



## Programming in Spark


## Main Modules in the Spark Ecosystem


## Hands-on: Data Processing in Spark

 **Reading:** Exploring SparkSQL and Spark DataFrames  
20 min

 **Video:** Exploring SparkSQL and Spark DataFrames  
6 min

 **Reading:** Instructions for Configuring VirtualBox for Spark Streaming  
10 min

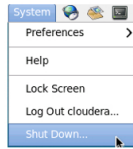
 **Reading:** Analyzing Sensor Data with Spark Streaming  
30 min

 **Video:** Analyzing Sensor Data with Spark Streaming  
7 min

✔ **Quiz: SparkSQL and Spark Streaming**  
4 questions

Spark Streaming requires more than one executor. This Reading describes how to configure VirtualBox so that the Cloudera VM has more than one virtual processor.

Step 1. **Stop the Cloudera VM.** Before we can change the settings for the Cloudera VM, the VM needs to be powered off. If the VM is running, click on *System* in the top toolbar, and then click on *Shutdown*:



Next, click on *Shut down*:



**Step 2. Change number of processors.** Once the Cloudera VM is powered off, select the Cloudera VM in the list of virtual machines in the VirtualBox Manager:



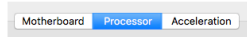
Next, click on *Settings*:



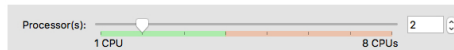
Next, click on *System*:



Next, click on *Processor*:



The default number of processors is one. Change this to two or more:



Finally, click on *OK* and start the Cloudera VM.

Mark as completed