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

* Pig Latin is script language and, at the moment, there is no IDE or text editor that can help in maintaining and refactoring Pig code as well as it might be needed.
* Java is a first-class language in Hadoop and will always give the developer more options.
* When run against much bigger set of data, five millions of records, script finished in two minutes (roughly) comparing to native Map/Reduce time of around forty seconds.

1. What are advantages of pig over MapReduce?

Pig Latin uses a lot fewer lines of code than the Java MapReduce script.

We don't have to import any libraries.

Easier to read for someone with a little SQL background.

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

**Pig** is a platform for analyzing large data sets that consists of a high-level language for expressing data analysis programs, coupled with infrastructure for evaluating these programs. The salient property of Pig programs is that their structure is amenable to substantial parallelization, which in turns enables them to handle very large data sets.

Importance:

* 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.
* Pig got its name because it’s omnivorous – it will happily consume any data you feed it: structured, semi-structured, or unstructured.
* Pig provides the common data operations (filters, joins, ordering, etc.) and nested data types (e.g. tuples, bags, and maps) missing from MapReduce.
* It’s easy to learn (especially if you’re familiar with SQL) and opens Hadoop to data professionals who may not be software engineers.
* PigLatin reads like a series of steps (e.g. join this data to that data, then filter the result…) so it is easy to write, and even better, it is easy to read.
* Pig is easily extensible by UDFs – including Python, Java, JavaScript, and Ruby – so you can use them to load, aggregate, or do sophisticated analysis.
* 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.

1. What are the modes of Pig execution?

* 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.
* Mapreduce Mode - To run Pig in mapreduce mode, you need access to a Hadoop cluster and HDFS installation.

1. What is grunt shell in Pig?

* Using Grunt shell, we can run your Pig scripts in the shell. In addition to that, there are certain useful shell and utility commands provided by 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.
* Pig Latin statements can span multiple lines and must end with a semi-colon ( ; ).

1. Is Pig latin commands case sensitive?

Yes

1. What is a data flow language?

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.