

# **GROUP TASK – 01**

## **BIG DATA PROCESS MAPPING**

A Structured Approach to Understanding the Flow of Big Data

### **1. Introduction**

Big Data Process Mapping refers to the systematic representation of how large volumes of data move through different stages — from collection to decision-making.

It helps organizations:

- Understand data flow
- Identify bottlenecks
- Improve efficiency
- Ensure data quality
- Optimize performance

Instead of just studying Big Data concepts, process mapping focuses on how data actually travels inside a system.

### **2. What is Process Mapping?**

Process Mapping is a visual representation of:

Input → Processing → Output

In Big Data systems, this includes:

- Data sources
- Storage systems
- Processing engines
- Analytics tools
- Visualization platforms

It ensures clarity in handling complex data environments.

### **3. Stages of Big Data Process Mapping**

Big Data generally follows these major stages:

#### **→Stage 1: Data Generation (Sources)**

Data is generated from multiple sources such as:

- Social media platforms
- IoT devices
- Sensors
- Websites
- Mobile applications
- Banking systems

Example: Millions of searches on Google Search generate massive data every second.

### →Stage 2: Data Collection

Data is collected using:

- APIs
- Log files
- Streaming platforms
- Web scraping
- Message queues

Tools used:

- Apache Kafka
- Flume

This stage ensures raw data enters the system.

### →Stage 3: Data Storage

Since data volume is huge, distributed storage is used.

Common storage systems:

- Hadoop Distributed File System (HDFS)
- Cloud storage
- NoSQL databases

Cloud providers:

- Amazon Web Services
- Microsoft Azure
- Google Cloud

### →Stage 4: Data Processing

Processing can be:

1. Batch Processing
2. Real-Time Processing

Popular frameworks:

- Apache Hadoop
- Apache Spark

For example: Streaming data from Twitter is processed in real-time to detect trends.

### →**Stage 5: Data Cleaning & Transformation**

Before analysis:

- Remove duplicates
- Handle missing values
- Normalize formats
- Transform into structured form

This step ensures data quality.

### →**Stage 6: Data Analysis**

Using:

- Machine Learning
- Data Mining
- Statistical analysis

Organizations extract insights such as:

- Customer behavior
- Sales trends
- Fraud detection

### →**Stage 7: Data Visualization & Decision Making**

Results are shown through:

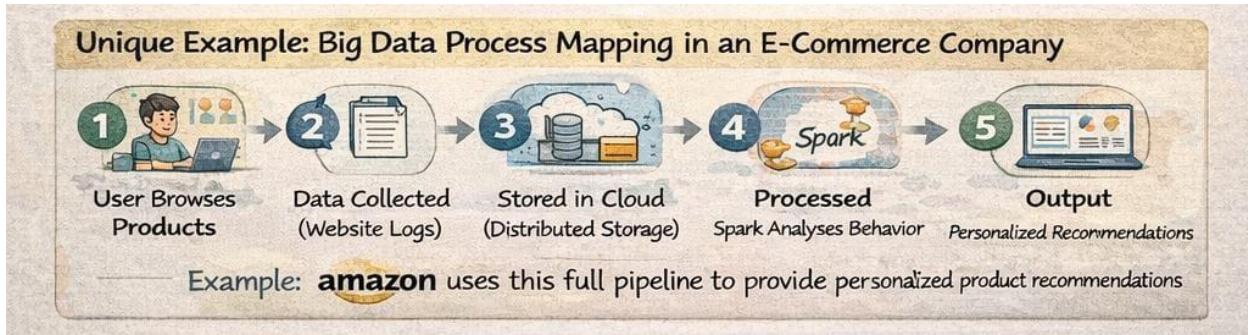
- Dashboards
- Graphs
- Reports

Business decisions are taken based on insights.

#### 4. Complete Big Data Process Flow (Simple Representation)



## 5. Unique Example: Big Data Process Mapping in an E-Commerce Company



Let's map it practically:

→ Step 1: User Browses Product

Data generated (clicks, search history)

→ Step 2: Data Collected

Through website logs

→ Step 3: Stored in Cloud

Using distributed storage

→ Step 4: Processed

Spark analyzes purchase behavior

→ Step 5: Analyzed

ML predicts recommended products

→ Step 6: Output

Personalized recommendation shown

Example company: Amazon uses this full pipeline.

This practical mapping makes your report unique.

## 6. Importance of Big Data Process Mapping

✓ Improves efficiency

✓ Reduces errors

- ✓ Enhances scalability
- ✓ Better system design
- ✓ Faster business decisions

## 7. Challenges in Big Data Process Mapping

- ✗ Managing unstructured data
- ✗ Real-time processing complexity
- ✗ Data security
- ✗ High infrastructure cost
- ✗ Data integration issues

### **Conclusion:**

Big Data Process Mapping provides a clear blueprint of how massive data flows through an organization. It transforms raw data into meaningful insights through structured stages such as collection, storage, processing, analysis, and visualization.

Understanding process mapping is essential for designing scalable, efficient, and intelligent data systems.

Thus, Big Data Process Mapping is not just a technical process — it is the backbone of data-driven decision making.