Group Member: Tung Doan, Ibrahim Dawha, Dakota Hernandez, Shuxiao Xie

12/20/2017

Apache Cassandra – Tutorial

**PART I – Background and Properties**

**Four Parts of Cassandra Tutorial**

* History and Background
* Installation
* Speed Contest *Vs* MongoDB
* Generate A Speed Layer

**History and Background**

* Apache Cassandra - An open source, distributed /decentralized database, provides highly available service with no single point of failure.
* Notable Features:
  + **Elastic scalability** - Cassandra is highly scalable; it allows to add more hardware to accommodate more customers and more data as per requirement.
  + **Always on architecture -** Cassandra has no single point of failure and it is continuously available for business-critical applications that cannot afford a failure.
  + **Fast linear-scale performance** - Cassandra is linearly scalable. Therefore, it maintains a quick response time.
  + **Flexible data storage** - Dynamically accommodate structured, semi-structured, and unstructured according to need.
  + **Easy data distribution** - Cassandra provides the flexibility to distribute data where you need by replicating data across multiple data centers.
  + **Transaction support** - Cassandra supports properties like Atomicity, Consistency, Isolation, and Durability (ACID).
  + **Fast writes** - run on cheap commodity hardware. It performs blazingly fast writes and can store hundreds of terabytes of data, without sacrificing the read efficiency.

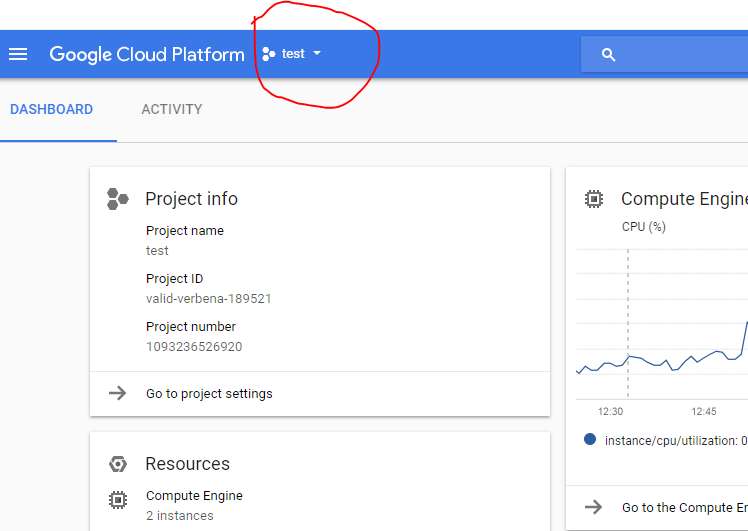
**Pros:**

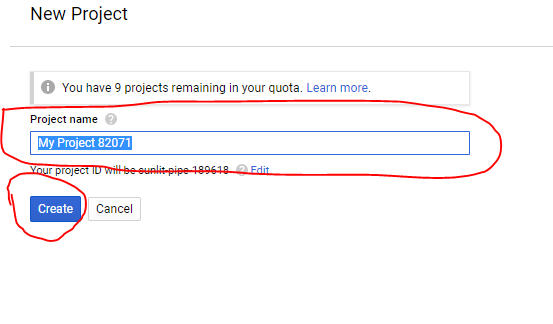
* It is scalable, fault-tolerant, and consistent.
* It is a column-oriented database.
* Its distribution design is based on Amazon’s Dynamo and its data model on Google’s Bigtable.
* Created at Facebook, it differs sharply from relational database management systems.
* Cassandra implements a Dynamo-style replication model with no single point of failure, but adds a more powerful “column family” data model.
* Cassandra is being used by some of the biggest companies such as Facebook, Twitter, Cisco, Rackspace, ebay, Twitter, Netflix, and more.

