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

Pig is a platform to analyze large data sets. Pig consists of high-level language for expressing data analysis programs, coupled with infrastructure for evaluating these programs. Pig programs support parallelization, which enables them to handle very large data sets. Pig's infrastructure layer consists of a compiler that produces sequences of Map-Reduce programs, for which large-scale parallel implementations already exist i.e. the Hadoop subproject.

1. What are advantages of pig over MapReduce?
2. Ease of programming: Complex tasks comprised of multiple interrelated data transformations are explicitly encoded as data flow sequences, making them easy to write, understand, and maintain.
3. Optimization opportunities: Pig permits the system to optimize their execution automatically, allowing the user to focus on semantics rather than efficiency.
4. Pig Latin script has fewer lines of code that is easy to write and understand.
5. Pig’s multi-query approach combines certain types of operations together in a single pipeline, reducing the number of times data is scanned. This means 1/20th the lines of code and 1/16th the development time when compared to writing raw MapReduce.
6. It process any type of data: structured, semi-structured, or unstructured.
7. Pig provides nested data types (e.g. tuples, bags, and maps) missing from MapReduce.
8. Pig insulates your code from changes to the Hadoop Java API, so your jobs won’t suddenly break due to an update. It also manages all details of submitting jobs and running complex data flows.
9. What is pig engine and what is its importance?

Pig Engine parses, optimizes, and automatically executes Pig Latin scripts as a series of MapReduce jobs on a Hadoop cluster.

It 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.

1. What are the modes of Pig execution?

Pig has two execution modes or types. They are:

1. Local Mode
2. MapReduce Mode

Local Mode:

In Local Mode of Pig execution, all the input data will be taken from local file system. After execution it provides output on top of local file system. In local mode, Pig runs in a single JVM and accesses the local filesystem. This mode of Pig is suitable only for small datasets and when trying out Pig. To start the local mode of execution, the following command is used.

# pig -x local

MapReduce Mode/HDFS Mode/ Clustered Mode:

In this mode Pig will take the input form HDFS paths only, and after processing data it will put output files on top of HDFS. In MapReduce mode of execution, Pig translates queries into MapReduce jobs and runs them on a Hadoop Cluster.

1. What is grunt shell in Pig?

Grunt acts as a command interpreter where you can interactively enter Pig Latin at the Grunt command line and immediately see the response. This method is helpful for prototyping during initial development and with what-if scenarios.

1. What are the features of Pig Latin language?

A Pig Latin statement is an operator that takes a relation as input and produces another relation as output.

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.

Pig processes Pig Latin statements as follows:

First, Pig validates the syntax and semantics of all statements.

Next, if Pig encounters a DUMP or STORE, Pig will execute the statements.

You can execute Pig Latin statements:

1. Using grunt shell or command line
2. In MapReduce mode or local mode
3. Either interactively or in batch
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 (see [Parameter Substitution](http://pig.apache.org/docs/r0.10.0/cont.html#Parameter-Sub)) and all other Pig Latin keywords (see [Reserved Keywords](http://pig.apache.org/docs/r0.10.0/basic.html#Reserved-Keywords)) are case insensitive

1. 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