Quiz 1: DSC 208 Data Management for Analytics

Questions and Explanations

Question 1. Which of these structured data models has native support for mixed types in columns?

a) Matrix

b) Relation

c) All of the three

d) DataFrame

Answer: d) DataFrame

- *Explanation:* DataFrames, commonly found in libraries like Pandas or Spark, are designed to handle heterogeneous data types within different columns, similar to a spreadsheet or a table in a relational database. Matrices typically require all elements to be of the same numeric type. Relations (in the context of relational databases) enforce strict data types for each column, though they can vary *between* columns.

Question 2. In what way is data access control often a challenge in the data sourcing stage?

a) Limits access to some ML software

b) Prevents use of some data

c) Raises computational resource costs

d) Requires downsampling of data

Answer: b) Prevents use of some data

- *Explanation:* Data access control mechanisms, while crucial for security and privacy, can often restrict data scientists from accessing necessary datasets due to permissions, compliance regulations, or internal policies. This directly impacts the ability to source and utilize all potentially relevant data for an ML project.

Question 3. Which type of integrity constraint ensures that a value in one table must match a value in another table?

a) Functional dependency

b) Domain integrity

c) Referential integrity

d) Entity integrity

Answer: c) Referential integrity

- *Explanation:* Referential integrity is enforced using foreign keys, which establish a link between data in two tables. It ensures that a foreign key value in the referencing table corresponds to an existing primary key value in the referenced table, preventing orphaned records and maintaining consistency across related data.

Question 4. What is the purpose of a primary key in a relational database table?

a) To improve performance

b) To ensure data is ordered

c) To ensure data is unique

d) To enforce referential integrity

Answer: c) To ensure data is unique

- *Explanation:* The fundamental purpose of a primary key is to uniquely identify each record (row) in a table. While it can be used to improve performance (often by indexing) and is essential for referential integrity, its core role is to guarantee uniqueness.

Question 5. What is the purpose of a foreign key in a relational database table?

a) To improve performance

b) To ensure data is ordered

c) To enforce referential integrity

d) To ensure data is unique

Answer: c) To enforce referential integrity

- *Explanation:* As discussed in Question 3's explanation, a foreign key's primary role is to enforce referential integrity. It creates a link between two tables, ensuring that relationships between data are valid and consistent.