

Slide # 26

```
hive> create table salesjan2009_clustered(
  > productid string, salesamount int, paymenttype string, customername string, city string, region string, country string)
  > clustered by (country) into 10 buckets
  > stored as ORCFILE;
OK
Time taken: 182.887 seconds
```

Slide # 27

```
hive> insert into salesjan2009_clustered
  > select * from salesjan2009_ext;
Query ID = root_20200524044648_60a7c2c8-282d-466d-b31a-2e49a9291452
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.
Status: Running (Executing on YARN cluster with App id application_1590192907351_0032)

-----
VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 .....  SUCCEEDED      1          1          0          0          0          0
-----
VERTICES: 01/01 [=====>>] 100%  ELAPSED TIME: 11.99 s
-----
Loading data to table default.salesjan2009_clustered
Table default.salesjan2009_clustered stats: [numFiles=1, numRows=998, totalSize=16364, rawDataSize=557882]
OK
Time taken: 213.913 seconds
```

Slide # 28

```
[root@sandbox-hdp ~]# hive
log4j:WARN No such property [maxFileSize] in org.apache.log4j.DailyRollingFileAppender.

Logging initialized using configuration in file:/etc/hive/2.6.5.0-292/0/hive-log4j.properties
hive> describe formatted salesjan2009_clustered;
OK
# col_name                data_type                comment

productid                 string
salesamount               int
paymenttype               string
customername              string
city                      string
region                    string
country                   string

# Detailed Table Information
Database:                  default
Owner:                     root
CreateTime:                Sun May 24 04:40:49 UTC 2020
LastAccessTime:            UNKNOWN
Protect Mode:              None
Retention:                 0
Location:                  hdfs://sandbox-hdp.hortonworks.com:8020/apps/hive/warehouse/salesjan2009_clustered
Table Type:                MANAGED_TABLE
Table Parameters:
    COLUMN_STATS_ACCURATE  {"BASIC_STATS\":"true\"}
    numFiles                1
    numRows                998
    rawDataSize             557882
    totalSize               16364
    transient_lastDdlTime   1590295639

# Storage Information
SerDe Library:             org.apache.hadoop.hive.ql.io.orc.OrcSerde
InputFormat:               org.apache.hadoop.hive.ql.io.orc.OrcInputFormat
OutputFormat:              org.apache.hadoop.hive.ql.io.orc.OrcOutputFormat
Compressed:                No
Num Buckets:               10
Bucket Columns:            [country]
Sort Columns:              []
Storage Desc Params:
    serialization.format    1
Time taken: 2.918 seconds, Fetched: 37 row(s)
```

Slide # 29

```
[root@sandbox-hdp ~]# hdfs dfs -ls /apps/hive/warehouse/salesjan2009_clustered
Found 1 items
-rwxrwxrwx  1 root hadoop      16364 2020-05-24 04:47 /apps/hive/warehouse/salesjan2009_clustered/000000_0
```

```
[root@sandbox-hdp ~]# hdfs dfs -cat /apps/hive/warehouse/salesjan2009_clustered/000000_0
ORC
PJ[Rbf`xT$P[R\b( e+H< `9
E?P?
"
AmexVisa[P=
"
AverenaaPI
"
AardalZurichP?
"
AKZurichXP]
/
"
United StatesP;MQ;NC1D>$ $ HOB)9yx{#0;m0Tv,gC?z#P1S4S
j9^a\cH3ilw=KG:OlLrP*b\<=#'KEt5/au*CoV\T2K:BJ[W*|oZ
'P-&g
R&
```

Slide # 31

```
hive> create external table SalesJan2009_ext1(
> productid string, salesamount int, paymenttype string, customername string, city string, region string, country string)
> row format delimited
> fields terminated by ','
> stored as textfile
> location '/user/anup/Hive/warehouse';
OK
Time taken: 188.61 seconds
hive> load data inpath '/user/anup/Hive/warehouse/SalesJan2009.csv'
> into table SalesJan2009_ext1;
Loading data to table default.salesjan2009_ext1
Table default.salesjan2009_ext1 stats: [numFiles=1, totalSize=65835]
OK
Time taken: 183.996 seconds
```

```
hive> select country, count(*) as count from SalesJan2009_ext1
> group by (country)
> order by count;
Query ID = root_20200524052744_32678768-d5b9-4701-be48-aff313bbcf54
Total jobs = 1
Launching Job 1 out of 1
```

Slide # 32

| VERTICES | STATUS | TOTAL | COMPLETED | RUNNING | PENDING | FAILED | KILLED |
|--|-----------|-------|-----------|---------|---------|--------|--------|
| Map 1 | SUCCEEDED | 1 | 1 | 0 | 0 | 0 | 0 |
| Reducer 2 | SUCCEEDED | 1 | 1 | 0 | 0 | 0 | 0 |
| Reducer 3 | SUCCEEDED | 1 | 1 | 0 | 0 | 0 | 0 |
| VERTICES: 03/03 [=====>>] 100% ELAPSED TIME: 14.28 s | | | | | | | |
| OK | | | | | | | |
| Argentina | 1 | | | | | | |
| Bahrain | 1 | | | | | | |
| Bermuda | 1 | | | | | | |
| Bulgaria | 1 | | | | | | |
| Cayman Isls | 1 | | | | | | |
| China | 1 | | | | | | |
| Costa Rica | 1 | | | | | | |
| Dominican Republic | 1 | | | | | | |

United Kingdom 100
United States 463
Time taken: 33.511 seconds, Fetched: 56 row(s)