



Big Data(BCS061/BCDS-601/KOE-097)



Unit –1 Introduction to Big Data

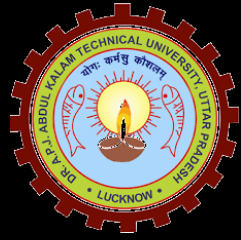
I	Introduction to Big Data: Types of digital data, history of Big Data innovation, introduction to Big Data platform, drivers for Big Data, Big Data architecture and characteristics, 5 Vs of Big Data, Big Data technology components, Big Data importance and applications, Big Data features – security, compliance, auditing and protection, Big Data privacy and ethics, Big Data Analytics, Challenges of conventional systems, intelligent data analysis, nature of data, analytic processes and tools, analysis vs reporting, modern data analytic tools.
---	---



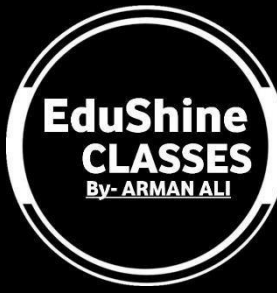
Edushine Classes



Download Notes : <https://rzp.io/rzp/i1wHz0Xm>



Big Data(BCS061/BCDS-601/KOE-097)



✓ What is Digital Data?

Digital Data means **any information** that is stored or processed using **computers**. It is the data we use every day on our phones, computers, or online.

☞ Example:

- A photo you click with your phone
- A message you send on WhatsApp
- A YouTube video
- A Google search result

All of these are digital data.

📁 Types of Digital Data (3 main types):

1. Structured Data

- ✓ This data is **organized in proper format** like rows and columns.
- Easy to search and store in databases (like Excel or MySQL).

❓ Example:

A table of student records (Name, Roll No., Marks)



Big Data(BCS061/BCDS-601/KOE-097)



2. Unstructured Data

✗ This data has **no fixed format**, and it's difficult to organize.

🔍 **Example:**

A selfie or photo

3. Semi-Structured Data

↻ This data is **partly organized**. Not in full table form, but still has **some structure** using tags or keys.

🔍 **Example:**

JSON file: { "name": "Arman", "age": 20 }

★ **Big Data Platform – Key Points:**

- A toolset to **store, process, and analyze** big data
- Handles **huge volume** of data (more than normal software can)
- Works with **structured, unstructured, and semi-structured** data
- Helps in **fast data processing** and finding useful insights
- Used in **business, healthcare, banking, social media**, etc.
- Popular platforms: **Hadoop, Spark, AWS, Google BigQuery**



Big Data(BCS061/BCDS-601/KOE-097)



✦ What are "Drivers for Big Data"?

Drivers are the **reasons** or **factors** that caused the **growth of Big Data**. They are like the **fuel** behind the rise of Big Data.

