Install Hadoop/Oozie on AWS VMs:

- Create EC2 Instance:
  - select ubuntu OS
  - architecture 64 bit
  - instance type -t2.xlarge 16gb
  - key pair vockey
  - network setting -Edit
  - select type as all traffic and source type anywhere
  - storage 80GB
  - click launch instance

- Please add private ip addresses as below on all machines at /etc/hosts using root

  ex:

  172.31.80.189 master
  172.31.80.153 slave1
  172.31.89.24 slave2

- Install softwares on all machines. Below are commands

  ```
  sudo apt-get update
  sudo apt-get install openjdk-8-jdk -y
  sudo apt-get install unzip
  sudo apt-get install maven -y
  ```

- Generate public key of Master and add it to all machines authorized_keys file

  ```
  ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa
  cat ~/.ssh/id_rsa.pub
  ```

  Note: added key manually to slave machines

- Upload and untar Hadoop tar on master and copy to slaves

  Ex: scp hadoop.tar.gz slave1:~

  scp hadoop.tar.gz slave2:~

- create below users on master

  ```
  sudo addgroup hadoop

  sudo adduser ubuntu hadoop

  sudo adduser --ingroup hadoop oozie

  sudo adduser oozie sudo
  ```

- untar Hadoop tar and create directories for hadoop

  ```
  sudo tar zxvf hadoop.tar.gz --directory /usr/local

  rm hadoop.tar.gz

  sudo mkdir -p /usr/local/hdata/dfs/name

  sudo mkdir -p /usr/local/hdata/dfs/data
  ```

- change permissions of /usr/local folder

  master machine:

  sudo chown -R ubuntu:hadoop /usr/local/

  slaves machine:

  sudo chown -R ubuntu /usr/local/

- add below environmental variables to bashrc file on all machines.

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

  export HADOOP_HOME=/usr/local/hadoop-2.10.2

  export PATH=$PATH:$HADOOP_HOME/bin

  export PATH=$PATH:$HADOOP_HOME/sbin

  export HADOOP_MAPRED_HOME=$HADOOP_HOME

  export HADOOP_COMMON_HOME=$HADOOP_HOME

  export HADOOP_HDFS_HOME=$HADOOP_HOME

  export YARN_HOME=$HADOOP_HOME

  export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/native

  export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/native"

  export HADOOP_CLASSPATH=${JAVA_HOME}/lib/tools.jar

  command: source ~/.bashrc

  - On Master format HDFS

    hdfs namenode -format

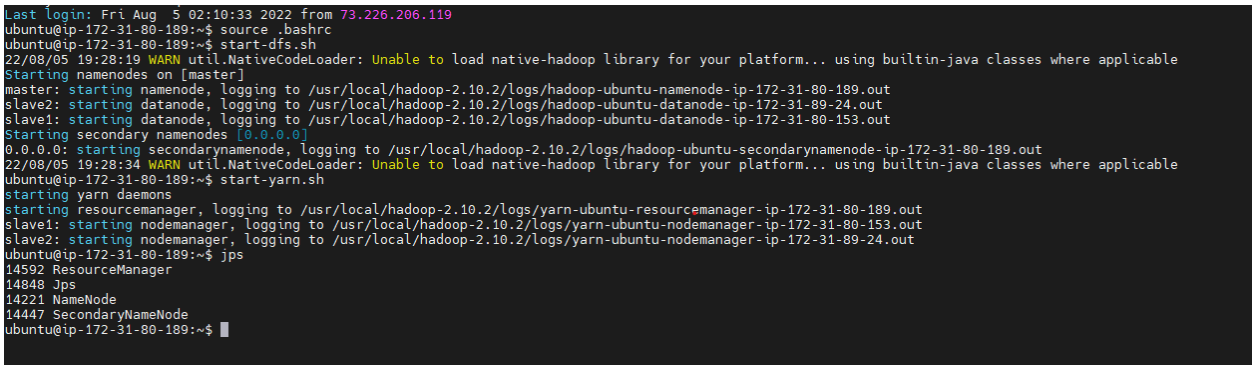
    you can check Hadoop setup complete or not with below

    Master:

    start-dfs.sh

    start-yarn.sh

    jps



slaves:

jps

```
ubuntu@ip-172-31-89-24:~$ source .bashrc
ubuntu@ip-172-31-89-24:~$ jps
14011 DataNode
14174 NodeManager
14366 Jps
ubuntu@ip-172-31-89-24:~$
```

Then stop Hadoop nodes.

