

Deepakjot Singh

Data Engineer

604-614-9479 | deepakjotsingh8@gmail.com | [linkedin.com/in/deepakjotsingh](https://www.linkedin.com/in/deepakjotsingh) | [Github](#) | [Portfolio](#)

PROFESSIONAL SUMMARY

- Data engineer skilled in Python, SQL, and PySpark, building distributed analytics pipelines on AWS (Glue, Athena) and Azure using Terraform and Airflow, standardizing data schemas and optimizing query performance.

TECHNICAL SKILLS

Languages: Python, SQL, C#, JavaScript, TypeScript

Libraries: PySpark, Scikit-learn, Pandas, NumPy, TensorFlow, PyTorch, Matplotlib, Seaborn

Platforms & Tools: AWS (S3, EC2, RDS, Amplify, Glue, Athena), Microsoft Azure, Databricks, Airflow, Parquet, Google BigQuery, Power BI, Github

PROJECTS

ResearchFlow | *PySpark, AWS (Glue, Athena, Lambda, EventBridge, S3), Terraform* Jul 2026 – Present

- Built an end-to-end **AWS** pipeline ingesting **10 years** of data across **505 S&P 500** tickers, producing **2.5M** rows across two strategies with incompatible schemas.
- Standardized schemas via distributed **PySpark** on **AWS Glue**, automated via **EventBridge** and **Lambda**. **Athena** SQL analysis revealed **40%** signal agreement between strategies.

Athletics Performance Analytics | *PySpark, SQL, Databricks, AWS S3* Sep 2024 – Oct 2024

- Built an end-to-end **PySpark** pipeline on **Databricks** using **AWS S3**, **Spark SQL**, and visualization tools to transform and analyze IPL datasets, enhancing query speed and reporting efficiency by **30%**.

Algorithmic Trading Strategist | *Python, GARCH, Twitter API, Scikit-learn* May 2024 – Sep 2024

- Developed an unsupervised learning trading strategy on **S&P 500** data using **Python** and **Scikit-learn**, resulting in improvement in portfolio performance through feature engineering and machine learning optimization.

WORK EXPERIENCE

Business Analyst Jan 2026 – Present
Seaspan Delta, BC

- Built advanced **SQL** queries and **PySpark**-based **machine learning** pipelines on **Microsoft Azure** to power **Power BI** dashboards for real-time business visibility, optimizing script and query runtime by **60%**.

GenAI Tool Developer May 2025 – Nov 2025
Centre for Teaching, Learning and Technology, UBC Vancouver, BC

- Developed GenAI-powered **H5P** tools using **TypeScript**, **Node**, **React**, and **RAG** pipelines; integrated into 10+ UBC courses, reducing manual question creation by **70%** and improving student engagement.

XR Software Developer Jan 2025 – April 2025
Faculty of Medicine, UBC Vancouver, BC

- Designed standalone VR medical simulations in **Unity** using **C#**, delivering 3+ modules used by **50+** students across British Columbia, enhancing engagement through immersive, faculty-informed educational experiences.

Software Developer May 2023 – Dec 2024
The University of British Columbia Vancouver, BC

- Developed a full-stack app using **JavaScript**, **React**, **Node**, **MongoDB** on **AWS EC2** and **Amplify**, leveraging **S3** for storage, **RDS** for database, hosted on **UBC's website** for **40,000+** students.
- Built a secure admin system with **React**, **Node**, **MongoDB**, managing **30+** user-added markers using **JWT** authentication, **Bcrypt** hashing, and optimized APIs documented with **Swagger**.
- Processed **4 million** tweets using **Python** and **Detoxify**, with parallel, batch processing and memory optimization to reduce toxicity score runtime from **14** hours to **42** minutes leading to **95%** improvement in efficiency.

Undergraduate Teaching Assistant Jun 2024 – Dec 2024
Sauder School of Business, UBC Vancouver, BC

- Led practical sessions for COMM 414, teaching **SQL**, **Python**, and Google **BigQuery** to **100+** students, covering fundamental machine learning algorithms, and introducing both frequentist and Bayesian statistical approaches.

EDUCATION

Bachelor of Science in Mathematics Sep 2020 – May 2025
University of British Columbia Vancouver, BC