



CIKLUM

SPEAKER'S
CORNER

IN PARTNERSHIP WITH



OCTOBER, 25

17:30 (EEST)

UKRAINIAN

WHY DATA QUALITY MANAGEMENT IS VITAL



MYKOLA POBYIVOVK
DATA TECH LEAD AT CIKLUM

Meet Ciklum



We empower companies to meet their digital initiatives by providing end-to-software, integration and innovation services

Our services:

Digital commerce

Intelligent
Automation

ITO & Managed
service

Data & Analytics

Cloud

Engineering
Services

2002
founded

4000+
professionals

20+
offices

300+
clients

Leading companies choose us:



Speaker:

Mykola Pobyivovk

Data tech Lead at Ciklum

- More than 13+ years of experience in IT, GMBA
- 12+ years of experience working with Microsoft Data Platform
- Co-organizer #UDCK and SQL Saturday Kyiv (2016, 2017, 2018,2019)
- Mentor at Lviv IT Cluster
- Ukrainian defenders volunteer



Fundraising for tactical medicine

My volunteer project, DiyTu (ДійТи), aims to support the Ukrainian army!

Since the start of the full-scale war we have acquired valuable equipment worth 6 million UAH for the Armed Forces.

We stand for our warriors and urge you to do the same: support us with donations, so that we may cover as many of their requests as possible.



FB: [diytu.volunteer](https://www.facebook.com/diytu.volunteer)

Agenda

01

How Poor Data Quality Affects Businesses

02

Data Quality Management in Data Engineering

03

Data Quality Process

04

Demo

Agenda

05

Best Practices of Data Quality Management

06

Summary

Is Data a new oil?

- Decision-making
- Go-to-market activities
- ML\AI
- R&D
- Analytics(Behavioral,Predictive, Customer Sentiment and so on)
- Product experiences
- Risk mitigation
- Customer Segmentation
- Fraud detection



How Poor Data Quality Affects Businesses

An Experian study revealed that 55% of business leaders lack trust in their business assets.

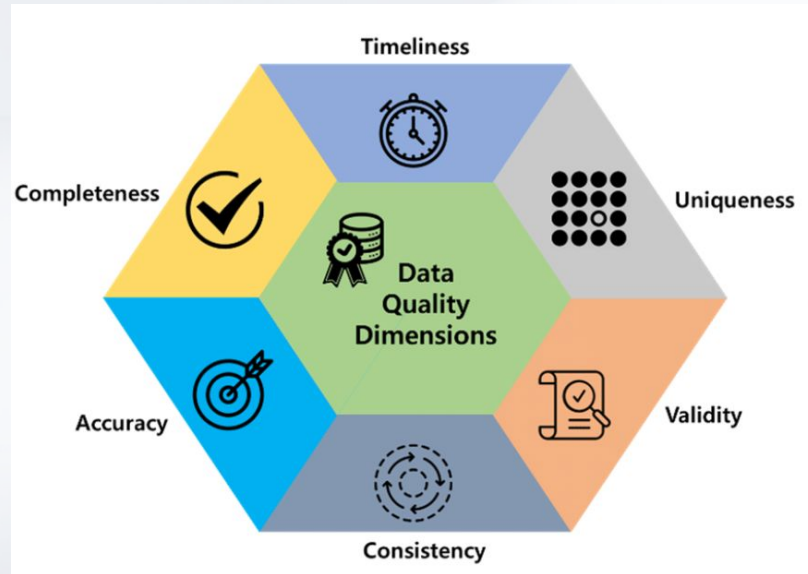


SOURCE: Created by author with data from the 2021 Experian Research

Data Quality Management in Data Engineering

Data quality management, or DQM, is a technique used by data organizations to maintain the quality of data. In other words, DQM helps to keep the data clean from any errors, keep it up-to-date and well-managed, and ready to use by the experts for further process.