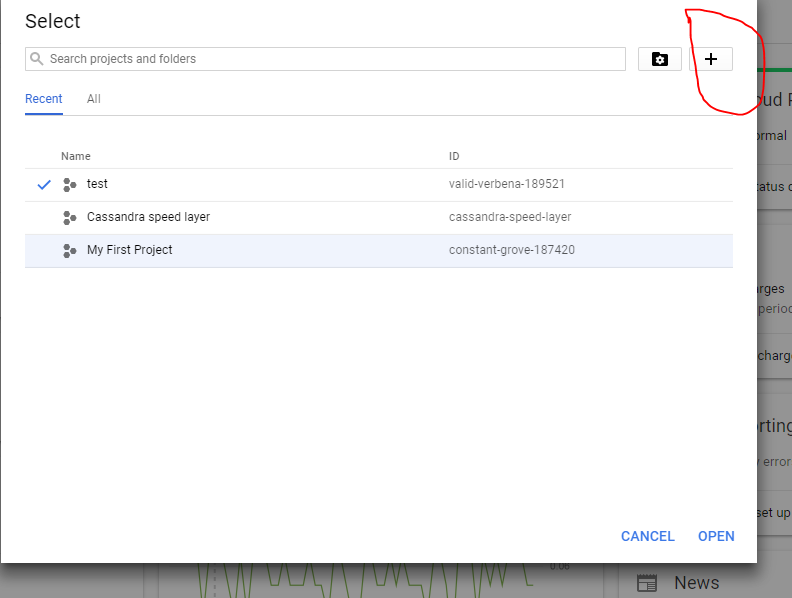
**PART II – Installation**

SETUP GOOGLE CLOUD

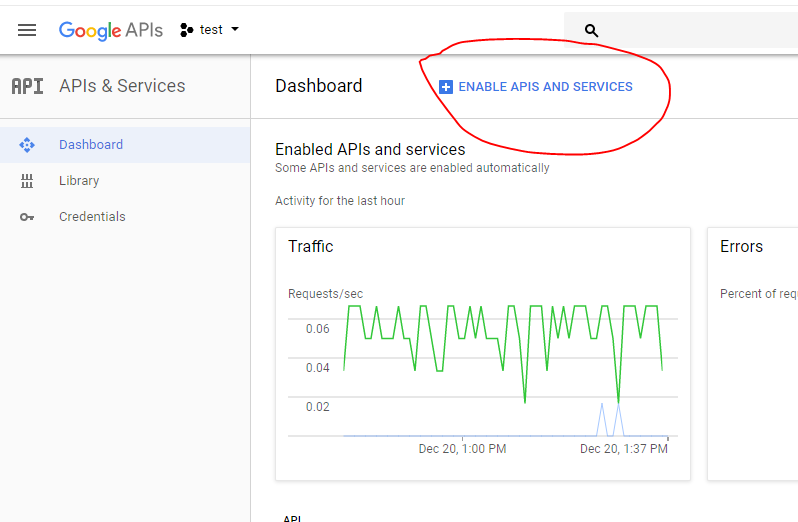
1. go to “https;//cloud.google.com” and sign up for the free trial with your google account
2. After signing up go to “<http://console.cloud.google.com>” and create a new project by following the steps below

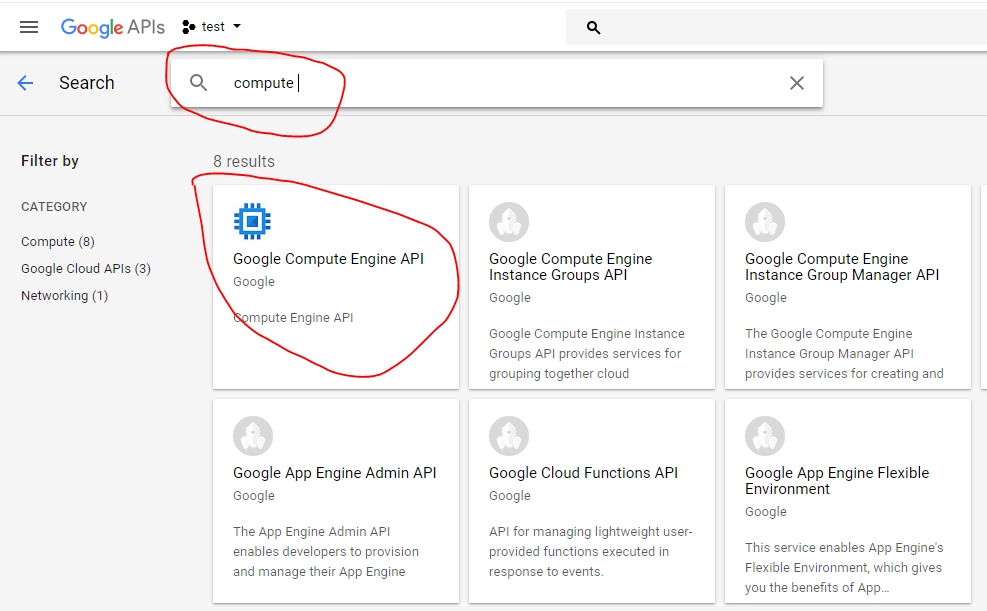
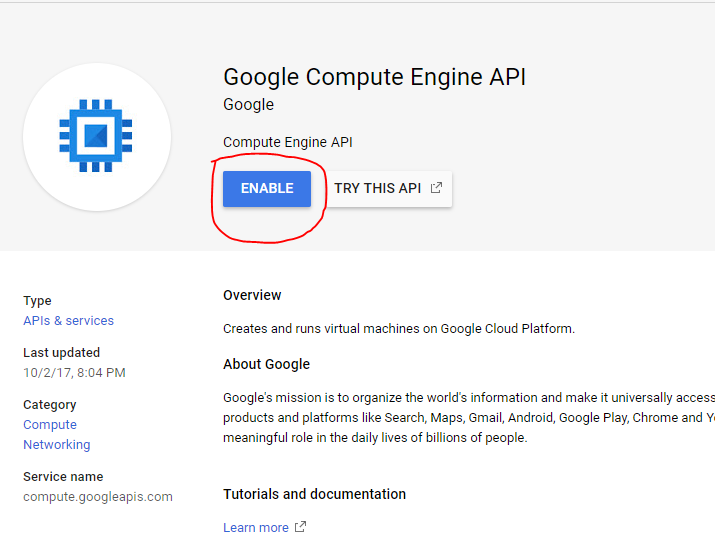




