

Hive Understanding

Dr. A. Suresh, INSOFE

HIVE

The image shows a Jupyter Notebook interface. On the left, there's a sidebar with icons for File, Edit, View, Run, Kernel, Tabs, Settings, Help, and a file browser. The file browser shows a directory structure under '/'. In the center, there are two terminal tabs. The left tab shows a local session with commands like 'ssh' and 'hive'. The right tab shows a session on a remote host ('ilab9798') connected via SSH, displaying the output of 'hive' commands. The 'hive' command output includes a warning about a missing log4j property and a list of databases.

```
suresha@ilab9798:~$ ssh suresha@172.16.0.118
suresha@172.16.0.118's password:
Last login: Thu Jun 16 09:14:29 2022 from 172.16.0.67
[suresha@dn2 ~]$ hive
log4j:WARN No such property [maxFileSize] in org.apache.log4j.DailyRollingFileAppender.

Logging initialized using configuration in file:/etc/hive/2.6.5.0-292/0/hive-log4j.properties
hive> show databases;
OK
aj1
b
b104
b104_4174
b104_4179
b104_kiran
b104_sruthi
b109
b91_3963
b91_abhay
b91_hive
b91_madhu_db
b91_ravi
b92_hive
b92_hive_anita
```

HIVE Query

```
suresha@dn2:~          suresha@dn2:~          X
Time taken: 1.711 seconds, Fetched: 56 row(s)
hive> use org;
OK
Time taken: 0.29 seconds
hive> show tables;
OK
emp
emp1
emp7
salary
salary11
salary7
Time taken: 0.268 seconds, Fetched: 6 row(s)
hive> select count(*) from emp;
Query ID = suresha_20220616091728_5a799145-df72-49e4-92b6-e1cc42543e75
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_1654413090031_0002, Tracking URL = http://snn.insofe.edu.in:8088/proxy/application_1654413090031_0002/
Kill Command = /usr/hdp/2.6.5.0-292/hadoop/bin/hadoop job -kill job_1654413090031_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-06-16 09:17:44,198 Stage-1 map = 0%,  reduce = 0%
2022-06-16 09:17:50,666 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.14 sec
2022-06-16 09:17:57,086 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.76 sec
MapReduce Total cumulative CPU time: 4 seconds 760 msec
Ended Job = job_1654413090031_0002
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1    Cumulative CPU: 4.76 sec    HDFS Read: 8227 HDFS Write: 3 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 760 msec
OK
12
```

HIVE Query

```
$ suresha@dn2:~          X $ suresha@dn2:~          X  
  
MapReduce Jobs Launched:  
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.76 sec HDFS Read: 8227 HDFS Write: 3 SU  
CESS  
Total MapReduce CPU Time Spent: 4 seconds 760 msec  
OK  
12  
Time taken: 29.752 seconds, Fetched: 1 row(s)  
hive> select  
  > * from emp;  
OK  
100      zzz      zzz      Addrz      Cityz  
101      aaa      aaa      addr1      city1  
102      bbb      bbb      addr2      city2  
103      ccc      ccc      addr3      city3  
104      ddd      ddd      addr4      city4  
105      eee      eee      addr5      city5  
106      fff      fff      addr6      city6  
107      ggg      ggg      addr7      city7  
108      hhh      hhh      addr8      city8  
109      iii      iii      addr9      city9  
110      jjj      jjj      addr10     city10  
NULL      NULL     NULL     NULL      NULL  
Time taken: 0.158 seconds, Fetched: 12 row(s)  
hive>
```

Table Creation

```
$ suresha@dn2:~          $ suresha@dn2:~          $  
  
hive> create database batch109_org;  
OK  
Time taken: 0.196 seconds  
hive> use batch109_org;  
OK  
Time taken: 0.268 seconds  
hive> create table emp1(empid int,Fname string,Lname string,Address string,City string) row format delimited fi  
elds terminated by ',';  
OK  
Time taken: 0.311 seconds  
hive> show tables;  
OK  
emp1  
Time taken: 0.262 seconds, Fetched: 1 row(s)  
hive> insert into emp1 values(100,'zzz','zzz','addrz','cityz');  
Query ID = suresha_20220616092458_167acd22-a96a-4d30-b87c-8dbb9c544b92  
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_1654413090031_0003, Tracking URL = http://snn.insofe.edu.in:8088/proxy/application_165441309  
0031_0003/  
Kill Command = /usr/hdp/2.6.5.0-292/hadoop/bin/hadoop job -kill job_1654413090031_0003  
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0  
2022-06-16 09:25:08,728 Stage-1 map = 0%, reduce = 0%  
2022-06-16 09:25:15,071 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.87 sec  
MapReduce Total cumulative CPU time: 2 seconds 870 msec  
Ended Job = job_1654413090031_0003  
Stage-4 is selected by condition resolver.
```

