SRINATH BULUSU

Data Engineer
Email: srinathbulusu28@gmail.com
Phone: +1-8178198120
Github: https://github.com/srinath2805
LinkedIN: https://www.linkedin.com/in/srinath-bulusu-665b341a4/
Portfolio: srinathbulusu.com

PROFESSIONAL SUMMARY:

- **Data Engineer** with over 6 years of experience in designing, developing, and maintaining scalable **Data Warehouses** and Modern Data Solutions Architecture. Proven expertise in high and low-level design, production support in 24x7 environments, and root cause analysis. Skilled in building robust ETL/ELT pipelines using **Databricks**,**Snowflake**,**Snowpark** and **Azure/AWS** cloud environments to support business .
- Skilled in CI/CD automation (**Jenkins, GitHub**) and working within Agile/Scrum teams to drive engineering excellence. Adept at bridging the gap between complex software engineering principles and business intelligence needs, supporting self serve analytics and data driven decision making.
- Hands on experience designing, documenting and orchestrating **robust data pipelines** by collaborating with cross functional teams of 100+ users and supporting teams with data onboarding and quality checks using cloud platforms supporting batch and continous data workflows.

EXPERIENCE:

Upen Group | Data Engineer | August 2025 – Present

1.Enterprise Data Architecture & Cloud Migration (Client: Bank of America)

- Built a robust ETL pipeline using **Databricks** and **PySpark** ,**Python** to ingest and process high volume banking transaction data.Loaded data into 3 layers
- Engineered complex features, such as 48 hour rolling transaction totals and velocity checks, to identify suspicious patterns indicative of money laundering.
- Orchestrated end-to-end data workflows using Airflow, ensuring timely data availability for downstream fraud detection models.

Capgemini Technology Services | Senior Software Engineer | May 2021 – Jan 2024

1.Enterprise Data Architecture & Cloud Migration (Client: Prudential Financial)

- Architected a scalable **data ingestion framework** for **100+ cross-functional teams**, enabling structured onboarding of financial and healthcare data into **Snowflake**, ensuring 100% data quality compliance.
- Engineered high-performance **ETL/ELT pipelines** using **Snowpark (Python)** and **SQL** to transform **15TB+** of raw data across six architectural layers, optimizing storage and retrieval speeds.
- Designed serverless automation workflows using **AWS Lambda** and **S3 triggers** to ingest multi-format files (Parquet, JSON, XML), implementing **auto-recovery mechanisms** that reduced pipeline failure rates by **40%.**
- Implemented **Change Data Capture (CDC)** strategies to facilitate real-time analytics and maintain data integrity across distributed systems.
- Established robust security governance by designing **Role-Based Access Control (RBAC)** policies and dynamic data masking, ensuring compliance with strict healthcare data privacy regulations.
- Standardized deployment cycles by integrating **GitHub**,**Jenkins** for version control**,Flyway** for database version control within **CI/CD pipelines**, ensuring consistent environments across Dev, QA and Production.
- Implemented Snowflake **micro partitioning** and **clustering** to **optimize query performance**, reduce scan costs, and improve large-scale analytics efficiency.Enabled secure Snowflake **data sharing** to other accounts supporting data access while maintaining governance and access controls.

2. Data Warehouse Optimization (Client: Medica HealthCare)

- Optimized data warehousing capabilities by designing complex **Snowflake ETL workflows**, **utilizing Streams and Tasks** to automate SCD Type 1 & 2 historical data tracking.
- Created complex snowflake views from raw source tables in snowflake using multi table joins, CTEs and mappings to standardize data.Worked on **reducing query latency optimizing SQL queries by restructuring joins, reducing data scans, warehouse scaling** ,**caching**, **performance tuning**, applying efficient filtering to improve query performance and execution time
- Automated job execution using Atomic scheduling tool,aligning technical workflows with workforce system operations.
- Worked on flattening **JSON** data within raw tables to extract columns.Written transformation logic for multiple use cases and created member,med claim and pharmacy views from raw data.
- Created **resource monitors** on warehouses for tracking credit consumption and saving costs for the organization.

Upen Group | Data Analyst | Aug 2019 – Apr 2021

Performance Analytics & Reporting

- Improved reporting latency by **87%** (reducing delay from 2 months to 1 week) by migrating manual reporting procees to automated cloud-based logic, enabling faster decision-making for leadership.

