INNER , EXTERNAL TABLE :

----------------------------------------------->

PARTITION AND NON PARTITION TABLE:

BY DEFAULT EACHE TABLE IS NON PARTITION TABLE

PARTITION IS SUB DIRECTORY IN TABLEL DIRECTORY IS CALLED AS PARATITIONS.

ADV OF PARTITION IS NO NEED TO SCAN ALL THE TABLE DATA.

=>CREATE TABLE EMP(id int,name string,sal int,sex string,dno int) row format delimited fields terminated by ',';

==> load data local inpath 'emp1' inot tableemp;

==> select \* from emp;

==> select \* from emp where sex='f';

NOTE :

ABOVE WICH QUIRY IS FAST FIRST ONCE IS FAST BECAUSE ALL ROW IS READ AND GIVEN OUTPUT.

BUT SECOND ONE IS READ ADN CHECK THE CRIATERIA THAN GIVE OUPUT . SO FIRST ONE IS GIVEN OUTPUT FAST.

======>

hive> create table pall(id int,name string,sal int,sex string,dno int) row format delimited fields terminated by ',';

OK

Time taken: 1.931 seconds

hive> load data local inpath 'emp' into table pall;

Loading data to table default.pall

Table default.pall stats: [numFiles=1, totalSize=226]

OK

Time taken: 4.149 seconds

hive> load data local inpath 'mep' into table pall;

Loading data to table default.pall

Table default.pall stats: [numFiles=2, totalSize=267]

OK

Time taken: 3.179 seconds

hive> select \* from pall;

OK

101 aaa 70000 m 12

102 bbb 90000 f 12

103 ccc 10000 m 11

104 ddd 40000 f 12

105 eeee 70000 m 13

106 de 80000 f 13

107 io 90000 m 14

108 yu 10000 f 14

109 aaa 60000 m 11

110 pio 60000 f 14

123 djd 90000 m 15

122 asasd 10000 15 NULL

101 subbu 70000 m 12

103 suji 80000 f 13

Time taken: 1.093 seconds, Fetched: 14 row(s)

==>hive> create table pall\_h(id int,name string,sal int,sex string,dno int) partitioned by (s string);

OK

==>hive> insert overwrite table pall\_h partition(s='f') select \* from pall where sex='f';

==>OK

Time taken: 128.965 seconds

hive> select \* from pall\_h;

OK

102 bbb 90000 f 12 f

104 ddd 40000 f 12 f

106 de 80000 f 13 f

108 yu 10000 f 14 f

110 pio 60000 f 14 f

103 suji 80000 f 13 f

Time taken: 0.672 seconds, Fetched: 6 row(s)

Query ID = cloudera\_20180213153232\_3ff73f0c-3ff6-46e7-9a0b-cd0c6c710a72

Total jobs = 3

Launching Job 1 out of 3

==>Total MapReduce CPU Time Spent: 10 seconds 850 msec

OK

Time taken: 78.341 seconds

==>insert overwrite table pall\_h partition(s='m') select \* from pall where sex='m';

hive> select \* from pall\_h;

OK

102 bbb 90000 f 12 f

104 ddd 40000 f 12 f

106 de 80000 f 13 f

108 yu 10000 f 14 f

110 pio 60000 f 14 f

103 suji 80000 f 13 f

101 aaa 70000 m 12 m

103 ccc 10000 m 11 m

105 eeee 70000 m 13 m

107 io 90000 m 14 m

109 aaa 60000 m 11 m

123 djd 90000 m 15 m

101 subbu 70000 m 12 m

Time taken: 0.296 seconds, Fetched: 13 row(s)

==>I WAND TO DNO DATA PARTITIONS

==>hive> create table sp\_sc(id int,name string,sal int,sex string,dno int) partitioned by (d int) row format delimited fields terminated by ',';

OK

==>insert overwrite table sp\_sc partition(d=11) select \* from pall where dno=11;

Time taken: 89.154 seconds

hive> select \* from sp\_sc;

OK

103 ccc 10000 m 11 11

109 aaa 60000 m 11 11

Time taken: 0.276 seconds, Fetched: 2 row(s)

hive> insert overwrite table sp\_sc partition(d=12) select \* from pall where dno=12;

Query ID = cloudera\_20180213161414\_d0ff7d4d-eaf6-4844-9b7a-968da3db12db

Time taken: 59.446 seconds

hive> select \* from sp\_sc;

OK

103 ccc 10000 m 11 11

109 aaa 60000 m 11 11

101 aaa 70000 m 12 12

102 bbb 90000 f 12 12

104 ddd 40000 f 12 12

101 subbu 70000 m 12 12

Time taken: 0.38 seconds, Fetched: 6 row(s)

hive> insert overwrite table sp\_sc partition(d=13) select \* from pall where dno=13;

hive> select \* from sp\_sc;

OK

103 ccc 10000 m 11 11

109 aaa 60000 m 11 11

101 aaa 70000 m 12 12

102 bbb 90000 f 12 12

104 ddd 40000 f 12 12

101 subbu 70000 m 12 12

105 eeee 70000 m 13 13

106 de 80000 f 13 13

103 suji 80000 f 13 13

Time taken: 0.35 seconds, Fetched: 9 row(s)

====================IN HDFS FIND THE RESULT====================

[cloudera@quickstart ~]$ hadoop fs -ls /user/hive/warehouse/pall

Found 2 items

-rwxrwxrwx 1 cloudera supergroup 226 2018-02-13 15:21 /user/hive/warehouse/pall/emp

-rwxrwxrwx 1 cloudera supergroup 41 2018-02-13 15:21 /user/hive/warehouse/pall/mep

[cloudera@quickstart ~]$ hadoop fs -ls /user/hive/warehouse/pall\_h

[cloudera@quickstart ~]$ hadoop fs -ls /user/hive/warehouse/pall\_h

Found 2 items

drwxrwxrwx - cloudera supergroup 0 2018-02-13 15:34 /user/hive/warehouse/pall\_h/s=f

drwxrwxrwx - cloudera supergroup 0 2018-02-13 15:38 /user/hive/warehouse/pall\_h/s=m

[cloudera@quickstart ~]$ hadoop fs -ls /user/hive/warehouse/pall\_h/s=f

Found 1 items

-rwxrwxrwx 1 cloudera supergroup 113 2018-02-13 15:33 /user/hive/warehouse/pall\_h/s=f/000000\_0

[cloudera@quickstart ~]$ hadoop fs -cat /user/hive/warehouse/pall\_h/s=f/000000\_0

102bbb90000f12

104ddd40000f12

106de80000f13

108yu10000f14

110pio60000f14

103suji80000f13

[cloudera@quickstart ~]$ hadoop fs -cat /user/hive/warehouse/pall\_h/s=m/000000\_1

cat: `/user/hive/warehouse/pall\_h/s=m/000000\_1': No such file or directory

[cloudera@quickstart ~]$ hadoop fs -ls /user/hive/warehouse/pall\_h/s=m

Found 1 items

-rwxrwxrwx 1 cloudera supergroup 135 2018-02-13 15:38 /user/hive/warehouse/pall\_h/s=m/000000\_0

[cloudera@quickstart ~]$ hadoop fs -cat /user/hive/warehouse/pall\_h/s=m/000000\_0

101aaa70000m12

103ccc10000m11

105eeee70000m13

107io90000m14

109aaa60000m11

123djd90000m15

101subbu70000m12

[cloudera@quickstart ~]$ hadoop fs -cat /user/hive/warehouse/pall\_h/

cat: `/user/hive/warehouse/pall\_h': Is a directory

[cloudera@quickstart ~]$ hadoop fs -ls /user/hive/warehouse/sp\_sc

Found 3 items

drwxrwxrwx - cloudera supergroup 0 2018-02-13 16:09 /user/hive/warehouse/sp\_sc/d=11

drwxrwxrwx - cloudera supergroup 0 2018-02-13 16:15 /user/hive/warehouse/sp\_sc/d=12

drwxrwxrwx - cloudera supergroup 0 2018-02-13 16:17 /user/hive/warehouse/sp\_sc/d=13

[cloudera@quickstart ~]$ hadoop fs -ls /user/hive/warehouse/sp\_sc/d=11

Found 1 items

-rwxrwxrwx 1 cloudera supergroup 38 2018-02-13 16:09 /user/hive/warehouse/sp\_sc/d=11/000000\_0

[cloudera@quickstart ~]$ hadoop fs -ls /user/hive/warehouse/sp\_sc/d=11/000000\_0

-rwxrwxrwx 1 cloudera supergroup 38 2018-02-13 16:09 /user/hive/warehouse/sp\_sc/d=11/000000\_0

[cloudera@quickstart ~]$ hadoop fs -cat /user/hive/warehouse/sp\_sc/d=11/000000\_0

103,ccc,10000,m,11

109,aaa,60000,m,11

=====IN STATIC PARTITION MULTIPEL COLUMNS IS NOR WORKING============

=====================DYNAMIC PARTITION=============================

==>hive> create table spa\_sca(id int,name string,sal int,sex string,dno int) partitioned by(d int, s string);

OK

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

hive> insert overwrite table spa\_sca partition(d, s) select \* from pall;

FAILED: SemanticException [Error 10096]: Dynamic partition strict mode requires at least one static partition column. To turn this off set hive.exec.dynamic.partition.mode=nonstrict

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

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

hive> insert overwrite table spa\_sca partition(d, s) select \* from pall;

FAILED: SemanticException [Error 10044]: Line 1:23 Cannot insert into target table because column number/types are different 's': Table insclause-0 has 7 columns, but query has 5 columns.

hive> insert overwrite table spa\_sca partition(d, s) select id,name,sal,sex,dno from pall;

FAILED: SemanticException [Error 10044]: Line 1:23 Cannot insert into target table because column number/types are different 's': Table insclause-0 has 7 columns, but query has 5 columns.

hive> insert overwrite table spa\_sca partition(d, s) select id,name,sal,sex,dno,dno,sex from pall;

Moving data to: hdfs://quickstart.cloudera:8020/user/hive/warehouse/spa\_sca/.hive-staging\_hive\_2018-02-13\_16-48-31\_762\_7442566470662907238-1/-ext-10000

Loading data to table default.spa\_sca partition (d=null, s=null)

Time taken for load dynamic partitions : 4895

Loading partition {d=\_\_HIVE\_DEFAULT\_PARTITION\_\_, s=15}

Loading partition {d=14, s=m}

Loading partition {d=13, s=m}

Loading partition {d=13, s=f}

Loading partition {d=11, s=m}

Loading partition {d=14, s=f}

Loading partition {d=12, s=m}

Loading partition {d=15, s=m}

Loading partition {d=12, s=f}

Time taken for adding to write entity : 12

Partition default.spa\_sca{d=11, s=m} stats: [numFiles=1, numRows=2, totalSize=38, rawDataSize=36]

