Setup IAM and Networking for your Dataflow Jobs | Google Cloud Skills Boost

Qwiklabs : 7-9 minutes

Overview

In this lab, you will learn to set up IAM permissions and use private IP adresses for your Datafow jobs.

Objectives

- Use IAM permissions that affect whether a job can be launched.
- Use Private IP addresses for your Dataflow jobs.

Setup and requirements

For each lab, you get a new Google Cloud project and set of resources for a fixed time at no cost.

1. Sign in to Qwiklabs using an **incognito window**.
2. Note the lab's access time (for example, `1:15:00`), and make sure you can finish within that time.
   There is no pause feature. You can restart if needed, but you have to start at the beginning.
3. When ready, click **Start lab**.
4. Note your lab credentials (**Username** and **Password**). You will use them to sign in to the Google Cloud Console.
5. Click **Open Google Console**.
6. Click **Use another account** and copy/paste credentials for **this** lab into the prompts.
   If you use other credentials, you'll receive errors or **incur charges**.
7. Accept the terms and skip the recovery resource page.

Activate Cloud Shell

Cloud Shell is a virtual machine that contains development tools. It offers a persistent 5-GB home directory and runs on Google Cloud. Cloud Shell provides command-line access to your Google Cloud resources. `gcloud` is the command-line tool for Google Cloud. It comes pre-installed on Cloud Shell and supports tab completion.

1. Click the **Activate Cloud Shell** button () at the top right of the console.
2. Click **Continue**.
   It takes a few moments to provision and connect to the environment. When you are connected, you are

also authenticated, and the project is set to your *PROJECT_ID*.

Sample commands

- List the active account name:

gcloud auth list

(Output)

Credentialed accounts: - <myaccount>@<mydomain>.com (active)

(Example output)

Credentialed accounts: - google1623327_student@qwiklabs.net

- List the project ID:

gcloud config list project

(Output)

[core] project = <project_ID>

(Example output)

[core] project = qwiklabs-gcp-44776a13dea667a6 **Note:** Full documentation of **gcloud** is available in the gcloud CLI overview guide.

Task 1. Create a Cloud Storage bucket

1. In Cloud Shell, to set up your variables, run the following command:

   PROJECT=`gcloud config list --format 'value(core.project)'` USER_EMAIL=`gcloud config list account --format "value(core.account)"` REGION={{{ project_0.default_region | "REGION" }}}

2. Create a Cloud Storage bucket:

   gsutil mb -p $PROJECT -b on gs://$PROJECT

Click *Check my progress* to verify the objective. Create a Cloud Storage bucket.

Task 2. Launch a Dataflow job

In this task, you try to run a Dataflow job. It will initially fail because of the lack of IAM permissions. After you assign the required role, the job runs successfully.

1. Firstly, verify the IAM roles associated with the account:

gcloud projects get-iam-policy $PROJECT \ --format='table(bindings.role)' \ --flatten="bindings[].members" \ --filter="bindings.members:$USER_EMAIL"

2. Attempt to launch a Dataflow job:

gcloud dataflow jobs run job1 \ --gcs-location gs://dataflow-templates-{{{ project_0.default_region | "REGION" }}}/latest/Word_Count \ --region $REGION \ --staging-location gs://$PROJECT/tmp \ --parameters inputFile=gs://dataflow-samples/shakespeare/kinglear.txt,output=gs://$PROJECT/results/outputs

This will fail as expected because of missing IAM permissions.

3. Add the Dataflow Admin role to the user account:

gcloud projects add-iam-policy-binding $PROJECT --member=user:$USER_EMAIL --role=roles/dataflow.admin

4. Launch the Dataflow job again:

gcloud dataflow jobs run job1 \ --gcs-location gs://dataflow-templates-{{{ project_0.default_region | "REGION" }}}/latest/Word_Count \ --region $REGION \ --staging-location gs://$PROJECT/tmp \ --parameters inputFile=gs://dataflow-samples/shakespeare/kinglear.txt,output=gs://$PROJECT/results/outputs

5. In the **Google Cloud Console**, on the **Navigation menu**, click **Dataflow > Jobs**, and you will see your dataflow job *job1*.

Please wait for about 5 minutes for your job to complete before you proceed.

Click *Check my progress* to verify the objective. Launch a Dataflow job.

Task 3. Launch in private IPs

In this task, you first try to launch a Dataflow job with the **--disable-public-ips** directive. It will fail in the first attempt because the network does not have Private Google Access (PGA) turned on. You configure PGA and re-run the command to launch the job.

1. In Cloud Shell, to launch a Dataflow job using the **--disable-public-ips** directive, run the following command: gcloud dataflow jobs run job2 \ --gcs-location gs://dataflow-templates-{{{ project_0.default_region | "REGION" }}}/latest/Word_Count \ --region $REGION \ --staging-location gs://$PROJECT/tmp \ --parameters inputFile=gs://dataflow-samples/shakespeare/kinglear.txt,output=gs://$PROJECT/results/outputs --disable-public-ips

This job will fail as expected because PGA is not turned on.

2. To verify, go to the **Google Cloud Console**, on the **Navigation menu**, click **Dataflow > Jobs**, and notice that job2 failed.

3. Click on **job2**, then scroll to the bottom to click on "SHOW" next to *Logs* to see the cause of error.

4. In Cloud Shell, run the following commands to give the user the required role to enable PGA, and then enable PGA:

gcloud projects add-iam-policy-binding $PROJECT --member=user:$USER_EMAIL --role=roles/compute.networkAdmin gcloud compute networks subnets update default \ --region=$REGION \ --enable-private-ip-google-access

5. Repeat step 1:

   gcloud dataflow jobs run job2 \ --gcs-location gs://dataflow-templates-{{{ project_0.default_region | "REGION" }}}/latest/Word_Count \ --region $REGION \ --staging-location gs://$PROJECT/tmp \ --parameters inputFile=gs://dataflow-samples/shakespeare/kinglear.txt,output=gs://$PROJECT/results/outputs --disable-public-ips

6. In the **Google Cloud Console**, on the **Navigation menu**, click **Compute Engine > VM Instances**, and notice that the VM launched has no external IP address.

Note: The VM instance will be deleted once the job status will change to succeeded.

Click *Check my progress* to verify the objective. Launch in Private IPs.

Congratulations!

This concludes the lab. In the lab, you used the correct IAM roles to launch a Dataflow job. Next, you changed the subnet to Private Google Access and launched the VMs that do not use an external IP address as part of your Dataflow job.

End your lab

When you have completed your lab, click **End Lab**. Qwiklabs removes the resources you've used and cleans the account for you.

You will be given an opportunity to rate the lab experience. Select the applicable number of stars, type a comment, and then click **Submit**.

The number of stars indicates the following:

- 1 star = Very dissatisfied
- 2 stars = Dissatisfied
- 3 stars = Neutral
- 4 stars = Satisfied
- 5 stars = Very satisfied

You can close the dialog box if you don't want to provide feedback.

For feedback, suggestions, or corrections, please use the **Support** tab.