CDP Public Cloud

GCP Onboarding Quickstart

Date published: 2019-08-22
Date modified:

CLOUDƎRA

Legal Notice

Contents

GCP quickstart

In this quickstart, we will show you step by step how to connect CDP to your Google Cloud Platform (GCP) account, so that you can begin to provision clusters and workloads.
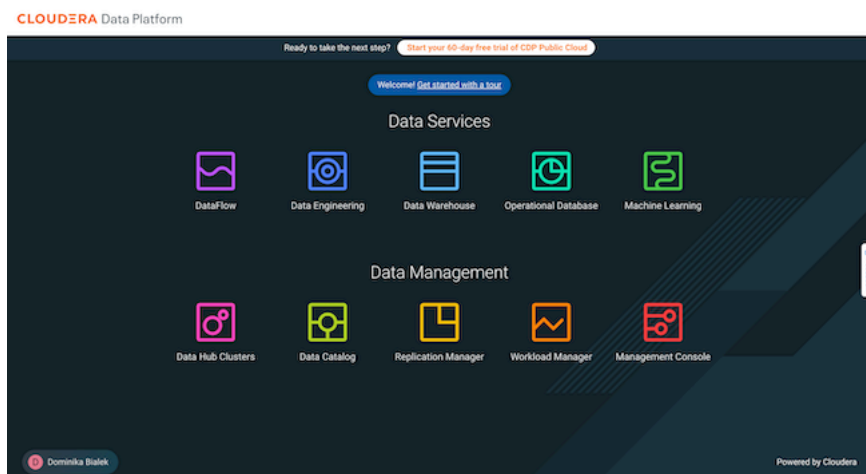
To complete this quickstart, you will need access:

- The CDP console pictured above
- The GCP console

> **Note:** This GCP onboarding quickstart is intended for simple CDP evaluation deployments only. It may not work for scenarios where GCP resources such as VPC network, firewall rules, storage accounts, and so on, are pre-created or GCP accounts have restrictions in place.

The steps that we will perform are:

Step 0: Verify the GCP prerequisites

Step 1: Create a provisioning credential

Step 2: Create GCP prerequisites

Step 3: Register a GCP environment in CDP

Verify GCP cloud platform prerequisites

Before getting started with the Google Cloud Platform (GCP) onboarding quickstart, review and acknowledge the following:

- This GCP onboarding quickstart is intended for simple CDP evaluation deployments only. It may not work for scenarios where GCP resources such as VPC network, firewall rules, storage accounts, and so on, are pre-created or GCP accounts have restrictions in place.
- Users running the GCP onboarding quickstart should have CDP Admin role or Power User role in CDP subscription.
- The following APIs should be enabled in the project that you would like to use to run the quickstart: GCP APIs.

- In order to run the quickstart, the Google APIs service agent user must be set to Owner. For instructions on how to check or update the Google APIs service agent permissions, see Grant permissions to the Google APIs service account.

> **Note:** The Grant permissions to the Google APIs service account instructions show how to assign the Role Administrator role, but you should assign the Owner role instead.

    If your organization's policies don't allow you to assign the Owner role, and you are required to use granular roles or permissions instead, you should make sure to assign, among other roles, the Role Administrator role (or equivalent granular permissions allowing access to the Deployment Manager).

- This GCP onboarding quickstart uses a Deployment Manager template that automatically creates the required resources such as the VPC network, firewall rules, storage buckets, service accounts, and so on.
- CDP Public Cloud relies on several GCP services that should be available and enabled in your project of choice. Verify if you have enough quota for each GCP service to set up CDP. See list of GCP resources used by CDP.

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

Create a provisioning credential

The first step is to create a provisioning credential. The CDP credential is the mechanism that allows CDP to create resources inside of your GCP account.

Steps

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 on the "Create Credential" button.
5. Select the (Google Cloud Platform) tab.
6. Give your credential a name and description.

7. Copy the script provided under "Create Service Account":

This script can create the service account that is a prerequisite for the CDP credential.

