**Task 2**

* **What are the 3 Vs of Data and explain each one in detail?**

The 3 Vs of data are **Volume**, **Velocity**, and **Variety**.

1. Volume:

The data that is present today is there in various formats like videos, music, images, text, etc. It is very common to have a storage system of Terabytes and Petabytes for enterprises. As the database grows the applications and architecture built to support the data needs to be reevaluated quite often. Sometimes the same data is re-evaluated with multiple angles and even though the original data is the same the new found intelligence creates explosion of the data. The big volume indeed represents Transaction-based data stored through the years. Unstructured data streaming in from social media and Satellites as well as increasing amounts of sensor and machine-to-machine data being collected. In the past, too much data volume was a storage issue. But with decreasing storage costs, other issues emerge, including how to determine relevance within large data volumes and how to use

analytics to create value from relevant data.

2. Velocity:

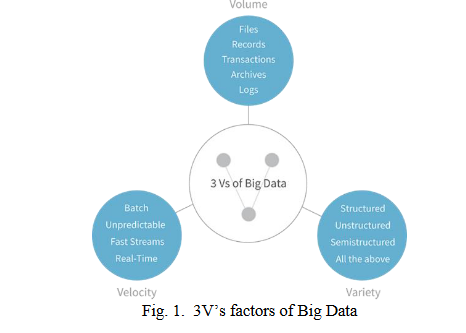
Data is streaming in at unprecedented speed and must be dealt with in a timely manner. The data growth and social media explosion have changed how we look at the data. There was a time when we used to believe that data of yesterday is recent. The matter of the fact newspapers is still following that logic. However, news channels and radios have changed how fast we receive the news. Today, people reply on social media to update them with the latest happening. On social media sometimes a few seconds old messages (a tweet, status updates etc.) is not something interests users. They often discard old messages and pay attention to recent updates. The data movement is now almost real time and the update window has reduced to fractions of the seconds. This high velocity data represents RFID tags, sensors and smart metering are driving the need to deal with torrents of data in near-real time. Reacting quickly enough to deal with data velocity is a challenge for most organizations.

3. Variety:

Data can be stored in multiple formats. Structured, numeric data and also as database, csv, excel, access or for the matter of the fact, it can be stored in a simple text file. Sometimes the data is not even in the traditional format as we want, it may be in the form of picture, video, SMS, pdf or something we might have not thought about it. Information created from line-of-business applications. Unstructured text documents, email, audio, video, stock ticker data and financial transactions. Managing, merging, processing and governing different varieties of data is something many organizations still grapple with. For

example, it is the need of the organization to arrange it and make it meaningful. It will be easy to do so if we have data in the same format, however it is not the case most of the time.

The real world has data in many different formats and that is the challenge we need to overcome with the Big Data.



* **List capabilities of Business Intelligence systems.**

The key capabilities of BI systems are:

**1. Analysis and Insights:** BI processes vast amounts of data to forecast, budget, plan, and stay current. Competitive analysis helps companies understand the competition and benchmark competitor performance. This business intelligence enables product and service differentiation.

**2. Multi-Platform, Multi-User:** BI applications work online and in mobile environments. Tools improve system performance so enterprises can distribute more information to targeted users faster. In multi-terabyte data warehouses, these tools provide excellent query performance.

**3. Speed and Competitive Edge:** BI can perform faster reporting, analysis and planning because of access to global data. The system's analysis capabilities make it possible to react to market or other conditions quickly.

**4. Scalability**: Many systems offer user scalability to support advanced reporting and analysis. Dashboards and reports are available to many users, not just restricted to the organisation's data analysts or executives.

**5**. **Connection:** The ability to manage and meld access to various data sources provides a 360-degree view of your business and your company that is not possible in a siloed data environment.

**6. Collaboration**: Tools enable data-informed improvements in various business functions like marketing, finance, sales, operations, finance, support, HR and customer care individually and together.

**7. Visualisation**: Advanced interactive dashboard representations of data using simple user interfaces offer the ability to visualise information in a graphical format to understand data more insightfully.

