Shamsheer

Data Engineer | BI Developer | Azure Data Engineer | Fabric Data Engineer | Analytics Engineer

Contact: +44 7436 189570| **Email**: Shamsheer1893@outlook.com | **LinkedIn:** www.linkedin.com/in/shamsheer1893

Notice Period: Available immediately

PROFESSIONAL SUMMARY

Data and BI Engineer with **5 years** of experience across **financial, insurance,** and **health** sector services, delivering TB-scale data platforms for FTSE-listed clients including **Aviva plc, Nationwide Building Society, AXA Health** and **MetLife** processing **500GB** of financial data daily. My focus is designing Medallion Architecture, Data Lakehouse, and Metadata-Driven pipelines using **Databricks, Microsoft Fabric, Azure Data Factory, PySpark, Python, SQL, Delta Synapse Analytics and Power BI** following Kimball Methodology for dimensional warehouse design**.** Delivered measurable impact across all engagements: **45% faster** analytics, **60% reduction** in manual testing, **£58,000+** in annual infrastructure savings, **40% faster** data delivery**,** and **98%** pipeline success rate. **MSc** in IT & Management from Sheffield Hallam University and the **Microsoft Certified: Fabric Data Engineer Associate DP-700** and **Power BI Data Analyst Associate PL-300.**

KEY ACHIEVEMENTS

- Reduced ETL runtime by 40% through PySpark optimisation, saving **100+** compute hours a month (Aviva)
- Cut compute costs by 50% through Databricks resource allocation, roughly **£58,000** saved a year (Nationwide)
- Improved analytics speed by 45% across **20+** enterprise workflows, enabling real-time insight at TB-scale (Aviva)
- Achieved **98%** pipeline success rate and 40% faster data delivery across all integrated systems (Nationwide)
- Reduced query latency by **50%**+ across TB scale SAP BW and financial data via Databricks and Synapse (AXA Health)
- Removed **60%** of manual testing effort, recovering **250** engineering hours per year (Aviva Insurance Limited)

TECHNICAL SKILLS

Azure Services: Azure Data Factory (ADF), Azure Databricks, Microsoft Fabric, Azure SQL Server, Azure ADLS Gen2, Azure Functions, Azure Logic Apps, Azure Synapse Analytics, Azure Event Hubs, Azure Stream Analytics, Azure Key Vault.
Cloud Platforms: Snowflake, AWS S3, Amazon Redshift, Google BigQuery.
Programming: Python, PySpark, dbt, Advanced SQL (window functions, CTEs, query tuning, performance optimisation) T-SQL, (pandas, Excel, API ingestion, automation, testing), DAX, MDX, Power Query (M)
Storage & Databases: Delta Lake, Cosmos DB, SQL Pools, OneLake, Azure SQL Database, SQL Server, MySQL, PostgreSQL.
Enterprise Systems: SAP BW, SAP BO, Microsoft Dataverse, Microsoft 365, CRM.
ETL & Orchestration: Apache Airflow, SSIS, Change Data Capture (CDC), ELT Pipelines, Apache Kafka, ODI, Alteryx.
Data Visualisation & BI: Power BI, Tableau, Qlik, Looker, dbt semantic models, SSAS and SSRS, Row Level Security
Architecture & Modelling: Medallion Architecture, Data Lakehouse, Kimball dimensional modelling, star schemas, SCD Type 1 and 2, Metadata-Driven Architecture, Bronze / Silver / Gold Layers, OLTP (Normalised Schema Design) and OLAP
Governance & Compliance: FCA and PRA Reporting, Solvency II, UK GDPR, Data Protection Act 2018, Information Governance, Data Quality, Reconciliation, Unity Catalog, Microsoft Purview, RBAC, IAM, Metadata and Data Lineage
Domain & Analytics: Claims and Policy Data, Underwriting and Pricing Analytics, Loss Ratio and Premium Analysis, Provider Network Analytics, Actuarial and Finance Reporting, Customer Retention and Segmentation, Executive MI
DevOps & Delivery: Azure DevOps, Git, GitHub, CI/CD, automated testing and deployment, Agile, Scrum, JIRA, Source-to-Target Mapping, Technical Documentation, Stakeholder Engagement.

