GROUP TASK (Module -2)

Big Data Process Mapping: Groups select a real-world big data system (like Google Maps, Amazon recommendations, or smart city sensors) and map out the entire data flow: data sources, storage, processing, and output.

Introduction

Big Data Process Mapping refers to the analysis of how large-scale systems collect, store, process, and transform massive amounts of data into meaningful outputs. In modern digital platforms, millions of users generate data every second through their interactions with applications, devices, and online services. This data must be handled efficiently using advanced technologies to provide accurate, fast, and personalized services. According to the given case study, big data systems follow a pipeline consisting of data collection, storage, processing, and output generation.

Case Study 1: Google Maps Big Data Process Mapping

1. Data Collection in Google Maps

The first stage of the big data process is data collection. Google Maps continuously collects data from various sources to provide accurate navigation and traffic information.

The primary source of data is smartphones and GPS-enabled devices. When users enable location services, their devices send information such as GPS coordinates, movement patterns, and travel speed to Google servers. Each search for directions or location generates additional data that contributes to system learning.

Apart from user devices, Google Maps also collects data from:

- Satellites and aerial imaging systems
- Road sensors and traffic cameras
- Government transport systems
- Weather and construction updates
- Partner organizations

2. Data Storage in Google Maps

After collection, the data is transmitted to Google's cloud infrastructure for storage. Since the data volume is extremely large, traditional storage methods are not sufficient. Google Maps uses distributed databases and data centers located worldwide to manage data efficiently.

The stored data includes:

- Current user locations
- Historical travel records
- Digital road maps
- Satellite images
- Traffic patterns
- Speed and route data

The system replicates data across multiple servers to ensure reliability and fault tolerance. This means that if one server fails, data can still be accessed from another server. Long-term data is stored in data warehouses for future analysis and system improvement.

3. Data Processing and Analysis in Google Maps

Once the data is stored, Google Maps processes it using real-time analytics and machine learning models. This stage converts raw data into useful information.

Real-time processing enables the system to monitor traffic conditions instantly. For example, when multiple users slow down in a specific area, the system detects congestion or accidents. Machine learning algorithms analyze both current and historical data to predict traffic patterns, peak hours, and accident-prone zones.

Route optimization algorithms calculate the best path by considering:

- Distance
- Traffic density
- Road closures
- User preferences
- Weather conditions

This intelligent processing allows Google Maps to provide accurate navigation services even in complex urban environments.

4. Data Output and User Interaction in Google Maps

After processing, Google Maps delivers results through its interface. The system provides:

- Optimized routes
- Estimated travel time
- Live traffic updates

- Accident alerts
- Alternate paths

These outputs are continuously updated as new data arrives. When users follow suggested routes, their movements generate additional data that is fed back into the system. This creates a continuous feedback loop that improves system accuracy over time.

Thus, Google Maps operates as a self-improving big data system.

Case Study 2: Amazon Recommendation System

1. Data Collection in Amazon

Amazon collects detailed information about customer behavior on its platform. Every action performed by users generates data.

Examples include:

- Product searches
- Clicking items
- Reading reviews
- Adding products to cart
- Making purchases
- Scrolling behavior

Amazon also collects feedback data such as ratings, reviews, returns, and customer service interactions. Over time, this data helps build digital profiles of customers and understand their preferences and buying patterns.

2. Data Storage in Amazon

The collected data is stored in Amazon's distributed cloud storage systems capable of handling billions of transactions daily.

Stored data includes:

- Customer profiles
- Purchase history
- Product information
- Payment records
- Transaction logs

To ensure security, sensitive data is encrypted using advanced security protocols. Backup and redundancy mechanisms protect data from hardware failures and cyber threats. Historical data is maintained for trend analysis and business planning.

3. Data Processing and Recommendation Generation

Amazon uses artificial intelligence and machine learning techniques to analyze stored data and generate recommendations.

One major technique used is collaborative filtering, which identifies users with similar interests and recommends products based on shared behavior. The system also applies predictive analytics and data mining to discover relationships between products.

For example, if customers who buy smartphones also purchase earphones or cases, the system learns this pattern and recommends related products to future buyers. Deep learning models further improve recommendation accuracy by analyzing user demographics, browsing behavior, and purchase history.

4. Data Output and Personalization in Amazon

The results of data processing are delivered as personalized recommendations to users. These recommendations appear in:

- Homepage suggestions
- Product pages
- Email notifications
- Mobile alerts

Sections such as “Recommended for You”, “Customers Also Bought”, and “Frequently Bought Together” are dynamically generated.

