1. HADOOP IN LAYMAN’S TERM:

According to layman Hadoop has two major components , they are (i) HDFS and (ii) Map reduce

HDFS is used for storing whereas Mapreduce is used for processing.

In mapreduce , we have two major technology they are

1. Pig
2. Hive

* Pig
* hive

Map reduce

(Processing)

HDFS

(Storage)

Hadoop components

1. HADOOP COMPONENTS:

The Hadoop Ecosystem comprises of 4 core components –

**1) Hadoop Common-**

Apache Foundation has pre-defined set of utilities and libraries that can be used by other modules within the Hadoop ecosystem. For example, if HBase and Hive want to access HDFS they need to make of Java archives (JAR files) that are stored in Hadoop Common.

### ****2) Hadoop Distributed File System (HDFS)**** -

The default big data storage layer for Apache Hadoop is [HDFS](https://www.dezyre.com/hadoop-course/hdfs). HDFS is the “Secret Sauce” of Apache Hadoop components as users can dump huge datasets into HDFS and the data will sit there nicely until the user wants to leverage it for analysis. HDFS component creates several replicas of the data block to be distributed across different clusters for reliable and quick data access. HDFS comprises of 3 important components-NameNode, DataNode and Secondary NameNode. HDFS operates on a Master-Slave architecture model where the NameNode acts as the master node for keeping a track of the storage cluster and the DataNode acts as a slave node summing up to the various systems within a Hadoop cluster.

## **3) MapReduce- Distributed Data Processing Framework of Apache Hadoop**

[MapReduce](https://www.dezyre.com/hadoop-course/mapreduce) is a Java-based system created by Google where the actual data from the HDFS store gets processed efficiently. MapReduce breaks down a big data processing job into smaller tasks. [MapReduce](https://www.dezyre.com/article/hadoop-mapreduce-vs-apache-spark-who-wins-the-battle/83) is responsible for the analysing large datasets in parallel before reducing it to find the results. In the Hadoop ecosystem, Hadoop MapReduce is a framework based on YARN architecture. YARN based Hadoop architecture, supports parallel processing of huge data sets and MapReduce provides the framework for easily writing applications on thousands of nodes, considering fault and failure management.

## **4)YARN**

[YARN](https://www.dezyre.com/article/hadoop-2-0-yarn-framework-the-gateway-to-easier-programming-for-hadoop-users/84) forms an integral part of Hadoop 2.0.YARN is great enabler for dynamic resource utilization on Hadoop framework as users can run various Hadoop applications without having to bother about increasing workloads.

Data integration

(SQOOP/REST/ODBC)

Query

(HIVE)

Machine learning

(MAHOUT)

Scripting

(pig)

Distributed processing

(MAP REDUCE)

Nosql database

(HBASE)

Distributed Storage

(HDFS)

Workflow

And scheduling

(OOZIE)

Coordination

(ZOO

KEEPER)

Management and monitoring

(AMBARI)

1. REASON TO LEARN HADOOP:
2. **No signs of slowing down.** the Big Data revolution will continue to grow.
3. Personal reasons like newjobs,skill development, supply and demand.
4. Demand for bigdat is extremely high and it is difficult to prove your skill .
5. Facebook,yahoo,Apache hbase,adobe Mozilla,ebay,nokia,huawie etc are using bigdata.
6. IBM,cisco and oracle together announces 26,488 vacany that required bigdata expertise in past tweleve months
7. 64% of it hiring managers requires bigdata knowledge.