Venkata Narendra Kumar Sykam

(940) 354-4796
venkatanksykam@gmail.com
https://linkedin.com/in/venkatanksykam
https://sykamsolutions.com/portfolio/
https://github.com/SykamVenkata
Dallas Fort Worth Metroplex, Texas

Senior Azure Data Engineer with 8+ years of experience architecting scalable data pipelines, cloud data lakehouse, enterprise data warehouse solutions on Microsoft Azure. Led Databricks-to-Microsoft Fabric migration and processing 2M+ hourly connected vehicle telemetry events using Azure Databricks, OneLake, Azure Data Factory, Python, SQL, PySpark. Skilled in data governance, FinOps cost optimization, prod ML feature pipelines.

TECHNICAL SKILLS

Programming: Python, SQL, T-SQL, PySpark, Spark SQL, KQL, DAX, Java

Azure and Cloud Platforms: Azure Databricks, Delta Live Tables, Azure Data Factory / ADF, Azure Synapse Analytics, Microsoft Fabric (Lakehouse, Warehouse, SQL Analytics Endpoints, Notebooks, OneLake), Snowflake, ADLS Gen2, Azure Event Hubs, Azure Stream Analytics, Azure Functions, Azure Cosmos DB (NoSQL), Azure SQL Database, Azure Key Vault, Azure Monitor, Log Analytics

Data Engineering and Modeling: Data Pipelines, ETL/ELT, Semi-Structured Data (JSON), Dimensional Modeling, Fact and Dimension Tables, Slowly Changing Dimensions / SCD, Star Schema, Snowflake Schema, Data Warehouse, Data Lakehouse, Medallion Architecture, Delta Lake, dbt, Apache Spark, Apache Kafka, Apache Airflow, Change Data Capture / CDC

Data Governance and DevOps: Data Quality, Data Governance, Data Observability, Data Lineage, Data Contracts, Unity Catalog, Microsoft Purview (Azure Purview), Alation, Azure DevOps CI/CD (Azure Repos, Azure Pipelines), GitHub Actions, Terraform, GitHub Copilot

PROFESSIONAL EXPERIENCE

Company: Virat Solutions Inc | Role: Senior Azure Data Engineer | Location: Frisco, TX, USA | Jan 2026 - Present

- Improved data pipeline deployment velocity by 40% across 5 Agile engineering teams, as measured by bi-weekly release metrics, by building reusable data products, Spark notebooks, modular Azure data lakehouse architectures using Azure Databricks, ADLS Gen2, Microsoft Fabric Lakehouse (OneLake shortcuts, SQL analytics endpoints), Delta Lake, Azure Data Factory.
- Designed scalable dimensional data models to integrate structured relational tables with semi-structured JSON payloads stored in Azure Cosmos DB.
- Reduced Power BI semantic model query latency from 18 seconds to under 6 seconds, as measured by SQL query profiling, by designing star schema fact and dimension tables with Slowly Changing Dimensions (SCD Type 1 and Type 2) across Azure Synapse Analytics and Snowflake using dbt, stored procedures, and optimized ETL/ELT data pipelines.
- Set up data contracts, schema validation and automated data quality checks with stakeholders, and enforced RBAC, PII masking and lineage through Unity Catalog and Microsoft Purview.
- Engineered data pipelines using Azure Event Hubs, Stream Analytics, Apache Kafka, Python, PySpark to ingest 2M+ hourly connected vehicle IoT telemetry events into Medallion Architecture (Bronze, Silver, Gold) Delta Lake tables.
- Built automated workflow orchestration and incremental Change Data Capture (CDC) pipelines across Azure Data Factory and Azure Functions to handle REST API ingestion, schema drift, and event-driven triggers.
- Reduced monthly Databricks compute spend using cluster auto-scaling, partition pruning, Z-Order clustering and Parquet compaction on multi-terabyte Delta Lake workloads.
- Owned 24/7 production support and maintained 99.9% SLA compliance using Azure Monitor, Log Analytics, automated alerting.
- Standardized infrastructure as code, automated unit testing, documentation, multi-stage YAML release pipelines using Terraform, Git branching workflows, Azure Repos, Azure Pipelines (Azure DevOps CI/CD), GitHub Actions across Dev, UAT, Prod environments.
- Accelerated downstream Artificial Intelligence (AI), Machine Learning (ML) model readiness by building curated ML-ready data layers/automated feature pipelines in Azure Databricks.

Company: Decision Minds Inc | Role: Senior Data Engineer | Location: San Jose, CA, USA | Jan 2025 - Nov 2025

- Scaled unstructured and semi-structured data ingestion to 500K+ daily member interaction records, as measured by pipeline throughput logs, by building Python and Azure Databricks data pipelines with vector embeddings on ADLS Gen2 for AI workloads.
- Resolved recurring nightly Spark compute memory bottlenecks on multi-terabyte jobs by tuning cluster configurations, PySpark transformations, and Spark SQL partitioning in Azure Databricks.
- Developed end-to-end ETL/ELT data pipelines to ingest transactional, CRM, and semi-structured JSON data from SQL Server, Azure Cosmos DB, and third-party REST APIs into Medallion Architecture Delta Lake layers.
- Architected dimensional data modeling layers with fact and dimension tables, star schemas, and Slowly Changing Dimensions (SCD Type 1 and Type 2) using Delta Live Tables (DLT), stored procedures, and dbt for enterprise financial reporting.
- Orchestrated scheduled batch data pipelines and event-driven workflows using Azure Data Factory, Apache Airflow, and Azure Functions to meet strict morning production support SLAs.
- Led Databricks-to-Fabric migration and production cutover, moving legacy analytical datasets into Microsoft Fabric Lakehouse, Fabric Warehouse and OneLake with Fabric Data Pipelines, and re-pointing Power BI Direct Lake semantic models.
- Implemented fine-grained data governance, Microsoft Entra ID access controls, RLS, column-level data lineage, and automated data quality rules using Unity Catalog and Microsoft Purview for financial compliance.
- Built ML-ready feature data pipelines and automated data observability monitors using Python, Azure Monitor, and Azure Log Analytics for reliable downstream Artificial Intelligence (AI), Machine Learning (ML) and analytics models.