Stop-yarn.sh

Stop-dfs.sh

Oozie installation and setup

- **We will setup oozie on only Master machine**
- **Please execute below**

wget https://dlcdn.apache.org/oozie/5.2.1/oozie-5.2.1.tar.gz
wget https://dlcdn.apache.org/oozie/5.2.1/oozie-5.2.1.tar.gz
sudo mkdir /usr/local/temp
sudo chmod 777 -R /usr/local/temp
tar zxvf oozie-5.2.1.tar.gz --directory /usr/local/temp
cd /usr/local/temp/oozie-5.2.1/bin/
./mkdistro.sh -Dhadoop.version=2.10.2 -DskipTests

cd /usr/local/temp/oozie-5.2.1/distro/target/oozie-5.2.1-distro/oozie-5.2.1
mkdir libext
cd libext
wget http://archive.cloudera.com/gplextras/misc/ext-2.2.zip

- Please copy hadoop jars to oozie

cp $HADOOP_HOME/share/hadoop/common/*.jar .
cp $HADOOP_HOME/share/hadoop/common/lib/*.jar .
cp $HADOOP_HOME/share/hadoop/hdfs/*.jar .
cp $HADOOP_HOME/share/hadoop/hdfs/lib/*.jar .
cp $HADOOP_HOME/share/hadoop/mapreduce/*.jar .
cp $HADOOP_HOME/share/hadoop/mapreduce/lib/*.jar .
cp $HADOOP_HOME/share/hadoop/yarn/*.jar .
cp $HADOOP_HOME/share/hadoop/yarn/lib/*.jar .

- Move installation to /usr/local folder

sudo mkdir /usr/local/oozie
cd /usr/local/temp/oozie-5.2.1/distro/target/oozie-5.2.1-distro
sudo mv oozie-5.2.1 /usr/local/oozie

sudo chown -R oozie:hadoop /usr/local/temp
sudo chown -R oozie:hadoop /usr/local/oozie

- Add oozie environmental variables to bashrc file with oozie user

su oozie
cd ~
nano .bashrc

export OOZIE_HOME=/usr/local/oozie/oozie-5.2.1
export PATH=$PATH:$OOZIE_HOME/bin

source .bashrc

- Add below to oozie-site.xml on master with oozie user

File: $OOZIE_HOME/conf/oozie-site.xml

```
<property>

<name>oozie.service.HadoopAccessorService.hadoop.configurations</name>

<value>*=/usr/local/hadoop-2.10.2/etc/hadoop</value>

</property>

<property>

<name>oozie.service.WorkflowAppService.system.libpath</name>

<value>hdfs:///user/${user.name}/share/lib</value>

</property>
```

- Add below to core-site.xml on master with ubuntu user

File : $HADOOP_HOME/etc/hadoop/core-site.xml

```
<!-- OOZIE -->
  <property>
        <name>hadoop.proxyuser.oozie.hosts</name>
        <value>*</value>
  </property>
  <property>
        <name>hadoop.proxyuser.oozie.groups</name>
        <value>*</value>
  </property>
```

- Run oozie-setup.sh on master with oozie user

- Start nodes on master with ubuntu user

start-dfs.sh
start-yarn.sh

hdfs dfs -mkdir -p /user/oozie
hdfs dfs -chown oozie /user/oozie
hdfs dfs -chmod -R 777 /

