Assignment1

Q1: Lengthwise Word Count Problem

Upload the file to hdfs using put command

[tanveer7p6164@cxln4 Assignment]$ hdfs dfs -put assign-text1.txt

[tanveer7p6164@cxln4 Assignment]$

vendor\_id string

pickup\_datetime string

dropoff\_datetime string

passenger\_count int

trip\_distance decimal(9,6)

pickup\_longitude decimal(9,6)

pickup\_latitude decimal(9,6)

rate\_code int

store\_and\_fwd\_flag string

dropoff\_longitude decimal(9,6)

dropoff\_latitude decimal(9,6)

payment\_type string

fare\_amount decimal(9,6)

extra decimal(9,6)

mta\_tax decimal(9,6)

tip\_amount decimal(9,6)

tolls\_amount decimal(9,6)

total\_amount decimal(9,6)

trip\_time\_in\_secs int

What is the total Number of trips ( equal to the number of rows)?

Select count(0) from taxidata = 10000

What is the total revenue generated by all the trips? The fare is stored in the column total\_amount.

Select sum(total\_amount) from taxidata; 160546.81

What fraction of the total is paid for tolls? The toll is stored in tolls\_amount.=2496.99 fraction = 1.55 percent

Select (sum(tolls\_amount)/sum(total\_amount))\*100 from taxidata;

Select (sum(tolls\_amount) from taxidata;

What fraction of it is driver tips? The tip is stored in tip\_amount.

Select (sum(tip\_amount)/sum(total\_amount))\*100 from taxidata; fractio of driver tips=10.785203393328089

What is the average trip amount?

Select avg(fare\_amount) from taxidata;

What is the average distance of the trips? Distance is stored in the column trip\_distance.

Select average(trip\_distance) from taxidata; = 3.25

How many different payment types are used?

Select distinct(payment\_type) from taxidata; 4

For each payment type, display the following details: • Average fare generated • Average tip • Average tax – tax is stored in column mta\_tax

Select payment\_type,avg(fare\_amount),avg(tip\_amount),avg(mta\_tax) from taxidata group by payment\_type,fare\_amount,tip\_amount,mta\_tax

Select payment\_type,avg(fare\_amount) from taxidata group by payment\_type,fare\_amount;

Select payment\_type,avg(tip\_amount) from taxidata group by payment\_type,tip\_amount;

Select payment\_type,avg(mta\_tax) from taxidata group by payment\_type,mta\_tax;

select h24 as hour,avg(total\_amount) as avg\_revenue, from (select hour(pickup\_datetime) as h24,total\_amount from taxidata) ff group by h24 order by avg\_revenue desc