

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`

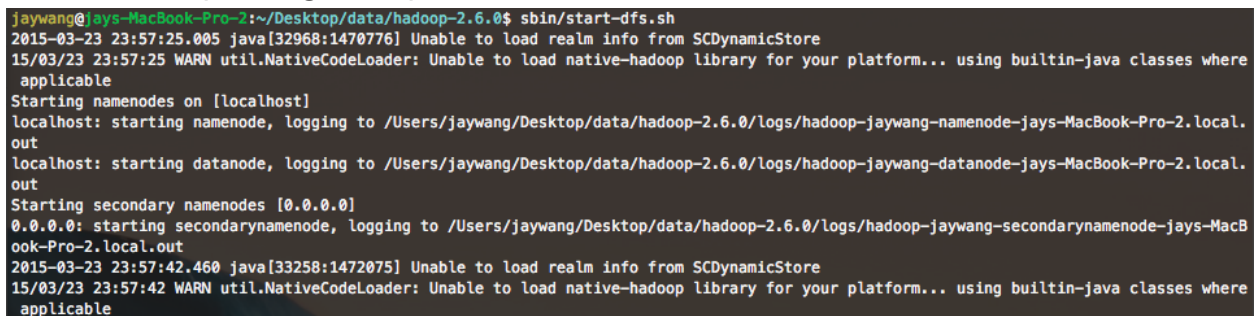
Command: `cat ~/.ssh/id_dsa.pub >> ~/.ssh/authorized_keys`

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-MacBook-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

Hadoop Overview Datanodes Snapshot Startup Progress Utilities -

Overview 'localhost:9000' (active)

Started:	Mon Mar 23 23:57:31 EDT 2015
Version:	2.6.0, re3496499ecb8d220fba99dc5ed4c99c8f9e33bb1
Compiled:	2014-11-13T21:10Z by jenkins from (detached from e349649)
Cluster ID:	CID-ee3d8ff1-8579-49a9-a16f-57ba16abf638
Block Pool ID:	BP-1042811528-10.190.40.24-1427138232653

Summary

Security is off.

Safemode is off.

97 files and directories, 67 blocks = 164 total filesystem object(s).

Heap Memory used 25.59 MB of 81.06 MB Heap Memory. Max Heap Memory is 987.5 MB.

Non Heap Memory used 37.32 MB of 38.57 MB Committed Non Heap Memory. Max Non Heap Memory is 130 MB.

Configured Capacity:	464.62 GB
DFS Used:	51.54 MB
Non DFS Used:	207.47 GB
DFS Remaining:	257.09 GB
DFS Used%:	0.01%
DFS Remaining%:	55.33%
Block Pool Used:	51.54 MB

3) Start up YARN.

In the same directory, run the following command.

Command: `sbin/start-yarn.sh`

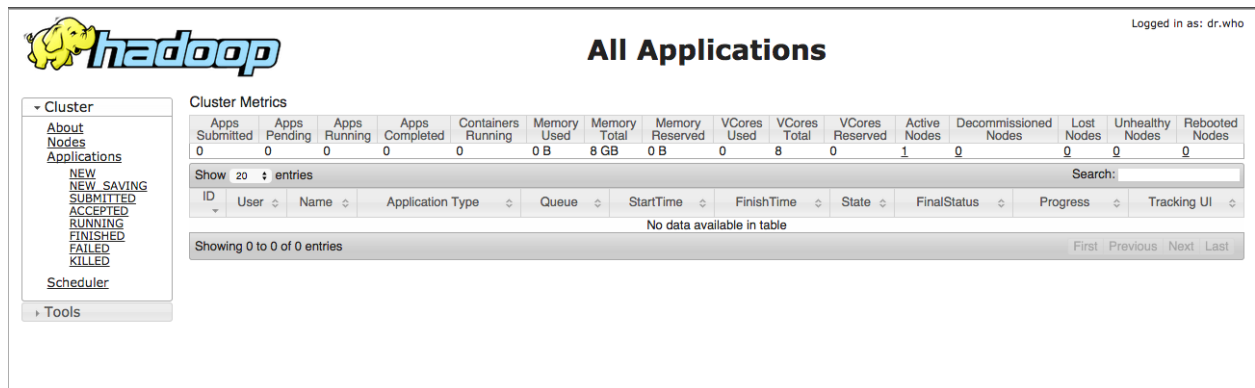
Screenshot 3 (starting YARN)

```
jaywang@jays-MacBook-Pro-2:~/Desktop/data/hadoop-2.6.0$ sbin/start-yarn.sh
starting yarn daemons
starting resourcemanager, logging to /Users/jaywang/Desktop/data/hadoop-2.6.0/logs/yarn-jaywang-resourcemanager-jays-MacBook-Pro-2.local.out
localhost: starting nodemanager, logging to /Users/jaywang/Desktop/data/hadoop-2.6.0/logs/yarn-jaywang-nodemanager-jays-MacBook-Pro-2.local.out
```

