

Q. Explain Apache Hadoop.

-
- Apache Hadoop project develops open-source SW for reliable, scalable & distributed computing.
 - Apache Hadoop SW library is a framework.
 - It allows distributed processing of large data sets across clusters of computers using simple programming models.

Modules in Apache Hadoop

① MapReduce -

- This is a system for parallel processing of large sets.

2

② Hadoop YARN -

This is a framework for job scheduling & cluster resource management.

③ HDFS - (Hadoop Distributed File System)

- HDFS is a unique design technique that provides storage for extremely large files with streaming data.

④ Hadoop Common -

- These are common utilities that support the other Hadoop modules.

Q. Explain MapReduce.

-
- MapReduce is a programming model for processing large datasets using parallel, distributed & clustered compute nodes.



Apache Pig consists of a data flow language called Pig Latin. It also provides a runtime environment to execute Pig Latin programs

- Q Explain MapReduce Algorithm
- MapReduce Algorithm contains two important tasks.
- ① Map - Map task converts the set of data into key-value pairs.
 - ② Reduce - Reduce task shuffles the key-value pairs & combine the similar data to produce the desired results.