

Group Task-2

Big Data Process Mapping : Google Maps

Introduction

Google Maps is one of the most widely used real-world big data systems, serving billions of users globally. This report maps the end-to-end big data flow of Google Maps and highlights how modern big data technologies transform raw data into real-time, intelligent services.

1. Data Sources

Google Maps collects data from diverse and distributed sources, making it a classic example of big data characterized by high volume, velocity, and variety.

Primary data sources include:

- **Smartphones and user devices:** GPS location, movement speed, navigation queries, and search patterns.
- **Satellite imagery:** High-resolution geographic images used for terrain and map generation.

2. Data Ingestion

The ingestion layer is responsible for collecting and transmitting raw data into Google's infrastructure.

Key ingestion mechanisms:

- Real-time streaming from mobile devices.
- API-based integration for third-party data.

3. Data Storage

Due to massive data scale, Google Maps relies on distributed and highly scalable storage systems.

Storage technologies include:

- Distributed file systems like Google File System (GFS) and Colossus.
- NoSQL databases such as Bigtable.

4. Data Processing

The processing layer transforms raw data into usable intelligence using distributed computing and AI.

Batch Processing:

- Map building and updates.
- Satellite image stitching.

Real-Time Processing:

- Live traffic congestion analysis.
- Incident detection and rerouting.

Key technologies and methods:

- Distributed computing models similar to MapReduce.
- Graph algorithms like Dijkstra and A* for shortest path routing.

5. Analytics and Intelligence

Google Maps applies advanced analytics and artificial intelligence to generate meaningful insights.

Major analytics functions:

- Traffic congestion prediction.
- Estimated Time of Arrival (ETA) calculation.

AI integration includes:

- Deep learning for image recognition.
- Natural Language Processing (NLP) for place search.

6. Output and User Services

The processed data is delivered to users in the form of real-time, intelligent outputs.

End-user services:

- Turn-by-turn navigation.
- Live traffic visualization.

Business and ecosystem outputs:

- Location-based advertising.
- Business visibility and recommendations.

7. Feedback Loop

Google Maps continuously improves through a feedback-driven learning system.

Feedback sources:

- User corrections and reports.
- Navigation success signals.

Challenges

Despite its effectiveness, Google Maps faces several big data challenges:

- Managing massive real-time global data streams.

Conclusion

Google Maps is a powerful example of a real-world big data system that demonstrates the full lifecycle of big data processing. It integrates diverse data sources such as smartphones, satellites, and IoT sensors, processes them using distributed storage and advanced analytics, and delivers real-time intelligent navigation services. Through continuous data ingestion, machine learning-driven processing, and feedback loops, Google Maps transforms raw geographic data into actionable insights for billions of users. This process mapping illustrates how modern big data architectures enable scalable, intelligent, and adaptive digital services, making Google Maps a benchmark system in the big data domain.