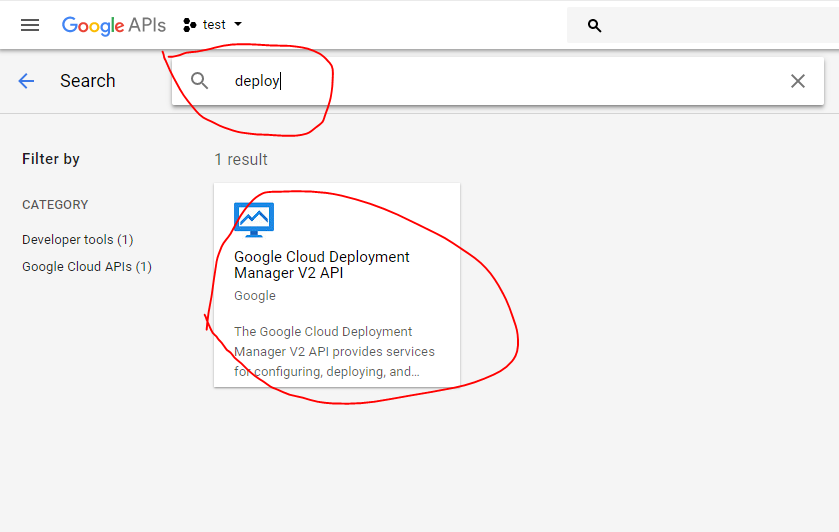


3. Enable the compute engine and deploy manager api from the console



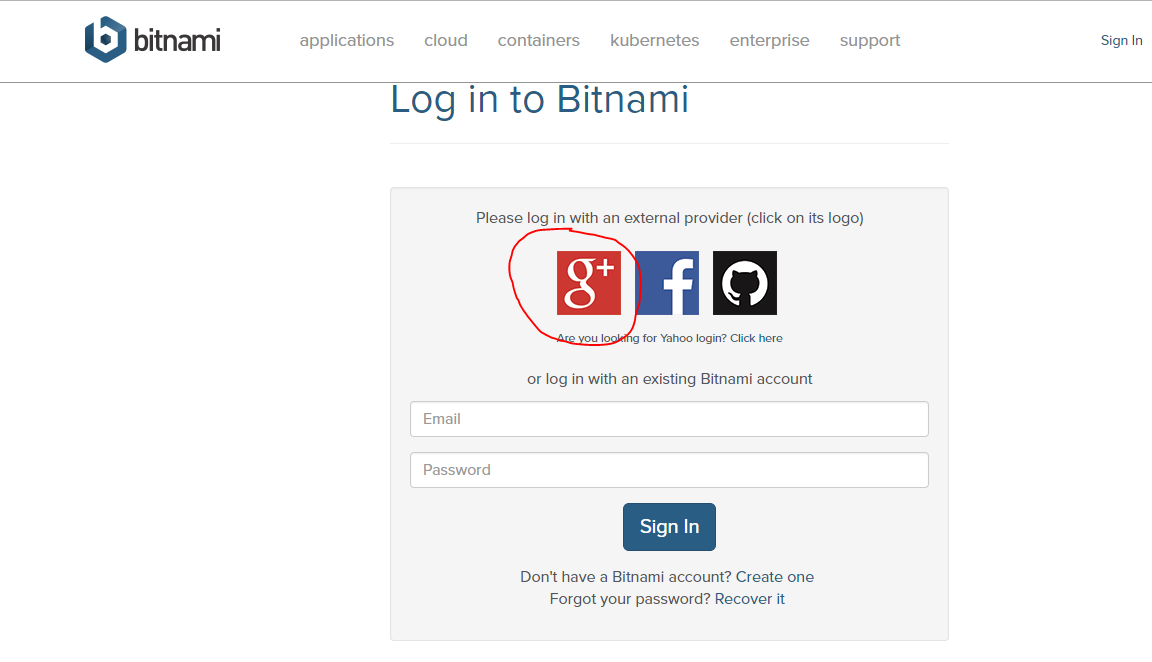


**4.** Repeat the same steps for the deployment manager

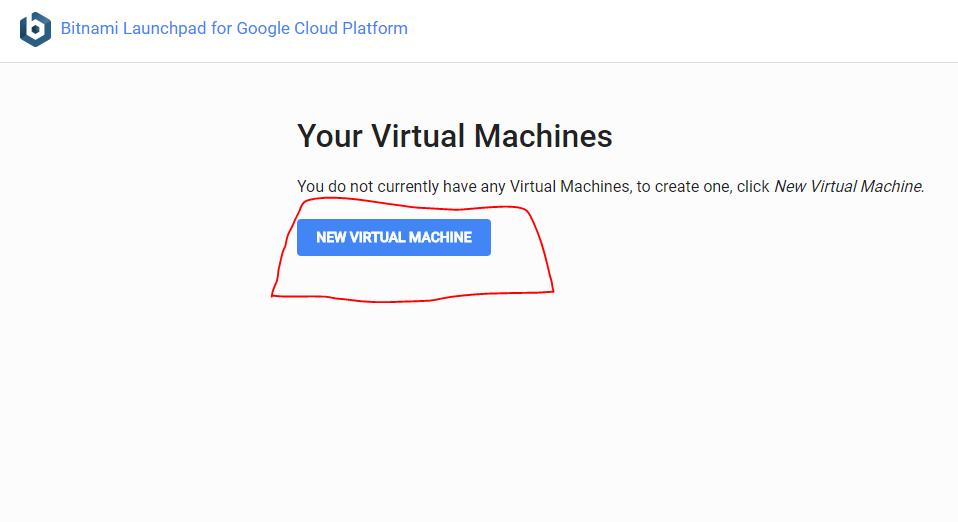


