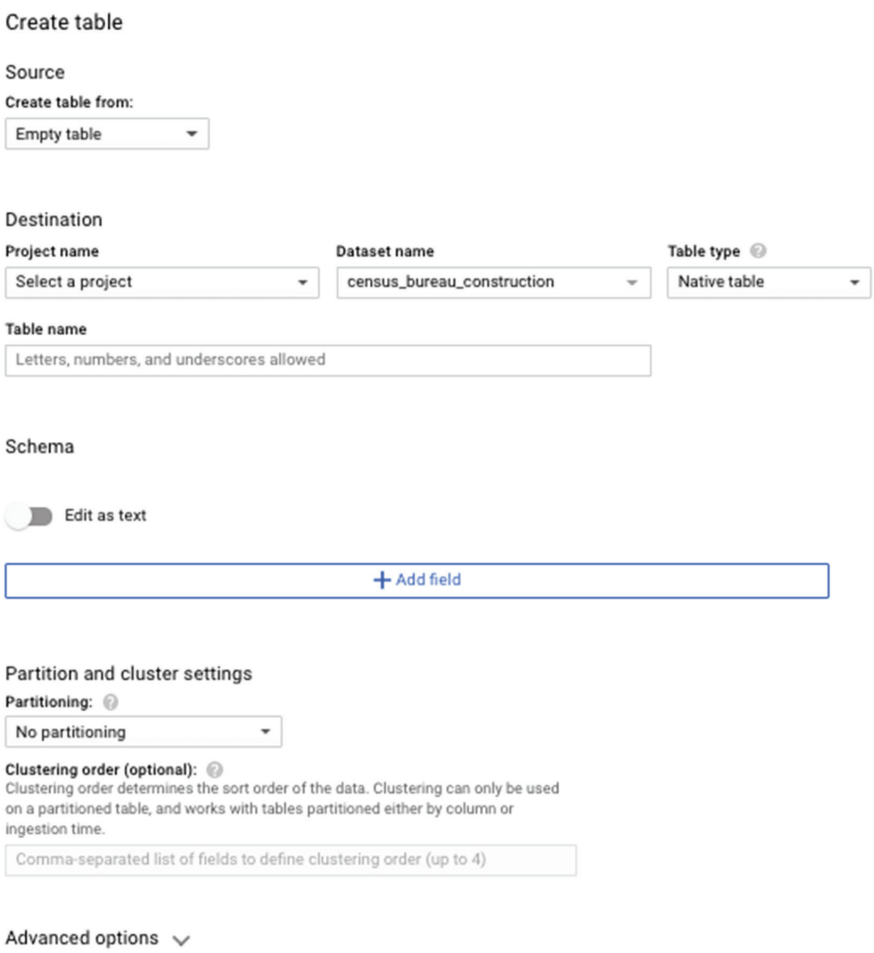
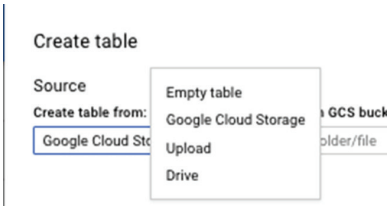
FIGURE 13.17 Creating a table in BigQuery



The Create Table From field indicates where to find the source data, if any. This provides a way to create a table based on data in an existing table, but defaults to an empty table (see Figure 13.18).

FIGURE 13.18 Data can be imported from multiple kinds of locations.



You will also need to specify the file format of the file that will be imported. The options include CSV, JSON, Avro, Parquet, PRC, and Cloud Datastore Backup (see Figure 13.19).

FIGURE 13.19 File format options for importing

Provide destination information, including project, dataset name, table type, and table name. Table type may be native type or external table. If the table is external, the data is kept in the source location, and only metadata about the table is stored in BigQuery. This is used when you have large data sets and do not want to load them all into BigQuery.

After specifying all parameters, click Create Table to create the table and load the data.

To load data from the command line, use the `bq load` command. Its structure is as follows:

```
bq load --autodetect --source_format=[FORMAT] [DATASET].[TABLE] [PATH_TO_SOURCE]
```

The `--autodetect` parameter has `bq load` automatically detect the table schema from the source file. An example command is as follows:

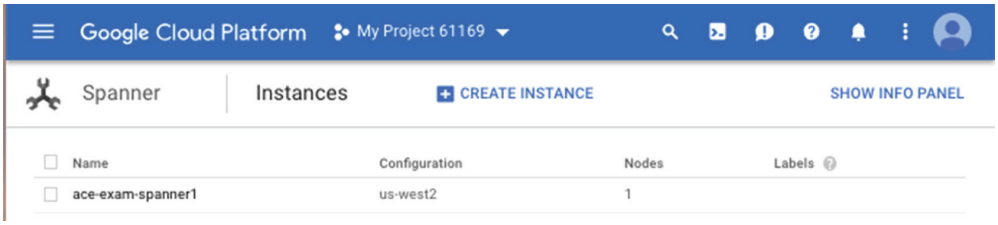
```
bq load --autodetect --source_format=CSV mydataset.mytable gs://ace-exam-biquery/
mydata.csv
```

Importing and Exporting Data: Cloud Spanner

Cloud Spanner users can import and export data using Cloud Console.

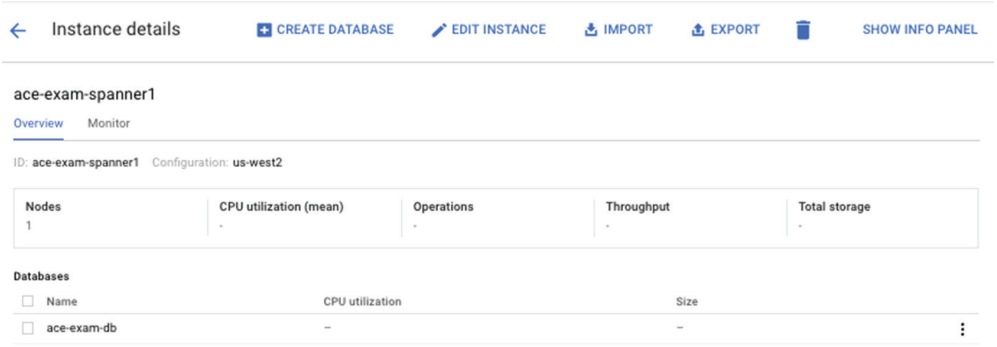
To export data from Cloud Spanner, navigate to the Cloud Spanner section of the console. You will see a list of Spanner instances, as shown in Figure 13.20.

FIGURE 13.20 Listing of Spanner instances



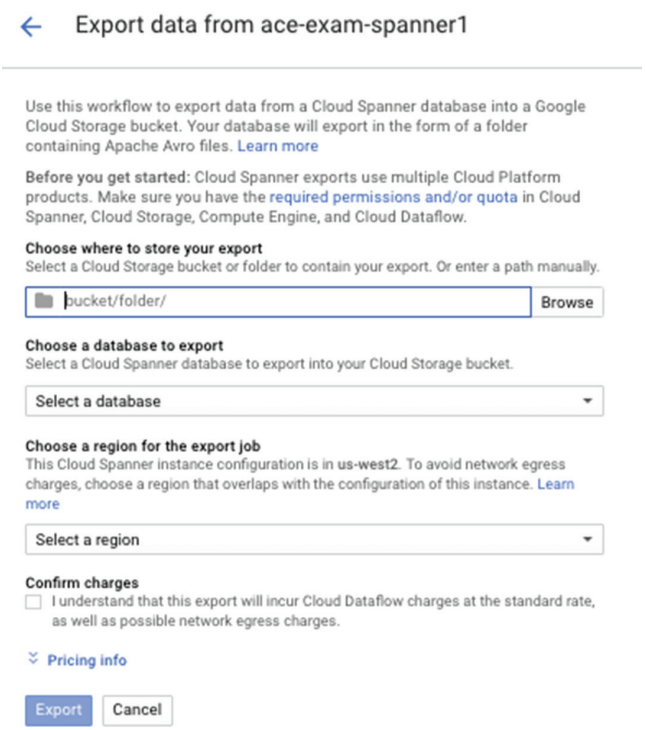
Click the name of the instance that is the source of data to export. This will show the Instance Detail page (see Figure 13.21).

FIGURE 13.21 Details of Spanner instance, with Import and Export tabs



Click Export to show the Export form, as shown in Figure 13.22. You will need to enter a destination bucket, the database to export, and a region to run the job. Notice, you need to confirm that there will be charges for running Cloud Dataflow, and there may be data egress charges for data sent between regions.

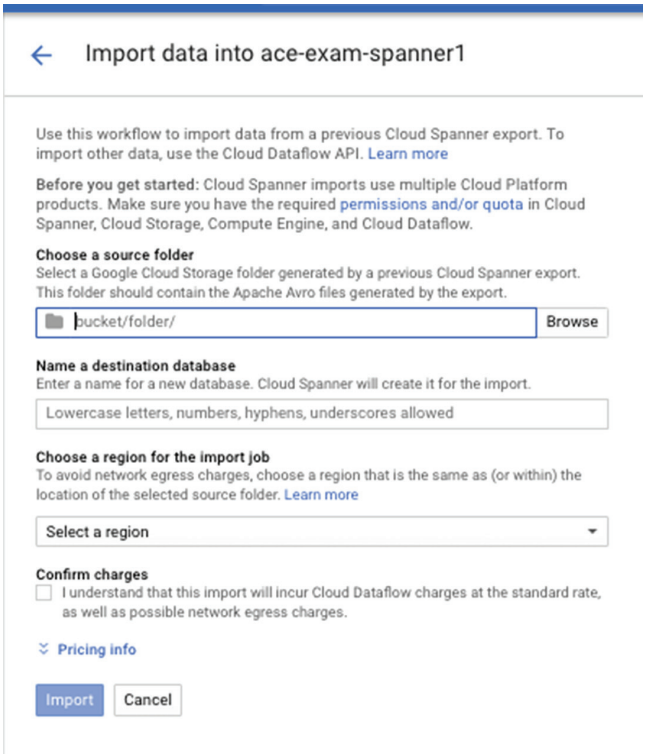
FIGURE 13.22 Export form for Cloud Spanner



To import data, click the Import tab to display the Import form (see Figure 13.23). You will need to specify a source bucket, a destination database, and a region to run a job.

Cloud Spanner does not have a `gcloud` command to export data, but you can use Dataflow to export data. The details of constructing Dataflow jobs is outside the scope of this section. For more details, see the Cloud Dataflow documentation at `https://cloud.google.com/dataflow/docs/`.

FIGURE 13.23 Import form for Cloud Spanner



Importing and Exporting Data: Cloud Bigtable

Unlike other GCP databases, Cloud Bigtable does not have an Export and Import option in the Cloud Console or in `gcloud`. You have two other options: using a Java application for importing and exporting or using the HBase interface to execute HBase commands. The HBase commands are not included in Google documentation and will not be described in detail here. For more information, see the HBase documentation at `https://hbase.apache.org/book.html/`.

To export a Bigtable table, you will need to download a JAR file, which is a compiled program for the Java VM. The command to download the file is as follows:

```
curl -f -O http://repo1.maven.org/maven2/com/google/cloud/bigtable/bigtable-beam-import/1.6.0/bigtable-beam-import-1.6.0-shaded.jar
```

To execute the export program, issue a command in the form of the following:

```
java -jar bigtable-beam-import-1.6.0-shaded.jar export \
    --runner=dataflow \
    --project=[PROJECT_ID] \
    --bigtableInstanceId=[INSTANCE_ID] \
    --bigtableTableId=[TABLE_ID] \
    --destinationPath=gs://[BUCKET_NAME]/[EXPORT_PATH] \
    --tempLocation=gs://[BUCKET_NAME]/[TEMP_PATH] \
    --maxNumWorkers=[10x_NUMBER_OF_NODES] \
    --zone=[DATAFLOW_JOB_ZONE]
```

You will need to specify the appropriate `project_id`, `table_id`, bucket information, a zone to run the Dataflow job, and the maximum number of workers to dedicate the export.

The following is an example of the `export` command:

```
java -jar bigtable-beam-import-1.6.0-shaded.jar export \
    --runner=dataflow \
    --project=my-project \
    --bigtableInstanceId=ace-exam-instance \
    --bigtableTableId=ace-exam-table1 \
    --destinationPath=gs://ace-exam-bucket1/ace-exam-table1 \
    --tempLocation=gs://my-export-bucket/jar-temp \
    --maxNumWorkers=30 \
                --zone=us-west2-a
```

To import data, you can use the same JAR file, but you will need to specify `import` instead of `export` in the command. There are some changes to the parameters, too, which are explained next. The `import` command structure is as follows:

```
java -jar bigtable-beam-import-1.6.0-shaded.jar import \
    --runner=dataflow \
    --project=[PROJECT_ID] \
    --bigtableInstanceId=[INSTANCE_ID] \
    --bigtableTableId=[TABLE_ID] \
    --sourcePattern='gs://[BUCKET_NAME]/[EXPORT_PATH]/part-*' \
    --tempLocation=gs://[BUCKET_NAME]/[TEMP_PATH] \
    --maxNumWorkers=[3x_NUMBER_OF_NODES] \
    --zone=[DATAFLOW_JOB_ZONE]
```

In addition to the parameters specified in the `export` command, the `import` command takes parameters to describe filename patterns that describe the files to import. Bigtable exports can be large enough to require multiple files to store all the data. You will also need to specify a bucket that can be used for temporary storage during the import.

The following is an example of the `import` command:

```
java -jar bigtable-beam-import-1.6.0-shaded.jar import \
    --runner=dataflow \
    --project=my-project \
    --bigtableInstanceId= ace-exam-instance \
    --bigtableTableId= ace-exam-table1 \
    --sourcePattern='gs://my-export-bucket/my-table/part-*' \
    --tempLocation=gs://my-export-bucket/jar-temp \
    --maxNumWorkers=10 \
    --zone=us-west2-a
```

Importing and Exporting Data: Cloud Dataproc

Cloud Dataproc is not a database like Cloud SQL or Bigtable; rather, it is a data analysis platform. These platforms are designed more for data manipulation, statistical analysis, machine learning, and other complex operations than for data storage and retrieval. Cloud Dataproc is not designed to be a persistent store of data. For that you should use Cloud Storage or persistent disks to store the data files you want to analyze.

Cloud Dataproc does have Import and Export commands to save and restore cluster configuration data. These commands are available, in beta, using `gcloud`.

To ensure you have access to beta commands in `gcloud`, issue the following command:

```
gcloud components install beta
```

The command to export a Dataproc cluster configuration is as follows:

```
gcloud beta dataproc clusters export [CLUSTER_NAME] --destination=[PATH_TO_EXPORT_FILE]
```

An example is as follows:

```
gcloud beta dataproc clusters export ace-exam-dataproc-cluster
--destination=gs://ace-exam-bucket1/mydataproc.yaml
```

To import a configuration file, use the `import` command:

```
gcloud beta dataproc clusters import [SOURCE_FILE]
```

For example, to import the file created in the previous `export` example, you could use the following:

```
gcloud beta dataproc clusters import gs://ace-exam-bucket1/mydataproc.yaml
```

Importing and exporting data are common operations. GCP provides console and command-line tools for most database services. There are also beta commands for exporting and importing cluster configuration data for Dataproc.

