Table Properties 1 - Skipping header and Footer

Table properties can be **Active** or **Passive**.

Passive Table Properties do nothing, they just tag a table, like we can give creator name, date of creation while creating a table. They are simply used for tagging purpose.
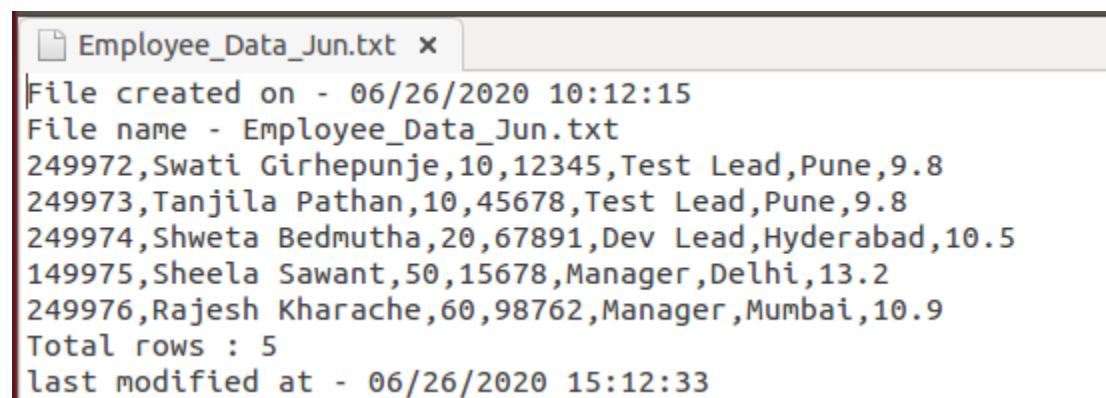
But Active Table Properties are those which helps in data processing or which can be responsible for selecting a specific data from the table.

Skip header and footer records in Hive

In many of the source files we have observed some lines present on the top (header) and some lines at the bottom (footer), which are not actual data but contains so information about the data or the file.

we can skip these header and footer lines while creating table using table properties.

Consider below file as a source file

```
Employee_Data_Jun.txt  ✕
File created on - 06/26/2020 10:12:15
File name - Employee_Data_Jun.txt
249972,Swati Girhepunje,10,12345,Test Lead,Pune,9.8
249973,Tanjila Pathan,10,45678,Test Lead,Pune,9.8
249974,Shweta Bedmutha,20,67891,Dev Lead,Hyderabad,10.5
149975,Sheela Sawant,50,15678,Manager,Delhi,13.2
249976,Rajesh Kharache,60,98762,Manager,Mumbai,10.9
Total rows : 5
last modified at - 06/26/2020 15:12:33
```

- We created table mentioning property *"skip.header.line.count"="2"* and *"skip.footer.line.count"="2"*

```
create table if not exists Tb_Employee (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 ("skip.header.line.count"="2",
"skip.footer.line.count"="2");
```

```
hive> create table if not exists Tb_Employee (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 ("skip.header.line.count"="2", "skip.footer.line.count"="2");
OK
Time taken: 42.92 seconds
hive>
```

- We loaded the above file into table created using *load data inpath* command

```
hive> load data inpath '/user/cloudera/Employee_Data_Jun.txt' into table TB_EMPLOYEE;
Loading data to table default.tb_employee
chgrp: changing ownership of 'hdfs://quickstart.cloudera:8020/user/hive/warehouse/tb_employee/Employee_Data_Jun.txt': Permission denie
d. user=root is not the owner of inode=Employee_Data_Jun.txt
chmod: changing permissions of 'hdfs://quickstart.cloudera:8020/user/hive/warehouse/tb_employee/Employee_Data_Jun.txt': Permission den
ied. user=root is not the owner of inode=Employee_Data_Jun.txt
Table default.tb_employee stats: [numFiles=1, totalSize=385]
OK
Time taken: 3.665 seconds
```

- Select command on table

```
hive> Select * from tb_employee;
OK
tb_employee.id  tb_employee.emp_name    tb_employee.department_id       tb_employee.manager_id  tb_employee.designation tb_employee.lo
cation  tb_employee.years_of_experience
249972  Swati Girhepunje        10      12345   Test Lead       Pune    9.8
249973  Tanjila Pathan  10      45678   Test Lead       Pune    9.8
249974  Shweta Bedmutha 20      67891   Dev Lead        Hyderabad       10.5
149975  Sheela Sawant   50      15678   Manager Delhi   13.2
249976  Rajesh Kharache 60      98762   Manager Mumbai  10.9
Time taken: 0.648 seconds, Fetched: 5 row(s)
hive>
```

- We can see all the table properties on a table using below command.

```
show tblproperties table_name;
```

```
hive> show tblproperties TB_EMPLOYEE;
OK
prpt_name       prpt_value
COLUMN_STATS_ACCURATE   true
numFiles        1
skip.footer.line.count  2
skip.header.line.count  2
totalSize       385
transient_lastDdlTime   1593164153
Time taken: 4.835 seconds, Fetched: 6 row(s)
hive>
```

What Will Happen if we load above data without using skip footer ?

We can observe in below screenshot that, last 2 lines are extra with null values in the Table. As the data in source file, in the last 2 lines(footer) is not matching with the data type we provided (while creating table) for the first fields (int) and also there is no ',' (filed separator), so all other columns are also filled with null for these last 2 lines.

```
hive> create table if not exists Tb_Employee_skip_header (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 ("skip.header.line.count"="2");
OK
Time taken: 0.065 seconds
hive> load data inpath '/user/cloudera/Employee_Data_Jun.txt' into table TB_EMPLOYEE_SKIP_HEADER;
Loading data to table default.tb_employee_skip_header
chgrp: changing ownership of 'hdfs://quickstart.cloudera:8020/user/hive/warehouse/tb_employee_skip_header/Employee_Data_Jun.txt': Perm
ission denied. user=root is not the owner of inode=Employee_Data_Jun.txt
chmod: changing permissions of 'hdfs://quickstart.cloudera:8020/user/hive/warehouse/tb_employee_skip_header/Employee_Data_Jun.txt': Pe
rmission denied. user=root is not the owner of inode=Employee_Data_Jun.txt
Table default.tb_employee_skip_header stats: [numFiles=1, totalSize=385]
OK
Time taken: 1.24 seconds
hive> Select * from TB_EMPLOYEE_SKIP_HEADER;
OK
tb_employee_skip_header.id      tb_employee_skip_header.emp_name        tb_employee_skip_header.department_id   tb_employee_skip_heade
r.manager_id    tb_employee_skip_header.designation     tb_employee_skip_header.location        tb_employee_skip_header.years_of_exper
ience
249972  Swati Girhepunje        10      12345   Test Lead       Pune    9.8
249973  Tanjila Pathan  10      45678   Test Lead       Pune    9.8
249974  Shweta Bedmutha 20      67891   Dev Lead        Hyderabad       10.5
149975  Sheela Sawant   50      15678   Manager Delhi   13.2
249976  Rajesh Kharache 60      98762   Manager Mumbai  10.9
NULL    NULL    NULL    NULL    NULL    NULL    NULL
NULL    NULL    NULL    NULL    NULL    NULL    NULL
Time taken: 0.18 seconds, Fetched: 7 row(s)
hive>
```