

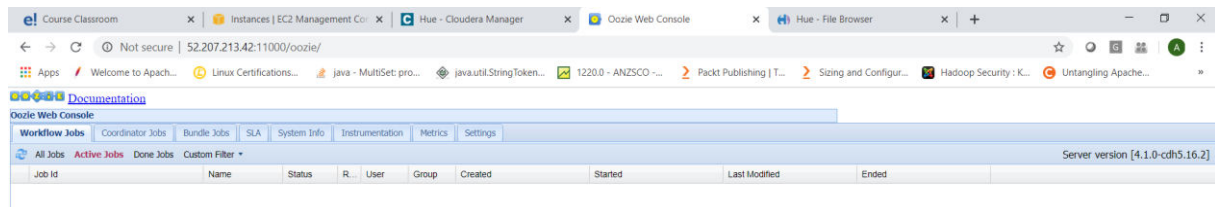
Oozie: Working with workflows in CDH

Add oozie service in your CDH cluster

To enable oozie web console follow instructions here:

https://docs.cloudera.com/documentation/enterprise/5-14-x/topics/admin_oozie_console.html

enable oozie web console by enabling property in Oozie>configuration



Refer : <https://oozie.apache.org/docs/4.2.0/index.html>

Creating workflow from HUE:

To create a workflow that will run some simple Hive Queries.

Step 1:

From CLI connect to Hive CLI/Beeline (or use Hue) and create a database (mydb) and a table(testtbl2).

Table can be created as follows: [use mydb;create table testtbl2 (id int,name string)]

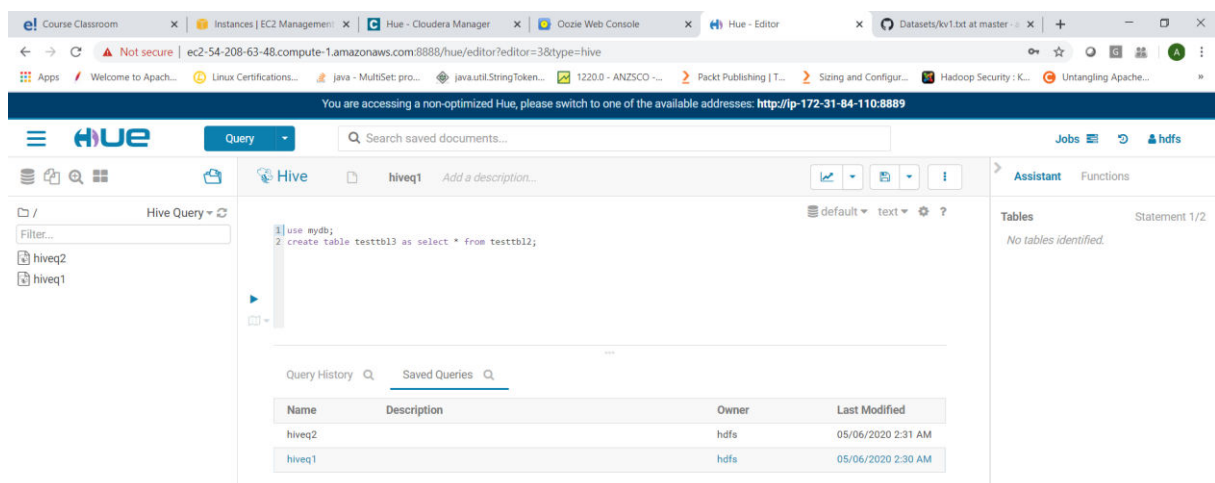
Now load data into this table by using Hue and uploading file into table directory i.e. /user/hive/warehouse/mydb.db/testtbl2/

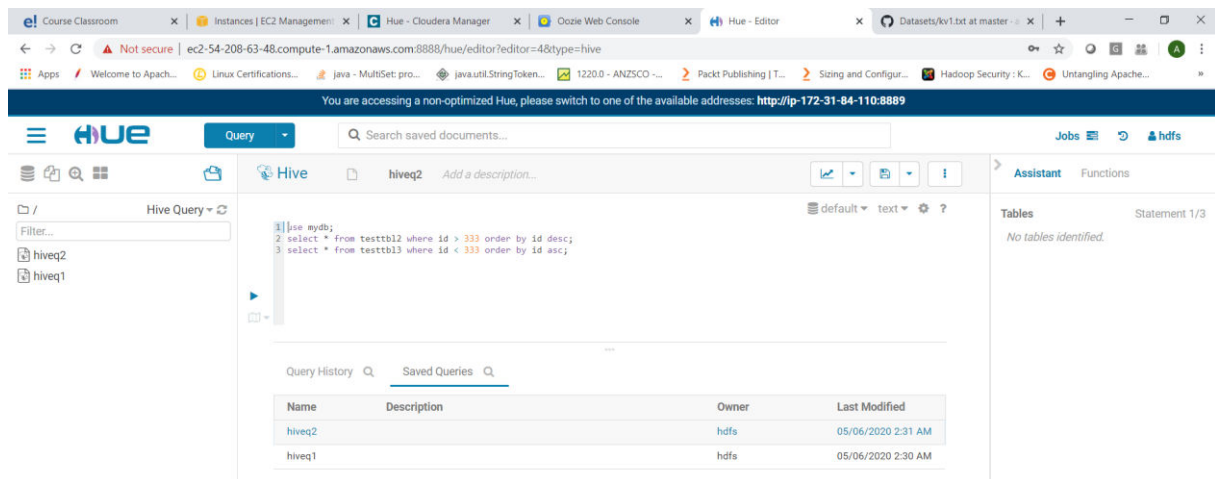
File can be found here: <https://github.com/ajaykuma/Datasets/blob/master/kv1.txt>

Step 2:

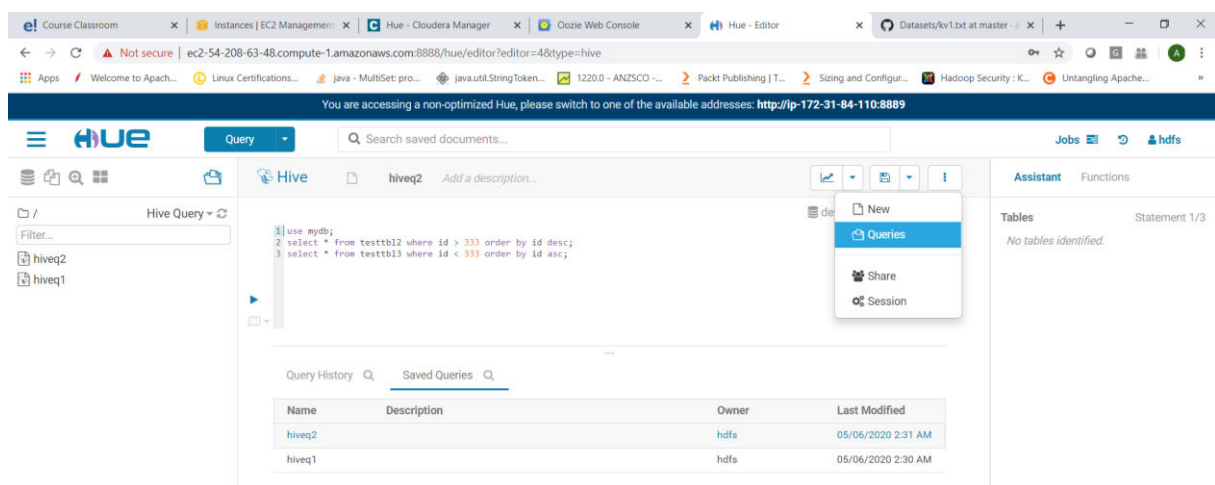
Using Hue create queries and save them.

