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

Pig programs are written in a query language known as Pig Latin which is similar to the SQL query. To execute the query, there is a need for an execution engine. The Pig engine converts the queries into MapReduce jobs and thus MapReduce acts as the execution engine and is needed to run the programs.

2. What are advantages of pig over MapReduce?

* Pig programs are written in a query language hence development is much faster in Pig.
* Pig is simple to learn and use as compared to Map Reduce.
* Pig data flow language i.e pig Latin. For Map Reduce, Java is by default supported programming language.
* Pig supports Multiquery (it is when PIG tries to minimize the number on MapReduce Jobs by doing more stuff in a single Job).
* If we want to perform Small filter function using MapReduce, we need to write min three class to run it, but in pig it only few lines of data flow commands.

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

It acts as interpreter between pig latin and Mapreduce job. It is creating environment to execute pig script into series of map reduce job in parllel manner.

4. What are the modes of Pig execution?

* MapReduce Mode: This is the default mode, which requires access to a Hadoop cluster and HDFS installation. Since, this is a default mode, it is not necessary to specify -x flag (you can execute pig OR pig -x MapReduce). The input and output in this mode are present on HDFS.
* Local Mode: With access to a single machine, all files are installed and run using a local host and file system. Here the local mode is specified using ‘-x flag’ (pig -x local). The input and output in this mode are present on local file system.

5. What is grunt shell in Pig?

Grunt is Pig’s interactive shell. It enables users to enter Pig Latin interactively and provides a shell for users to interact with HDFS. Enter Grunt by typing: pig (MapReduce), pig –x local (local mode).

6. What are the features of Pig Latin language?

* Pig Latin is a dataflow language.
* Pig can be executed in two modes (local and MapReduce mode).
* Pig supports UDFs.
* Pig Latin is schema independent.

7. Is Pig latin commands case sensitive?

The names of relations and fields are case sensitive. The names of Pig Latin functions are case sensitive. The names of parameters and all other Pig Latin keywords are case insensitive.

8. What is a data flow language?

To access the external data, every luggage must follow rules and regulations. The instructions are flowing through data by executing different control statement, but data doesn’t get moved. Data flow language can get a stream of data which passes from one instruction to another instruction to be process. Pig can easily process those conditions, jumps, loops.

And it allows the user to describe how data from one or more inputs should be read, processed, and then stored to one or more outputs in parallel.