



SAI SUMANTH U

Edison, NJ 08820 | (929)367-0483 | saisumanthu85@gmail.com

Professional Summary

Results-driven **Data Engineer** with 6 + years of experience in **designing, developing, and optimizing large-scale data engineering solutions** across industries like **finance, healthcare, and technology** . Expertise in **Azure, Snowflake, Databricks, and SQL** , specializing in **real-time data processing, ETL/ELT workflows, and cloud data architectures** . Skilled in **Azure Data Factory (ADF), Synapse Analytics, Apache Spark (PySpark, Spark-SQL, Scala), and Kafka** for high-performance data integration. Strong background in **DevOps, CI/CD automation, and infrastructure management** , utilizing **Terraform and Azure DevOps** . Proficient in **data security compliance (HIPAA), role-based access control (RBAC), and encryption** . Developed **machine learning-driven insights** using **Azure ML and Databricks MLflow** , supporting fraud detection, risk assessment, and business intelligence. Adept at **collaborating with cross-functional teams** to deliver **scalable, secure, and high-performing data solutions** that drive innovation and efficiency across multiple domains.

Skills

Programming Languages: Python, C++, Java, SQL

Data Integration & ETL/ELT: Informatica, DBT, Fivetran, Apache Kafka, Airflow, Snowflake, Azure Data Factory

Data Visualization: Tableau, Power BI, Matplotlib, Pandas

DevOps Tools: Terraform, Ansible, Kubernetes

Big Data & Analytics: Azure Databricks, Azure Synapse Analytics, PolyBase Apache Spark, PySpark, Hadoop, Delta Lake, Hive

Cloud Platforms: Microsoft Azure, AWS, GCP

APIs & Integration: REST APIs, SOAP APIs, JSON, XML

Work History

Data Engineer

Fiserv

Feb 2023 - Current

NJ

- **Designed and implemented scalable data ingestion pipelines** using **Azure Data Factory (ADF)** to extract, transform, and load (**ETL**) data from diverse sources like **Oracle, SAP HANA, MySQL, and manual files** into **Azure Data Lake Storage (ADLS Gen2)**
- **Developed and optimized big data processing workflows** using **Azure Databricks, PySpark, and Spark SQL** , ensuring efficient handling of high-volume datasets and performance tuning for real-time analytics.
- **Automated ETL processes** by implementing **scheduling, triggers, and monitoring mechanisms** in **Azure Data Factory (ADF)** and **Databricks workflows** , reducing manual intervention and

ensuring data accuracy.

- **Managed and stored structured and semi-structured data** in **Azure SQL DB** and **Snowflake**, ensuring data integrity, availability, and security through effective schema design and governance.
- **Developed and optimized SQL queries and stored procedures** for data transformation and analysis, improving query performance in **Azure Synapse Analytics, Snowflake, and PostgreSQL** environments.
- **Implemented real-time data streaming solutions** using **Kafka and Spark Streaming**, enabling low-latency data processing and analytics for live event-driven applications.
- **Ensured data security and governance** by implementing **RBAC (Role-Based Access Control), encryption, masking techniques, and Snowflake Secure Views**, following compliance and best practices.
- **Optimized cloud infrastructure** by configuring **Azure Kubernetes Service (AKS), Docker containers, Terraform, and CI/CD pipelines in Azure DevOps**, improving scalability and deployment efficiency.
- **Performed data validation, logging, and troubleshooting** using **Datadog, Snowflake resource monitors, and ADF monitoring tools** to track ETL performance, reduce failures, and enhance cost management.
- **Created dashboards and reports** using **Power BI, Tableau, and Matplotlib**, delivering actionable insights by visualizing key performance indicators (KPIs) and trends in business data.

