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

Conference Proceedings   May 08, 2025 - Florida, USA

Advanced Data Engineering, Analytics, and AI: Strategies for Innovation and Efficiency

Preyaa Atri

Senior Data Engineer at LG Electronics, USA

In today's data-centric landscape, the ability to effectively manage and utilize vast amounts of data is paramount for fostering innovation and enhancing operational efficiency. This expert note, presented by Preyaa Atri, a seasoned Senior Data Engineer with over a decade of experience, delves into the sophisticated strategies and methodologies in data engineering, analytics, and AI that drive success across diverse industries. Drawing on her extensive professional background and numerous published research works, Preyaa will provide a comprehensive and detailed exploration of the following key areas:

Optimizing Data Processing Workflows:
- **Cost Reduction through Integration:** Detailed techniques for integrating Apache Airflow and Dataform to reduce data processing costs by up to 60%. Practical examples of how these integrations streamline workflows, automate repetitive tasks, and enhance data pipeline efficiency.
- **Cloud-Based ETL Pipelines:** In-depth case studies on the implementation of ETL pipelines using cloud platforms like Google Cloud and AWS. Discussion on best practices for designing scalable and resilient data transformation processes that ensure high performance and reliability.

Leveraging Machine Learning for Enhanced Decision-Making
- **Predictive Analytics Models:** Insights into building machine learning models that predict hardware malfunctions with over 80% accuracy. Examination of the algorithms used, data preprocessing techniques, and model evaluation metrics that contribute to high predictive performance.
- **Optimizing Recommendation Engines:** Strategies for enhancing the efficiency of recommendation engines by 33%. Detailed overview of collaborative filtering, content-based filtering, and hybrid recommendation systems, along with real-world applications and results.

Innovative Use of Big Data Ecosystems
- **Big Data Tools and Technologies:** Comprehensive analysis of using Hadoop, Spark, Kafka, and other big data tools to manage and analyze large datasets. Discussion on data ingestion, processing, and real-time analytics, highlighting the benefits and challenges of each tool.
- **Data Interoperability and Integration:** Strategies for ensuring data interoperability across diverse platforms. Detailed methodologies for automating schema expansion from Parquet to Big Query, facilitating seamless data integration and transformation.

Cost-Effective Data Warehouse Management
- **Data Warehouse Optimization:** Techniques for optimizing data warehouses on platforms like Google Cloud, achieving up to 30% cost reduction. Practical examples of optimizing storage, indexing, and query performance to support efficient data retrieval and analysis.
- **Business Intelligence and Analytics:** Practical examples of managing data storage, processing, and retrieval to support business intelligence and analytics. Discussion on the use of tools like Big Query, Composer, and Dataflow to enhance data visibility and decision-making.

Advanced Cloud Services and Security Measures
- **Leveraging Cloud Services:** Best practices for utilizing cloud services such as BigQuery, Dataflow, and Vertex AI to drive innovation. Detailed exploration of cloud-based data processing, machine learning model deployment, and real-time analytics.
- **Data Security and Protection:** Implementing comprehensive data protection measures to ensure the security of data at rest and in transit. Strategies for encryption, access control, and compliance with data privacy regulations.

Future Trends and Innovations
- **Emerging Trends in Data Engineering and AI:** Exploration of emerging trends such as generative AI, advanced analytics, and automated machine learning. Discussion on the impact of these technologies on data engineering practices and their potential to drive future innovations.
- **Scalable and Flexible Data Strategies:** Preparing for the future by adopting scalable and flexible data strategies. Insights into building adaptable data architectures that can evolve with technological advancements and changing business needs.