BIG DATA ANALYTICS

LAB ASSIGNMENT 8

Mahesh Pachare

FINAL YEAR B.TECH

IT 191080054

Aim :

To set up and install HBase and Oozie. To execute basic commands.

Theory :

- HBase:
  - HBase is a distributed column-oriented database built on top of the Hadoop file system. It is an open-source project and is horizontally scalable.
  - HBase is a data model that is similar to Google's big table designed to provide quick random access to huge amounts of structured data. It leverages the fault tolerance provided by the Hadoop File System (HDFS).
  - It is a part of the Hadoop ecosystem that provides random real-time read/write access to data in the Hadoop File System.
  - One can store the data in HDFS either directly or through HBase. Data consumers read/access the data in HDFS randomly using HBase. HBase sits on top of the Hadoop File System and provides read and write access.

- Features of HBase :
  - HBase is linearly scalable.
  - It has automatic failure support.
  - It provides consistent reading and writing.
  - It integrates with Hadoop, both as a source and a destination.
  - It has easy Java API for clients.

- It provides data replication across clusters.

- Applications of HBase
  - It is used whenever there is a need to write heavy applications.
  - HBase is used whenever we need to provide fast random access to available data.
  - Companies such as Facebook, Twitter, Yahoo, and Adobe use HBase internally.

- Oozie

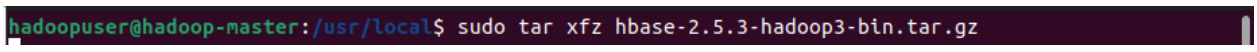
  - Oozie is designed to simplify the management of complex Hadoop workflows by providing a centralized system for defining, scheduling, and executing those workflows. It allows users to specify dependencies between jobs, and to chain together multiple jobs into a single workflow. Workflows can be triggered by time-based or event-based triggers and can be monitored and managed using a web-based console or REST API.
  - One of the key benefits of Oozie is its flexibility. It supports a wide range of actions, including MapReduce jobs, Pig scripts, Hive queries, Sqoop imports and exports, and shell commands. This makes it easy to incorporate a variety of Hadoop tools and technologies into your workflows.
  - Oozie also provides support for conditional execution, error handling, and retries, which can help to ensure the reliability and robustness of your workflows. It also integrates with other Hadoop ecosystem projects, such as HCatalog and HBase, making it easy to use with a variety of other technologies.
  - Overall, Oozie is a powerful tool for managing and automating complex Hadoop workflows. Its flexibility, scalability, and reliability make it a popular choice for many organizations working with Hadoop.
  - It consists of two parts:
    - Workflow engine:The responsibility of a workflow engine is to store and run workflows composed of Hadoop jobs e.g.,MapReduce,Pig, hive

        - Coordinator engine: It runs workflow jobs based on predefined schedules and availability of data.
    - Features of Oozie include:
        - Having a client API and command line interface which can be used to launch, control and monitor jobs from Java applications.
        - Using its Web Service APIs one can control jobs from anywhere.
        - Having provisions to execute jobs that are scheduled to run periodically.
        - Having provisions to send email notifications upon completion of jobs.

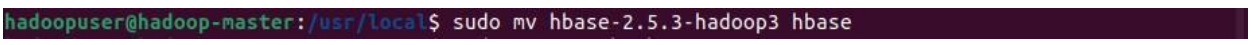
EXECUTION STEPS AND OUTPUT :

- **DO THE FOLLOWING STEPS IN ALL THE MACHINES**
- **Installing and Setting Up HBase**
    - Downloading binary version of HBase from the following link: https://downloads.apache.org/hbase/2.5.3/

    - Running the following command to unzip the file



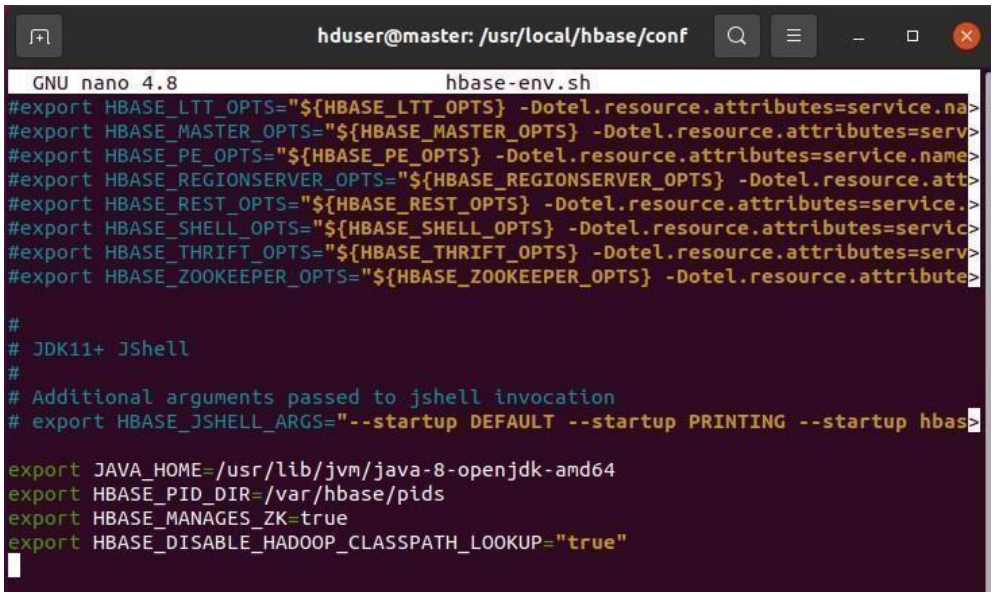
    - Renaming the folder as hbase using following command:



    - Updating bashrc file by adding following line

```
hadoopuser@hadoop-master:/usr/local$ sudo nano ~/.bashrc
hadoopuser@hadoop-master:/usr/local$ source ~/.bashrc
hadoopuser@hadoop-master:/usr/local$
```

