

G. Pulla Reddy Engineering College (Autonomous): Kurnool

Department of Computer Science and Engineering

Big Data Technologies Lab (Scheme 2020)

Hive DML Commands

LOAD

SELECT

INSERT

DELETE

UPDATE

Steps to execute Lab program

1. Open terminal, type command **hive**
2. Enter following in hive

```
SET hive.support.concurrency=true;
SET hive.txn.manager=org.apache.hadoop.hive ql.lockmgr.DbTxnManager;
SET hive.enforce.bucketing=true;
SET hive.exec.dynamic.partition.mode=nostrict;
SET hive.compactor.initiator.on=true;
SET hive.compactor.worker.threads=1;
```
3. Create the following table in hive prompt

```
CREATE TABLE student_table (rollno int, name string,cgpa float)
PARTITIONED BY (admission_date date)
CLUSTERED BY(rollno) INTO 3 BUCKETS
STORED AS ORC TBLPROPERTIES ('transactional'='true');
```

4. Insert values into the table created

```
INSERT INTO student_table partition (admission_date='2020-04-01')
VALUES (120, 'sreedhar',9.8);
```

```
INSERT INTO student_table partition (admission_date='2020-04-02')
VALUES (121, 'sree',9.6);
```

```
INSERT INTO student_table partition (admission_date='2020-04-03')  
VALUES (122, 'chowdam',9.2);
```

```
INSERT INTO student_table partition (admission_date='2020-04-04')  
VALUES (123, 'sreedharc',8.8);
```

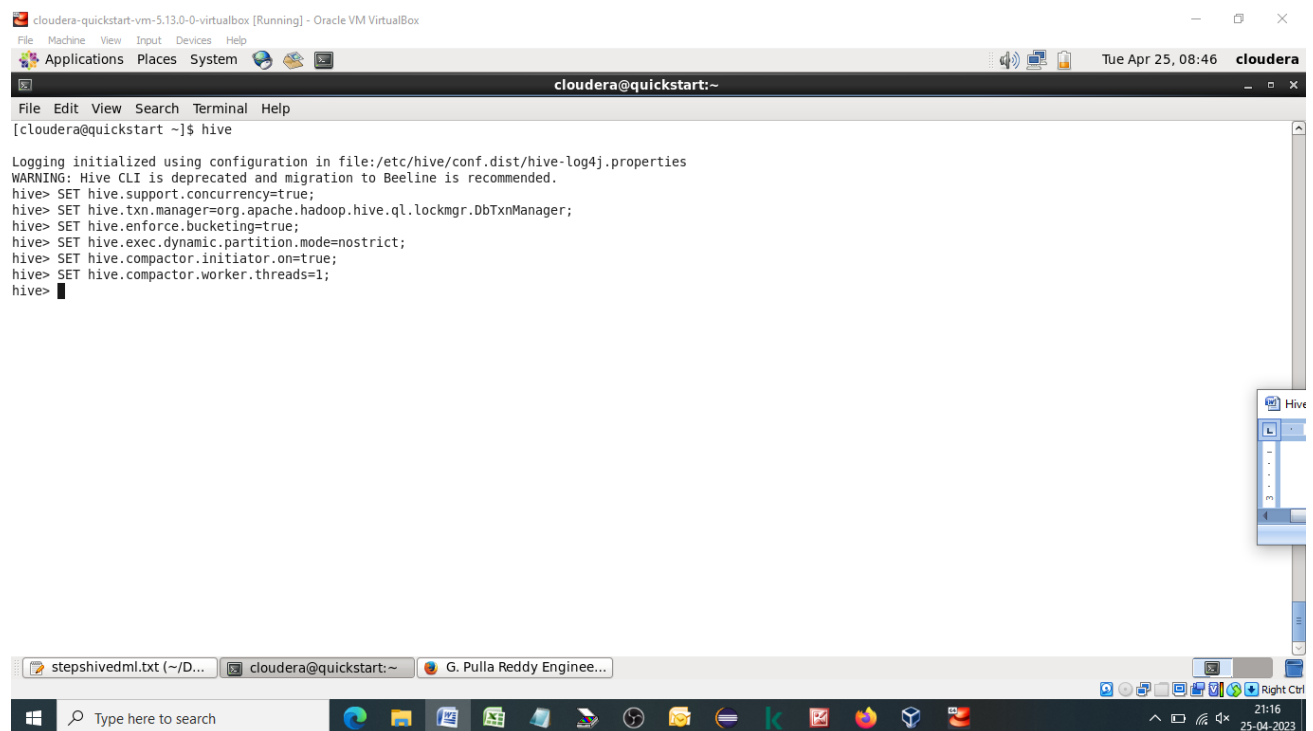
5. Update the table using the following

```
update student_table set cgpa=10 where rollno = 123;
```

