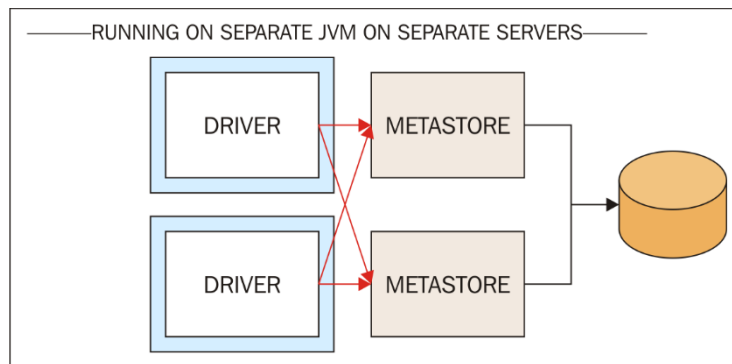
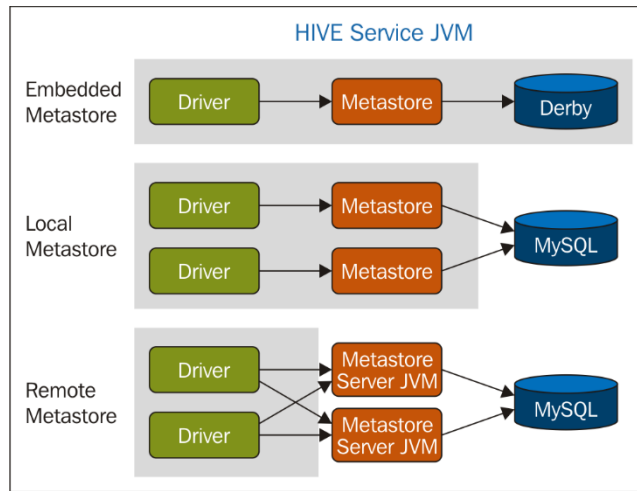


Chapter 1: Developing Hive



Chapter 2: Services in Hive

```
[root@localhost ~]# hive --service hiveserver2 &
[1] 3713
[root@localhost ~]#
```

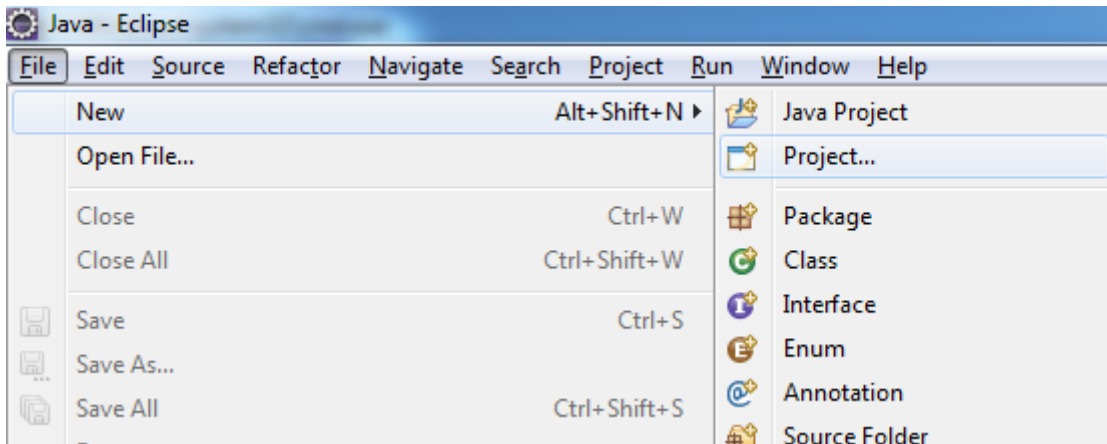
```
[root@localhost ~]#
[root@localhost ~]# jps
3066 NodeManager
2642 DataNode
3713 RunJar
2970 ResourceManager
2549 NameNode
2826 SecondaryNameNode
3806 Jps
[root@localhost ~]#
```

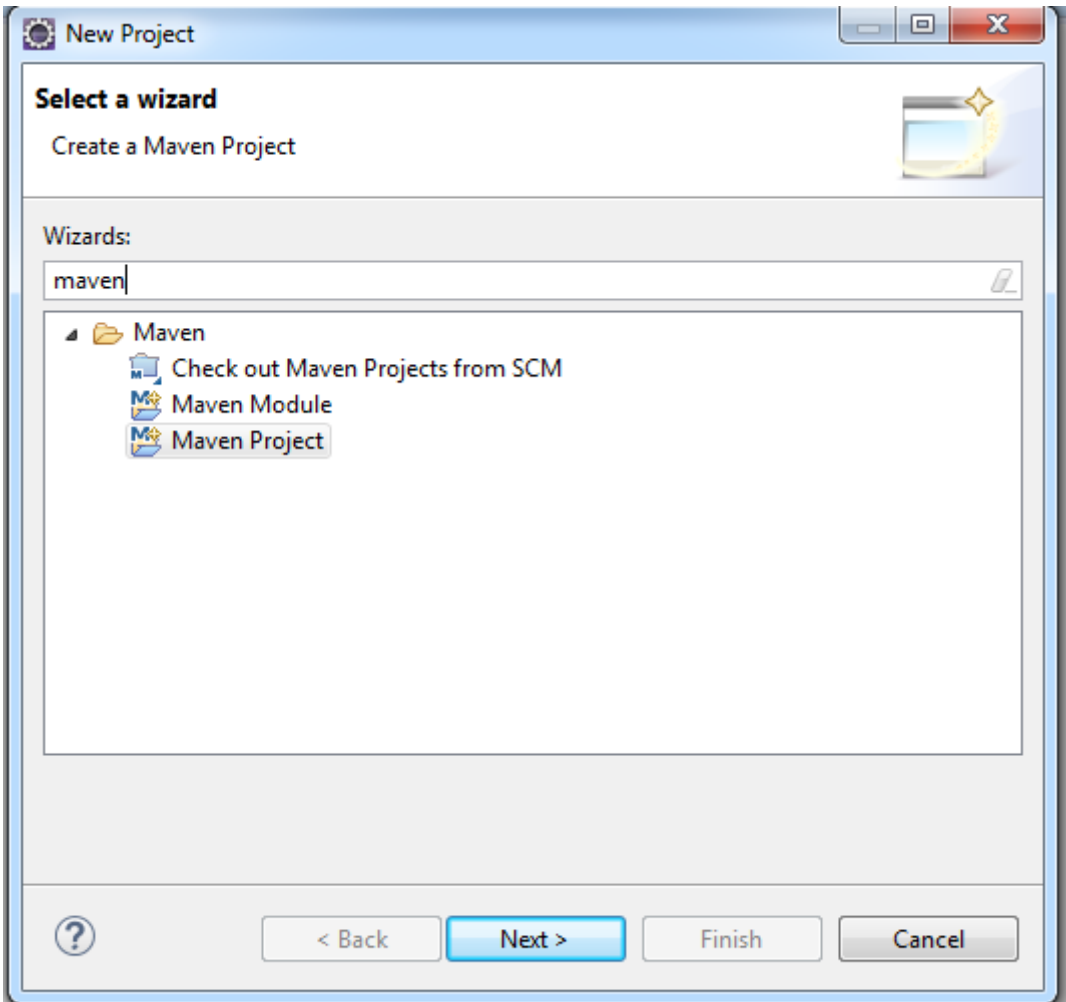
```
[root@localhost ~]# beeline
Beeline version 1.2.1 by Apache Hive
beeline> !connect jdbc:hive2://localhost:10000 scott tiger org.apache.hive.jdbc.HiveDriver
Connecting to jdbc:hive2://localhost:10000
Connected to: Apache Hive (version 1.2.1)
Driver: Hive JDBC (version 1.2.1)
Transaction isolation: TRANSACTION_REPEATABLE_READ
0: jdbc:hive2://localhost:10000>
```

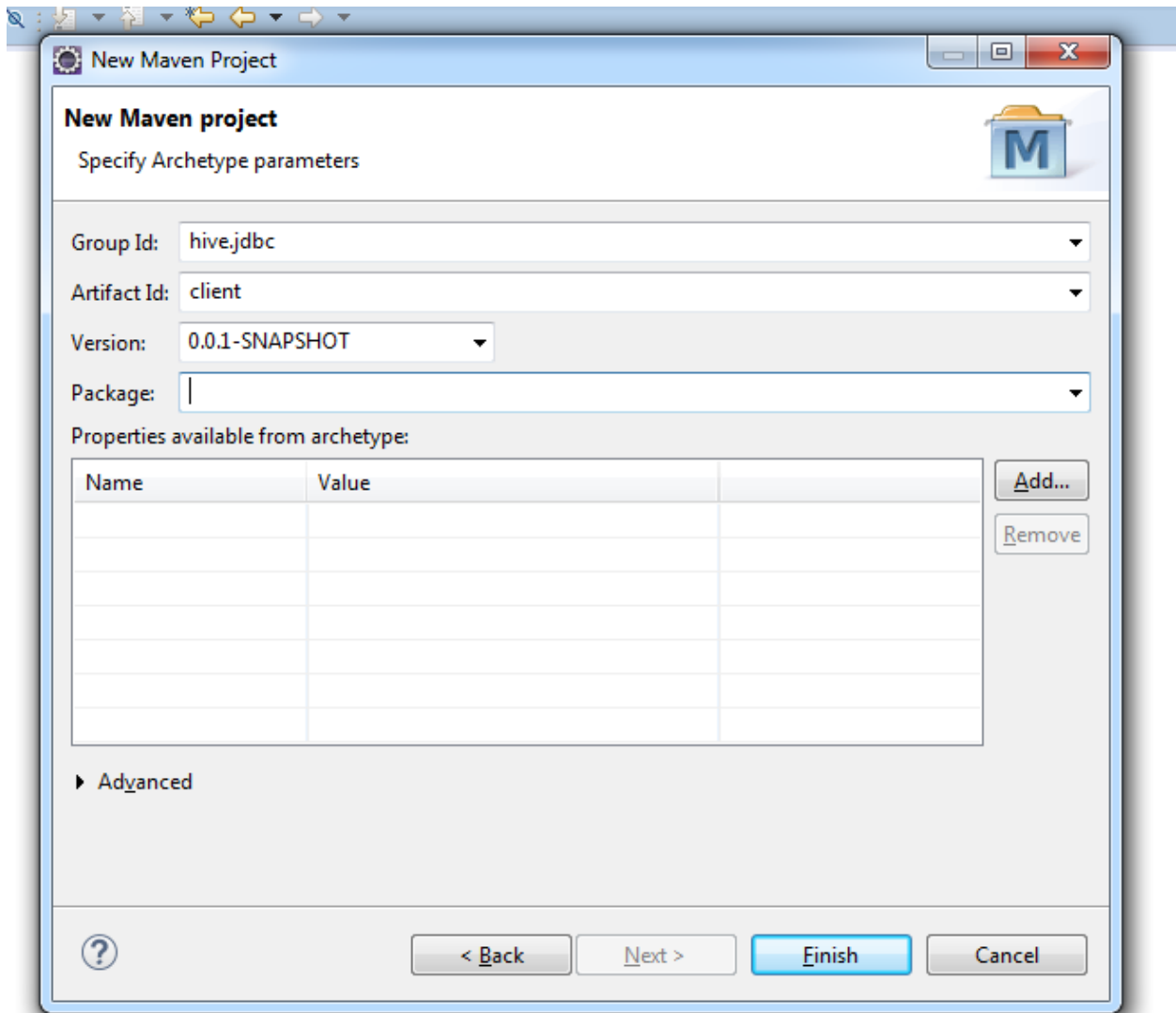
```
[root@localhost ~]# beeline -u 'jdbc:hive2://localhost:10000/default' -n root -p [REDACTED] -d org.apache.hive.jdbc.HiveDriver -e "select * from sales;"
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/opt/spark-1.6.0-bin-hadoop2.6/lib/spark-assembly-1.6.0-hadoop2.6.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/opt/hadoop-2.6.0/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
Connecting to jdbc:hive2://localhost:10000/default
Connected to: Apache Hive (version 1.2.1)
Driver: Hive JDBC (version 1.2.1)
Transaction isolation: TRANSACTION_REPEATABLE_READ
OK
+-----+-----+-----+-----+-----+-----+
| sales.id | sales.fname | sales.state | sales.zip | sales.ip | sales.pid |
+-----+-----+-----+-----+-----+-----+
| 0 | Zena | Tennessee | 21550 | 192.168.56.101 | PI_09 |
| 1 | Elaine | Alaska | 06429 | 192.168.56.101 | PI_03 |
| 2 | Sage | Nevada | 08899 | 192.168.56.102 | PI_03 |
| 3 | Cade | Missouri | 11233 | 192.168.56.103 | PI_06 |
| 4 | Abra | New Jersey | 21550 | 192.168.56.101 | PI_09 |
| 5 | Stone | Nebraska | 03560 | 192.168.56.104 | PI_08 |
| 6 | Regina | Tennessee | 21550 | 192.168.56.105 | PI_10 |
| 7 | Donovan | New York | 95234 | 192.168.56.106 | PI_05 |
| 8 | Aileen | Illinois | 68284 | 192.168.56.106 | PI_02 |
| 9 | Mariam | Hawaii | 95234 | 192.168.56.107 | PI_07 |
+-----+-----+-----+-----+-----+-----+