- Editing hbase-env.sh file

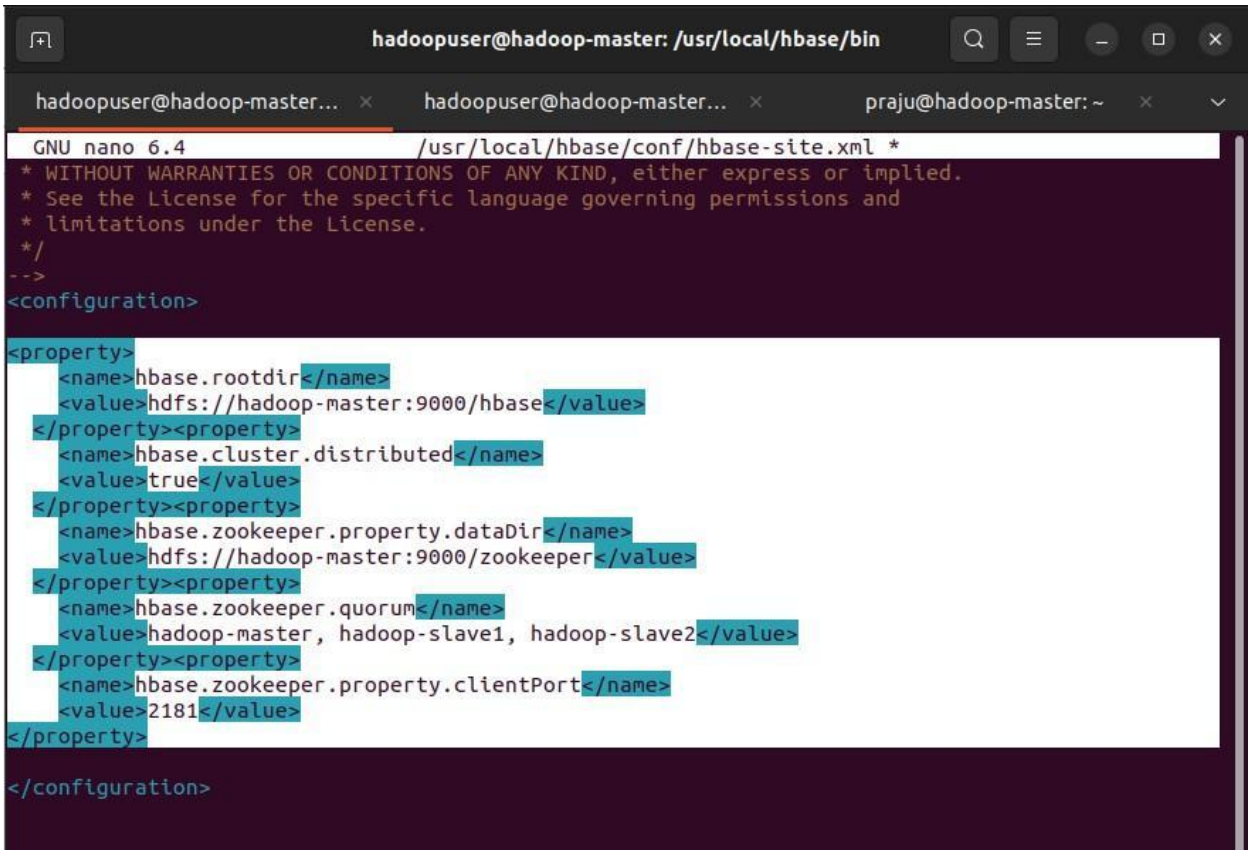


- Running the following command to verify HBase installation:

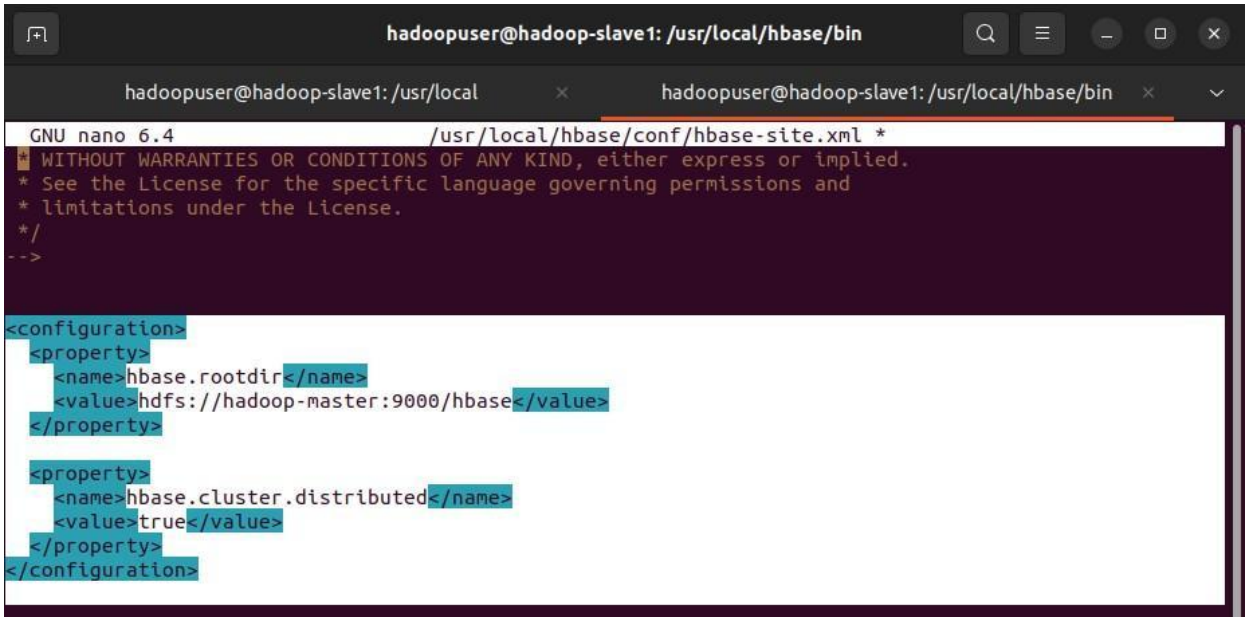
```
hadoopuser@hadoop-master:/usr/local/hbase/bin$ sudo ./start-hbase.sh
running master, logging to /usr/local/hbase/bin/../logs/hbase-root-master-hadoop-master.out
hadoopuser@hadoop-master:/usr/local/hbase/bin$ hbase version
HBase 2.5.3-hadoop3
Source code repository git://bd1c794451e7/home/taklwu/hbase-rm/output/hbase revision=d385524561
f771dcb402905c2bdcaeb4a8fecbdb
Compiled by taklwu on Wed Feb  1 10:46:16 UTC 2023
From source with checksum 8c75feee25791553f34e76ee4ba922a647fe9d35f891d3bcf6c1c83b7164d18f8e0fc
367318ea8f94ea97bd5280775aee35633fc1b424ccc50feb3848f50ad9f
```

- Editing hbase-site.xml file

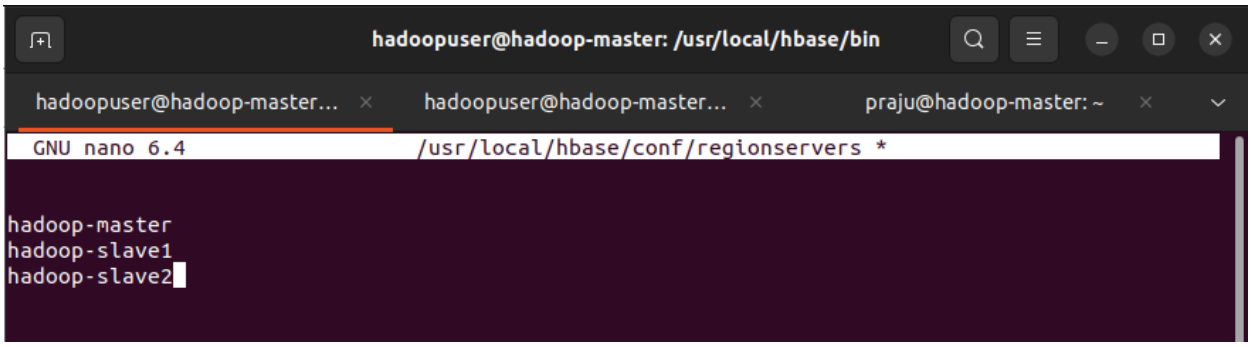
  On Master Machine:



