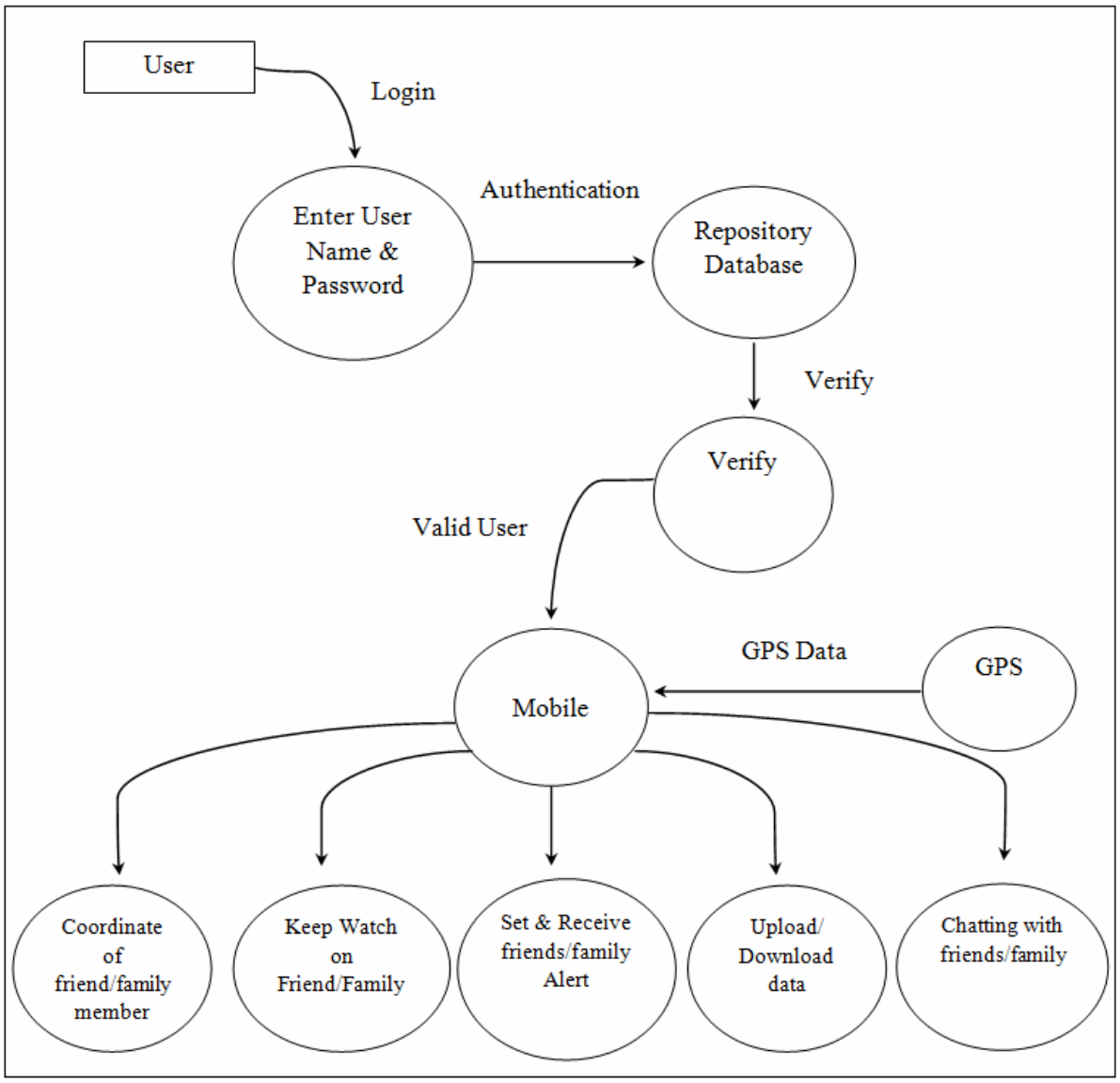
GROUP TASK-2

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.

Big Data Process Mapping Report: Google Maps



1. Introduction

Big Data refers to extremely large, fast-growing, and complex datasets that cannot be processed using traditional database systems. Modern applications such as navigation platforms, e-commerce recommendations, and smart cities rely on Big Data to deliver real-time, intelligent services.

