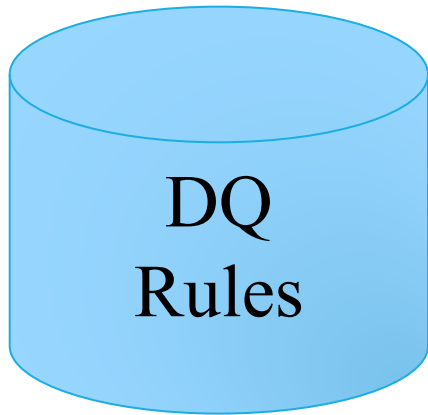




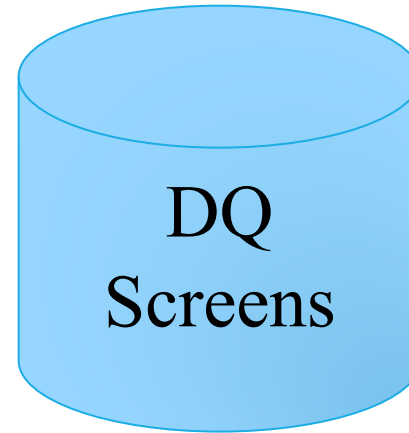
Data Quality Rules and Screens

School of Information Studies
Syracuse University

Data Quality Databases



- A database of rules and actions for each rule (accept, allow, warn, or fix).
- Keep rules simple.



- A database of actual data that failed a DQ screen.
- Includes information about which rule triggered the screen.

Five Criteria of Data Quality

- **Five criteria:** accuracy, precision, timeliness, completeness, consistency.
- For any set of data we determine a **data reliability** score based on one or more of the five factors.
- We write this score as a formula representing the percentage reliability of the data. Score should be between 0 and 1.
- This information is for the business user.
- Who determines this score? It's why you need data governance!

Example Incoming Validation DQ Rules: GPA

- DQ rule 1: accurate GPA
 - GPA must be a between 0.000 and 4.000.
 - Actions: REJECT.
- DQ rule 2: precise GPA
 - GPA must be three decimal places.
 - Actions: WARN.
- DQ rule 3: complete GPA
 - GPA must not be null.
 - Actions: FIX—Replace null with 0.000.
- Whenever a rule is triggered, it is written to the DQ screens database.

Three Types of DQ Rules

1. Incoming data validation

- Rules are checked as staged data enters the DW (NDS, ODS, or DDS). Most common.

2. Cross-reference validation

- Rules to check incoming data against the data already in the DW.
- E.g., Monday's website visitors against a running average of the last three Mondays +/- 15%

3. Data warehouse internal validation

- Check DW data against itself, typically for aggregates.
- E.g., yearly sales in summary table matches actual number of sales by year.