

1. Ensure your system has java.  
Java 8 is present in your system.
2. <https://archive.apache.org/dist/hadoop/common/hadoop-3.2.0/hadoop-3.2.0.tar.gz>
3. move to your desired location with good memory space
4. now extract with 7zip
5. setup system variables
  - HADOOP\_HOME
  - PATH (both for bin and sbin)
6. Modify the file in etc folder.
  - core-site
  - mapred-site
  - hdfs-site
  - yarn-site
  - hadoop-env
7. core-site

```
<property>
  <name>fs.defaultFS</name>
  <value>hdfs://localhost:9000</value>
</property>
```
8. mapred-site  
Make copy of mapred-site.xml.template and rename to mapred-site.xml

```
<property>
  <name>mapreduce.framework.name</name>
  <value>yarn</value>
</property>
```
9. Create a "data" folder in hadoop directory inside data folder create datanode (any files pushed in data node will be stored as blocks) and namenode (will be used by namenode for fsimages and editlogs)
  - data
    - datanode
    - namenode
10. hdfs-site

```
<property>
  <name>dfs.replication</name>
  <value>1</value>
</property>
<property>
  <name>dfs.namenode.name.dir</name>
  <value>F:\Software\Apache\Hadoop\hadoop-3.2.0\data\namenode</value>
</property>
<property>
  <name>dfs.datanode.data.dir</name>
  <value>F:\Software\Apache\Hadoop\hadoop-3.2.0\data\datanode</value>
</property>
```
11. yarn-site

```
<property>
  <name>yarn.nodemanager.aux-services</name>
  <value>mapreduce_shuffle</value>
</property>
<property>
  <name>yarn.nodemanager.auxservices.mapreduce.shuffle.class</name>
  <value>org.apache.hadoop.mapred.ShuffleHandler</value>
</property>
```
12. hadoop-env  
set JAVA\_HOME=D:\InstalledPrograms\Java

13. below 3.2.xx  
Hadoop needs windows OS specific files which does not come with default download of hadoop  
To include those files, replace the bin folder in hadoop directory with the bin folder provided in the below link.  
download the bin file based on your hadoop version.  
  
above 3.2.xx  
copy the files in the download from the bin and place it in hadoop bin folder  
  
<https://github.com/cdarlint/winutils>
14. check hadoop installation in command prompt  
hadoop version
15. Format the NameNode in command prompt  
hdfs namenode -format
16. now in command port start all the hadoop services (start-all.cmd which in sbin)  
  
start-all (or) start-dfs.cmd , start-yarn.cmd
17. now check the resource manager , namenode and data node  
  
-current jobs, successful and failed jobs, go to this link in browser-  
<http://localhost:8088/cluster>  
  
name node and data node details  
<http://localhost:9870/>
18. checking hadoop working properly .  
create as samplefile in windows pc which need to be pushed to hadoop  
sample.txt  
F:\Software\Apache\Hadoop\temp\sample.txt  
  
hadoop fs -mkdir /TestFolder  
  
hadoop fs F:\Software\Apache\Hadoop\temp\sample.txt /TestFolder  
  
check the file is added to hdfs  
  
hadoop fs -ls /TestFolder/\*  
  
to see data  
  
hadoop dfs -cat /TestFolder/sample.txt
19. for more details on Hadoop commands  
<https://hadoop.apache.org/docs/r3.1.0/hadoop-project-dist/hadoop-hdfs/HDFSCommands.html>