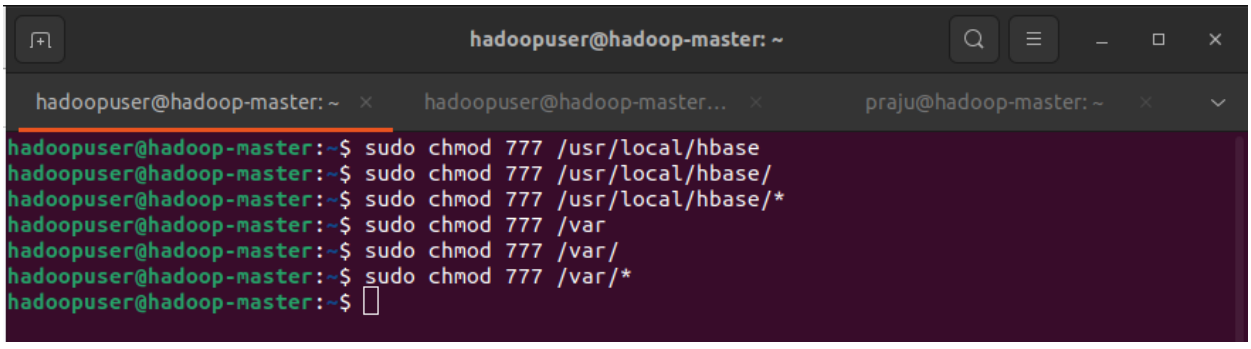
  On both Slave Machine :



- Editing regionservers file (Only in MAster Machine)



- Running the following commands in all the machines to give permissions:



- On Master Node, starting Hadoop Cluster

```
hadoopuser@hadoop-master:~$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as hadoopuser in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [hadoop-master]
Starting datanodes
Starting secondary namenodes [hadoop-master]
Starting resourcemanager
Starting nodemanagers
```

- On Master Node, starting HBase Cluster

```
hadoopuser@hadoop-master:/usr/local/hbase/bin$ ./start-hbase.sh
hadoop-master: running zookeeper, logging to /usr/local/hbase/bin/../logs/hbase-hadoopuser-zook
eeper-hadoop-master.out
hadoop-slave1: running zookeeper, logging to /usr/local/hbase/bin/../logs/hbase-hadoopuser-zook
eeper-hadoop-slave1.out
hadoop-slave2: running zookeeper, logging to /usr/local/hbase/bin/../logs/hbase-hadoopuser-zook
eeper-hadoop-slave2.out
hadoop-master: /usr/local/hbase/bin/hbase-daemon.sh: line 249: /var/hbase/pids/hbase-hadoopuser
-zookeeper.pid: Permission denied
hadoop-slave1: /usr/local/hbase/bin/hbase-daemon.sh: line 249: /var/hbase/pids/hbase-hadoopuser
-zookeeper.pid: Permission denied
hadoop-slave2: /usr/local/hbase/bin/hbase-daemon.sh: line 249: /var/hbase/pids/hbase-hadoopuser
-zookeeper.pid: Permission denied
running master, logging to /usr/local/hbase/logs/hbase-hadoopuser-master-hadoop-master.out
/usr/local/hbase/bin/hbase-daemon.sh: line 249: /var/hbase/pids/hbase-hadoopuser-master.pid: Pe
rmission denied
```

- Jps on Master Node

```
hadoopuser@hadoop-master:/usr/local/hbase/bin$ jps
24754 HRegionServer
24499 HQuorumPeer
23013 DataNode
23223 SecondaryNameNode
23880 NodeManager
22810 NameNode
23709 ResourceManager
24605 HMaster
26303 Jps
```

- Jps on Slave Node

```
hadoopuser@hadoop-slave1:~$ jps
15317 NodeManager
16186 Jps
14683 DataNode
15500 HQuorumPeer
15661 HRegionServer
hadoopuser@hadoop-slave1:~$
```

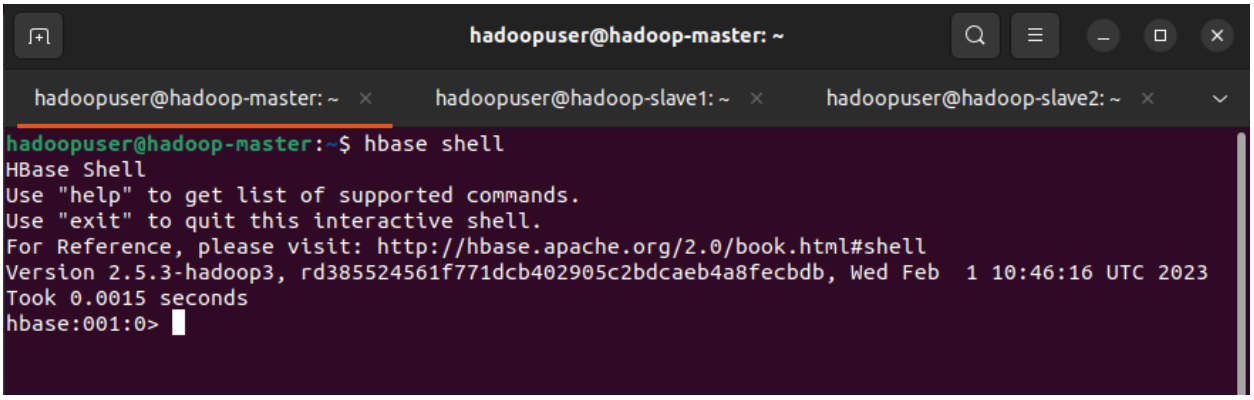
```
hadoopuser@hadoop-slave2:~$ jps
15664 NodeManager
15457 DataNode
16532 Jps
15932 HQuorumPeer
16079 HRegionServer
hadoopuser@hadoop-slave2:~$
```

- Executing basic commands using HBase Shell

  After the successful installation of HBase on top of Hadoop, we get an

interactive shell to execute various commands and perform several operations. Using these commands, we can perform multiple operations on data tables that can give better data storage efficiencies and flexible interaction by the client.

- Starting Hbase shell



- Version Command

  This command will display the currently used HBase version in command mode

```
hbase:001:0> version
2.5.3-hadoop3, rd385524561f771dcb402905c2bdcaeb4a8fecbdb, Wed Feb  1 10:46:16 UTC 2023
Took 0.0160 seconds
```

- Status Command

  This command will give details about the system status like the number of servers in the cluster, active server count, and average load value.

```
hbase:004:0> status
1 active master, 0 backup masters, 2 servers, 0 dead, 1.0000 average load
Took 12.4670 seconds
```

- Create table Command

  Syntax: create <tablename>, <columnfamilyname>

```
hbase:013:0> create 'employee', 'employee_col_family';
Created table employee
Took 10.1795 seconds
=> Hbase::Table - employee
```

- List all tables command

```
hbase:014:0> list
TABLE
employee
1 row(s)
Took 0.9539 seconds
=> ["employee"]
```

- Describe table

```
hbase:016:0> describe 'employee'
Table employee is ENABLED
employee, {TABLE_ATTRIBUTES => {METADATA => {'hbase.store.file-tracker.impl' => 'DEFAULT'}}}
COLUMN FAMILIES DESCRIPTION
{NAME => 'employee_col_family', INDEX_BLOCK_ENCODING => 'NONE', VERSIONS => '1', KEEP_DELETED_C
ELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', TTL => 'FOREVER', MIN_VERSIONS => '0', REPLICAT
ION_SCOPE => '0', BLOOMFILTER => 'ROW', IN_MEMORY => 'false', COMPRESSION => 'NONE', BLOCKCACHE
 => 'true', BLOCKSIZE => '65536 B (64KB)'}

1 row(s)
Quota is disabled
Took 2.2993 seconds
```

This command describes the named table.