A powerful real-world example of a Big Data system is **Google Maps**, developed by **Google**. Google Maps provides services such as real-time navigation, traffic congestion alerts, route optimization, location search, and estimated arrival times (ETA) to billions of users worldwide.

This report explains the **complete Big Data process mapping of Google Maps**, clearly describing:

- Data Sources
- Data Ingestion and Collection
- Data Storage
- Data Processing and Analytics
- Machine Learning Role
- Output and User Services

---

2. Why Google Maps is a Big Data System

Google Maps qualifies as a Big Data system because it handles:

- **Huge data volume** from millions of users every second
- **High-velocity real-time data** such as live traffic updates
- **Wide variety of data formats** including GPS signals, images, text, and sensor data

Traditional systems cannot manage this scale, making Big Data technologies essential.

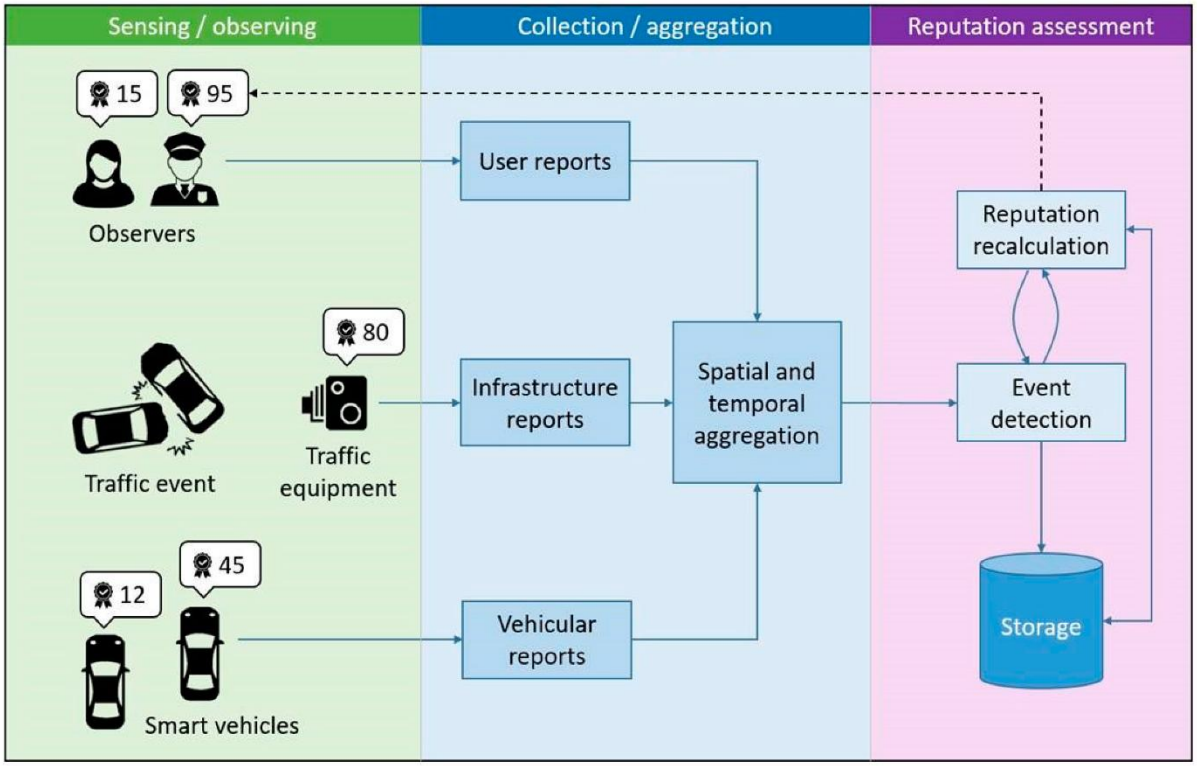
---

3. Overall Big Data Flow in Google Maps

The complete Big Data flow can be represented as:

Data Sources → Data Ingestion → Distributed Storage → Data Processing & Analytics → Machine Learning Models → User Output

4. Data Sources



4.1 Primary Data Sources

Google Maps collects data from multiple real-world sources:

a) GPS and Mobile Device Data

- Location coordinates (latitude & longitude)
- Speed and direction of movement
- Travel patterns of users

