BDT Lab Assignment 07

Name: Devanshu Surana

PRN: 1032210755

Roll No.: BDT2(50)

Batch: 2

1. Create text file and show the contents.

```
[cloudera@quickstart ~]$ cd /home/cloudera/Documents
[cloudera@quickstart Documents]$ cat Devanshu
101,Devanshu,Pune,50000
102,Prachiti,Pune,12000
103,Pranav,Pune,51000
104,Abhilash,Solapur,25000
```

2. Start Hive and Show available DB

```
[cloudera@quickstart Documents]$ hive

Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.properties
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive> show databases;
OK
default
Time taken: 0.896 seconds, Fetched: 1 row(s)
hive>
```

3. Create table with same schema

```
hive> create table employee(emp_id int,Name string,City string,Salary int)
    > row format delimited fields terminated by ',';
OK
Time taken: 0.861 seconds
```

```
hive> desc employee;
OK
emp_id              int
name                string
city                string
salary              int
Time taken: 0.231 seconds, Fetched: 4 row(s)
hive>
```

4. Load employee database in Hadoop and display contents

```
[cloudera@quickstart Documents]$ hadoop fs -put employee
[cloudera@quickstart Documents]$ hadoop fs -ls
Found 2 items
-rw-r--r--   1 cloudera cloudera         90 2023-10-11 01:52 Devanshu
-rw-r--r--   1 cloudera cloudera         90 2023-10-11 01:54 employee
[cloudera@quickstart Documents]$ hadoop fs -put employee
put: `employee': File exists
[cloudera@quickstart Documents]$ hadoop fs -cat employee
101,Devanshu,Jalgaon,50000
102,Jay,Rajkot,12000
103,Vansh,Jaipur,51000
104,Ayush,Nagpur,25000
```

5. Load data into employee table on hive and fetch data

```
hive> load data inpath 'employee' overwrite into table employee;
Loading data to table default.employee
chgrp: changing ownership of
'hdfs://quickstart.cloudera:8020/user/hive/warehouse/employee/employee': User does
not belong to supergroup
Table default.employee stats: [numFiles=1, numRows=0, totalSize=90, rawDataSize=0]
OK
Time taken: 0.789 seconds
hive> select *from employee;
OK
```

101,Devanshu,Pune,50000

102,Prachiti,Pune,12000

103,Pranav,Pune,51000

104,Abhilash,Solapur,25000

Time taken: 0.416 seconds, Fetched: 4 row(s)

hive>

6. Check schema of employee table and perform other operations

hive> desc employee;

OK

emp_id	int
name	string
city	string
salary	int

Time taken: 0.12 seconds, Fetched: 4 row(s)

hive> select emp_id, name from employee;

OK

101	Devanshu
102	Prachiti
103	Pranav
104	Abhilash

Time taken: 0.121 seconds, Fetched: 4 row(s)

hive>

Aggregate Functions:

1. SUM:

hive> SELECT SUM(Salary) AS TotalSalary FROM employee;

Query ID = cloudera_20231011021313_fb5dfc76-3ebd-4f18-977a-8bc9ae40e668

Total jobs = 1

Launching Job 1 out of 1

Number of reduce tasks determined at compile time: 1

In order to change the average load for a reducer (in bytes):

set hive.exec.reducers.bytes.per.reducer=<number>

In order to limit the maximum number of reducers:

set hive.exec.reducers.max=<number>

In order to set a constant number of reducers:

set mapreduce.job.reduces=<number>

Starting Job = job_1696914829987_0001, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1696914829987_0001/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1696914829987_0001

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1

2023-10-11 02:13:37,981 Stage-1 map = 0%, reduce = 0%

2023-10-11 02:13:58,756 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.71 sec

2023-10-11 02:14:26,478 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 6.09 sec

MapReduce Total cumulative CPU time: 6 seconds 90 msec

Ended Job = job_1696914829987_0001

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 6.09 sec HDFS Read: 7052 HDFS Write: 7 SUCCESS

Total MapReduce CPU Time Spent: 6 seconds 90 msec

OK

138000

Time taken: 70.392 seconds, Fetched: 1 row(s)

hive>

2. AVERAGE:

hive> SELECT AVG(Salary) AS AverageSalary FROM employee;

Query ID = cloudera_20231011022121_e95567ae-5607-48d4-92e1-f61dad6ad0a1

Total jobs = 1

Launching Job 1 out of 1

Number of reduce tasks determined at compile time: 1

In order to change the average load for a reducer (in bytes):

set hive.exec.reducers.bytes.per.reducer=<number>

In order to limit the maximum number of reducers:

set hive.exec.reducers.max=<number>

In order to set a constant number of reducers:

set mapreduce.job.reduces=<number>

Starting Job = job_1696914829987_0002, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1696914829987_0002/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1696914829987_0002

