Chapter 8
Engineering Data Quality Automation, Validation, and Trust in Large-Scale Data Systems for Sustainable Industrial Processes

Devkiran Narayana
https://orcid.org/0009-0001-7058-1221
Independent Researcher, USA

ABSTRACT

In the era of sustainable industrial processes, the reliability and quality of data play a pivotal role in driving informed decision-making, process optimization, and regulatory compliance. Large-scale data systems, however, face challenges related to data heterogeneity, volume, velocity, and veracity. This chapter explores engineering practices for ensuring high-quality data through automation, validation, and trust mechanisms within data pipelines. Emphasis is placed on designing automated data quality checks, implementing real-time validation frameworks, and embedding trust through provenance tracking and governance-by-design principles. The chapter also highlights the role of emerging technologies such as AI-driven anomaly detection, blockchain-based auditability, and IoT-enabled monitoring for enhancing transparency in industrial workflows. A case study with quantitative results demonstrates how automated data quality engineering significantly improves operational efficiency, reduces compliance risks, and advances sustainability goals. By aligning data

DOI: 10.4018/979-8-3373-2205-6.ch008

quality strategies with industrial sustainability objectives, the chapter provides a holistic framework for building resilient, trustworthy, and scalable data systems that enable industries to achieve both economic and environmental performance targets.

INTRODUCTION

In the era of digital transformation, data has become one of the most critical assets driving decision-making, innovation, and operational efficiency across industries. From manufacturing and energy to healthcare and logistics, industrial processes are increasingly reliant on massive volumes of data generated through connected systems, sensors, IoT devices, and enterprise applications. While data availability has grown exponentially, ensuring its quality, reliability, and trustworthiness has emerged as one of the most pressing challenges. Low-quality data leads to erroneous insights, suboptimal decisions, compliance violations, and wasted resources. Hence, the focus on engineering data quality has become not just a technical priority but also a cornerstone for sustainable industrial processes.

Large-scale data systems, characterized by distributed architectures, heterogeneous sources, and high-volume real-time streams, pose unique challenges to quality management. Unlike traditional databases where data could be controlled and validated at ingestion, modern pipelines are dynamic and often involve multiple actors, technologies, and environments. This complexity increases the likelihood of errors such as missing values, duplicates, inconsistencies, schema drifts, and delayed updates. To address these challenges, organizations are turning to automation, validation frameworks, and governance strategies that can ensure data trustworthiness at scale.

Importance of Data Quality in Industrial Sustainability

Sustainable industrial processes depend on accurate, reliable, and timely data. For example, energy-intensive sectors such as steel production, oil refining, or cement manufacturing rely heavily on real-time monitoring of emissions, energy consumption, and resource utilization. Poor-quality data in such contexts can lead to incorrect reporting of carbon footprints, inefficient resource allocation, and failure to comply with environmental regulations. Similarly, predictive maintenance in manufacturing plants depends on clean sensor data to accurately forecast equipment failures. Inaccurate data could lead to unplanned downtime, excessive maintenance costs, or safety hazards.

The push for Industry 4.0 and smart factories has made data governance and quality engineering central to achieving long-term sustainability goals. High-quality data allows industries to design circular economy models, track carbon emissions,

Development of Hybrid Cellulose Bio Nanocomposite From Banana and Jute Fiber