Query > Editor > Hive

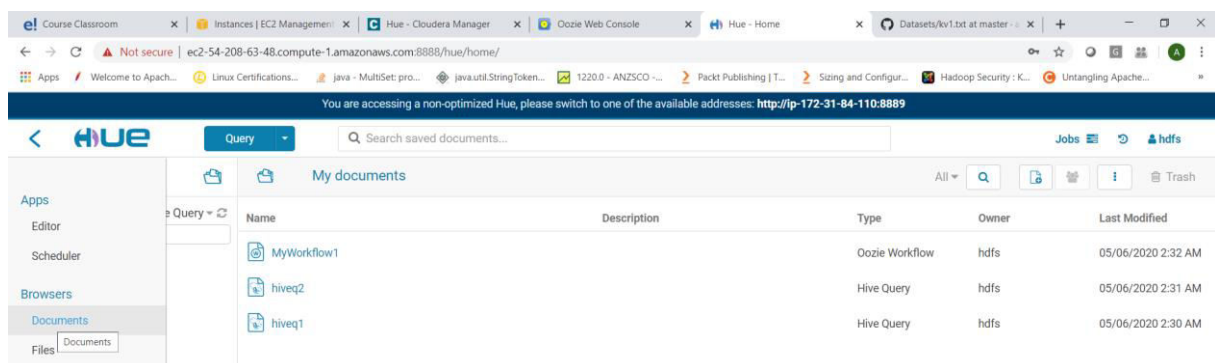




You can find your saved queries in



You can also find queries, workflows etc in

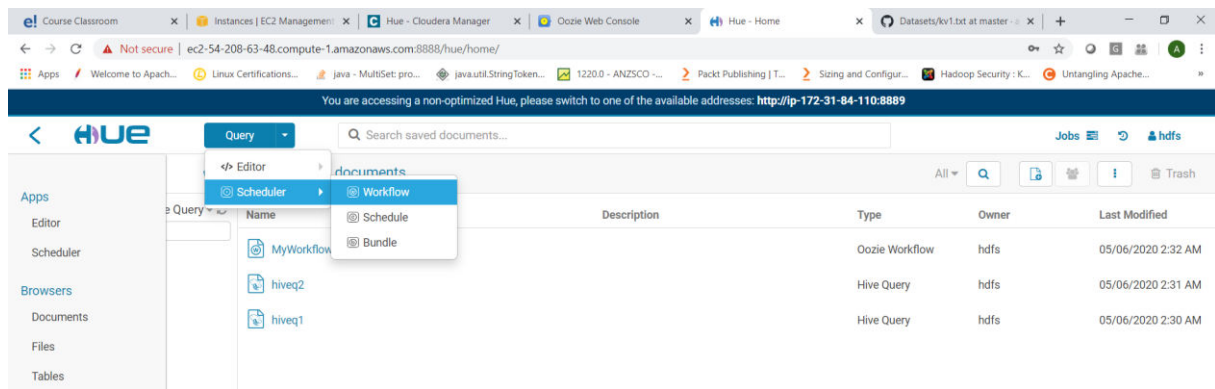


Query> editor> scheduler>workflow

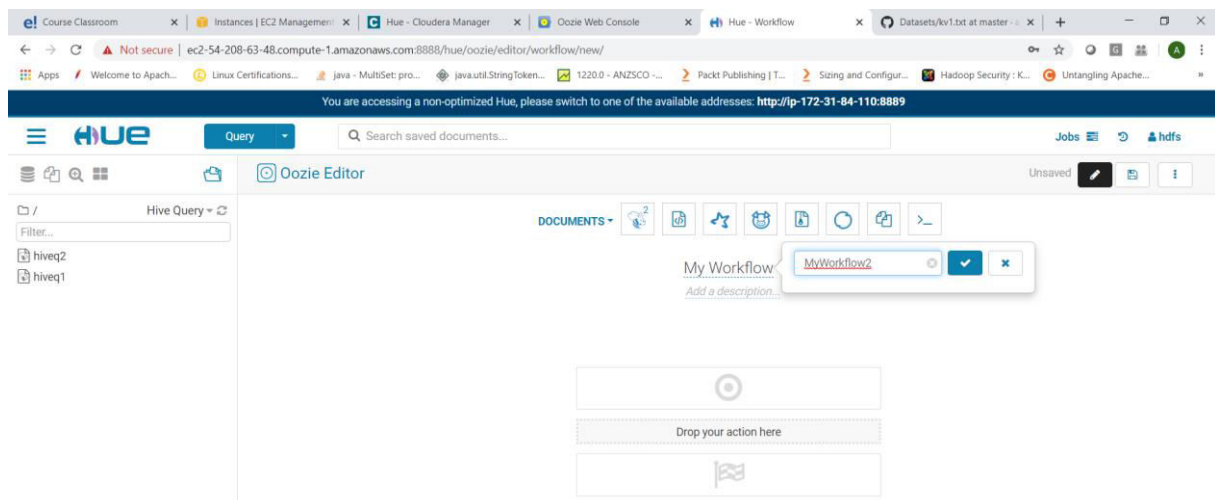
We will create a simple workflow which runs multiple hive queries.

https://oozie.apache.org/docs/4.2.0/DG_Hive2ActionExtension.html

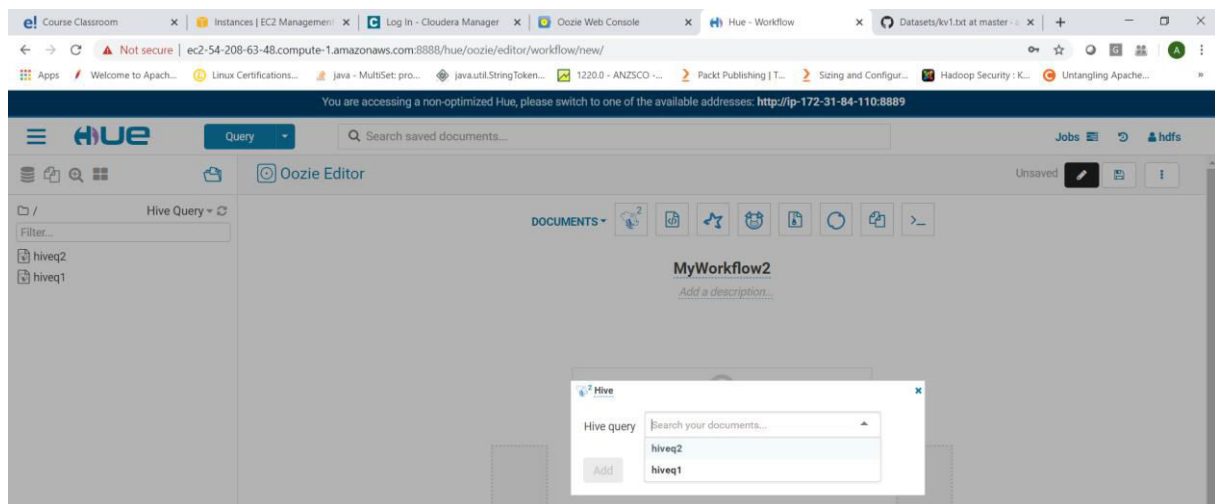
Workflow1



Edit and give a name to your workflow

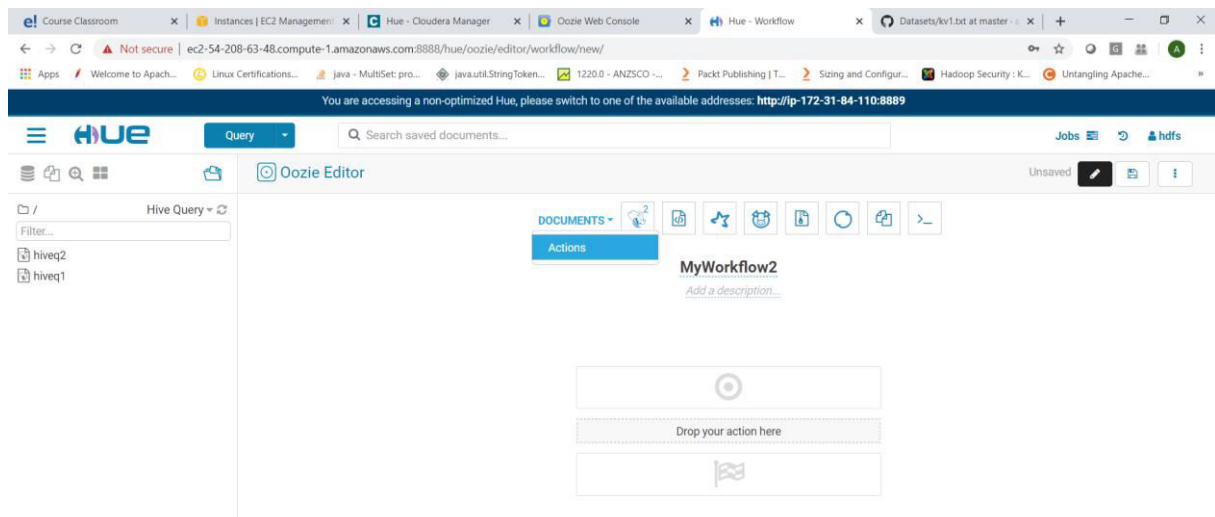


Click on option that says HiveServer2 and drag and drop it in “Drag your action here”

