

Prepared by Dr. Wesam Ahmed

IT402
Internet of Things

Lecture 4



Data Handling and Analytics

Lecture 4

IT402

Internet of Things

Prepared by
Dr. Wesam Ahmed

04

Learning Objectives:

- 1.1 Describe the big data.
- 1.2 Explain the big data architecture
- 1.3 Gain knowledge about the big data challenges
- 1.4 The big data and IoT

04

Introduction to Big Data

Data is a collection of numbers, characters, images, or other symbols that **a computer uses to determine a specific action**. This data can be **stored or transmitted via electrical signals**. With the proliferation of **sensor networks, cyber-physical systems, and the IoT in everyday life**, there has been a significant increase in data collection. For instance, data is gathered **every time a user logs in to an online platform**, opens an app on their mobile phone, visits a website, or **conducts a search on a search engine**. The origins of Big Data can be traced back to the establishment of data centers and relational databases. However, it **gained widespread attention in the middle of the first decade of the 2000s with the emergence of online services** such as Facebook and YouTube, resulting in a significant increase in data production.

04

Introduction to Big Data

The IoT has further increased the production of Big Data by enabling the collection of data on consumer usage trends and product performance through connected devices. Additionally, advancements in machine learning have led to even more data being generated.

Big Data refers to datasets that are too vast for commonly used software tools to handle. The data has an enormous volume and is expanding rapidly, making it impossible for traditional data management systems to store or manage.

The term “Big Data” refers to techniques used for analyzing or working with datasets that are too large or complex to be processed using standard data processing applications.

04

Introduction to Big Data

Big Data analytics is a cutting-edge analytical approach that deals with massive, **diverse datasets which comprise structured, semi-structured, and unstructured data from multiple sources**. The size of these datasets **ranges from terabytes to zettabytes**. The **analysis of Big Data enables** analysts, researchers, and business users to make informed decisions quickly by utilizing previously inaccessible or underutilized data.



04

Why Big Data?

Big Data is a popular tool used by companies to improve their operations, enhance customer service, and develop targeted marketing campaigns, all with the ultimate goal of increasing revenue. By analyzing both historical and real-time data, companies can assess market requirements, understand customer needs, and adjust their products accordingly.

The **combination of Big Data and machine learning** enables companies to develop marketing plans and predict customer requirements, allowing them to respond quickly to demand and generate more revenue.

Following are several reasons why Big Data is important :

04

Why Big Data?

Following are several reasons why Big Data is important :

- **Valuable Insights:** Big Data provides organizations with the opportunity to **gain valuable insights and knowledge that can drive informed decision-making**. By analyzing large and diverse datasets, organizations can uncover patterns, trends, and correlations that were previously hidden. **These insights can be used to identify new business opportunities, optimize processes**, improve customer experiences, and make data-driven strategic decisions.
- **Real-Time Decision-Making:** With the advancements in Big Data technologies, organizations can analyze data **in real time or near real-time**. This allows businesses to make timely decisions **based on the most up-to-date information**. For example, in industries such as finance and healthcare, real-time data analysis can enable fraud detection, **risk management, and patient monitoring, leading to more proactive and effective decision-making**.

04

Types of Data

Big Data consists of various data types, high velocity and extensible. Data comes in three different forms: structured, semi-structured, and unstructured .

- **Structured:** Structured data is data that has been collected and sorted systematically and can be easily retrieved. Structured data is easy to process because it has a predefined schema.
- Examples of structured data include **financial transactions, customer information, sensor readings** with fixed attributes, and stock market data.
- To manage structured data, we can use SQL. An “Employee” table in a database as shown in Table 4.1 is an example of structured data.

04

Types of Data

Employee_ID	Employee_Name	Gender	Department	Salary in Lakhs
2365	Rajesh Kulkarni	Male	Finance	650000
3398	Pratibha Joshi	Female	Admin	650000
7465	Shushil Roy	Male	Admin	500000
7500	Shubhojit Das	Male	Finance	500000
7699	Priya Sane	Female	Finance	550000

Table 4.1.
Examples of
structured data

04

Types of Data

- **Semi-Structured:** Semi-structured data is a combination of structured and unstructured data. It contains elements of both, with some organization or metadata but without a rigid schema. Semi-structured data is often represented in formats like XML, JSON, comma-separated values (CSV), and log files.
- **Examples of semi-structured data** include web server logs, XML documents, and **data from IoT devices**.