Streaming Data to Cloud Pub/Sub

So far in this chapter you have spent most of your time on moving data into and around Cloud Storage, along with importing and exporting data to databases. Let's now turn your attention to working with Cloud Pub/Sub, the messaging queue.

As a Cloud Engineer, you may need to create message queues for application developers. Although developers will most likely write services that use Pub/Sub, Cloud Engineers should be able to test Pub/Sub topics and subscriptions. We discussed how to create message queues in Chapter 12. Here our focus will be on creating messages on topics and receiving those messages through subscriptions.

The `gcloud pubsub` commands you will use are `create`, `publish`, and `pull`. To create a topic, you use the following command:

```
gcloud pubsub topics create [TOPIC_NAME]
```

The command to create a subscription is as follows:

```
gcloud pubsub subscriptions create --topic [TOPIC_NAME] [SUBSCRIPTION_NAME]
```

For example, to create a topic called `ace-exam-topic1` and a subscription to that topic called `ace-exam-sub1`, you can use these commands:

```
gcloud pubsub topics create ace-exam-topic1
gcloud pubsub subscriptions create --topic=ace-exam-topic1 ace-exam-sub1
```

Now, to test whether the message queue is working correctly, you can send data to the topic using the following command:

```
cloud pubsub topics publish [TOPIC_NAME] --message [MESSAGE]
```

and then read that message from the subscription using the following:

```
gcloud pubsub subscriptions pull --auto-ack [SUBSCRIPTION_NAME]
```

To write a message to the topic and read it from the subscription you just created, you can use the following:

```
gcloud pubsub topics publish ace-exam-topic1 --message "first ace exam message"
gcloud pubsub subscriptions pull --auto-ack ace-exam-sub1
```

Real World Scenario

Decoupling Services Using Message Queues

One of the challenges with distributed systems is that sometimes one service cannot keep up with the inflow of data. This can create a backlog in services that depend on the lagging service.

For example, a sudden spike in traffic on a retail site may put a high load on an inventory tracking service, which updates inventory as customers add or remove items from their baskets. The inventory program may be slow to respond to a service that added an item to the cart. If that service is waiting for a response from the inventory service, it too will be delayed. This kind of synchronous communication is problematic when distributed systems are under load.

A better option is to decouple the direct connection between services. For example, the user interface could write a message to a Pub/Sub topic each time an item is added or removed from a customer's basket. The inventory management service can subscribe to this topic and update the inventory system as new messages come in. If the inventory system slows down, it will not affect the user interface because it is writing to a Pub/Sub topic, which can scale along with the load generated by the user interface.

Summary

In this chapter, you looked at the different ways you can load data into storage, database, and message queue systems. Cloud Storage is organized around objects in buckets. The `gsutil` command and Cloud Console can be used to upload data as well as move it between buckets. You saw that the `gsutil cp` command can be used to copy files between Cloud Storage and VMs.

The database services provide import and export utilities. Some, such as Cloud SQL and BigQuery, make these services available from both Cloud Console and the command line. Others, like Bigtable and Cloud Dataproc, have command-line options only.

Cloud Pub/Sub can be used to decouple applications and improve resiliency to spikes in load. You saw how to create a topic and subscriptions and how to push data to the message queue, where it can be read by subscribers.

Know that Cloud Spanner uses the Dataflow service for importing and exporting. There can be additional charges when using Dataflow and moving data between regions. There is no `gcloud` command for importing or exporting Cloud Spanner databases.

Exam Essentials

Know how to load data into and move data around Cloud Storage. Cloud Storage is widely used for a variety of use cases, including long-term storage and archiving, file transfers, and data sharing. Understand the structure of `gsutil` commands, which is different from `gcloud`. `gsutil` commands start with `gsutil` followed by an operation, such as `copy` or `make bucket`. Be sure to know the syntax of the copy (`cp`), move (`mv`), and make bucket (`mb`) commands. You can copy files from Cloud Storage to VMs, and vice versa. Also, know that the `gsutil acl ch -u` command is used to change permissions on objects.

Understand how import and export work with Cloud SQL. Importing and exporting data from databases are common operations. You can use the `gsutil acl ch` command to change permissions on a Cloud Storage bucket. You can perform imports and exports from the console and from the command line.

Know that you can export entities from a Cloud Datastore. Exports and imports are done at the level of namespaces. There isn't a Cloud Console option for exporting and importing from Datastore.

Understand how to export and import data from BigQuery. BigQuery has a range of options for the source of data to import. Data can be compressed when exported to save on space. BigQuery can export data in multiple formats, including CSV, JSON, and Avro. Know that the `bq` command is used for importing and exporting from the command line.

Remember that Bigtable and Cloud Dataproc are different from other import and export functions. Bigtable does not have a console or command-line feature to import or export data. A Java program is run from the command line to import or export data from Bigtable. Cloud Dataproc is different in that it is not designed as a persistent data store. It is a data analysis tool. When you export from Dataproc, you are exporting the cluster configuration, not data in the cluster.

Know that Pub/Sub is used to send messages between services. Pub/Sub allows for greater resiliency to fluctuations in load. If one service lags, its work can accumulate in a Pub/Sub queue without forcing the service that generates that data to wait.

Review Questions

You can find the answers in the Appendix.

1. Which of the following commands is used to create buckets in Cloud Storage?
   A. `gcloud storage buckets create`
   B. `gsutil storage buckets create`
   C. `gsutil mb`
   D. `gcloud mb`

2. You need to copy files from your local device to a bucket in Cloud Storage. What command would you use? Assume you have Cloud SDK installed on your local computer.
   A. `gsutil copy`
   B. `gsutil cp`
   C. `gcloud cp`
   D. `gcloud storage objects copy`

3. You are migrating a large number of files from a local storage system to Cloud Storage. You want to use the Cloud Console instead of writing a script. Which of the following Cloud Storage operations can you perform in the console?
   A. Upload files only
   B. Upload folders only
   C. Upload files and folders
   D. Compare local files with files in the bucket using the `diff` command

4. A software developer asks for your help exporting data from a Cloud SQL database. The developer tells you which database to export and which bucket to store the export file in, but hasn't mentioned which file format should be used for the export file. What are the options for the export file format?
   A. CSV and XML
   B. CSV and JSON
   C. JSON and SQL
   D. CSV and SQL

5. A database administrator has asked for an export of a MySQL database in Cloud SQL. The database administrator will load the data into another relational database and would to do it with the least amount of work. Specifically, the loading method should not require the database administrator to define a schema. What file format would you recommend for this task?
   A. SQL
   B. CSV
   C. XML
   D. JSON

6. Which command will export a MySQL database called `ace-exam-mysql1` to a file called `ace-exam-mysql-export.sql` in a bucket named `ace-exam-buckete1`?
   - **A.** `gcloud storage export sql ace-exam-mysql1 gs://ace-exam-buckete1/ace-exam-mysql-export.sql \ --database=mysql`
   - **B.** `gcloud sql export ace-exam-mysql1 gs://ace-exam-buckete1/ace-exam-mysql-export.sql \ --database=mysql`
   - **C.** `gcloud sql export sql ace-exam-mysql1 gs://ace-exam-buckete1/ace-exam-mysql-export.sql \ --database=mysql`
   - **D.** `gcloud sql export sql ace-exam-mysql1 gs://ace-exam-mysql-export.sql/ ace-exam-buckete1/ \ --database=mysql`

7. As part of a compliance regimen, your team is required to back up data from your Datastore database to an object storage system. Your data is stored in the default namespace. What command would you use to export the default namespace from Datastore to a bucket called `ace-exam-bucket1`?
   - **A.** `gcloud datastore export --namespaces="(default)" gs://ace-exam-bucket1`
   - **B.** `gcloud datastore export --namespaces="(default)" ace-exam-bucket1`
   - **C.** `gcloud datastore dump --namespaces="(default)" gs://ace-exam-bucket1`
   - **D.** `gcloud datastore dump --namespaces="(default)" ace-exam-bucket1`

8. As required by your company's policy, you need to back up your Datastore database at least once per day. An auditor is questioning whether or not Datastore export is sufficient. You explain that the Datastore export command produces what outputs?
   - **A.** A single entity file
   - **B.** A metadata file
   - **C.** A metadata file and a folder with the data
   - **D.** A metadata file, an entity file, and a folder with the data

9. Which of the following file formats is not an option for an export file when exporting from BigQuery?
   - **A.** CSV
   - **B.** XML
   - **C.** Avro
   - **D.** JSON

10. Which of the following file formats is not supported when importing data into BigQuery?
    - **A.** CSV
    - **B.** Parquet
    - **C.** Avro
    - **D.** YAML

11. You have received a large data set from an Internet of Things (IoT) system. You want to use BigQuery to analyze the data. What command-line command would you use to make data available for analysis in BigQuery?

    A. `bq load --autodetect --source_format=[FORMAT] [DATASET].[TABLE] [PATH_TO_SOURCE]`

    B. `bq import --autodetect --source_format=[FORMAT] [DATASET].[TABLE] [PATH_TO_SOURCE]`

    C. `gloud BigQuery load --autodetect --source_format=[FORMAT] [DATASET].[TABLE] [PATH_TO_SOURCE]`

    D. `gcloud BigQuery load --autodetect --source_format=[FORMAT] [DATASET].[TABLE] [PATH_TO_SOURCE]`

12. You have set up a Cloud Spanner process to export data to Cloud Storage. You notice that each time the process runs you incur charges for another GCP service, which you think is related to the export process. What other GCP service might be incurring charges during the Cloud Spanner export?

    A. Dataproc

    B. Dataflow

    C. Datastore

    D. `bq`

13. As a developer on a project using Bigtable for an IoT application, you will need to export data from Bigtable to make some data available for analysis with another tool. What would you use to export the data, assuming you want to minimize the amount of effort required on your part?

    A. A Java program designed for importing and exporting data from Bigtable

    B. `gcloud bigtable table export`

    C. `bq bigtable table export`

    D. An import tool provided by the analysis tool

14. You have just exported from a Dataproc cluster. What have you exported?

    A. Data in Spark DataFrames

    B. All tables in the Spark database

    C. Configuration data about the cluster

    D. All tables in the Hadoop database

15. A team of data scientists has requested access to data stored in Bigtable so that they can train machine learning models. They explain that Bigtable does not have the features required to build machine learning models. Which of the following GCP services are they most likely to use to build machine learning models?

    A. Datastore

    B. Dataflow

    C. Dataproc

    D. DataAnalyze

16. The correct command to create a Pub/Sub topic is which of the following?

A. `gcloud pubsub topics create`

B. `gcloud pubsub create topics`

C. `bq pubsub create topics`

D. `cbt pubsub topics create`

17. Which of the following commands will create a subscription on the topic `ace-exam-topic1`?

A. `gcloud pubsub create --topic=ace-exam-topic1 ace-exam-sub1`

B. `gcloud pubsub subscriptions create --topic=ace-exam-topic1`

C. `gcloud pubsub subscriptions create --topic=ace-exam-topic1 ace-exam-sub1`

D. `gsutil pubsub subscriptions create --topic=ace-exam-topic1 ace-exam-sub1`

18. What is one of the direct advantages of using a message queue in distributed systems?

A. It increases security.

B. It decouples services, so if one lags, it does not cause other services to lag.

C. It supports more programming languages.

D. It stores messages until they are read by default.

19. To ensure you have installed beta `gcloud` commands, which command should you run?

A. `gcloud components beta install`

B. `gcloud components install beta`

C. `gcloud commands install beta`

D. `gcloud commands beta install`

20. What parameter is used to tell BigQuery to automatically detect the schema of a file on import?

A. `--autodetect`

B. `--autoschema`

C. `--detectschema`

D. `--dry_run`

21. The compression options deflate and snappy are available for what file types when exporting from BigQuery?

A. Avro

B. CSV

C. XML

D. Thrift

Chapter

14

Networking in the Cloud: Virtual Private Clouds and Virtual Private Networks

THIS CHAPTER COVERS THE FOLLOWING OBJECTIVES OF THE GOOGLE ASSOCIATE CLOUD ENGINEER CERTIFICATION EXAM:

- ✓ **2.4 Planning and configuring network resources**
- ✓ **4.5 Managing networking resources**

Official Google Cloud Certified: Associate Cloud Engineer: Study Guide
By Dan Sullivan

In this chapter we turn our attention to networking, starting with virtual private clouds (VPCs). You will learn how to create VPCs with default and custom subnets. You'll learn about creating custom network configurations in Compute Engine for cases when default network configurations do not meet your needs. Next, we will show how to configure firewall rules and create virtual private networks (VPNs).

Creating a Virtual Private Cloud with Subnets

VPCs are software versions of physical networks that link resources in a project. GCP automatically creates a VPC when you create a project. You can create additional VPCs and modify the VPCs created by GCP.

VPCs are global resources, so they are not tied to a specific region or zone. Resources, such as Compute Engine virtual machines (VMs) and Kubernetes Engine clusters, can communicate with each other, assuming traffic is not blocked by a firewall rule.

VPCs contain subnetworks, call *subnets*, which are regional resources. Subnets have a range of IP addresses associated with them. Subnets provide private internal addresses. Resources use these addresses to communicate with each other and with Google APIs and services.

In addition to VPCs associated with projects, you can create a shared VPC within an organization. The shared VPC is hosted in a common project. Users in other projects who have sufficient permissions can create resources in the shared VPC. You can also use VPC peering for interproject connectivity, even if an organization is not defined.

In this section, you will create a VPC with subnets using Cloud Console and `gcloud`, and then turn your attention to creating a shared VPC.

Creating a Virtual Private Cloud with Cloud Console

To create a VPC in Cloud Console, navigate to the VPC page, as shown in Figure 14.1.

Clicking Create VPC opens the form to create a VPC, as shown in Figure 14.2. Figure 14.2 shows that you can assign a name and description to a new VPC. It also shows a list of subnets that will be created in the VPC. When a VPC is created, subnets are created in each region. GCP chooses a range of IP addresses for each subnet when creating an auto mode network.

