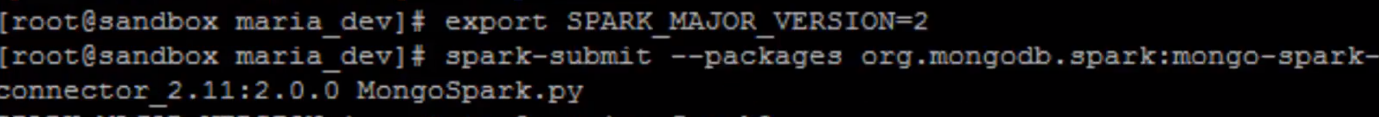
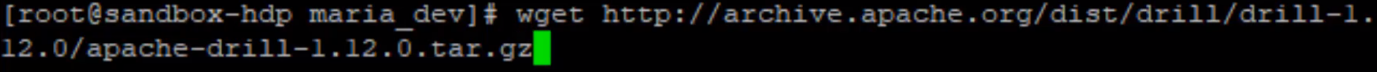
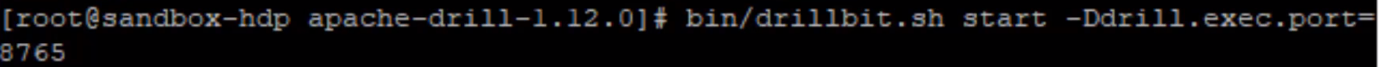
Horton Works Sandbox up and running,

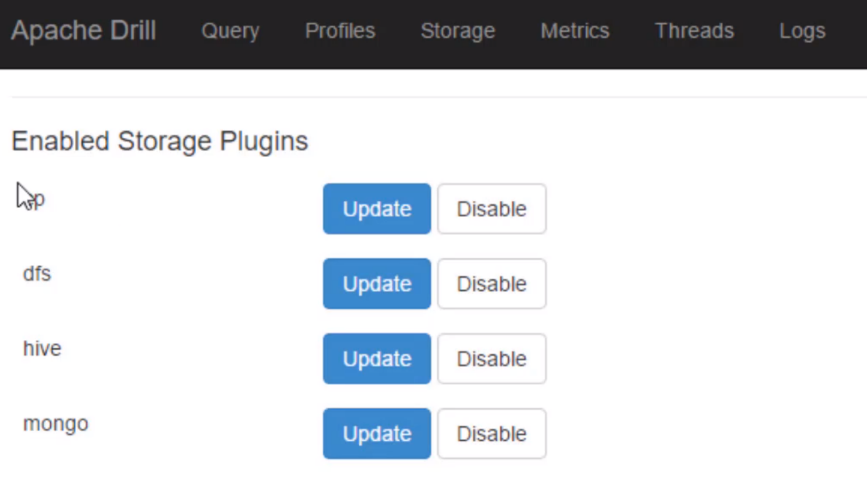
1. Login to Ambari (login as Admin)

Get Mongo DB up and Running

1. MongoDB Service -> Start service
2. Hive View -> Import data using Hive.
   1. CREATE DATABASE movielens; -> Execute
   2. Upload Table, check delimiter and assign them titles.
3. Get Data into MongoDB database -> Using Putty
   1. Login as Root.
   2. Ensure that you have u.user file under Files View.
   3. (Using Apache Spark integrated with MongoDB to import data from HDFS into MongoDB)
   4. Ensure that you are using Spark Version 2.
4. Install 1.12 of Apache Dril
   1. Decompress drill

tar -xvf apache-drill-1.12.0.tar.gz

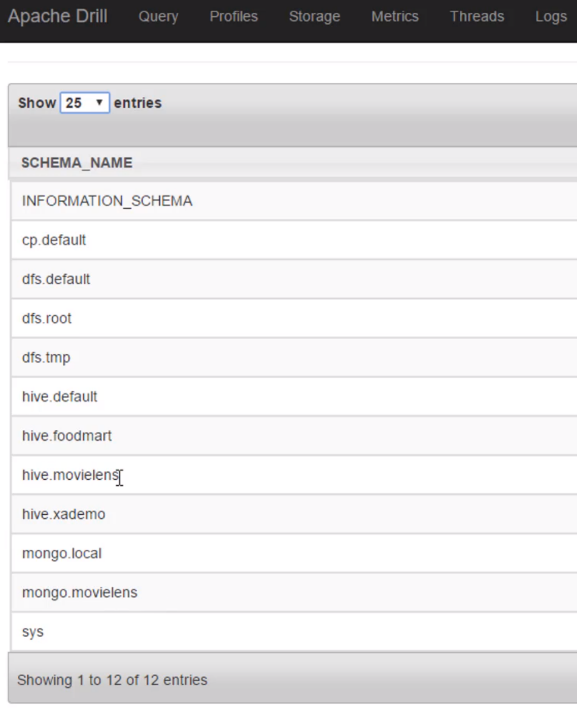
* 1. Need to script in the bin folder called drillbit
  2. Start the drillbit. Override the port that Drill run on. Communicate with drill from our web browser that is running on our host system.
  3. HDP is running within a Docker Container within the virtual machine. But we will be using 8765 for convienience.
  4. Go to 127.0.0.1:8765 to access the drill UI.
  5. Hit the Storage Tab there, and ensure that you have Hive and Mongo Plugins



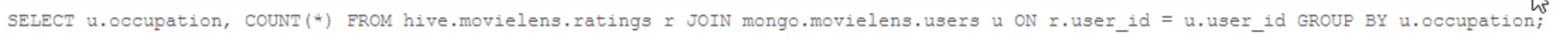
* 1. Update hive plugin.
  2. Change ‘hive.metastore.uris’ url into : ’thrift://localhost:9083’ then Update.

Set up HiveDatabase with Drill

Using Drill. There is no learning curve cuz its actual SQL.

See Databases: SHOW DATABASES.

How many ratings provided by each occupation



SELECT u.occupation,

COUNT(\*)

FROM hive.movielens.ratings r

JOIN mongo.movielens.users u ON r.user\_id = u.user\_id

GROUP BY u.occupation;

How to clean up

Shutdown drillbit -> bin/drilbit.sh stop

Shutdown mongodb -> Ambari UI-> MongoDB Stop