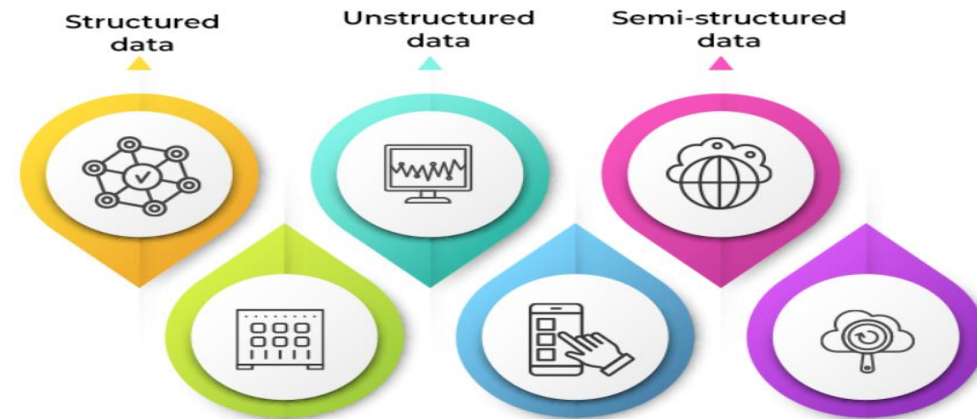
XML files or JSON documents (personal data stored in an XML file)

```
<rec><name>Prashant Rao</name><sex>Male</sex><age>35</age></rec>  
<rec><name>Seema R.</name><sex>Female</sex><age>41</age></rec>  
<rec><name>Satish Mane</name><sex>Male</sex><age>29</age></rec>  
<rec><name>Subrato Roy</name><sex>Male</sex><age>26</age></rec>  
<rec><name>Jeremiah J.</name><sex>Male</sex><age>35</age></rec>
```

04

Types of Data

- **Unstructured:** Unstructured data refers to data that **does not have a specific format or organization**. It is often in the form of **text, audio, video, images, social media posts, emails, and documents**.
- The example of unstructured data in day-to-day life are **text files and multimedia content like audio, video, images, output returned by Google searches**.



04

Big Data Architecture

Big Data architecture refers to the overall framework and components that are designed to store, process, and analyze large volumes of data. It involves a combination of hardware, software, and network infrastructure that enables organizations to effectively handle Big Data and derive meaningful insights. Big Data architecture is a method for collecting, storing, and analyzing massive amounts of data for business .

The typical Big Data architecture is as shown in Figure 4.1. The architecture should be designed to handle the challenges of Big Data—such as data volume, variety, velocity, and veracity—while providing flexibility, scalability, and security to support the organization's data-driven initiatives. A robust architecture saves money for the company by enabling trend forecasting and better decision-making.

04

Big Data Architecture

Big Data architecture comprises the following components:

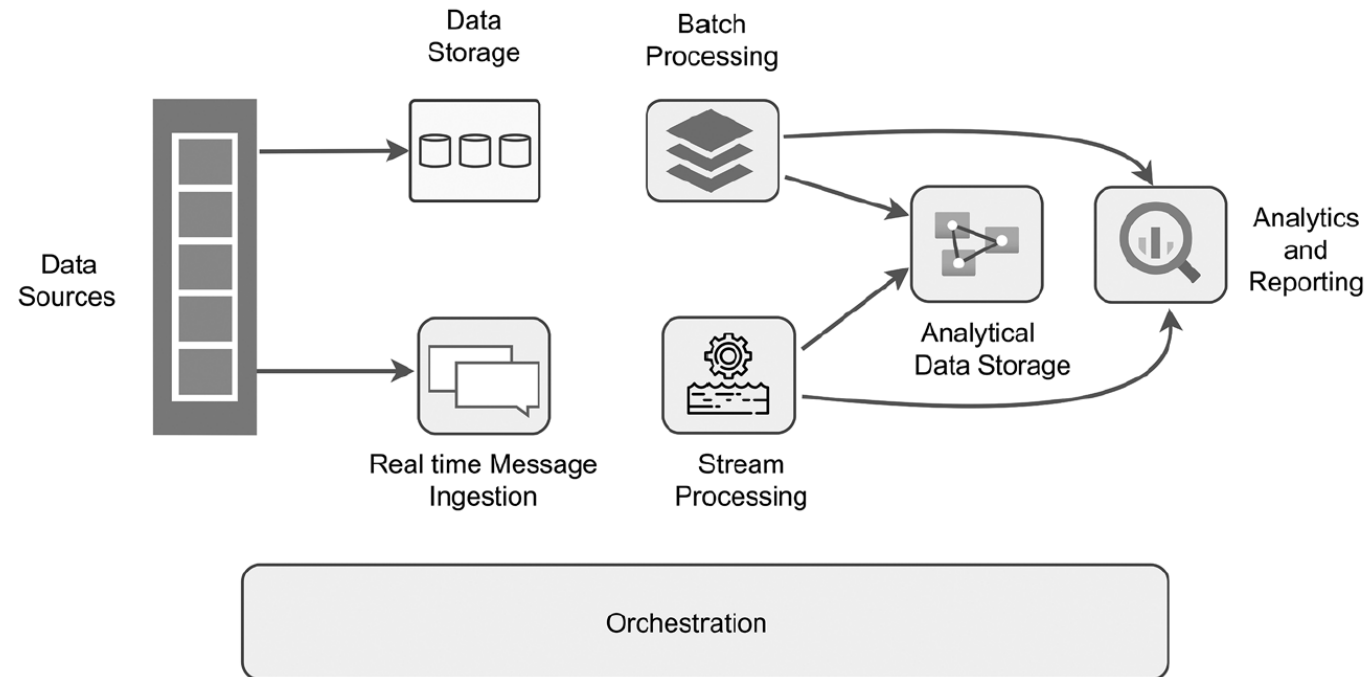


Figure 4.1
Big Data
architecture

04

Big Data Architecture

Big Data architecture comprises the following components:

- **Data Sources:** Data source refers to the location where the information is acquired. The process of Big Data begins with data sources. Data sources may differ depending on the application. There are two types of data sources: file data sources and machine data sources. A file data source is any entity that resembles a data repository, such as an email archive, file sharing. The machine data source is information created by a computer process or application activity that does not require human intervention. Examples of machine data sources are computers, mobile phones, embedded systems, and other networked devices.

04

Big Data Architecture

Big Data architecture comprises the following components:

- **Real-Time Message Ingestion:** Real-time data ingestion is the process of transmitting data from a live source into an object storage system. This process involves the storage of new messages in a data store, which serves as a message buffer and supports scalable processing. In addition to various message queuing semantics, they provide stable and reliable delivery.
- Examples include Apache Kafka, Azure Event Hubs, and Apache Flume.
-

04

Big Data Architecture

Platform	Description
Apache Kafka	Open-source distributed streaming platform for building real-time data pipelines. Handles high throughput and ensures reliable delivery.
Azure Event Hubs	Cloud-based event ingestion service by Microsoft; integrates easily with Azure storage and analytics tools.
Apache Flume	Tool for collecting, aggregating, and moving large amounts of log data to storage (like HDFS). Often used in Hadoop ecosystems.

04

Big Data Architecture

- **Batch Processing:** A batch-processing system is required for advanced analytics because it filters, aggregates, and processes data. These are frequently long batch processes that read data from storage, process it, and then output the results to new files. The preferred batch-processing approach is Apache Hadoop .
- **Stream Processing:** Real-time message intake and stream processing share many similarities. Stream processing involves the management of continuous data streams or windows and the writing of data to output sinks. **Stream processing** means analyzing and acting on data continuously as it flows into the system — *in real time* — instead of waiting for all the data to be collected first (like in batch processing).
- Popular software used for stream processing includes Apache Spark, Apache Flink.

