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

* Pig is an application that works on top of MapReduce, Yarn or Tez.
* Pig is written in Java and compiles Pig Latin scripts into to MapReduce jobs. Pig as a compiler that takes Pig Latin scripts and transforms them into Java.
* While analysing data using Pig, aggregation is done using map tasks

2. What are advantages of pig over MapReduce?

* An **advantage PIG** has **over MapReduce** is that the former is more concise. A 200 lines Java code written for **MapReduce** can be reduced to 10 lines of **PIG** code.
* MapReduce requires programmers must think in terms of map and reduce functions. Most probably Java programmers are required. Pig provides high-level language that can be used by: Data Analysts ,Data Scientists

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

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

4. What are the modes of Pig execution?

Pig can run in 2 modes

– Local mode

– MapReduce mode/Hadoop mode.

5. What is grunt shell in Pig?

•Interactive Shell for executing Pig Commands.

•Used when script file is not provided.

•Can execute scripts from Grunt via run or exec commands.

6. What are the features of Pig Latin language?

Pig Latin is a command based scripting language and designed specifically for data transformation and flow.

* + - Dataflow Language
    - Quick Start and Interoperability
    - Nested Data Model
    - UDFs as First-Class Citizens
    - Parallelism Required
    - Debugging Environment

7. Is Pig latin commands case sensitive?

No, pig latin commands are not case sensitive.

Eg:LOAD /file1.txt/ is same as load /file1.txt/

8. What is a data flow language?

 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.

 It is trivial to achieve parallel execution of simple, "embarrassingly parallel" data analysis tasks. Complex tasks comprised of multiple interrelated data transformations are explicitly encoded as data flow sequences, making them easy to write, understand, and maintain.