***Advanced Operating Systems Project Phase 1 Report***

**Members:**

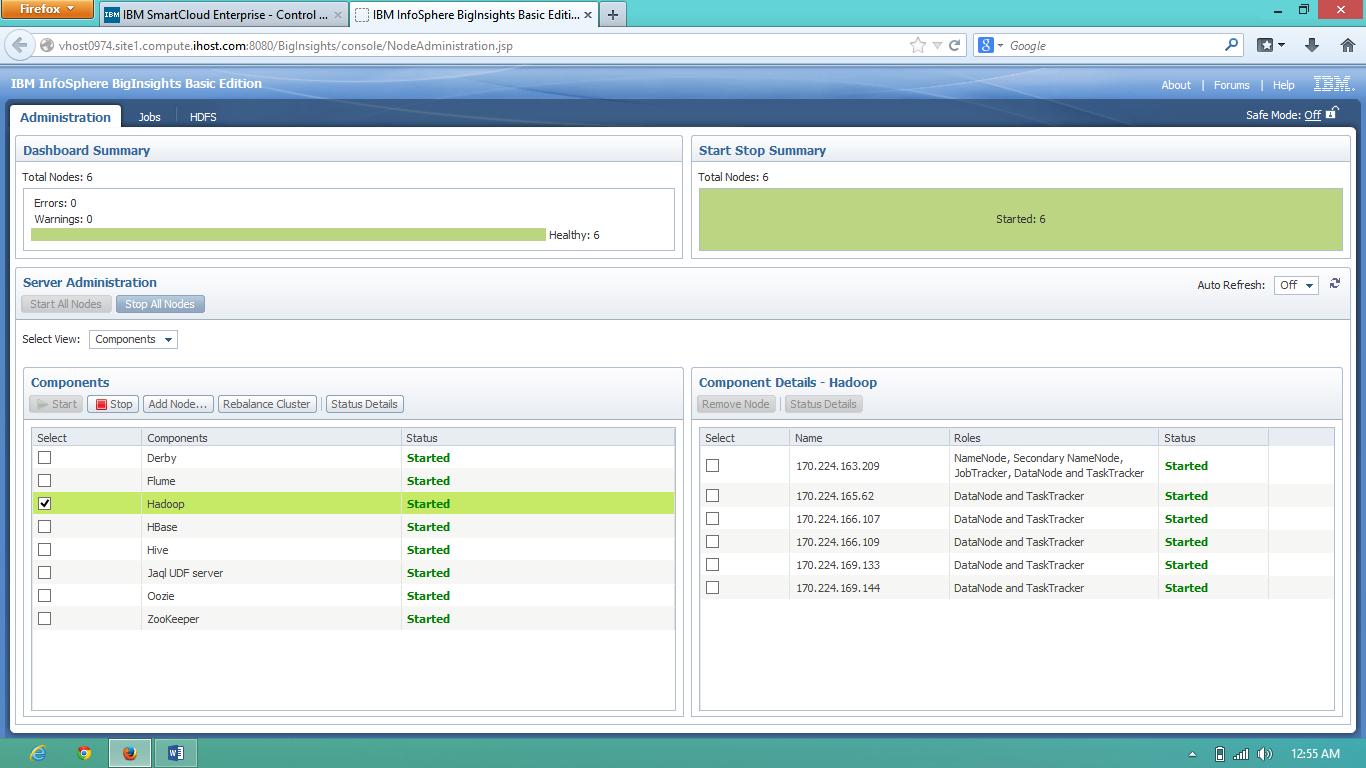
Nayab Rasool Shaik

Md Imrul Mukit Siddique

Venkata Vamsi Siva Krishna Vura

**Screenshot of BigInsights Web Console where Hadoop as the component**

**with 6 healthy nodes**



**Running the word count example:**

Using username "idcuser".

Authenticating with public key "imported-openssh-key"

[idcuser@vhost0974 ~]$ $HADOOP\_HOME/bin/hadoop dfs -mkdir /user

[idcuser@vhost0974 ~]$ $HADOOP\_HOME/bin/hadoop dfs -mkdir /user/idcuser

[idcuser@vhost0974 ~]$ $HADOOP\_HOME/bin/hadoop dfs -mkdir /user/idcuser/cs5531

[idcuser@vhost0974 ~]$ $HADOOP\_HOME/bin/hadoop dfs -copyFromLocal /home/idcuser/cs5531/Input/testinput.dat /user/idcuser/cs5531