8. Navigate to the GCP console.

9. Verify that you are in the project that you would like to use for CDP. Switch projects if needed:
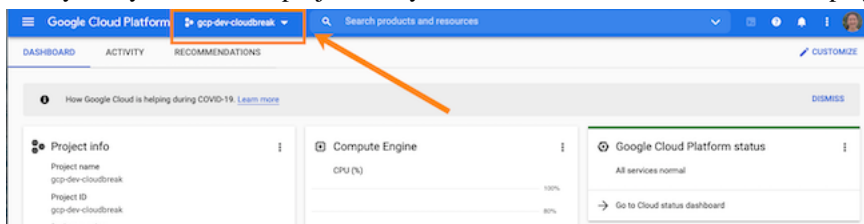
10. Open the Cloud Shell (available from upper right corner):
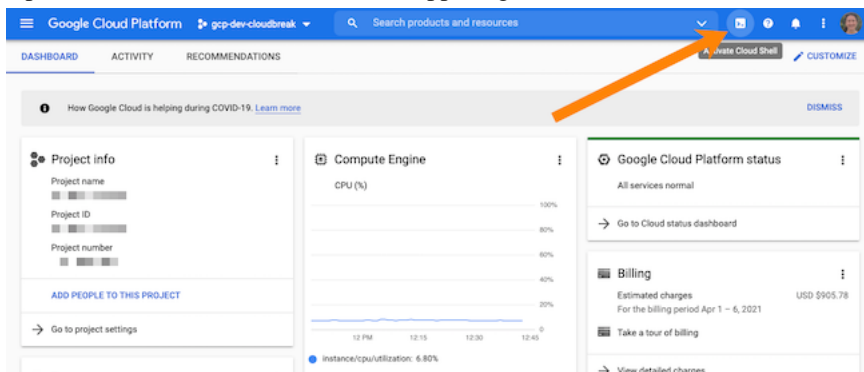
The Cloud Shell window opens in the bottom of the browser window.

11. Paste the script directly into the Cloud Shell terminal.

12. When prompted, click Authorize.

13. The script will run and then end by prompting you to download the credential file to your local machine. Click Download to download the file:
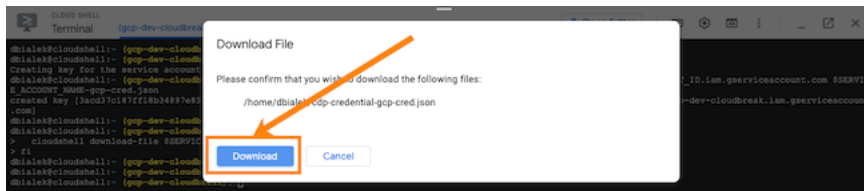
14. Head back to the CDP console and upload the JSON credential file you just downloaded from the GCP console:

15. Click the "Create" button and you're done!

Create GCP prerequisites

The second step is to create the CDP prerequisites in your GCP project. To get this done quickly, we will use a script that creates a VPC network, a subnet, firewall rules, service accounts, storage buckets, and so on.

Steps

1. Navigate to the browser tab with the GCP console.

2. Click on the project navigation widget:

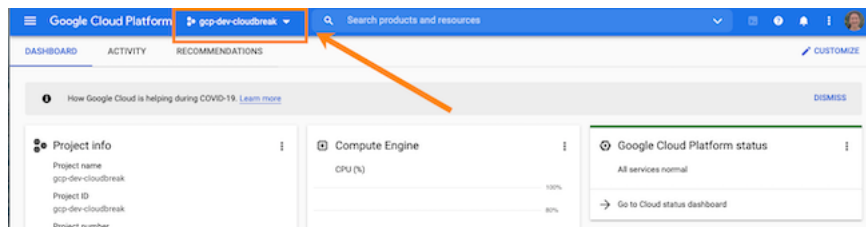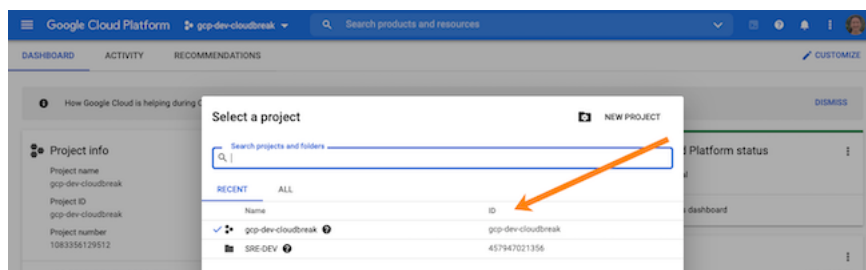**3.** A window appears, allowing you to select a project. Copy the ID of the project where you created your credential earlier:

You will need it in a later step

4. In the GCP console, download the following files gcp-script.sh and gcp_param.yml:

```
wget https://docs.cloudera.com/cdp-public-cloud/cloud/quickstart-files/g
cp-script.sh
wget https://docs.cloudera.com/cdp-public-cloud/cloud/quickstart-files/
gcp_param.yml
```

This is a script that creates all of the prerequisite resources required for CDP.

