

ORC File Format

- Create table that stored data in ORC format -
 - **create table cases_orc(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 '\$'
stored as orc;
- Put file to HDFS -
 - **hdfs dfs -put /home/itv180149/Hive/Partitions/Hospital_Data/case.txt /user/itv180149**
- Load data into Hive table -
 - **load data inpath "/user/itv180149/case.txt" into table cases_orc;**
 - We get an error because the data isn't in the format we are looking for. To successfully load this, we need to create a temporary table first and then insert the data from over there.
- Create temporary table -
 - **create table cases_text(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"="");

- Load data into Hive table -
 - **load data inpath "/user/itv180149/case.txt" into table cases_text;**
- Now let's drop the earlier ORC table -
 - **drop table cases_orc;**
- Create ORC table with data from temporary table -
 - **create table cases_orc stored as orc as select * from cases_text;**
- Describe table -
 - **describe formatted cases_orc;**
 - We have an ORC file format over here.