**Deploying cassandra on Bitnami**

1. Go to the URL <https://bitnami.com/sign_in> and login with your own google account:



2. Select virtual machines from the top right of the screen and select create a new virtual machine

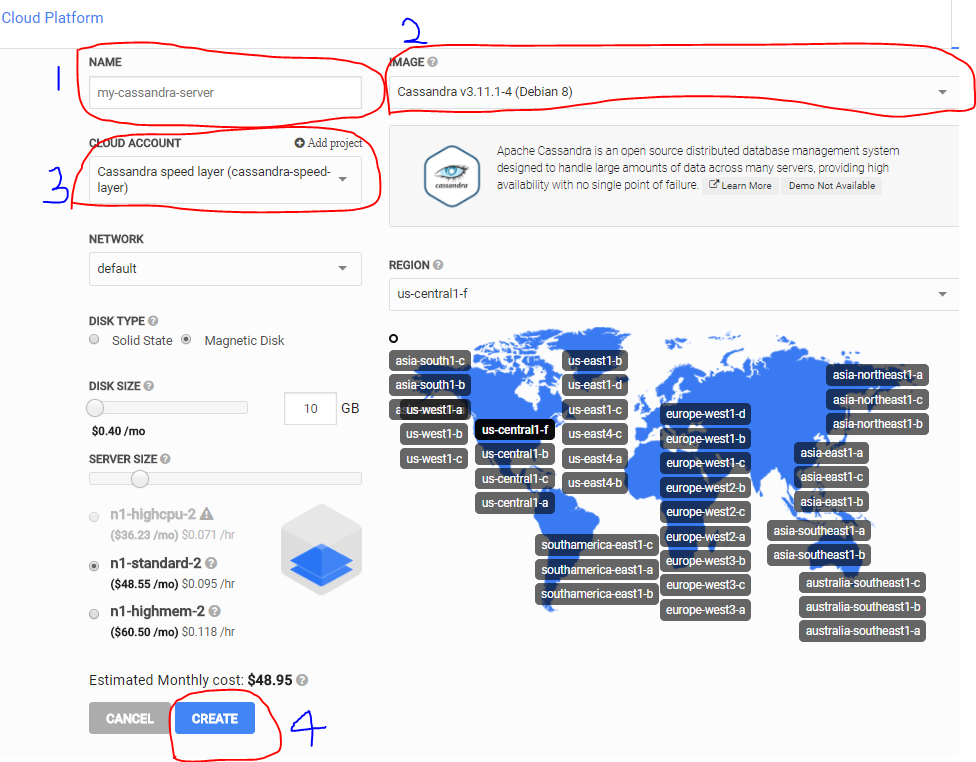


3. a). Type in the name you want to name your Cassandra vm (Step 1)