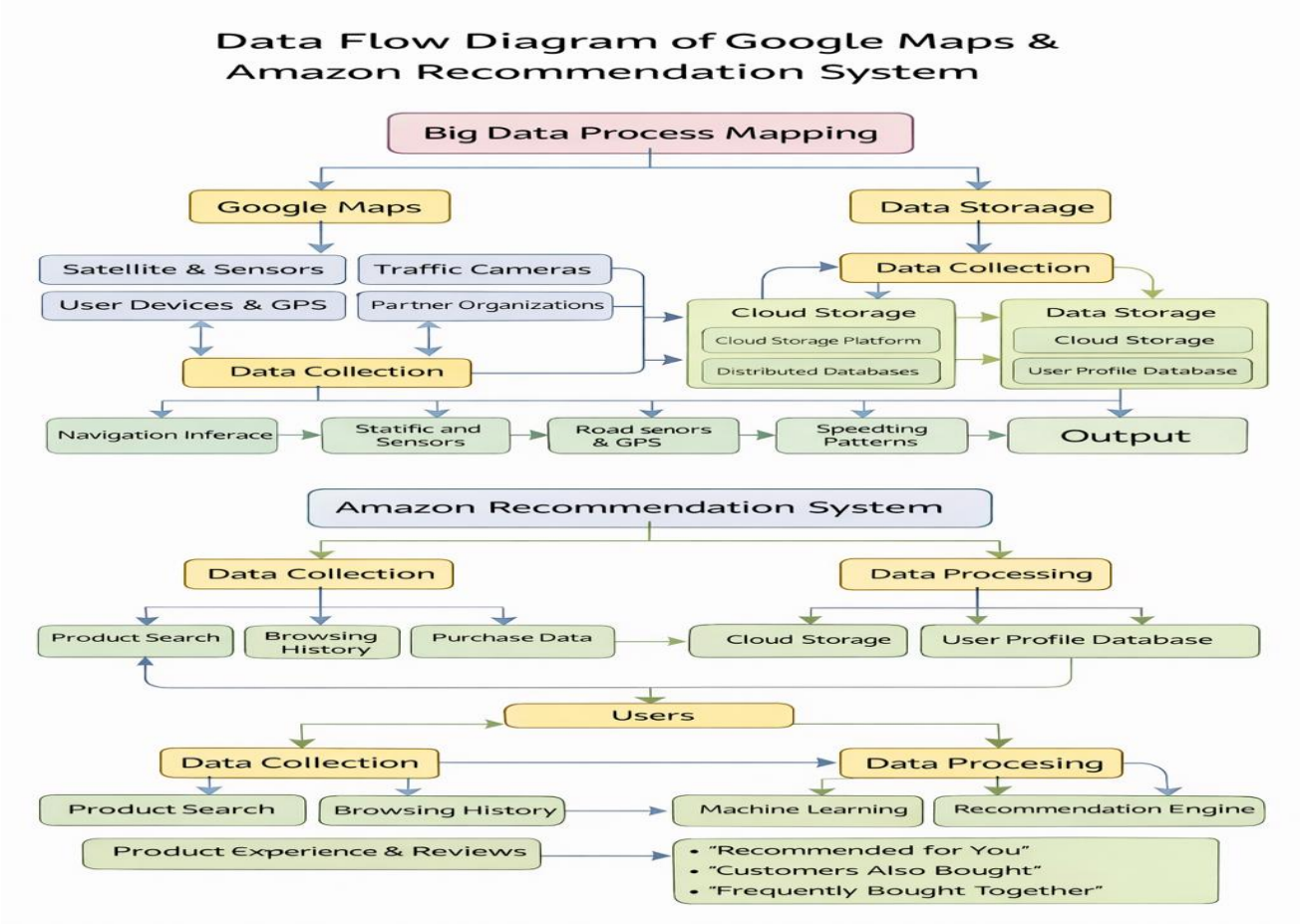
Importance of Big Data Process Mapping

Big data process mapping helps organizations understand data flow and improve system efficiency. It provides:

- Better decision-making
- Improved service accuracy
- Real-time analytics
- Personalized user experience
- System optimization

These benefits make big data essential for modern digital platforms.

User interaction with recommendations generates new data, which is again stored and processed. This creates a continuous self-learning cycle that enhances personalization and improves customer experience.



Conclusion

Big data process mapping provides a structured understanding of how modern systems manage massive volumes of data to deliver intelligent services. The case study of Google Maps and Amazon demonstrates how data flows from collection to output through advanced technologies such as cloud computing, distributed databases, and machine learning models.

Google Maps uses big data to provide navigation and traffic services, while Amazon uses big data to deliver personalized product recommendations. As technology continues to advance, big data systems will become more powerful and essential in shaping the future of digital platforms.