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1

2023-10-11 02:22:06,093 Stage-1 map = 0%, reduce = 0%

2023-10-11 02:22:35,806 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.14 sec

2023-10-11 02:23:35,502 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 8.72 sec

MapReduce Total cumulative CPU time: 8 seconds 720 msec

Ended Job = job_1696914829987_0002

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 8.72 sec HDFS Read: 7351 HDFS Write: 8 SUCCESS

Total MapReduce CPU Time Spent: 8 seconds 720 msec

OK

34500.0

Time taken: 129.959 seconds, Fetched: 1 row(s)

hive>

3. COUNT

hive> SELECT COUNT(*) AS TotalEmployees FROM employee;

Query ID = cloudera_20231011022525_e0008578-4c07-4780-b4cc-465986bd1923

Total jobs = 1

Launching Job 1 out of 1

Number of reduce tasks determined at compile time: 1

In order to change the average load for a reducer (in bytes):

set hive.exec.reducers.bytes.per.reducer=<number>

In order to limit the maximum number of reducers:

set hive.exec.reducers.max=<number>

In order to set a constant number of reducers:

set mapreduce.job.reduces=<number>

Starting Job = job_1696914829987_0003, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1696914829987_0003/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1696914829987_0003

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1

2023-10-11 02:27:39,909 Stage-1 map = 0%, reduce = 0%

2023-10-11 02:28:30,091 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.47 sec

2023-10-11 02:28:52,850 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.78 sec

MapReduce Total cumulative CPU time: 5 seconds 780 msec

Ended Job = job_1696914829987_0003

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 5.78 sec HDFS Read: 7079 HDFS Write: 2 SUCCESS

Total MapReduce CPU Time Spent: 5 seconds 780 msec

OK

4

Time taken: 220.364 seconds, Fetched: 1 row(s)

hive>

4. MIN:

hive> SELECT MIN(Salary) AS MinSalary FROM employee;

Query ID = cloudera_20231011023030_e33d091c-1808-4da4-9c33-ebebeedea8d1

Total jobs = 1

Launching Job 1 out of 1

Number of reduce tasks determined at compile time: 1

In order to change the average load for a reducer (in bytes):

set hive.exec.reducers.bytes.per.reducer=<number>

In order to limit the maximum number of reducers:

set hive.exec.reducers.max=<number>

In order to set a constant number of reducers:

set mapreduce.job.reduces=<number>

Starting Job = job_1696914829987_0004, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1696914829987_0004/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1696914829987_0004

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1

2023-10-11 02:31:11,859 Stage-1 map = 0%, reduce = 0%

2023-10-11 02:31:30,748 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.33 sec

2023-10-11 02:31:53,073 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.54 sec

MapReduce Total cumulative CPU time: 5 seconds 540 msec

Ended Job = job_1696914829987_0004

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1   Cumulative CPU: 5.54 sec   HDFS Read: 7145 HDFS Write: 6 SUCCESS

Total MapReduce CPU Time Spent: 5 seconds 540 msec

OK

12000

Time taken: 73.043 seconds, Fetched: 1 row(s)

hive>

5. MAX:

hive> SELECT MAX(Salary) AS MaxSalary FROM employee;

Query ID = cloudera_20231011023333_9595cc60-c85e-4d2a-9b3a-c2501a36b44c

Total jobs = 1

Launching Job 1 out of 1

Number of reduce tasks determined at compile time: 1

In order to change the average load for a reducer (in bytes):

  set hive.exec.reducers.bytes.per.reducer=<number>

In order to limit the maximum number of reducers:

  set hive.exec.reducers.max=<number>

In order to set a constant number of reducers:

  set mapreduce.job.reduces=<number>

Starting Job = job_1696914829987_0005, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1696914829987_0005/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1696914829987_0005

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1

2023-10-11 02:33:21,623 Stage-1 map = 0%, reduce = 0%

2023-10-11 02:33:29,328 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.17 sec

2023-10-11 02:33:39,208 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 2.74 sec

MapReduce Total cumulative CPU time: 2 seconds 740 msec

Ended Job = job_1696914829987_0005

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1   Cumulative CPU: 2.74 sec   HDFS Read: 7138 HDFS Write: 6 SUCCESS

Total MapReduce CPU Time Spent: 2 seconds 740 msec

OK

51000

Time taken: 29.297 seconds, Fetched: 1 row(s)

hive>