

Vamsi
Data Engineer

Email ID: vamsiv1369@gmail.com

Contact No: +44 776 790 5753

CAREER SUMMARY

Azure Data Engineer with 4 years of experience in architecting and delivering enterprise-scale data solutions across cloud environments. Demonstrated expertise in building resilient data pipelines using **ADF, Databricks, Synapse, and Microsoft Fabric** for large-volume data processing. Strong capability in working with distributed computing frameworks using **Python, PySpark, and SQL** to transform complex datasets into actionable insights. Experienced in implementing modern data platforms including **Lakehouse, DWH, and hybrid integration architectures**. Adept at collaborating with cross-functional teams to translate business requirements into efficient, secure, and scalable data solutions.

PROFESSIONAL SUMMARY

- ✓ Experienced in managing end-to-end stakeholder engagement including requirement gathering, solution design, project coordination, and delivery of enterprise data platforms aligned with objectives and reporting needs.
- ✓ Expertise in Data Engineering ecosystem, including building scalable ETL/ELT pipelines, integrating multiple heterogeneous systems, and enabling reliable batch and real-time data processing using Azure cloud services.
- ✓ Proficient in **Python, PySpark, SQL, Spark SQL**, and performance optimization techniques such as partitioning, indexing, caching, broadcast joins, query tuning, and distributed computing strategies for large datasets.
- ✓ Hands-on experience in designing and implementing **OLTP, OLAP, Data Warehouse, Data Lake, and Delta Lake architectures**, ensuring scalable, structured, and high-performance data storage and processing systems.
- ✓ Extensive experience in **Data Analytics and Reporting solutions**, building dashboards and analytical models using Power BI, Tableau, SSRS, and supporting business decision-making through actionable insights.
- ✓ Strong background in **Data Governance frameworks**, including metadata management, data lineage tracking, security enforcement, auditing, and compliance using Microsoft Purview and Unity Catalog.

TECHNICAL SKILLS

- ❖ **Cloud Platforms:** Azure, AWS, Snowflake
- ❖ **BI & Data Analytics:** Power BI, Tableau, Excel, SSRS, Looker, Qlik
- ❖ **Data Warehousing & Storage:** ADLS, S3, One Lake, MS SQL Server, SAP, CRM, Dataverse, Redshift
- ❖ **ETL / ELT Tools:** SSIS, Azure Data Factory (ADF), Databricks, Fabric, Synapse, Power Query
- ❖ **Programming & Scripting:** Python, PySpark, SQL, Spark SQL, Big Query
- ❖ **Automation & DevOps:** CI/CD, Azure DevOps, Git, Azure Functions, Logic Apps
- ❖ **Others:** Jira, SSMS, Visual Studio, Pandas, Azure Monitor

PROFESSIONAL EXPERIENCE

Client: Recruit First

Role: Data Engineer, Duration: Aug/2024 to Mar/2026

- ✓ Worked closely with stakeholders and technical teams to gather data requirements, understand reporting needs, and design scalable enterprise data architecture solutions aligned with business goals and analytics strategy.
- ✓ Developed and maintained complex ETL/ELT pipelines using **Azure Data Factory** and **Databricks**, enabling ingestion of large-scale structured and semi-structured data from enterprise systems into **ADLS**.
- ✓ Designed and implemented data processing workflows using **Synapse Analytics, Spark notebooks**, and **Microsoft Fabric dataflows**, enabling unified data transformation, orchestration, and analytics across domains.
- ✓ Built optimized transformation logic using **Python, PySpark, and Spark Data Frames**, improving performance of distributed data processing jobs by reducing execution time and efficiently handling large datasets at scale.
- ✓ Implemented **Delta Lake architecture** with **Medallion layers (Bronze, Silver, Gold)** to ensure high-quality, version-controlled, and ACID-compliant data pipelines for enterprise-grade analytics workloads.
- ✓ Integrated external systems using **REST APIs**, handling pagination, authentication mechanisms, retry logic, rate limiting, and network failure scenarios to ensure reliable and fault-tolerant data ingestion processes.
- ✓ Implemented enterprise-level data governance solutions using **Azure Purview** and **Unity Catalog**, ensuring metadata management, data lineage tracking, access control, and compliance with organizational policies.
- ✓ Developed interactive and dynamic dashboards using **Power BI**, enabling business users to visualize KPIs, monitor operational metrics, and support data-driven decision-making processes effectively.

