Tez Usage Guide

Walking Through How To Use Tez

Note: This guide assumes that you have setup and installed Tez, following the instructions from the Tez Installation Guide.

---

1) **Enter your localhost via ssh.**

Command: `ssh localhost`

If this command doesn't work, it means you don't have localhost setup. There are two commands to setup localhost.

Command: `ssh-keygen -t dsa -P '' -f ~/.ssh/id_dsa`
Command: `cat ~/.ssh/id_dsa.pub >> ~/.ssh/authorized_keys`

2) **Start up HDFS.**

First navigate to your hadoop directory and run the following command.

Command: `sbin/start-dfs.sh`

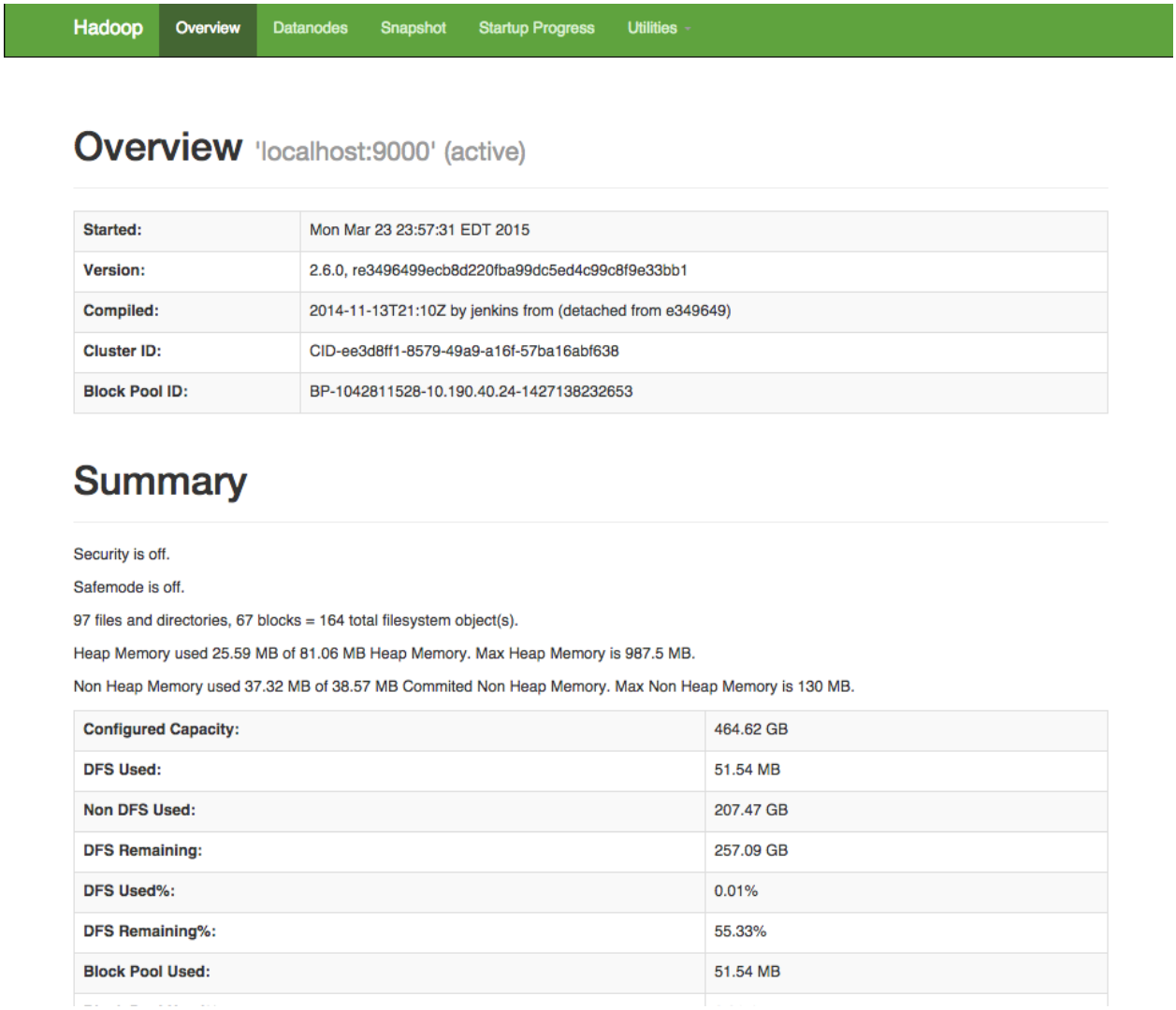
Screenshot 1 (starting HDFS)

```
jaywang@jays-MacBook-Pro-2:~/Desktop/data/hadoop-2.6.0$ sbin/start-dfs.sh
2015-03-23 23:57:25.005 java[32968:1470776] Unable to load realm info from SCDynamicStore
15/03/23 23:57:25 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where
 applicable
Starting namenodes on [localhost]
localhost: starting namenode, logging to /Users/jaywang/Desktop/data/hadoop-2.6.0/logs/hadoop-jaywang-namenode-jays-MacBook-Pro-2.local.
out
localhost: starting datanode, logging to /Users/jaywang/Desktop/data/hadoop-2.6.0/logs/hadoop-jaywang-datanode-jays-MacBook-Pro-2.local.
out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /Users/jaywang/Desktop/data/hadoop-2.6.0/logs/hadoop-jaywang-secondarynamenode-jays-MacB
ook-Pro-2.local.out
2015-03-23 23:57:42.460 java[33258:1472075] Unable to load realm info from SCDynamicStore
15/03/23 23:57:42 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where
 applicable
```

