<u>Experiment 8: HDFS</u>

Aim : To demonstrate the running status of HDFS file system

Procedure:

1. **Install Java 8 if not already installed.**

2. **Download Hadoop 3.3.1.**

3. **Extract the downloaded Hadoop archive.**

4. **Set environment variables for Java and Hadoop paths.**

5. **Configure HDFS by updating core-site.xml and hdfs-site.xml.**

6. **Format HDFS using hdfs namenode -format.**

7. **Verify HDFS is running using jps command.**

Code:

```
!apt-get install openjdk-8-jdk-headless -qq
```

```
!wget -q https://downloads.apache.org/hadoop/common/hadoop-
3.3.1/hadoop-3.3.1.tar.gz
```

```
!tar xf hadoop-3.3.1.tar.gz
```

```
import os
os.environ['JAVA_HOME'] = '/usr/lib/jvm/java-8-openjdk-amd64'
os.environ['HADOOP_HOME'] = '/content/hadoop-3.3.1'
os.environ['HADOOP_CONF_DIR'] = '/content/hadoop-3.3.1/etc/hadoop'
os.environ['PATH'] = os.environ['HADOOP_HOME'] + '/bin:' +
os.environ['PATH']

# Configure HDFS
!cp /content/hadoop-3.3.1/etc/hadoop/core-site.xml /content/hadoop-
3.3.1/etc/hadoop/core-site.xml.bak
!cp /content/hadoop-3.3.1/etc/hadoop/hdfs-site.xml /content/hadoop-
3.3.1/etc/hadoop/hdfs-site.xml.bak

!echo -e
'<configuration><property><name>fs.default.name</name><value>hdfs://loc
```

```
alhost:9000</value></property></configuration>' > /content/hadoop-
3.3.1/etc/hadoop/core-site.xml
!echo -e
'<configuration><property><name>dfs.replication</name><value>1</value><
/property></configuration>' > /content/hadoop-3.3.1/etc/hadoop/hdfs-
site.xml

# Format HDFS
!hdfs namenode -format


# Verify HDFS is running
!jps
```

Output:

```
Re-format filesystem in Storage Directory root= /tmp/hadoop-
root/dfs/name; location= null ? (Y or N) Y
2023-09-01 04:16:59,475 INFO namenode.FSImage: Allocated new
BlockPoolId: BP-150640189-172.28.0.12-1693541819454
2023-09-01 04:16:59,475 INFO common.Storage: Will remove files:
[/tmp/hadoop-root/dfs/name/current/seen_txid, /tmp/hadoop-
root/dfs/name/current/fsimage_0000000000000000000, /tmp/hadoop-
root/dfs/name/current/VERSION, /tmp/hadoop-
root/dfs/name/current/fsimage_0000000000000000000.md5]
2023-09-01 04:16:59,515 INFO common.Storage: Storage directory
/tmp/hadoop-root/dfs/name has been successfully formatted.
2023-09-01 04:16:59,556 INFO namenode.FSImageFormatProtobuf: Saving
image file /tmp/hadoop-
root/dfs/name/current/fsimage.ckpt_0000000000000000000 using no
compression
2023-09-01 04:16:59,732 INFO namenode.FSImageFormatProtobuf: Image file
/tmp/hadoop-root/dfs/name/current/fsimage.ckpt_0000000000000000000 of
size 399 bytes saved in 0 seconds .
2023-09-01 04:16:59,745 INFO namenode.NNStorageRetentionManager: Going
to retain 1 images with txid >= 0
2023-09-01 04:16:59,794 INFO namenode.FSNamesystem: Stopping services
started for active state
2023-09-01 04:16:59,795 INFO namenode.FSNamesystem: Stopping services
started for standby state
2023-09-01 04:16:59,801 INFO namenode.FSImage: FSImageSaver clean
checkpoint: txid=0 when meet shutdown.
2023-09-01 04:16:59,801 INFO namenode.NameNode: SHUTDOWN_MSG:
/************************************************************
SHUTDOWN_MSG: Shutting down NameNode at ba73d57e1e43/172.28.0.12
************************************************************/
```

```
2320 Jps
```

Re-format filesystem in Storage Directory root= /tmp/hadoop-
root/dfs/name; location= null ? (Y or N) N
Format aborted in Storage Directory root= /tmp/hadoop-root/dfs/name;
location= null
2023-09-01 04:38:05,514 INFO namenode.FSNamesystem: Stopping services
started for active state
2023-09-01 04:38:05,515 INFO namenode.FSNamesystem: Stopping services
started for standby state
2023-09-01 04:38:05,517 INFO util.ExitUtil: Exiting with status 1:
ExitException
2023-09-01 04:38:05,521 INFO namenode.NameNode: SHUTDOWN_MSG:
/***
SHUTDOWN_MSG: Shutting down NameNode at ba73d57e1e43/172.28.0.12
***/
7432 Jps