**8. Trusted Data and Accuracy:** Reports can be highly customised, and KPIs monitored using more than one data source. Real-time generated reports offer relevant data, which helps organisations, and their employees make better decisions. These reports provide insights, access, accuracy, and relevance.

**9. Decision-Making Support**: Companies gain a competitive edge when they can leverage the existing data at the right time to make accurate decisions faster.

**10. Customer Satisfaction**: BI can help you identify what services or products you're lacking and improve customer satisfaction by making necessary changes. Reports help you understand customer behaviour, develop user personas, and use real-time data on the customer's feedback to make corrective changes and improve customer service and, therefore, satisfaction.

**11. Employee Satisfaction:** Using BI data, you can assess team members' strengths and weaknesses and assign relevant training modules to support success. BI tools can automatically recognise positive behaviour while regularly tracking worker contributions and improvement.

**12. Strategic and KPI Targeting**: BI assists companies in gaining a competitive edge by helping them find new opportunities and build smarter strategies. Use the data to identify market trends and help improve profit margins for the company. Reports based on tracking established KPIs ensure the enterprise stays on course to match or exceed goals.

* **Different types of data with example for each type**

In statistics, data are the raw materials we use in our analysis.

There are 4 types of data:

* Nominal
* Ordinal
* Discrete
* Continuous

**1.** **Nominal Data:** Nominal data are qualitative data with no inherent ranking based on size or magnitude.

Example **–** Ethnicity

* Asian
* African
* Latin

**2**. **Ordinal Data:** Ordinal data are qualitative data with an order based on size or magnitude.

Example - Self-reported customer satisfaction of survey respondents.

* Very much unsatisfied
* Somewhat unsatisfied
* Neutral
* Somewhat satisfied
* Very much satisfied

**3. Discrete Data:** Discrete data are quantitative data representing distinct countable values. Discrete data usually takes the form of whole numbers or integers.

Example - The number of customers who visit salon each hour.

**4. Continuous data:** Continuous data are quantitative data where the data can take on decimals or fractional values. They represent measurements. Continuous data can take on an infinite set of values within a range. ‌You usually need a measuring device like a stopwatch or scale to collect continuous data.

Example - The time it takes a baker to make one cake.

* **Define data visualization:**

Data visualization is the graphical representation of data and information through charts, graphs, maps, and other visual elements. It is a powerful way to transform raw data into a more understandable and insightful format, making it easier to identify patterns, trends, relationships, and anomalies within the data. Data visualization plays a crucial role in simplifying complex information, helping with decision-making, and effectively communicating findings.

Here are some key points to understand what data visualization is:

* **Visual representation**: Data visualization uses visual elements like bars, lines, pie slices, colours, and symbols to represent data points. These visuals make it easier for people to grasp complex data quickly.
* **Simplification**: It simplifies data by condensing large datasets into a more digestible format. For example, a long list of numbers can be transformed into a bar chart or line graph for easier comprehension.
* **Patterns and trends**: Data visualization helps reveal patterns and trends within the data that might not be immediately apparent when looking at the raw data. This is essential for data analysis and decision-making.
* **Communication**: It serves as a powerful tool for communicating insights and findings. Visuals are more engaging and memorable than tables or text, making it easier to convey information to a wide audience.

In essence, data visualization is a valuable tool that transforms data into a visual language that everyone can understand, helping organizations and individuals harness the power of data for better insights and informed decision-making.

* **What is a KPI and provide an example:**

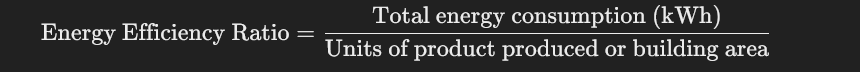
**Key Performance Indicator** (KPIs) significantly shape business strategies. They’re not just numbers on a board. Instead, they show how well a company is doing. By tracking areas like finance and customer service, KPIs help guide businesses. Metrics like Gross Profit Margin and First Contact Resolution Rate are essential. They act as navigators in the business world.

