Name : Tejas Redkar
PRN : 10322 10937, Panel-C
Roll No : BDT-22.
Batch: I [BDT-2]

BDT Lab Assignment - 8

* Problem Statement

Create pig database & perform data analytics on it.

* Objectives :

1) To learn pig concept
2) To perform data analytics on it.

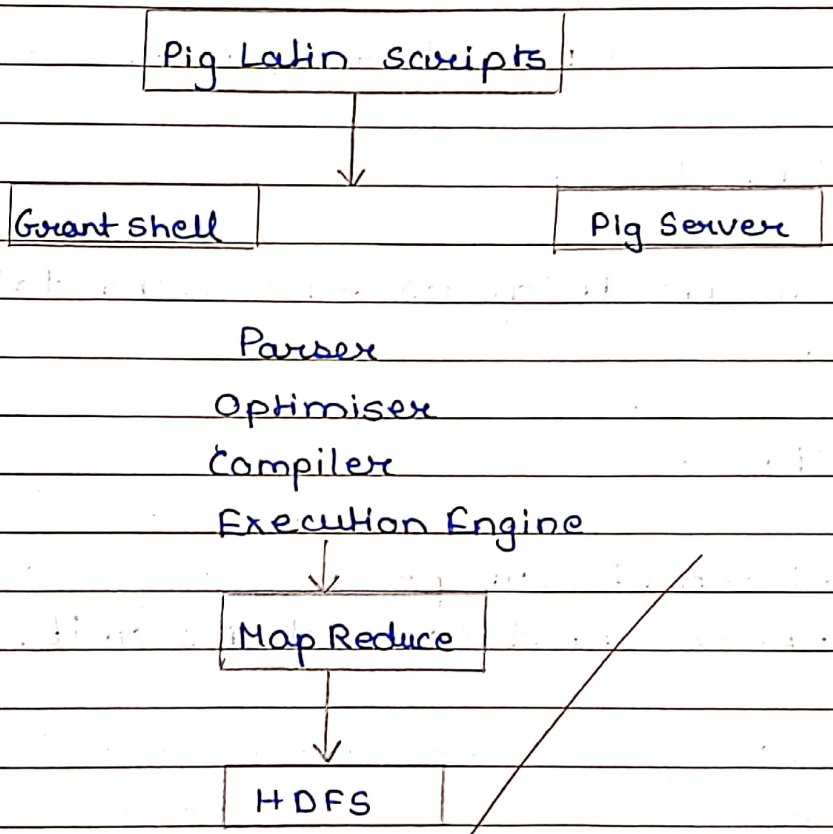
* Theory:

Pig is a high-level platform for processing & analyzing large datasets in Apache Hadoop. It provides an abstraction over Hadoop MapReduce, making it easier to work large-scale data. Pig Latin is the language used for writing Pig scripts.

Pig Architecture :

Pig Latin : The scripting language for defining data transformation & analysis.

- Pig Execution Environment: Pig scripts are executed. It supports local mode & MapReduce mode.
- UDFs: Custom functions to perform specific data processing tasks.



* Pig functions:

- Load: Loads the data
- FILTER: Filters records based on condition
- GROUP: Groups the data
- FOREACH: - Applies operations to each record
- JOIN: Combines Data
- STORE: Saves Data.

* Platform : 64-bit Open Source Windows

* Conclusion : Hence, I learned to create Pig Latin Program to perform data analytics

* FAQ's

Q1) Write a Pig scripts to perform JOIN operation.

Ans
- orders = LOAD 'orders_data' USING PigStorage(',') AS (order_id : int, order_data : char array, customer_id : int);
- customers = LOAD 'customers_data' USING PigStorage(',') AS (customer_id : int, customer_name : char array);
- joined_data = JOIN orders BY customer_id, customers BY customer_id;
- STORE joined_data INTO 'output_data';

Q2) Explain complex data types in Pig.

Ans Pig Tuple : An ordered set of fields.
Example : ' (1, 'Alice') ;

Bag : An unordered collection of tuples
Example : '{(1, 'Alice'), (2, 'Bob')};

Map : A key value pair collection
Example : ['name # Alice , id # 1];

3) State examples of Pig technology which can be used with hadoop.

Ans Pig can be used with hadoop thorough various mechanisms, including Pig on Tez, integration with HCatalog for metadata, custom UDFs, & Pig storage functions for different data formats.