IPTC 2024
IPTC-23258

Data Quality Transformation, Best Practices And Tools

R. Al-Jawad, Saudi Aramco Exploration; P. Benyeogor, S. Albarrak, M. Hakami, K. Hakami, Saudi Aramco

Abstract

Objectives/Scope: Data quality plays a crucial role in decision-making processes. With the increasing volume and complexity of data available today, the importance of ensuring that data is accurate, consistent, complete, and relevant has become even more critical. The objective of this paper is to demonstrate how Saudi Aramco operationalized data quality using industry best practices for maintaining high quality data, and tools we have deployed to achieve this goal.

Methods, Procedures, Process: In recognition of the importance of reliable data in making decisions, we undertook a thorough assessment of our data quality standing to evaluate the most effective solution for optimizing data quality. After evaluating several commercial options, none were deemed suitable for measuring all six dimensions of data quality, so we decided to create our own solution: Exploration Data Quality (EDQ). EDQ was designed to identify complex data quality issues, such as overlapping, incorrect order, duplication, and unrealistic time gaps between records. In addition, we implemented a comprehensive data quality framework that encompassed data governance policies, and provided our employees with training on data quality best practices.

Results, Observations, Conclusions: The inclusion of EDQ within our data quality architecture has enabled us to establish an uninterrupted and sustainable process for expeditiously validating data to a high standard, and providing it to fulfill operational requirements with confidence. This approach yielded a notable improvement in overall data quality, increasing from 82 percent to 95.6 percent, with a target of reaching 99.3 percent, equating to level 4 of the Six Sigma industry standard, through continued collaboration with the stakeholders. The data visualization module of the EDQ system offers real time insights into the status of each data type for the business and data management. This integration has enabled live monitoring of data quality and has reinforced adherence to global best practices.

Novel/Additive Information: This paper presents a fresh data quality management approach that effectively utilizes the three essential components of data management: people, processes, and technology. Additionally, it illustrates that data quality management is a collective responsibility, not just that of the data management professionals.