*//Create a temporary table*

Use IMDB

create table new\_temp\_table (col\_value STRING);

*//inserting data into Address Table*

LOAD DATA INPATH '/user/maria\_dev/project/Address.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table address

SELECT

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

regexp\_extract(col\_value, '^(?:([^,]\*),?){2}', 1) Address\_Line1,

regexp\_extract(col\_value, '^(?:([^,]\*),?){3}', 1) Address\_Line2,

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

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

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

from new\_temp\_table;

*//insert data into AddressType table*

LOAD DATA INPATH '/user/maria\_dev/project/AddressType.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table addresstype

SELECT

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

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

from new\_temp\_table;

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

*//inserting data into Artist Table*

LOAD DATA INPATH '/user/maria\_dev/project/Artist.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table artist

SELECT

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

from new\_temp\_table;

*//insert data into Artist\_Address table*

LOAD DATA INPATH '/user/maria\_dev/project/Artist\_Address.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table artistaddress

SELECT

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

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

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

from new\_temp\_table;

*//insert data into Artist\_Trademark table*

LOAD DATA INPATH '/user/maria\_dev/project/Artist\_Trademark.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table artisttrademark

SELECT

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

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

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

from new\_temp\_table;

*//insert data into ArtistType table*

LOAD DATA INPATH '/user/maria\_dev/project/Artist\_Trademark.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table artisttype

SELECT

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

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

from new\_temp\_table;

*//insert data into Award table*

LOAD DATA INPATH '/user/maria\_dev/project/Award.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table award

SELECT

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

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

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

from new\_temp\_table;

*//insert data into Award\_Category table*

LOAD DATA INPATH '/user/maria\_dev/project/Artist\_Category.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table awardcategory

SELECT

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

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

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

from new\_temp\_table;

*//insert data into BoxOffice table*

LOAD DATA INPATH '/user/maria\_dev/project/BoxOffice.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table boxoffice

SELECT

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

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

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

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

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

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

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

from new\_temp\_table;

*//insert data into Cast table*

LOAD DATA INPATH '/user/maria\_dev/project/Cast1.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table cast1

SELECT

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

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

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

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

from new\_temp\_table;

*//insert data into Certificate table*

LOAD DATA INPATH '/user/maria\_dev/project/Certificate.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table certificate

SELECT

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

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

from new\_temp\_table;

*//insert data into City table*

LOAD DATA INPATH '/user/maria\_dev/project/City.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table city

SELECT

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

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

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

from new\_temp\_table;

*//insert data into Comment table*

LOAD DATA INPATH '/user/maria\_dev/project/Comment.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table comment

SELECT

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

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

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

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

from new\_temp\_table;

*//insert data into Country table*

LOAD DATA INPATH '/user/maria\_dev/project/Country.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table country

SELECT

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

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

from new\_temp\_table;

*//insert data into Critic table*

LOAD DATA INPATH '/user/maria\_dev/project/Critic.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table critic

SELECT

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

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

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

from new\_temp\_table;

*//inserting data into Episode Table*

LOAD DATA INPATH '/user/maria\_dev/project/Episode.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table episode

SELECT

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

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

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

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

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

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

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

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

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

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

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

from new\_temp\_table;

*//insert data into Genre table*

LOAD DATA INPATH '/user/maria\_dev/project/Genre.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table genre

SELECT

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

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

from new\_temp\_table;

*//insert data into Language table*

LOAD DATA INPATH '/user/maria\_dev/project/Language.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table language

SELECT

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

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

from new\_temp\_table;

*//insert data into List table*

LOAD DATA INPATH '/user/maria\_dev/project/List.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table list

SELECT

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

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

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

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

from new\_temp\_table;

*//insert data into ListBridge table*

LOAD DATA INPATH '/user/maria\_dev/project/ListBridge.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table listbridge

SELECT

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

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

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

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

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

from new\_temp\_table;

*//insert data into MessageBoard table*

LOAD DATA INPATH '/user/maria\_dev/project/MassageBoard.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table messageboard

SELECT

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

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

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

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

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

from new\_temp\_table;

*//insert data into MessageBoardReadBy table*

LOAD DATA INPATH '/user/maria\_dev/project/MassageBoardReadBy.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table messageboardreadby

SELECT

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

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

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

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

from new\_temp\_table;

*//insert data into Movie table*

LOAD DATA INPATH '/user/maria\_dev/project/Movie.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table movie

SELECT

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

from new\_temp\_table;

*//insert data into Movie\_Genre table*

LOAD DATA INPATH '/user/maria\_dev/project/Movie\_genre.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table moviegenre

SELECT

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

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

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

from new\_temp\_table;

*//insert data into Movie\_Screen table*

LOAD DATA INPATH '/user/maria\_dev/project/Movie\_Screen.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table moviescreen

SELECT

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

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

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

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

from new\_temp\_table;

*//insert data into NewsArticle table*

LOAD DATA INPATH '/user/maria\_dev/project/NewsArticle.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table newsarticle

SELECT

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

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

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

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

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

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

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

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

from new\_temp\_table;

*//insert data into Nominee table*

LOAD DATA INPATH '/user/maria\_dev/project/Nominee.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table nominee

SELECT

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

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

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

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

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

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

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

from new\_temp\_table;

*//insert data into Organization table*

LOAD DATA INPATH '/user/maria\_dev/project/Organization.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table organization

SELECT

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

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

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

from new\_temp\_table;

*//insert data into ParentGuide table*

LOAD DATA INPATH '/user/maria\_dev/project/ParentGuide.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table parentguide

SELECT

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

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

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

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

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

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

from new\_temp\_table;

