



Cloud1 Data Quality Analysis

ARCHITECTS OF DIGITALIZATION

CLOUD1

Why data quality is important

According to several studies:

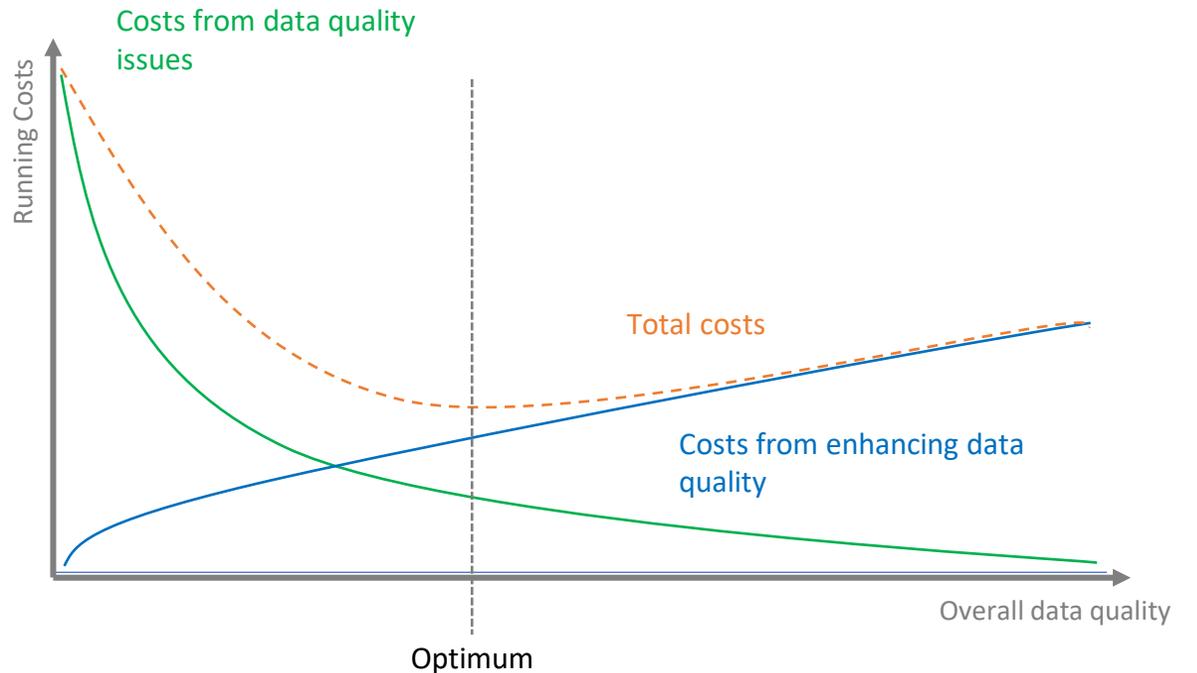
- Bad data costs companies an estimated 15% of their turnover
- Nearly a third of data analysts spend more than 40% of their working time verifying and validating the data used in analysis
- Knowledge workers waste 50% of their time hunting for information, finding and correcting errors, and finding confirmatory data sources to replace those they don't trust
- It is estimated that up to 20-30% of operational costs arise from bad data

