CSCI 3287 Design and Analysis of Data Systems

Homework # 7 -- Cassandra Lab

April 20, 2020

## Overview

This Project is worth 100 points (out of 1000) toward your final grade. It is due on Sunday, April 26 at 11:59 p.m.  Your assignment submission should be a document saved and submitted as a PDF file via the link found in the assignment section of "Week 14" in Moodle which is the same place where you found this file.

This assignment will give you hands-on practice in working with the Apache Cassandra "NoSQL" database software.

## Objectives

- Students will deploy and run the Apache Cassandra NoSQL database engine

- Students will gain exposure to CQL – Cassandra Query Language

- Students will experience deployment of Apache Cassandra in GCP (the Google Cloud Platform)

- Students will launch and use the Cassandra CQL Shell in GCP

## Sections

This assignment consists of TWO sections.

Part One      Students sign up with GCP (Google Cloud Platform) and obtain free credits

Part Two      Students use free QwikLabs credits to complete QwikLabs course "GSP704 - Deploying an Open Source Cassandra™ Database using the GCP Marketplace"

## Submission

Create a document and save it as a PDF file.  Submit your document via the link in Moodle in Week Fourteen for Homework # 7.

You may work with a partner on this assignment.  Each partner must submit their own submission.  Be sure to include your partner's name on your submission.

# Part One Instructions:

When signing up for your free GCP & Qwiklabs credits, you must use your email address that is composed of your identikey ID (for example, mine is "alpa4161") followed by "@colorado.edu".

This is a valid Google account through which the GCP folks will verify that you are indeed a CU student and are therefore authorized to receive free GCP/QwikLabs credits.

**Step 1 -** Navigate to the following page and enter in your student credentials

https://edu.google.com/programs/students/benefits/?modal_active=none

Click on the "Access Benefits" button and follow the instructions.

User = *identikey*@colorado.edu

Pass = Identikey password

## Get credits and discounts

Explore what you can do in Google Cloud with credits and certification discounts — **no credit card or other payment instrument required**. *If you're a higher education student in one of these countries, you may be eligible for benefits.*

**Access benefits**

### 200 Qwiklabs credits

Get access to the Google Cloud catalog in Qwiklabs for hands-on practice using interactive labs.

### On-demand training courses

Gain a deeper understanding of Google Cloud with 13 professional-level fundamental courses on Coursera at no cost.

### $50 in Google Cloud credits

Receive $50 in Google Cloud credits for career development in your courses and clubs. Credits expire 12 months after activation.

## **Step 2** – Verify your student status

EDU VERIFICATION FORM

# Verify your education credentials

This form will collect your email address based on the Google Account you are signed in to. Before proceeding, ensure you are signed in with the account you want your edu benefits applied to. You can change accounts by selecting the profile picture in the top right corner of your screen.

* Required Fields.

First name *

| Sam |

Legal name registered with your institution

Last name *

| Stone |

Legal name registered with your institution

Role *

| Student |

Email address

| YourIdentiKey@colorado.edu |

Country *

| United States |

Academic institution *

| University of Colorado Boulder (Boulder, CO) |

Begin typing and select institution from the provided list

Date of birth *

| August | | 27 | | 1954 |

How did you hear about us? *

| A faculty or professor |

Sign up for Google Cloud emails with news, product updates, event information, special offers, and more. You can unsubscribe at any time. *

○ Yes
⦿ No

Information on this form, as well as your Google Account email address, will be sent to SheerID to verify your student or faculty status. Information provided will be processed by SheerID in accordance with their privacy policy. Your information is also governed by the Google privacy policy and any redemption of credits will be governed by the Terms of Service for each respective program. *

☑ I agree

Submit

**Step 3** - Redeem your $50 in Google Cloud Credits and 200 Qwiklabs Credits

Here are your student benefits.

Redeem the benefits that are right for your current cloud journey. There is no need to select all at once - most have an expiration date once redeemed and you can always return to this page to access more benefits in the future.

**$50 in Google Cloud credits**

Receive $50 in Google Cloud credits for career development in your courses and clubs. Credits expire 12 months after activation.

Redeem credits    *Automatic redemption*

**200 Qwiklabs credits**

Get access to the Google Cloud catalog in Qwiklabs for hands-on practice using interactive labs.

Redeem access    *Automatic redemption*

**On-demand training courses**

Gain a deeper understanding of Google Cloud with 13 professional-level fundamental courses on Coursera at no cost.

Request courses    *Additional application required*

**$1000 in research credits**

PhD candidates are eligible to receive up to $1000 in Google Cloud credits per research proposal.