```

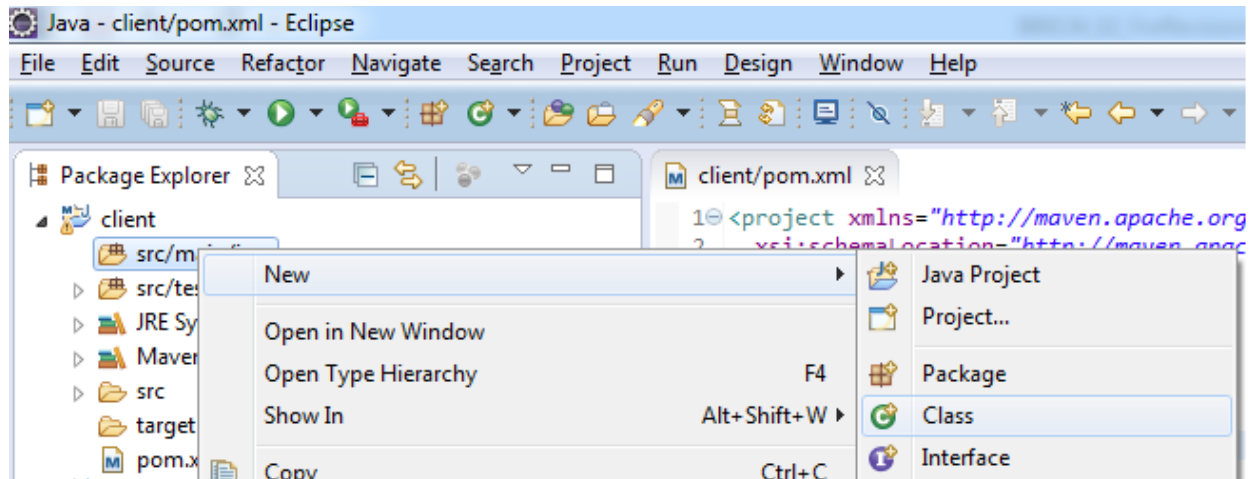
```
[root@localhost ~]# beeline -u 'jdbc:hive2://localhost:10000/default' -n root -p xxx -d org.apache.hive.jdbc.HiveDriver -f /opt/hivescript
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/opt/spark-1.6.0-bin-hadoop2.6/lib/spark-assembly-1.6.0-hadoop2.6.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/opt/hadoop-2.6.0/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
0: jdbc:hive2://localhost:10000/default> select fname from sales;
OK
+-----+
| fname |
+-----+
| Zena  |
| Elaine|
| Sage  |
| Cade  |
| Abra  |
| Stone |
| Regina|
| Donovan|
| Aileen|
| Mariam|
+-----+
10 rows selected (0.332 seconds)
0: jdbc:hive2://localhost:10000/default>
```

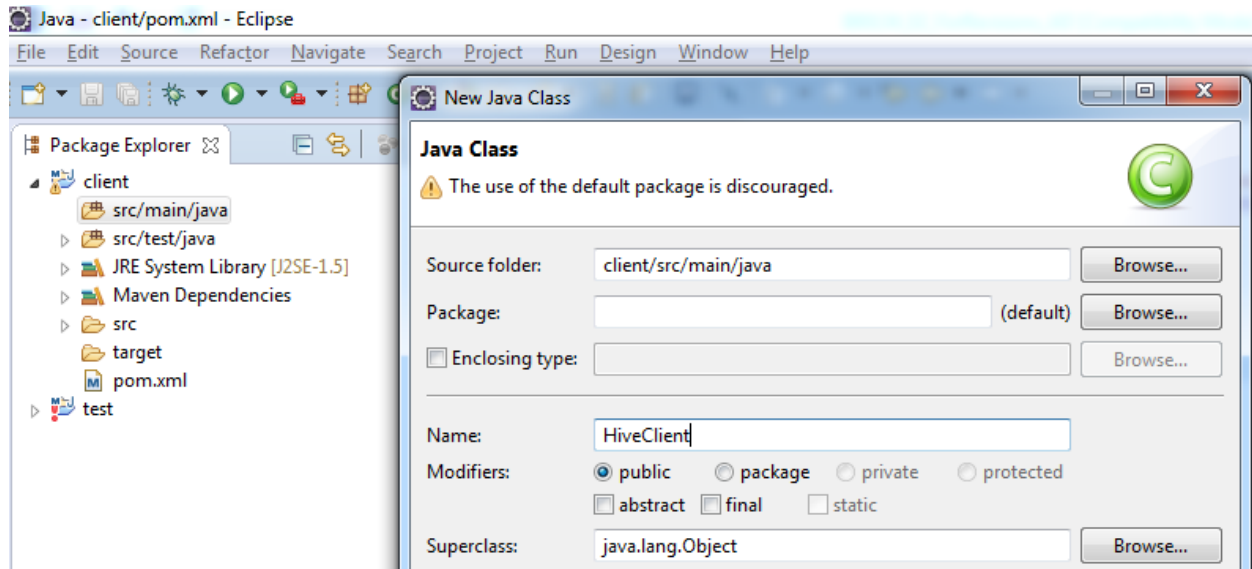






```
client/pom.xml ☒
1 <project xmlns="http://maven.apache.org/POM/4.0.0" xmlns:
2   xsi:schemaLocation="http://maven.apache.org/POM/4.0.0
3   <modelVersion>4.0.0</modelVersion>
4
5   <groupId>hive.jdbc</groupId>
6   <artifactId>client</artifactId>
7   <version>0.0.1-SNAPSHOT</version>
8   <packaging>jar</packaging>
9
10  <dependencies>
11    <dependency>
12      <groupId>org.apache.hive</groupId>
13      <artifactId>hive-jdbc</artifactId>
14      <version>1.2.1</version>
15    </dependency>
16    <dependency>
17      <groupId>org.apache.hadoop</groupId>
18      <artifactId>hadoop-common</artifactId>
19      <version>2.6.0</version>
20    </dependency>
21    <dependency>
22      <groupId>junit</groupId>
23      <artifactId>junit</artifactId>
24      <version>3.8.1</version>
25      <scope>test</scope>
26    </dependency>
27  </dependencies>
```





```
[root@localhost ~]# hive --service metastore &
[2] 3929
[root@localhost ~]#
```

```
Post-processed 'zookeeper/js/base64.js' as 'zookeeper/js/base64.ce5e02af31e5.js'
Post-processed 'zookeeper/help/index.html' as 'zookeeper/help/index.7570dbb625f31'
```

```
907 static files copied to '/usr/local/hue/build/static', 907 post-processed.
make[1]: Leaving directory `/usr/local/hue/apps'
--- Setting up Desktop database
make -C /usr/local/hue/desktop syncdb
make[1]: Entering directory `/usr/local/hue/desktop'
make[1]: Nothing to be done for `syncdb'.
make[1]: Leaving directory `/usr/local/hue/desktop'
```

```
[root@localhost ~]#
[root@localhost ~]# jps
3066 NodeManager
3929 RunJar
2642 DataNode
3713 RunJar
2970 ResourceManager
2549 NameNode
2826 SecondaryNameNode
4024 Jps
[root@localhost ~]#
```

HUE Query Editors Data Browsers Workflows Search Security File Browser Job Browser admin

Hive Editor My Queries Saved Queries History

Assist Settings

Hive
 Impala
 DB Query
 Pig
 Job Designer
 Spark (beta)

DATABASE

default

Table name...

sales

```
1 Example: SELECT * FROM tablename, or press CTRL + space
```

Execute Save as... Explain or create a New query

Recent queries Query Log Columns Results Chart

Time	Query	Result
03/12/2016 9:39:01 AM	show databases;	See results...

HUE Query Editors Data Browsers Workflows Search Security File

Hive Editor Query Editor My Queries Saved Queries History

Assist Settings

DATABASE

default

Table name...

sales

```
1 SELECT fname, lname FROM sales LIMIT 10;
```

Execute Save as... Explain or create a New query

Recent queries Query Log Columns Results Chart

	fname	lname
0	Zena	Ross
1	Elaine	Bishop
2	Sage	Carroll
3	Cade	Singleton
4	Abra	Wright
5	Stone	Palmer
6	Regina	Bryant
7	Donovan	Aguirre
8	Aileen	Mendoza
9	Mariam	Henson

Assist Settings

SETTINGS

Add

FILE RESOURCES

Add

UDFS

Add

OPTIONS

Enable parameterization

1

Execute Save as... Explain or create a New query

Recent queries			Query	Log	Columns	Results	Chart
Time	Query	Result					
03/12/2016 10:18:54 AM	SELECT fname, lname FROM sales LIMIT 10;	See results...					
03/12/2016 10:17:42 AM	SELECT fname, lname FROM sales LIMIT 10;	See results...					
03/12/2016 10:13:09 AM	load data inpath "/sample_10000.txt" into table sales;	See results...					
03/12/2016 10:11:24 AM	load data inpath "/sample_10000.txt" into table sales;	See results...					
03/12/2016 10:11:13 AM	load data inpath "/sample_1000.txt" into table sales;	See results...					
03/12/2016 9:39:01 AM	show databases;	See results...					

Chapter 3: Understanding the Hive Data Model

Hadoop Overview Datanodes Snapshot Startup Progress Utilities

Browse Directory

/user/hive/warehouse/sales_bucketed

Go!

Permission	Owner	Group	Size	Replication	Block Size	Name
-rwxr-xr-x	centos	supergroup	98.14 KB	1	128 MB	000000_0
-rwxr-xr-x	centos	supergroup	98.37 KB	1	128 MB	000001_0
-rwxr-xr-x	centos	supergroup	98 KB	1	128 MB	000002_0
-rwxr-xr-x	centos	supergroup	98.14 KB	1	128 MB	000003_0
-rwxr-xr-x	centos	supergroup	98.18 KB	1	128 MB	000004_0
-rwxr-xr-x	centos	supergroup	98.56 KB	1	128 MB	000005_0
-rwxr-xr-x	centos	supergroup	97.83 KB	1	128 MB	000006_0
-rwxr-xr-x	centos	supergroup	97.8 KB	1	128 MB	000007_0
-rwxr-xr-x	centos	supergroup	98.13 KB	1	128 MB	000008_0
-rwxr-xr-x	centos	supergroup	98.02 KB	1	128 MB	000009_0

