Name: Shobharani

Student ID: S1365603

Reading Check 3

Chapter4

15points

**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. List the **3 Leading Cloud Environments** that is common within industry? (3 points)

Three leading cloud environments common within the industry are:

Amazon Web Services (AWS)

Microsoft Azure

Google Cloud Platform (GCP)

1. Building out a solid ETL pipeline is crucial in the ingestion phase of the Modern Data Stack, what are the **3 disruptors** that can happen during development (3 points)

Three disruptors that can happen during the development of an ETL (Extract, Transform, Load) pipeline in the ingestion phase of the Modern Data Stack are:

a. Changes in data schema: Modifications to the structure or format of the source data can disrupt the ETL process, requiring adjustments to the pipeline.

b. Connectivity issues: Interruptions or inconsistencies in data connectivity between source and destination systems can disrupt data ingestion and processing.

c. Data quality issues: Poor data quality, such as missing values, duplicates, or inconsistencies, can impact the effectiveness and reliability of the ETL pipeline.

1. What are the **4 KPIs** that used when evaluating your Data Warehouse? (4point)

Four KPIs used when evaluating your Data Warehouse include:

Query Performance: Time taken to execute queries.

Load Times: Efficiency of data ingestion processes.

Data Quality: Accuracy, completeness, and reliability of the data stored.

Cost Efficiency: Overall cost of data storage and processing.

1. **True of False** Data Lakes follow a *Top-Down* approach when attempting to ingest data from various sources (2 point)

False. Data Lakes typically follow a Bottom-Up approach when attempting to ingest data from various sources.

1. Name the **3 Failures** that will cause a Data Swamp to occur within your storage (3 points)

Three failures that can cause a Data Swamp to occur within your storage include:

Lack of Governance: Without proper data management and governance, data lakes can become unorganized.

Poor Data Quality Control: Failing to monitor and maintain the quality of ingested data.

Inadequate Metadata: Insufficient or missing metadata can make it difficult to understand, find, and use the stored data effectively.