

LAB 1 INSTRUCTIONS

CKME 134 – BIG DATA ANALYTICS TOOLS

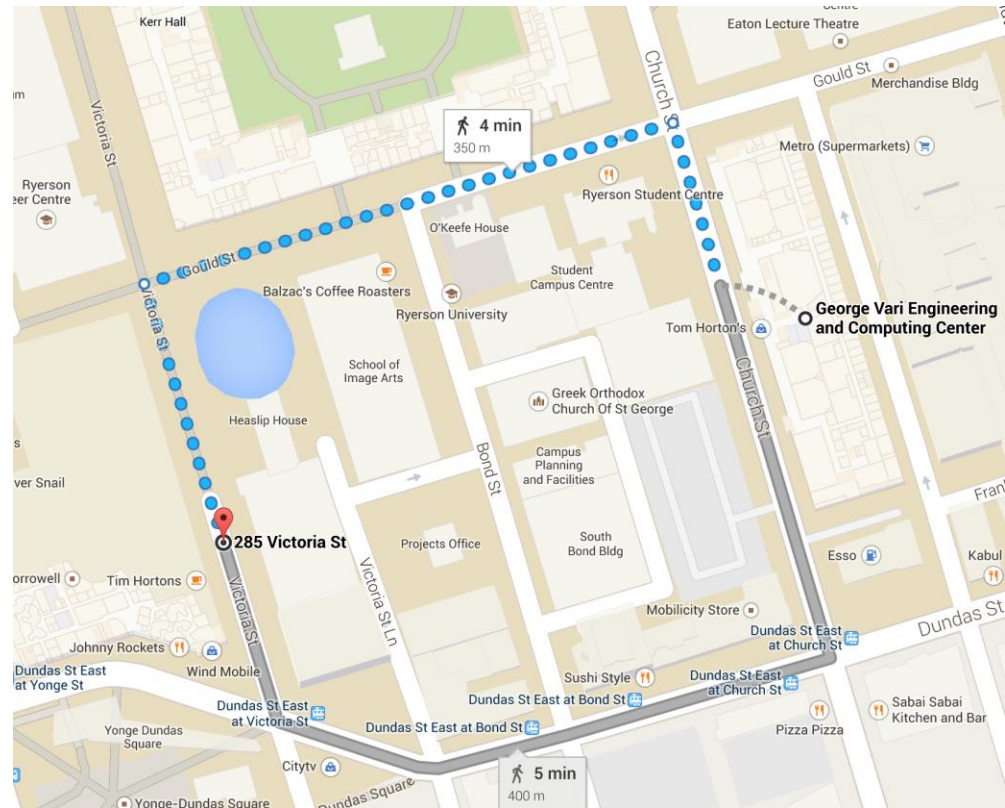
RYERSON UNIVERSITY

SPRING 2015

Instructor: Shaohua Zhang

General Course Information

- Instructor
 - ▣ Shaohua Zhang
 - ▣ Ryerson shaohua.zhang@ryerson.ca
 - ▣ Personal shaohua.zhang@live.com
- GA
 - ▣ Behjat Soltanifar
 - ▣ behjat.soltanifar@ryerson.ca
- Lectures
 - ▣ 6:30~8:30
 - ▣ ENGLG06
- Lab
 - ▣ 8:30~9:30
 - ▣ 285 Victoria St (403/404)
 - Take the elevator to 4FL



Lab & Assignments

- Lab Computer

- Username: datastudent
- Password: datastudent

- Lab 1

- Virtualbox
- Hadoop setup

- **Assignment 1**

- Pick a big data job title that fits your interests
- Find 3 companies (U.S. or Canada) that offer the job position you identified
- Read the job postings and summarize the skillsets required by the position
 - Send to Instructor's email: shaohua.zhang@ryerson.ca by putting "[Assignment1]" in the title



Lab 1 – Environment Setup

Lab 1 – Install VirtualBox

- Installation of VirtualBox

- ▣ <https://www.virtualbox.org/wiki/Downloads>

Our Lab Director Shannon has already downloaded and installed VirtualBox for us 😊

Lab 1 – Install Putty

- Windows users need to install Putty in order to use SSH to access hadoop
 - ▣ <http://www.chiark.greenend.org.uk/~sgtatham/putty/>

Our Lab Director Shannon has already downloaded and installed Putty for us 😊

Lab 1 – Install Hortonworks HDP Sandbox

❑ Download HDP hadoop vm

image: <http://hortonworks.com/products/hortonworks-sandbox/>

❑ Tutorial for Windows: <http://hortonworks.com/wp-content/uploads/unversioned/pdfs/InstallingHortonworksSandbox2onWindowsusingVB.pdf>

- The HDP VM Sandbox has been downloaded to C drive on windows.
- After installation and importing the instance, you start the vm and will see the screen on the right

```
tonworks Sandbox with HDP 2.2  
p://hortonworks.com
```

```
initiate your Hortonworks Sandbox session,  
ase open a browser and enter this address  
the browser's address field:  
p://127.0.0.1:8888/
```

```
can access SSH by $ ssh root@127.0.0.1 -p 2222
```