Data Quality Dimensions



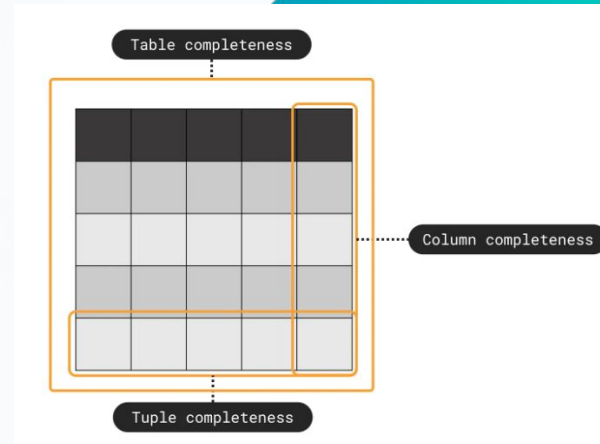
- Consistency
- Accuracy
- Completeness
- Validity
- Uniqueness
- Timeliness.

Completeness

Completeness shows the presence of sufficient data to draw conclusions. Therefore, each data entry should be complete and all the required fields must be filled out.

The completeness is often related to the number of missing values in a record (tuple completeness), columns (column completeness) or table (relation completeness).

ContractID	First Name	Last Name	Currency Code	Started From	Finished To
1	Петро	Карпенко	UAH	NULL	01-01-25
2	Тарас	Пилипенко	UAH	01-01-00	01-01-18
5	Іван	Перекотиполо	UAH	01-01-01	01-01-12
4	Андрій	Непийпиво	Гривня	01-06-21	01-01-25
5	Іван	Перекотиполо	UAH	01-01-18	01-01-30
6	Петро	Карась	UAH	NULL	NULL
...					



Accuracy

EXAMPLE

Accuracy

ContractID	First Name	Last Name	Currency Code	Started From	Finished To
1	Петро	Карленко	UAH	NULL	01-01-25
2	Тарас	Пилипенко	UAH	01-01-00	01-01-18
5	Іван	Перекотиполо	UAH	01-01-01	01-01-12
4	Андрій	Непийпиво	Гривня	01-06-21	01-01-25
5	Іван	Перекотиполо	UAH	01-01-18	01-01-30
6	Петро	Карась	UAH	NULL	NULL
...					

The quantity and types of mistakes in a data set are measured by accuracy. The data should be free from any spelling errors, capitalization issues, missing details, or any other issues that may decrease the quality of the data.

Consistency

This metric looks at how consistent the data is across different sources. It's essential to have consistent data for it to be useful for decision-making.

EXAMPLE

ContractID	First Name	Last Name	Currency Code	Started From	Finished To
1	Петро	Карпенко	UAH	NULL	01-01-25
2	Тарас	Пилипенко	UAH	01-01-00	01-01-18
5	Іван	Перекотипале	UAH	01-01-01	01-01-12
4	Андрій	Непийпиво	Гривня	01-06-21	01-01-25
5	Іван	Перекотипале	UAH	01-01-18	01-01-30
6	Петро	Карась	UAH	NULL	NULL
...					

Validity

Validity gauges how closely data complies with requirements. In other words, information placed into a field should be of the right type; otherwise, it may not be used. It's critical to ensure that each piece of data complies with the necessary data types and formats.

EXAMPLE

ContractID	First Name	Last Name	Currency Code	Started From	Finished To
1	Петро	Карпенко	UAH	NULL	01-01-25
2	Тарас	Пилипенко	UAH	01-01-00	01-01-18
5	Іван	Перекотиполе	UAH	01-01-01	01-01-12
4	Андрій	Непийпиво	Гривня	01-06-21	01-01-25
5	Іван	Перекотиполе	UAH	01-01-18	01-01-30
6	Петро	Карась	UAH	NULL	NULL
...					

Uniqueness

This dimension captures the number of duplicated records. However, there are many different levels of duplications and each requires different methods to measure.

EXAMPLE

ContractID	First Name	Last Name	Currency Code	Started From	Finished To
1	Петро	Карпенко	UAH	NULL	01-01-25
2	Тарас	Пилипенко	UAH	01-01-00	01-01-18
5	Іван	Перекотиполе	UAH	01-01-01	01-01-12
4	Андрій	Непийпиво	Гривня	01-06-21	01-01-25
5	Іван	Перекотиполе	UAH	01-01-18	01-01-30
6	Петро	Карась	UAH	NULL	NULL
...					