- Reduced incremental data refresh latency by 32%, by implementing Change Data Capture (CDC) pipelines from core banking databases into OneLake and ADLS Gen2.
- Eliminated manual workspace promotion errors across dev, test, and prod environments by configuring automated Fabric deployment pipelines and Azure DevOps CI/CD workflows.

Company: Decision Minds Inc | Role: Azure Data Engineer | Location: Bangalore, KA, India | Jun 2022 - Apr 2023

- Reduced enterprise data warehouse reporting latency by 15%, by re-engineering legacy transformations into parameterized Azure Data Factory, ADLS Gen2, and Azure Databricks Medallion Architecture ETL/ELT data pipelines.
- Reduced daily data ingestion delays by building automated Python scripting and SQL validation scripts to detect and resolve upstream schema changes before batch execution.
- Designed dimensional data warehouse schemas in Azure Synapse Analytics using dbt, T-SQL, stored procedures.
- Optimized multi-terabyte warehouse query concurrency across concurrent BI users by tuning Synapse table distribution keys, indexing, and materialized views.
- Implemented dynamic data masking, Azure Key Vault secret management, RBAC to protect sensitive member banking data.
- Built governed analytical datasets, DAX measures in Power BI for reliable self-service reporting across finance and risk teams.
- Cataloged end-to-end data lineage across Microsoft Purview to support regulatory compliance audits and engineering onboarding.

Company: W3Softech India Private Limited | Role: Risk Data Engineer | Location: Hyderabad, TS, India | Mar 2020 - May 2022

- Increased risk transaction data processing throughput by 30% by migrating legacy batch workloads to distributed Apache Spark and Databricks pipelines on cloud object storage.
- Saved 13 hours of manual reconciliation effort per week by automating data quality validation using SQL views, Parquet storage layers, and Python scripts.
- Built multi-stage ETL/ELT pipelines with incremental data loads and dimensional schemas (SCD Type 2) using PySpark, Spark SQL, and Python across structured and semi-structured JSON banking feeds.
- Established production support runbooks, automated validation rules, and Azure DevOps CI/CD release standards to keep risk reporting pipelines compliant with regulatory SLAs.
- Accelerated historical regulatory audit queries from minutes to seconds by integrating Snowflake cloud data warehouse tables with partitioned cloud storage layers.

Company: Genpact India Private Limited | Role: Associate Data Engineer | Location: Hyderabad, TS, India | Jul 2018 - Mar 2020

- Reduced downstream SQL query latency by 20% on high-volume payment event logs by building real-time streaming data ingestion pipelines using Apache Kafka, Python, and Java.
- Improved multi-table SQL join performance by implementing data partitioning, bucketing, and star schema modeling on Hadoop and Apache Spark clusters.
- Migrated historical payment transaction logs into centralized analytical warehouse tables using incremental Spark ETL data pipelines and automated Python data quality scripts.
- Accelerated daily reporting availability by 2 hours by optimizing workflow orchestration, dependency triggers across multi-source production batch jobs.

Company: W3Softech India Private Limited |Role: Jr Data Warehouse Engineer |Location: Hyderabad, TS, India | Jun 2013 - Jul 2015

- Improved reporting query latency by 14%, tuning T-SQL queries, stored procedures, indexing, table partitioning in SQL Server.
- Reduced daily batch processing delays by building automated SSIS ETL data pipelines for high-volume banking transaction feeds.
- Designed relational data warehouse tables and dimensional schemas with SCD Type 1 and Type 2 tracking in SQL Server.
- Executed automated data validation, reconciliation, data quality checks for accurate data delivery in reporting systems.

EDUCATION:	*Master's Degree, Computer Science, Texas A&M University Kingsville, USA, May 2023 - Dec 2024* *M.Tech, Computer Science and Engineering, Lovely Professional University, LPU, India, Aug 2015 - Dec 2017* *Bachelors, Information Technology, PVPSIT, Jawaharlal Nehru Technological University Kakinada, India, Aug 2009 - Apr 2013*
PUBLICATION:	*AVTAR Cyber Shield: A Comprehensive Framework for Machine Learning-Enhanced Cybersecurity in Cyber-Physical Environment*
CERTIFICATIONS:	*Microsoft Certified Fabric Analytics Engineer Associate Verify* *Microsoft Certified Azure Databricks Data Engineer Associate Verify* *AWS Certified Data Engineer Associate Verify*
PROJECTS:	Azure Banking Risk Lakehouse *(ADF, Databricks, Azure OpenAI):* Data contracts, CECL staging, BCBS 239 gate GitHub Credit Union Member 360 on Microsoft Fabric *(PySpark, dbt, MLflow):* Identity resolution, churn AUC 0.88, AML GitHub Real-time Vehicle Telemetry Lakehouse *(Event Hubs, Structured Streaming):* Late data, exactly-once MERGE, 13 DQ rules GitHub