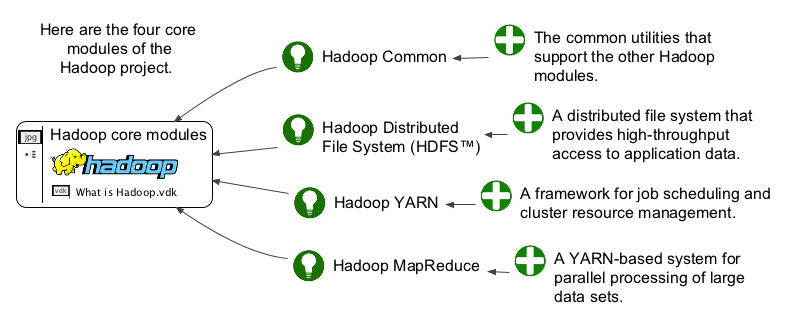
**Hadoop in layman's term, Components and why Hadoop:**

In simple terms, Hadoop is a collection of open source programs relating to Big Data analysis which is freely available for use, reuse and modify for anyone who is interested in it. Big Data scientists call Hadoop the ***backbone*** of their operations. They realized that reading data from bulk storage devices took longer than reading it from small storage devices of multiple numbers working simultaneously.  
  
**Hadoop is named after the toy elephant belonging to the son of one of the key developers.**



Mainly Hadoop comprises of 4 modules:

* Hadoop Common
* HDFS
* YARN
* Map Reduce.  
   Each module has a specific task assigned to it for facilitating Big Data Analytics.

**Hadoop Common** provides Java based user tools that are to be used for accessing and retrieving data stored in a Hadoop file system.

**Hadoop Distributed File System** enables data to be stored in a form that can be easily accessed   
and retrieved. The data will be stored across several interconnected devices which can be reached using Map Reduce.

**Map Reduce** basically does two primary functions: It reads data from databases and puts them into a format that is suitable for Big Data Analytics. Secondly, it breaks down the data into meaningful information that can be used for interpretation.

*At least 90% of the Fortune 500 Companies have integrated****Hadoop tutorials****and training program for their engineers to make better use of Big Data.*

**YARN** manages the system resources when the analytics are being conducted on the data stored in linked devices.

**Benefits of Hadoop 2.0 YARN**

* It offers improved cluster utilization
* Highly scalable
* Beyond Java
* Novel programming models and services
* Agility

## **Data Access Components of Hadoop Ecosystem-**[**Pig and Hive**](https://www.dezyre.com/article/difference-between-pig-and-hive-the-two-key-components-of-hadoop-ecosystem/79)

[**Apache Pig**](https://www.dezyre.com/hadoop-course/pig) is a convenient tools developed by Yahoo for analysing huge data sets efficiently and easily. It provides a high level data flow language Pig Latin that is optimized, extensible and easy to use. The most outstanding feature of Pig programs is that their structure is open to considerable parallelization making it easy for handling large data sets.

[**​ Hive**](https://www.dezyre.com/hadoop-course/hive) developed by Facebook is a data warehouse built on top of Hadoop and provides a simple language known as HiveQL similar to SQL for querying, data summarization and analysis. Hive makes querying faster through indexing.

**​Zookeeper** is the king of coordination and provides simple, fast, reliable and ordered operational services for a Hadoop cluster. Zookeeper is responsible for synchronization service, distributed configuration service and for providing a naming registry for distributed systems.

**Main reasons why Big data Technology:**

* It is flexible. More data systems can be added, edited or deleted when required.
* It is cost-effective and practical. More storage units can be added by procuring readily-available storage from IT vendors.
* It is open source, providing ample flexibility for corporations to customize it the way they want for effective use. Unlike bespoke off-shelf software systems that are rigid and complex to customize.
* Commercial versions like [Cloudera](https://www.cloudera.com/)are available in the market which further simplify the process of installing and setting up the Hadoop framework.