A Hadoop cluster is a group of interconnected computers (nodes) that work together to store and process large-scale datasets using the Hadoop Distributed File System (HDFS) and the MapReduce programming paradigm.

The cluster is the basis of the Apache Hadoop framework and makes it possible to store and process data in a distributed way.

Components of a Hadoop Cluster

1. NameNode: The NameNode is the central component of the Hadoop cluster. It manages the file system namespace and metadata, including the mapping of data blocks to DataNodes (data storage nodes). The NameNode keeps track of the location of data blocks in the cluster and ensures data reliability through replication.
2. DataNodes: DataNodes are worker nodes responsible for storing and managing the actual data in the Hadoop cluster. Each DataNode stores one or more data blocks and communicates with the NameNode to report the status of the data blocks it manages.
3. Secondary NameNode (Deprecated): The Secondary NameNode assists the primary NameNode by periodically merging changes to the file system's metadata. It is not a backup or failover NameNode. Note that the Secondary NameNode has been deprecated in Hadoop 2.x and replaced with the HDFS High Availability (HA) feature.
4. ResourceManager: The ResourceManager is responsible for resource allocation and job scheduling in the YARN (Yet Another Resource Negotiator) framework, which manages resources across the Hadoop cluster. It receives job requests from clients, negotiates resources with NodeManagers, and monitors job progress.
5. NodeManagers: NodeManagers run on each individual node in the cluster and manage the resources (CPU, memory) on that node. They are responsible for launching and monitoring containers that run application tasks (including MapReduce tasks) on the

cluster.

6. JobTracker (Deprecated): In older versions of Hadoop (Hadoop 1.x), the JobTracker was responsible for resource management and job scheduling in the MapReduce framework. However, in Hadoop 2.x and later, the JobTracker has been replaced with the ResourceManager and ApplicationMaster.
7. ApplicationMaster: The ApplicationMaster is responsible for managing the lifecycle of a specific application running on the Hadoop cluster. Each application (e.g., a MapReduce job or a Spark application) has its own ApplicationMaster, which negotiates resources with the ResourceManager and coordinates the tasks across NodeManagers.