Sai Varun Kumar Namburi

Boston, MA | 857-265-1349 | saivarun.namburi@gmail.com| linkedin.com/in/sai-varun-kumar-namburi| GitHub

Result-driven Data Engineer with over **2.5 years** of hands-on experience in driving **data transformation** initiatives within fast-paced environments. Proficient in **AWS technologies, Python, and SQL**, adept at orchestrating **end-to-end ETL/ELT pipelines** to ensure data accuracy, compliance, and transparency, and demonstrated expertise in automating workflows, optimizing data processing

EDUCATION

Northeastern University | Boston, USA | *GPA 3.82/4.00* *Sep 2022 - May 2024*
Master of Science in Data Analytics Engineering

Malla Reddy College of Engineering and Technology | Hyderabad, India | *GPA 8.72/10.00* *Aug 2016 - Aug 2020*
Bachelor of Technology in Computer Science and Engineering

WORK EXPERIENCE

Professional 1 Data Engineer | DXC Technology | Bangalore, India *Nov 2021 - Aug 2022*
- Rolled out intricate data transformation logic in **AWS Glue** using **Python**, and **Pyspark** ensuring **90% accuracy** and consistency in big data transformations which are loaded into **Snowflake Data warehouse**
- Automated **Glue notebooks** execution via data pipeline using **Apache Airflow**, reducing manual intervention by **30%**, ensuring timely data ingestion into **Redshift**
- Revolutionized data vault with **AWS Glue, orchestrating ETL/ELT pipeline** lineage, achieved 95% accuracy in tracking data flow transformations, ensuring compliance and transparency
- Enhanced **AWS-based data analytics** solution by leveraging services like **S3, SNS, Lambda, and Redshift**, resulting in a 40% improvement and 25% cost reduction, delivering actionable insights for informed decision-making.

Associate Professional Data Engineer | DXC Technology | Bangalore, India *May 2020 - Oct 2021*
- Engineered and deployed a robust **Data Quality and Data Ingestion** tool in **AWS EMR**, utilizing **Pyspark** & **Spark SQL** scripts for performing quality checks, which helped to increase the data consistency by 80%
- Employed **RDS Postgres** to establish a robust database for **a Data Quality tool**, resulting in **a 65% acceleration** in **data retrieval**
- Implemented **Django Rest framework-based APIs**, facilitating a significant reduction of 70% in errors for metadata retrieval
- Initiated development of **5+ Power BI** dashboards to visualize data quality and anomalies for insightful visualization, empowered business stakeholders to make informed decisions, resulting in a 35% increase in data governance and business decisions
- Implemented robust **CI/CD pipelines using AWS Code Build** to streamline codebase and deployment processes, utilizing **GitHub** for version control to ensure seamless collaboration and code integration within the development team

TECHNICAL SKILLS

- Programming Languages : Python, PySpark, Scala, Shell Scripting, Parquet, Avro, JSON
- Databases : RDS SQL, PostgreSQL, Redshift, Snowflake, MariaDB
- Software & BI Tools : MS Excel, **Talend Open Studio**, GitHub, Docker, Code Build, Tableau, Power BI
- Cloud Technologies : AWS, SQS, Kinesis Data Stream, AWS Glue, IAM, Airflow, S3, Lambda Functions, EMR, SNS

ACADEMIC PROJECTS

Sales Order Processing – Real-time CDC using AWS (Github) *Feb 2024 - Mar 2024*
- Spearheaded the execution of a real-time change **CDC** solution leveraging **AWS** services including **DynamoDB, Kinesis Data Streams, and Kinesis Firehose**, resulting in a **50% reduction** in data processing latency and increased **data accuracy by 30%**
- Integrated **AWS Lambda and Glue** for streamlined data transformation and cataloging, enabling seamless querying of processed data through **Athena** resulting in **40% improved decision-making** efficiency and **actionable insights** for stakeholders

DoorDash Delivery Processing using AWS (Github) *Jan 2024 - Feb 2024*
- Engineered an end-to-end solution utilizing AWS services like **S3, Lambda, and Event Bridge**, **SNS** streamlining the processing of daily delivery data, achieving a **30% reduction** in processing time and a **25% improvement** in data accuracy
- Designed and deployed a streamlined **CI/CD pipeline with AWS Code Build**, resulting in a **40% increase** in deployment efficiency and **enhanced scalability** for fluctuating data volumes.

CERTIFICATIONS / LEADERSHIP ACTIVITIES

- **Azure Data Engineer Certified** (Microsoft DP-203) *May 2023 – May 2024*
- **Apache Airflow** Fundamentals Certified (Astronomer) *Jul 2023 – Present*
- Mentored **40+ students** in creating **conceptual designs of data warehouses** and building them in **star, snowflake, and star-flake schemas**. Led hands-on exercises sourcing operational data from diverse databases, utilizing ETL techniques in **Talend**