Once starting HDFS, you should be able to access http://localhost:50070/.

This will look something like:

Screenshot 2

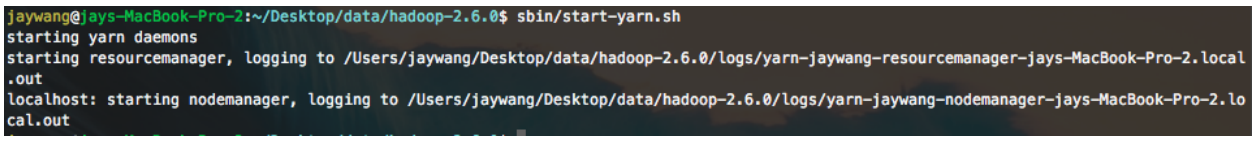


3) Start up YARN.

In the same directory, run the following command.

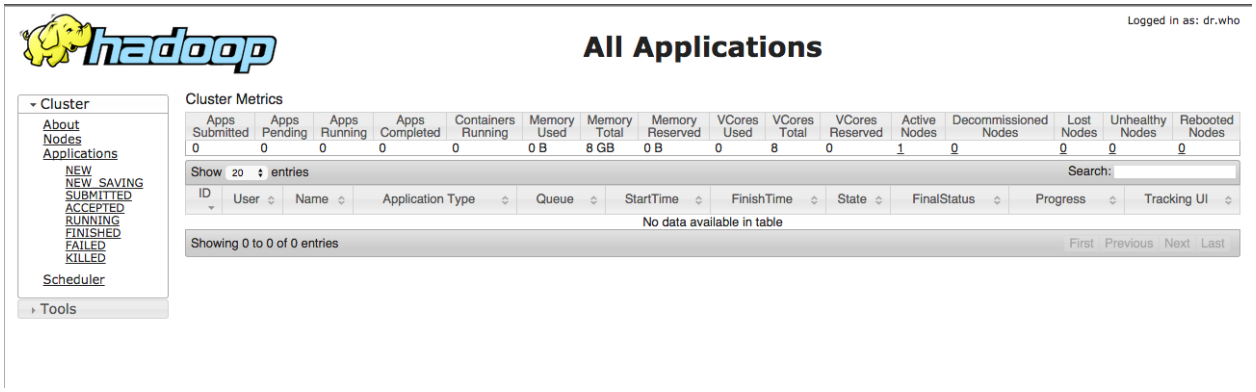
Command: `sbin/start-yarn.sh`

Screenshot 3 (starting YARN)



After starting YARN, you should be able to access: `http://localhost:8088/`.
This will look something like:

Screenshot 4 (YARN cluster)



4) Run your Hive job.