04

Big Data Architecture

Stream Processing vs. Batch Processing

Feature	Stream Processing	Batch Processing
Data type	Continuous, real-time	Static, collected data
Processing time	Immediately as data arrives	After all data is collected
Examples	Live analytics, fraud detection, IoT monitoring	End-of-day reports, backups
Tools	Apache Flink, Kafka Streams, Spark Streaming	Hadoop, Spark Batch, MapReduce

04

Big Data Architecture

- **Analytical Data Store:** We must collect data after processing it to **analyze the entire dataset**. The analytical data store is crucial because it allows **for detailed analysis and centrally stores all our processed data**. Analyzing transactions is not ideal. Depending on our needs, it may be a cloud-based data warehouse or a relational database. **An Analytical Data Store** is the place where processed and cleaned data is collected and stored for analysis. It acts as **the central repository** where all the final, ready-to-use data lives — so analysts, data scientists, and business tools can query it efficiently.
- **Analytics and Reporting:** After importing and processing data from various data sources, we require a data analysis tool. **The data is analyzed using a variety of data analytics and visualization tools, which produce dashboards or reports after the analysis**. These reports support corporate **decision-making based on data**.

04

Big Data Architecture

- **Orchestration:** Orchestration is required to **move data between various systems**, and automation carries this out. Data ingestion, transformation, batch and stream processing, loading into an analytical data storage system, and insight-generating analysis all require repeatable workflows. **Orchestration** means **automating and managing the flow of data between different parts of the data pipeline — from ingestion → processing → storage → analytics**.
- It ensures that all **data-related tasks run in the correct order**, on schedule, and without manual intervention.

04

Characteristics of Big Data

- Big Data can exhibit characteristics commonly referred to as the “Five Vs”: volume, velocity, variety, veracity, and value, as shown in Figure 4.2.

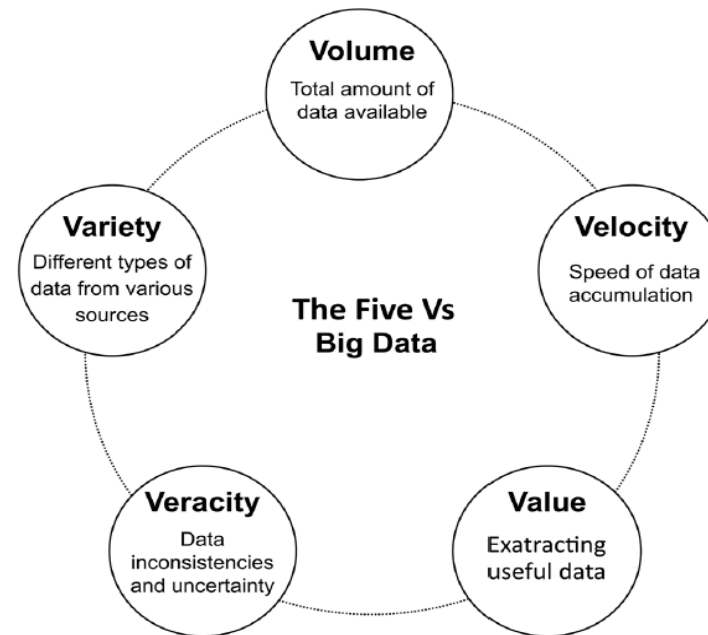


Figure 4.2 :
Characteristics
of Big Data

04

Characteristics of Big Data

- These characteristics define the **unique aspects of Big Data** and highlight the challenges and opportunities associated with it.
- Understanding the characteristics of Big Data helps **organizations adopt appropriate technologies, methodologies, and strategies** to manage, process, and analyze large volumes of data effectively.
- **Volume:** Managing this **massive quantity of information** requires **scalable storage** solutions and distributed computing frameworks. A **consistent workflow** is essential for data ingestion, **transformation, and batch and stream processing**, as well as loading data into an analytical data storage system and conducting analysis **to generate insights**.
- **Variety:** There are various ways that Big Data is produced. **Varieties of data produced by diverse sources** are referred to as being varied. **Structured, unstructured, or semi-structured** are all possible.

04

Characteristics of Big Data

Velocity: Various sources generate a vast amount of data every day. The term “velocity” refers to this unceasing, enormous flow of information . Velocity refers to the speed at which data is generated, transmitted, and processed. With the growth of real-time applications—such as IoT sensors.