6. Delete

```
DELETE FROM student_table WHERE rollno = 123;
```

Screenshots of the above steps



cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

Applications Places System

cloudera@quickstart:~

File Edit View Search Terminal Help

Time taken: 487.744 seconds

hive> select * from hivetable;

OK

2 10 2023-04-02

Time taken: 0.635 seconds, Fetched: 1 row(s)

hive>

hive> CREATE TABLE student_table (rollno int, name string, cgpa float)

> PARTITIONED BY (admission_date date)

> CLUSTERED BY(rollno) INTO 3 BUCKETS

> STORED AS ORC TBLPROPERTIES ('transactional'='true');

OK

Time taken: 0.477 seconds

hive> INSERT INTO student_table partition (admission_date='2020-04-01') VALUES (120, 'sreedhar',9.8);

Query ID = cloudera_20230425080808_f1a6c9c1-fc1b-4ec9-8282-6fc53fd5a784

Total jobs = 1

Launching Job 1 out of 1

Number of reduce tasks determined at compile time: 3

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.reducers=<number>

Starting Job = job_1682428922186_0005, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1682428922186_0005/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1682428922186_0005

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 3

2023-04-25 08:09:20,382 Stage-1 map = 0%, reduce = 0%

2023-04-25 08:10:20,955 Stage-1 map = 0%, reduce = 0%

2023-04-25 08:10:26,529 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 5.09 sec

Hive Enable and Use A... Gmail - Mozilla Firefox stepshivedml.txt (~/.D... cloudera@quickstart:~

Type here to search

Document1 - Microsoft Word

Home Insert Page Layout References Mailings

Calibri (Body) 11

Paste

Clipboa...

Font Paragraph

Hive DML Commands Lab Program

Page: 1 of 1 Words: 10

20:40 25-04-2023

cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

Applications Places System

cloudera@quickstart:~

File Edit View Search Terminal Help

eedhar',9.8);

Query ID = cloudera_20230425080808_f1a6c9c1-fc1b-4ec9-8282-6fc53fd5a784

Total jobs = 1

Launching Job 1 out of 1

Number of reduce tasks determined at compile time: 3

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.reducers=<number>

Starting Job = job_1682428922186_0005, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1682428922186_0005/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1682428922186_0005

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 3

2023-04-25 08:09:20,382 Stage-1 map = 0%, reduce = 0%

2023-04-25 08:10:20,955 Stage-1 map = 0%, reduce = 0%

2023-04-25 08:10:26,529 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 5.09 sec

2023-04-25 08:11:28,140 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 6.0 sec

2023-04-25 08:12:19,733 Stage-1 map = 100%, reduce = 22%, Cumulative CPU 8.23 sec

2023-04-25 08:12:21,802 Stage-1 map = 100%, reduce = 44%, Cumulative CPU 10.51 sec

2023-04-25 08:12:23,464 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 12.77 sec

2023-04-25 08:12:55,231 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 33.2 sec

MapReduce Total cumulative CPU time: 33 seconds 500 msec

Ended Job = job_1682428922186_0005

Loading data to table default.student_table partition (admission_date=2020-04-01)

Partition default.student_table(admission_date=2020-04-01) stats: [numFiles=3, numRows=1, totalSize=1206, rawDataSize=0]

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 3 Cumulative CPU: 33.5 sec HDFS Read: 17225 HDFS Write: 1467 SUCCESS

Total MapReduce CPU Time Spent: 33 seconds 500 msec

OK

Time taken: 277.976 seconds

hive>

stepshivedml.txt (~/.D... cloudera@quickstart:~ G. Pulla Reddy Enginee...

Type here to search

Document1 - Microsoft Word

Home Insert Page Layout References Mailings

Calibri (Body) 11

Paste

Clipboa...

