

INSTALLATION OF HADOOP

INSTALLATION STEPS:

- 1.INSTALL JAVA
- 2.INSTALL AND UNZIP HADOOP
- 3.SETTING UP ENVIRONMENT VARIABLES

User variables for shash

Variable	Value
FLUME_CONF	D:\Shashank\Study\Flume\conf
FLUME_HOME	D:\Shashank\Study\Flume
HADOOP_HOME	D:\Shashank\Study\hadoop-2.9.2
HIVE_HOME	D:\Shashank\Study\Hive-3.1.2
JAVA_HOME	C:\Program Files\Java\jdk1.8.0_221
OneDrive	C:\Users\shash\OneDrive

System variables

Variable	Value
Path	C:\Windows\System32;D:\Softwares\Python\Scripts\;D:\Softw...
PATHEXT	.COM;.EXE;.BAT;.CMD;.VBS;.VBE;.JS;.JSE;.WSF;.WSH;.MSC;.PY;.PYW
platformcode	KV
PROCESSOR_ARCHITECTU...	AMD64
PROCESSOR_IDENTIFIER	Intel64 Family 6 Model 142 Stepping 12, GenuineIntel
PROCESSOR_LEVEL	6
PROCESSOR_REVISION	8~0~

4.EDITING CONFIGURATION FILES Core-site.xml

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

Hdfs-site.xml

```
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>C:\hadoop-3.3.6\data\namenode</value>
  </property>
  <property>
    <name>dfs.datanode.data.dir</name>
    <value>C:\hadoop-3.3.6\data\datanode</value>
  </property>
</configuration>
```

Mapred-site.xml

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

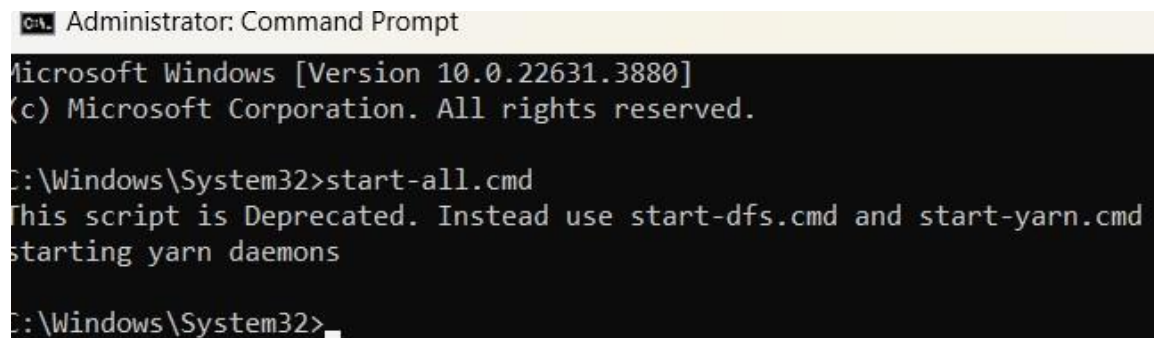
Yarn-site.xml

```
<configuration>
  <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>
</configuration>
```

5.REPLACE BIN FOLDER**6.FORMAT NAMENODE AND DATANODE**

hadoop namenode -format

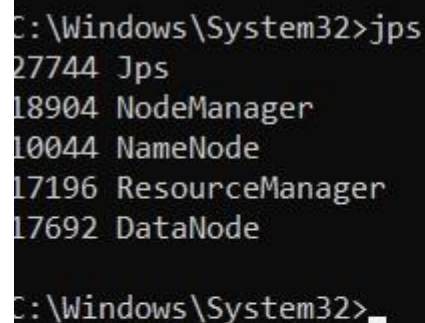
hadoop datanode -format

7.START HADOOP

```
Administrator: Command Prompt
Microsoft Windows [Version 10.0.22631.3880]
(c) Microsoft Corporation. All rights reserved.

C:\Windows\System32>start-all.cmd
This script is Deprecated. Instead use start-dfs.cmd and start-yarn.cmd
starting yarn daemons

C:\Windows\System32>_
```




```
C:\Windows\System32>jps
27744 Jps
18904 NodeManager
10044 NameNode
17196 ResourceManager
17692 DataNode

C:\Windows\System32>_
```

8.RUNNING HADOOP

localhost:8088/cluster



Cluster

- About
- Nodes
- Node Labels
- Applications
 - NEW
 - NEW SAVING
 - SUBMITTED
 - ACCEPTED
 - RUNNING
 - FINISHED
 - FAILED
 - KILLED
- Scheduler

Tools

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running
0	0	0	0	<m

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned N
1	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Alloc
Capacity Scheduler	[memory-mb (unit=Mi), vcores]	<memory:1024, vCores:1>

Show 20 entries

ID	User	Name	Application Type	Application Tags	Queue	Application Priority	StartTime	LaunchTime	FinishTime
----	------	------	------------------	------------------	-------	----------------------	-----------	------------	------------

Showing 0 to 0 of 0 entries

Namenode information

localhost:9870/dfshealth.html#tab-overview

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

Overview 'localhost:9000' (✓active)

Started:	Tue Aug 13 07:54:43 +0530 2024
Version:	3.3.6, r1be78238728da9266a4f88195058f08fd012bf9c
Compiled:	Sun Jun 18 13:52:00 +0530 2023 by ubuntu from (HEAD detached at release-3.3.6-RC1)
Cluster ID:	CID-9ff10afd-9f2a-4744-9bc6-f45b2553ad17
Block Pool ID:	BP-468074218-172.16.11.91-1722913037250