to return the subscription ID and tenant ID:

```
az account list | jq '.[] | { "name": .name, "subscriptionId": .id, "tenantId": .tenantId, "state": .state }'
```

The output of this command is shown below:

```
{
  "name": "Solutions Engineering",
  "subscriptionId": "xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx",
  "tenantId": "xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx",
  "state": "Enabled"
}
```

Make a note of the subscriptionId and tenantId values. You will need them later.



Note: In case you have more than one subscription, make sure to only make a note of the subscription that you would like to use for CDP.

4. Run the command below to create an app in Azure AD and assign the "Contributor" role at the subscription.



Note: Replace {subscriptionId} in the command with the subscription ID value from the previous step.

```
az ad sp create-for-rbac --name http://cloudbreak-app --role Contributor
--scopes /subscriptions/{subscriptionId}
```

The output of this command is shown below:

```

msf5@kali:~$ as msf create-for-rbac --name http://cloudbreak-appl --role Contributor --scopes /subscriptions/
Creating a role assignment under the scope of "/subscriptions/..."
{
  "appId": "...",
  "displayName": "cloudbreak-appl",
  "name": "http://cloudbreak-appl",
  "password": "...",
  "tenant": "..."
}

```

Deploy the Azure quickstart template

The Azure quickstart template is a customized ARM template that deploys essential Azure resources for the CDP environment.

Procedure

1. Click [Deploy to Azure](#) to begin ARM template deployment of CDP prerequisites in your Azure subscription.
2. Log in to Azure to create the resources in your subscription that are required for CDP deployment. These resources include VNet, ADLS Gen2, and 4 user managed identities.
3. On the **Custom deployment** screen, click Create new under the Resource group field and then give the resource group a name (it should only contain letters, numbers, and hyphens).

- Under **Settings**, provide an Environment Name in the corresponding field. The Environment Name should be short (between five and twelve characters) and should include only lowercase characters and hyphens (no underscores).

Custom deployment
Deploy from a custom template

TEMPLATE

Customized template
10 resources

Edit template Edit paramet... Learn more

BASICS

Subscription * azure-se-cdp-sandbox-env

Resource group * (New) azure-quickstart-test1
Create new

Location * (US) Central US

SETTINGS

Environment Name ① cdpazureqs

Virtual Network Name ① [parameters('environmentName')]

Storage Account Name ① [parameters('environmentName')]

Data Access Identity Name ① [concat(parameters('environmentName'), '-DataAccessIdentity')]

Logger Identity Name ① [concat(parameters('environmentName'), '-LoggerIdentity')]

Assumer Identity Name ① [concat(parameters('environmentName'), '-AssumerIdentity')]

Ranger Audit Identity Name ① [concat(parameters('environmentName'), '-RangerIdentity')]

TERMS AND CONDITIONS

Purchase

- Accept the terms and conditions, and click Purchase.
An ARM script begins to run in the background, creating the resources required for a CDP environment. This may take around 10 minutes.
- When your resource group is up, navigate to the **Overview** page of the resource group.
- Copy and paste the following values into a note, as you will need them in the next task:
 - Subscription ID: Your subscription ID is found at the top of the resource group **Overview** page.
 - Resource group: The name of the resource group that you created.

Assign roles

Azure Resource Manager templates do not support role assignments at a scope other than resource groups. Perform the following role assignments through UI or CLI.

Before you begin

Make sure that you have your note from the previous step, where you copied values for the Subscription ID and resource group name.

Procedure

- Once you have values for the subscription ID, resource group name, storage account, environment name, and all four managed identities, click [here](#) to download a script.
- Create a new file in Cloud Shell with the same name, and copy the content of the script there.

3. Replace the following values in the script with the values you have collected thus far:

```
#!/bin/sh

export SUBSCRIPTIONID="<REPLACE WITH YOUR AZURE SUBSCRIPTION ID>"
export RESOURCEGROUPNAME="<REPLACE WITH EXISTING RESOURCE GROUP NAME>"
export STORAGEACCOUNTNAME=$(az storage account list -g $RESOURCEGROUPNAME --subscription $SUBSCRIPTIONID | jq '.[] | select(.name | test("StorageAccount")) | .name' | tr -d ' ')
export ASSUMER_OBJECTID=$(az identity list -g $RESOURCEGROUPNAME --subscription $SUBSCRIPTIONID | jq '.[] | select(.name | test("AssumerIdentity")) | .principalId' | tr -d ' ')
export DATAACCESS_OBJECTID=$(az identity list -g $RESOURCEGROUPNAME --subscription $SUBSCRIPTIONID | jq '.[] | select(.name | test("DataAccessIdentity")) | .principalId' | tr -d ' ')
```