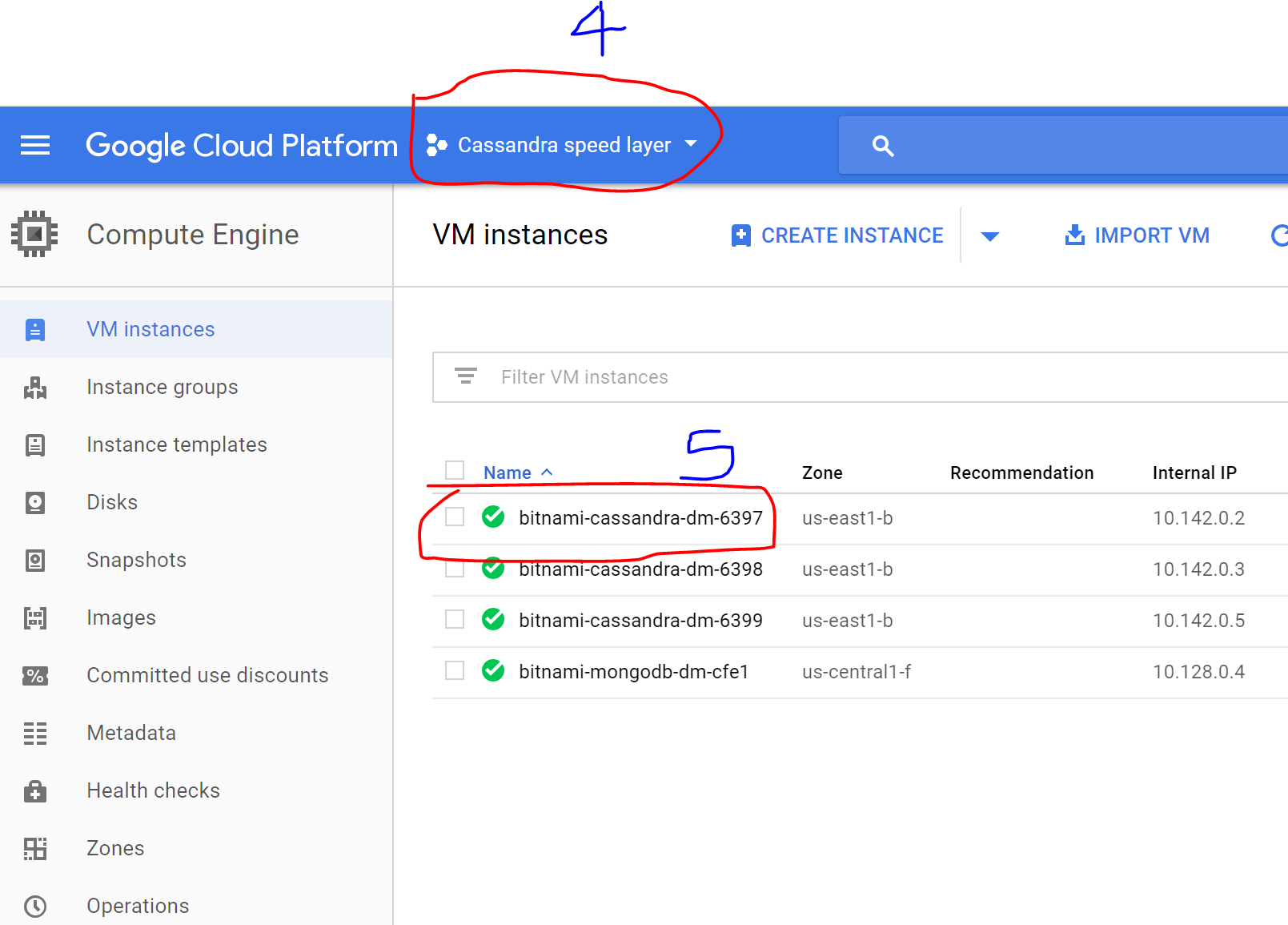
b). Select Cassandra from the list of virtual machine instances (Step 2)