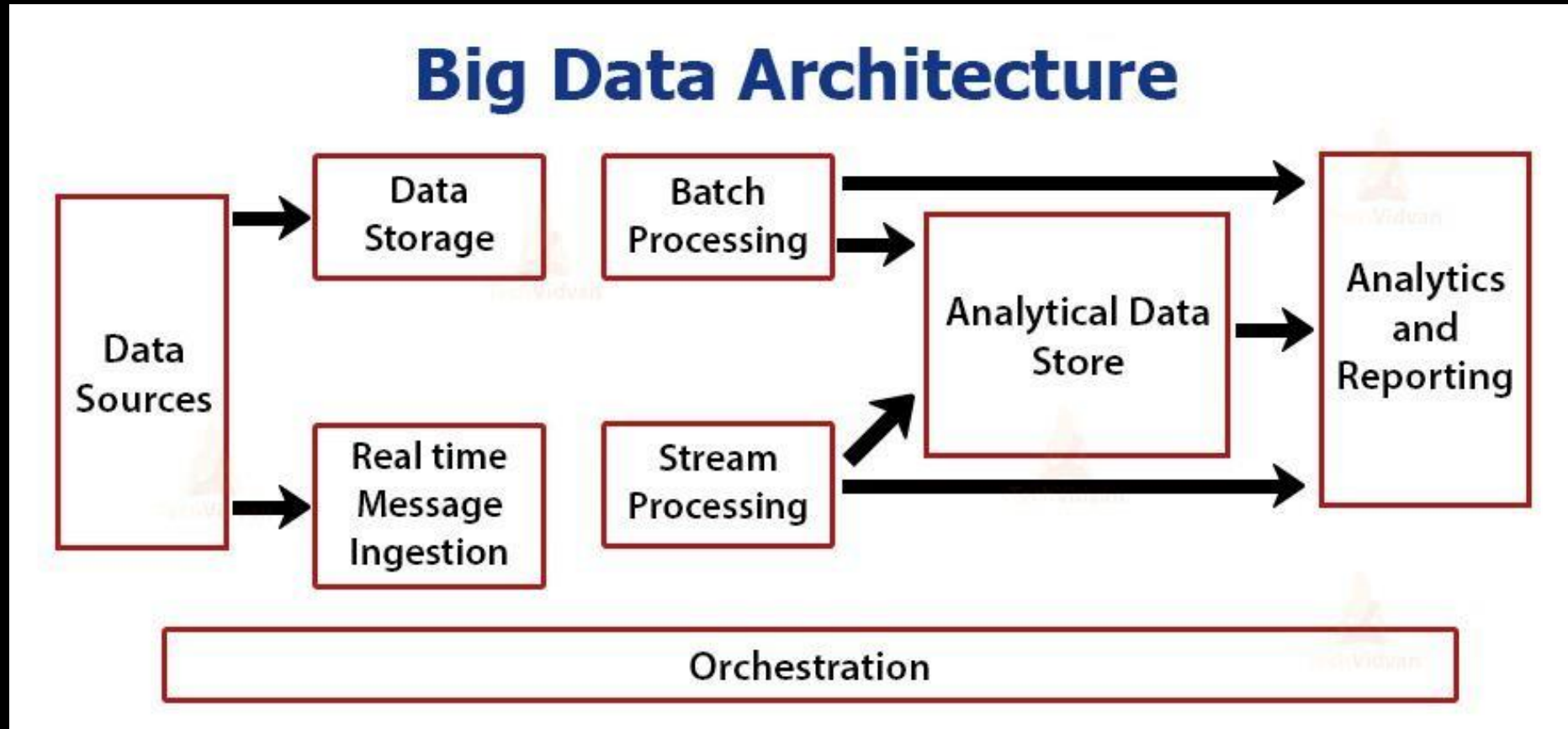
- 📱 **Social Media** – Billions of posts, photos, videos daily
- ➡️📱 **Mobile Apps** – Continuous data from usage and clicks
- 🌐 **IoT Devices** – Smart gadgets sending data all the time
- 🛒 **Online Shopping** – Data from user behavior and purchases
- ☁️ **Cloud Storage** – Easy and cheap way to store big data
- 💾 **Cheap Storage Devices** – Hard disks and cloud are affordable
- ❓ **AI & Analytics** – Companies want to use data for smart decisions



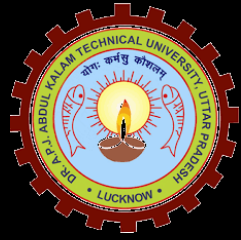
Big Data(BCS061/BCDS-601/KOE-097)



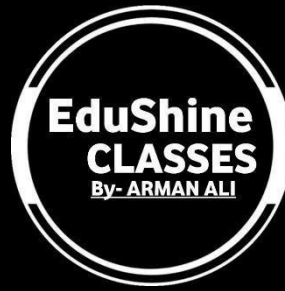
❖ Big Data Architecture : V.V.VIMP



This architecture shows how **Big Data flows** from start to end — from data collection to final results like reports.



Big Data(BCS061/BCDS-601/KOE-097)



1. Data Sources

- The starting point.
- Data comes from websites, apps, sensors, social media, etc.

2. Data Storage

- Big data is stored here.
- Think of it like a huge locker or warehouse for data.

3. Batch Processing

- This processes **large data in chunks**, not instantly.
- Example: Processing sales data of the whole day at night.

4. Real-Time Message Ingestion

- Collects data **as soon as it is created** (live).
- Example: A sensor sending temperature every second.

5. Stream Processing

- Processes data **immediately** as it comes.
- Example: YouTube showing live viewer count.

6. Analytical Data Store

- Processed data is stored here, ready for analysis.
- Like a cleaned and organized shelf.

7. Analytics and Reporting

- Final step: Data is used to create **reports, dashboards, or graphs**.
- Helps in making **smart decisions**

8. Orchestration

Manages all the steps smoothly.
Think of it as a **controller** that keeps everything running properly.