Chapter 5: Hive Data Manipulation Language

```
hive> select * from sales;
OK
Time taken: 0.146 seconds
```

```
hive> LOAD DATA LOCAL INPATH '/tmp/sales.txt' INTO TABLE sales;
Loading data to table default.sales
Table default.sales stats: [numFiles=1, numRows=0, totalSize=1014721, rawDataSize=0]
OK
Time taken: 0.805 seconds
```

```
hive> select count(*) from sales;
OK
10000
Time taken: 19.353 seconds, Fetched: 1 row(s)
```

```
hive> select * from sales;
OK
Time taken: 0.146 seconds
```

```
hive> LOAD DATA INPATH '/sales.txt' INTO TABLE sales;
Loading data to table default.sales
Table default.sales stats: [numFiles=1, numRows=0, totalSize=1014721, rawDataSize=0]
OK
Time taken: 1.682 seconds
```

```
hive> select count(*) from sales;
OK
10000
Time taken: 19.353 seconds, Fetched: 1 row(s)
```

```
hive> LOAD DATA INPATH '/sales.txt' OVERWRITE INTO TABLE sales ;
Loading data to table default.sales
Table default.sales stats: [numFiles=1, numRows=0, totalSize=1014721, rawDataSize=0]
OK
Time taken: 1.29 seconds
```

```
hive> select count(*) from sales;
OK
10000
Time taken: 22.67 seconds, Fetched: 1 row(s)
```

```
hive> select count(*) from sales;
OK
2
Time taken: 20.086 seconds, Fetched: 1 row(s)
```

```
hive> select count(*) from sales_rgn;
OK
5000
Time taken: 20.086 seconds, Fetched: 1 row(s)
```

```
hive> select count(*) from sales;
OK
5002
Time taken: 19.925 seconds, Fetched: 1 row(s)
```

```
hive> select count(*) from sales;
OK
103
Time taken: 19.703 seconds, Fetched: 1 row(s)
```

```
hive> select count(*) from sales;
OK
5000
Time taken: 21.019 seconds, Fetched: 1 row(s)
```

```
hive> select count(*) from sales;
OK
1
Time taken: 21.019 seconds, Fetched: 1 row(s)
```

```
hive> select * from sales where id = 1;
OK
1      John   Terry   H-43 Sector-23  Delhi   India   10.10.10.10   P_1   15-11-1985
Time taken: 0.268 seconds, Fetched: 2 row(s)
```

```
hive> select * from sales where id = 2;
OK
2      Terry   John    H-43 Sector-23  Delhi   India   10.10.10.10
Time taken: 0.155 seconds, Fetched: 1 row(s)
```

```
hive> Create table employees (name string, age int, job string) clustered by (age) into 2 buckets stored as orc;
OK
Time taken: 0.308 seconds
```

```
hive> SELECT * FROM employees;
OK
John    30      IT
Jerry   35      Sales
Time taken: 0.11 seconds, Fetched: 2 row(s)
```

```
hive> Create table department (name string, age int, deptno int) Partitioned by (datestamp string);
OK
Time taken: 0.108 seconds
```

```
hive> select * from department;
OK
Jason   20      10      2015-10-23
Nelson  30      20      2015-10-23
Time taken: 0.152 seconds, Fetched: 2 row(s)
```

Chapter 6: Hive Extensibility Features

```

0      Zena      Tennessee      21550      192.168.56.101      PI_09
1      Elaine   Alaska      06429      192.168.56.101      PI_03
2      Sage     Nevada      08899      192.168.56.102      PI_03
3      Cade     Missouri     11233      192.168.56.103      PI_06
4      Abra     New Jersey   21550      192.168.56.101      PI_09
5      Stone    Nebraska     03560      192.168.56.104      PI_08
6      Regina   Tennessee    21550      192.168.56.105      PI_10
7      Donovan  New York     95234      192.168.56.106      PI_05
8      Aileen   Illinois     68284      192.168.56.106      PI_02
9      Mariam   Hawaii      95234      192.168.56.107      PI_07

```

```

mysql> select * from TBLS;
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| TBL_ID | CREATE_TIME | DB_ID | LAST_ACCESS_TIME | OWNER | RETENTION | SD_ID | TBL_NAME | TBL_TYPE | VIEW_EXPANDED_TEXT | VIEW_ORIGINAL_TEXT |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| 11 | 1455771821 | 1 | 0 | root | 0 | 11 | sales | MANAGED_TABLE | NULL | NULL |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

```

```

mysql> select * from SERDES;
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| SERDE_ID | NAME | SLIB |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| 11 | NULL | org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

```

```

hive> SELECT * FROM sales_view WHERE pid = 'PI_02' or pid = 'PI_03';
OK
1      Elaine   Alaska      06429      192.168.56.101      PI_03
8      Aileen   Illinois     68284      192.168.56.106      PI_02
Time taken: 0.986 seconds, Fetched: 2 row(s)

```

```

mysql> select * from TBLS;
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| TBL_ID | CREATE_TIME | DB_ID | LAST_ACCESS_TIME | OWNER | RETENTION | SD_ID | TBL_NAME | TBL_TYPE | VIEW_EXPANDED_TEXT | VIEW_ORIGINAL_TEXT |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| 19 | 1455870935 | 1 | 0 | root | 0 | 43 | sales_part | MANAGED_TABLE | NULL | NULL |
| 23 | 1455917422 | 1 | 0 | root | 0 | 59 | sales_part_state | MANAGED_TABLE | NULL | NULL |
| 36 | 1456027593 | 1 | 0 | root | 0 | 86 | sales_buck | MANAGED_TABLE | NULL | NULL |
| 41 | 1456039175 | 1 | 0 | root | 0 | 91 | web_logs | MANAGED_TABLE | NULL | NULL |
| 47 | 1456041934 | 1 | 0 | root | 0 | 97 | messages | EXTERNAL_TABLE | NULL | NULL |
| 52 | 1456276770 | 1 | 0 | root | 0 | 102 | sales | MANAGED_TABLE | NULL | NULL |
| 56 | 1456277389 | 1 | 0 | root | 0 | 106 | sales_view | VIRTUAL_VIEW | SELECT 'sales'.id, 'sales'.fname, 'sales'.state, 'sales'.zip, 'sales'.ip, 'sales'.pid FROM 'default'.'SALES' WHERE 'sales'.ip = '192.168.56.101' or 'sales'.ip='192.168.56.106' | SELECT * FROM SALES WHERE ip = '192.168.56.101' or ip='192.168.56.106' |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

```


Browse Directory

/user/hive/warehouse/sales_part_state

Permission	Owner	Group	Size	Replication	Block Size	Name
drwxr-xr-x	root	supergroup	0 B	0	0 B	state=Alaska
drwxr-xr-x	root	supergroup	0 B	0	0 B	state=Hawaii
drwxr-xr-x	root	supergroup	0 B	0	0 B	state=Illinois
drwxr-xr-x	root	supergroup	0 B	0	0 B	state=Missouri
drwxr-xr-x	root	supergroup	0 B	0	0 B	state=Nebraska
drwxr-xr-x	root	supergroup	0 B	0	0 B	state=Nevada
drwxr-xr-x	root	supergroup	0 B	0	0 B	state=New Jersey
drwxr-xr-x	root	supergroup	0 B	0	0 B	state=New York
drwxr-xr-x	root	supergroup	0 B	0	0 B	state=Tennessee

```
mysql> show tables;
+-----+
| Tables_in_metastore_db |
+-----+
| BUCKETING_COLS         |
| CDS                    |
| COLUMNS_V2            |
| DATABASE_PARAMS       |
| DBS                    |
| FUNCS                  |
| FUNC_RU                |
| GLOBAL_PRIVS           |
| IDXS                   |
| INDEX_PARAMS           |
| PARTITIONS             |
| PARTITION_KEYS         |
| PARTITION_KEY_VALS     |
| PARTITION_PARAMS       |
| PART_COL_PRIVS         |
| PART_COL_STATS         |
+-----+
```

```
mysql> select * from PARTITIONS;
```

PART_ID	CREATE_TIME	LAST_ACCESS_TIME	PART_NAME	SD_ID	TBL_ID
25	1455870997	0	pid=PI_03	44	19
26	1455870997	0	pid=PI_06	45	19
27	1455870997	0	pid=PI_08	46	19
28	1455870998	0	pid=PI_07	47	19
29	1455870998	0	pid=PI_02	48	19
30	1455870998	0	pid=PI_10	49	19
31	1455870998	0	pid=PI_09	50	19
32	1455870998	0	pid=PI_05	51	19
37	1455917531	0	state=Nevada	60	23
38	1455917532	0	state=New York	61	23
39	1455917532	0	state=New Jersey	62	23
40	1455917532	0	state=Alaska	63	23
41	1455917533	0	state=Tennessee	64	23
42	1455917533	0	state=Missouri	65	23
43	1455917533	0	state=Hawaii	66	23
44	1455917533	0	state=Illinois	67	23
45	1455917534	0	state=Nebraska	68	23

```
mysql> select * from PARTITION_KEYS;
```

TBL_ID	PKEY_COMMENT	PKEY_NAME	PKEY_TYPE	INTEGER_IDX
19	NULL	pid	string	0
23	NULL	state	string	0

```
mysql> select * from PARTITION_KEY_VALS;
```

PART_ID	PART_KEY_VAL	INTEGER_IDX
25	PI_03	0
26	PI_06	0
27	PI_08	0
28	PI_07	0
29	PI_02	0
30	PI_10	0
31	PI_09	0
32	PI_05	0
37	Nevada	0
38	New York	0
39	New Jersey	0
40	Alaska	0
41	Tennessee	0
42	Missouri	0
43	Hawaii	0
44	Illinois	0
45	Nebraska	0


```
hive> create table sales_buck (id int, fname string, state string, zip string, ip string, pid string) clustered by (id) into 50 buckets row format de
limited fields terminated by '\t';
OR
Time taken: 0.238 seconds
hive> set hive.enforce.bucketing=true;
hive> insert into table sales_buck select * from sales;
Query ID = root_20160224073617_42330cfc-4cd8-42f1-b69b-7a6cdc4565f1
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 50
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_1456273553886_0004, Tracking URL = http://localhost:8088/proxy/application_1456273553886_0004/
Kill Command = /opt/hadoop-2.6.0/bin/hadoop job -kill job_1456273553886_0004
Hadoop job information for Stage-1: number of mappers: 1, number of reducers: 50
2016-02-24 07:36:41,958 Stage-1 map = 0%, reduce = 0%
```

Browse Directory

/user/hive/warehouse/sales_buck

Permission	Owner	Group	Size	Replication	Block Size	Name
-rwxr-xr-x	root	supergroup	87 B	3	256 MB	000000_1
-rwxr-xr-x	root	supergroup	106 B	3	256 MB	000001_1
-rwxr-xr-x	root	supergroup	90 B	3	256 MB	000002_1
-rwxr-xr-x	root	supergroup	95 B	3	256 MB	000003_2
-rwxr-xr-x	root	supergroup	99 B	3	256 MB	000004_1
-rwxr-xr-x	root	supergroup	93 B	3	256 MB	000005_1
-rwxr-xr-x	root	supergroup	94 B	3	256 MB	000006_0
-rwxr-xr-x	root	supergroup	105 B	3	256 MB	000007_0
-rwxr-xr-x	root	supergroup	109 B	3	256 MB	000008_0
-rwxr-xr-x	root	supergroup	88 B	3	256 MB	000009_0

```
mysql> select * from BUCKETING_COLS;
+-----+-----+-----+
| SD_ID | BUCKET_COL_NAME | INTEGER_IDX |
+-----+-----+-----+
| 109 | id | 0 |
+-----+-----+-----+
```

fname	ip	rownum
Abra	192.168.56.101	1
Elaine	192.168.56.101	2
Zena	192.168.56.101	3
Sage	192.168.56.102	4
Cade	192.168.56.103	5
Stone	192.168.56.104	6
Regina	192.168.56.105	7
Aileen	192.168.56.106	8
Donovan	192.168.56.106	9
Mariam	192.168.56.107	10

fname	ip	ranknum	rank_window_1
Abra	192.168.56.101	1	1
Elaine	192.168.56.101	1	2
Zena	192.168.56.101	1	3
Sage	192.168.56.102	4	1
Cade	192.168.56.103	5	1
Stone	192.168.56.104	6	1
Regina	192.168.56.105	7	1
Aileen	192.168.56.106	8	1
Donovan	192.168.56.106	8	2
Mariam	192.168.56.107	10	1

fname	ip	densenum	dense_rank_window_1
Abra	192.168.56.101	1	1
Elaine	192.168.56.101	1	2
Zena	192.168.56.101	1	3
Sage	192.168.56.102	2	1
Cade	192.168.56.103	3	1
Stone	192.168.56.104	4	1
Regina	192.168.56.105	5	1
Aileen	192.168.56.106	6	1
Donovan	192.168.56.106	6	2
Mariam	192.168.56.107	7	1

fname	ip	row_number_window_0	rank_window_1	dense_rank_window_2
Abra	192.168.56.101	1	1	1
Elaine	192.168.56.101	2	1	1
Zena	192.168.56.101	3	1	1
Sage	192.168.56.102	4	4	2
Cade	192.168.56.103	5	5	3
Stone	192.168.56.104	6	6	4
Regina	192.168.56.105	7	7	5
Aileen	192.168.56.106	8	8	6
Donovan	192.168.56.106	9	8	6
Mariam	192.168.56.107	10	10	7

fname	ip	cume_dist
Abra	192.168.56.101	0.3333333333333333
Elaine	192.168.56.101	0.6666666666666666
Zena	192.168.56.101	1.0
Sage	192.168.56.102	1.0
Cade	192.168.56.103	1.0
Stone	192.168.56.104	1.0
Regina	192.168.56.105	1.0
Aileen	192.168.56.106	0.5
Donovan	192.168.56.106	1.0
Mariam	192.168.56.107	1.0