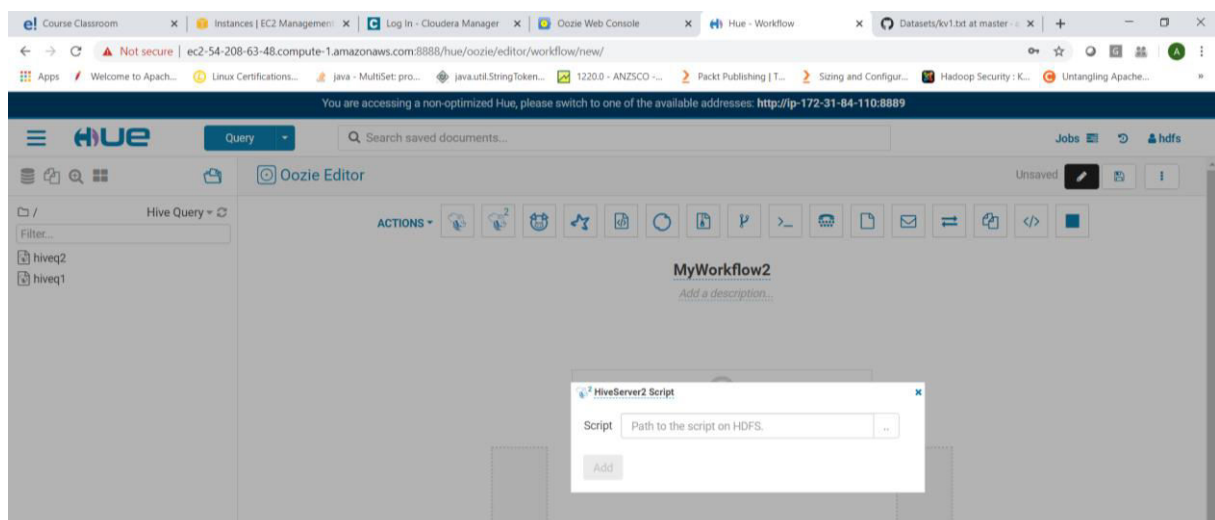
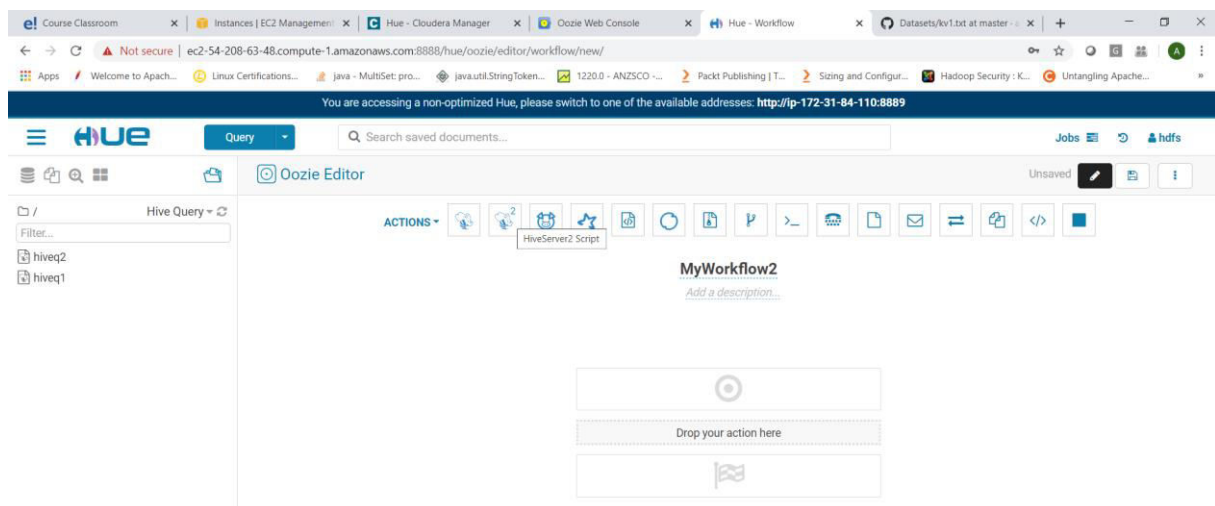


Choose your hiveq1 repeat the same for hiveq2

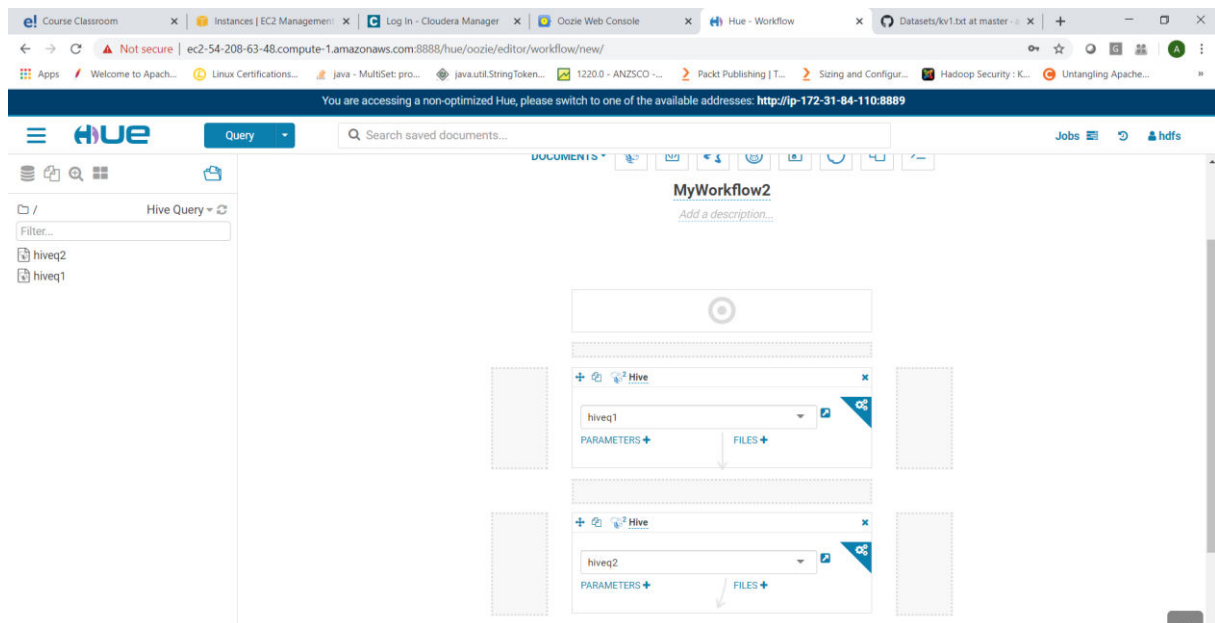
Optionally you expand options by clicking on actions as shown [This is to be used if you have saved your queries in some folder on HDFS] . If not skip next 3 screenshots.



Click on second option, ie HiveServer2 Script and drag it to box which says “Drop your action here”

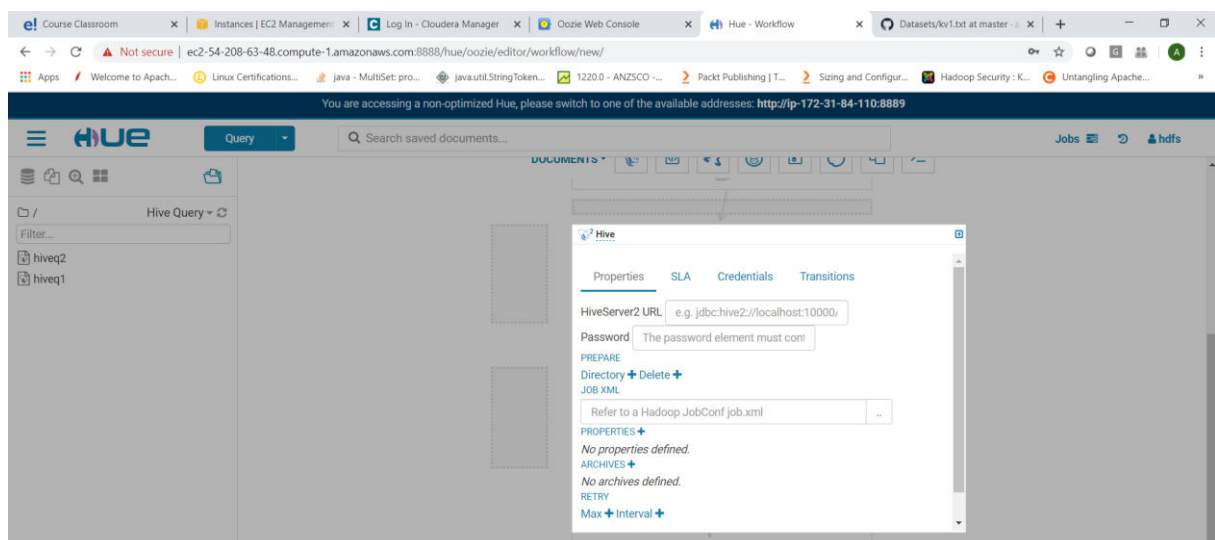


Now your workflow would look like this:



We can edit settings here or add files/parameters as needed.

If we click on settings it shows



Now your workflow is ready.

We can see the workflow by

Course Classroom | Instances | EC2 Management | Log In - Cloudera Manager | Oozie Web Console | Hue - Workflow | Datasets/kv1.txt at master

Not secure | ec2-54-208-63-48.compute-1.amazonaws.com:8888/hue/oozie/editor/workflow/edit/?workflow=5

You are accessing a non-optimized Hue, please switch to one of the available addresses: <http://p-172-31-84-110:8889>

HUE Query Search saved documents...

Oozie Editor

MyWorkflow1

Workflow diagram showing a sequence of tasks: Start, Hive (hiveq1), Hive (hiveq2), and End.

Jobs hdfs

Workflow Settings menu: New, Workflows, Schedule, Settings, Workspace, Share

Course Classroom | Instances | EC2 Management | Log In - Cloudera Manager | Oozie Web Console | Hue - Workflow | Datasets/kv1.txt at master

Not secure | ec2-54-208-63-48.compute-1.amazonaws.com:8888/hue/oozie/editor/workflow/edit/?workflow=5

You are accessing a non-optimized Hue, please switch to one of the available addresses: <http://p-172-31-84-110:8889>

HUE Query Search saved documents...