Big Data(BCS061/BCDS-601/KOE-097)



Simple Flow:

Data Sources → Storage/Real-time → Processing → Analytics → Results

★ 5 Vs of Big Data (Easy & Short): Characteristics or Properties (V.V.IMP)

1. 📊 Volume

- Huge amount of data
- Example: Facebook stores billions of posts, photos, and videos.

2. ⚡ Velocity

- Speed of data coming in
- Example: Live updates from YouTube views, sensors, stock markets.

3. 🎭 Variety

- Different types of data
- Example: Text, images, videos, audios, PDFs, etc.

4. 🎯 Veracity

- Correctness or truth of data
- Data must be accurate and reliable (no fake or wrong info).

5. 💰 Value

- Usefulness of data
- Data should give us helpful insights or benefits.



Big Data(BCS061/BCDS-601/KOE-097)



★ Big Data Technology Components : (IMP)

These are the **main parts (tools/technologies)** used to handle Big Data — from collecting to analyzing it.

✓ 1. Data Sources

- 📥 Where data comes from
 - 📖 Example: Mobile apps, websites, social media, sensors

✓ 2. Data Storage

- 💾 Stores large amounts of data safely
 - 📖 Example: HDFS (Hadoop Distributed File System), NoSQL databases (like MongoDB)

✓ 3. Data Processing

- ⚙️ Processes the data to make it useful

Two types:

- Batch Processing:** Data processed in bulk (e.g., daily reports)
 - 👉 Tool: **Hadoop MapReduce**
- Real-time Processing:** Data processed instantly (e.g., live chats)
 - 👉 Tool: **Apache Spark, Apache Storm**



Big Data(BCS061/BCDS-601/KOE-097)



✓ 4. Data Analysis

Finds patterns and useful information

? Tools: Apache Hive, Pig, Spark SQL, etc.

✓ 5. Data Visualization

? Shows data in charts, graphs, dashboards

? Tools: Tableau, Power BI, Google Data Studio

✓ 6. Data Security & Privacy

A Keeps data safe from hacking, loss, or misuse

? Tools: Kerberos, SSL, data encryption

✓ 7. Data Management & Orchestration

? Manages the flow of data between all components

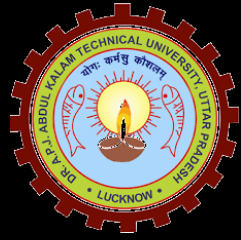
? Tools: Apache Oozie, Apache NiFi

Big Data technology includes tools for **collecting, storing, processing, analyzing, showing, and protecting** huge data.



-

[Download Notes : https://rzp.io/rzp/i1wHz0Xm](https://rzp.io/rzp/i1wHz0Xm)



Big Data(BCS061/BCDS-601/KOE-097)



✦ Importance of Big Data (Easy & Short):

- ✓ Helps in better decision making
- 📈 Improves business growth and planning
- 🎯 Understands customer behavior and needs
- ⚙️ Supports automation and AI
- 🔒 Helps in fraud detection and security
- ⌚ Saves time and money through faster analysis

✦ Applications of Big Data (Where it's used):

- 🛒 E-commerce – Recommends products (like Amazon)
- 🚑 Healthcare – Tracks diseases, improves treatment
- 🏦 Banking – Detects fraud and manages risks
- 📱 Social Media – Analyzes trends and user activity
- 🚗 Transport – Optimizes routes and traffic (like Ola, Uber)
- 🏢 Business – Analyzes market and customer data



Big Data(BCS061/BCDS-601/KOE-097)



🔑 Big Data Security (Short Notes)

Big Data Security means protecting large volumes of data from unauthorized access, misuse, or attacks.

✅ Steps to Secure Big Data:

- ✓ **Secure Data Sources** – Use authentication to protect where data comes from.
- ✓ **Encrypt Data** – Lock data during storage and transfer.
- ✓ **Access Control** – Only authorized users can access data.
- ✓ **Firewalls & Detection** – Block threats and monitor attacks.
- ✓ **Monitoring & Auditing** – Keep logs of access and changes.
- ✓ **Data Masking** – Hide sensitive parts of data.
- ✓ **Regular Updates** – Fix software bugs and close security gaps.
- ✓ **Tool Security** – Secure big data tools like Hadoop, Spark.
- ✓ **Backup & Recovery** – Keep backups to restore lost data.
- ✓ **Legal Compliance** – Follow data protection laws (e.g., GDPR).



A What is Big Data Privacy?

Big Data Privacy means **protecting personal and sensitive information** in big data from being misused or exposed. It ensures that **individuals' data is used legally and ethically**, with their **consent and control**.

