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

Pig is high level procedural language for querying semi structured data. Pig is intended for parallel data processing like Map-Reduce language. Pig aims at dataflow operations instead of control flow operations. Since pig latin script is built on top Map reduce, Any pig command which you execute in grunt shell or using pig script will be translated into Map reduce program. Without map-reduce, we cannot execute an y pig script.

1. What are advantages of pig over MapReduce?

Pig: data flow operations. More focus on data analysis

MapReduce: control flow operations

Pig: can write with in 10 pig statements for a complex requirement

MapReduce : such complex requirement could be written 1000 lines of code

Pig: Debugging would be easy as it is 10 lines of codes

MapReduce: Debugging takes lot of time as codes are more than 1000 line of code

Pig: Pig can easily do the computations over MapReduce

Pig: It uses simple built in syntax/operators to reduce development effort and lines of codes involved

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

Pig engine is the place where we execute pig script. Pig engine consists of parser, optimizer, and compiler and execution engine.

Pig Parser:-

1. Initially pig scripts are handled by the parser.
2. It checks the syntax of the script, does type checking and other miscellaneous check
3. The output of the parser will be a DAG( directed acyclic graph ) which represents the pig latin statements and logical operators
4. In the DAG, logical operators of the scripts are represented as the nodes and the data flows are represented as edges

Optimizer: - The logical plan (DAG) is passed to the logical optimizer which carries out the logical optimizations such as projection and pushdown

Compiler:- The compiler compiles the optimized logical plan into a series of map reduce jobs.

Execution engine:- Finally MapReduce jobs are submitted to hadoop in a sorted order and produce the desired result

1. What are the modes of Pig execution?

There are 2 mode of pig execution. Pig local and pig mapreduce

Pig local – if we execute script with pig local mode, It will refer local file system

Pig –x local text.pig

Pig MapReduce – If we execute script with pig mapreduce mode, it will refer hdfs file sytem

Pig –x mapreduce text.pig or pig –x text.pig

1. What is grunt shell in Pig?

Grunt shell in pig is interactive command mode where we execute pig statements one after another.

1. *What are the features of Pig Latin language?*

*Pig Latin is a dataflow language. Each processing step results in a new data set, or*relation.

Pig can operate on data with or without metadata.

Pig is intended to be a language for parallel data processing.

Pig can be used for data analytics

Pig statement simplifies the use of hadoop by allowing SQL-like queries to a distributed dataset

7. Is Pig latin commands case sensitive?

No. It is not case sensitive.

8. What is a data flow language?

Pig Latin is a dataflow language. *Each processing step results in a new data set, or*relation. Next statement in a script will work on the results of new dataset which was produced by first step.