Integrate Presto with Hive and Cassandra

1. Load up HortonWorks Docker Sandbox and login as Root
2. Download the presto package. <https://prestodb.io/docs/current/installation/deployment.html>. And copy link address
3. wget + RIGHT CLICK to paste the copied link and install
4. tar -xvf presto-server-0.165.tar.gz to expand the it into presto server directory.

They don’t provide you any configuration at all. Its up to you to create those file. Create your own etc files. It is because of your different environments.

Load in etc files

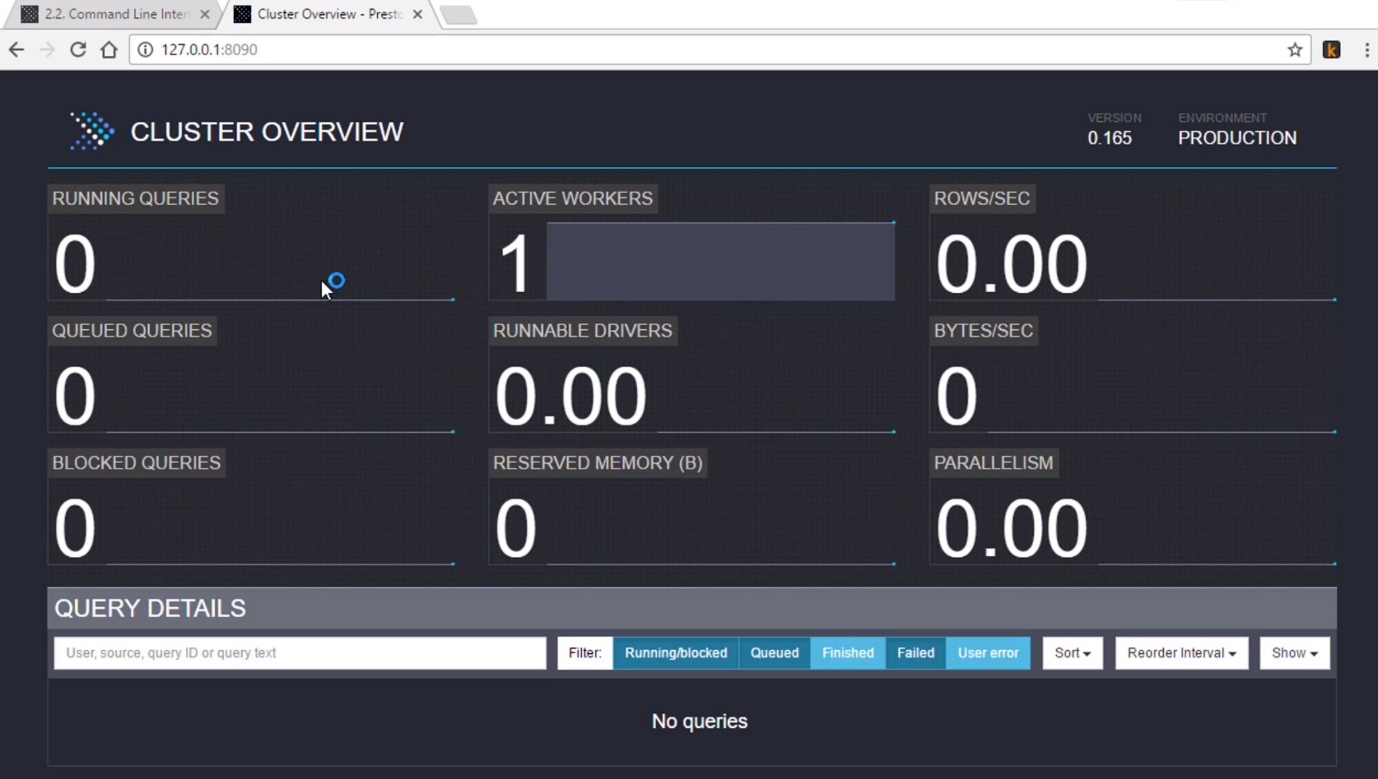
1. wget <http://media.sundog-soft.com/hadoop/presto-hdp-config.tgz>.
2. Unpack by tar -xvs presto-hdp-config.tgz
3. Less <package> to see what is inside.
   1. less node.properties, need unique IDs
   2. hive.properties -> Hive database connector

