



## Hadoop Big Data Analytics

This course is intended to give a holistic understanding of Hadoop Big Data Analytics. The workshop will cover

- Introduction to Hadoop
- What is BigData
- Motivation for Hadoop
- Hadoop Demons: name node, data node, job tracker and task tracker
- Perform Hadoop essential commands ( data in and out of hdfs)
- Flume- ecosystem \_ twitter analysis
- Perform twitter analysis ( capture live tweets with desired keywords)
- Difference between Hive and traditional databases
- Hbase- ecosystem

# Hadoop Big Data Analytics

---

## OBJECTIVE

This course is recommended as an advancement to Processing BigData on Hadoop. While Processing BigData was geared more towards developers who wanted to understand how to use Hadoop to handle and arrange BigData. This course will go on to teach how to perform analytical operations to gain insights from data processed through Hadoop. BigData Analytics on Hadoop will teach you all you need to learn about BigData Analytics on Hadoop. Big Data is a popular term used to describe the exponential growth, availability and use of information, both structured and unstructured. It is imperative that organizations and IT leaders focus on the ever-increasing volume, variety and velocity of information that forms Big Data. Hadoop is the core platform for structuring Big Data, and solves the problem of making it useful for Analytics.

## Why is Big Data Analytics on Hadoop important?

- Businesses are now aware of the large volumes of data that they generate in their day to day transactions. They have also realized that this Big Data can provide very valuable insights once analyzed
- the massive volume of Big Data and its unstructured format make it difficult to analyze Big Data. Hadoop brings the ability to cheaply process large amounts of data, regardless of structure.
- Knowledge about Big Data Analytics on Hadoop will also prove to be a huge Resume builder for Students who are aiming to work in the IT Industry
- if you are a developer who is uncertain about how Big Data Analytics can be performed on Hadoop, this course will clear things up and save you a lot of time and effort
- if you are a business that is planning to shift to Hadoop, then this is the right course to train your employees.
- Big Data Analytics on Hadoop will prove to be an answer to many questions at once.
- The session will be handled by very experienced trainers who not only have immense knowledge but are also loaded with valuable experience

## HARDWARE AND SOFTWARE

1. Participants should bring their laptop (preferably Windows 7 or higher/ Mac OS installed).
2. Operating System (any of the following):
  - Mac OS X with [XQuartz](#)
  - Windows (Version XP or later) is required.
3. Minimum 4 GB RAM on the system is advisable.

## PRE-REQUISITE

1. Participants should have basic knowledge of Java, basic programming skills and should be in able to understand the scripting language.
2. High-speed Internet connection will be provided at the training venue.

## COURSE OUTLINE

### Day 1: Hadoop Foundation

#### Session 1:

- **Module 1 : Introduction to Hadoop**
- **Module 2 : What is Big Data**
- **Module 3 : Motivation for Hadoop**
- **Module 4 : Hadoop Distribute file system**
- **Module 5 : How Hadoop works**
- **Module 6 : Hadoop Demons: name node, data node, job tracker and task tracker**

#### Session 2:

- **Lab 1 : Deploy single node Hadoop**
- **Lab 2 : Test your Hadoop clusters run**
- **Lab 3 : Perform Hadoop essential commands ( data in and out of hdfs)**

### Day 2: Hadoop on premise

#### Session 1:

##### **Module 7: Flume- ecosystem \_ twitter analysis**

- **What is Flume**
- **Install and configure Flume**
- **Perform twitter analysis ( capture live tweets with desired keywords)**

#### Session 2:

##### **Module 8: Hive**

- **What is Hive**
- **Difference between Hive and traditional databases**
- **Data types in Hive**
- **Install and configure Hive**
- **Interact with hive**

#### Session 3:

##### **Module 9: Hbase- ecosystem**

- **What is Hbase**
- **Install and configure Hbase**
- **Run sample program of Hbase**

# Hadoop Big Data Analytics

---

## COURSE SCHEDULE

### Day 1: Hadoop Foundation

The day will primarily cover introduction to Hadoop, Big Data and Hadoop Demons.

Topic	Session	From	To
<b>Introduction to Hadoop, What is Big Data</b>	1	9 AM	10:15 AM
<b>Hadoop Distribute file system</b>	1	10:30 AM	11:45 AM
<b>Hadoop Demons... cont.</b>	1	12:00 PM	1:15 PM
<b>Lab 1,2,3 : Perform Hadoop essential Commands.... Cont.</b>	2	2:15 PM	4:30 PM

### Day 2: Hadoop on Premise

Day is primarily devoted to concept building Flume, Hive and H Base – Ecosystem

Topic	Session	From	To
Module 7: Flume- ecosystem _ twitter analysis...cont	1	9 AM	11:00 AM
Hive...cont.	2	11:15 AM	1:00 PM
Hbase- ecosystem...cont	3	2:00 PM	4:30 PM

# Hadoop Big Data Analytics

---

## ABOUT INSTRUCTOR



**Mr. Gowtham Balaji** has 6+ years of rich experience in Internet of Things (IOT), Hadoop Big Data Analytics, Data Science and Cloud Computing. He is a certified trainer from IBM.

**Professional Training Experiences (Colleges and Corporate):**

He has conducted many training programmes in reputed institutions for the benefit of student community as well as Faculty Development Programmes like BITS Pilani Hyderabad, IIT Bhilai etc., As part of executive development on the technical advancement he engaged on corporate organizations training. Professional guidance with practical approach made successful delivery is his value addition to the participants. He has successfully delivered many training through MSME.

