



SNS COLLEGE OF TECHNOLOGY

Coimbatore-35

An Autonomous Institution

Accredited by NBA – AICTE and Accredited by NAAC – UGC with 'A++' Grade

Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai



DEPARTMENT OF MECHATRONICS ENGINEERING

UNIT 1 – INTRODUCTION TO BIG DATA

EVOLUTION OF BIG DATA

&

BEST PRACTICES FOR BIG DATA ANALYTICS



EVOLUTION OF BIG DATA



The evolution of big data is a complex and ongoing process that has unfolded over several decades. Here's a brief overview of its key stages:

1. Early Days (1960s-1980s):

1. The concept of big data can be traced back to the early days of computing when traditional databases and storage systems were introduced.
2. During this period, data was relatively small in scale compared to today, and the emphasis was on developing efficient data storage and retrieval mechanisms.

2. Data Warehousing (1990s):

1. As organizations started accumulating larger volumes of data, the need for centralized data storage and management became apparent.
2. Data warehousing emerged as a solution, where large repositories of structured data were created to support decision-making processes.

3. Internet and E-commerce (late 1990s-early 2000s):

1. The rise of the internet and e-commerce led to a significant increase in the volume of data being generated. This included user logs, clickstream data, and transaction records.
2. Companies began to recognize the value of analyzing this data to gain insights into customer behavior and business trends.



EVOLUTION OF BIG DATA



1.Hadoop and MapReduce (mid-2000s):

1. Hadoop, an open-source framework, was introduced to process and analyze large datasets using a distributed computing model.
2. MapReduce, a programming model for processing and generating large datasets, became a key component of the Hadoop ecosystem.

2.NoSQL Databases (mid-2000s):

1. The limitations of traditional relational databases in handling unstructured and semi-structured data became evident.
2. NoSQL databases, such as MongoDB and Cassandra, emerged to address the need for scalable and flexible storage solutions for diverse data types.

3.Cloud Computing (late 2000s-present):

1. Cloud platforms, like Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform, played a crucial role in the evolution of big data.
2. Cloud computing offered scalable and on-demand resources, making it easier for organizations to store and process massive amounts of data without significant upfront investments.



BEST PRACTICES FOR BIG DATA ANALYTICS



Implementing big data analytics effectively involves a combination of technological, organizational, and procedural considerations. Here are some best practices for big data analytics:

- **Define Clear Objectives:**
Clearly define the objectives and goals of your big data analytics initiative. Understand what insights or value you are seeking to derive from your data.
- **Understand Your Data:**
Conduct a thorough assessment of your data sources, including their quality, volume, velocity, and variety. Understand the structure and format of your data.
- **Scalable Infrastructure:**
Choose a scalable and flexible infrastructure, such as cloud computing services, that can handle the volume and complexity of big data. This allows you to scale resources up or down based on your needs.
- **Data Governance and Security:**
Establish robust data governance policies to ensure data quality, integrity, and security. Implement access controls, encryption, and auditing mechanisms to protect sensitive data.
- **Data Integration:**
Ensure seamless integration of data from various sources. Use ETL (Extract, Transform, Load) processes to clean, transform, and integrate data for analysis.
-



BEST PRACTICES FOR BIG DATA ANALYTICS



- **Choose the Right Tools and Technologies:**
Select appropriate tools and technologies based on your specific use case. Consider technologies like Hadoop, Spark, and NoSQL databases, as well as machine learning frameworks for advanced analytics.
- **Data Modeling and Analysis:**
Develop effective data models and analytical algorithms to derive meaningful insights from your data. Leverage statistical methods, machine learning, and other advanced analytics techniques.
- **User-Friendly Interfaces:**
Provide user-friendly interfaces for data analysts and business users. Visualization tools can help in presenting complex insights in an understandable and actionable format.
- **Collaboration and Communication:**
Encourage collaboration between data scientists, analysts, and business stakeholders. Effective communication is crucial for ensuring that insights are translated into business value.
- **Performance Monitoring:**
Implement robust monitoring and performance tracking mechanisms to identify bottlenecks, optimize queries, and ensure the overall efficiency of your big data analytics processes.
- **Regular Maintenance and Optimization:**
Regularly maintain and optimize your big data infrastructure. This includes updating software, revisiting data models, and ensuring that your system is aligned with evolving business requirements.



BEST PRACTICES FOR BIG DATA ANALYTICS



Training and Skill Development:

- Invest in training and skill development for your team. Big data analytics often requires specialized skills, and keeping your team up-to-date with the latest technologies and methodologies is crucial.

Iterative Approach:

- Adopt an iterative approach to big data analytics. Start with small, manageable projects and gradually scale up as you gain insights and experience.

Feedback and Continuous Improvement:

- Collect feedback from users and stakeholders to understand the effectiveness of your big data analytics solutions. Use this feedback for continuous improvement and optimization.