1. Why Map-reduce program is needed in Pig Programming?

Ans:

Pig(engine) is implemented on top of Hadoop MapReduce paradigm.

At the end, all pig commands/scripts are compiled into Map-Reduce programs by the Pig engine.

1. What are advantages of pig over MapReduce?

Ans:

1. It is easy for non-java programmers. It is much easier to read for someone without Java background
2. Easy of coding compared to MapReduce and have several operators to write complex logic. For ex: In Mapreduce, it is difficult to implement join b/w two data sets, whereas in Pig, it is simple to write.
3. Provides standard data-processing operations to achieve complex functionality
4. What is pig engine and what is its importance?

Ans:

Pig Engine executes the data flows in parallel on Hadoop. This includes a language, Pig Latin, to express the data flows.

This is to parse, optimize the Pig Latin statements and converts them into series of MapReduce programs.

1. What are the modes of Pig execution?

Ans:

There are 2 modes

1. Hadoop/MapReduce

This is to run Pig scripts on the HDFS files. It read/write the files from/to HDFS

Pig or Pig –x mapreduce

1. Local

This to run Pig scripts on Local file system. It doesn’t require HDFS/Hadoop.

Pig or pig –x Local

1. What is grunt shell in Pig?

Ans:

This is shell where Pig Latin statements are executed. This is the interactive shell for Pig Latin commands.

1. What are the features of Pig Latin language?

Ans:

This provides the standard data-processing operations like FILTER, ORDER, JOIN etc

1. Is Pig latin commands case sensitive?

Ans: Both

The names of relations are case-sensitive

Keywords ( like LOAD, STORE, FILTER, AS etc..) are case-insensitive

Functions like PigStore/COUNT are case-senstive

8. What is a data flow language?

Ans:

It is a programming paradigm that models the program as Directed Graph of the data flowing between operations.

Pig Latin describes a Directed Acyclic Graph, where edges are data flows and nodes are operators that process the data.