Neeraj
BI & Data Engineer
Email ID: neerajk210601@gmail.com
https://www.linkedin.com/in/neeraj-k-86406a417 **Phone: +44 7721952628**

Professional Summary

Results-oriented Azure Data Engineer with 4+ years of experience delivering end-to-end data solutions across cloud platforms. Skilled in building scalable **ETL pipelines**, designing **Data Lakes, Delta Lakes, and DWH architectures**, and developing actionable **BI dashboards**. Proven expertise in **Python, PySpark, SQL**, and **Azure cloud services**, enabling robust data processing, analytics, and governance. Adept at collaborating with stakeholders to understand business requirements, optimize workflows, and implement enterprise-grade data solutions. Strong focus on **data quality, governance, and performance optimization** in large-scale data environments.

Key Highlights:

- Proven expertise in designing and implementing end-to-end **Azure data solutions**, including **ADF, Databricks, Synapse Analytics, Delta Lake**, and **Data Lake architectures**, ensuring scalable and reliable data processing.
- Strong programming and data engineering skills using **Python, PySpark, Spark SQL, and SQL**, applying performance optimization, transformations, and complex aggregations for multi-terabyte datasets efficiently.
- Experienced in **data governance, security, and compliance** using **Purview, Unity Catalog, and metadata management**, ensuring traceability, audit readiness, and controlled access across enterprise datasets.
- Skilled in building **interactive BI dashboards and analytics solutions** with **Power BI, Tableau, and Excel**, enabling stakeholders to monitor KPIs, track trends, and make informed business decisions.

Technical Skills

- Azure
- Synapse, Databricks, Data Factory, CosmosDB, ADLS, Azure SQL DB, Unity Catalog.
- ADF
- Databricks
- SSIS
- SSAS
- PySpark
- Power BI
- Tableau
- Delta Lake
- Cosmos DB
- RedShift
- Azure SQL DB
- Python
- SQL
- PySpark
- Unity Catalog
- Airflow
- DevOps
- Git
- Agile Methodologies
- Cross-functional Teamwork
- Client Relations

Professional Experience

Client: L1 Capital (Contract)

Role: Data Engineer | Duration: Jan2025 to Apr2026

- Collaborated with stakeholders and technical teams to gather requirements, design data ingestion pipelines, and implement scalable **ADF**, **Databricks**, and **Delta Lake** solutions, enabling reliable enterprise reporting.
- Developed robust **ETL pipelines** using **Databricks** and **PySpark**, handling complex data transformations, joins, aggregations, incremental loads, and large-scale data processing across structured and unstructured datasets.
- Optimized large-scale data processing workflows using **Spark pools and Fabric dataflows**, implementing caching, partitioning, and tuning strategies to significantly reduce runtime while improving resource utilization.
- Designed and maintained **Data Lake** and **Delta Lake** architectures, ensuring schema enforcement, incremental updates, time-travel, and versioned historical data management for downstream analytics and reporting.
- Integrated external systems via **REST APIs**, implementing robust error handling, retries, pagination, and logging mechanisms to ensure continuous ingestion of business-critical data from multiple sources.
- Applied enterprise **data governance policies** using Purview and Unity Catalog, enabling metadata management, lineage tracking, access control, and compliance adherence across all enterprise datasets.

- Built interactive dashboards using **Power BI**, translating large datasets into meaningful KPIs, metrics, visualizations, and actionable insights to drive business decision-making and operational efficiency.
- Automated deployment processes using **Azure DevOps, ARM templates, and Databricks Asset Bundles**, ensuring consistent, reliable, and repeatable CI/CD pipelines for all development and production workflows.
- Monitored production pipelines using **Azure Monitoring and logging tools**, proactively identifying performance bottlenecks, resolving incidents, and optimizing pipeline efficiency for high-volume data processing.
- Created comprehensive documentation for pipeline architecture, technical workflows, and operational processes, ensuring knowledge sharing, maintainability, and compliance with enterprise standards.

Client: Max Healthcare Employer: Truefy IT Consulting

Role: Data Developer | Duration: Jun2023 - Apr2024

- Engaged with business and technical stakeholders to capture requirements, design multi-source data ingestion pipelines, and implement scalable **ADF**, **Databricks**, and **Synapse Analytics** solutions.
- Developed high-performance ETL/ELT pipelines to process data from **CosmosDB, S3, Redshift, and CRM systems**, transforming and loading it efficiently into Azure-based data storage for analytics and dashboards.
- Built distributed data processing workflows using **PySpark and Spark SQL**, optimizing joins, aggregations, and transformations for multi-terabyte datasets to improve analytics performance and reporting speed.
- Designed and implemented **Delta Lake** environments with ACID compliance, incremental load capabilities, and historical data tracking, ensuring reliability, traceability, and accurate analytical processing.
- Tuned SQL queries and optimized database performance on **Azure SQL DB and Redshift**, reducing query runtime, improving data retrieval speed, and enhancing overall reporting efficiency.
- Implemented workflow orchestration using **Airflow and ADF**, managing pipeline dependencies, error notifications, retries, and logging to ensure end-to-end reliability of production data workflows.
- Created detailed **Power BI** dashboards, enabling stakeholders to monitor KPIs, track trends, analyze performance, and make informed decisions by translating raw datasets into actionable visual insights.
- Applied enterprise data governance and security practices using **Unity Catalog**, ensuring proper access control, metadata management, auditing, and regulatory compliance across all pipelines and datasets.
- Developed **CI/CD pipelines** using **Azure DevOps**, automating deployment of ETL processes, dashboards, and data pipelines with version control, reducing errors and improving overall deployment speed.
- Provided production support and monitoring for data pipelines, proactively identifying bottlenecks, resolving incidents, and continuously optimizing workflows to ensure reliable, high-performance analytics.

