

Sneha Mariam Santhosh

Data Engineer

+1 647-269-8404

snehams.santhosh@gmail.com

[linkedin.com/in/sneha-santhosh](https://www.linkedin.com/in/sneha-santhosh)

Toronto, Canada

PROFESSIONAL SUMMARY

Seasoned Data Engineer with 5 years of hands-on expertise in designing and implementing enterprise-scale data solutions, showcasing a proven track record in **Azure** cloud platforms and **Databricks**, real-time analytics, and ML pipeline optimization across retail, manufacturing, energy and finance sectors.

Passionate about revolutionizing data engineering practices to deliver exceptional business intelligence and analytics solutions consistently.

My technical expertise includes:

Data Platform Architecture

- Designed and delivered Azure-based lakehouse solutions using **Databricks**, **Delta Lake**, **Delta Live Tables**, and **Unity Catalog**, aligned with **medallion architecture** and best practices.
- Applied Spark optimization techniques (adaptive query execution, Z-ordering, partition pruning) with a strong focus on performance and maintainability.

Pipeline Engineering & Problem Solving

- Engineered 100+ resilient **ETL/ELT pipelines** using **Azure Data Factory**, **Apache Airflow**, and **Python**, incorporating complex dependencies, CDC logic, and retry mechanisms.
- Demonstrated strong **analytical thinking** and **root cause analysis** skills to resolve production issues and improve reliability (99.9% data accuracy).

Cloud-Native Development & Adaptability

- Developed robust **CI/CD pipelines** using **Azure DevOps**, promoting automation, reusability, and faster time to production.
- Embraced continuous learning and adaptability in rapidly evolving tech environments.

Programming & Optimization

- Expert in **Python**, **PySpark**, and **SQL**, delivering modular and scalable data transformations.
- Optimized performance with **broadcast joins**, **bucketing**, and **caching**, achieving up to 90% improvement in processing efficiency.

Data Governance, Security & Ownership

- Implemented governance strategies using **Unity Catalog**, **RBAC**, and metadata automation—ensuring data integrity, compliance (GDPR/CCPA), and **enterprise trust**.
- Took full ownership of access control, encryption standards, and lineage tracking across data assets.

Advanced Analytics & Collaboration with Data Science

- Supported **MLOps pipelines** (MLflow) with feature stores and automated model deployment.
- Proactively collaborated with data scientists, providing clean, production-ready data for modeling and experimentation.

Business Intelligence & User Focus

- Designed **dimensional models** and **star schemas** to support high-performance BI systems.

- Delivered actionable insights through **Power BI, Tableau**, and embedded dashboards always with a strong focus on **end-user needs** and **clarity**.

Cross-Functional Collaboration & Agile Delivery

- Actively engaged in **Agile/Scrum teams**, leading sprint planning, demos, and retrospectives using **JIRA, Confluence**, and **Azure Boards**.
- Known for **clear communication, knowledge sharing**, and technical documentation that improves onboarding and sustainability of solutions.

Database Systems & Cloud Storage Efficiency

- Worked across diverse systems: **SQL Server, Snowflake, Cosmos DB** and **Netezza**.
- Optimized cost and performance by leveraging modern file formats (**Parquet, Delta, Iceberg**) and partitioning strategies.

TECHNICAL SKILLS

Cloud Platforms : Azure Databricks (Delta Live Tables, Unity Catalog), Azure Data Factory (ADF), Azure Data Lake Storage (ADLS Gen2), Azure Synapse Analytics, Microsoft Fabric, Apache Spark

Programming & Automation : Python, PySpark, SQL, T-SQL, Azure DevOps, Git

Data Engineering & Integration : ETL/ELT Pipelines, Delta Lake, Medallion Architecture (Bronze/Silver/Gold), Change Data Capture (CDC), Data Modeling, Source-to-Target Mapping, Data Migration, Data Validation & Cleansing

Workflow Orchestration & Monitoring : Apache Airflow, Azure Monitor, Spark UI, Job and Cluster Configuration, Performance Tuning

Data Quality, Governance & Security : Unity Catalog, Data Lineage, Metadata Management, Data Profiling, RBAC, Access Control, Data Quality Frameworks, Azure AD Authentication

Databases & Business Intelligence : Microsoft SQL Server, Snowflake, Power BI

PROFESSIONAL EXPERIENCE

Landmark Group - Data Labs Landmark

May 2023 - April 2025

Data Engineer

- **Contributed to the design, development, testing, and deployment of scalable data and AI products**, partnering with business stakeholders, product owners, and data analysts to translate requirements into actionable insights aligned with business objectives.
- **Built and maintained 50+ robust ETL/ELT pipelines using Azure Databricks, Data Factory, and Delta Lake**, processing structured, semi-structured, and unstructured data sources, reducing manual data preparation by 80%.
- **Developed reusable frameworks and automation scripts for data quality validation**, profiling, and anomaly detection, accelerating development cycles and reducing manual intervention in testing workflows.
- **Applied data modeling and profiling techniques** to design conceptual, logical, and physical data architecture supporting advanced analytics, reporting, and data science initiatives across enterprise systems.
- **Enforced metadata management, lineage tracking, and centralized access control using Unity Catalog**, improving data governance, auditing capabilities, and regulatory compliance alignment.
- **Automated CI/CD deployments** for data pipelines using Azure DevOps, reducing release cycles by 50% and enhancing deployment efficiency in a digital-first environment.
- **Optimized Spark jobs** through partitioning, caching, and adaptive query optimization, improving batch and streaming data processing performance by 90%.

- **Participated in Agile ceremonies** using JIRA, collaborating with cross-functional teams to gather requirements, deliver on sprint commitments, and maintain technical documentation for knowledge management.
- **Led data engineering initiatives** supporting inventory demand forecasting, predictive analytics, and GA4 integration, delivering curated datasets to data scientists for model development and feature engineering.

Infosys Limited

January 2021 - April 2023

Client : Siemens Gamesa, AMS Osram, Huhtamaki

- Developed and maintained **70+ end-to-end ETL/ELT pipelines** using **Azure Data Factory, Databricks, Logic Apps, and Python**, enabling data-driven decisions across **renewable energy, manufacturing, and packaging** sectors.
- Engineered scalable **Azure data lake solutions**, processing **2TB+ weekly** from diverse sources including **SAP, APIs, SQL databases, and ADLS Gen2**.
- Integrated data into **SQL Server** and **Snowflake**, enhancing performance and accessibility for analytics and reporting teams.
- Achieved **40% reduction in pipeline downtime** through in-depth root cause analysis and proactive SLA-driven monitoring improvements.
- Automated **API-based ingestion** using **PowerShell** and **Python**, cutting refresh times by **40%** and streamlining workflows.
- Deployed **CI/CD pipelines** with **Azure DevOps**, improving release reliability by **40%** and accelerating delivery cycles.
- Managed **Apache Airflow DAGs** for orchestration, reducing batch failures by **25%** via intelligent retries and proactive monitoring.
- Authored **optimized SQL queries and stored procedures**, improving reporting performance and reducing data latency by **70%**.
- Led production deployments and ensured **99.9% system availability**, supporting operational continuity and compliance.
- Contributed actively to **Agile delivery**—sprint planning, backlog grooming, demos, and documentation using **JIRA** and **Confluence**.

EDUCATION

Bachelor of Engineering in Computer Science

APJ Abdul Kalam Technological University - 2020

CERTIFICATION

Microsoft Azure Data Fundamentals

Databricks Certified Data Engineer Associate - Pursuing

Microsoft Certified: Fabric Data Engineer Associate - Pursuing