Instructions for Downloading Hands On Datasets

This Reading describes the steps to download the datasets and scripts necessary for the Hands On exercises in this course. Before proceeding, make sure have you downloaded and installed VirtualBox and the Cloudera virtual machine. Instructions for this process can be found in the Reading called *Downloading and Installing the Cloudera VM*.

Step 1.**Start the Cloudera VM.** Most of the Hands On exercises in this course use the Cloudera Virtual Machine, so we will download the datasets onto the VM. Start the VM in VirtualBox and perform the remaining steps in the VM.

Step 2.**Open a web browser.** Open a web browser by clicking on the web browser icon in the top toolbar.

https://d3c33hcgiwev3.cloudfront.net/imageAssetProxy.v1/Zc2s0wpTEea72QqIV6G4Sw_c49276790fe7fd0f65defdbcb24b12e9_web-icon.png?expiry=1595289600000&hmac=PIgLlNusXStJ1kDoeL-XEkWFBpcjR1h0JXZj50vIJjg

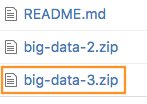
In the web browser, enter the following for the URL:



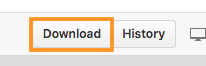
1

http://github.com/words-sdsc/coursera

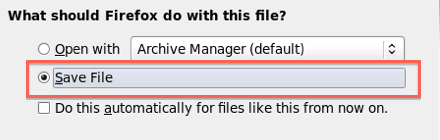
Step 3.**Download the datasets.** Click on *big-data-3.zip*:



Click on the *Download* button:



In the dialog, select *Save* *File*:



Click *OK*, and the file big-data-3.zip file will be downloaded to the Downloads directory.

Step 4. **Uncompress the datasets.** Open a terminal shell by clicking on the terminal shell icon in the top toolbar.



In the terminal, run:



1

2

cd Downloads

unzip -o big-data-3.zip

Step 5. **Install tools.** Change directories to *big-data-3* and run *setup.sh* to install tools and libraries.



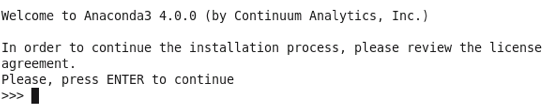
1

2

cd big-data-3

./setup.sh

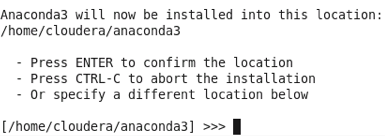
During the setup process, Anaconda will give you a series of prompts. First, press *enter* to continue the installation:



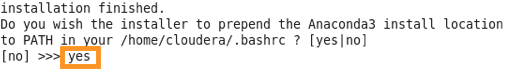
Next, read and accept the license:

https://d3c33hcgiwev3.cloudfront.net/imageAssetProxy.v1/RuD-OlXEEeaX4QpLJOK7gQ_725952c670089d2d9619ed65513b1dfa_anaconda-license.png?expiry=1595289600000&hmac=5ksX63As7om5KiiBFhGXiZhkDGEQKDDtyr9DA0D3Ixo

Next, press *enter* to accept the default installation location:



Next, enter *yes* when it asks if you want to prepend the install location to PATH:



The setup of tools and datasets should continue.

Finally,

source $HOME/.bashrc: