

**Exp-1:****Downloading and installing Hadoop on Ubuntu, Understanding different Hadoop modes, Startup scripts, Configuration files****Aim:**

To successfully install, configure, and run Hadoop on a local system using a single-node setup.

**Procedure:****1. Install Java and SSH:**

- Update your package lists and install OpenJDK 8 and SSH.

```
sudo apt update
```

```
sudo apt install openjdk-8-jdk
```

```
java -version # Verify Java installation
```

```
sudo apt install ssh
```

**2. Create Hadoop User:**

- Add a dedicated user for Hadoop and generate SSH keys for passwordless SSH.

```
sudo adduser hadoop
```

```
su - hadoop # Switch to Hadoop user
```

```
ssh-keygen -t rsa
```

```
cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
```

```
chmod 640 ~/.ssh/authorized_keys
```

```
ssh localhost # Test SSH connection to localhost
```

**3. Download and Install Hadoop:**

- Download the latest Hadoop version (3.3.6), extract the tarball, and move it to the desired location.

```
wget https://downloads.apache.org/hadoop/common/hadoop-3.3.6/hadoop-3.3.6.tar.gz
```

```
tar -xvzf hadoop-3.3.6.tar.gz
```

```
mv hadoop-3.3.6 hadoop
```

**4. Configure Environment Variables:**

- **Update .bashrc to include Hadoop and Java paths.**

```
nano ~/.bashrc
```

```
# Add the following lines at the end
```

```
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
```

```
export HADOOP_HOME=$HOME/hadoop
export PATH=$PATH:$HADOOP_HOME/bin:$HADOOP_HOME/sbin
source ~/.bashrc # Apply changes
```

## 5. Edit Hadoop Configuration Files:

- Modify configuration files to set up the necessary Hadoop directories and services.
- **core-site.xml:**

```
nano $HADOOP_HOME/etc/hadoop/core-site.xml
```

```
# Add between <configuration></configuration>:
```

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

- **hdfs-site.xml:**

```
nano $HADOOP_HOME/etc/hadoop/hdfs-site.xml
```

**Add:**

```
<property>
  <name>dfs.replication</name>
  <value>1</value>
</property>
<property>
  <name>dfs.namenode.name.dir</name>
  <value>file:///home/hadoop/hadoopdata/hdfs/namenode</value>
</property>
<property>
  <name>dfs.datanode.data.dir</name>
  <value>file:///home/hadoop/hadoopdata/hdfs/datanode</value>
</property>
```

- **mapred-site.xml:**

```
cp $HADOOP_HOME/etc/hadoop/mapred-site.xml.template
  $HADOOP_HOME/etc/hadoop/mapred-site.xml
```

```
nano $HADOOP_HOME/etc/hadoop/mapred-site.xml
```

**Add:**

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

- **yarn-site.xml:**

```
nano $HADOOP_HOME/etc/hadoop/yarn-site.xml
```

**Add:**

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

**6. Format the NameNode:**

- Format the HDFS NameNode.

```
hdfs namenode -format
```

**7. Start Hadoop:**

- Start Hadoop services (NameNode, DataNode, ResourceManager, and NodeManager).

```
start-all.sh
```

```
jps # Verify running services
```

**8. Access Web Interfaces:**

- Verify that Hadoop is running by accessing the following URLs:

- **NameNode:** <http://localhost:9870>
- **Resource Manager:** <http://localhost:8088>

**9. Stop Hadoop Cluster:**

- Stop all Hadoop services.

```
stop-all.sh
```

```
yzm318@Ubuntu:~$ hadoop version
Hadoop 3.4.0
Source code repository git@github.com:apache/hadoop.git -r bd8b77f398f626bb7791783192ee7a5dfaec760
Compiled by root on 2024-03-04T06:29Z
Compiled on platform linux-aarch_64
Compiled with protoc 3.21.12
From source with checksum f7fe694a3613358b38812ae9c31114e
This command was run using /home/sai/hadoop-3.4.0/share/hadoop/common/hadoop-common-3.4.0.jar
yzm318@Ubuntu:~$
```

```
yzm318@Ubuntu:~$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as yzm318 in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [Ubuntu]
2024-10-10 11:30:50,873 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
Starting resourcemanager
Starting nodemanagers
```

```
yzm318@Ubuntu:~$ jps
4192 SecondaryNameNode
4870 Jps
4503 NodeManager
3945 DataNode
3818 NameNode
4379 ResourceManager
yzm318@Ubuntu:~$
```

Overview 'localhost:9000' (✓active)

|                |  |
|----------------|--|
| Started:       | Thu Oct 10 11:38:29 +0530 2024   |
| Version:       | 3.4.0, rbd8b77f398f626bb7791783192ee7a5dfaec760                                  |
| Compiled:      | Mon Mar 04 11:59:00 +0530 2024 by root from (HEAD detached at release-3.4.0-RC3) |
| Cluster ID:    | CID-653f4afa-bc4d-4111-9842-8c068261eaad   |
| Block Pool ID: | BP-750355565-127.0.1.1-1724908368015   |

Summary

|   |                  |
|---|------------------|
| Security is off.  |                  |
| Safemode is off.  |                  |
| 135 files and directories, 82 blocks (82 replicated blocks, 0 erasure coded block groups) = 217 total filesystem object(s). |                  |
| Heap Memory used 60.02 MB of 235 MB Heap Memory. Max Heap Memory is 871.5 MB.   |                  |
| Non Heap Memory used 53.17 MB of 54.84 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.                    |                  |
| Configured Capacity:  | 28.87 GB         |
| Configured Remote Capacity:   | 0 B              |
| DFS Used:   | 24.1 MB (0.08%)  |
| Non DFS Used:   | 20.08 GB         |
| DFS Remaining:  | 7.27 GB (25.18%) |

**RESULT:**

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