For example, your script should look similar to this:

```
#!/bin/sh

export SUBSCRIPTIONID="jfs85ls8-sik8-8329-fq0m-jqo7v06dk6sy"
export RESOURCEGROUPNAME="myCDPresourcegroup"
export STORAGEACCOUNTNAME=$(az storage account list -g $RESOURCEGROUPNAME --subscription $SUBSCRIPTIONID | jq '.[] | select(.name | test("StorageAccount")) | .name' | tr -d ' ')
export ASSUMER_OBJECTID=$(az identity list -g $RESOURCEGROUPNAME --subscription $SUBSCRIPTIONID | jq '.[] | select(.name | test("AssumerIdentity")) | .principalId' | tr -d ' ')
export DATAACCESS_OBJECTID=$(az identity list -g $RESOURCEGROUPNAME --subscription $SUBSCRIPTIONID | jq '.[] | select(.name | test("DataAccessIdentity")) | .principalId' | tr -d ' ')
export LOGGER_OBJECTID=$(az identity list -g $RESOURCEGROUPNAME --subscription $SUBSCRIPTIONID | jq '.[] | select(.name | test("LoggerIdentity")) | .principalId' | tr -d ' ')
export RANGER_OBJECTID=$(az identity list -g $RESOURCEGROUPNAME --subscription $SUBSCRIPTIONID | jq '.[] | select(.name | test("RangerIdentity")) | .principalId' | tr -d ' ')
# Assign Managed Identity Operator role to the assumerIdentity principal at subscription scope
az role assignment create --assignee $ASSUMER_OBJECTID --role 'fla07417-d97a-45cb-824c-7a7467783830' --scope "/subscriptions/$SUBSCRIPTIONID"
# Assign Virtual Machine Contributor role to the assumerIdentity principal at subscription scope
az role assignment create --assignee $ASSUMER_OBJECTID --role '9980e02c-c2be-4d73-94e8-173bdc7cf3c' --scope "/subscriptions/$SUBSCRIPTIONID"
# Assign Storage Blob Data Contributor role to the assumerIdentity principal at logs filesystem scope
az role assignment create --assignee $ASSUMER_OBJECTID --role 'ba92f5b4-2d11-453d-a403-e96b0029c9fe' --scope "/subscriptions/$SUBSCRIPTIONID/resourceGroups/$RESOURCEGROUPNAME/providers/Microsoft.Storage/storageAccounts/$STORAGEACCOUNTNAME/blobServices/default/containers/logs"
# Assign Storage Blob Data Contributor role to the loggerIdentity principal at logs/backup filesystem scope
az role assignment create --assignee $LOGGER_OBJECTID --role 'ba92f5b4-2d11-453d-a403-e96b0029c9fe' --scope "/subscriptions/$SUBSCRIPTIONID/resourceGroups/$RESOURCEGROUPNAME/providers/Microsoft.Storage/storageAccounts/$STORAGEACCOUNTNAME/blobServices/default/containers/logs"
az role assignment create --assignee $LOGGER_OBJECTID --role 'ba92f5b4-2d11-453d-a403-e96b0029c9fe' --scope "/subscriptions/$SUBSCRIPTIONID/resourceGroups/$RESOURCEGROUPNAME/providers/Microsoft.Storage/storageAccounts/$STORAGEACCOUNTNAME/blobServices/default/containers/backups"
# Assign Storage Blob Data Owner role to the dataAccessIdentity principal at logs/data/backup filesystem scope
az role assignment create --assignee $DATAACCESS_OBJECTID --role 'b7e6dc6d-f1e8-4753-8033-0f276bb0955b' --scope "/subscriptions/$SUBSCRIPTIONID/resourceGroups/$RESOURCEGROUPNAME/providers/Microsoft.Storage/storageAccounts/$STORAGEACCOUNTNAME/blobServices/default/containers/data"
az role assignment create --assignee $DATAACCESS_OBJECTID --role 'b7e6dc6d-f1e8-4753-8033-0f276bb0955b' --scope "/subscriptions/$SUBSCRIPTIONID/resourceGroups/$RESOURCEGROUPNAME/providers/Microsoft.Storage/storageAccounts/$STORAGEACCOUNTNAME/blobServices/default/containers/logs"
```

```
az role assignment create --assignee $DATAACCESS_OBJECTID --role 'b7e6dc6d-f1e8-4753-8033-0f276bb0955b' --scope "/subscriptions/$SUBSCRIPTIONID/resourceGroups/$RESOURCEGROUPNAME/providers/Microsoft.Storage/storageAccounts/$STORAGEACCOUNTNAME/blobServices/default/containers/backups"
# Assign Storage Blob Data Contributor role to the rangerIdentity principal at data/backup filesystem scope
az role assignment create --assignee $RANGER_OBJECTID --role 'ba92f5b4-2d11-453d-a403-e96b0029c9fe' --scope "/subscriptions/$SUBSCRIPTIONID/resourceGroups/$RESOURCEGROUPNAME/providers/Microsoft.Storage/storageAccounts/$STORAGEACCOUNTNAME/blobServices/default/containers/data"
az role assignment create --assignee $RANGER_OBJECTID --role 'ba92f5b4-2d11-453d-a403-e96b0029c9fe' --scope "/subscriptions/$SUBSCRIPTIONID/resourceGroups/$RESOURCEGROUPNAME/providers/Microsoft.Storage/storageAccounts/$STORAGEACCOUNTNAME/blobServices/default/containers/backups"
```