Open a new tab in your terminal and navigate to your hive directory. Run your Hive script, making sure that in your script you set the execution engine to tez. This is done with the line:

Line: `set hive.execution.engine = tez;`

Please reference one of the **_tez.sql* scripts to see an example.

Run your Hive script. Here is what my command looks like (yours may be different):
Command: `bin/hive -f queries/overweight_zips_tez.sql`

Note that you can also run your script in the Hive console. I have just chosen to run it in batch mode for convenience.

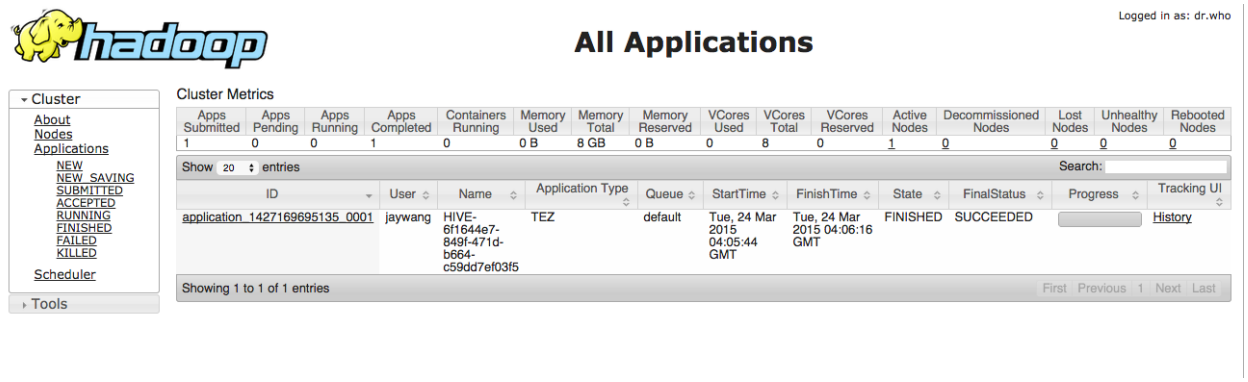
Here is an excerpt from what the console output should look like:

Screenshot 5 (Console output from Tez job)

```
OK
Time taken: 0.039 seconds
Total jobs = 1
Launching Job 1 out of 1


Status: Running (application id: application_1427169695135_0001)

Map 1: -/-      Map 2: -/-      Reducer 3: 0/1
Map 1: 0/1      Map 2: 0/1      Reducer 3: 0/1
Map 1: 1/1      Map 2: 0/1      Reducer 3: 0/1
Map 1: 1/1      Map 2: 1/1      Reducer 3: 0/1
Map 1: 1/1      Map 2: 1/1      Reducer 3: 1/1
Status: Finished successfully
Loading data to table default.least_overweight_zips
rmr: DEPRECATED: Please use 'rm -r' instead.
Deleted hdfs://localhost:9000/Users/jaywang/Desktop/data/hive_tables/least_overweight_zips
Table default.least_overweight_zips stats: [numFiles=0, numRows=100, totalSize=0, rawDataSize=2187]
OK
Time taken: 1.959 seconds
Total jobs = 1
Launching Job 1 out of 1


Status: Running (application id: application_1427169695135_0001)

Map 1: -/-      Reducer 2: 0/1
Map 1: 0/1      Reducer 2: 0/1
Map 1: 1/1      Reducer 2: 0/1
Map 1: 1/1      Reducer 2: 1/1
Status: Finished successfully
OK
Average population density per zip code:        1254.6468685961365
Time taken: 1.275 seconds, Fetched: 1 row(s)
Total jobs = 1
Launching Job 1 out of 1


Status: Running (application id: application_1427169695135_0001)

Map 1: -/-      Reducer 2: 0/1
Map 1: 0/1      Reducer 2: 0/1
Map 1: 1/1      Reducer 2: 0/1
Map 1: 1/1      Reducer 2: 1/1
Status: Finished successfully
OK
Average population density per zip code (top overweight):       953.66536454547
Time taken: 0.817 seconds, Fetched: 1 row(s)
Total jobs = 1
Launching Job 1 out of 1


Status: Running (application id: application_1427169695135_0001)

Map 1: -/-      Reducer 2: 0/1
Map 1: 1/1      Reducer 2: 0/1
Map 1: 1/1      Reducer 2: 1/1
Status: Finished successfully
OK
Average population density per zip code (least overweight):     1454.0622953780796
Time taken: 0.608 seconds, Fetched: 1 row(s)
```

