Name:

Student ID:

Reading Check 1

Chapters 1&2

**Instructions:** Please **highlight the answer(s)** that you are choosing. Once completed, please submit your highlighted version back into the assignment task. Ensure you **replace the name and student ID at the top of this page**. Please reach out to me if you have trouble uploading your assignment.

1. **In common theory, Big Data is described with 3 V’s; Volume, Velocity and Variety. The author mentioned that Analyst Doug Laney coined 4 V’s. What is considered the 4th V not mentioned previously? (1 point)**

The 4th V coined by Doug Laney is Veracity. It is described as the reliability or trustworthiness of data. Doug Laney emphasizes the importance of data being accurate and clean to maintain high data integrity.

1. **What are the 4 benefits of the Modern Data Stack (4 points)**

4 simplified benefits of the modern data stack:

Scalability: Cloud-based architecture that handles growing data volumes efficiently.

Speed: Enables faster insights and real-time decision-making.

Flexibility: Seamless data integration across disparate sources.

Security: Strong supporting functions for data governance and access controls.

1. **What key part of a high level architecture of the Modern Data Stack does Flexibility fall under? (1point)**

All transformations involve processing different data sources. This is important at the part of the system where data is taken and prepared, called data integration, the intake layer.

Flexibility usually falls under **Data Transformation** or **Integration Services** within the Modern Data Stack.

Example: A system that can take both structured data from a database and unstructured data such as email or video.

1. **Out of the following list, what are considered data sources? (1 point)**
2. **ERP**
3. **Operational Apps**
4. **Logs**
5. **All of the above**

All of the above. ERP systems, operational applications, and logs are all considered data sources in a modern data architecture.

1. **Name the 4 Supporting Functions of the Modern Data Stack (4 points)**

The 4 Supporting Functions are typically Data Discovery, Data Governance, Entitlements & Security, Data Observability.

Data Discovery: Identifying and locating data within an organization's systems. Example: Using tools to find customer information or sales data.

Data Governance: Establishing policies and processes to ensure data quality, security, and compliance. Example: Creating data quality standards and access controls.

Entitlements & Security: Managing access rights and protecting data from unauthorized access or breaches. Example: Implementing role-based access controls and encryption.

Data Observability: Monitoring data pipelines and systems to ensure reliability and performance. Example: Tracking data ingestion rates and error rates in pipelines.

1. **What are the 3 leading industry cloud environments for Data Warehousing? (3 points)**
   1. Amazon Redshift: Amazon's fully managed cloud data warehousing service, known for its worldwide scalability, high performance, and cost-effectiveness.
   2. Snowflake: A cloud-based data warehousing platform that separates storage and computing, allowing for independent scaling and offering flexibility and cost-effectiveness.
   3. Google BigQuery: Google's cloud-based data warehouse service, providing scalable resources for data processing and analytics, with a focus on ease of use and performance.
2. **What ecosystem helps automate the process of data ingestion from transactional databases, applications and other sources? (1 point)**

Data Pipeline Tools is an ecosystem of collaborative frameworks for delivering data from interactive databases, applications, and other resources. Examples of such tools are Apache NiFi, Apache Airflow, Talend, Stitch, and Fivetran. These tools facilitate the easy integration of data from various sources into the system for automated analysis. For example, they enable the automatic extraction of sales data from store systems, after which it is integrated without manual interception for analysis