Veracity: In managing Big Data, organizations need to consider the uncertainty of the data due to its inconsistency and incompleteness. The term “veracity” pertains to the degree of accuracy, quality, trustworthiness of the data .

Value: Without knowledge, data is meaningless. There is no point in just collecting and storing Big Data unless it is analyzed and provides valuable knowledge. “Data importance” is the most important thing to be addressed in the future. It represents the usefulness and actionable insights that data provides.

04

Challenges of Big Data

- Big Data refers to the **increasing volume, velocity, and variety of data generated by organizations**, individuals, and machines. **Handling, analyzing, and gaining insights from Big Data** can be challenging due to its sheer size, complexity, and diversity. To achieve **effective decision-making and BI**, it is important to address Big Data challenges such as the following:
 - **Storage and Processing**: Lack of computing power and storage space for large amounts of data.
 - **Data Quality**: Ensuring **the accuracy, completeness, and relevance of data** is crucial for maintaining data quality.

04

Challenges of Big Data

- **Data Integration**: The process of **data integration involves consolidating data from various sources** to create a unified perspective. Different systems often use **different formats** or **data structures**, making integration difficult.
- **Privacy and Security**: Ensuring the **confidentiality and protection of sensitive data**. Data breaches can lead to **financial loss, legal penalties, and loss of trust**.
- **Scalability**: The ability of a system to **handle increasing or decreasing data loads without** performance loss. As data grows, systems must scale efficiently; otherwise, performance slows down. **It is the capacity of a system to maintain performance levels when the amount of data or number of users increases**

04

Relationship Between Big Data and IoT

The business landscape is being transformed by two technological trends: Big Data and the IoT. IoT encompasses physical devices, buildings, and vehicles that are embedded with software and sensors to facilitate data exchange and connection. On the other hand, Big Data refers to the vast amounts of information produced from various sources, including IoT devices. The two trends are interdependent, with IoT generating large amounts of data that is collected by sensors and processed by centralized systems. This data includes environmental variables like temperature, humidity, and pressure used to control and monitor the devices.

Big Data analytics enables companies to analyze and derive insights from IoT-generated data. IoT data is usually unstructured and requires analysis to produce meaningful insights. Big Data analytics provides the necessary tools and techniques for processing and analyzing this data.

04

Relationship Between Big Data and IoT

The resulting insights can be used to optimize customer behavior, operational efficiency, and other business-critical areas. Another significant benefit of the IoT and Big Data is improved decision-making.

The data generated by IoT devices provides real-time insights into customer behavior, product performance, and other key areas that can affect business operations. Big Data analytics provides the tools needed to analyze this data and generate insights that can be used to make informed decisions. Furthermore, the IoT and Big Data are transforming the retail industry. Retailers are using IoT devices to track customer behavior, optimize inventory levels, and improve customer experience. IoT systems often operate in real time or near-real time environments in which data needs to be processed and analyzed in a timely manner. Big Data analytics facilitates real-time data processing, allowing organizations to capture and analyze streaming data as it is generated. This enables proactive decision-making, instant responses to events, and timely actions based on the analyzed data.

04

Relationship Between Big Data and IoT

The **integration of Big Data analytics with IoT systems** is crucial for unlocking the full potential of IoT data and realizing its value. The **relationship between Big Data and the IoT is becoming increasingly important** as organizations seek to leverage the benefits of these technologies. By combining these technologies, **companies can gain valuable insights into customer behavior, market trends, and operational processes that can improve business outcomes.**

A practical example is the use of **smart sensors in water systems across regions** like Africa and cities such as San Francisco. These systems **monitor water quality and quantity**, demonstrating that investments in smart water management can significantly reduce healthcare costs. The biggest challenges for the **faster implementation of smart water meters are the complexity of the water supply system and financial constraints .**

04

Applications of Big Data and IoT

Big Data analytics is a methodology that involves gathering, modifying, analyzing, and managing vast quantities of data to generate valuable insights. In order to gain a competitive edge and foster growth, businesses are increasingly adopting Big Data technologies for analysis. Apache Hadoop, Apache Spark, and Kafka are widely used Big Data analytics tools. The benefits of Big Data analytics are being leveraged across various sectors such as finance, retail, education, government, manufacturing, energy, and utilities.