[idcuser@vhost0974 ~]$ $HADOOP\_HOME/bin/hadoop jar /mnt/biginsights/opt/ibm/biginsights/IHC/hadoop-examples-1.0.3.jar wordcount /user/idcuser/cs5531/testinput.dat /user/idcuser/cs5531/Output

13/09/19 03:09:56 INFO input.FileInputFormat: Total input paths to process : 1

13/09/19 03:09:56 INFO mapred.JobClient: Running job: job\_201309162310\_0030

13/09/19 03:09:57 INFO mapred.JobClient: map 0% reduce 0%

13/09/19 03:10:05 INFO mapred.JobClient: map 33% reduce 0%

13/09/19 03:10:07 INFO mapred.JobClient: map 51% reduce 0%

13/09/19 03:10:10 INFO mapred.JobClient: map 61% reduce 0%

13/09/19 03:10:13 INFO mapred.JobClient: map 72% reduce 0%

13/09/19 03:10:16 INFO mapred.JobClient: map 83% reduce 0%

13/09/19 03:10:19 INFO mapred.JobClient: map 93% reduce 0%

13/09/19 03:10:22 INFO mapred.JobClient: map 99% reduce 0%

13/09/19 03:10:23 INFO mapred.JobClient: map 100% reduce 0%

13/09/19 03:10:34 INFO mapred.JobClient: map 100% reduce 100%

13/09/19 03:10:34 INFO mapred.JobClient: Job complete: job\_201309162310\_0030

13/09/19 03:10:34 INFO mapred.JobClient: Counters: 29

13/09/19 03:10:34 INFO mapred.JobClient: Job Counters

13/09/19 03:10:34 INFO mapred.JobClient: Data-local map tasks=3

13/09/19 03:10:34 INFO mapred.JobClient: SLOTS\_MILLIS\_MAPS=51327

13/09/19 03:10:34 INFO mapred.JobClient: Launched map tasks=3

13/09/19 03:10:34 INFO mapred.JobClient: Total time spent by all reduces waiting after reserving slots (ms)=0

13/09/19 03:10:34 INFO mapred.JobClient: Total time spent by all maps waiting after reserving slots (ms)=0

13/09/19 03:10:34 INFO mapred.JobClient: Launched reduce tasks=1

13/09/19 03:10:34 INFO mapred.JobClient: SLOTS\_MILLIS\_REDUCES=14176

13/09/19 03:10:34 INFO mapred.JobClient: File Input Format Counters

13/09/19 03:10:34 INFO mapred.JobClient: Bytes Read=287323058

13/09/19 03:10:34 INFO mapred.JobClient: File Output Format Counters

13/09/19 03:10:34 INFO mapred.JobClient: Bytes Written=20473017

13/09/19 03:10:34 INFO mapred.JobClient: FileSystemCounters

13/09/19 03:10:34 INFO mapred.JobClient: FILE\_BYTES\_WRITTEN=65968767

13/09/19 03:10:34 INFO mapred.JobClient: HDFS\_BYTES\_READ=287323436

13/09/19 03:10:34 INFO mapred.JobClient: FILE\_BYTES\_READ=43833103

13/09/19 03:10:34 INFO mapred.JobClient: HDFS\_BYTES\_WRITTEN=20473017

13/09/19 03:10:34 INFO mapred.JobClient: Map-Reduce Framework

13/09/19 03:10:34 INFO mapred.JobClient: Virtual memory (bytes) snapshot=4237762560

13/09/19 03:10:34 INFO mapred.JobClient: Reduce input groups=325029

13/09/19 03:10:34 INFO mapred.JobClient: Combine output records=726183

13/09/19 03:10:34 INFO mapred.JobClient: Map output records=9254838

13/09/19 03:10:34 INFO mapred.JobClient: CPU time spent (ms)=73580

13/09/19 03:10:34 INFO mapred.JobClient: Map input records=2159320

13/09/19 03:10:34 INFO mapred.JobClient: Reduce shuffle bytes=11811982

13/09/19 03:10:34 INFO mapred.JobClient: Combine input records=9640443

13/09/19 03:10:34 INFO mapred.JobClient: Spilled Records=1066761

13/09/19 03:10:34 INFO mapred.JobClient: SPLIT\_RAW\_BYTES=378

13/09/19 03:10:34 INFO mapred.JobClient: Map output bytes=324211336

13/09/19 03:10:34 INFO mapred.JobClient: Reduce input records=340578

13/09/19 03:10:34 INFO mapred.JobClient: Physical memory (bytes) snapshot=1494781952

