Congratulations! You passed!
TO PASS 80% or higher

Keep Learning

GRADE
100%

Techniques: Data Quality Assessments

LATEST SUBMISSION GRADE
100%

1. What features distinguish intrinsic data quality from fitness for use data quality (all that apply): **1 / 1 point**

- [ ] Fitness for use data quality focuses on data completeness (missingness) whereas intrinsic data quality focuses on data accuracy
- [ ] Intrinsic data quality measures are called intrinsic because they apply only to a local database whereas fitness for use data quality measures can be shared with other data partners
- [x] Intrinsic data quality can be pre-specified whereas fitness for use data quality must be defined with each project

Correct
Correct! This is the core difference

- [x] Intrinsic data quality measures are easier to specify since they are less complex

Correct
Correct!

2. A data quality program measures the same data quality measures after each data refresh. Following the most-recent refresh, a large jump in the number of data quality issues compared to the last refresh was observed at one of the sites. Which of the following could explain this observation: **1 / 1 point**

- [x] New data tables were added to the ETL process

Correct
Correct!

- [x] Additional data quality rules were added to the program

Correct
Correct!

- [ ] A new data partner was added to the network
- [x] The source database was changed

Correct
Correct!

3. SQL alone is not often used for computing data quality measures. Some of the reasons for not using SQL only are (all that apply): **1 / 1 point**

- [ ] SQL is a difficult language to learn so not many people know how to program using SQL.
- [x] SQL is optimized for extracting data, not for computation.

Correct

- [ ] SQL has a large set of aggregation functions that makes it difficult to know now best to program data quality calculations.
- [x] SQL cannot perform the same range of computations as can a programming language.

Correct
SQL is not "Turing Complete" whereas nearly all programming languages are. "Turing Complete" is a term used to describe the breadth of computations that a language can support.

4. Which of the following statements are false? **1 / 1 point**

- [ ] Data quality measures are grouped into data quality dimensions.
- [ ] A data quality measure may appear in multiple data quality dimensions.
- [x] A data quality rule uses a data quality dimension to determine if a data quality measure is acceptable.

Correct
The criteria used by a DQ rule is set by the person who writes the rule based on their needs/use case.

- [x] A data quality dimension determines if the ETL programming was done correctly

Correct
DQ dimensions are simply a was of categorizing data quality issues, not to identify the source of those issues.

5. Which of the following statements is false? **1 / 1 point**

- ( ) Data profiling tools that provide insights into data quality by revealing errors in ETL mappings.
- ( ) Data profiling tools that provide insights into data quality by revealing unusual distributions of data values.
- ( ) Data profiling tools that provide insights into data quality by revealing issues between two or more variables.
- (•) Data profiling tools that provide insights into data quality by determining if data are acceptable for use.

Correct
This is the role of DQ rules