Deployment of programs on Hadoop

The execution of the Map reduce program :

Cat books.json |python mapper_Index.py |sort +0 –1 | python reducer_Index.py

```
woe       bible-kjv.txt,shakespeare-caesar.txt
woman     milton-paradise.txt
woodhouse         austen-emma.txt
word      austen-emma.txt
work      whitman-leaves.txt
workman           edgeworth-parents.txt
world     shakespeare-caesar.txt
would     austen-emma.txt,chesterton-thursday.txt,melville-moby_dick.txt,milton-paradise.txt
write     shakespeare-caesar.txt
wrote     blake-poems.txt,whitman-leaves.txt
yea       blake-poems.txt
years     chesterton-thursday.txt
yet       austen-emma.txt
you       milton-paradise.txt,whitman-leaves.txt
young     chesterton-thursday.txt,shakespeare-caesar.txt
youngest          chesterton-thursday.txt
your      austen-emma.txt
biadmin:~/Desktop> cat  books.json |python mapper_Index.py |sort +0 -1 |python reducer_Index.py
```

Checking the creation of Input-dir :

Hadoop fs -ls /user/biadmin/

```
biadmin@bivm:/opt/ibm/biginsights/bin> hadoop fs -ls /user/biadmin/
Found 9 items
drwx------    - biadmin biadmin          0 2018-01-07 18:30 /user/biadmin/.staging
-rw-r--r--    1 biadmin biadmin        292 2018-01-07 17:14 /user/biadmin/MapReduce
-rw-r--r--    1 biadmin biadmin        580 2018-01-07 17:15 /user/biadmin/MapReducer
drwxr-xr-x    - biadmin biadmin          0 2018-01-07 16:35 /user/biadmin/input-dir
-rw-r--r--    1 biadmin biadmin    1154214 2017-12-28 07:22 /user/biadmin/lab
-rw-r--r--    1 biadmin biadmin    1154214 2017-12-28 07:22 /user/biadmin/lab.pdf
drwxr-xr-x    - biadmin biadmin          0 2017-12-27 20:32 /user/biadmin/myTestDir
drwxr-xr-x    - biadmin biadmin          0 2017-12-28 06:53 /user/biadmin/myTestDir2
-rw-r--r--    1 biadmin biadmin         78 2017-12-28 10:37 /user/biadmin/test
```

Make scripts executable :

Chmod u+x /home/biadmin/Desktop/mapper_Index.py

Chmod u+x /home/biadmin/Desktop/mapper_Index.py

```
biadmin@bivm:/opt/ibm/biginsights/bin> chmod u+x /home/biadmin/Desktop/mapper_Index.py
biadmin@bivm:/opt/ibm/biginsights/bin> chmod u+x /home/biadmin/Desktop/reducer_Index.py
```

Hadoop Streaming command execution :

