Big Data Process Mapping – Google Maps System

Introduction

- Google Maps is a large-scale big data system used worldwide for navigation, traffic monitoring, and location services.
- It processes petabytes of data daily from millions of users.
- The system uses real-time and historical data to provide:
    - Route suggestions
    - Live traffic updates
    - Estimated Time of Arrival (ETA)
    - Nearby place recommendations
- It is a perfect example of a Big Data ecosystem because it handles:
    - High Volume (billions of GPS signals)
    - High Velocity (real-time updates)
    - High Variety (text, maps, images, GPS signals, reviews)

Data Sources in Google Maps

A. User-Generated Data

- GPS location from smartphones
- Search queries (e.g., "restaurants near me")
- Route requests
- User reviews and ratings
- Traffic reports (accident, roadblock)

B. Device & Sensor Data

- Mobile phone GPS sensors
- Accelerometer data (detect movement speed)
- Car navigation systems
- Smart traffic cameras

C. External Data Sources

- Government road maps
- Satellite imagery

- Weather data

- Business listings

Data Storage Layer

Need for Storage in Big Data

- Data collected is extremely large.

- It must be stored securely and efficiently.

- Storage should allow fast access and scalability.

Storage Technologies Used

Google Maps uses distributed storage systems such as:

- Google Bigtable – For structured data storage

- Google File System – For large-scale distributed file storage

- Cloud data centers across the world

Types of Data Stored

- Road network graphs

- Traffic history data

- User activity logs

- Images and satellite data

- Business listings

Storage Architecture Diagram

Incoming Data

↓

Data Cleaning & Validation

↓

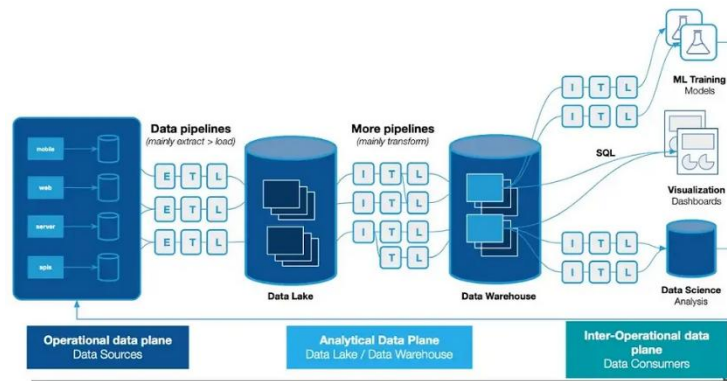Distributed Storage (Cloud Servers)

↓

Data Replication for Backup
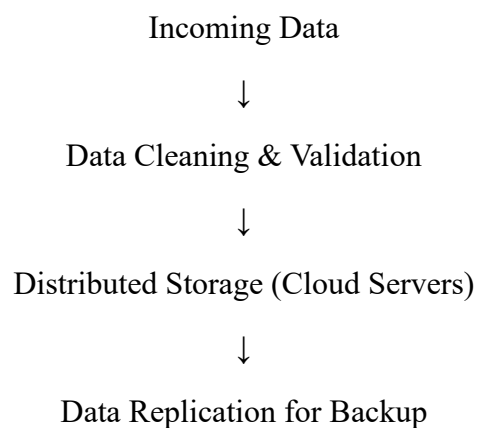
Explanation

- Data is first cleaned to remove incorrect GPS signals.

- It is then stored in distributed servers.

- Replication ensures data safety.

- Cloud storage ensures scalability.

Data Processing Layer

Real-Time Processing

- Google Maps processes traffic data every few seconds.

- It calculates:

    - Current vehicle speed

    - Traffic congestion

    - Road block detection

- Real-time stream processing frameworks are used.

Batch Processing

- Historical traffic data is analyzed.

- Peak traffic hours are identified.

- Long-term route optimization is done.

Machine Learning in Processing

- Machine learning models predict:

    - Future traffic conditions

    - Best route suggestions

    - ETA accuracy

- Algorithms used:

    - Graph algorithms

3

- o Shortest path algorithms

  - o Predictive analytics

Output Layer And Complete Data Flow

Output to Users

- Navigation directions

- Real-time traffic alerts

- Alternate route suggestions

- Nearby business recommendations

- Estimated arrival time

Output Channels

- Mobile app interface

- Web application

- API for third-party apps

Big Data Characteristics in Google Maps

- Volume → Billions of GPS records daily

- Velocity → Real-time updates every few seconds

- Variety → Text, images, maps, reviews

- Veracity → Data cleaning improves accuracy

- Value → Accurate navigation and time saving

Conclusion

- Google Maps is a complete big data ecosystem.

- It integrates data collection, distributed storage, large-scale processing, and intelligent output.

- It demonstrates how big data systems improve daily life through predictive analytics and real-time computation.

- The system continuously learns and improves using user interactions.