IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.67
Volume 5, Issue 2, February 2025

Robust Data Pipelines for AI Workloads: Architectures, Challenges, and Future Directions

Giriraj Agarwal
Sr. Manager - Projects – Cognizant,
https://orcid.org/0009-0006-1042-6568

Abstract: *The exponential growth of data and the increasing complexity of AI workloads necessitate robust data pipelines that ensure scalability, reliability, and efficiency. This paper explores the design and implementation of robust data pipelines tailored for AI applications. We discuss advanced architectural strategies, including distributed ETL processes, real-time data streaming, and automated data quality management, that underpin the effective training and deployment of AI models. Through a comprehensive review of current literature, practical case studies, and experimental evaluations, we identify key challenges such as data heterogeneity, latency, and fault tolerance. We propose a unified framework that integrates modern data ingestion tools, scalable processing frameworks (e.g., Apache Spark), and monitoring systems to ensure high throughput and low latency for AI workloads. Finally, we outline future research directions, including enhanced data governance, integration of edge computing, and the incorporation of explainable AI techniques to further optimize pipeline performance and reliability.*

Keywords: Data ingestion, AI, Data Engineering, ETL

DOI: 10.48175/IJARSCT-23391