Table Creation

```
suresha@dn2:~          X  suresha@dn2:~          X
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1   Cumulative CPU: 2.87 sec   HDFS Read: 5058 HDFS Write: 97 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 870 msec
OK
Time taken: 18.441 seconds
hive> insert into empl values(100,'zzz','zzz','addrz','cityz');
Query ID = suresha_20220616092517_a9630698-4b47-451b-9346-1f0962ff7091
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_1654413090031_0004, Tracking URL = http://snn.insofe.edu.in:8088/proxy/application_1654413090031_0004/
Kill Command = /usr/hdp/2.6.5.0-292/hadoop/bin/hadoop job -kill job_1654413090031_0004
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2022-06-16 09:25:27,314 Stage-1 map = 0%,  reduce = 0%
2022-06-16 09:25:33,684 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.07 sec
MapReduce Total cumulative CPU time: 3 seconds 70 msec
Ended Job = job_1654413090031_0004
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 directory hdfs://nn.insofe.edu.in:8020/apps/hive/warehouse/batch109_org.db/empl/.hive-staging_hive_2022-06-16_09-2
5-17_185_4723281556065089514-1-ext-10000
Loading data to table batch109_org.emp1
Table batch109_org.emp1 stats: [numFiles=2, numRows=2, totalSize=48, rawDataSize=46]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1   Cumulative CPU: 3.07 sec   HDFS Read: 5058 HDFS Write: 97 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 70 msec
OK
Time taken: 18.473 seconds
hive> select * from empl;
OK
100      zzz      zzz      addrz      cityz
100      zzz      zzz      addrz      cityz
Time taken: 0.154 seconds, Fetched: 2 row(s)
hive>
```

Table Creation

The screenshot shows a Jupyter Notebook interface. On the left, there's a file browser with a sidebar containing icons for various file types like CSV, IPYNB, and PDF. The main area has two terminal tabs both titled "suresha@dn2:~". The left terminal tab displays the output of the command "ls -al", listing numerous files and directories with their permissions, ownership, and timestamps. The right terminal tab displays the output of the command "cat myinput.txt", which contains a list of address and city pairs. The file "myinput.txt" is highlighted in the file browser.

```
[suresha@dn2 ~]$ ls -al
total 272
drwx----- 7 suresha suresha 274 Apr 30 17:01 .
drwxr-xr-x 130 root root 4096 Jun  8 21:38 ..
-rw----- 1 suresha suresha 8423 May  1 10:14 .bash_history
-rw-r--r-- 1 suresha suresha 18 Apr  1 2020 .bash_logout
-rw-r--r-- 1 suresha suresha 193 Apr  1 2020 .bash_profile
-rw-r--r-- 1 suresha suresha 231 Apr  1 2020 .bashrc
-rw-r--r-- 1 suresha suresha 16 Mar 10 20:11 bbb.txt
-rw-r--r-- 1 suresha suresha 16 Mar 10 20:11 b.txt
drwxrwxr-x 3 suresha suresha 18 Feb  9 12:22 .cache
drwxrwxr-x 3 suresha suresha 18 Feb  9 12:22 .config
-rw-rw-r-- 1 suresha suresha 8753 Jun 16 09:28 .hivehistory
drwxr-xr-x 4 suresha suresha 39 May 23 2021 .mozilla
-rw-rw-r-- 1 suresha suresha 214709 Feb 27 11:28 myinput3.csv
-rw-rw-r-- 1 suresha suresha 602 Mar 16 20:28 myinput.txt
drwxrwxr-x 2 suresha suresha 40 Feb  9 12:53 .oracle_jre_usage
-rw----- 1 suresha suresha 43 Mar  3 18:30 .scala_history
drwx----- 2 suresha suresha 25 Feb 10 12:07 .ssh

[suresha@dn2 ~]$ cat myinput.txt
101 aaa aaa addr1 city1
102 bbb bbb addr2 city2
103 ccc ccc addr3 city3
104 ddd ddd addr4 city4
105 eee eee addr5 city5
106 fff fff addr6 city6
107 ggg ggg addr7 city7
108 hhh hhh addr8 city8
109 iii iii addr9 city9
110 jjj jjj addr10 city10
101 aaa aaa addr1 city1
102 bbb bbb addr2 city2
```

