

## Apache Pig

Apache Pig is a platform for analysing large data sets. It provides a high-level language, called Pig Latin, which abstracts the complexities of writing MapReduce programs in Java. Pig Latin scripts are translated into MapReduce jobs by the Pig runtime environment, allowing users to perform data processing tasks without having to write complex Java programs.

**Pig Latin:** This is the high-level scripting language used in Apache Pig. It resembles SQL and is designed to make it easy for users to express data processing tasks concisely.

**Grunt Shell:** Grunt is the interactive shell provided by Pig. It allows users to enter Pig Latin commands interactively and execute them directly.

- `pig -x local`
- `fs -ls`

**Task 1:** Create a text file and dump the data in PIG

- `X = LOAD 'path_of_file' USING PigStorage(',') AS (id:int, name:chararray, occupation:chararray, age:int);`
- `dump X;`
- `describe X;`

**Task 2:** Perform cross join between two data sets (A, B)

- `result = cross A, B;`
- `dump result;`

**Task 3:** Create a demo text file with id, name, age salary and filter the records less than 20000 salary;

`result = filter t by salary < 20000;`

`Dump result;`