After starting YARN, you should be able to access: <http://localhost:8088/>.

This will look something like:

Screenshot 4 (YARN cluster)



hadoop

Logged in as: dr.who

All Applications

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	Vcores Used	Vcores Total	Vcores Reserved	Active Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes
0	0	0	0	0	0 B	8 GB	0 B	0	8	0	1	0	0	0	0

Show 20 entries

Search:

ID	User	Name	Application Type	Queue	StartTime	FinishTime	State	FinalStatus	Progress	Tracking UI
No data available in table										

Showing 0 to 0 of 0 entries

First Previous Next Last

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



All Applications

Logged in as: dr.who

Cluster

About Nodes

Applications

NEW
NEW SAVING
SUBMITTED
ACCEPTED
RUNNING
FINISHED
FAILED
KILLED

Scheduler

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total	VCores Reserved	Active Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes
1	0	0	1	0	0 B	8 GB	0 B	0	8	0	1	0	0	0	0

Show 20 entries

Search:

ID	User	Name	Application Type	Queue	StartTime	FinishTime	State	FinalStatus	Progress	Tracking UI
application_1427169695135_0001	jaywang	HIVE-6f1644e7-849f-471d-b664-c59dd7ef03f5	TEZ	default	Tue, 24 Mar 2015 04:05:44 GMT	Tue, 24 Mar 2015 04:06:16 GMT	FINISHED	SUCCEEDED		History

Showing 1 to 1 of 1 entries

First Previous 1 Next Last

Tools

Tip: Sometimes Hive will throw very arcane errors such as `FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.tez.TezTask`.

Since there is not much to go off that, you can configure the logger to output more detailed logs to the console.

Navigate to your `hive/conf` directory and copy the contents of `hive-log4j.properties.template` into a `hive-log4j.properties` file. Change the value of `hive.root.logger` to be either `INFO` or `DEBUG` (depending on what you want), and set the output to `console`.

Line: `hive.root.logger=INFO,console`

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)



Logged in as: dr.who

Cluster

About

Nodes

Applications

NEW

NEW_SAVING

SUBMITTED

ACCEPTED

RUNNING

FINISHED

FAILED

KILLED

Scheduler

Tools

Application Overview

User: jaywang

Name: HIVE-6f1644e7-849f-471d-b664-c59dd7ef03f5

Application Type: TEZ

Application Tags:

State: FINISHED

FinalStatus: SUCCEEDED

Started: 24-Mar-2015 00:05:44

Elapsed: 32sec

Tracking URL: History

Diagnostics: Session stats:submittedDAGs=5, successfulDAGs=5, failedDAGs=0, killedDAGs=0

Application Metrics

Total Resource Preempted: <memory:0, vCores:0>

Total Number of Non-AM Containers Preempted: 0

Total Number of AM Containers Preempted: 0

Resource Preempted from Current Attempt: <memory:0, vCores:0>

Number of Non-AM Containers Preempted from Current Attempt: 0

Aggregate Resource Allocation: 105871 MB-seconds, 70 vcore-seconds

ApplicationMaster

Attempt Number	Start Time	Node	Logs
1	24-Mar-2015 00:05:44	10.190.38.245:8042	logs

6) Reading Logs.

