

PROFILE SUMMARY

Associate Data Engineer with 3+ years of experience in developing cloud-based data pipelines using Microsoft Azure. Proficient in Azure Data Factory, Azure Databricks, ADLS Gen2, PySpark, and SQL. Experienced in building metadata-driven pipelines to automate data migration and incremental loading from on-premises systems to Azure SQL Database. Skilled in organizing data into bronze, silver, and gold layers for better structure and reuse. Focused on delivering high-quality, validated data to support business reporting and analytics.

FUNCTIONAL AREAS

- ETL (Extract, Transform, Load) Development
- Data Lake Structuring using Medallion Architecture
- Data Quality Checks and Validation
- Curated Data Delivery for Reporting/Analytics Teams
- Metadata-driven and Incremental Data Pipelines
- Data Pipeline Monitoring, Logging, and Optimization
- On-Premises to Cloud (Azure SQL) Migration

TECHNICAL SKILLS

- Programming & Scripting:** Python, PySpark, SQL, Spark SQL
- ETL Tools** – Azure Data Factory (ADF), Azure Databricks
- Big Data & Processing Frameworks:** Apache Spark, Delta Lake, Medallion Architecture
- Metadata-Driven Design:** JSON-based configuration, dynamic datasets, table-driven pipelines, watermark logic
- Data Quality & Monitoring:** Row count validation, checksums, exception handling, logging, alerting
- Cloud Platform & Storage:** Microsoft Azure – ADLS Gen2, Azure SQL Database, Azure Key Vault
- Data Warehousing & Curation:** Gold Layer Curation on Azure Databricks, Partitioned & Validated Datasets
- Version Control & CI/CD:** Azure DevOps, Git

WORK EXPERIENCE

Ensono Technologies LLP, Chennai

- Associate Data Engineer** 12/2022 – Present
 - Engineered a Medallion Lakehouse architecture (Bronze → Silver → Gold) using Azure Databricks and Delta Lake for a mid-sized US-based e-commerce platform.
 - Automated metadata-driven ingestion pipelines in Azure Data Factory to ingest data from SQL Server, REST APIs, and SFTP into ADLS Gen2 (Bronze layer).
 - Implemented watermark-based incremental logic and dynamic pipeline selection using JSON configs and control tables, minimizing hardcoding.
 - Applied schema enforcement, deduplication, and validation in silver layer using PySpark with Delta features (Merge, Time Travel, Schema Evolution).
 - Delivered curated Gold KPIs (e.g., Sales Revenue, Refund %, Customer Retention), enabling a 25% increase in sales forecast accuracy.
 - Validated data quality using row count checks, checksums, and null value thresholds; captured audit logs per run for traceability.
 - Governed data access and schema control using Unity Catalog with RBAC and lineage tracking.
 - Maintained 99% SLA compliance through SLA control tables, Power BI dashboards, and Logic App-based email alerts for failures.
 - Deployed ADF pipelines and Databricks notebooks using Azure DevOps CI/CD pipelines integrated with Change Request workflows and rollback readiness via Delta Time Travel.
 - Tuned Spark performance using broadcast joins, partitioning, and Z-Ordering, reducing job runtime by 30% and compute cost by 15%.
 - Supported QA validation, UAT testing, backfill reprocessing, and production cutovers using L2 support runbooks and operational workflows.

EDUCATION

- B. Tech. in Mechanical Engineering** 06/2018 – 05/2022
 - Hindustan Institute of Technology and Science – 9.21

CERTIFICATIONS

- Azure** – Fundamentals - AZ 900, Data Fundamentals - DP 900 and AI Fundamentals - AI 900.
- AWS** – Cloud Practitioner Foundational - CLF 01.
- Databricks** – Generative AI Fundamentals, Lakehouse Fundamentals.