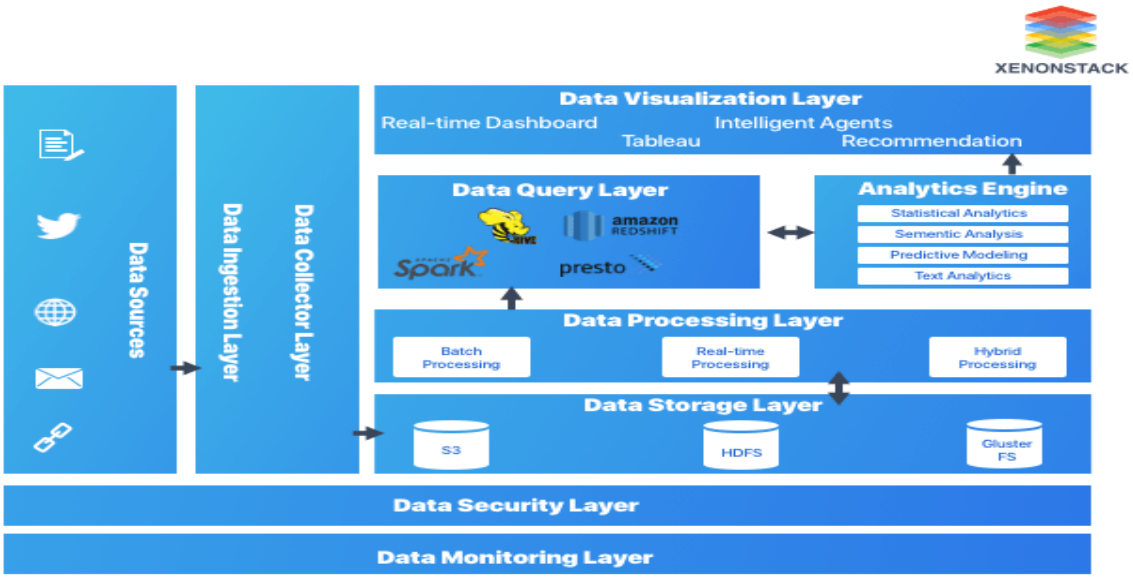
Example:
When users drive with Google Maps enabled, anonymous location data helps detect traffic congestion.

---

b) Crowdsourced User Data

- Accident reports
- Road closures
- Traffic jams
- User reviews and ratings
- Photos of place

5. Data Ingestion and Collection



5.1 Storage Mechanism

Due to massive data size, Google Maps uses **distributed cloud storage systems**:

- Data is stored across multiple servers and data centers
- Structured data (roads, locations)
- Unstructured data (images, videos, reviews)

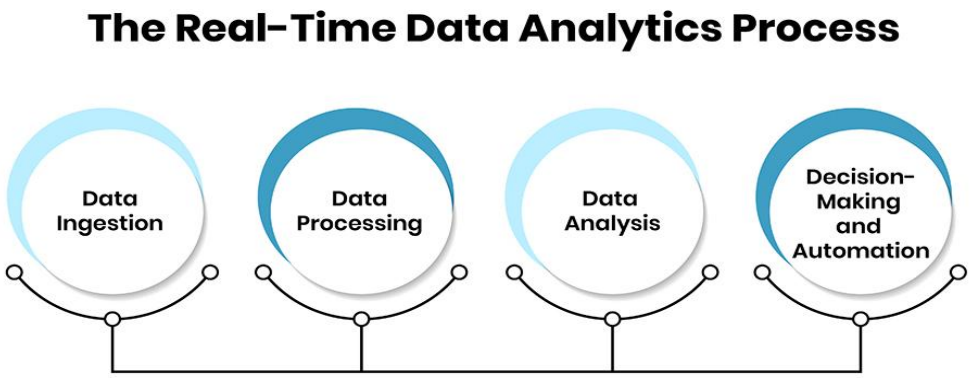
Characteristics of Storage

- High scalability
- Fault tolerance
- Fast data access

Importance of Storage

- Handles petabytes of data
- Ensures data safety and reliability
- Supports real-time queries

6. Data Processing



6.1 Processing Techniques

Google Maps uses both **batch processing** and **real-time stream processing**:

