Table

Description automatically generated with low confidence

1. Describe the basic features of the relational data model, and discuss their importance to the end user and the designer.
   1. ACID compliance
   2. Normalized
   3. Designed to get data securely into databases
2. Explain how the entity relationship (ER) model helped produce a more structured relational database design environment.
   1. Help standardize data and remove
3. What does it mean cardinality within the context of entity relationship diagram?
   1. Basically describes the relationship (1:1, 1:many, many:many)
4. What three data anomalies are likely to be the result of data redundancy? How can such anomalies be eliminated?
   1. Update, deletion, and addition anomalies. These can be eliminated by data normalization
5. What does it mean data integrity within the context of relational database applications?
   1. Completeness, accuracy and consistency
6. What does it mean referential integrity within the context of relational database applications?
   1. Relationships between tables (foreign keys)
7. Define and discuss the concept of partial and transitive dependencies.
   1. Partial is when a column only is defined by a subset of a tables primary key (2nf fixes this)
   2. Transitive dependencie is when one defines another thing which defines the third, 3nf removes this
8. Is it possible that a relation be in 3NF but not in 2NF? Explain
   1. No because for it to be in 3NF it has to be 2NF AND have no transactional functional dependencies
9. Illustrate by an example a scenario where an attribute has unique values in the different rows yet it can’t be used practically as a primary key in the database relations/tables.
10. Explain the difference between structured and unstructured data
    1. Specific format and data integrit
11. What is the difference between relation and relationship within the context of relational data modeling and Entity Relationship Diagram
12. We use SQL to retrieve data from a database, can we use SQL as a general purpose programming language to clean, preprocess, and visualize data to generate information reports? Explain.
13. Discuss the difference between OLAP and OLTP
    1. In summary, OLTP systems are designed for transaction processing, while OLAP systems are designed for data analysis and decision-making. They have different data models, processing requirements, and performance characteristics, and are optimized for different types of workloads.
14. What is the star-schema that is used in data warehousing?
    1. Fact tbale (all keys) radiate out to values
15. Imagine you loaded your dataset into a Python DataFrame object, which method on the dataframe object could you execute to find rows/columns with missing values?
    1. Is null
16. What graph/chart that you could use in Python to visualize the data and get an idea if there are outliers/dirty/messy/invalid data values?
    1. Box plot
    2. histrogram
17. When do you use the HAVING clause when you have a SELECT statement in SQL that has the GROUP BY clause?
    1. Having helps filter values by the result of the group by
18. Explain the difference between natural join and outer join for relational tables.
    1. Inner vs outer basically
19. Explain the differences between the four database types we studied this quarter; consider the features and benefits of each type.
20. Illustrate by example the use of inverted index for information retrieval.
21. Precision and recall are widely used statistics to assess the effectiveness of the IR system, explain.
    1. Precisions is relevant divided by all returned
    2. Recall is returned relevant devided by all relevant
22. Discuss the difference between fuzzy queries and exact matches within the context of NoSQL database engines.
    1. Fuzzy queries are not exact add variances
    2. Fuzzy is one character
       1. “child~” - fuzzy
       2. “\*child\*” -regex
23. Which operator is used in a WHERE clause for the SELECT/SQL statement in order to search for a specific pattern in column values?
    1. LIKE