

# Introduction to Apache Hadoop



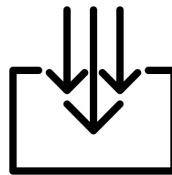
Apache Hadoop is an open source framework for storing and processing data in parallel and running applications in a distributed manner.



Replication



Higher  
Scalability



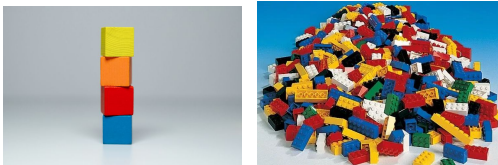
Higher  
Concurrency



Fault  
Tolerance



Open source



Structured &  
Unstructured data



Ecosystem

*MapReduce* is a programming model for writing applications that can process *Big Data* in parallel on multiple nodes.



**Hadoop Common** refers to the collection of **common** utilities and libraries that support other **Hadoop** modules.

Apache Hadoop YARN is the resource management and job scheduling technology in the open source Hadoop distributed

MAPREDUCE

**HDFS** is a distributed file system that handles large data sets running on commodity hardware.

YARN

HDFS

HADOOP COMMONS