It gives more information about column families present in the mentioned table.

- Inserting data into the table using put command

  Syntax: put '<table name>','row1','<colfamily:colname>','<value>'

```
hbase:018:0> put 'employee', '1', 'employee_col_family: name', 'Prajwal'
Took 0.7583 seconds
hbase:019:0> put 'employee', '1', 'employee_col_family: dept', 'Finance'
Took 0.0147 seconds
hbase:020:0> put 'employee', '1', 'employee_col_family: salary', '20000'
Took 0.0545 seconds
hbase:021:0> put 'employee', '2', 'employee_col_family: name', 'Tushar'
Took 0.0183 seconds
hbase:022:0> put 'employee', '2', 'employee_col_family: dept', 'Sales'
Took 0.0714 seconds
hbase:023:0> put 'employee', '2', 'employee_col_family: salary', '30000'
Took 0.1231 seconds
```

- Reading data from table using get command

  Syntax: get '<table name>','row1'

```
hbase:025:0> get 'employee', '1'
COLUMN                  CELL
 employee_col_family: de timestamp=2023-05-08T01:49:07.250, value=Finance
 pt
 employee_col_family: na timestamp=2023-05-08T01:48:48.642, value=Prajwal
 me
 employee_col_family: sa timestamp=2023-05-08T01:49:24.482, value=20000
 lary
1 row(s)
Took 0.7587 seconds
```

- Reading a specific column from table

  Syntax: get 'table name', 'rowid', {COLUMN ⇒ 'column family:column name '}

```
hbase:026:0> get 'employee', '1', {COLUMN => 'employee_col_family: name'}
COLUMN                  CELL
 employee_col_family: na timestamp=2023-05-08T01:48:48.642, value=Prajwal
 me
1 row(s)
Took 0.0447 seconds
```

- Deleting table

```
hbase:027:0> list
TABLE
employee
1 row(s)
Took 0.1579 seconds
=> ["employee"]
```

```
hbase:029:0> disable 'employee'
Took 0.9403 seconds
hbase:030:0> drop 'employee'
Took 0.8531 seconds
hbase:031:0> list
TABLE
0 row(s)
Took 0.1788 seconds
=> []
hbase:032:0>
```

- Installing and Setting Up Oozie
  - Downloading Apache Oozie 5.2.0 from the below link.: https://archive.apache.org/dist/oozie/5.2.0/

```
hadoopuser@hadoop-master:/usr/local$ sudo wget https://archive.apache.org/dist/oozie/5.2.0/oozi
e-5.2.0.tar.gz
--2023-05-08 02:17:44--  https://archive.apache.org/dist/oozie/5.2.0/oozie-5.2.0.tar.gz
Resolving archive.apache.org (archive.apache.org)... 138.201.131.134, 2a01:4f8:172:2ec5::2
Connecting to archive.apache.org (archive.apache.org)|138.201.131.134|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 2816120 (2.7M) [application/x-gzip]
Saving to: ‘oozie-5.2.0.tar.gz’

oozie-5.2.0.tar.gz      100%[=============================>]   2.69M  3.14MB/s    in 0.9s

2023-05-08 02:17:46 (3.14 MB/s) - ‘oozie-5.2.0.tar.gz’ saved [2816120/2816120]

hadoopuser@hadoop-master:/usr/local$
```

  - Extracting the downloaded Apache Oozie tar

```
hadoopuser@hadoop-master:/usr/local$ sudo tar -xzf oozie-5.2.0.tar.gz
hadoopuser@hadoop-master:/usr/local$ ls
apache-hive-3.1.2-bin.tar.gz  hbase                          oozie-5.2.0          share
bin                           hbase-2.5.3-hadoop3-bin.tar.gz  oozie-5.2.0.tar.gz   src
etc                           include                         pig-0.17.0
games                         lib                             pig-0.17.0.tar.gz
hadoop                        man                             sbin
hadoopuser@hadoop-master:/usr/local$
```

  - Checking if Maven is installed

```
hadoopuser@hadoop-master:/usr/local$ sudo apt install maven
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
The following package was automatically installed and is no longer required:
  tcpd
Use 'sudo apt autoremove' to remove it.
The following additional packages will be installed:
  libaopalliance-java libcdi-api-java libcommons-lang3-java
  libgeronimo-annotation-1.3-spec-java libgeronimo-interceptor-3.0-spec-java libguice-java
  libjansi-java libmaven-parent-java libmaven-resolver-java libmaven-shared-utils-java
  libmaven3-core-java libplexus-cipher-java libplexus-classworlds-java
  libplexus-component-annotations-java libplexus-interpolation-java
  libplexus-sec-dispatcher-java libplexus-utils2-java libsisu-inject-java libsisu-plexus-java
```

```
hadoopuser@hadoop-master:/usr/local$ mvn -version
Apache Maven 3.6.3
Maven home: /usr/share/maven
Java version: 1.8.0_362, vendor: Private Build, runtime: /usr/lib/jvm/java-8-openjdk-amd64/jre
Default locale: en_IN, platform encoding: UTF-8
OS name: "linux", version: "5.19.0-41-generic", arch: "amd64", family: "unix"
hadoopuser@hadoop-master:/usr/local$
```

