**PIG CLI Commands**

1. **LOAD Data from HDFS into relation without schema and read:**

*orders = LOAD '/user/root/sqoopimport/orders' USING PigStorage(',');*

*or*

*orders = LOAD '/user/root/sqoopimport/orders'*

1. **LOAD Data from HDFS into relation with schema and read:**

*orders = LOAD '/user/root/sqoopimport/orders' USING PigStorage(',') AS (order\_id: int,*

*order\_date: chararray, order\_customer\_id: int, order\_status: chararray);*

1. **LOAD Data from HIVE table (with schema only):**

*(open pig session with pig -useHCatalog to use HCatLoader class)*

*orders = LOAD 'akhil.orders' USING org.apache.hive.hcatalog.pig.HCatLoader();*

1. **APPLYING TRANFORMATION on Data in relation with schema:**

*ordersdate = FOREACH orders GENERATE order\_date AS orderdate;*

1. **APPLYING TRANSFORMATION on data without schema;**

*ordersdate = FOREACH orders GENERATE $1;*

1. **GROUP DATA with schema:**

*ordersgroup = GROUP orders ALL;*

*ordersgroupby = GROUP orders BY order\_status;*

1. **GROUP DATA without schema:**

ordersgroupby = GROUP orders BY $0;

1. **COUNT the rows:**

ordersgroupallcount = FOREACH ordersgroupall GENERATE COUNT\_STAR(orders) AS cnt;

ordersgroupbycount = FOREACH ordersgroupby GENERATE COUNT\_STAR(orders) AS cnt;

1. **FILTER data with schema:**

ordersfilternull = FILTER orders BY (order\_status == '');

ordersfilternotnull = FILTER orders BY (order\_status != '');

1. **FILTER data without schema:**

orderfilternull = FILTER orders BY ($3 IS NULL);

orderfilternotnull = FILTER orders BY ($3 IS NOT NULL);

1. **ORDER or SORTING:**

orderssort = ORDER orders BY order\_status asc;

orderssort = ORDER orders BY order\_status desc;

1. **DISTINCT:**

orders = LOAD 'akhil.orders' USING org.apache.hive.hcatalog.pig.HCatLoader();

ordersstatus = FOREACH orders GENERATE order\_status;

ordersstatusdist = DISTINCT ordersstatus;

1. **STORING Data from PIG into HDFS**:

STORE orders INTO '/user/akhilhdp' USING PigStorage('|');

or

STORE orders INTO '/user/akhilhdp' USING PigStorage(',');

or

STORE orders INTO '/user/akhilhdp' USING BinStorage('|');

or

STORE orders INTO '/user/akhilhdp' USING BinStorage(',');

or

STORE orders INTO '/user/akhilhdp' USING JsonStorage();

1. **STORING Data from PIG into HIVE:**

* First create a table in HIVE with exact same schema as the pig relation.
* Provide DATABASENAME.TABLENAME while importing and make sure to use HCatStorer class not HCatLoader class.

STORE orders INTO 'akhil.orders' USING org.apache.hive.hcatalog.pig.HCatStorer();

--------------------------------------------------------------------------------------------------------------------------------

--------------------------------------------------------------------------------------------------------------------------------

1. **Open pig session exec type tez:**

pig -x tez;

* open pig session as exce type tez and hcatalog for loading Hive data:
* Tez execution engine is not supported in the cloudera hadoop cluster.

pig -x tez -useHCatalog;

open pig session exec type mapreduce;

pig -x mapreduce;

1. **set number of parallels or reducers are pig level:**

set default\_parallel 4;

set default exec type as tez:

cd /etc/pig/conf

view pig.properties

set exec type = tez