PROFESSIONAL EXPERIENCE

Data & BI Engineer

SunTech | Aviva plc ***Nov 2024 to Present***

- Collaborated directly with **Aviva's stakeholders** and product managers and Claims, Underwriting, Actuarial and Finance stakeholders in an **Agile** team, across Aviva's governed general insurance data estate.
- Lead a **data modernisation** project using **Microsoft Fabric, Databricks** and **Synapse**, designing reusable data lake layers and automated pipelines to improve analytics speed and accuracy by **45%** across **20+** enterprise workflows.
- Architect Aviva's Data Lakehouse on **Medallion Architecture**, implementing **Bronze, Silver, and Gold** layers in **ADLS Gen2** and Delta Lake, ensuring high data availability and query optimisation for downstream analytics.
- Used **PySpark** for distributed data processing and performance optimisation, achieving a **40%** reduction in ETL runtime and saved **100+ compute hours** monthly through partitioning, caching, and cluster tuning in **Databricks**.
- Designed **Metadata-Driven pipelines** using **Delta Lake, Purview, ADLS, and ADF**, enabling parameterized, scalable ETL architecture that increased reusability across 20+ workflows.
- Built **dbt models** on Databricks and Microsoft Fabric that turn raw data into clean Gold-layer tables, with **99%** data accuracy and **60%** fewer data issues, and giving the business a trusted base for analytics and **future AI work.**
- Own the **semantic layer** across Databricks and Microsoft Fabric**,** using **Kimball dimensional** design to agree a one definition for every key metric, so Claims, Underwriting, Finance and Actuarial all report the same numbers.
- Developed **Python based automation frameworks** to perform **ETL validation**, metadata checks and schema drift detection, removing **60%** of manual testing and giving the team back around **250 engineering hours** a year.
- Used **AI coding assistants** [GitHub Copilot, Claude] to draft PySpark, SQL and unit tests faster, always reviewing and testing the output before it reached production.
- Orchestrate **20+** production ETL workflows in **Apache Airflow**, authoring Python **DAGs** with dependency management, retry logic and SLA-based alerting across Databricks, ADF and Delta Lake.
- Built **Power BI semantic models** over the Fabric Lakehouse and Synapse, using **DAX, Power Query** and **row-level security**, tuned in Tabular Editor and DAX Studio, reducing manual reporting effort by **50%.**
- Enforced **data governance** with **IAM roles, RBAC** and **Unity Catalog** for centralised metadata, and automate releases through **CI/CD** in **Azure DevOps**.

Data & BI Engineer

Capgemini | Nationwide Building Society ***Oct 2021 to Nov 2022***.

- Tuned **PySpark** batch and streaming jobs on **CosmosDB** datasets, achieving **50%** faster performance, a **98%** pipeline success rate and **40%** quicker delivery**,** saving roughly **£58,000 a year** in compute costs
- Led **end-to-end project implementation** using **Azure Databricks, Azure Synapse, and Unity Catalog**, integrating multiple data sources and standardizing data ingestion pipelines for enterprise analytics.
- Designed **data warehouse** and **data lake structures** for supporting tiered data zones **bronze, silver, gold** following Medallion Architecture, improving how data was organised, scaled and access for analysis.
- **Migrated** Nationwide's legacy **SQL**-based **ETL** workflows into modern **ADF** and **Databricks pipelines**, reducing manual intervention and improving data flow reliability, reusability, and consistency overall**.**
- Built ELT into the **Enterprise Data Warehouse (EDW)** on **Azure SQL**, transforming **SAP** and **SQL Server** data with **dbt**, **T-SQL** stored procedures and views, and improving data completeness by **30%.**
- Built **Metadata-Driven orchestration frameworks** using **ADF, Delta Lake, and Purview**, enabling fully dynamic pipelines that eliminated hard-coded dependencies and improved maintainability.
- Delivered end to end **Power BI solutions** covering ingestion, semantic modelling with reusable **DAX** measures and **Power Query** transformations, dashboards, apps and paginated reports, alongside **CXO-level Power BI** and **Tableau** dashboards with drill-down, reducing manual report production **48%**.
- Developed and maintained **SSRS** parameterised reports, subscriptions and **SSAS models**, and migrated legacy Crystal Reports content into Power BI and **SSRS**, retiring 48 duplicated reports and cutting maintenance effort by **45%.**
- Developed and deployed **Python-based data validation and reconciliation scripts**, automating quality checks and ensuring consistent data integrity throughout the ETL lifecycle.
- Enforced **data governance** with **IAM, RBAC** and **Microsoft Purview** lineage with releases via **CI/CD** in **Azure DevOps**.