*//insert data into Person table*

LOAD DATA INPATH '/user/maria\_dev/project/Person.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table person

SELECT

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

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

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

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

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

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

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

from new\_temp\_table;

*//insert data into Photo table*

LOAD DATA INPATH '/user/maria\_dev/project/Photo.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table photo

SELECT

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

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

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

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

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

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

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

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

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

from new\_temp\_table;

*//insert data into PloyKeyword table*

LOAD DATA INPATH '/user/maria\_dev/project/PlotKeyword.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table plotkeyword

SELECT

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

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

from new\_temp\_table;

*//insert data into PlotKeyword\_Bridge table*

LOAD DATA INPATH '/user/maria\_dev/project/PlotKeyword\_Bridge.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table plotKeyword\_Bridge

SELECT

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

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

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

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

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

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

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

from new\_temp\_table;

*//insert data into PlotSummary table*

LOAD DATA INPATH '/user/maria\_dev/project/PlotSummary.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table plotsummary

SELECT

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

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

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

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

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

from new\_temp\_table;

*//insert data into PollBridge table*

LOAD DATA INPATH '/user/maria\_dev/project/PollBridge.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table pollbridge

SELECT

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

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

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

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

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

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

from new\_temp\_table;

*//insert data into Polls table*

LOAD DATA INPATH '/user/maria\_dev/project/Polls.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table polls

SELECT

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

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

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

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

from new\_temp\_table;

*//insert data into PollVote table*

LOAD DATA INPATH '/user/maria\_dev/project/PollVote.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table pollvote

SELECT

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

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

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

from new\_temp\_table;

*//insert data into ProductionCompany table*

LOAD DATA INPATH '/user/maria\_dev/project/ProductionCompany.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table productioncompany

SELECT

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

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

from new\_temp\_table;

*//insert data into ProductionCompanyBridge table*

LOAD DATA INPATH '/user/maria\_dev/project/ProductionCompanyBridge.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table productioncompanybridge

SELECT

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

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

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

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

from new\_temp\_table;

*//insert data into ReadBy table*

LOAD DATA INPATH '/user/maria\_dev/project/ReadBy.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table readby

SELECT

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

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

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

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

from new\_temp\_table;

*//insert data into ReleaseDate table*

LOAD DATA INPATH '/user/maria\_dev/project/ReleaseDate.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table releasedate

SELECT

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

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

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

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

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

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

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

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

from new\_temp\_table;

*//insert data into Replythread table*

LOAD DATA INPATH '/user/maria\_dev/project/ReplyThread.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table replythread

SELECT

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

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

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

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

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

from new\_temp\_table;

*//insert data into Reviews table*

LOAD DATA INPATH '/user/maria\_dev/project/Reviews.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table reviews

SELECT

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

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

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

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

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

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

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

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

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

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

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

from new\_temp\_table;

*//insert data into Screen table*

LOAD DATA INPATH '/user/maria\_dev/project/Screen.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table screen

SELECT

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

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

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

from new\_temp\_table;

*//insert data into Season table*

LOAD DATA INPATH '/user/maria\_dev/project/Season.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table season

SELECT

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

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

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

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

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

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

from new\_temp\_table;

*//insert data into Seat table*

LOAD DATA INPATH '/user/maria\_dev/project/Seat.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table seat

SELECT

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

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

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

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

from new\_temp\_table;

*//insert data into Seattype table*

LOAD DATA INPATH '/user/maria\_dev/project/SeatType.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table seattype

SELECT

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

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

from new\_temp\_table;

*//insert data into ShowTiming table*

LOAD DATA INPATH '/user/maria\_dev/project/ShowTiming.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table showtiming

SELECT

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

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

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

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

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

from new\_temp\_table;

*//insert data into State table*

LOAD DATA INPATH '/user/maria\_dev/project/State.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table state

SELECT

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

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

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

from new\_temp\_table;

*//insert data into TechnicalSpecs table*

LOAD DATA INPATH '/user/maria\_dev/project/TechnicalSpecs.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table technicalspecs

SELECT

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

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

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

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

from new\_temp\_table;

*//insert data into TechnicalSpecs table*

LOAD DATA INPATH '/user/maria\_dev/project/TechnicalSpecs.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table technicalspecs

SELECT

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

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

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

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

from new\_temp\_table;

*//insert data into Theatre table*

LOAD DATA INPATH '/user/maria\_dev/project/Theatre.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table theatre

SELECT

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

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

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

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

from new\_temp\_table;

*//insert data into Ticket table*

LOAD DATA INPATH '/user/maria\_dev/project/Ticket.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table ticket

SELECT

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

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

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

from new\_temp\_table;

*//insert data into TradeMark table*

LOAD DATA INPATH '/user/maria\_dev/project/TradeMark.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table trademark

SELECT

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

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

from new\_temp\_table;

*//insert data into Trailer table*

LOAD DATA INPATH '/user/maria\_dev/project/Trailer.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table trailer

SELECT

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

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

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

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

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

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

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

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

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

from new\_temp\_table;

*//insert data into Trivia table*

LOAD DATA INPATH '/user/maria\_dev/project/Trivia.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table trivia

SELECT

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

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

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

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

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

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

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

from new\_temp\_table;

*//insert data into TVShow table*

LOAD DATA INPATH '/user/maria\_dev/project/TVShow.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table tvshow

SELECT

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

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

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

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

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

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

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

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

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

from new\_temp\_table;

*//insert data into USer table*

LOAD DATA INPATH '/user/maria\_dev/project/User.csv' OVERWRITE INTO TABLE new\_temp\_table;

insert overwrite table user

SELECT

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

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

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

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

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

from new\_temp\_table;