- **Real-time processing**
  - Live traffic updates
  - Accident detection
  - Route recalculation
- **Batch processing**
  - Map updates
  - Road network optimization
  - Historical traffic analysis

Role of Machine Learning

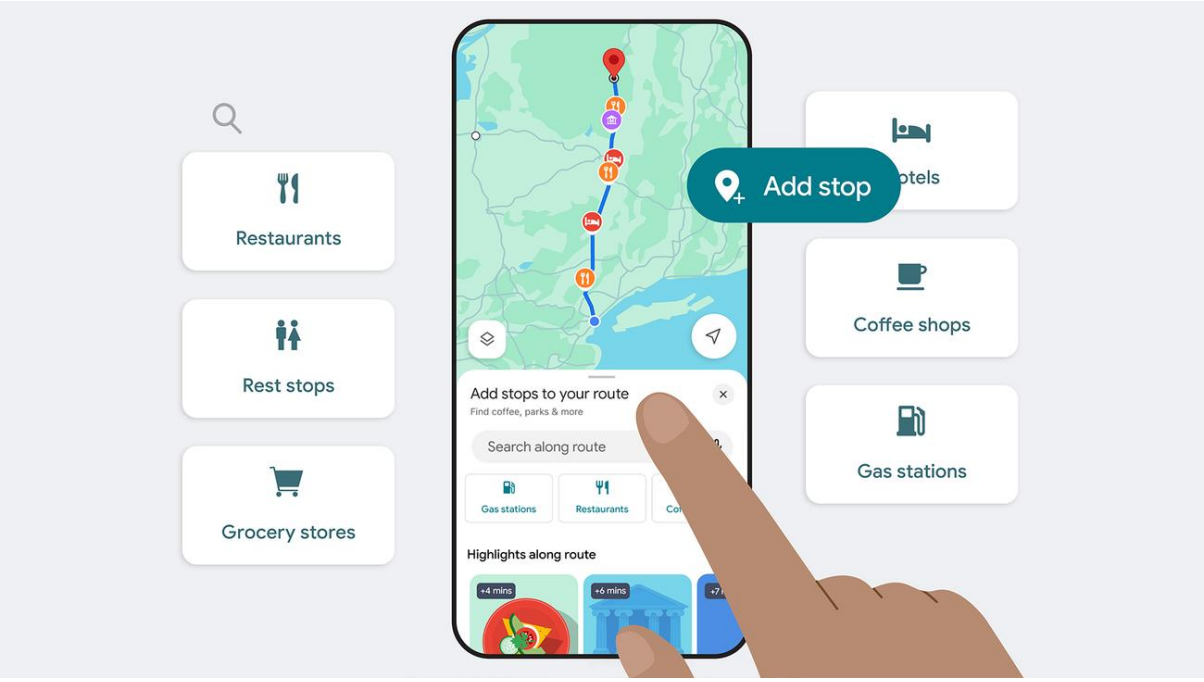
- Predicts traffic congestion
- Suggests fastest routes
- Estimates arrival time (ETA)

Importance of Data Processing

- Converts raw data into useful insights
- Enables quick decision-making
- Improves navigation accuracy

---

7. Output (User Services)



7.1 Final Outputs Provided to Users

After processing, Google Maps delivers the following outputs:

- Real-time traffic conditions
- Fastest and alternate routes
- Estimated travel time (ETA)
- Turn-by-turn navigation
- Nearby places (restaurants, hospitals, fuel stations)

Importance of Output

- Enhances user travel experience
- Saves time and fuel
- Improves road safety

Conclusion

Google Maps is a clear and practical example of how Big Data systems operate in real-world applications. By continuously collecting massive amounts of data from multiple sources such as GPS-enabled devices, user contributions, satellite imagery, and public datasets, Google Maps is able to maintain accurate and up-to-date mapping information. This data is stored in distributed cloud systems and processed using real-time and batch analytics to handle high volume, velocity, and variety.

Through the use of machine learning and advanced data processing techniques, Google Maps converts raw data into meaningful outputs such as real-time traffic updates, optimized routes, and accurate estimated arrival times. Overall, this Big Data process mapping shows how effective data flow management enables intelligent decision-making, improves user experience, and supports efficient transportation systems in everyday life.

One-line conclusion:
Google Maps effectively uses Big Data to transform large-scale real-time data into smart, reliable navigation services.
