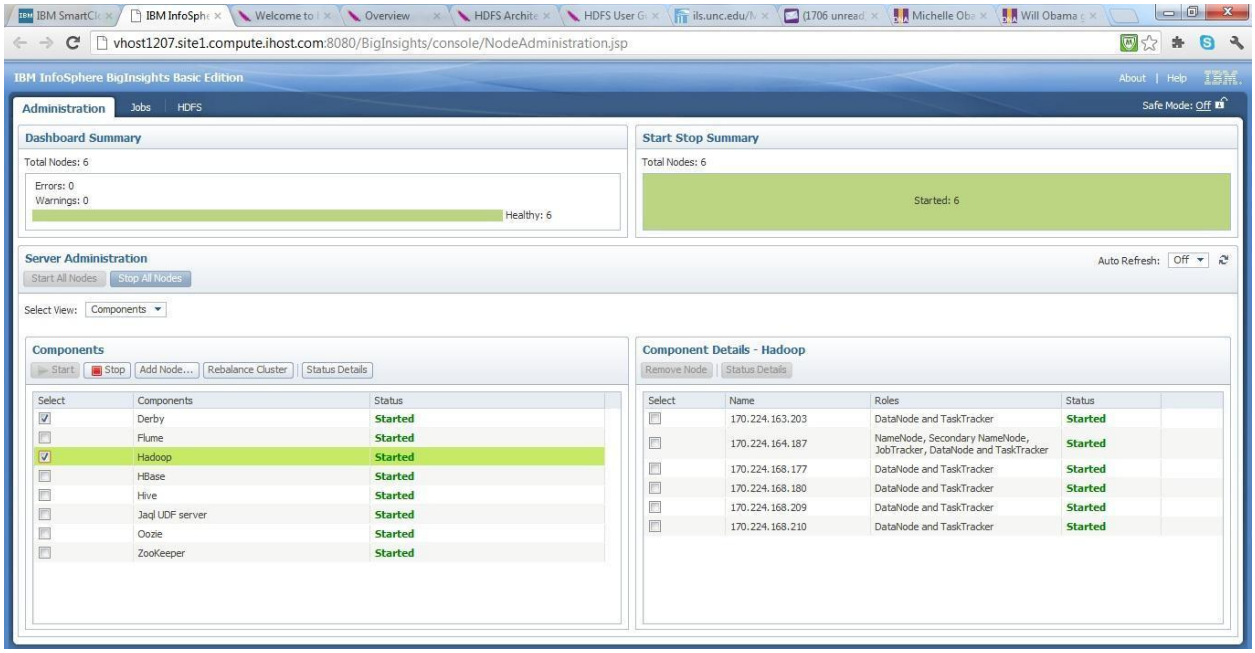
COMP-SCI 5531 0001- Advanced Operating Systems FS2012

Project Phase1

Team members:

1. Kalyana Venkata Ramana Gelli
2. Samyuktha Anumolu

Question 1:



Question 2:

Running WordCount:

hadoop jar hadoop-0.20.2-examples.jar wordcount /user/idcuser/cs5531/testinput.dat /user/idcuser/cs5531/output

Output obtained: 7 files

Concatenated output data is in output file (attached with this submission).

Question 3:

a)

In order to enable fair scheduler in hadoop, we have to follow the below steps.

1. Copy the *hadoop-*-fairscheduler.jar* from*HADOOP_HOME/contrib/fairscheduler* to *HADOOP_HOME/lib*.
2. Compile fair scheduler, by executing *ant package* in source folder and copy the *build/contrib/fair-scheduler/hadoop-*-fairscheduler.jar* to*HADOOP_HOME/lib*
3. Set the property of "mapred.jobtracker.taskScheduler" in Hadoop config file *HADOOP_CONF_DIR/mapred-site.xml* to value "org.apache.hadoop.mapred.FairScheduler"

From code perspective:

Source Code Download: ***ftp://apache.mirrors.pair.com/hadoop/common/hadoop-0.20.2/*** hadoop-0.20.2.tar.gz

Path of the code files:

hadoop-0.20.2\src\contrib\fairscheduler\src\java\org\apache\hadoop\mapred

Create an object of class FairScheduler (Source: FairScheduler.java) and call the method start(). This enables FairScheduling in Hadoop. DeficitComparator is the one which sorts jobs by deficit since the feature of fair scheduling is to assign resources to jobs such that all jobs get, on average, an equal share of resources over time.

The syntax for calling start() method can be found in TestFairScheduler.java (JUnit Test suite) at path "hadoop-0.20.2\src\contrib\fairscheduler\src\test\org\apache\hadoop\mapred\" which invokes FairScheduling in Hadoop.

b) There is a variable called useFifo in FairScheduler class. In order to revert to FIFO in hadoop this variable should be set to true. When it is set to true FifoJobComparator will be called instead of DeficitComparator (which is used in case of FairScheduling) and jobs are sorted by submit time.

Ideally from UI, when FIFO is selected the request passes a QueryString parameter called setFifo with its value true. Hence when servlet (FairSchedulerServlet) is run it sets the variable useFifo in FairScheduler class to true and hadoop is reverted to FIFO scheduling.