fname	id	quartile
Abra	4	1
Elaine	1	2
Zena	0	3

fname	id	quartile
Abra	4	1
Elaine	1	1
Zena	0	2

ip	fname	percent_rank
192.168.56.101	Abra	0.0
192.168.56.101	Elaine	0.5
192.168.56.101	Zena	1.0
192.168.56.102	Sage	0.0
192.168.56.103	Cade	0.0
192.168.56.104	Stone	0.0
192.168.56.105	Regina	0.0
192.168.56.106	Aileen	0.0
192.168.56.106	Donovan	1.0
192.168.56.107	Mariam	0.0

fname	ip	count_window_0
Abra	192.168.56.101	3
Elaine	192.168.56.101	3
Zena	192.168.56.101	3
Sage	192.168.56.102	1
Cade	192.168.56.103	1
Stone	192.168.56.104	1
Regina	192.168.56.105	1
Aileen	192.168.56.106	2
Donovan	192.168.56.106	2
Mariam	192.168.56.107	1

fname	ip	zip	pid	count	window_0
Elaine	192.168.56.101	06429	PI_03	1	
Abra	192.168.56.101	21550	PI_09	2	
Zena	192.168.56.101	21550	PI_09	2	
Sage	192.168.56.102	08899	PI_03	1	
Cade	192.168.56.103	11233	PI_06	1	
Stone	192.168.56.104	03560	PI_08	1	
Regina	192.168.56.105	21550	PI_10	1	
Aileen	192.168.56.106	68284	PI_02	1	
Donovan	192.168.56.106	95234	PI_05	1	
Mariam	192.168.56.107	95234	PI_07	1	

fname	pid	count	window_0
Abra	PI_09	1	
Elaine	PI_03	2	
Zena	PI_09	3	
Sage	PI_03	1	
Cade	PI_06	1	
Stone	PI_08	1	
Regina	PI_10	1	
Aileen	PI_02	1	
Donovan	PI_05	2	
Mariam	PI_07	1	

fname	ip	pid	count	window_0
Elaine	192.168.56.101	PI_03	1	
Abra	192.168.56.101	PI_09	1	
Zena	192.168.56.101	PI_09	2	
Sage	192.168.56.102	PI_03	1	
Cade	192.168.56.103	PI_06	1	
Stone	192.168.56.104	PI_08	1	
Regina	192.168.56.105	PI_10	1	
Aileen	192.168.56.106	PI_02	1	
Donovan	192.168.56.106	PI_05	1	
Mariam	192.168.56.107	PI_07	1	

fname	ip	count	window_0
Abra	192.168.56.101	1	
Elaine	192.168.56.101	2	
Zena	192.168.56.101	3	
Sage	192.168.56.102	1	
Cade	192.168.56.103	1	
Stone	192.168.56.104	1	
Regina	192.168.56.105	1	
Aileen	192.168.56.106	1	
Donovan	192.168.56.106	2	
Mariam	192.168.56.107	1	

```
fname ip count_window_0
Abra 192.168.56.101 1
Elaine 192.168.56.101 2
Zena 192.168.56.101 3
Sage 192.168.56.102 1
Cade 192.168.56.103 1
Stone 192.168.56.104 1
Regina 192.168.56.105 1
Aileen 192.168.56.106 1
Donovan 192.168.56.106 2
Mariam 192.168.56.107 1
```

```
fname ip count_window_0
Abra 192.168.56.101 3
Elaine 192.168.56.101 3
Zena 192.168.56.101 3
Sage 192.168.56.102 1
Cade 192.168.56.103 1
Stone 192.168.56.104 1
Regina 192.168.56.105 1
Aileen 192.168.56.106 2
Donovan 192.168.56.106 2
Mariam 192.168.56.107 1
```

```
fname ip count_window_0
Abra 192.168.56.101 3
Elaine 192.168.56.101 2
Zena 192.168.56.101 1
Sage 192.168.56.102 1
Cade 192.168.56.103 1
Stone 192.168.56.104 1
Regina 192.168.56.105 1
Aileen 192.168.56.106 2
Donovan 192.168.56.106 1
Mariam 192.168.56.107 1
```

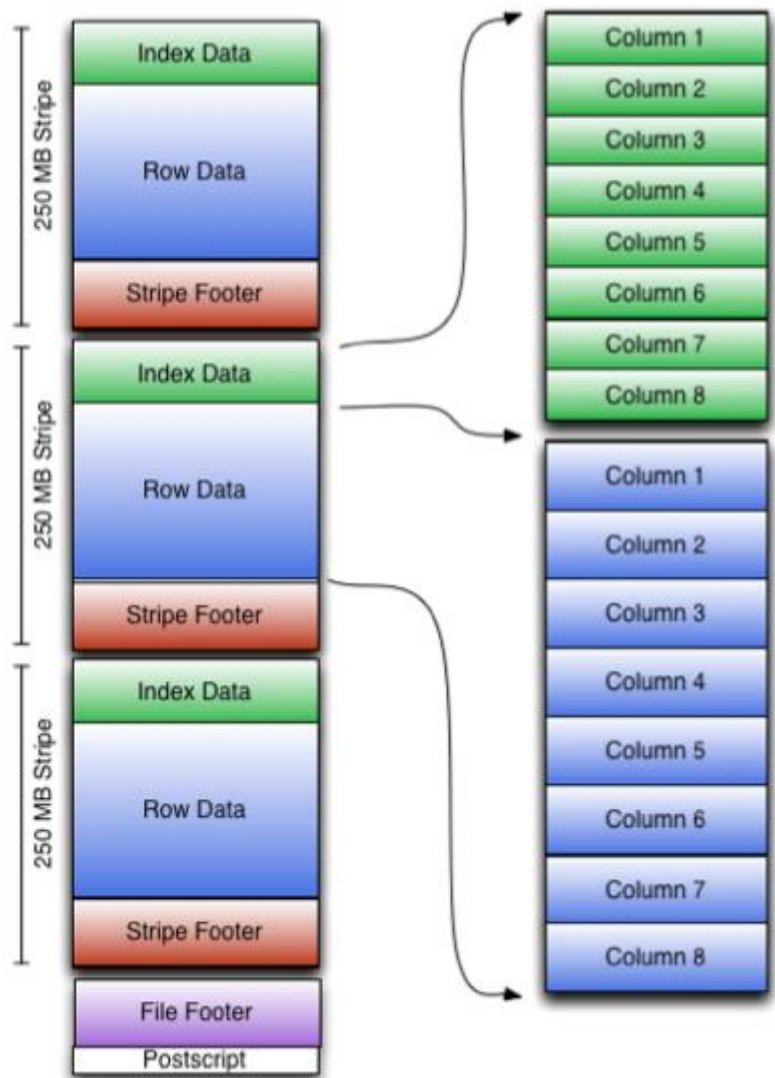
```
fname ip zip _c3 count_window_1
Stone 192.168.56.104 03560 1 1
Elaine 192.168.56.101 06429 3 1
Sage 192.168.56.102 08899 1 1
Cade 192.168.56.103 11233 1 1
Regina 192.168.56.105 21550 1 3
Zena 192.168.56.101 21550 3 3
Abra 192.168.56.101 21550 3 3
Aileen 192.168.56.106 68284 2 1
Mariam 192.168.56.107 95234 1 2
Donovan 192.168.56.106 95234 2 2
```

fname	pid	lead_window_0
Abra	PI_09	PI_03
Elaine	PI_03	PI_09
Zena	PI_09	NULL
Sage	PI_03	NULL
Cade	PI_06	NULL
Stone	PI_08	NULL
Regina	PI_10	NULL
Aileen	PI_02	PI_05
Donovan	PI_05	NULL
Mariam	PI_07	NULL

fname	pid	lag_window_0
Abra	PI_09	NULL
Elaine	PI_03	PI_09
Zena	PI_09	PI_03
Sage	PI_03	NULL
Cade	PI_06	NULL
Stone	PI_08	NULL
Regina	PI_10	NULL
Aileen	PI_02	NULL
Donovan	PI_05	PI_02
Mariam	PI_07	NULL

fname	ip	pid
Abra	192.168.56.101	PI_09
Elaine	192.168.56.101	PI_09
Zena	192.168.56.101	PI_09
Sage	192.168.56.102	PI_03
Cade	192.168.56.103	PI_06
Stone	192.168.56.104	PI_08
Regina	192.168.56.105	PI_10
Aileen	192.168.56.106	PI_02
Donovan	192.168.56.106	PI_02
Mariam	192.168.56.107	PI_07

fname	ip	pid
Abra	192.168.56.101	PI_09
Elaine	192.168.56.101	PI_03
Zena	192.168.56.101	PI_09
Sage	192.168.56.102	PI_03
Cade	192.168.56.103	PI_06
Stone	192.168.56.104	PI_08
Regina	192.168.56.105	PI_10
Aileen	192.168.56.106	PI_02
Donovan	192.168.56.106	PI_05
Mariam	192.168.56.107	PI_07



Chapter 7: Joins and Join Optimization

```
hive> SELECT COUNT(*) FROM Sales;
OK
10000
Time taken: 24.154 seconds, Fetched: 1 row(s)
```

```
SELECT COUNT(*) FROM Sales_orc;
OK
51
Time taken: 19.793 seconds, Fetched: 1 row(s)
```

```
hive> SELECT COUNT(*) FROM Sales_info;
OK
101
Time taken: 19.925 seconds, Fetched: 1 row(s)
```

```
hive> SELECT a.* FROM Sales a JOIN Sales_orc b ON a.id = b.id;
OK
0      Zena      Ross      41228 West India Ln.    Powell Tennessee      21550  192.168.56.127  PI_09
50     Dorothy Duke  47365 West Corona Blvd. Jeffersonville Colorado  50217  192.168.56.206  PI_01
48     Jocelyn Warner 44429 Ogdensburg St.   Chester Washington   80832  192.168.56.201  PI_01
46     Bradley Mcgowan 65530 South Hackensack Ct. Fullerton Illinois      73467  192.168.56.141  PI_02
```

```
hive> SELECT a.* FROM Sales a JOIN Sales_orc b ON a.id <> b.id;
FAILED: SemanticException [Error 10017]: Line 1:44 Both left and right aliases encountered in JOIN 'id'
hive>
```

```
hive> SELECT a.* FROM Sales a, Sales_orc b where a.id = b.id;
OK
0      Zena      Ross      41228 West India Ln.    Powell Tennessee      21550  192.168.56.127  PI_09
50     Dorothy Duke  47365 West Corona Blvd. Jeffersonville Colorado  50217  192.168.56.206  PI_01
48     Jocelyn Warner 44429 Ogdensburg St.   Chester Washington   80832  192.168.56.201  PI_01
46     Bradley Mcgowan 65530 South Hackensack Ct. Fullerton Illinois      73467  192.168.56.141  PI_02
44     Hoyt      Howard  27401 North Qatar Ave.  Chicopee      New York      21279  192.168.56.80   PI_02
```

```
hive> SELECT a.*, b.* FROM Sales a JOIN Sales_orc b ON a.id = b.id;
OK
0      Zena      Ross      41228 West India Ln.    Powell Tennessee      21550  192.168.56.127  PI_09  0      Zena
e      21550  192.168.56.127  PI_09
50     Dorothy Duke  47365 West Corona Blvd. Jeffersonville Colorado  50217  192.168.56.206  PI_01  50
nville Colorado      50217  192.168.56.206  PI_01
```