4. Run the Cloud Shell script: `sh azure_msi_role_assign.sh`

Create or locate an SSH Key

CDP requires that you provide a public SSH key for admin access to VM instances.

You can find more information on SSH key requirement in the topic [SSH key](#). If you need to create one, you can do so by running `ssh-keygen -t rsa`.

When you complete this step, you have created all of the Azure resources required for this quickstart.

Create a CDP credential

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

Procedure

1. Log in to the CDP web interface.
2. From the CDP home screen, click the Management Console icon.
3. In the Management Console, select Shared Resources > Credentials from the navigation pane.
4. Click in the Create Credential button to create a new credential.
5. Select the Azure tab, name your credential, under "Credential Type", select "App based", and enter the values you previously collected for subscription ID, app ID, and password.

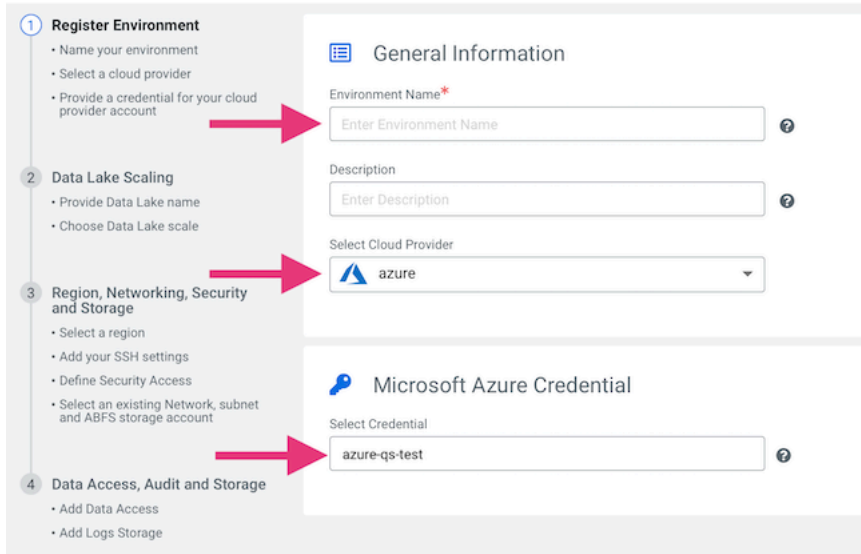
Register a CDP environment

When you register an environment, you set properties related to data lake scaling, networking, security, and storage. You will need your Azure environment name, resource group name, storage account name, and virtual network name from your resource group.

Procedure

1. In the CDP Management Console, navigate to Environments and click Register Environment.
2. Provide an Environment Name and description. The name can be any valid name.
3. Choose Azure as the cloud provider.

4. Under Microsoft Azure Credential, choose the credential you created in the previous task.



1 Register Environment

- Name your environment
- Select a cloud provider
- Provide a credential for your cloud provider account

2 Data Lake Scaling

- Provide Data Lake name
- Choose Data Lake scale

3 Region, Networking, Security and Storage

- Select a region
- Add your SSH settings
- Define Security Access
- Select an existing Network, subnet and ABFS storage account

4 Data Access, Audit and Storage

- Add Data Access
- Add Logs Storage

General Information

Environment Name*

Enter Environment Name ?

Description

Enter Description ?

Select Cloud Provider

azure

Microsoft Azure Credential

Select Credential

azure-qs-test ?

5. Click Next.
6. Under **Data Lake Settings**, give your new data lake a name. The name can be any valid name. Choose the latest data lake version.