5. Run the provided bash script using the following command:

```
bash gcp-script.sh <prefix> <region> <project-id> <IP-CIDR-to-whitelist>
```

Replace the following with actual values:

- Prefix - A prefix to prepend to names of all resources that the script creates for you. For example, if your chosen prefix is "cloudera", CDP prepends "cloudera-" to the names of all created resources.
- Region - A GCP region where you would like to deploy your environment. For a list of supported regions, see Supported GCP regions.
- Project ID - The project ID that you obtained in an earlier step. This is the project where you will deploy the resources required for a CDP environment.
- CIDR-to-whitelist - The IPv4 CIDR range for SSH and UI access.

For example:

```
bash gcp-script.sh test us-east4 gcp-dev 73.221.71.0/24
```

6. The script creates a new deployment called <prefix>-cdp-deployment in the Deployment Manager and creates resources in your GCP account:
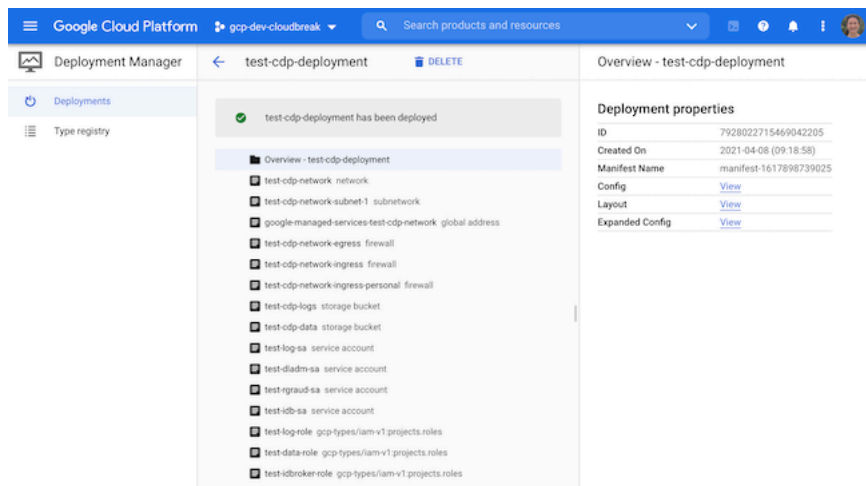
The script does the following:

1. Verify that the correct number of arguments were supplied.
2. Replace the Deployment Manager config parameters with those arguments that you supplied.
3. Check for the existence of a custom IAM role with the provided prefix (i.e. check if you have run this script before with the same prefix).
4. Run the Deployment Manager (which creates resources such as a VPC network, a subnet, firewall rules, service accounts, storage buckets, and so on).
5. Add policy bindings to the created service accounts
6. Change GCS bucket permissions for created service accounts
7. Create a VPC peering to servicenetworking.googleapis.com (for purposes of the CloudSQL DB that CDP creates).

Once the deployment is ready, you will see a message "<prefix>-cdp-deployment has been deployed". At that point, you can proceed to the next step.

> **Note:** If you need to delete all of the prerequisites created by the script, you can delete the deployment using the Deployment Manager UI (using the "Delete" button visible in the above screenshot).

Register a CDP environment

The third (and last) step is to register your GCP environment in CDP. You will:

• Use the credential created in Step1.
• Point CDP to the resources created in Step 2.

You have two options for performing the environment registration step:

• Option 1: CDP web interface
• Option 2: CDP CLI

Prerequisites

You need an RSA key pair. You will be asked to provide a public key and you will use the matching private key for admin access to CDP instances.

Register a GCP environment using CDP web interface

1. Switch back to the browser window with the CDP console.
2. Navigate to the Management Console > Environments.

3. Click the Register Environment button.
4. In the General Information section provide the following:

- Environment name - Provide a name for the environment.
- Select Cloud Provider - Select Google Cloud.

5. Under Select Credential, select the credential that you created in Step 1.

6. Click Next.
7. In the Data Lake Settings section, provide the following:

- Data Lake Name - Enter a name for the Data Lake that CDP creates for your environment.
- Data Lake Version - The latest version should be pre-selected by default.