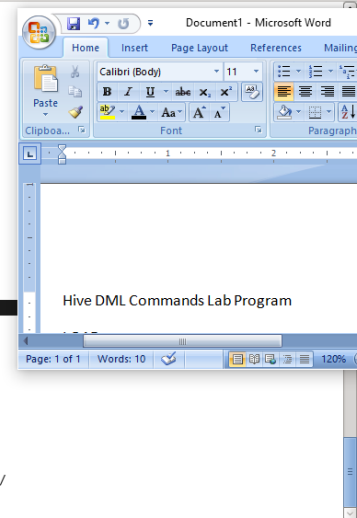
Font Paragraph

Hive DML Commands Lab Program

Page: 1 of 1 Words: 10

20:44 25-04-2023

```
cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
cloudera@quickstart:~$
File Edit View Search Terminal Help
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1682428922186_0005
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 3
2023-04-25 08:09:20,382 Stage-1 map = 0%, reduce = 0%
2023-04-25 08:10:20,955 Stage-1 map = 0%, reduce = 0%
2023-04-25 08:10:26,529 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 5.09 sec
2023-04-25 08:11:28,140 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 6.0 sec
2023-04-25 08:12:19,733 Stage-1 map = 100%, reduce = 22%, Cumulative CPU 8.23 sec
2023-04-25 08:12:21,802 Stage-1 map = 100%, reduce = 44%, Cumulative CPU 10.51 sec
2023-04-25 08:12:23,464 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 12.77 sec
2023-04-25 08:12:55,231 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 33.2 sec
MapReduce Total cumulative CPU time: 33 seconds 500 msec
Ended Job = job_1682428922186_0005
Loading data to table default.student_table partition (admission_date=2020-04-01)
Partition default.student_table(admission_date=2020-04-01) stats: [numFiles=3, numRows=1, totalSize=1206, rawDataSize=0]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 3 Cumulative CPU: 33.5 sec HDFS Read: 17225 HDFS Write: 1467 SUCCESS
Total MapReduce CPU Time Spent: 33 seconds 500 msec
OK
Time taken: 277.976 seconds
hive> INSERT INTO student_table partition (admission_date='2020-04-02') VALUES (121, 'sree',9.6);
Query ID = cloudera_20230425081414_3e609a8d-0117-4a22-8bbb-3c7f4a69aabe
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 3
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_1682428922186_0006, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1682428922186_0006/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1682428922186_0006
stepshivedml.txt (~D... cloudera@quickstart:~ G. Pulla Reddy Enginee...
```



```
cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
cloudera@quickstart:~$
File Edit View Search Terminal Help
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_1682428922186_0007, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1682428922186_0007/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1682428922186_0007
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 3
2023-04-25 08:19:22,729 Stage-1 map = 0%, reduce = 0%
2023-04-25 08:19:44,494 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.73 sec
2023-04-25 08:20:31,493 Stage-1 map = 100%, reduce = 22%, Cumulative CPU 7.1 sec
2023-04-25 08:20:34,211 Stage-1 map = 100%, reduce = 44%, Cumulative CPU 9.96 sec
2023-04-25 08:20:37,218 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 13.05 sec
2023-04-25 08:20:41,908 Stage-1 map = 100%, reduce = 78%, Cumulative CPU 15.88 sec
2023-04-25 08:20:44,820 Stage-1 map = 100%, reduce = 89%, Cumulative CPU 19.21 sec
2023-04-25 08:20:47,188 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 22.52 sec
MapReduce Total cumulative CPU time: 22 seconds 520 msec
Ended Job = job_1682428922186_0007
Loading data to table default.student_table partition (admission_date=2020-04-04)
Partition default.student_table(admission_date=2020-04-04) stats: [numFiles=3, numRows=1, totalSize=1212, rawDataSize=0]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 3 Cumulative CPU: 22.52 sec HDFS Read: 17226 HDFS Write: 1473 SUCCESS
Total MapReduce CPU Time Spent: 22 seconds 520 msec
OK
Time taken: 118.902 seconds
hive> select * from student_table;
OK
120 sreedhar 9.8 2020-04-01
121 sree 9.6 2020-04-02
123 sreedharc 8.8 2020-04-04
Time taken: 0.277 seconds, Fetched: 3 row(s)
hive>
stepshivedml.txt (~D... cloudera@quickstart:~ G. Pulla Reddy Enginee...
```

cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

Applications Places System

cloudera@quickstart:~

```
File Edit View Search Terminal Help
hive> select * from student_table;
OK
120 sreedhar 9.8 2020-04-01
121 sree 9.6 2020-04-02
123 sreedharc 8.8 2020-04-04
Time taken: 0.277 seconds, Fetched: 3 row(s)
hive> update student_table set cgpa=10 where rollno = 123;
Query ID = cloudera_20230425082222_8b5f62d7-f472-4180-b399-250651dccc4e
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 3
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_1682428922186_0008, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1682428922186_0008/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1682428922186_0008
Hadoop job information for Stage-1: number of mappers: 9; number of reducers: 3
2023-04-25 08:23:01,353 Stage-1 map = 0%, reduce = 0%
2023-04-25 08:24:02,937 Stage-1 map = 0%, reduce = 0%
2023-04-25 08:24:55,334 Stage-1 map = 44%, reduce = 0%, Cumulative CPU 26.65 sec
2023-04-25 08:24:56,534 Stage-1 map = 67%, reduce = 0%, Cumulative CPU 30.67 sec
2023-04-25 08:25:56,710 Stage-1 map = 67%, reduce = 0%, Cumulative CPU 30.67 sec
2023-04-25 08:26:27,433 Stage-1 map = 67%, reduce = 7%, Cumulative CPU 31.41 sec
2023-04-25 08:26:29,132 Stage-1 map = 67%, reduce = 15%, Cumulative CPU 32.03 sec
2023-04-25 08:26:30,423 Stage-1 map = 67%, reduce = 22%, Cumulative CPU 32.83 sec
2023-04-25 08:26:43,199 Stage-1 map = 89%, reduce = 22%, Cumulative CPU 48.6 sec
2023-04-25 08:26:44,915 Stage-1 map = 100%, reduce = 22%, Cumulative CPU 49.21 sec
2023-04-25 08:26:48,320 Stage-1 map = 100%, reduce = 52%, Cumulative CPU 50.98 sec
2023-04-25 08:26:49,584 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 51.89 sec
```

[stephivedml.txt (~/.D...)] cloudera@quickstart:~ G. Pulla Reddy Enginee...

Type here to search

20:56 25-04-2023

cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

Applications Places System

cloudera@quickstart:~

```
File Edit View Search Terminal Help
2023-04-25 08:24:02,937 Stage-1 map = 0%, reduce = 0%
2023-04-25 08:24:55,334 Stage-1 map = 44%, reduce = 0%, Cumulative CPU 26.65 sec
2023-04-25 08:24:56,534 Stage-1 map = 67%, reduce = 0%, Cumulative CPU 30.67 sec
2023-04-25 08:25:56,710 Stage-1 map = 67%, reduce = 0%, Cumulative CPU 30.67 sec
2023-04-25 08:26:27,433 Stage-1 map = 67%, reduce = 7%, Cumulative CPU 31.41 sec
2023-04-25 08:26:29,132 Stage-1 map = 67%, reduce = 15%, Cumulative CPU 32.03 sec
2023-04-25 08:26:30,423 Stage-1 map = 67%, reduce = 22%, Cumulative CPU 32.83 sec
2023-04-25 08:26:43,199 Stage-1 map = 89%, reduce = 22%, Cumulative CPU 48.6 sec
2023-04-25 08:26:44,915 Stage-1 map = 100%, reduce = 22%, Cumulative CPU 49.21 sec
2023-04-25 08:26:48,320 Stage-1 map = 100%, reduce = 52%, Cumulative CPU 50.98 sec
2023-04-25 08:26:49,584 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 51.89 sec
2023-04-25 08:27:07,912 Stage-1 map = 100%, reduce = 78%, Cumulative CPU 59.39 sec
2023-04-25 08:27:09,235 Stage-1 map = 100%, reduce = 89%, Cumulative CPU 62.44 sec
2023-04-25 08:27:10,448 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 66.44 sec
MapReduce Total cumulative CPU time: 1 minutes 6 seconds 440 msec
Ended Job = job_1682428922186_0008
Loading data to table default.student_table partition (admission_date=null)
Time taken for load dynamic partitions : 570
Loading partition (admission_date=2020-04-04)
Time taken for adding to write entity : 0
Partition default.student_table(admission_date=2020-04-04) stats: [numFiles=4, numRows=1, totalSize=1985, rawDataSize=0]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 9 Reduce: 3 Cumulative CPU: 66.44 sec HDFS Read: 73499 HDFS Write: 884 SUCCESS
Total MapReduce CPU Time Spent: 1 minutes 6 seconds 440 msec
OK
Time taken: 276.95 seconds
hive> select * from student_table;
OK
120 sreedhar 9.8 2020-04-01
121 sree 9.6 2020-04-02
123 sreedharc 10.0 2020-04-04
Time taken: 0.196 seconds, Fetched: 3 row(s)
hive>
```

[stephivedml.txt (~/.D...)] cloudera@quickstart:~ G. Pulla Reddy Enginee...

