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

As a developer we do not write any map reduce code in pig programming. But still map reduce is needed to run the pig programs. It is the pig compiler converts the optimized pig code into map reduce jobs and submits to execution engine for execution. Then, the Map Reduce jobs are executed and the required result is produced.

2. What are advantages of pig over Map Reduce?

Map Reduce is low level of programming and Pig is a high-level language .Pig is simple to learn and use as compared to Map Reduce. Pig latin is a data flow language.

For a map reduce task to perform any data processing which need hundreds of lines of code(usually in java) can be accomplished using pig latin commands by writing very few lines of pig code to achieve the same task.

There is no need for compilation. On execution, every Apache Pig operator is converted internally into a Map Reduce job. Pig provides inbuilt optimization for MR jobs whereas in map reduce developer needs to take care of optimization.

3. What is pig engine and what is its importance?

PigEngine accepts the Pig Latin scripts as input and converts those scripts into MapReduce jobs. Developer with no concern of writing typical map reduce codes, can write only pig latin script and execute. Pig engine acts as interpreter between pig latin script and map reduce jobs. It creates environment to execute pig scripts into series of map reduce jobs.

4. What are the modes of Pig execution?

Apache Pig can be started either in cluster mode(hdfs mode) or in local mode.

i. To start Apache Pig in local mode, developers should use the command “pig -x local”. In local mode only the files within the local file system can be accessed and processed locally.

ii. If no option is specified then Pig by default is started in the cluster mode (OR) using command "pig -x mapredce" to start the pig in cluster mode. The cluster mode allows Pig to access data and process data files present on HDFS.

5. What is grunt shell in Pig?

Grunt shell is used to run Apache Pig in interactive mode either in local mode or cluster mode. In this shell, developer can enter the Pig Latin commands/scripts and get the output.

6. What are the features of Pig Latin language?

7. Is Pig latin commands case sensitive?

No. Pig latin commands/keywords are case insensitive.

8. What is a data flow language?

Dataflow programming is a programming paradigm that models a program as a directed graph of the dataflowing between operations, thus implementing dataflow principles and architecture. Data flow language can get a stream of data which passes from one instruction to another instruction to be processed. The instructions are flowing through data by executing different control statements but data doesn't get moved.