
Big Data Essentials Course Preparation Guide

Computer

- You will need to bring a laptop that has an SSH client, and (optionally) a development environment setup for the class

SSH Installation

- An SSH client software should be installed to connect to the training VM from your laptop
 - On Linux or Mac OS X machines, SSH client is built-in. To check, type (on a terminal)


```
$ ssh
```

 - If this shows some usage help and exits, SSH is installed
 - otherwise, use your package manager to download SSH
 - On Windows machines, popular SSH clients such as Putty and WinSCP, available at <http://www.putty.org/> and <http://winscp.net> respectively, can be installed and used

Setting up a Development Environment

For performing

- Apache Hadoop HDFS
- Apache Pig
- Apache Hive
- Apache Sqoop
- Apache Flume

exercises, you will **not** require any special software. These exercises would be performed on the training machines you will log in (via SSH). You will be able to write Pig and Hive scripts/queries on the Hue web interface, too.

Optional Development Environment

The default implementations we will provide to you are hosted in **github**, meaning that you need to have **git** installed on your computer. Git can be downloaded from <http://git-scm.org>

The example programs are **Apache Maven** projects, so a developer will typically need Apache Maven as well.

Apache Maven can be downloaded from <http://maven.apache.org/download.cgi>, and the installation instructions are available here:

<http://maven.apache.org/download.cgi#Installation>