Apply for credits    *Additional application required*

Now google will ask you for more information so that they can verify your student status

for your free credits.

# We need your ID & class schedule

Documents will be reviewed in a few minutes by staff at SheerID, a trusted partner. How does verifying work?

## Upload your school-issued document that shows:

- Your full name
- Your school
- Date proving enrollment for the current term (or within the last 3 months)

## Acceptable examples include:

- School ID card with expiration date
- Class schedule
- Tuition receipt

Drag & drop your files here or browse

File must be .jpg .png .pdf .gif or .bmp. You can upload multiple files together and extraneous information can be redacted.

Your information will be transferred to the United States and will be treated according to SheerID's privacy policy.

Verification services powered by SheerID SheerID FAQs

Once you hear back from their verification process, they will provide you with a key code that you must enter in order to activate your credits.

# Part Two Instructions:

For Part Two of this homework assignment, you will use your free Qwiklabs credits to complete an online lab using the Cassandra database in the cloud.

1.      To access the online lab, first, go to https://google.qwiklabs.com/catalog

2.      Search for the "Cassandra" course GSP704, Deploying an Open Source Cassandra™ Database using the GCP Marketplace
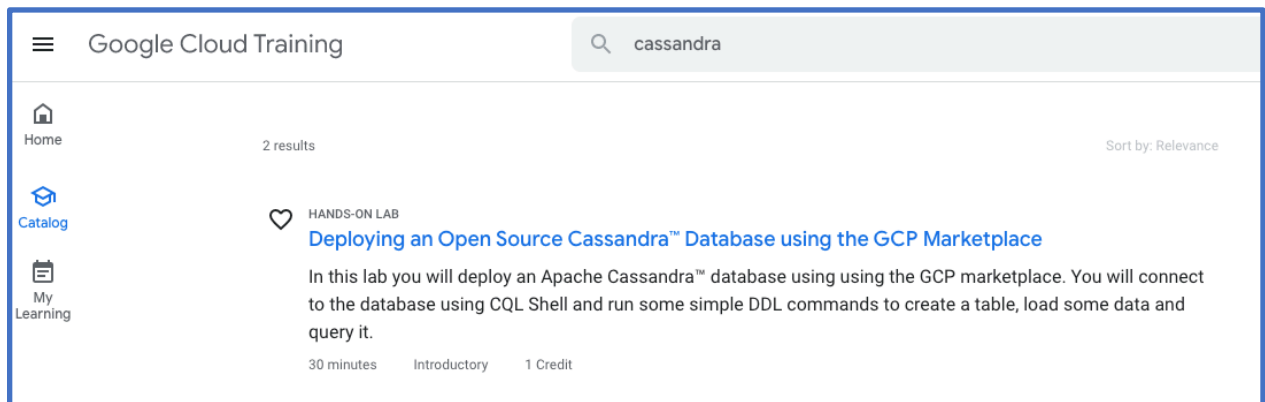


3.      Follow the directions and complete the lab. The directions will appear on your screen when you enter the lab.  For your review, I have copied the directions into Appendix "A" below for your reference. You should read through the directions before starting.

4.      Keep an eye on the time.  The lab allows you 30 minutes.  You should be able to complete the lab in 15 minutes or less.

5.      Once you finish the lab, the terminal session in GCP remains open for you to play. Feel free to use the cqlsh terminal window to play with the CQL language to see what it can and can't do (compared to SQL.)

6.      Answer the following questions for your homework submission:

QUESTIONS:

1.      In step 4 of the lab, you spun up a VM running Cassandra.  What is your deployment name?

2.      In what GCP zone is it running?

3.      How many CPUs in your VM?

4.      How much memory did it allocate to your VM?

5.      What is the IP address of your VM?

6.	What command do you enter to connect to the Cassandra shell?   What was your username and password?

7.	What is the name of the keyspace you created?

8.	What is the name of the table you created?

9.	What is the replication factor for the keyspace you created?

	In the instructions it told you to exit the cqlsh.   If ou have time remaining (out of your 30 minutes lab) go back to the SSH dropdown, open a new cqlsh terminal, log in again.

10.	Run the query:  `Select * from space_flights.catalog;` again

11.	Modify the query:  `select count(*) from space_flights.catalog;` How many rows are in the table?

12.	Run the query: `select spacecraft_name, journey_id from space_flights.catalog`

13.	Modify your query to select all the rows where the spacecraft_name is like "%mercury%".  What happens?    (Not all SQL commands will work in CQL).

