**1. Why MapReduce program is needed in Pig Programming?**

Pig is an abstraction over MapReduce. In other words, all Pig scripts internally are converted into Map and Reduce tasks to get the task done.

So Map reduce is the layer on top of which Pig is built.Moreover Map reduce is the means by which Hadoop can communicate with HDFS.so

MapReduce program is needed in Pig Programming

**2.Advantages of Pig Latin over Map Reduce**

1.Coding is Highly reduced(number of lines of code is 1/20th of that required in Map reduce) for example

In Pig Latin joins and ordering codes comprise of 8-9 lines of code and take few minutes to write and debug. The same code in MapReduce will span hundred lines of code and takes hours to develop.

2. Pig Latin provides all of the standard data-processing operations, such as join, filter, group by, order by, union, etc.

3.Time consumed is also less(1/16th of the time is only required)

4.MapReduce requires programmers: • Programmers must think in terms of map and reduce functions. Most probably Java programmers are required.

5. Pig provides high-level language that can be used by: • Data Analysts • Data Scientists

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

It acts as Interpreter between pig latin script and map reduce jobs.It creates environment to execute Pig scripts into series of map reduce job in parallel manner

**4. What are the modes of Pig execution?**

There are 2 modes of pig execution

1.Local Mode

2.Map Reduce Mode

Local Mode

In this mode, all the files are installed and run from your local host and local file system. There is no need of Hadoop or HDFS. This mode is generally used for testing purpose.

This can be started by **pig –x local**

Map Reduce Mode

MapReduce mode is where we load or process the data that exists in the Hadoop File System (HDFS) using Apache Pig. In this mode, whenever we execute the Pig Latin statements to process the data, a MapReduce job is invoked in the back-end to perform a particular operation on the data that exists in the HDFS.

This can be started by **pig –x mapreduce or by just Pig as Map reduce mode is the default mode**

**5. What is grunt shell in Pig?**

Grunt shell is an interactive shell for running Pig Commands .

It is used when script is not provided

It has line editing facilities like other command line application

It also has a completion mechanism like eclipse which can be done by using **TAB KEY**

**Eg:**grunt > m=foreach a ge

And if we press the tab key grunt > m=foreach a generate will automatically completed

**6.Features Of Pig Latin**

Apache Pig is a high-level procedural language for querying large semi-structured datasets using Hadoop and the MapReduce Platform.

• Pig simplifies the use of Hadoop by allowing SQL-like queries to a distributed dataset.

• Pig provides an engine for executing data flows in parallel on Hadoop. It includes a language, Pig Latin, for expressing these data flows.

• Pig Latin includes operators for many of the traditional data operations (join, sort, filter, etc.), as well as the ability for users to develop their own functions for reading, processing, and writing data.

• Pig runs on Hadoop. It makes use of both the Hadoop Distributed File System, HDFS, and Hadoop’s processing system, MapReduce.

• It allows users to describe how data from one or more inputs should be read, processed, and then stored to one or more outputs in parallel.

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

**7. Is Pig latin commands case sensitive?**

The names (aliases) of relations and fields are case sensitive. The names of Pig Latin functions are case sensitive. The names of parameters (see Parameter Substitution) and all other Pig Latin keywords are case insensitive.

In the example below, note the following:

1. The names (aliases) of relations A, B, and C are case sensitive.
2. The names (aliases) of fields f1, f2, and f3 are case sensitive.
3. Function names PIGSTORAGE and COUNT are case sensitive.
4. Keywords LOAD, USING, AS, GROUP, BY, FOREACH, GENERATE, and DUMP are case insensitive. They can also be written as load, using, as, group, by, etc.
5. In the FOREACH statement, the field in relation B is referred to by positional notation ($0).

**8. What is a data flow language?**

In a dataflow language, you have a stream of data which is passed from instruction to instruction to be processed. Conditional execution, jumps and procedure calls route the data to different instructions. This could be seen as data flowing through otherwise static instructions like how electrical signals flow through circuits or water flows through pipes. A dataflow "if" statement would route the data to the correct branch.

Pig provides developers many operators which can be applied on data one after another to get final output.

Once data is loaded, it flows through all Pig operators.

This is the reason Pig is called as data flow language.