ASSIGNMENT # 3 – MapReduce

Slide # 73

```
[root@sandbox-hdp /]# cd /home/hdfs
[root@sandbox-hdp hdfs]# ls
map-reduce-sample.jar   SalesJan2009.csv
[root@sandbox-hdp hdfs]# 
```

Slide # 74

```
[hdfs@sandbox-hdp ~]$ hdfs dfs -copyFromLocal SalesJan2009.csv /test_data/
[hdfs@sandbox-hdp ~]$ hdfs dfs -ls /test_data/
Found 7 items
-rw-r--r--   1 hdfs hdfs      65835 2020-05-23 12:23 /test_data/SalesJan2009.csv
-rw-r--r--   1 root hdfs          0 2020-05-23 05:44 /test_data/hello
-rw-r--r--   1 root hdfs         21 2020-05-23 06:35 /test_data/hello1.dat
-rw-r--r--   1 root hdfs         26 2020-05-23 07:04 /test_data/hello3.dat
-rw-r--r--   1 root hdfs         13 2020-05-23 08:39 /test_data/how.dat
-rw-r--r--   1 root hdfs         29 2020-05-23 08:47 /test_data/how2.dat
-rw-r--r--   1 root hdfs          0 2020-05-23 08:54 /test_data/test.dat
```

Slide # 75

```
-rw-r--r--   1 root hdfs          0 2020-05-23 08:54 /test_data/test.dat
[hdfs@sandbox-hdp ~]$ hadoop jar map-reduce-sample.jar /test_data/SalesJan2009.csv /test_data/mp_output
20/05/23 12:27:24 INFO client.RMProxy: Connecting to ResourceManager at sandbox-hdp.hortonworks.com/172.18.0.
```

```
20/05/23 12:30:25 INFO mapreduce.Job:  map 0% reduce 0%
20/05/23 12:31:14 INFO mapreduce.Job:  map 50% reduce 0%
20/05/23 12:31:16 INFO mapreduce.Job:  map 100% reduce 0%
20/05/23 12:31:25 INFO mapreduce.Job:  map 100% reduce 100%
20/05/23 12:31:27 INFO mapreduce.Job: Job job_1590192907351_0004 completed successfully
20/05/23 12:31:28 INFO mapreduce.Job: Counters: 49
```

Slide # 76

```
  File System Counters
          FILE: Number of bytes read=12824
          FILE: Number of bytes written=485646
          FILE: Number of read operations=0
          FILE: Number of large read operations=0
          FILE: Number of write operations=0
          HDFS: Number of bytes read=98989
          HDFS: Number of bytes written=43
          HDFS: Number of read operations=9
          HDFS: Number of large read operations=0
          HDFS: Number of write operations=2
  Job Counters
          Launched map tasks=2
          Launched reduce tasks=1
          Data-local map tasks=2
          Total time spent by all maps in occupied slots (ms)=96327
          Total time spent by all reduces in occupied slots (ms)=8289
          Total time spent by all map tasks (ms)=96327
          Total time spent by all reduce tasks (ms)=8289
          Total vcore-milliseconds taken by all map tasks=96327
          Total vcore-milliseconds taken by all reduce tasks=8289
          Total megabyte-milliseconds taken by all map tasks=24081750
          Total megabyte-milliseconds taken by all reduce tasks=207225
  Map-Reduce Framework
          Map input records=998
          Map output records=998
          Map output bytes=10822
          Map output materialized bytes=12830
          Input split bytes=236
          Combine input records=0
          Combine output records=0
          Reduce input groups=4
          Reduce shuffle bytes=12830
          Reduce input records=998
```

Slide # 77

```
[hdfs@sandbox-hdp ~]$ hdfs dfs -ls /test_data/
Found 8 items
-rw-r--r--   1 hdfs hdfs        65835 2020-05-23 12:23 /test_data/SalesJan2009.csv
-rw-r--r--   1 root hdfs            0 2020-05-23 05:44 /test_data/hello
-rw-r--r--   1 root hdfs           21 2020-05-23 06:35 /test_data/hello1.dat
-rw-r--r--   1 root hdfs           26 2020-05-23 07:04 /test_data/hello3.dat
-rw-r--r--   1 root hdfs           13 2020-05-23 08:39 /test_data/how.dat
-rw-r--r--   1 root hdfs           29 2020-05-23 08:47 /test_data/how2.dat
drwxr-xr-x   - hdfs hdfs            0 2020-05-23 12:31 /test_data/mp_output
-rw-r--r--   1 root hdfs            0 2020-05-23 08:54 /test_data/test.dat
[hdfs@sandbox-hdp ~]$ hdfs dfs -ls /test_data/mp_output/
Found 2 items
-rw-r--r--   1 hdfs hdfs            0 2020-05-23 12:31 /test_data/mp_output/_SUCCESS
-rw-r--r--   1 hdfs hdfs           43 2020-05-23 12:31 /test_data/mp_output/part-00000
[hdfs@sandbox-hdp ~]$ hdfs dfs -cat /test_data/mp_output/_SUCCESS
[hdfs@sandbox-hdp ~]$ hdfs dfs -cat /test_data/mp_output/part-00000
Amex       110
Diners     89
Mastercard         277
Visa       522
```