Get Docker image from :
https://blog.clairvoyantsoft.com/cloduera-quickstart-vm-using-docker-on-mac-2308acd196f2

After running the docker container, connect to docker container by following command

```
docker exec -it eb6d220784f6 /bin/bash
docker volume create my_volume
docker commit container_id my_custom_image
```

And run YARN resourcemanager by command :-

```
yarn resourcemanager
```

Access the yarn on http://localhost:8088/cluster

root:-

```
mkdir sde_project_files
mkdir sde_project_files/raw_DATA
mkdir sde_project_files/scripts
```

Local:-

```
cd Documents/SDE_project/scripts
```

Place all files one by one by commenting and uncommenting

```
./file_share.sh
```

Root:-

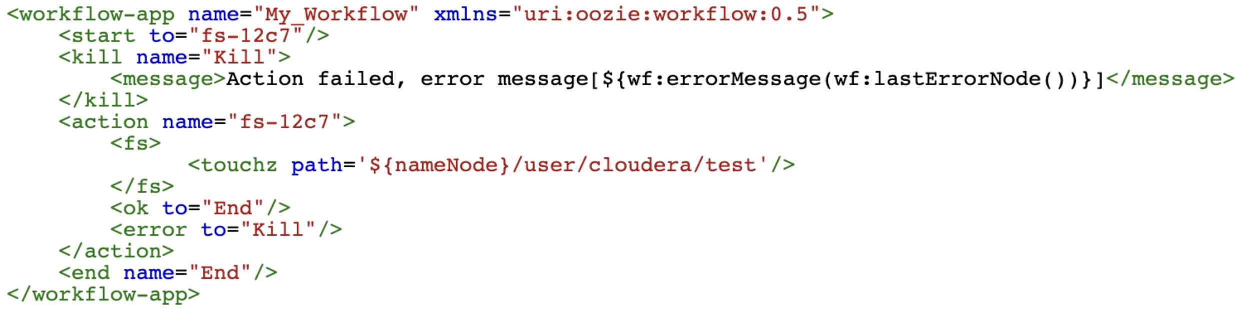
```
hdfs dfs -put /sde_project_files/* /user/cloudera/
hdfs dfs -put /sde_project_files/raw_DATA/sparkify_log_small.json /user/cloudera/raw/
hdfs dfs -put /sde_project_files/scripts/file_avail_check.sh /user/cloudera/scripts/
```

```
CREATE DATABASE sde_project;

USE sde_project;

-- Create a Hive table for the location data
CREATE TABLE IF NOT EXISTS sde_project.yellow_taxi_cab (
   LocationID INT,
   Borough STRING,
   Zone STRING,
   service_zone STRING
)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
STORED AS TEXTFILE
```

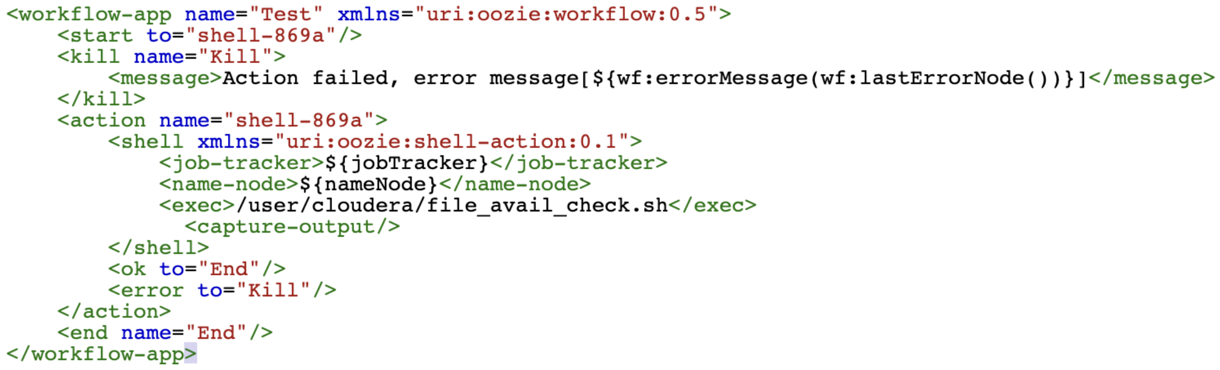
TBLPROPERTIES("skip.header.line.count"="1");

LOAD DATA INPATH 'hdfs:///user/cloudera/taxi_zone_lookup.csv' INTO TABLE sde_project.yellow_taxi_cab;

Creating workflow



```
<workflow-app name="My_Workflow" xmlns="uri:oozie:workflow:0.5">
    <start to="fs-12c7"/>
    <kill name="Kill">
        <message>Action failed, error message[${wf:errorMessage(wf:lastErrorNode())}]</message>
    </kill>
    <action name="fs-12c7">
        <fs>
              <touchz path='${nameNode}/user/cloudera/test'/>
        </fs>
        <ok to="End"/>
        <error to="Kill"/>
    </action>
    <end name="End"/>
</workflow-app>
```



```
<workflow-app name="Test" xmlns="uri:oozie:workflow:0.5">
    <start to="shell-869a"/>
    <kill name="Kill">
        <message>Action failed, error message[${wf:errorMessage(wf:lastErrorNode())}]</message>
    </kill>
    <action name="shell-869a">
        <shell xmlns="uri:oozie:shell-action:0.1">
            <job-tracker>${jobTracker}</job-tracker>
            <name-node>${nameNode}</name-node>
            <exec>/user/cloudera/file_avail_check.sh</exec>
              <capture-output/>
        </shell>
        <ok to="End"/>
        <error to="Kill"/>
    </action>
    <end name="End"/>
</workflow-app>
```

http://127.0.0.1:8888/oozie/list_oozie_workflows/

Checking Oozie workflow and status:-
oozie jobs -oozie http://localhost:11000/oozie -len 5 -jobtype coordinator
oozie jobs -oozie http://localhost:11000/oozie -len 5 -jobtype workflow
To get oozie info :-
oozie job -info <workflow_job_id>

To kill a oozie job :-
oozie job -kill 0000004-231120174206131-oozie-oozi-C -oozie http://localhost:11000/oozie

“File_share.sh” -

#!/bin/bash

```
# Replace these variables with your actual values
# CONTAINER_ID="f119ef7ff47d"

LOCAL_FILE_PATH1="/Users/bhawnabhoria/Documents/SDE_project/raw_DATA/sparkify_log_small.json"
LOCAL_FILE_PATH2="/Users/bhawnabhoria/Documents/SDE_project/scripts/file_avail_check.sh"

CONTAINER_ID=00668727f73b
CONTAINER_FILE_PATH1="/sde_project_files/raw_DATA/sparkify_log_small.json"
CONTAINER_FILE_PATH2="/sde_project_files/scripts/file_avail_check.sh"

# Copy the file to the Docker container
docker cp "$LOCAL_FILE_PATH1" "$CONTAINER_ID":"$CONTAINER_FILE_PATH1"
docker cp "$LOCAL_FILE_PATH2" "$CONTAINER_ID":"$CONTAINER_FILE_PATH2"
```

“File_availability_check .sh” :-

```
#!/bin/bash

# HDFS file path to check
hdfs_file_path="/user/cloudera/raw/sparkify_log_small.json"

# Check if the file exists in HDFS
hadoop fs -test -e $hdfs_file_path

# $? stores the exit status of the last command
if [ $? -eq 0 ]; then
   echo "File exists in HDFS: $hdfs_file_path"
   exit 0  # Exit with success status
else
   echo "File does not exist in HDFS: $hdfs_file_path"
   exit 1  # Exit with failure status
fi
```