Timeliness

The timeliness dimension is a measurement of the delay between an event occurring and the data being available to the business. It is important to remember that the data is still valid, it is just late.

EXAMPLE

ContractID	First Name	Last Name	Currency Code	Started From	Finished To
1	Петро	Карпенко	UAH	NULL	01-01-25
2	Тарас	Пилипенко	UAH	01-01-00	01-01-18
5	Іван	Перекотиполе	UAH	01-01-01	01-01-12
4	Андрій	Непийпиво	Гривня	01-06-21	01-01-25
5	Іван	Перекотиполе	UAH	01-01-18	01-01-30
6	Петро	Карась	UAH	NULL	NULL
...					

Data quality rules repository

Why we need a dedicated Data Quality Rules Repository

- Poor Data Quality is a fact
- Most organizations tackle Data Quality ad-hoc
- Maintaining Data Quality ad-hoc is inefficient, time consuming and inconsistent
- Reusability of patterns, rules and components

Data quality rules Examples

Data Quality Dimension	Object Type	Object Name	Data Quality Rule(Example)
Consistency	Column attribute	Table1.Currency	Currency needs to be in the list of allowed ISO Values
Validity	Column attribute	Table1.DateFrom	Should be more than 1995 year
Consistency	Column attribute	Table1.Gender	Gender needs to be M or F.
Completeness	Column attribute	Table1.Name	Name is mandatory
Validity	Column attribute	Table1.Email	E-mail address needs to be in right format

Data quality process

Data profiling

Data Quality Analysis

- Validation

Data Cleansing

- Matching
- Eliminate errors, duplicates, inconsistencies

Data Quality Monitoring

- Reporting and Monitoring

Examples of DQ tools: Informatica data quality, Microsoft data quality service, Talend Open Studio for DQ and so on



Demo



Best practices of data quality management

- Implement A Single Source Of Truth
- Draw Meaningful Insights From Issues
- Evaluate Data On A Regular Basis
- Perform Data Governance
- Automating data entry
- Create A Data Dictionary

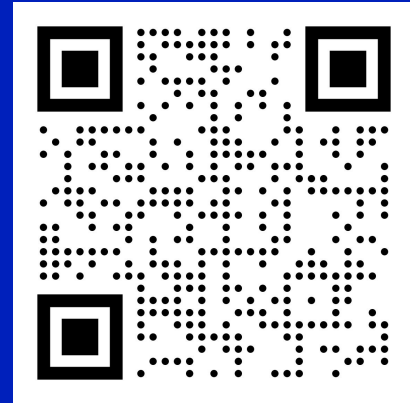
Summary

References

- <https://www.udemy.com/course/data-warehouse-etl-testing-data-quality-management-a-z/>
- <https://app.datacamp.com/learn/courses/introduction-to-data-quality>
- <https://blog.det.life/data-quality-management-in-data-engineering-why-how-and-what-7fdc2b8861e5>
- <https://www.metaplane.dev/blog/data-quality-metrics-for-data-warehouses>
- <https://www.scnsoft.com/blog/guide-to-data-quality-management>
- <https://www.erp-information.com/data-quality-metrics>
- <https://edqcdncmsprod.azureedge.net/491de5/globalassets/white-papers/global-data-management-research-2021.pdf>
- <https://firsteigen.com/blog/6-key-data-quality-metrics-you-should-be-tracking/>
- https://grindsuccess.com/data-quality-management/?expand_article=1
- <https://blog.det.life/data-quality-management-in-data-engineering-why-how-and-what-7fdc2b8861e5>
- <https://towardsdatascience.com/a-comprehensive-framework-for-data-quality-management-b110a0465e83>
- <https://christianlauer90.medium.com/3-best-practices-for-data-quality-management-462730f3a835>

Any questions?

Share your feedback!



Thank you!



Mykola Pobyivovk
nipo@ciklum.com | www.ciklum.com