**Migrate data from Hive table to Redis**

**Pre-Requisites:**

* Install git
* Install Maven
* Install Redis
* EMR Cluster
* Need data in Have Table

**Install Git:**

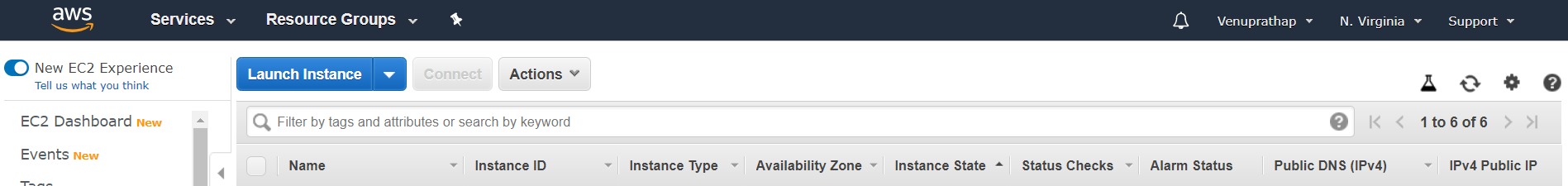
Install git on EMR Cluster

yum install git -y

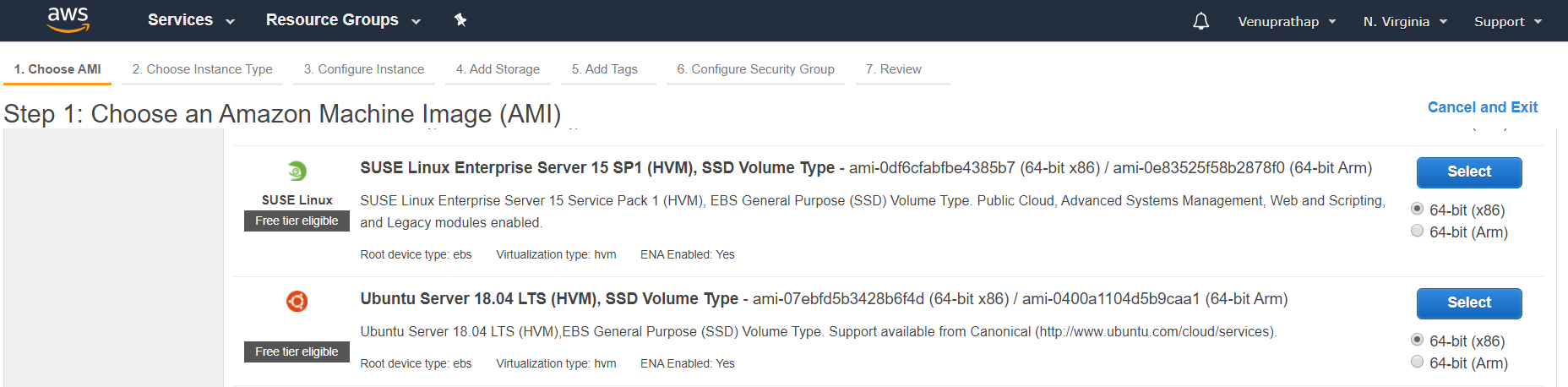
**Install Redis:**

Here we installed Redis in Ubuntu Server

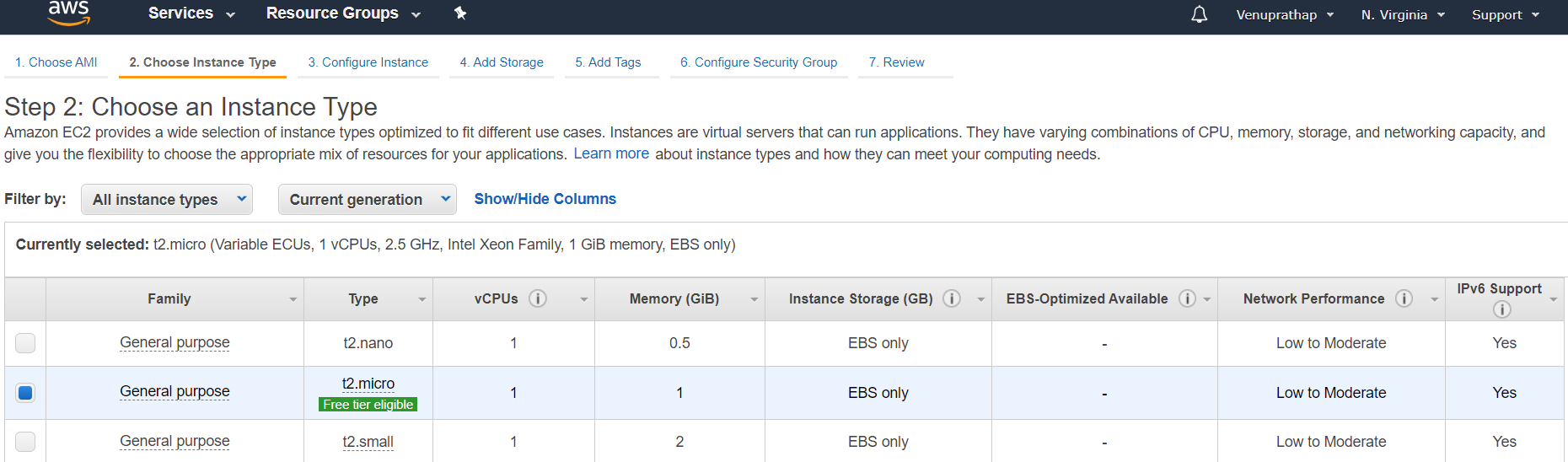
Goto EC2 Service 🡪 Instances



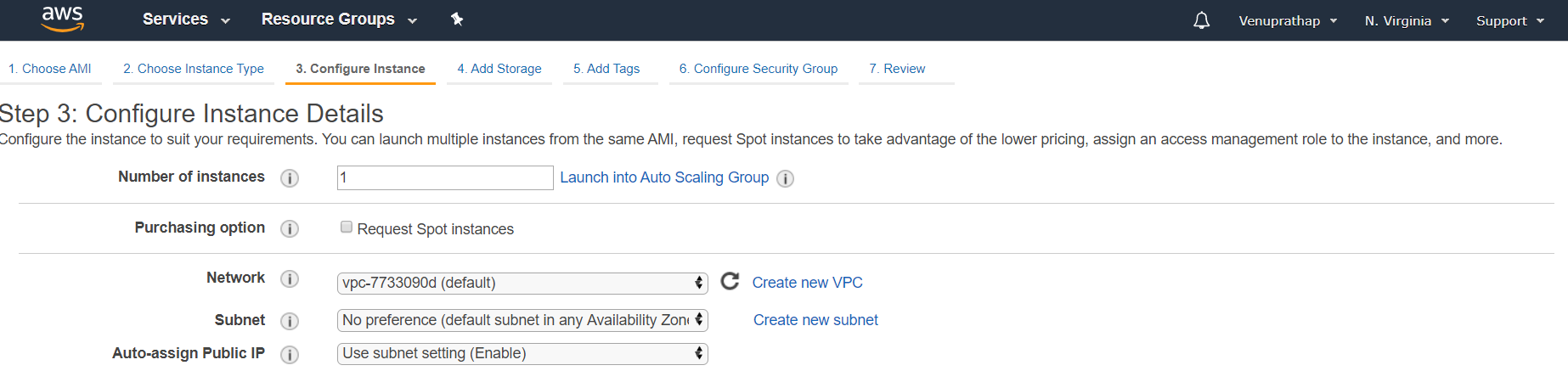
