

Agile Implementation of Enterprise Data Lake: A Governance-First Methodology for Sustainable Data Management

Pabitra Saikia

Truist Bank

pabitras@gmail.com

Abstract: Data in organizations and the necessity for data-driven decisions is increasing at an exponential rate. This has made the implementation of enterprise data lakes become mandatory for keeping structured and unstructured data in centralized stores. Implementation of data lake in linear and waterfall approaches commonly result in imperatives like long durations, scope increases, and non-comparability with business needs. This paper outlines a detailed framework for the implementation of enterprise data lakes in an iterative and agile manner with sound data governance foundations. The application of agile methodologies in building data lake architecture with iterative development, continuous stakeholder engagement, and adaptive planning can address the inherent complexities of large-scale data infrastructure projects.

Furthermore, we establish that effective data governance is not an afterthought but a foundational requirement that must be embedded from inception. Based on review of implementation patterns, architectural choices, and governance mechanisms, this study provides practical guidance for deploying enterprise data lakes that deliver rapid value without compromising data quality, security, or compliance. Findings indicate that agile methods adapted to data lake environments and anchored in governance-first principles can cut time-to-value while enabling sustainable, organization-wide data management (Inmon et al., 2019; Katal et al., 2021; Lwakatare et al., 2019).

Keywords: Data Lake, Agile Implementation, Data Governance, Enterprise Architecture, Data Management, DevOps, Data Quality

