

TIRUPATHI RAO LUKALAPU

Location: IL | Phone: +1 (312) 221-7726 | [Email](#) | [LinkedIn](#)

Professional Summary

Data Engineer with 4+ years of expertise in building and optimizing large-scale, cloud-native data pipelines in healthcare, finance and telecommunications. Proven ability to leverage Databricks, PySpark, SQL, and AWS/Azure to process multi-terabyte datasets, reduce costs by over 25%, and improve data processing performance by up to 40%. Holds Databricks Data Engineer Associate and Microsoft Fabric Data Engineer Associate certifications. Focused on delivering reliable data infrastructure that enables real-time analytics and data-driven decision-making.

Technical Skills

Programming Skills: Python (NumPy, Pandas, Matplotlib, seaborn, Scikit-learn, Tensorflow), SQL, Pyspark, R, Bash

Big Data/Frameworks: Apache Spark, Kafka, Hive, Hadoop, HDFS, Airflow

Databases/Storage: MongoDB, AWS S3, PostgreSQL, Delta Lake, Elasticsearch, SQL Server

Orchestration Tools: Apache Airflow, Azure Data Factory

Cloud Platforms: Azure (ADF, Databricks, ADLS Gen 2), AWS (S3, EC2, EMR), Snowflake, CloudWatch

Monitoring & Visualization Tools: Power BI (DAX, Data Modeling, UX Best Practices), Tableau, Grafana, Kibana,

Concepts & Practices: Data Modeling (Star/Snowflake), ETL/ELT, Data Lakehouse, Data Governance (HIPAA, GDPR), Test-Driven Development (TDD), Agile/Scrum

Experience

Kinaxis

Sep 2024 – Present

Data Engineer

Chicago, United States

- Designed and implemented scalable, governed data pipelines on Azure (ADF, Databricks, ADLS Gen2) to ingest and process network telemetry and customer experience data, forming the foundation for enterprise analytics.
- Architected Delta Lake tables within a medallion lakehouse framework to ensure data quality, support ACID transactions, and enable incremental data processing for downstream consumption.
- Orchestrated complex data workflows using Apache Airflow, improving pipeline reliability to 99.9% and reducing manual intervention by automating monitoring and alerting for 20+ critical data jobs.
- Developed PySpark transformation logic and automated data quality checks to maintain high fidelity of critical business data, supporting analytics and machine learning initiatives.
- Collaborated with data governance and platform teams to implement access controls and data lineage tracking, enhancing security and discoverability across the data estate.

CitiusTech

Jun 2022 – July 2023

Data Engineer

Hyderabad, India

- Built and maintained high-volume ETL pipelines using Databricks and Apache Spark, processing 25 million patient records daily (JSON, Parquet, CSV) to support a unified healthcare analytics platform.
- Optimized Spark jobs through intelligent partitioning, predicate pushdown, and memory tuning, achieving a 35% reduction in data load time and improving downstream Power BI dashboard performance.
- Engineered a reusable data quality framework using Great Expectations and PySpark, automating the detection of anomalies across 50+ data sources and reducing time spent on data validation by 15 hours per week.
- Migrated legacy on-premise SQL Server workloads to a cloud-native Databricks and Delta Lake architecture, cutting infrastructure costs by 25% and improving data freshness for end-users from daily to near-real-time.
- Implemented automated data validation frameworks with PySpark and SQL, ensuring 99.8% data accuracy for patient outcomes and hospital performance reporting.
- Partnered with clinical and business analysts in agile sprints to define KPIs and deliver interactive dashboards, directly influencing strategies to improve patient care and operational efficiency.

Vision Software Solution

Dec 2020 – May 2022

Junior Data Engineer

Hyderabad, India

- Automated the financial data lifecycle by developing Python (Pandas) and SQL scripts to ingest and clean data from 4+ disparate source systems, reducing manual intervention by 80%.
- Engineered standardized data models and KPI definitions, slashing the monthly financial reporting cycle from 3 days to 5

hours, 60% efficiency gain and improving reporting accuracy for leadership.

- Built the company's first centralized data repository by designing a star-schema data warehouse in PostgreSQL, consolidating fragmented reports and establishing a 'single source of truth' for financial data.
- Created Power BI dashboards to visualize revenue trends and departmental budget adherence, providing consistent, self-service insights that replaced fragmented manual reports.

Projects

Real-Time Data Streaming Pipeline / *Apache Spark, Amazon S3, Snowflake, Snowpipe*

Jan 2021

- Developed a cloud-based real-time pipeline integrating Spotify APIs with AWS Lambda, Glue, and Snowflake. The project emphasized seamless data integration, schema design, and scalable transformation logic, aligning closely with enterprise-grade healthcare data ingestion patterns.
- Automated Snowflake ingestion using Snowpipe, reducing manual intervention and data availability lag from hours to minutes.
- Enabled near real-time access to over 30K+ curated records, supporting downstream analytics and improving data readiness by 90%.

E-commerce Data Pipeline on Azure / *SQL, Azure Data Factory, Azure Databricks, Apache Spark*

Dec 2024

- Designed and implemented an end-to-end data pipeline to ingest approximately 100 GB of daily e-commerce sales data, perform critical data transformations and enrichments, and load the processed data into a data lake for efficient analytical consumption.
- Orchestrated the e-commerce pipeline using Apache Airflow to automate 15+ workflows and enforce DAG-based execution.
- Leveraged Azure Data Lake Storage Gen2 for secure and scalable data storage, orchestrated the pipeline using Azure Data Factory with 15+ data pipelines, and executed data transformations and aggregations within Azure Databricks using Apache Spark and Delta Lake for optimized performance.
- Enabled efficient processing of approximately 100 GB of daily e-commerce data, resulting in a 30% improvement in data processing time and providing timely insights for business intelligence reporting and analysis. Demonstrated strong Azure data engineering expertise in building and managing scalable data pipelines.

Certifications

- Databricks Certified Data Engineer Associate.
- Microsoft Fabric Certified Data Engineer Associate.
- Oracle cloud database services 2025 certified professional.

Education

Lindsey Wilson College

Aug 2023 – May 2025

Masters in Computer/Information Technology Administration and Management

Columbia, Kentucky