Oozie Editor

MyWorkflow1

Workflow Settings dialog:

- Variables: `oozie.use.system.libpath` = true
- + Add parameter
- Workspace: hue-oozie-1588757490.7
- Hadoop Properties: + Add property
- Show graph arrows: ☒
- Version

Course Classroom | Instances | EC2 Management | Log In - Cloudera Manager | Oozie Web Console | Hue - File Browser | Datasets/kv1.txt at master

Not secure | ec2-54-208-63-48.compute-1.amazonaws.com:8888/hue/filebrowser/view=/user/hue/oozie/workspaces/hue-oozie-1588757490.7

You are accessing a non-optimized Hue, please switch to one of the available addresses: <http://p-172-31-84-110:8889>

HUE Query Search saved documents...

File Browser

Search for file name Actions Move to trash Upload New

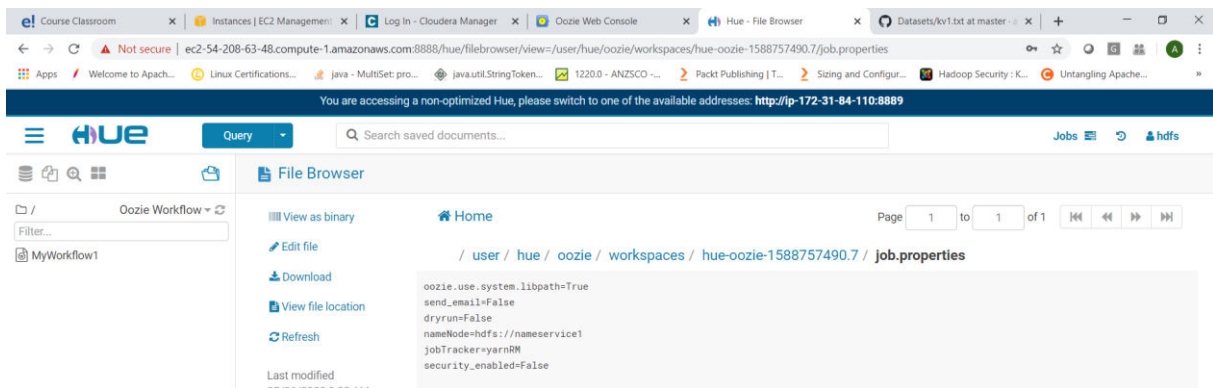
Home / user / hue / oozie / workspaces / hue-oozie-1588757490.7 Trash

Name	Size	User	Group	Permissions	Date
.		hue	hue	drwxrwxrwt	May 06, 2020 02:45 AM
hive-6s11.sql	134 bytes	hdfs	hue	-rw-r--r--	May 06, 2020 02:32 AM
hive-8bdc.sql	72 bytes	hdfs	hue	-rw-r--r--	May 06, 2020 02:32 AM
job.properties	129 bytes	hdfs	hue	-rw-r--r--	May 06, 2020 02:32 AM
lib		hdfs	hue	drwxr-xr-x	May 06, 2020 02:31 AM
workflow.xml	1.1 KB	hdfs	hue	-rw-r--r--	May 06, 2020 02:32 AM

Show 45 of 5 items Page 1 of 1

This shows our queries that we added to workflow.

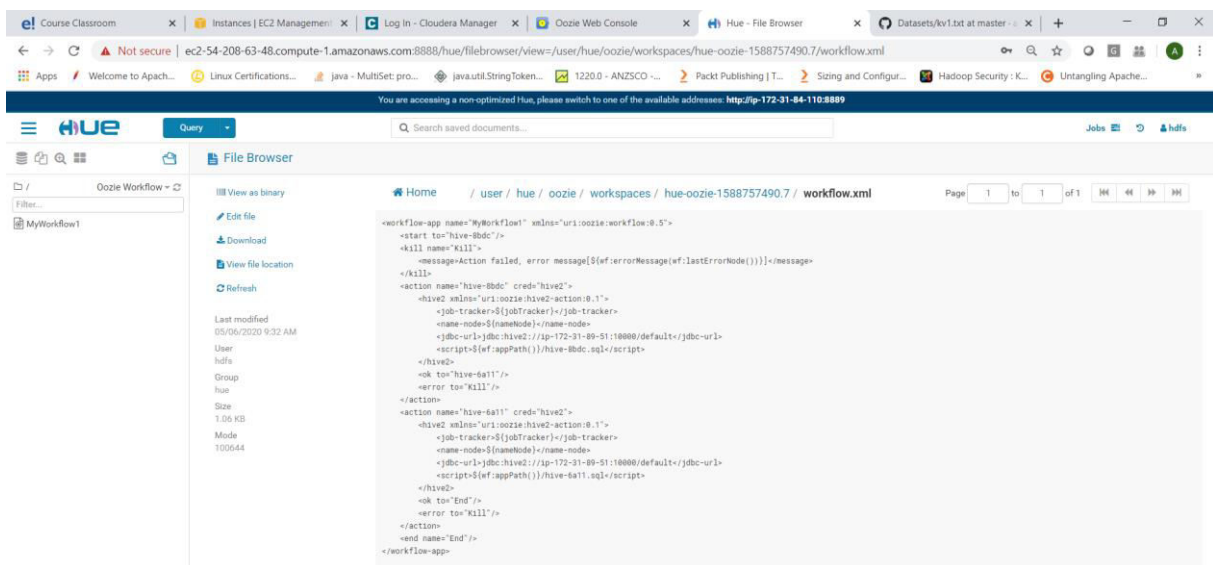
job.properties shows



The screenshot shows the Hue File Browser interface. The left sidebar displays the file tree with 'MyWorkflow1' selected. The main pane shows the 'job.properties' file. The file content is as follows:

```
oozie.use.system.libpath=True
send_email=False
dryrun=False
nameNode=hdfs://nameservice1
jobTracker=yarnRM
security_enabled=False
```

workflow.xml

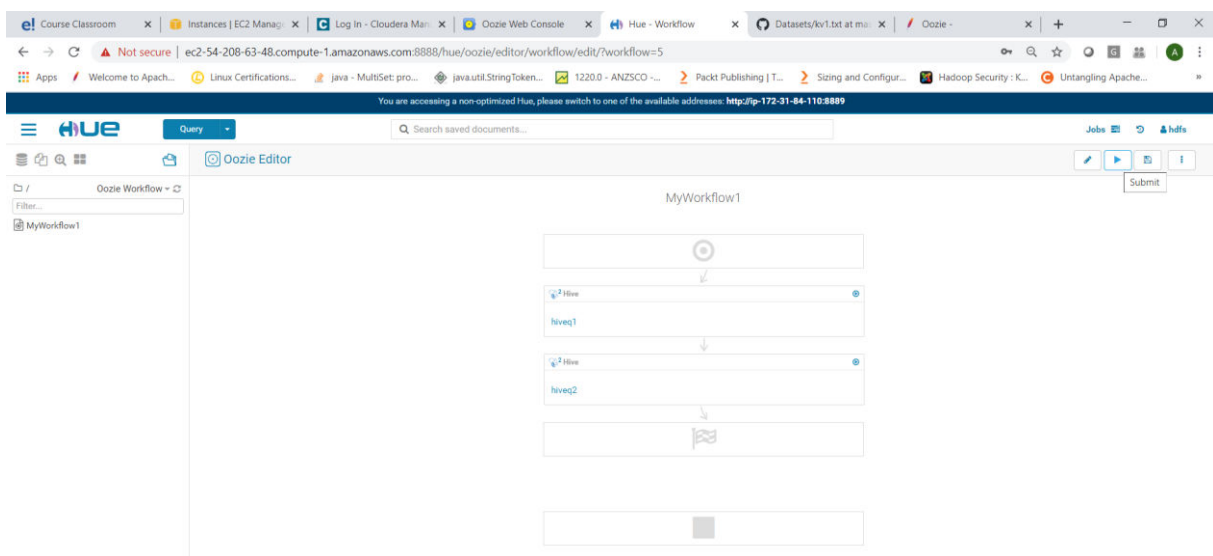


The screenshot shows the Hue File Browser interface with the 'workflow.xml' file selected. The file content is as follows:

