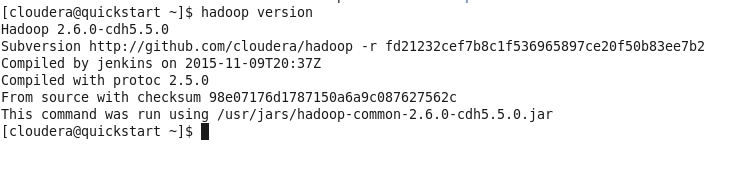
We are here planning to use Apache Flink with Hadoop

First check Hadoop version



Now download the latest apache flink for Hadoop version 2.6

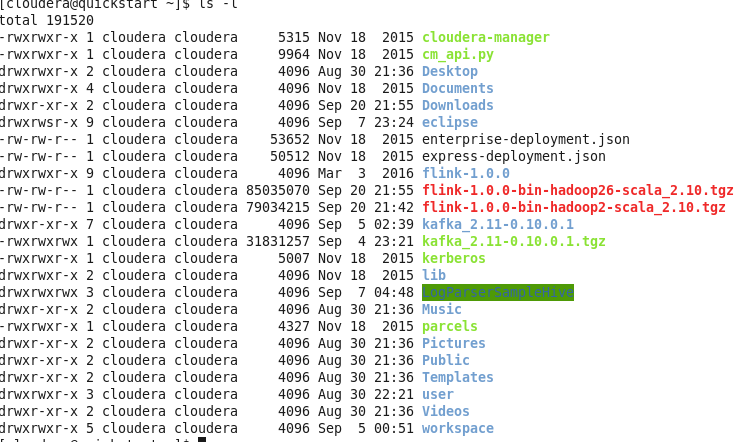
<http://archive.apache.org/dist/flink/flink-1.0.0/>

I have downloaded flink-1.0.0-bin-hadoop26-scala\_2.10.tgz file

Put this file under /home/cloudera/

Untar this file using below command

|  |
| --- |
| tar -xvzf flink-1.0.0-bin-hadoop26-scala\_2.10.tgz |



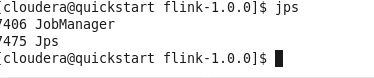
You will see flink-1.0.0 directory after un tar this file

Now go to Flink directory and start Flink in local mode

|  |
| --- |
| [cloudera@quickstart flink-1.0.0]$ bin/start-local.sh |



Verify by issuing jps command



If you found JobManager it means Flink is running

If netcat is not installed on your system , get it installed using command

|  |
| --- |
| sudo yum install nc |

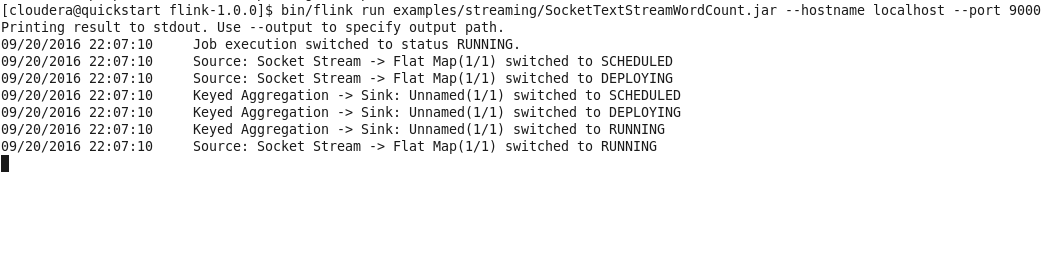
Now lock 9000 port by nc command



Now open a new terminal

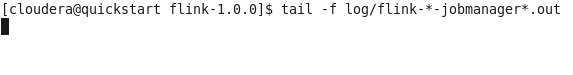
Switch to flink folder and issue below command

|  |
| --- |
| bin/flink run examples/streaming/SocketTextStreamWordCount.jar --hostname localhost --port 9000 |