- Compiling Apache Oozie to create binary files for the distro Command:

```
hduser@master:/usr/local/oozie-5.2.0$ sudo ./bin/mkdistro.sh -DskipTests
[INFO] Scanning for projects...
[WARNING]
[WARNING] Some problems were encountered while building the effective model for org.apache.oozie:oozie-main:pom:5.2.0
[WARNING] 'parent.relativePath' of POM org.apache.oozie:oozie-main:5.2.0 (/usr/local/oozie-5.2.0/pom.xml) points at org.apache.oozie:
ructure @ line 21, column 13
[WARNING]
[WARNING] It is highly recommended to fix these problems because they threaten the stability of your build.
[WARNING]
[WARNING] For this reason, future Maven versions might no longer support building such malformed projects.
[WARNING]
[INFO] ------------------------------------------------------------------------
[INFO] Reactor Build Order:
[INFO]
[INFO] Apache Oozie Main                                                  [pom]
[INFO] Apache Oozie Fluent Job                                            [pom]
[INFO] Apache Oozie Fluent Job API                                        [jar]
[INFO] Apache Oozie Client                                                [jar]
[INFO] Apache Oozie Share Lib Oozie                                       [jar]
[INFO] Apache Oozie Share Lib HCatalog                                    [jar]
[INFO] Apache Oozie Share Lib Distcp                                      [jar]
[INFO] Apache Oozie Core                                                  [jar]
[INFO] Apache Oozie Share Lib Streaming                                   [jar]
[INFO] Apache Oozie Share Lib Pig                                         [jar]
[INFO] Apache Oozie Share Lib Git                                         [jar]
[INFO] Apache Oozie Share Lib Hive                                        [jar]
[INFO] Apache Oozie Share Lib Hive 2                                      [jar]
[INFO] Apache Oozie Share Lib Sqoop                                       [jar]
[INFO] Apache Oozie Examples                                              [jar]
[INFO] Apache Oozie Share Lib Spark                                       [jar]
[INFO] Apache Oozie Share Lib                                             [pom]
[INFO] Apache Oozie Docs                                                  [war]
[INFO] Apache Oozie WebApp                                                [war]
[INFO] Apache Oozie Tools                                                 [jar]
[INFO] Apache Oozie MiniOozie                                             [jar]
[INFO] Apache Oozie Fluent Job Client                                     [jar]
[INFO] Apache Oozie Server                                                [jar]
[INFO] Apache Oozie Distro                                                [jar]
[INFO] Apache Oozie ZooKeeper Security Tests                              [jar]
[INFO]
[INFO] --------------------< org.apache.oozie:oozie-main >---------------------
[INFO] Building Apache Oozie Main 5.2.0                                 [1/25]
[INFO] --------------------------------[ pom ]---------------------------------
[INFO]
[INFO] --- maven-enforcer-plugin:1.4.1:enforce (clean) @ oozie-main ---
```

```
[INFO] ------------------------------------------------------------------------
[INFO] Reactor Summary for Apache Oozie Main 5.2.0:
[INFO]
[INFO] Apache Oozie Main ................................... SUCCESS [  1.199 s]
[INFO] Apache Oozie Fluent Job ............................. SUCCESS [  0.066 s]
[INFO] Apache Oozie Fluent Job API ......................... SUCCESS [  6.960 s]
[INFO] Apache Oozie Client ................................. SUCCESS [  3.767 s]
[INFO] Apache Oozie Share Lib Oozie ........................ SUCCESS [  2.117 s]
[INFO] Apache Oozie Share Lib HCatalog ..................... SUCCESS [  1.588 s]
[INFO] Apache Oozie Share Lib Distcp ....................... SUCCESS [  0.534 s]
[INFO] Apache Oozie Core ................................... SUCCESS [ 12.643 s]
[INFO] Apache Oozie Share Lib Streaming .................... SUCCESS [  2.252 s]
[INFO] Apache Oozie Share Lib Pig .......................... SUCCESS [  2.179 s]
[INFO] Apache Oozie Share Lib Git .......................... SUCCESS [  2.171 s]
[INFO] Apache Oozie Share Lib Hive ......................... SUCCESS [  1.841 s]
[INFO] Apache Oozie Share Lib Hive 2 ....................... SUCCESS [  2.386 s]
[INFO] Apache Oozie Share Lib Sqoop ........................ SUCCESS [  2.071 s]
[INFO] Apache Oozie Examples ............................... SUCCESS [  2.071 s]
[INFO] Apache Oozie Share Lib Spark ........................ SUCCESS [  4.558 s]
[INFO] Apache Oozie Share Lib .............................. SUCCESS [ 16.370 s]
[INFO] Apache Oozie Docs ................................... SUCCESS [  0.873 s]
[INFO] Apache Oozie WebApp ................................. SUCCESS [  5.197 s]
[INFO] Apache Oozie Tools .................................. SUCCESS [  1.529 s]
[INFO] Apache Oozie MiniOozie .............................. SUCCESS [  1.416 s]
[INFO] Apache Oozie Fluent Job Client ...................... SUCCESS [  0.818 s]
[INFO] Apache Oozie Server ................................. SUCCESS [  3.448 s]
[INFO] Apache Oozie Distro ................................. SUCCESS [ 23.468 s]
[INFO] Apache Oozie ZooKeeper Security Tests ............... SUCCESS [  2.655 s]
[INFO] ------------------------------------------------------------------------
[INFO] BUILD SUCCESS
[INFO] ------------------------------------------------------------------------
[INFO] Total time:  01:45 min
[INFO] Finished at: 2023-04-22T21:56:43+05:30
[INFO] ------------------------------------------------------------------------

Oozie distro created, DATE[2023.04.22-16:24:57GMT] VC-REV[unavailable], available at [/usr/local/oozie-5.2.0/distro/target]
```

- Creating a libext directory under the Oozie directory

```
hduser@master:/usr/local$ cd /usr/local/oozie-5.2.0/distro/target/oozie-5.2.0-distro/oozie-5.2.0
hduser@master:/usr/local/oozie-5.2.0/distro/target/oozie-5.2.0-distro/oozie-5.2.0$ sudo mkdir libext
hduser@master:/usr/local/oozie-5.2.0/distro/target/oozie-5.2.0-distro/oozie-5.2.0$
```

- Downloading the ext-2.2.zip

```
hduser@master:/usr/local/oozie-5.2.0/distro/target/oozie-5.2.0-distro/oozie-5.2.0$ cd libext/
hduser@master:/usr/local/oozie-5.2.0/distro/target/oozie-5.2.0-distro/oozie-5.2.0/libext$ sudo wget http://archive.cloudera.com/gplextras/misc/ext-2.2.zip
--2023-04-22 21:59:02--  http://archive.cloudera.com/gplextras/misc/ext-2.2.zip
Resolving archive.cloudera.com (archive.cloudera.com)... 151.101.152.167
Connecting to archive.cloudera.com (archive.cloudera.com)|151.101.152.167|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 6800612 (6.5M) [application/zip]
Saving to: 'ext-2.2.zip'

ext-2.2.zip                         100%[=============================================================================================================>]   6.49M  2.98MB/s    in 2.2s

2023-04-22 21:59:04 (2.98 MB/s) - 'ext-2.2.zip' saved [6800612/6800612]
```