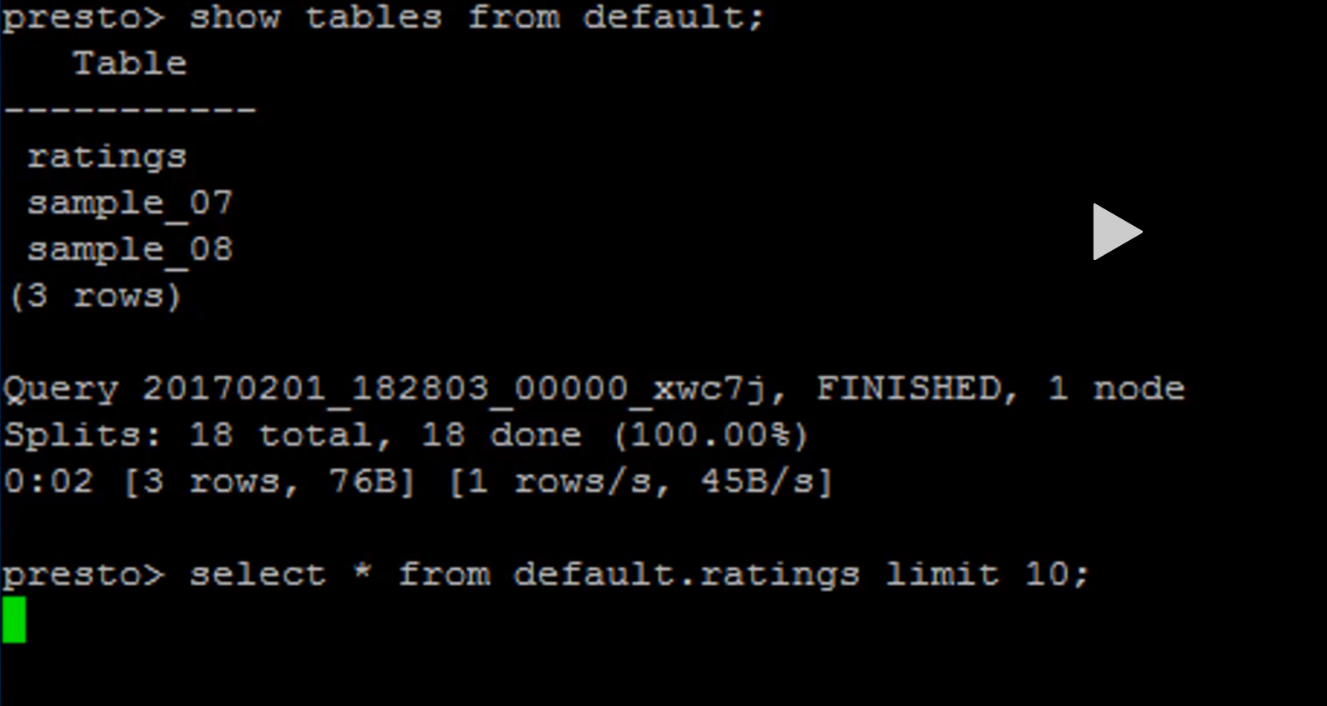
Download and install the command line interface for Presto., copy link address and ensure that you are inside the bin folder of presto-server-00.165

And download the interface by wget+RIGHT CLICK, Rename it as mv presto-cli-0.165-executable.jar presto

* chmod +x presto to make it into an exe

Ensure that you are in the presto directory.

1. Start the presto by typing “bin/launcher start”
2. Connect to the web interface. See if you see the Presto Dashboard
3. Connect to it using ‘bin/presto –server 127.0.01:8090 --cataloghive’

See the tables inside by: show tables from default;

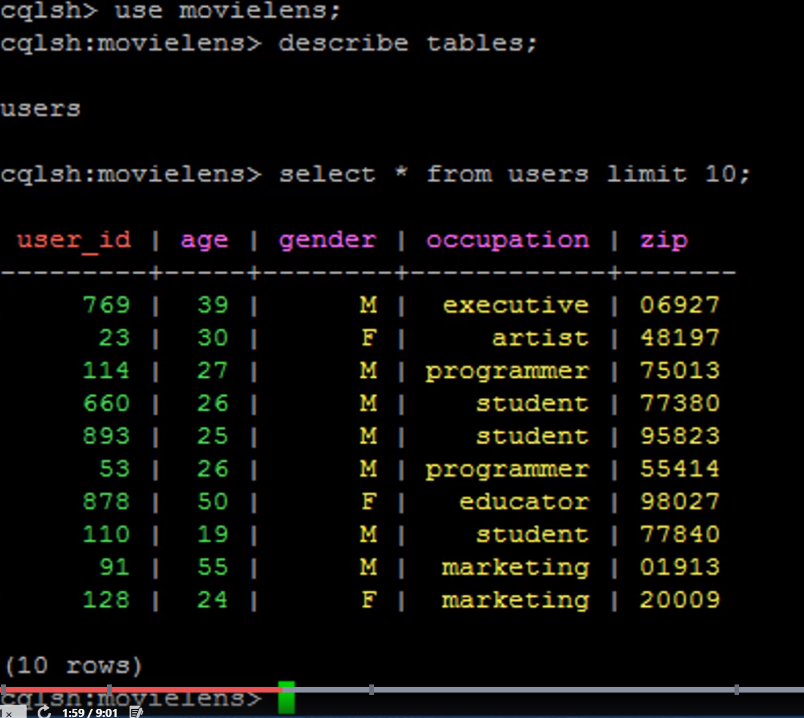
Perform more queries to test your database data availability.

And quit by quit

To stop presto: bin/launcher stop

**Connect Cassandra, and Query both Cassandra and Hive using Presto**

1. Enable Python 2.7 ‘scl enable python27 bash’
2. ‘service cassandra start’
3. Enable thrift service on Cassandra to connect with Presto : ‘nodetool enablethrift’
4. CQLSH ‘cqlsh –cqlversion=”3.4.0” to get into the Cassandra shell



Connected that to Presto together with the data living in Hive and query that together. Merge that together.

1. Set up a config file with presto.
   1. ‘cd etc/catalog’
   2. ‘vi cassandra.properties’ + press I for insert
      1. connector.name=Cassandra  
         cassandra.contact-points = 127.0.0.1
      2. Esc + :wq
   3. cd ../..
   4. Start Presto ‘bin/launcher start’
   5. Command line Interface ‘bin/presto –server 127.0.0.:8090 –catalog hive, cassandra’
      1. “show tables from cassandra.movielens;”
      2. “describe cassandra.movielens.users;”
      3. ‘select \* from cassandra.movielens.users limit 10;’
      4. ‘select \* from hive.default.ratings limit 10;’
2. Tie them together
   1. ‘select u.occupation, count(\*) from hive.default ratings r join cassandra.movielens.users u on r.user\_id = u.user\_id group by u.occupation;’

Clean up

1. Quit out of command line interface
2. ‘bin/launcher stop’
3. ‘service cassandra stop’