Partition default.spa\_sca{d=12, s=f} stats: [numFiles=1, numRows=2, totalSize=38, rawDataSize=36]

Partition default.spa\_sca{d=12, s=m} stats: [numFiles=1, numRows=2, totalSize=40, rawDataSize=38]

Partition default.spa\_sca{d=13, s=f} stats: [numFiles=1, numRows=2, totalSize=38, rawDataSize=36]

Partition default.spa\_sca{d=13, s=m} stats: [numFiles=1, numRows=1, totalSize=20, rawDataSize=19]

Partition default.spa\_sca{d=14, s=f} stats: [numFiles=1, numRows=2, totalSize=37, rawDataSize=35]

Partition default.spa\_sca{d=14, s=m} stats: [numFiles=1, numRows=1, totalSize=18, rawDataSize=17]

Partition default.spa\_sca{d=15, s=m} stats: [numFiles=1, numRows=1, totalSize=19, rawDataSize=18]

Partition default.spa\_sca{d=\_\_HIVE\_DEFAULT\_PARTITION\_\_, s=15} stats: [numFiles=1, numRows=1, totalSize=22, rawDataSize=21]

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Cumulative CPU: 7.18 sec HDFS Read: 4183 HDFS Write: 719 SUCCESS

Total MapReduce CPU Time Spent: 7 seconds 180 msec

OK

Time taken: 58.948 seconds

hive> select \* from spa\_sca;

OK

103 ccc 10000 m 11 11 m

109 aaa 60000 m 11 11 m

102 bbb 90000 f 12 12 f

104 ddd 40000 f 12 12 f

101 aaa 70000 m 12 12 m

101 subbu 70000 m 12 12 m

106 de 80000 f 13 13 f

103 suji 80000 f 13 13 f

105 eeee 70000 m 13 13 m

108 yu 10000 f 14 14 f

110 pio 60000 f 14 14 f

107 io 90000 m 14 14 m

123 djd 90000 m 15 15 m

122 asasd 10000 15 NULL NULL 15

Time taken: 0.421 seconds, Fetched: 14 row(s)

===========================IN HDFS OUTPUT FOR DYNAMIC DATA================

[cloudera@quickstart ~]$ hadoop fs -ls /user/hive/warehouse/spa\_sca/

Found 6 items

drwxrwxrwx - cloudera supergroup 0 2018-02-13 16:49 /user/hive/warehouse/spa\_sca/d=11

drwxrwxrwx - cloudera supergroup 0 2018-02-13 16:49 /user/hive/warehouse/spa\_sca/d=12

drwxrwxrwx - cloudera supergroup 0 2018-02-13 16:49 /user/hive/warehouse/spa\_sca/d=13

drwxrwxrwx - cloudera supergroup 0 2018-02-13 16:49 /user/hive/warehouse/spa\_sca/d=14

drwxrwxrwx - cloudera supergroup 0 2018-02-13 16:49 /user/hive/warehouse/spa\_sca/d=15

drwxrwxrwx - cloudera supergroup 0 2018-02-13 16:49 /user/hive/warehouse/spa\_sca/d=\_\_HIVE\_DEFAULT\_PARTITION\_\_

[cloudera@quickstart ~]$ hadoop fs -ls /user/hive/warehouse/spa\_sca/d=\_\_HIVE\_DEFAULT\_PARTITION\_

ls: `/user/hive/warehouse/spa\_sca/d=\_\_HIVE\_DEFAULT\_PARTITION\_': No such file or directory

[cloudera@quickstart ~]$ hadoop fs -cat /user/hive/warehouse/spa\_sca/d=\_\_HIVE\_DEFAULT\_PARTITION\_

cat: `/user/hive/warehouse/spa\_sca/d=\_\_HIVE\_DEFAULT\_PARTITION\_': No such file or directory

[cloudera@quickstart ~]$ hadoop fs -ls /user/hive/warehouse/spa\_sca/d=11

Found 2 items

drwxrwxrwx - cloudera supergroup 0 2018-02-13 16:39 /user/hive/warehouse/spa\_sca/d=11/s=f

drwxrwxrwx - cloudera supergroup 0 2018-02-13 16:49 /user/hive/warehouse/spa\_sca/d=11/s=m

[cloudera@quickstart ~]$ hadoop fs -ls /user/hive/warehouse/spa\_sca/d=11/s=f

Found 1 items

-rwxrwxrwx 1 cloudera supergroup 0 2018-02-13 16:39 /user/hive/warehouse/spa\_sca/d=11/s=f/000000\_0

[cloudera@quickstart ~]$ hadoop fs -cat /user/hive/warehouse/spa\_sca/d=11/s=f/000000\_0

[cloudera@quickstart ~]$ hadoop fs -cat /user/hive/warehouse/spa\_sca/d=11/s=m

cat: `/user/hive/warehouse/spa\_sca/d=11/s=m': Is a directory

[cloudera@quickstart ~]$ hadoop fs -ls /user/hive/warehouse/spa\_sca/d=11/s=m

Found 1 items

-rwxrwxrwx 1 cloudera supergroup 38 2018-02-13 16:49 /user/hive/warehouse/spa\_sca/d=11/s=m/000000\_0

[cloudera@quickstart ~]$ hadoop fs -cat /user/hive/warehouse/spa\_sca/d=11/s=m/000000\_0

103ccc10000m11

109aaa60000m11

[cloudera@quickstart ~]$ hadoop fs -ls /user/hive/warehouse/spa\_sca/d=12/

Found 2 items

drwxrwxrwx - cloudera supergroup 0 2018-02-13 16:49 /user/hive/warehouse/spa\_sca/d=12/s=f

drwxrwxrwx - cloudera supergroup 0 2018-02-13 16:49 /user/hive/warehouse/spa\_sca/d=12/s=m

[cloudera@quickstart ~]$ hadoop fs -ls /user/hive/warehouse/spa\_sca/d=12/s=f

Found 1 items

-rwxrwxrwx 1 cloudera supergroup 38 2018-02-13 16:49 /user/hive/warehouse/spa\_sca/d=12/s=f/000000\_0

[cloudera@quickstart ~]$ hadoop fs -cat /user/hive/warehouse/spa\_sca/d=12/s=f/000000\_0

102bbb90000f12

104ddd40000f12

[cloudera@quickstart ~]$ hadoop fs -cat /user/hive/warehouse/spa\_sca/d=12/s=m/000000\_0

101aaa70000m12

101subbu70000m12

===============THIS IS ONE TYPE OF EXCERSICE OF DATA FLOW OF PROCESSIGN===========

==>hive> create table dt\_mo(dt string,amount int)row format delimited fields terminated by ',';

OK

Time taken: 0.212 seconds

hive> load data local inpath 'subbu' into table dt\_mo;

Loading data to table default.dt\_mo

Table default.dt\_mo stats: [numFiles=1, totalSize=408]

OK

Time taken: 0.893 seconds

hive> select \* from dt\_mo;

OK

01/01/2011 45000

01/01/2011 46000

01/23/2011 70000

02/01/2011 45000

03/01/2011 46000

03/23/2011 90000

04/01/2011 45000

04/01/2011 46000

05/23/2011 70000

06/01/2011 45000

06/01/2011 46000

06/23/2011 70000

07/01/2011 45000

08/01/2011 46000

08/23/2011 70000

09/01/2011 45000

10/01/2011 46000

10/23/2011 70000

11/01/2011 45000

11/01/2011 46000

11/23/2011 70000

12/01/2011 45000

12/01/2011 46000

12/23/2011 70000

Time taken: 0.207 seconds, Fetched: 24 row(s)

hive> create table dt\_mo1 (dt Array<string>,amount int);

OK

hive> insert into table dt\_mo1 select split(dt, '/'),amount from dt\_mo;

OR

hive> insert overwrite table dt\_mo1 select split(dt, '/'),amount from dt\_mo;

hive> select \* from dt\_mo1;

OK

["01","01","2011"] 45000

["01","01","2011"] 46000

["01","23","2011"] 70000

["02","01","2011"] 45000

["03","01","2011"] 46000

["03","23","2011"] 90000

["04","01","2011"] 45000

["04","01","2011"] 46000

["05","23","2011"] 70000

["06","01","2011"] 45000

["06","01","2011"] 46000

["06","23","2011"] 70000

["07","01","2011"] 45000

["08","01","2011"] 46000

["08","23","2011"] 70000

["09","01","2011"] 45000

["10","01","2011"] 46000

["10","23","2011"] 70000

["11","01","2011"] 45000

["11","01","2011"] 46000

["11","23","2011"] 70000

["12","01","2011"] 45000

["12","01","2011"] 46000

["12","23","2011"] 70000

Time taken: 0.258 seconds, Fetched: 24 row(s)

hive> select concat(dt[2],'-',dt[0],'-',dt[1]) ,amount from dt\_mo1;

OK

2011-01-01 45000

2011-01-01 46000

2011-01-23 70000

2011-02-01 45000

2011-03-01 46000

2011-03-23 90000

2011-04-01 45000

2011-04-01 46000

2011-05-23 70000

2011-06-01 45000

2011-06-01 46000

2011-06-23 70000

2011-07-01 45000

2011-08-01 46000

2011-08-23 70000

2011-09-01 45000

2011-10-01 46000

2011-10-23 70000

2011-11-01 45000

2011-11-01 46000

2011-11-23 70000

2011-12-01 45000

2011-12-01 46000

2011-12-23 70000

Time taken: 0.477 seconds, Fetched: 24 row(s)

hive>create table dt\_mo2 like dt\_mo;

hive> insert into table dt\_mo2 select dt,amount from dt\_mo

> union all

> select concat(substr(dt,1,9),'2') as dt,amount+2000 as amount from dt\_mo

> union all

> select concat(substr(dt,1,9),'3') as dt,amount+10000 as amount from dt\_mo;

Query ID = cloudera\_20180213213333\_78fed37c-787f-433e-9b38-3d4eae880217

Total jobs = 3

hive> select \* from dt\_mo2;

