Getting Started

|  |  |
| --- | --- |
| **Java Info** | /home/[user-name]/workspace/[JobTitle]/src |

**Installation**

Install VMware, if you already have it, please update it now.

Next, copy the VM image onto your computer, somewhere safe.

Install VM image into VMware by using the Import option, under File.

Make sure you allocate the appropriate resources during setup, or just after.

Start up services by running the script in Terminal:

$ ./start\_services.sh

**Increase Memory in MapReduce and YARN in Ambari**

1. If you haven’t already, launch your internet browser within your virtual environment. Now, type localhost:8080 and hit Enter**.** This should take you promptly to your Ambari online platform. If not, please refer back to “Starting Up in Ambari.” NOTE: This will not work until services are started.
2. Once you have launched Ambari, login with “admin,” password “admin.”
3. Notice the list of services in the left-hand menu, and select YARN.
4. This should bring up a summary page. Look for and select “Configs,” next to “Heatmaps” and “Summary.”
5. Scroll down until you see the “Memory” section and locate the “Node” slider.

# Now, drag slider to 3840mb, save, and then look for the orange-yellow button labeled, “Restart” and click it.

7. Now, in the same left-hand menu you selected YARN from, select MapReduce2.

# 8. This should bring up a summary page. Look for and select “Configs,” next to “Summary.”

9. Then, in the MapReduce section look for the MaprReduce Framework and change the “Map Memory” slider to 1.5GB, save, and then look for the orange-yellow button labeled, “Restart” and click it.

1. Once Amabri states that all components have successfully restarted, you are finished.

# Using Firefox

# Launch Firefox from the terminal using the command: firefox

# Then, in your address bar type the following to access your course tools.

# Amber - localhost:8080

# YARN Tool - localhost:8888

Hue - localhost:8000

NOTE: Some services must be started before browsing to localhost.

**Eclipse**

Automatically import packages to satisfy errors using: CNTL + SHIFT + 0

You can find all your Eclipse projects in ~/workspace folder.

**Starting up with Ambari**

To start ambari services you need to first open terminal and then command”

ambari-server start

ambari-server stop

ambari-server status

**General Notes**

This training course utilizes a Virtual Machine running the CentOS Linux distribution. This VM has CDH Apache Hadoop installed in Pseudo-Distributed mode. Pseudo-Distributed mode is a method of running Hadoop whereby all Hadoop daemons run on the same machine. It is, essentially, a cluster consisting of a single machine. It works just like a larger Hadoop cluster, the only key difference being that the block replication factor is set to 1, since there is only a single DataNode available.

**Getting Started**

The VM is set to automatically log in as the [username]. Should you log out at any time, you can log back in as the user with the password.

**Working with the Virtual Machine**

Should you need it, the root password is hadoop. You may be prompted for this if, for example, you want to change the keyboard layout.

In some command‐line steps in the exercises, you will see lines like this:

$ hdfs dfs -put shakespeare

/user/[username]/shakespeare

The dollar sign ($) at the beginning of each line indicates the Linux shell

prompt. The actual prompt will include additional information (e.g.,

[[username]@localhost workspace]$ ) but this is omitted from

these instructions for brevity.

**Points to note during the exercises**

For most exercises, three folders are provided. Which you use will depend on how you would like to work on the exercises:

fixme: contains minimal skeleton code for the Java classes you’ll need to write. These are best for those with Java experience.

hints: contains Java class fixme that include additional hints about what’s required to complete the exercise. These are best for developers with limited Java experience.

solution: Fully implemented Java code which may be run “as-‐is”, or you may wish to compare your own solution to the examples provided.

As the exercises progress, and you gain more familiarity with Hadoop and MapReduce, we provide fewer step-‐by-‐step instructions; as in the real world, we merely give you a requirement and it’s up to you to solve the problem! You should feel free to refer to the hints or solutions provided, ask your instructor for assistance, or consult with your fellow students!

There are additional challenges for some of the Labs. If you finish the main exercise, please attempt the additional steps.

**END**