Ex.No – 09   **Date –** 27/08/2024

SET UP A SINGLE HADOOP CLUSTER AND SHOW THE PROCESS USING WEB UI

AIM:

To set-up one node Hadoop cluster.

PROCEDURE:

1. System Update

2. Install Java

3. Add a dedicated Hadoop user

4. Install SSH and setup SSH certificates

5. Check if SSH works

6. Install Hadoop

7. Modify Hadoop config files

8. Format Hadoop filesystem

9. Start Hadoop

10. Check Hadoop through web UI

11. Stop Hadoop

THEORY

Hadoop is an Apache open source framework written in java that allows distributed processing of large datasets across clusters of computers using simple programming models.

A Hadoop frame-worked application works in an environment that provides distributed storage and computation across clusters of computers. Hadoop is designed to scale up from a single server to thousands of machines, each offering local computation and storage.

HADOOP ARCHITECTURE

Hadoop framework includes following four modules:

Hadoop Common: These are Java libraries and utilities required by other Hadoop modules.
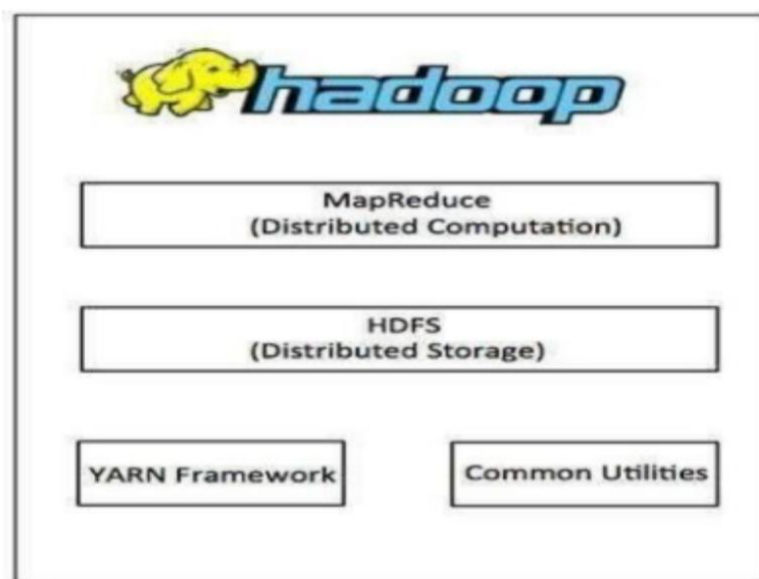
These libraries provide filesystem and OS level abstractions and contain the necessary Java files and scripts required to start Hadoop.

Hadoop YARN: This is a framework for job scheduling and cluster resource management.

Hadoop Distributed File System (HDFS): A distributed file system that provides high-throughput access to application data.

Hadoop MapReduce: This is a YARN-based system for parallel processing of large data sets.

We can use following diagram to depict these four components available in Hadoop framework.
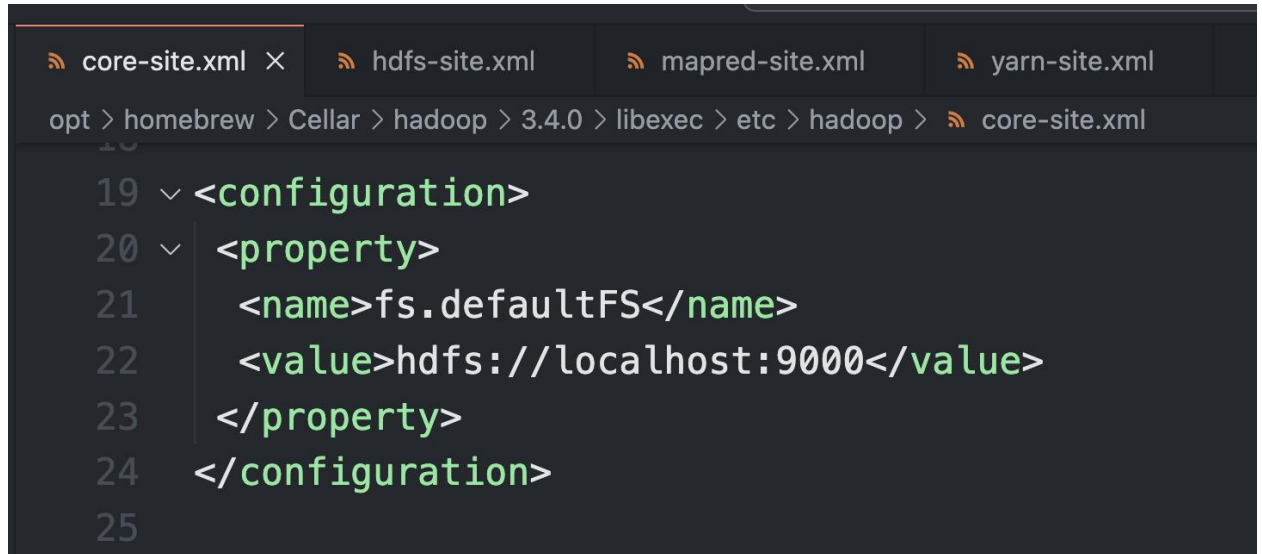