Hadoop jar /opt/ibm/biginsights/IHC/share/hadoop/tools/lib/hadoop-streaming-2.2.0.jar –mapper /home/biadmin/Desktop/mapper_Index.py -reducer /home/biadmin/Desktop/reducer_Index.py –input /user/biadmin/input-dir/books.json –output output-dir

```
biadmin@bivm:/opt/ibm/biginsights/bin> hadoop jar /opt/ibm/biginsights/IHC/share/hadoop/tools/lib/hadoop-streaming-2.2.0.jar -mapper /home/biadmin/Desktop/ma
pper_Index.py -reducer /home/biadmin/Desktop/reducer_Index.py -input /user/biadmin/input-dir/books.json -output output-dir
packageJobJar: [/var/ibm/biginsights/hadoop/tmp/biadmin/hadoop-unjar4197111830951781944/] [] /tmp/streamjob8922024520695247229.jar tmpDir=null
18/01/07 18:43:36 INFO mapred.FileInputFormat: Total input paths to process : 1
18/01/07 18:43:36 INFO mapred.JobClient: Running job: job_201801050944_0039
18/01/07 18:43:37 INFO mapred.JobClient:  map 0% reduce 0%
18/01/07 18:45:26 INFO mapred.JobClient:  map 100% reduce 0%
18/01/07 18:45:49 INFO mapred.JobClient:  map 100% reduce 100%
18/01/07 18:45:51 INFO mapred.JobClient: Job complete: job_201801050944_0039
18/01/07 18:45:51 INFO mapred.JobClient: Counters: 30
18/01/07 18:45:51 INFO mapred.JobClient:   File System Counters
18/01/07 18:45:51 INFO mapred.JobClient:     FILE: BYTES_READ=32119
18/01/07 18:45:51 INFO mapred.JobClient:     FILE: BYTES_WRITTEN=606458
18/01/07 18:45:51 INFO mapred.JobClient:     HDFS: BYTES_READ=10128
18/01/07 18:45:51 INFO mapred.JobClient:     HDFS: BYTES_WRITTEN=23264
18/01/07 18:45:51 INFO mapred.JobClient:   org.apache.hadoop.mapreduce.JobCounter
18/01/07 18:45:51 INFO mapred.JobClient:     TOTAL_LAUNCHED_MAPS=2
18/01/07 18:45:51 INFO mapred.JobClient:     TOTAL_LAUNCHED_REDUCES=1
18/01/07 18:45:51 INFO mapred.JobClient:     DATA_LOCAL_MAPS=2
18/01/07 18:45:51 INFO mapred.JobClient:     SLOTS_MILLIS_MAPS=207842
18/01/07 18:45:51 INFO mapred.JobClient:     SLOTS_MILLIS_REDUCES=22762
18/01/07 18:45:51 INFO mapred.JobClient:     FALLOW_SLOTS_MILLIS_MAPS=0
18/01/07 18:45:51 INFO mapred.JobClient:     FALLOW_SLOTS_MILLIS_REDUCES=0
18/01/07 18:45:51 INFO mapred.JobClient:   org.apache.hadoop.mapreduce.TaskCounter
18/01/07 18:45:51 INFO mapred.JobClient:     MAP_INPUT_RECORDS=10
18/01/07 18:45:51 INFO mapred.JobClient:     MAP_OUTPUT_RECORDS=1159
18/01/07 18:45:51 INFO mapred.JobClient:     MAP_OUTPUT_BYTES=29795
18/01/07 18:45:51 INFO mapred.JobClient:     MAP_OUTPUT_MATERIALIZED_BYTES=32125
18/01/07 18:45:51 INFO mapred.JobClient:     SPLIT_RAW_BYTES=222
18/01/07 18:45:51 INFO mapred.JobClient:     COMBINE_INPUT_RECORDS=0
18/01/07 18:45:51 INFO mapred.JobClient:     COMBINE_OUTPUT_RECORDS=0
18/01/07 18:45:51 INFO mapred.JobClient:     REDUCE_INPUT_GROUPS=529
18/01/07 18:45:51 INFO mapred.JobClient:     REDUCE_SHUFFLE_BYTES=32125
18/01/07 18:45:51 INFO mapred.JobClient:     REDUCE_INPUT_RECORDS=1159
18/01/07 18:45:51 INFO mapred.JobClient:     REDUCE_OUTPUT_RECORDS=529
18/01/07 18:45:51 INFO mapred.JobClient:     SPILLED_RECORDS=2318
18/01/07 18:45:51 INFO mapred.JobClient:     CPU_MILLISECONDS=123000
18/01/07 18:45:51 INFO mapred.JobClient:     PHYSICAL_MEMORY_BYTES=922079232
```

```
18/01/07 18:45:51 INFO mapred.JobClient:     PHYSICAL_MEMORY_BYTES=922079232
18/01/07 18:45:51 INFO mapred.JobClient:     VIRTUAL_MEMORY_BYTES=5234331648
18/01/07 18:45:51 INFO mapred.JobClient:     COMMITTED_HEAP_BYTES=3145728000
18/01/07 18:45:51 INFO mapred.JobClient:   org.apache.hadoop.mapred.FileInputFormat$Counter
18/01/07 18:45:51 INFO mapred.JobClient:     BYTES_READ=9906
18/01/07 18:45:51 INFO mapred.JobClient:   org.apache.hadoop.mapred.FileOutputFormat$Counter
18/01/07 18:45:51 INFO mapred.JobClient:     BYTES_WRITTEN=23264
18/01/07 18:45:51 INFO mapred.JobClient:   org.apache.hadoop.mapreduce.lib.input.FileInputFormatCounter
18/01/07 18:45:51 INFO mapred.JobClient:     BYTES_READ=6604
18/01/07 18:45:51 INFO streaming.StreamJob: Output directory: output-dir
```