Table Creation

The screenshot shows a Jupyter Notebook interface with a sidebar containing file and folder icons. The main area has two tabs open:

- suresha@dn2:~**: This tab displays a list of 110 rows of data from a file named "myinput109.txt". The columns are labeled 102 through 110 and contain values such as bbb, ccc, ddd, etc., followed by address and city names.
- suresha@dn2:~**: This tab shows terminal output. It starts with the command [suresha@dn2 ~]\$ scp myinput109.txt suresha@172.16.0.118:/home/suresh. It then prompts for a password. When the user types "myinput109.txt", it responds with "No such file or directory".

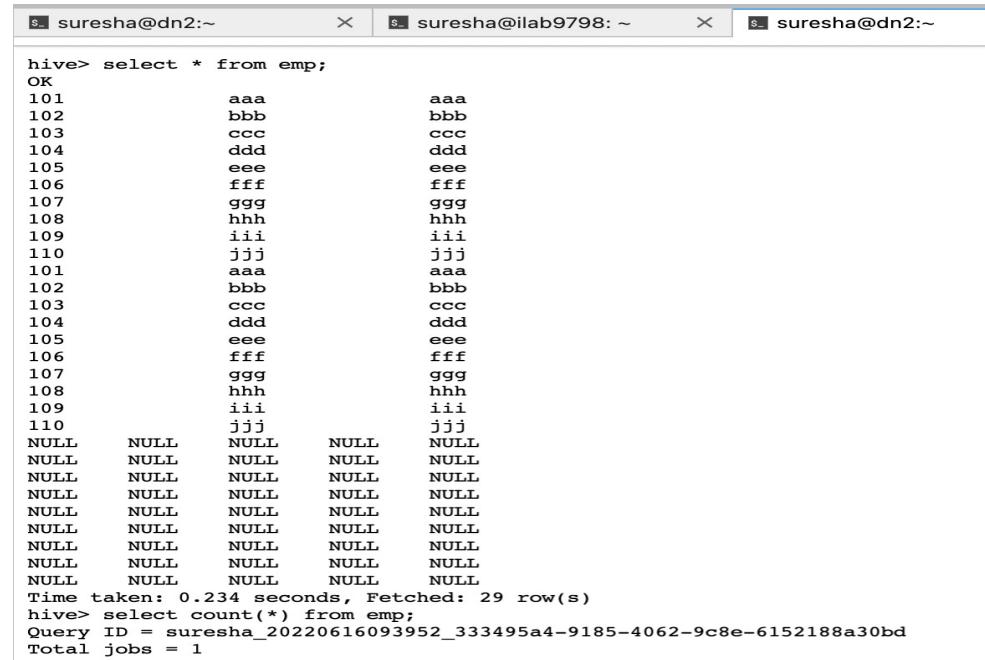
Column	Value 1	Value 2	Value 3	Value 4	Value 5
102	bbb	bbb	addr2	city2	
103	ccc	ccc	addr3	city3	
104	ddd	ddd	addr4	city4	
105	eee	eee	addr5	city5	
106	fff	fff	addr6	city6	
107	ggg	ggg	addr7	city7	
108	hhh	hhh	addr8	city8	
109	iii	iii	addr9	city9	
110	jjj	jjj	addr10	city10	

