

This Reading describes the steps to download the dataset and scripts necessary for the Spark 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. **Download data for KNIME.** Download the weather data to use with the KNIME hands on activities.

daily\_weather.csv

Step 2. **Start the Cloudera VM.** The Spark Hands On exercises in this course use the Cloudera Virtual Machine, so we will download the weather data onto the VM. Start the VM in VirtualBox and perform the remaining steps in the VM.

Step 3. **Open a 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>

Some browsers in the Cloudera VM may not be displaying the Generic code blocks. If you do not see a URL above, please enter the following URL:

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

Step 4. **Download the dataset.** Click on *big-data-4.zip*:

 [big-data-2.zip](#)

 [big-data-3.zip](#)

 [big-data-4.zip](#)

Click on the *Download* button:

