Group Task-II
BIG DATA PROCESS MAPPING

Group task: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 systems are designed to collect, manage, and analyze extremely large volumes of data generated at high speed from multiple sources. These systems operate using distributed computing, cloud storage, and machine learning algorithms to produce real-time results.
For this group task, we selected Google Maps as our real-world Big Data system. It is one of the most widely used navigation and location-based service platforms globally. Google Maps processes billions of location requests daily and integrates satellite data, live traffic signals, user inputs, and geographic databases to provide optimized routes and location recommendations.
This report maps the complete data flow of Google Maps, including:

- Data Sources
- Data Storage
- Data Processing
- Data Output

1.Data Sources (Data Collection Stage)
Google Maps collects data from multiple channels. Each source contributes different types of data.
A. User GPS Data
Smartphones continuously send GPS coordinates when location services are enabled.
Explanation:

- GPS sensors determine latitude and longitude.
- The app collects speed, direction, and time stamps.
- This data helps detect traffic congestion and travel speed.
- Millions of users together create real-time traffic patterns.

Data Type: Structured (coordinates, speed values stored in numeric format) B. Search Queries:When users search for places like “restaurants near me” or “petrol pump,” the system records the query.
Explanation:

- Search terms are logged.

- Time and location of search are recorded.
- Helps understand popular locations and demand patterns.
- Used to improve search accuracy and recommendations.

Data Type: Semi-Structured (text queries with metadata such as timestamp and user ID)

C. Satellite Imagery

Google Maps integrates satellite images to provide visual geographic representation.

Explanation:

- High-resolution images are captured using satellites.
- Images are updated periodically.
- Used to map terrain, buildings, and road networks.
- AI models detect new roads and infrastructure changes.

Data Type: Unstructured (image files)

2.Data Storage:

After collection, data Is stored In distributed cloud systems.

A. Structured Storag:Stores GPS values, ratings, traffic speed I' databases for quick access.
B. Semi-Structured Storage:Search logs and API data stored in ""exib'e NoSQL systems.
C. Unstructured Storage:Images and videos stored in large distributed file systems.

Purpose:Ensures scalability, fault tolerance, and fast retrieval.

3. Data Processing: Once stored, data is processed using advanced analytics and machine learning.

1.Real-Time Traffic Analysis:Live GPS signals are processed instantly.

Explanation:

- Aggregates speed data from thousands of devices.
- Detects slow-moving clusters.
- Identifies traffic jams and accidents.
- Updates routes dynamically.

2.Route Optimization Algorithms:Shortest path algorithms compute optimal routes.

Explanation:

- Uses graph-based algorithms.
- Considers traffic, distance, and road conditions.
- Provides fastest and alternative routes.
- Recalculates if user changes direction.

4.Data Output: After processing, results are delivered to users.
1.Navigation Routes:Users receive optimized travel routes.
Explanation:

- Shows fastest path.
- Provides turn-by-turn instructions.
- Updates if traffic changes.

2 .Real-Time Traffic Updates:Maps use color coding to display traffic conditions. Explanation:

- Green → Smooth traffic
- Orange → Moderate traffic
- Red → Heavy congestion
- Helps users choose better routes.

3. Estimated Time of Arrival (ETA):Predicted arrival time is displayed.
Explanation:

- Based on speed, distance, and traffic.
- Updated continuously.
- Improves travel planning.

4. Location Recommendations:Users see nearby suggestions. Explanation:

- Restaurants, hospitals, fuel stations shown.
- Personalized based on search history.
- Influenced by ratings and popularity.

Advantages:

- Real-time navigation and traffic updates.
- Personalized recommendations for users.
- Scalable system handling billions of users.
- High accuracy in ETA and traffic predictions.
- Fault-tolerant due to distributed storage.
- Continuous improvement through machine learning.

Applications:

- Navigation for private vehicles, taxis, and deliveries.
- Urban planning and traffic management for cities.
- Business analytics for location-based marketing.
- Logistics and supply chain optimization.
- Emergency response route planning.
- Tourism and virtual exploration using Street View.

Data Flow Mapping
Process:Data is collected continuously.Raw data is validated and filtered.

Process Mapping Diagram (Text format) :

DATA SOURCES
↓
User GPS | Satellites | IoT Sensors | Government Data
↓
DATA INGESTION
↓
Streaming APIs & Batch Uploads
↓
DATA STORAGE
↓
Distributed Storage | NoSQL | In-memory Cache
↓
DATA PROCESSING
↓
Real-time Analytics | Batch Processing | ML Models
↓
OUTPUT
↓
Navigation | Traffic Alerts | APIs | Smart City Insights
↓
Feedback loop(User interaction generate new data)

Conclusion: Google Maps is a practical example of a Big Data system that combines high-volume data collection, distributed storage, advanced processing, and intelligent output.

Key points:

- Integrates structured, semi-structured, and unstructured data from multiple sources.
- Uses real-time analytics and machine learning to provide optimized navigation and personalized suggestions.
- Continuously improves through a feedback loop, learning from user behavior and historical trends.

This system highlights the importance of scalability, fault tolerance, and intelligent data processing in modern Big Data applications. It demonstrates how raw data can be transformed into actionable insights that enhance user experience and operational efficiency.