	See https://docs.datastax.com/en/archived/cql/3.3/cql/cql_reference/cqlSelect.html

14.	Modify your query to try to sort the output using a "ORDER BY".   What happens?

# Deploying an Open Source Cassandra™ Database using the GCP Marketplace

30 minutes1 Credit

Rate Lab

**GSP704**

Google Cloud Self-Paced Labs

# Overview

In this lab you will deploy an Apache Cassandra™ database using using the GCP marketplace. You will connect to the database using CQL Shell and run some simple DDL commands to create a table, load some data and query it.

# Objectives

In this lab, you learn how to perform the following tasks:

- Deploy Apache Cassandra™ using the GCP marketplace

- Connect to Cassandra™ using CQL Shell

- Create keyspaces and tables using CQL

- Read and write data using CQL

# Setup and Requirements

## Qwiklabs setup

**Before you click the Start Lab button**

Read these instructions. Labs are timed and you cannot pause them. The timer, which starts when you click Start Lab, shows how long Cloud resources will be made available to you.

This Qwiklabs hands-on lab lets you do the lab activities yourself in a real cloud environment, not in a simulation or demo environment. It does so by giving you new, temporary credentials that you use to sign in and access the Google Cloud Platform for the duration of the lab.

**What you need**

To complete this lab, you need:

- Access to a standard internet browser (Chrome browser recommended).
- Time to complete the lab.
  **Note:** If you already have your own personal GCP account or project, do not use it for this lab.

  **Note:** If you are using a Pixelbook please open an Incognito window to run this lab.

# Google Cloud Platform Console

**How to start your lab and sign in to the Console**

1. Click the **Start Lab** button. If you need to pay for the lab, a pop-up opens for you to select your payment method. On the left is a panel populated with the temporary credentials that you must use for this lab.

2. Copy the username, and then click **Open Google Console**. The lab spins up resources, and then opens another tab that shows the **Choose an account** page.
   *Tip:* Open the tabs in separate windows, side-by-side.

3. On the Choose an account page, click **Use Another Account**.
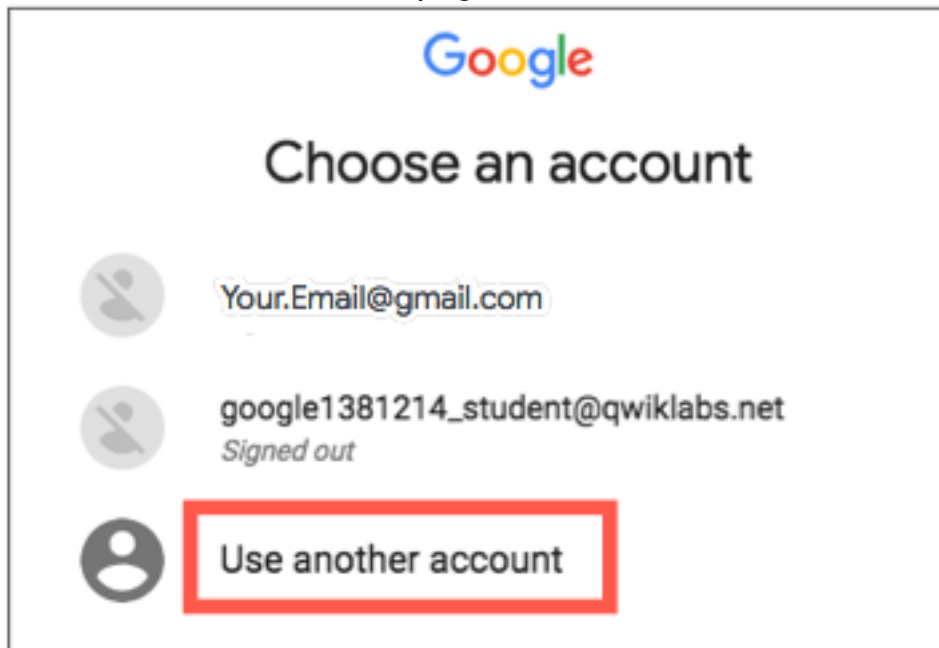
4. The Sign in page opens. Paste the username that you copied from the Connection Details panel. Then copy and paste the password.

   *Important:* You must use the credentials from the Connection Details panel. Do not use your Qwiklabs credentials. If you have your own GCP account, do not use it for this lab (avoids incurring charges).

5. Click through the subsequent pages:

   - Accept the terms and conditions.
   - Do not add recovery options or two-factor authentication (because this is a temporary account).
   - Do not sign up for free trials.

