

Professional Summary

Strategic data engineer with 4+ years of experience in optimizing data process and driving business value through actionable insights. Proven ability to reduce operational costs, improve decision-making speed, and minimizing risks, while fostering collaboration within diverse teams to support business growth.

Technical Skills

Languages & Tools: Python, R, PySpark, MySQL, MongoDB, Cosmos DB, Data Modeling and Warehousing
Platforms: AWS, Azure Analysis Services, Data Factory and Data Lake
Tools: GitHub, MATLAB, Adobe Analytics, Docker, Kubernetes
Analytics: Power BI, Tableau, Excel, Data Mining, SSIS, SSMS, CI/CD
Data distribution: Kafka, Spark, Redshift, BigQuery, Snowflake

Education

Texas A&M University, Master of Science, Computer Science

CGPA: 3.83/4.00

Work Experience

Data Engineer, NextRow Digital

July 2021 – Dec 2022

- Optimized SQL queries and automated data retrieval from disparate sources, reducing query execution time by 40% using indexing strategies, materialized views, and efficient partitioning in Azure Data Lake, Synapse, and Runbook
- Advised internal customers on dashboarding personalization, experimentation and best practices and clearly communicated insights discovered from data in a snackable manner for easy interpretation by business users
- Crafted more than 30 interactive reports using Python and Power BI to uncover data patterns and insights that led to a 15% increase in user engagement
- Mentored new team members on best practices, product functionality, and the use of BI tools, emphasizing engineering principles. This ensured execution and integration of data pipelines, data modeling, and ETL processes
- Reduced cloud costs and improved report performance by optimizing 20 Azure Analysis Services data models with memory-efficient configurations, saving 500MB per model
- Utilized Azure DevOps services such as Azure Repos, Boards and Git to collaborate on code development, built and deployed applications
- Led the data engineering team, working with stakeholders to identify customer needs, and collaborated with data scientists to deploy ML models into production using MLflow and Feature Store

Software Developer, Meslova Systems

Sept 2018 - June 2021

- Led an onsite team of 5 members and acted as a project SME to oversee production deployment, resolve high priority go live issues and trained client teams on new features deployed
- Created data governance and security models, mapping 30+ documents and 20 data dictionaries. Streamlined pipelines and metrics in Confluence, boosting collaboration and reducing onboarding by 30%
- Optimized complex SQL queries, improving data retrieval speed by 5x (from 5s to 1s) by minimizing joins, indexing, and using subqueries enhancing overall application performance
- Redesigned database schemas into BCNF to streamline customer order management, reducing redundancy and ensuring data integrity, which improved query performance by 30% and enabled accurate real-time inventory tracking

Projects

- Created enterprise BI solutions, data visualization tools, and key metrics for manufacturing departments. Established a unified data warehouse and documented metric definitions for consistent usage
- Engineered and automated 10+ data pipelines in Azure Data Factory and Databricks, saving 4+ hours per week, while optimizing ETL workflows using Apache Airflow and Spark for efficient data processing
- Designed real-time data ingestion pipelines using Kafka and Azure Event Hubs, reducing data processing time by 40%, enabling faster machine learning model inference and analytics
- Designed and implemented scalable data lake and warehouse solutions using AWS S3, Glue, Redshift, and Snowflake, improving data retrieval and ETL time by 50% through efficient partitioning and schema optimization
- Enabled real-time fraud detection by developing a Kafka-based event processing pipeline with highly scalable and low-latency Cosmos DB, identifying transaction anomalies in milliseconds to minimize financial risk and prevent fraud