Type here to search

20:57 25-04-2023

```
cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System cloudera@quickstart:~
File Edit View Search Terminal Help
2023-04-25 08:27:10,448 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 66.44 sec
MapReduce Total cumulative CPU time: 1 minutes 6 seconds 440 msec
Ended Job = job_1682428922186_0008
Loading data to table default.student_table partition (admission_date=null)
Time taken for load dynamic partitions : 570
Loading partition {admission_date=2020-04-04}
Time taken for adding to write entity : 0
Partition default.student_table{admission_date=2020-04-04} stats: [numFiles=4, numRows=1, totalSize=1985, rawDataSize=0]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 9 Reduce: 3 Cumulative CPU: 66.44 sec HDFS Read: 73499 HDFS Write: 884 SUCCESS
Total MapReduce CPU Time Spent: 1 minutes 6 seconds 440 msec
OK
Time taken: 276.95 seconds
hive> select * from student_table;
OK
120 sreedhar 9.8 2020-04-01
121 sree 9.6 2020-04-02
123 sreedhar 10.0 2020-04-04
Time taken: 0.196 seconds, Fetched: 3 row(s)
hive> DELETE FROM student_table WHERE rollno = 123;
Query ID = cloudera_20230425082828_138da02b-1987-4680-9e39-a31eb253df09
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 3
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_1682428922186_0009, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1682428922186_0009/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1682428922186_0009
stepshivedml.txt (~/.D... cloudera@quickstart:~ G. Pulla Reddy Enginee...
Type here to search 20:58 25-04-2023
```

```
cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System cloudera@quickstart:~
File Edit View Search Terminal Help
2023-04-25 08:30:42,343 Stage-1 map = 22%, reduce = 0%, Cumulative CPU 24.37 sec
2023-04-25 08:30:43,946 Stage-1 map = 33%, reduce = 0%, Cumulative CPU 25.46 sec
2023-04-25 08:30:45,932 Stage-1 map = 44%, reduce = 0%, Cumulative CPU 26.62 sec
2023-04-25 08:30:50,266 Stage-1 map = 56%, reduce = 0%, Cumulative CPU 27.98 sec
2023-04-25 08:30:51,929 Stage-1 map = 67%, reduce = 0%, Cumulative CPU 29.56 sec
2023-04-25 08:31:52,441 Stage-1 map = 67%, reduce = 0%, Cumulative CPU 29.56 sec
2023-04-25 08:32:22,098 Stage-1 map = 67%, reduce = 15%, Cumulative CPU 31.28 sec
2023-04-25 08:32:28,071 Stage-1 map = 67%, reduce = 22%, Cumulative CPU 32.44 sec
2023-04-25 08:32:43,033 Stage-1 map = 89%, reduce = 22%, Cumulative CPU 49.57 sec
2023-04-25 08:32:45,875 Stage-1 map = 100%, reduce = 22%, Cumulative CPU 50.79 sec
2023-04-25 08:32:47,159 Stage-1 map = 100%, reduce = 37%, Cumulative CPU 50.96 sec
2023-04-25 08:32:49,136 Stage-1 map = 100%, reduce = 52%, Cumulative CPU 51.57 sec
2023-04-25 08:32:50,875 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 52.38 sec
2023-04-25 08:33:06,468 Stage-1 map = 100%, reduce = 89%, Cumulative CPU 61.58 sec
2023-04-25 08:33:07,757 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 64.56 sec
MapReduce Total cumulative CPU time: 1 minutes 4 seconds 560 msec
Ended Job = job_1682428922186_0009
Loading data to table default.student_table partition (admission_date=null)
Time taken for load dynamic partitions : 586
Loading partition {admission_date=2020-04-04}
Time taken for adding to write entity : 1
Partition default.student_table{admission_date=2020-04-04} stats: [numFiles=5, numRows=0, totalSize=2500, rawDataSize=0]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 9 Reduce: 3 Cumulative CPU: 64.56 sec HDFS Read: 70772 HDFS Write: 626 SUCCESS
Total MapReduce CPU Time Spent: 1 minutes 4 seconds 560 msec
OK
Time taken: 292.775 seconds
hive> select * from student_table;
OK
120 sreedhar 9.8 2020-04-01
121 sree 9.6 2020-04-02
Time taken: 0.692 seconds, Fetched: 2 row(s)
hive>
stepshivedml.txt (~/.D... cloudera@quickstart:~ G. Pulla Reddy Enginee...
Type here to search 21:13 25-04-2023
```