docker cp socialmedia.csv awesome\_gates:/socialmedia.csv

docker cp products.csv awesome\_gates:/products.csv

docker cp demographics.csv awesome\_gates:/demographics.csv

**SOCIALMEDIA – CSV**

create table SocialMedia (StoreName string, StoreType string, Latitude string, Longitude string, Address string, Zipcode string, CreatedAt string, Text string, TextType string, SentimentType string, SentimentScore string, Product string, Service string, CalculatedRating int, Source string, EffectiveNovalty string, HouseholdsPerZip string, IncomePerHousehold string, PersonPerHousehold string, CustomerTendency string, SourceCredibility string, SourcePerceptionIndex string)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY '\t'

LINES TERMINATED BY '\n'

STORED AS TEXTFILE;

load data local inpath '/socialmedia.csv' overwrite into table socialmedia;

**PRODUCTS – CSV**

create table products (ProdName string, LaunchDate string, Description string)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY '\t'

LINES TERMINATED BY '\n'

STORED AS TEXTFILE;

load data local inpath '/products.csv' overwrite into table products;

**DEMOGRAPHICS – CSV**

Create table Demographics (ZipCode string, CurrentPopulation string, Population2010 string, HouseholdsPerZIP string, AverageHouseValue string, AverageIncomePerHousehold string, PersonsPerHousehold string, WhitePopulation string, BlackPopulation string, HispanicPopulation string, AsianPopulation string, AmericanIndianPopulation string, HawaiianPopulation string, OtherPopulation string, MalePopulation string, FemalePopulation string, MedianAge string, MaleMedianAge string, FemaleMedianAge string)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY '\t'

LINES TERMINATED BY '\n'

STORED AS TEXTFILE;

load data local inpath '/demographics.csv' overwrite into table demographics;

**STORE DIMENSION**

create table StoreDim (StoreID int, StoreName string, StoreType string, Latitude string, Longitude string, Address string, Zipcode string, HouseholdsPerZip string, IncomePerHousehold string, PersonPerHousehold string)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY '\t'

LINES TERMINATED BY '\n'

STORED AS TEXTFILE;

insert into StoreDim select distinct rank() over (order by StoreName asc) as StoreID, StoreName, StoreType, Latitude, Longitude, Address, Zipcode, HouseholdsPerZip, IncomePerHousehold, PersonPerHousehold from socialmedia;

**DATE DIMENSION**

create table DateDim (TimeID int, Year int, Month int, Day int, Hour int, Minute int, Second int)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY '\t'

LINES TERMINATED BY '\n'

STORED AS TEXTFILE;

create table tempDateDim (Timeid int, Year int, Month int, Day int, Hour int, Minute int, Second int, Createdat string)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY '\t'

LINES TERMINATED BY '\n'

STORED AS TEXTFILE;

insert into tempDateDim select distinct rank() over (order by Createdat asc) as Timeid, substr(createdat,1,4) as Year, substr(createdat,6,2) as Month, substr(createdat,9,2) as Day, split(split(createdat,’ ’)[1],’:’)[0] as Hour, split(split(createdat,’ ’)[1],’:’)[1] as Minute, split(split(createdat,’ ’)[1],’:’)[2] as Second, **Createdat** from socialmedia;

insert into datedim select Timeid, Year, Month, Day, Hour, Minute, Second from tempdatedim;

**PRODUCT DIMENSION**

create table ProductDim (ProdID int, ProdName string, LaunchDate string, Description string)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY '\t'

LINES TERMINATED BY '\n'

STORED AS TEXTFILE;

insert into productdim select distinct rank() over (order by ProdName asc) as ProdID, ProdName, LaunchDate, Description from products;

**SENTIMENT DIMENSION**

create table SentimentDim (SentimentID int, SentimentType string)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY '\t'

LINES TERMINATED BY '\n'

STORED AS TEXTFILE;

insert into sentimentdim select distinct rank() over (order by SentimentType asc) as SentimentID, SentimentType from socialmedia;

**SOURCE DIMENSION**

create table SourceDim (SourceID int, Source string)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY '\t'

LINES TERMINATED BY '\n'

STORED AS TEXTFILE;

insert into sourcedim select distinct rank() over (order by Source asc) as SourceID, Source from socialmedia;

**SERVICE DIMENSION**

create table ServiceDim (ServiceID int, Service string)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY '\t'

LINES TERMINATED BY '\n'

STORED AS TEXTFILE;

insert into servicedim select distinct rank() over (order by Service asc) as ServiceID, Service from socialmedia;

**TEXT JUNK DIMENSION**

create table TextJunkDim (TextID int, Text string)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY '\t'

LINES TERMINATED BY '\n'

STORED AS TEXTFILE;

insert into textjunkdim select distinct rank() over (order by Text asc) as TextID, Text from socialmedia;

**TEXT TYPE FLAG JUNK DIMENSION**

create table TextTypeFlagJunkDimension(TextTypeFlagID int, TextType string)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY '\t'

LINES TERMINATED BY '\n'

STORED AS TEXTFILE;

insert into texttypeflagjunkdimension select distinct rank() over (order by TextType asc) as TextTypeFlagID, TextType from socialmedia;

**STORE FACT:**

create table FactStore (StoreId int, TimeId int, TextID int, TextTypeFlagID int, SentimentID int, ProdID int, ServiceID int, SourceID int, SentimentScore float, CalculatedRating float, EffectiveNovalty float, CustomerTendency float, SourceCredibility float, SourcePerceptionIndex float)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY '\t'

LINES TERMINATED BY '\n'

STORED AS TEXTFILE;

insert into FactStore

select st.storeid as StoreId,

d.timeid as TimeId,

td.textid as TextID,

ttf.TextTypeFlagID as TextTypeFlagID,

snd.sentimentid as SentimentID,

pd.prodid as ProdID,

srd.serviceid as ServiceID,

sod.sourceid as SourceID,

m.SentimentScore as SentimentScore,

m.rating as CalculatedRating,

m.EffectiveNovalty as EffectiveNovalty,

m.CustomerTendency as CustomerTendency,

m.SourceCredibility as SourceCredibility,

m.SourcePerceptionIndex as SourcePerceptionIndex

from socialmedia m inner join

storedim st on m.storename=st.storename inner join

datedim d on unix\_timestamp(concat(d.year, '-', d.month, '-', d.day, ' ', d.hour, ':', d.minute, ':', d.second))=unix\_timestamp(concat(substr(m.createdat,1,4), '-', substr(m.createdat,6,2), '-', substr(m.createdat,9,2), ' ', split(split(m.createdat,’ ’)[1],’:’)[0], ':', split(split(m.createdat,’ ’)[1],’:’)[1], ':', split(split(m.createdat,’ ’)[1],’:’)[2])) inner join

textjunkdim td on m.text=td.text inner join

texttypeflagjunkdimension ttf on m.texttype=ttf.texttype inner join

sentimentdim snd on m.SentimentType=snd.SentimentType inner join

productdim pd on m.Product=pd.ProdName inner join

servicedim srd on m.Service=srd.Service inner join

sourcedim sod on m.Source=sod.Source;