You should also see the Tez job submitted to your cluster at: http://localhost:8088/cluster.

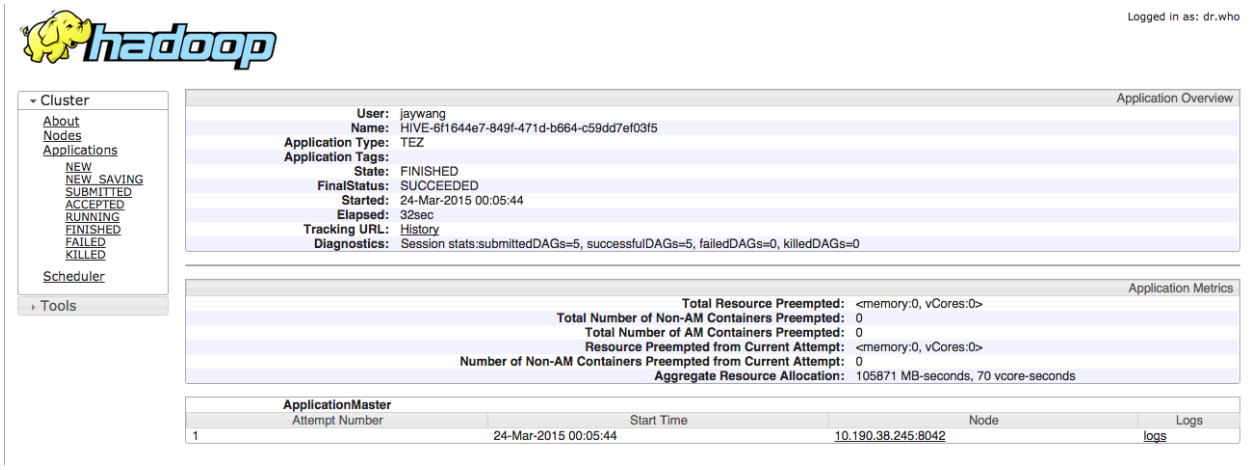
Screenshot 6



5) Drilling into jobs.

To get more information about the job, you can click the application id to drill in.

