# Big data Hadoop, HDFS, YARN, ApplicationMaster, others

Before we start digging on big data mystery, we’d better understand some major concepts. Let’s read the explanation below. Here I just give a quick overview how we can use Hadoop. The new version of Hadoop has two components HDFS and YARN. The rest of tools are built on the top of YARN. YARN is the big architecture change. It isolates the HDFS(resources) from the detail logic(MapReduce, hive, pig, etc.). therefore, there are two steps to make your code working on distributed environment. the first step is to assign the data to HDFS cluster. The second step is to override the Map+Reduce algorithm if you chose MapReduce solution or create a schema on Hive and use query to access the data. The query will be automatically converted to MapReduce jobs no longer to write any MapReduce code.

Hadoop is divided into two core components

* HDFS - a distributed file system
* YARN - a cluster resource management technology