Key KPI takeaways are as follows:

* Comprehending KPI meaning is essential for optimizing business strategy.
* Strategic and operational improvements are driven by effectively managed KPIs.
* KPIs are multifaceted, tracking diverse aspects from customer behaviour to operational efficiency.
* Performance metrics are distinguished by their direct linkage to business goals.
* Successful KPI management entails setting realistic targets and monitoring progress.
* Leading and lagging indicators complement each other for balanced performance insight.
* Understanding the role of KPIs in various functions can enhance departmental performance.

Example:

**Energy Efficiency Ratio**

* A KPI used in manufacturing or real estate industries to track how much energy is used relative to output (e.g., product units produced, or square feet managed).
* **Formula:** 
* **Example:** A factory might track energy consumption relative to the number of units produced to identify opportunities for reducing waste and optimizing energy use.
* **What is a BI system?**
* BI System Design, also known as Business Intelligence System Design, is the process of creating and configuring a robust framework that allows organizations to gather, analyse, and visualize data in a structured and meaningful way. This strategic approach enables businesses to make informed decisions and gain valuable insights for optimizing their operations.
* At its core, BI System Design involves designing and implementing a system that integrates various data sources, such as databases, data warehouses, and external sources, into a reliable and centralized platform. This platform acts as a foundation for collecting and organizing data from disparate sources, transforming raw data into actionable information.
* A well-designed BI system focuses on creating user-friendly interfaces and interactive dashboards, empowering users with the ability to access, explore, and analyse data without technical expertise. By utilizing intuitive visualization techniques, BI System Design assists users in uncovering patterns, trends, and correlations within their data, enabling them to identify opportunities and make data-driven decisions.
* Additionally, BI System Design ensures data accuracy, reliability, and consistency by implementing robust data governance practices. This involves defining data quality standards, establishing data validation processes, and implementing data security measures to protect sensitive information.
* **What are the 5 Cs of Data for data preparation and the purpose of each?**

**1. Clean:** A cleaned data refers to data that is free from inconsistencies, irrelevant information and errors. This also includes, removing duplicates, correcting spelling errors, filling in missing values with appropriate entries and executing unnecessary data.

**Purpose:** Clean data assists in generating insights efficiently and accurately and also ensures a clean data is provided for analysis purposes.

**2. Consistent**: Data consistency ensures a standard format is followed throughout the system making it easier to analyse, compare and observe the data.

**Purpose:** Consistent data ensures quick analysis of the data as there is one uniform pattern followed across whole dataset. In addition, seamless integration and aggregation of data across systems is also ensured.

**3. Conformed:** Conformed data refers to a set of common protocols or standards that are set for analysing and reporting purposes. It ensures data across different systems follows a unified structure, thus making it easier to analyse and integrate.

**Purpose:** Without conforming data, analyses may produce incomplete or conflicting results. It also assists in scalability, as data from new sources can be easily integrated.

**4. Current:** Current data refers to the up-to-date data made available for analysis. This involves determining how quickly the new data is collected and updated.

**Purpose**: With the help of current data, decisions are made based on the latest information available. Outdated data can impact companies where trends and factors change frequently.

**5. Comprehensive:** Comprehensive data refers to all that data that is required to make a decision is available. It ensures no critical data is missing and that all the necessary data dimensions are captured.

**Purpose:** Comprehensive data ensures that every relevant aspect of a situation is considered. It helps decision makers get a clear picture of the task they are involved in, thus, reducing the risk of making decisions based on incomplete or partial information.

* **What are some Key Success Factors of a Successful BI Program and explain each factor?**

Listed below are some key success factors of a successful BI program:

**1. Cohesive BI Strategy:** Developing a cohesive BI strategy that aligns with the organization’s overall business strategy is crucial. This involves defining the vision, goals, and roadmap for BI initiatives and ensuring that they are communicated throughout the organization. A cohesive strategy ensures that all BI efforts are aligned and focused on delivering value to the organization, preventing fragmentation and promoting a unified approach to data-driven decision-making.

