Apache Hadoop

Functional Area: Parallel Horizontally Scalable Data Processing:Batch

Overview:

Hadoop uses Java

Tutorial:

Hadoop Installation Instructions

Ubuntu 14.04 latest version on 64 bit

Software Requirements: Java Development Kit  
 Linux OS or Ubuntu 14.04 on Virtual Box  
 Windows- Cygwin

Installation Tutorial

1. Install Java from Oracle
2. Download a stable release of Hadoop from <http://www.apache.org/dyn/closer.cgi/hadoop/common/>
3. Uncompress the archive:  
    tar xvzf hadoop-2.4.1-bin.tar.gz
4. Edit the file conf/hadoop-env.sh and add :  
    export JAVA\_HOME=<JDK DIRECTORY>
5. Export/set the Hadoop path in ~/.bashrc  
    export HADOPP\_INSTALL=/path/to your installation  
    PATH=$PATH:$HADOOP\_INSTALL/bin:$HADOOP\_INSTALL/sbin
6. Install ssh server  
    sudo –apt get install openssh-server
7. Generate ssh key and authorize the keys  
    ssh-keygen –t rsa –P ‘ ’  
    cat ~/.ssh/id\_rsa.pub > ~/.ssh/authorized\_keys