

Hadoop-docs

Running a Word Count Example

- To start Hadoop - `start-all.sh`
- Check if Hadoop has started:
 - URL for Name Node - `localhost:9870`
 - URL for Cluster info - `localhost:8088`
- Change directory to Downloads - `cd Downloads`
- Run `hadoop fs -mkdir /input`
- Run `hadoop fs -put input.txt /input`
- Run `hadoop jar WordCount-1.0-SNAPSHOT.jar org.example.WC_Runner /input/input.txt /output`
- Check the result at
 - `localhost:9870`
 - Utilities -> Browse file system
 - `/output/part-00000`
- Run `stop-all.sh`
- Run `hdfs datanode -format`
- Run `hdfs namenode -format`

Troubleshooting

- If `localhost:9870` is unable to connect
 - Run `jps`
 - Check if the list has `NameNode`
 - If not then
 - Stop Hadoop `stop-all.sh`
 - Run `hdfs namenode -format`
 - Run `start-all.sh`
- If `DataNode` hasn't started
 - Run `stop-all.sh`
 - Run `hdfs datanode -format`
 - Run `hdfs namenode -format`
 - Run `start-all.sh`