Step 1: Download & Untar SPARK

Download the version 1.0.2 of spark

Untar the downloaded file to any location (say C:\spark-1.0.2)

Step 2: Download SBT msi (needed for Windows)

Download sbt.MSI & execute it.
You may need to restart the machine so that command line can identify the sbt command
Step 3: Package Spark using SBT

C:\spark-1.0.2>sbt assembly

Note: This step takes enormous amount of time. Please be patient

Step 4: Download SCALA

Spark 1.0.2 needs Scala 2.10. This is extremely important to note. And you can read the README.MD file in the SPARK folder to find the correct scala version needed for your spark.

Download and unzip the scala to any location (say C:\ scala-2.10.1)

Set SCALA_HOME environment variable & set the PATH variable to the bin directory of scala

Verify the scala version (and thus the download)

Step 5: Start the spark shell

C:\spark-1.0.2\bin>spark-shell
Sample program in SPARK

1) Create a data set of 1...10000 integers

```
scala> val data = 1 to 10000
```

2) Use Spark Context to create an RDD [Resilient Distributed Dataset] from that data

```
scala> val distData = sc.parallelize(data)
```

3) Perform a filter mechanism on that data

```
scala> distData.filter(_ < 10).collect()
```