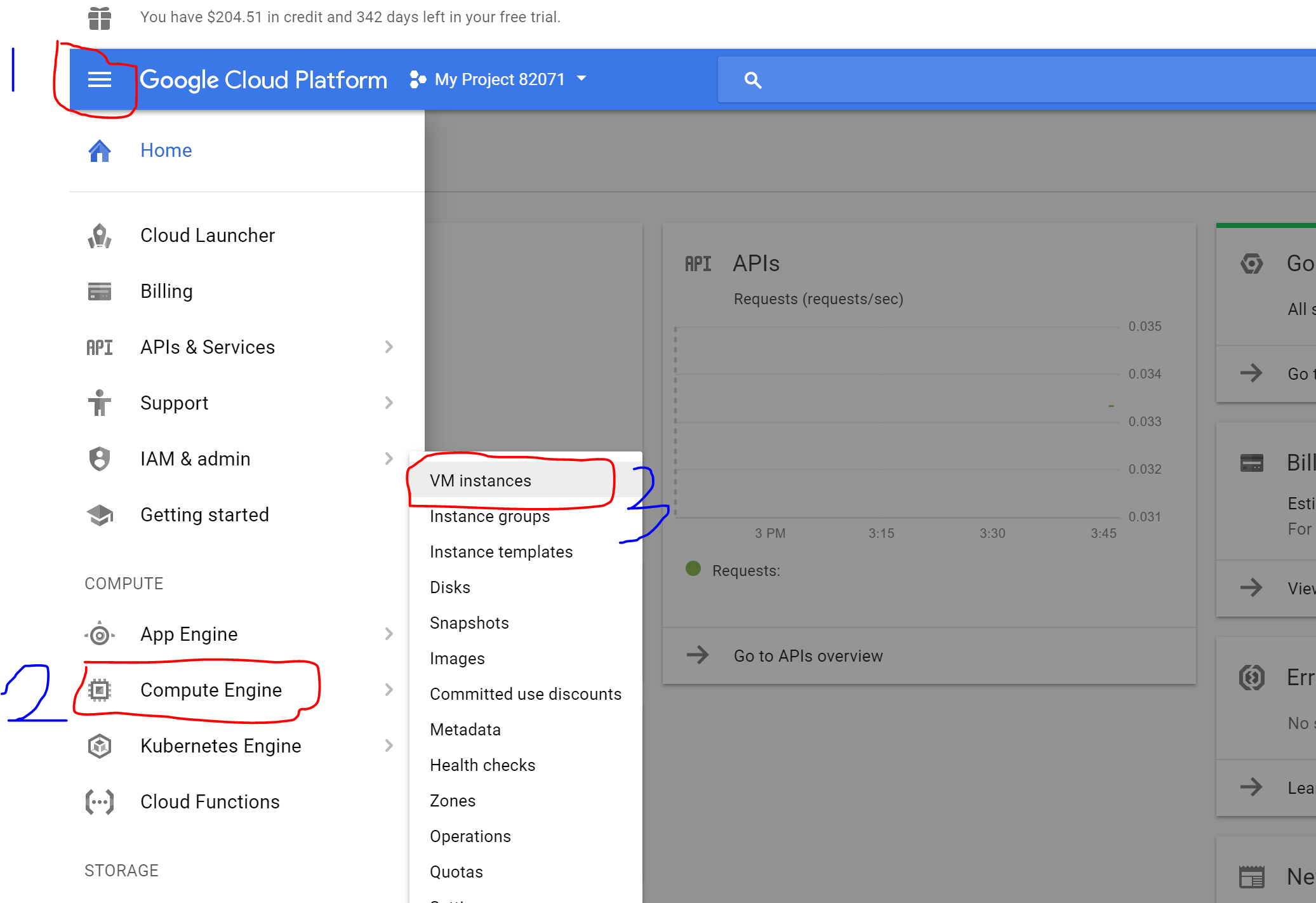
c). Select the google cloud project (project we just created) (Step 3)