⚠️ **Big Data Privacy Concerns:**

1. Lack of User Consent

Data is collected without asking or informing users.

2. Data Misuse

Personal data may be sold or used for other purposes.

3. Re-identification Risk

Even anonymized data can sometimes be traced back to individuals.

4. Data Breaches

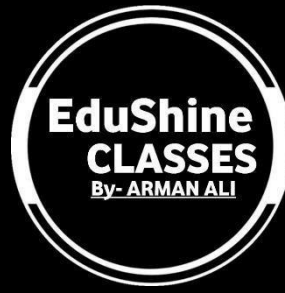
Hackers can steal large amounts of private information.

5. Tracking and Surveillance

User behavior is tracked across devices and platforms.



Big Data(BCS061/BCDS-601/KOE-097)



6. Data Sharing with Third Parties

Companies share data with advertisers or partners without permission

7. Lack of Transparency

Users don't know how their data is being used or stored.

What is Big Data Analytics?

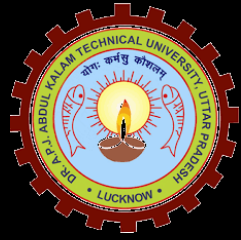
Big Data Analytics means analyzing large and complex data to find useful patterns, trends, and insights. It helps companies make better decisions, improve services, and understand customer behavior.

✓ Advantages of Big Data Analytics:

- Better Decision Making
- Customer Understanding
- Fraud Detection
- Cost Reduction
- Innovation

✗ Disadvantages of Big Data Analytics:

- Privacy Issues
- High Cost
- Data Overload
- Security Risks
- Wrong Analysis



Big Data(BCS061/BCDS-601/KOE-097)



What is a Conventional System?

A conventional system is a traditional data processing system like relational databases (RDBMS) that can handle small to medium amounts of structured data.

Challenges of Conventional Systems (in Big Data context):

- **Limited Storage** – Cannot handle very large data.
- **Slow Processing** – Takes time to process big data.
- **Cannot Handle Variety** – Works well only with structured data.
- **Scalability Issues** – Hard to expand as data grows.
- **High Cost** – Becomes expensive to upgrade or manage big data.
- **Real-time Analysis Not Possible** – Can't give instant insights.



Big Data(BCS061/BCDS-601/KOE-097)



❖ What is Intelligent Data Analysis? (Easy Explanation)

Intelligent Data Analysis (IDA) means using smart methods like machine learning, AI, and statistical tools to understand data and find hidden patterns automatically.

It's more than just collecting or viewing data — it helps in **predicting trends, finding problems, and making decisions** without human guessing.

✓ Key Features of Intelligent Data Analysis:

- i. **Automated Learning** – Learns from past data to improve over time.
- ii. **Pattern Detection** – Finds trends, relationships, or unusual data.
- iii. **Predictive Analysis** – Can forecast future events (e.g., sales, fraud).
- iv. **Decision Support** – Helps businesses make smart choices.
- v. **Handles Big Data** – Works well with large and complex data sets.

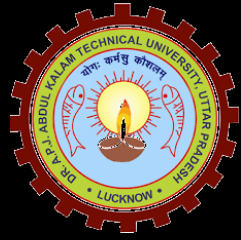
🔧 Tools Used:

Machine Learning (like Decision Trees, Clustering)

Artificial Intelligence

Data Mining

Neural Networks



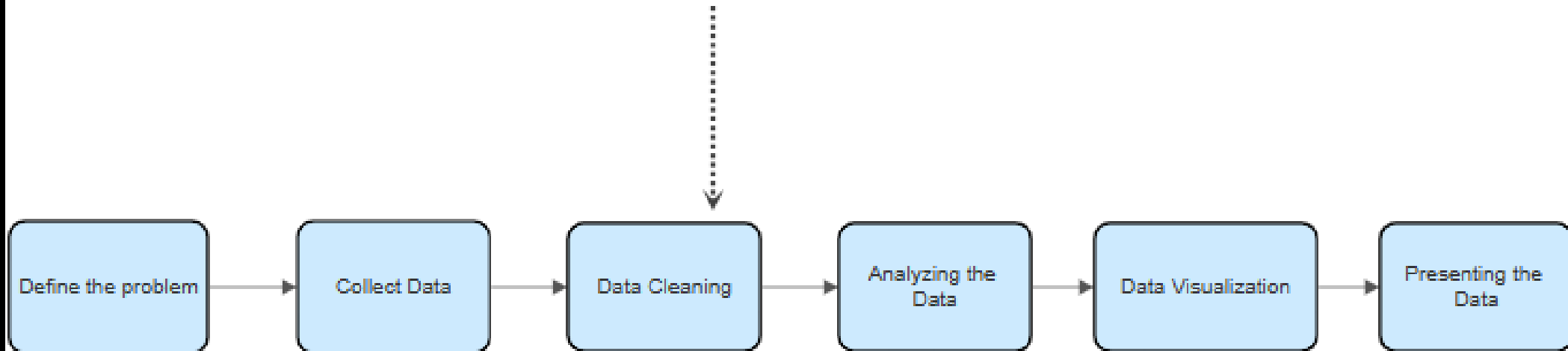
Big Data(BCS061/BCDS-601/KOE-097)



