Ans Hadoop Architecture.

Hadoop is open source framework. Hadoop works on a Map Reduce programming algorithm.

Hadoop architecture mainly consist of

- Map Reduce
- HDFS
- Yarn

Map Reduce: Map Reduce is used to perform the distributed processing in parallel in Hadoop cluster which makes Hadoop working so fast.

It has mainly two task

1) Map Task
2) Reduce Task

First phase is Map and then next phase is Reduce

Input is provided to Map() function

→ Map Task
- Record Reader
- Map
- Combiner

→ Reduce Task
- Sort
- Reduce
- Output

* HDFS

Hadoop Distributed file System is utilized for storage permission is Hadoop Cluster. HDFS is designed in such a way that it can store large data.

HDFS follow master - slave architecture

1.) Name Node (Master)
2.) Data Node (Slave)

Name Node: It works as a master in Hadoop cluster It is used to store Meta data. Namenode instructs the data node with the operation like create, replicate.

Data node: Data node works as a slave node. They are mainly utilized for storing the data in a Hadoop cluster.

file block in HDFS: Data is stored always in form of block. Block data is divided in multiple blocks of 128MB.

* Yarn (Yet Another resource Negotiator)

Yarn is a framework on which map reduce works. Yarn performs 2 operations that are Job scheduling and resource management. Job scheduler keeps track of which job has more priority, dependencies. job timing etc. Resource manager manages all the resources.