

## Lect09 hadoop commands and output

### *# verify existence of files*

```
hduser@ubuntu:~$ ls -l /tmp/gutenberg/
total 3604
-rw-r--r-- 1 hduser hadoop 674566 Feb  3 10:17 pg20417.txt
-rw-r--r-- 1 hduser hadoop 1573112 Feb  3 10:18 pg4300.txt
-rw-r--r-- 1 hduser hadoop 1423801 Feb  3 10:18 pg5000.txt
hduser@ubuntu:~$
```

### *# copy files to HDFS*

```
hduser@ubuntu:~$ bin/hadoop dfs -copyFromLocal /tmp/gutenberg /user/hduser/gutenberg
hduser@ubuntu:~$ bin/hadoop dfs -ls
```

Found 1 items

```
drwxr-xr-x - hduser supergroup          0 2010-05-08 17:40 /user/hduser/gutenberg
```

```
hduser@ubuntu:~$ bin/hadoop dfs -ls /user/hduser/gutenberg
```

Found 3 items

```
-rw-r--r--  3 hduser supergroup      674566 2011-03-10 11:38 /user/hduser/gutenberg/pg20417.txt
-rw-r--r--  3 hduser supergroup     1573112 2011-03-10 11:38 /user/hduser/gutenberg/pg4300.txt
-rw-r--r--  3 hduser supergroup     1423801 2011-03-10 11:38 /user/hduser/gutenberg/pg5000.txt
hduser@ubuntu:~$
```

### *# run MapReduce job*

```
hduser@ubuntu:~$ bin/hadoop jar contrib/streaming/hadoop-*streaming*.jar \
-file /home/hduser/mapper.py -mapper /home/hduser/mapper.py \
-file /home/hduser/reducer.py -reducer /home/hduser/reducer.py \
-input /user/hduser/gutenberg/* -output /user/hduser/gutenberg-output
```

### *# increase number of Reduce tasks -- on the fly!!*

```
hduser@ubuntu:~$ bin/hadoop jar contrib/streaming/hadoop-*streaming*.jar -D mapred.reduce.tasks=16 ...
```

### *# homework: investigate contrast between mapred.reduce.tasks and mapred.map.tasks*

### *# console output*

```
hduser@ubuntu:~$ bin/hadoop jar contrib/streaming/hadoop-*streaming*.jar -mapper /home/hduser/mapper.py -reducer /home/hduser/reducer.py -input /user/hduser/gutenberg/* -output /user/hduser/gutenberg-output
additionalConfSpec_:null
null=@@@userJobConfProps_.get(stream.shipped.hadoopstreaming
packageJobJar: [/app/hadoop/tmp/hadoop-unjar54543/]
[] /tmp/streamjob54544.jar tmpDir=null
[...] INFO mapred.FileInputFormat: Total input paths to process : 7
[...] INFO streaming.StreamJob: getLocalDirs(): [/app/hadoop/tmp/mapred/local]
[...] INFO streaming.StreamJob: Running job: job_200803031615_0021
[...]
[...] INFO streaming.StreamJob: map 0% reduce 0%
[...] INFO streaming.StreamJob: map 43% reduce 0%
[...] INFO streaming.StreamJob: map 86% reduce 0%
[...] INFO streaming.StreamJob: map 100% reduce 0%
[...] INFO streaming.StreamJob: map 100% reduce 33%
[...] INFO streaming.StreamJob: map 100% reduce 70%
[...] INFO streaming.StreamJob: map 100% reduce 77%
[...] INFO streaming.StreamJob: map 100% reduce 100%
[...] INFO streaming.StreamJob: Job complete: job_200803031615_0021
[...] INFO streaming.StreamJob: Output: /user/hduser/gutenberg-output
hduser@ubuntu:~$
```