

ASSIGNMENT 11.3

Explain in brief static and dynamic partitioning with an example

STATIC AND DYNAMIC PARTITIONING

Partition is horizontally dividing the data into number of slice in a equal and manageable manner. Every partition is stored as directory within data warehouse table.

Hive partition is supported for Multiple columns in a table. In Hive we can apply Hive Partition concept on Managed tables and External tables. If we not created dynamic partition for hive, Hive also creates an automatic partition scheme when the table is created

Example Hive Partition

```
create table cityreport(cityid string, creport string, ctover string)
partitioned by (city string)
row format delimited
fields terminated by '|'
stored as textfile;
```

Static Partitioning: In Static Partition, we know the partition column before itself. Now when we load data there it makes the difference.

**LOAD DATA LOCAL INPATH [path_name] OVERWRITE INTO
TABLE [table_name] PARTITION(partition_column='value'...). Here
we have to give the partition column value explicitly whenever we
want to create new partition as shown below:**

CODE.

```
File Edit View Search Terminal Help
e:null, serializationLib:org.apache.hadoop.hive.serde2.LazyLazySimpleSerDe, parameters:{serialization.format= , field.delim=
Time taken: 18.097 seconds
root@ubuntu:/home/GAURAV/HIVE# nano UserLog.txt
root@ubuntu:/home/GAURAV/HIVE# hive -e "LOAD DATA LOCAL INPATH 'UserLog.txt' OVERWRITE INTO TABLE T_USER_LOG PARTITION(dt='2016-04-29',country='UK');"
Hive history file=/tmp/root/hive/job_log_root_201604252107_817941186.txt
Copying data from file:/home/GAURAV/HIVE/UserLog.txt
Copying file: file:/home/GAURAV/HIVE/UserLog.txt
Loading data to table default.t_user_log partition (dt=2016-04-29, country=UK)
OK
Time taken: 25.392 seconds
root@ubuntu:/home/GAURAV/HIVE# hive -e "LOAD DATA LOCAL INPATH 'UserLog.txt' OVERWRITE INTO TABLE T_USER_LOG PARTITION(dt='2016-04-29',country='US');"
Hive history file=/tmp/root/hive/job_log_root_201604252108_1007926206.txt
Copying data from file:/home/GAURAV/HIVE/UserLog.txt
Copying file: file:/home/GAURAV/HIVE/UserLog.txt
Loading data to table default.t_user_log partition (dt=2016-04-29, country=US)
OK
Time taken: 19.501 seconds
root@ubuntu:/home/GAURAV/HIVE# hive -e "LOAD DATA LOCAL INPATH 'UserLog.txt' OVERWRITE INTO TABLE T_USER_LOG PARTITION(dt='2016-04-28',country='CHINA');"
Hive history file=/tmp/root/hive/job_log_root_201604252109_1005707855.txt
Copying data from file:/home/GAURAV/HIVE/UserLog.txt
Copying file: file:/home/GAURAV/HIVE/UserLog.txt
Loading data to table default.t_user_log partition (dt=2016-04-28, country=CHINA)
OK
Time taken: 18.901 seconds
root@ubuntu:/home/GAURAV/HIVE# hive -e "LOAD DATA LOCAL INPATH 'UserLog.txt' OVERWRITE INTO TABLE T_USER_LOG PARTITION(dt='2016-04-28',country='JAPAN');"
Hive history file=/tmp/root/hive/job_log_root_201604252109_1003935071.txt
Copying data from file:/home/GAURAV/HIVE/UserLog.txt
Copying file: file:/home/GAURAV/HIVE/UserLog.txt
Loading data to table default.t_user_log partition (dt=2016-04-28, country=JAPAN)
OK
Time taken: 18.635 seconds
root@ubuntu:/home/GAURAV/HIVE# hive -e "LOAD DATA LOCAL INPATH 'UserLog.txt' OVERWRITE INTO TABLE T_USER_LOG PARTITION(dt='2016-04-27',country='INDIA');"
Hive history file=/tmp/root/hive/job_log_root_201604252110_1000240724.txt
Copying data from file:/home/GAURAV/HIVE/UserLog.txt
Copying file: file:/home/GAURAV/HIVE/UserLog.txt
Loading data to table default.t_user_log partition (dt=2016-04-27, country=INDIA)
OK
Time taken: 18.902 seconds
root@ubuntu:/home/GAURAV/HIVE#
```

OUTPUT:

Contents of directory /user/hive/warehouse/t_user_log

Goto :

[Go to parent directory](#)

Name	Type	Size	Replication	Block Size	Modification Time	Permission	Owner	Group
dt=2016-04-27	dir				2016-04-25 21:10	rwxr-xr-x	root	supergroup
dt=2016-04-28	dir				2016-04-25 21:10	rwxr-xr-x	root	supergroup
dt=2016-04-29	dir				2016-04-25 21:09	rwxr-xr-x	root	supergroup

