Hadoop Installation

Pre Requisites: Java (jdk)
Hadoop (any version 2.x / 3.x)

⇒ Hadoop in its 3 Operating Modes.

- Local/Standalone (Single Monolithic Java process)
- Pseudo Distributed (Single Machine with Distributed Env")
- Fully Distributed (not possible in Windows)

OS: Linux, Ubuntu, Windows

STEP 1: Download JDK 8/11 (8 is preferable)
Install it, set {Path & Variable name} (Environment Variable)

Ex: C:\Program Files\Java\jdk 1.8.x

User Variables:
new > JAVA-HOME
paste Java path

path, new, paste Java

System Variables:
Path > edit > new
paste Java path.

STEP 2: Download Hadoop for Windows (From Manual link (or) apache hadoop (Google))
Download → 3.2.4 → binary → link tar file

Extract the Downloaded Hadoop folder

It consists Hadoop 2.6.x → bin, etc – hadoop, include, lib

STEP 3: hadop 2.6.x > etc > hadoop > hadoop-env ⇒ Edit
On that file

```
@echo off
.
.
set JAVA-HOME = " paste the Java-Jdk Path "   (Save it)
.                Ex: c:\Java\jdk1.7.0        (don't include bin in path)
.
```

Env Var Set the new variable as HADOOP-HOME
C:\hadoop\bin

User Variable

Add path ⇒ Edit ⇒ Add this path here > ok.
Add 2 paths set ⇒ C:\hadoop\bin

System Variable.
C:\hadoop\sbin

STEP 4: Verify Java & Hadoop are installed/not

cmd ⇒ C:\Users\kela> javac ↵

C:\Users\kela> hadoop ↵

C:\Users\kela> hadoop version ↵

Successfully installed.

STEP 5: Configure Hadoop (XML files)

Goto C:\hadoop\etc\hadoop\

Edit
1. core-site.xml
2. hdfs-site.xml
3. mapred-site.xml
4. yarn-site.xml

1. core-site.xml

```
<configuration>
 <property>
  <name>fs.defaultFS</name>
  <value>hdfs://localhost:9000</value>
 </property>
</configuration>
```

2. hdfs-site.xml

```
<configuration>
 <property>
  <name>dfs.replication</name>
  <value>1</value>
 </property>
 <property>
  <name>dfs.namenode.name.dir</name>
  <value>C:\hadoop\data\namenode</value>
 </property>
 <property>
  <name>dfs.datanode.data.dir</name>
  <value>C:\hadoop\data\datanode</value>
 </property>
</configuration>
```

In parallel we do

Goto C:\hadoop ↵

Create 1 folder as data

click on it

Create 2 folder as

namenode @

datanode @@

Copy the paths in xml files

3. mapred-site.xml

```
<configuration>
 <property>
  <name>mapreduce.framework.name</name>
  <value>yarn</value>
 </property>
</configuration>
```

④ yarn-site.xml

```
<configuration>
  <property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
  </property>
</configuration>
```

```
<property>
<name>yarn.nodemanager.auxservices.mapreduce.class
</name>
<value>org.apache.hadoop.mapred.ShuffleHandler
</value>
</property>
```

STEP6: Replace bin folder.

Goto c:\hadoop ⇒ delete bin folder

download winutils.exe from Google/github & copy that exe file in bin folder of hadoop.

Step7: Setup done, check at cmd prompt

```
C:\Users\User> hdfs namenode -format
```

It starts the process.

Step8: Commands.

```
C:\Users\User> cd \
C:\> cd hadoop
C:\hadoop> cd sbin
C:\hadoop\sbin> start-all.cmd
```

It opens multiple windows in parallel.

———— o ————

Start Hadoop Services

(1) Start HDFS

cmd> start-dfs.cmd

Name & Data Nodes started

(2) Start YARN

cmd> start-yarn.cmd

Starts Resource & Node Managers.

(3) Check running daemons

cmd> jps

We can see NameNode, DataNode, SecondaryNode, Node & Resource Managers.

(4) Check Web Interfaces.

URL localhost:8088/cluster

(5) HDFS basic Commands.

Create Dir in HDFS

cmd:\> hdfs dfs -mkdir /user

.. .. /admin.

Copy file from local system to HDFS