**HANDS-ON ACTIVITY**

SCENARIO: COPY/IMPORT DATA FROM RDBMS DIRECT INTO HIVE WAREHOUSE. CREATE PARTITIONED TABLE USING DATA FROM NON-PARTITIONED TABLE.

DATASET: WE WILL USE MOVIELENS DATABASE (RDBMS)

STEP ONE: IMPORT/ TRANSFORM DATA FROM RDBMS (movielens) INTO HDFS THEN DATA WILL MOVE AUTOMATICALLY INTO HIVE WAREHOUSE DIRECTORY. AND THE INTERNAL/MANAGED TABLE WILL BE CREATE BY THE NAME OF 'non\_movies'

sqoop import --connect jdbc:mysql://localhost/movielens --driver com.mysql.jdbc.Driver --username root --password password --query "SELECT \* FROM movies WHERE \$CONDITIONS limit 100"

--split-by id

--target-dir /user/maria\_dev/test/ --hive-import

--hive-database default

--hive-table non\_movies

--create-hive-table

--fields-terminated-by ','

Goto Ambari -> Click on File View -> you can see imported data under hive warehouse directory "apps/hive/warehouse/movies"

Explanation : If we use --hive tag its means Data will be store in the hdfs directory then automatically move into HIVE Warehouse directory.

STEP 2: GOTO  HIVE VIEW. IN THE WORKSHEET RUN THE BELOW LINE. BELOW LINE  WILL ALTER TABLE PROPERTIES.

ALTER TABLE non\_movies SET TBLPROPERTIES ('EXTERNAL' = 'TRUE')

//we make our new table external because when we use sqoop default is internal table.

STEP 3: WE ARE GOING TO MOVE  DATA  FROM  NON-PARTITION TABLE INTO PARTITION TABLE. TO DO SO  WE WILL CREATE PARTITIONED TABLE.

BELOW IS THE HIVE QUERY WILL CREATE HIVE EXTERNAL TABLE WITH PARTITION PROPERTIES

SET hive.exec.dynamic.partition = true;

SET hive.exec.dynamic.partition.mode = nonstrict;

CREATE EXTERNAL TABLE parititon\_movies (

id INT,

title string,

)

PARTITIONED BY (release\_date string)

ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

LINES TERMINATED BY '\n'

STORED AS TEXTFILE

LOCATION "/user/maria\_dev/partition\_movies ";

STEP 4: NOW BELOW HIVE QUERY WILL COPY DATA FROM NON-PARTITION TABLE INTO PARTITION TABLE.

SET hive.exec.dynamic.partition = true;

SET hive.exec.dynamic.partition.mode = nonstrict;

INSERT OVERWRITE TABLE parititon\_movies

PARTITION (release\_date) select id, title, release\_date

FROM non\_movies;

STEP 5: BELOW QUERY WILL PRINT OUT TABLE PROPERTIES.

GOTO HIVE VIEW IN THE WORKSHEET TYPE DOWN BELOW QUERY

DESCRIBE formatted non\_movies;

DESCRIBE formatted parititon\_movies;