1. Why MapReduce program is needed in Pig Programming?

Pig is an application that works on top of MapReduce. Pig is written in Java and compiles Pig Latin scripts into to MapReduce jobs. In simple terms Map Reduce is low level of programming and Pig is a high-level language for expressing data analysis programs which internally create sequence of Map Reduce Programs. Pig is simple to learn and use as compared to Map Reduce. Pig data flow language i.e pig Latin. For MapReduce, Java is by default supported programming language. However support for other language is also available. Pig provides inbuilt optimization for MR jobs whereas in map reduce developer needs to take care of optimization.

1. What are advantages of pig over MapReduce?

* Pig is application that runs on top of MapReduce and abstracts Java MapReduce jobs away from developers.
* Pig Latin uses a lot fewer lines of code than the Java MapReduce script.
* The Pig Latin script is easier to read for someone without a Java background.
* MapReduce jobs can written in Pig Latin.
* Java is a great and powerful language, but it has a higher learning curve than Pig Latin. Therefore, using a higher-level language, like Pig Latin, enables many more developers to write MapReduce jobs.

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

Pig Engine acts as interpreter between Pig Latin script and MapReduce Jobs. It creates environment to execute Pig scripts into series of mapreduce jobs in parallel manner.If it is compared to Mapreduce many features are add on in the Pig. In MapReduce its very difficult to join multiple data sets. Development cycle is very long. Depending on the task, pig converts the code into map or the reduce. Easy to join multiple tables and run many sql queries like join, filter, group by, order by, union and many more.

1. What are the modes of Pig execution?

Pig has two execution modes or exectypes:

* **Local Mode** - To run Pig in local mode, you need access to a single machine; all files are installed and run using your local host and file system. Specify local mode using pig -x local.
* **Mapreduce Mode** - To run Pig in mapreduce mode, you need access to a Hadoop cluster and HDFS installation. Mapreduce mode is the default mode; you can, *but don't need to*, specify it using the pig OR pig -x mapreduce.

1. What is grunt shell in Pig?

The Grunt shell of Apache Pig is mainly used to write Pig Latin scripts. The Grunt shell provides a set of utility commands. These include utility commands such as clear, help, history, quit, and set; and commands such as exec, kill, and run to control Pig from the Grunt shell.

1. What are the features of Pig Latin language?

A Pig Latin statement is an operator that takes a [relation](https://pig.apache.org/docs/r0.7.0/piglatin_ref2.html#Relations%2C+Bags%2C+Tuples%2C+Fields) as input and produces another relation as output. This applies to all Pig Latin operators except LOAD and STORE which read data from and write data to the file system. Pig Latin statements can span multiple lines and must end with a semi-colon (;). Pig Latin statements are generally organized in the following manner:

1. A LOAD statement reads data from the file system.
2. A series of "transformation" statements process the data.
3. A STORE statement writes output to the file system; or, a DUMP statement displays output to the screen.
4. 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 and all other Pig Latin keywords are case insensitive.

1. What is a data flow language?

In computer programming, dataflow programming is a programming paradigm that models a program as a directed graph of the data flowing between operations, thus implementing dataflow principles and architecture. Dataflow programming languages share some features of functional languages, and were generally developed in order to bring some functional concepts to a language more suitable for numeric processing.

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.