

Professional Summary

Results-oriented **BI and Data Engineer** with 4 years of hands-on experience in designing, developing, and maintaining scalable data platforms. Skilled in **DWH, One Lake, DataLake** architecture. Adept at gathering business requirements, collaborating with stakeholders, and delivering end-to-end ETL/ELT pipelines. Experienced in **data governance**, GDPR performance optimization, and building analytics-ready datasets for reporting and business intelligence. Passionate about enabling organizations to make data-driven decisions through modern cloud-based solutions.

Career Objective

To leverage my expertise in **Azure, Databricks, ADF**, and **Python/PySpark** to design and implement high-performance, scalable data platforms. I aim to drive data-driven decision-making, ensure data quality, and optimize cloud-based analytics solutions for strategic business outcomes.

Technical Skills

- Cloud Platforms: Azure, AWS, Snowflake
- Data Analytics & BI Tools: Power BI, Tableau, Excel, Looker, Qlik, SSRS
- DWH & Data Storage: ADLS, S3, OneLake, MS SQL, SAP, Dataverse, Redshift, CRM
- ETL/ELT Tools: ADF, SSIS, Databricks, Synapse Analytics, Fabric, Power Query
- Programming & Scripting: Python, PySpark, SQL, Spark SQL, BigQuery
- Automation & DevOps: Azure DevOps, CI/CD, Git, Azure Functions, Logic Apps
- Data Governance: Azure Purview, Unity Catalog, Auditing, Data Lineage
- Others: Jira, SSMS, Visual Studio, Pandas, Azure Monitoring, REST APIs

Education

Bachelor in Business Administration, Lovely Professional University, 2023
MSc in Business Analytics, Coventry University, 2025

Certifications

Microsoft Certified: Azure Data Engineer Associate (DP-203)
Microsoft Certified: Power BI Data Analyst Associate (PL-300)

Work Experience

Organization: Infinite Business Solutions, Duration: May 2025 to Mar 2026

Project Role: Data Engineer

- Partnered with business stakeholders and product owners to gather data requirements, design scalable **ADF pipelines**, and implement end-to-end **Databricks workflows**, ensuring alignment with enterprise data strategy.
- Developed complex **ETL pipelines** using **Azure Databricks** and **ADF**, integrating data from **CRM, Dataverse**, and external APIs into **ADLS**, implementing incremental loads, transformations, and validation for analytics datasets.
- Designed **Synapse Analytics SQL Pools** and **Fabric Dataflows**, creating optimized Spark-based pipelines to process large datasets efficiently while maintaining high performance across multiple production workloads.
- Built and optimized data transformation logic using PySpark and Python, implementing **joins, aggregations, window functions**, and **partitioning** strategies to improve efficiency of structured and semi-structured data.
- Implemented **Delta Lake** architecture in **ADLS**, leveraging **schema evolution, time travel, ACID compliance**, and **efficient partitioned storage** for highly reliable and scalable enterprise data workflows.
- Integrated REST APIs with proper pagination handling, robust error handling, and automated retry mechanisms to manage network failures while ensuring continuous data ingestion into enterprise pipelines.
- Established data governance using **Azure Purview** and **Unity Catalog**, implementing data lineage, auditing, classification, and role-based access control to meet compliance and security standards.
- **Developed Power BI dashboards** to effectively track KPIs, trends, and operational metrics, enabling stakeholders to make informed, **data-driven decisions** across multiple business units and departments.

- Implemented CI/CD pipelines using **Azure DevOps**, ARM templates, and **Databricks Asset Bundles**, automating deployments and maintaining version control across development, testing, and production environments.
- Monitored and maintained production pipelines using **Azure Monitoring**, efficiently resolving incidents, optimizing performance, and ensuring 24/7 availability and reliability of enterprise data platforms.

Client: ITC Hotels, **Employer:** ITC (Contract), **Duration:** May 2023 to Aug 2024

Project Role: Associate Data Engineer

- Collaborated with business users and analysts to design and implement data pipelines for ingesting transactional and financial data from legacy, web applications and **Cosmos DB** into **ADLS** for centralized analytics.
- Developed ETL/ELT workflows using **ADF** and **Databricks**, processing large volumes of structured and semi-structured data while ensuring data quality, consistency, and schema enforcement across environments.
- Designed and implemented **Delta Lake** solutions with time travel, ACID transactions, and schema evolution, enabling highly reliable batch and streaming data processing for enterprise analytics.
- Created optimized **data models** in Azure SQL DB, Synapse, and Redshift, significantly improving query performance, efficiency, and scalability for analytics workloads and reporting dashboards.
- Automated complex **workflow orchestration** using Airflow and ADF triggers, enabling highly efficient scheduling, dependency handling, and proactive monitoring of critical enterprise data pipelines.
- Developed interactive and visually intuitive **dashboards using Power BI**, allowing stakeholders to effectively track KPIs, operational metrics, and evolving business trends with near real-time data updates.
- Implemented robust CI/CD pipelines using **Azure DevOps** and Git, standardizing deployment processes and ensuring consistent code promotion across development, QA, and production environments.
- Applied enterprise-grade data governance practices using **Unity Catalog**, ensuring proper access control, metadata management, and compliance with internal policies and industry regulations.
- Optimized **ETL workflows** and SQL queries using partitioning, indexing, and caching, dramatically reducing runtime while significantly improving overall pipeline efficiency and performance.
- Provided production support by monitoring pipelines and troubleshooting issues using **Azure Monitoring**, ensuring high availability, reliability, and data accuracy for business-critical operations.

Client: Petronas, **Employer:** ITC (Contract), **Duration:** Jun 2022 to May 2023

Project Role: MSBI Developer

- Gathered requirements from business and IT stakeholders to design end-to-end **ETL** solutions, migrating legacy **SAP BW** and relational databases into **Azure Data Lake** and **SQL Data Warehouses**.
- Developed ETL pipelines using **SSIS**, **ADF**, and **Databricks**, transforming and consolidating large volumes of structured and semi-structured data into centralized repositories for analytics and reporting.
- Built scalable **transformation logic using PySpark and Python**, implementing highly efficient data aggregation, complex joins, validations, and robust error-handling mechanisms to ensure high-quality outputs.
- Designed **Azure SQL DB** and SSAS data models to fully support complex OLAP workloads, enabling comprehensive multidimensional analytics for finance, operations, and clinical reporting requirements.
- Implemented **Delta Lake** in ADLS, carefully managing schema evolution, partitioning strategies, and versioning to support robust historical data tracking, scalable storage, and reliable enterprise-wide analytics.
- Developed interactive and visually intuitive **Power BI dashboards** and reports to effectively visualize operational, financial, and clinical metrics, enabling actionable insights for senior management and operational teams.
- Managed source code using **Git**, enforcing proper branching strategies, version control, and collaborative development practices across multiple cross-functional teams for streamlined and efficient project delivery.
- Optimized ETL workflows and SQL queries using indexing, partitioning, and advanced query tuning techniques, significantly reducing execution times while improving overall system performance and reliability.
- Integrated **SAP and external systems** into modern cloud platforms, ensuring seamless, reliable, and highly consistent availability of business-critical data for downstream analytics, reporting, and operational insights.