TIRUPATHI RAO LUKALAPU

Chicago, Illinois 60616

📞 312-774-9864 ✉ Tirupathiraolukalapu08@gmail.com

Professional Summary

Results-driven Data Engineer with 3+ years of experience designing, building, and optimizing scalable data pipelines across healthcare and fintech domains. Proven expertise in ETL/ELT development, data modeling, and orchestration using tools like PySpark, Snowflake, dbt, Apache Airflow, and Azure Data Factory. Skilled in cloud platforms including AWS and Azure, with a strong foundation in SQL, Python, and CI/CD practices. Adept at working in Agile teams to deliver high-quality data solutions, enforce data governance, and enable analytics through curated data products. Committed to data quality, performance optimization, and enabling business insights through automation and collaboration with cross-functional teams.

Education

Lindsey Wilson College   **Aug 2023 – May 2025**
Masters in Computer/Information Technology Administration and Management   *Columbia, Kentucky*

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, Snowflake, Azure Data Lake Storage Gen2, AWS S3, PostgreSQL
Orchestration Tools: Apache Airflow, Azure Data Factory
Cloud Platforms: Azure (ADF, Databricks, ADLS Gen 2), AWS (S3, EC2, EMR), Snowflake
Visualization Tools: Power BI, Tableau
Other: Data Cleaning, Data Modeling, Patient Data Analytics, KPI Reporting, Github, Excel, Agile Environment, CI/CD

Experience

Innovaccer Inc.   **September 2023 – March 2025**
Data Engineer   *California, United States*

- Built and maintained scalable ETL pipelines using PySpark and AWS Glue to ingest and transform large-scale healthcare data across multiple vendor sources.
- Designed reusable data pipeline components and implemented workflow orchestration with Apache Airflow, improving operational efficiency and data freshness.
- Developed ELT workflows with dbt and optimized Snowflake schemas, resulting in 40% faster query execution and improved reporting accuracy.
- Applied data quality validation using Great Expectations and implemented automated monitoring and alerting for data anomalies.
- Worked in Agile teams to support product features and data requirements across analytics and technology teams, following CI/CD practices using Git and Jenkins.
- Collaborated with stakeholders to ensure HIPAA-compliant data governance, version control, and secure cloud architecture on AWS and Databricks.

Teamtech Solutions   **March 2021 – June 2023**
Data Engineer   *Hyderabad, Telangana*

- Designed and developed data pipelines using Azure Data Factory to extract, transform, and load structured and unstructured financial datasets into Snowflake and SQL Server.
- Integrated internal databases and third-party APIs to support loan analytics and fraud detection workflows, reducing eligibility processing time by 50%.
- Created reusable data models, stored procedures, UDFs, and views; improved data lineage tracking and metadata documentation across data assets.
- Implemented CI/CD pipelines and version control with Azure DevOps and Git; actively supported data testing and production deployment processes.
- Conducted data profiling and validation using Python and SQL to ensure data quality, consistency, and accuracy in compliance with partner bank standards.
- Supported Power BI dashboard development by enabling curated datasets and ensuring data availability through monitored pipelines.

Projects

Real-Time Data Streaming Pipeline | *Apache Spark, Amazon S3, Snowflake, Snowpipe*   **August 2024**
- 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*   **February 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.
- 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.
- 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.