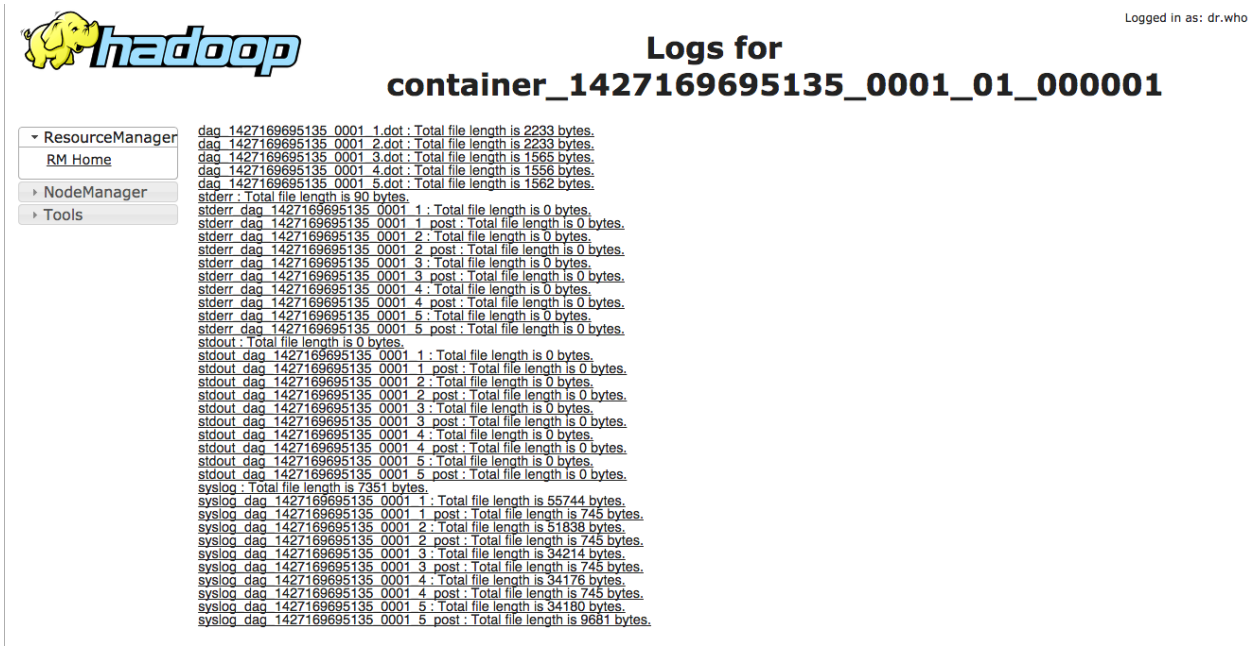
Screenshot 7 (Drilling into Tez job)



6) Reading Logs.

Clicking logs gives you detailed log information from your Tez job.

Screenshot 8 (Tez Logs)



7) **Once you are done running your jobs, shut down your systems.**

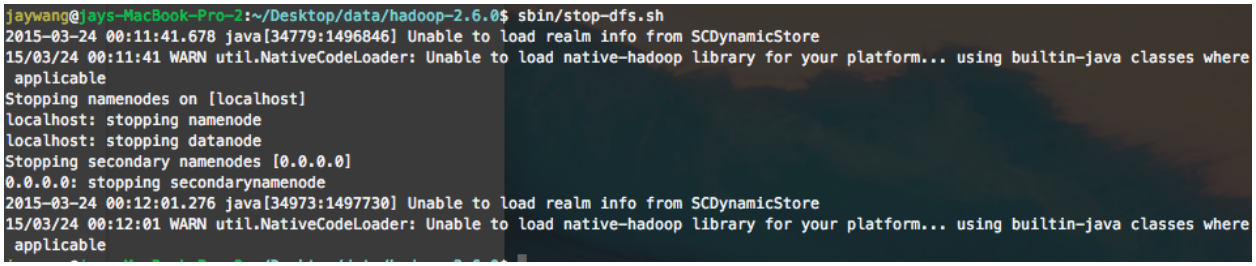
Start by shutting down YARN.
Command: `sbin/stop-yarn.sh`

Screenshot 9 (Shutting down YARN)

8) **Shut down HDFS.**

Command: `sbin/stop-dfs.sh`

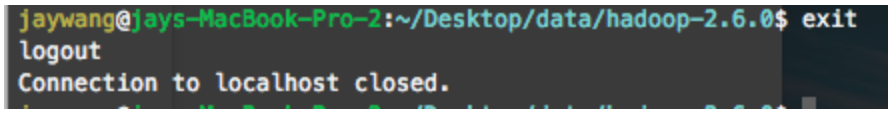
Screenshot 10 (Shutting down HDFS)



9) Exit Localhost.

Command: `exit`

Screenshot 11



That's it for using Tez. I hope this guide was helpful. If you have any questions, shoot me an email at jayduke15@gmail.com.