Install Ubuntu choose Mac UK key board

Install VMWare tools

Increase screen resolution

Change lock so it doesn’t need password

sudo apt-get update

sudo apt-get upgrade

install kafkacat

sudo nano /etc/sudoers

Add at the bottom:

oxclo ALL=(ALL) NOPASSWD:ALL

sudo apt-get install openjdk-8-jdk

sudo apt-get install build-essential checkinstall

sudo apt-get install libreadline-gplv2-dev libncursesw5-dev libssl-dev libsqlite3-dev tk-dev libgdbm-dev libc6-dev libbz2-dev

sudo apt-get install python-pip

sudo addgroup hadoop

sudo usermod -a -G hadoop oxclo

sudo apt-get install openssh-server

add the following to the bottom of /etc/sysctl.conf to disable ipv6

net.ipv6.conf.all.disable\_ipv6 = 1

net.ipv6.conf.default.disable\_ipv6 = 1

net.ipv6.conf.lo.disable\_ipv6 = 1

You will need to restart before this comes into effect.

cd Downloads

wget <http://mirror.vorboss.net/apache/hadoop/common/hadoop-2.7.3/hadoop-2.7.3.tar.gz>

untar

sudo mv hadoop-2.7.3/ /usr/local/hadoop

sudo chown -R oxclo:hadoop /usr/local/hadoop/

Add the following to /etc/environment

PATH="/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/usr/games:/usr/local/hadoop/bin:/usr/local/hadoop/sbin"

HADOOP\_HOME=/usr/local/hadoop

JAVA\_HOME=/usr/lib/jvm/java-8-openjdk-amd64

HADOOP\_INSTALL=/usr/local/hadoop

HADOOP\_MAPRED\_HOME=/usr/local/hadoop

HADOOP\_COMMON\_HOME=/usr/local/hadoop

HADOOP\_HDFS\_HOME=/usr/local/hadoop

YARN\_HOME=/usr/local/hadoop

HADOOP\_COMMON\_LIB\_NATIVE\_DIR=/usr/local/hadoop/lib/native

HADOOP\_OPTS="-Djava.library.path=/usr/local/hadoop/lib"

sudo mkdir -p /app/hadoop/tmp

sudo chown oxclo:hadoop /app/hadoop/tmp

**ssh-keygen -t rsa -P ""**

**cat $HOME/.ssh/id\_rsa.pub >> $HOME/.ssh/authorized\_keys**

modify core-site.xml

<configuration>

<property>

<name>hadoop.tmp.dir</name>

<value>/app/hadoop/tmp</value>

<description>A base for other temporary directories.</description>

</property>

<property>

<name>fs.default.name</name>

<value>hdfs://localhost:54310</value>

<description>The name of the default file system. A URI whose

scheme and authority determine the FileSystem implementation. The

uri's scheme determines the config property (fs.SCHEME.impl) naming

the FileSystem implementation class. The uri's authority is used to

determine the host, port, etc. for a filesystem.</description>

</property>

</configuration>

cp /usr/local/hadoop/etc/hadoop/mapred-site.xml.template /usr/local/hadoop/etc/hadoop/mapred-site.xml

edit it:

<configuration>

<property>

<name>mapred.job.tracker</name>

<value>localhost:54311</value>

<description>The host and port that the MapReduce job tracker runs

at. If "local", then jobs are run in-process as a single map

and reduce task.

</description>

</property>

</configuration>

hdfs-site.xml

<property>

<name>dfs.replication</name>

<value>1</value>

<description>Default block replication.

The actual number of replications can be specified when the file is created.

The default is used if replication is not specified in create time.