There is wide variety of application areas of Big Data analytics, as illustrated in Figure 4.3.

04

Applications of Big Data and IoT

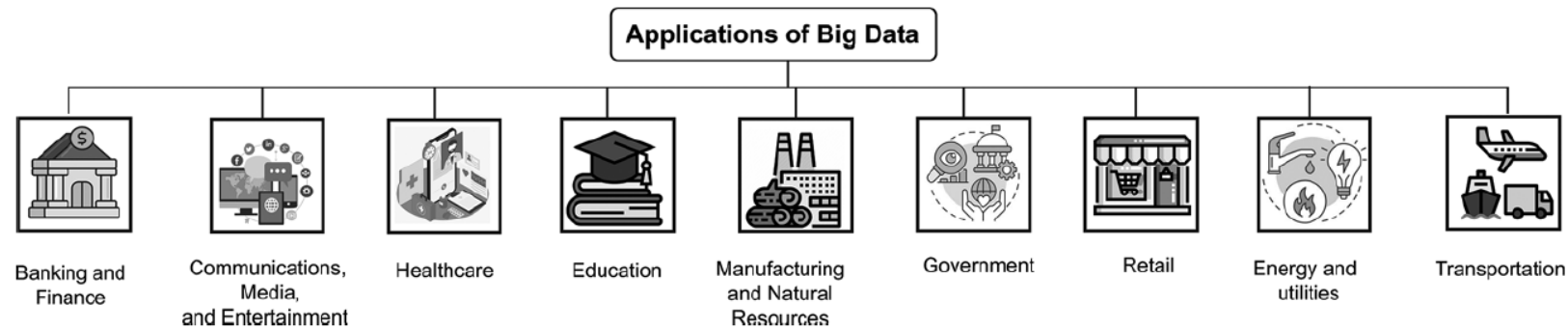


Figure 4.3:
Applications of
Big Data

04

Applications of Big Data and IoT

❑ Big Data Applications in the Banking and Finance Sector:

Big Data and IoT (Internet of Things) **have completely changed how banks and financial institutions operate. Big Data** helps banks analyze massive amounts of information — **from transactions, market data, and customer interactions. IoT** provides data from connected devices like mobile apps, ATMs, and payment systems. Together, they solve major challenges such as **fraud detection, credit risk evaluation, trading analysis, and customer data management.** By analyzing customer behavior and payment habits, **banks can enhance digital banking, optimize risk management, increase ROI, and improve customer satisfaction.**

04

Applications of Big Data and IoT

❑ Big Data Uses in Communications, Media, and Entertainment:

Big Data is transforming the **communications, media, and entertainment industry** by helping companies understand **audience preferences and trends**. By analyzing data such as **search history, watch time, social media activity, and viewer reactions**, media companies can **create and market content that aligns with audience interests**.

In addition, **Big Data and IoT technologies** enable **the development of targeted advertisements**, improving **efficiency and return on investment for advertisers**. Media producers can now use data-driven insights to **predict the popularity** of shows, actors, and storylines, allowing for smarter content creation and reducing reliance on intuition.

04

Applications of Big Data and IoT

❑ Big Data Uses in Healthcare Sectors:

The challenges in healthcare are to **lower rising costs and improve patient outcomes**. **Big Data is essential for predictive analytics in healthcare**. By using predictive analytics of a patient's medical history, physicians can **provide superior patient care**, patient protection. Clinics can make decisions that **improve patient health** because analytics tools can predict which **patients are at risk for which diseases**. The healthcare industry has struggled to keep up with the rapidly growing patient base. **Each patient now has a personal set of medical records, including medication lists, medical reports, lab test results, and more.**

Many tools are available that can analyze patient data and advise physicians on appropriate actions. Patients' health is monitored by a variety of wearable sensors, including those that measure heart rate, blood pressure, and pulse. Doctors can monitor the data, reducing unnecessary hospitalization of patients. Medical researchers use Big Data analytics to predict the most successful treatments for cancer patients based on their cure rates .

