

Sneha Mariam Santhosh

Azure Data Engineer

+1 647-269-8404

sneham.santhosh@gmail.com

linkedin.com/in/sneha-santhosh

Toronto, Canada

SUMMARY

Azure Data Engineer with 5 years of experience delivering scalable data solutions across retail and finance. Skilled in ETL/ELT pipeline design, real-time processing, and data governance using Azure technologies, ensuring high data quality and operational efficiency.

- Designed and implemented ETL/ELT pipelines using Azure Databricks, Data Factory
- Proficient in Lakehouse and Medallion architectures for modern data platforms
- Enabled batch and real-time data processing with scalable pipelines
- Developed robust data validation frameworks and monitoring systems
- Strong command over SQL, PySpark, and Power BI for analytics and reporting
- Automated deployments and CI/CD workflows using Azure DevOps
- Ensured data governance, lineage tracking, and compliance in production systems

TECHNICAL SKILLS

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

Programming & Automation: Python, PySpark, SQL, T-SQL, PowerShell, CI/CD (Azure DevOps)

Data Engineering & Integration: ETL / ELT Pipelines, Delta Lake, Real-time & Batch Processing, Source to Target Mapping, Data Modeling, Data Migration, Data Validation & Cleansing

Workflow Orchestration & Monitoring: Apache Airflow, Azure Monitor, Spark UI, Failure Handling & Alerting

Data Governance & Security: Unity Catalog, Data Lineage, Data Quality, Metadata Management, Privacy & Compliance

Databases & BI Tools: SQL Server, Snowflake, Azure SQL, Power BI

PROFESSIONAL EXPERIENCE

Landmark Group

May 2023 - April 2025

Data Engineer

- Collaborated with business stakeholders, product owners, and data analysts to gather requirements, define KPIs, and translate business needs into scalable data engineering solutions.

- Designed and executed data models (conceptual, logical, and physical) to support reporting, analytics, and data science use cases in the retail domain.
- Built and maintained 50+ robust ETL/ELT pipelines using Azure Databricks, Data Factory, and Delta Lake, reducing manual data prep by 80% and saving 60+ engineering hours/month.
- Established a comprehensive data quality framework, integrating validation, profiling, and anomaly detection checks, reducing production data issues by 95%.
- Enforced metadata management and governance controls using Unity Catalog, enabling centralized access control, auditing, and lineage tracking across 30+ Delta tables.
- Automated CI/CD deployments for data pipelines using Python and Azure DevOps, cutting release cycles by 50% and ensuring consistent delivery to production.
- Tuned Spark jobs through partitioning, caching, and adaptive query optimization to improve performance by 90% and reduce cloud compute costs.
- Proactively monitored pipeline health using Azure Monitor, Log Analytics, and Spark UI, leading to an 85% reduction in SLA breaches and operational incidents.
- Curated and maintained 100+ certified datasets in Power BI, accelerating real-time analytics adoption and C-level reporting.
- Led strategic data engineering initiatives for inventory management and fast fashion analytics, enabling real-time stock tracking and demand forecasting.

Infosys Limited

Client - **Siemens Gamesa, Renewable Energy**

Data Engineer

October 2022 – April 2023

- Managed and monitored 30+ Azure-based data pipelines, ensuring 99.9% data availability and timely ingestion across reporting systems.
- Conducted root cause analysis on recurring failures, reducing pipeline downtimes by 40% and improving SLA adherence.
- Developed automation scripts using Azure Databricks (PySpark) and Python, cutting manual interventions by 70% and accelerating issue resolution time by 60%.

Client - **AMS Osram**

Global leader in optical solutions

Data Lake Engineer

March 2022 - October 2022

- Partnered with business stakeholders to gather requirements and design Azure-based data lake solutions supporting manufacturing analytics.
- Developed 10+ data ingestion and transformation pipelines in Azure Databricks, processing 2TB+ of data weekly from CSV, JSON, and Parquet sources in ADLS Gen2.
- Integrated transformed datasets into SQL Server and Snowflake, improving data accessibility for business intelligence teams.

- Automated data extraction from Swagger APIs via PowerShell scripts, cutting refresh time by 40% and reducing manual touchpoints.
- Deployed CI/CD pipelines using Azure DevOps, increasing release reliability by 40% and reducing rollback incidents.
- Managed Apache Airflow DAGs to orchestrate daily workflows, decreasing batch failures by 25% through alerting and retry logic.
- Worked in Agile sprints, contributing to backlog grooming, sprint demos, and technical documentation using Jira and Confluence.

Client - **Huhtamaki, Espoo**

Global provider of sustainable packaging solutions

Data Engineer

January 2021 – March 2022

- Designed and developed 20+ ETL pipelines using Azure Data Factory, Logic Apps, and Databricks to support financial and operational reporting.
- Integrated 15+ data sources (APIs, SAP ERP, SQL) into SQL Server to enable enterprise-level analytics.
- Authored optimized SQL queries and stored procedures, reducing data latency by 70% and enhancing reporting accuracy.
- Led production deployment and monitoring of critical data workflows, ensuring data quality and system stability.
- Conducted structured root cause analyses to resolve long-standing data discrepancies and performance bottlenecks.

EDUCATION

Bachelor of Engineering in Computer Science

APJ Abdul Kalam Technological University - 2020

CERTIFICATION

Microsoft Azure Data Fundamentals