7. Under **Data Access and Audit**, choose the following:

- Assumer Identity: <resourcegroup-name>-<envName>-AssumerIdentity
- Storage Location Base: data@<storageaccount-name>
- Data Access Identity: <resourcegroup-name>-<envName>-DataAccessIdentity
- Ranger Audit Role: <resourcegroup-name>-<envName>-RangerIdentity



Warning: Ensure that you have entered the correct location base. If the name entered does not match the actual location base created by the quickstart script, environment registration will fail.

For example:

Data Access and Audit

Provide an existing location where workload data will be stored.

Assumer Identity*

azure-quickstart-test - cdpazureqs-AssumerIdentity  


Storage Location Base*

abfs:// data@cdpazureqs  .dfs.core.windows.net 

Data Access Identity*

azure-quickstart-test - cdpazureqs-DataAccessIdentity  

Ranger Audit Identity*

azure-quickstart-test - cdpazureqs-RangerIdentity  

8. For Data Lake Scale, choose Light Duty.

The screenshot displays the Azure portal configuration interface for a Data Lake environment. On the left, a vertical navigation pane shows four steps: 1. Register Environment (completed), 2. Data Lake Scaling (active), 3. Region, Networking, Security and Storage, and 4. Data Access, Audit and Storage. A red arrow points from step 2 to the 'Data Lake Settings' section. This section contains two input fields: 'Data Lake Name' with the value 'azure-qs-test-dl' and 'Data Lake version' with a dropdown menu set to 'Runtime 7.1.0'. Below this is the 'Scale' section, which prompts the user to 'Choose a scale and a purpose of this environment from a pre-defined Data Lake template'. Under the 'Scale' heading, the 'Light Duty' option is selected with a radio button, and the 'Secure Access' option is also visible with a shield icon.

9. Click Next.

10. Under Select Region, choose your desired region. This should be the same region you created an SSH key in previously.

11. Under Select Resource Group, choose your resource group <resourcegroup-name>.

12. For the Select Network field, select the name of the "Virtual Network" resource that was created when you deployed the ARM template to create the resource group. The name of the Virtual Network should be the same as your environment name, but you can verify this in the Azure portal on the Overview page of your resource group. In the following Select Subnets field, ensure all three subnets are selected.

13. Slide the Enable Public Endpoint Access Gateway toggle to the Enabled position. Select any of your three subnets.

14. The Create Public IPs toggle can remain in the default Enabled position.

15. In the next drop-down input box, keep Flexible Server selected.

16. Under the Encryption section, leave both the Enable encryption at host and Enable Customer-Managed Keys toggles in their default deselected state.

17. Under the Proxies section, leave the Select Proxy Configuration field in its default "Do not use Proxy Configuration" state.

18. Under **Security Access Settings**, select Create New Security Groups for the Security Access Type.

The screenshot shows the 'Region, Networking, Security and Storage' step of the Azure CDP configuration wizard. The left sidebar lists the steps: 1. Register Environment, 2. Data Lake Scaling, 3. Region, Networking, Security and Storage (selected), and 4. Data Access, Audit and Storage. The main content area is divided into three sections: 'Region, Location' with a 'Select Region' dropdown set to 'Central US - Central US'; 'Network' with 'Select Network' set to 'cdpazureqs' and 'Select Subnets' set to 'default', plus toggle switches for 'Enable Cluster Connectivity Manager' and 'Don't Create Public Ip'; and 'Security Access Settings' with 'Select Security Access Type' set to 'Create New Security Groups' and 'Access CIDR' set to '0.0.0.0/0'. Red arrows point from the sidebar to the 'Region, Location' and 'Security Access Settings' sections.

19. Under **SSH Settings**, paste the public SSH key that you created earlier.

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

21. Click Next.

22. Under **Logs**, choose the following:

- Logger Identity: <resourcegroup-name>-<envName>-LoggerIdentity
- Logs Location Base: logs@<storageaccount-name>
- Backup Location Base: backups@<storageaccount-name>



Warning: Ensure that you have entered the correct location base. If the name entered does not match the actual location base created by the quickstart script, environment registration will fail.


For example:



Logs

Provide an existing location where log files will be stored.

Logger Identity*

azure-quickstart-test - cdpazurereqs-LoggerIdentity  

Logs Location Base*

abfs:// logs@cdpazurereqs  dfe.core.windows.net 

Backup Location Base (Optional)

abfs:// backups@cdpazurereqs  dfe.core.windows.net 

23. Under the Telemetry section, leave the inputs as default (Observability Enabled, Deployment Cluster Logs Collection Disabled).

24. Before finishing, click the Show CLI Command button and then Copy the full command for future reference. This can be useful for e.g. analyzing any errors in the deployment process.

25. Click Register Environment.