<https://dataladder.com/what-is-data-accuracy/>

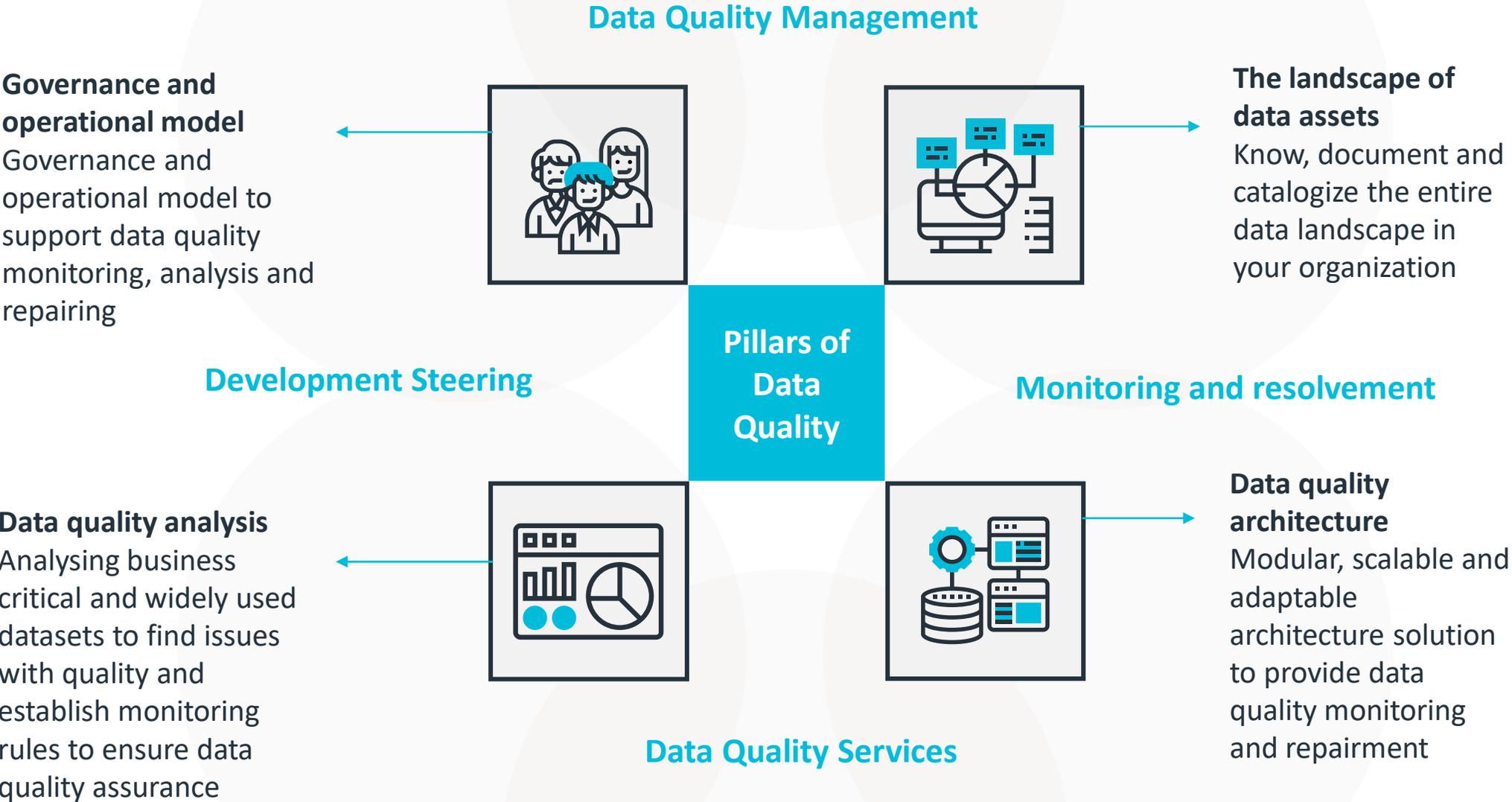
Optimum for data quality effort

100 % quality is not achievable

- Data is always changing and always flowing
- Fixing old errors do not make new data batches correct
- 100% quality 100% of the time is not possible
- Optimum can be found balancing costs from data quality and costs generating quality assurance