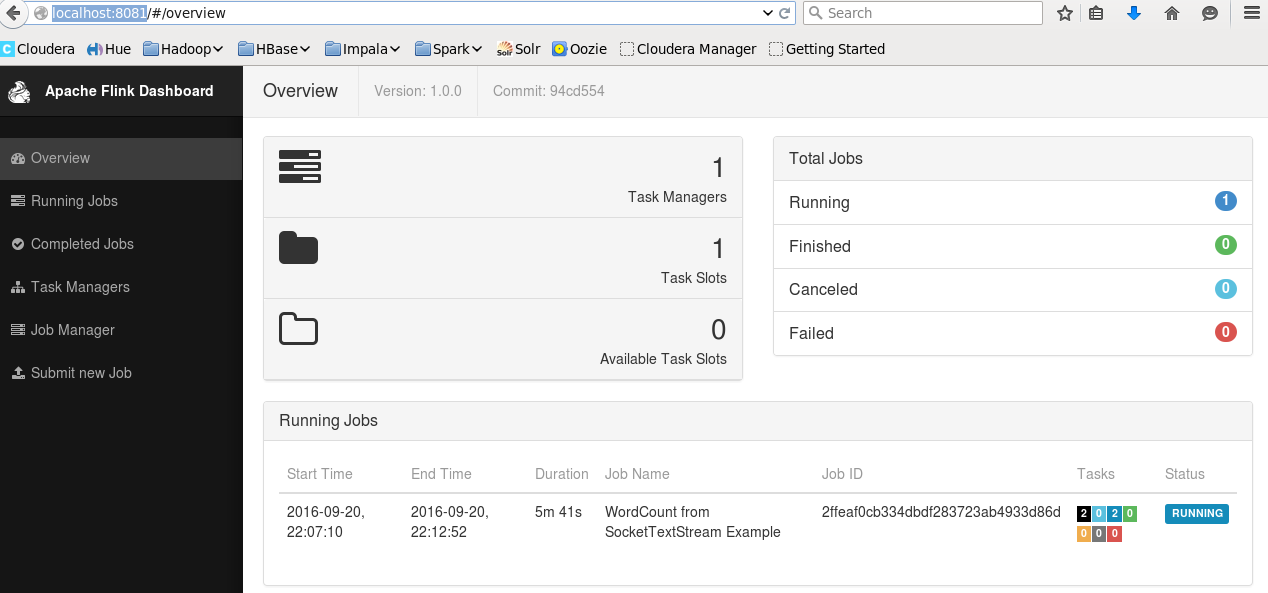
Now open a new terminal and switch to flink folder again and issue below command to check the log

|  |
| --- |
| tail -f log/flink-\*-jobmanager\*.out |



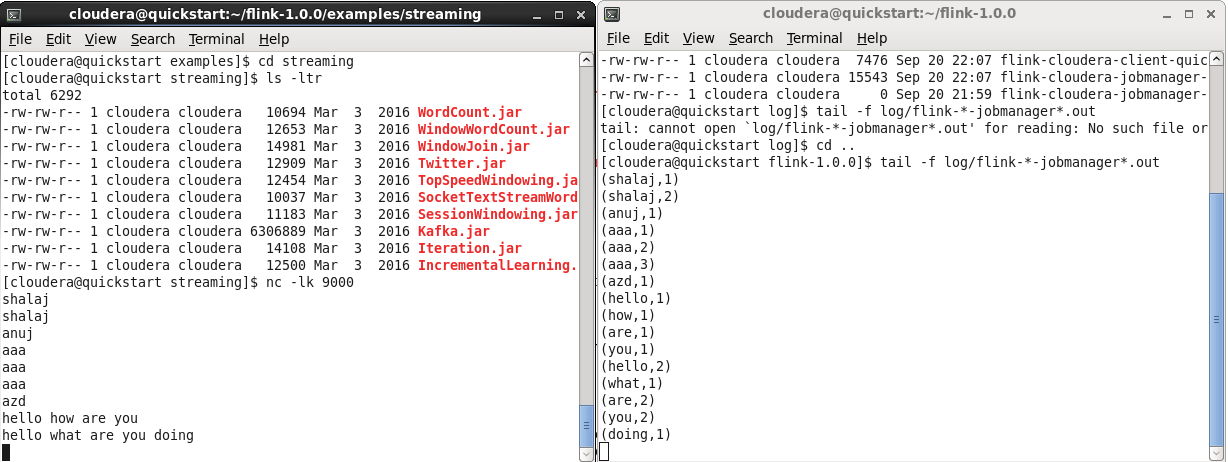
Now we have started wordcount job and also checking log file but we haven’t started giving input yet let’s check Web dashboard

Open <http://localhost:8081>



Here you can see one job is running and you can check the more information by clicking on Running Jobs link at bottom

Now start giving the input in netcat terminal and check the log at the same time on log terminal



You can see as we are entering some input the word count is happened on other window in real time

Apache flink streaming is based on Records level while spark stream in based on time window

Flink keep all the records in memory and do the processing/aggregation considering the other records

While spark creates micro batch and process the data on certain window period and that is why we get low latency from flink while medium latecy from Spark