



Data Cleansing and Conforming

School of Information Studies
Syracuse University

Cleanse and Conform Systems



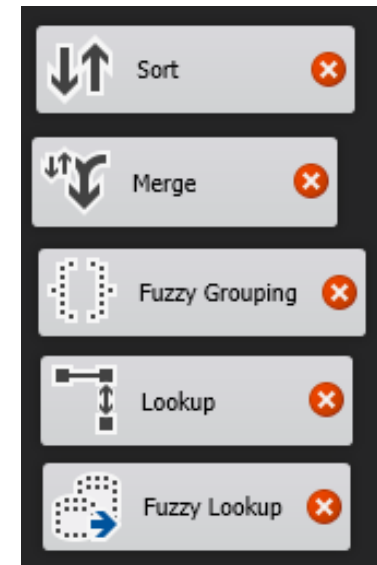
Data-Cleansing System

- Balance these conflicting goals:
 - Fix dirty data yet maintain data accuracy.
- Quality screens act as diagnostic filters:
 - **Column screens:** test data in fields
 - **Structure screens:** test data relationships, lookups
 - **Business rule screens:** test business logic
- Responding to quality events:
 - Fix (e.g., replace NULL w/value)
 - Log error and continue or abort (depending on severity)



Deduplication System

- When dimensions are derived from several sources
 - E.g., customer information merges from several lines of business
- **Survivorship:** the process of combining a set of matched records into unified image of authoritative data
- **Master data management:** centralized facilities to store master copies of data



Conforming System

- Responsible for creating conformed dimensions and facts.
- Typically conformed dimensions are managed in one place and distributed as a copy into the required dimensional model.