**2.** **Customer-Centric Insights**: Focusing on customer-centric insights can drive better customer experiences and satisfaction. This involves analysing customer data to understand their needs, preferences and behaviours.

**3.** **Regular Audits and Reviews**: Conducting regular audits and reviews of the BI program helps in identifying areas for improvement. This includes assessing data quality, user satisfaction and the overall impact of the BI initiatives.

**4.** **Leveraging Data Science and Advanced Analytics:** Integrating data science and advanced analytics capabilities into the BI program enables organizations to extract deeper insights from data. This includes utilizing techniques such as clustering, regression analysis, and natural language processing. Leveraging advanced analytics allows organizations to uncover hidden patterns and correlations in data, leading to more robust insights and informed strategies.

**5.** **Social media and Sentiment Analysis:** Integrating social media data and sentiment analysis into the BI program provides insights into customer opinions, market trends, and brand perception. This can help organizations respond to public sentiment and enhance customer engagement.Leveraging social media insights improves the organization’s ability to engage with customers and adapt strategies based on real-time feedback, leading to better customer relationships.

**6.** **Integration with Business Processes:** Ensuring that BI insights are integrated into everyday business processes allows organizations to act on data in real-time. This includes embedding analytics into workflows and operational systems.Integrating BI with business processes enhances operational efficiency, as decisions can be made based on up-to-date insights without the need for separate analysis sessions.

**7.** **Crisis and Contingency Planning:** Developing a crisis and contingency plan that outlines how the BI program will respond to unexpected challenges, such as data breaches or system failures, is crucial for maintaining operational continuity.Having a clear crisis plan enhances the organization’s resilience, ensuring that it can quickly adapt to and recover from disruptions while minimizing impacts on decision-making.

**8. Use of Natural Language Processing (NLP):** Incorporating NLP capabilities in BI tools allows users to interact with data using natural language queries, making data analysis more accessible to non-technical users. NLP simplifies data interaction, allowing users to generate insights without needing advanced analytical skills, thereby increasing engagement and usage of BI tools.

**9**. **Developing a Robust Data Backup Strategy:** Establishing a data backup and recovery strategy ensures that critical data is protected and can be restored in case of system failures or data loss incidents. This includes regular backups and testing recovery procedures. A robust backup strategy minimizes the risk of data loss and ensures business continuity, allowing organizations to quickly recover and maintain operations in adverse situations.

**10.** **Enhancement of Data Visualization Standards:** Establishing standards for data visualization ensures consistency in how data is presented across the organization. This includes guidelines for colours, charts, and layouts to facilitate clear communication of insights.Enhanced visualization standards improve the effectiveness of data presentation, making it easier for users to interpret insights and draw actionable conclusions.

**11.** **Utilization of Data Pipelines:** Implementing data pipelines to automate the process of collecting, transforming, and loading data into BI systems ensures a consistent flow of high-quality data.Data pipelines enhance efficiency and reliability in data management, reducing manual interventions and ensuring that users have access to the most up-to-date information.

**12.** **Emphasis on Data Privacy and Security:** Prioritizing data privacy and security involves implementing measures to protect sensitive data from unauthorized access and breaches. This includes encryption, access controls, and compliance with data protection regulations (like GDPR).Emphasizing data privacy and security builds trust among customers and stakeholders, ensuring that sensitive information is handled responsibly and mitigating the risks of data breaches

**13.** **Cloud Integration**: Leveraging cloud-based BI solutions can offer scalability, flexibility, and cost savings. Cloud integration also facilitates remote access and collaboration.

**14.** **Data Quality Tools**: Utilizing data quality tools helps in identifying and correcting data issues. These tools can automate data cleansing processes and ensure that data is accurate and reliable.

**15. Incorporation of Blockchain for Data Integrity:** Utilizing blockchain technology enhances data integrity by providing a decentralized, tamper-proof record of transactions and data changes.Blockchain improves trust in data accuracy and provenance, making it especially valuable for industries where data security and authenticity are critical.