

Patrick DENG

+1 647 514 4114 | keqi.deng@outlook.com | kdeng4.github.io | linkedin.com/in/deng-patrick

Data & MLOps Engineer (Toronto) | GCP: BigQuery, Dataproc | Airflow | PySpark | Python | SQL
ELT at 70M+ records per day | Walmart Connect (Ads) CRM data warehousing | Identity resolution (2%) | Email recs (ALS/CF, \$10M annualized) | GenAI (LangGraph)

Professional Experience

MLOps & Data Engineer at Walmart Canada | Mississauga ON

2024.07 - Present

- Owned end-to-end **Walmart Connect (Ads) CRM** data pipelines and warehouse architecture (ingestion to curated marts), enabling reliable activation and measurement for marketing and ads stakeholders.
- Automated ELT at **70M+ records per day** using BigQuery, Airflow, and Python, improving reliability, scalability, and operational maintainability.
- Built a **customer identity-resolution** layer (**2%** increase in identifiable transactions) and standardized the transaction “data story” across department reporting and executive consumption, improving consistency for downstream analytics and modelling.
- Extended transaction data from executive-level sales to item-level purchases mapped to customers via the identity graph, enabling model training datasets and deeper customer behaviour analysis.
- Built and shipped an internal **GenAI agent** using **LangGraph** for task-specific query generation and data retrieval, accelerating self-serve analytics and cross-team data access.
- Designed and operationalized **CRM email item recommendations** for Customer Office using collaborative filtering (ALS/CF); improved engagement and conversion and drove **\$800k per month** incremental impact (**about \$10M** annualized estimate).
- Built customer segmentation (clustering and dimensionality reduction) to support targeted campaigns and lifecycle strategies; partnered with marketing to operationalize segments in CRM.
- Led basket continuity and mission analytics using statistical methods on purchase sequences (for example, repeat behaviour, co-occurrence, and temporal patterns) to quantify how baskets evolve over time and link shifts to ongoing business events.
- Modelled omni-channel transitions between store and e-commerce using Markov chains, estimating transition and absorption rates and measuring post-transition department breadth and spend to quantify cannibalization versus incremental growth.
- Prototyped purchase connectivity across channels using Spark ALS (implicit signals) to surface cross-channel gaps and opportunities between store and online.
- Delivered high-visibility analytics (for example, new store opening impact and geographic strategy assessment) and presented findings in department town halls, translating complex pipelines and models into clear, decision-ready storytelling.

Engineer & Analyst, Data, Strategy at Dane Creek Capital Corp. | Toronto ON

2021.07 - 2024.07

- Built analytics pipelines on **GCP** orchestrated by **Airflow**, integrating multi-source sales and operations data into standardized datasets for reporting and decision-making.
- Defined a KPI and metrics layer and delivered executive dashboards in **Metabase**, improving visibility into sales performance and reducing manual reporting.
- Deployed and administered **self-hosted Metabase** (access control, refresh schedules, reliability), improving data accessibility while reducing dependence on paid BI tooling.
- Partnered with stakeholders to translate goals into measurable analyses; delivered marketing and customer analytics using **Python/SQL** to inform campaigns and improve engagement and retention.
- Supported loyalty programme launch by defining core metrics and measurement approach; guided iteration using data insights.

Professional Skills

Core Stack	GCP (BigQuery, Dataproc, Cloud Storage, Cloud Functions), Airflow, PySpark, Python, SQL, Linux, Git
Data Engineering	ETL/ELT, Data warehousing, Data modelling (dimensional marts), Batch pipelines, Orchestration, Data quality and validation, Monitoring and SLAs, Identity resolution
Platforms	Docker, CI/CD, Kubernetes, Kafka, REST APIs
Databases & BI	PostgreSQL, MS SQL Server, MySQL, Cassandra, Metabase, Looker, Tableau, Power BI
ML/AI (Applied)	Recommender systems (ALS, collaborative filtering), Clustering, Dimensionality reduction, Experimentation (A/B testing), PyTorch, GenAI (LangGraph), LLM APIs (GPT, Gemini)
Other	Azure, R, Java, C#, MATLAB, LaTeX

Education

3.8 / 4.0	M.Sc. in Computer Science , Georgia Institute of Technology Atlanta, GA (Remote)	2025
3.9 / 4.0	M.Eng. in Chemical Engineering , University of Calgary Calgary, AB	2020
3.8 / 4.0	B.Eng. in Chemical Engineering , University of Calgary / Southwest Petroleum University (Cohort) Calgary, AB	2018

Publications

Patrick Deng, Alex Greenhalgh, Brian Yu, David A. Joyner, Nicholas Lytle.

“Exploring transitions of graduates from an online master’s in computer science program to doctoral programs.”
SIGCSE TS 2026 (Papers Track), accepted, Feb 2026.