- brew install hadoop

```
(base) madhumithak@madhumithas-MacBook-Air ~ % cd /opt/homebrew/Cellar/hadoop/3.4.0/libexec/etc/hadoop
(base) madhumithak@madhumithas-MacBook-Air hadoop % code hadoop-env.sh
(base) madhumithak@madhumithas-MacBook-Air hadoop % code core-site.xml
(base) madhumithak@madhumithas-MacBook-Air hadoop % code hdfs-site.xml
(base) madhumithak@madhumithas-MacBook-Air hadoop % code mapred-site.xml
(base) madhumithak@madhumithas-MacBook-Air hadoop % code yarn-site.xml
(base) madhumithak@madhumithas-MacBook-Air hadoop %
```

- hadoop-env.sh file

```
# The java implementation to use. By default, this environment
# variable is REQUIRED on ALL platforms except OS X!
export JAVA_HOME=/Library/Java/JavaVirtualMachines/jdk-18.0.2.1.jdk/Contents/Home

# The language environment in which Hadoop runs. Use the English
# environment to ensure that logs are printed as expected.
export LANG=en_US.UTF-8
```
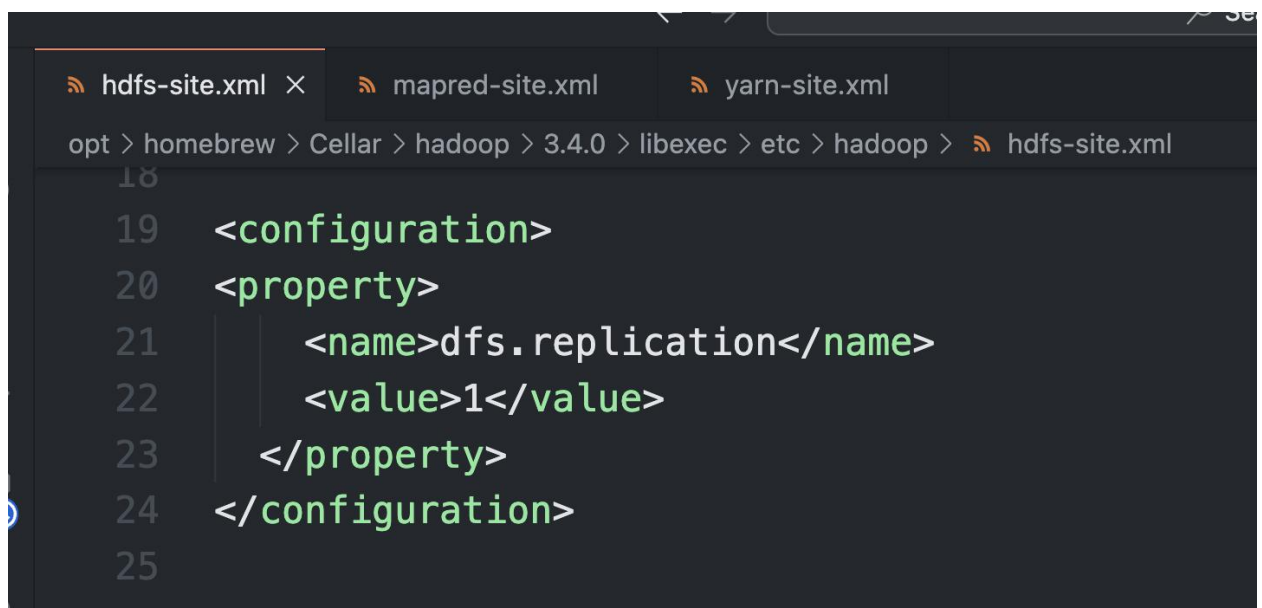
- Core-site.xml file