Client: Delivery Logistics Employer: Truefy IT Consulting

Role: Data Analyst | Duration: Sep2022 – Jun2023

- Collaborated with business analysts and IT teams to gather requirements, design ETL workflows, and implement **SSIS, ADF, and Databricks solutions** for enterprise data integration across multiple systems.
- Developed ETL pipelines integrating data from SAP BW and SAP BO systems, performing transformations, aggregations, and cleansing to ensure high-quality, reliable data for reporting and analytics.
- Built OLAP cubes and multidimensional models using **SSAS**, enabling fast analytical queries, interactive reporting, and advanced business intelligence for operational and strategic decision-making.
- Implemented **Python and PySpark scripts** to perform large-scale data transformations, join operations, and aggregations, improving pipeline efficiency and reducing execution time for high-volume datasets.
- Designed and maintained **Delta Lake and Data Lake architectures**, enabling incremental processing, time-travel, schema enforcement, and historical data versioning for scalable analytics and reporting environments.
- Optimized **SQL** queries on **Azure SQL DB**, significantly improving performance, reducing execution time, and enabling faster, more efficient access to analytical datasets for business stakeholders.
- Created interactive **Power BI** dashboards, enabling business users to monitor KPIs, explore trends, visualize insights, and gain actionable intelligence from complex enterprise datasets in real time.
- Applied **Git** version control for ETL scripts, notebooks, and pipeline configurations, ensuring collaborative development, code traceability, seamless integration, and adherence to organizational coding standards.
- Performed rigorous data validation, testing, and debugging of pipelines, ensuring data accuracy, completeness, consistency, integrity, and full compliance with internal and external reporting standards.

- Provided production support for pipelines and reporting solutions, monitoring performance, resolving incidents promptly, and implementing optimization measures for high availability and reliability.

Client: Axis Bank Employer: Truefy IT Consulting

Role: BI Developer | Duration: Mar2022 – Sep2022

- Gathered reporting and analytics requirements from business teams and implemented **SSIS ETL workflows**, transforming raw datasets into structured formats for downstream analytics and operational reporting.
- Developed detailed reports using **SSRS**, providing operational metrics, KPI tracking, interactive dashboards, and actionable insights to business stakeholders for timely, accurate, and data-driven decision-making.
- Designed and optimized **OLAP cubes** using **SSAS**, enabling fast, multidimensional queries, supporting analytical reporting, generating insights, and improving performance for various business processes and departments.
- Created visually interactive dashboards using **Power BI** and **Tableau**, providing actionable insights, trends, KPI visualizations, drill-down analysis, and real-time analytics for business leadership and operational teams.
- Performed comprehensive data cleansing, transformation, and enrichment using **Alteryx**, ensuring datasets were accurate, standardized, complete, reliable, and ready for advanced analytics and reporting purposes.
- Integrated data from **SAP ERP** systems, improving consistency, reliability, timeliness, accuracy, usability, and accessibility of reporting data used across different business units and processes.
- Optimized ETL jobs and **SQL queries** for better runtime performance, reliability, efficiency, scalability, and reduced pipeline execution times, improving overall data availability for reporting.
- Collaborated closely with cross-functional teams including business analysts, IT, and management to deliver high-quality, scalable, effective data solutions aligned with business objectives and strategic goals.
- Applied rigorous data validation, testing, and auditing practices to ensure high-quality, accurate, consistent, fully compliant, and reliable datasets, enabling trusted analytics and reporting for business stakeholders.
- Provided ongoing production support, troubleshooting pipeline issues, implementing effective fixes, and ensuring seamless availability and reliability of analytics and reporting workflows to stakeholders.

Educational Qualifications

Masters in Big Data and Data Science Technology – Northumbria University, London, UK

Bachelors in Electronics Communication Engineering – JNTU Hyderabad, India

Certifications

Microsoft Certified: Azure Data Engineer Associate (DP-203)
Microsoft Certified: Power BI Data Analyst Associate (PL-300)
Microsoft Certified: Fabric Data Analytics Associate (DP-600)
Microsoft Certified: Fabric Data Engineer Associate (DP-700)
Microsoft Certified: Azure Databricks Data Engineer Associate (DP-750)
365-Data Science: Machine Learning in Python (CC-C0B85729C0)
365-Data Science: SQL + Tableau + Python (CC-460AC4F7D9)
365-Data Science: Power BI (CC-DD482A05BB)
365-Data Science: Introduction to R Programming (CC-C08CA187BD)
365-Data Science: Deep Learning with TensorFlow 2(CC-8B3B0C681E)
Career Labs: Junior Software Developer (CERT_1473686)
Hexa Software: Intern Data Science (HXA20SE030213)