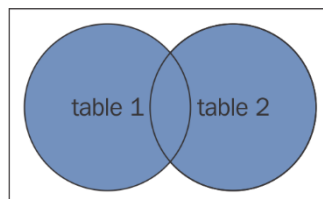
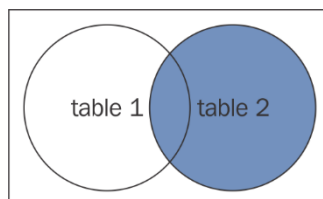
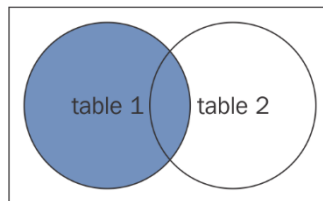
```
hive> SELECT a.fname, b.lname FROM Sales a JOIN Sales_orc b ON a.id = b.id;
OK
Zena      Ross
Dorothy Duke
Jocelyn Warner
Bradley Mcgowan
Hoyt      Howard
Wynter Galloway
Fiona      Patton
Marvin      Gamble
```



```
hive> SELECT a.* FROM Sales a JOIN Sales_orc b ON a.id = b.id and a.fname = b.fname;
OK
0      Zena  Ross  41228 West India Ln.  Powell Tennessee  21550  192.168.56.127  PI_09
50     Dorothy Duke  47365 West Corona Blvd. Jeffersonville Colorado  50217  192.168.56.206  PI_01
48     Jocelyn Warner  44429 Ogdensburg St.  Chester Washington  80832  192.168.56.201  PI_01
46     Bradley Mcgowan  65530 South Hackensack Ct.  Fullerton Illinois  73467  192.168.56.141  PI_02
44     Hoyt  Howard  27401 North Qatar Ave.  Chicopee New York  21279  192.168.56.80  PI_02
42     Wynter Galloway  68748 West Nicaragua Ave.  Aguadilla Utah  61982  192.168.56.157  PI_04
40     Fiona  Patton  53793 North Norfolk Island Ln.  Hornell South Carolina  03962  192.168.56.70  PI_05
```

```
hive> SELECT a.fname, b.lname, c.address FROM Sales a JOIN Sales_orc b ON a.id = b.id join Sales_info c ON c.id = b.id;
OK
Zena  Ross  41228 West India Ln.
Dorothy Duke  47365 West Corona Blvd.
Jocelyn Warner  44429 Ogdensburg St.
Bradley Mcgowan  65530 South Hackensack Ct.
Hoyt  Howard  27401 North Qatar Ave.
Wynter Galloway  68748 West Nicaragua Ave.
```

```
hive> SELECT a.fname, b.lname, c.address FROM Sales a JOIN Sales_orc b ON a.id = b.id join Sales_info c ON c.address = b.address;
OK
Zena  Ross  41228 West India Ln.
Dorothy Duke  47365 West Corona Blvd.
Jocelyn Warner  44429 Ogdensburg St.
Bradley Mcgowan  65530 South Hackensack Ct.
Hoyt  Howard  27401 North Qatar Ave.
```



```
hive> SELECT * FROM Sales a LEFT OUTER JOIN Sales_orc b ON a.id = b.id;
OK
0      Zena  Ross  41228 West India Ln.  Powell Tennessee  21550  192.168.56.127  PI_09  0      Zena  Ross
e      21550  192.168.56.127  PI_09
50     Dorothy Duke  47365 West Corona Blvd. Jeffersonville Colorado  50217  192.168.56.206  PI_01  50     Dorothy
nville Colorado  50217  192.168.56.206  PI_01
48     Jocelyn Warner  44429 Ogdensburg St.  Chester Washington  80832  192.168.56.201  PI_01  48     Jocelyn Warner
on      80832  192.168.56.201  PI_01
46     Bradley Mcgowan  65530 South Hackensack Ct.  Fullerton Illinois  73467  192.168.56.141  PI_02  46
t.     Fullerton Illinois  73467  192.168.56.141  PI_02
44     Hoyt  Howard  27401 North Qatar Ave.  Chicopee New York  21279  192.168.56.80  PI_02  44     Hoyt
```

```
hive> SELECT * FROM Sales a RIGHT OUTER JOIN Sales_orc b ON a.id = b.id;
OK
0      Zena  Ross  41228 West India Ln.  Powell Tennessee  21550  192.168.56.127  PI_09  0      Zena  Ross
e      21550  192.168.56.127  PI_09
1      Elaine Bishop  15903 North North Adams Blvd.  Hawaiian Gardens Alaska  06429  192.168.56.105  PI_03  1
Blvd.  Hawaiian Gardens Alaska  06429  192.168.56.105  PI_03
2      Sage Carroll  6880 Greenland Ct.  Guayanilla Nevada  08899  192.168.56.40  PI_03  2      Sage  Carroll
evada  08899  192.168.56.40  PI_03
3      Cade Singleton  64021 South Bulgaria Ave.  Derby Missouri  11233  192.168.56.171  PI_06  3
aria Ave.  Derby Missouri  11233  192.168.56.171  PI_06
4      Abra Wright  50155 South Mongolia Ln.  Port Jervis New Jersey  17751  192.168.56.52  PI_09  4
Port Jervis New Jersey  17751  192.168.56.52  PI_09
```

```
hive> SELECT * FROM Sales a FULL OUTER JOIN Sales_orc b ON a.id = b.id;
9979  Amaya  Madden  89534 West Tokelau Blvd.  Moorhead New York  10087  192.168.56.0  PI_04  NULL
ULL   NULL   NULL
9980  Mira  Phelps  74956 North Turkmenistan Ct.  Carolina North Carolina  62789  192.168.56.6  PI_04  NULL
ULL   NULL   NULL
9981  Cain  Shepherd  65814 West Christmas Island Blvd.  Oneonta North Carolina  17014  192.168.56.140  PI_06
ULL   NULL   NULL  NULL
9982  Ria  Hansen  23700 West Bradford Blvd.  Missoula West Virginia  99380  192.168.56.121  PI_06  NULL
ULL   NULL   NULL
9983  Tate  Roach  18461 North Japan Way  Wheeling California  16368  192.168.56.38  PI_05  NULL  NULL
ULL   NULL
9984  Dolan  Briggs  46722 East Congo Blvd.  Ansonia Missouri  58463  192.168.56.247  PI_06  NULL  NULL  NULL
ULL
9985  MacKenzie  Swanson  56966 South Chad Ln.  Hialeah Washington  82211  192.168.56.221  PI_02  NULL  NULL
```

```
hive> SELECT * FROM Sales a LEFT OUTER JOIN Sales_orc b ON a.id = b.id WHERE a.fname = 'John';
OK
9898  John  Norman  88176 South Cambodia Ave.  Laguna Woods Delaware  87857  192.168.56.123  PI_06  NULL  NULL
ULL   NULL   NULL
5552  John  Leonard  92838 Cameroon Ave.  Brookfield Massachusetts  82157  192.168.56.183  PI_06  NULL  NULL  NULL
ULL   NULL
5438  John  Case  45250 North Micronesia Blvd.  La Puente Idaho  27577  192.168.56.172  PI_10  NULL  NULL  NULL
ULL   NULL
9587  John  Rivers  49544 South Norton Ln.  Scarborough Missouri  71896  192.168.56.91  PI_05  NULL  NULL  NULL
ULL   NULL
7759  John  Conley  72203 South Areceibo Ln.  Chattanooga South Carolina  91570  192.168.56.37  PI_04  NULL  NULL
ULL   NULL  NULL
5109  John  Hancock  75125 India Way  Methuen North Dakota  39688  192.168.56.121  PI_02  NULL  NULL  NULL  NULL
ULL
3841  John  Flowers  98976 North Bolivia St.  San Luis Obispo Texas  86173  192.168.56.71  PI_06  NULL  NULL  NULL  NULL
ULL
3411  John  Russo  41922 South Sierra Leone Ave.  Chesapeake Wyoming  43952  192.168.56.98  PI_06  NULL  NULL  NULL
ULL   NULL
971   John  Merritt  55106 South Armenia Way Ypsilanti California  13701  192.168.56.7  PI_09  NULL  NULL  NULL
ULL   NULL
```