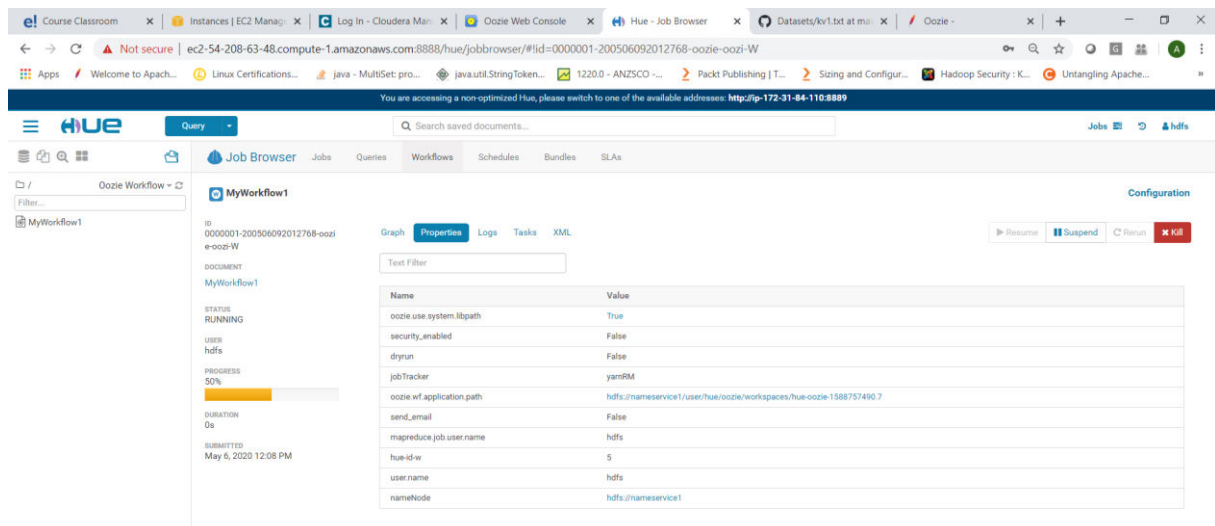
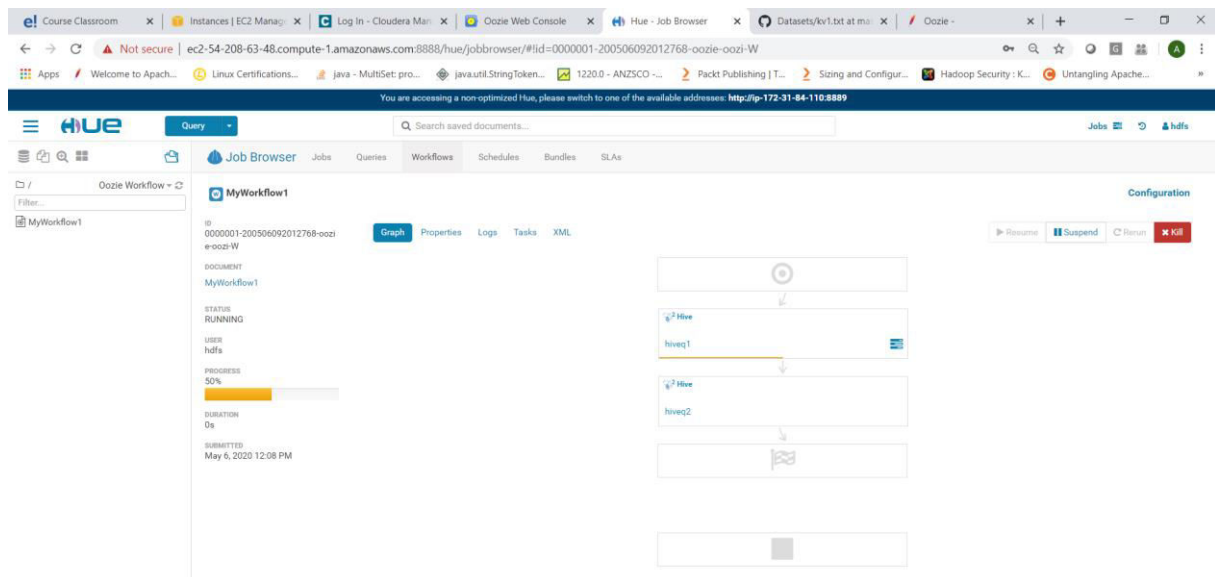
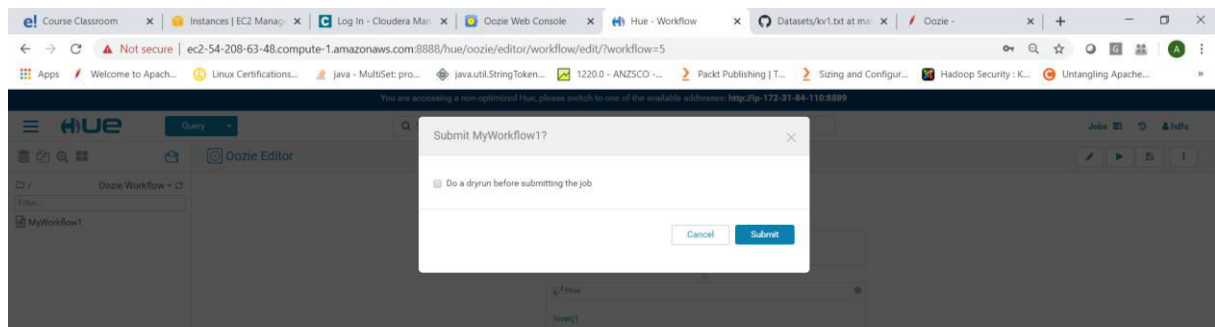
```
<workflow-app name="MyWorkflow1" xmlns="uri:oozie:workflow:0.5">
  <start to="hive-00dc"/>
  <kill name="Kill">
    <message>Action failed, error message[${error:errorMessage}]/></message>
  </kill>
  <action name="hive-00dc" cred="hive2">
    <hive2 xmlns="uri:oozie:hive2-action:0.1">
      <job-tracker[${jobTracker}]/></job-tracker>
      <name-node[${nameNode}]/></name-node>
      <jdbc-ur1-jdbc:hive2://ip-172-31-89-51:10000/default/></jdbc-ur1>
      <script[${appPath}]/hive-00dc.sql/></script>
    </hive2>
    <ok to="hive-0a11"/>
    <error to="Kill"/>
  </action>
  <action name="hive-0a11" cred="hive2">
    <hive2 xmlns="uri:oozie:hive2-action:0.1">
      <job-tracker[${jobTracker}]/></job-tracker>
      <name-node[${nameNode}]/></name-node>
      <jdbc-ur1-jdbc:hive2://ip-172-31-89-51:10000/default/></jdbc-ur1>
      <script[${appPath}]/hive-0a11.sql/></script>
    </hive2>
    <ok to="End"/>
    <error to="Kill"/>
  </action>
</workflow-app>
```

Refer <workflow.xml>

We can submit this workflow



The screenshot shows the Hue Oozie Editor interface. The workflow diagram for 'MyWorkflow1' is displayed. It starts with a 'Start' node, followed by a 'hive' node (labeled 'hiveq1'), then another 'hive' node (labeled 'hiveq2'), and finally an 'End' node. The workflow is shown in a visual flow diagram style.



Course Classroom | Instances | EC2 Manag... | Log In - Cloudera Man... | Oozie Web Console | Hue - Job Browser | Datasets/kv1.txt at m... | Oozie -

ec2-54-208-63-48.compute-1.amazonaws.com:8888/hue/jobbrowser/#lid=job_1588756771156_0008

You are accessing a non-optimized Hue, please switch to one of the available addresses: <http://ip-172-31-64-110-8889>

Query Search saved documents...

Jobs Queries Workflows Schedules Bundles SLAs

Filter... Oozie Workflow

MyWorkflow1

0000001-200506092012768-oozie-oozi-W > oozie:launcher:T-hive2:W=MyWorkflow1:A=hive-6a11-ID=0000001-200506092012768-oozie-oozi-W

job_1588756771156_0008

NAME oozie:launcher:T-hive2:W=MyW...

Log Tasks Metadata Counters

Filter by name Succeeded Running Failed Map Reduce

Type	Id	Elapsed Time	Progress	State	Start Time	Successful Attempt	Finish Time
MAP	task_1588756771156_000000	54.2s	100	SUCCEEDED	May 6, 2020 12:09 PM	attempt_1588756771156_000000_0	May 6, 2020 12:10 PM

STATUS SUCCEEDED

USER hdfs

PROGRESS 100%

MAP 0% 0 / 1

REDUCE 0% 0 / 0

DURATION 24s

SUBMITTED 05/06/20 03:09:28

Click on Task shows

Course Classroom | Instances | EC2 Manag... | Log In - Cloudera Man... | Oozie Web Console | Hue - Job Browser | Datasets/kv1.txt at m... | Oozie -

ec2-54-208-63-48.compute-1.amazonaws.com:8888/hue/jobbrowser/#lid=task_1588756771156_000000

You are accessing a non-optimized Hue, please switch to one of the available addresses: <http://ip-172-31-64-110-8889>

Query Search saved documents...

Jobs Queries Workflows Schedules Bundles SLAs

Filter... Oozie Workflow

MyWorkflow1

application_1588756771156_0008 > task_1588756771156_000000

task_1588756771156_000000

Log Attempts Counters

TYPE MAP

PROGRESS 100%

STATE SUCCEEDED

START TIME May 6, 2020 12:09 PM

SUCCESSFUL ATTEMPT attempt_1588756771156_0000...

FINISH TIME May 6, 2020 12:10 PM

ELAPSED TIME 54.2s

2020-05-06 10:10:25,708 INFO [Thread-90] org.apache.hadoop.mapreduce.v2.app.rm.RMContainerAllocator: Setting job diagnostics to

2020-05-06 10:10:25,731 INFO [Thread-90] org.apache.hadoop.mapreduce.v2.app.rm.RMContainerAllocator: History url is http://ip-172-31-64-110-8889/jobhistory/job/job_1588756771156_0008

2020-05-06 10:10:25,737 INFO [Thread-90] org.apache.hadoop.mapreduce.v2.app.rm.RMContainerAllocator: Waiting for application to be successfully unregistered.