8. In the Data Access and Audit section, provide the following:

- Assumer Service Account - Select <prefix>-idb-sa. The prefix is what you provided in Step 2.
- Storage Location Base - Enter <prefix>-cdp-data
- Data Access Service Account - Select <prefix>-dladm-sa
- Ranger Audit Service Account - Select <prefix>-rgraud-sa
- Under IDBroker Mappings:

  a. Click Add
  b. Under User or Group, select your user
  c. Under Service Account, enter the following service account name <prefix>-dladm-sa

9. Click Next.
10. Under Region, Location, select the same region that you provided in Step 2.

11. In the Network section, provide the following:

- Select Network - Select the network called <prefix>-cdp-network
- Select Subnets - Select <prefix>-cdp-network-subnet-1
- Ensure that CCM is enabled.
- Enable Create Public IPs.

12. Under Security Access Settings, under Select Security Access Type, select Do not create firewall rule.

13. Under SSH Settings, paste your RSA public key.

14. Under Add Tags add tags if necessary.

15. Click Next.

16. Under Logs, provide the following:

- Logger Service Account - Select <prefix>-log-sa
- Logs Location Base - Enter <prefix>-cdp-logs
- Backups Location Base - Enter <prefix>-cdp-backup

17. Click Register Environment.
18. Once your environment is created, its status will change to Available and the Data Lake status will change to Running.

Once your environment is running, you can start creating Data Hub clusters.

Register a GCP environment using CDP CLI

1. Install and configure CDP CLI. If you haven't, refer to CLI client setup.
2. Open the terminal app on your computer.
3. Create your environment using the following command. Replace the following with actual values:

- <NAME_OF_YOUR_CDP_CREDENTIAL> - Replace this with the actual name that you provided on the CDP web UI in step 1.
- <REGION> - Replace this with the ID of the region selected in step 2.
- <RSA_PUBLIC_KEY> - Replace this with your RSA public key. You will use the matching private key for admin access to CDP instances.
- <PREFIX> - Replace this with the prefix specified in step 2.
- <PROJECT_ID> - Replace this with the ID of the GCP project specified in step 2.

```
cdp environments create-gcp-environment  --environment-name "<PREFIX>-cdp-
env" \
    --credential-name "<NAME_OF_YOUR_CDP_CREDENTIAL>" \
    --region "<REGION>" \
    --public-key "<RSA_PUBLIC_KEY>" \
    --log-storage storageLocationBase="gs://<PREFIX>-cdp-logs/",service
AccountEmail="<PREFIX>-log-sa@<PROJECT_ID>.iam.gserviceaccount.com", bac
kupStorageLocationBase="gs://<PREFIX>-cdp-backup" \
```

```
    --existing-network-params networkName="<PREFIX>-cdp-network",subnetN
ames="<PREFIX>-cdp-network-subnet-1",sharedProjectId="<PROJECT_ID>"  \
    --enable-tunnel \
    --use-public-ip
```

4. Find your user CRN using the following command:

```
user_crn=$(cdp iam get-user | jq -r .user.crn)
```

5. Set the IDBroker mappings between users and service accounts using the following command. Replace the following with actual values:

- <PREFIX> - Same as used earlier
- <PROJECT_ID> - Same as used earlier
- <USER_CRN> - Replace with your user CRN.

```
cdp environments set-id-broker-mappings \
    --environment-name "<PREFIX>-cdp-env" \
    --baseline-role "<PREFIX>-rgraud-sa@<PROJECT_ID>.iam.gserviceaccount.c
om" \
    --data-access-role "<PREFIX>-dladm-sa@<PROJECT_ID>.iam.gserviceaccount
.com" \
    --mappings accessorCrn="<USER_CRN>",role="<PREFIX>-dladm-sa@<PROJECT
_ID>.iam.gserviceaccount.com"
```

6. Create the Data Lake using the following command. Replace the following with actual values:

- <PREFIX> - Same as used earlier
- <PROJECT_ID> - Same as used earlier

```
cdp datalake create-gcp-datalake --datalake-name "<PREFIX>-cdp-dl" \
    --environment-name "<PREFIX>-cdp-env" \
    --cloud-provider-configuration "serviceAccountEmail=<PREFIX>-idb-sa
@<PROJECT_ID>.iam.gserviceaccount.com",storageLocation="gs://<PREFIX>-cdp-
data"
```

7. Once your environment is created, its status will change to Available and the Data Lake status will change to Running.

Once your environment is running, you can start creating Data Hub clusters.