❖ Data Analytics Process : (V.V.VIMP)

Here is the six steps involve in Data Analytics process let discuss -

Six Steps of Data Analysis Process





Big Data(BCS061/BCDS-601/KOE-097)



1. Define the Problem

- Clearly understand what question or issue you want to solve using data.

2. Collect Data

- Gather data from different sources like websites, databases, sensors, etc.

3. Data Cleaning

- Remove incorrect, duplicate, or missing data to improve quality.

4. Analyze the Data

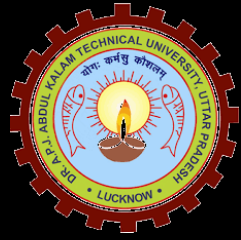
- Use statistical or machine learning tools to find patterns and insights.

5. Data Visualization

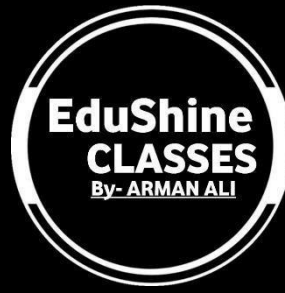
- Show results using graphs, charts, or dashboards to make them easy to understand.

6. Presenting the Data

- Share the final results and insights with others (e.g., in reports or meetings).



Big Data(BCS061/BCDS-601/KOE-097)



❖ Tools used in Analytics:

Step	Tools/Examples
Data Collection	Apache Flume, Kafka, APIs
Data Cleaning	Python (Pandas), OpenRefine
Data Storage	HDFS, Hive, MongoDB, Cloud Storage
Data Analysis	R, Python (NumPy, scikit-learn), Spark MLlib
Visualization	Tableau, Power BI, Matplotlib, Excel

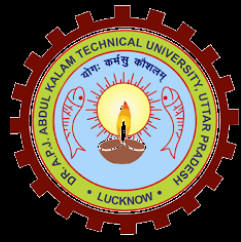


Big Data(BCS061/BCDS-601/KOE-097)

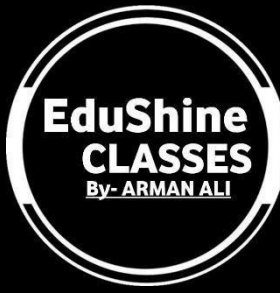


❖ Difference between Analysis and Reporting :

Point	Analysis	Reporting
Purpose	To explore data and find insights	To summarize data and show results
Focus	Discovering patterns, trends, and causes	Presenting facts and figures clearly
Approach	Investigative and deep	Descriptive and straightforward
Outcome	New knowledge and understanding	Regular updates or snapshots
Tools	Data mining, statistics, machine learning	Charts, dashboards, summaries
Timeframe	Can be real-time or periodic	Usually periodic (daily, weekly)
Examples	Why did sales increase last month?	Sales report for last month
User	Data scientists, analysts	Managers, stakeholders



Big Data(BCS061/BCDS-601/KOE-097)



❖ Modern Data Analytics Tools :

1. Hadoop

Helps to **store and handle huge data** by using many computers together.

2. Spark

A fast tool that helps to **work with big data quickly**.

3. Tableau

Makes **easy and colorful pictures (charts)** to show data clearly.

4. Power BI

Like Tableau, it helps make **simple reports and charts**.

5. Python

A computer language that helps you **look at data and find patterns**.

6. R

Another computer language good for **studying numbers and making graphs**.

7. Cloud Services (like AWS)

Let you **store and use big data on the internet** without buying your own computers.



Big Data(BCS061/BCDS-601/KOE-097)



8. Jupyter Notebook

A tool where you can write code and see results quickly in one place.

Thank You...



Download Notes : <https://rzp.io/rzp/i1wHz0Xm>