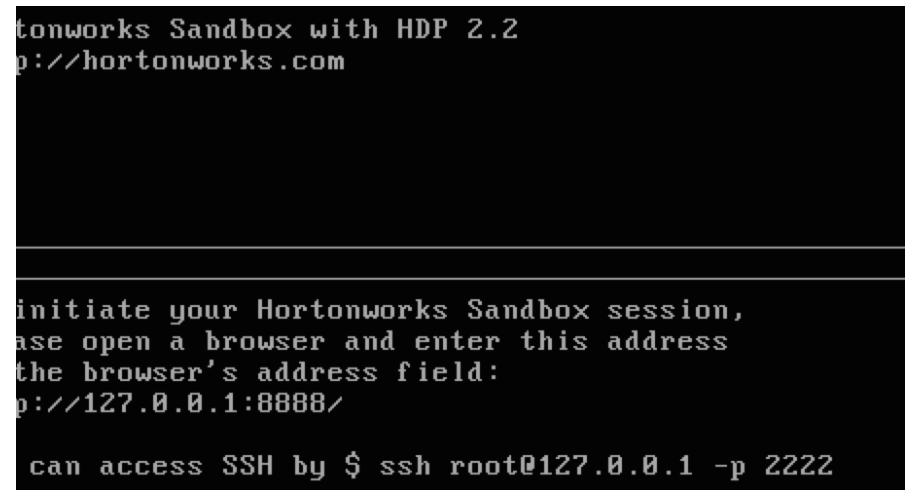
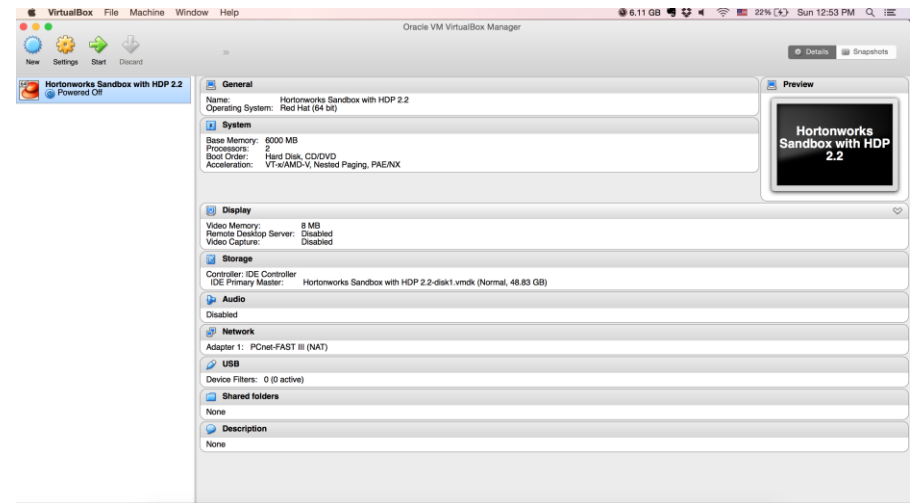
Lab 1 – Log On and Test Hive

1. Open VirtualBox
2. Start the HDP 2.2 VM
3. Open Putty and connect to 127.0.0.1 port 2222

- ▣ Username: root
- ▣ Password: hadoop

4. Test Hive

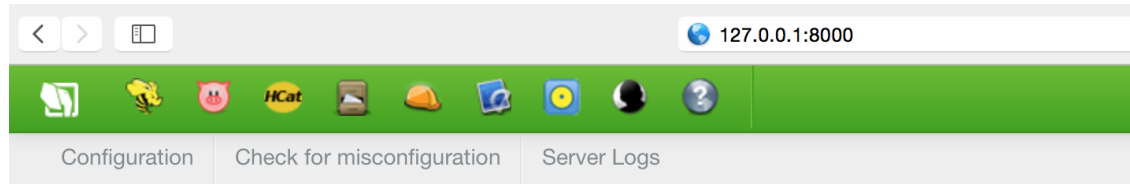
- ▣ \$ hive



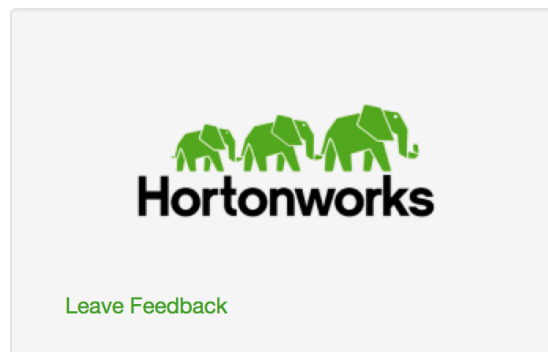
Lab 1 – Access Hadoop via Browser Using Hue

□ Access Hue via Browser

▣ <http://127.0.0.1:8000>



Hortonworks Sandbox with HDP 2.2



Component	Version	
Hue	2.6.1-2041	
HDP	2.2.0	
Hadoop	2.6.0	
Pig	0.14.0	
Hive-Hcatalog	0.14.0	
Oozie	4.1.0	
Ambari	1.7-169	<div>Enable</div>