* Click on Launch Instance



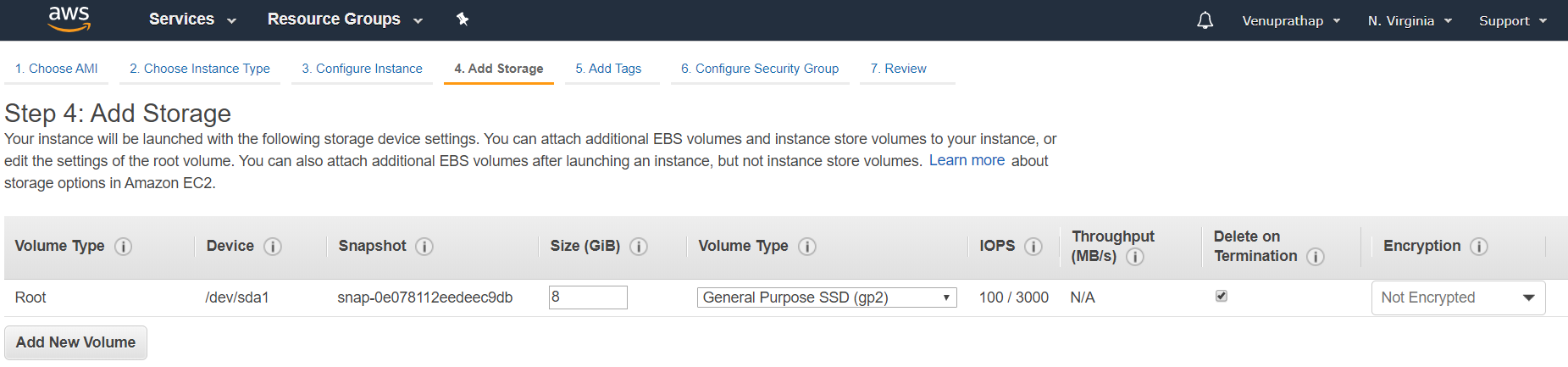
* Take Ubuntu Server 18.04 and Select



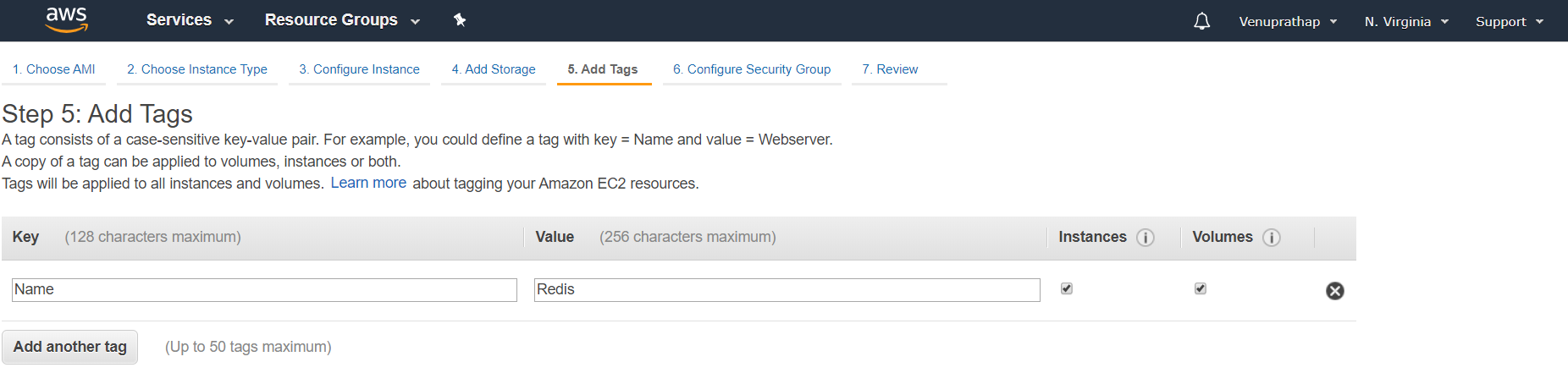
* Click Next



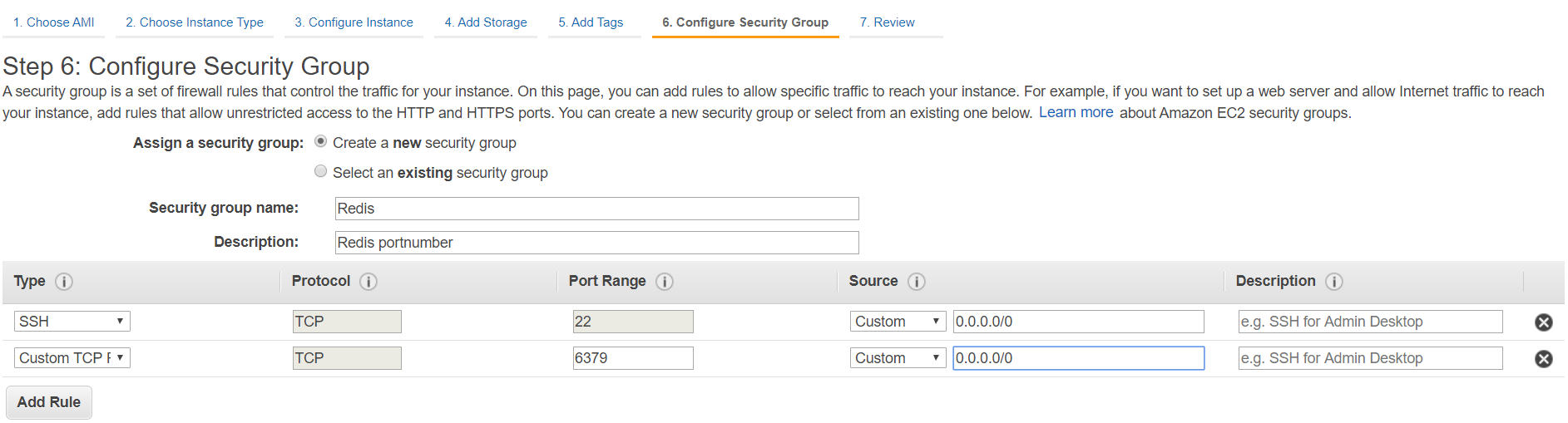
* Click Next



