Data quality rule

Definition

A data quality rule is a rule used to validate data values or data records.

Synonyms

- Data validation rule
- Data quality rule specification

Purpose

The purpose of a data quality rule is to prevent, monitor or report insufficient quality of data values and data records.

Life cycle

Phase	Activity
Plan	* Derive data quality rules from business rules * Collect/identify data quality rules * Specify data quality rules * Document data quality rules as metadata * Establish data quality rules
Do	* Implement/apply data quality rules in a database or application.
Check	* Test data quality rules * Evaluate data quality rules
Act	* Adapt/maintain data quality rules

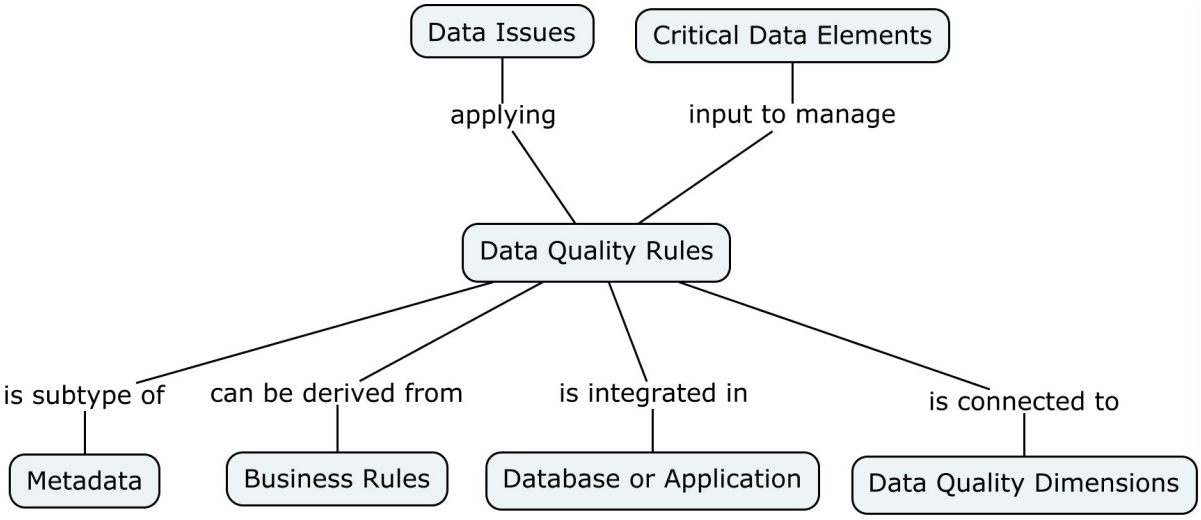
Characteristics and requirements

Characteristic	Requirement
Accessibility	Data quality rules are accessible to data users. It ensures that data users can ask questions about it and provide feedback on rules.
Unambiguity	Data quality rules are unambiguous. It ensures that they can be implemented in a database or application.
Maintainability	Data quality can efficiently being maintained.

Relations

Data quality rule	is child of	metadata
Data quality rule	is child of	rule
Data quality rule	is element of a	data quality management system
Data quality rule	is associated with specific	data quality dimensions
Data quality rule	can be applied to	data values
Data quality rule	can be applied to	accuracy

Data quality rule	can be applied to	completeness
Data quality rule	can be applied to	consistency
Data quality rule	can be applied to	data records
Data quality rule	can be applied to	uniqueness
Data quality rule	can be integrated in a	database
Data quality rule	can be integreted in an	application
Data quality rule	is derived from a	business rule
Data quality rule	prevents	data issues
Data quality rule	is applied firstly to	critical data elements
Data quality rule	is needed for	data quality monitoring
Data quality rule	is needed for	data cleansing



Classification A

In table 1 Data quality rules are classified in three categories, and their subcategories.

Table 1: Category, subcategories, and examples.

Category and subcategory	**Example**
Simple data element content rules. **These are considered “simple” because you only need to inspect the contents of a single data element and check to see if the content meets the rules.**	
Valid values, range, data type, pattern, and domain.	
Optional versus mandatory (evaluates completeness).	
Reasonable distribution of values.	In a customer database, you would expect a fairly even distribution of birthdays; a much larger number of birthdays on a given day of the year probably indicates a problem.
Cross data element validation rules. **The rules require inspecting values in multiple data elements (typically in a single data file) to determine whether the data meets the quality rules.**	

Category and subcategory	Example
Valid values that depend on other column values	An overall list of location codes might pass the simple-column content rules, but only a smaller list of locations is valid if, for example, the region code is set to “West.”
Optional becomes mandatory when other columns contain certain data.	The Value of Collateral field may be optional, but if the loan type is “mortgage,” a positive value must be filled into the Value of Collateral field.
Mandatory becomes null when other columns contain certain data.	The Writing Insurance Agent Name field might normally be mandatory, but if the Origination Point is “web” (indicating the customer applied for the policy online), the Writing Insurance Agent Name field must then be blank.
Cross-table validation rule. As the name suggests, these Data quality rules check columns (and combinations of columns) across tables.	An example cross-validates the name of a city with the name of a state in an address table—that is, Minneapolis is not in Wisconsin.
Cross data files validation rules. As the name suggests, these Data quality rules check data elements (and combinations of data elements) across data files.	
Mandatory presence of foreign-key relationships	If an account must have a customer, then the account table must have value in the Customer ID column that matches a value in the Customer ID column of the Customer table.
Optional presence of foreign-key relationships depending on other data	If the Loan Type is “mortgage” in the Loan table, there must be a matching value for Loan_ID in the Collateral table. On the other hand, if the Loan_Type is “unsecured,” then there must not be a matching value for Loan_ID in the Collateral table because “unsecured” means there is no collateral for the loan.
Columns in different tables are consistent	If the Collateral_Value column contains a value above a certain level, the Appraisal_Type must be “in person” because of the high value of the property.