OK

01/01/2011 45000

01/01/2012 47000

01/01/2013 55000

01/01/2011 46000

01/01/2012 48000

01/01/2013 56000

01/23/2011 70000

01/23/2012 72000

01/23/2013 80000

02/01/2011 45000

02/01/2012 47000

02/01/2013 55000

03/01/2011 46000

03/01/2012 48000

03/01/2013 56000

03/23/2011 90000

03/23/2012 92000

03/23/2013 100000

04/01/2011 45000

04/01/2012 47000

04/01/2013 55000

04/01/2011 46000

04/01/2012 48000

04/01/2013 56000

05/23/2011 70000

05/23/2012 72000

05/23/2013 80000

06/01/2011 45000

06/01/2012 47000

06/01/2013 55000

06/01/2011 46000

06/01/2012 48000

06/01/2013 56000

06/23/2011 70000

06/23/2012 72000

06/23/2013 80000

07/01/2011 45000

07/01/2012 47000

07/01/2013 55000

08/01/2011 46000

08/01/2012 48000

08/01/2013 56000

08/23/2011 70000

08/23/2012 72000

08/23/2013 80000

09/01/2011 45000

09/01/2012 47000

09/01/2013 55000

10/01/2011 46000

10/01/2012 48000

10/01/2013 56000

10/23/2011 70000

10/23/2012 72000

10/23/2013 80000

11/01/2011 45000

11/01/2012 47000

11/01/2013 55000

11/01/2011 46000

11/01/2012 48000

11/01/2013 56000

11/23/2011 70000

11/23/2012 72000

11/23/2013 80000

12/01/2011 45000

12/01/2012 47000

12/01/2013 55000

12/01/2011 46000

12/01/2012 48000

12/01/2013 56000

12/23/2011 70000

12/23/2012 72000

12/23/2013 80000

Time taken: 0.381 seconds, Fetched: 72 row(s)

hive> create table dt\_mo4 (dt Array<string>,amount int);

OK

hive> insert overwrite table dt\_mo4 select split(dt,'/'),amount from dt\_mo2;

Query ID = cloudera\_20180213215858\_041ac62f-54b8-485e-b663-4cb5a03f50cf

Total MapReduce CPU Time Spent: 9 seconds 500 msec

OK

Time taken: 94.441 seconds

hive> select \* from dt\_mo4;

OK

["01","01","2011"] 45000

["01","01","2012"] 47000

["01","01","2013"] 55000

["01","01","2011"] 46000

["01","01","2012"] 48000

["01","01","2013"] 56000

["01","23","2011"] 70000

["01","23","2012"] 72000

["01","23","2013"] 80000

["02","01","2011"] 45000

["02","01","2012"] 47000

["02","01","2013"] 55000

["03","01","2011"] 46000

["03","01","2012"] 48000

["03","01","2013"] 56000

["03","23","2011"] 90000

["03","23","2012"] 92000

["03","23","2013"] 100000

["04","01","2011"] 45000

["04","01","2012"] 47000

["04","01","2013"] 55000

["04","01","2011"] 46000

["04","01","2012"] 48000

["04","01","2013"] 56000

["05","23","2011"] 70000

["05","23","2012"] 72000

["05","23","2013"] 80000

["06","01","2011"] 45000

["06","01","2012"] 47000

["06","01","2013"] 55000

["06","01","2011"] 46000

["06","01","2012"] 48000

["06","01","2013"] 56000

["06","23","2011"] 70000

["06","23","2012"] 72000

["06","23","2013"] 80000

["07","01","2011"] 45000

["07","01","2012"] 47000

["07","01","2013"] 55000

["08","01","2011"] 46000

["08","01","2012"] 48000

["08","01","2013"] 56000

["08","23","2011"] 70000

["08","23","2012"] 72000

["08","23","2013"] 80000

["09","01","2011"] 45000

["09","01","2012"] 47000

["09","01","2013"] 55000

["10","01","2011"] 46000

["10","01","2012"] 48000

["10","01","2013"] 56000

["10","23","2011"] 70000

["10","23","2012"] 72000

["10","23","2013"] 80000

["11","01","2011"] 45000

["11","01","2012"] 47000

["11","01","2013"] 55000

["11","01","2011"] 46000

["11","01","2012"] 48000

["11","01","2013"] 56000

["11","23","2011"] 70000

["11","23","2012"] 72000

["11","23","2013"] 80000

["12","01","2011"] 45000

["12","01","2012"] 47000

["12","01","2013"] 55000

["12","01","2011"] 46000

["12","01","2012"] 48000

["12","01","2013"] 56000

["12","23","2011"] 70000

["12","23","2012"] 72000

["12","23","2013"] 80000

Time taken: 0.21 seconds, Fetched: 72 row(s)

hive> select concat(dt[2],'-',dt[0],'-',dt[1]),amount from dt\_mo4;

OK

2011-01-01 45000

2012-01-01 47000

2013-01-01 55000

2011-01-01 46000

2012-01-01 48000

2013-01-01 56000

2011-01-23 70000

2012-01-23 72000

2013-01-23 80000

2011-02-01 45000

2012-02-01 47000

2013-02-01 55000

2011-03-01 46000

2012-03-01 48000

2013-03-01 56000

2011-03-23 90000

2012-03-23 92000

2013-03-23 100000

2011-04-01 45000

2012-04-01 47000

2013-04-01 55000

2011-04-01 46000

2012-04-01 48000

2013-04-01 56000

2011-05-23 70000

2012-05-23 72000

2013-05-23 80000

2011-06-01 45000

2012-06-01 47000

2013-06-01 55000

2011-06-01 46000

2012-06-01 48000

2013-06-01 56000

2011-06-23 70000

2012-06-23 72000

2013-06-23 80000

2011-07-01 45000

2012-07-01 47000

2013-07-01 55000

2011-08-01 46000

2012-08-01 48000

2013-08-01 56000

2011-08-23 70000

2012-08-23 72000

2013-08-23 80000

2011-09-01 45000

2012-09-01 47000

2013-09-01 55000

2011-10-01 46000

2012-10-01 48000

2013-10-01 56000

2011-10-23 70000

2012-10-23 72000

2013-10-23 80000

2011-11-01 45000

2012-11-01 47000

2013-11-01 55000

2011-11-01 46000

2012-11-01 48000

2013-11-01 56000

2011-11-23 70000

2012-11-23 72000

2013-11-23 80000

2011-12-01 45000

2012-12-01 47000

2013-12-01 55000

2011-12-01 46000

2012-12-01 48000

2013-12-01 56000

2011-12-23 70000

2012-12-23 72000

2013-12-23 80000

Time taken: 0.214 seconds, Fetched: 72 row(s)

hive> create table sales\_dt(dt string,amount int);

OK

Time taken: 0.307 seconds

hive> insert overwrite table sales\_dt select concat(dt[2],'-',dt[0],'-',dt[1]),amount from dt\_mo4;

hive> select \*from sales\_dt;

OK

2011-01-01 45000

2012-01-01 47000

2013-01-01 55000

2011-01-01 46000

2012-01-01 48000

2013-01-01 56000

2011-01-23 70000

2012-01-23 72000

2013-01-23 80000

2011-02-01 45000

2012-02-01 47000

2013-02-01 55000

2011-03-01 46000

2012-03-01 48000

2013-03-01 56000

2011-03-23 90000

2012-03-23 92000

2013-03-23 100000

2011-04-01 45000

2012-04-01 47000

2013-04-01 55000

2011-04-01 46000

2012-04-01 48000

2013-04-01 56000

2011-05-23 70000

2012-05-23 72000

2013-05-23 80000

2011-06-01 45000

2012-06-01 47000

2013-06-01 55000

2011-06-01 46000

2012-06-01 48000

2013-06-01 56000

2011-06-23 70000

2012-06-23 72000

2013-06-23 80000

2011-07-01 45000

2012-07-01 47000

2013-07-01 55000

2011-08-01 46000

2012-08-01 48000

2013-08-01 56000

2011-08-23 70000

2012-08-23 72000

2013-08-23 80000

2011-09-01 45000

2012-09-01 47000

2013-09-01 55000

2011-10-01 46000

2012-10-01 48000

2013-10-01 56000

2011-10-23 70000

2012-10-23 72000

2013-10-23 80000

2011-11-01 45000

2012-11-01 47000

2013-11-01 55000

2011-11-01 46000

2012-11-01 48000

2013-11-01 56000

2011-11-23 70000

2012-11-23 72000

2013-11-23 80000

2011-12-01 45000

2012-12-01 47000

2013-12-01 55000

2011-12-01 46000

2012-12-01 48000

2013-12-01 56000

2011-12-23 70000

2012-12-23 72000

2013-12-23 80000

Time taken: 0.207 seconds, Fetched: 72 row(s)

hive> create table dollos(dt string,amont int)partitioned by(y int,m int,d int) row format delimited fields terminated by',';

OK

Time taken: 0.374 seconds

hive> insert overwrite table dollos partition(y, m, d) select dt,amount,year(dt),month(dt),day(dt) from sales\_dt;

Query ID = cloudera\_20180213231717\_bc96b1cd-b276-4e97-8939-0d48d1b4d48b

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\_1518461294181\_0020, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1518461294181\_0020/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1518461294181\_0020

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0

2018-02-13 23:17:49,210 Stage-1 map = 0%, reduce = 0%

2018-02-13 23:18:14,250 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 9.21 sec

MapReduce Total cumulative CPU time: 9 seconds 210 msec

Ended Job = job\_1518461294181\_0020

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/dollos/.hive-staging\_hive\_2018-02-13\_23-17-13\_202\_69936126987861718-1/-ext-10000

Loading data to table default.dollos partition (y=null, m=null, d=null)

Time taken for load dynamic partitions : 31372

Loading partition {y=2013, m=7, d=1}

Loading partition {y=2013, m=8, d=23}

Loading partition {y=2012, m=1, d=1}

Loading partition {y=2011, m=1, d=23}

Loading partition {y=2012, m=1, d=23}

Loading partition {y=2011, m=3, d=1}

Loading partition {y=2013, m=1, d=23}

Loading partition {y=2012, m=11, d=23}

Loading partition {y=2013, m=11, d=1}

Loading partition {y=2013, m=6, d=23}

Loading partition {y=2011, m=8, d=1}

Loading partition {y=2012, m=8, d=1}

Loading partition {y=2011, m=11, d=1}

Loading partition {y=2013, m=6, d=1}

Loading partition {y=2013, m=9, d=1}

Loading partition {y=2013, m=3, d=23}

Loading partition {y=2012, m=6, d=23}

Loading partition {y=2013, m=11, d=23}

Loading partition {y=2012, m=6, d=1}

Loading partition {y=2013, m=12, d=23}

Loading partition {y=2011, m=3, d=23}

Loading partition {y=2012, m=7, d=1}

Loading partition {y=2013, m=10, d=23}

Loading partition {y=2011, m=6, d=23}

Loading partition {y=2013, m=8, d=1}

Loading partition {y=2011, m=6, d=1}

Loading partition {y=2012, m=10, d=1}

Loading partition {y=2011, m=12, d=1}

Loading partition {y=2012, m=11, d=1}

Loading partition {y=2012, m=12, d=23}

Loading partition {y=2012, m=10, d=23}

Loading partition {y=2011, m=4, d=1}

Loading partition {y=2012, m=3, d=23}

Loading partition {y=2011, m=11, d=23}

Loading partition {y=2011, m=7, d=1}

Loading partition {y=2013, m=1, d=1}

Loading partition {y=2013, m=2, d=1}

Loading partition {y=2012, m=2, d=1}

Loading partition {y=2011, m=2, d=1}

Loading partition {y=2013, m=12, d=1}

Loading partition {y=2012, m=9, d=1}

Loading partition {y=2013, m=5, d=23}

Loading partition {y=2011, m=9, d=1}

Loading partition {y=2011, m=8, d=23}

Loading partition {y=2011, m=1, d=1}

Loading partition {y=2012, m=12, d=1}

Loading partition {y=2011, m=5, d=23}

Loading partition {y=2012, m=8, d=23}

Loading partition {y=2011, m=12, d=23}

Loading partition {y=2013, m=10, d=1}

Loading partition {y=2013, m=3, d=1}

Loading partition {y=2013, m=4, d=1}

Loading partition {y=2011, m=10, d=23}

Loading partition {y=2012, m=3, d=1}

Loading partition {y=2012, m=4, d=1}

Loading partition {y=2012, m=5, d=23}

Loading partition {y=2011, m=10, d=1}

Time taken for adding to write entity : 75

Partition default.dollos{y=2011, m=1, d=1} stats: [numFiles=1, numRows=2, totalSize=34, rawDataSize=32]

