Queries Hive

1. DIRECTOR

LOAD DATA INPATH '/imdbproject/Director.csv' OVERWRITE INTO TABLE temp\_directors;

CREATE TABLE director (DirectorID INT, DirFirstName STRING,DirLastName STRING,Gender STRING,DateOfBirth STRING,BirthPlace STRING,Bio STRING,Nickname STRING,Height STRING,StarSign STRING,Image BINARY);

insert overwrite table director

SELECT

regexp\_extract(col\_value, '^(?:([^,]\*),?){1}', 1) DriverID,

regexp\_extract(col\_value, '^(?:([^,]\*),?){2}', 1) DirFirstName,

regexp\_extract(col\_value, '^(?:([^,]\*),?){3}', 1) DirLastName,

regexp\_extract(col\_value, '^(?:([^,]\*),?){4}', 1) Gender,

regexp\_extract(col\_value, '^(?:([^,]\*),?){5}', 1) DateOfBirth,

regexp\_extract(col\_value, '^(?:([^,]\*),?){6}', 1) BirthPlace,

regexp\_extract(col\_value, '^(?:([^,]\*),?){7}', 1) Bio,

regexp\_extract(col\_value, '^(?:([^,]\*),?){8}', 1) Nickname,

regexp\_extract(col\_value, '^(?:([^,]\*),?){9}', 1) Height,

regexp\_extract(col\_value, '^(?:([^,]\*),?){10}', 1) StarSign,

regexp\_extract(col\_value, '^(?:([^,]\*),?){11}', 1) Image

from temp\_directors;

1. Movie

create table temp\_movies (col\_value STRING);

LOAD DATA INPATH '/imdbproject/movie.csv' OVERWRITE INTO TABLE temp\_movies;

CREATE TABLE movie (MovieORTVID INT, Type STRING,Name STRING,ParentalGuidence STRING,Popularity STRING,Description STRING,PlotKeyword STRING,Language STRING,OriginCountry STRING,FilimingDates STRING,Budget BIGINT,NoOfWeeksRunning BIGINT,WeekEndCollection BIGINT,GrossCollection BIGINT,Image BINARY,Trailer BINARY,Rating FLOAT);

insert overwrite table movie

SELECT

regexp\_extract(col\_value, '^(?:([^,]\*),?){1}', 1) MovieORTVID,

regexp\_extract(col\_value, '^(?:([^,]\*),?){2}', 1) Type,

regexp\_extract(col\_value, '^(?:([^,]\*),?){3}', 1) Name,

regexp\_extract(col\_value, '^(?:([^,]\*),?){4}', 1) ParentalGuidence,

regexp\_extract(col\_value, '^(?:([^,]\*),?){5}', 1) Popularity,

regexp\_extract(col\_value, '^(?:([^,]\*),?){6}', 1) Description,

regexp\_extract(col\_value, '^(?:([^,]\*),?){7}', 1) PlotKeyword,

regexp\_extract(col\_value, '^(?:([^,]\*),?){8}', 1) Language,

regexp\_extract(col\_value, '^(?:([^,]\*),?){9}', 1) OriginCountry,

regexp\_extract(col\_value, '^(?:([^,]\*),?){10}', 1) FilimingDates,

regexp\_extract(col\_value, '^(?:([^,]\*),?){11}', 1) Budget,

regexp\_extract(col\_value, '^(?:([^,]\*),?){12}', 1) NoOfWeeksRunning,

regexp\_extract(col\_value, '^(?:([^,]\*),?){13}', 1) WeekEndCollection,

regexp\_extract(col\_value, '^(?:([^,]\*),?){14}', 1) GrossCollection,

regexp\_extract(col\_value, '^(?:([^,]\*),?){15}', 1) Image,

regexp\_extract(col\_value, '^(?:([^,]\*),?){16}', 1) Trailer,

regexp\_extract(col\_value, '^(?:([^,]\*),?){17}', 1) Rating

from temp\_movies;

1. Directormovie

create table temp\_directormovie (col\_value STRING);

LOAD DATA INPATH '/imdbproject/DirectorMovie.csv' OVERWRITE INTO TABLE temp\_directormovie;

CREATE TABLE directormovie (DirectorID INT, MovieORTVID INT);

insert overwrite table directormovie

SELECT

regexp\_extract(col\_value, '^(?:([^,]\*),?){1}', 1) DirectorID,

regexp\_extract(col\_value, '^(?:([^,]\*),?){2}', 1) MovieORTVID

from temp\_directormovie;

1. DVDAndBluRay

create table temp\_dvdandbluray (col\_value STRING);

LOAD DATA INPATH '/imdbproject/DVDAndBluRay.csv' OVERWRITE INTO TABLE temp\_dvdandbluray;

CREATE TABLE dvdandbluray (DVDAndBluRay INT, MovieORTVID INT,Link STRING);

insert overwrite table dvdandbluray

SELECT

regexp\_extract(col\_value, '^(?:([^,]\*),?){1}', 1) DVDAndBluRay,

regexp\_extract(col\_value, '^(?:([^,]\*),?){2}', 1) MovieORTVID,

regexp\_extract(col\_value, '^(?:([^,]\*),?){3}', 1) Link

from temp\_dvdandbluray;

1. Genre

create table temp\_genre (col\_value STRING);

LOAD DATA INPATH '/imdbproject/Genre.csv' OVERWRITE INTO TABLE temp\_genre;

CREATE TABLE genre (GenreID INT, GenreCategoryName STRING);

insert overwrite table genre

SELECT

regexp\_extract(col\_value, '^(?:([^,]\*),?){1}', 1) GenreID,

regexp\_extract(col\_value, '^(?:([^,]\*),?){2}', 1) GenreCategoryName

from temp\_genre;