Checking the creation of output-dir :

Hadoop fs –ls output-dir

```
biadmin@bivm:/opt/ibm/biginsights/bin> hadoop fs -ls output-dir
Found 3 items
-rw-r--r--   1 biadmin biadmin          0 2018-01-07 18:45 output-dir/_SUCCESS
drwx--x--x   - biadmin biadmin          0 2018-01-07 18:43 output-dir/_logs
-rw-r--r--   1 biadmin biadmin      23264 2018-01-07 18:45 output-dir/part-00000
```

Display the content of the part :

hadoop fs –cat output-dir/part-00000

```
woe:bible-kjv.txt,shakespeare-caesar.txt
woman:milton-paradise.txt
woodhouse:austen-emma.txt
word:austen-emma.txt
work:whitman-leaves.txt
workman:edgeworth-parents.txt
world:shakespeare-caesar.txt
would:melville-moby_dick.txt,chesterton-thursday.txt,milton-paradise.txt,austen-emma.txt,chesterton-thursday.txt
write:shakespeare-caesar.txt
wrote:whitman-leaves.txt,blake-poems.txt
yea:blake-poems.txt
years:chesterton-thursday.txt
yet:austen-emma.txt
you:milton-paradise.txt,whitman-leaves.txt
young:shakespeare-caesar.txt,chesterton-thursday.txt
youngest:chesterton-thursday.txt
your:austen-emma.txt
biadmin@bivm:~/Desktop> hadoop fs -cat output-dir/part-00000
```

Creating the Table :

Create 'Books', 'Words' , 'Files'

```
hbase(main):013:0> create 'Books', 'Words', 'Files'
0 row(s) in 1.0120 seconds

=> Hbase::Table - Books

hbase(main):015:0> scan 'Books'
ROW                              COLUMN+CELL
0 row(s) in 0.0230 seconds
```

Import the part to the Table:

/opt/ibm/biginsights/hbase/bin/hbase org.apache.hadoop.hbase.mapreduce.ImportTsv –Dimporttsv !;separator=' :' – Dimporttsv.colums=HBASE ,ROW_KEY,Files Books ouput-dir/part-00000

```
^Cbiadmin@bivm:~/Desktop> /opt/ibm/biginsights/hbase/bin/hbase org.apache.hadoop.hbase.mapreduce.ImportTsv -Dimporttsv.separator=':' -Dimporttsv.columns=HBAS
ROW_KEY,Files Books output-dir/part-00000
18/01/07 20:01:25 INFO zookeeper.ZooKeeper: Client environment:zookeeper.version=3.4.5--1, built on 02/05/2013 08:46 GMT
18/01/07 20:01:25 INFO zookeeper.ZooKeeper: Client environment:host.name=bivm.ibm.com
18/01/07 20:01:25 INFO zookeeper.ZooKeeper: Client environment:java.version=1.7.0
18/01/07 20:01:25 INFO zookeeper.ZooKeeper: Client environment:java.vendor=IBM Corporation
18/01/07 20:01:25 INFO zookeeper.ZooKeeper: Client environment:java.home=/opt/ibm/biginsights/jdk/jre
18/01/07 20:01:25 INFO zookeeper.ZooKeeper: Client environment:java.class.path=/opt/ibm/biginsights/IHC/:/opt/ibm/biginsights/IHC/:/opt/ibm/biginsights/hadoo
p-conf:/opt/ibm/biginsights/hbase/conf:/opt/ibm/biginsights/lib/biginsights-gpfs.jar:/opt/ibm/biginsights/IHC/hadoop-core.jar::/opt/ibm/biginsights/jdk/lib/t
ools.jar:/opt/ibm/biginsights/hbase:/opt/ibm/biginsights/hbase/lib/activation-1.1.jar:/opt/ibm/biginsights/hbase/lib/antlr-runtime-3.4.jar:/opt/ibm/biginsigh
ts/hbase/lib/aopalliance-1.0.jar:/opt/ibm/biginsights/hbase/lib/asm-3.1.jar:/opt/ibm/biginsights/hbase/lib/avro-1.7.4.jar:/opt/ibm/biginsights/hbase/lib/comm
ons-beanutils-1.8.0.jar:/opt/ibm/biginsights/hbase/lib/commons-cli-1.2.jar:/opt/ibm/biginsights/hbase/lib/commons-codec-1.7.jar:/opt/ibm/biginsights/hbase/li
b/commons-collections-3.2.1.jar:/opt/ibm/biginsights/hbase/lib/commons-compress-1.4.1.jar:/opt/ibm/biginsights/hbase/lib/commons-configuration-1.6.jar:/opt/i
bm/biginsights/hbase/lib/commons-daemon-1.0.5.jar:/opt/ibm/biginsights/hbase/lib/commons-digester-1.8.jar:/opt/ibm/biginsights/hbase/lib/commons-el-1.0.jar:/
opt/ibm/biginsights/hbase/lib/commons-httpclient-3.1.jar:/opt/ibm/biginsights/hbase/lib/commons-io-2.4.jar:/opt/ibm/biginsights/hbase/lib/commons-lang-2.6.ja
r:/opt/ibm/biginsights/hbase/lib/commons-logging-1.1.1.jar:/opt/ibm/biginsights/hbase/lib/commons-math-2.2.jar:/opt/ibm/biginsights/hbase/lib/commons-net-3.1
.jar:/opt/ibm/biginsights/hbase/lib/core-3.1.1.jar:/opt/ibm/biginsights/hbase/lib/findbugs-annotations-1.3.9-1.jar:/opt/ibm/biginsights/hbase/lib/guardium-pr
oxy.jar:/opt/ibm/biginsights/hbase/lib/guava-12.0.1.jar:/opt/ibm/biginsights/hbase/lib/guice-3.0.jar:/opt/ibm/biginsights/hbase/lib/guice-servlet-3.0.jar:/op
t/ibm/biginsights/hbase/lib/hadoop-core.jar:/opt/ibm/biginsights/hbase/lib/hadoop-distcp-2.2.0.jar:/opt/ibm/biginsights/hbase/lib/hbase-backup-0.96.0.jar:/op