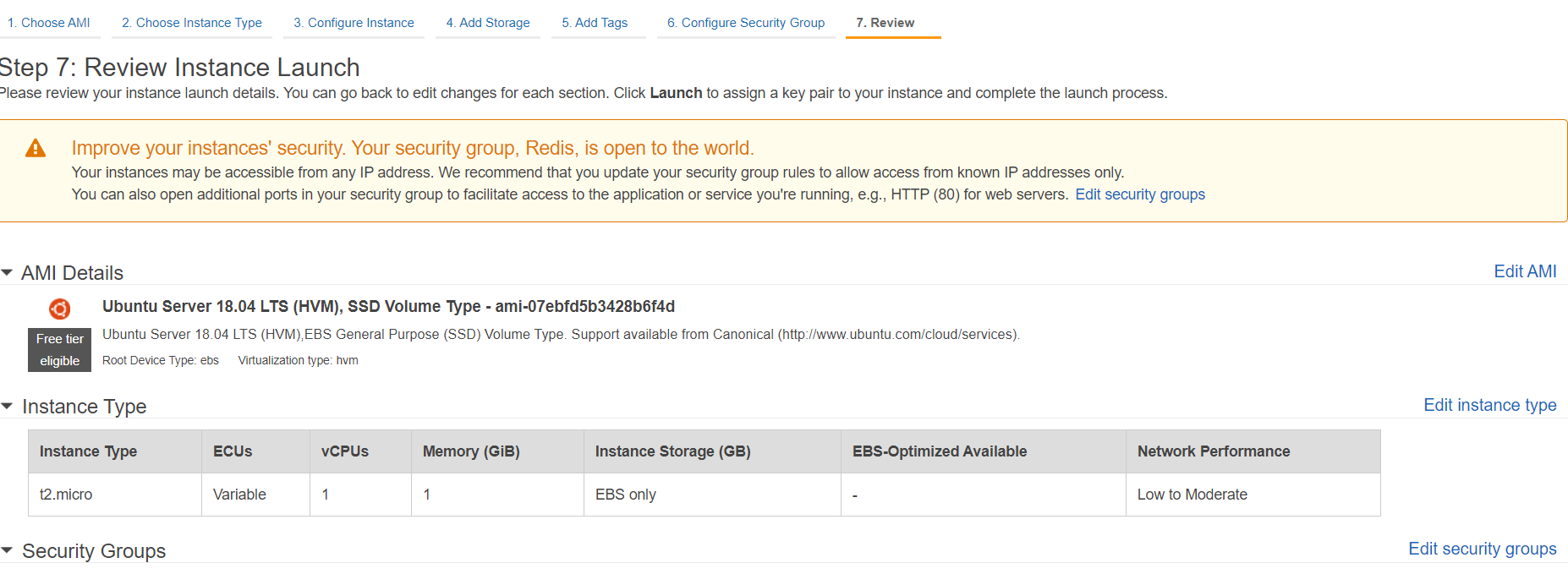
* Click on Next



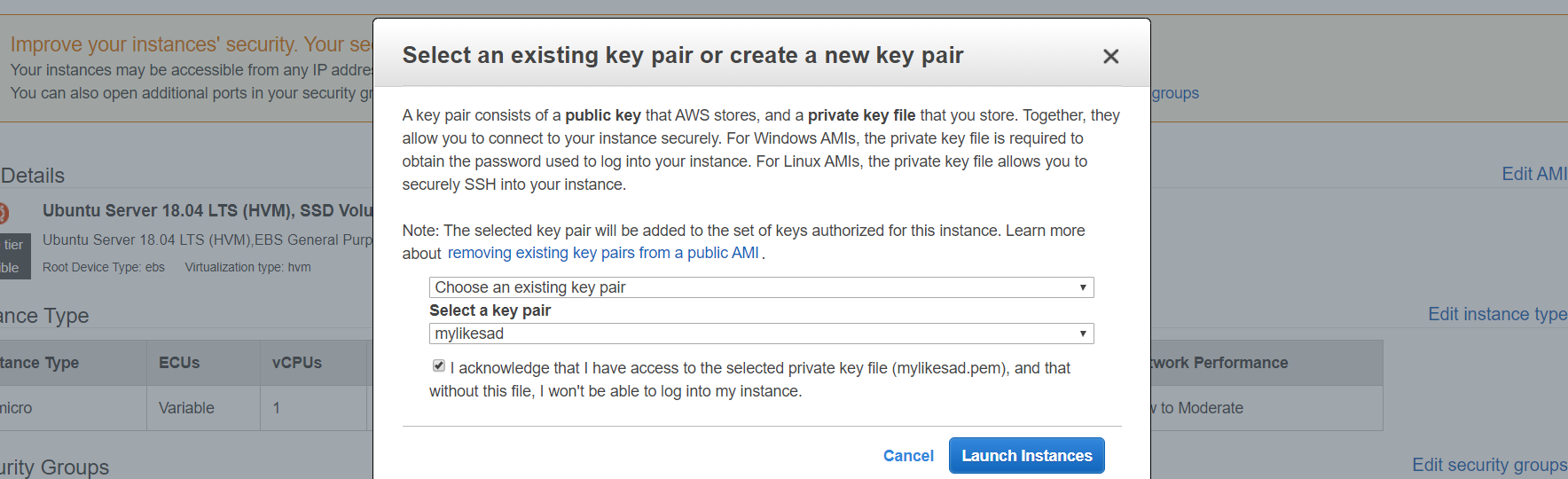
* Click on Next



Allow port number “6379” and Click on Review and Launch



* Click on Launch



* Click on Launch Instances
* Click on View Instances

**Now connect to our Instance and Install Redis:**

**Step 1 – Prerequsities**

sudo apt-get update

sudo apt-get upgrade -y

**Step 2 – Installing Redis**

Run below command from the terminal to install Redis on your machine:

sudo apt-get install redis-server -y

Next is to enable Redis to start on system boot. Also restart Redis service once

sudo systemctl enable redis-server.service

**Step 3 – Configure Redis**

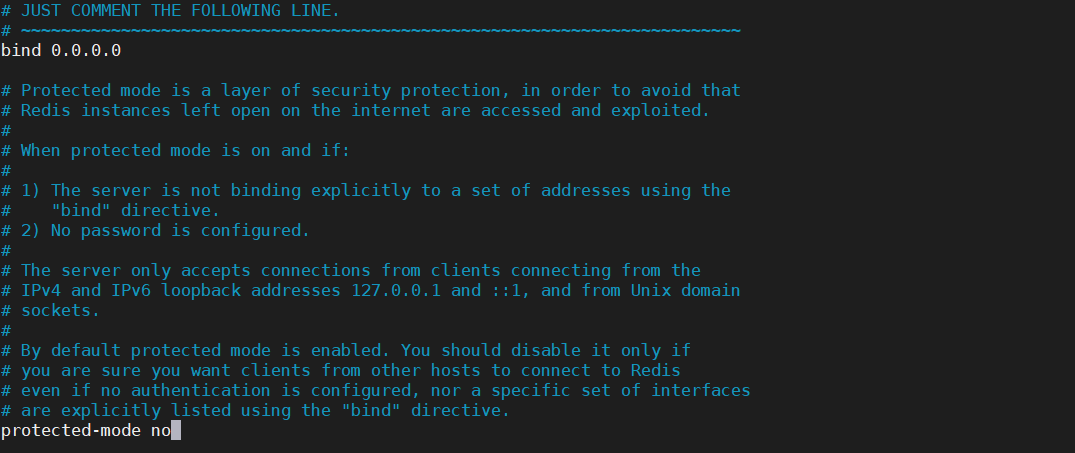
Redis can be started without a configuration file using a built-in default configuration. But to make any extra parameter changes you can use its configuration file that is: /etc/redis/redis.conf. Edit the Redis configuration file in a text editor to make changes

In this file we need to change below parameters:

bind 127.0.01 ::1 🡪 bind 0.0.0.0

protected-mode no

sudo vi /etc/redis/redis.conf



Save the configuration file and restart the Redis service:

sudo systemctl restart redis-server.service

**Step 4 – Install Redis PHP Extension**

Now, if you need to use Redis from PHP application, you also need to install Redis PHP extension on your Ubuntu system. Run below command to install:

sudo apt-get install redis-server -y

**Step 5 – Test Connection to Redis Server**

Use redis-cli tool to verify the connection between the Redis server.

redis-cli



Goto to EMR Cluster clone source code from github:

Connect to hadoop user and clone

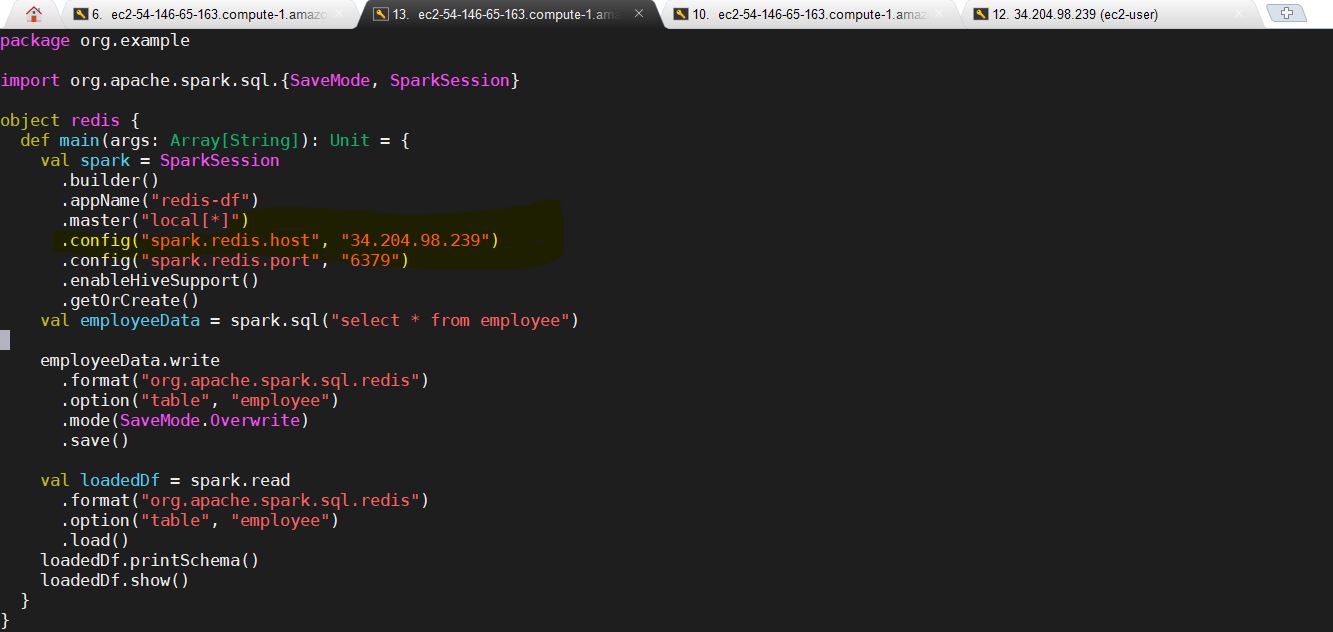
git clone <https://github.com/Naresh240/migrate-data-hive-redis.git>

