1. Why MapReduce program is needed in Pig Programming?

Pig is a framework which is designed by using java language , which internally uses mapreduce predefined functions for execution of commands . When we observe architecture ,the compiler component compiles the optimized logical plan into a series of MapReduce jobs .That’s why mapreduce program is needed.

1. What are advantages of pig over MapReduce?
2. Pig does not require specialists in java programming
3. Length of the program is decreased in pig

For example : Word count program in mapreduce is of 60 to 70 lines

Where in pig it is 4-5 lines.

1. It is used by Data analysts and Data scientists for fast analysis.
2. Easy to learn, lot of people annoy to learn java ,for them pig will be the definite choice.
3. It got advantage of parallel processing of data also because internally mapreduce is used .
4. What is pig engine and what is its importance?

It is like a interpreter which executes pig scripts into series of mapreduce jobs in parallel manner. MapReduce jobs are submitted to Hadoop in a sorted order.

1. What are the modes of Pig execution?
2. MapReduce/Hadoop Mode

It uses HDFS while io operations in case of Map reduce mode.

Command “pig or pig –x mapreduce” is used to run Pig in MapReduce Mode.

1. Local Mode

It uses local file system while io operations in case of Map reduce mode.

Command “ pig –x local” is used to run Pig in MapReduce Mode.

1. What is grunt shell in Pig?

Pig uses another shell called grunt shell. There is a reason behind the name, as pig grunts so they kept shell name as grunt .It is interactive Shell for executing Pig Commands and it is used when script file is not provided . pig scripts are excuted from Grunt via run or exec commands

1. What are the features of Pig Latin language?

Pig Latin is a dataflow language which aims at DataFlow operations instead of control flow operations, hence more focus is on data analysis. In each processing step results in a new data set, or relation.

Pig Latin cannot decide whether it is case-sensitive. Keywords in Pig Latin are not case-sensitive; for example, LOAD is equivalent to load. But relation and field names are. So A = load 'foo'; is not equivalent to a = load 'foo';. UDF names are also case-sensitive, thus COUNT is not the same UDF as count.

1. Is Pig latin commands case sensitive?

Pig Latin cannot decide whether it is case-sensitive. Keywords in Pig Latin are not case-sensitive; for example, LOAD is equivalent to load. But relation and field names are. So A = load 'foo'; is not equivalent to a = load 'foo';. UDF names are also case-sensitive, thus COUNT is not the same UDF as count.

1. What is a data flow language?

Pig Latin is a dataflow language. . In each processing step results in a new data set, or relation which is stored in a variable. Data flows between the every operations

Example:

A= load ‘some file’;

It loads into ‘A’;

B= FILTER A BY (condition);

After filtering it loads the new data into B;

This is the data flow language property.