13/09/19 03:10:34 INFO mapred.JobClient: Total committed heap usage (bytes)=1200047104

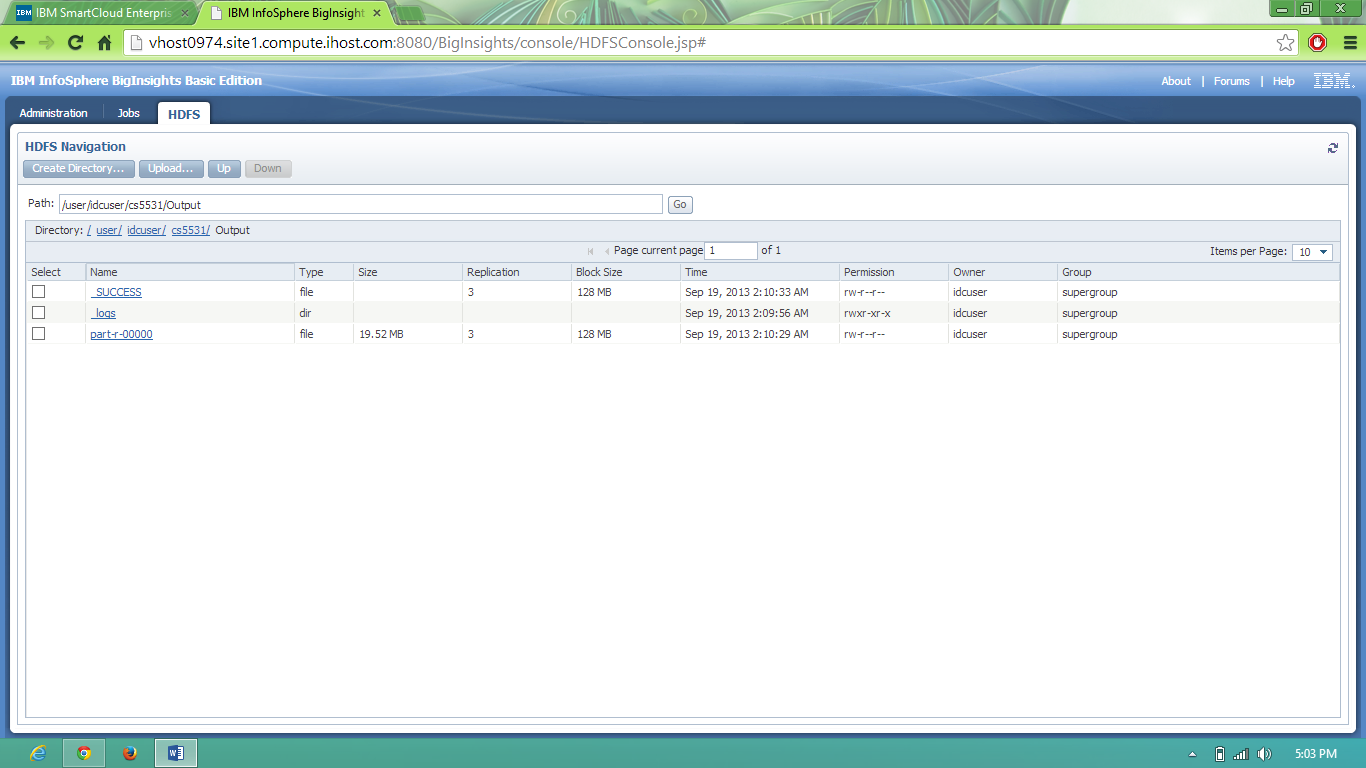
13/09/19 03:10:34 INFO mapred.JobClient: Reduce output records=325029

13/09/19 03:10:34 INFO mapred.JobClient: Map output materialized bytes=22044971

[idcuser@vhost0974 ~]$

**Output is in the attachment**

**Screenshot of output:**



**How can you enable Hadoop to use the Fair Scheduler for running jobs?**  
Answer: There are two steps to follow to enable fair scheduler:  
    1. Copy jar file to {hadoop\_home}/lib directory. Or add the location of the jar file in the classpath entry.  
    2. Change the property value of task scheduler in the mapred-site.xml file.  
Then restart the cluster.  
  
**How can you revert to FIFO scheduling within the Fair Scheduler?**  
Answer: There are two configuration xml file- mapred-site.xml and fair-scheduler.xml. In the second xml file there is a property named "defaultPoolSchedulingMode" which takes "fifo" or "fair" as value. Depending on the value the scheduler will run the corresponding scheduler.