

Installation of Hadoop

Sang Shin

JPassion.com

“Learn with Passion!”



Topics

- Hadoop operational modes
- Hadoop installation options

Hadoop Operational Modes

Hadoop Operational Modes

- Local (Standalone) operation
 - > All Hadoop daemons (NameNode, DataNode, JobTracker, Task Tracker) are running on a single Java process (JVM)
 - > Uses local file system for storage
- Pseudo-distributed operation (test & development)
 - > Each Hadoop daemon runs in a separate Java process (JVM) but physically on a single machine
- Fully-distributed operation (production)
 - > Each Hadoop daemon runs on a different machine
 - > The power of Hadoop – parallel processing, scalability, automatic replication management, fault-tolerance, data consistency, workflow management – gets realized only in this mode

Hadoop Installation Options

Installation Options of Hadoop

- VM Image
 - > All Hadoop technologies pre-installed and pre-configured via VM for out of the box Hadoop experience
 - > Typically single node cluster configuration
 - > Pseudo-distributed operational mode
 - > Two options: Cloudera QuickStart VM, Hortonworks Sandbox
 - > Easiest way to get started with Hadoop
- Hadoop package
 - > A single package that contains most Hadoop technologies
- Manual installation
 - > You download each Hadoop technology individually and do manual installation and configuration for each

Hadoop Distributions

- Apache Hadoop
 - > <http://hadoop.apache.org/>
- Cloudera
 - > <http://cloudera.com/>
 - > The most popular distribution
 - > Integrated with many popular Hadoop ecosystem technologies
- Hortonworks
 - > <http://hortonworks.com>
- MapR
 - > <http://mapr.com>

Cloudera QuickStart VM

- CDH 5.3.x
 - > Cloudera Manager does not get started automatically
 - > Requires more memory for Cloudera Manager
 - > Adds Hadoop Spark support
- CDH 4.7.x
 - > Cloudera Manager does get started automatically
 - > We are going to use CDH 4.7.x for our lab – please be assured there will not any negative learning experience!

Lab:

Exercise 1, 2, 3
5901_hadoop_installation.zip



Learn with Passion!
JPassion.com

