

International Conference on Artificial Intelligence and Cloud Computing (ICAICC-2025)

Conference Proceedings

May 10, 2025 (Virtual)

Integrating DevOps, Data Engineering, and ML Workflows in Databricks and AWS Environments

Satyadeepak Bollineni

DevOps CI/CD engineering, Advanced Data Engineering, and Building Machine Learning Workflows Tailored to Databricks Services and AWS Infrastructure, USA

This paper presents a comprehensive framework for the integration of DevOps continuous integration/continuous deployment (CI/CD) methodologies with advanced data engineering practices and machine learning (ML) workflows specifically optimized for Databricks services and AWS infrastructure. We examine how traditional development pipelines can be transformed through automated testing, deployment, and monitoring systems designed for data-intensive applications. The research explores specialized techniques for data preparation, transformation, and feature engineering that leverage the distributed computing capabilities of Databricks runtime environments. We detail the architecture and implementation of end-to-end ML pipelines that seamlessly connect data ingestion from diverse AWS sources through model training, deployment, and production monitoring within the Databricks ecosystem. Through real-world case studies, we demonstrate how this integrated approach significantly reduces development cycles, improves model performance, and enhances operational reliability. Our findings provide valuable insights for organizations seeking to build robust, scalable, and efficient data platforms that combine the strengths of modern DevOps practices, enterprise-grade data engineering, and production-ready machine learning systems.