- Copying all Hadoop libraries into the libext folder

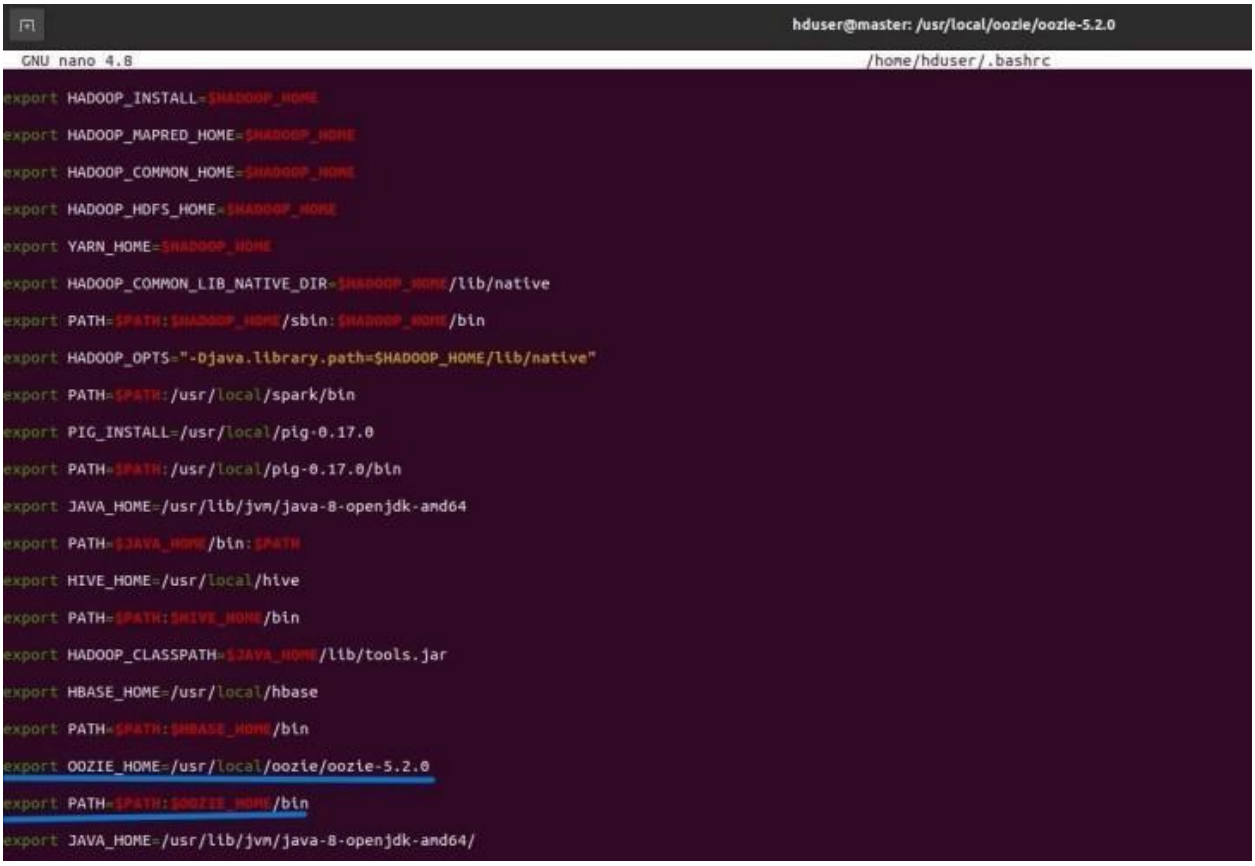
```
hduser@master:/usr/local/oozie-5.2.0/distro/target/oozie-5.2.0-distro/oozie-5.2.0/libext$ sudo cp /usr/local/hadoop/share/hadoop/common/*.jar /usr/local/oozie-5.2.0/distro/target/oozie-5.2.0-distro/oozie-5.2.0/libext/
hduser@master:/usr/local/oozie-5.2.0/distro/target/oozie-5.2.0-distro/oozie-5.2.0/libext$ sudo cp /usr/local/hadoop/share/hadoop/common/lib/*.jar /usr/local/oozie-5.2.0/distro/target/oozie-5.2.0-distro/oozie-5.2.0/libext/
hduser@master:/usr/local/oozie-5.2.0/distro/target/oozie-5.2.0-distro/oozie-5.2.0/libext$ sudo cp /usr/local/hadoop/share/hadoop/hdfs/*.jar /usr/local/oozie-5.2.0/distro/target/oozie-5.2.0-distro/oozie-5.2.0/libext/
hduser@master:/usr/local/oozie-5.2.0/distro/target/oozie-5.2.0-distro/oozie-5.2.0/libext$ sudo cp /usr/local/hadoop/share/hadoop/hdfs/lib/*.jar /usr/local/oozie-5.2.0/distro/target/oozie-5.2.0-distro/oozie-5.2.0/libext/
hduser@master:/usr/local/oozie-5.2.0/distro/target/oozie-5.2.0-distro/oozie-5.2.0/libext$ sudo cp /usr/local/hadoop/share/hadoop/mapreduce/*.jar /usr/local/oozie-5.2.0/distro/target/oozie-5.2.0-distro/oozie-5.2.0/libext/
hduser@master:/usr/local/oozie-5.2.0/distro/target/oozie-5.2.0-distro/oozie-5.2.0/libext$ sudo cp /usr/local/hadoop/share/hadoop/mapreduce/sources/*.jar /usr/local/oozie-5.2.0/distro/target/oozie-5.2.0-distro/oozie-5.2.0/libext/
hduser@master:/usr/local/oozie-5.2.0/distro/target/oozie-5.2.0-distro/oozie-5.2.0/libext$ sudo cp /usr/local/hadoop/share/hadoop/yarn/*.jar /usr/local/oozie-5.2.0/distro/target/oozie-5.2.0-distro/oozie-5.2.0/libext/
hduser@master:/usr/local/oozie-5.2.0/distro/target/oozie-5.2.0-distro/oozie-5.2.0/libext$ sudo cp /usr/local/hadoop/share/hadoop/yarn/lib/*.jar /usr/local/oozie-5.2.0/distro/target/oozie-5.2.0-distro/oozie-5.2.0/libext/
hduser@master:/usr/local/oozie-5.2.0/distro/target/oozie-5.2.0-distro/oozie-5.2.0/libext$
```

- Creating a directory name oozie under /usr/local and Moving oozie-5.2.0 directory into /usr/local/oozie directory

```
hduser@master:/usr/local$ sudo mkdir /usr/local/oozie
hduser@master:/usr/local$ cd oozie-5.2.0/distro/target
```

```
hduser@master:/usr/local$ cd /usr/local/oozie-5.2.0/distro/target/oozie-5.2.0-distro
hduser@master:/usr/local/oozie-5.2.0/distro/target/oozie-5.2.0-distro$ sudo mv oozie-5.2.0 /usr/local/oozie
hduser@master:/usr/local/oozie-5.2.0/distro/target/oozie-5.2.0-distro$
```

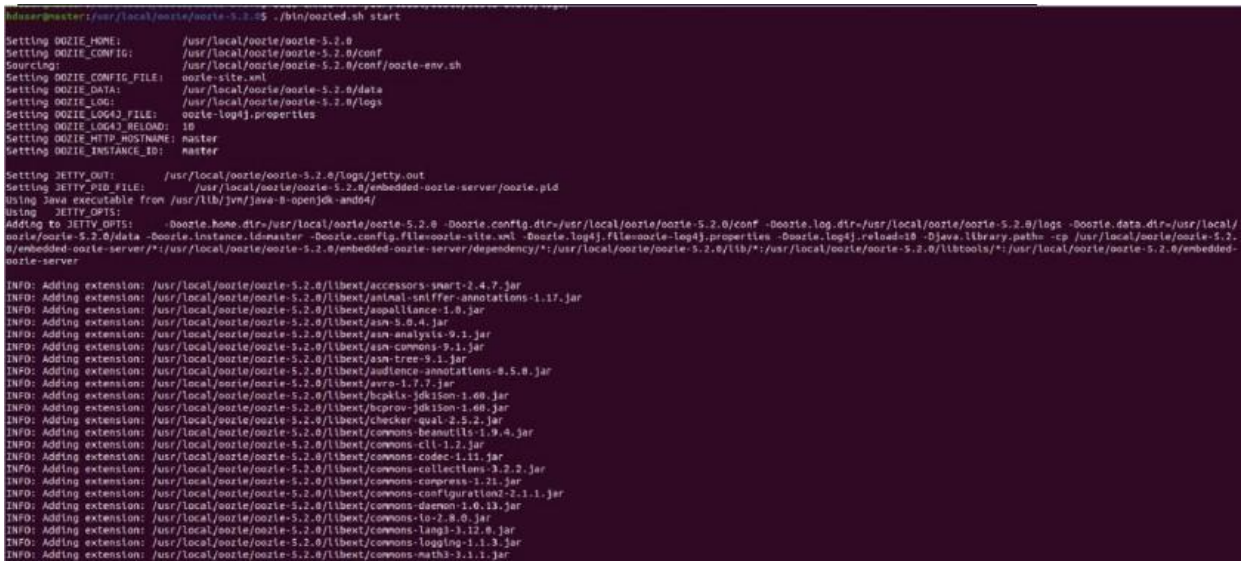
- Adding OOZIE_HOME path on the ".bashcr" file



