SRINIVAS UNIVERSITY

INSTITUTE OF ENGINEERING AND TECHNOLOGY

SUBJECT: FUNDAMENTALS OF ARTIFICIAL INTELIGENTS AND MACHINE LEARNING

GROUP TASK

GROUP MEMBERS
MOHAMMAD SHABAS TB
MUHAMMED NASEEM
MUHAMMAD ZAHIER
ENOCH PRAISE

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.

A powerful real-world example of a big data system is Google Maps, a web-based navigation and mapping platform developed by Google. It provides services such as route planning, live traffic updates, nearby place suggestions, satellite imagery, and street-level views. Behind this simple interface lies a complex big data ecosystem that continuously collects, stores, processes, and delivers massive volumes of information. Mapping the entire data flow of Google Maps helps us understand how big data systems operate in real time. The process can be divided into four major stages: data sources, data storage, data processing, and data output.

The first stage in the data flow is data collection from multiple diverse sources. Google Maps gathers data from satellites, GPS devices, smartphones, vehicles, government agencies, local businesses, and user contributions. Satellite imagery provides geographical and topographical information about land, buildings, rivers, and roads. GPS signals from millions of smartphones help determine user locations and movement patterns. When users enable location services, anonymized data such as speed, direction, and travel routes are sent to Google's servers. Additionally, users contribute by reporting accidents, road closures, traffic jams, construction work, and incorrect map details. Businesses also provide information about their names, locations, operating hours, and contact details. Public transportation agencies supply bus and train schedules. This enormous inflow of structured, semi-structured, and unstructured data represents the "Volume" and "Variety" characteristics of big data. The data arrives continuously every second, reflecting the "Velocity" aspect. Thus, Google Maps receives terabytes of data daily from millions of distributed sources across the globe.

The second stage is data ingestion and storage. Once data is collected, it is transmitted securely to Google's distributed data centers located worldwide. Because the data is extremely large and constantly growing, traditional databases are not sufficient. Instead, Google uses distributed storage systems that break data into smaller chunks and store them across multiple servers. Technologies like distributed file systems and NoSQL databases are used to manage structured and unstructured data efficiently. Data replication is applied so that copies of the same data are stored in different locations, ensuring reliability and fault tolerance. If one server fails, another can immediately take over. Real-time traffic data is stored in fast-access memory systems to allow immediate updates, while historical data such as old traffic patterns and past route statistics are stored in long-term storage systems for analysis. This layered storage strategy ensures scalability, high availability, and quick retrieval. Data indexing techniques are used so that map tiles, road networks, and business listings can be quickly accessed when requested by users.

The third stage is data processing and analytics, which forms the core intelligence of Google Maps. Raw data by itself is not useful unless it is cleaned, organized, and analyzed. First, data cleaning removes duplicate, inconsistent, or incorrect entries. Machine learning algorithms detect anomalies, such as unusual traffic speeds or false reports. After cleaning, the data is integrated into a unified geographic information system (GIS). Real-time processing

frameworks analyze incoming GPS signals to calculate traffic density. For example, if many devices on a road segment are moving slowly, the system interprets this as traffic congestion. Algorithms compare current speeds with historical averages to determine the severity of traffic conditions. Route optimization algorithms then calculate the fastest or shortest path using graph theory techniques. These algorithms consider multiple variables such as distance, traffic, road closures, toll charges, and estimated time of arrival. Predictive analytics also plays a major role. By analyzing past data, the system can predict future traffic patterns at specific times of the day or week. Artificial intelligence models continuously learn from new data to improve accuracy. Image processing techniques analyze satellite images and Street View photos to detect new roads or changes in infrastructure. Thus, this stage transforms raw big data into meaningful insights using advanced computing techniques.

The fourth and final stage is data output and user interaction. Once processed, the information is delivered to users through mobile apps and web interfaces. When a user searches for a destination, the system instantly retrieves map tiles, calculates routes, and displays estimated travel time. Live traffic is shown using color-coded lines, such as green for smooth traffic and red for congestion. Voice navigation provides turn-by-turn directions using real-time updates. If an accident occurs ahead, the system automatically reroutes the user. Businesses appear with ratings, reviews, and photos. Public transport schedules are displayed with live arrival times. All of this output must be delivered within milliseconds to ensure a seamless user experience. The system also collects feedback from user interactions, such as route selections and travel confirmations, which are fed back into the data pipeline for continuous improvement. This creates a feedback loop where output becomes new input, enhancing the system's intelligence over time.

In addition to these four core stages, Google Maps incorporates strong data security and privacy mechanisms. User data is anonymized and encrypted during transmission. Access controls ensure that sensitive information is protected. Compliance with international data protection regulations is maintained. The system also employs monitoring tools to track performance, detect cyber threats, and maintain uptime. Scalability is another essential feature. During peak travel hours or major events, data traffic increases significantly. Cloud infrastructure automatically scales computing resources up or down based on demand, ensuring uninterrupted service.

Overall, Google Maps represents a complete big data ecosystem. The data flow begins with massive and diverse data collection from satellites, GPS devices, and user inputs. It moves to distributed storage systems capable of handling enormous volume. It is then processed using machine learning, real-time analytics, and predictive algorithms. Finally, meaningful and actionable information is delivered instantly to users worldwide. The system operates continuously, handling high velocity data streams while maintaining accuracy and reliability. This example clearly demonstrates how big data technologies transform raw information into intelligent services that impact millions of people daily. By mapping this process, we understand how modern big data systems integrate data sources, storage mechanisms, advanced processing techniques, and user-centric outputs into a seamless and highly efficient digital platform.