- run below to setup Hadoop jars with oozie and start oozie

oozie-setup.sh sharelib create -fs hdfs://master:8020
oozied.sh start

```
Setting up oozie DB

Validate DB Connection
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/oozie/oozie-5.2.1/lib/slf4j-log4j12-1.6.6.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/oozie/oozie-5.2.1/lib/slf4j-reload4j-1.7.36.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/oozie/oozie-5.2.1/libext/slf4j-reload4j-1.7.36.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
DONE
DB schema exists

The SQL commands have been written to: /tmp/ooziedb-5280909819085144590.sql

Existing PID file found during start.
Removing/clearing stale PID file.
Oozie server started . PID: 14992.
oozie@ip-172-31-80-189:/home/ubuntu$
```

Run sample jobs on oozie

- Please extract oozie sample code

  cd /usr/local/temp/oozie-5.2.1/examples/target
  tar xzvf oozie-examples-5.2.1-examples.tar.gz

- change job.properties to have master details(change localhost to master)

```
# See the License for the specific language governing permissions and
# limitations under the License.
#

nameNode=hdfs://master:8020
resourceManager=master:8032
queueName=default
examplesRoot=examples

oozie.wf.application.path=${nameNode}/user/${user.name}/${examplesRoot}/apps/map-reduce/workflow.xml
outputDir=map-reduce
oozie@ip-172-31-80-189:/usr/local/temp/oozie-5.2.1/examples/target$
```

- Create folder at hdfs
  hdfs dfs -put examples /user/oozie

- stop oozie and start Hadoop

  oozied.sh stop (with oozie user)

  stop-yarn.sh
  stop-dfs.sh

  (restart instances)

  start-dfs.sh
  start-yarn.sh

- run job at oozie
  cd /usr/local/temp/oozie-5.2.1/examples/target

  oozied.sh start
  oozie job -oozie http://master:11000/oozie -config ./examples/apps/map-reduce/job.properties -run

```
oozie@ip-172-31-80-189:/usr/local/temp/oozie-5.2.1/examples/target$ oozie job -oozie http://master:11000/oozie -config ./examples/apps/map-reduce/job.properties -run
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/oozie/oozie-5.2.1/lib/slf4j-log4j12-1.6.6.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/oozie/oozie-5.2.1/lib/slf4j-reload4j-1.7.36.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/oozie/oozie-5.2.1/libext/slf4j-reload4j-1.7.36.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
job: 0000000-220805194228784-oozie-oozi-W
oozie@ip-172-31-80-189:/usr/local/temp/oozie-5.2.1/examples/target$
```

```
oozie@ip-172-31-80-189:/usr/local/temp/oozie-5.2.1/examples/target$ hdfs dfs -ls  /user/oozie
22/08/05 20:00:47 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 3 items
drwxr-xr-x   - oozie supergroup          0 2022-08-05 01:47 /user/oozie/examples
drwxr-xr-x   - oozie supergroup          0 2022-08-05 19:50 /user/oozie/oozie-oozi
drwxr-xr-x   - oozie supergroup          0 2022-08-05 01:36 /user/oozie/share
oozie@ip-172-31-80-189:/usr/local/temp/oozie-5.2.1/examples/target$ hdfs dfs -ls  /user/oozie/examples
22/08/05 20:01:08 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 4 items
drwxr-xr-x   - oozie supergroup          0 2022-08-05 01:38 /user/oozie/examples/apps
drwxr-xr-x   - oozie supergroup          0 2022-08-05 01:38 /user/oozie/examples/input-data
drwxr-xr-x   - oozie supergroup          0 2022-08-05 19:50 /user/oozie/examples/output-data
drwxr-xr-x   - oozie supergroup          0 2022-08-05 01:38 /user/oozie/examples/src
oozie@ip-172-31-80-189:/usr/local/temp/oozie-5.2.1/examples/target$ hdfs dfs -ls  /user/oozie/examples/input-data
22/08/05 20:01:26 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
drwxr-xr-x   - oozie supergroup          0 2022-08-05 01:38 /user/oozie/examples/input-data/rawLogs
drwxr-xr-x   - oozie supergroup          0 2022-08-05 01:38 /user/oozie/examples/input-data/text
oozie@ip-172-31-80-189:/usr/local/temp/oozie-5.2.1/examples/target$ hdfs dfs -ls  /user/oozie/examples/output-data
22/08/05 20:01:38 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 1 items
drwxr-xr-x   - oozie supergroup          0 2022-08-05 19:50 /user/oozie/examples/output-data/map-reduce
oozie@ip-172-31-80-189:/usr/local/temp/oozie-5.2.1/examples/target$ hdfs dfs -ls  /user/oozie/examples/output-data/map-reduce
22/08/05 20:01:52 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r--   2 oozie supergroup          0 2022-08-05 19:50 /user/oozie/examples/output-data/map-reduce/_SUCCESS
-rw-r--r--   2 oozie supergroup       1547 2022-08-05 19:50 /user/oozie/examples/output-data/map-reduce/part-00000
oozie@ip-172-31-80-189:/usr/local/temp/oozie-5.2.1/examples/target$ hdfs dfs -get  /user/oozie/examples/output-data/map-reduce/part-00000
```

