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

Pig is a high-level platform that makes many Hadoop data analysis issues easier to execute. A program written in Pig Latin is a data flow language, which need an execution engine to execute the query. So, when a program is written in Pig Latin, Pig compiler converts the program into MapReduce jobs.

2. What are advantages of pig over MapReduce?

No java required. we see that we don't have to import any libraries. Line of code to write is reduced by huge margine. Sql queries are easier to understand. Also we can run Pig in local as well. Where the entire pig job runs as a single JVM picking the local unix path for execution.

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

It’s a component of pig. Pig engine acts as a interpreter between pig latin script and mapreduce jobs. It creates environment to execute pig scripts into series of map reduce jobs in parallel.

4. What are the modes of Pig execution?

1. Local Mode: No need to start or install hadoop. Pig scripts run in local system. By default pig store data in local FS.

Command to run in local: pig –x local

1. MapReduce Mode: need to start hadoop. Pig scripts are run and stored in HDFS.

Type command pig –x mapreduce to run pig in mapreduce mode.

Make sure job history server is running.

Mr-jobhistory-daemon.sh start historyserver

It help us to view previous MR job details.

5. What is grunt shell in Pig?

Interacting shell for executing pig commands.

Used when script file is not provided.

Can execute script from grunt using run or exec commands.

To start Grunt, users should use pig –x local command . This command will prompt Grunt shell. To exit from grunt shell, press CTRL+D or just type exit.

6. What are the features of Pig Latin language?

Pig latin is a dataflow language which need an execution engine to execute the query. It included operators for many of the traditional data operations (join, sort, filter etc) as well as the ability for users to write their own functions for reading, processing and writing data.

Its describes a directed acyclic graph (DAG), where the edges are data flows and the nodes are operators that process the data.

Pig latin just requires few lines of scripts compared to MapReduce programs.

It includes both atomic data types like int and float etc and also complex data types like tuble, bag and map.

7. Is Pig latin commands case sensitive?

Pig script is both case sensitive and case insensitive.

User defined functions, the field name, and relations are case sensitive i.e. EMPLOYEE is not same as employee or M=LOAD ‘data’ is not same as M=LOAD ‘Data’.

Whereas Pig script keywords are case insensitive i.e. LOAD is same as load.

8. What is a data flow language?

To access the external data, every language must follow many rules and regulations. The instruction are flowing through data by executing different control statements, but data doesn’t get moved. Dataflow language can get a stream of data which process those conditions , jumps, loops and process the data in efficient manner.