Partition default.dollos{y=2011, m=1, d=23} stats: [numFiles=1, numRows=1, totalSize=17, rawDataSize=16]

Partition default.dollos{y=2011, m=10, d=1} stats: [numFiles=1, numRows=1, totalSize=17, rawDataSize=16]

Partition default.dollos{y=2011, m=10, d=23} stats: [numFiles=1, numRows=1, totalSize=17, rawDataSize=16]

Partition default.dollos{y=2011, m=11, d=1} stats: [numFiles=1, numRows=2, totalSize=34, rawDataSize=32]

Partition default.dollos{y=2011, m=11, d=23} stats: [numFiles=1, numRows=1, totalSize=17, rawDataSize=16]

Partition default.dollos{y=2011, m=12, d=1} stats: [numFiles=1, numRows=2, totalSize=34, rawDataSize=32]

Partition default.dollos{y=2011, m=12, d=23} stats: [numFiles=1, numRows=1, totalSize=17, rawDataSize=16]

Partition default.dollos{y=2011, m=2, d=1} stats: [numFiles=1, numRows=1, totalSize=17, rawDataSize=16]

Partition default.dollos{y=2011, m=3, d=1} stats: [numFiles=1, numRows=1, totalSize=17, rawDataSize=16]

Partition default.dollos{y=2011, m=3, d=23} stats: [numFiles=1, numRows=1, totalSize=17, rawDataSize=16]

Partition default.dollos{y=2011, m=4, d=1} stats: [numFiles=1, numRows=2, totalSize=34, rawDataSize=32]

Partition default.dollos{y=2011, m=5, d=23} stats: [numFiles=1, numRows=1, totalSize=17, rawDataSize=16]

Partition default.dollos{y=2011, m=6, d=1} stats: [numFiles=1, numRows=2, totalSize=34, rawDataSize=32]

Partition default.dollos{y=2011, m=6, d=23} stats: [numFiles=1, numRows=1, totalSize=17, rawDataSize=16]

Partition default.dollos{y=2011, m=7, d=1} stats: [numFiles=1, numRows=1, totalSize=17, rawDataSize=16]

Partition default.dollos{y=2011, m=8, d=1} stats: [numFiles=1, numRows=1, totalSize=17, rawDataSize=16]

Partition default.dollos{y=2011, m=8, d=23} stats: [numFiles=1, numRows=1, totalSize=17, rawDataSize=16]

Partition default.dollos{y=2011, m=9, d=1} stats: [numFiles=1, numRows=1, totalSize=17, rawDataSize=16]

Partition default.dollos{y=2012, m=1, d=1} stats: [numFiles=1, numRows=2, totalSize=34, rawDataSize=32]

Partition default.dollos{y=2012, m=1, d=23} stats: [numFiles=1, numRows=1, totalSize=17, rawDataSize=16]

Partition default.dollos{y=2012, m=10, d=1} stats: [numFiles=1, numRows=1, totalSize=17, rawDataSize=16]

Partition default.dollos{y=2012, m=10, d=23} stats: [numFiles=1, numRows=1, totalSize=17, rawDataSize=16]

Partition default.dollos{y=2012, m=11, d=1} stats: [numFiles=1, numRows=2, totalSize=34, rawDataSize=32]

Partition default.dollos{y=2012, m=11, d=23} stats: [numFiles=1, numRows=1, totalSize=17, rawDataSize=16]

Partition default.dollos{y=2012, m=12, d=1} stats: [numFiles=1, numRows=2, totalSize=34, rawDataSize=32]

Partition default.dollos{y=2012, m=12, d=23} stats: [numFiles=1, numRows=1, totalSize=17, rawDataSize=16]

Partition default.dollos{y=2012, m=2, d=1} stats: [numFiles=1, numRows=1, totalSize=17, rawDataSize=16]

Partition default.dollos{y=2012, m=3, d=1} stats: [numFiles=1, numRows=1, totalSize=17, rawDataSize=16]

Partition default.dollos{y=2012, m=3, d=23} stats: [numFiles=1, numRows=1, totalSize=17, rawDataSize=16]

Partition default.dollos{y=2012, m=4, d=1} stats: [numFiles=1, numRows=2, totalSize=34, rawDataSize=32]

Partition default.dollos{y=2012, m=5, d=23} stats: [numFiles=1, numRows=1, totalSize=17, rawDataSize=16]

Partition default.dollos{y=2012, m=6, d=1} stats: [numFiles=1, numRows=2, totalSize=34, rawDataSize=32]

Partition default.dollos{y=2012, m=6, d=23} stats: [numFiles=1, numRows=1, totalSize=17, rawDataSize=16]

Partition default.dollos{y=2012, m=7, d=1} stats: [numFiles=1, numRows=1, totalSize=17, rawDataSize=16]

Partition default.dollos{y=2012, m=8, d=1} stats: [numFiles=1, numRows=1, totalSize=17, rawDataSize=16]

Partition default.dollos{y=2012, m=8, d=23} stats: [numFiles=1, numRows=1, totalSize=17, rawDataSize=16]

Partition default.dollos{y=2012, m=9, d=1} stats: [numFiles=1, numRows=1, totalSize=17, rawDataSize=16]

Partition default.dollos{y=2013, m=1, d=1} stats: [numFiles=1, numRows=2, totalSize=34, rawDataSize=32]

Partition default.dollos{y=2013, m=1, d=23} stats: [numFiles=1, numRows=1, totalSize=17, rawDataSize=16]

Partition default.dollos{y=2013, m=10, d=1} stats: [numFiles=1, numRows=1, totalSize=17, rawDataSize=16]

Partition default.dollos{y=2013, m=10, d=23} stats: [numFiles=1, numRows=1, totalSize=17, rawDataSize=16]

Partition default.dollos{y=2013, m=11, d=1} stats: [numFiles=1, numRows=2, totalSize=34, rawDataSize=32]

Partition default.dollos{y=2013, m=11, d=23} stats: [numFiles=1, numRows=1, totalSize=17, rawDataSize=16]

Partition default.dollos{y=2013, m=12, d=1} stats: [numFiles=1, numRows=2, totalSize=34, rawDataSize=32]

Partition default.dollos{y=2013, m=12, d=23} stats: [numFiles=1, numRows=1, totalSize=17, rawDataSize=16]

Partition default.dollos{y=2013, m=2, d=1} stats: [numFiles=1, numRows=1, totalSize=17, rawDataSize=16]

Partition default.dollos{y=2013, m=3, d=1} stats: [numFiles=1, numRows=1, totalSize=17, rawDataSize=16]

Partition default.dollos{y=2013, m=3, d=23} stats: [numFiles=1, numRows=1, totalSize=18, rawDataSize=17]

Partition default.dollos{y=2013, m=4, d=1} stats: [numFiles=1, numRows=2, totalSize=34, rawDataSize=32]

Partition default.dollos{y=2013, m=5, d=23} stats: [numFiles=1, numRows=1, totalSize=17, rawDataSize=16]

Partition default.dollos{y=2013, m=6, d=1} stats: [numFiles=1, numRows=2, totalSize=34, rawDataSize=32]

Partition default.dollos{y=2013, m=6, d=23} stats: [numFiles=1, numRows=1, totalSize=17, rawDataSize=16]

Partition default.dollos{y=2013, m=7, d=1} stats: [numFiles=1, numRows=1, totalSize=17, rawDataSize=16]

Partition default.dollos{y=2013, m=8, d=1} stats: [numFiles=1, numRows=1, totalSize=17, rawDataSize=16]

Partition default.dollos{y=2013, m=8, d=23} stats: [numFiles=1, numRows=1, totalSize=17, rawDataSize=16]

Partition default.dollos{y=2013, m=9, d=1} stats: [numFiles=1, numRows=1, totalSize=17, rawDataSize=16]

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Cumulative CPU: 10.9 sec HDFS Read: 5101 HDFS Write: 4040 SUCCESS

Total MapReduce CPU Time Spent: 10 seconds 900 msec

OK

hive> select \* from dollos;

OK

2011-01-01 45000 2011 1 1

2011-01-01 46000 2011 1 1

2011-01-23 70000 2011 1 23

2011-10-01 46000 2011 10 1

2011-10-23 70000 2011 10 23

2011-11-01 45000 2011 11 1

2011-11-01 46000 2011 11 1

2011-11-23 70000 2011 11 23

2011-12-01 45000 2011 12 1

2011-12-01 46000 2011 12 1

2011-12-23 70000 2011 12 23

2011-02-01 45000 2011 2 1

2011-03-01 46000 2011 3 1

2011-03-23 90000 2011 3 23

2011-04-01 45000 2011 4 1

2011-04-01 46000 2011 4 1

2011-05-23 70000 2011 5 23

2011-06-01 45000 2011 6 1

2011-06-01 46000 2011 6 1

2011-06-23 70000 2011 6 23

2011-07-01 45000 2011 7 1

2011-08-01 46000 2011 8 1

2011-08-23 70000 2011 8 23

2011-09-01 45000 2011 9 1

2012-01-01 47000 2012 1 1

2012-01-01 48000 2012 1 1

2012-01-23 72000 2012 1 23

2012-10-01 48000 2012 10 1

2012-10-23 72000 2012 10 23

2012-11-01 47000 2012 11 1

2012-11-01 48000 2012 11 1

2012-11-23 72000 2012 11 23

2012-12-01 47000 2012 12 1

2012-12-01 48000 2012 12 1

2012-12-23 72000 2012 12 23

2012-02-01 47000 2012 2 1

2012-03-01 48000 2012 3 1

2012-03-23 92000 2012 3 23

2012-04-01 47000 2012 4 1

2012-04-01 48000 2012 4 1

2012-05-23 72000 2012 5 23

2012-06-01 47000 2012 6 1

2012-06-01 48000 2012 6 1

2012-06-23 72000 2012 6 23

2012-07-01 47000 2012 7 1

2012-08-01 48000 2012 8 1

2012-08-23 72000 2012 8 23

2012-09-01 47000 2012 9 1

2013-01-01 55000 2013 1 1

2013-01-01 56000 2013 1 1

2013-01-23 80000 2013 1 23

2013-10-01 56000 2013 10 1

2013-10-23 80000 2013 10 23

2013-11-01 55000 2013 11 1

2013-11-01 56000 2013 11 1

2013-11-23 80000 2013 11 23

2013-12-01 55000 2013 12 1

2013-12-01 56000 2013 12 1

2013-12-23 80000 2013 12 23

2013-02-01 55000 2013 2 1

2013-03-01 56000 2013 3 1

2013-03-23 100000 2013 3 23

2013-04-01 55000 2013 4 1

2013-04-01 56000 2013 4 1

2013-05-23 80000 2013 5 23

2013-06-01 55000 2013 6 1

2013-06-01 56000 2013 6 1

2013-06-23 80000 2013 6 23

2013-07-01 55000 2013 7 1

2013-08-01 56000 2013 8 1

2013-08-23 80000 2013 8 23

2013-09-01 55000 2013 9 1

Time taken: 0.655 seconds, Fetched: 72 row(s)

