Abhiram Kankipati

Jersey City, NJ, 07302 | abhiram.kankipati@gmail.com | +1-757-300-6216 | linkedin | github | portfolio

Summary

Senior Data & MLOps Engineer with a proven track record of building and operationalizing robust data platforms on AWS and Azure. With 8 years of expertise, I specialize in architecting high-performance databases (SQL & NoSQL) and engineering scalable data pipelines with Spark. I bridge the gap between data and deployment by creating automated MLOps workflows that have reduced model release cycles from months to days.

Skills

Big Data & Streaming: Apache Spark, Hadoop, HDFS, Hive, Kafka, AWS EMR, Glue, DBT
Database Systems: Snowflake, Redshift, Azure SQL, Oracle, Teradata, SQL Server, MongoDB, HBase
MLOps & Automation: Kubernetes, Docker, GitLab CI/CD, Airflow, AWS Sagemaker, Azure AI Foundry
Programming & APIs: Python (FastAPI, Flask, Pandas), Scala, PL/SQL, RESTful APIs, API Gateway
Data Visualization: Tableau, Power BI, ThoughtSpot, Looker Studio

Experience

Data Engineer, Bank of New York Mellon – New York City, NY October 2023 - Present

- Architected and deployed a scalable data warehousing solution on Snowflake, processing over 5TB of daily financial data and improving query performance for analytics teams by 40%.
- Developed and operationalized a CI/CD pipeline for machine learning models using GitLab, Docker, and Kubernetes, reducing the model deployment lifecycle from 2 months to under a week.
- Engineered complex, performance-tuned PL/SQL packages and automated ETL workflows, enhancing data transformation efficiency and ensuring 99.9% data accuracy for regulatory reporting.
- Optimized Spark jobs for large-scale data processing on Azure, cutting down execution time by 35% and significantly reducing cloud compute costs.
- Collaborated with the Internal Audit division to design and prototype a self-service data analytics platform using Power BI and Python, empowering auditors to independently investigate anomalies and reducing data request turnaround times by 60%.
- Conceptualized and developed a GenAI-powered agent for audit teams using a Retrieval-Augmented Generation (RAG) architecture to instantly query internal policy documentation, significantly accelerating the evidence-gathering phase of the audit lifecycle.
- Architected and implemented a scalable data modeling framework in Snowflake using dbt, building efficient incremental models that materialized as Apache Iceberg tables. This strategy reduced data processing latency by over 70% and enabled reliable, high-performance time-travel queries for critical financial auditing.
- Orchestrate cross-functional data initiatives, implementing Azure SSO authentication and maintaining comprehensive documentation for stakeholders.

Data Engineer, Citi Group – Dallas, TX August 2022 - October 2023

- Designed and implemented a cross-platform data validation system using Spark and Python, which verified the integrity of over 1 million customer statements daily against a Teradata production warehouse.
- Automated complex ETL processes for a fraud detection system, ingesting data from Kafka streams into MongoDB and reducing manual data handling by 90%.
- Managed and optimized multi-terabyte Oracle and Azure SQL databases, implementing robust indexing and partitioning strategies that improved API response times by 25%.
- Engineered and automated a robust data pipeline using Apache Spark and Python to process and aggregate risk data from disparate sources, including Oracle and MongoDB, ensuring the high-quality, reliable data needed for regulatory reporting.
- Simplified database operations by implementing automated validation processes, reducing manual intervention, and enhancing data integrity across platforms.

Student Developer, ODU IT Services – Norfolk, VA January 2021 - July 2022

- Built a foundational Big Data ETL pipeline on an AWS EMR cluster using Spark and Hadoop to analyze student performance data, uncovering key metrics that informed university retention strategies.
- Developed a proof-of-concept data warehouse with Redshift and RDS, integrating multiple university data sources to enable faster, more efficient analytical reporting for faculty."
- Streamlined AWS resource deployment by authoring CloudFormation templates, reducing infrastructure setup time by over 50% and establishing a reproducible environment for data projects.

Python Developer, TCS – Bengaluru, India May 2017 - December 2020

- Engineered the company's first automated, serverless ETL pipeline on AWS using Glue, S3, Lambda, and SQS, processing client supply chain data from diverse geographies with high availability.
- Contributed to the design of a mission-critical SAP HANA database using star schema methodology, optimizing data models for complex BI queries that served a global user base.
- Developed and maintained a key Python-based web application using Flask for a major client's supply chain management system, directly improving data integration and reporting efficiency.
- Automated the extraction and transformation of complex datasets from diverse client web sources and internal databases using Python and SQL, creating curated data feeds that directly powered critical business intelligence reports and dashboards.

Education

Master of Science in Computer Science, Old Dominion University – Norfolk, VA January 2021 - May 2022

- GPA: 3.87

Bachelor of Technology in Mechanical Engineering, Gayatri Vidyaparishad College of Engineering – Vizag, India August 2013 - April 2017

Certifications

Deep Learning Foundations: NLP with TensorFlow, LinkedIn Learning
Algorithms and Data Structures, Udemy
SQL Advanced Certificate, HackerRank

Projects

PC Video game Recommendation Engine and Analysis

- Engineered a content-based recommendation system by developing an automated web scraping pipeline with Scrapy and Selenium to acquire gaming data from multiple online sources.
- Implemented NLP techniques using NLTK and built a cosine similarity matrix to accurately match users with games, demonstrating early proficiency in applied machine learning.

Employee Resource and Login Patterns

- Applied data mining and unsupervised learning (K-Means clustering) to analyze complex login and resource usage patterns from enterprise-level datasets.
- Developed a predictive model using XGBoost to forecast peak resource demand, providing insights that could increase operational efficiency and reduce costs.

Pretrained Encyclopedia: Weakly Supervised Knowledge-Pretrained Language Model

- Conducted research on large-scale language models, implementing and fine-tuning BERT-based architectures for complex question-answering tasks on established academic datasets.
- Designed and built a custom Bidirectional Transformer model, demonstrating a fundamental understanding of attention mechanisms and modern NLP model architecture.

Accessible Digital Menu for Blind People

- Developed a browser extension that leveraged a Convolutional Neural Network (CNN) to perform information extraction on unstructured menu images, enhancing digital accessibility. Optimized model inference time by leveraging GPU acceleration with CUDA, reducing processing latency by over 70% to ensure a real-time user experience.