![TutorialKart]

# Apache Hadoop Tutorial – Learn Hadoop Ecosystem with Examples

Apache Hadoop Tutorial – Learn Hadoop Ecosystem to store and process huge amounts of data with simplified examples.

## What is Hadoop ?

Hadoop is a set of big data technologies used to store and process **huge amounts of data**. It is helping institutions and industry to realize big data use cases. It is designed to run on data that is stored in **cheap** and old commodity hardware where hardware failures are common. It is **flexible** in such a way that you may scale the commodity hardware for **distributed processing**.
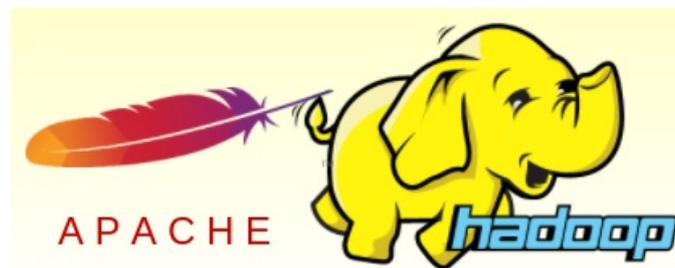


## Need for Hadoop

We shall start with the data storage. Traditional Relational Databases like MySQL, Oracle etc. have limitations on the size of data they can store, scalability, speed (real-time), running sophisticated machine learning algorithms, etc . These limitations could be overcome, but with a huge cost. Also, as the organizational data, sensor data or financial data is growing day by day, and industry wants to work on Big Data projects,

## Prerequisites to learn Hadoop

Following are the concepts that would be helpful in understanding Hadoop :

- Relational Database – Having an understanding of Queries (MySQL)
- Programming Languages – Java, Python
- Basic Linux Commands (like running shell scripts)

## Kinds of Data Hadoop deals with !

Hadoop is a good fit for **data that is available in batches**, the data batches that are inherent with behaviors. A good example would be medical or health care. Most of the wearable and smart phones are becoming smart enough to monitor your body and are gathering huge amount of data. These data have patterns and behavior of the parameters hidden in them. Finding out these behaviors and integrating them into solutions like medical diagnostics is meaningful.

Hadoop is not a good fit for mission critical systems. They ought to be kept in the traditional Relational Database systems. And it has to be noted that Hadoop is not a replacement for Relational Database Management Systems. It is only a choice based on the kind of data we deal with and consistency level required for a solution/application.

## Use Cases for Hadoop

- Data Analytics
- Threat Analysis
- Finding Trends

## What does Hadoop consists of ?

Hadoop consists of following two components :

1. HDFS (Hadoop File System) – An Open-source data storage File System.
2. MapReduce – The Processing API

When a Hadoop project is deployed in production, some of the following projects/libraries go along with the standard Hadoop.

- HBase
- Hive
- Pig

## Databases for Hadoop

Following are the list of database choices for working with Hadoop :

- XML
- HDFS (an alternative file system that Hadoop uses)
- HBase
- MongoDB

## Apache Hadoop Tutorial

We shall provide you with the detailed concepts and simplified examples to get started with Hadoop and start developing Big Data applications for yourself or for your organization.

### As a first step, Setup Hadoop on your computer.

Install Hadoop on your Ubuntu Machine – Apache Hadoop Tutorial

Install Hadoop on your MacOS – Apache Hadoop Tutorial

## With Hadoop installed on your computer, we shall learn about the components of Hadoop

HDFS

MapReduce 1.0

What is new in MapReduce 2.0

## Writing Hadoop applications

Word Count Example Program

## Tune your Hadoop applications

Tune MapReduce

## Hadoop Interview Questions

Most Frequently asked Hadoop Interview Questions

Top 10 Hadoop Interview Questions

Top 25 Hadoop Interview Questions

Top 50 Hadoop Interview Questions

### Learn Apache Hadoop

⊩ Hadoop Tutorial

⊩ Install Hadoop on Ubuntu

⊩ Hadoop MapReduce 1.0