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

Ans : Pig is an abstraction over MapReduce. All Pig scripts internally converted into Map and Reduce tasks to get the task done. Pig was built to make programming MapReduce applications easier.

2. What are advantages of pig over MapReduce?

Ans: Below are the advantage of pig over MapReduce

* Pig is a data flow language whereas MapReduce is a data processing paradigm.
* It is a high level language whereas MapReduce is low level and rigid.
* Performing a Join operation in Apache Pig is pretty simple compare to MapReduce.
* Basic knowledge of SQL require to work conveniently with Pig Whereas for MapReduce Java is must.
* Pig uses multi-query approach, thereby reducing the length of the codes to a great extent. Almost 20 times less compare to MapReduce.
* There is no need for compilation. On execution, every Apache Pig operator is converted internally into a MapReduce job. MapReduce jobs have a long compilation process.

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

Ans : Pig Engine is an interpreter between Pig Latinscript and MapReduce Jobs.

Pig Engine creates environments to execute Pig scripts into series of MapReduce jobs in parallel manner.

Pig parses, compiles, optimizes, and fires MapReduce statements.

MapReduce accesses HDFS and returns the results.



4. What are the modes of Pig execution?

Ans: **Pig Execution mode: MapReduce mode , Local mode**

a. MapReduce/Hadoop Mode: In this mode Pig jobs run as a series of MapReduce jobs picking the input and output paths from HDFS. Input file has to be copied in HDFS in case of Map reduce mode.

Command: pig –x mapreduce

b.Local Mode: In this mode the entire Pig job runs as a single JVM picking the local Unix path for execution.Input file has to be kept in local file system

Command: pig -x local

**Mode of Accessing approaches:**

1. Batch mode: submit a script directly
2. Interactive mode: Grunt, the pig shell

5. What is grunt shell in Pig?

Ans : Grunt Shell is an Interactive Shell or command line for executing Pig Commands.

The grunt shell execute PigLatin statements to quickly test out data flows on your data step by step without having to execute complete scripts. It is used when script file is not provided.

Scripts can execute from Grunt via run or exec commands.

6. What are the features of Pig Latin language?

Ans: Pig Latin is a command based scripting language and designed specifically for

data transformation and flow. Features of Pig Latin language are as follows:

* **Rich set of operators** − It provides many operators to perform operations like join, sort, filer, etc.
* **Ease of programming** − Pig Latin is similar to SQL and it is easy to write a Pig script.
* **Optimization opportunities** − The tasks in Apache Pig optimize their execution automatically.
* **Extensibility** − Using the existing operators, custom functions can created to read, process, and write data.
* **UDF’s** − Pig provides the facility to create User-defined Functions in other programming languages such as Java and invoke or embed them in Pig Scripts.
* **Handles all kinds of data** − Apache Pig analyzes all kinds of data, both structured as well as unstructured. It stores the results in HDFS.
* **Multi-query approach –** Less line of codes. For example, an operation that would require 200 lines of code in Java can be easily done by typing as less as just 10 LoC in Apache Pig.

7. Is Pig latin commands case sensitive?

Ans : Pig latin commands are both ,case sensitive and case insensitive.

* 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.

Example :

1. The names/aliases of relations A, B, and C are case sensitive.
2. The names/aliases of fields f1, f2, and f3 are case sensitive.
3. Function names 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?

Ans: A Dataflow language is a language in which a stream of data passes from one instruction to another instruction to be processed. Pig easily process those conditions, jumps, loops and process the data in efficient manner.

Dataflow flow languages are inherently parallel, because the operations rely on inputs that when met will cause the operation to execute. This means unlike a normal program where one operation is followed by the next operation, in a dataflow program operations will execute as long as the inputs are met and thus there is no set order.

Often dataflow programming languages use a large hash table where the keys are the data of the program and the values of the table are pointers to the operations of the program. This makes multicore programs easier to create in a dataflow programming language, since each core would only need the hash table to work.