==================================HDFS IN SIDE OUTPUT=================

[cloudera@quickstart ~]$ hadoop fs -ls /user/hive/warehouse/dollos/

Found 3 items

drwxrwxrwx - cloudera supergroup 0 2018-02-13 23:18 /user/hive/warehouse/dollos/y=2011

drwxrwxrwx - cloudera supergroup 0 2018-02-13 23:18 /user/hive/warehouse/dollos/y=2012

drwxrwxrwx - cloudera supergroup 0 2018-02-13 23:18 /user/hive/warehouse/dollos/y=2013

[cloudera@quickstart ~]$ hadoop fs -ls /user/hive/warehouse/dollos/y=2011

^[[2Found 12 items

drwxrwxrwx - cloudera supergroup 0 2018-02-13 23:18 /user/hive/warehouse/dollos/y=2011/m=1

drwxrwxrwx - cloudera supergroup 0 2018-02-13 23:18 /user/hive/warehouse/dollos/y=2011/m=10

drwxrwxrwx - cloudera supergroup 0 2018-02-13 23:18 /user/hive/warehouse/dollos/y=2011/m=11

drwxrwxrwx - cloudera supergroup 0 2018-02-13 23:18 /user/hive/warehouse/dollos/y=2011/m=12

drwxrwxrwx - cloudera supergroup 0 2018-02-13 23:18 /user/hive/warehouse/dollos/y=2011/m=2

drwxrwxrwx - cloudera supergroup 0 2018-02-13 23:18 /user/hive/warehouse/dollos/y=2011/m=3

drwxrwxrwx - cloudera supergroup 0 2018-02-13 23:18 /user/hive/warehouse/dollos/y=2011/m=4

drwxrwxrwx - cloudera supergroup 0 2018-02-13 23:18 /user/hive/warehouse/dollos/y=2011/m=5

drwxrwxrwx - cloudera supergroup 0 2018-02-13 23:18 /user/hive/warehouse/dollos/y=2011/m=6

drwxrwxrwx - cloudera supergroup 0 2018-02-13 23:18 /user/hive/warehouse/dollos/y=2011/m=7

drwxrwxrwx - cloudera supergroup 0 2018-02-13 23:18 /user/hive/warehouse/dollos/y=2011/m=8

drwxrwxrwx - cloudera supergroup 0 2018-02-13 23:18 /user/hive/warehouse/dollos/y=2011/m=9

[cloudera@quickstart ~]$ hadoop fs -ls /user/hive/warehouse/dollos/y=2011/m=1

Found 2 items

drwxrwxrwx - cloudera supergroup 0 2018-02-13 23:18 /user/hive/warehouse/dollos/y=2011/m=1/d=1

drwxrwxrwx - cloudera supergroup 0 2018-02-13 23:18 /user/hive/warehouse/dollos/y=2011/m=1/d=23

[cloudera@quickstart ~]$ hadoop fs -ls /user/hive/warehouse/dollos/y=2011/m=1/d=1

Found 1 items

-rwxrwxrwx 1 cloudera supergroup 34 2018-02-13 23:18 /user/hive/warehouse/dollos/y=2011/m=1/d=1/000000\_0

[cloudera@quickstart ~]$ hadoop fs -cat /user/hive/warehouse/dollos/y=2011/m=1/d=1/000000\_0

2011-01-01,45000

2011-01-01,46000

hive> select \* from dollos where dt ='2017-08-07';

hive> select \* from dollos where m=1;

hive> select \* from dollos where y=2013 and m=1;

hive>select \* from dollos where y=2013 and m=1 and d=3;

hive> select \* from dollos where y=2013 and m=1 and d>15;

hive >select \* from dollos where (y=2003 and m>=6)

or y>2003 and y<2009) or (y=2009 and m<8);

==========================BUCKETING IN HIVE=================================

HIVE TABLE TYEPS:

INNER AND EXTERNAL.

PARTITIONED AND NON-PARTITIONED.

INNER NON-PARTITIONDE.

INNER PARTITIONED.

EXTERNAL NON-PARTITIONED.

EXTERNAL PARTITIONED.

--------------------------------------------------------------------------------------------

BUCKETING TABLES:

Which divides the table data into multiple data files [buckeck].

Each buckect is a data file.

Single bucket can have multiple keys.

but A key is available with only one bucket.

-------------------------------------------------

set hive.enforce.bucketing=true;

Adv:

Sampling technique.

-------------------------------------------------------------------------------------------------------------

hive> create table charitra(id int,name string,sal int,sex string,dno int)

> row format delimited fields terminated by',';

OK

hive> load data local inpath 'lolo' into table charitra;

Loading data to table default.charitra

Table default.charitra stats: [numFiles=1, totalSize=96]

OK

Time taken: 3.201 seconds

hive> select \*from charitra;

OK

101 aaa 40000 m 11

102 bbb 50000 f 12

103 ccc 90000 m 12

104 ddd 100000 f 11

105 eee 20000 m 13

Time taken: 0.382 seconds, Fetched: 5 row(s)

hive> create table cortion(sal int,avg int);

OK

Time taken: 0.169 seconds

hive> insert into table cortion

> select sal,avg from

> charitra l join (select avg(sal) as avg from charitra) r;

hive> select \* from cortion;

OK

40000 60000

50000 60000

90000 60000

100000 60000

20000 60000

Time taken: 0.281 seconds, Fetched: 5 row(s)

hive> alter table cortion add columns (start string);

OK

Time taken: 0.575 seconds

hive> select \* from cortion;

OK

40000 60000 NULL

50000 60000 NULL

90000 60000 NULL

100000 60000 NULL

20000 60000 NULL

Time taken: 0.241 seconds, Fetched: 5 row(s)

hive> insert overwrite table cortion select sal,avg, if(sal>=avg,'Above','Below') from cortion;

Time taken: 64.672 seconds

hive> select \* from cortion;

OK

40000 60000 Below

50000 60000 Below

90000 60000 Above

100000 60000 Above

20000 60000 Below

Time taken: 0.304 seconds, Fetched: 5 row(s)

hive> select stat,count(\*) from coriton group by cortion;

hive> select start,count(\*) from cortion group by start;

Total MapReduce CPU Time Spent: 9 seconds 20 msec

OK

Above 2

Below 3

Time taken: 82.758 seconds, Fetched: 2 row(s)

---------------------ADDING FROM EXITING TABLE ONE NEW COLUMN VALUES BASED ON SALARY-------------

hive> select \* from subbu\_part;

OK

1000 subbu m 100000 11

1100 raju f 2000000 12

1200 shashi f 300000 13

1300 gopal m 400000 14

1500 suri f 500000 11

1600 mani m 60000 15

1700 yadav m 70000 13

1800 chandu f 80000 16

1900 ranii f 90000 17

2000 suji f 2000000 18

2001 lakhxmi f 210000 19

1000 dhilipa f 22000 20

1100 teju m 230000 21

1200 geetha f 240000 22

Time taken: 2.355 seconds, Fetched: 14 row(s)

hive> alter table subbu\_part add columns(tax int):

> ;

FAILED: ParseException line 1:43 extraneous input ':' expecting EOF near '<EOF>'

hive> alter table subbu\_part add columns(tax int);

OK

Time taken: 13.989 seconds

hive> select \*from subbu\_part;

OK

1000 subbu m 100000 11 NULL

1100 raju f 2000000 12 NULL

1200 shashi f 300000 13 NULL

1300 gopal m 400000 14 NULL

1500 suri f 500000 11 NULL

1600 mani m 60000 15 NULL

1700 yadav m 70000 13 NULL

1800 chandu f 80000 16 NULL

1900 ranii f 90000 17 NULL

2000 suji f 2000000 18 NULL

2001 lakhxmi f 210000 19 NULL

1000 dhilipa f 22000 20 NULL

1100 teju m 230000 21 NULL

1200 geetha f 240000 22 NULL

Time taken: 0.26 seconds, Fetched: 14 row(s)

hive> insert overwrite table subbu\_part select id,name,sex,sla,dno,sal\*0.1 form subuu\_part;

FAILED: NullPointerException null

hive> insert overwrite table subbu\_part select id,name,sex,sla,dno,sal\*0.1 from subuu\_part;

FAILED: SemanticException [Error 10001]: Line 1:74 Table not found 'subuu\_part'

hive> describe subbu\_part;

OK

id int

name string

sex string

sal int

deptno int

tax int

Time taken: 0.42 seconds, Fetched: 6 row(s)

hive> insert overwrite table subbu\_part select id,name,sex,sal,deptno,sal\*0.1 from subuu\_part;

FAILED: SemanticException [Error 10001]: Line 1:77 Table not found 'subuu\_part'

hive> insert overwrite table subbu\_part select id,name,sex,sal,deptno,sal\*0.1 from subbu\_part;

Query ID = cloudera\_20180207153838\_2c6ca88f-bc9c-4268-b479-231aceeb9908

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\_1518028021574\_0001, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1518028021574\_0001/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1518028021574\_0001

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0

2018-02-07 15:39:39,214 Stage-1 map = 0%, reduce = 0%

2018-02-07 15:40:47,474 Stage-1 map = 0%, reduce = 0%

2018-02-07 15:41:55,913 Stage-1 map = 0%, reduce = 0%

2018-02-07 15:42:56,661 Stage-1 map = 0%, reduce = 0%

2018-02-07 15:43:58,907 Stage-1 map = 0%, reduce = 0%

2018-02-07 15:44:18,329 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 9.41 sec

MapReduce Total cumulative CPU time: 9 seconds 410 msec

Ended Job = job\_1518028021574\_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/subbu\_part/.hive-staging\_hive\_2018-02-07\_15-38-17\_297\_1329953132384139765-1/-ext-10000

