

**What is Pig in Hadoop?**

Pig is a scripting platform that runs on Hadoop clusters designed to process and analyze large datasets. Pig is extensible, self-optimizing, and easily programmed.

Programmers can use Pig to write data transformations without knowing Java. Pig uses both structured and unstructured data as input to perform analytics and uses HDFS to store the results.

**Components of Pig**

There are two major components of the Pig:

* Pig Latin script language
* A runtime engine

**Pig Latin script language:**

The Pig Latin script is a procedural data flow language. It contains syntax and commands that can be applied to implement business logic. Examples of Pig Latin are LOAD and STORE.

**A runtime engine:**

The runtime engine is a compiler that produces sequences of MapReduce programs. It uses HDFS to store and retrieve data. It is also used to interact with the Hadoop system (HDFS and MapReduce).

The runtime engine parses, validates, and compiles the script operations into a sequence of MapReduce jobs.

**step1) Start the cloudera**

Step 2 : Now Open the terminal. And start Pig by typing pig on terminal.



Step 3: now pig get started

Step 4: now load the file mydata.txt file,

A=LOAD ‘/user/cloudera/Training/pig/mydata.txt’ AS (c1:int,c2:int,c3:int);

And then dump,

dump A;





Step 5: LOAD B

B=LOAD 'user/cloudera/Training/pig/mydata.txt.





Step 6:now check Schema of A.



Step 7: ILLUSTRATE A





File name : STUDENT.txt

All the data available in this file



Step 8: pig/home/cloudera/Documents/student.pig

Pig script file.





Display the patricular data:

Step 9:

Mention the data which we want to check. It will be display in listing form.