t/ibm/biginsights/hbase/lib/hbase-client-0.96.0.jar:/opt/ibm/biginsights/hbase/lib/hbase-common-0.96.0.jar:/opt/ibm/biginsights/hbase/lib/hbase-common-0.96.0
-tests.jar:/opt/ibm/biginsights/hbase/lib/hbase-examples-0.96.0.jar:/opt/ibm/biginsights/hbase/lib/hbase-hadoop2-compat-0.96.0.jar:/opt/ibm/biginsights/hbase
/lib/hbase-hadoop-compat-0.96.0.jar:/opt/ibm/biginsights/hbase/lib/hbase-it-0.96.0.jar:/opt/ibm/biginsights/hbase/lib/hbase-it-0.96.0-tests.jar:/opt/ibm/bigi
nsights/hbase/lib/hbase-prefix-tree-0.96.0.jar:/opt/ibm/biginsights/hbase/lib/hbase-protocol-0.96.0.jar:/opt/ibm/biginsights/hbase/lib/hbase-server-0.96.0.ja
```

```
18/01/07 20:01:28 INFO mapreduce.TableOutputFormat: Created table instance for Books
18/01/07 20:01:28 INFO input.FileInputFormat: Total input paths to process : 1
18/01/07 20:01:29 INFO mapred.JobClient: Running job: job_201801050944_0062
18/01/07 20:01:30 INFO mapred.JobClient:   map 0% reduce 0%
18/01/07 20:01:53 INFO mapred.JobClient:   map 100% reduce 0%
18/01/07 20:01:56 INFO mapred.JobClient: Job complete: job_201801050944_0062
18/01/07 20:01:56 INFO mapred.JobClient: Counters: 19
18/01/07 20:01:56 INFO mapred.JobClient:    File System Counters
18/01/07 20:01:56 INFO mapred.JobClient:      FILE: BYTES_WRITTEN=241743
18/01/07 20:01:56 INFO mapred.JobClient:      HDFS: BYTES_READ=22330
18/01/07 20:01:56 INFO mapred.JobClient:    org.apache.hadoop.mapreduce.JobCounter
18/01/07 20:01:56 INFO mapred.JobClient:      TOTAL_LAUNCHED_MAPS=1
18/01/07 20:01:56 INFO mapred.JobClient:      DATA_LOCAL_MAPS=1
18/01/07 20:01:56 INFO mapred.JobClient:      SLOTS_MILLIS_MAPS=14085
18/01/07 20:01:56 INFO mapred.JobClient:      SLOTS_MILLIS_REDUCES=0
18/01/07 20:01:56 INFO mapred.JobClient:      FALLOW_SLOTS_MILLIS_MAPS=0
18/01/07 20:01:56 INFO mapred.JobClient:      FALLOW_SLOTS_MILLIS_REDUCES=0
18/01/07 20:01:56 INFO mapred.JobClient:    org.apache.hadoop.mapreduce.TaskCounter
18/01/07 20:01:56 INFO mapred.JobClient:      MAP_INPUT_RECORDS=529
18/01/07 20:01:56 INFO mapred.JobClient:      MAP_OUTPUT_RECORDS=529
18/01/07 20:01:56 INFO mapred.JobClient:      SPLIT_RAW_BYTES=124
18/01/07 20:01:56 INFO mapred.JobClient:      SPILLED_RECORDS=0
18/01/07 20:01:56 INFO mapred.JobClient:      CPU_MILLISECONDS=640
18/01/07 20:01:56 INFO mapred.JobClient:      PHYSICAL_MEMORY_BYTES=178130944
18/01/07 20:01:56 INFO mapred.JobClient:      VIRTUAL_MEMORY_BYTES=1771032576
18/01/07 20:01:56 INFO mapred.JobClient:      COMMITTED_HEAP_BYTES=1048576000
18/01/07 20:01:56 INFO mapred.JobClient:    ImportTsv
18/01/07 20:01:56 INFO mapred.JobClient:      Bad Lines=0
18/01/07 20:01:56 INFO mapred.JobClient:    File Input Format Counters
18/01/07 20:01:56 INFO mapred.JobClient:      Bytes Read=22206
18/01/07 20:01:56 INFO mapred.JobClient:    org.apache.hadoop.mapreduce.lib.output.FileOutputFormat$Counter
18/01/07 20:01:56 INFO mapred.JobClient:      BYTES_WRITTEN=0
```

