

Data Engineer and Analyst with **over two years of experience** in transforming, optimizing, and analyzing complex datasets. Skilled in **ETL development, query optimization, and workflow automation** using Apache Airflow. I've worked across government and academic sectors, leveraging **cloud platforms like AWS, Azure and Snowflake** to ensure efficient data processing and insightful reporting. My expertise lies in **building scalable data pipelines, enhancing system performance, and developing dashboards that turn raw data into meaningful stories**. I thrive in **collaborative environments**, balancing technical problem-solving with clear communication to bridge the gap between business needs and data-driven solutions.

Skills

- **ETL & Workflow Automation:** Apache Airflow, Microsoft SSIS
 - **Data Processing:** Apache PySpark, Pandas, NumPy , Databricks
 - **Cloud Platforms:** Snowflake, AWS, Azure , GCP
 - **Programming:** Python (PySpark, SQLAlchemy, Pandas), R (dplyr, ggplot2), SQL (PostgreSQL, MySQL, MS SQL Server), SAS
- **Database Management:** RDBMS (PostgreSQL, MySQL), NoSQL (MongoDB)
 - **Business Intelligence:** Power BI, Tableau, Qlik Sense, Excel
 - **Query Optimization:** SQL query tuning, indexing, and performance tuning.
 - **Data Modeling & Analysis:** DAX, Calculated Columns, Measures, Advanced Excel

Work Experience

Data Engineer

State of Florida Department of Health [Division of Emergency Preparedness and Community Support (EMS)]

May 2024 - Present
Florida, USA

- Developed a vehicle tracking system for the **Vehicle Change History project** to enhance disaster planning capabilities. This involved adding new elements to the Florida Data Dictionary, resulting in more reliable and robust disaster response strategies.
- Executed database updates for the **Sex elements project** to ensure data accuracy and compliance with reporting standards. Leveraged existing 'Gender' data to update millions of records, thereby sharpening analytics and fulfilling reporting requirements.
- Designed and implementation of ETL pipelines utilizing **AWS Glue**. Automated the extraction, transformation, and loading of data from diverse sources into **Snowflake**, reducing manual data entry time by 25%.
- Optimized large-scale data processing workflows using **PySpark on AWS EMR**, handling over 10 million records and achieving a 30% increase in processing speed
- Enhanced SQL query performance and database indexing in **PostgreSQL**, leading to a 30% reduction in query execution time and more efficient reporting processes.
- Integrated AWS Glue jobs with **Lambda functions** to enhance event-driven data processing, ensuring real-time updates across systems.
- Orchestrated data workflows using **Apache Airflow**, improving scheduling reliability and reducing data latency by 20%.
- Monitored and managed ETL job runs using **AWS CloudWatch** to keep jobs running smoothly, maintaining a 99% success rate in job executions.

Data Analyst

State of Florida Department of Health [Division of Medical Quality Assurance (MQA)]

Oct 2023 – May2024
Florida, USA

- Collaborated with stakeholders to define data requirements for the V3.4 to V3.5 Remediation project, ensuring alignment with the new V3.5 standards for the MQA Operational Data Store (ODS).
- Developed scalable **AWS Glue pipelines** to automate data extraction from various healthcare systems, increasing data processing efficiency by 35%.
- Managed the migration of legacy systems to **Snowflake** data warehouse, improving the scalability, availability, and security of healthcare data.
- Used **Spark on AWS Lambda (SoAL) framework** to streamline the transformation of medical data into reporting-ready formats, ensuring data quality and improving retrieval times by 25%.
- Designed **Power BI** dashboards for MQA to enhance **statewide submission quality monitoring**, resulting in a 30% improvement in data quality metrics.
- Created **DAX** formulas using Calculated Columns, Measures, and the CALCULATE () function for accurate data modeling and numerical analysis, leading to the development of interactive dashboards that revamped data visualization and decision -making efficiency by 15%.
- Prepared a financial report during the state emergency using **Excel** Pivot, VBA, and DAX formulas, leading to a 30% improvement in resource management efficiency.
- Generated weekly **ad-hoc analysis reports** for management using **SQL Server** and Excel, resulting in a 20% improvement in decision-making efficiency.
- Automated reporting workflows using **SQL and Python**, streamlining the delivery of critical performance metrics to stakeholders, which amended data retrieval speed by 10% and ensured 99% data accuracy and quality.

Data Research Analyst

George Mason University

Nov 2022 - May 2023
Virginia, USA

- Developed and optimized data processing pipelines using **ADF, Python, and Databricks** to handle large volumes of academic and demographic data for the **student cohort project**, achieving a 30% increase in efficiency and reducing manual intervention.
- Assisted in optimizing the data warehouse on **Azure Synapse Analytics** by leveraging ADF for data ingestion and transformation, improving query performance by 40%.
- Developed student cohort trends data for visualization in **Power BI**, enabling predictive insights that enhanced data-driven decision-making for university administration.
- Prepared professional **paginated reports in Power BI** to compile detailed findings and analyses for stakeholders.

Education

George Mason University

Master's in Data Analytics Engineering

Aug 2021- May 2023
Virginia, USA

Gokaraju Rangaraju Institute of Engineering & Technology

B.Tech, Computer Science & Engineering

Aug 2017 - Jul 2021
Telangana, India