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

While running Pig script on MapReduce mode we need MapReduce jobs to be running.

For complex programs which cannot be handled by Pig scripting we need map reduce program to be developed

and used as there are limited numbers of predefined actions defined on Pig scripting.

So to execute and develop complex kind of actions on data we need MapReduce program.

1. What are advantages of pig over MapReduce?

Pig scripting takes less time to be developed so less time to maintain it.

Learning pig scripting is easier than that of Java used for MapReduce.

It provides standard data processing operations, like FILTER, JOIN, GROUP, etc.

It is a data flow language so users thinks about data need not worry about control structures of Java.

Enjoys everything that Hadoop offers, parallelization, fault-tolerance with many relational database features.

 Pig is one of the best tool to make the large unstructured data to structured.

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

It provides an execution engine for the pig Latin script language on Hadoop cluster.

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

1. What are the modes of Pig execution?

Local mode(works on local file system) and MapReduce Mode(works on HDFS file system).

1. What is grunt shell in Pig?

Used for writing the pig latin script for data processing.

It also used for invoking any FileSystem Shell command from within the Grunt shell.

1. What are the features of Pig Latin language?

It is a data flow language, focus is on data not on control structure for the programmer.

Predefined operator available for data processing so no need to write thousands of line for processing simple data.

Easy to learn and maintain the language.

1. Is Pig latin commands case sensitive?

Pig latin keywords are case insensitive whereas in relation it is case sensitive.

For keyword : LOAD and load is treated as same.

For Relation : A = load ‘f1.txt’ and a = load ‘f1.txt’ both are treated as different

1. What is a data flow language?

Data flow language is that where execution of statements takes place one after another in a simple pipeline.

In this stream of data which is passed from instruction to instruction to be processed with no jumps or loops.