

MPI & MAPREDUCE TUTORIALS

15-440 - Fall 2017

VIRUTAL MACHINES INFORMATION

We have prepared a 4-node virtual machine (VM) cluster for your 15-440 assignments this semester. The first node, which you will login to most often, is named:

your-andrewid-n01.qatar.cmu.local

The other 3 nodes are similarly named, replacing **-n01** with **-n02**, **-n03**, and **-n04**.

Each Linux VM has 4 vCPUs, 4GiB RAM, and a 40GB partition (mounted at /hadoop).

The initial passwords for the system are as follows:

Username: hadoop

Password: N/A

You can reach the Hadoop JobTracker web interface here:

your-andrewid-n01.qatar.cmu.local

Because these VMs are on CMU-Qatar's private network, they are not directly accessible from off-campus. If you want to connect from home, you will need to first connect to <https://vpnconnect.qatar.cmu.edu> before attempting to connect to the VMs. This will install *Cisco AnyConnect VPN* software and provide encrypted tunnel so you can reach private network resources.

MPI TEST: HELLO WORLD

- 1) SSH into the cluster using the information in the previous section (login to the **01 node**).
- 2) Copy the "HelloWorld.c" program from the webpage using the following command:

```
wget http://web2.qatar.cmu.edu/~mhhammou/15440-f17/recitations/HelloWorld.c
```
- 3) Copy the **machinefile** from the webpage using the following command:

```
wget http://web2.qatar.cmu.edu/~mhhammou/15440-f17/recitations/machinefile
```
- 4) Open the **machinefile** using **vim**:

```
vim machinefile
```
- 5) Edit the machine file so that it has the correct names for your cluster nodes. For example, if your Andrew ID is *ahmad*, replace

```
andrewid-n01.qatar.cmu.local:1
```

with:

```
ahmad-n01.qatar.cmu.local:1
```

Do so for the two nodes.
- 6) To save and exit, hit **escape**, then enter a **colon** (the ":" character), then type "**wq**"
- 7) Now, to compile your HelloWorld.c program, use the following command:

```
mpicc HelloWorld.c -o HelloWorld
```
- 8) For MPI to run, you must copy the **object file**, "**HelloWorld**" to the other machines. Simply run the following command to copy it to node 2 (we'll only be using two machines for this test):

```
scp -p "HelloWorld" andrewid-n02.qatar.cmu.local:/home/hadoop/
```
- 9) Finally, use the following command to run the MPI program:

```
mpiexec -f machinefile -n 2 ./HelloWorld
```

MAPREDUCE TEST: WORD COUNT

- 1) As before, log in to your **01 machine**.
- 2) Copy a “WordCount.sh” shell file from the webpage using the following command:

```
wget http://web2.qatar.cmu.edu/~mhhamou/15440-f17/recitations/WordCount.sh
```
- 3) Run the shell file downloaded using the following command:

```
sh WordCount.sh
```

This will set up things for you.
- 4) All the files are compiled and now ready for you. All you have to do is use the following command to run the program:

```
hadoop jar WordCount.jar WordCount /user/hadoop/wordcount/input  
/user/hadoop/wordcount/output
```
- 5) To view the output, run the following command:

```
hadoop dfs -text /user/hadoop/wordcount/output/part-00000
```
- 6) You can use the MapReduce Web UI to view information about your job:

```
http://your_andrewid-n01.qatar.cmu.local:50030/jobtracker.jsp
```