</description>

</property>

sudo mkdir -p /usr/local/hadoop\_store/hdfs/namenode

sudo mkdir -p /usr/local/hadoop\_store/hdfs/datanode

sudo chown -R oxclo:hadoop /usr/local/hadoop\_store

hdfs namenode –format

start-dfs.sh

hadoop fs -mkdir /user/

hadoop fs -mkdir /user/hduser

echo test > localfile

hadoop fs -put localfile /user/hduser/remotefile

hadoop fs -cat /user/hduser/remotefile

sudo apt-get install git

cd Downloads

wget http://d3kbcqa49mib13.cloudfront.net/spark-2.0.0.tgz

build/sbt -Pyarn -Phadoop-2.7 -Phive assembly package

mv spark-2.0.0 spark

Test with spark-2.0.0/bin/pyspark

sudo apt-get install python-dev

pip install --upgrade pip

sudo pip install numpy

sudo pip install python-dateutil

sudo pip install scipy

sudo apt-get install mosquitto mosquitto-clients

sudo pip install mosquito

sudo pip install paho-mqtt

sudo pip install python-dateutil

Installed pycharms and sublime text

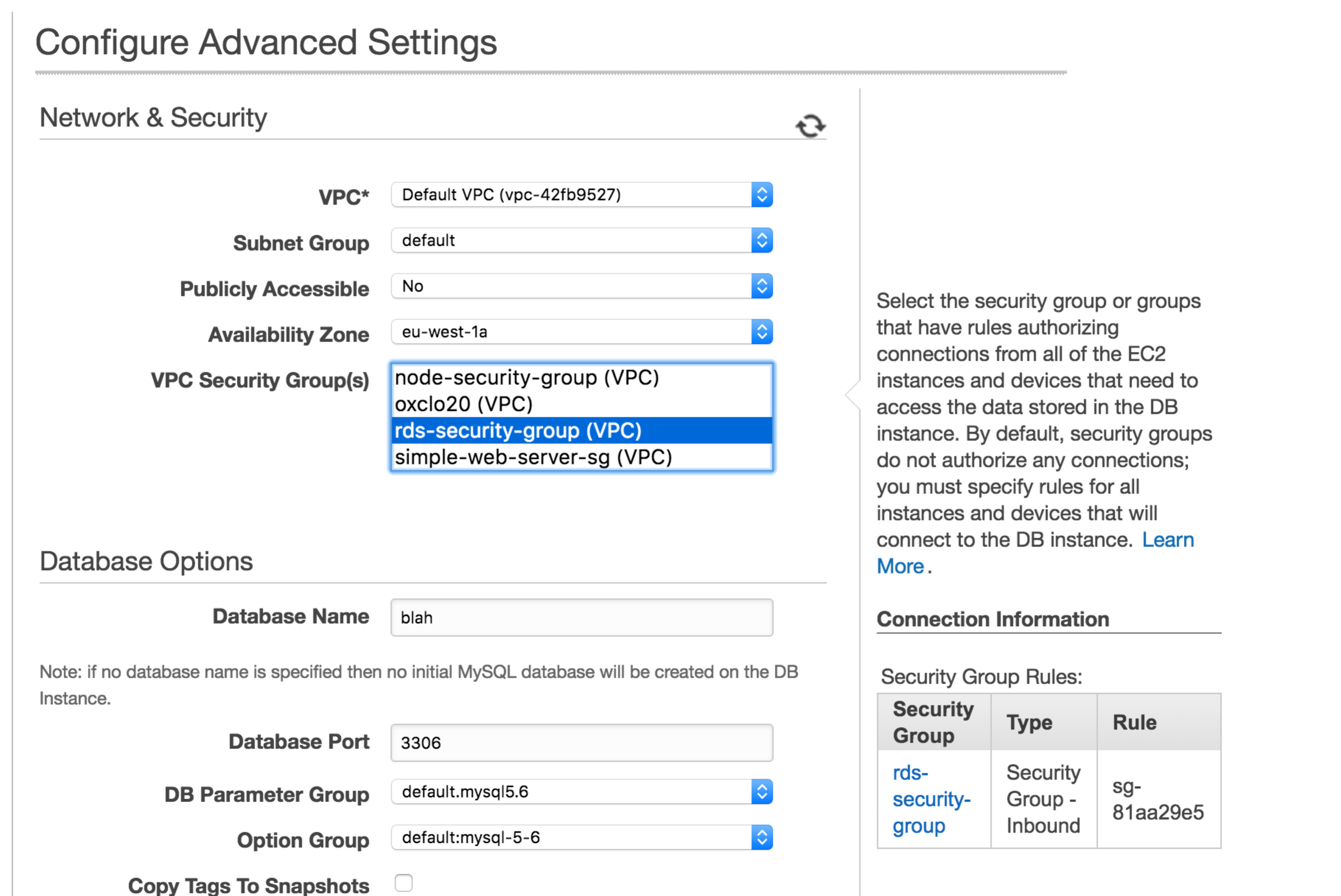
sudo apt install chromium-browser

Cloud setup

Deleted all the security groups starting oxclo

(painful)

Created a mysql database running in RDS



Then I created a new EC2 instance in the node group. Installed mysql client

create table user ( firstname varchar(100), lastname varchar(100), age int not null, primary key (firstname, lastname));

create user 'node' identified by 'node';

grant select on blah.user to 'node';

insert into user (firstname, lastname, age) values ("Paul", "Smith", 43);

plus a couple of others

sudo add-apt-repository ppa:fkrull/deadsnakes-python2.7

sudo apt-get update

sudo apt install python2.7

Deleted Launch configurations

Need to install Cassandra

CQLSH\_NO\_BUNDLED=true

sudo pip install Cassandra-driver

nightmare with Cassandra and python 2.7.12 ☹

wget http://launchpadlibrarian.net/109052632/python-support\_1.0.15\_all.deb sudo dpkg -i python-support\_1.0.15\_all.deb

**echo "deb http://debian.datastax.com/community stable main" | sudo tee -a /etc/apt/sources.list.d/cassandra.sources.list**

**curl -L http://debian.datastax.com/debian/repo\_key | sudo apt-key add -**

**sudo apt-get install dsc22 cassandra-tools**