- Setup oozie

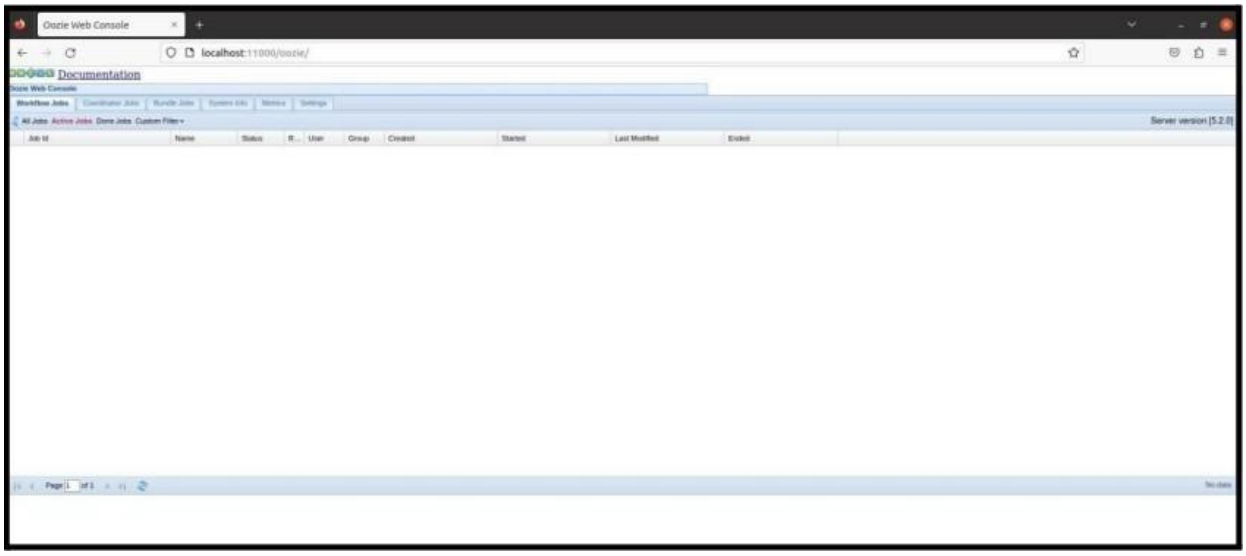
```
hduser@master:/usr/local/oozie/oozie-5.2.0$ sudo ./bin/oozie-setup.sh

INFO: Adding extension: /usr/local/oozie/oozie-5.2.0/libext/accessors-smart-2.4.7.jar
INFO: Adding extension: /usr/local/oozie/oozie-5.2.0/libext/animal-sniffer-annotations-1.17.jar
INFO: Adding extension: /usr/local/oozie/oozie-5.2.0/libext/aopalliance-1.0.jar
INFO: Adding extension: /usr/local/oozie/oozie-5.2.0/libext/asm-5.0.4.jar
INFO: Adding extension: /usr/local/oozie/oozie-5.2.0/libext/asm-analysis-9.1.jar
INFO: Adding extension: /usr/local/oozie/oozie-5.2.0/libext/asm-commons-9.1.jar
INFO: Adding extension: /usr/local/oozie/oozie-5.2.0/libext/asm-tree-9.1.jar
INFO: Adding extension: /usr/local/oozie/oozie-5.2.0/libext/audience-annotations-0.5.0.jar
INFO: Adding extension: /usr/local/oozie/oozie-5.2.0/libext/avro-1.7.7.jar
INFO: Adding extension: /usr/local/oozie/oozie-5.2.0/libext/bcpkix-jdk15on-1.60.jar
INFO: Adding extension: /usr/local/oozie/oozie-5.2.0/libext/bcprov-jdk15on-1.60.jar
INFO: Adding extension: /usr/local/oozie/oozie-5.2.0/libext/checker-qual-2.5.2.jar
INFO: Adding extension: /usr/local/oozie/oozie-5.2.0/libext/commons-beanutils-1.9.4.jar
INFO: Adding extension: /usr/local/oozie/oozie-5.2.0/libext/commons-cli-1.2.jar
INFO: Adding extension: /usr/local/oozie/oozie-5.2.0/libext/commons-codec-1.11.jar
INFO: Adding extension: /usr/local/oozie/oozie-5.2.0/libext/commons-collections-3.2.2.jar
INFO: Adding extension: /usr/local/oozie/oozie-5.2.0/libext/commons-compress-1.21.jar
INFO: Adding extension: /usr/local/oozie/oozie-5.2.0/libext/commons-configuration2-2.1.1.jar
INFO: Adding extension: /usr/local/oozie/oozie-5.2.0/libext/commons-daemon-1.0.13.jar
INFO: Adding extension: /usr/local/oozie/oozie-5.2.0/libext/commons-io-2.8.0.jar
INFO: Adding extension: /usr/local/oozie/oozie-5.2.0/libext/commons-lang3-3.12.0.jar
```

- Starting the Oozie server



- Checking Oozie services

- The Oozie server has started. Verifying it from the web browser. http://localhost:11000/oozie



- Output of jps command

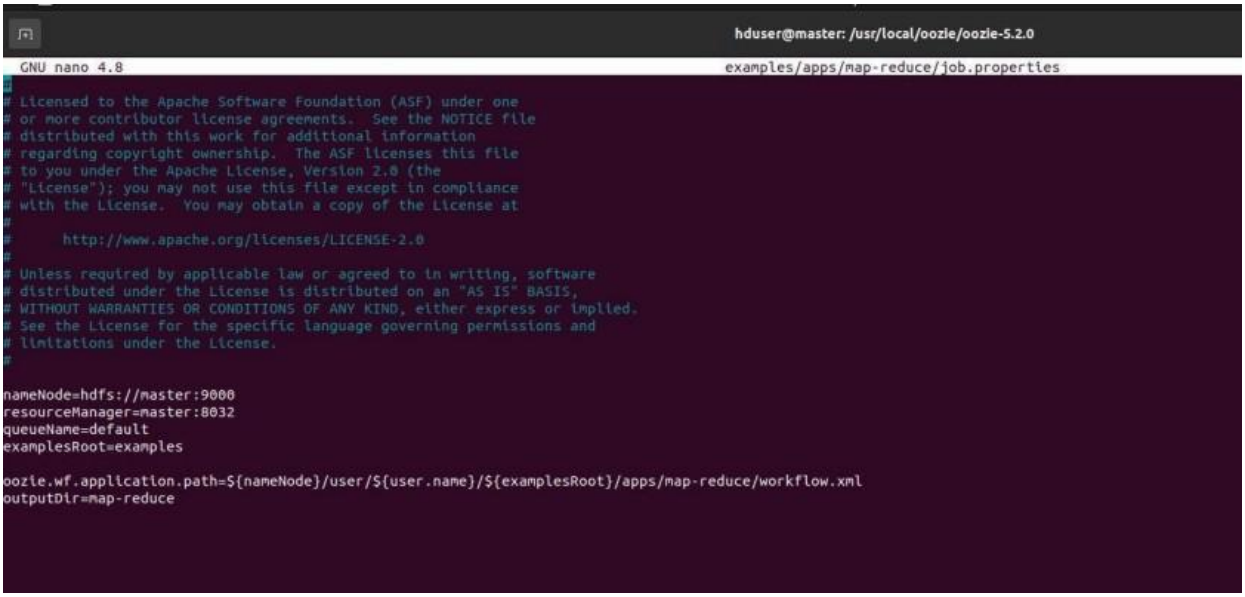
```
hduser@master:/usr/local/oozie/oozie-5.2.0$ jps
2834 SecondaryNameNode
3047 ResourceManager
5367 EmbeddedOozieServer
2442 NameNode
3197 NodeManager
5870 Jps
2590 DataNode
```

- Executing Oozie workflow job:
  - Oozie examples are bundled within the Oozie distribution in the oozie- examples.tar.gz file.