FIGURE 14.1 The VPC section of the Cloud Console

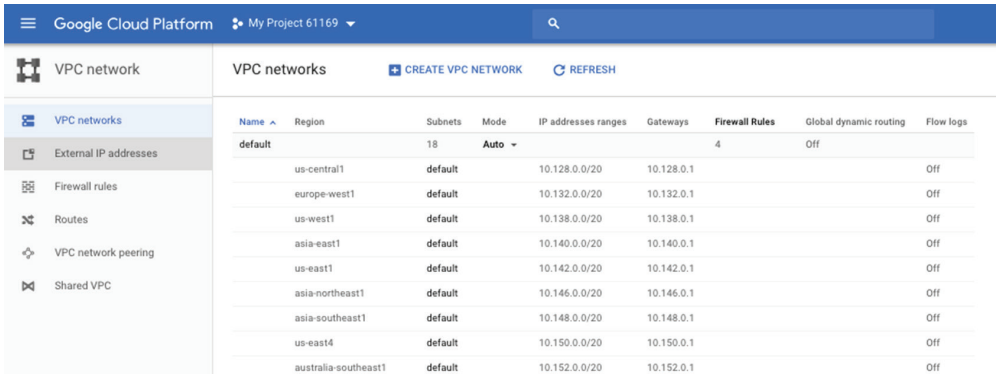
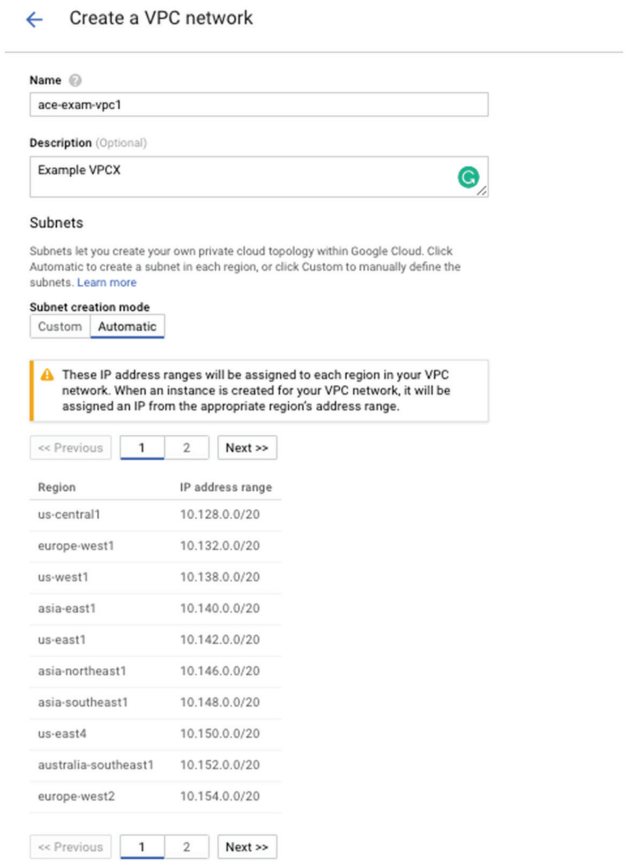


FIGURE 14.2 Form to create a VPC in Cloud Console, part 1



Alternatively, you can create one or more custom subnets by selecting the Custom tab in the Subnet section (Figure 14.3). This displays another form that allows you to specify a region and an IP address range. The IP range is specified in classless inter-domain routing (CIDR) notation. (See the sidebar "CIDR Notation Overview" on page 342 for details on how to specify IP address using that notation.) You can turn off Private Google Access. That allows VMs on the subnet to access Google services without assigning an external IP address to the VM. You can also turn on logging of network traffic by setting the Flow Logs option to on.

FIGURE 14.3 Creating a custom subnet

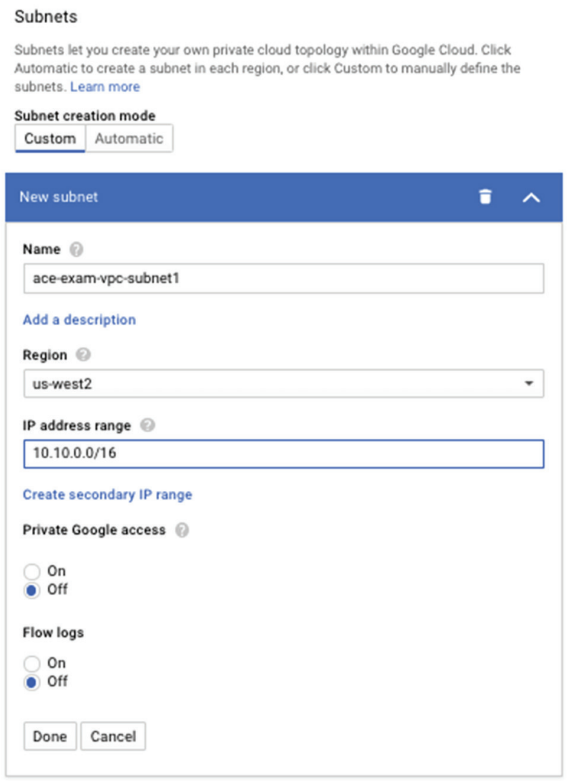
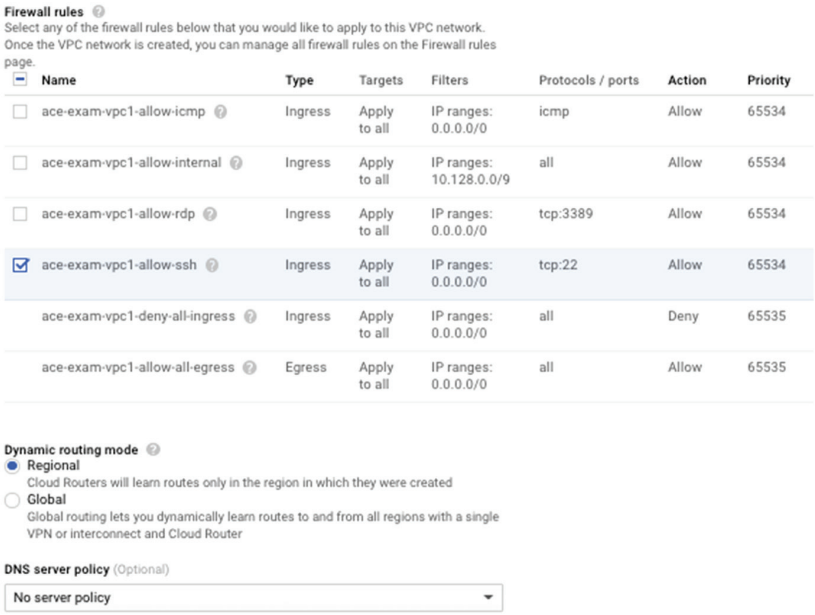


Figure 14.4 shows the second part of the VPC form, which includes firewall rules, dynamic routing setting, and a DNS server policy. The Firewall Rules section lists rules that can be applied to the VPC. In the example in Figure 14.4, the rule allows ingress, which is incoming TCP traffic on port 22, to allow for SSH access. The IP range of 0.0.0.0/0 allows traffic from all source IP addresses.

FIGURE 14.4 Form to create a VPC in Cloud Console, part 2

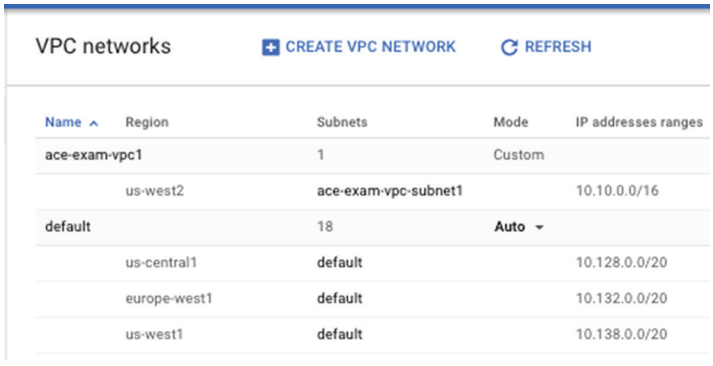


The dynamic routing option determines what routes are learned. Regional routing will have Google Cloud Routers learn routes within the region. Global routing will enable Google Cloud Routers to learn routes on all subnetworks in the VPC.

The optional DNS server policy lets you choose a DNS policy that enables DNS name resolution provided by GCP or makes changes to name resolution order. (See Chapter 15 for more details.)

Once you have specified the parameters and created a VPC, it will appear in the VPC listing and show information about the VPC and its subnets, as shown in Figure 14.5.

FIGURE 14.5 Listing of VPCs and subnets



Creating a Virtual Private Cloud with gcloud

The gcloud command to create a VPC is `gcloud compute networks create`. For example, to create a VPC in the default project with automatically generated subnets, you would use the following command:

```
gcloud compute networks create ace-exam-vpc1 --subnet-mode=auto
```

You can also configure custom subnets by creating a VPC network specifying the custom option and then creating subnets in that VPC. The first command to create a custom VPC called ace-exam-vpc1 is as follows:

```
gcloud compute networks create ace-exam-vpc1 --subnet-mode=custom
```

Next, you can create a subnet using the `gcloud compute networks subnet create` command. This command requires that you specify a VPC, the region, and the IP range. You can optionally turn on the Private Google Access and Flow Logs settings by adding the appropriate flags.

Here is an example command to create a subnet called `ace-exam-vpc-subnet1` in the ace-exam-vpc1 VPC. This subnet is created in the us-west2 region with an IP range of 10.10.0.0/16. The Private IP Access and Flow Logs settings are turned on.

```
gcloud beta compute networks subnets create ace-exam-vpc-subnet1 --network=ace-
exam-vpc1 --region=us-west2 --range=10.10.0.0/16 --enable-private-ip-google-
access --enable-flow-logs
```

Understanding CIDR Notation

When you specify ranges of IP addresses, you use something called *classless inter-domain routing* (CIDR). The name stems from early IP networks that were defined into three fixed classes: A, B, and C. A classless network address structure was created to overcome the limitations of a class-based routing structure, particularly the lack of flexibility in creating different-sized subnets.

CIDR uses variable-length subnet masking (VLSM) to allow network administrators to define networks with the number of addresses that they need, not the fixed number that were allocated to the older class model interdomain routine.

CIDR addresses consist of two sets of numbers, a network address for identifying a subnet and a host identifier. These numbers are written out using CIDR notation, which consists of a network address and a network mask. Example network addresses, according to the RFC1918 specification are:

10.0.0.0

172.16,0.0

192.168.0.0

CIDR notation adds a slash (/) and a number indicating how many bits of an IP address to allocate to the network mask, which determines which addresses are within the block of the address and which are not.

For example, 192.168.0.0/16 means that 16 bits of the 32 bits of an IP address are used to specify the network, and 16 bits are used to specify the host address. With 16 bits, you can create 2^{16} or 65,536 addresses.

The CIDR block 172.16.0.0/12 indicates that 12 bits are used for specifying the network, and 20 bits are used to specify host addresses. With 20 bits, you can create up to 1,048,576 addresses. In general, the smaller the number after the slash, the more addresses are available. You can experiment with CIDR block options using a CIDR calculator such as the one at `www.subnet-calculator.com/cidr.php`.

Creating a Shared Virtual Private Cloud Using gcloud

If you want to create a shared VPC, you can use the `gcloud` command `gcloud compute shared-vpc`.

Before executing commands to create a shared VPC, you will need to assign an org member the Shared VPC Admin role at the organization level or the folder level. To assign the Shared VPC Admin role, which uses the descriptor `roles/compute.xpnAdmin`, you would issue this command:

```
gcloud organizations add-iam-policy-binding [ORG_ID]
    --member='user:[EMAIL_ADDRESS]'
    --role="roles/compute.xpnAdmin"
```

`[ORG_ID]` is the organization identifier of the organization using the policy. You can find an organization ID with the command `gcloud organizations list`. If you prefer to assign the Shared VPC Admin role to a folder, you can use this command:

```
gcloud beta resource-manager folders add-iam-policy-binding [FOLDER_ID]
--member='user:[EMAIL_ADDRESS]'
--role="roles/compute.xpnAdmin"
```

`[FOLDER_ID]` is the identifier of the folder of the policy. You can get folder IDs by using this command:

```
gcloud beta resource-manager folders list --organization=[ORG_ID]
```

For more on roles and privileges, see Chapter 17.

Once you have set the Shared VPC Admin role at the organization level, you can issue the `shared-vpc` command:

```
gcloud compute shared-vpc enable [HOST_PROJECT_ID]
```

If you are sharing the VPC at the folder level, use this command:

```
gcloud beta compute shared-vpc enable [HOST_PROJECT_ID]
```

Now that the shared VPC is created, you can associate projects using the `gcloud compute shared-vpc associate-projects` command. At the organization level, you can use this command:

```
gcloud compute shared-vpc associated-projects add [SERVICE_PROJECT_ID] \
    --host-project [HOST_PROJECT_ID]
```

At the folder level, the command to associate folders is as follows:

```
gcloud beta compute shared-vpc associated-projects add [SERVICE_PROJECT_ID] \
    --host-project [HOST_PROJECT_ID]
```

Alternatively, VPC peering can be used for interproject traffic when an organization does not exist. VPC peering is implemented using the `gcloud compute networks peerings create` command. For example, you peer two VPCs by specifying peerings on each network. Here's an example:

```
gcloud compute networks peerings create peer-ace-exam-1 \
    --network ace-exam-network-A \
    --peer-project ace-exam-project-B \
    --peer-network ace-exam-network-B \
    --auto-create-routes
```

And then create a peering on the other network using:

```
gcloud compute networks peerings create peer-ace-exam-1 \
     --network ace-exam-network-B \
     --peer-project ace-exam-project-A \
     --peer-network ace-exam-network-A \
     --auto-create-routes
```

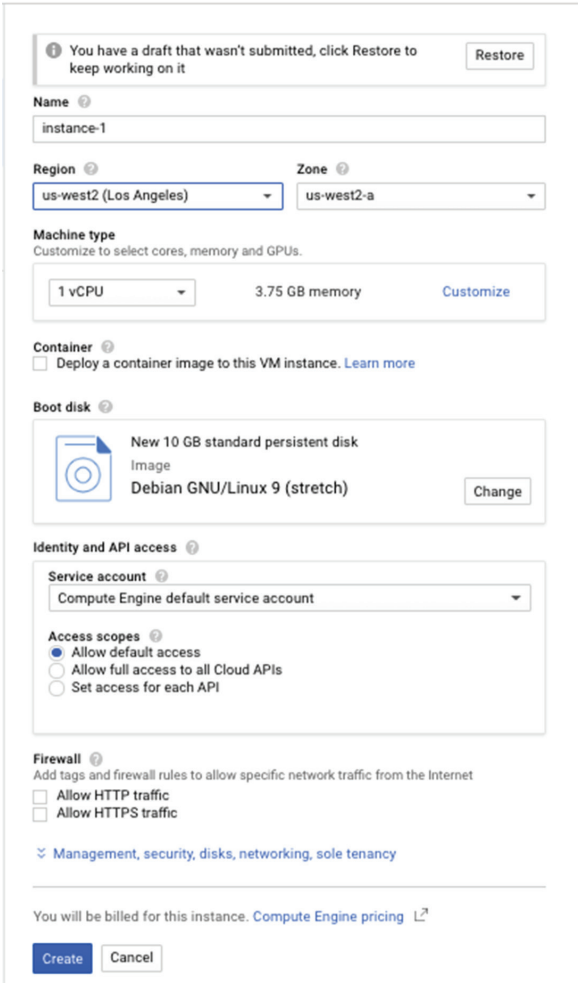
This peering will allow private traffic to flow between the two VPCs.

Deploying Compute Engine with a Custom Network

You can deploy a VM with custom network configurations using the console and the command line.

Navigate to the Compute Engine section of the console and open the Create Instance form, like that shown in Figure 14.6.

FIGURE 14.6 Preliminary form to create an instance in Cloud Console



Click Management, Security, Disks, Networking, Sole Tenancy to expand the optional forms and then click the Networking tab to display a form similar to Figure 14.7.

Note that in this form, you can set network tags. Click Add Network Interface to display a form like that shown in Figure 14.8. In this form you can choose a custom network. In this example, we are choosing ace-exam-vpc1, which we created earlier in the chapter. We also select a subnet in that form.