Loading data to table default.subbu\_part

Table default.subbu\_part stats: [numFiles=1, numRows=14, totalSize=402, rawDataSize=388]

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Cumulative CPU: 10.47 sec HDFS Read: 4831 HDFS Write: 477 SUCCESS

Total MapReduce CPU Time Spent: 10 seconds 470 msec

OK

Time taken: 373.788 seconds

hive> select \*from subbu\_part;

OK

1000 subbu m 100000 11 10000

1100 raju f 2000000 12 200000

1200 shashi f 300000 13 30000

1300 gopal m 400000 14 40000

1500 suri f 500000 11 50000

1600 mani m 60000 15 6000

1700 yadav m 70000 13 7000

1800 chandu f 80000 16 8000

1900 ranii f 90000 17 9000

2000 suji f 2000000 18 200000

2001 lakhxmi f 210000 19 21000

1000 dhilipa f 22000 20 2200

1100 teju m 230000 21 23000

1200 geetha f 240000 22 24000

Time taken: 0.532 seconds, Fetched: 14 row(s)

hive> insert overwrite table subbu\_part select id,name,sex,sal,deptno,sal\*0.1,net \*0.2 as net from subuu\_part;

FAILED: SemanticException [Error 10001]: Line 1:93 Table not found 'subuu\_part'

hive> insert overwrite table subbu\_part select id,name,sex,sal,deptno,sal\*0.1,net \*0.2 as net from subbu\_part;

FAILED: SemanticException [Error 10004]: Line 1:72 Invalid table alias or column reference 'net': (possible column names are: id, name, sex, sal, deptno, tax)

hive> alter table subbu\_part add columns (net int);

OK

Time taken: 0.502 seconds

hive> select \* From subbu\_part;

OK

1000 subbu m 100000 11 10000 NULL

1100 raju f 2000000 12 200000 NULL

1200 shashi f 300000 13 30000 NULL

1300 gopal m 400000 14 40000 NULL

1500 suri f 500000 11 50000 NULL

1600 mani m 60000 15 6000 NULL

1700 yadav m 70000 13 7000 NULL

1800 chandu f 80000 16 8000 NULL

1900 ranii f 90000 17 9000 NULL

2000 suji f 2000000 18 200000 NULL

2001 lakhxmi f 210000 19 21000 NULL

1000 dhilipa f 22000 20 2200 NULL

1100 teju m 230000 21 23000 NULL

1200 geetha f 240000 22 24000 NULL

Time taken: 0.25 seconds, Fetched: 14 row(s)

hive> insert overwrite table subbu\_part select id,name,sex,sal,deptno,sal\*0.1,net \*0.2 as net from subbu\_part;

Query ID = cloudera\_20180207155555\_a8b4250d-8a87-42e8-89d5-66b27790d79e

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\_1518028021574\_0002, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1518028021574\_0002/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1518028021574\_0002

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0

2018-02-07 15:56:16,211 Stage-1 map = 0%, reduce = 0%

2018-02-07 15:56:43,596 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 7.04 sec

MapReduce Total cumulative CPU time: 7 seconds 40 msec

Ended Job = job\_1518028021574\_0002

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/subbu\_part/.hive-staging\_hive\_2018-02-07\_15-55-43\_226\_6558404378865262189-1/-ext-10000

Loading data to table default.subbu\_part

Table default.subbu\_part stats: [numFiles=1, numRows=14, totalSize=444, rawDataSize=430]

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Cumulative CPU: 8.05 sec HDFS Read: 5161 HDFS Write: 519 SUCCESS

Total MapReduce CPU Time Spent: 8 seconds 50 msec

OK

Time taken: 69.077 seconds

hive> select \* from subbu\_part;

OK

1000 subbu m 100000 11 10000 NULL

1100 raju f 2000000 12 200000 NULL

1200 shashi f 300000 13 30000 NULL

1300 gopal m 400000 14 40000 NULL

1500 suri f 500000 11 50000 NULL

1600 mani m 60000 15 6000 NULL

1700 yadav m 70000 13 7000 NULL

1800 chandu f 80000 16 8000 NULL

1900 ranii f 90000 17 9000 NULL

2000 suji f 2000000 18 200000 NULL

2001 lakhxmi f 210000 19 21000 NULL

1000 dhilipa f 22000 20 2200 NULL

1100 teju m 230000 21 23000 NULL

1200 geetha f 240000 22 24000 NULL

Time taken: 0.238 seconds, Fetched: 14 row(s)

hive> insert overwrite table subbu\_part select id,name,sex,sal,deptno,sal\*0.1,net \*0.2 from subbu\_part;

Query ID = cloudera\_20180207155959\_093dd339-25ae-4636-9efb-d0e1ba6e249c

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\_1518028021574\_0003, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1518028021574\_0003/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1518028021574\_0003

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0

2018-02-07 16:00:16,324 Stage-1 map = 0%, reduce = 0%

2018-02-07 16:00:55,816 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 8.3 sec

MapReduce Total cumulative CPU time: 8 seconds 300 msec

Ended Job = job\_1518028021574\_0003

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/subbu\_part/.hive-staging\_hive\_2018-02-07\_15-59-42\_575\_7332482266323255364-1/-ext-10000

Loading data to table default.subbu\_part

Table default.subbu\_part stats: [numFiles=1, numRows=14, totalSize=444, rawDataSize=430]

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Cumulative CPU: 8.3 sec HDFS Read: 5208 HDFS Write: 519 SUCCESS

Total MapReduce CPU Time Spent: 8 seconds 300 msec

OK

Time taken: 80.158 seconds

hive> select \* from subbu\_part;

OK

1000 subbu m 100000 11 10000 NULL

1100 raju f 2000000 12 200000 NULL

1200 shashi f 300000 13 30000 NULL

1300 gopal m 400000 14 40000 NULL

1500 suri f 500000 11 50000 NULL

1600 mani m 60000 15 6000 NULL

1700 yadav m 70000 13 7000 NULL

1800 chandu f 80000 16 8000 NULL

1900 ranii f 90000 17 9000 NULL

2000 suji f 2000000 18 200000 NULL

2001 lakhxmi f 210000 19 21000 NULL

1000 dhilipa f 22000 20 2200 NULL

1100 teju m 230000 21 23000 NULL

1200 geetha f 240000 22 24000 NULL

Time taken: 0.271 seconds, Fetched: 14 row(s)

hive> insert overwrite table subbu\_part select id,name,sex,sal,deptno,sal\*0.1,sal\*0.2 from subbu\_part;

Query ID = cloudera\_20180207160101\_2d465ec9-97dd-463c-b4de-a91713c8e27b

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\_1518028021574\_0004, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1518028021574\_0004/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1518028021574\_0004

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0

2018-02-07 16:02:07,309 Stage-1 map = 0%, reduce = 0%

2018-02-07 16:02:47,481 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 8.47 sec

MapReduce Total cumulative CPU time: 8 seconds 470 msec

Ended Job = job\_1518028021574\_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: hdfs://quickstart.cloudera:8020/user/hive/warehouse/subbu\_part/.hive-staging\_hive\_2018-02-07\_16-01-33\_665\_5419331357493987940-1/-ext-10000

Loading data to table default.subbu\_part

Table default.subbu\_part stats: [numFiles=1, numRows=14, totalSize=488, rawDataSize=474]

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Cumulative CPU: 8.47 sec HDFS Read: 5217 HDFS Write: 563 SUCCESS

Total MapReduce CPU Time Spent: 8 seconds 470 msec

OK

Time taken: 79.124 seconds

hive> select \*From subbu\_part;

OK

1000 subbu m 100000 11 10000 20000

1100 raju f 2000000 12 200000 400000

1200 shashi f 300000 13 30000 60000

1300 gopal m 400000 14 40000 80000

1500 suri f 500000 11 50000 100000

1600 mani m 60000 15 6000 12000

1700 yadav m 70000 13 7000 14000

1800 chandu f 80000 16 8000 16000

1900 ranii f 90000 17 9000 18000

2000 suji f 2000000 18 200000 400000

2001 lakhxmi f 210000 19 21000 42000

1000 dhilipa f 22000 20 2200 4400

1100 teju m 230000 21 23000 46000

1200 geetha f 240000 22 24000 48000

Time taken: 0.246 seconds, Fetched: 14 row(s)

hive> describe subbu\_part;

OK

id int

name string

sex string

sal int

deptno int

tax int

net int

Time taken: 0.339 seconds, Fetched: 7 row(s)

-------------------------SUME DOINGTRANSFORMATION-----------------------------

hive> create table transformation(id int,name string,sal int,sex string,dno int) row format delimited fields terminated by ',';

OK

Time taken: 1.287 seconds

hive> load data local inpath 'student\_local' into table transformation;

Loading data to table default.transformation

Table default.transformation stats: [numFiles=1, totalSize=212]

OK

Time taken: 1.348 seconds

hive> select \* from transformation;

OK

101 sunil 10000 m 11

102 menaka 20000 f 12

103 subbu 30000 m 13

104 suji 40000 f 14

105 gopal 50000 m 15

106 laxmi 60000 f 16

107 shashi 70000 m 17

108 dhilipa 80000 f 18

109 suri 90000 m 19

120 teja 100000 f 20

Time taken: 0.229 seconds, Fetched: 10 row(s)

hive> create table etab(id int,name string,sal int,sex string,dname string,grade string);

OK

Time taken: 0.207 seconds

hive> insert overwrite table etab select id,name,sal,if(sex='f','female','male'),if(dno=11,'marketing',if(dno=12,'hr',if(dno=13,'finance',if(dno=14,'telecom','others')))),if(sal>=70000,'A',if(sal>=50000,'B',if(sal>=30000,'C','D'))) from newtransformation;

FAILED: SemanticException [Error 10001]: Line 1:231 Table not found 'newtransformation'

hive> insert overwrite table etab select id,name,sal,if(sex='f','female','male'),if(dno=11,'marketing',if(dno=12,'hr',if(dno=13,'finance',if(dno=14,'telecom','others')))),if(sal>=70000,'A',if(sal>=50000,'B',if(sal>=30000,'C','D'))) from transformation;

Query ID = cloudera\_20180207162424\_3b5285fe-4074-45c2-a241-11aa29ab7781

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\_1518028021574\_0005, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1518028021574\_0005/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1518028021574\_0005

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0

2018-02-07 16:24:33,953 Stage-1 map = 0%, reduce = 0%

2018-02-07 16:25:16,519 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 10.66 sec

MapReduce Total cumulative CPU time: 10 seconds 660 msec

Ended Job = job\_1518028021574\_0005

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/etab/.hive-staging\_hive\_2018-02-07\_16-24-00\_931\_1629379045633023644-1/-ext-10000