- Developed and validated STMs with complex **SQL transformation logic** to build scorecards, ranking providers on quality compliance and identifying gaps in care.
- Designed emergency data tables to centralize critical patient information, significantly improving response times for urgent care reporting for faster access to urgent care data, improved reporting accuracy for emergency cases.

ACADEMIC PROJECTS:

1.ProActiveCare: AI-Driven Patient Risk Prediction Platform

- Engineered an end-to-end machine learning pipeline to predict ICU patient outcomes, achieving 99.7% accuracy and 95% recall by optimizing **Random Forest** and **Gradient Boosting** models.
- Built a data ingestion framework on **Databricks** using Auto Loader and Delta Lake to clean and transform raw clinical data across medallion architecture (Bronze, Silver, Gold).
- Addressed data imbalance significantly by applying **SMOTE** (Synthetic Minority Over-sampling Technique) and advanced feature imputation, enhancing the model's ability to detect high-risk cases.
- Automated real-time risk alerts by integrating model outputs with external notification workflows, enabling proactive clinical intervention.

2.Personalized Fitness Tracker with AI-Based Insights

- Developed a Full Stack web application incorporating AI-driven analytics to deliver personalized workout recommendations and real-time health tracking.
- Integrated Natural Language Processing (**NLP**) to enable an intelligent conversational interface for effortless exercise logging and user interaction.
- Designed a secure backend architecture with **Node.js** and **Express.js**, implementing JWT authentication to protect user health data.
- Built dynamic dashboards in React to visualize user progress, using machine learning algorithms to adaptively adjust fitness goals based on historical activity data.

3.Home Credit Default Risk Prediction

- Developed comprehensive risk assessment models using Deep Learning (CNN) and ensemble methods to evaluate loan repayment likelihood.
- Conducted extensive Quantitative Analysis on customer demographics and financial history, performing rigorous data cleaning to handle anomalies and missing values.
- Enhanced model performance through advanced Feature Engineering and hyperparameter tuning, successfully identifying key indicators of default risk to support financial decision-making.

4.Traffic Accident Analysis & Prediction

- Leveraged Long Short-Term Memory (LSTM) networks and Gradient Boosting to forecast traffic accident likelihood based on weather, road conditions and temporal factors.
- Processed and cleaned 5 years of large-scale accident data (FARS), ensuring high data integrity for statistical modeling.
- Designed interactive Tableau dashboards to visualize accident hotspots and time-series patterns, providing actionable insights for road safety officials to allocate resources effectively.

SKILLS:

- Data Engineering &Cloud: AWS(Lambda,S3,Glue), Snowflake(Snowpark,Streams,Tasks), Databricks(PySpark,Delta Lake), Airflow.
- Languages & Scripting: Advanced SQL, Python, JavaScript, C.
- Databases: PostgreSQL, Microsoft SQL Server,Google BigQuery,AWS Redshift,SSMS,Oracle SQL.
- ETL & Data Analysis: ETL/ELT Pipeline Design, XML, JSON, Parquet, Alteryx.
- DevOps & CI/CD: Jenkins Pipelines, GitHub, Flyway (Database Version Control).
- Visualization: Tableau, Power BI, Informatica.
- Machine Learning: Linear/Logistic Regression, Random Forest, Gradient Boosting (XGBoost), CNN, RNN.
- Other Tools: JIRA, Microsoft Office (Excel, PowerPoint, Word), AutoCAD.

EDUCATION:

1. Masters Degree in Data Science at University of Texas at Arlington,TX,USA with 4.0 /4.0 GPA.
2. Bachelors Degree in Mechanical Engineering at Jawaharlal Nehru Technological University with 8.03/10 CGPA.

CERTIFICATIONS:

1. Python Essentials by Coursera (2021).
2. Snowflake Cloud Data Practitioner by Coursera (2021).
3. Associate Cloud Engineer from Google Cloud Platform(2021).

AWARDS & RECOGNITION:

1. **Rising Star in Q2 FY-23** for scaling up quickly to deliver high performance in the project.
2. Promotion as **Senior Software Engineer** at Capgemini within 1 year for delivering efficient solutions.