AUTOMATING DATA QUALITY MANAGEMENT

A resource guide

SCHEMA VALIDATION

- What is it? Ensures data adheres to a predefined data model or schema
- How does it do it? Maintains consistency by aligning data types with schema definitions
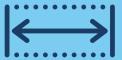- Example: Validating a JSON file against a JSON schema

SYNTAX CHECKING

- What is it? Automates the detection of syntax errors in data files
- How does it do it? Checks for missing or extra delimiters, incorrect field separators, or misplaced quotation marks
- Example: Scanning a CSV file for comma inconsistencies or missing line terminators

RANGE VERIFICATION

- What is it? Checks if data values fall within specified ranges or constraints
- How does it do it? Identifies out-of-range or invalid values based on predefined rules
- Example: Verifying that an age value is within a valid range (e.g., 0-120 years)

PATTERN RECOGNITION

- What is it? Validates data against predefined patterns or regular expressions
- How does it do it? Useful for enforcing specific formats or structures (e.g., email addresses, phone numbers)
- Example: Ensuring email addresses follow a valid pattern (name@domain.com)