CSV SerDe

- Let's try and read a simple CSV file - -

  - **create table hospital_csv(Hospital_Code int, Hospital_Type_Code char(1), City_Code_Hospital int, Hospital_Region_Code char(1), Address string)**

    row format delimited

    fields terminated by ','

    stored as textfile;

- Put file to HDFS -

  - **hdfs dfs -put /home/itv180149/Hive/File_Formats/hospital.csv /user/itv180149**

- Let's try and read the data -

  - **load data inpath "/user/itv180149/hospital.csv" into table hospital_csv;**

  - The data isn't read perfectly.

  - Notice how the state after the comma in the address field is not read by Hive because we provided the comma separator. Use a CSV serde in Hive.

- Using the CSV serde that comes packaged in Hive -

  - **create table hospital_csv(Hospital_Code int, Hospital_Type_Code char(1), City_Code_Hospital int, Hospital_Region_Code char(1), Address string)**

    row format serde 'org.apache.hadoop.hive.serde2.OpenCSVSerde'

    stored as textfile;

- Put file to HDFS -

  - **hdfs dfs -put /home/itv180149/Hive/File_Formats/hospital.csv /user/itv180149**

- Load the data -

  - **load data inpath "/user/itv180149/hospital.csv" into table hospital_csv;**

- Let's try and read the data -

  - **select * from hospital_csv limit 2;**

- Check the table description -

  - **describe formatted hospital_csv;**

- Documentation link -

  - https://cwiki.apache.org/confluence/display/hive/csv+serde