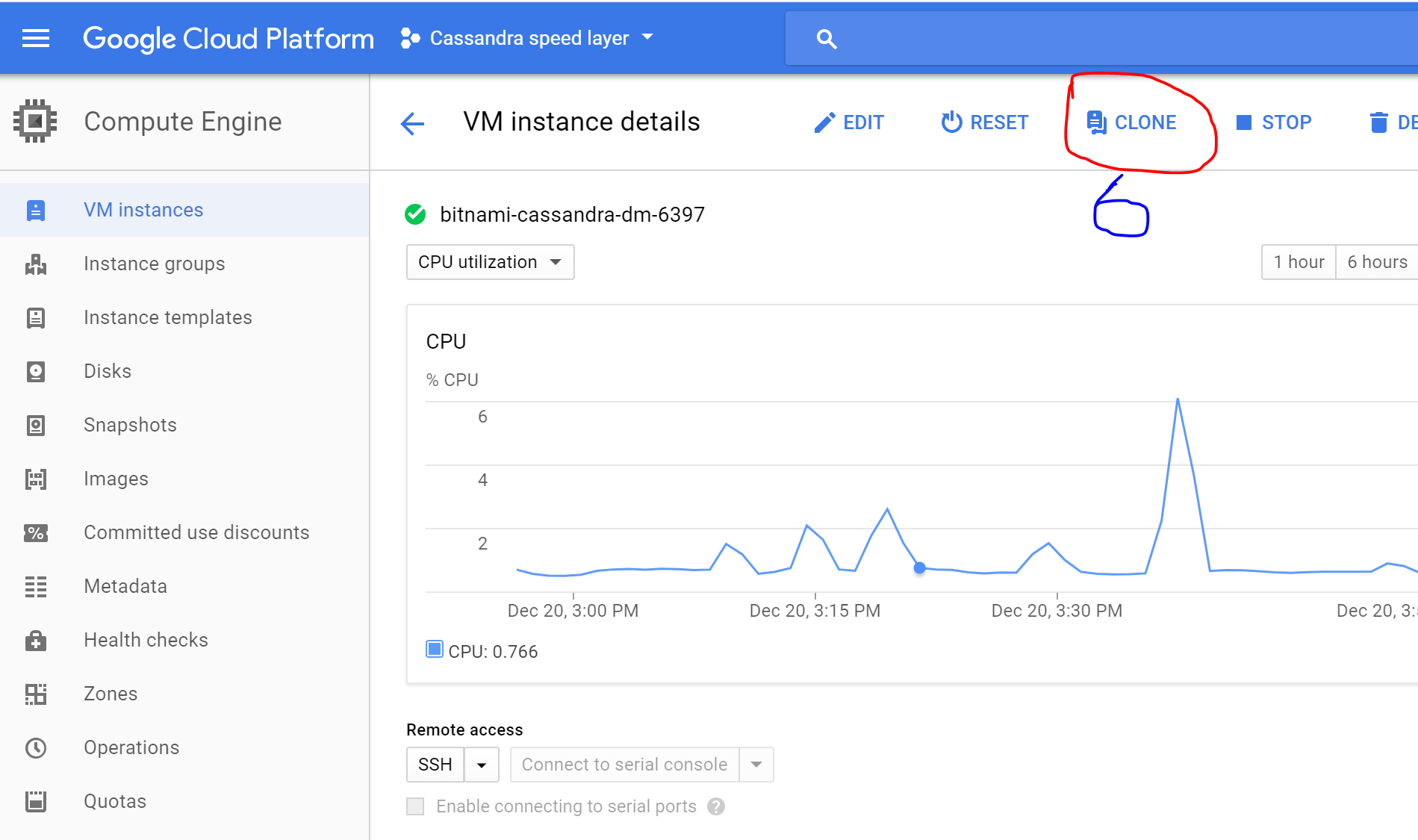
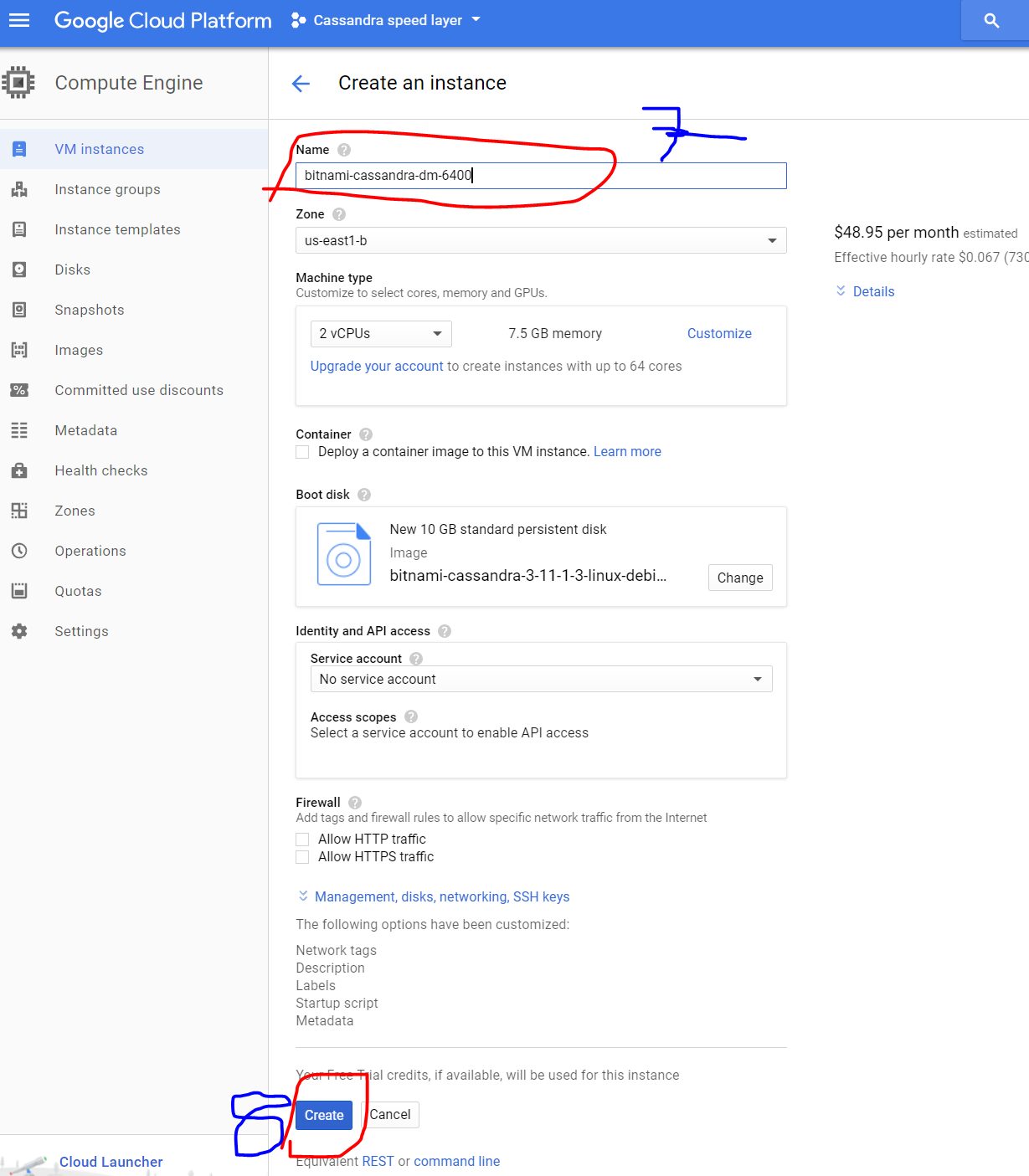
d) Select create to create the project (Step 4)



