

Querying Partitioned Table

- Now, when we query the partition table based on the partition values, then only the records in that particular partition are scanned.
 - For example, if we query the table and count the records where the department is “radiotherapy” -
 - **select count(*) from cases_part where department='radiotherapy';**
 - In this case, only the records from the radiotherapy directory were scanned.
- Since we have another partition within this directory, we can include that in the WHERE clause as well to further reduce the number of records to be scanned.
 - **select count(*) from cases_part where department='radiotherapy' and severity_of_illness='Extreme';**
- To prevent from accidentally running a query that scans all the records in the entire table without mentioning the partitions, we can run Hive in “strict” mode -
 - **set hive.mapred.mode=strict;**
 - This will now throw an error -
 - **select count(*) from cases_part;**
 - Hive’s strict mode prevents querying on partitioned tables without using a WHERE clause that filters on the partitions. The above query will run fine with WHERE clause -
 - **select count(*) from cases_part where department='radiotherapy';**

- We can turn off the strict mode as follows-
 - **set hive.mapred.mode=nonstrict;**
- Create a non-partitioned managed table with the same cases data and see the difference in the querying.
 - **create table cases_non_part(Case_Id int, Hospital_Code int, Patient_Id int, Ward_Type char(1), Ward_Facility_Code char(1), Bed_Grade int, Type_of_Admission string, Visitors_with_Patient int, Age_Group string, Admission_Deposit float, Stay string, Department string, Severity_of_Illness string)**
row format delimited
fields terminated by '\$'
TBLPROPERTIES ("serialization.null.format"="");
- Let's put the data in HDFS -
 - **hdfs dfs -put /home/itv180149/Hive/Partitions/Hospital_Data/case.txt /user/itv180149**
- Let's load data into this non-partitioned Hive table -
 - **load data inpath "/user/itv180149/case.txt" into table cases_non_part;**
- Count the number of records that belong to the 'radiotherapy' or 'anesthesia' departments in the non-partitioned and the partitioned tables -
 - Running the query on the non-partitioned table -
 - **select count(*) from cases_non_part where department in ('radiotherapy', 'anesthesia');**

- Running the query on the partitioned table -
 - **select count(*) from cases_part where department in ('radiotherapy', 'anesthesia');**