Time taken: 27.702 seconds, Fetched: 9 row(s)

```
hive> SELECT * FROM Sales a RIGHT OUTER JOIN Sales_orc b ON a.id = b.id WHERE a.fname = 'John';
OK
971   John  Merritt  55106 South Armenia Way Ypsilanti California  13701  192.168.56.7  PI_09  971   John
i     California  13701  192.168.56.7  PI_09
Time taken: 20.568 seconds, Fetched: 1 row(s)
```

```
hive> SELECT a.* FROM Sales a LEFT SEMI JOIN Sales_orc b ON a.id = b.id;
OK
0      Zena      Ross      41228 West India Ln.      Powell Tennessee      21550      192.168.56.127      PI_09
50     Dorothy Duke 47365 West Corona Blvd. Jeffersonville Colorado      50217      192.168.56.206      PI_01
48     Jocelyn Warner 44429 Ogdensburg St. Chester Washington      80832      192.168.56.201      PI_01
46     Bradley McGowan 65530 South Hackensack Ct. Fullerton Illinois      73467      192.168.56.141      PI_02
44     Hoyt Howard 27401 North Qatar Ave. Chicopee New York      21279      192.168.56.80      PI_02
42     Wynter Galloway 68748 West Nicaragua Ave. Aguadilla Utah      61982      192.168.56.157      PI_04
40     Fiona Patton 53793 North Norfolk Island Ln. Hornell South Carolina      03962      192.168.56.70      PI_05
38     Marvin Gamble 98294 South Liberia St. Manitowoc Missouri      41805      192.168.56.136      PI_03
36     Merritt Waters 73679 North Norfolk Island St. San Juan Capistrano Louisiana      45934      192.168.56.157      PI_02
```

```
hive> SELECT a.* FROM Sales a WHERE a.id IN (SELECT b.id FROM Sales_orc b);
OK
0      Zena      Ross      41228 West India Ln.      Powell Tennessee      21550      192.168.56.127      PI_09
50     Dorothy Duke 47365 West Corona Blvd. Jeffersonville Colorado      50217      192.168.56.206      PI_01
48     Jocelyn Warner 44429 Ogdensburg St. Chester Washington      80832      192.168.56.201      PI_01
46     Bradley McGowan 65530 South Hackensack Ct. Fullerton Illinois      73467      192.168.56.141      PI_02
44     Hoyt Howard 27401 North Qatar Ave. Chicopee New York      21279      192.168.56.80      PI_02
42     Wynter Galloway 68748 West Nicaragua Ave. Aguadilla Utah      61982      192.168.56.157      PI_04
40     Fiona Patton 53793 North Norfolk Island Ln. Hornell South Carolina      03962      192.168.56.70      PI_05
38     Marvin Gamble 98294 South Liberia St. Manitowoc Missouri      41805      192.168.56.136      PI_03
36     Merritt Waters 73679 North Norfolk Island St. San Juan Capistrano Louisiana      45934      192.168.56.157      PI_02
```

```
hive> SELECT a.*, b.* FROM Sales a LEFT SEMI JOIN Sales_orc b ON a.id = b.id;
FAILED: SemanticException [Error 10009]: Line 1:12 Invalid table alias 'b'
```

```
hive> SELECT a.* FROM Sales a LEFT SEMI JOIN Sales_orc b ON a.id = b.id WHERE b.id = 1;
FAILED: SemanticException [Error 10004]: Line 1:72 Invalid table alias or column reference 'b':
p_id, dop)
```

```
hive> SELECT * FROM Sales JOIN Sales_orc;
4999      Jamal      Howell      26029 Pitcairn Ln.      Stafford      South Carolina      26305
asthampton      California      53026      192.168.56.152      PI_01
4999      Jamal      Howell      26029 Pitcairn Ln.      Stafford      South Carolina      26305
ullerton      Illinois      73467      192.168.56.141      PI_02
4999      Jamal      Howell      26029 Pitcairn Ln.      Stafford      South Carolina      26305
rth      Virginia      57375      192.168.56.10      PI_05
4999      Jamal      Howell      26029 Pitcairn Ln.      Stafford      South Carolina      26305
ashington      80832      192.168.56.201      PI_01
4999      Jamal      Howell      26029 Pitcairn Ln.      Stafford      South Carolina      26305
ichigan 69621      192.168.56.191      PI_01
4999      Jamal      Howell      26029 Pitcairn Ln.      Stafford      South Carolina      26305
nville Colorado      50217      192.168.56.206      PI_01
4999      Jamal      Howell      26029 Pitcairn Ln.      Stafford      South Carolina      26305
i      California      13701      192.168.56.7      PI_09
Time taken: 32.72 seconds, Fetched: 520000 row(s)
```

```
hive> SELECT * FROM Sales JOIN Sales_orc WHERE Sales.id = 1;
1      Elaine Bishop 15903 North North Adams Blvd. Hawaiian Gardens      Alaska      06429
t.      Fullerton      Illinois      73467      192.168.56.141      PI_02
1      Elaine Bishop 15903 North North Adams Blvd. Hawaiian Gardens      Alaska      06429
omersworth      Virginia      57375      192.168.56.10      PI_05
1      Elaine Bishop 15903 North North Adams Blvd. Hawaiian Gardens      Alaska      06429
hester Washington      80832      192.168.56.201      PI_01
1      Elaine Bishop 15903 North North Adams Blvd. Hawaiian Gardens      Alaska      06429
ilford Michigan      69621      192.168.56.191      PI_01
1      Elaine Bishop 15903 North North Adams Blvd. Hawaiian Gardens      Alaska      06429
effersonville Colorado      50217      192.168.56.206      PI_01
1      Elaine Bishop 15903 North North Adams Blvd. Hawaiian Gardens      Alaska      06429
psilanti      California      13701      192.168.56.7      PI_09
Time taken: 30.331 seconds, Fetched: 52 row(s)
```



```
hive> SELECT * FROM Sales CROSS JOIN Sales_orc;
4999   Jamal   Howell  26029  Pitcairn Ln.      Stafford      South Carolina  26305
ashington      80832   192.168.56.201  PI_01
4999   Jamal   Howell  26029  Pitcairn Ln.      Stafford      South Carolina  26305
ichigan 69621   192.168.56.191  PI_01
4999   Jamal   Howell  26029  Pitcairn Ln.      Stafford      South Carolina  26305
nville  Colorado      50217   192.168.56.206  PI_01
4999   Jamal   Howell  26029  Pitcairn Ln.      Stafford      South Carolina  26305
i       California  13701   192.168.56.7    PI_09
Time taken: 35.41 seconds, Fetched: 520000 row(s)
```

```
hive> SELECT * FROM Sales a CROSS JOIN Sales_orc b JOIN Location c on a.id = c.id;
1      Elaine  Bishop  15903  North North Adams Blvd.  Hawaiian Gardens  Alaska  06429  192.168.56.105
omersworth  Virginia      57375   192.168.56.10  PI_05  1      Elaine  Bishop  15903  North North Adams
56.105  PI_03
1      Elaine  Bishop  15903  North North Adams Blvd.  Hawaiian Gardens  Alaska  06429  192.168.56.105
hester  Washington  80832   192.168.56.201  PI_01  1      Elaine  Bishop  15903  North North Adams Blvd.
I_03
1      Elaine  Bishop  15903  North North Adams Blvd.  Hawaiian Gardens  Alaska  06429  192.168.56.105
ilford  Michigan      69621   192.168.56.191  PI_01  1      Elaine  Bishop  15903  North North Adams Blvd.
I_03
1      Elaine  Bishop  15903  North North Adams Blvd.  Hawaiian Gardens  Alaska  06429  192.168.56.105
effersonville  Colorado      50217   192.168.56.206  PI_01  1      Elaine  Bishop  15903  North North Adams
56.105  PI_03
1      Elaine  Bishop  15903  North North Adams Blvd.  Hawaiian Gardens  Alaska  06429  192.168.56.105
psilanti  California  13701   192.168.56.7    PI_09  1      Elaine  Bishop  15903  North North Adams
56.105  PI_03
Time taken: 31.478 seconds, Fetched: 572 row(s)
```

```
hive> set hive.auto.convert.join=true;
```

```
hive> SELECT /*+ MAPJOIN(Sales_orc)*/ a.fname, b.lname FROM Sales a JOIN Sales_orc b ON a.id = b.id;
Query ID = hadoop_20160211234040_f59e1ffc-7908-4ca3-b086-d0e6d0ace54c
Total jobs = 1
Execution log at: /tmp/hadoop/hadoop_20160211234040_f59e1ffc-7908-4ca3-b086-d0e6d0ace54c.log
2016-02-11 11:41:02 Starting to launch local task to process map join; maximum memory = 518979584
2016-02-11 11:41:03 Dump the side-table for tag: 1 with group count: 52 into file: file:/tmp/hadoop/bff
1_535059690879355711-1/-local-10003/HashTable-Stage-3/MapJoin-mapfile41--.hashtable
2016-02-11 11:41:03 Uploaded 1 File to: file:/tmp/hadoop/bff9849e-5768-4216-b002-125c786e33a5/hive_2016
le-Stage-3/MapJoin-mapfile41--.hashtable (1648 bytes)
2016-02-11 11:41:03 End of local task; Time Taken: 1.716 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_1455043728844_0066, Tracking URL = http://node1:8088/proxy/application_1455043728844_0066
Kill Command = /opt/hadoop-2.6.0/bin/hadoop job -kill job_1455043728844_0066
Hadoop job information for Stage-3: number of mappers: 2; number of reducers: 0
2016-02-11 23:41:12,845 Stage-3 map = 0%, reduce = 0%
2016-02-11 23:41:24,828 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 1.79 sec
MapReduce Total cumulative CPU time: 1 seconds 790 msec
Ended Job = job_1455043728844_0066
MapReduce Jobs Launched:
Stage-Stage-3: Map: 2 Cumulative CPU: 3.56 sec HDFS Read: 297078 HDFS Write: 725 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 560 msec
OK
Zena Ross
Dorothy Duke
Jocelyn Warner
Bradley McGowan
Hoyt Howard
Wynter Galloway
Time taken: 40.022 seconds, Fetched: 52 row(s)
```

```
hive> SELECT a.fname, b.lname FROM Sales a JOIN Sales_orc b ON a.id = b.id;
Query ID = hadoop_20160212000505_2d585c51-690c-4d5a-b533-1865332feb7f
Total jobs = 1
Execution log at: /tmp/hadoop/hadoop_20160212000505_2d585c51-690c-4d5a-b533-1865332feb7f.log
2016-02-12 12:05:42 Starting to launch local task to process map join; maximum memory = 518979584
2016-02-12 12:05:43 Dump the side-table for tag: 1 with group count: 52 into file: file:/tmp/hadoop/bff984
9_8850161040585641990-1/-local-10003/HashTable-Stage-3/MapJoin-mapfile51--.hashtable
2016-02-12 12:05:43 Uploaded 1 File to: file:/tmp/hadoop/bff9849e-5768-4216-b002-125c786e33a5/hive_2016-02
ble-Stage-3/MapJoin-mapfile51--.hashtable (1648 bytes)
2016-02-12 12:05:43 End of local task; Time Taken: 1.649 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_1455043728844_0067, Tracking URL = http://node1:8088/proxy/application_1455043728844_0067/
Kill Command = /opt/hadoop-2.6.0/bin/hadoop job -kill job_1455043728844_0067
Hadoop job information for Stage-3: number of mappers: 2; number of reducers: 0
2016-02-12 00:05:51,672 Stage-3 map = 0%, reduce = 0%
2016-02-12 00:06:03,715 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 3.55 sec
MapReduce Total cumulative CPU time: 3 seconds 550 msec
Ended Job = job_1455043728844_0067
MapReduce Jobs Launched:
Stage-Stage-3: Map: 2 Cumulative CPU: 3.55 sec HDFS Read: 297090 HDFS Write: 725 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 550 msec
OK
Zena Ross
Dorothy Duke
Jocelyn Warner
Bradley McGowan
Hoyt Howard
Wynter Galloway
Time taken: 30.108 seconds, Fetched: 52 row(s)
```

```

hive> set hive.auto.convert.join=false;
hive> SELECT a.* FROM Sales a JOIN Sales_orc b ON a.id = b.id and a.fname = b.fname;
Query ID = hadoop_20160227013636_46cb8b3b-23fd-453f-b5fb-44399a3347fd
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_1456515901041_0001, Tracking URL = http://node1:8088/proxy/application
Kill Command = /opt/hadoop-2.6.0/bin/hadoop job -kill job_1456515901041_0001
Hadoop job information for Stage-1: number of mappers: 3; number of reducers: 1
2016-02-27 01:36:52,350 Stage-1 map = 0%, reduce = 0%
2016-02-27 01:37:07,556 Stage-1 map = 33%, reduce = 0%, Cumulative CPU 1.44 sec
2016-02-27 01:37:08,706 Stage-1 map = 67%, reduce = 0%, Cumulative CPU 3.35 sec
2016-02-27 01:37:09,782 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 5.23 sec

```

```

hive> desc Sales;
OK
id                int
fname             string
lname            string
address           string
city              string
state             string
ip                string
p_id              string
dop               string
Time taken: 0.172 seconds, Fetched: 9 row(s)

```

```

hive> desc Sales_orc;
OK
id                int
fname             string
lname            string
address           string
city              string
state             string
ip                string
p_id              string
dop               string
Time taken: 0.161 seconds, Fetched: 9 row(s)

```

Chapter 8: Statistics in Hive

sales.id	sales.fname	sales.lname	sales.address	sales.city	sales.state	sales.ip	sales.pid
0	Zena	Ross	41228 West India Ln.	Powell	Tennessee	192.168.56.127	PI_09
1	Elaine	Bishop	15903 North North Adams Blvd.	Hawaiian Gardens	Alaska	192.168.56.105	PI_03
2	Sage	Carroll	6880 Greenland Ct.	Guayanilla	Nevada	192.168.56.40	PI_03
3	Cade	Singleton	64021 South Bulgaria Ave.	Derby	Missouri	192.168.56.171	PI_06
4	Abra	Wright	50155 South Mongolia Ln.	Port Jervis	New Jersey	192.168.56.52	PI_09
5	Stone	Palmer	12191 West Armenia Way	Henderson	Nebraska	192.168.56.85	PI_08
6	Regina	Bryant	29073 Henderson Ct.	Texarkana	Tennessee	192.168.56.5	PI_10
7	Donovan	Aguirre	77718 East Farmington Blvd.	Winston-Salem	New York	192.168.56.114	PI_05
8	Aileen	Mendoza	46855 East Russian Federation Way	Schenectady	Illinois	192.168.56.51	PI_02
9	Mariam	Henson	84567 Gambia Ct.	Owensboro	Hawaii	192.168.56.214	PI_07

```
0: jdbc:hive2://localhost:10000> desc formatted sales;
```

col_name	data_type	comment
# col_name	data_type	comment
id	int	
fname	string	
lname	string	
address	string	
city	string	
state	string	
ip	string	
pid	string	
	NULL	NULL
# Detailed Table Information	NULL	NULL
Database:	default	NULL
Owner:	root	NULL
CreateTime:	Tue Feb 23 16:48:31 IST 2016	NULL
LastAccessTime:	UNKNOWN	NULL
Protect Mode:	None	NULL
Retention:	0	NULL
Location:	hdfs://localhost:8020/user/hive/warehouse/sales	NULL
Table Type:	MANAGED_TABLE	NULL
Table Parameters:	NULL	NULL
	COLUMN_STATS_ACCURATE	true
	numFiles	1
	totalSize	814
	transient_lastDdlTime	1456226754
	NULL	NULL
# Storage Information	NULL	NULL
SerDe Library:	org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe	NULL
InputFormat:	org.apache.hadoop.mapred.TextInputFormat	NULL
OutputFormat:	org.apache.hadoop.hive ql.io.HiveIgnoreKeyTextOutputFormat	NULL
Compressed:	No	NULL
Num Buckets:	-1	NULL
Bucket Columns:	[]	NULL
Sort Columns:	[]	NULL
Storage Desc Params:	NULL	NULL
	field.delim	\t
	serialization.format	\t

col_name	data_type	comment
# col_name	data_type	comment
id	int	
fname	string	
lname	string	
address	string	
city	string	
state	string	
ip	string	
pid	string	
# Detailed Table Information	NULL	NULL
Database:	default	NULL
Owner:	root	NULL
CreateTime:	Tue Feb 23 16:48:31 IST 2016	NULL
LastAccessTime:	UNKNOWN	NULL
Protect Mode:	None	NULL
Retention:	0	NULL
Location:	hdfs://localhost:8020/user/hive/warehouse/sales	NULL
Table Type:	MANAGED_TABLE	NULL
Table Parameters:	NULL	NULL
	COLUMN_STATS_ACCURATE	true
	numFiles	1
	numRows	10
	rawDataSize	793
	totalSize	814
	transient_lastDdlTime	1436228262
# Storage Information	NULL	NULL
serde Library:	org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe	NULL

```
0: jdbc:hive2://localhost:10000> DESCRIBE FORMATTED sales_part PARTITION(pid='PI_09');
```

