

Share on your Social Media



# Big Data Analytics Tutorial

Published On: July 3, 2024

## Big Data Analytics Tutorial

Every day, the amount of data grows. Learning about big data analytics has become necessary today. With the help of this big data analytics tutorial, we are going to explain its fundamentals.

[Big Data Analytics Tutorial PDF](#)

## Introduction to Big Data Analytics

**Big Data:** "Big Data" refers to an approach for analyzing vast and varied data collections. It finds essential information such as market trends, user preferences, hidden patterns, and unidentified correlations.

**Big Data Analytics:** To extract insights from massive datasets, it makes use of sophisticated analytics techniques like statistical analysis, machine learning, data mining, and predictive modeling. Making sense of the deluge of data available to us now is the difficulty. This is the context in which big data analytics are useful.

## Use Cases of Big Data Analytics

Big data analytics helps businesses use the enormous volumes of data at their disposal and

## Featured Articles

 **Want to know more about becoming an expert in IT?**

[Click Here to Get Started](#) >>

100% Placement Assurance

AUTHORISED CERTIFICATION PARTNER

IBM

Quick Enquiry

## Related Courses at SLA

- **Big Data Analytics Training in OMR**
- **Big Data Analytics Training in Chennai**

## Related Posts



**C and C++ Tutorial**  
Published On: August 1, 2024

transform it into insights that can be used to spur innovation and commercial expansion.

Big Data Analytics aims to help businesses in the following ways:

- Make better business decisions
- Boost productivity
- Enhance customer satisfaction and services
- Ensure particular sectors survive in a cutthroat global marketplace.

## Steps Involved in the Big Data Analytics

Big Data analytics is an effective technology that helps unlock the potential of enormous and intricate information. To enhance comprehension, let us dissect it into essential steps:

**Data Collection:** Numerous sources, including social media, sensors, online platforms, business transactions, website logs, and others, are used to gather data.

The data can be classified as follows:

- Unstructured (text documents, images, and videos)
- Semi-structured (log files)
- Structured (predefined organizations, such as databases).

**Data Cleaning:** Replacing missing data, fixing errors, and getting rid of duplicates are all part of data cleaning and pre-processing. This phase contains the following:

- Proceed to clean up the gathered data
- Ensuring it is appropriate for analysis and free of errors.

In most cases, errors, missing numbers, inconsistencies, and noisy data can be found in the collected raw data. Finding and fixing flaws in data is the process of cleaning it up so that it is reliable and consistent.

C and C++ Tutorial C is a high-level, procedural, general-purpose programming language. Whereas C++, a...



### ASP DOTNET Tutorial

Published On: July 31, 2024

ASP DOTNET Tutorial Microsoft created the web framework known as ASP.NET. It is employed in...



### Artificial Intelligence Tutorial

Published On: July 30, 2024

Artificial Intelligence Tutorial Artificial intelligence (AI) is significant since it enhances many facets of society...



### Appium Testing Tutorial

Published On: July 30, 2024

Appium Testing Tutorial Designed to make the UI automation of many app platforms easier, Appium...

To prepare the data for additional analysis, pre-processing procedures may also include feature extraction, normalization, and data transformation.

**Data Analysis:** Various methods and algorithms are employed to analyze data and extract valuable insights.

This step covers the following:

- Prescriptive analytics (making decisions or recommendations based on the analysis)
- Diagnostic analytics (finding patterns and relationships)
- Predictive analytics (predicting future trends or outcomes)
- Descriptive analytics (summarizing data to better understand its characteristics)

**Data Visualization:** Using interactive dashboards, graphs, and charts to display data visually is a phase in the process.

To improve the clarity and usability of data analysis insights, data visualization techniques are employed to visually represent the data using charts, graphs, dashboards, and other graphical formats.

**Interpretation and Decision Making:** After gaining insights through data analytics and visualization, stakeholders evaluate the results to make well-informed decisions.

This step involves the following:

- Developing new goods and services
- Improving customer experiences
- Streamlining business operations
- Guiding strategic planning.

