

[◀ Back to Week 1](#)[✕ Lessons](#)

This Course: Big Data Integration and Processing

[Prev](#)[Next](#)

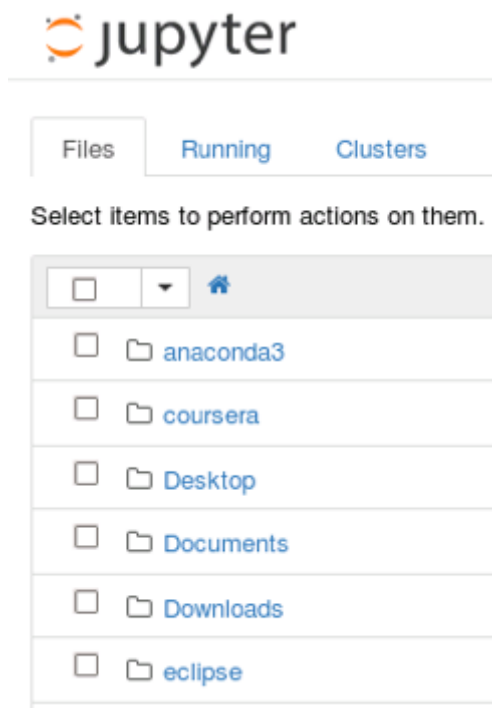
If you have not done so, please download the datasets referred to in the instructions in this link:

<https://www.coursera.org/learn/big-data-integration-processing/supplement/r8sXi/instructions-for-downloading-hands-on-datasets>

Many of the Hands On exercises in this course are executed in a Jupyter Notebook. To start Jupyter, open a new terminal window and run *pyspark*:

```
[cloudera@quickstart ~]$ pyspark
[I 11:53:34.330 NotebookApp] The port 8888 is already in use, trying another random port.
[I 11:53:34.347 NotebookApp] Serving notebooks from local directory: /home/cloudera
[I 11:53:34.347 NotebookApp] 0 active kernels
[I 11:53:34.348 NotebookApp] The Jupyter Notebook is running at: http://localhost:8889/
```

This will start Firefox (if not already running), and open a new tab showing the current directory:



You can find the notebooks for this course in the *Downloads/big-data-3* directory. Click on *Downloads* and then click on *big-data-3*.



Files **Running** Clusters

Select items to perform actions on them.

<input type="checkbox"/>	▼	🏠
<input type="checkbox"/>	📁	mongodb
<input type="checkbox"/>	📁	notebooks
<input type="checkbox"/>	📁	setup
<input type="checkbox"/>	📁	spark-sql
<input type="checkbox"/>	📁	spark-streaming

You can open Jupyter Notebooks by clicking on the `.ipynb` files.

When you are done, Jupyter can be stopped by entering `control-c` in the terminal window and then entering `yes`.

Mark as completed

