ASSIGNMENT26.5

Q.How many kinds of tables are present in hive and explain the difference between them with a demo

|  |
| --- |
| **There are 2 types of tables in Hive –**   * **Internal** * **External** |
|  |

|  |
| --- |
| **Internal table** |
|  |

|  |
| --- |
| Internal table are like normal database table where data can be stored and queried on. On dropping these tables the data stored in them also gets deleted and data is lost forever. So one should be careful while using internal tables as one drop command can destroy the whole data.   * Open new terminal and fire up hive by just typing **hive**. |
|  |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| Create table on weather data. |
|  |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| * ROW FORMAT should have delimiters used to terminate the fields and lines like in the above example the fields are terminated with comma (“,”). |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| * Load the Data in Table from local path |
| * Input wether.txt file |
|  |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| * Data can be loaded in 2 ways in Hive either from local file or from HDFS to Hive. To load the data from local to Hive use the following command |
|  |

|  |
| --- |
| **External Tables** |
| * It stores data on HDFS so even if we will drop our table data will still there on HDFS. File and table link is there but that is read only. * As an example if you create an external table called “**amandeep\_test**” in HIVE using HIVE-QL and link the table to file “**flat\_file.txt**”, then deleting “**amandeep\_test**” from HIVE will not delete “**flat\_file.txt**” from HDFS. * External table files are accessible to anyone who has access to HDFS file structure and therefore security needs to be managed at the HDFS file/folder level. * Meta data is maintained on master node and deleting an external table from HIVE, only deletes the metadata not the data/file. |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| **Creating external table** |
|  |

|  |
| --- |
| * Open new terminal and fire up hive by just typing hive.      * Create table on weather data. |
| |  | | --- | | ROW FORMAT should have delimiters used to terminate the fields and lines like in the above example the fields are terminated with comma (“,”). | |  | |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
|  |

|  |
| --- |
|  |
| Before loading data we need to copy our file from local to HDFS so using put we can do that. |

|  |
| --- |
| * Now Load the data from HDFS to Hive using the following command:   Input weather1.txt file :    Now showing the data in our table |
|  |
|  |

The above is exactly same as that we have in our txt file.

**So we can conclude as :**

* When we drop an internal table,it drops the data and it also drops the metadata
* .
* When we drop an external table, it only drops the metadata
* .
* On dropping the external table, the data does not get deleted from HDFS.
* Thus it is evident that the external table just a pointer on HDFS data.

**Use EXTERNAL tables when:**

* The data is also used outside of Hive. For example, the files are read and processed by an existing program that doesn’t lock the files.
* Data needs to remain in the underlying location even after a DROP TABLE. This can apply if we are pointing multiple schemas(tables or views) at a single data set or if we are iterating through various possible schemas.
* Hive should not own data and control setting,dirs,etc.,we have another program or process that will do those things.
* We are not creating table based on existing table(As select).

**Use INTERNAL tables when:**

* The data is temporary
* We want Hive to completely manage the lifecycle of the table and data.