Table Creation

```
suresha@dn2:~          × | suresha@ilab9798: ~      × | suresha@dn2:~      × | myinput.txt      × |  
1 101 aaa aaa addr1 city1  
2 102 bbb bbb addr2 city2  
3 103 ccc ccc addr3 city3  
4 104 ddd ddd addr4 city4  
5 105 eee eee addr5 city5  
6 106 fff fff addr6 city6  
7 107 ggg ggg addr7 city7  
8 108 hhh hhh addr8 city8  
9 109 iii iii addr9 city9  
10 110 jjj jjj addr10 city10  
11 101 aaa aaa addr1 city1  
12 102 bbb bbb addr2 city2  
13 103 ccc ccc addr3 city3  
14 104 ddd ddd addr4 city4  
15 105 eee eee addr5 city5  
16 106 fff fff addr6 city6  
17 107 ggg ggg addr7 city7  
18 108 hhh hhh addr8 city8  
19 109 iii iii addr9 city9|  
20 110 jjj jjj addr10 city10  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30
```

Internal Table Creation

```
hive> create table emp(empid int,Fname string,Lname string,Address string,City string) row format delimited fields terminated by
' ';
OK
Time taken: 0.733 seconds
hive> load data local inpath '/home/suresha/myinput.txt'into table emp;
Loading data to table batch109_org.emp
Table batch109_org.emp stats: [numFiles=1, numRows=0, totalSize=602, rawDataSize=0]
OK
Time taken: 0.928 seconds
```



The screenshot shows three terminal windows side-by-side, all belonging to the user 'suresha'. The first window shows the creation of the 'emp' table and the loading of data from a local file. The second window shows the result of a SELECT * query, which returns 29 rows of data. The third window shows the result of a COUNT(*) query, which returns 29 rows.

```
suresha@dn2:~          suresha@ilab9798: ~          suresha@dn2:~
hive> select * from emp;
OK
101      aaa      aaa
102      bbb      bbb
103      ccc      ccc
104      ddd      ddd
105      eee      eee
106      fff      fff
107      ggg      ggg
108      hhh      hhh
109      iii      iiii
110      jjj      jjjj
101      aaa      aaa
102      bbb      bbb
103      ccc      ccc
104      ddd      ddd
105      eee      eee
106      fff      fff
107      ggg      ggg
108      hhh      hhh
109      iii      iiii
110      jjj      jjjj
NULL    NULL     NULL     NULL     NULL
Time taken: 0.234 seconds, Fetched: 29 row(s)
hive> select count(*) from emp;
Query ID = suresha_20220616093952_333495a4-9185-4062-9c8e-6152188a30bd
Total jobs = 1
```

Internal Table Creation

```
$ suresha@dn2:~          × | $ suresha@ilab9798: ~          × | $ suresha@dn2:~          × |   myinput.txt          × |  
NULL    NULL    NULL    NULL    NULL  
Time taken: 0.234 seconds, Fetched: 29 row(s)  
hive> select count(*) from emp;  
Query ID = suresha_20220616093952_333495a4-9185-4062-9c8e-6152188a30bd  
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_1654413090031_0005, Tracking URL = http://snn.insofe.edu.in:8088/proxy/application_1654413090031_0005/  
Kill Command = /usr/hdp/2.6.5.0-292/hadoop/bin/hadoop job -kill job_1654413090031_0005  
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1  
2022-06-16 09:39:59,542 Stage-1 map = 0%, reduce = 0%  
2022-06-16 09:40:04,922 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.25 sec  
2022-06-16 09:40:11,276 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.07 sec  
MapReduce Total cumulative CPU time: 5 seconds 70 msec  
Ended Job = job_1654413090031_0005  
MapReduce Jobs Launched:  
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 5.07 sec HDFS Read: 8525 HDFS Write: 3 SUCCESS  
Total MapReduce CPU Time Spent: 5 seconds 70 msec  
OK  
29  
Time taken: 19.731 seconds, Fetched: 1 row(s)
```

