Name: Shobharani Polasa

Student ID: S1365603

Reading Check 2

Chapter3

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 OLTP Database (Online Transaction Processing) applications** that is common within industry environments? (3 points)

Three common OLTP (Online Transaction Processing) database applications widely used in industry environments are:

Oracle Database: Known for its robustness and extensive features that support complex transactions.

Microsoft SQL Server: Popular in enterprise environments, offering comprehensive data management and business intelligence tools.

MySQL: An open-source RDBMS that is widely used for web applications and online transactions.

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 common disruptors that can occur during the development of an ETL (Extract, Transform, Load)

* Performance Bottlenecks: ETL processes often deal with large volumes of data, and performance bottlenecks can occur during extraction, transformation, or loading stages
* Data Quality Issues: Inconsistent, missing, or erroneous data can significantly disrupt the ETL process.
* Changes in Source Data Schema: Source systems may undergo changes in their data schema, such as the addition, modification, or removal of fields.

1. What are the **4 main challenges** that can occur during the Data Replication process? (4point)

Four main challenges that can occur during the data replication process are:

Consistency Issues: Ensuring data remains consistent across different systems can be challenging.

Performance Overhead: Replication can introduce latency and degrade the performance of the source systems.

Conflict Resolution: Handling conflicts in data when the same records are modified in multiple locations.

Bandwidth and Storage: Replication can consume significant network bandwidth and storage resources.

1. Out of the following list, what are questions to ask when strategizing a Reverse ETL? (1 point)
2. Which operational systems are required for this enrichment of data?
3. When should we update the data in the flow?
4. What value does this process add to the business?
5. All of the above

Answer: d. All of the above

1. Name the **4 Tips/Strategies** determine what data has to g o through the replication and transformation phase (4 points)

Four tips or strategies to determine what data has to go through the replication and transformation phase include:

Identify Business Needs: Understand what data is critical for business operations and decision-making.

Data Quality Assessment: Evaluate the quality of data to ensure only accurate and relevant data is processed.

Compliance and Security: Consider legal and security requirements that might dictate data handling processes.

Scalability and Performance: Ensure the selected data can be processed efficiently without impacting system performance.