Cloud1 Pillars of Data Quality

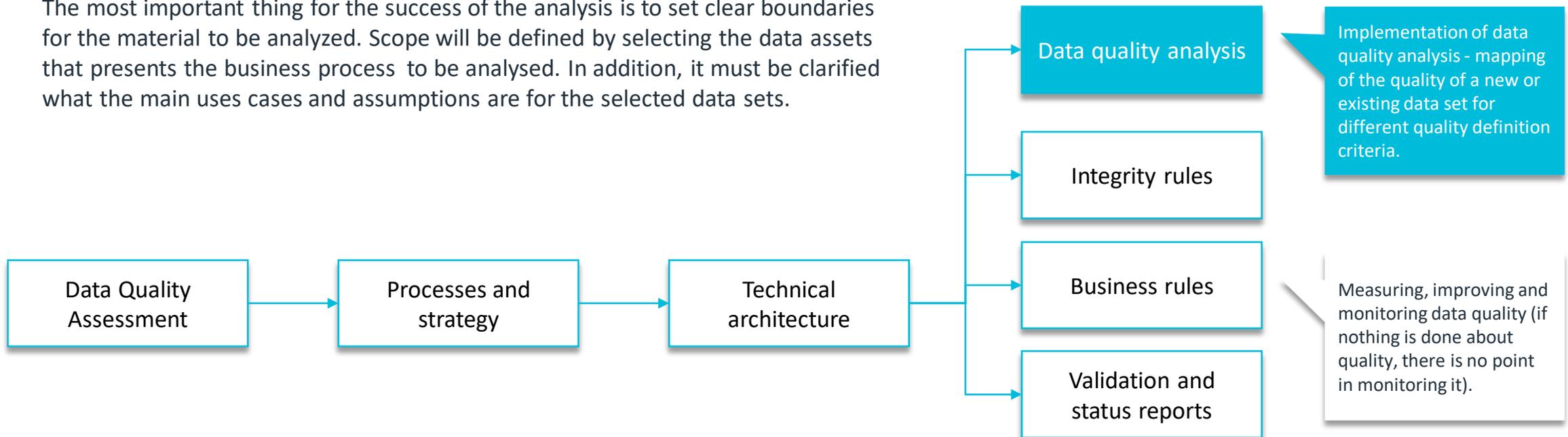


Framework

How should you proceed with improving data quality?

Data quality analysis delves into the quality of data sets guided by business vision and information. Data quality analysis should be performed in close collaboration with business experts, utilizing the latest analysis tools.

The most important thing for the success of the analysis is to set clear boundaries for the material to be analyzed. Scope will be defined by selecting the data assets that presents the business process to be analysed. In addition, it must be clarified what the main uses cases and assumptions are for the selected data sets.