1. GenreMovie

create table temp\_genremovie (col\_value STRING);

LOAD DATA INPATH '/imdbproject/GenreMovie.csv' OVERWRITE INTO TABLE temp\_genremovie;

CREATE TABLE genremovie (MovieORTVID INT,GenreID INT);

insert overwrite table genremovie

SELECT

regexp\_extract(col\_value, '^(?:([^,]\*),?){1}', 1) MovieORTVID,

regexp\_extract(col\_value, '^(?:([^,]\*),?){2}', 1) GenreID

from temp\_genremovie;

1. AwardCategory

create table temp\_awardcategory (col\_value STRING);

LOAD DATA INPATH '/imdbproject/AwardCategory.csv' OVERWRITE INTO TABLE temp\_awardcategory;

CREATE TABLE awardcategory (AwardCategoryID INT,AwardHeaderID INT,AwardCategoryName STRING);

insert overwrite table awardcategory

SELECT

regexp\_extract(col\_value, '^(?:([^,]\*),?){1}', 1) AwardCategoryID,

regexp\_extract(col\_value, '^(?:([^,]\*),?){2}', 1) AwardHeaderID,

regexp\_extract(col\_value, '^(?:([^,]\*),?){3}', 1) AwardCategoryName

from temp\_awardcategory;

1. AwardDetails

create table temp\_awarddetails (col\_value STRING);

LOAD DATA INPATH '/imdbproject/AwardDetails.csv' OVERWRITE INTO TABLE temp\_awarddetails;

CREATE TABLE awarddetails (AwardDetailID INT,AwardCategoryID INT,Status STRING,StarID INT,DirectorID INT,WriterID INT,ProducerID INT);

insert overwrite table awarddetails

SELECT

regexp\_extract(col\_value, '^(?:([^,]\*),?){1}', 1) AwardDetailID,

regexp\_extract(col\_value, '^(?:([^,]\*),?){2}', 1) AwardCategoryID,

regexp\_extract(col\_value, '^(?:([^,]\*),?){3}', 1) Status,

regexp\_extract(col\_value, '^(?:([^,]\*),?){4}', 1) StarID,

regexp\_extract(col\_value, '^(?:([^,]\*),?){5}', 1) DirectorID,

regexp\_extract(col\_value, '^(?:([^,]\*),?){6}', 1) WriterID,

regexp\_extract(col\_value, '^(?:([^,]\*),?){7}', 1) ProducerID,

from temp\_awarddetails;

1. AwardHeader

create table temp\_awardheader (col\_value STRING);

LOAD DATA INPATH '/imdbproject/AwardHeader.csv' OVERWRITE INTO TABLE temp\_awardheader;

CREATE TABLE awarddheader (AwardHeaderID INT,MovieORTVID INT,AwardName STRING,Year BIGINT);

insert overwrite table awardheader

SELECT

regexp\_extract(col\_value, '^(?:([^,]\*),?){1}', 1) AwardHeaderID,

regexp\_extract(col\_value, '^(?:([^,]\*),?){2}', 1) MovieORTVID,

regexp\_extract(col\_value, '^(?:([^,]\*),?){3}', 1) AwardName,

regexp\_extract(col\_value, '^(?:([^,]\*),?){4}', 1) Year

from temp\_awardheader;

1. Filming Location Header

create table temp\_filminglocationheader (col\_value STRING);

LOAD DATA INPATH '/imdbproject/FilimingLocationHeader.csv' OVERWRITE INTO TABLE temp\_filminglocationheader;

CREATE TABLE filminglocationheader (FilimingLocationHeaderID INT,MovieORTVID INT,CountryName STRING);

insert overwrite table filminglocationheader

SELECT

regexp\_extract(col\_value, '^(?:([^,]\*),?){1}', 1) FilimingLocationHeaderID,

regexp\_extract(col\_value, '^(?:([^,]\*),?){2}', 1) MovieORTVID,

regexp\_extract(col\_value, '^(?:([^,]\*),?){3}', 1) CountryName

from temp\_filminglocationheader;

1. Film Location Detail

create table temp\_filmlocationdetail (col\_value STRING);

LOAD DATA INPATH '/imdbproject/FilmLocationDetail.csv' OVERWRITE INTO TABLE temp\_filmlocationdetail;

CREATE TABLE filmlocationdetail (FilimingLocationDetailId INT,FilimingLocationHeaderID INT,City STRING,State STRING);

insert overwrite table filmlocationdetail

SELECT

regexp\_extract(col\_value, '^(?:([^,]\*),?){1}', 1) FilimingLocationDetailId,

regexp\_extract(col\_value, '^(?:([^,]\*),?){2}', 1) FilimingLocationHeaderID,

regexp\_extract(col\_value, '^(?:([^,]\*),?){3}', 1) City,

regexp\_extract(col\_value, '^(?:([^,]\*),?){4}', 1) State

from temp\_filmlocationdetail;

1. Movieandshowtimestickets

create table temp\_movieandshowtimestickets (col\_value STRING);

LOAD DATA INPATH '/imdbproject/MovieAndShowtimesTickets.csv' OVERWRITE INTO TABLE temp\_movieandshowtimestickets;

CREATE TABLE movieandshowtimestickets (MovieAndTVID INT,ShowtimesAndTicketsID INT);

insert overwrite table movieandshowtimestickets

SELECT

regexp\_extract(col\_value, '^(?:([^,]\*),?){1}', 1) MovieAndTVID,

regexp\_extract(col\_value, '^(?:([^,]\*),?){2}', 1) ShowtimesAndTicketsID,

from temp\_movieandshowtimestickets;