

## Group Task – Module 2

# Data Essentials, Types, Big Data, processing, and Ethics

## 1)Big-Data Process mapping.

### Overview of the system

- Google maps is developed by Google and is one of the most widely used navigation platforms in the world.
- It uses big data technologies to:
  - provide real-time traffic updates
  - suggest shortest / fastest routes
  - show nearby places
  - predict travel time.
  - provide satellite and street views
- This system handles massive data characterized by:
  - Volume - Billions of location data points.
  - Velocity - Real-time traffic updates.
  - Variety - GPS data, image, text reviews, maps.

### Complete Data Flow mapping :

We can divide the big data process into five main stages.

- Data Sources
- Data collection.
- Data storage
- Data processing.
- output / user interface.

## 1)Data Source

Google maps collects data from multiple sources

User GPS Data:

- Location Coordinates from smartphones,

- speed and movement patterns.
- Route searches.

Traffic sensing:

- Road cameras
- smart traffic lights.
- IOT road sensors.

Satellite and street view Images:

- Satellite imagery.
- street view cameras mounted on vehicles.

User contributions:

- Reviews and ratings.
- photos.
- Business information updates.

Government & Public Data:

- Road maps.
- construction updates.
- Accident reports.

This creates large and continuous data streams

## 2) Data Collection :

Data is collected through.

- mobile apps.
- Web Platform
- APPS
- IOT device.

Google uses distributed systems to handle millions of data per second.

Key characteristics:

- real-time streaming data
- Batch data
- structured and unstructured data.

## 3) Data storage

- Because the data volume is extremely large, traditional databases are not sufficient.

Google uses:

- Distributed storage systems.
- Cloud-based infrastructure
- Big data storage technologies.

Data types stored:

- Structured data (Coordinates, timestamps)
- semi-structured data (JSON, Logs)

The storage system must be:

- Scalable
- Fault-tolerant.
- Secure.

## 4)Data Processing

This is the most important stage in big data systems

Real-Time Processing:

- Detect traffic congestion.
- Identify accidents.
- Calculate estimated time of arrival

Batch Processing:

- Analyze historical traffic patterns.
- Improve route prediction models.

Machine Learning Algorithm:

- Predict traffic conditions
- Suggest alternate routes
- Recommend nearby places.

## 5)Output/User Interface

After processing, the system provides:

- Faster route suggestions:
- Real-time traffic color indications
- Distance and time estimation.
- Voice navigation.
- Nearby recommendations.

The final output is delivered to users through:

- Mobile App
- Web application.