FIGURE 14.7 Networking configuration form

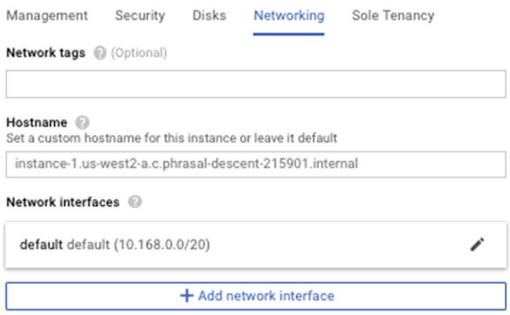
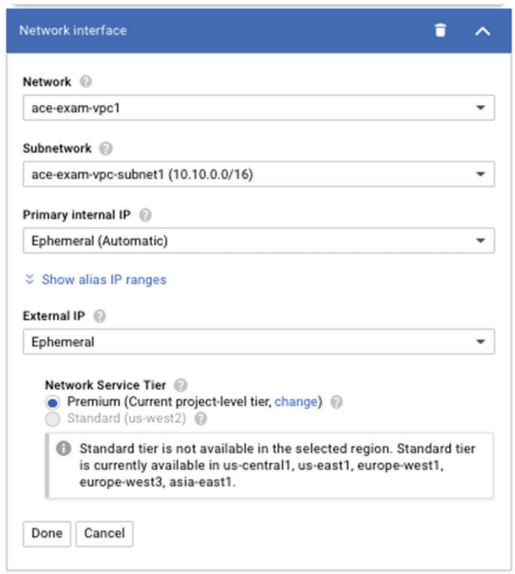


FIGURE 14.8 Form to add a custom network interface



From this form, you can also specify a static IP address or choose a custom ephemeral address using the Primary Internal IP setting. The External IP drop-down allows you to have an ephemeral external IP.

You can also create an instance to run in a particular subnet using the `gcloud compute instances create` command with Subnet and Zone parameters.

```
gcloud compute instances create [INSTANCE_NAME] --subnet [SUBNET_NAME] --zone
[ZONE_NAME]
```

Creating Firewall Rules for a Virtual Private Cloud

Firewall rules are defined at the network level and used to control the flow of network traffic to VMs.

Firewall rules allow or deny a kind of traffic on a port; for example, a rule may allow TCP traffic to port 22. They also are applied to traffic in one direction, either incoming (ingress) or outgoing (egress) traffic. It is important to note that the firewall is stateful which means if traffic is allowed in one direction and a connection established, it is allowed in the other direction. Firewalls rulesets are stateful so if a connection is allowed, like establishing a SSH connection on port 22, then all later traffic matching this rule is permitted as long as the connection is active. An active connection is one with at least one packet exchanged every ten minutes.

Structure of Firewall Rules

Firewall rules consist of several components:

- **Direction:** Either ingress or egress.
- **Priority:** Highest-priority rules are applied; any rule with a lower priority that matches are not applied. Priority is specified by an integer from 0 to 65535. 0 is the highest priority, and 65535 is lowest.
- **Action:** Either allow or deny. Only one can be chosen.
- **Target:** An instance to which the rule applies. Targets can be all instances in a network, instances with particular network tags, or instances using a specific service account.
- **Source/destination:** Source applies to ingress rules and specifies source IP ranges, instances with particular network tags, or instances using a particular service account. You can also use combinations of source IP ranges and network tags and combinations of source IP ranges and service accounts used by instances. The IP address 0.0.0.0/0 indicates any IP address. The Destination parameter uses only IP ranges.
- **Protocol and port:** A network protocol such as TCP, UDP, or ICMP and a port number. If no protocol is specified, then the rule applies to all protocols.
- **Enforcement status:** Firewall rules are either enabled or disabled. Disabled rules are not applied even if they match. Disabling is sometimes used to troubleshoot problems with traffic getting through when it should not or not getting through when it should.

All VPCs start with two implied rules: One allows egress traffic to all destinations (IP address 0.0.0.0/0), and one denies all incoming traffic from any source (IP address 0.0.0.0/0). Both implied rules have priority 65535, so you can create other rules with higher deny or allow traffic as you need. You cannot delete an implied rule.

When a VPC is automatically created, the default network is created with four network rules. These rules allow the following:

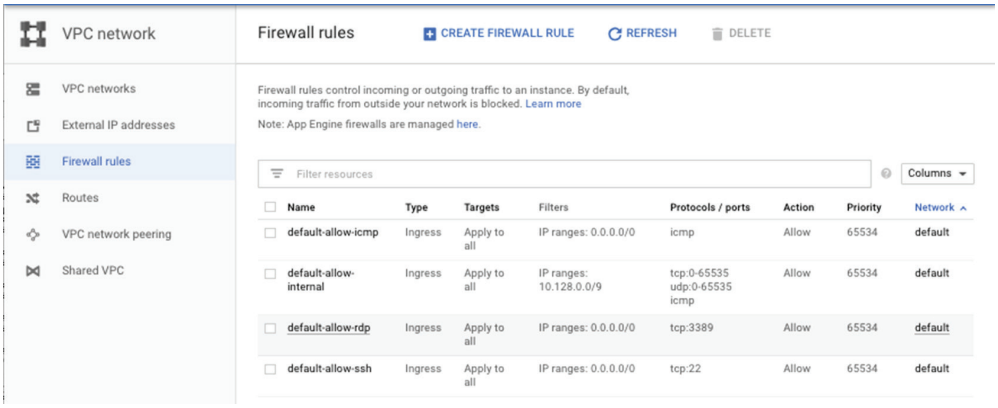
- Incoming traffic from any VM instance on the same network
- Incoming TCP traffic on port 22, allowing SSH
- Incoming TCP traffic on port 3389, allowing Microsoft Remote Desktop Protocol (RDP)
- Incoming Internet Control Message Protocol (ICMP) from any source on the network

The default rules all have priority 65534.

Creating Firewall Rules Using Cloud Console

To create or edit firewall rules, navigate to the VPC section of the console and select the Firewall option from the VPC menu (see Figure 14.9).

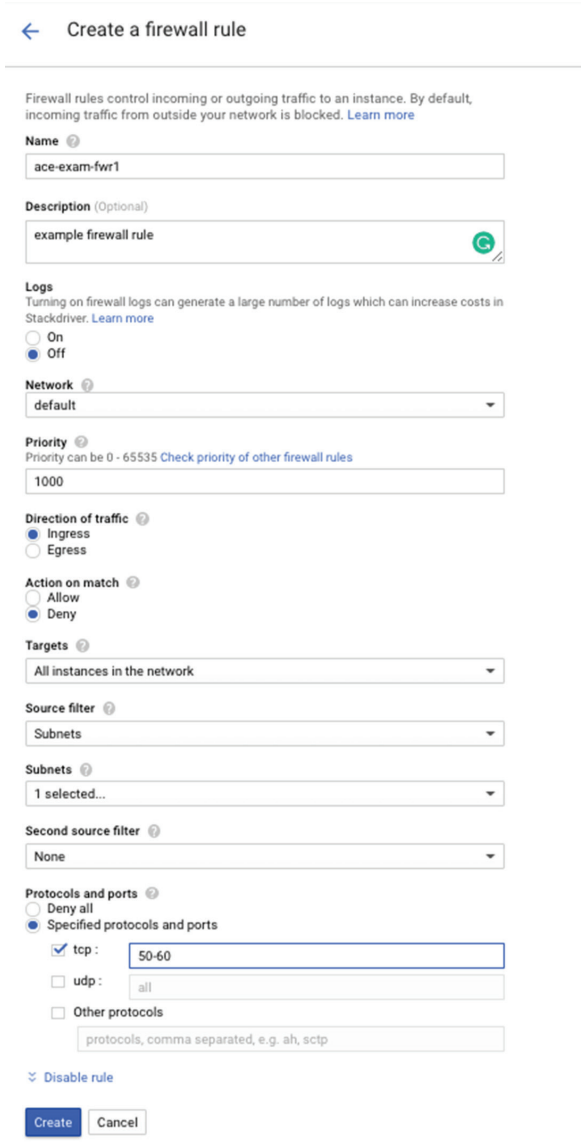
FIGURE 14.9 List of firewall rules in the VPC section of Cloud Console



Click Create Firewall Rule to create a new firewall rule. This shows a form similar to Figure 14.10.

In this form, you specify a name and description of the firewall rule. You can choose to turn logging on or off. If it is on, logging information will be captured in Stackdriver. (See Chapter 18 for more on Stackdriver logging.) You also need to specify the network in the VPC to apply the rule to.

FIGURE 14.10 Create firewall rule form

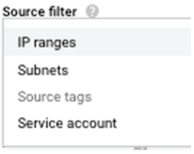


Next, you will need to specify a priority, direction, action, targets, and sources. Priority can be integers in the range from 0 to 65535. Direction can be ingress or egress. Action can be allow or deny. Targets are chosen from a drop-down list; the options are shown in Figure 14.11.

FIGURE 14.11 List of target types

If you choose tags or service accounts, you will be able to specify the tags or the name of the service account. You can also specify source filters as either IP ranges, subnets, source tags, or service accounts. GCP allows a second source filter if you'd like to use a combination of conditions. A list of source filters is shown in Figure 14.12.

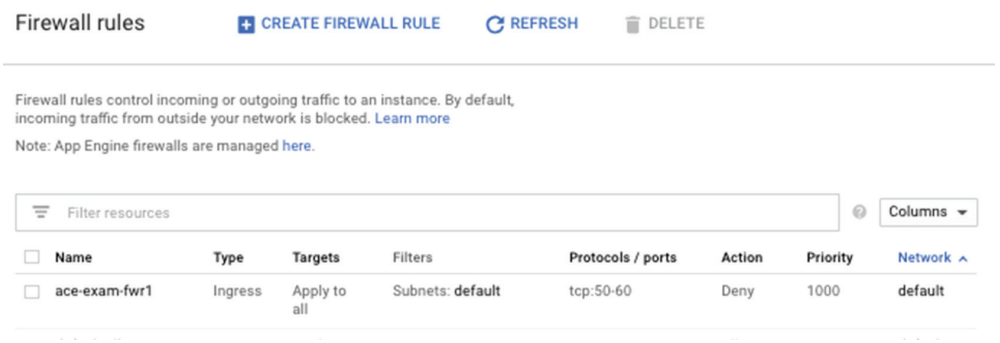
FIGURE 14.12 List of source filter types



Finally, you specify protocol and ports by choosing between the Allow All and Specified Protocols and Ports options. If you choose the latter, you can specify protocols and ports.

Figure 14.13 shows the listing of the firewall rule created using the parameters specified in Figure 14.10.

FIGURE 14.13 Listing of firewall rule created using earlier configuration



Creating Firewall Rules Using gcloud

The command for working with firewall rules from the command line is `gcloud compute firewall-rules`. With this command, you can create, delete, describe, update, and list firewall rules.

There are a number of parameters used with `gcloud compute firewall-rules create`:

- `--action`
- `--allow`
- `--description`
- `--destination-ranges`
- `--direction`
- `--network`
- `--priority`
- `--source-ranges`
- `--source-service-accounts`
- `--source-tags`
- `--target-service-accounts`
- `--target-tags`

For example, to allow all TCP traffic on ports 20000 to 25000, use this:

```
gcloud compute firewall-rules create ace-exam-fwr2 --network ace-exam-vpc1
--allow tcp:20000-25000
```

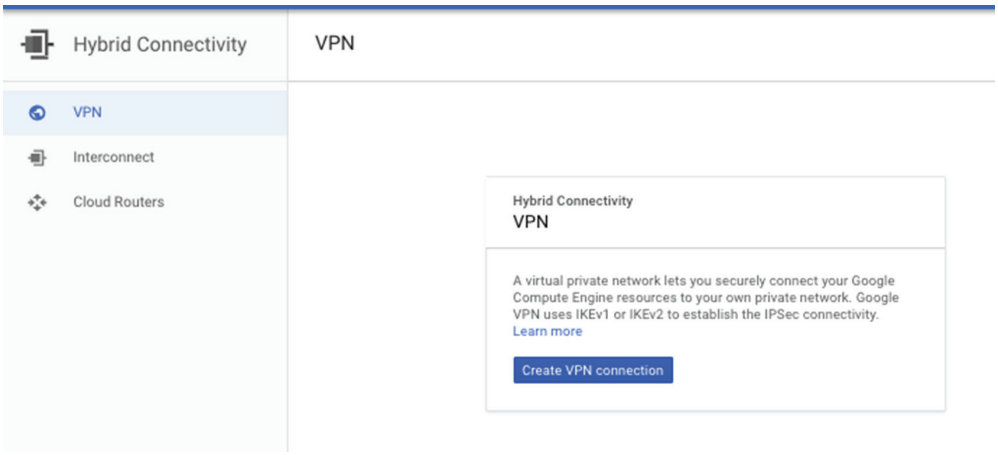
Creating a Virtual Private Network

VPNs allow you to securely send network traffic from the Google network to your own network. You can create a VPN using Cloud Console or the command line.

Creating a Virtual Private Network Using Cloud Console

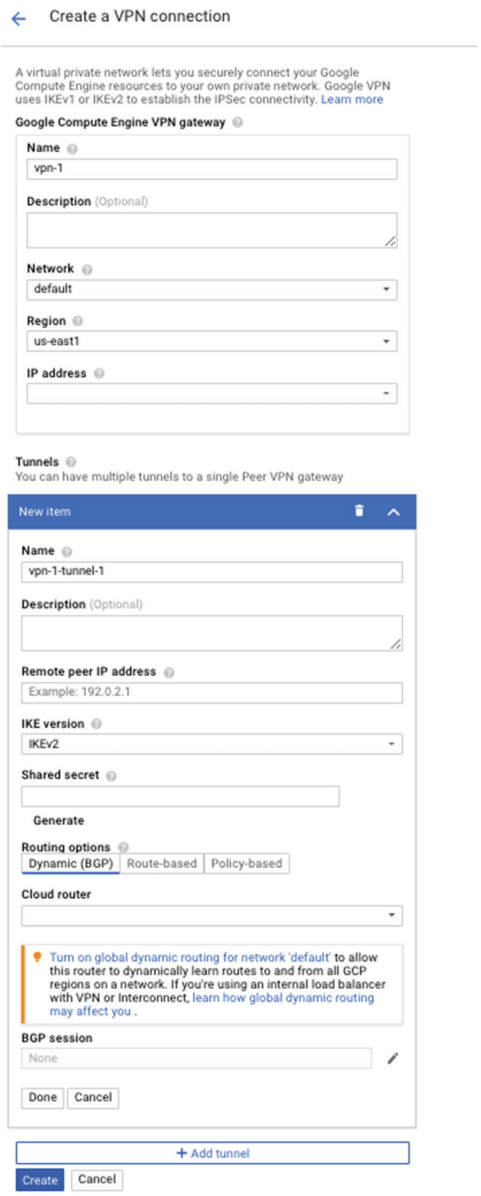
To create a VPN using Cloud Console, navigate to the Hybrid Connectivity section of the console, as shown in Figure 14.14.

FIGURE 14.14 Hybrid Connectivity section of Cloud Console



Click Create VPN Connection to display the form shown in Figure 14.15.

FIGURE 14.15 Create a VPN connection form



In this form, you specify a name and description of the VPN. In the section labeled Google Compute Engine VPN Gateway, you configure the GCP end of the VPN connection. This includes specifying a network, the region containing the network, and a static IP address. If you have not created an IP address, you create one by selecting Create IP Address from the drop-down menu for the IP Address parameter. It will display a dialog like the one in Figure 14.16.