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

Screenshot 8 (Tez Logs)



Logged in as: dr.who

Logs for container_1427169695135_0001_01_000001

ResourceManager

RM Home

NodeManager

Tools

dag_1427169695135_0001_1.dot : Total file length is 2233 bytes.

dag_1427169695135_0001_2.dot : Total file length is 2233 bytes.

dag_1427169695135_0001_3.dot : Total file length is 1565 bytes.

dag_1427169695135_0001_4.dot : Total file length is 1556 bytes.

dag_1427169695135_0001_5.dot : Total file length is 1562 bytes.

stderr_dag_1427169695135_0001_1 : Total file length is 90 bytes.

stderr_dag_1427169695135_0001_1 post : Total file length is 0 bytes.

stderr_dag_1427169695135_0001_2 : Total file length is 0 bytes.

stderr_dag_1427169695135_0001_2 post : Total file length is 0 bytes.

stderr_dag_1427169695135_0001_3 : Total file length is 0 bytes.

stderr_dag_1427169695135_0001_3 post : Total file length is 0 bytes.

stderr_dag_1427169695135_0001_4 : Total file length is 0 bytes.

stderr_dag_1427169695135_0001_4 post : Total file length is 0 bytes.

stderr_dag_1427169695135_0001_5 : Total file length is 0 bytes.

stderr_dag_1427169695135_0001_5 post : Total file length is 0 bytes.

stdout : Total file length is 0 bytes.

stdout_dag_1427169695135_0001_1 : Total file length is 0 bytes.

stdout_dag_1427169695135_0001_1 post : Total file length is 0 bytes.

stdout_dag_1427169695135_0001_2 : Total file length is 0 bytes.

stdout_dag_1427169695135_0001_2 post : Total file length is 0 bytes.

stdout_dag_1427169695135_0001_3 : Total file length is 0 bytes.

stdout_dag_1427169695135_0001_3 post : Total file length is 0 bytes.

stdout_dag_1427169695135_0001_4 : Total file length is 0 bytes.

stdout_dag_1427169695135_0001_4 post : Total file length is 0 bytes.

stdout_dag_1427169695135_0001_5 : Total file length is 0 bytes.

stdout_dag_1427169695135_0001_5 post : Total file length is 0 bytes.

syslog : Total file length is 7351 bytes.

syslog_dag_1427169695135_0001_1 : Total file length is 55744 bytes.

syslog_dag_1427169695135_0001_1 post : Total file length is 745 bytes.

syslog_dag_1427169695135_0001_2 : Total file length is 51838 bytes.

syslog_dag_1427169695135_0001_2 post : Total file length is 745 bytes.

syslog_dag_1427169695135_0001_3 : Total file length is 34214 bytes.

syslog_dag_1427169695135_0001_3 post : Total file length is 745 bytes.

syslog_dag_1427169695135_0001_4 : Total file length is 34176 bytes.

syslog_dag_1427169695135_0001_4 post : Total file length is 745 bytes.

syslog_dag_1427169695135_0001_5 : Total file length is 34180 bytes.

syslog_dag_1427169695135_0001_5 post : Total file length is 9681 bytes.

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)

```
jaywang@jays-MacBook-Pro-2:~/Desktop/data/hadoop-2.6.0$ sbin/stop-yarn.sh
stopping yarn daemons
stopping resourcemanager
localhost: stopping nodemanager
localhost: nodemanager did not stop gracefully after 5 seconds: killing with kill -9
no proxyserver to stop
```

8) Shut down HDFS.

Command: `sbin/stop-dfs.sh`

Screenshot 10 (Shutting down HDFS)

```
jaywang@jays-MacBook-Pro-2:~/Desktop/data/hadoop-2.6.0$ sbin/stop-dfs.sh
2015-03-24 00:11:41.678 java[34779:1496846] Unable to load realm info from SCDynamicStore
15/03/24 00:11:41 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Stopping namenodes on [localhost]
localhost: stopping namenode
localhost: stopping datanode
Stopping secondary namenodes [0.0.0.0]
0.0.0.0: stopping secondarynamenode
2015-03-24 00:12:01.276 java[34973:1497730] Unable to load realm info from SCDynamicStore
15/03/24 00:12:01 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
```

9) Exit Localhost.

Command: `exit`

Screenshot 11

```
jaywang@jays-MacBook-Pro-2:~/Desktop/data/hadoop-2.6.0$ exit
logout
Connection to localhost closed.
```

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.