Internal Table Creation

File Edit View Run Kernel Tabs Settings Help

+ X ↗ C

/

Name	Last Modified
itsnew	4 months ago
new10	2 months ago
op.txt	3 months ago
WordCountSpark	4 months ago
WordCountSpark1	2 months ago
a.txt	2 months ago
aa.txt	4 months ago
examples.desktop	4 years ago
j.txt	4 months ago
myinput.txt	6 minutes ago
myinput1.txt	4 minutes ago
myinput3.csv	3 months ago
rdd.ipynb	3 months ago
rdd1.ipynb	3 months ago
rdd2.ipynb	3 months ago
suresha@172.16...	4 months ago
temp_data.txt	3 months ago
Untitled.ipynb	4 months ago
untitled.txt	2 months ago
Untitled1.ipynb	4 months ago

```
s suresha@dn2:~ suresha@ilab9798: ~ Terminal 35 myinput1.txt myinput.txt
create table salary(emplid int,Basesalary float,Grosssalary float) row format delimited fields terminated by ',';
OK
Time taken: 0.565 seconds
hive> load data local inpath '/home/suresha/myinput3.csv'into table salary;
Loading data to table batch109_org.salary
Table batch109_org.salary stats: [numFiles=1, numRows=0, totalSize=214709, rawDataSize=0]
OK
Time taken: 0.976 seconds
hive> select count(*) from salary;
Query ID = suresha_20220616095019_4998f744-0873-4d10-993d-ff517d4aaaac
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_1654413090031_0006, Tracking URL = http://snn.insofe.edu.in:8088/proxy/application_1654413090031_0006/
Kill Command = /usr/hdp/2.6.5.0-292/hadoop/bin/hadoop job -kill job_1654413090031_0006
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-06-16 09:50:30,896 Stage-1 map = 0%, reduce = 0%
2022-06-16 09:50:37,245 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.25 sec
2022-06-16 09:50:42,546 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.85 sec
MapReduce Total cumulative CPU time: 4 seconds 850 msec
Ended Job = job_1654413090031_0006
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.85 sec HDFS Read: 222623 HDFS Write: 5 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 850 msec
OK
9901
Time taken: 24.944 seconds, Fetched: 1 row(s)
hive>
```

External Table Creation

```
[suresha@dn2 ~]$ hdfs dfs -ls /user/suresha
Found 11 items
drwx-----  - suresha suresha          0 2022-06-16 11:50 /user/suresha/.Trash
drwxr-xr-x  - suresha suresha          0 2022-02-15 21:33 /user/suresha/.hiveJars
drwx-----  - suresha suresha          0 2022-06-16 09:50 /user/suresha/.staging
drwxr-xr-x  - suresha suresha          0 2022-03-10 20:14 /user/suresha/hdfs10
drwxr-xr-x  - suresha suresha          0 2022-02-17 19:35 /user/suresha/hdfs2
drwxr-xr-x  - suresha suresha          0 2022-02-25 12:07 /user/suresha/hdfs5
drwxr-xr-x  - suresha suresha          0 2022-03-10 20:38 /user/suresha/hdfsnew
-rw-r--r--  50 suresha suresha        614 2022-03-10 19:58 /user/suresha/myinput.txt
-rwxr-xr-x   3 suresha suresha       204 2022-02-17 20:22 /user/suresha/myinput1.txt
-rwxr-xr-x   3 suresha suresha  214709 2022-03-16 20:57 /user/suresha/myinput3.csv
drwxr-xr-x  - suresha suresha          0 2022-02-09 13:31 /user/suresha/newfile
[suresha@dn2 ~]$
```

External Table Creation

```
s suresha@dn2:~ x | s suresha@ilab9798: ~ x | $ Terminal 35 x

,' location '/user/suresha/';
OK
Time taken: 0.363 seconds
hive> load data inpath 'hdfs:/user/suresha/myinput3.csv' into table salary;
Loading data to table batch109_org.salary
Table batch109_org.salary stats: [numFiles=11, totalSize=645445]
OK
Time taken: 0.957 seconds
hive> select count(*) from salary;
Query ID = suresha_20220616115536_8f833adf-a5b1-4e4c-9b6a-d6bbc3c32677
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_1654413090031_0007, Tracking URL = http://snn.insofe.edu.in:8088/proxy/application_1654413090031_0007/
Kill Command = /usr/hdp/2.6.5.0-292/hadoop/bin/hadoop job -kill job_1654413090031_0007
Hadoop job information for Stage-1: number of mappers: 3; number of reducers: 1
2022-06-16 11:55:43,765 Stage-1 map = 0%, reduce = 0%
2022-06-16 11:55:50,105 Stage-1 map = 67%, reduce = 0%, Cumulative CPU 4.64 sec
2022-06-16 11:55:51,164 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 6.7 sec
2022-06-16 11:55:55,391 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 9.46 sec
MapReduce Total cumulative CPU time: 9 seconds 460 msec
Ended Job = job_1654413090031_0007
MapReduce Jobs Launched:
Stage-Stage-1: Map: 3 Reduce: 1 Cumulative CPU: 9.46 sec    HDFS Read: 231162 HDFS Write: 5 SUCCESS
Total MapReduce CPU Time Spent: 9 seconds 460 msec
OK
9944
Time taken: 20.867 seconds, Fetched: 1 row(s)
hive>
```