**Data Storage and Management:** Now, the data needs to be kept in a format that makes it simple to access and analyze.

Large volumes of data may be too much for traditional databases to handle, which is why many businesses choose cloud-based storage options like *Amazon S3* or distributed storage systems like *Hadoop Distributed File System (HDFS)*.

**Continuous Learning and Improvement:** The practice of continuously gathering, purifying, and evaluating data to find undiscovered insights is known as big data analytics. It gives companies a competitive edge and aids in improved decision-making.

[\*\*Big Data Analytics Syllabus PDF\*\*](#)

## Types of Big Data Analytics

Typical forms of big data analytics include the following:

### Descriptive Analytics

In business-related datasets, descriptive analytics returns insights such as *"What takes place in my business?"*

Overall, this helps create reports that include a company's income, profit, and sales data by summarizing previous information. It also helps with social media metrics tabulation. It is capable of complete, accurate, real-time data processing and powerful visualization.

### Diagnostic Analytics

Using data, diagnostic analytics finds the underlying causes. It responds by asking, *"Why is it happening?"*

**Example:** Drill-down, data mining, and data recovery are a few typical examples.

Organizations utilize diagnostic analytics because they provide a thorough insight into a particular problem. It can identify the underlying reasons and

separate any contradicting data.

## **Predictive Analytics**

To predict future events, this type of analytics examines data from both the past and the present. So, it responds with something like, *“What will happen in the future?”*.

Predictive analytics looks at current data and makes predictions using machine learning, artificial intelligence, and data mining. It is capable of determining trends in the market, in customers, and so on.

**Example:** PayPal sets the guidelines that Bajaj Finance must abide by to protect its clients from fraudulent transactions.

## **Prescriptive Analytics**

With the help of perspective analytics, one can formulate a strategy decision and get a response to the question, *“What do I need to do?”*

Descriptive and predictive analytics are both compatible with perspective analytics. It depends on AI and machine learning for the most part.

**For example:** Perspective Analytics uses a set of algorithms in the airline sector that automatically adjust ticket pricing in response to demand.

## **Big Data Analytics Interview Questions**

## **Tools and Technologies of Big Data Analytics**

The following are a few often-used big data analytics tools:

### **Hadoop**

Hadoop is a framework that facilitates large data analytics and makes big data management possible. It is the best tool for storing and analyzing

big data.

## **MongoDB**

It's a database created specifically to handle, access, and store vast amounts of unstructured data. It is best for handling unstructured data.

## **Talend**

Talend's integrations with Hadoop, Spark, and NoSQL databases facilitate the efficient processing and analysis of vast amounts of data by organizations. This is best for managing and integrating data.

## **Cassandra**

Massive volumes of data are managed via the open-source distributed NoSQL database management system Cassandra across multiple commodity computers. It is best for handling data chunks.

## **Spark**

Apache Spark is a prominent distributed computing framework in e-commerce, banking, healthcare, and telecommunications because it offers a single platform for big data analytics. It is utilized for processing and analyzing massive volumes of data in real-time.

## **Storm**

Apache Storm enables organizations to process and analyze real-time data streams on a massive scale. It is best for a variety of use cases in sectors like banking, telecommunications, e-commerce, and the Internet of Things.

## **Kafka**

To effectively satisfy their data processing needs, organizations can build scalable, fault-tolerant, real-time data pipelines and streaming applications using Apache Kafka, a flexible and

potent event streaming platform. This distributed streaming infrastructure makes fault-tolerant storage possible.

## Big Data Analytics – Characteristics

The “Five V’s,” which are frequently used to summarize the qualities of big data, are as follows:

**Volume:** Volume is the term for the vast amount of data that is created and saved every second via social media, financial transactions, videos, Internet of Things devices, and customer logs.

**Velocity:** The velocity of data has increased dramatically with the creation and use of IoT devices and real-time data streams, that can analyze data quickly to produce insightful data.

**Variety:** Big Data covers a variety of data formats, including semi-structured (JSON and XML), unstructured (text, photos, and videos), and structured (found in databases) data.

**Veracity:** It describes how accurate and reliable the data is. Three key challenges in big data analytics include ensuring data quality, resolving data conflicts, and handling data ambiguity.

