

# SURANJIT BANIK

Senior Data Engineer | Azure Databricks Specialist

Toronto, ON | 647 616 9494 | suran.cse@gmail.com | linkedin.com/in/suranjitbanik | mastermindlab.github.io

## PROFESSIONAL SUMMARY

Senior Data Engineer with 5+ years building on Azure Databricks: Delta Lake, Lakeflow Declarative Pipelines, Unity Catalog, Databricks Apps, Asset Bundles, and Genie. Delivers production lakehouse pipelines, AI powered data products, and CI/CD automation across insurance, digital media, and analytics. Early adopter of agentic development on Databricks using Claude Code, the Databricks MCP server, and Omnigent.

## CORE SKILLS

**Databricks:** Delta Lake, Lakeflow Declarative Pipelines (DLT), Unity Catalog, Databricks Apps, Workflows, Asset Bundles, Auto Loader, Genie, Lakebase

**Languages:** Python, PySpark, Spark SQL, SQL

**Cloud:** Azure (ADLS, DevOps, ADF, Logic Apps), AWS S3, GCP BigQuery

**AI / Agentic Tooling:** Claude Code, Databricks MCP server, Databricks CLI, Omnigent, Claude Cowork

**Tools:** GitHub, Jira, Confluence, Streamlit

**Strengths:** Medallion Architecture, Metadata Driven Ingestion, Data Governance, CI/CD, ETL Design, Performance Tuning

## PROFESSIONAL EXPERIENCE

**Senior Data Engineer** | Zurich Canada, Toronto, ON

Apr 2024 to Present

- Built an AI powered policy and claims app with Databricks Genie and Lakebase, letting underwriters query live data in natural language and persist open claims notes to a Lakebase OLTP store, cutting insight turnaround from days to seconds.
- Developed a Databricks App for bordereaux (BDX) file intake with Claude LLM powered column mapping that strips special characters, converts headers to snake\_case, and auto maps columns to the target schema, eliminating manual mapping fixes in lakehouse ingestion.
- Built declarative medallion pipelines (bronze, silver, gold) on Lakeflow with data quality expectations, CDC, Unity Catalog governance, and Auto Loader for schema resilient ingestion across insurance data platforms.
- Designed config driven PySpark ETL frameworks reused across 30+ brokers data, cutting new source onboarding time by roughly 60%.
- Modernized CI/CD with Databricks Asset Bundles and Azure DevOps across dev, qa, uat, and prod, improving release velocity and environment portability.
- Optimized Spark jobs with Predictive Optimization, liquid clustering (CLUSTER BY) on serverless compute, and join strategy tuning, reducing runtimes by roughly 40%.
- Adopted agentic engineering workflows pairing Claude Code with the Databricks MCP server and CLI to scaffold pipelines, generate tests, and debug jobs, shipping multiple projects ahead of schedule.

**Data Engineer** | GroupM, Toronto, ON

Aug 2021 to Apr 2024

- Built Databricks and PySpark pipelines on GCP integrating BigQuery, Bigtable, and Cloud Functions, improving reporting performance by 60%.
- Delivered multi cloud solutions across GCP, Azure, and AWS and led data governance and compliance for enterprise data warehousing.

**Data Engineer** | Clue, Toronto, ON

Jan 2021 to Jul 2021

- Designed graph data models with Neo4j and Cypher and built ETL pipelines on AWS (S3, Athena, SageMaker, DataBrew) for digital marketing analytics.

**Data Analyst** | Clue, Toronto, ON

Feb 2020 to Dec 2020

- Built Azure analytics services ingesting Salesforce data for media campaign reporting with near zero downtime on real time pipelines.

**Digital Analyst** | Mindshare, Toronto, ON

May 2019 to Feb 2020

- Architected Datorama solutions for Nestle and CPG clients, streamlining marketing data integration and campaign ROI measurement.

## EDUCATION

**Master of Engineering, Computer Engineering**

University of Ottawa

**Bachelor of Engineering, Computer Science**

Visvesvaraya Technological University

## CERTIFICATIONS AND PROFESSIONAL DEVELOPMENT

**Databricks Certified Data Engineer Associate** | AWS Certified Cloud Practitioner

Valid through Dec 2026

Databricks Data + AI Summit, San Francisco: hands on training on Lakeflow, Lakebase, and Genie

2026