

Calvin Du

587-888-7313 | calvingdu@gmail.com | [linkedin.com/in/calvingdu](https://www.linkedin.com/in/calvingdu) | github.com/calvingdu

TECHNICAL SKILLS

Programming/Scripting: Python (Scikit-learn, PyTorch, Numpy, Pandas), SQL, Julia, Java, R, Linux/Bash
Tools: Docker, Airflow, Prefect, Spark/PySpark, Kafka, Kubernetes, Elasticsearch, Kibana, Tableau, PowerBI, Git
Cloud & Database: Databricks, AWS (S3, EC2), Azure (ADLS), MySQL, PostgreSQL, NoSQL (MongoDB)

WORK EXPERIENCE

Machine Learning Engineer Intern

May 2024 – Present

Royal Bank of Canada

Toronto, ON

- Developed classification and predictive models with Sklearn and Pytorch to predict transaction settlement delays, leveraging feature engineering and time forecasting to achieve 94% accuracy
- Improved data retrieval times by 64% with efficient data modeling for over 100 financial sources, optimizing API query performance using Databricks data warehouse and deltalakes
- Enabled real-time data processing with Kafka and Spark Streaming, reducing data latency by 74%

Data Engineer Intern

Jan 2023 – Apr 2024

Royal Bank of Canada

Toronto, ON

- Developed an end-to-end scalable on-prem and cloud data platform with Databricks, PySpark, AWS S3, Azure ADLS, and Airflow, achieving 50x faster data processing for 400+ pipelines and 50+ data assets
- Created an in-house package that encompassed data transformation logic and utilized PySpark and database utilities with Azure SQL Database for both batch and real-time processing
- Co-owned a data quality platform utilizing GreatExpectations and Soda achieving a 70% reduction in data quality issues with data visualization dashboards using Elasticsearch and Kibana

PROJECTS & OPEN-SOURCE DEVELOPMENT

Investment Instrument Data Platform | Python, Apache Airflow, Apache Spark, PostgreSQL, Docker

- Architected a data platform using Apache Airflow for orchestrating ETL processes with dynamic DAG creation to upload into a PostgreSQL database applying data across 10+ financial, news and economic sources
- Utilized Apache Spark for leveraging distributed computing to enhance the scalability of ETL and ML deployment
- Implemented DevOps best practices to streamline development utilizing Docker containers, CI integrations (GitHub Actions), and pre-commit hooks for collaboration and deployment optimization

GreatExpectations Open-Source Contribution | Python, Spark, SQL, Git

- Created and enhanced custom expectations using Python, Spark and SQL engines to validate the data integrity of the GreatExpectations library, a cutting-edge data validation and testing tool
- Recognized as a GX Champion for exceptional code contributions and involvement in community support

Canadian Bank Transaction Tracker | Python, Prefect, MongoDB, Docker, GreatExpectations, Scikit-learn

- Engineered a data pipeline with Prefect, efficiently processing and storing 3000+ monthly bank transactions
- Developed a categorization machine learning model with Scikit-learn with dynamic transaction matching achieving 76% accuracy and reducing manual categorization time by 50%

EXTRACURRICULARS

VP Finance

May 2022 – June 2023

Enactus UBC

Vancouver, BC

- Led a team of 8 Finance Managers, coordinating budgets and finances for 6 sub-groups with 150+ members and managing \$23,000+ in expenses to successfully end the year with \$15000+ net income
- Directed 7 Fundraisers and 23 proposals for sponsorships securing \$38,655, a 3x increase from the previous year
- Awarded John Dobson Founders Award (\$2500) for exceptional leadership and contributions to Vancouver

EDUCATION

University Of British Columbia

Vancouver, BC

Bachelor of Science in Statistics and Economics, Minor in Data Science

Sep 2020 – Dec 2025

- Relevant Courses:** Applied Machine Learning, Applied Linear Algebra, Statistical Inference, Advanced Econometrics, Quantitative Economics with Data Science Applications, Computational Methods in Economics