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

Ans :- Pig is a scripting language whose main functionality is to analyze large data sets which is not possible using traditional method (like using RDBMS and querying onto dataset )their already exist a map reduce framework which does this job perfectly by first diviing large datasets into multiple spilts and assigning to nodes (taken by HDFS ) and then analyzing by mapper and reducer programs

So Map-Reduce is much needed and so Pig runs on top of it ,Pig programs are written in a query language known as Pig Latin that is similar to the SQL query language. To execute the query, there is a need for an execution engine. The Pig engine converts the queries into MapReduce jobs and thus MapReduce acts as the execution engine and is needed to run the programs.

Main advantages of using Pig

providing easy method to write code for analyzing data

also enabling user to perform complex query like joining of two datasets

development cycle is much lesser here

2. What are advantages of pig over MapReduce?

Answer :

1. It is easier to write a code in Pig compared to MapReduce.
2. Even an analyst or person from non programing background can easily write code in Pig , as it is similar to a query language and contains all the useful Relational and evaluation and built In function module .
3. Development Time :- A single command can perform a task which may need multiple lines of code, so Development tme is less .

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

Asnwer:

 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?

Answer :

Pig can be used in two modes

1. Local mode \_ - when executed in this mode , it uses local file system

Command = pig –x local

1. MapReduce - when executed in this mode , it uses Hadoop distributed file system

Command = pig

Or

pig –x mapreduce

There is no other differences in the pig commands which needed to be executed .

5.What is grunt shell in Pig?

Answer :

6. What are the features of Pig Latin language?

Answer :

Features

1. Pig is a DataFlow Language, a user specifies a sequence of steps where each step specifies only a single, highlevel data transformation.
2. 2.2 Quick Start and Interoperability, Pig is designed to support ad-hoc data analysis. If a user has a data file obtained, say, from a dump of the search engine logs, she can run Pig Latin queries over it directly

7. Is Pig latin commands case sensitive?

Answer

It is difficult to say whether Apache Pig is case sensitive or not.

Considering user defined functions, relations and field names in pig are case sensitive i.e. the function  COUNT is not the same as function count or X=load ‘foo’ is not same as x=load ‘foo’.

While , keywords in Apache Pig are case insensitive i.e. LOAD is same as load.

8. What is a data flow language?

Answer : In Pig Latin, a user specifies a sequence of steps where each step specifies only a single, highlevel data transformation. This is stylistically different from the SQL approach where the user specifies a set of declarative constraints that collectively define the result. While the SQL approach is good for non-programmers and/or small data sets, experienced programmers who must manipulate large data sets often prefer the Pig Latin approach