

Neha Singh

New Delhi, India, • neha97official@gmail.com • +91 9810935304 [Neha Singh](#) | [LinkedIn](#)

Data Engineer | 4+ years of Experience

PROFESSIONAL EXPERIENCE

AHEAD

Senior Analyst

Gurugram, India

2024-Present

- Engineered 10+ end-to-end ETL pipelines in Azure Data Factory (ADF) to ingest data from Azure SQL and Blob Storage, improving delivery speed by 35%.
- Streamlined large-scale data transformations using Databricks (PySpark), reducing job runtime by 40% on datasets exceeding 5 million records.
- Designed and configured 6+ reusable ADF pipelines across environments, decreasing manual adjustments by 80%.
- Secured pipelines using Azure Key Vault, reducing credential exposure risk to zero.
- Collaborated with analytics, product, and engineering teams to define 8+ data models and build 6 dashboards that support weekly stakeholder decisions.
- Integrated ADF with Databricks notebooks to enable scalable processing and incremental data loads, reducing daily pipeline runtime by 35% and ensuring timely data refresh.
- Conducted data quality checks, logging, and monitoring to ensure high pipeline reliability and traceability.
- Developed complex SQL queries, views, and stored procedures in Azure SQL Database to support analytical workflows.
- Built automated monitoring for zero-record detection and pipeline failures, reducing manual checks by 90%.
- Migrated 6 pipelines to Unity Catalog, modifying 18+ table paths, 12 notebooks, and 4 cluster configs.
- Conducted row-level validation on 250,000+ records across environments to verify data accuracy.

EXL

Executive

Greater Noida, India

2023-2024

- Deployed and managed 10+ Azure Virtual Machines, storage accounts, and VNets using Azure Portal and ARM templates, achieving a 25% improvement in deployment consistency across environments.
- Developed SQL scripts to generate internal reports, perform ITSM trend analysis, and monitor cloud resource utilization reducing analysis time by 30% and supporting cost-optimization initiatives.
- Built and tested foundational data pipelines using SQL, Python, and PySpark on Azure Databricks, enabling automated data ingestion and transformation workflows.
- Validated data transformations and resolved recurring pipeline errors in collaboration with senior data engineers, improving data accuracy and reliability.

CADENTIAL TECHNOLOGIES Pvt Ltd.

Analyst

Bangalore, India

2021-2023

- Developed and optimized SQL queries to extract, filter, and analyze large datasets, improving internal reporting efficiency by 30%.
- Leveraged AWS services, including S3 and EC2, to manage and organize datasets, enhancing data accessibility and storage reliability.
- Maintained and updated data logs and reports in collaboration with senior analysts, ensuring 100% data accuracy across business operations.
- Cleaned, validated, and standardized multi-source data to support analytical dashboards and trend analysis.

PROJECT

Project: Scalable ETL Pipeline using Azure & Databricks

Tools: Azure Data Factory, Azure Databricks, PySpark, SQL, Azure Blob Storage

- Engineered a **scalable ETL pipeline** to process over **100,000 CSV records** from **Azure Blob Storage**, using **PySpark** in **Azure Databricks** to transform and load data into **Delta Lake tables** with schema management.
- Orchestrated the workflow using **Azure Data Factory**, implementing **parameterized pipelines**, **scheduled triggers**, and **fail-safe mechanisms** to ensure reliable, production-grade data ingestion.
- Embedded **custom logging** and **Azure alerts** to track pipeline health, identify failures, and monitor performance in real time—resulting in **50% faster root cause detection**.
- Applied retry logic to Spark jobs to reduce data failure rate by **40%** and consistently maintain **data refresh cycles under 5 minutes**.

SKILLS

- **Data Engineering & ETL** - Azure Data Factory (ADF), Databricks (PySpark, Delta Lake), Airflow (DAGs), Snowflake, ETL Design, Data Pipeline Automation
- **Cloud & Storage** - Azure Blob, Data Lake, Synapse, Key Vault • AWS S3 • GCP (BigQuery Learning).
- **Data Modeling & Governance** - Star & Snowflake Schemas, Unity Catalog, Data Lineage, Metadata Management.
- **Streaming & Batch Processing** - Apache Kafka, Event Hub, Spark Structured Streaming
- **Monitoring & Quality** - SQL Performance Tuning, Data Validation Frameworks, Azure Monitor, Airflow Alerts.
- **Programming & Automation** - Python, PySpark, Pandas, Numpy, Matplotlib, MySQL.
- **Visualization & BI** - Power BI, Databricks SQL, Snowflake, Excel Dashboards.
- **Version Control & CI/CD** - Git, GitHub.

EDUCATION

Bachelor of Arts
Ramjas College, University of Delhi

2016 – 2019

ADDITIONAL INFORMATION

- **Languages:** English, Hindi (Native)
- **Certifications:**
 - [Microsoft Certified: Azure Data Engineer Associate \(DP-203\)](#)
 - [Microsoft Certified: Azure Fundamentals \(AZ-900\)](#)