After a few moments, the GCP console opens in this tab.

**Note:** You can view the menu with a list of GCP Products and Services by clicking the **Navigation menu** at the top-left, next to "Google Cloud Platform".
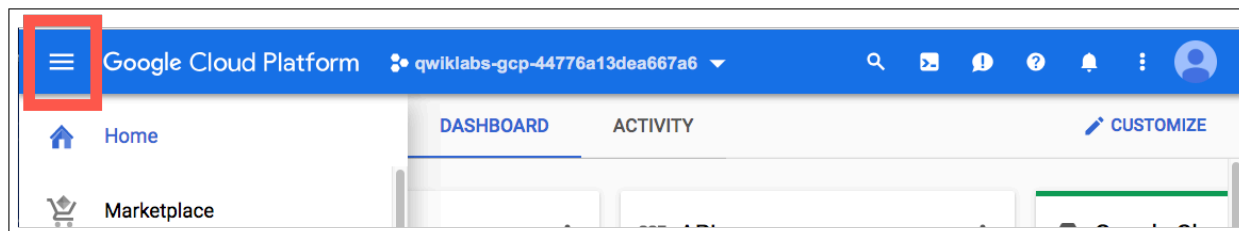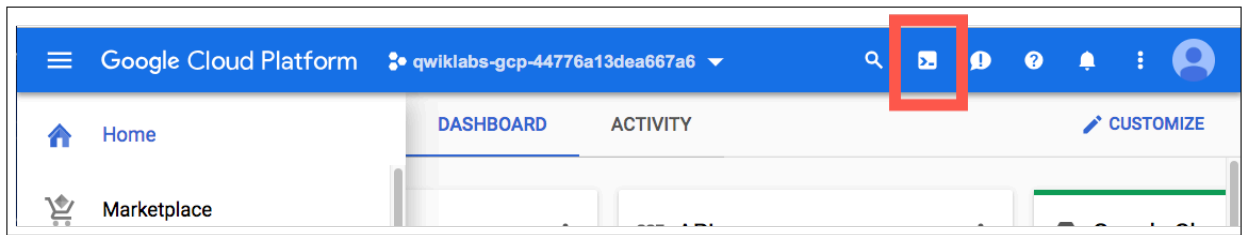


# Activate Google Cloud Shell

Google Cloud Shell is a virtual machine that is loaded with development tools. It offers a persistent 5GB home directory and runs on the Google Cloud. Google Cloud Shell provides command-line access to your GCP resources.

1. In GCP console, on the top right toolbar, click the Open Cloud Shell button.

   NOTE:  the button is labeled "Activate"

