CAP 457: INTRODUCTION TO BIG DATA - LABORATORY

CONTINUOUS ASSESSMENTS (C.A)-1

ST_NAME: - EKHLAKH AHMAD
REG NO.: - 12209166
ROLL NO.: - RD2215B50
SECTION: - D2215
GROUP: - 2
Date: - 23/02/2023

SET II

Question Number	Question Statement	Course Outcome	Bloom's level	Marks per Question
Q1	Explain the HIVE commands: 1. Create a database 2. Table creation (Eg: Library) 3. Insert records (Min:7 records) 4. View all the records from table 5. Apply the aggregate functions (min, max, avg) 6. Using alter table command to change the column positions	CO1	L1: Understand	15

```
hive> create database library;
OK
Time taken: 0.19 seconds
```

```
hive> use library;
OK
Time taken: 0.058 seconds
hive> create table library(id int, name string, salary float)row format delimite
d fields terminated by ',';
OK
Time taken: 0.106 seconds
hive> show tables;
OK
library
Time taken: 0.051 seconds, Fetched: 1 row(s)
```

```
hive> insert into library values(1,'ekhlakh',10000),(2,'harsh',12000),(3,'satish
',15000),(4,'saroj',14000),(5,'hamid',11000);
Query ID = cloudera_20230223062828_aaab1ad8-a82f-40e4-ab15-9e9dd291be54
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1677172468763_0001, Tracking URL = http://quickstart.cloudera
:8088/proxy/application_1677172468763_0001/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1677172468763_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2023-02-23 06:29:02,269 Stage-1 map = 0%,  reduce = 0%
2023-02-23 06:29:11,111 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.41 se
c
MapReduce Total cumulative CPU time: 1 seconds 410 msec
Ended Job = job_1677172468763_0001
Stage-4 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
Stage-5 is filtered out by condition resolver.
Moving data to: hdfs://quickstart.cloudera:8020/user/hive/warehouse/library.db/l
ibrary/.hive-staging_hive_2023-02-23_06-28-40_841_4638907254667640290-1/-ext-100
00
Loading data to table library.library
Table library.library stats: [numFiles=1, numRows=5, totalSize=83, rawDataSize=7
8]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1   Cumulative CPU: 1.41 sec   HDFS Read: 4293 HDFS Write: 1
54 SUCCESS
Total MapReduce CPU Time Spent: 1 seconds 410 msec
OK
Time taken: 37.311 seconds
```

```
hive> select * from library;
OK
1       ekhlakh 10000.0
2       harsh   12000.0
3       satish  15000.0
4       saroj   14000.0
5       hamid   11000.0
Time taken: 1.312 seconds, Fetched: 5 row(s)
```

```
hive> select min(salary) from library;
Query ID = cloudera_20230223063131_7e27c3c9-8abc-43c2-ab30-46211c6b440f
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_1677172468763_0002, Tracking URL = http://quickstart.cloudera
:8088/proxy/application_1677172468763_0002/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1677172468763_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-02-23 06:31:34,742 Stage-1 map = 0%,  reduce = 0%
2023-02-23 06:31:45,031 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.32 se
c
2023-02-23 06:31:57,561 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.68
sec
MapReduce Total cumulative CPU time: 2 seconds 680 msec
Ended Job = job_1677172468763_0002
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 2.68 sec   HDFS Read: 7714 HD
FS Write: 8 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 680 msec
OK
10000.0
Time taken: 39.899 seconds, Fetched: 1 row(s)
```

```
hive> select max(salary) from library;
Query ID = cloudera_20230223063434_7e4c33e0-0acf-448c-9e99-28e92bc1f8cc
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_1677172468763_0003, Tracking URL = http://quickstart.cloudera
:8088/proxy/application_1677172468763_0003/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1677172468763_0003
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-02-23 06:34:47,882 Stage-1 map = 0%,  reduce = 0%
2023-02-23 06:34:56,105 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.22 se
c
2023-02-23 06:35:08,766 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.73
sec
MapReduce Total cumulative CPU time: 2 seconds 730 msec
Ended Job = job_1677172468763_0003
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 2.73 sec   HDFS Read: 7797 HD
FS Write: 8 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 730 msec
OK
15000.0
Time taken: 39.706 seconds, Fetched: 1 row(s)
```

```
hive> select avg(salary) from library;
Query ID = cloudera_20230223063636_6fc07a15-0a7d-49d7-996d-048147b55398
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_1677172468763_0004, Tracking URL = http://quickstart.cloudera
:8088/proxy/application_1677172468763_0004/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1677172468763_0004
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-02-23 06:36:25,636 Stage-1 map = 0%,  reduce = 0%
2023-02-23 06:36:33,614 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.21 se
c
2023-02-23 06:36:43,859 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.51
sec
MapReduce Total cumulative CPU time: 2 seconds 510 msec
Ended Job = job_1677172468763_0004
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 2.51 sec   HDFS Read: 8254 HD
FS Write: 8 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 510 msec
OK
12400.0
Time taken: 35.697 seconds, Fetched: 1 row(s)
```

```
hive> desc library;
OK
id                      int
name                    string
salary                  float
Time taken: 0.216 seconds, Fetched: 3 row(s)
hive> alter table library change column salary salary float after id;
OK
Time taken: 0.416 seconds
hive> desc library;
OK
id                      int
salary                  float
name                    string
Time taken: 0.119 seconds, Fetched: 3 row(s)
hive>
```

Q2	Hive Queries: 1. Create a database 2. Table creation (Eg: Employee) 3. Load data from local to hadoop 4. View all the records in the table 5. Apply the aggregate functions (avg, count and sum) 6. Using alter table command to change the column positions	CO2	L3: Apply	15

```
hive> use employee;
OK
Time taken: 0.1 seconds
hive> show tables;
OK
employee
Time taken: 0.041 seconds, Fetched: 1 row(s)
hive> load data local inpath '/home/cloudera/Desktop/emp.txt' into table employee;
Loading data to table employee.employee
Table employee.employee stats: [numFiles=1, totalSize=107]
OK
Time taken: 0.579 seconds
hive> select * from employee;
OK
1       ekhlakh ahmad   50000
2       saurabh kumar   60000
3       satish kumar    70000
4       sonapl parmar   80000
5       saroj kumar     90000
Time taken: 0.081 seconds, Fetched: 5 row(s)
hive>
```