Data & BI Engineer

Capgemini | AXA Health Insurance *Nov 2020 – Oct 2021*

- Worked with **AXA Health Insurance's stakeholders** and SMEs to unpick a complex **SAP BW migration,** mapping the data into **Azure Databricks**, while keeping it true to the agreed business rules.
- Migrated **40+** SAP BW **AXA Health Insurance's** data objects into **Delta Lake**, building checks to compare row counts and key financial figures against the source system, so reports matched exactly after go-live.
- Delivered a **modern data stack migration** integrating **Databricks, Azure Synapse,** and **Delta Lake** with **SAP BW** data, streamlining analytics workflows and reducing query latency by more than **50%** efficiently**.**
- Optimised **PySpark ETL jobs** at AXA Health Insurance to handle SAP extracts with high cardinality, applying caching, partitioning, and broadcasting techniques to enhance overall job performance by **42% consistently**.
- Designed and implemented a **Data Lakehouse architecture** for AXA Health Insurance incorporating **Bronze, Silver and Gold** integrating SAP BW with **Azure SQL**, enabling unified analytics and improving enterprise data accuracy.
- Modelled **claims**, **policies** and **membership data** into fact and dimension tables, supporting reporting on provider networks and treatment costs for operations and finance teams.
- Built the transformation layer of an **Azure SQL** Enterprise Data Warehouse with **T-SQL** stored procedures, views and star schemas, improving query performance by **35%.**
- Developed **Power BI dashboards,** paginated reports and **semantic models** with **SSRS** aligned to **SAP reporting**, with **DAX measures** and **row level security** to protect sensitive data, reducing manual MI production by **30%**.
- Deployed **CI/CD pipelines** for **ADF and Databricks** via **Azure DevOps**, standardising deployments and improving rollback efficiency, Reduced **ADF pipeline runtime by 40%** through performance tuning and **job optimisation.**
- Built **Metadata-Driven Architecture** ingestion pipelines on using Delta Lake, ADF, and Purview integration, keeping them scalable, reusable, and easy to track as schemas evolved with **ServiceNow** and **Azure Monitor** for monitoring.
- Established robust **data governance** controls using **IAM** and **RBAC** roles to segregate user access effectively, ensuring strict compliance, traceability, and adherence to enterprise-level data standards.

Business Intelligence Developer

Capgemini | MetLife *Sept 2019 - Nov 2020*

- Engaged with **MetLife's business users and stakeholders** to collect BI requirements, define measurable metrics, and standardize dashboards across **operational and financial domains** for unified visibility.
- Delivered new **reporting project** for MetLife using **Power BI and Tableau**, integrating multi-source enterprise data into centralised analytics layers for improved data accessibility.
- Developed **SSIS ETL workflows** for MetLife covering data cleansing, aggregation, and staging operations, improving reporting data quality and consistency by nearly **30% organisation-wide**.
- Created **Python-based automation scripts** for report refresh validation, automated logging, and exception handling, significantly reducing manual oversight and system downtime efficiently.
- Designed **Metadata-Driven Architecture SSIS pipelines** for MetLife to handle dynamic source mapping and schema variations, improving maintainability, scalability, and version control accuracy.
- Designed **Power BI and Tableau dashboards** for MetLife's CXO-level executive reporting, offering interactive insights, drill-down Visualisation, and strategic business intelligence accessibility.
- Built **SSAS** data warehouse cubes following **Kimball Methodology**, applying **Star Schema, Snowflake Schemas** with **SCD Type 1** and **Type 2** for historical reporting, trend monitoring, and advanced **KPI tracking**

EDUCATION

- MSc in IT & Management Sheffield Hallam University, Sheffield, United Kingdom.
- Bachelor of Technology in Electrical & Electronics Engineering, JNTU-A, India.

CERTIFICATION

- ✦ Microsoft Certified: Fabric Data Engineer Associate DP-700
- ✦ Power BI Data Analyst Associate PL-300