Apache Spark

http://spark.apache.org/downloads.html

Java

http://www.java.com/en/download/manual.jsp

You can find older versions here: http://www.oracle.com/technetwork/java/javase/arch...

Python

https://www.continuum.io/downloads

Sample Data

1) Download NY State Baby Names in CSV format from:

http://www.healthdata.gov/dataset/baby-names-begin...

(I renamed the csv file to baby_names.csv to make it more convenient to call from Spark drivers)

2) Uber NYC data

For background information
https://github.com/fivethirtyeight/uber-tlc-foil-response

For raw CSV download
https://raw.githubusercontent.com/fivethirtyeight/uber-tlc-foil-response/master/Uber-Jan-Feb-FOIL.csv