Contents of directory `/user/hive/warehouse/t_user_log/dt=2016-04-28`

Goto :

[Go to parent directory](#)

Name	Type	Size	Replication	Block Size	Modification Time	Permission	Owner	Group
country=CHINA	dir				2016-04-25 21:09	rwxr-xr-x	root	supergroup
country=JAPAN	dir				2016-04-25 21:10	rwxr-xr-x	root	supergroup

Contents of directory `/user/hive/warehouse/t_user_log/dt=2016-04-28/country=CHINA`

Goto :

[Go to parent directory](#)

Name	Type	Size	Replication	Block Size	Modification Time	Permission	Owner
UserLog.txt	file	0.11 KB	1	64 MB	2016-04-25 21:09	rw-r--r--	root

File: `/user/hive/warehouse/t_user_log/dt=2016-04-28/country=CHINA/UserLog.txt`

Goto :

[Go back to dir listing](#)

[Advanced view/download options](#)

```

1001  John  Google
1001  John  Facebook
1002  Eric  Instagram
1002  Eric  Google
1003  Annie Facebook
1004  Sara  Instagram
  
```

All partitions in hive is there as directories. Loading in hive is instantaneous process and it won't trigger a Map/Reduce job. That's why our file is stored as UserLog.txt instead of 00000_o file. Please follow the article as I will show in dynamic partition where we will LOAD table using another table where Map/reduce job is triggered.

DYNAMIC PARTITIONING:

Let us see now the CODE of Dynamic Partitioning. We will create new table T_USER_LOG_DYN for dynamic partition and also as we told earlier that we will load this table using a new table, let's create another table T_USER_LOG_SRC.

Data for the source table

1001	John	Google	2016-04-27	US
1001	John	Facebook	2016-04-27	US
1002	Eric	Instagram	2016-04-28	US
1002	Eric	Google	2016-04-28	UK
1003	Annie	Facebook	2016-04-28	UK
1004	Sara	Instagram	2016-04-28	US
1005	Wei	Google	2016-04-29	CHINA
1006	Ming	Facebook	2016-04-29	CHINA
1007	Li	Instagram	2016-04-29	CHINA
1008	Sota	Google	2016-04-29	JAPAN
1009	Yuto	Facebook	2016-04-27	JAPAN
1010	Ryota	Instagram	2016-04-28	JAPAN
1011	Gaurav	Google	2016-04-29	INDIA
1012	Anu	Facebook	2016-04-27	INDIA
1013	Maya	Instagram	2016-04-28	INDIA
1014	Rohit	Google	2016-04-28	INDIA
1015	Michel	Facebook	2016-04-28	AUSTRALIA
1016	Ricky	Instagram	2016-04-28	AUSTRALIA

hive script for table DDL

```
DROP TABLE IF EXISTS T_USER_LOG_DYN;  
  
CREATE TABLE T_USER_LOG_DYN (USER_ID INT  
                             ,NAME STRING  
                             ,SITE STRING  
                             )  
PARTITIONED BY (DT STRING,COUNTRY STRING)  
ROW FORMAT DELIMITED  
FIELDS TERMINATED BY '\\t'  
STORED AS TEXTFILE  
;  
  
DROP TABLE IF EXISTS T_USER_LOG_SRC;  
  
CREATE TABLE T_USER_LOG_SRC (USER_ID INT  
                              ,NAME STRING  
                              ,SITE STRING  
                              ,DT STRING  
                              ,COUNTRY STRING  
                              )  
ROW FORMAT DELIMITED  
FIELDS TERMINATED BY '\\t'  
STORED AS TEXTFILE  
;  
  
LOAD DATA LOCAL INPATH 'UserLogSrc.txt' OVERWRITE INTO TABLE T_USER_LOG_SRC;
```

setting hive properties

SET hive.exec.dynamic.partition= true;

SET hive.exec.dynamic.partition.mode= nonstrict

```
root@ubuntu:/home/GAURAV/NEVER# hive -e 'INSERT OVERWRITE TABLE T_USER_LOG_DYN PARTITION(DT,COUNTRY) SELECT USER_ID,NAME,SITE,DT,COUNTRY FROM T_USER_LOG_SRC';  
Hive history file=/tmp/root/hive/job_log_root_201804261320_95404991.txt  
FAILED: Error in semantic analysis: dynamic partition strict mode requires at least one static partition column. To turn this off set hive.exec.dynamic.partition.mode=nonstrict  
root@ubuntu:/home/GAURAV/NEVER# hive -e 'INSERT OVERWRITE TABLE T_USER_LOG_DYN PARTITION(DT='2018-04-26',COUNTRY) SELECT USER_ID,NAME,SITE,DT,COUNTRY FROM T_USER_LOG_SRC';  
Hive history file=/tmp/root/hive/job_log_root_201804261322_95818172.txt  
FAILED: Error in semantic analysis: dynamic partition is disabled. Either enable it by setting hive.exec.dynamic.partition=true or specify partition column values  
root@ubuntu:/home/GAURAV/NEVER#  
root@ubuntu:/home/GAURAV/NEVER#
```


While loading the table, the partition will be created dynamically on all partition columns if `hive.exec.dynamic.partition.mode= nonstrict` is set.

