login as: training

training@192.168.99.100's password:

Last login: Fri Mar 24 00:23:45 2017 from 192.168.99.1

Appliance: Cloudera-Training-VM-4.2.1.p appliance 4.2

Hostname: localhost.localdomain

IP Address:

**Write a set of Hive scripts that load these files into a staging external table called, Teller.**

**Create a managed partitioned table, called PtTeller that partitions the rows in the table Teller**

**dynamically based on the teller’s name.**

**Creating Teller Table**

hive> Create external table Teller(Name string,Bills array<int>)

> Row format delimited

> Fields terminated by ','

> Collection Items terminated by ' '

> Lines terminated by '\n'

> Stored as textfile

> LOCATION '/user/training/hive/Teller';

**Loading Data into Teller Table**

hive> load data inpath '/user/training/hive/bankdata' into table Teller;

**Teller Table Contents**

hive> Select \* from Teller;

Saeed [20,5,5,10,1,1,1,1,1,5,10,10,10]

Steve [1,1,1,1,1,5,10,20,10,10,10,10]

Mo [1,1,5,1,10,5,10,20,10,20,1,1,10,10,10]

Todd [1,1,1,1,1,5,10,20,10,10,10,10,10,10]

Jon [1,1,1,1,1,5,10,20,10,10,10,10,5,20,1]

Jon [20,5,5,10,1,1,1,1,1,5,5,5,5,5,5]

Steve [1,1,1,1,1,5,10,20,10,10,10,10]

Todd [20,5,5,10,1,1,1,1,1,5,10,10,10]

Mo [1,1,1,1,1,5,10,20,10,10]

Saeed [1,1,5,1,10,5,10,20,10,20,1,1,10,10,10]

Jon [1,1,1,1,1,5,10,20,10,10,10,10,5,20,1]

Mo [1,1,1,1,1,5,10,20,10,10,10,20,5,5,10,1,1,1,1,1,5,10,10,10]

Ron [1,1,1,1,1,5,10,20,10,10,10,10]

Ron [1,1,5,1,10,5,10,20,10,20,1,1,10,10,10]

Jon [1,1,1,1,1,5,10,20,10,10,10,10,10,10]

Mo [20,5,5,10,1,1,1,1,1,5,5,5,5,5,5]

Mo [1,1,1,1,1,5,10,20,10,10,10,10]

Jon [1,1,1,1,1,5,5,5,5,5,5]

Todd [20,10,10,10,10]

**Creating ptTeller Partition table**

hive> Create table ptTeller(Bills array<int>)

> Partitioned by(Name String)

> Row format delimited

> Fields terminated by ','

> LOCATION'/user/training/hive/ptTeller';

**Setting Dynamic Partition**

hive> set hive.exec.dynamic.partition.mode=nonstrict;

hive> set hive.exec.dynamic.partition=true;

**Inserting Data into ptTeller**

hive> INSERT OVERWRITE TABLE ptTeller PARTITION(Name)SELECT Bills,Name FROM Teller;

**ptTeller Table Contents**

hive> Select \* from ptTeller;

[1,1,1,1,1,5,10,20,10,10,10,10,5,20,1] Jon

[20,5,5,10,1,1,1,1,1,5,5,5,5,5,5] Jon

[1,1,1,1,1,5,10,20,10,10,10,10,5,20,1] Jon

[1,1,1,1,1,5,10,20,10,10,10,10,10,10] Jon

[1,1,1,1,1,5,5,5,5,5,5] Jon

[1,1,5,1,10,5,10,20,10,20,1,1,10,10,10] Mo

[1,1,1,1,1,5,10,20,10,10] Mo

[1,1,1,1,1,5,10,20,10,10,10,20,5,5,10,1,1,1,1,1,5,10,10,10] Mo

[20,5,5,10,1,1,1,1,1,5,5,5,5,5,5] Mo

[1,1,1,1,1,5,10,20,10,10,10,10] Mo

[1,1,1,1,1,5,10,20,10,10,10,10] Ron

[1,1,5,1,10,5,10,20,10,20,1,1,10,10,10] Ron

[20,5,5,10,1,1,1,1,1,5,10,10,10] Saeed

[1,1,5,1,10,5,10,20,10,20,1,1,10,10,10] Saeed

[1,1,1,1,1,5,10,20,10,10,10,10] Steve

[1,1,1,1,1,5,10,20,10,10,10,10] Steve

[1,1,1,1,1,5,10,20,10,10,10,10,10,10] Todd

[20,5,5,10,1,1,1,1,1,5,10,10,10] Todd

[20,10,10,10,10] Todd

**PART A**

**Working with table Teller, write a set of Hive scripts that print the answers the following**

**questions:**

**Q1: The total amount of money and total number of bills collected by all tellers.**

hive> SELECT distinct c.sum,c.count

> FROM (Select name, name1 FROM Teller LATERAL VIEW explode(bills) subs as name1) a

> JOIN

> (SELECT SUM(b.name1) sum,count(b.name1) count

> FROM

> (Select name, name1 FROM Teller LATERAL VIEW explode(bills) subs as name1)b)c;

1692 257

**Q2:The teller name and the total amount of money collected and the number of bills and the average money for each teller (as done in the MR job).**

hive> select name, sum(name1) as sum,avg(name1) as avg,count(name1) as count

> from

> (Select name, name1 FROM Teller LATERAL VIEW explode(bills) subs as name1)a

> group by name;

Jon 422 6.0285714285714285 70

Mo 480 6.315789473684211 76

Ron 195 7.222222222222222 27

Saeed 195 6.964285714285714 28

Steve 160 6.666666666666667 24

Todd 240 7.5 32

**Q3: Print the teller name and the number of bills for those tellers that have collected the same number of bills.**

hive> select a.name,a.count1 from

> (select c.name,count(name3) as count1 FROM Teller c LATERAL VIEW explode(c.bill s)subs as name3 group by c.name)a

> join

> (select d.name,count(name3) as count1 FROM Teller d LATERAL VIEW explode(d.bill s)subs as name3 group by d.name)b

> on a.count1=b.count1

> where

> a.name!=b.name;

**PART B**

**Working with the PtTeller table, write a set of scripts that print the answers to the following**

**questions:**

**Q1: The number of partitions in this table.**

hive> select count(distinct Name) from ptTeller;

6

**Q2: The teller name and the total amount of money collected and the number of bills and**

**the average money for each teller (as done in the MR job).**

hive> select name, sum(name1) as sum,avg(name1) as avg,count(name1) as count

> from

> (Select name, name1 FROM ptTeller LATERAL VIEW explode(bills) subs as name1)a

> group by name;

Jon 422 6.0285714285714285 70

Mo 480 6.315789473684211 76

Ron 195 7.222222222222222 27

Saeed 195 6.964285714285714 28

Steve 160 6.666666666666667 24

Todd 240 7.5 32

**Q3: The teller or the tellers’ names that have collected the maximum amount of money.**

hive> select name, sum(name1) as sum

> from

> (Select name, name1 FROM ptTeller LATERAL VIEW explode(bills) subs as name1)a

> group by name

> sort by sum desc limit 1;

Mo 480