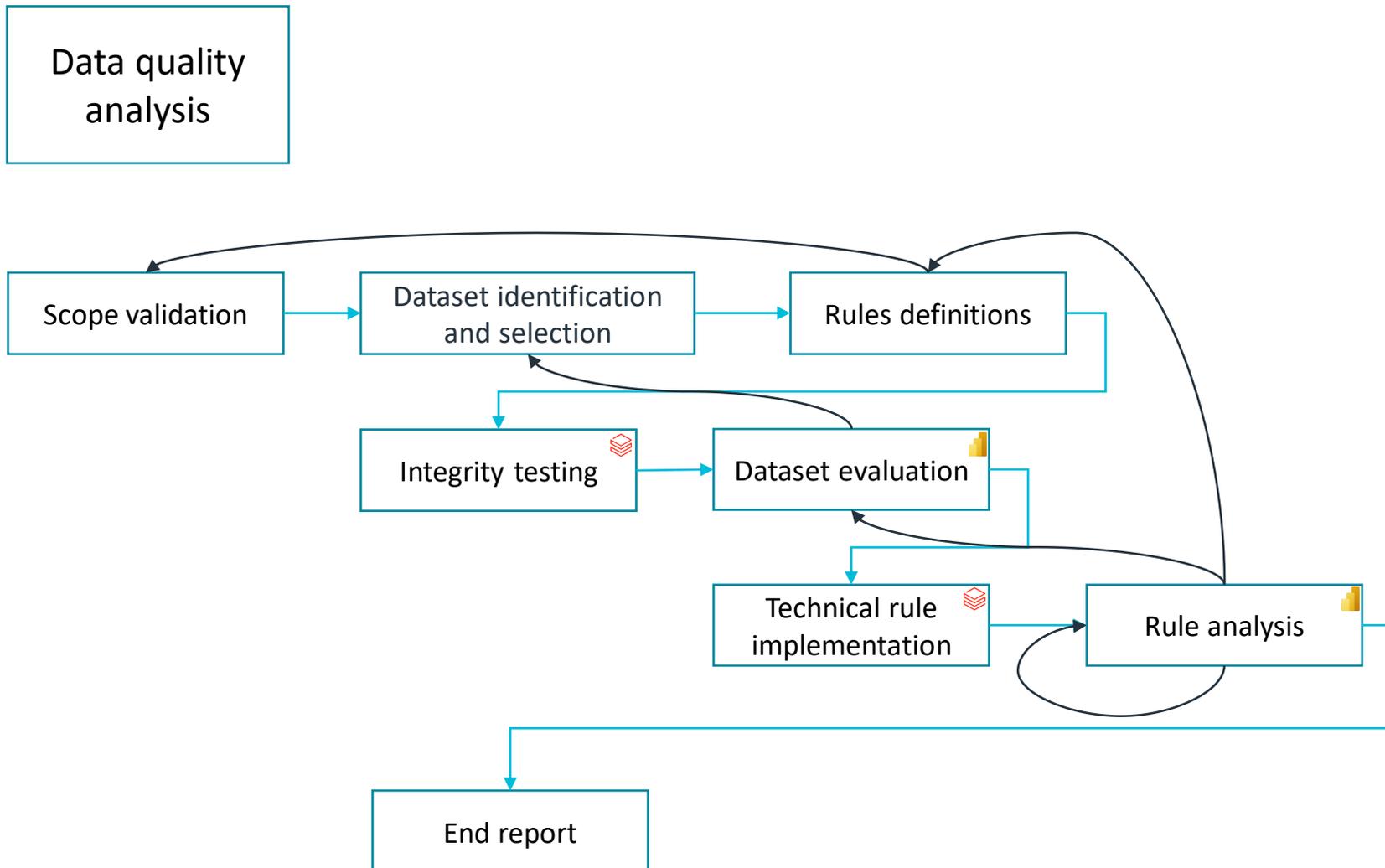




Data quality analysis

- Much too often reporting projects gets into trouble because the state of the data. Undocumented and undiscovered problems in the data sets keeps them from deploying promised business values. And this is only the best-case scenario. In other case the data leaves the data platform and ends up to be used. Users then questions the data's precision and reliability. Erroneous data contributes to flawed decisions.
- To tackle this problem a preliminary data quality analyses for new systems and business critical data assets should be performed. Before accepting a new ERP, logistic, ordering or similar module into production the end result, data, needs to be validated. This is rarely done and system changes and updates or even entire new systems goes into production without any standardize data quality checks.
- Data quality rules also doesn't come from a thin air. Investigation work needs to be done to verify what can and should be measured and monitored. A data analysis is a perfect way to establish the rules and gain needed information of new data assets for data stewardship.

Steps of data quality analysis



- Data analysis delves into the quality of the data sets, supported by perspective provided from use cases. It is performed in tight collaboration with business experts utilizing modern analysis tools.
- In order to make the data analysis to be possible, the exact scope must first be known:
 - Which data asset is to be analyzed in order to gather correct data sets
 - What are the main purposes and / or assumptions that exist for the data sets and use cases

Cloud1 Data Quality Analysis

Why

In the cloud, the amount of data has multiplied and the pace of development is accelerating. At the same time, the need for data quality has increased. Avoid the high cost of poor quality data to your organization by investigating data before it hits the development phase or in the worst case production.

What

The agile data quality deep diving project is implemented in close collaboration with the organization's own business data experts. We combine the needs and goals of both business and IT for data quality. During the project, we will familiarize you with the data quality of your data assets, and deepen your understanding and knowledge of data quality in your organization.

Tools and roles needed for the project

- Access to the data for the analysis (Preferred: SQL or Data Lake)
- Azure tools to be available (Databricks, Power BI, Storage Account, Key Vault)
- Roles (Project Owner, Owner(s) of selected data sets, business process expert)

Project duration

- 2-4 weeks

Methods

- Workshops, analysis and final report

What the organization gets

- Estimate of the level of data quality
- Proposals for action
- Tool suggestions
- A concrete roadmap to improve data quality

Price: 19 000€