core-site.xml ✕   hdfs-site.xml   mapred-site.xml   yarn-site.xml

opt > homebrew > Cellar > hadoop > 3.4.0 > libexec > etc > hadoop > core-site.xml

```
19 ∨ <configuration>
20 ∨  <property>
21      <name>fs.defaultFS</name>
22      <value>hdfs://localhost:9000</value>
23    </property>
24   </configuration>
25
```

- Hdfs-site.xml file

hdfs-site.xml ✕   mapred-site.xml   yarn-site.xml

opt > homebrew > Cellar > hadoop > 3.4.0 > libexec > etc > hadoop > hdfs-site.xml

```
18
19    <configuration>
20    <property>
21        <name>dfs.replication</name>
22        <value>1</value>
23      </property>
24    </configuration>
25
```

Mapred-site.xml file

📄 mapred-site.xml ✕   📄 yarn-site.xml

opt > homebrew > Cellar > hadoop > 3.4.0 > libexec > etc > hadoop > 📄 mapred-site.xml

```
19  <configuration>
20  <property>
21          <name>mapreduce.framework.name</name>
22          <value>yarn</value>
23      </property>
24      <property>
25      <name>mapreduce.application.classpath</name>
26    <value>
27  $HADOOP_MAPRED_HOME/share/hadoop/mapreduce/*:$HADOOP_MAPRED_HOME/share/hadoop
28    </value>
29      </property>
30  </configuration>
31
```

- Yarn-site.xml file

📄 yarn-site.xml ✕

opt > homebrew > Cellar > hadoop > 3.4.0 > libexec > etc > hadoop > 📄 yarn-site.xml

```
14    -->
15    <configuration>
16
17    <!-- Site specific YARN configuration properties -->
18    <property>
19        <name>yarn.nodemanager.aux-services</name>
20        <value>mapreduce_shuffle</value>
21      </property>
22      <property>
23        <name>yarn.nodemanager.env-whitelist</name>
24       <value>
25  JAVA_HOME,HADOOP_COMMON_HOME,HADOOP_HDFS_HOME,HADOOP_CONF_DIR,CLASSP
26      </value>
27      </property>
28  </configuration>
29  |
```

- RSA key

```
(base) madhumithak@madhumithas-MacBook-Air hadoop % ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa
Generating public/private rsa key pair.
/Users/madhumithak/.ssh/id_rsa already exists.
Overwrite (y/n)? y
Your identification has been saved in /Users/madhumithak/.ssh/id_rsa
Your public key has been saved in /Users/madhumithak/.ssh/id_rsa.pub
The key fingerprint is:
SHA256:wUti4Zinq8j8wcXllwclvEYEwQt91TrvQIBWTjfD0Yo madhumithak@madhumithas-MacBook-Air.local
The key's randomart image is:
+---[RSA 3072]----+
|      +oB=oB+     |
|     = Bo=+.oo    |
|    o B.Bo+ o     |
|    .+o+ Eo=      |
|    .o .So..o     |
|  . .. . ... .    |
|   o.        o    |
|o. ..        .    |
|.oo.         |
+----[SHA256]-----+
(base) madhumithak@madhumithas-MacBook-Air hadoop % 
```

```
+----[SHA256]-----+
(base) madhumithak@madhumithas-MacBook-Air hadoop % cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
(base) madhumithak@madhumithas-MacBook-Air hadoop % hadoop namenode -format
WARNING: Use of this script to execute namenode is deprecated.
WARNING: Attempting to execute replacement "hdfs namenode" instead.

2024-08-08 08:17:20,214 INFO namenode.NameNode: STARTUP_MSG:
/************************************************************
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = madhumithas-MacBook-Air.local/127.0.0.1
STARTUP_MSG:   args = [-format]
STARTUP_MSG:   version = 3.4.0
```