Loading data to table default.etab

Table default.etab stats: [numFiles=1, numRows=10, totalSize=313, rawDataSize=303]

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Cumulative CPU: 10.66 sec HDFS Read: 4731 HDFS Write: 382 SUCCESS

Total MapReduce CPU Time Spent: 10 seconds 660 msec

OK

Time taken: 80.995 seconds

hive> select \* from etab;

OK

101 sunil 10000 male marketing D

102 menaka 20000 female hr D

103 subbu 30000 male finance C

104 suji 40000 female telecom C

105 gopal 50000 male others B

106 laxmi 60000 female others B

107 shashi 70000 male others A

108 dhilipa 80000 female others A

109 suri 90000 male others A

120 teja 100000 female others A

Time taken: 0.126 seconds, Fetched: 10 row(s)

-------------------------OTHER MODEL DONING SAME ABOVE JOB---------------------------------------

hive> create table etab1(id int,name string,sal int,sex string,dname string,grade string);

OK

Time taken: 0.172 seconds

hive> insert overwrite table etab1 select id,name,sal,if(sex='m','male','female'),if(dno=11,'chittoor',if(dno=12,'punganur',if(dno=13,'anantapur',if(dno=14,'palamaner',if(dno=15,'nellore',if(dno=16,'gunture',if(dno=17,'eastgodaavari',if(dno=18,'westgodavari',if(dno=19,'vijayawada',if(dno=20,'vizag')))))))))),if(sal>=80000,'A',if(sal>=60000,'B',if(sal>=40000,'C',if(sal>=20000,'D','E')))) from transformation;

FAILED: SemanticException [Error 10015]: Line 1:276 Arguments length mismatch ''vizag'': The function IF(expr1,expr2,expr3) accepts exactly 3 arguments.

hive> insert overwrite table etab1 select id,name,sal,if(sex='m','male','female'),if(dno=11,'chittoor',if(dno=12,'punganur',if(dno=13,'anantapur',if(dno=14,'palamaner',if(dno=15,'nellore',if(dno=16,'gunture',if(dno=17,'eastgodaavari',if(dno=18,'westgodavari',if(dno=19,'vijayawada','others'))))))))),if(sal>=80000,'A',if(sal>=60000,'B',if(sal>=40000,'C',if(sal>=20000,'D','E')))) from transformation;

Query ID = cloudera\_20180207164242\_feddee9d-da05-42cf-8ec3-1b4ccd61b17b

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\_1518028021574\_0006, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1518028021574\_0006/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1518028021574\_0006

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0

2018-02-07 16:43:23,164 Stage-1 map = 0%, reduce = 0%

2018-02-07 16:44:11,053 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 8.43 sec

MapReduce Total cumulative CPU time: 8 seconds 430 msec

Ended Job = job\_1518028021574\_0006

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/etab1/.hive-staging\_hive\_2018-02-07\_16-42-47\_538\_2613196538563308554-1/-ext-10000

Loading data to table default.etab1

Table default.etab1 stats: [numFiles=1, numRows=10, totalSize=341, rawDataSize=331]

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Cumulative CPU: 8.74 sec HDFS Read: 5198 HDFS Write: 411 SUCCESS

Total MapReduce CPU Time Spent: 8 seconds 740 msec

OK

Time taken: 89.863 seconds

hive> select \* from etab1;

OK

101 sunil 10000 male chittoor E

102 menaka 20000 female punganur D

103 subbu 30000 male anantapur D

104 suji 40000 female palamaner C

105 gopal 50000 male nellore C

106 laxmi 60000 female gunture B

107 shashi 70000 male eastgodaavari B

108 dhilipa 80000 female westgodavari A

109 suri 90000 male vijayawada A

120 teja 100000 female others A

Time taken: 0.263 seconds, Fetched: 10 row(s)

---------------------------------------------HOW TO SOLVE THIS IS ------------------------------------

hive> create table ram\_yadav(id int,city string,name string) row format delimited fields terminated by',';

OK

Time taken: 0.287 seconds

hive> load data local inpath'trns' into table ram\_yadav;

Loading data to table default.ram\_yadav

Table default.ram\_yadav stats: [numFiles=1, totalSize=40]

OK

Time taken: 1.401 seconds

hive> select \* from ram\_yadav;

OK

1 s aaa

2 c bbb

3 m ccc

4 t ddd

5 a eee

Time taken: 0.207 seconds, Fetched: 5 row(s)

hive> create table ubbus(id int,name string,city\_name string);

OK

Time taken: 0.251 seconds

hive> insert overwrite table ubbus select id,name,if(city=s,'singapur',if(city=c,'chittoor',if(city=m,'amaravati',if(city=t,'tirupati','others'))))from ubbus;

FAILED: SemanticException [Error 10004]: Line 1:47 Invalid table alias or column reference 'city': (possible column names are: id, name, city\_name)

hive> insert overwrite table ubbus select id,name,if(city\_name=s,'singapur',if(city\_name=c,'chittoor',if(city\_name=m,'amaravati',if(city\_name=t,'tirupati','others'))))from ubbus;

FAILED: SemanticException [Error 10004]: Line 1:57 Invalid table alias or column reference 's': (possible column names are: id, name, city\_name)

hive> create table ubbus(id int,name string,dname string);

FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.DDLTask. AlreadyExistsException(message:Table ubbus already exists)

hive> drop table ubbus;

OK

Time taken: 1.975 seconds

hive> create table ubbus(id int,name string,dname string);

OK

Time taken: 0.336 seconds

hive> insert overwrite table ubbus select id,name,if(city=s,'singapur',if(city=c,'chittoor',if(city=m,'amaravati',if(city=t,'tirupati','others'))))from ubbus;

FAILED: SemanticException [Error 10004]: Line 1:47 Invalid table alias or column reference 'city': (possible column names are: id, name, dname)

hive> drop table ubbus;

OK

Time taken: 0.407 seconds

hive> create table ubbus(id int,dname string,name string);

OK

Time taken: 0.227 seconds

hive> insert overwrite table ubbus select id,name,if(city=s,'singapur',if(city=c,'chittoor',if(city=m,'amaravati',if(city=t,'tirupati','others'))))from ubbus;

FAILED: SemanticException [Error 10004]: Line 1:47 Invalid table alias or column reference 'city': (possible column names are: id, dname, name)

hive> drop table ubbus;

OK

Time taken: 0.395 seconds

hive> create table ubbus(id int,cityname string,name string);

OK

Time taken: 0.201 seconds

hive> insert overwrite table ubbus select id,name,if(city=s,'singapur',if(city=c,'chittoor',if(city=m,'amaravati',if(city=t,'tirupati','others'))))from ubbus;

FAILED: SemanticException [Error 10004]: Line 1:47 Invalid table alias or column reference 'city': (possible column names are: id, cityname, name)

hive> drop table ubbus;

OK

Time taken: 0.3 seconds

hive> create table ubbus(id int,city string,name string);

OK

Time taken: 0.223 seconds

hive> insert overwrite table ubbus select id,name,if(city=s,'singapur',if(city=c,'chittoor',if(city=m,'amaravati',if(city=t,'tirupati','others'))))from ubbus;

FAILED: SemanticException [Error 10004]: Line 1:52 Invalid table alias or column reference 's': (possible column names are: id, city, name)

hive> drop table ubbus;

OK

Time taken: 0.345 seconds

hive> create table ubbus(id int,cname string,name string);

OK

Time taken: 0.194 seconds

hive> insert overwrite table ubbus select id,name,if(city=s,'singapur',if(city=c,'chittoor',if(city=m,'amaravati',if(city=t,'tirupati','others'))))from ubbus;

FAILED: SemanticException [Error 10004]: Line 1:47 Invalid table alias or column reference 'city': (possible column names are: id, cname, name)

hive> drop table ubbus;

--------------------------------THIS IS ADD WITH FULL NAME ------------------------------

hive> create table subbu\_spark(id int,name string,c string) row format delimited fields terminated by ',';

OK

Time taken: 0.817 seconds

hive> load data local inpath 'ddd' into table subbu\_spark;

Loading data to table default.subbu\_spark

Table default.subbu\_spark stats: [numFiles=1, totalSize=30]

OK

Time taken: 2.496 seconds

hive> select \*from subbu\_spark;

OK

1 subbu c

2 chandu a

3 rani w

Time taken: 1.76 seconds, Fetched: 3 row(s)

hive> create table scala\_data(id int,name string, c string);

OK

Time taken: 0.222 seconds

hive> insert overwrite table scala\_data select id,name,if(c='c',"chitttoor"),if(c='a',"amaravathi"),if(c='w',"walk");

FAILED: NullPointerException null

hive> insert overwrite table scala\_data select id,name,if(c='c',"chitttoor",if(c='a',"amaravathi",if(c='w',"walk")));

FAILED: NullPointerException null

hive> insert overwrite table scala\_data select id,name,if(c='c','chitttoor',if(c='a','amaravathi',if(c='w','walk')));

FAILED: NullPointerException null

hive> insert overwrite table scala\_data select id,name,if(c='c','chittoor','amaravathi','walk') from scala\_data;

FAILED: SemanticException [Error 10015]: Line 1:49 Arguments length mismatch ''walk'': The function IF(expr1,expr2,expr3) accepts exactly 3 arguments.

hive> insert overwrite table scala\_data select id,name,if(c='c','chittoor','amaravathi') from scala\_data;

Query ID = cloudera\_20180212113434\_9b560200-7c7a-42ec-b793-758127d6a907

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\_1518461294181\_0001, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1518461294181\_0001/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1518461294181\_0001

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0

2018-02-12 11:35:39,529 Stage-1 map = 0%, reduce = 0%

2018-02-12 11:36:15,192 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 9.45 sec

MapReduce Total cumulative CPU time: 9 seconds 450 msec

Ended Job = job\_1518461294181\_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/scala\_data/.hive-staging\_hive\_2018-02-12\_11-34-57\_486\_2682200579171866055-1/-ext-10000

Loading data to table default.scala\_data

Table default.scala\_data stats: [numFiles=1, numRows=0, totalSize=0, rawDataSize=0]

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Cumulative CPU: 9.45 sec HDFS Read: 3569 HDFS Write: 44 SUCCESS

Total MapReduce CPU Time Spent: 9 seconds 450 msec

OK

Time taken: 84.98 seconds

hive> select \* from scala\_data;

OK

Time taken: 0.33 seconds

hive> insert overwrite table scala\_data select id,name,if(c='c','chitttoor',if(c='a','amaravathi',if(c='w','walk'))) from scala\_data;

FAILED: SemanticException [Error 10015]: Line 1:92 Arguments length mismatch ''walk'': The function IF(expr1,expr2,expr3) accepts exactly 3 arguments.

hive> insert overwrite table scala\_data select id,name,if(c='c','chitttoor',if(c='a','amaravathi','other')) from scala\_data;