- ✓ Automated deployment pipelines using **Azure DevOps CI/CD, ARM templates, and Databricks Asset Bundles**, ensuring smooth, consistent, and error-free deployment across multiple environments.
- ✓ Provided production support, monitored pipelines using **Azure Monitoring tools**, resolved incidents, and ensured continuous availability and stability of data platforms in production environments.

Client: Blue Dart

Role: Data Engineer, Duration: Oct/2021 to Sep/2022

- ✓ Collaborated with stakeholders to gather data requirements, analyse reporting needs, and design scalable cloud-based data warehouse architecture supporting enterprise analytics and business intelligence solutions.
- ✓ Developed robust **data ingestion pipelines** using **Azure Data Factory and Apache Airflow**, enabling automated extraction of data from relational databases, APIs, and cloud storage systems into centralized data lakes.
- ✓ Designed and optimized large-scale data transformation workflows using **Databricks and PySpark**, ensuring efficient processing of structured and semi-structured datasets with improved performance and reduced latency.
- ✓ Implemented **Synapse Analytics SQL pools (Dedicated and Serverless)** to support high-performance analytical queries, enabling business users to run complex reporting workloads on large datasets.
- ✓ Integrated multiple data sources such as **Cosmos DB, Azure SQL DB, and ADLS**, ensuring seamless data flow and consistency across distributed cloud-based data ecosystems and platforms efficiently.
- ✓ Designed and implemented **Delta Lake** architecture, enabling incremental data processing, ACID transactions, and efficient historical data management for enterprise analytics systems at scale reliably.
- ✓ Developed reusable **Python** and **SQL** scripts for data validation, cleansing, and transformation to ensure high-quality, consistent, and reliable datasets across enterprise data pipelines effectively.
- ✓ Created interactive dashboards using **Power BI**, enabling real-time analytics, KPI tracking, and business reporting for stakeholders across different departments and functions with actionable insights and visibility.
- ✓ Automated CI/CD deployment pipelines using **Azure DevOps** and **Git**, ensuring smooth and reliable delivery of data pipelines across development, testing, and production environments efficiently and consistently.
- ✓ Monitored and optimized data workflows, ensuring high availability, scalability, and performance tuning of enterprise data systems for continuous business operations and stability across platforms.

Client: Razorpay

Role: Data Engineer, Duration: Feb/2021 to Oct/2021

- ✓ Designed and implemented real-time data ingestion pipelines using **Azure Data Factory and Databricks streaming**, enabling near real-time analytics and event-driven data processing across enterprise systems.
- ✓ Built scalable event-driven architectures using **Azure Functions** and **Logic Apps**, supporting automated workflows and real-time data processing for business-critical applications efficiently and reliably.
- ✓ Developed **PySpark streaming** applications, enabling continuous data ingestion, transformation, and processing of high-volume data streams with low latency and high reliability in production environments.
- ✓ Led migration of on-premise data systems to **Azure Data Lake Storage** and **Snowflake**, ensuring secure, scalable, and cost-effective cloud-based data infrastructure across enterprise environments and platforms.
- ✓ Developed reusable **REST API-based** ingestion frameworks, enabling seamless integration of external systems and third-party applications into enterprise data pipelines with flexibility and scalability.
- ✓ Implemented enterprise data governance frameworks using **Unity Catalog** and **Azure Purview**, ensuring compliance, security, and metadata management across all datasets consistently and effectively.
- ✓ Designed and implemented **CI/CD pipelines** using **Azure DevOps**, enabling automated deployment, version control, and environment management for data engineering solutions efficiently and reliably.
- ✓ Built real-time dashboards using **Power BI**, enabling business users to monitor live metrics and make data-driven decisions based on streaming data insights, KPIs, and actionable trends across operations.
- ✓ Optimized **Spark streaming** jobs for performance, scalability, and fault tolerance, ensuring efficient handling of high-throughput data workloads at scale, in real time, and reliably across distributed systems and environments.
- ✓ Provided monitoring and production support for real-time systems, ensuring high availability and resolving issues in streaming data pipelines with minimal downtime and business impact.

