KUSH PATEL

Data Engineer | SQL · PySpark · Microsoft Fabric · Azure Data Factory · Databricks

Vancouver, BC +1 (778) 637-1602 dharma.patel552@gmail.com linkedin.com/in/kush-patel-48885719b github.com/KushPatel29
kushpatel29.github.io

Open to remote and relocation across Canada · Legally eligible to work in Canada · Available immediately

SUMMARY

- Analytics professional with 4+ years across finance and operations, including Azure/Fabric pipeline ownership as sole analyst at a perishable-goods distributor. Integrated ERP, POS, workforce and API feeds using SQL and PySpark.
- Public engineering evidence demonstrates quarantine/replay, failure gates, migration reconciliation and scheduled execution. Snowflake/AWS runs use personal accounts; Airflow, Kafka, Terraform and Kubernetes evidence is project/CI execution, separate from enterprise operations.

TECHNICAL SKILLS

Pipelines: Azure Data Factory, Fabric pipelines/notebooks/Dataflows Gen2, Azure Synapse, SSIS, SQL Server Agent; REST ingestion, incremental loads and schema-drift handling.

Processing & models: SQL/T-SQL, PySpark, Python/pandas, Delta Lake, SQL Server, Fabric Lakehouse/Warehouse; star schemas, SCD Type 1/2 and Power BI semantic models.

Reliability: Contracts, quarantine/replay, quality gates, control totals, freshness, idempotent reruns, exception logs, alerts, cutover checks and rollback runbooks.

Project platforms: Databricks, Snowflake/dbt/Streams/Tasks/Snowpark, DuckDB, AWS Athena/S3/Glue; Airflow scheduler, Redpanda/Kafka, Terraform and Kubernetes in CI; GitHub Actions, Docker and pytest.

PROFESSIONAL EXPERIENCE

Two Rivers Specialty Meats

Dec 2023 – Jun 2026 · North Vancouver, BC

Data Analyst — Operations, Logistics & Enterprise Reporting

- As sole analyst, integrated ERP, Square POS, workforce and REST feeds into Bronze/Silver/Gold datasets using Microsoft Fabric, Azure Data Factory/Synapse, SQL and PySpark.
- Maintained SSIS and SSRS while moving priority SQL Server pipelines to Fabric notebooks; validated cutover with row counts, checksums and control totals.
- Automated schema, completeness, uniqueness and referential-integrity checks, with anomaly thresholds, exception logs and pipeline alerts.
- Built source-to-mart GL/P&L and inventory controls, cutting discrepancies an estimated ~45% and reporting errors ~30% (internal directional estimates, not audited).
- Designed star-schema marts and SCDs; served curated data through Power BI with RLS/RBAC and column masking.
- Delivered FEFO and lot-level traceability, customer inactivity and rep-coverage reporting, and SKU demand forecasts as Python web apps supporting procurement and lower perishable waste.

Shivam Investments

Dec 2020 – Jul 2022 · Vadodara, India (hybrid)

Financial Analyst — Data & Reporting

- Gathered requirements from portfolio and risk stakeholders; delivered Power BI dashboards with auditable measure definitions and reconciled GL/P&L figures across sources.
- Automated recurring preparation with Python and SQL views, replacing spreadsheet consolidation and cutting manual effort an estimated ~40% (internal directional estimate).

SELECTED PROJECTS — public code and CI-tested results; synthetic data is labelled

Supply Chain Control Tower — PySpark, Delta Lake, Databricks, SQL, Power BI · synthetic data

github.com/KushPatel29/supply-chain-control-tower · live: kush-network-risk-decision-room.streamlit.app

A medallion pipeline defended in three layers: schema contracts before Bronze, row quarantine with replay, and a quality gate on Gold, each proven in CI by injecting the failure it exists to catch. Benchmarked at 10M rows and run as a Databricks job, with all 36 Gold totals matching the local build to the cent.

Legacy-to-Fabric Migration — SSIS, SSRS, PySpark, Microsoft Fabric, Python · synthetic data

github.com/KushPatel29/legacy-to-fabric-migration · live: kush-asset-management-decision-board.streamlit.app

Cutover that has to prove itself: the move from SSIS/SSRS to Fabric waits on a row-for-row GO/NO-GO gate across 107 artifacts comparing row counts, keys, values, control totals and checksums. The gate is cent-exact and conjunctive, an empty output never passes, and tests break the data on purpose to prove it catches dropped rows, shifted values and errors that cancel out.

Canada Wildfire Risk — Python, requests, Parquet, GitHub Actions · real public data

github.com/KushPatel29/wildfire-prediction-app · live: wildfire-prediction-app1.streamlit.app

A scheduled pipeline that keeps running: pulls CWFIS station observations and an Open-Meteo forecast twice a day since September, rebuilds a seven-day forecast for 753 grid cells, and publishes it to a release and a public data branch. A weekly job grades published forecasts against satellite hotspots and commits the result. Found and fixed an index-alignment bug that had put 97.7% of fires in the wrong cell.

Real-time Transaction Monitoring — Kafka (Redpanda), Python, Terraform, Kubernetes · synthetic data

github.com/KushPatel29/aml-transaction-monitoring

A stream held to its batch: two windowed rules run in event time over 100,795 deliveries that arrive late and twice; at-least-once delivery with idempotent, window-keyed alerts on a compacted topic. Terraform deploys the broker, a two-replica consumer group and its jobs to a Kubernetes cluster in CI, a pod is killed mid-stream, and the run passes only if the stream flags exactly what the batch flags: 51 of 51 and 1,270 of 1,270.

Supply Chain Analytics — dbt — dbt Core, Databricks, Snowflake, DuckDB · synthetic data

github.com/KushPatel29/supply-chain-analytics-dbt

Same project, three engines: built on DuckDB, on a Databricks serverless SQL warehouse and inside Snowflake with dbt Projects on Snowflake: 180 of 180 nodes green and all 41 mart totals equal across all three, with contracts, incremental facts and SCD2 snapshots. An AWS Athena run on private S3 data reconciles 32 warehouse measures to Python, under a 10 MiB scan cap. Airflow runs the build nightly under its scheduler in CI; separate Snowflake Streams/Tasks and Snowpark checks reconcile 94,758 retail lines after duplicates and corrections.

A twice-daily public-data forecast pipeline and weekly outcome grading are available at kushpatel29.github.io. Receipts distinguish scheduled CI from manual cloud runs.

EDUCATION

Northeastern University — Master of Professional Studies in Analytics (Applied Machine Intelligence), GPA 3.76/4.0 · Vancouver, BC — **Sep 2022 – Jun 2024**

Gujarat Technological University — Bachelor of Engineering, Computer Science · India — **Aug 2017 – Aug 2021**

CERTIFICATIONS

Databricks Fundamentals · UC Davis SQL for Data Science · IBM Python for Data Science, AI & Development · Google Data Analytics Professional Certificate. Preparing for the Microsoft DP-600 (Fabric Analytics Engineer) exam.