Query ID = cloudera\_20180212113838\_3a020666-0a06-4a06-9b3b-c45f046fe54e

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\_1518461294181\_0002, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1518461294181\_0002/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1518461294181\_0002

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0

2018-02-12 11:39:33,806 Stage-1 map = 0%, reduce = 0%

2018-02-12 11:39:54,396 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 5.61 sec

MapReduce Total cumulative CPU time: 5 seconds 610 msec

Ended Job = job\_1518461294181\_0002

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/scala\_data/.hive-staging\_hive\_2018-02-12\_11-38-55\_642\_8063320174723396155-1/-ext-10000

Loading data to table default.scala\_data

Table default.scala\_data stats: [numFiles=1, numRows=0, totalSize=0, rawDataSize=0]

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Cumulative CPU: 5.61 sec HDFS Read: 3661 HDFS Write: 44 SUCCESS

Total MapReduce CPU Time Spent: 5 seconds 610 msec

OK

Time taken: 62.404 seconds

hive> select \* from scala\_data;

OK

Time taken: 0.239 seconds

hive> dfs -cat hdfs://quickstart.cloudera:8020/user/hive/warehouse/scala\_data/.hive-staging\_hive\_2018-02-12\_11-38-55\_642\_8063320174723396155-1/-ext-10000;

cat: `hdfs://quickstart.cloudera:8020/user/hive/warehouse/scala\_data/.hive-staging\_hive\_2018-02-12\_11-38-55\_642\_8063320174723396155-1/-ext-10000': No such file or directory

Command failed with exit code = 1

Query returned non-zero code: 1, cause: null

hive> create table scala\_data1(id int,name string, c string,city string);

OK

Time taken: 0.329 seconds

hive> insert overwrite tabel scala\_data1 select id,name,c string,if(c='c','chittoor','c='a','amaravathi','c='w','walk') from scala\_data1;

NoViableAltException(26@[])

FAILED: ParseException line 1:17 cannot recognize input near 'tabel' 'scala\_data1' 'select' in destination specification

hive> insert overwrite table scala\_data1 select id,name,c string,if(c='c','chittoor','c='a','amaravathi','c='w','walk') from scala\_data1;

FAILED: NullPointerException null

hive> insert overwrite table scala\_data1 select id,name,if(c='c','chittoor','c='a','amaravathi','c='w','walk') from scala\_data1;

FAILED: NullPointerException null

hive> drop table scala\_data;

OK

Time taken: 0.585 seconds

hive> drop table scala\_data1;

OK

Time taken: 0.394 seconds

hive> create table ss\_dd(id int,name strign,sex string)row fromat delimited fields terminated by ',';

NoViableAltException(26@[])

FAILED: ParseException line 1:31 cannot recognize input near 'strign' ',' 'sex' in column type

hive> create table ss\_dd(id int,name string,sex string)row fromat delimited fields terminated by ',';

NoViableAltException(220@[])

FAILED: ParseException line 1:49 cannot recognize input near 'row' 'fromat' 'delimited' in table row format specification

hive> create table ss\_dd(id int,name string,sex string)row format delimited fields terminated by ',';

OK

Time taken: 0.224 seconds

hive> load data local inpath 'ddd' into table ss\_dd;

Loading data to table default.ss\_dd

Table default.ss\_dd stats: [numFiles=1, totalSize=42]

OK

Time taken: 1.028 seconds

hive> select \* from ss\_dd;

OK

1 subbu f

2 chandu m

3 rani f

4 remesh m

NULL NULL NULL

Time taken: 0.218 seconds, Fetched: 5 row(s)

hive> select id,name,sex from ss\_dd group by id,name,sex;

Query ID = cloudera\_20180212115959\_97ba0844-bfca-45b8-abb1-bf58af21f477

Total jobs = 1

Launching Job 1 out of 1

Number of reduce tasks not specified. Estimated from input data size: 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\_1518461294181\_0003, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1518461294181\_0003/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1518461294181\_0003

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

2018-02-12 11:59:53,061 Stage-1 map = 0%, reduce = 0%

2018-02-11 22:31:22,919 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 7.36 sec

2018-02-11 22:31:55,999 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 12.42 sec

2018-02-11 22:31:58,590 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 15.32 sec

2018-02-12 12:01:34,674 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 15.32 sec

MapReduce Total cumulative CPU time: 15 seconds 320 msec

Ended Job = job\_1518461294181\_0003

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 15.32 sec HDFS Read: 7019 HDFS Write: 50 SUCCESS

Total MapReduce CPU Time Spent: 15 seconds 320 msec

OK

NULL NULL NULL

1 subbu f

2 chandu m

3 rani f

4 remesh m

Time taken: 124.491 seconds, Fetched: 5 row(s)

hive> create table sss\_ddd(id int,name string,sex string) row format delimited fields terminated by ',' tblproperties("skip.header.line.count"="1");

OK

Time taken: 0.375 seconds

hive> insert into table sss\_ddd select id,name,sex from ss\_dd;

Query ID = cloudera\_20180212121515\_a5af0635-315d-46d6-a9d3-4438d56f1cfa

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\_1518461294181\_0004, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1518461294181\_0004/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1518461294181\_0004

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0

2018-02-12 12:16:20,633 Stage-1 map = 0%, reduce = 0%

2018-02-11 22:47:17,798 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 4.78 sec

MapReduce Total cumulative CPU time: 4 seconds 780 msec

Ended Job = job\_1518461294181\_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: hdfs://quickstart.cloudera:8020/user/hive/warehouse/sss\_ddd/.hive-staging\_hive\_2018-02-12\_12-15-49\_507\_6220069277206165799-1/-ext-10000

Loading data to table default.sss\_ddd

Table default.sss\_ddd stats: [numFiles=1, numRows=5, totalSize=50, rawDataSize=45]

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Cumulative CPU: 5.8 sec HDFS Read: 3244 HDFS Write: 121 SUCCESS

Total MapReduce CPU Time Spent: 5 seconds 800 msec

OK

Time taken: -48504.212 seconds

hive> select \* from sss\_ddd;

OK

2 chandu m

3 rani f

4 remesh m

NULL NULL NULL

Time taken: 0.23 seconds, Fetched: 4 row(s)

hive> create table sss\_ddd(id int,name string,sex string) row format delimited fields terminated by ',' tblproperties("skip.header.line.count"="0");

FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.DDLTask. AlreadyExistsException(message:Table sss\_ddd already exists)

hive> create table sss\_ddd1(id int,name string,sex string) row format delimited fields terminated by ',' tblproperties("skip.header.line.count"="0");

OK

Time taken: 0.394 seconds

hive> insert into table sss\_ddd1 select id,name,sex from ss\_dd;

Query ID = cloudera\_20180212121818\_719b2e74-9184-4a2c-9d01-cb6b14e069af

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\_1518461294181\_0005, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1518461294181\_0005/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1518461294181\_0005

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0

2018-02-12 12:19:01,413 Stage-1 map = 0%, reduce = 0%

2018-02-12 12:19:32,492 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 7.36 sec

MapReduce Total cumulative CPU time: 7 seconds 360 msec

Ended Job = job\_1518461294181\_0005

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/sss\_ddd1/.hive-staging\_hive\_2018-02-12\_12-18-27\_834\_6927367249156950747-1/-ext-10000

Loading data to table default.sss\_ddd1

Table default.sss\_ddd1 stats: [numFiles=1, numRows=5, totalSize=50, rawDataSize=45]

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Cumulative CPU: 7.7 sec HDFS Read: 3249 HDFS Write: 122 SUCCESS

Total MapReduce CPU Time Spent: 7 seconds 700 msec

OK

Time taken: 71.622 seconds

hive> select\* from sss\_ddd1;

OK

1 subbu f

2 chandu m

3 rani f

4 remesh m

NULL NULL NULL

Time taken: 0.223 seconds, Fetched: 5 row(s)

hive> load data local inpath 'qqq' overwrite into table sss\_ddd1;

Loading data to table default.sss\_ddd1

Table default.sss\_ddd1 stats: [numFiles=1, numRows=0, totalSize=41, rawDataSize=0]

OK

Time taken: 1.524 seconds

hive> select \* from sss\_ddd1;

OK

1 subbu f

2 chandu m

3 rani f

4 remesh m

Time taken: 0.261 seconds, Fetched: 4 row(s)

hive> crete table su\_che(id int,name string,sex string);

NoViableAltException(26@[])

FAILED: ParseException line 1:0 cannot recognize input near 'crete' 'table' 'su\_che'

hive> create table su\_che(id int,name string,sex string);

OK

Time taken: 0.307 seconds

hive> insert overwrit table su\_che select id,name,if(sex='m','male','female') from sss\_ddd1;

NoViableAltException(138@[])

FAILED: ParseException line 1:0 cannot recognize input near 'insert' 'overwrit' 'table' in insert clause

hive> insert overwrite table su\_che select id,name,if(sex='m','male','female') from sss\_ddd1;

Query ID = cloudera\_20180212124444\_d5abc29c-65c7-46f4-9a6a-27e46b6b6837

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\_1518461294181\_0006, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1518461294181\_0006/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1518461294181\_0006

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0

2018-02-12 12:45:11,951 Stage-1 map = 0%, reduce = 0%

2018-02-12 12:46:07,526 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 6.87 sec

MapReduce Total cumulative CPU time: 6 seconds 870 msec

Ended Job = job\_1518461294181\_0006

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/su\_che/.hive-staging\_hive\_2018-02-12\_12-44-36\_049\_7962811065227835468-1/-ext-10000

Loading data to table default.su\_che

Table default.su\_che stats: [numFiles=1, numRows=4, totalSize=57, rawDataSize=53]

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Cumulative CPU: 9.16 sec HDFS Read: 3642 HDFS Write: 127 SUCCESS

Total MapReduce CPU Time Spent: 9 seconds 160 msec

OK

Time taken: 97.788 seconds

hive> select \* from su\_che;

OK

1 subbu female

2 chandu male

3 rani female

4 remesh male

Time taken: 0.269 seconds, Fetched: 4 row(s)

-------------------------------------HEAR NAMES REMOVING AT A TIME LOADING TO HIVE--------------------------

[cloudera@quickstart ~]$ cat rrr

id,name,sex,deptno

1,subbu,m,10

12,gopa,m,11

13,lasya,f,13

14,chitra,f,14

hive> create table sayra(id int,name string,sex string,dno int) row format delimited fields terminated by ',' tblproperties ("skip.header.line.count"="1");

OK

Time taken: 0.35 seconds

hive> load data local inpath 'rrr' into table sayra;

Loading data to table default.sayra

Table default.sayra stats: [numFiles=1, totalSize=74]

OK

Time taken: 1.014 seconds

hive> select \* from sayra;

OK

1 subbu m 10

12 gopa m 11

13 lasya f 13

14 chitra f 14

Time taken: 0.259 seconds, Fetched: 4 row(s)