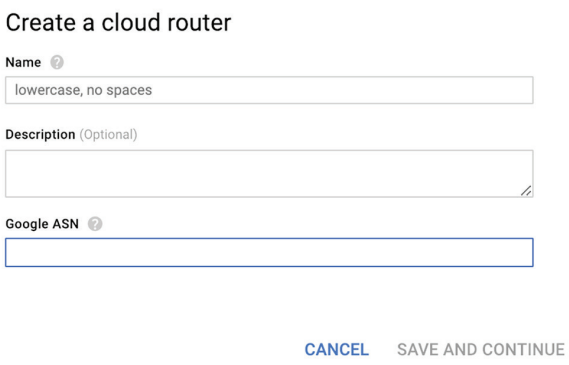
FIGURE 14.16 Creating a static IP address

In the Tunnels section, you configure the other network endpoint in the VPN. You specify a name, description, and IP address of the VPN gateway on your network. You can specify which version of the Internet Key Exchange (IKE) protocol to use. You will have to specify a shared secret, which is a secret string of characters, which your browser can create for you if you click Generate. You will need this shared secret when configuring your VPN endpoint.

In the Routing Options section, you can choose Dynamic, Route-Based, or Policy-Based Routing.

Dynamic routing uses the BGP protocol to learn routes in your networks. You will need to select or create a cloud router. If you have not created one, you can select Create Cloud Router from the drop-down list in the Cloud Router parameter. This will display a form like Figure 14.17.

FIGURE 14.17 Creating a cloud router



In that form, provide a name and description. You'll also need to specify a private autonomous system number (ASN) used by the BGP protocol. The ASN is a number in the range 64512–65534 or 4000000000–4294967294. Each cloud router you create will need a unique ASN.

If you choose route-based routing, you will need to enter the IP ranges of the remote network. If you choose policy-based routing, you will need to enter remote IP ranges, local subnetworks that will use the VPN, and local IP ranges.

Real World Scenario

Analytics in the Cloud

Data science and data analysis are increasingly important to businesses. To derive insights from these practices, you need both the data and the tools. Data about customers, sales, and other kinds of transactions are often stored in a database in a company's data center. The tools analysts want to use, such as Spark and machine learning services, are readily available in the cloud. Many organizations have security practices to protect data and would not allow an analyst, for example, to download some data and then copy it over an unsecure Internet connection to the cloud. Instead, network and cloud engineers would create a VPN between the company's data center and GCP. This would ensure that network traffic between the data center and the cloud is encrypted. Analysts get access to the data and tools they need, and the information security professionals in the organization are able to protect the confidentiality and integrity of the data.

Creating a Virtual Private Network Using gcloud

To create a VPN at the command line, you can use these three commands:

- `gcloud compute target-vpn-gateways`
- `gcloud compute forwarding-rule`
- `gcloud compute vpn-tunnels`

The format of the `gcloud compute target-vpn-gateways` command is as follows:

```
gcloud compute vpn-tunnels create NAME --peer-address=PEER_ADDRESS --shared-
secret=SHARED_SECRET --target-vpn-gateway=TARGET_VPN_GATEWAY
```

`NAME` is the name of the tunnel. `PEER_ADDRESS` is the IPv4 address of the remote tunnel endpoint. `SHARED_SECRET` is a secret string. `TARGET_VPN_GATEWAY` is a reference to the target VPN gateway IP.

The format of `gcloud compute forwarding-rule` is as follows:

```
gcloud compute forwarding-rules create NAME --TARGET_SPECIFICATION=VPN_GATEWAY
```

NAME is the name of the forwarding rule. TARGET_SPECIFICATION is one of several target types, including target-instance, target-http-proxy, and --target-vpn-gateway. For additional details, see the documentation at https://cloud.google.com/sdk/gcloud/reference/compute/forwarding-rules/create.

The format of the gcloud compute vpn-tunnels command is as follows:

```
gcloud compute vpn-tunnels create NAME --peer-address=PEER_ADDRESS --shared-
secret=SHARED_SECRET --target-vpn-gateway=TARGET_VPN_GATEWAY
```

NAME is the name of the VPN tunnel, PEER_ADDRESS is the IPv4 address of the remote tunnel, SHARED_SECRET is a secret string, and TARGET_VPN_GATEWAY is a reference to a VPN gateway.

Summary

This chapter reviewed how to create VPCs and VPNs. VPCs define networks in the Google Cloud to link your GCP resources. VPNs in GCP are used to link your GCP networks to your internal networks. We discussed how to create VPCs, shared VPCs, and subnets. There was a description of CIDR notation. You also learned how to configure VMs with custom network connections. Next we reviewed firewall rules and how to create them. The chapter concluded with discussing the steps required to create a VPN.

Exam Essentials

Know that VPCs are logical data centers in the cloud and VPNs are secure connections between your VPC subnets and your internal network. Your cloud resources are in a VPC. VPCs have subnets and routing rules for routing traffic between subnets. You control the flow of traffic using firewall rules.

Know that VPCs create subnets in each region when in auto mode. You can create additional subnets. Each subnet has a range of IP addresses. Firewall rules are applied to subnets, also called networks. Routers can be configured to learn just regional routes or global routes.

Understand how to read and calculate CIDR notation. CIDR notation represents a subnet mask and the size of available IP address in the IP range. The smaller the subnet mask size, which is the number after the slash in a CIDR block, the more IP addresses are available. The format of the CIDR address is an IP address followed by a slash, followed by the size of the subnet mask, for example 10.0.0.0/8.

Know that VPCs can be created using gcloud commands. A VPC can be created with gcloud compute networks create. A shared VPC can be created using gcloud beta

`compute shared-vpc`. Shared VPCs can be shared at the network or folder level. You will need to bind identity and access management (IAM) policies at the organizational or folder level to enable Shared VPC Admin roles. VPC peering can be used for interproject connectivity.

Understand that you can add network interfaces to a VM. You can configure these interfaces to use a particular subnet. You can assign ephemeral or static IP addresses.

Know that firewall rules control the flow of network traffic. Firewall rules consist of direction, priority, action, target, source/destination, protocols and port, and enforcement status. Firewall rules are applied to a subnet.

Know how to create a VPN with Cloud Console. VPNs route traffic between your cloud resources and your internal network. VPNs include gateways, forwarding rules, and tunnels.

Review Questions

You can find the answers in the Appendix.

1. Virtual private clouds have a _____ scope.
   - **A.** Zonal
   - **B.** Regional
   - **C.** Super-regional
   - **D.** Global

2. You have been tasked with defining CIDR ranges to use with a project. The project includes 2 VPCs with several subnets in each VPC. How many CIDR ranges will you need to define?
   - **A.** One for each VPC
   - **B.** One for each subnet
   - **C.** One for each region
   - **D.** One for each zone

3. The legal department needs to isolate its resources on its own VPC. You want to have network provide routing to any other service available on the global network. The VPC network has not learned global routes. What parameter may have been missed when creating the VPC subnets?
   - **A.** DNS server policy
   - **B.** Dynamic routing
   - **C.** Static routing policy
   - **D.** Systemic routing policy

4. The command to create a VPC from the command line is:
   - **A.** `gcloud compute networks create`
   - **B.** `gcloud networks vpc create`
   - **C.** `gsutil networks vpc create`
   - **D.** `gcloud compute create networks`

5. You have created several subnets. Most of them are sending logs to Stackdriver. One subnet is not sending logs. What option may have been misconfigured when creating the subnet that is not forwarding logs?
   - **A.** Flow Logs
   - **B.** Private IP Access
   - **C.** Stackdriver Logging
   - **D.** Variable-Length Subnet Masking

6. At what levels of the resource hierarchy can a shared VPC be created?
   - **A.** Folders and resources
   - **B.** Organizations and project
   - **C.** Organizations and folders
   - **D.** Folders and subnets

7. You are using Cloud Console to create a VM that you want to exist in a custom subnet you just created. What section of the Create Instance form would you use to specify the custom subnet?
   - **A.** Networking tab of the Management, Security, Disks, Networking, Sole Tenancy section
   - **B.** Management tab of the Management, Security, Disks, Networking, Sole Tenancy section
   - **C.** Sole Tenancy tab of Management, Security, Disks, Networking, Sole Tenancy
   - **D.** Sole Tenancy tab of Management, Security, Disks, Networking

8. You want to implement interproject communication between VPCs. Which feature of VPCs would you use to implement this?
   - **A.** VPC peering
   - **B.** Interproject peering
   - **C.** VPN
   - **D.** Interconnect

9. You want to limit traffic to a set of instances. You decide to set a specific network tag on each instance. What part of a firewall rule can reference the network tag to determine the set of instances affected by the rule?
   - **A.** Action
   - **B.** Target
   - **C.** Priority
   - **D.** Direction

10. What part of a firewall rule determines whether a rule applies to incoming or outgoing traffic?
    - **A.** Action
    - **B.** Target
    - **C.** Priority
    - **D.** Direction

11. You want to define a CIDR range that applies to all destination addresses. What IP address would you specify?
    - **A.** 0.0.0.0/0
    - **B.** 10.0.0.0/8
    - **C.** 172.16.0.0/12
    - **D.** 192.168.0.0/16

12. You are using `gcloud` to create a firewall rule. Which command would you use?
    A. `gcloud network firewall-rules create`
    B. `gcloud compute firewall-rules create`
    C. `gcloud network rules create`
    D. `gcloud compute rules create`

13. You are using `gcloud` to create a firewall rule. Which parameter would you use to specify the subnet it should apply to?
    A. `--subnet`
    B. `--network`
    C. `--destination`
    D. `--source-ranges`

14. An application development team is deploying a set of specialized service endpoints and wants to limit traffic so that only traffic going to one of the endpoints is allowed through by firewall rules. The service endpoints will accept any UDP traffic and each endpoint will use a port in the range of 20000–30000. Which of the following commands would you use?
    A. `gcloud compute firewall-rules create fwr1 --allow=udp:20000-30000 --direction=ingress`
    B. `gcloud network firewall-rules create fwr1 --allow=udp:20000-30000 --direction=ingress`
    C. `gcloud compute firewall-rules create fwr1 --allow=udp`
    D. `gcloud compute firewall-rules create fwr1 --direction=ingress`

15. You have a rule to allow inbound traffic to a VM. You want it to apply only if there is not another rule that would deny that traffic. What priority should you give this rule?
    A. 0
    B. 1
    C. 1000
    D. 65535

16. You want to create a VPN using Cloud Console. What section of Cloud Console should you use?
    A. Compute Engine
    B. App Engine
    C. Hybrid Connectivity
    D. IAM & Admin

17. You are using Cloud Console to create a VPN. You want to configure the GCP end of the VPN. What section of the Create VPN form would you use?
    - **A.** Tunnels
    - **B.** Routing Options
    - **C.** Google Compute Engine VPN
    - **D.** IKE Version

18. You want the router on a tunnel you are creating to learn routes from all GCP regions on the network. What feature of GCP routing would you enable?
    - **A.** Global dynamic routing
    - **B.** Regional routing
    - **C.** VPC
    - **D.** Firewall rules

19. When you create a cloud router, what kind of unique identifier do you need to assign for the BGP protocol?
    - **A.** IP address
    - **B.** ASN
    - **C.** Dynamic load routing ID
    - **D.** None of the above

20. You are using `gcloud` to create a VPN. Which command would you use?
    - **A.** `gcloud compute target-vpn-gateways only`
    - **B.** `gcloud compute forwarding-rule and gcloud compute target-vpn-gateways only`
    - **C.** `gcloud compute vpn-tunnels only`
    - **D.** `gcloud compute forwarding-rule, gcloud compute target-vpn-gateways, and gcloud compute vpn-tunnels`

Chapter

15

Networking in the Cloud: DNS, Load Balancing, and IP Addressing

THIS CHAPTER COVERS THE FOLLOWING OBJECTIVES OF THE GOOGLE ASSOCIATE CLOUD ENGINEER CERTIFICATION EXAM:

- ✓ **2.4 Planning and configuring network resources**
- ✓ **3.5 Deploying and implementing networking resources**
- ✓ **4.5 Managing network resources**

Official Google Cloud Certified: Associate Cloud Engineer: Study Guide
By Dan Sullivan

This chapter continues the focus on networking, specifically configuring the Domain Name System (DNS), load balancing, and managing IP addresses. Cloud DNS is a managed service providing authoritative domain naming services. It is designed for high availability, low latency, and scalability. Load balancing services in Google Cloud Platform (GCP) offer five types of load balancers to address a range of needs. In this chapter, you will see how HTTP(S), SSL Proxy, TCP Proxy, Network TCP/UDP, and Internal TCP/UDP Network differ and when to use each. Cloud engineers should also be familiar with managing IP addresses, in particular managing Classless Inter-Domain Routing (CIDR) blocks and understanding how to reserve IP addresses. This chapter, in combination with Chapter 14, provides an overview of the networking topics covered on the Associate Cloud Engineer exam.

Configuring Cloud DNS

Cloud DNS is a Google service that provides domain name resolution. At the most basic level, DNS services map domain names, such as `example.com`, to IP addresses, such as 35.20.24.107. A managed zone contains DNS records associated with a DNS name suffix, such as aceexamdns1.com. DNS records contain specific details about a zone. For example, an A record maps a hostname to IP addresses in IPv4. AAAA records are used in IPv6 to map names to IPv6 addresses. CNAME records hold the canonical name, which contains alias names of a domain. In this section, you will learn how to configure DNS services in GCP, which consists of creating zones and adding records.

Creating DNS Managed Zones Using Cloud Console

To create a managed zone using Cloud Console, navigate to the Network Services section of the console. Click Cloud DNS. This displays a form like that in Figure 15.1.

Click Create Zone to display a form like the one shown in Figure 15.2.

First, select a zone type, which can be public or private.

Public zones are accessible from the Internet. These zones provide name servers that respond to queries from any source. Private zones provide name services to your GCP resources, such as virtual machines (VMs) and load balancers. Private zones respond only to queries that originate from resources in the same project as the zone.

In the form, provide a zone name and description. Specify the DNS name, which should be the suffix of a DNS name, such as `aceexamdns1.com`.

