Participation W5

Exercise 1:

Define Big Data, and discuss its basic characteristics. Discuss the ways that data governance can address the common challenges in managing data.

Sample Answer

Definition of Big Data

Big Data refers to extremely large datasets that are beyond the ability of traditional data processing software to manage and process efficiently. These datasets are characterized by their enormous volume, high velocity, and wide variety of data types. Big Data often includes structured, semi-structured, and unstructured data gathered from various sources.

Basic Characteristics of Big Data

- Volume: This refers to the immense amount of data generated every second. It's not just the quantity of data but also the scale. Large organizations, social media platforms, and IoT devices contribute to this vast volume of data.
- Velocity: This is the speed at which new data is generated and the pace at which data moves. With the rise of real-time information streams, the velocity of data flow is higher than ever.
- Variety: Data comes in various formats - from structured, numeric data in traditional databases to unstructured text documents, emails, videos, audios, stock ticker data, and financial transactions.
- Veracity: This pertains to the quality and accuracy of data. Due to the diverse sources and types of data, ensuring high veracity can be challenging.
- Value: It's essential to distill the enormous volume of data into valuable insights. The ultimate goal of processing Big Data is to make informed decisions based on the analyses.

Data Governance in Addressing Challenges in Managing Data

Data governance refers to the overall management of the availability, usability, integrity, and security of the data employed in an organization. It encompasses a set of processes, policies, standards, and metrics that ensure the effective and efficient use of information in enabling an organization to achieve its goals.

- Ensuring Data Quality and Accuracy: Data governance programs establish processes to ensure that data is accurate and used consistently across an organization. This includes setting standards for data entry, storage, and maintenance.
- Compliance and Security: Data governance helps ensure that an organization complies with data-related regulations like GDPR or HIPAA. It sets policies for data access, encryption, and secure storage, thereby reducing the risk of data breaches and legal penalties.
- Efficient Data Management: By establishing clear guidelines and policies for data management, data governance helps organizations avoid data silos, ensuring that data is accessible and usable across different departments.
- Standardization and Integration: Data governance ensures that data across the organization is standardized, making it easier to integrate and use data from different sources. This is crucial in a Big Data environment where data comes in various formats and from multiple sources.
- Accountability and Transparency: Establishing clear roles and responsibilities as part of data governance ensures accountability in data management processes. This transparency is essential for building trust in data and the insights derived from it.
- Data Privacy: Data governance frameworks establish guidelines for handling personal and sensitive data, ensuring privacy and ethical use of data.

In summary, data governance plays a critical role in managing the challenges posed by Big Data. It helps in ensuring data quality, compliance, security, and efficient use of data, thereby enabling organizations to leverage Big Data for strategic decision-making.

Exercise 2:

What is meant by the term BI (Business Intelligence)? Describe and provide examples of the three different ways in which organizations use business intelligence (BI).

Specify the BI applications available to users for data analysis, and provide examples of how each one might be used to solve a business problem at RMIT?

Sample Answer

Definition of Business Intelligence (BI)

Business Intelligence (BI) refers to the technologies, applications, strategies, and practices used to collect, integrate, analyze, and present an organization's raw data to create insightful and actionable business information. The primary goal of BI is to support better business decision-making.

Three Ways Organizations Use Business Intelligence

1. Data-Driven Decision Making:

   Example: A retail company uses BI to analyze customer buying patterns, inventory levels, and supply chain efficiency. By understanding these patterns, the company can optimize stock levels, tailor marketing campaigns, and improve customer satisfaction.

2. Performance Metrics and Benchmarking:

   Example: A healthcare provider uses BI tools to track performance against industry standards. This includes monitoring patient wait times, treatment outcomes, and staff efficiency. By benchmarking against these metrics, they can identify areas for improvement.

3. Reporting and Analysis:

   Example: A financial services firm employs BI for regulatory reporting and market analysis. They generate reports on compliance, market trends, and customer portfolio performance, helping them to navigate regulatory environments and make informed investment decisions.

BI Applications for Data Analysis at RMIT

1. Data Visualization Tools (e.g., Tableau, Microsoft Power BI):
   - Usage: These tools can be used by RMIT to visualize student enrollment trends, course popularity, and academic performance metrics. This visualization aids in understanding patterns and making strategic decisions about course offerings and resource allocation.
2. Self-Service BI Tools (e.g., QlikView, TIBCO Spotfire):
   - Usage: RMIT can enable non-technical staff to generate ad-hoc reports on student demographics, course evaluations, or research output. This empowers staff to access information quickly without relying on IT departments, facilitating immediate and informed decision-making.
3. Predictive Analytics Tools (e.g., SAS, IBM SPSS):
   - Usage: These can be used for predicting future trends in student admissions or identifying which students might be at risk of dropping out. By analyzing historical data, RMIT can implement proactive measures to enhance student retention and success.
4. OLAP (Online Analytical Processing) Tools:
   - Usage: OLAP tools enable RMIT to perform multidimensional analysis of large volumes of data from different perspectives. This can be used for complex analyses, like comparing the performance of various departments over multiple years.

5. Data Mining Tools:
    - Usage: These tools can be employed to uncover hidden patterns and correlations in large datasets, such as discovering common factors among high-achieving students or identifying the most effective teaching methods.

In summary, BI encompasses various tools and applications that enable organizations like RMIT to make informed decisions based on data. By leveraging these tools, RMIT can enhance operational efficiency, improve student and staff experiences, and strategically plan for the future.