

📖 01-gcp-prerequisites.md

🔗 About

This module includes all prerequisites for running the Serverless Spark lab-

- [1. Declare variables](#)
- [2. Enable Google Dataproc and Vertex AI APIs](#)
- [3. Network Configuration](#)
- [4. Create a User Managed Service Account](#)
- [5. Grant IAM permissions for UMSA](#)

🔗 0. Prerequisites

🔗 1. Create a project new project or select an existing project.

Note the project number and project ID.
We will need this for the rest of the lab

🔗 2. IAM Roles needed to execute the prereqs

Grant yourself **Security Admin** role.
This is needed for the networking setup and UMSA

🔗 3. Attach cloud shell to your project.

Open Cloud shell or navigate to shell.cloud.google.com.
Run the below command to set the project to cloud shell terminal:

```
gcloud config set project <enter your project id here>
```

🔗 1. Declare variables

We will use these throughout the lab.
Run the below in cloud shell copied to the project you selected-

```
PROJECT_ID= #Project ID
REGION= #Region to be used
```

```
#User Managed Service Account
UMSA="serverless-spark"
```

```
# Note: Lowercase letters, numbers, hyphens allowed. All network names must be unique within the project
VPC=
SUBNET=
FIREWALL=
```

2. Enable Google Dataproc and Vertex AI APIs

From cloud shell, run the below-

```
gcloud services enable dataproc.googleapis.com
gcloud services enable aiplatform.googleapis.com
```

3. Network Configuration

Run the commands below to create the networking entities required for the hands on lab.

3.1. Create a VPC

```
gcloud compute networks create $VPC \
  --subnet-mode=custom \
  --bgp-routing-mode=regional \
  --mtu=1500
```

b) List VPCs with:

```
gcloud compute networks list
```

c) Describe your network with:

```
gcloud compute networks describe $VPC
```

3.2. Create a subnet for Serverless Spark with private google access

```
gcloud compute networks subnets create $SUBNET \
  --network=$VPC \
  --range=10.0.0.0/24 \
  --region=$REGION \
  --enable-private-ip-google-access
```

3.3. Create firewall rules

Intra-VPC, allow all communication

```
gcloud compute firewall-rules create $FIREWALL \
  --project=$PROJECT_ID \
  --network=projects/$PROJECT_ID/global/networks/$VPC \
  --description="Allows connection from any source to any instance on the network using custom protocols." \
  --direction=INGRESS \
  --priority=65534 \
  --source-ranges=10.0.0.0/9 \
  --action=ALLOW --rules=all
```

4. Create a User Managed Service Account

```
gcloud iam service-accounts create $UMSA \
  --description="User Managed Service Account for Serverless Spark" \
  --display-name "Serverless Spark SA"
```

5. Grant IAM Permissions for UMSA

5.1.a. Basic role for UMSA

```
gcloud projects add-iam-policy-binding $PROJECT_ID \  
  --member serviceAccount:$UMSA@$PROJECT_ID.iam.gserviceaccount.com --role roles/viewer
```

5.1.b. Storage Admin role for UMSA

```
gcloud projects add-iam-policy-binding $PROJECT_ID \  
  --member serviceAccount:$UMSA@$PROJECT_ID.iam.gserviceaccount.com --role roles/storage.admin
```

5.1.c. Dataproc Editor role for UMSA

```
gcloud projects add-iam-policy-binding $PROJECT_ID \  
  --member serviceAccount:$UMSA@$PROJECT_ID.iam.gserviceaccount.com --role roles/dataproc.editor
```

5.1.d. Dataproc Worker role for UMSA

```
gcloud projects add-iam-policy-binding $PROJECT_ID \  
  --member serviceAccount:$UMSA@$PROJECT_ID.iam.gserviceaccount.com --role roles/dataproc.worker
```

5.1.e. BigQuery Data Editor role for UMSA

```
gcloud projects add-iam-policy-binding $PROJECT_ID \  
  --member serviceAccount:$UMSA@$PROJECT_ID.iam.gserviceaccount.com --role roles/bigquery.dataEditor
```

5.1.f. BigQuery User role for UMSA

```
gcloud projects add-iam-policy-binding $PROJECT_ID \  
  --member serviceAccount:$UMSA@$PROJECT_ID.iam.gserviceaccount.com --role roles/bigquery.user
```