FIGURE 15.1 Network Services Cloud DNS page

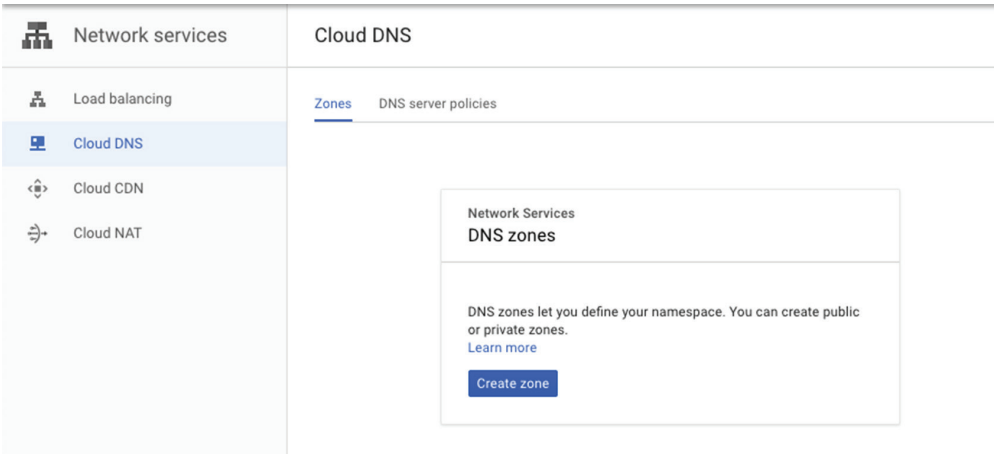
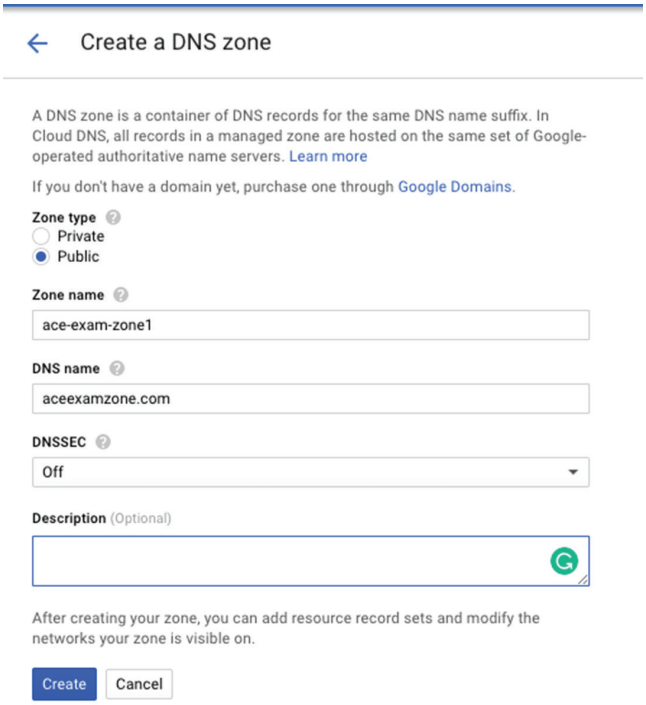


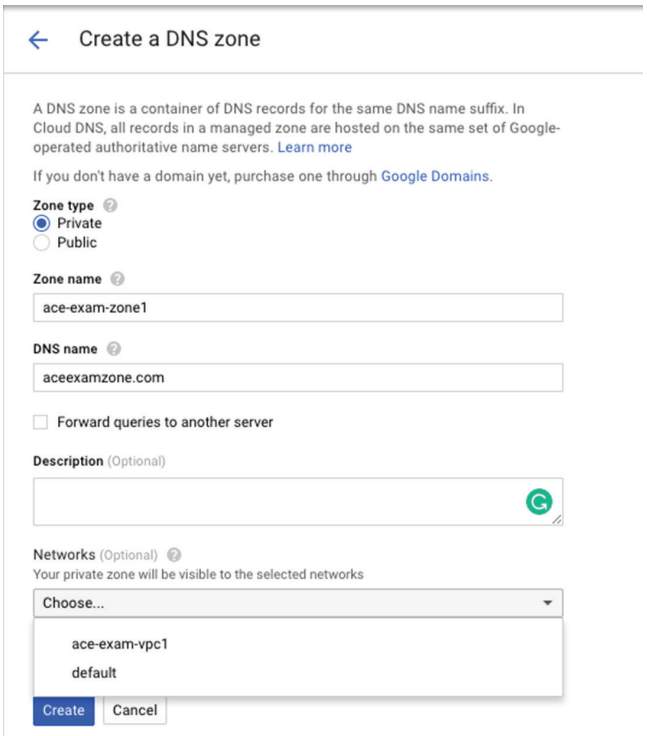
FIGURE 15.2 Create a public DNS zone.



You can enable DNSSEC, which is DNS security. It provides strong authentication of clients communicating with DNS services. DNSSEC is designed to prevent spoofing (a client appearing to be some other client) and cache poisoning (a client sending incorrect information to update the DNS server).

If you choose to create a private zone, a form such as Figure 15.3 appears.

FIGURE 15.3 Create a private DNS zone.



In addition to the parameters set for a public zone, you will need to specify the networks that will have access to the private zone.

After creating zones, the Cloud DNS page will list the zones, as shown in Figure 15.4.

Click the name of a zone to see its details. As shown in Figure 15.5, the zone details include a list of records associated with the zone. When a zone is created, NS and SOA records are added. NS is a *name server* record that has the address of an authoritative server that manages the zone information. SOA is a *start of authority* record, which has authoritative information about the zone. You can add other records, such as A and CNAME records.

FIGURE 15.4 List of DNS zones

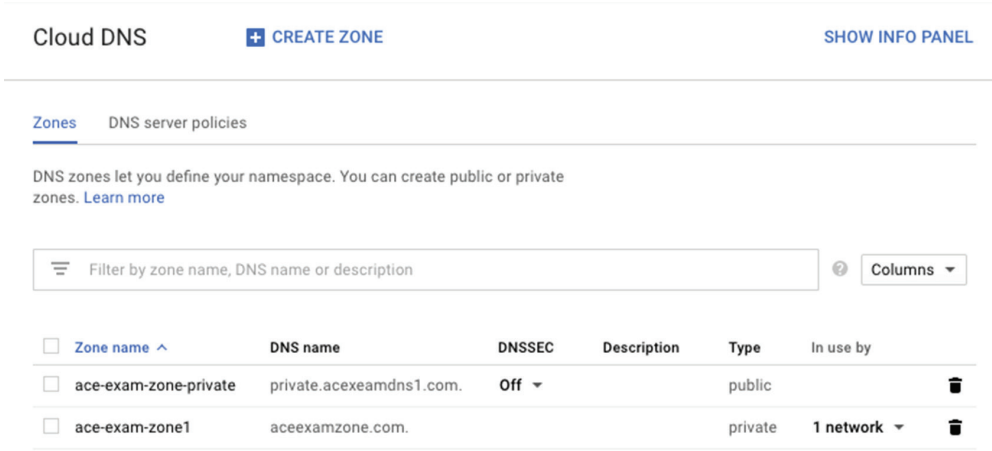
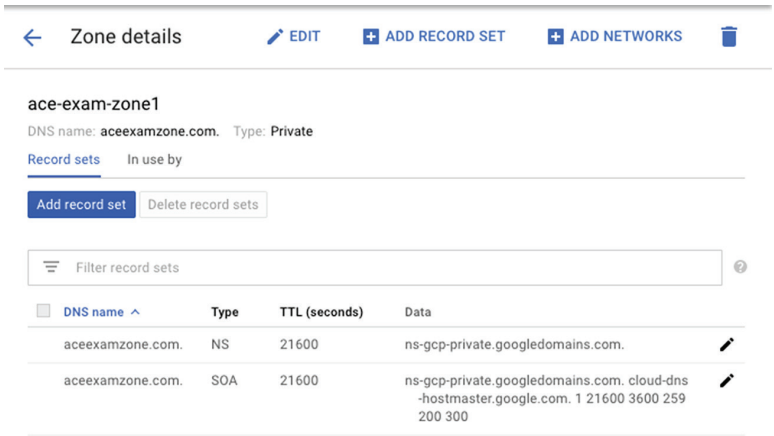


FIGURE 15.5 List of records in a DNS zone

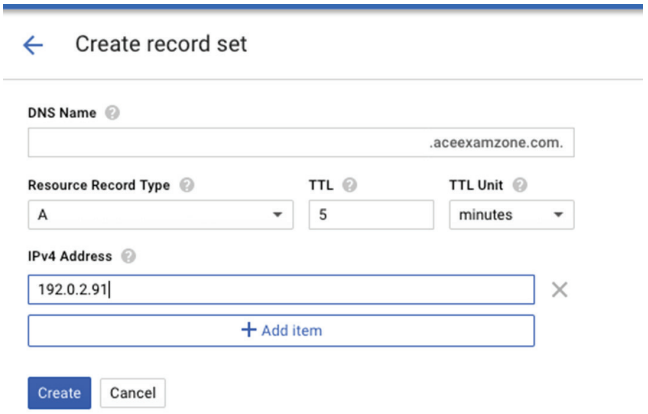


To add an A record, click Add Record Set to display a form like that in Figure 15.6.

Select A as a resource record type and specify an IPv4 address of the server that maps domain names to IP addresses for this zone.

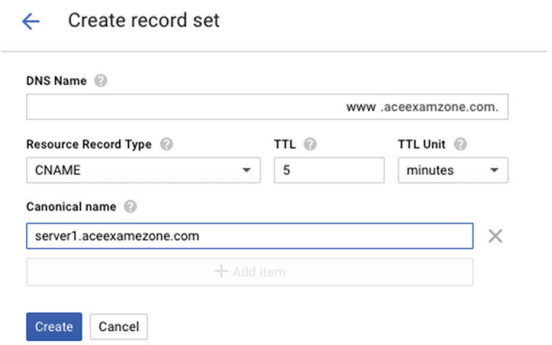
The TTL, known as time to live, and TTL Unit parameters specify how long the record can live in a cache. This is the period of time DNS resolvers should cache the data before querying for the value again. DNS resolvers perform lookup operations mapping domain names to IP addresses. If you want to specify multiple IP addresses in the record, click Add Item to add other IP addresses.

FIGURE 15.6 Create an A record set.



You can also add canonical name records using the Add Record Set form. In this case, select CNAME as the Resource Record Type, as shown in Figure 15.7.

FIGURE 15.7 Create a CNAME record.



The CNAME record takes a name, or alias, of a server. The DNS name and TTL parameters are the same as in the A record example.

Also, DNS Forwarding is now available, which allows your DNS queries to be passed to an on-premise DNS server if you are using Cloud VPN or Interconnect.

Creating a DNS Managed Zones Using gcloud

To create DNS zones and add records, you will use gcloud beta dns managed-zones and gcloud dns record-sets transaction.

To create a managed public zone called ace-exam-zone1 with the DNS suffix aceexamzone .com, you use this:

```
gcloud beta dns managed-zones create ace-exam-zone1 --description= --dns-
name=aceexamzone.com.
```

To make this a private zone, you add the --visibility parameter set to private.

```
gcloud beta dns managed-zones create ace-exam-zone1 --description= --dns-
name=aceexamzone.com. --visibility=private --networks=default
```

To add an A record, you start a transaction, add the A record information, and then execute the transaction.

Transactions are started with `gcloud dns record-sets transaction start`. Record sets are added using `gcloud dns record-sets transaction add`, and transactions are completed using `gcloud dns record-sets-transaction execute`. Together, the steps are as follows:

```
gcloud dns record-sets transaction start --zone=ace-exam-zone1
gcloud dns record-sets transaction add 192.0.2.91 --name=aceexamzone.com. --ttl=300
--type=A --zone=ace-exam-zone1
gcloud dns record-sets transaction execute --zone=ace-exam-zone1.
```

To create a CNAME record, we would use similar commands:

```
gcloud dns record-sets transaction start --zone=ace-exam-zone1
gcloud dns record-sets transaction add server1.aceexamezone.com. --
name=www2.aceexamzone.com. --ttl=300 --type=CNAME --zone=ace-exam-zone1
gcloud dns record-sets transaction execute --zone=ace-exam-zone1
```

Configuring Load Balancers

Load balancers distribute workload to servers running an application. In this section, we will discuss the different types of load balancers and how to configure them.

Types of Load Balancers

Load balancers can distribute load within a single region or across multiple regions. The several load balancers offered by GCP are characterized by three features:

- Global versus regional load balancing
- External versus internal load balancing
- Traffic type, such as HTTP and TCP

Global load balancers are used when an application is globally distributed. Regional load balancers are used when resources providing an application are in a single region. There are three global load balancers:

- HTTP(S), which balances HTTP and HTTPS load across a set of backend instances
- SSL Proxy, which terminates SSL/TLS connections, which are secure socket layer connections. This type is used for non-HTTPS traffic.
- TCP Proxy, which terminates TCP sessions at the load balancer and then forwards traffic to backend servers.

The regional load balancers are as follows:

- Internal TCP/UDP, which balances TCP/UDP traffic on private networks hosting internal VMs
- Network TCP/UDP, which enables balancing based on IP protocol, address, and port. This load balancer is used for SSL and TCP traffic not supported by the SSL Proxy and TCP Proxy load balancers, respectively.

External load balancers distribute traffic from the Internet, while internal load balancers distribute traffic that originates within GCP. The Internal TCP/UDP load balancer is the only internal load balancer. The HTTP(S), SSL Proxy, TCP Proxy, and Network TCP/UDP load balancers are all external.

You will need to consider the traffic type too when choosing a load balancer. HTTP and HTTPS traffic needs to use external global load balancing. TCP traffic can use external global, external regional, or internal regional load balancers. UDP traffic can use either external regional or internal regional load balancing.

Real World Scenario

Load Balancing and High Availability

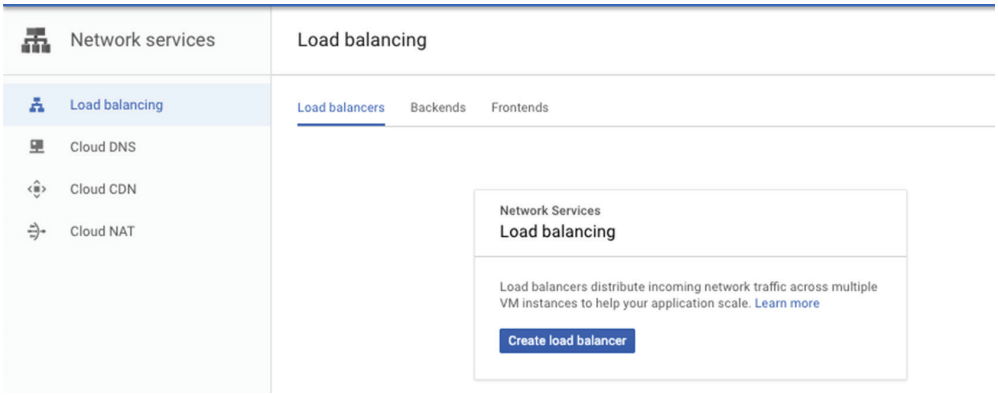
Applications that need to be highly available should use load balancers to distribute traffic and to monitor the health of VMs in the backend. A company offering API access to customer data will need to consider how to scale up and down in response to changes in load and how to ensure high availability.

The combination of instance groups (Chapter 6) and load balancers solves both problems. Instance groups can manage autoscaling, and load balancers can perform health checks. If a VM is not functioning, the health checks will fail and take the failed VM out of rotation for traffic. Users of the API are less likely to get failed response codes when instance groups keep an appropriate number of VMs active and load balancers prevent any traffic from being routed to failed servers.