2. Click **Continue**.
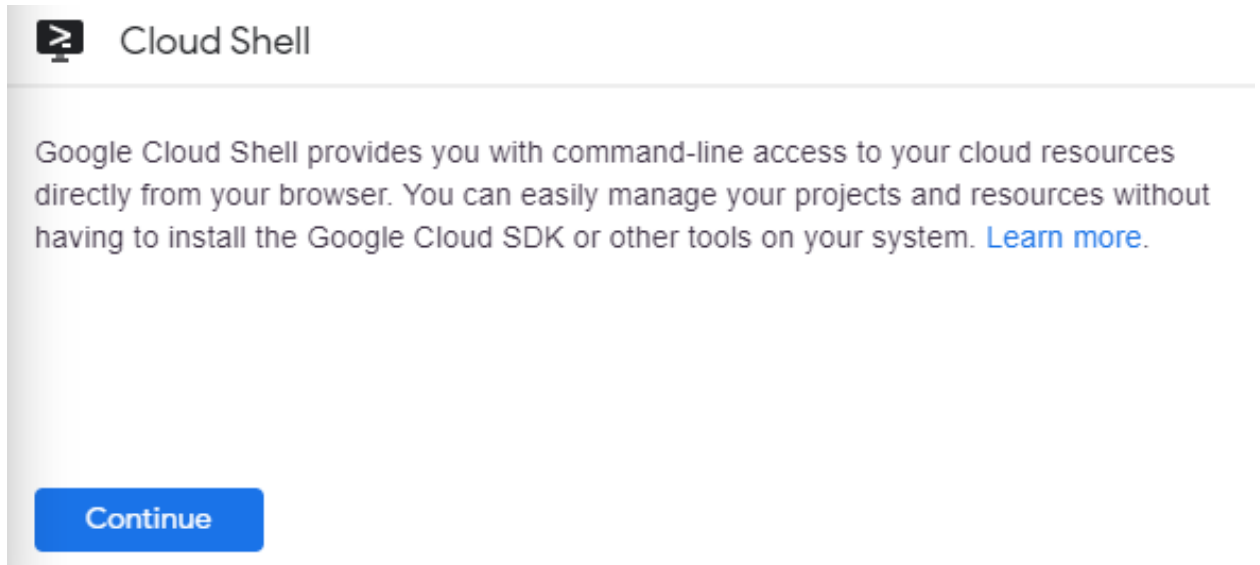


It takes a few moments to provision and connect to the environment. When you are connected, you are already authenticated, and the project is set to your *PROJECT_ID*. For example:
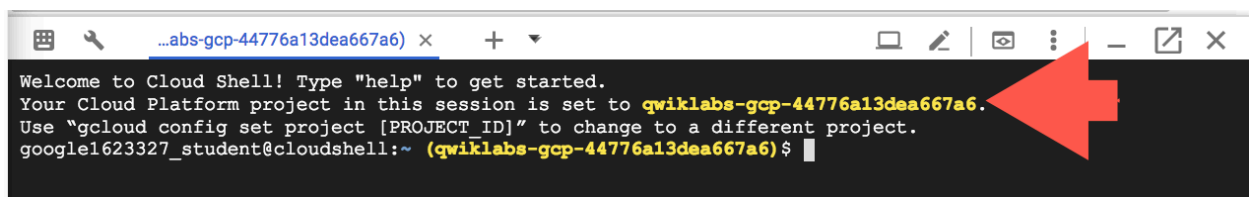


**gcloud** is the command-line tool for Google Cloud Platform. It comes pre-installed on Cloud Shell and supports tab-completion.

You can list the active account name with this command:

```
gcloud auth list
```

Output:

```
Credentialed accounts:
```

```
- <myaccount>@<mydomain>.com (active)
```

Example output:

```
Credentialed accounts:
 - google1623327_student@qwiklabs.net
```

You can list the project ID with this command:

```
gcloud config list project
```

Output:

```
[core]
project = <project_ID>
```

Example output:

```
[core]
project = qwiklabs-gcp-44776a13dea667a6
```

Full documentation of **gcloud** is available on [Google Cloud gcloud Overview](#).

