

KUSHAL SHAH

DATA ENGINEER - Databricks, Azure, & Pyspark

✉ kushal28shah@gmail.com ☎ +1 (289)-541-3283 📍 [Brampton](#) **in** [LinkedIn](#)

SKILLS

- **Programming & Querying:** Python, PySpark, SQL, data processing, anomaly detection and API integration
- **Data Engineering:** ETL development, pipeline automation, schema enforcement, SCD type 2 and medallion
- **Cloud Platforms:** Azure Data Factory, Databricks, Data Lake Gen2, Delta Lake, Unity Catalog and warehouse
- **Governance :** Metadata handling, access control, schema evolution, ACID transactions and delta time travel
- **Orchestration & Deployment:** Databricks jobs, workflow automation, dependency management and CI/CD
- **Analytics & Optimization:** Power BI dashboards, geospatial visualization, query tuning, KPI optimization

WORK EXPERIENCE

Data Engineer

Arihant eSolutions

September 2024 – Present

Brampton, ON

- Developed ETL/ELT pipelines using Azure Data Factory and Databricks, processing 500GB+ data daily, achieving 25% storage cost reduction through optimized compression.
- Engineered SQL queries in Databricks for data transformation, reducing query execution time by 40% through indexing and query optimization techniques.
- Designed Power BI dashboards visualizing 15+ KPIs across departments, enabling real-time monitoring and collaborated with cross-functional teams.
- Teamed with DevOps and BI teams implementing CI/CD pipelines using Azure DevOps, automating deployments and reducing manual effort by 50% across 10+ projects.
- Established data quality frameworks using Unity Catalog, implementing validation rules that reduced data errors by 35% across 20+ data sources.
- Orchestrated data ingestion workflows processing 2TB+ monthly data from multiple sources using Azure Data Factory and Databricks Delta Lake architecture.
- Optimized Databricks cluster configurations, reducing compute costs by 30% while maintaining performance standards for 50+ concurrent workloads.
- Constructed dimensional data models using star schema design patterns, supporting analytics queries with 60% faster response times for end users.
- Automated data pipeline monitoring using Azure Monitor and Databricks alerts, achieving 99.5% uptime and reducing incident response time by 45%.
- Migrated legacy data systems to Azure cloud infrastructure, completing 8 database migrations with zero data loss and 20% improvement in latency.

PROJECT EXPERIENCE

City Micro-Climate Anomaly & Alert Lakehouse

Azure Databricks, Delta Lake, PySpark, Unity Catalog, Power BI, Python

- Architected 4-layer Medallion data platform on Azure Data Lake Gen2 using Delta Lake, enabling ACID transactions, schema evolution, and time travel across 6 GTA cities.
- Orchestrated real-time API ingestion from Open-Meteo Weather and Air Quality sources at 15-minute intervals, processing 17,280 records daily with rule-based simulation.
- Partnered with team members engineering Bronze-to-Gold transformations using PySpark, implementing schema enforcement and city-based partitioning for 10x query gain.
- Implemented SCD Type 2 in `dim_city` using surrogate keys and temporal tracking, designed star schema fact table supporting 150M+ annual rows with incremental Delta checkpointing.
- Configured Unity Catalog for metadata governance, automated ETL workflows using Databricks Jobs with dependency graphs, error handling, and Slack notification integration.

EDUCATION

Computer Engineer Technician

Sheridan College, Brampton, ON, Canada

August 2023 – September 2025

CERTIFICATIONS

- **Databricks Certified Data Engineer – Databricks, November 2025**