Configuring Load Balancers Using Cloud Console

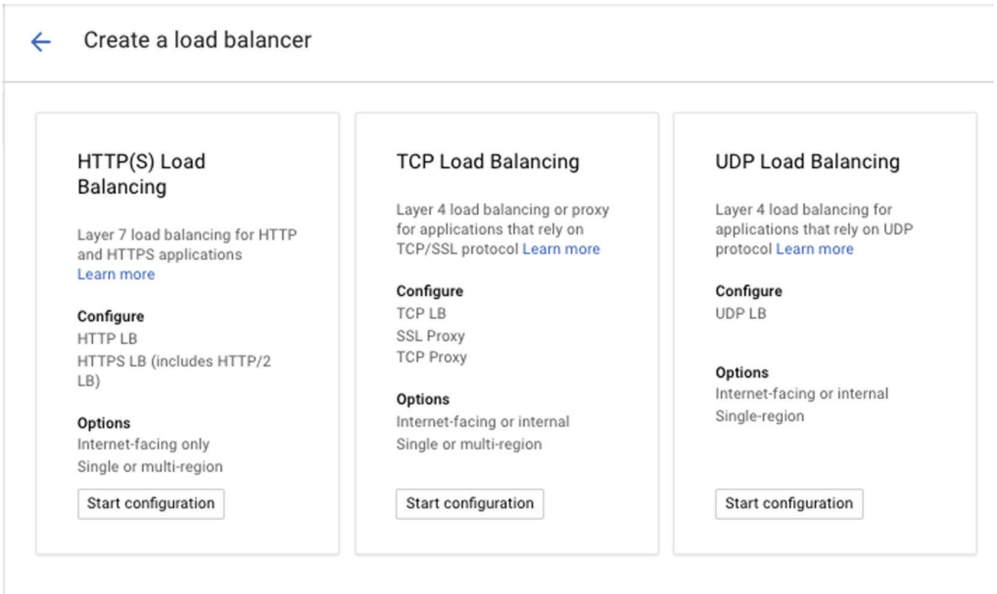
To create a load balancer in Cloud Console, navigate to the Network Services section and select Load Balancing, as shown in Figure 15.8.

FIGURE 15.8 Network services, load balancing section



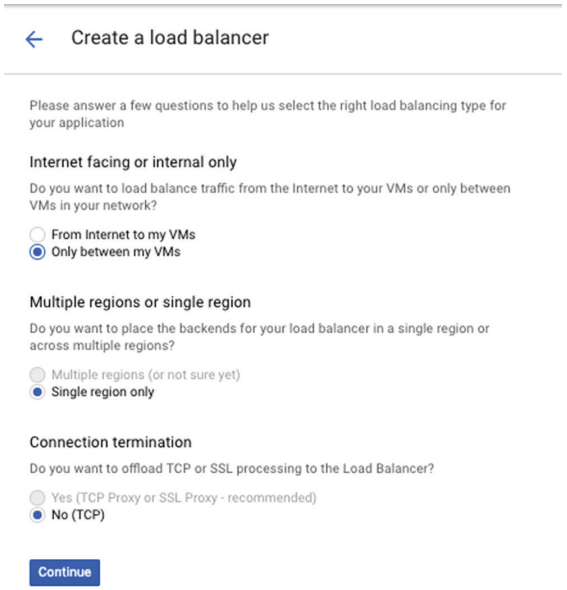
The first step to creating a load balancer is deciding on the type. In this example, you will create a TCP load balancer (see Figure 15.9).

FIGURE 15.9 Create A Load Balancer options



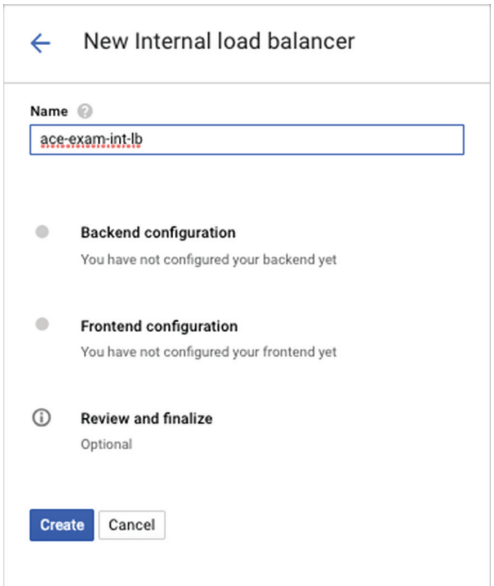
After selecting the TCP load balancer option, a form like Figure 15.10 appears. Select Only Between My VMs for private load balancing. This load balancer will be used in a single region, and you will not offload TCP or SSL processing.

FIGURE 15.10 Creating a TCP balancer



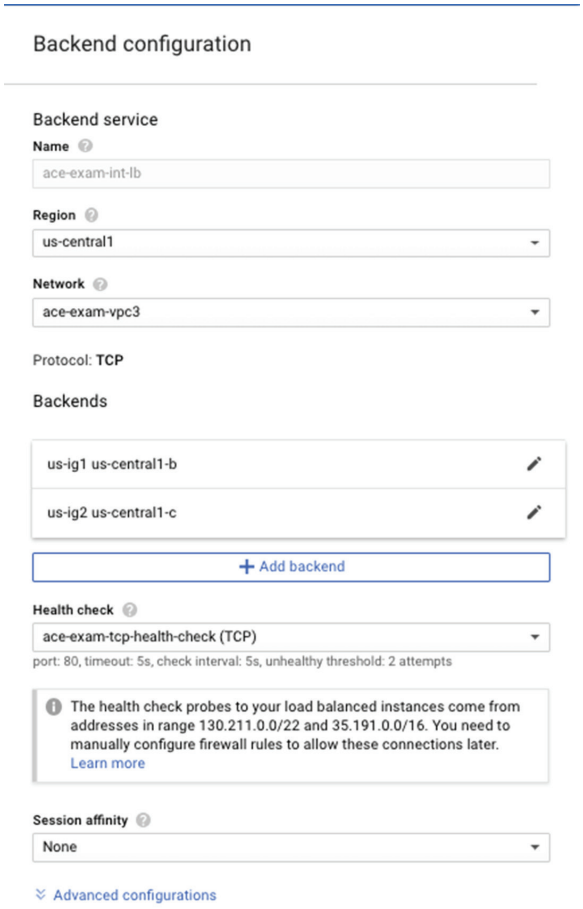
Now, you will begin a three-step process, as shown in Figure 15.11. You will configure the backend and the frontend and then review the configuration before creating the load balancer.

FIGURE 15.11 Three-step process to configure a load balancer



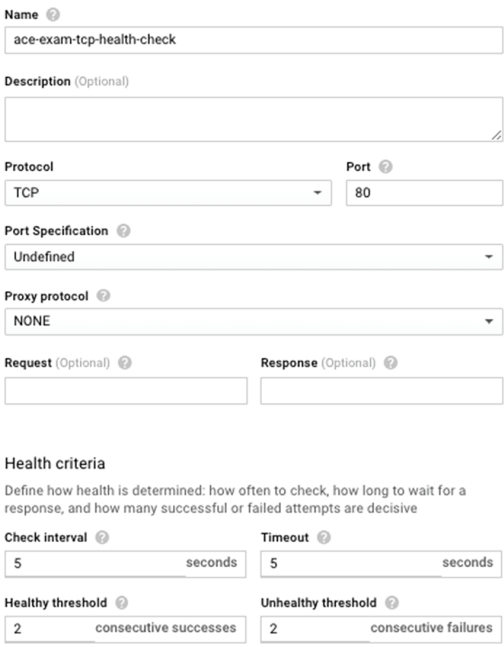
To configure the backend, you specify a name, a region, the network, and the backends. Backends are VMs that will have load distributed to them. In this example, two existing VMs are specified as backends (see Figure 15.12).

FIGURE 15.12 Configuring the backend



You can configure a health check for the backend. This will bring up a separate form, as shown in Figure 15.13.

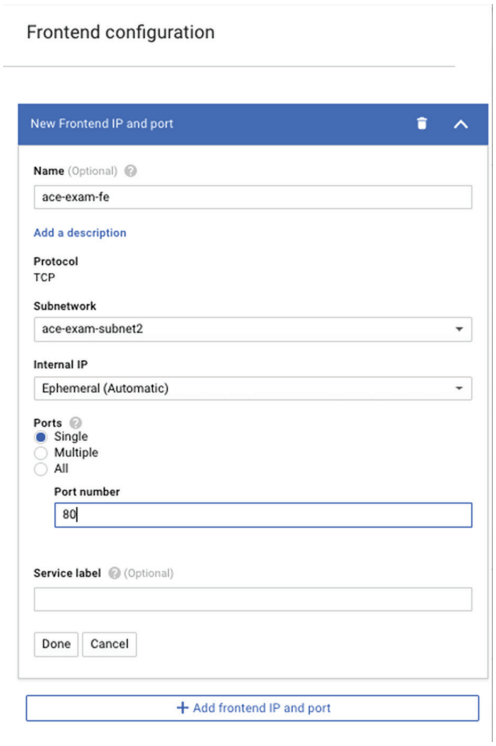
FIGURE 15.13 Creating a health check



In the health check, you specify a name, a protocol and port, and a set of health criteria. In this case, you check backends every 5 seconds and will wait for a response for up to 5 seconds. If you have two consecutive periods where the health check fails, then the server will be considered unhealthy and taken out of the load balancing rotation.

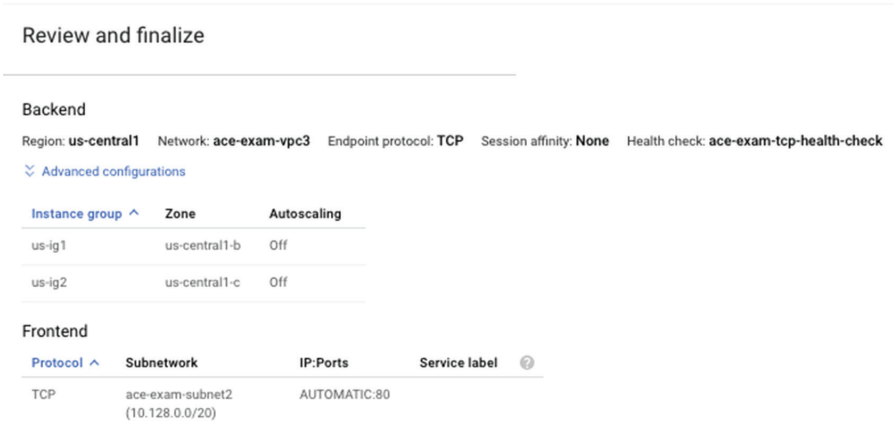
Next, you configure the frontend using the form in Figure 15.14. You specify a name, subnetwork, and an internal IP configuration, which in this case is ephemeral (see "Managing IP Addresses" on page 375 for more on types of IP addresses). You also specify the port that will have its traffic forwarded to the backend. In this example, you are forwarding traffic on port 80.

FIGURE 15.14 Configuring the frontend



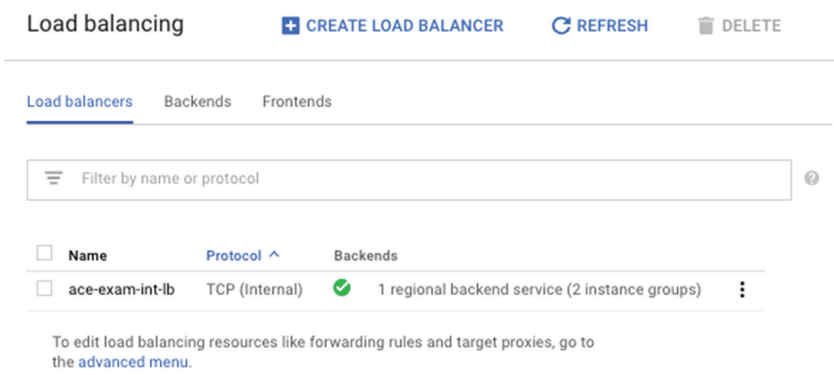
The last step prior to creating the frontend is to review the configuration, as shown in Figure 15.15.

FIGURE 15.15 Reviewing the load balancer configuration



After creating the load balancer, you will see the list of existing load balancers in the console (see Figure 15.16).

FIGURE 15.16 Listing of load balancers



Configuring Load Balancers Using gcloud

In this section, we will review the steps needed to create a network load balancer. These are good options when you need to load balance protocols in addition to HTTP(S).

The gcloud compute forward-rules command is used to forward traffic that matches an IP address to the load balancer.

```
gcloud compute forwarding-rules create ace-exam-lb --port=80
--target-pool ace-exam-pool
```

This command routes traffic to any VM in the ace-exam-pool to the load balancer called ace-exam-lb.

Target pools are created using the gcloud compute target-pools create command. Instances are added to the target pool using the `gcloud compute target-pools add-instances` command. For example, to add VMs ig1 and ig2 to the target pool called ace-exam-pool, use the following command:

```
gcloud compute target-pools add-instances ace-exam-pool --instances ig1,ig2
```

Managing IP Addresses

The exam topics for the Associate Cloud Engineer certification specifically identifies two IP address–related topics: expanding CIDR blocks and reserving IP addresses.

It is also important to understand the difference between ephemeral and static IP addresses. Static IP addresses are assigned to a project until they are released. These are used if you need a fixed IP address for a service, such as a website. Ephemeral IP addresses exist only as long as the resource is using the IP address, such as on a VM running an application only accessed by other VMs in the same project. If you delete or stop a VM, ephemeral addresses are released.

Expanding CIDR Blocks

CIDR blocks define a range of IP addresses that are available for use in a subnet. If you need to increase the number of addresses available, for example, if you need to expand the size of clusters running in a subnet, you can use the `gcloud compute networks subnets expand-ip-range` command. It takes the name of the subnet and a new prefix length. The prefix length determines the size of the network mask.

For example, to increase the number of addresses in ace-exam-subnet1 to 65,536, you set the prefix length to 16:

```
gcloud compute networks subnets expand-ip-range ace-exam-subnet1 --prefix-length 16
```

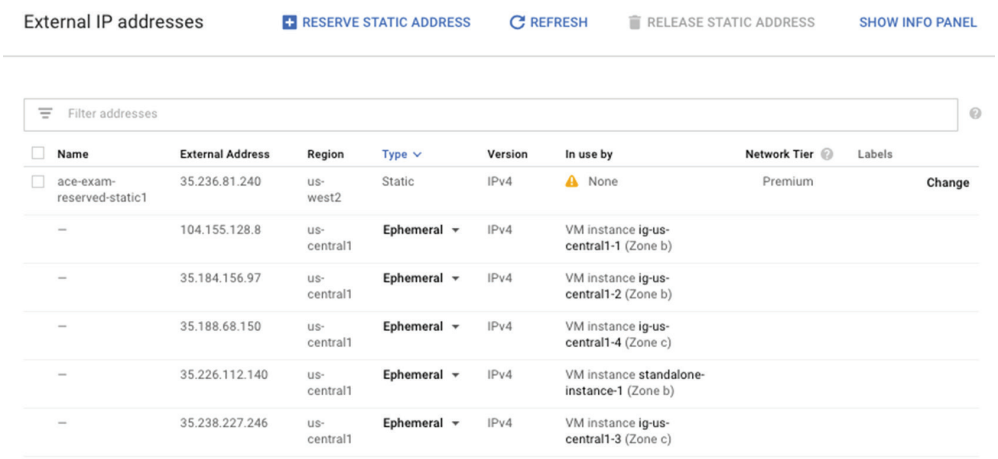
This assumes the prefix length was larger than 16 prior to issuing this command. The `expand-ip-range` command is used only to increase the number of addresses. You cannot decrease them, though. You would have to re-create the subnet with a smaller number of addresses.