# Deploy Cassandra™

1. Browse to the GCP Marketplace in the left hand navigation menu.
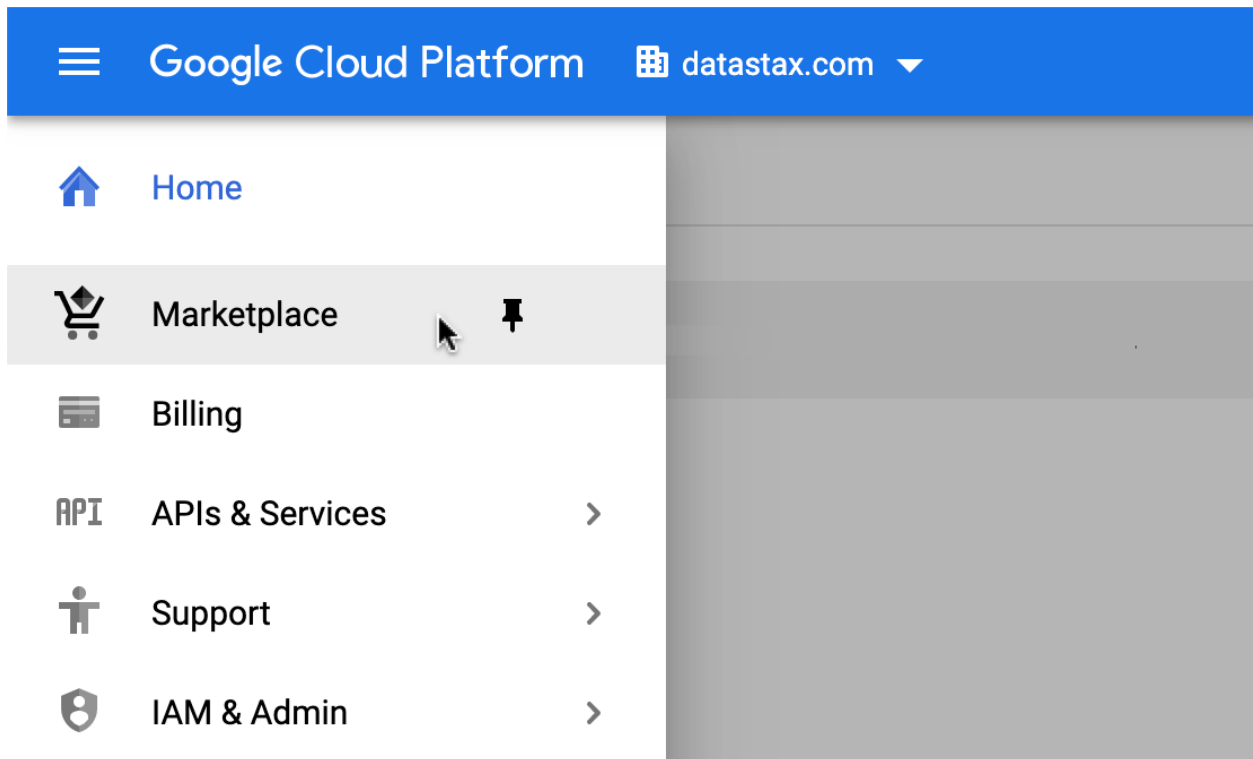
2. Search for "apache cassandra" and click on the search result named "Cassandra Certified by Bitnami".
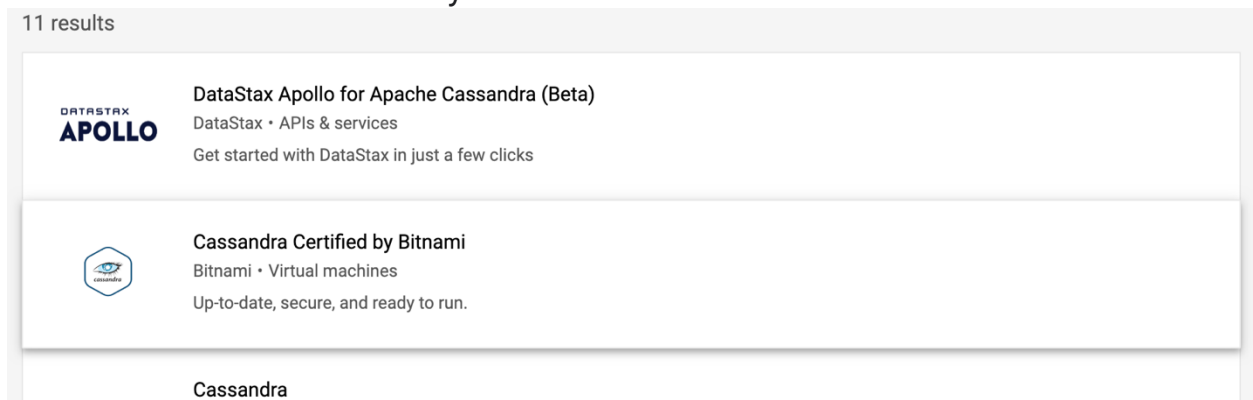


3. Click the "LAUNCH" button.

Cassandra Certified by Bitnami

Bitnami

Estimated costs: $41.45/month

Up-to-date, secure, and ready to run.

LAUNCH

4.  Leave all the default values and click "Deploy".

**Deployment name**

cassandra-1

**Zone** ❓

us-east4-a ▼

**Machine type** ❓

2 vCPUs ▼       4 GB memory        Customize

## Boot Disk

**Boot disk type** ❓

Standard Persistent Disk ▼

**Boot disk size in GB** ❓

10

## Networking

**Network interfaces**

**default** default (10.150.0.0/20)        ✏️

➕ Add network interface

ℹ️ You have reached the maximum number of one network interface

Deploy

When the database has finished deploying you will see the details in Deployment Manager.

<mark>Click *Check my progress* to verify the objective.</mark>

<mark>Deploy Cassandra Database using the GCP Marketplace</mark>

<mark>Check my progress</mark>

I never saw this…

AAHHH – it's in the instructions, NOT in the window.

# Connect to the database

1.  Note the Admin user and Admin password on the right hand side.

2. To connect to the database, open up an SSH session by clicking the SSH dropdown menu and selecting "Open in browser window".