Checking that the table is loaded :

Scan 'Books'

```
woe          column=Files:, timestamp=1515373284224, value=bible-kjv.txt,shakespeare-caesar.txt\x09
woman        column=Files:, timestamp=1515373284224, value=milton-paradise.txt\x09
woodhouse    column=Files:, timestamp=1515373284224, value=austen-emma.txt\x09
word         column=Files:, timestamp=1515373284224, value=austen-emma.txt\x09
work         column=Files:, timestamp=1515373284224, value=whitman-leaves.txt\x09
workman      column=Files:, timestamp=1515373284224, value=edgeworth-parents.txt\x09
world        column=Files:, timestamp=1515373284224, value=shakespeare-caesar.txt\x09
would        column=Files:, timestamp=1515373284224, value=melville-moby_dick.txt,chesterton-thursday.txt,milton-paradise.txt,aus
             ten-emma.txt,chesterton-thursday.txt\x09
write        column=Files:, timestamp=1515373284224, value=shakespeare-caesar.txt\x09
wrote        column=Files:, timestamp=1515373284224, value=whitman-leaves.txt,blake-poems.txt\x09
yea          column=Files:, timestamp=1515373284224, value=blake-poems.txt\x09
years        column=Files:, timestamp=1515373284224, value=chesterton-thursday.txt\x09
yet          column=Files:, timestamp=1515373284224, value=austen-emma.txt\x09
you          column=Files:, timestamp=1515373284224, value=milton-paradise.txt,whitman-leaves.txt\x09
young        column=Files:, timestamp=1515373284224, value=shakespeare-caesar.txt,chesterton-thursday.txt\x09
youngest     column=Files:, timestamp=1515373284224, value=chesterton-thursday.txt\x09
your         column=Files:, timestamp=1515373284224, value=austen-emma.txt\x09
529 row(s) in 0.6810 seconds

hbase(main):017:0> scan 'Books'
```

Scheduling some commands using Cron : Executing **Crontab –e** which opens the vi editor

20 16 08 01 01 / cat /home/biadmin/Desktop/books.json >> /home/biadmin/Desktop/schedule.txt

58 3 8 1 1 ls /home/biadmin/Desktop >> /home/biadmin/Desktop/schedule.txt

39 16 08 01 01 hadoop jar /opt/ibm/biginsights/IHC/share/hadoop/tools/lib/hadoop-streaming-2.2.0.jar -mapper /home/biadmin/Desktop/mapper_Index.py –reducer /home/biadmin/Desktop/reducer_Index.py -input /user/biadmin/input-dir/books.json -output out | /opt/ibm/biginsights/IHC/bin/hadoop >> /home/biadmin/Desktop/schedule.txt

51 16 08 01 01 hadoop fs -cat output-dir/part-00000 >> /home/biadmin/Desktop/schedule.txt

Testing the research in Hbase ("you" word) :

Get 'Books' , 'you'

```
hbase(main):002:0> get 'Books', 'you'
COLUMN                              CELL
 Files:                             timestamp=1515373284224, value=milton-paradise.txt,whitman-leaves.txt\x09
1 row(s) in 0.0890 seconds
```

Writing the Script Shell :

```bash
#!/bin/bash
echo "PLease enter the word you are looking for: "
read word
echo "get 'Books', '$word'" | /opt/ibm/biginsights/hbase/bin/hbase shell
```

Changing the permisions :

Chmod u+x /home/biadmin/Desktop/test.sh

```
biadmin@bivm:/opt/ibm/biginsights/bin> chmod u+x /home/biadmin/Desktop/test.sh
```

Executing the Script Shell :

./test.sh

```
biadmin@bivm:~/Desktop> ./test.sh
PLease enter the word you are looking for:
you
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 0.96.0, rd4281a01692003ff4ba6240277f647317b1acb0b, Mon Nov 17 10:58:35 PST 2014

get 'Books', 'you'
COLUMN                              CELL
 Files:                             timestamp=1515373284224, value=milton-paradise.txt,whitman-leaves.txt\x09
1 row(s) in 0.1100 seconds
```