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

* Pig is an abstraction over MapReduce it is beacause all Pig scripts were internally converted into Map as well Reduce tasks in order to make the work done.
* Top layer of Pig is built with map reduce.
* Map reduce job is to make Hadoop to communicate with HDFS.
* Map Reduce program is compulsory in all Pig Programming.

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

* Coding is used to reduce the number of lines in coding (number of lines of code is 1/20th of that required in Map reduce)

For example:

* Pig Latin joins and order codes which have only 8-9 lines of code and will take only few minutes to write and debug.
* When same work is done via Map Reduce then it will take more number of lines and time to develop.
* Pig Latin provides all standard data-processing operations like join, filter, group by, order by, union, etc.
* Time consumed will be much less when compared to normal map reduce ( only 1/16th of time will be required)
* Map Reduce requires programmers who needs to know java programs as well map and reduce functions. Most probably Java programmers are required.
* Pig provides high-level language that can be used:
* • Data Analysts
* • Data Scientists

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

* It is an Interpreter between pig latin script and map reduce.
* Environment is created to execute Pig scripts with map reduce job as parallel process.

**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 Local mode, the files are installed and were made to run from local host and local file system. In this case Hadoop or HDFS were not needed. And were mostly used for testing purpose.
* It was started by using **pig –x local**

**Map Reduce Mode**

* Here the data is loaded or processed that will exists in Hadoop File System (HDFS) using Apache Pig.
* When we execute Pig Latin statement to process data, MapReduce will run in the back-end to perform some particular operation on data which will exists in the HDFS.
* It is initiated by using pig –x mapreduce or by using Pig as Map reduce as default mode

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

* Grunt shell is used as intermediate shell to run Pig Commands.
* Used when there is no script provided.
* Like command line application it has line editing facilities.
* like eclipse it has completion mechanism done using **TAB KEY**
* **Eg:** grunt > m=foreach a ge
* When tab key is pressed then the key grunt > m

**6.Features Of Pig Latin**

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

• Pig simplified by using 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.

•It includes operators for traditional data operations (like join, sort, filter, etc.), also for users to develop their own functions for reading, processing, and writing data.

• Pig can be run on Hadoop. uses both Hadoop Distributed File System (HDFS), and Hadoop’s processing system (MapReduce).

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

• It describes directed acyclic graph (DAG), where the edges were data flows and nodes were operators which are used to process the data.

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

* The names of relations as well the fields are case sensitive.
* Parameters (and their Parameter Substitution) and Pig Latin keywords are case insensitive.

For example :

1. Names of relations A, B, and C are case sensitive.
2. As well Names of fields f1, f2, and f3 are case sensitive.
3. Function names of 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.

**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.
* It follows conditional execution, jumps and procedure calls to route the data to different instructions and this is how the data flow through or else for instructions like how electrical signals flow through circuits or water flows through pipes "if" statement will be used in order to the correct branch.
* Developer were provided with many operators which are applied one after another as data to get final output.
* When the data is loaded it will flow through all Pig operators and so Pig is called as data flow language.