    Expanding this file will create an examples/ directory in the local file system.

```
hduser@master:/usr/local/oozie/oozie-5.2.0$ sudo tar -xzf oozie-examples.tar.gz
```

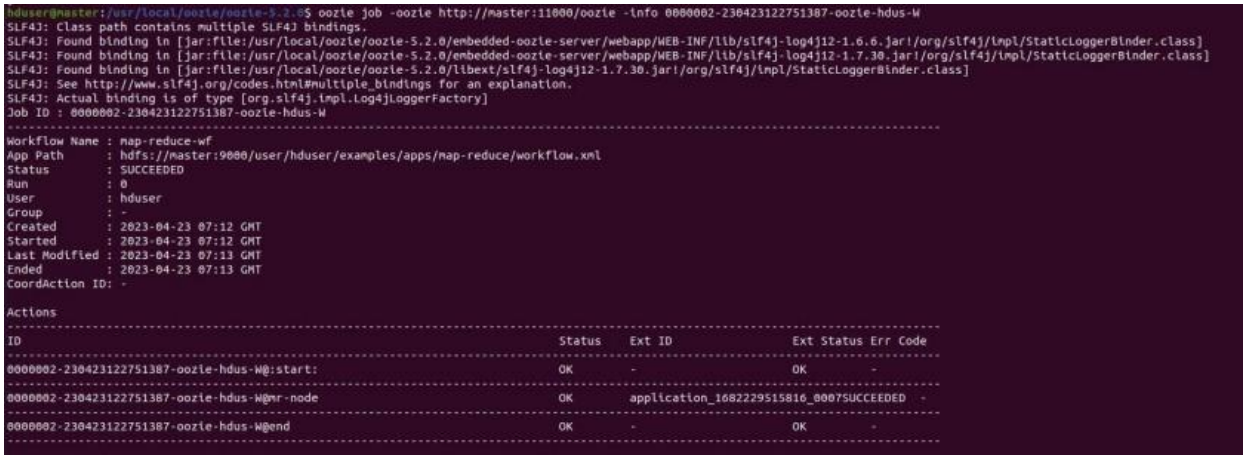
  - Editing job.properties file



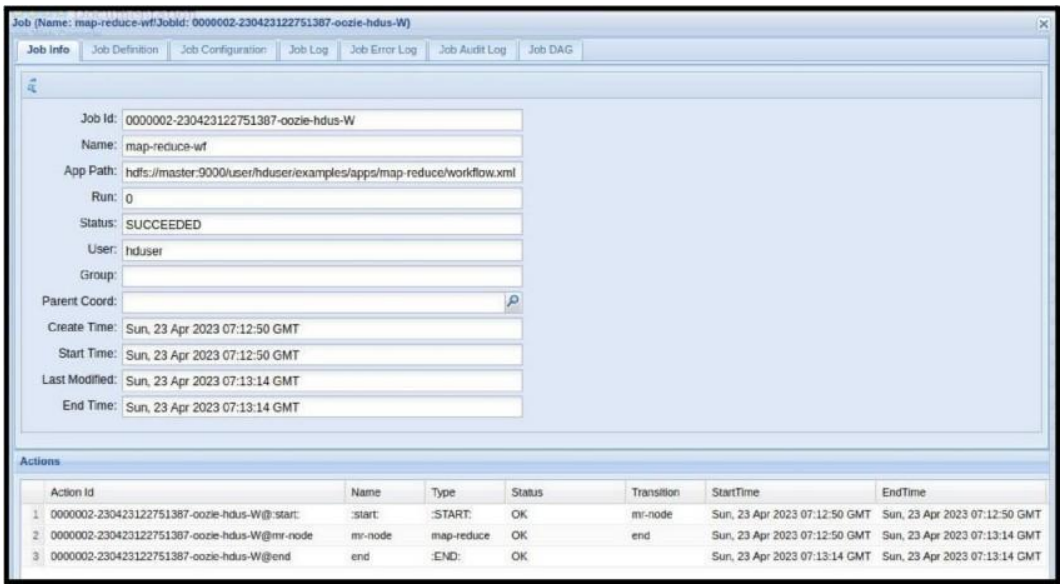
- Copying the examples/ directory to the user HOME directory in HDFS:

- Running the job

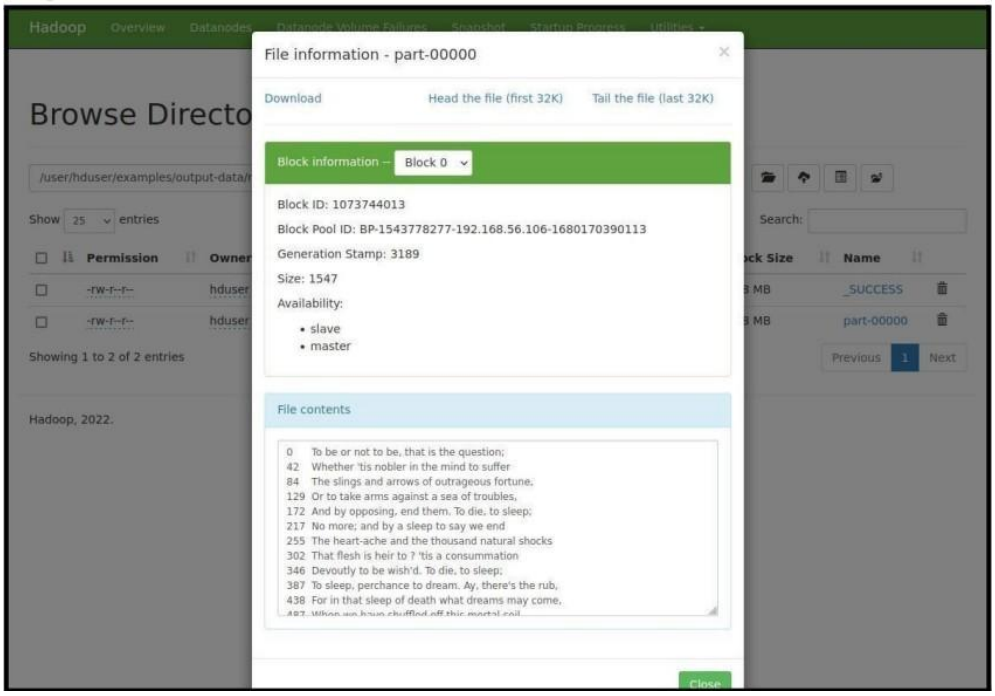
- Checking the workflow job status



- On web browser



- Output file in HDF



CONCLUSION :

From this experiment, I learned about HBase and Oozie. I understand that HBase is a NoSQL, column-oriented database. I learned basic commands which can be used to insert, delete, and update data in the HBase database that is to interact with the database. I also learned about Oozie and installed it on Linux. Also, executed the Oozie workflow job