cd migrate-data-hive-redis/

Need to Update **spark.redis.host** in below file (Give Redis IP-address)

cd src/main/scala/org/example/

vi redis.scala



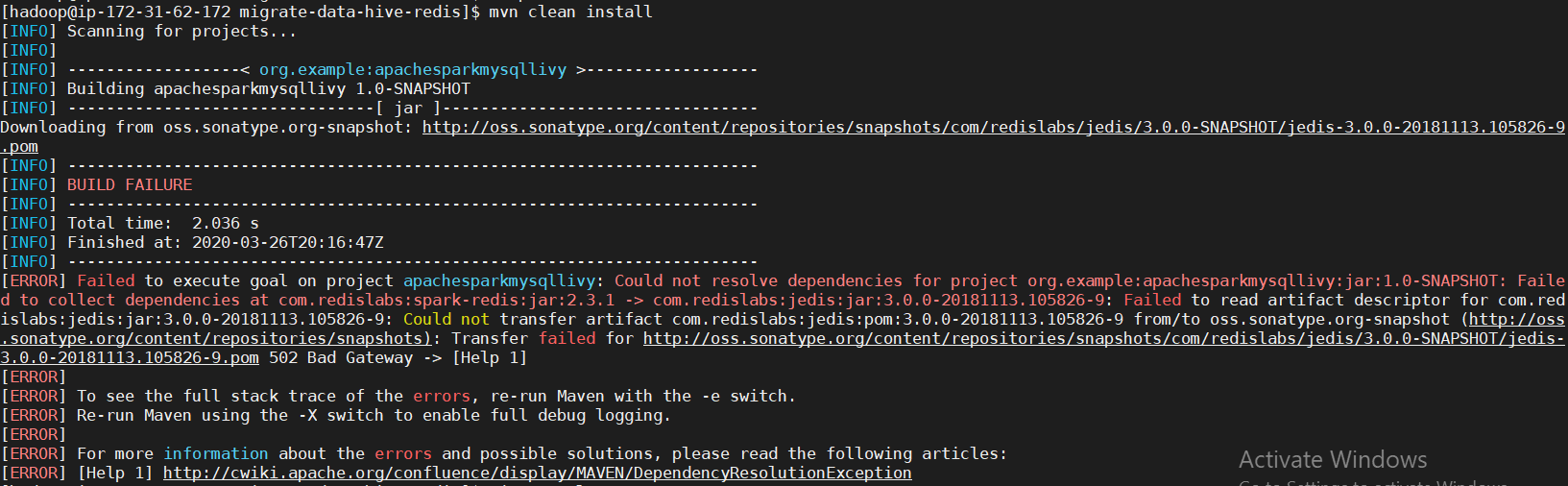
**Build Artifacts:**

mvn clean install

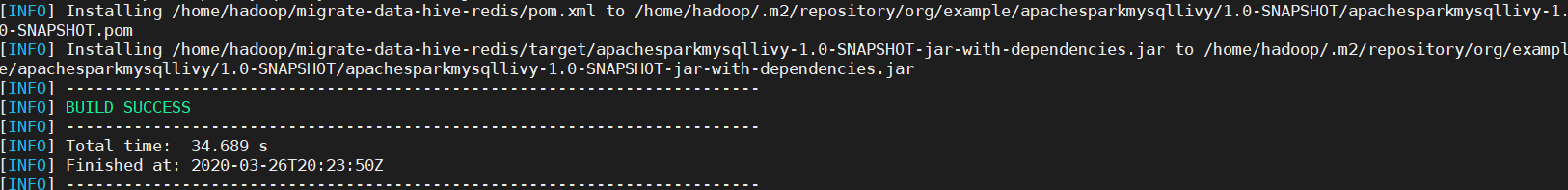
While running this command we may get below error. If we get same error we need to clean-ip below directory

cd /home/hadoop/.m2/repository/com/redislabs

rm -rf ./\*



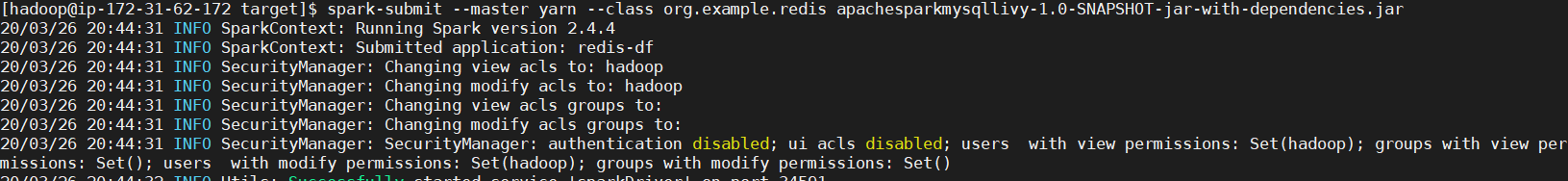
Again give “**mvn clean install**” command

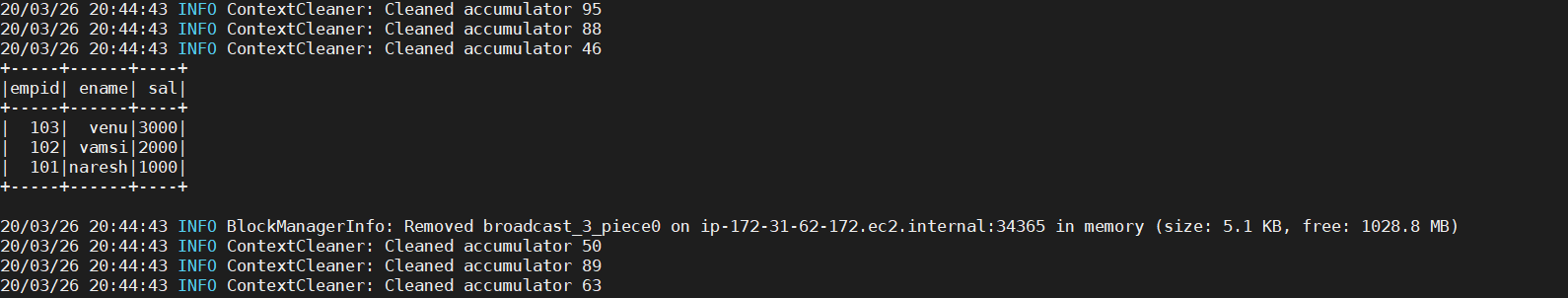


Now we need to run Spark-submit job:

Note: Please run this job in target directory:

spark-submit --master yarn --class org.example.redis apachesparkmysqllivy-1.0-SNAPSHOT-jar-with-dependencies.jar





See here we read data read from redis while running the spark-submit job🡪Everything is there in Code

**Now we can see our data in Redis:**

Goto Redis server:

Check employee data in Redis:

keys employee\*



Here we can see each Row which we pushed:

hgetall employee:77411f024cbc4a37813d178cde56242b

hgetall employee:9ea5e52fb24c4e0e936d2b9ff16af167

hgetall employee:ce6e563b26de4fe49badc2602ae0c138

