



Introduction to Big Data

Introduction

Big data is a hot trend and everyone seems to be talking about it, big data can be characterized as data that has high volume, high variety and high velocity. Data includes numbers, text, images, audio, video, or any other kind of information you might store on your computer.

Big Data is the term for a collection of datasets so large and complex that they become difficult to process using on-hand database management tools or traditional data processing applications. The challenges include capture, curation, storage, search, sharing, transfer, analysis, and visualization.

Big Data examples includes social media network analyzing their members data to learn more about them and connect them with content and advertising relevant to their interests, or search engines looking at the relationship between queries and results to give better answers to users questions.

One of the best known methods for turning raw data into useful information is by what is known as MapReduce. MapReduce is a method for taking a large data set and performing computations on it across multiple computers, in parallel. It serves as a model for how program, and is often used to refer to the actual implementation of this model. MapReduce consists of two parts: The Map function does sorting and filtering, taking data and placing it inside of categories so that it can be analyzed. The Reduce function provides a summary of this data by combining it all together.



Objectives

- In This Course You will learn an introduction to data formats, technologies, and techniques. Fundamentals of Relational Databases. Fundamentals of NoSQL Databases. An overview of principles and technologies for working with Big Data.

Duration: 8 hours

Target Audience

This Course targets audience are:

- IT managers looking to better manage data analysis. Businesses looking to organize and analyze large amounts of vital data in order to improve business insights.
- Managers who wants to reach business goals and improve agility.
- IT professionals looking to implement new data analysis tools.
- Data Analyst, SQL Developers, Database Administrator, Database Developers, IT Professionals, and any other person who would like to get familiar with big data technology.

Prerequisites

- Internet Access
- Basic Computer Skills



Contents – Day 1

- **Introduction to Big Data**
- **Fundamentals of Databases**
- **Foundations for Big Data**
- **Storing Big Data**
- **Processing Big Data**
- **Tools and Techniques to Analyze Big Data**
- **Developing a Big Data Strategy**
- **Implementing a Big Data Solution**
- **Introduction to the Google Approach**
 - **What is a cluster**
 - **What is a Node**
 - **The Google File System -GFS**
 - **Google's Big Table**
- **MapReduce Introduction**
- **Apache Hadoop Introduction**
- **Summary**