

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 you have 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*.

1. Start 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.

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

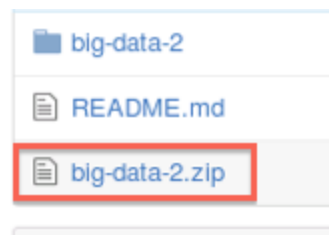


In the web browser, enter the following for the URL:

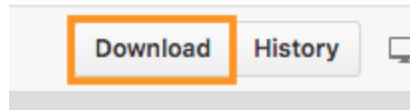
1

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

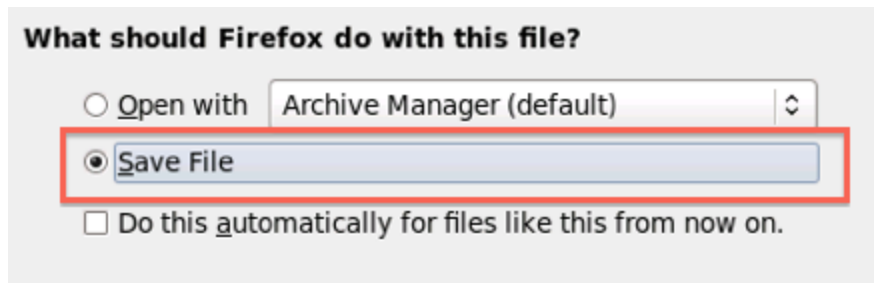
3. Download datasets. Click on *big-data-2.zip*:



Click on the *Download* button:



In the dialog, select *Save File*:



Click *OK*, and the file *big-data-2.zip* file will be downloaded to the Downloads directory.

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



In the terminal, run:

```
cd Downloads
unzip -o big-data-2.zip
```

5. **Install tools.** Finally, change directories to *big-data-2* and run *setup.sh* to install tools and libraries.

```
./setup.sh
cd big-data-2
```