Main components of HADOOP ecosystem are –

1. Hadoop Common- Pre-defined set of libraries and utilities that can be used by hadoop ecosystem as defined by Apache.
2. HDFS – Known as Hadoop distributed filesystem. It is the big data storage layer for Hadoop. Huge amount of data can be stored in HDFS and latter on retrieved for analysis. HDFS will create replicas of data block which will be distributed across different cluster for reliable and quick data access. HDFS consists of 3 components – Name Node, Data Node and Secondary Name Node. Name Node is the master node which keeps tracks of data clusters. Data Node is the node which actually stores the data. Secondary Name Node is used when Name Node is down.
3. MapReduce – MapReduce breaks down big data processing job in to smaller tasks. It is responsible to do processing of large dataset in parallel and then reduce it to final result. Map job sends a query for processing to different clusters and reduce collects output from all the map job and process them to give the final result.
4. YARN – It keeps track of all the resource across clusters. Each application asks resource manager what resources it has and then process accordingly. It can also be used to monitor mapreduce tasks. It can be used to write application to run across distributed architecture.
5. PIG – It runs mapreduce operation across datasets. It gives SQL like syntaxes to write mapreduce code.
6. HIVE – It is similar to PIG. It allows to create tables and load external files in tables using SQL. Then it creates mapreduce jobes in java.
7. SPARK – It does the same thing as HADOOP. It runs calculations on data and store results across a distributed file system. Deep down it also uses HDFS to store data.