DSBDAL

Assignment No:- 12

* Aim:- Design a distributed application using map Reduce which process a log file of system.

* Theory:- Big data.
- Big data can be termed, as that dosed load of data that can be hardly processed using the traditional data processing units.

• Hadoop:-
- Hadoop is a big data framework designed and deployed by apache foundation. It is an open source software utility that works in the network of computer in parallel to find solution to big data process it using the map Reduce algorithm.

- Hadoop consists the master slave architecture. The Hadoop consist node & Data node.
The name node is master and data node is slave.
The data is firstly divided into sumb block i.e. also segments.

- Then that 64mb blocks are transferred to name node. The name node transfer or give the infrastructure to the data node.

- Then the data is stored into data nodes respectively.

- The big data is often characterized by the value three.

- Three large volume of data.
- The wide variety of data.
- The variety at which much as the data.

- Examples of how big data is used by organization -

1) In the energy industry, big data helps oil and gas companies identify potential drilling location and monitor pipeline operations, likewise we it to track the electrical grids.

2) financial sevice firms users big data system for the risk management.

3) Manufactures and transportation companies rely on big data to manage their supply chains and optimize delivery routes.

* mapreduce :-

It consists of two distinct token map and reduce. As the name mapreduce suggests the reducer phase takes place after mapper phase has been completed.

So, the first the map job where a block of data is read & processed to the produce key value pairs as intermedidte outputs.

* conclusion :-

Hence, We learnt about the map reduce log files.