- Start the server

```
(base) madhumithak@madhumithas-MacBook-Air hadoop % start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as madhumithak 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 [madhumithas-MacBook-Air.local]
2024-08-08 08:18:15,499 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where
applicable
Starting resourcemanager
Starting nodemanagers
(base) madhumithak@madhumithas-MacBook-Air hadoop % 
```

- Open localhost

Overview 'localhost:9000' (✓active)

Started:	Thu Aug 08 08:18:08 +0530 2024
Version:	3.4.0, rbd8b77f398f626bb7791783192ee7a5dfaeec760
Compiled:	Mon Mar 04 12:05:00 +0530 2024 by root from (HEAD detached at release-3.4.0-RC3)
Cluster ID:	CID-1273d0fe-e152-45ca-8a48-72e83f9d8eb3
Block Pool ID:	BP-2046741561-127.0.0.1-1723085241244

Summary

Security is off.

Safemode is off.

1 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 1 total filesystem object(s).

Heap Memory used 58.1 MB of 102 MB Heap Memory. Max Heap Memory is 2 GB.

Non Heap Memory used 50.63 MB of 53.38 MB Commited Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	228.27 GB
Configured Remote Capacity:	0 B

Non Heap Memory used 50.63 MB of 53.38 MB Commited Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	228.27 GB
Configured Remote Capacity:	0 B
DFS Used:	4 KB (0%)
Non DFS Used:	199.75 GB
DFS Remaining:	28.52 GB (12.49%)
Block Pool Used:	4 KB (0%)
DataNodes usages% (Min/Median/Max/stdDev):	0.00% / 0.00% / 0.00% / 0.00%
Live Nodes	1 (Decommissioned: 0, In Maintenance: 0)
Dead Nodes	0 (Decommissioned: 0, In Maintenance: 0)
Decommissioning Nodes	0
Entering Maintenance Nodes	0
Total Datanode Volume Failures	0 (0 B)
Number of Under-Replicated Blocks	0
Number of Blocks Pending Deletion (including replicas)	0
Block Deletion Start Time	Thu Aug 08 08:18:08 +0530 2024
Last Checkpoint Time	Thu Aug 08 08:17:21 +0530 2024
Last HA Transition Time	Never
Enabled Erasure Coding Policies	RS-6-3-1024k

NameNode Journal Status

Current transaction ID: 1

Journal Manager	State
FileJournalManager(root=/tmp/hadoop-madhumithak/dfs/name)	EditLogFileOutputStream(/tmp/hadoop-madhumithak/dfs/name/current/edits_inprogress_0000000000000000001)

NameNode Storage

Storage Directory	Type	State
/tmp/hadoop-madhumithak/dfs/name	IMAGE_AND_EDITS	Active

DFS Storage Types

Storage Type	Configured Capacity	Capacity Used	Capacity Remaining	Block Pool Used	Nodes In Service
DISK	228.27 GB	4 KB (0%)	28.52 GB (12.49%)	4 KB	1

Hadoop, 2024.

- Stopping server

```
(base) madhumithak@madhumithas-MacBook-Air hadoop % stop-all.sh
WARNING: Stopping all Apache Hadoop daemons as madhumithak in 10 seconds.
WARNING: Use CTRL-C to abort.
Stopping namenodes on [localhost]
Stopping datanodes
Stopping secondary namenodes [madhumithas-MacBook-Air.local]
2024-08-08 08:20:30,194 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where
applicable
Stopping nodemanagers
Stopping resourcemanager
(base) madhumithak@madhumithas-MacBook-Air hadoop %
```

RESULT:
Thus Hadoop is installed and Set-up Hadoop Successfully.