Tirupathi Rao Lukalapu

312-774-9864 Tirupathiraolukalapu08@gmail.com linkedin.com github.com

Professional Summary

Data Engineer with 4+ years of expertise in designing scalable, governed data pipelines for healthcare and fintech. Proficient in cloud-native architecture (AWS, Snowflake, Databricks) and workflow automation (Python, SQL, Airflow) for real-time and batch processing. Leads cross-functional data initiatives from raw ingestion to actionable insights while enforcing DevOps rigor and regulatory compliance (HIPAA/GDPR). Strategic partner in optimizing infrastructure for reliability and cost efficiency. Passionate about building systems where technical excellence accelerates business outcomes.

Education

Lindsey Wilson College **Aug 2023 – May 2025**

Masters in Computer/Information Technology Administration and Management *Columbia, Kentucky*

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, Azure Data Lake Storage Gen2, 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, Apache Superset
Other: Data Cleaning, Data Modeling, Git, Jira, Confluence, KPI Reporting, Github, Excel, Agile Environment, CI/CD, DataOps, Test-Driven Development

Professional Experience

Innovaccer **December 2023 – April 2025**

Data Engineer *California, United states*

- Built HIPAA-compliant data pipelines using AWS Glue, Python, and Spark, processing 10TB+ of EHR data monthly, improving data availability by 35%.
- Automated data validation with Python reducing errors by 25% and accelerating data source onboarding by 40%.
- Migrated on-premise data to AWS S3/Redshift, cutting storage costs by 25% while boosting query speed by 40%.
- Implemented Medallion Architecture in Delta Lake, enabling self-service analytics for 50+ researchers and reducing ad-hoc requests by 20%.
- Pioneered FHIR-compliant data integrations for EHR systems (Epic, Cerner), standardizing clinical data from 50+ hospitals into Innovaccer's unified data model, accelerating analytics readiness by 35%
- Optimized cloud infrastructure costs by 20% through auto-scaling AWS Glue jobs and partitioning Delta Lake tables, aligning with Innovaccer's focus on "scalable, sustainable solutions"
- Co-led agile workshops with clinical stakeholders to define data requirements for population health analytics, directly supporting Innovaccer's mission to "activate data flow for better health outcomes"

Capgemini **October 2021 – June 2023**

Data Engineer *Hyderabad, India*

- Engineered real-time AML monitoring pipelines using Spark Streaming and Kafka, reducing fraud detection latency to less than 2 seconds and aligning with Capgemini's emphasis on "risk exposure mitigation"
- Optimized Snowflake data warehouse through partitioning/materialized views, slashing report generation from 2 hours to 12 minutes.
- Modernized legacy data warehouses to Snowflake using Azure Data Factory, achieving 99.9% pipeline reliability and supporting Capgemini's "cloud-first" client transformation initiatives
- Built CI/CD pipelines with Azure DevOps, reducing deployment failures by 35% and accelerating releases to hourly cycles.
- Implemented dbt models to standardize transformations, boosting team productivity by 15% while ensuring GDPR compliance.
- Reduced Snowflake costs by $18K/year via auto-scaling and query tuning.

Teamtech Solutions **June 2020 – August 2021**

Data Analyst *Hyderabad, India*

- Built startup's first analytics infrastructure from zero using Python/SQL, unifying 7+ data sources (Salesforce, Stripe, Mixpanel) into a central Redshift warehouse, enabling data-driven decisions across product/sales teams.
- Created 10+ executive dashboards in Power BI tracking MRR, CAC, and feature adoption, directly influencing a pivot that reduced churn by 15% and retained $120K in annual revenue.
- Automated manual revenue reporting with Python scripts, reducing CFO's financial close time from 3 days → 4 hours and freeing capacity for Series A fundraising.
- Identified upsell opportunities through cohort analysis of 2,500+ free-tier users, driving targeted campaigns that converted 18% to paid plans ($45K ARR increase).
- Trained non-technical teams on self-service analytics (Power BI/Excel), increasing data adoption by 50% and reducing ad-hoc requests by 30%.

Projects

Real-Time Data Streaming Pipeline | *Apache Spark, Amazon S3, Snowflake, Snowpipe* **March 2025**

- 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.
- Stored raw JSON data in Amazon S3 and transformed 100% of records using AWS Glue (PySpark), optimizing schema consistency and processing time by 30%.
- 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 | *Azure Data Factory, Azure Databricks, Apache Spark, SQL* **April 2025**

- 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.