

**Worksheet – Vocabulary****Vocabulary**

Discuss the defining characteristics of each of the following:

<b><i>data</i></b>	<i>Data refers to raw facts and figures that are collected and recorded but not yet processed or organized into a meaningful form. Data can be numbers, characters, symbols, or images that, on their own, do not convey any specific meaning.</i>
<b><i>database</i></b>	<i>an organized collection of structured information, or data, typically stored electronically in a computer system</i>
<b><i>database management system (DBMS)</i></b>	<i>software system that enables users to define, create, maintain and control access to the database</i>
<b><i>CRUD</i></b>	<i>The four basic operations to perform on persistent computer storage: Create – add new records Read – retrieve existing records Update – modify existing data Delete – remove existing records</i>
<b><i>database schema</i></b>	<i>The schema is the blueprint that outlines the construction of the database's architecture. It provides the structural framework, defining how data is organized and the relationships among data entities. It includes the definition of tables, fields, relationships, views, indexes, and other elements.</i>
<b><i>data dictionary</i></b>	<i>a centralized repository of metadata that provides a comprehensive description of data elements within a system, including object names, data types, sizes, and classifications</i>

**Worksheet – Vocabulary**

<b><i>data validation</i></b>	<i>Validation is a key component of maintaining data integrity, preventing the entry of incorrect or invalid data, helping to maintain the accuracy and reliability of the database content.</i>
<b><i>data integrity</i></b>	<i>Data integrity refers to the maintenance and assurance of data accuracy and consistency over its entire lifecycle. Data needs to be recorded exactly as intended and remain unchanged upon later retrieval.</i>
<b><i>data redundancy</i></b>	<i>Redundancy refers to the condition where the same piece of data is stored in more than one place. While databases aim to minimize unnecessary data redundancy to optimize storage and performance, they also use controlled redundancy for data backup, fault tolerance, and high availability</i>
<b><i>data governance</i></b>	<i>Governance aims to maintain high-quality data that is secure and easily accessible. It involves the overall management of the availability, usability, integrity, and security. It encompasses the processes, policies, standards, and metrics.</i>
<b><i>big data</i></b>	<i>Big Data refers to extremely large data sets that may be analyzed computationally to reveal patterns, trends, and associations. It is characterized by 3 Vs: <b>Volume:</b> The quantity of generated and stored data. <b>Velocity:</b> The speed at which new data is generated and moves around. <b>Variety:</b> The different types of data (structured, unstructured, and semi-structured).</i>
<b><i>information</i></b>	<i>Information is data that has been processed, organized, or structured in a way that is meaningful and useful to the person who receives it. Information is typically the result of analysis performed on data, turning it into a form that can inform decision-making or provide insight.</i>