External Table Creation

S suresha@dn2:~ X S suresha@ilab9798: ~ X S Terminal 35 X myinput3.csv X

Delimiter: ,

	100	1000	1000
9881	9981	9882000	9882000
9882	9982	9883000	9883000
9883	9983	9884000	9884000
9884	9984	9885000	9885000
9885	9985	9886000	9886000
9886	9986	9887000	9887000
9887	9987	9888000	9888000
9888	9988	9889000	9889000
9889	9989	9890000	9890000
9890	9990	9891000	9891000
9891	9991	9892000	9892000
9892	9992	9893000	9893000
9893	9993	9894000	9894000
9894	9994	9895000	9895000
9895	9995	9896000	9896000
9896	9996	9897000	9897000
9897	9997	9898000	9898000
9898	9998	9899000	9899000
9899	9999	9900000	9900000
9900	10000	9901000	9901000

Join Query

```
s suresha@dn2:~ X | s suresha@ilab9798: ~ X | s Terminal 35 X | myinput3.csv X |  
  
hive> select emp.empid,emp.Fname,Grosssalary from emp inner join salary on emp.empid=salary.empid;  
Query ID = suresha_20220616120335_def46c3a-4762-4f1b-b6a1-3f317aa8d810  
Total jobs = 1  
Execution log at: /tmp/suresha/suresha_20220616120335_def46c3a-4762-4f1b-b6a1-3f317aa8d810.log  
2022-06-16 12:03:39 Starting to launch local task to process map join; maximum memory = 1046478848  
2022-06-16 12:03:40 Dump the side-table for tag: 0 with group count: 10 into file: file:/tmp/suresha/99444d2d-2b61-4ae8-8e19-58693e21e553/hive_2022-06-16_12-03-35_811_2154124648895580162-1-local-10003/HashTable-Stage-3/MapJoin-mapfile00--.hashtable  
2022-06-16 12:03:40 Uploaded 1 File to: file:/tmp/suresha/99444d2d-2b61-4ae8-8e19-58693e21e553/hive_2022-06-16_12-03-35_811_2154124648895580162-1-local-10003/HashTable-Stage-3/MapJoin-mapfile00--.hashtable (523 bytes)  
2022-06-16 12:03:40 End of local task; Time Taken: 1.244 sec.  
Execution completed successfully  
MapredLocal task succeeded  
Launching Job 1 out of 1  
Number of reduce tasks is set to 0 since there's no reduce operator  
Starting Job = job_1654413090031_0009, Tracking URL = http://snn.insofe.edu.in:8088/proxy/application_1654413090031_0009/  
Kill Command = /usr/hdp/2.6.5.0-292/hadoop/bin/hadoop job -kill job_1654413090031_0009  
Hadoop job information for Stage-3: number of mappers: 3; number of reducers: 0  
2022-06-16 12:03:48,518 Stage-3 map = 0%, reduce = 0%  
2022-06-16 12:03:54,863 Stage-3 map = 67%, reduce = 0%, Cumulative CPU 6.06 sec  
2022-06-16 12:03:55,925 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 9.37 sec  
MapReduce Total cumulative CPU time: 9 seconds 370 msec  
Ended Job = job_1654413090031_0009  
MapReduce Jobs Launched:  
Stage-Stage-3: Map: 3 Cumulative CPU: 9.37 sec HDFS Read: 236817 HDFS Write: 488 SUCCESS  
Total MapReduce CPU Time Spent: 9 seconds 370 msec  
OK  
101      2000.0  
101      2000.0  
102      3000.0  
102      3000.0  
103      4000.0  
103      4000.0  
104      5000.0  
104      5000.0
```