2020-05-06 10:10:26,738 INFO [Thread-90] org.apache.hadoop.mapreduce.v2.app.rm.RMContainerAllocator: Final Status: PendingBeds:0 ScheduledBeds:0 AssignedMaps:1 AssignedReds:0

2020-05-06 10:10:26,739 INFO [Thread-90] org.apache.hadoop.mapreduce.v2.app.RMAppMaster: Deleting staging directory hdfs://nameservice1/user/hdfs/.staging/job_1588756771156_0008

2020-05-06 10:10:26,749 INFO [Thread-90] org.apache.hadoop.ipc.Server: Stopping server on 39396

2020-05-06 10:10:26,752 INFO [Thread-90] org.apache.hadoop.ipc.Server: Stopping IPC Server listener on 39396

2020-05-06 10:10:26,753 INFO [Thread-90] org.apache.hadoop.ipc.Server: Stopping IPC Server responder

2020-05-06 10:10:26,753 INFO [Thread-90] org.apache.hadoop.mapreduce.v2.app.TaskHeartbeatHandler: TaskHeartbeatHandler thread interrupted

2020-05-06 10:10:26,753 INFO [Thread-90] org.apache.hadoop.mapreduce.v2.app.TaskHeartbeatHandler: PingChecker thread interrupted

2020-05-06 10:10:26,754 INFO [Thread-90] org.apache.hadoop.mapreduce.v2.app.RMAppMaster: Job end notification started for jobID: job_1588756771156_0008

2020-05-06 10:10:26,755 INFO [Thread-90] org.mortbay.Log: Job end notification attempts left 0

2020-05-06 10:10:26,755 INFO [Thread-90] org.mortbay.Log: Job end notification trying http://ip-172-31-64-110-8889/oozie/calback?ip=0000001-200506092012768-oozie-oozi-W:hive-6a11&status=SUCCEEDED succeeded

2020-05-06 10:10:26,767 INFO [Thread-90] org.mortbay.Log: Job end notification to http://ip-172-31-64-110-8889/oozie/calback?ip=0000001-200506092012768-oozie-oozi-W:hive-6a11&status=SUCCEEDED succeeded

2020-05-06 10:10:26,767 INFO [Thread-90] org.mortbay.Log: Job end notification succeeded for job_1588756771156_0008

2020-05-06 10:10:31,767 INFO [Thread-90] org.apache.hadoop.ipc.Server: Stopping server on 35776

2020-05-06 10:10:31,768 INFO [Thread-90] org.apache.hadoop.ipc.Server: Stopping IPC Server listener on 35776

2020-05-06 10:10:31,768 INFO [Thread-90] org.apache.hadoop.ipc.Server: Stopping IPC Server responder

2020-05-06 10:10:31,778 INFO [Thread-90] org.mortbay.Log: Stopped HttpServer2SelectChannelConnectorWtSafeStartup@8.8.8.8

Course Classroom | Instances | EC2 Manag... | Log In - Cloudera Man... | Oozie Web Console | Hue - Job Browser | Datasets/kv1.txt at m... | Oozie -

ec2-54-208-63-48.compute-1.amazonaws.com:8888/hue/jobbrowser/#lid=task_1588756771156_000000

You are accessing a non-optimized Hue, please switch to one of the available addresses: <http://ip-172-31-64-110-8889>

Query Search saved documents...

Jobs Queries Workflows Schedules Bundles SLAs

Filter... Oozie Workflow

MyWorkflow1

application_1588756771156_0008 > task_1588756771156_000000

task_1588756771156_000000

Log Attempts Counters

Assigned Container Id	Progress	Elapsed Time	State	Rack	Node Http Address	Type	Start Time	Id	Finish Time
container_e09_1588756771156_000000_01_000002	100	54.2s	SUCCEEDED	/rack2	ip-172-31-64-110-8889	MAP_ATTEMPT	May 6, 2020 12:09 PM	attempt_1588756771156_000000_0	May 6, 2020 12:10 PM

TYPE MAP

PROGRESS 100%

STATE SUCCEEDED

START TIME May 6, 2020 12:09 PM

SUCCESSFUL ATTEMPT attempt_1588756771156_0000...

FINISH TIME May 6, 2020 12:10 PM

ELAPSED TIME 54.2s

Job Browser

Jobs Queries Workflows Schedules Bundles SLAs

0000001-200506092012768-oozie-oozi-W

oozie:launcher:T=hive2:W=MyWorkflow1:A=hive-6a11-ID=0000001-200506092012768-oozie-oozi-W

Job: 1588756771156_0008

NAME: oozie:launcher:T=hive2:W=MyW...

TYPE: MAPREDUCE

STATUS: SUCCEEDED

USER: hdfs

PROGRESS: 100%

MAP: 0% 0 / 1

REDUCE: 0% 0 / 0

DURATION: 24s

SUBMITTED: 05/06/20 03:09:28

Text Filter

Name	Value
dfs.namenode.read-lock-reporting.threshold.ms	5000
dfs.datanode.data.dir	file:///@hadoop.tmp.dir/dfs/data
mapred.output.committer.class	org.apache.oozie.action.hadoop.OozieLauncherOutputCommitter
dfs.namenode.checkpoint.time	1000000
s3.replication	3
mapreduce.output.fileoutputformat.compress.type	BLOCK
mapreduce.jobtracker.jobhistory.lru.cache.size	5
hadoop.http.filter.initializers	org.apache.hadoop.yarn.server.webproxy.amfilter.AmFilterInitializer
mapreduce.reduce.shuffle.memory.limit.percent	0.25
yarn.nodemanager.keytab	/etc/krb5.keytab
fs.s3a.readahead.range	64K
dfs.data.transfer.server.tcpnodelay	true
nfs.mountd.port	4242
dfs.namenode.top.num.users	10

Looking at all workflows

Job Browser

Jobs Queries Workflows Schedules Bundles SLAs

user:hdfs

Showing 1 to 2 of 2

Id	Name	User	Type	Status	Progress	Group	Started	Duration
0000001-200506092012768-oozie-oozi-W	MyWorkflow1	hdfs	workflow	SUCCEEDED	100%		May 6, 2020 12:08 PM	2m, 20s
0000000-200506092012768-oozie-oozi-W	MyWorkflow1	hdfs	workflow	SUCCEEDED	100%		May 6, 2020 11:32 AM	1m, 46s

From oozie console, check Workflow Jobs> Done Jobs . Optionally you can also explore other tabs.

Oozie Web Console

Workflow Jobs Coordinator Jobs Bundle Jobs SLA System Info Instrumentation Metrics Settings

All Jobs Active Jobs Done Jobs Custom Filter

Server version [4.1.0-cdh5.16.2]

Job ID	Name	Status	R...	User	Group	Created	Started	Last Modified	Ended
0000001-200506092012768-oozie-oozi-W	MyWorkflow1	SUCCE...	0	hdfs		Wed, 06 May 2020 10:08:08	Wed, 06 May 2020 10:08:08	Wed, 06 May 2020 10:10:28	Wed, 06 May 2020 10:10:28
0000000-200506092012768-oozie-oozi-W	MyWorkflow1	SUCCE...	0	hdfs		Wed, 06 May 2020 09:32:44	Wed, 06 May 2020 09:32:44	Wed, 06 May 2020 09:34:30	Wed, 06 May 2020 09:34:30

Double click on Job ID

Course Classroom | Instances | EC2 Manag... | Log In - Cloudera Man... | Oozie Web Console | Hue - Job Browser | Datasets/Kv1.txt at m... | Oozie -

Not secure | 52.207.213.42:11000/oozie/

Apps | Welcome to Apache... | Linux Certifications... | java - MultiSet: pro... | java.util.StringToken... | 1220.0 - ANZSCO... | Packt Publishing | T... | Sizing and Configur... | Hadoop Security : K... | Untangling Apache...

Job (Name: MyWorkflow1/JobId: 0000001-200506092012768-oozie-oozi-W)

Job Info | Job Definition | Job Configuration | Job Log | Job DAG

Job Id: 0000001-200506092012768-oozie-oozi-W
Name: MyWorkflow1
App Path: hdfs://nameservice1/user/hue/oozie/workspaces/hue-oozie-1588757490.7
Run: 0
Status: SUCCEEDED
User: hdfs
Group:
Parent Coord:
Create Time: Wed, 06 May 2020 10:08:08 GMT
Start Time: Wed, 06 May 2020 10:08:08 GMT
Last Modified: Wed, 06 May 2020 10:10:28 GMT
End Time: Wed, 06 May 2020 10:10:28 GMT

Server version [4.1.0-cdh5.16.2]

Ended
Wed, 06 May 2020 10:10:28 ...
Wed, 06 May 2020 09:34:30 ...

