#### FutureSystems VM Assignment

Abderrezak Allalen

Varun Patil

April 6th 2015

Indiana University

Using OpenStack

**FurtureGrid and OpenStack:**

After registering for FutureGrid acount, a user name a password is created. The next step was to generate private and public keys. Once the keys are generated, the public key is then copied and pasted into the “My Portal Account >> SSH Keys>>Add public key”. After 24 hours the account is all set up and the access to resources step should show valid status.

Once the steps mentioned above are completed, The FutureGrid India OpenStack is ready to be used. After logging in to FutureGrid India Cluster using the private key created previously, we load OpenStack and then load OpenStack Nova security credentials. We can see the available virtual images and available instances using the nova commands provided to us. Next we create our own Nova ssh key pair for our Futuregrid started instance, then add security group for SSH connection. Next we set the net list variable for using the internal network and boot OpenStack Nova virtual compute node. If the VM is active, we log into the VM with our nova ssh private key.

Once all the steps mentioned in the second paragraph are completed, we start the two VM instances and login to both of them.

We can log into the system by using the following command:  
ssh -i ~/.ssh/$USER-fg464-vm-sshPrivatekey [ubuntu@10.23.0.XX](mailto:ubuntu@10.23.0.XX)  
We obtain the information for the correct node instances and the IP addresses on the Futuregrid system by using the ‘nova list’ command. Thus, we can replace the XX in the mentioned command with IP addresses that have been mentioned in the nova list.  
Once we log into our system, we execute the following commands:  
sudo vi /etc/hosts  
10.23.0.251 <username>-001  
192.23.0.252 <username>-002  
This enables us to add the IPs and the hostnames for all nodes.  
  
**Hadoop Configuration:**

We proceed to edit the $HADOOP\_HOME/nodes on the first login nodes with the help of the following script:  
vi ~/software/hadoop-1.1.2/nodes  
We proceed to run the following commands to start Hadoop:  
bash> cd $HADOOP\_HOME  
bash> ./MultiNodesOneClickStartUp.sh $JAVA\_HOME nodes  
These commands are run form the first node in the node file and enable the running of Hadoop and MapReduce daemons on the node.  
Hadoop status has to be checked to verify if the Hadoop was setup and started properly.  
Once Hadoop is set up and running, we can edit, compile and run the WordCount and the PageRank algorithms on the Hadoop system. We make sure the programs are with administrator permissions to make sure that there are no further “Permission denied” errors.  
The steps to run WordCount on Hadoop are as follows:  
cd ~/MoocHomeworks/HadoopWordCount

cp input/Word\_Count\_input.txt input/Word\_Count\_input2.txt  
hadoop fs -rmr output  
hadoop fs -rmr input  
hadoop fs -put input  
./build.sh  
hadoop jar WordCount.jar WordCount input output  
hadoop fs -cat output/\*   
After this, the output of this execution is stored in WordCount-Output.txt and we note the time required for execution.

**Analysis of PageRank on OpenStack:**

The PageRank algorithm was executed twice on the OpenStack by using one and two nodes respectively.  
The execution timings for the program have been tabulated in the following table:

|  |  |
| --- | --- |
| Number of Nodes | Execution Time (seconds) |
| 1 | 56.293 |
| 2 | 54.721 |

Thus we observe that, as the number of nodes increases the execution time required to run the same program reduces. The URLs are split into multiple nodes and this increases the efficiency as we can see based on our execution of PageRank.

**Experience using OpenStack:**

The experience of using OpenStack Cloud with FutureGrid systems was very positive and helped us learn a lot about how efficient Hadoop is on large clusters.  
There was a marked improvement in the performance of the WordCount and PageRank programs on OpenStack Cloud as compared to their execution on Oracle VM.  
All the instructions in the Lecture\_Openstack\_tutorial\_B649\_2015.ppt were helpful and enabled us to understand the individual steps involved in the execution very effectively.