

Sumanth Kumar Sivadi

DATA ENGINEER

Georgia, USA | Phone no: (470) 252-6919 | Mail ID: sivadisumanthkumar@gmail.com | [LinkedIn](#)

SUMMARY

Data Engineer with 3+ years of experience in the healthcare and financial services domains, specializing in the design and development of scalable data pipelines, ETL/ELT processes, and cloud-based data platforms. Highly proficient in Python, SQL, and PySpark, with hands-on experience across Azure, AWS, Databricks, and Snowflake data warehousing solutions. Strong expertise in data modeling (Star and Snowflake schemas), data quality validation, and performance optimization to support analytics and reporting. Proven ability to thrive in Agile environments, delivering reliable, production-ready data solutions that enable dashboards, regulatory reporting, and actionable business insights.

SKILLS

| | |
|---|---|
| Programming Language: | Python, SQL, Java |
| Big Data & Processing Tools: | Hadoop, MapReduce, HDFS, Hive, NIFI, Apache Kafka, Apache Spark, PySpark |
| Cloud Technologies: | Microsoft Azure (Azure Data Factory, Azure Synapse Analytics, Azure Data Lake Storage, Azure Blob Storage, Azure SQL Database, Databricks, Microsoft Azure SQL Managed Instance), AWS (AWS Glue, Amazon Redshift, Amazon S3 & Lake Formation, Amazon RDS/Aurora, EMR, Amazon Athena, Amazon QuickSight), Microsoft Fabric |
| Database: | MySQL, MongoDB, PostgreSQL, SQL Server |
| Packages: | Pandas, NumPy, Matplotlib, SciPy, Seaborn, Scikit-Learn |
| Data Visualization Tools: | Tableau, Power BI, Advanced Excel, Statistics |
| IDEs: | Visual Studio Code |
| Other Technical Skills: | Data Mining, Data warehousing, Data transformation, KPI, Data Modelling (Star Schema, Snowflake Schema), dbt, delta lake, ETL/ ELT Design, A/B Testing, Data Wrangling, Jira, Git, GitHub, CI/CD Concepts, Prompting |
| Methodologies: | SDLC, Agile, Waterfall |

RELEVANT EXPERIENCE

- Data Engineer, WellStar Health System | GA**

Aug 2024 – Current

 - Developed and maintained Python- and SQL-based ETL workflows to process high-volume EHR, claims, and patient encounter data, implementing data quality checks, schema evolution handling, and partitioning strategies for large healthcare datasets.
 - Assisted in building scalable data ingestion pipelines using Apache Spark (PySpark) and Azure Data Factory, automating ingestion from EHR systems, clinical applications, and downstream healthcare systems, reducing manual intervention and improving pipeline reliability.
 - Contributed to data warehouse modeling in Snowflake and Azure Synapse using Star and fact-dimension schemas, optimizing query performance for clinical operations, quality metrics, and financial reporting dashboards.
 - Optimized Power BI report performance by reducing model complexity, tuning relationships, and leveraging aggregations, improving clinical and operational dashboard load times by 30%.
 - Leveraged MapReduce for batch processing of raw healthcare logs and clinical event data, transforming semi-structured data into structured datasets for downstream analytics.
 - Implemented data quality validation and reconciliation logic in Azure Databricks to detect duplicate, missing, or inconsistent patient, encounter, and claims records, reducing data inconsistencies by 15%.
- Data Engineer, Cognizant | India**

Dec 2021 – Jul 2023

 - Developed and optimized Tableau dashboards to visualize payment authorizations, settlements, fraud alerts, and regulatory compliance metrics, improving report load times.
 - Implemented robust data quality validations, schema evolution handling, and partitioning strategies, reducing processing errors by 20% and ensuring reliable downstream analytics.
 - Built scalable data ingestion workflows using PySpark and AWS Glue, automating batch and near real-time data collection from payment gateways, core banking systems, and downstream reporting platforms, reducing manual intervention by 35%.
 - Designed and implemented MongoDB collections to store semi-structured transaction metadata, customer profiles, and audit logs, enabling flexible schema evolution without impacting core ETL pipelines.
 - Processed high-volume banking and payment transaction logs using Hadoop HDFS, enabling reliable storage and batch processing for millions of daily records.
 - Monitored and tuned Apache NiFi pipelines using Provenance, metrics, and alerts, proactively identifying bottlenecks and preventing delays in high-volume transaction processing.
 - Ensured secure handling of sensitive financial data in Python workflows using encryption libraries and secure credential management, complying with internal governance standards.
 - Configured automated notifications and alerting within CI/CD pipelines to notify stakeholders of build failures or errors in real time.

EDUCATION

- Master of Science in Applied Computer Science – Northwest Missouri State University, MO, USA**
- Bachelor of Technology in Electronics and Communication Engineering – Audisankara College of Engineering and Technology, India**