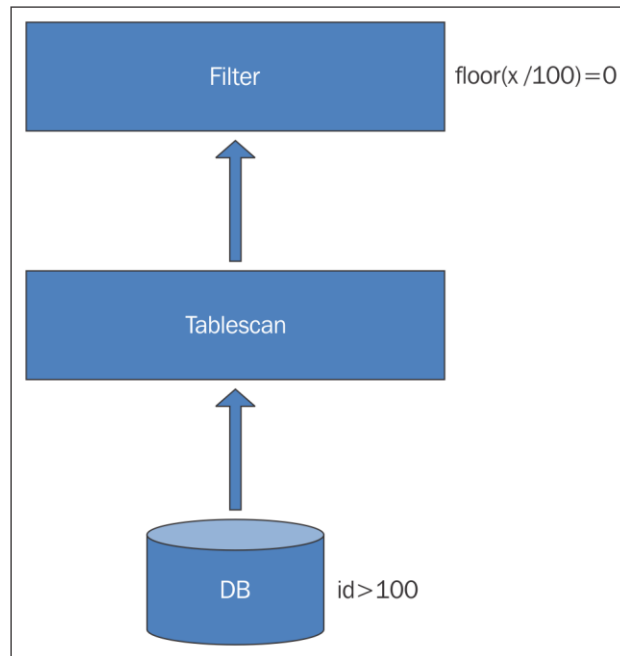
col_name	data_type	comment
# col_name	data_type	comment
id	int	
fname	string	
lname	string	
address	string	
city	string	
state	string	
ip	string	
	NULL	NULL
# Partition Information	NULL	NULL
# col_name	data_type	comment
	NULL	NULL
pid	string	
	NULL	NULL
# Detailed Partition Information	NULL	NULL
Partition Value:	[PI_09]	NULL
Database:	default	NULL
Table:	sales_part	NULL
CreateTime:	Tue Feb 23 17:33:30 IST 2016	NULL
LastAccessTime:	UNKNOWN	NULL
Protect Mode:	None	NULL
Location:	hdfs://localhost:8020/user/hive/warehouse/sales_part/pid=PI_09	NULL
Partition Parameters:	NULL	NULL
	COLUMN_STATS_ACCURATE	true
	numFiles	1
	numRows	2
	rawDataSize	139
	totalSize	141
	transient_lastDdlTime	1436229093
# Storage Information	NULL	NULL
serde Library:	org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe	NULL

Chapter 9: Functions in Hive

```
hive> select * from table_with_array_datatype;
OK
Noida    [201301,201303,201307]
Delhi    [110001,110002,110003]
Time taken: 0.065 seconds, Fetched: 2 row(s)
hive> select city,array_contains(pins,110001) from table_with_array_datatype;
OK
Noida    false
Delhi    true
Time taken: 0.107 seconds, Fetched: 2 row(s)
```

```
hive> SELECT cast('1000' as INT);
OK
1000
Time taken: 0.429 seconds, Fetched: 1 row(s)
hive> SELECT cast('Hi John' as INT);
OK
NULL
Time taken: 0.315 seconds, Fetched: 1 row(s)
hive> SELECT cast('1000.5' as INT);
OK
1000
Time taken: 0.454 seconds, Fetched: 1 row(s)
```

Chapter 10: Hive Tuning



```
hive> set hive.optimize.ppd=false;
hive> explain select a.* from sales_orc a join sales b on a.id = b.id where a.id > 100;
OK
STAGE DEPENDENCIES:
  Stage-4 is a root stage
  Stage-3 depends on stages: Stage-4
  Stage-0 depends on stages: Stage-3
```

```
STAGE PLANS:
  Stage: Stage-4
    Map Reduce Local Work
      Alias -> Map Local Tables:
        a
          Fetch Operator
            limit: -1
          Alias -> Map Local Operator Tree:
            a
              TableScan
                alias: a
                Statistics: Num rows: 52 Data size: 4569 Basic stats: COMPLETE Column stats: NONE
              Filter Operator
                predicate: id is not null (type: boolean)
                Statistics: Num rows: 26 Data size: 2284 Basic stats: COMPLETE Column stats: NONE
              HashTable Sink Operator
                keys:
                  0 id (type: int)
                  1 id (type: int)
```

```

Stage: Stage-3
Map Reduce
Map Operator Tree:
  TableScan
    alias: b
    Statistics: Num rows: 10000 Data size: 252858 Basic stats: COMPLETE Column stats: NONE
  Filter Operator
    predicate: id is not null (type: boolean)
    Statistics: Num rows: 5000 Data size: 126429 Basic stats: COMPLETE Column stats: NONE
  Map Join Operator
    condition map:
      Inner Join 0 to 1
    keys:
      0 id (type: int)
      1 id (type: int)
    outputColumnNames: _col0, _col1, _col2, _col3, _col4, _col5, _col6, _col7, _col8
    Statistics: Num rows: 5500 Data size: 139071 Basic stats: COMPLETE Column stats: NONE
  Filter Operator
    predicate: (_col0 > 100) (type: boolean)
    Statistics: Num rows: 1833 Data size: 46348 Basic stats: COMPLETE Column stats: NONE
  Select Operator
    expressions: _col0 (type: int), _col1 (type: string), _col2 (type: string), _col3 (type:
(type: string), _col7 (type: string), _col8 (type: string)
    outputColumnNames: _col0, _col1, _col2, _col3, _col4, _col5, _col6, _col7, _col8
    Statistics: Num rows: 1833 Data size: 46348 Basic stats: COMPLETE Column stats: NONE
  File Output Operator
    compressed: false
    Statistics: Num rows: 1833 Data size: 46348 Basic stats: COMPLETE Column stats: NONE
    table:
      input format: org.apache.hadoop.mapred.TextInputFormat
      output format: org.apache.hadoop.hive ql.io.HiveIgnoreKeyTextOutputFormat
      serde: org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe

Local Work:
Map Reduce Local Work

```

```

Stage: Stage-0
Fetch Operator
  limit: -1
Processor Tree:
  ListSink

Time taken: 1.856 seconds, Fetched: 65 row(s)

```

```

hive> set hive.optimize.ppd=true;
hive> explain select a.* from sales_orc a join sales b on a.id = b.id where a.id > 100;
OK
STAGE DEPENDENCIES:
  Stage-4 is a root stage
  Stage-3 depends on stages: Stage-4
  Stage-0 depends on stages: Stage-3

```

STAGE PLANS:

```
Stage: Stage-4
Map Reduce Local Work
  Alias -> Map Local Tables:
    a
      Fetch Operator
        limit: -1
  Alias -> Map Local Operator Tree:
    a
      TableScan
        alias: a
        Statistics: Num rows: 52 Data size: 4569 Basic stats: COMPLETE Column stats: NONE
      Filter Operator
        predicate: (id > 100) (type: boolean)
        Statistics: Num rows: 17 Data size: 1493 Basic stats: COMPLETE Column stats: NONE
      HashTable Sink Operator
        keys:
          0 id (type: int)
          1 id (type: int)
```

```
Stage: Stage-3
Map Reduce
  Map Operator Tree:
    TableScan
      alias: b
      Statistics: Num rows: 10000 Data size: 252858 Basic stats: COMPLETE Column stats: NONE
    Filter Operator
      predicate: (id > 100) (type: boolean)
      Statistics: Num rows: 3333 Data size: 84277 Basic stats: COMPLETE Column stats: NONE
    Map Join Operator
      condition map:
        Inner Join 0 to 1
      keys:
        0 id (type: int)
        1 id (type: int)
      outputColumnNames: _col0, _col1, _col2, _col3, _col4, _col5, _col6, _col7, _col8
      Statistics: Num rows: 3666 Data size: 92704 Basic stats: COMPLETE Column stats: NONE
    Select Operator
      expressions: _col0 (type: int), _col1 (type: string), _col2 (type: string), _col3 (type: string), _col7 (type: string), _col8 (type: string)
      outputColumnNames: _col0, _col1, _col2, _col3, _col4, _col5, _col6, _col7, _col8
      Statistics: Num rows: 3666 Data size: 92704 Basic stats: COMPLETE Column stats: NONE
    File Output Operator
      compressed: false
      Statistics: Num rows: 3666 Data size: 92704 Basic stats: COMPLETE Column stats: NONE
      table:
        input format: org.apache.hadoop.mapred.TextInputFormat
        output format: org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat
        serde: org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe

Local Work:
  Map Reduce Local Work
```

```
Stage: Stage-0
Fetch Operator
  limit: -1
Processor Tree:
  ListSink

Time taken: 0.238 seconds, Fetched: 62 row(s)
```

```
hive> set mapreduce.input.fileinputformat.split.maxsize = 10000;
hive> set mapreduce.input.fileinputformat.split.minsize = 10000;
```

```
hive> set mapreduce.input.fileinputformat.split.maxsize = 5000;
hive> set mapreduce.input.fileinputformat.split.minsize = 5000;
```

```
4970 Tasha Mcgowan 66152 South Brazil Ct. Port Huron Iowa
7490 Shay Madden 1728 East Malta Ln. Winnemucca Iowa
4980 Peter Harper 79440 South Bahamas Blvd. Parker Alaska
4990 Cleo Bowen 86121 East Argentina Blvd. Pembroke Pines
0 Zena Ross 41228 West India Ln. Powell Tennessee
Time taken: 0.742 seconds, Fetched: 1000 row(s)
```

```
9987 Malcolm Howard 45083 East Auburn Way East Lansing Nevada 33691
6497 Clio Fitzpatrick 64429 East Clovis Ct. Nampa Hawaii 45515
9997 Brielle Humphrey 32559 West Solomon Islands Ct. Providence
1997 Wanda Hull 22547 Charlotte Amalie Blvd. Hornell Wyoming 86484
Time taken: 0.11 seconds, Fetched: 2000 row(s)
```

```
6502 Amber Thompson 65119 South Guadeloupe St. New York Kentucky
3002 Quemby Becker 52425 West Northern Mariana Islands Way Columbus South Dakota
7302 Cole Robinson 82239 East Liberia Ct. Rutland Oregon 48314 192.168.56.50
4602 Madaline Gill 78128 North Georgia St. Moreno Valley Michigan 28840
Time taken: 0.144 seconds, Fetched: 100 row(s)
```

```
hive> SELECT * FROM Sales_orc TABLESAMPLE(BUCKET 3 OUT OF 1000 ON id);
OK
2002 Aline Gordon 421 South Afghanistan Ln. Hackensack Texas
6002 Ramona Duran 11580 East Kyrgyzstan Ct. Port Arthur Maine
8002 Megan Hood 14500 North Newton Ln. Danbury Kentucky 95235
9002 Clinton Harvey 48466 North Central African Republic Ln. Abilene
1002 Lysandra Monroe 43220 East Belize Way Aberdeen District
7002 Kato Herrera 60413 West Boulder City St. Fayetteville Iowa
4002 Lee Stone 30287 South Bosnia and Herzegovina Way Anchorage
5002 Rhiannon Fields 18530 West Reunion Ct. Elizabeth City Delaware
2 Sage Carroll 6880 Greenland Ct. Guayanilla Nevada 08899
3002 Quemby Becker 52425 West Northern Mariana Islands Way Columbus
Time taken: 0.172 seconds, Fetched: 10 row(s)
```

```
9319 Clementine Mendez 17945 West Nigeria Ave. Orem Illinois 67450
9389 Troy Shields 26994 South Botswana Way Reading New Hampshire 81450
9519 Rae Frederick 74992 West Latvia Ln. Grand Rapids Wyoming 99677
9529 Channing Hansen 9886 West Australia Ln. Beaver Falls Alaska 74070
9789 Hoyt Rice 17080 East Congo Ln. Brown Deer West Virginia 83715
Time taken: 0.161 seconds, Fetched: 916 row(s)
```

```
Total MapReduce CPU Time Spent: 26 seconds 870 msec
OK
968
```

```
Total MapReduce CPU Time Spent: 26 seconds 800 msec
OK
983
```

```
Total MapReduce CPU Time Spent: 26 seconds 870 msec
OK
988
```



```
Total MapReduce CPU Time Spent: 28 seconds 190 msec
OK
1015
```

```
7490 Shay Madden 1728 East Malta Ln. Winnemucca Iowa 02739
4980 Peter Harper 79440 South Bahamas Blvd. Parker Alaska 43854
4990 Cleo Bowen 86121 East Argentina Blvd. Pembroke Pines South
0 Zena Ross 41228 West India Ln. Powell Tennessee 21550
Time taken: 0.203 seconds, Fetched: 1000 row(s)
```

```
hive> SELECT * FROM Sales_orc TABLESAMPLE(10%);
NoViableAltException(289@[])
  at org.apache.hadoop.hive.ql.parse.HiveParser_FromClauseParser.splitSample(HiveParse
  at org.apache.hadoop.hive.ql.parse.HiveParser_FromClauseParser.tableSample(HiveParse
  at org.apache.hadoop.hive.ql.parse.HiveParser_FromClauseParser.tableSource(HiveParse
  at org.apache.hadoop.hive.ql.parse.HiveParser_FromClauseParser.fromSource(HiveParser
  at org.apache.hadoop.hive.ql.parse.HiveParser_FromClauseParser.joinSource(HiveParser
  at org.apache.hadoop.hive.ql.parse.HiveParser_FromClauseParser.fromClause(HiveParser
  at org.apache.hadoop.hive.ql.parse.HiveParser.fromClause(HiveParser.java:44428)
```

```
9959 Plato Vazquez 48213 South Chico Way Decatur Arizona 14411 192.168.56.149
9969 Micah Talley 71206 North Swaziland St. Christiansted Texas 14789
9979 Amaya Madden 89534 West Tokelau Blvd. Moorhead New York
9989 Plato Gamble 79304 North Lebanon Way Atlanta Wyoming 82746 192.168.56.92
9999 Jerome Bentley 18289 South Bulgaria Ave. Bellingham Tennessee
Time taken: 0.14 seconds, Fetched: 10000 row(s)
```