3. Open up the CQL Shell by typing the following into the command line, replacing `<<user>>` and `<<pass>>` with the username and password you just made a note of.

```
cqlsh -u <<user>> -p <<pass>>
```

```
                                        ***
ifinlayson@cassandra-1-vm:~$ cqlsh -u cassandra -p DfxSsKZAQa1S
Connected to My Cluster at 127.0.0.1:9042.
[cqlsh 5.0.1 | Cassandra 3.11.5 | CQL spec 3.4.4 | Native protocol v4]
Use HELP for help.
cassandra@cqlsh> █
```

# Create a keyspace and table

1. Once you are logged in to the database, create a keyspace.

```
CREATE KEYSPACE space_flights WITH replication = {'class': 'SimpleStrategy',
'replication_factor': 1};
```

2. Create a table in the keyspace

```
CREATE TABLE space_flights.catalog (

    spacecraft_name text,

    journey_id timeuuid,

    start timestamp,

    end timestamp,

    active boolean,

    summary text,

    PRIMARY KEY ((spacecraft_name), journey_id)

    ) WITH CLUSTERING ORDER BY (journey_id desc);
```

You can view the keyspace and table you created using the following commands.

```
DESCRIBE KEYSPACE space_flights;



DESCRIBE TABLE space_flights.catalog;
```

Note that the output includes default values for a number of settings that you did not specify when you ran the CREATE statements.

```
cassandra@cqlsh> DESCRIBE KEYSPACE space_flights;

CREATE KEYSPACE space_flights WITH replication = {'class': 'SimpleStrategy', 'replication_factor': '1'}  AND durable_writes = true;

CREATE TABLE space_flights.catalog (
    spacecraft_name text,
    journey_id timeuuid,
    active boolean,
    end timestamp,
    start timestamp,
    summary text,
    PRIMARY KEY (spacecraft_name, journey_id)
) WITH CLUSTERING ORDER BY (journey_id DESC)
    AND bloom_filter_fp_chance = 0.01
    AND caching = {'keys': 'ALL', 'rows_per_partition': 'NONE'}
    AND comment = ''
    AND compaction = {'class': 'org.apache.cassandra.db.compaction.SizeTieredCompactionStrategy', 'max_threshold': '32', 'min_threshold': '4'}
    AND compression = {'chunk_length_in_kb': '64', 'class': 'org.apache.cassandra.io.compress.LZ4Compressor'}
    AND crc_check_chance = 1.0
    AND dclocal_read_repair_chance = 0.1
    AND default_time_to_live = 0
    AND gc_grace_seconds = 864000
    AND max_index_interval = 2048
    AND memtable_flush_period_in_ms = 0
    AND min_index_interval = 128
    AND read_repair_chance = 0.0
    AND speculative_retry = '99PERCENTILE';

cassandra@cqlsh> ▮
```

Click *Check my progress* to verify the objective.

Create a keyspace and table

Check my progress

# Insert data into the table

1.  Insert the following data

```
INSERT INTO space_flights.catalog (spacecraft_name, journey_id, start, end,
active, summary) VALUES ('vostok1', 805b1a00-5673-11a8-8080-808080808080,
'1961-4-12T06:07:00+0000', '1961-4-12T07:55:00+0000', False, 'First manned
spaceflight. Completed one Earth orbit.');

INSERT INTO space_flights.catalog (spacecraft_name, journey_id, start, end,
active, summary) VALUES ('mercury-redstone3', 2396fc00-68cd-11a8-8080-
808080808080, '1961-5-5T14:34:00+0000', '1961-5-5T14:49:00+0000', False,
'First American manned suborbital spaceflight (altitude 187 kilometres, 116
miles).');
```

```
INSERT INTO space_flights.catalog (spacecraft_name, journey_id, start, end,
active, summary) VALUES ('mercury-redstone4', 2d2f1800-a53c-11a8-8080-
808080808080, '1961-7-21T12:20:00+0000', '1961-7-21T12:35:00+0000', False,
'Second American manned suborbital flight (altitude 118.26mi, 190km).');

INSERT INTO space_flights.catalog (spacecraft_name, journey_id, start, end,
active, summary) VALUES ('vostok2', 5c2ac800-b191-11a8-8080-808080808080,
'1961-8-6T05:00:00+0000', '1961-8-7T05:01:00+0000', False, 'Day-long flight.
Completed 17 Earth orbits. Brief manual control by pilot.');

INSERT INTO space_flights.catalog (spacecraft_name, journey_id, start, end,
active, summary) VALUES ('mercury-atlas6', 8c7b3200-4d82-11a9-8080-
808080808080, '1962-2-20T15:47:00+0000', '1962-2-20T20:42:00+0000', False,
'First American manned orbital flight. Completed three orbits.');

INSERT INTO space_flights.catalog (spacecraft_name, journey_id, start, end,
active, summary) VALUES ('mercury-atlas7', e9d69600-9685-11a9-8080-
808080808080, '1962-5-24T13:45:00+0000', '1962-5-24T18:41:00+0000', False,
'First manual retrofire. Earth photography and study of liquids in weightless
conditions.');

INSERT INTO space_flights.catalog (spacecraft_name, journey_id, start, end,
active, summary) VALUES ('vostok3', ff31b400-d46d-11a9-8080-808080808080,
'1962-8-11T08:30:00+0000', '1962-8-15T06:52:00+0000', False, 'First instance
of two manned spacecraft in orbit simultaneously.');

INSERT INTO space_flights.catalog (spacecraft_name, journey_id, start, end,
active, summary) VALUES ('vostok4', 403fcc00-d533-11a9-8080-808080808080,
'1962-8-12T08:02:00+0000', '1962-8-15T06:59:00+0000', False, 'First instance
of two manned spacecraft in orbit simultaneously.');

INSERT INTO space_flights.catalog (spacecraft_name, journey_id, start, end,
active, summary) VALUES ('mercury-atlas8', 977b6200-fe3b-11a9-8080-
808080808080, '1962-10-3T13:15:00+0000', '1962-10-3T22:28:00+0000', False,
'First flawless Mercury mission.');
```

