Architecture of Hadoop are as follow:

HDFS (Hadoop Distributed File System) :- It is the storage component of Hadoop that stores data in the form of files.Each file is divided into blocks of 128MB (configurable) and stores them on different machines in the cluster. It has a master-slave architecture with two main components: Name Node and Data Node. Name node is the master node and there is only one per cluster. Its task is to know where each block belonging to a file is lying in the cluster. Data node is the slave node that stores the blocks of data and there are more than one per cluster. Its task is to retrieve the data as and when required. It keeps in constant touch with the Name node through heartbeats.

MapReduce :- To handle Big Data, Hadoop relies on the MapReduce algorithm introduced by Google and makes it easy to distribute a job and run it in parallel in a cluster. It essentially divides a single task into multiple tasks and processes them on different machines.

YARN :- YARN or Yet Another Resource Negotiator manages resources in the cluster and manages the applications over Hadoop. It allows data stored in HDFS to be processed and run by various data processing engines such as batch processing, stream processing, interactive processing, graph processing, and many more. This increases efficiency with the use of YARN.

HBase :- HBase is a Column-based NoSQL database. It runs on top of HDFS and can handle any type of data. It allows for real-time processing and random read/write operations to be performed in the data.

Pig :- Pig was developed for analyzing large datasets and overcomes the difficulty to write map and reduce functions. It consists of two components: Pig Latin and Pig Engine. Internally, the code written in Pig is converted to MapReduce functions and makes it very easy for programmers who aren't proficient in Java.

Hive:- Hive is a distributed data warehouse system developed by Facebook. It allows for easy reading, writing, and managing files on HDFS. It has its own querying language for the purpose known as Hive Querying Language (HQL) which is very similar to SQL. This makes it very easy for programmers to write MapReduce functions using simple HQL queries.

Sqoop:- A lot of applications still store data in relational databases, thus making them a very important source of data. Therefore, Sqoop plays an important part in bringing data from Relational Databases into HDFS

Flume:- It is an open-source, reliable, and available service used to efficiently collect, aggregate, and move large amounts of data from multiple data sources into HDFS. It can collect data in real-time as well as in batch mode. It has a flexible architecture and is fault-tolerant with multiple recovery mechanisms.

Kafka:- There are a lot of applications generating data and a commensurate number of applications consuming that data. But connecting them individually is a tough task. That's where Kafka comes in. It sits between the applications generating data (Producers) and the applications consuming data (Consumers).

Oozie:- Oozie is a workflow scheduler system that allows users to link jobs written on various platforms like MapReduce, Hive, Pig, etc. Using Oozie you can schedule a job in advance and can create a pipeline of individual jobs to be executed sequentially or in parallel to achieve a bigger task.

Zookeeper:- In a Hadoop cluster, coordinating and synchronizing nodes can be a challenging task. Therefore, Zookeeper is the perfect tool for the problem. It is an open source, distributed, and centralized service for maintaining configuration information, naming, providing distributed synchronization, and providing group services across the cluster.