Data Engineer

Apr 2022 - Jan 2023

Change Healthcare

TN

- **Developed and optimized scalable ETL pipelines** using **Azure Data Factory (ADF) and Azure Synapse Analytics**, extracting, transforming, and loading data from diverse sources like **Azure Data Lake Storage (ADLS Gen2), SQL Server, Cosmos DB, and APIs** into Azure-based data warehouses.
- **Built and deployed real-time data streaming solutions** using **Azure Event Hubs, Azure Stream Analytics, and Azure Data Explorer**, processing high-velocity data for analytics and reporting.
- **Designed and developed data processing applications** using **Azure Databricks (PySpark, Spark-SQL, Scala)**, implementing complex **data transformations, aggregations, and business logic**, optimizing performance with partitioning, bucketing, and indexing.
- **Implemented DevOps and automation solutions** using **Azure DevOps, Azure Kubernetes Service (AKS), Terraform, and CI/CD pipelines**, managing cloud infrastructure and automating deployments.
- **Ensured data security and compliance and RBAC policies**, implementing **data masking, encryption, Azure Key Vault, Managed Identities, and secure data sharing in Synapse and Snowflake**.
- **Developed and scheduled workflows in Azure Data Factory (ADF) and Apache Airflow** for automating data pipelines, monitoring performance, and ensuring pipeline reliability with alerts and logging in **Azure Monitor**.
- **Designed and optimized Azure-based data warehouses** using **Azure Synapse Analytics**,

Snowflake, and Azure SQL Database , leveraging **PolyBase, Data Partitioning, and Materialized Views** for performance tuning.

- **Performed advanced data profiling, validation, and integrity checks** across **structured, semi-structured, and unstructured data** , using **Azure Purview, SQL, and PySpark** for anomaly detection and data quality improvements.
- **Implemented machine learning and predictive analytics solutions** by integrating **Azure Machine Learning (AML), Azure Cognitive Services, and Databricks MLflow** with data pipelines for real-time insights.
- **Collaborated with cross-functional teams** including **data analysts, data scientists, and business stakeholders** to ensure Azure-based data solutions met business needs, delivering accurate and timely insights using **Power BI and Synapse Analytics** .

Data Engineer

Feb 2018 - Jul 2021

Xbreach Technologies

Hyd, TG

- **Designed and implemented a modern analytics platform** leveraging **Azure and Snowflake** to support **real-time data processing and visualization** , enabling data-driven decision-making across the organization.
- **Developed and optimized ETL workflows** using **SSIS** , implementing various **data transformations, error handling, event handlers, and automated notifications** to improve data pipeline reliability.
- **Created and maintained parameterized reports, sub-reports, and ad-hoc reports** using **SQL Server Reporting Services (SSRS)** , optimizing **caching and report performance** for business users.
- **Developed and optimized SSIS packages for seamless data flow between OLTP and OLAP systems** , improving **ETL efficiency and execution times** .
- **Performed complex SQL queries and stored procedure optimization** for generating **state compliance reports** , improving data retrieval and reporting efficiency.
- **Designed and developed multi-dimensional OLAP cubes** using **MDX scripts in SQL Server Analysis Services (SSAS)** to enable high-performance analytical reporting.
- **Executed bulk data migration from heterogeneous data sources (Flat Files, XML, Excel, and MS Access)** to **SQL Server using SSIS** , ensuring data integrity and consistency.
- **Developed and scheduled SSIS packages with package configurations, event handling, and error logging mechanisms** , enhancing troubleshooting and system monitoring.
- **Implemented complex transformations in SSIS** , including **Lookup Transformations, Merge Joins, Derived Columns, Conditional Splits, and Data Conversions** , streamlining ETL processing.
- **Deployed and maintained SQL Server Reporting Services (SSRS) reports** using **expressions, global variables, and custom code** , ensuring accurate **data visualization and compliance reporting** .

Education

Master of Science: Computer Science

University of Dayton

Dec 2022

Dayton, OH

Certifications

- Microsoft Certified Azure Fundamentals
- Databricks Lakehouse Fundamentals
- Power BI Beginner to Pro Workshop