Actions

Action Id	Name	Type	Status	Transition	StartTime	EndTime
1 0000001-200506092012768-oozie-oozi-W@start	start	START	OK	hive-8bdc	Wed, 06 May 2020 10:08:09 ...	Wed, 06 May 2020 10:08:09 ...
2 0000001-200506092012768-oozie-oozi-W@hive-8bdc	hive-8bdc	hive2	OK	hive-6a11	Wed, 06 May 2020 10:08:09 ...	Wed, 06 May 2020 10:09:22 ...
3 0000001-200506092012768-oozie-oozi-W@hive-6a11	hive-6a11	hive2	OK	End	Wed, 06 May 2020 10:09:22 ...	Wed, 06 May 2020 10:10:27 ...
4 0000001-200506092012768-oozie-oozi-W@End	End	END	OK		Wed, 06 May 2020 10:10:27 ...	Wed, 06 May 2020 10:10:27 ...

Explore other tabs

Course Classroom | Instances | EC2 Manag... | Log In - Cloudera Man... | Oozie Web Console | Hue - Job Browser | Datasets/Kv1.txt at m... | Oozie -

Not secure | 52.207.213.42:11000/oozie/

Apps | Welcome to Apache... | Linux Certifications... | java - MultiSet: pro... | java.util.StringToken... | 1220.0 - ANZSCO... | Packt Publishing | T... | Sizing and Configur... | Hadoop Security : K... | Untangling Apache...

Job (Name: MyWorkflow1/JobId: 0000001-200506092012768-oozie-oozi-W)

Job Info | Job Definition | Job Configuration | Job Log | Job DAG

```
<workflow-app name="MyWorkflow1" xmlns="un:oozie:workflow:0.5">
  <start to="hive-8bdc"/>
  <kill name="kill"/>
  <message>Action failed, error message[${wf.errorMessage(wf.lastErrorNode)}]</message>
  </kill>
  <action name="hive-8bdc" cred="hive2">
    <hive2 xmlns="un:oozie:hive2-action:0.1">
      <job-tracker>${jobTracker}</job-tracker>
      <name-node>${nameNode}</name-node>
      <jdbc-url>jdbc:hive2://ip-172-31-89-51:10000/default</jdbc-ur<
      <script>${wf.appPath()}/hive-8bdc.sql</script>
    </hive2>
    <ok to="hive-6a11"/>
    <error to="kill"/>
  </action>
  <action name="hive-6a11" cred="hive2">
    <hive2 xmlns="un:oozie:hive2-action:0.1">
      <job-tracker>${jobTracker}</job-tracker>
      <name-node>${nameNode}</name-node>
      <jdbc-ur<
      <script>${wf.appPath()}/hive-6a11.sql</script>
    </hive2>
    <ok to="End"/>
    <error to="kill"/>
  </action>
  <end name="End"/>
</workflow-app>
```

Server version [4.1.0-cdh5.16.2]

Ended
Wed, 06 May 2020 10:10:28 ...
Wed, 06 May 2020 09:34:30 ...

Action Id	Name	Type	Status	Transition	StartTime	EndTime
1 0000001-200506092012768-oozie-oozi-W@start	start	START	OK	hive-8bdc	Wed, 06 May 2020 10:08:09 ...	Wed, 06 May 2020 10:08:09 ...
2 0000001-200506092012768-oozie-oozi-W@hive-8bdc	hive-8bdc	hive2	OK	hive-6a11	Wed, 06 May 2020 10:08:09 ...	Wed, 06 May 2020 10:09:22 ...
3 0000001-200506092012768-oozie-oozi-W@hive-6a11	hive-6a11	hive2	OK	End	Wed, 06 May 2020 10:09:22 ...	Wed, 06 May 2020 10:10:27 ...
4 0000001-200506092012768-oozie-oozi-W@End	End	END	OK		Wed, 06 May 2020 10:10:27 ...	Wed, 06 May 2020 10:10:27 ...

Course Classroom | Instances | EC2 Manag... | Log In - Cloudera Man... | Oozie Web Console | Hue - Job Browser | Datasets/Kv1.txt at m... | Oozie -

Not secure | 52.207.213.42:11000/oozie/

Apps | Welcome to Apache... | Linux Certifications... | java - MultiSet: pro... | java.util.StringToken... | 1220.0 - ANZSCO... | Packt Publishing | T... | Sizing and Configur... | Hadoop Security : K... | Untangling Apache...

Job (Name: MyWorkflow1/JobId: 0000001-200506092012768-oozie-oozi-W)

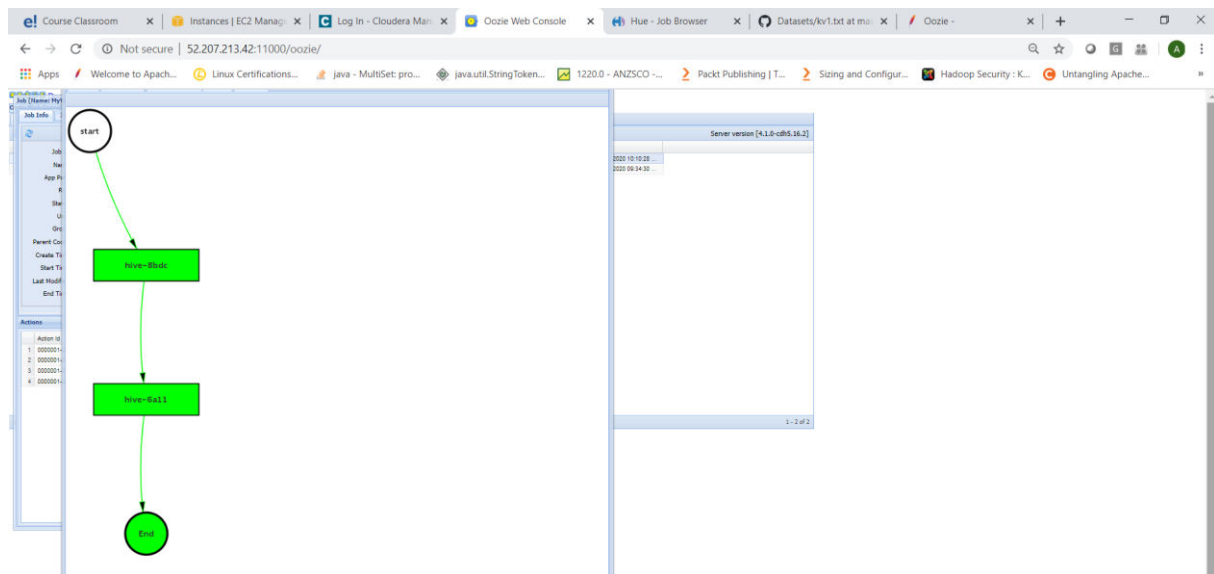
Job Info | Job Definition | Job Configuration | Job Log | Job DAG

```
<configuration>
  <property>
    <name>hue-id-w</name>
    <value>5</value>
  </property>
  <property>
    <name>user.name</name>
    <value>hdfs</value>
  </property>
  <property>
    <name>oozie.use.system.libpath</name>
    <value>True</value>
  </property>
  <property>
    <name>mapreduce.job.user.name</name>
    <value>hdfs</value>
  </property>
  <property>
    <name>security_enabled</name>
    <value>False</value>
  </property>
  <property>
    <name>nameNode</name>
    <value>hdfs://nameservice1</value>
  </property>
  <property>
    <name>dryrun</name>
    <value>False</value>
  </property>
</configuration>
```

Server version [4.1.0-cdh5.16.2]

Ended
Wed, 06 May 2020 10:10:28 ...
Wed, 06 May 2020 09:34:30 ...

Action Id	Name	Type	Status	Transition	StartTime	EndTime
2 0000001-200506092012768-oozie-oozi-W@hive-8bdc	hive-8bdc	hive2	OK	hive-6a11	Wed, 06 May 2020 10:08:09 ...	Wed, 06 May 2020 10:09:22 ...
3 0000001-200506092012768-oozie-oozi-W@hive-6a11	hive-6a11	hive2	OK	End	Wed, 06 May 2020 10:09:22 ...	Wed, 06 May 2020 10:10:27 ...
4 0000001-200506092012768-oozie-oozi-W@End	End	END	OK		Wed, 06 May 2020 10:10:27 ...	Wed, 06 May 2020 10:10:27 ...



Now check from Hive CLI, as a result of your query, you should new table

```
hive> show tables;
OK
testtbl1
testtbl2
Time taken: 0.195 seconds, Fetched: 2 row(s)
hive> show tables;
OK
testtbl1
testtbl2
testtbl3
Time taken: 0.018 seconds, Fetched: 3 row(s)
hive> drop table testtbl3;
OK
Time taken: 0.242 seconds
hive> show tables;
OK
testtbl1
testtbl2
testtbl3
Time taken: 0.017 seconds, Fetched: 3 row(s)
hive>
```

We can run this workflow from command line too.