Classification B

In Table 2 Data quality rules are classified in ten categories.

Table 2: Category of Data quality rule, description, and example

Category	Description	Example
Domain List	A domain list rule defines a list of values that a data element is allowed to have.	The Gender data element can have 'M' or 'F'.
Domain Pattern List	A domain pattern list rule defines a list of patterns that a data element is allowed to conform to. The patterns are defined in the regular expression syntax.	An example pattern for a telephone number is as follows: (^[[::space]]*[0-9]{ 3 }[[::punct|:space:]]?[0-9]{ 4 }[[::space]]*$)
Domain Range	A domain range rule defines a range of values that a data element is allowed to have.	The value of the salary data element can be between 100 and 10000.

Category	Description	Example
Common Format / Pattern	A common format rule defines a known common format that a data element is allowed to conform to.	This rule type has many subtypes: Telephone Number, IP Address, SSN, URL, E-mail Address.
No Nulls	A no nulls rule specifies that the data element cannot have null values	The department_id data element for an employee in the Employees table cannot be null.
Functional Dependency	A functional dependency defines that the data in the data object may be normalized or derived	
Unique Key	A unique key data rule defines whether a data element or group of data elements are unique in the given data object.	The name of a department should be unique.
Referential	A referential data rule defines the type of a relationship (1:x) a value must have to another value.	The department_id data element of the Departments table should have a 1:n relationship with the department_id data element of the Employees table.
Name and address	A name and address data rule evaluate a group of data elements as a name or address	
Custom	A custom data rule applies a SQL expression that you specify to its input parameters.	VALID_DATE with two input parameters, START_DATE and END_DATE. A valid expression for this rule is: “THIS”.“END_DATE” > “THIS”.“START_DATE

Examples A

Table 3 shows how a Data quality rule is derived from a business rule.

Table 3: Example of a Data quality rules derived from Business Rule

Business Rule	Data quality rule
The marital status code may have values of single, married, widowed, and divorced. It may not be left blank. A value must be picked when entering a new customer. The values for widowed and divorced are tracked separately from single because risk factors are sensitive to whether the customer was previously married and is not married anymore.	“Customer.Mar_Stat_Cd” may be “S,” “M,” “W,” or “D.” Blank is considered an invalid value.

Examples B

Table 4 shows Data quality rules of two fields.

Table 4: Example of Data quality rules

Data element	Data quality rules
Email	* An email must contain the '@' sign. * '@' must be used only once. * An email must contain any or all of the following: letters, digits, non-alphabetic characters, such as, ! # $ % & ' * + – / = ?
Customer name	* The 'Customer's full name' field an email refers to must not be 'Null'. * Customer's full name must consist only of letters; no other characters allowed. * Only first letters in customer name, middle name (if any), and surname must be capitalized.

Tips

1. Involve subject matter experts from various department.
2. Be moderate with the number of rules. Define rules that really matter according to the subject matter expert. Don't create rules which already are implemented in the sources system as input check.
3. Favour a step-by-step approach.
4. Treat each field of the database individually and creates rules accordingly.
5. Decide between a centralized and local storage for your Data quality rules.

References

Data quality rule (2023). YouTube video. DAMA-NL.

DAMA (2017). DAMA-DMBOK. Data Management Body of Knowledge. 2nd Edition. Technics PublicationsLlc. August 2017. DAMA Dictionary of Data Management.

Data migration checklist: Planner + template for effective data migration planning — Data migration pro. (n.d.).

Data Migration Pro. https://www.datamigrationpro.com/data-quality-rules-management-repository

Data quality - Data rules. (2020, March 31). Datacadamia - Data and Co. https://datacadamia.com/data/quality/data_rule

Data quality and business rules explained: Expert interview with Ronald G. Ross — Data quality pro. (2020, February 29). Data Quality Pro. https://www.dataqualitypro.com/blog/business-rules-for-data-quality-ronald-g-ross

Data quality rule. (n.d.). ScienceDirect.com | Science, health and medical journals, full text articles and books. https://www.sciencedirect.com/topics/computer-science/data-quality-rule

Reeb, B. (n.d.). Data quality rules. IData Insights Blog. https://blog.idatainc.com/data-quality-rules

Tips to create effective Data quality rules. (2016, April 29). TDAN.com. https://tdan.com/tips-to-create-effective-data-quality-rules/24525

What are Data quality rules? How do they look, and why are they needed in data management? Our expert provides the answers! (2020, September 3). Data Quality Solutions: Data Quality Software & DQaaS | CDQ. https://www.cdq.ch/data-sharing/data-quality-rules

All, DQMS

From:
https://datamanagement.wiki/ - **Data Management Wiki**

Permanent link:
https://datamanagement.wiki/data_quality_management_system/data_quality_rule

Last update: **2024/03/08 13:33**