Table Properties 2 – Purge

- **What is Purge Table Property?**

Purge is a table property. If Purge is set to true, then in case of dropping/Truncating an internal table, the data won't go to trash, it will be permanently gone. And in case of Insert overwrite table, the previous data won't go to trash, it will be permanently deleted.

This property works only for internal table.

In normal scenario, when Purge is not set to True, then dropping an internal table, the data first goes to Trash directory, And if that was mistakenly done, **Hadoop admins can recover the data from Trash**.

- **Drop/Truncate Table –**

1. In case of Internal table, Drop table command removes the metadata and data of a table,
2. In case of external table, Drop table command removes metadata and data of a table, but from only Hive table, data will be still present in HDFS location.

   (As external table's data is completely managed by HDFS itself, so only the linkage between Hive and HDFS is removed)
3. The difference between Truncate and Drop is, Truncate removes only the data from a table, metadata or structure of a table will not be removed,

- **Drop Table Example :**

```
hive> show tables;
OK
dept
dept_det
emp
Time taken: 2.585 seconds, Fetched: 3 row(s)
hive> drop table emp;
OK
Time taken: 5.416 seconds
hive> show tables;
OK
dept
dept_det
Time taken: 0.071 seconds, Fetched: 2 row(s)
hive> Select * from emp;
FAILED: SemanticException [Error 10001]: Line 1:14 Table not found 'emp'
hive>
```

- **Truncate Table Example :**

```
hive> show tables;
OK
dept
dept_det
Time taken: 0.086 seconds, Fetched: 2 row(s)
hive> truncate table dept;
OK
Time taken: 1.285 seconds
hive> show tables;
OK
dept
dept_det
Time taken: 0.063 seconds, Fetched: 2 row(s)
hive> Select * from dept;
OK
Time taken: 2.308 seconds
hive> desc dept;
OK
department_id          int
department_name        string
location               string
Time taken: 0.455 seconds, Fetched: 3 row(s)
hive>
```

- **Creating a table with Purge Property:**

```
create table if not exists EMP_DET (EMP_ID int ,EMP_NAME string
,Department_ID int,Manager_ID int,Designation string,Location
string,Years_of_Experience double) row format delimited fields
terminated by ',' lines terminated by '\n' stored as textfile
tblproperties ("auto.purge"="true");
```

```
hive> create table if not exists EMP_DET (EMP_ID int ,EMP_NAME string ,Department_ID int,Manager_ID int,Designation string,Location st
ring,Years_of_Experience double) row format delimited fields terminated by ',' lines terminated by '\n' stored as textfile tblproperti
es ("auto.purge"="true");
OK
Time taken: 22.93 seconds
hive>
```

1. in case when Purge is set to True, then if we drop a table, the data will not go to Trash directory. It will be completely gone along with the metadata.
2. Same logic holds true for the Truncate command as well. Data will be completely gone, when Purge property is set to True. But metadata will be present as it is truncate command.

Insert Overwrite Query:

- The insert overwrite table query will overwrite any existing table or partition in Hive. If the table property set as 'auto.purge'='true', the previous data of the table is not moved to trash when insert overwrite query is run against the table.
- If we not set the 'auto.purge'='true' in the table properties and run the insert overwrite query frequently, it occupy the memory for the previous data in the trash and create the insufficient memory issue after some time.