Hadoop examples

Prerequisites

Run these examples after you finished the the [Cluster Setup](#). Be familiar with the following commands before testing these examples.

```
hdfs namenode -format # Format the name node. This command is need to be run only once.
start-dfs.sh && start-yarn.sh
# Start Hadoop, make sure you execute this command before running the following examples.
stop-dfs.sh && stop-yarn.sh # Stop Hadoop
jps # Check the nodes of Hadoop
hdfs dfsadmin -report # Report Hadoop
```

Example 1: grep

```
hdfs dfs -mkdir -p input # Make a *input* folder on HDFS
hdfs dfs -put /usr/local/hadoop/etc/hadoop/*.xml input
# Put the xml files on the local file system (i.e. /usr/local/hadoop/etc/hadoop/) to the HDFS *inp
ut* folder
hdfs dfs -ls input # Check the content of *input* folder
hdfs dfs -rm -r output # Delete the *output* folder if it exists
hadoop jar /usr/
local/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-*.jar grep input output 'dfs[a-z.]+'
 # Find out all the files which contains the pattern 'dfs[a-z.]+' using grep
hdfs dfs -cat output/* # Display the result
```

Example 2: wordcount

```
hdfs dfs -mkdir -p input # Make a *input* folder on HDFS
hdfs dfs -put /usr/local/hadoop/etc/hadoop/*.xml input
# Put the xml files on the local file system (i.e. /usr/local/hadoop/etc/hadoop/) to the HDFS *inp
ut* folder
hdfs dfs -ls input # Check the content of *input* folder
hdfs dfs -rm -r output # Delete the *output* folder if it exists
hadoop jar /usr/
local/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-*.jar wordcount input output
# Count the words of all the files in the *input* folder
hdfs dfs -cat output/* # Display the result
```

Example 3: [Electrical consumption](#)

Example 4: [wordcount in python](#)

Example 5: [wordcount in C#](#)