Output :

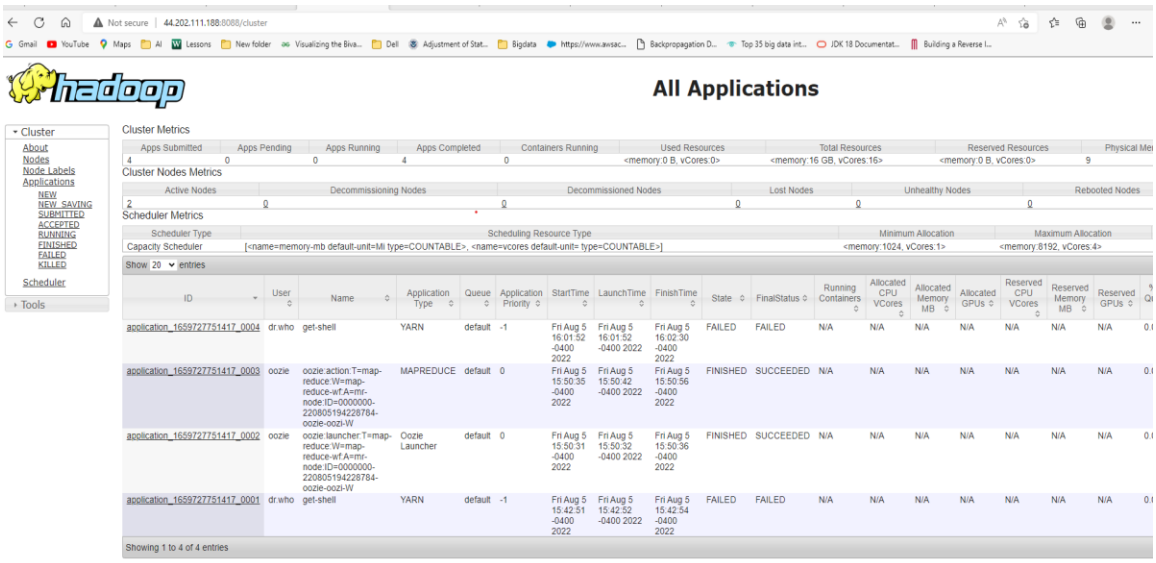
hdfs dfs -ls /user/oozie/examples
hdfs dfs -ls /user/oozie/examples/output-data/map-reduce
hdfs dfs -get /user/oozie/examples/output-data/map-reduce/part-00000

```
get: `part-00000': File exists
oozie@ip-172-31-80-189:/usr/local/temp/oozie-5.2.1/examples/target$ cat part-00000
0       To be or not to be, that is the question;
42      Whether 'tis nobler in the mind to suffer
84      The slings and arrows of outrageous fortune,
129     Or to take arms against a sea of troubles,
172     And by opposing, end them. To die, to sleep;
217     No more; and by a sleep to say we end
255     The heart-ache and the thousand natural shocks
302     That flesh is heir to ? 'tis a consummation
346     Devoutly to be wish'd. To die, to sleep;
387     To sleep, perchance to dream. Ay, there's the rub,
438     For in that sleep of death what dreams may come,
487     When we have shuffled off this mortal coil,
531     Must give us pause. There's the respect
571     That makes calamity of so long life,
608     For who would bear the whips and scorns of time,
657     Th'oppressor's wrong, the proud man's contumely,
706     The pangs of despised love, the law's delay,
751     The insolence of office, and the spurns
791     That patient merit of th'unworthy takes,
832     When he himself might his quietus make
871     With a bare bodkin? who would fardels bear,
915     To grunt and sweat under a weary life,
954     But that the dread of something after death,
999     The undiscovered country from whose bourn
1041    No traveller returns, puzzles the will,
1081    And makes us rather bear those ills we have
1125    Than fly to others that we know not of?
1165    Thus conscience does make cowards of us all,
1210    And thus the native hue of resolution
1248    Is sicklied o'er with the pale cast of thought,
1296    And enterprises of great pitch and moment
1338    With this regard their currents turn awry,
1381    And lose the name of action.
oozie@ip-172-31-80-189:/usr/local/temp/oozie-5.2.1/examples/target$
```