Static vs Dynamic Partition

```
create table emp_static_part(empid int,Fname string,Lname string,Address string)partitioned by (City string);
insert into table emp_static_part partition(City='city2') select empid,Fname,City from emp where City='city2';
select * from emp_static_part;
```

```
create table emp_dynamic_part(empid int,Fname string,Lname string,Address string)partitioned by (City string);
set hive.exec.dynamic.partition.mode=nonstrict;
insert into table emp_dynamic_part partition(City) select empid,Fname,City from emp where City='city2';
select * from emp_dynamic_part;
```

Static vs Dynamic Partition

```
set hive.exec.dynamic.partition=true;  
set hive.exec.dynamic.partition.mode=nonstrict;  
set hive.enforce.bucketing=true;
```

```
create table salary2(empid int,Basesalary float,Grosssalary float) row format delimited fields terminated by ',';
```

```
load data local inpath '/home/suresha/myinput3.csv'into table salary2;
```

Bucketing

```
suresha@dn2:~          X

9999    9900000.0      9900000.0
10000   9901000.0      9901000.0
Time taken: 0.521 seconds, Fetched: 9901 row(s)
hive> create table salary_bucket1(empid int,Basesalary float,Grosssalary float)clustered by(Basesalary) into 10 buckets;
OK
Time taken: 0.465 seconds
hive> insert into table salary_bucket1 select * from salary2;
Query ID = suresha_20220616140715_446c14d2-633a-45d5-af80-3fbe78b29ee0
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 10
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_1654413090031_0011, Tracking URL = http://snn.insofe.edu.in:8088/proxy/application_1654413090031_0011/
Kill Command = /usr/hdp/2.6.5.0-292/hadoop/bin/hadoop job -kill job_1654413090031_0011
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 10
2022-06-16 14:07:24,239 Stage-1 map = 0%,  reduce = 0%
2022-06-16 14:07:30,695 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.41 sec
2022-06-16 14:07:37,206 Stage-1 map = 100%,  reduce = 20%, Cumulative CPU 9.4 sec
2022-06-16 14:07:38,272 Stage-1 map = 100%,  reduce = 70%, Cumulative CPU 22.28 sec
2022-06-16 14:07:39,344 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 32.4 sec
MapReduce Total cumulative CPU time: 32 seconds 400 msec
Ended Job = job_1654413090031_0011
Loading data to table org109.salary_bucket1
Table org109.salary_bucket1 stats: [numFiles=10, numRows=9901, totalSize=244412, rawDataSize=234511]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 10  Cumulative CPU: 32.4 sec  HDFS Read: 259359 HDFS Write: 245057 SUCCESS
Total MapReduce CPU Time Spent: 32 seconds 400 msec
OK
Time taken: 27.363 seconds
hive>
```

Bucketing

```
hive> select max(Basesalary) from salary_bucket1 tablesample(bucket 5 out of 10 on rand())s;
Query ID = suresha_20220616141459_2706eccb-0c8e-46f8-9e2d-5fe7025be13d
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_1654413090031_0013, Tracking URL = http://snn.insofe.edu.in:8088/proxy/application_1654413090031_0013/
Kill Command = /usr/hdp/2.6.5.0-292/hadoop/bin/hadoop job -kill job_1654413090031_0013
Hadoop job information for Stage-1: number of mappers: 4; number of reducers: 1
2022-06-16 14:15:06,421 Stage-1 map = 0%,  reduce = 0%
2022-06-16 14:15:12,774 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 13.06 sec
2022-06-16 14:15:19,115 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 15.91 sec
MapReduce Total cumulative CPU time: 15 seconds 910 msec
Ended Job = job_1654413090031_0013
MapReduce Jobs Launched:
Stage-Stage-1: Map: 4  Reduce: 1  Cumulative CPU: 15.91 sec  HDFS Read: 269659 HDFS Write: 10 SUCCESS
Total MapReduce CPU Time Spent: 15 seconds 910 msec
OK
9877000.0
Time taken: 21.14 seconds, Fetched: 1 row(s)
L:--
```

*Q&A ?
Thank
You*