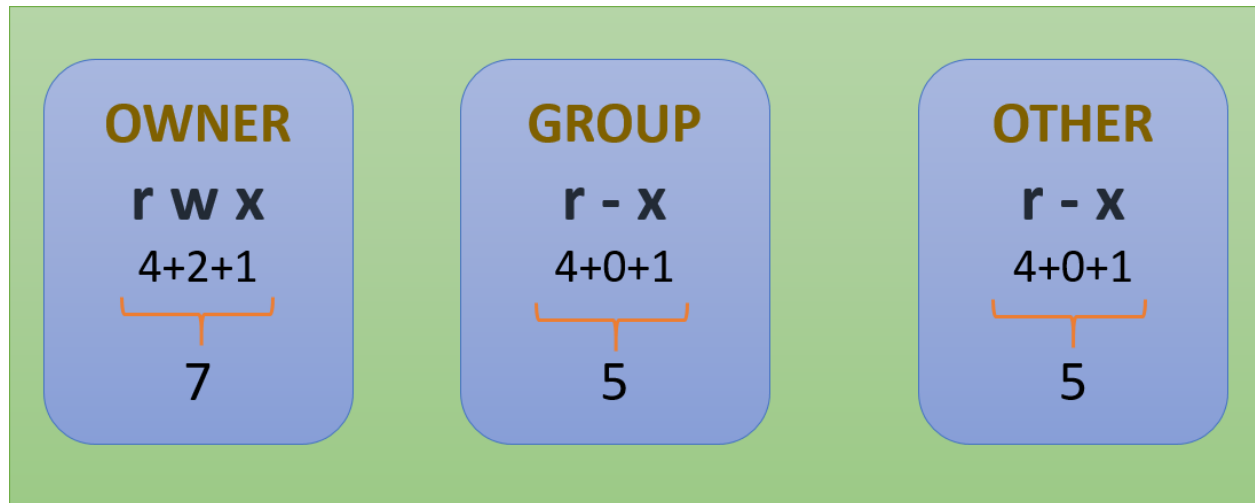
```
hive> SELECT * FROM Sales_orc TABLESAMPLE(0.1M);
NoViableAltException(26@[])
  at org.apache.hadoop.hive.ql.parse.HiveParser_FromClauseParser.splitSample(Hive
  at org.apache.hadoop.hive.ql.parse.HiveParser_FromClauseParser.tableSample(Hive
  at org.apache.hadoop.hive.ql.parse.HiveParser_FromClauseParser.tableSource(Hive
  at org.apache.hadoop.hive.ql.parse.HiveParser_FromClauseParser.fromSource(HiveP
  at org.apache.hadoop.hive.ql.parse.HiveParser_FromClauseParser.joinSource(HiveP
  at org.apache.hadoop.hive.ql.parse.HiveParser_FromClauseParser.fromClause(HiveP
  at org.apache.hadoop.hive.ql.parse.HiveParser.fromClause(HiveParser.java:44428)
```

```
hive> SELECT * FROM Sales_orc TABLESAMPLE(10 ROWS);
OK
5000 Tanner Dotson 29583 Martinsburg Ct. Galesburg California
7500 Gabriel Pace 32093 Montebello St. Half Moon Bay Oregon 49755
8750 Isaac Romero 91922 West Armenia Ct. Opelousas Alaska 91427
10 Silas Hughes 86635 North Ghana St. Beverly Virginia 08642
9960 Burke Myers 45437 Cypress St. Cary West Virginia 25715
5010 Shea Strickland 18243 East Burkina Faso Ln. Palm Springs
20 Grant Thompson 62331 North Great Falls Ln. South El Monte
30 Germaine Stafford 1889 East Russian Federation Ct.
5020 Michelle Gibbs 59445 Cocos (Keeling) Islands Ln. Jackson
40 Fiona Patton 53793 North Norfolk Island Ln. Hornell South Carolina
Time taken: 0.116 seconds, Fetched: 10 row(s)
```

9850	Morgan	West	97008	Clarksville Way	Salt Lake City	Kentucky	
470	Amber	Gallegos		87729 West Fallon St.	Enfield	Texas	60808
5240	Lucian	House	3166	West Belgium Way	Aguadilla	Alabama	47550
480	Ira	Walton	31380	West Macao Ln.	Portsmouth	Utah	41593
7620	Lani	Velez	24832	Mauritania Way	Dickinson	New Jersey	

Time taken: 0.155 seconds, Fetched: 100 row(s)

Chapter 11: Hive Security



```
[centos@ip-172-31-22-157 hadoop-2.6.0]$ hadoop fs -getfacl /stock-data
# file: /stock-data
# owner: centos
# group: supergroup
user::rw-
group::r--
other::r--
```

```
[centos@ip-172-31-22-157 hadoop-2.6.0]$ hadoop fs -ls /stock-data
-rw-r--r--  1 centos supergroup          0 2015-12-30 08:32 /stock-data
```

```
[centos@ip-172-31-22-157 hadoop-2.6.0]$ hadoop fs -setfacl -m user:mike:rw- /stock-data
[centos@ip-172-31-22-157 hadoop-2.6.0]$
[centos@ip-172-31-22-157 hadoop-2.6.0]$ hadoop fs -getfacl /stock-data
# file: /stock-data
# owner: centos
# group: supergroup
user::rw-
user:mike:rw-
group::r--
mask::rw-
other::r--
```

```
[centos@ip-172-31-22-157 hadoop-2.6.0]$ hadoop fs -ls /stock-data
-rw-rw-r--+ 1 centos supergroup          0 2015-12-30 08:32 /stock-data
```

```
[centos@ip-172-31-22-157 ~]$ hadoop fs -setfacl -x user:mike /stock-data
[centos@ip-172-31-22-157 ~]$ hadoop fs -getfacl /stock-data
# file: /stock-data
# owner: centos
# group: supergroup
user::rw-
group::r--
mask::r--
other::r--
```

```
<configuration>

  <property>
    <name>hive.metastore.pre.event.listeners</name>
    <value>org.apache.hadoop.hive.ql.security.authorization.AuthorizationPreEventListener</value>
  </property>

  <property>
    <name>hive.security.metastore.authorization.manager</name>
    <value>org.apache.hadoop.hive.ql.security.authorization.StorageBasedAuthorizationProvider</value>
  </property>

  <property>
    <name>hive.security.metastore.authenticator.manager</name>
    <value>org.apache.hadoop.hive.ql.security.HadoopDefaultMetastoreAuthenticator</value>
  </property>

  <property>
    <name>hive.security.metastore.authorization.auth.reads</name>
    <value>true</value>
  </property>
  .
  .
  .
</configuration>
```

```
[root@localhost hive-1.2.1]# jps
4316 ResourceManager
5487 RunJar
5570 RunJar
3986 DataNode
4160 SecondaryNameNode
6016 Jps
4413 NodeManager
3891 NameNode
```

Chapter 12: Hive Integration with Other Frameworks



Spark Master at spark://node1:7077

URL: spark://node1:7077
REST URL: spark://node1:6066 (cluster mode)
Alive Workers: 0
Cores in use: 0 Total, 0 Used
Memory in use: 0.0 B Total, 0.0 B Used
Applications: 0 Running, 0 Completed
Drivers: 0 Running, 0 Completed
Status: ALIVE

Workers

Worker Id	Address	State
-----------	---------	-------

Running Applications

Application ID	Name	Cores	Memory per Node
----------------	------	-------	-----------------

Completed Applications

Application ID	Name	Cores	Memory per Node
----------------	------	-------	-----------------



Spark Master at spark://node1:7077

URL: spark://node1:7077
REST URL: spark://node1:6066 (cluster mode)
Alive Workers: 1
Cores in use: 1 Total, 0 Used
Memory in use: 1024.0 MB Total, 0.0 B Used
Applications: 0 Running, 0 Completed
Drivers: 0 Running, 0 Completed
Status: ALIVE

Workers

Worker Id	Address	State
worker-20160227103452-192.168.56.101-49384	192.168.56.101:49384	ALIVE

Running Applications

Application ID	Name	Cores	Memory per Node	Submitted Time
----------------	------	-------	-----------------	----------------

Completed Applications

Application ID	Name	Cores	Memory per Node	Submitted Time
----------------	------	-------	-----------------	----------------

Master node1

Region Servers

[Base Stats](#)
[Memory](#)
[Requests](#)
[Storefiles](#)
[Compactions](#)

ServerName	Start time	Requests Per Second	Num. Regions
node1,16201,1450647697829	Sun Feb 28 13:51:37 IST 2016	0	5
Total: 1		0	5

```
hbase(main):031:0> create 'users', 'cf'
0 row(s) in 1.2910 seconds
```

```
=> Hbase::Table - users
hbase(main):032:0> put 'users', 'row1', 'cf:name', 'john'
0 row(s) in 0.0520 seconds

hbase(main):034:0* put 'users', 'row2', 'cf:name', 'mike'
0 row(s) in 0.0260 seconds

hbase(main):035:0> put 'users', 'row3', 'cf:name', 'honey'
0 row(s) in 0.0090 seconds
```

```
hbase(main):002:0> scan "users"
ROW                                COLUMN+CELL
row1                                column=cf:name, timestamp=1456665804693, value=john
row2                                column=cf:name, timestamp=1456665816303, value=mike
row3                                column=cf:name, timestamp=1456665830231, value=honey
3 row(s) in 0.4830 seconds
```

```
hive> CREATE EXTERNAL TABLE hbase_table_users(key string, name string)
> STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
> WITH SERDEPROPERTIES ("hbase.columns.mapping" = "cf:name")
> TBLPROPERTIES("hbase.table.name" = "users", "hbase.mapred.output.outputtable" = "users");
OK
Time taken: 0.247 seconds
```

```
hive> select * from hbase_table_users;
OK
hbase_table_users.key  hbase_table_users.name
row1      john
row2      mike
row3      honey
Time taken: 0.237 seconds, Fetched: 3 row(s)
```

```
hive> CREATE TABLE `Sales_drill` (
>   `id` int,
>   `fname` string,
>   `lname` string,
>   `address` string,
>   `city` string,
>   `state` string,
>   `ip` string,
>   `p_id` string,
>   `dop` string)
> row format delimited fields terminated by '\t' stored as textfile;
OK
Time taken: 0.835 seconds
```

```
hive> LOAD DATA LOCAL INPATH '/opt/data/sales_drill.txt' INTO TABLE Sales_drill;
Loading data to table default.sales_drill
Table default.sales_drill stats: [numFiles=4, numRows=0, totalSize=3044163, rawDataSize=0]
OK
Time taken: 0.522 seconds
```

```
hadoop@node1:/opt/Drill/apache-drill-1.4.0$ cd /opt/Drill/apache-drill-1.4.0
hadoop@node1:/opt/Drill/apache-drill-1.4.0$ bin/drill-embedded
drill-override.conf: 30: unbalanced close brace ')' with no open brace (if you use quotes)
apache drill 1.4.0
"a drill is a terrible thing to waste"
0: jdbc:drill:zk=local> select id, fname, lname from sales_drill where id <= 10;
```

```
0: jdbc:drill:zk=local> select id, fname, lname from sales_drill where id <= 10;
OK
0      Zena    Ross
1      Elaine  Bishop
2      Sage    Carroll
3      Cade    Singleton
4      Abra   Wright
5      Stone  Palmer
6      Regina Bryant
7      Donovan Aguirre
8      Aileen Mendoza
9      Mariam Henson
10     Silas   Hughes
0      Zena    Ross
1      Elaine  Bishop
2      Sage    Carroll
3      Cade    Singleton
4      Abra   Wright
5      Stone  Palmer
6      Regina Bryant
7      Donovan Aguirre
8      Aileen Mendoza
9      Mariam Henson
10     Silas   Hughes
Time taken: 1.264 seconds, Fetched: 22 row(s)
```