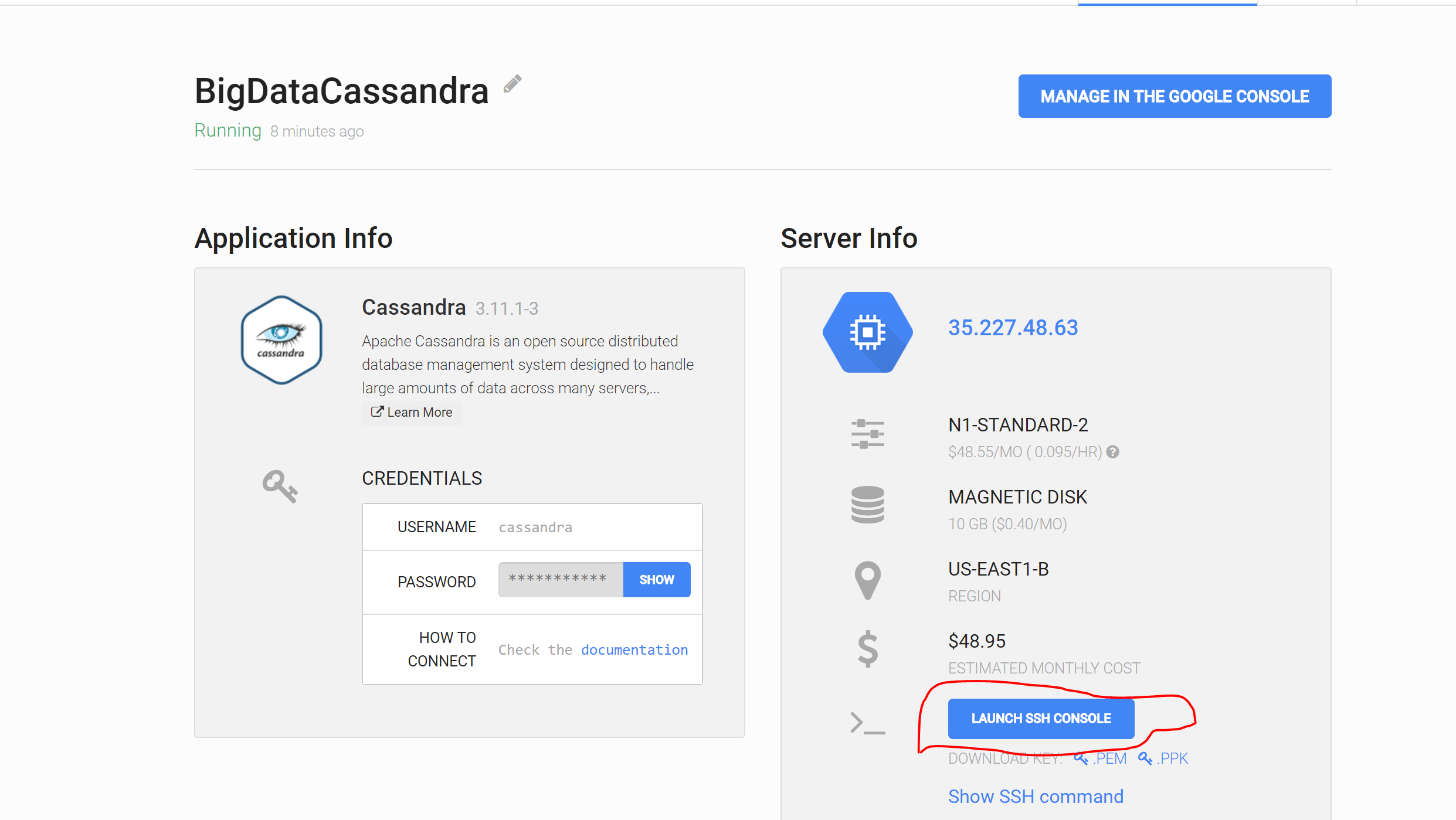
**CREATE MULTIPLE NODES STEPS**

Go to the google cloud console and follow the steps below to clone and create multiple cassandra nodes





**Multiple Node Cluster Setup**

1. Go to the bitnami console using “<https://google.bitnami.com>
2. Sign in and select virtual machines from the top right of the screen
3. Select the cassandra virtual machine you have setup and the screen below will show up\

4. There are multiple ways to do the parts from here, either with a file transfer service like filezilla setup steps here: “<https://docs.bitnami.com/google/infrastructure/cassandra/#filezilla>” or use the web console which is what i will be explaining.

1. Launch the web console by clicking on the circled link above
2. In the console type this command to **STOP** cassandra (this is important)
   1. sudo /opt/bitnami/ctlscript.sh stop cassandra
3. Install the emacs editor to the VM using the command below
   1. sudo apt-get install emacs
4. Run this command to open the cassandra configuration file using emacs
   1. emacs */opt/bitnami/cassandra/conf/cassandra.yaml*
5. Edit the following lines:
   1. cluster\_name : <Your name for the cluster> (*same for all nodes)*
   2. seeds: <Ip address for all seed nodes separated by commas >
   3. broadcast\_address: <host machine’s public ip address>
   4. listen\_address: <host machines private ip address>
   5. endpoint\_snitch: GoogleCloudSnitch
6. Save the file by typing <ctrl x and ctrl s> on your keyboard

5. Remove any data from cassandra using command below (Important step)

sudo rm -rf /opt/bitnami/cassandra/data/\*