Client: Aurobindo Pharma

Role: MI Analyst, Duration: July/2020 to Feb/2021

- ✓ Worked with SAP functional and technical teams to analyse legacy systems, understand data structures, and define migration strategies for transitioning enterprise data to modern cloud-based data platforms.

- ✓ Developed **ETL pipelines using SSIS and Azure Data Factory** to extract data from SAP BW and SAP BO systems, transforming and loading into **Azure Data Lake Storage (ADLS)** for analytics consumption.
- ✓ Built scalable data processing workflows using **Databricks and PySpark**, enabling high-performance transformation of large SAP datasets with optimized Spark job execution and reduced processing time.
- ✓ Designed and implemented **SSAS OLAP models**, enabling multidimensional analysis and supporting fast query performance for business intelligence and reporting requirements across enterprise systems.
- ✓ Integrated **SAP systems** with **Azure SQL DB** and **ADLS**, ensuring seamless data movement between legacy ERP systems and modern cloud-based analytics platforms efficiently, securely, and reliably.
- ✓ Implemented **Delta Lake** architecture for handling incremental data loads, historical tracking, and maintaining consistent data versions for enterprise reporting and analytics at scale, reliably, and efficiently.
- ✓ Developed **Power BI** reports and dashboards to validate migrated data, enabling business users to verify accuracy and gain insights from transformed datasets effectively, efficiently, and consistently across projects.
- ✓ Automated deployment processes using **Git** and **CI/CD pipelines**, ensuring consistent delivery of data pipelines and reducing manual intervention across development, testing, and production environments.
- ✓ Optimized data transformation logic and **SQL queries**, improving performance of migration workflows and ensuring efficient handling of large enterprise datasets across systems, processes, and environments.
- ✓ Provided production support and troubleshooting for migration-related issues, ensuring smooth transition and minimal downtime during enterprise data migration processes across environments, systems, and workloads.

Client: BCD Travel

Role: MI Analyst, Duration: Nov/2019 to July/2020

- ✓ Gathered business requirements from stakeholders and designed scalable **BI reporting solutions**, ensuring alignment with organizational KPIs, operational metrics, and decision-making requirements across departments.
- ✓ Developed and maintained **SSIS ETL packages**, enabling extraction and transformation of data from multiple on-premise systems into centralized data repositories for reporting purposes and analytics effectively.
- ✓ Designed and implemented **SSRS reports** for financial, operational, and business reporting, ensuring accuracy, performance optimization, and alignment with business requirements and standards consistently.
- ✓ Built **SSAS OLAP cubes**, enabling multidimensional analysis and improving query performance for business intelligence and analytical reporting systems across enterprise environments efficiently and effectively.
- ✓ Created interactive dashboards using **Power BI** and **Tableau**, providing real-time insights and visualizations for business users and executive management decision-making processes with clarity.
- ✓ Used **Alteryx workflows** for data blending, transformation, and preparation, supporting advanced analytics use cases and complex reporting requirements efficiently and at scale across enterprise domains.
- ✓ Integrated **SAP data** into BI systems, ensuring seamless flow of enterprise data into reporting and analytical platforms for business consumption and insights generation consistently and reliably.
- ✓ Optimized **SQL queries** and data models to improve performance of reporting systems and reduce data retrieval time significantly across enterprise environments and workloads efficiently and effectively.
- ✓ Implemented data validation and reconciliation processes to ensure consistency and accuracy between source systems and reporting layers across multiple platforms reliably, accurately, and consistently.

EDUCATION

Bachelors in Civil Engineering (Jun/2015 – May/2019)

Masters in MBA Project Management (Sep/2022 – Aug/2024)

CERTIFICATION

Microsoft Certified: Azure Data Engineer Associate (DP-203)

Microsoft Certified: Power BI Data Analyst Associate (PL-300)