

INDIVIDUAL TASK 2 UNDERSTANDING BIG DATA AROUND ME

Introduction

In the modern digital world, data is generated continuously through smartphones, online applications, GPS systems, banking transactions, social media platforms, and transportation services. Every digital interaction creates information. When this information becomes extremely large in size, moves at high speed, and exists in multiple formats, it is called Big Data.

Big Data refers to datasets that are so large and complex that traditional data processing systems are unable to manage them efficiently. Specialized technologies such as cloud computing, distributed databases, and real-time analytics systems are required to store and process such data.

To understand Big Data clearly, it is commonly explained using three main characteristics known as the 3 V's of Big Data:

1. Volume – The huge amount of data generated
2. Velocity – The speed at which data is created and processed
3. Variety – The different types and formats of data

This report explains these three characteristics in detail using Uber as a real-world example. Uber is a ride-hailing platform that connects drivers and passengers through a mobile application. It depends entirely on data for its operation.

Volume:

Volume refers to the massive quantity of data generated by a system. In Big Data environments, the size of data is measured in terabytes (TB), petabytes (PB), or even exabytes (EB).

The larger the number of users and transactions, the higher the volume of data.

Sources of Data Volume in Uber

Uber generates large amounts of data from multiple sources:

A. Ride Information

Each ride produces detailed records including:

- Pickup and drop-off locations
- Distance traveled
- Time taken
- Route followed
- Fare amount

B. GPS Tracking

During every ride:

- The driver's location updates every few seconds
- Speed and route information are recorded
- Traffic data is monitored

Millions of such GPS updates are recorded daily.

C. User and Driver Profiles

Uber stores:

- Names
- Contact information
- Vehicle details
- License information
- Background verification data
- Ratings and review history

D. Historical Data Storage

Uber maintains years of ride history for:

- Demand prediction
- Service improvement
- Legal compliance
- Business analysis

When all this information is combined across millions of daily rides worldwide, the total volume becomes extremely large. Managing such enormous data requires powerful storage systems and distributed cloud servers.

Thus, Uber clearly demonstrates the Volume aspect of Big Data.

Velocity:

Velocity refers to the speed at which data is generated, transmitted, and processed. In modern digital platforms, data flows continuously and must be analyzed instantly to provide real-time services.

3.2 Real-Time Data in Uber

Uber operates entirely in real-time. The following activities show high data velocity:

A. Ride Matching

When a passenger books a ride:

1. The app detects the user's GPS location instantly.
2. It searches for nearby available drivers.
3. It calculates estimated time of arrival (ETA).
4. It estimates the fare.
5. It sends ride requests to drivers.

Variety:

Variety refers to the different types and formats of data handled by a system. Big Data is not limited to simple numbers and text; it includes structured, semi-structured, and unstructured data.

4.2 Types of Data in Uber

1. Structured Data

This data is organized in tables and databases.

- Customer names and phone numbers
- Driver IDs
- Payment records
- Ride duration
- Fare calculations

2. Semi-Structured Data

This data does not fit perfectly into tables but has some organization.

- GPS coordinate logs
- App usage logs
- Traffic information
- JSON data files

3. Unstructured Data

This data has no fixed format.

- Customer feedback comments
- Support chat messages
- Voice call recordings
- Uploaded profile pictures
- Identity verification images

4. External Data Sources

Uber also integrates:

- Map data
- Weather data
- Road condition data
- City traffic reports

Managing and analyzing these different types of data together requires advanced data integration systems.

Thus, Uber clearly demonstrates the Variety aspect of Big Data.

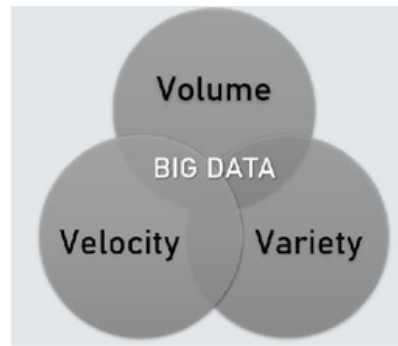


Figure 3 Characteristics of Big Data

Overall Analysis

First, considering Volume, Uber generates enormous amounts of data every single day. Each ride produces multiple data points including pickup and drop locations, time, route, distance, fare, payment details, driver information, and customer ratings. In addition to ride-related data, Uber continuously collects GPS coordinates, traffic conditions, driver availability, and historical ride records. Since millions of rides occur daily across different cities worldwide, the total data collected reaches terabytes per day and accumulates into petabytes over time. Managing such vast data requires advanced storage systems and distributed cloud infrastructure. This confirms that Uber strongly represents the Volume aspect of Big Data.

Second, in terms of Velocity, Uber operates in real time. The system must process data instantly when a user books a ride. Within seconds, the app detects the passenger's location, identifies nearby drivers, calculates estimated time of arrival (ETA), checks traffic conditions, estimates fare, and confirms the booking. GPS updates occur every few seconds, and surge pricing changes dynamically depending on demand. Notifications such as ride confirmations and payment receipts are delivered immediately. The speed at which Uber processes and responds to data highlights the importance of high-velocity data systems. Without rapid data processing, the platform would fail to provide efficient service.

Third, regarding Variety, Uber handles multiple types of data. Structured data includes customer profiles, payment records, and ride details stored in databases. Semi-structured data includes GPS logs and app activity records. Unstructured data includes feedback comments, chat messages, voice calls, and uploaded verification images. Uber also integrates external data sources such as maps, traffic reports, and weather information. Managing these different formats together requires advanced data integration and analytics systems. This clearly demonstrates the Variety dimension of Big Data.

When these three characteristics work together, they form the foundation of Uber's operational system. Volume ensures a large dataset for analysis, Velocity enables real-time decision-making, and Variety allows comprehensive understanding of customer behavior and city traffic patterns.

Conclusion

From the above analysis, it can be concluded that Uber is a practical and strong real-world example of Big Data in action. The company continuously generates massive amounts of information (Volume), processes it at extremely high speed (Velocity), and manages multiple forms of data (Variety).

Big Data enables Uber to match riders and drivers efficiently, calculate accurate fares, adjust surge pricing dynamically, improve route optimization, enhance safety measures, and provide better customer experience. Without Big Data technologies, such real-time and large-scale transportation services would not be possible.

Therefore, the concepts of Volume, Velocity, and Variety are not merely theoretical definitions found in textbooks. They are actively applied in everyday applications like Uber, demonstrating how Big Data supports modern digital systems and improves the efficiency of services we use daily.

This analysis clearly shows that Big Data plays a crucial role in shaping smart, technology-driven platforms in today's world.