

EXP.1: Downloading and installing Hadoop, Understanding different Hadoop modes, Startup scripts, Configuration files.

AIM:

To Download and install Hadoop, Understanding different Hadoop modes, Startup scripts, Configuration files.

PROCEDURE:

1. Install Java 8: Download Java 8 from the link:

<http://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html>

- a. Set environmental variables:
 - i. User variable:
 - Variable: JAVA_HOME
 - Value: C:/java
 - ii. System variable:
 - Variable: PATH
 - Value: C:/java/bin
- b. Check on cmd, see below:



```
Command Prompt
Microsoft Windows [Version 10.0.19045.4780]
(c) Microsoft Corporation. All rights reserved.

C:\Users\user>cd \

C:\>java -version
java version "1.8.0_202"
Java(TM) SE Runtime Environment (build 1.8.0_202-b08)
Java HotSpot(TM) 64-Bit Server VM (build 25.202-b08, mixed mode)

C:\>
```

2. Download Hadoop-3.3.6: download Hadoop 3.3.6 from the link:

<http://www.apache.org/dyn/closer.cgi/hadoop/common/hadoop-3.3.6/hadoop-3.3.6.tar.gz>

- a. Put extracted Hadoop-3.3.6 files into C drive. Note that do not put these extracted files into Cdrive, where you installed your Windows.
- b. **Download “hadoop-common-3.3.6-bin-master”** from the link: <https://github.com/amihalik/hadoop-common-3.3.6-bin/tree/master/bin>. You will see 11 filesthere. Paste all these files into the “bin” folder of Hadoop-3.3.6.
- c. Create a “data” folder inside Hadoop-3.3.6, and also create two more folders in the “data” folderas “data” and “name.”
- d. Create a folder to store temporary data during execution of a project, such as “D:/hadoop/temp.”
- e. Create a log folder, such as “D:/hadoop/userlog”
- f. Go to Hadoop-3.3.6 →etc → Hadoop and edit four files:
 - i. core-site.xml
 - ii. hdfs-site.xml
 - iii. mapred-site.xml

iv. yarn-site.xml

core-site.xml:

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<configuration>
<property>
  <name>fs.defaultFS</name>
  <value>hdfs://localhost:9000</value>
</property>
</configuration>
```

hdfs-site.xml:

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>file:///C:/hadoop-3.3.6/data/namenode</value>
  </property>
  <property>
    <name>dfs.datanode.data.dir</name>
    <value>file:///C:/hadoop-3.3.6/data/datanode</value>
  </property>
</configuration>
```

mapred-site.xml:

```
<?xml version="1.0"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<configuration>
<property>
  <name>mapreduce.framework.name</name>
  <value>yarn</value>
</property>
</configuration>
```

yarn-site.xml:

```
<?xml version="1.0"?>
<configuration>
  <property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
  </property>
</configuration>
```

- g. Go to the location: “Hadoop-3.3.6→etc→hadoop,” and edit
“hadoop-env.cmd” by writing
set JAVA_HOME=C:/java/jdk1.8.0_202
- h. Set environmental variables: Do: My computer → Properties → Advance
system settings →
Advanced → Environmental variables
 - i. User variables:
 - Variable: HADOOP_HOME
 - Value: C:/hadoop-3.3.6
 - ii. System variable
 - Variable: Path
 - Value: C:/hadoop-3.3.6/bin
C:/hadoop-3.3.6/sbin
C:/hadoop-3.3.6/share/hadoop/common/*
C:/hadoop-3.3.6/share/hadoop/hdfs
C:/hadoop-3.3.6/share/hadoop/hdfs/lib/*
C:/hadoop-3.3.6/share/hadoop/hdfs/*
C:/hadoop-3.3.6/share/hadoop/yarn/lib/*
C:/hadoop-3.3.6/share/hadoop/yarn/*
C:/hadoop-3.3.6/share/hadoop/mapreduce/lib/*
C:/hadoop-3.3.6/share/hadoop/mapreduce/*
C:/hadoop-3.3.6/share/hadoop/common/lib/*
- i. Check on cmd; see below.

```
C:\Windows\System32>hadoop version
Hadoop 3.3.6
Source code repository https://github.com/apache/hadoop.git -r 1be78238728da9266a4f88195058f08fd012bf9c
Compiled by ubuntu on 2023-06-18T08:22Z
Compiled on platform linux-x86_64
Compiled with protoc 3.7.1
From source with checksum 5652179ad55f76cb287d9c633bb53bbd
This command was run using /C:/hadoop-3.3.6/share/hadoop/common/hadoop-common-3.3.6.jar
```

- j. **Format name-node:** On cmd go to the location “Hadoop-2.6.0→bin” by
writing on cmd “cdhadoop-2.6.0\bin” and then “hdfs namenode –format”
- k. Start Hadoop. Go to the location: “D:\hadoop-2.6.0\sbin.” Run the following

