I think this video is pretty good. Presenter Jim Scott does not provide a tutorial, explain how to install Spark, or tell you what to do with it in the real world. He does explain how Spark fits into other data science tools. For me it helped cleared up Hive vs Pig vs Mahout vs Spark vs HDFS vs R vs Python vs Java, vs etc.

<https://www.youtube.com/watch?v=mL5dQ_1gkiA>

Thanks Paul.  This is good.  The MapR folks make me laugh,,, I've been to a number of "intro" courses by them and they never present the information as clear as it should be.  I cant tell if that means they dont really understand the material or they just assume everyone understands them.  Either way, I default to the Einstein quote "If you can't explain it simply, you don't understand it well enough"  The cloudera guy below understands it well enough and presents it very simply.

This one is also good for an introduction

<https://www.youtube.com/watch?v=rvDpBTV89AM>

And I like the Cloudera overview to get some context at a high level

<https://www.youtube.com/watch?v=SxAxAhn-BDU>

[hadoop@spark ~]$ wget <http://apache.arvixe.com/spark/spark-1.5.0/spark-1.5.0.tgz>

[hadoop@spark ~]$ tar xzvf spark-1.5.0.tgz

[hadoop@spark ~]$ nano .bashrc

export SPARK\_HOME=/home/hadoop/spark-1.5.0

export PATH=$PATH:$SPARK\_HOME/bin

[hadoop@spark ~]$ source .bashrc

[hadoop@spark ~]$ cd spark-1.5.0

[hadoop@spark spark-1.5.0]$ build/mvn -Pyarn -Phadoop-2.4 -Dhadoop.version=2.4.0 -DskipTests clean package

... a few thousand lines later ...

[INFO] BUILD SUCCESS

[INFO] ------------------------------------------------------------------------

[INFO] Total time: 19:09 min

[INFO] Finished at: 2015-09-26T16:25:27-06:00

[INFO] Final Memory: 456M/1640M

[INFO] ------------------------------------------------------------------------

[hadoop@spark spark-1.5.0]$

[hadoop@spark spark-1.5.0]$ spark-shell

Welcome to

      \_\_\_\_              \_\_

     / \_\_/\_\_  \_\_\_ \_\_\_\_\_/ /\_\_

    \_\ \/ \_ \/ \_ `/ \_\_/ ' \_/

   /\_\_\_/ .\_\_/\\_,\_/\_/ /\_/\\_\   version 1.5.0

      /\_/

Using Scala version 2.10.4 (Java HotSpot(TM) 64-Bit Server VM, Java 1.8.0\_60)

Type in expressions to have them evaluated.

Type :help for more information.

15/09/26 16:31:31 WARN MetricsSystem: Using default name DAGScheduler for source because spark.app.id is not set.

Spark context available as sc.

SQL context available as sqlContext.

scala> exit

[hadoop@spark spark-1.5.0]$