04

Applications of Big Data and IoT

❑ Big Data Applications in Education:

Big Data is revolutionizing the **education sector** by enabling institutions to monitor and analyze student performance across subjects. Through statistical analysis, schools and universities can identify factors affecting learning outcomes and design **personalized study plans** tailored to each student's learning style. **IoT devices** such as **smartboards, tablets, RFID attendance systems, and classroom sensors** collect real-time data on students' engagement, **attendance, and performance**. This data is then processed and analyzed using **Big Data analytics** to gain valuable insights. Big Data analytics helps educators **provide targeted support** and suggest appropriate interventions to improve academic performance. It also assists in **career guidance**, allowing teachers to recommend career paths based on students' interests and strengths. Big Data enables the creation of **customized lessons and materials**, enhancing learning effectiveness and student development.

04

Applications of Big Data and IoT

❑ Big Data Applications in Manufacturing and Natural Resources:

The integration of **Big Data and IoT technologies** is transforming both the **manufacturing** and **natural resources** industries by enabling smarter, data-driven operations. In natural resources sectors such as oil, gas, mining, and agriculture, **IoT sensors**, **satellite systems**, and **connected equipment** continuously collect vast amounts of **temporal**, **spatial**, and **textual data** about environmental conditions, machinery performance, and resource usage. This data is analyzed using **Big Data analytics** and **predictive modeling** to support decision-making, optimize production, and improve sustainability. For example, predictive analytics can forecast equipment failures, monitor soil and weather conditions in agriculture, or identify optimal drilling locations in mining.

04

Applications of Big Data and IoT

❑ Big Data Applications in Manufacturing and Natural Resources:

Similarly, in manufacturing, IoT-enabled machines and production lines generate real-time data on operations, quality, and energy consumption. **Big Data solutions** analyze this information to detect defects, enhance product quality, forecast demand, and streamline supply chains and delivery planning. These technologies also help evaluate **new production techniques**, improve **energy efficiency**, and reduce waste, leading to **cost savings and higher productivity**. Overall, the combination of **IoT for real-time data collection** and **Big Data for advanced analytics** enables industries to innovate, operate more efficiently, and make proactive, evidence-based decisions that enhance performance and sustainability.

04

Big Data Companies

There are several **top companies across various industries** that are utilizing both Big Data technology and IoT to drive innovation and gain competitive advantages. These top companies are leveraging **Big Data technology and IoT to drive innovation**, improve operational efficiency, enhance customer experiences, and create new business opportunities. **Following are some notable examples:**

- **Google:** Google's **self-driving car project**, Waymo, utilizes Big Data and IoT extensively. IoT sensors embedded in the vehicles **collect vast amounts of real-time data, including information about road conditions, traffic patterns, and pedestrian movement**. This data is then analyzed using Big Data analytics to enable autonomous driving and improve transportation efficiency.

04

Big Data Companies

- **Amazon:** Amazon uses Big Data and IoT to enhance its e-commerce operations. The Amazon Echo device, collects user data and interactions which are processed and analyzed using Big Data techniques .This enables personalized recommendations, targeted advertising, and voice-controlled home automation.
- **Microsoft:** Microsoft's Azure IoT platform offers a comprehensive suite of services for managing, analyzing, and deriving insights from IoT data. Azure IoT Hub securely connects and collects data from devices, while Azure IoT Suite provides pre-configured solutions for specific industries. Microsoft's Big Data analytics services, such as Azure Data Lake and Azure Machine Learning, enable organizations to process and analyze vast amounts of IoT data, uncover patterns, and derive actionable insights .

04

Big Data Companies

- **Tesla:** Tesla utilizes Big Data and IoT technologies in its electric vehicles. Tesla vehicles are equipped with a multitude of sensors and systems that collect data related to vehicle performance, battery usage, and driving behavior. This data is continuously analyzed using Big Data techniques to optimize vehicle performance, improve energy efficiency, and enhance the overall driving experience.
- **Cisco:** Cisco provides networking infrastructure and solutions for IoT deployments. Its IoT offerings include connectivity solutions, security solutions, and data management platforms. Cisco's Big Data analytics capabilities enable organizations to process, analyze, and gain insights from the massive amounts of data generated by IoT devices. This helps optimize IoT deployments, improve operational efficiency, and enhance security.

04

Summary

1. The big data architecture
2. IoT and big data
3. IoT and big data applications



Thank You