Workflow1-Command line

```
hdfs@ip-172-31-88-2:~$ hdfs dfs -ls /
Found 6 items
-rw-r--r-- 2 hdfs supergroup 1054212 2020-05-04 08:00 /auction.csv
drwxr-xr-x - hbase hbase 0 2020-05-06 09:19 /hbase
drwxr-xr-x - hdfs supergroup 0 2020-05-04 06:35 /mydata
drwxr-xr-x - hdfs supergroup 0 2020-05-04 06:51 /system
drwxrwxrwt - hdfs supergroup 0 2020-05-04 06:27 /tmp
drwxr-xr-x - hdfs supergroup 0 2020-05-04 06:31 /user
hdfs@ip-172-31-88-2:~$ hdfs dfs -ls /user/hdfs
Found 4 items
drwx----- - hdfs hdfs 0 2020-05-06 10:07 /user/hdfs/.Trash
drwx----- - hdfs hdfs 0 2020-05-06 10:10 /user/hdfs/.staging
drwxr-xr-x - hdfs hdfs 0 2020-05-06 09:34 /user/hdfs/hue-oozie-1588757490.7
drwxr-xr-x - hdfs hdfs 0 2020-05-06 10:10 /user/hdfs/oozie-oozi
hdfs@ip-172-31-88-2:~$ hdfs dfs -get /user/hdfs/hue-oozie-1588757490.7 .
hdfs@ip-172-31-88-2:~$ mv hue-oozie-1588757490.7 workflow1
hdfs@ip-172-31-88-2:~$ ls
workflow1
hdfs@ip-172-31-88-2:~$ ls workflow1/
hive-6all.sql hive-8bdc.sql job.properties lib workflow.xml
hdfs@ip-172-31-88-2:~$
```

```
hdfs@ip-172-31-88-2:~$ logout
root@ip-172-31-88-2:~# su - hdfs
hdfs@ip-172-31-88-2:~$ oozie admin -oozie http://172.31.88.2:11000/oozie -status
System mode: NORMAL
hdfs@ip-172-31-88-2:~$ oozie admin -oozie http://172.31.88.2:11000/oozie --version
Oozie server build version: 4.1.0-cdh5.16.2
hdfs@ip-172-31-88-2:~$ export OOOIE URL="http://172.31.88.2:11000/oozie"
hdfs@ip-172-31-88-2:~$ oozie admin -status
System mode: NORMAL
hdfs@ip-172-31-88-2:~$
```



```
hdfs@ip-172-31-88-2: ~/workflow$ ls lib
hdfs@ip-172-31-88-2: ~/workflow$ vi job.properties
hdfs@ip-172-31-88-2: ~/workflow$ vi workflow.xml
hdfs@ip-172-31-88-2: ~/workflow$ cd ..
hdfs@ip-172-31-88-2: ~$ ls workflow/
hive-6all.sql  hive-8bdc.sql  job.properties  lib  workflow.xml
hdfs@ip-172-31-88-2: ~$ hdfs dfs -mkdir /Workflow
hdfs@ip-172-31-88-2: ~$ hdfs dfs -put workflow/*.sql /Workflow
hdfs@ip-172-31-88-2: ~$ hdfs dfs -put workflow/job.properties /Workflow
hdfs@ip-172-31-88-2: ~$ hdfs dfs -put workflow/lib /Workflow/
hdfs@ip-172-31-88-2: ~$ hdfs dfs -put workflow/workflow.xml /Workflow
hdfs@ip-172-31-88-2: ~$ hdfs dfs -ls /Workflow
Found 5 items
-rw-r--r--  2 hdfs supergroup      134 2020-05-06 11:05 /Workflow/hive-6all.sql
-rw-r--r--  2 hdfs supergroup       72 2020-05-06 11:05 /Workflow/hive-8bdc.sql
-rw-r--r--  2 hdfs supergroup      177 2020-05-06 11:05 /Workflow/job.properties
drwxr-xr-x  - hdfs supergroup       0 2020-05-06 11:06 /Workflow/lib
-rw-r--r--  2 hdfs supergroup     1088 2020-05-06 11:06 /Workflow/workflow.xml
hdfs@ip-172-31-88-2: ~$ ls workflow/
hive-6all.sql  hive-8bdc.sql  job.properties  lib  workflow.xml
hdfs@ip-172-31-88-2: ~$ vi workflow/job.properties
hdfs@ip-172-31-88-2: ~$ hdfs dfs -ls /Workflow
Found 5 items
-rw-r--r--  2 hdfs supergroup      134 2020-05-06 11:05 /Workflow/hive-6all.sql
-rw-r--r--  2 hdfs supergroup       72 2020-05-06 11:05 /Workflow/hive-8bdc.sql
-rw-r--r--  2 hdfs supergroup      177 2020-05-06 11:05 /Workflow/job.properties
drwxr-xr-x  - hdfs supergroup       0 2020-05-06 11:06 /Workflow/lib
-rw-r--r--  2 hdfs supergroup     1088 2020-05-06 11:06 /Workflow/workflow.xml
hdfs@ip-172-31-88-2: ~$ oozie job -config workflow/job.properties -run
job: 00000002-200506092012768-oozie-oozi-W
hdfs@ip-172-31-88-2: ~$
```

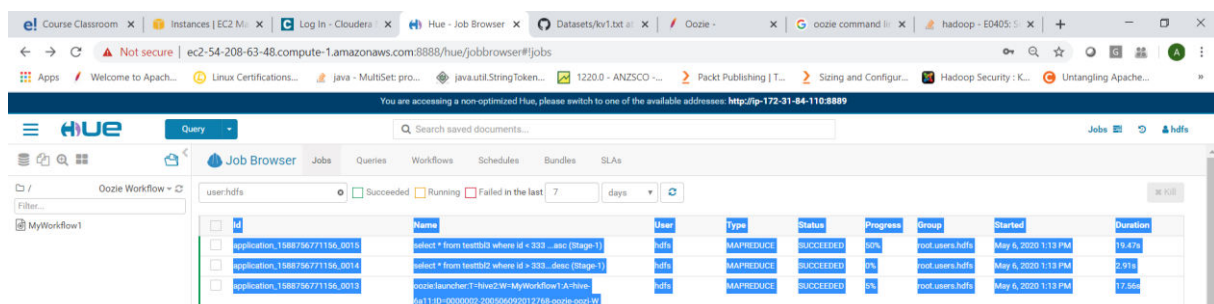
Job.properties contains

```
oozie.use.system.libpath=true
send_email=false
dryrun=false
nameNode=hdfs://nameservice1
jobTracker=jarn8M
security_enabled=false
oozie.wf.application.path=${nameNode}/Workflow
~
```

Workflow.xml contains

```
<workflow-app name="MyWorkflow1" xmlns="uri:oozie:workflow:0.5">
  <start to="hive-8bdc"/>
  <kill name="Kill1">
    <message>Action failed, error message[${wf:errorMessage(wf:lastErrorNode())}]</message>
  </kill>
  <action name="hive-8bdc" cred="hive2">
    <hive2 xmlns="uri:oozie:hive2-action:0.1">
      <job-tracker>${jobTracker}</job-tracker>
      <name-node>${nameNode}</name-node>
      <jdbc-url>jdbc:hive2://ip-172-31-89-51:10000/default</jdbc-url>
      <script>${wf:appPath()} /hive-8bdc.sql</script>
    </hive2>
    <ok to="hive-6all"/>
    <error to="Kill1"/>
  </action>
  <action name="hive-6all" cred="hive2">
    <hive2 xmlns="uri:oozie:hive2-action:0.1">
      <job-tracker>${jobTracker}</job-tracker>
      <name-node>${nameNode}</name-node>
      <jdbc-url>jdbc:hive2://ip-172-31-89-51:10000/default</jdbc-url>
      <script>${wf:appPath()} /hive-6all.sql</script>
    </hive2>
    <ok to="End"/>
    <error to="Kill1"/>
  </action>
  <end name="End"/>
</workflow-app>
```

Looking job run from hue



ID	Name	User	Type	Status	Progress	Group	Started	Duration
application_1588756771156_0013	select * from test03 where id < 233 ..._desc (Stage-1)	hdfs	MAPREDUCE	SUCCEEDED	100%	root.users.hdfs	May 6, 2020 1:13 PM	15.47s
application_1588756771156_0014	select * from test02 where id < 233 ..._desc (Stage-1)	hdfs	MAPREDUCE	SUCCEEDED	100%	root.users.hdfs	May 6, 2020 1:13 PM	2.51s
application_1588756771156_0013	oozie:launcher-T-hive2-W-MyWorkflow1-A-hive-6all1-00-0000002-200506092012768-oozie-oozi-W	hdfs	MAPREDUCE	SUCCEEDED	100%	root.users.hdfs	May 6, 2020 1:13 PM	17.55s

Finally verifying if hive new table was created


```
hdfs@ip-172-31-88-2:~$ hive
Logging initialized using configuration in jar:file:/opt/cloudera/parcels/CDH-5.16.2-1.cdh5.16.2.p0.8/jars/hive-common-1.1.0-cdh5.16.2.jar!/hive-log4j.properties
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive> use mydb;
OK
Time taken: 4.279 seconds
hive> show tables;
OK
testtbl1
testtbl2
testtbl3
Time taken: 0.28 seconds, Fetched: 3 row(s)
hive>
```

Workflow2

As previous example, we will create a workflow which will run a mapreduce job for wordcount.

Workflow3

We will create a workflow which will run a mapreduce job followed by a hive query.

Refer the files in folders.

https://github.com/ajaykuma/Hadoop_Spark_Setup/tree/master/Oozie_Setup