Venkata Soma Sekhar Data Engineer

| somasekhargowdsonti@gmail.com | New Jersey | +1(332) - 268 - 5305 | https://www.linkedin.com/in/v-soma-sekhar-gowd-sonti7/ |

Profile Summary

- **Data Engineer with around 3 years of experience** in designing, developing, and optimizing **scalable data pipelines** using **Apache Spark, Kafka, and Airflow**, reducing **processing time by 45%** and enhancing **real-time data flow for 10,000+ records daily**.
- **Proficient in cloud-based data solutions**, deploying and managing **AWS (Glue, Redshift, S3), Azure (Data Factory, Synapse), and GCP (BigQuery, Dataflow)** to improve **data retrieval speeds by 35%** and enable seamless cross-platform data integration.
- **Expert in implementing** data governance, security, and compliance frameworks, ensuring adherence to **GDPR, SOC 2, and HIPAA regulations**, improving **data integrity and security for 100+ stakeholders**.
- **Skilled in automating** data workflows using **Python, SQL, and Spark**, reducing **manual intervention by 50%** and enabling **real-time analytics**, leading to faster and more accurate decision-making
- **Experienced in database optimization and performance tuning**, leveraging **PostgreSQL, MySQL, and NoSQL databases (MongoDB, Cassandra)** to improve query execution and streamline data storage solutions.

Technical Skills

Programming Languages: Python, SQL, R, Scala, MATLAB
Databases: MySQL, MS SQL Server, PostgreSQL, Snowflake, BigQuery, MongoDB (NoSQL)
Big Data & Cloud: Apache Spark, Databricks, Azure Synapse, AWS, Azure, GCP, S3, GCS, Delta Lake, EC2
ETL & Data Integration: Apache Airflow, Matillion ETL, MSBI (SSIS, SSAS, SSRS)
Data Processing & Orchestration: Apache Flink, Apache Nifi
Version Control & DevOps: Git, Bitbucket, Jupyter, REST APIs, DevOps
Data Visualization: Tableau, Power BI, Qlik
Workflow & Project Management: Jira, Confluence, Agile, Data Pipelines
Analytics & Data Science: Time Series Analysis, A/B Testing, Regression, Hypothesis Testing, Probability & Statistics

Professional Experience

Data Engineer, *Honeywell* *Jan 2024 – Present | Remote, USA*

- **Architected and optimized** high-performance **big data processing frameworks** using **Apache Spark, Databricks, and Delta Lake**, improving **data transformation efficiency by 45%** and ensuring high-throughput batch and streaming workloads.
- **Formulated and automated** ETL workflows using **AWS Glue, Redshift, and Lambda**, eliminating **50% of manual data processing tasks** increasing efficiency, and ensuring **high pipeline reliability**.
- **Implemented data governance frameworks**, enforcing **RBAC, encryption standards, GDPR, and SOC 2 compliance**, securing sensitive **financial and IoT datasets** while maintaining regulatory adherence.
- **Engineered and managed** real-time **data ingestion pipelines** using **Apache Kafka and AWS Kinesis**, enabling **millisecond-latency streaming analytics** for industrial automation and predictive maintenance.
- **Integrated advanced business intelligence solutions**, connecting **Tableau, Power BI, Looker, and cloud-based data lakes**, developing **30+ automated dashboards** that provided real-time business insights and operational efficiency.
- **Worked cross-functionally with data architects, DevOps teams, and business leaders**, aligning **data strategies with scalable architecture solutions**, improving reporting accuracy, and reducing data pipeline downtime.

Data Engineer, *IBM* *Apr 2021 – Dec 2022 | India*

- **Designed and implemented** scalable **ETL pipelines** using **Apache Spark, Apache Airflow, and Python**, processing **10TB+ of structured and unstructured data**, optimizing **data extraction, transformation, and loading** workflows, and **reducing processing time by 40%**.
- **Enhanced database performance** for **PostgreSQL, MongoDB, and Snowflake**, implementing **advanced indexing, partitioning, and materialized views**, which **cut query execution time by 35%** and improved data retrieval speeds for analytics teams.
- **Constructed and deployed** real-time data streaming frameworks using **Apache Kafka and Apache Flink**, enabling **low-latency data processing** and improving data availability for real-time analytics in mission-critical applications.
- **Built and managed** large-scale **data warehouses** with **Google BigQuery and Amazon Redshift**, ensuring efficient storage, retrieval, and processing of high-volume datasets, improving **data accessibility for 100+ business reports**.
- **Automated and optimized** data validation, quality checks, and anomaly detection using **Python (Pandas, NumPy) and SQL scripts**, ensuring **25% fewer discrepancies** and maintaining high data integrity standards.
- **Collaborated closely with data scientists, analysts, and software engineers**, optimizing **machine learning feature engineering pipelines** and ensuring seamless integration of **AI/ML models** into production systems.
- **Deployed cloud-based data solutions** on **AWS (S3, Lambda, Glue) and GCP (Dataflow, BigQuery, Pub/Sub)**, implementing **cost-efficient storage and processing architectures**, reducing infrastructure overhead.
- **Established comprehensive documentation, standardized workflows, and implemented CI/CD pipelines** for **automated deployment of data pipelines**, improving maintainability, scalability, and **reducing debugging time by 30%**.

Education

Master of Science in Information Systems Technology *Jan 2023 – Apr 2024 | Delaware, USA*

Wilmington University

- **Relevant Courses:** SQL, Database Management, Data Mining, Applied Statistics, Data Visualization, Database Management, Cryptography, Applied Natural Language, Operating Systems, Object Oriented Programming, Python, R, OOP, DSA