**Value:** The capacity to transform massive data sets into insightful knowledge. It helps in extracting useful and applicable insights that can result in improved user experiences, new products, improved decision-making, and competitive advantages.

## Big Data Analytics Salary

## Big Data Analytics – Methodologies

The big data analytics approaches are as follows:

### Define Objectives

Clearly state the aims and objectives of the analysis.

For the procedure to be guided throughout, this step is crucial. It finds an objective for which insights you are looking for and which business problems you are trying to resolve.

## **Data Collection**

Collect pertinent information from multiple sources.

It comprises unstructured data from papers, emails, and social media, as well as semi-structured data from logs and JSON files.

## **Data Pre-Processing**

To guarantee the data's quality and consistency, it must be cleaned and pre-processed.

This entails fixing missing numbers, eliminating duplicates, correcting discrepancies, and formatting data so that it is usable.

## **Data Storage and Management**

Put the information in the proper storage system.

A NoSQL database, a standard relational database, or a distributed file system like the *Hadoop Distributed File System (HDFS)* could all fall under this category.

## **Exploratory Data Analysis (EDA)**

Finding patterns, identifying outliers, and identifying data features are all part of this step. Visualization methods, including box plots, scatter plots, and histograms, are frequently used.

## **Feature Engineering**

To boost the effectiveness of machine learning models, add new features or alter current ones. This can entail creating composite features, dimensionality reduction, or feature scaling.

## **Model Selection and Training**

Based on the characteristics of the data and the



nature of the problem, select appropriate machine learning methods. Train the models with labeled data if it is available.

## Model Evaluation

ROC curves, accuracy, precision, recall, F1-score, and recall can all be used to gauge how well the trained models perform. It helps in determining which model is most suitable for deployment.

## Deployment

Get the model up and running in a real-world setting. This can entail setting up monitoring tools, developing APIs for model inference, and integrating the model with existing systems.

## Monitoring and Maintenance

Adjust the analytics pipeline as necessary to account for evolving data characteristics or business requirements.

## Iterate

Analytics for big data is an iterative process. To make the models or procedures more accurate and efficient over time, analyze the data, get feedback, and make necessary updates.

**Big Data Analytics Training**

## Conclusion

We cover everything in our big data analytics tutorial. Learn them comprehensively in our **[big data analytics training in Chennai](#)**.

Share on your Social  
Media



## Softlogic Academy

# Softlogic Systems

### KK Nagar [Corporate Office]

No.10, PT Rajan Salai, K.K. Nagar, Chennai  
– 600 078.

**Landmark:** Karnataka Bank Building

**Phone:** [+91 86818 84318](tel:+918681884318)

**Email:** [enquiry@softlogicsys.in](mailto:enquiry@softlogicsys.in)

**Map:** [Google Maps Link](#)

### OMR

No. E1-A10, RTS Food Street  
92, Rajiv Gandhi Salai (OMR),  
Navalur, Chennai – 600 130.

**Landmark:** Adj. to AGS Cinemas

**Phone:** [+91 89256 88858](tel:+918925688858)

**Email:** [info@softlogicsys.in](mailto:info@softlogicsys.in)

**Map:** [Google Maps Link](#)

## Courses

Python

Software Testing

Full Stack Developer

Java

Power BI

Clinical SAS

Data Science

Embedded

Cloud Computing

Hardware and Networking

VBA Macros

## Navigation

[About Us](#)

[Blog Posts](#)

[Careers](#)

[Contact](#)

[Placement Training](#)

[Corporate Training](#)

[Hire With Us](#)

[Job Seekers](#)

[SLA's Recently Placed Students](#)

[Reviews](#)

[Sitemap](#)

## Important Links

[Disclaimer](#)

[Privacy Policy](#)

[Terms and Conditions](#)

## Social Media Links



## Review Sources

[Google](#)

[Trustpilot](#)

[Glassdoor](#)

[Mouthshut](#)

[Sulekha](#)

[Justdial](#)

[Ambitionbox](#)

Mobile App Development

Indeed

DevOps

Software Suggest

Sitejabber

Copyright © 2024 - Softlogic  
Systems. All Rights Reserved

SLA™ is a trademark of Softlogic Systems, Chennai.  
Unauthorised use prohibited.