# Query the table

1. Select all of the records from the table

```
select * from space_flights.catalog;
```

```
spacecraft_name   | journey_id                              | active | end                            | start                          | summary
------------------+-----------------------------------------+--------+--------------------------------+--------------------------------+----------------------------------------------------------------------
   mercury-atlas7 | e3d69600-9685-11a9-8080-808080808080    | False  | 1962-05-24 18:41:00.000000+0000 | 1962-05-24 13:45:00.000000+0000 | First manual retrofire. Earth photography and study of liquids in weightless conditions.
 mercury-redstone4 | 2d2f1800-a53c-11a8-8080-808080808080    | False  | 1961-07-21 12:35:00.000000+0000 | 1961-07-21 12:20:00.000000+0000 | Second American manned suborbital flight (altitude 118.26mi, 190km).
   mercury-atlas6 | 8c7b3200-4d82-11a9-8080-808080808080    | False  | 1962-02-20 20:42:00.000000+0000 | 1962-02-20 15:47:00.000000+0000 | First American manned orbital flight. Completed three orbits.
   mercury-atlas8 | 977b6200-fe3b-11a9-8080-808080808080    | False  | 1962-10-03 22:28:00.000000+0000 | 1962-10-03 13:15:00.000000+0000 | First flawless Mercury mission.
          vostok3 | ff31b400-d46d-11a9-8080-808080808080    | False  | 1962-08-15 06:52:00.000000+0000 | 1962-08-11 08:30:00.000000+0000 | First instance of two manned spacecraft in orbit simultaneously.
 mercury-redstone3 | 2396bc00-68cd-11a8-8080-808080808080    | False  | 1961-05-05 14:49:00.000000+0000 | 1961-05-05 14:34:00.000000+0000 | First American manned suborbital spaceflight (altitude 187 kilometres, 116 miles).
          vostok2 | 5c2ac800-b191-11a8-8080-808080808080    | False  | 1961-08-07 05:01:00.000000+0000 | 1961-08-06 05:00:00.000000+0000 | Day-long flight. Completed 17 Earth orbits. Brief manual control by pilot.
          vostok1 | 805b1a00-5673-11a8-8080-808080808080    | False  | 1961-04-12 07:55:00.000000+0000 | 1961-04-12 06:07:00.000000+0000 | First manned spaceflight. Completed one Earth orbit.
          vostok4 | 403fcc00-d533-11a9-8080-808080808080    | False  | 1962-08-15 06:59:00.000000+0000 | 1962-08-12 08:02:00.000000+0000 | First instance of two manned spacecraft in orbit simultaneously.

(9 rows)
```

2. Select one of the records by its primary key

```
SELECT * FROM space_flights.catalog WHERE spacecraft_name = 'vostok2';
```

```
spacecraft_name | journey_id                           | active | end                            | start                          | summary
----------------+--------------------------------------+--------+--------------------------------+--------------------------------+--------------------------------------------------------------------------
        vostok2 | 5c2ac800-b191-11a8-8080-808080808080 | False  | 1961-08-07 05:01:00.000000+0000 | 1961-08-06 05:00:00.000000+0000 | Day-long flight. Completed 17 Earth orbits. Brief manual control by pilot.

(1 rows)
```

3. Exit the CQL shell

```
cqlsh> exit
```

4. Close the SSH window.


Click *Check my progress* to verify the objective.

Insert data into the table

Check my progress