Here is a sample use case that can be run in hive.

<https://cwiki.apache.org/confluence/display/Hive/GettingStarted#GettingStarted-CreatingHiveTables>

Simple Example Use Cases

**MovieLens User Ratings**

Prior to starting this exercise Hadoop and Hive need to be running.

Since CentOS is in a manual install the zip utility has to be installed

Switch to the Root Unser

# sudo yum install unzip

Upload and unzip download the data files from MovieLens 100k on the [GroupLens datasets](http://grouplens.org/datasets/movielens/) page into hadoop using wet command.

wget <http://files.grouplens.org/datasets/movielens/ml-100k.zip>

unzip ml-100k.zip    -- unzips the file

hadoop fs -copyFromLocal ml-100k wk4

Log into hive

            $hive

Use this command to create a u\_data table with tab-delimited text file format.

CREATE TABLE u\_data (

  userid INT,

  movieid INT,

  rating INT,

  unixtime STRING)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY '\t'

STORED AS TEXTFILE;

Load u.data into the table that was just created. (omit local as data is in hdfs)

LOAD DATA INPATH 'wk4/ml-100k/u.data' OVERWRITE INTO TABLE u\_data;

Create a weekday\_mapper script  in python and add it to hadoop.

import sys

import datetime

for line in sys.stdin:

  line = line.strip()

  userid, movieid, rating, unixtime = line.split('\t')

  weekday = datetime.datetime.fromtimestamp(float(unixtime)).isoweekday()

  print '\t'.join([userid, movieid, rating, str(weekday)])

Add file to hadoop

scp /Users/Carla/Datasets/weekday\_mapper.py hadoop@192.168.93.163:/home/hadoop

Run the following command

CREATE TABLE u\_data\_new (

  userid INT,

  movieid INT,

  rating INT,

  weekday INT)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY '\t';

add FILE weekday\_mapper.py;

INSERT OVERWRITE TABLE u\_data\_new

SELECT

  TRANSFORM (userid, movieid, rating, unixtime)

  USING 'python weekday\_mapper.py'

  AS (userid, movieid, rating, weekday)

FROM u\_data;

 SELECT weekday, COUNT(\*)

FROM u\_data\_new

GROUP BY weekday