Tirupathi Rao Lukalapu

312-774-9864 Tirupathiraolukalapu08@gmail.com

Professional Summary

Healthcare Data Engineer with 3+ years of experience designing HIPAA-compliant data pipelines, clinical data integrations, and healthcare analytics platforms on Azure, AWS, Databricks, and Snowflake. Skilled in EHR (Epic, Cerner), HL7, FHIR APIs, and real-time ETL frameworks using Spark, Kafka, PySpark, and Airflow. Proven track record of processing 10TB+ of clinical and claims data monthly, improving data quality by 99.8%, and enabling predictive analytics to enhance patient outcomes.

Education

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

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

Technical Skills

Programming & Databases:Python (Pandas, NumPy, Scikit-learn), SQL, R, PostgreSQL, SQL Server, MongoDB
Cloud & Data Warehousing: Azure (Databricks, ADLS Gen2), AWS (S3, EMR, EC2), Snowflake, Redshift
Data Engineering & Big Data:Apache Spark, Kafka, PySpark, Hadoop, Hive, Delta Lake, AWS Glue
Orchestration Tools:Apache Airflow, Azure Data Factory
Healthcare & Data Governance: EHR Integration (Epic, Cerner), HL7, FHIR APIs, HIPAA, PHI Security, Payer-Provider Data Pipelines
DevOps & CI/CD: Azure DevOps, Terraform, Unity Catalog, dbt, DataOps
Visualization & Analytics: Power BI (DAX, Data Modeling), Tableau, Grafana, Kibana .

Professional Experience

Chicago Education Advocacy Cooperative **July 2025 – Present**

Data Engineer *Chicago, United States*

- Design HIPAA-compliant data pipelines on AWS (S3/Glue) for 50+ schools, ensuring zero PII breaches via encryption, RBAC, and audit trails.
- Develop scalable ETL workflows ingesting student records, curriculum data, and policy docs (Python/PySpark), optimizing costs 30% through partitioning/lifecycle policies.
- Build predictive model pipelines (Scikit-learn) identifying learning gaps, driving interventions that reduce dropout risk by 18%.
- Engineer real-time dashboards (Power BI/Shiny) for educators, slashing report generation from 8hrs to 15mins.).
- Automate data validation (Great Expectations) ensuring 99.8% accuracy in critical advocacy reports.
- Migrate legacy district reports to cloud-based dashboards (Power BI), serving 200+ educators with real-time literacy/math metrics.
- Develop data governance frameworks unifying records from 50+ schools into standardized schemas, reducing data reconciliation time by 65%.
- Prevent $15K+ potential breach costs through PII safeguards and cybersecurity training for staff.

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 under 2 seconds for financial clients
- Built scalable ETL frameworks using Azure Data Factory and Databricks, processing 10TB+ of healthcare data monthly with 99.9% pipeline reliability.
- Migrated legacy data warehouses to Snowflake, optimizing through partitioning and materialized views to slash report generation from 2 hours to 12 minutes.
- Developed Python-based data validation packages using Great Expectations, reducing data errors by 25% and improving data quality for critical business decisions.
- Implemented CI/CD pipelines with Azure DevOps, reducing deployment failures by 35% and enabling hourly release cycles for data applications.
- Designed real-time dashboards in Power BI tracking key financial metrics, enabling faster business decisions for executive leadership.
- Optimized complex SQL queries on large datasets, reducing query latency by 40% and improving analyst productivity across teams.

Capgemini **June 2020 – October 2021**

Data Analyst *Hyderabad, India*

- Created 10+ executive dashboards in Power BI tracking MRR, CAC, and feature adoption, directly influencing product pivots that reduced churn by 15%.
- Developed Python scripts for automated revenue reporting, reducing CFO's financial close time from 3 days to 4 hours and supporting Series A fundraising.
- Performed cohort analysis on 2,500+ users, identifying upsell opportunities that drove 18% conversion to paid plans ($45K ARR increase.)
- Conducted statistical analysis using Python (Pandas, NumPy, Scikit-learn) to identify trends and patterns in large financial datasets.
- Automated manual reporting processes using Python and SQL, freeing up 30% of team capacity for strategic analysis work .
- 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.