|  |  |  |
| --- | --- | --- |
| **Sr. No** | **Hive** | **Pig** |
|  | **Load CSV File :**  Load data local inpath '/home/cloudera/file.csv' into table demo; | **Load CSV File:**  Register /user/cloudera/piggybank.jar  Demo = load '/user/cloudera/file.csv' using PigStorage(',') as demo(var: chararray); |
|  | **How to deal with Multi delimiter:**  Create table demo(id int,name string)  Row format serde 'org.apachhe.hadoop.hive.contrib.serde2.MultiDelimitSerDe'  With SERDEPROPERTIES ("fields.delim" ="#?>")  **Multi limiter data with data containing delimiters:**  Crearte table table\_n (fields type)  Row format 'org.apache.hadoop.hive.serde2.regexserde'  With serdeproperties(Input.regex = '^ (\\d\*), ([^\\t\*)\\|([^\\t\*)'); | **How to deal with Multi delimiter:**  Mult = load '/user/cloudera/file.csv' as (x:chararray);  M = foreach Mult generate flattern(regex\_extract\_all(x,'(.\*)[,|](.\*)[#,/](.\*))  Dump M |
|  | **How to deal with fixed width records**  Add jar /.........hive.contrib.jar  create table hive\_fixed (e\_id string,ename string,sal string,dept string)  row format serde 'org.apache.hadoop.hive.contrib.serde2.regexserde'  WITH SERDEPROPERTIES(  "input.regex"="(.{7})(.{15})(.{6})(.{8})",  "output.format.string"="%1$s %2$s %3$s %4$s"  ) | **How to deal with fixed width records**  Register Piggybank.jar  Define Substring 'org.apache.pig.piggybank.evaluation.substring'  X = load '/ h' as (row : chararray);  Y= foreach x generate (int) trim (row,substring(2,7) as emp\_id)) |
|  | **How to deal with Null Values:**  Create table demo(var string)  Row format delimited fields terminated by ','  TBLPROPRTIRS('serialization.null.format'=' '); |  |
|  | **How to deal with XML Files if records are missing :**  **Add jar /...../hivexmlserde.jar'**  **Create table tab\_name()**  **Row format serde 'com.ibmss.spss.hive.serde2.xml.xmlserde'**  **With serdeproperties ("column.xpath.transaction\_id" = "/transaction/transction\_id.text()",.....)** | **How to deal with XML Files if records are missing :**  Register piggybank.jar  Define xpath org.apache.pig.piggybank.evaluaton.xml.xpath()  Define xpath\_all org.apache.pig.piggybank.evaluation.xml.xpathall()  Xml\_load = load '/user/cloudera/demo.xml' using org.apache.pig.piggybank.xmlloader ('Book') as (x:chararray);  Xml\_final = foreach Xml\_load generate xpath(x,title),xpath(x,author); |