If it is strict [which is by default], it will need at least one partition column to be defined in CODE.

CODE:

loading the table `T_USER_LOG_DYN` using the data from `T_USER_LOG_SRC` and creating dynamic partitions.

`LoadDynPartition.hql` loads the table dynamically as shown below:

```
File Edit View Search Terminal Help
root@ubuntu:/home/GAURAV/HIVE# more LoadDynPartition.hql
SET hive.exec.dynamic.partition = true;
SET hive.exec.dynamic.partition.mode = nonstrict;

INSERT OVERWRITE TABLE T_USER_LOG_DYN
PARTITION(DT,COUNTRY)
SELECT USER_ID,NAME,SITE,DT,COUNTRY FROM T_USER_LOG_SRC
;

root@ubuntu:/home/GAURAV/HIVE# hive -f LoadDynPartition.hql
Hive history file=/tmp/root/hive_job_log_root_201604260942_1630764239.txt
Total MapReduce jobs = 2
Launching Job 1 out of 2
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_201604260842_0003, Tracking URL = http://localhost:50030/jobdetails.jsp?jobid=job_201604260842_0003
Kill Command = /usr/lib/hadoop/bin/hadoop job -Dmapred.job.tracker=localhost:8021 -kill job_201604260842_0003
2016-04-26 09:43:13,163 Stage-1 map = 0%, reduce = 0%
2016-04-26 09:43:20,288 Stage-1 map = 100%, reduce = 0%
2016-04-26 09:43:26,429 Stage-1 map = 100%, reduce = 100%
Ended Job = job_201604260842_0003
Ended Job = 1561898483, job is filtered out (removed at runtime).
Moving data to: hdfs://localhost:8020/tmp/hive-root/hive_2016-04-26_09-42-46_540_3033688718707476412/-ext-10000
Loading data to table default.t user log dyn partition (dt=null, country=null)
Loading partition {dt=2016-04-27, country=INDIA}
Loading partition {dt=2016-04-27, country=JAPAN}
Loading partition {dt=2016-04-27, country=US}
Loading partition {dt=2016-04-28, country=AUSTRALIA}
Loading partition {dt=2016-04-28, country=INDIA}
Loading partition {dt=2016-04-28, country=JAPAN}
Loading partition {dt=2016-04-28, country=UK}
Loading partition {dt=2016-04-28, country=US}
Loading partition {dt=2016-04-29, country=CHINA}
Loading partition {dt=2016-04-29, country=INDIA}
Loading partition {dt=2016-04-29, country=JAPAN}
Partition default.t user log dyn(dt=2016-04-27, country=INDIA) stats: [num_files: 1, num_rows: 0, total_size: 18]
Partition default.t user log dyn(dt=2016-04-27, country=JAPAN) stats: [num_files: 1, num_rows: 0, total_size: 19]
Partition default.t user log dyn(dt=2016-04-27, country=US) stats: [num_files: 1, num_rows: 0, total_size: 36]
Partition default.t user log dyn(dt=2016-04-28, country=AUSTRALIA) stats: [num_files: 1, num_rows: 0, total_size: 42]
```

OUTPUT:

Contents of directory **/user/hive/warehouse/t_user_log_dyn**

Goto :

Go to parent directory

Name	Type	Size	Replication	Block Size	Modification Time	Permission	Owner	Group
dt=2016-04-27	dir				2016-04-26 09:43	rwxr-xr-x	root	supergroup
dt=2016-04-26	dir				2016-04-26 09:43	rwxr-xr-x	root	supergroup
dt=2016-04-25	dir				2016-04-26 09:43	rwxr-xr-x	root	supergroup

Contents of directory **/user/hive/warehouse/t_user_log_dyn/dt=2016-04-27**

Goto :

Go to parent directory

Name	Type	Size	Replication	Block Size	Modification Time	Permission	Owner	Group
country=INDIA	dir				2016-04-26 09:43	rwxr-xr-x	root	supergroup
country=JAPAN	dir				2016-04-26 09:43	rwxr-xr-x	root	supergroup
country=US	dir				2016-04-26 09:43	rwxr-xr-x	root	supergroup

Contents of directory **/user/hive/warehouse/t_user_log_dyn/dt=2016-04-27/country=JAPAN**

Goto :

Go to parent directory

Name	Type	Size	Replication	Block Size	Modification Time	Permission	Owner	Group
000000_0	file	0.02 KB	1	64 MB	2016-04-26 09:43	rw-r--r--	root	supergroup