30th November 2014   How to Install Apache Spark on Mac OS X Yosemite
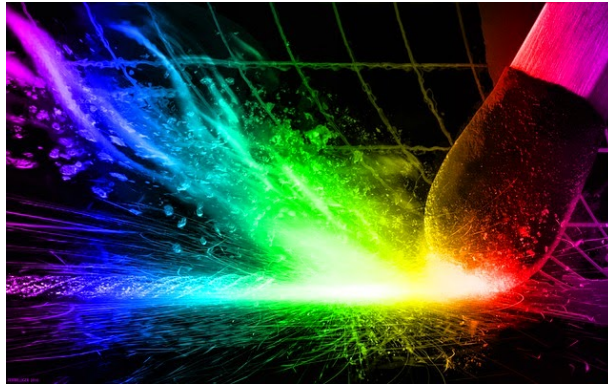

[http://4.bp.blogspot.com/-oGB2rkoQmqY/VHqUfh1rhcI/AAAAAAAAHyQ/tcpOjbPSr_k/s1600/match-89578_640.jpg]

Apache Spark™ is a fast and general engine for large-scale data processing.

- **Install Java**

  - Download Oracle Java SE Development Kit 7 or 8 at Oracle JDK downloads page [http://www.oracle.com/technetwork/java/javase/downloads/jdk7-downloads-1880260.html] .
  - Double click on .dmg file to start the installation
  - Open up the terminal.
  - Type *java -version*, should display the following

  java version "1.7.0_71"
  Java(TM) SE Runtime Environment (build 1.7.0_71-b14)
  Java HotSpot(TM) 64-Bit Server VM (build 24.71-b01, mixed mode)

- **Set JAVA_HOME**

  export JAVA_HOME=$(/usr/libexec/java_home)

- **Install Homebrew**

  ruby -e "$(curl -fsSL https://raw.githubusercontent.com/Homebrew/install/master/install [https://raw.githubusercontent.com/Homebrew/install/master/install])"

- **Install Scala**

brew install scala

- **Set SCALA_HOME**

  export SCALA_HOME=/usr/local/bin/scala
  export PATH=$PATH:$SCALA_HOME/bin

- **Download Spark from https://spark.apache.org/downloads.html**

  tar -xvzf spark-1.1.1.tar
  cd spark-1.1.1

- **Build and Install Apache Spark**

  sbt/sbt clean assembly

- **Fire up the Spark**

  For the Scala shell:
  ./bin/spark-shell

  For the Python shell:
  ./bin/pyspark

- **Run Examples**

  Calculate Pi:

  ./bin/run-example org.apache.spark.examples.SparkPi

  MLlib Correlations example:

  ./bin/run-example org.apache.spark.examples.mllib.Correlations

  MLlib Linear Regression example:

  ./bin/spark-submit
  --class org.apache.spark.examples.mllib.LinearRegression
  examples/target/scala-/spark-*.jar data/mllib/sample_linear_regression_data.txt*

References:

How to install Spark on Mac OS X

http://ondrej-kvasnovsky.blogspot.com/2014/06/how-to-install-spark-on-mac-os-x.html   [http://ondrej-
kvasnovsky.blogspot.com/2014/06/how-to-install-spark-on-mac-os-x.html]

How To Set $JAVA_HOME Environment Variable On Mac OS X
http://www.mkyong.com/java/how-to-set-java_home-environment-variable-on-mac-os-x/
[http://www.mkyong.com/java/how-to-set-java_home-environment-variable-on-mac-os-x/]

Homebrew - The missing package manager for OS X
http://brew.sh [http://brew.sh/]

Posted 30th November 2014 by Sudhir Chowbina

Labels: Apache Spark, Mac OS X, machine learning, MLlib, Python, Scala

View comments