- [illegible]

```
C:\> Administrator: Command Prompt
Microsoft Windows [Version 10.0.19045.4780]
(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>jps
11104 Jps
12868 DataNode
11288 ResourceManager
12456 NodeManager
5596 NameNode

C:\WINDOWS\system32>
```



```
16/05/22 12:42:08 INFO delegation.AbstractDelegat...
delegation tokens
16/05/22 12:42:08 INFO delegation.AbstractDelegat...
delegation tokens
16/05/22 12:42:08 INFO delegation.AbstractDelegat...
delegation tokens
16/05/22 12:42:09 INFO tokenRemover$scanInterval=60 min(s)
16/05/22 12:42:09 INFO Started HttpServer: 127.0.0.1:8042
16/05/22 12:42:09 INFO webapp.WebApps: Web app [/c...
16/05/22 12:42:09 INFO delegation.AbstractDelegat...
16/05/22 12:42:09 INFO tokenRemover$scanInterval=60 min(s)
16/05/22 12:42:09 INFO delegation.AbstractDelegat...
delegation tokens
16/05/22 12:42:09 INFO delegation.AbstractDelegat...
delegation tokens
16/05/22 12:42:11 INFO webapp.WebApps: Registered
16/05/22 12:42:11 INFO pb.RpcServerFactory$PBImpl: ...
16/05/22 12:42:11 INFO IPC Server: IPC Server: Starting Socket
16/05/22 12:42:11 INFO IPC Server: IPC Server: Starting Socket
16/05/22 12:42:11 INFO IPC Server: IPC Server: Starting Socket
16/05/22 12:42:11 INFO util.RackResolver: Resolved
16/05/22 12:42:12 INFO resource.ResourceTracker
16/05/22 12:42:12 INFO rmmode.RMNodeImpl: 132.72.2...
16/05/22 12:42:12 INFO capacity.CapacityScheduler:
8>

16/05/22 12:42:09 INFO http.HttpServerLog: http request log for http.requests.nodemanager is not defined
16/05/22 12:42:09 INFO http.HttpServer2: Added global filter 'safety' (class=org.apache.hadoop.http.HttpServer2$QuotingI
nputFilter)
16/05/22 12:42:09 INFO http.HttpServer2: Added filter static_user_filter (class=org.apache.hadoop.http.lib.StaticUserWeb
Filter$StaticUserFilter) to context node
16/05/22 12:42:09 INFO http.HttpServer2: Added filter static_user_filter (class=org.apache.hadoop.http.lib.StaticUserWeb
Filter$StaticUserFilter) to context static
16/05/22 12:42:09 INFO http.HttpServer2: Added filter static_user_filter (class=org.apache.hadoop.http.lib.StaticUserWeb
Filter$StaticUserFilter) to context logs
16/05/22 12:42:09 INFO http.HttpServer2: adding path spec: /node/*
16/05/22 12:42:09 INFO http.HttpServer2: adding path spec: /ws/*
16/05/22 12:42:09 INFO http.HttpServer2: Jetty bound to port 8042
16/05/22 12:42:09 INFO morthbay.log: jetty-6.1.26
16/05/22 12:42:09 INFO morthbay.log: Extract jar-file: /D:/hadoop-2.6.0/share/hadoop/yarn/hadoop-yarn-common-2.6.0-jar1/we
bapps/node to C:\Users\shantanu\AppData\Local\Temp\jetty-0_0_0_8042_node_19t9k\webapp
16/05/22 12:42:10 INFO morthbay.log: Started HttpServer2$SelectChannelConnectorWithSafeStartup@0.0.0.0:8042
16/05/22 12:42:10 INFO webapp.WebApps: Web app /node started at 8042
16/05/22 12:42:12 INFO webapp.WebApps: Registered webapp guide modules
16/05/22 12:42:12 INFO client.RMPProxy: Connecting to ResourceManager at 0.0.0.0:8031
16/05/22 12:42:12 INFO nodemanager.NodeStatusUpdaterImpl: Sending out 0 NM container statuses: []
16/05/22 12:42:12 INFO nodemanager.NodeStatusUpdaterImpl: Registering with RM using containers: []
16/05/22 12:42:12 INFO security.NMContainerTokenSecretManager: Rolling master-key for container-tokens, got key with id
-215775033
16/05/22 12:42:12 INFO security.NMContainersSecretManager: Rolling master-key for container-tokens, got key with id 20514
16/05/22 12:42:12 INFO nodemanager.NodeStatusUpdaterImpl: Registered with ResourceManager as 132.72.25.79:57068 with to
tal resource of <memory:8192, vCores:8>
16/05/22 12:42:12 INFO nodemanager.NodeStatusUpdaterImpl: Notifying ContainerManager to unlock new container-requests
```

Namenode information

localhost:9870/dfshealth.html#tab-overview

HadoopOverviewDatanodesDatanode Volume FailuresSnapshotStartup ProgressUtilities

Overview 'localhost:9000' (✓active)

Started:	Sun Sep 08 00:06:16 +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-50e55097-947b-42ab-ac22-401506263472
Block Pool ID:	BP-1769990285-192.168.56.1-1724916758896

Summary


Security is off.

Safemode is off.

75 files and directories, 27 blocks (27 replicated blocks, 0 erasure coded block groups) = 102 total filesystem object(s).

All Applications

localhost:8088/cluster



All App

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	Used Resources
0	0	0	0	0	<memory:0 B, vCores:0>

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes
1	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation
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	State	FinalStatus
No data											

Showing 0 to 0 of 0 entries

RESULT:

The step-by-step installation and configuration of Hadoop on Windows system have been successfully completed.