- Oozie console:
  http://<IP>:11000/oozie
  Note :use public ip address of Master machine

- Also the check resource manager: http://<IP>:8088

  Note :use public ip address of Master machine



Setup app for flight data :

- Copy data from locahost to Master EC2 at /home/ubuntu/
- Unzip Data.zip file
    - unzip Data.zip

      bzip2 -d 1987.csv.bz2
      bzip2 -d 1988.csv.bz2
      bzip2 -d 1989.csv.bz2
      bzip2 -d 1990.csv.bz2
      bzip2 -d 1991.csv.bz2
      bzip2 -d 1992.csv.bz2
      bzip2 -d 1993.csv.bz2
      bzip2 -d 1994.csv.bz2
      bzip2 -d 1995.csv.bz2
      bzip2 -d 1996.csv.bz2
      bzip2 -d 1997.csv.bz2
      bzip2 -d 1998.csv.bz2
      bzip2 -d 1999.csv.bz2
      bzip2 -d 2000.csv.bz2
      bzip2 -d 2001.csv.bz2
      bzip2 -d 2002.csv.bz2
      bzip2 -d 2003.csv.bz2
      bzip2 -d 2004.csv.bz2
      bzip2 -d 2005.csv.bz2
      bzip2 -d 2006.csv.bz2
      bzip2 -d 2007.csv.bz2
      bzip2 -d 2008.csv.bz2

- create directory flight app

      cd /usr/local/temp/oozie-5.2.1/examples/target/examples/apps

      /usr/local/temp/oozie-5.2.1/examples/target/examples/apps$ mkdir flightapp

Copy sample files to this folder

Cd flightapp
cp -R ../map-reduce/* .
cd lib
rm *

- copy code jar this lib folder

cp /home/ubuntu/flight.jar /usr/local/temp/oozie-5.2.1/examples/target/examples/apps/flightapp/lib

- change job.properties as below

nameNode=hdfs://master:8020

resourceManager=master:8032

queueName=default

examplesRoot=examples

oozie.wf.application.path=${nameNode}/user/${user.name}/${examplesRoot}/apps/flightapp/workflow.xml

outputDir=FlightApp

oozie@ip-172-31-80-189:/usr/local/temp/oozie-5.2.1/examples/target/examples/apps/flightapp$ vi job.properties

oozie@ip-172-31-80-189:/usr/local/temp/oozie-5.2.1/examples/target/examples/apps/flightapp$ cat job

job-with-config-class.properties job.properties

oozie@ip-172-31-80-189:/usr/local/temp/oozie-5.2.1/examples/target/examples/apps/flightapp$ cat job.properties

- copy workflow.xml to flightapp folder
- start oozie

oozied.sh start

- and setup share lib

oozie-setup.sh sharelib create -fs hdfs://master:8020

- upload data /home /ubuntu/Data folder to :/usr/local/temp/oozie-5.2.1/examples/target/examples/target/examples/input-data/flightappdata
- upload examples folder to hdfs
  hdfs dfs -put examples /user/oozie

- Run Job
  oozie job -oozie http://master:11000/oozie -config ./examples/apps/flightapp/job.properties -run