Reserving IP Addresses

Static external IP addresses can be reserved using Cloud Console or the command line. To reserve a static IP address using Cloud Console, navigate to the Virtual Private Cloud (VPC) section of the console and select External IP Addresses.

This will display a form like the one shown in Figure 15.17.

FIGURE 15.17 List of reserved static IP addresses



Click Reserve Static Address to display the form to reserve an IP address (see Figure 15.18).

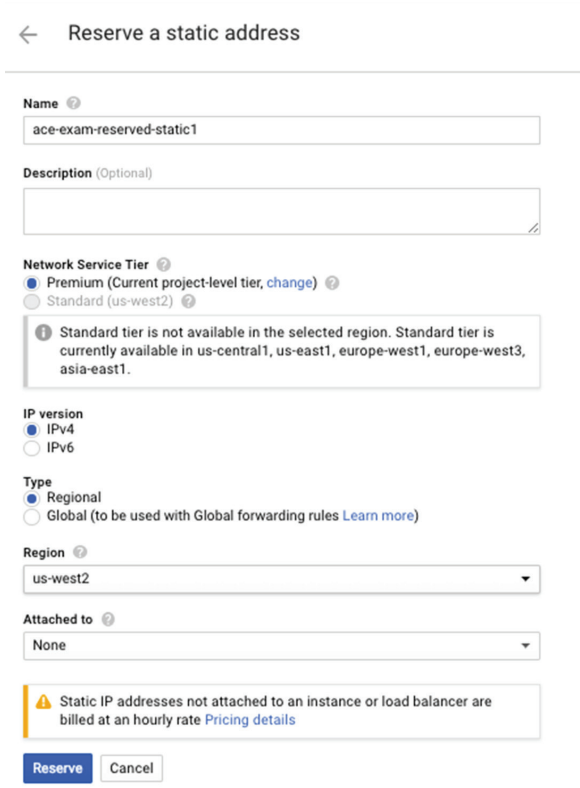
When reserving an IP address, you will need to specify a name and optional description. You may have the option of using the lower-cost Standard service tier for networking, which uses the Internet for some transfer of data. The Premium tier routes all traffic over Google's global network. You will also need to determine whether the address is in IPv4 or IPv6 and whether it's regional or global. You can attach the static IP address to a resource as part of the reservation process, or you can keep it unattached.

Reserved addresses stay attached to a VM when it is not in use and stay attached until released. This is different from ephemeral addresses, which are released automatically when a VM shuts down.

To reserve an IP address using the command line, use the `gcloud` command `gcloud beta compute addresses create`. For example, to create a static IP address in the us-west2 region, which uses the Premium tier, use this command:

```
gcloud beta compute addresses create ace-exam-reserved-static1 --region=us-west2
--network-tier=PREMIUM
```

FIGURE 15.18 Reserving a static IP address



Summary

The Associate Cloud Engineer exam may test your knowledge of Cloud DNS, load balancing, and managing IP addresses. Cloud DNS is an authoritative name service for mapping domain names to IP addresses. You can set up public or private DNS zones. You will also need to be familiar with load balancing and the different types of load balancers. Some load balancers are regional, and some are global. Some are for internal use only, and others support external sources of traffic. The chapter also reviewed how to expand the number of addresses available in a subnet and discussed how to reserve IP addresses.

Exam Essentials

Understand that Cloud DNS is used to map domain names to IP addresses. If you want to support queries from the Internet, use a public DNS zone. Use a private DNS zone only if you want to accept queries from resources in your project.

Know that DNS entries, like example.com, can have multiple records associated with them. The A record specifies the address of a DNS resolver that maps domain names to IP addresses. CNAME records store the canonical name of the domain.

Know how load balancers are distinguished. Load balancers are distinguished based on global versus regional load balancing, external versus internal load balancing, and the protocols supported. Global balancers distribute load across regions, while regional load balancers work within a region. Internal load balancers balance traffic only from within GCP, not external sources. Some load balancers are protocol-specific, such as HTTP and SSL load balancers.

Know the five types of load balancers and when they should be used. The five are: HTTP(S), SSL Proxy, TCP Proxy, Internal TCP/UDP, and Network TCP/UDP.

HTTP(S) balances HTTP and HTTPS load.

SSL Proxy terminates SSL/TLS connections.

TCP Proxy terminates TCP sessions.

Internal TCP/UDP balances TCP/UDP traffic on private networks hosting internal VMs

Network TCP/UDP load balancing is based on IP protocol, address, and port.

Understand that configuring a load balancer can require configuring both the frontend and backend. The network load balancer can be configured by specifying a forwarding rule that routes traffic to the load balancer to VMs in the target pool.

Know how to increase the number of IP addresses in a subnet. Use the `gcloud compute network subnets expand-ip-range` command to increase IP addresses in a subnet. The number of addresses can only increase. The `expand-ip-range` command cannot be used to decrease the number of addresses.

Know how to reserve an IP address using the console and the `gcloud beta compute address create` command. Reserved IP addresses continue to be available to your project even if they are not attached to a resource. Know the difference between Premium and Standard tier network services.

Review Questions

You can find the answers in the Appendix.

1. What record type is used to specify the IPv4 address of a domain?
   - **A.** AAAA
   - **B.** A
   - **C.** NS
   - **D.** SOA

2. The CEO of your startup just read a news report about a company that was attacked by something called cache poisoning. The CEO wants to implement additional security measures to reduce the risk of DNS spoofing and cache poisoning. What would you recommend?
   - **A.** Using DNSSEC
   - **B.** Adding SOA records
   - **C.** Adding CNAME records
   - **D.** Deleting CNAME records

3. What do the TTL parameters specify in a DNS record?
   - **A.** Time a record can exist in a cache before it should be queried again
   - **B.** Time a client has to respond to a request for DNS information
   - **C.** Time allowed to create a CNAME record
   - **D.** Time before a human has to manually verify the information in the DNS record

4. What command is used to create a DNS zone in the command line?
   - **A.** `gsutil dns managed-zones create`
   - **B.** `gcloud beta dns managed-zones create`
   - **C.** `gcloud beta managed-zones create`
   - **D.** `gcloud beta dns create managed zones`

5. What parameter is used to make a DNS zone private?
   - **A.** --private
   - **B.** --visibility=private
   - **C.** --private=true
   - **D.** --status=private

6. Which load balancers provide global load balancing?
   - **A.** HTTP(S) only
   - **B.** SSL Proxy and TCP Proxy only
   - **C.** HTTP(S), SSL Proxy, and TCP Proxy
   - **D.** Internal TCP/UDP, HTTP(S), SSL Proxy, and TCP Proxy

7. Which regional load balancer allows for load balancing based on IP protocol, address, and port?
   - **A.** HTTP(S)
   - **B.** SSL Proxy
   - **C.** TCP Proxy
   - **D.** Network TCP/UDP

8. You are configuring a load balancer and want to implement private load balancing. Which option would you select?
   - **A.** Only Between My VMs
   - **B.** Enable Private
   - **C.** Disable Public
   - **D.** Local Only

9. What two components need to be configured when creating a TCP Proxy load balancer?
   - **A.** Frontend and forwarding rule
   - **B.** Frontend and backend
   - **C.** Forwarding rule and backend only
   - **D.** Backend and forwarding rule only

10. A health check is used to check what resources?
    - **A.** Load balancer
    - **B.** VMs
    - **C.** Storage buckets
    - **D.** Persistent disks

11. Where do you specify the ports on a TCP Proxy load balancer that should have their traffic forwarded?
    - **A.** Backend
    - **B.** Frontend
    - **C.** Network Services section
    - **D.** VPC

12. What command is used to create a network load balancer at the command line?
    A. `gcloud compute forwarding-rules create`
    B. `gcloud network forwarding-rules create`
    C. `gcloud compute create forwarding-rules`
    D. `gcloud network create forwarding-rules`

13. A team is setting up a web service for internal use. They want to use the same IP address for the foreseeable future. What type of IP address would you assign?
    A. Internal
    B. External
    C. Static
    D. Ephemeral

14. You are starting up a VM to experiment with a new Python data science library. You'll SSH via the server name into the VM, use the Python interpreter interactively for a while and then shut down the machine. What type of IP address would you assign to this VM?
    A. Ephemeral
    B. Static
    C. Permanent
    D. IPv8

15. You have created a subnet called sn1 using 192.168.0.0 with 65,534 addresses. You realize that you will not need that many addresses, and you'd like to reduce that number to 254. Which of the following commands would you use?
    A. `gcloud compute networks subnets expand-ip-range sn1 --prefix-length=24`
    B. `gcloud compute networks subnets expand-ip-range sn1 --prefix-length=-8`
    C. `gcloud compute networks subnets expand-ip-range sn1 --size=256`
    D. There is no command to reduce the number of IP addresses available.

16. You have created a subnet called sn1 using 192.168.0.0. You want it to have 14 addresses. What prefix length would you use?
    A. 32
    B. 28
    C. 20
    D. 16

17. You want all your network traffic to route over the Google network and not traverse the public Internet. What level of network service should you choose?
    A. Standard
    B. Google-only
    C. Premium
    D. Non-Internet

18. You have a website hosted on a Compute Engine VM. Users can access the website using the domain name you provided. You do some maintenance work on the VM and stop the server and restart it. Now users cannot access the website. No other changes have occurred on the subnet. What might be the cause of the problem?
    - **A.** The restart caused a change in the DNS record.
    - **B.** You used an ephemeral instead of a static IP address.
    - **C.** You do not have enough addresses available on your subnet.
    - **D.** Your subnet has changed.

19. You are deploying a distributed system. Messages will be passed between Compute Engine VMs using a reliable UDP protocol. All VMs are in the same region. You want to use the load balancer that best fits these requirements. Which kind of load balancer would you use?
    - **A.** Internal TCP/UDP
    - **B.** TCP Proxy
    - **C.** SSL Proxy
    - **D.** HTTP(S)

20. You want to use Cloud Console to review the records in a DNS entry. What section of Cloud Console would you navigate to?
    - **A.** Compute Engine
    - **B.** Network Services
    - **C.** Kubernetes Engine
    - **D.** Hybrid Connectivity

Chapter

16

Deploying Applications with Cloud Launcher and Deployment Manager

THIS CHAPTER COVERS THE FOLLOWING OBJECTIVES OF THE GOOGLE ASSOCIATE CLOUD ENGINEER CERTIFICATION EXAM:

- ✓ **3.6 Deploying a solution using Cloud Launcher**
- ✓ **3.7 Deploying an application using Deployment Manager**

Official Google Cloud Certified: Associate Cloud Engineer: Study Guide
By Dan Sullivan

Throughout this exam guide you have learned how to deploy computing, storage, and networking resources, and now you will turn your attention to deploying applications. Cloud Launcher is Google Cloud Platform's (GCP's) marketplace, where you can find preconfigured applications that are ready to deploy into the Google Cloud.

Google has given Cloud Launcher a new name: Marketplace. The Associate Cloud Engineer Certification guide refers to the service as Cloud Launcher, so we will continue to refer to it as Cloud Launcher in this chapter. You will see how to use Deployment Manager to configure templates, which can launch your own custom applications into Google Cloud. Cloud Launcher and Deployment Manager let users deploy applications and necessary compute, storage, and network resources without having to configure those resources themselves.

Deploying a Solution Using Cloud Launcher

Cloud Launcher is a central repository of applications and data sets that can be deployed to your GCP environment. Working with the Cloud Launcher is a two-step process: browsing for a solution that fits your needs and then deploying the solution.

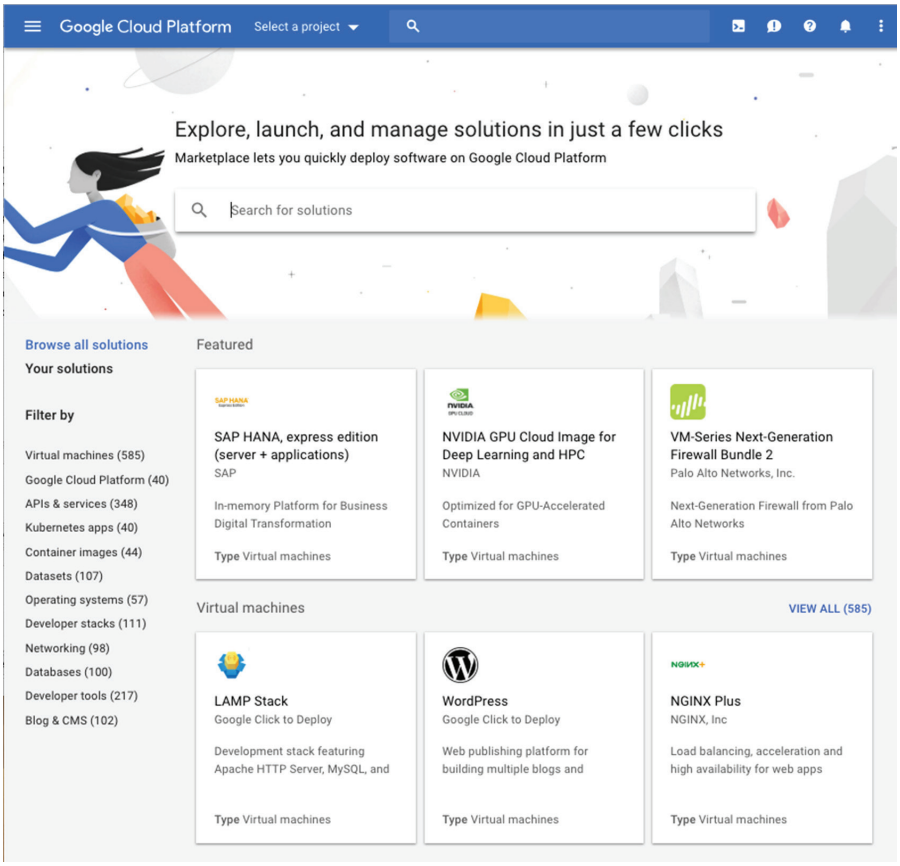
Browsing Cloud Launcher and Viewing Solutions

To view the solutions available in Cloud Launcher, navigate to the Marketplace section. Marketplace is another name for the Cloud Launcher page in Cloud Console. This will display a page like that shown in Figure 16.1.

The main page of Cloud Launcher shows some featured solutions.

The solutions shown in Figure 16.1 include SAP HANA, a NVIDIA deep learning application, and a Palo Alto networks firewall package. There are also some popular open source systems, including a Linux, Apache, MySQL, and PHP (LAMP) stack and a WordPress blog platform.

FIGURE 16.1 Cloud Launcher main page



You can either search or browse by filter to see the list of solutions. Figure 16.2 shows the list of categories of available solutions.

You can narrow the set of solutions displayed on the main page by choosing a particular category. For example, if you filter to see only data sets, you will see a list of datasets such as that shown in Figure 16.3.

FIGURE 16.2 Filtering by category

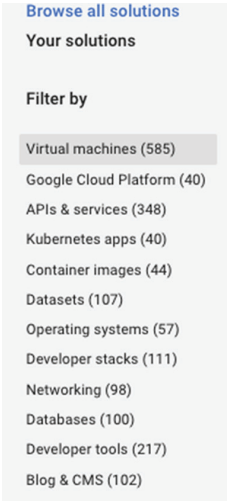


FIGURE 16.3 Data sets available in Cloud Launcher