

Yixi Zhao

<https://www.linkedin.com/in/yixi-zhao-816ab6189/>

☎ (+1) 647-818-8621

✉ zhaoyixi0306@gmail.com

Skills

Languages JavaScript | Java | Python | SQL | HTML | CSS
Technologies Git | Linux | Azure | AWS | Terraform | Jenkins | Power-shell | Agile/SCRUM | Databrick | Spark/Pandas
Certificate Microsoft Certified: Azure Data Fundamentals

Work Experience

05/2024– **Data Engineer Coop**, RBC INVESTMENT SERVICES, Toronto, Ontario.

- 08/2024
 - Used **Apache Airflow** to create and manage data workflows aligned with project requirements.
 - Enhanced **Airflow** framework to support conditional dataset dependencies along with cron scheduling.
 - Conducted **Data Quality(DQ) tests** for existing workflows to ensure data integrity and functionality.
 - Implemented **Unit tests** in existing framework to validate functionality and ensure code reliability..
 - Experienced in container-based deployments using **Docker**, **docker compose**, and **docker image**.
 - Developed command-line interfaces with **Click** to simplify workflow automation and user interaction.
 - Established secure access and managed permissions for Azure resources using **AD** and **Key Vault**.
 - Assisted the business team in generating reports by **aggregating** internal data and **scraping** public data.

09/2023– **Performance Analyst Student**, BLACKBERRY, Waterloo, Ontario.

- 04/2024
 - Designed **ETL** process to extract data from SQL server to AWS S3 bucket.
 - Built **streaming** and **batch** data pipelines using **Delta Live Tables** in Databricks.
 - Leveraged **Terraform** to deploy data pipelines, manage AWS infrastructure, and oversee Unity Catalog tables.
 - Executed **Continuous Integration and Delivery (CI/CD)** pipeline in **Bitbucket** with Atlassian.
 - Transitioned workflows from **Job Cluster** to **Serverless** to optimize costs and improve resource allocation.
 - Assisted in debugging existing workflows, addressing **schema loss**, **data quality**, and **memory leaks**.
 - Supported the team in tracking cluster performance and expenses, contributing to a **cost-saving strategy**.
 - Created user tagging strategies leveraging multi-layer golden tables for more precise **data categorization**.
 - Partnered with the product team to develop dashboards from golden and silver datasets within Databricks.

05/2022– **Research Assistant**, YORK UNIVERSITY, Toronto, Ontario.

- 08/2022
 - Used **Python** to extract keywords and sentences from target datasets.
 - Manually validated the outcomes and summarized the findings.
 - Proposed ideas and refined scripts based on the professor's guidance in weekly **scrum meetings**.
 - Utilized the **Python Pandas library** to perform **statistical analysis** and **generated graphs** based on datasets.
 - Assisted in drafting and refining the paper, and collaborated with the professor on publishing the work.

Projects

Ongoing **Privacy Leakage in Android App**, University of Waterloo.

- Experienced with build tools like **Maven**, **Ant**, and **Gradle** to create deployable artifacts from source code
- Leveraged **BeautifulSoup**, **REST APIs**, and **Selenium** for web scraping from public sources.
- Deployed **AndroidGuard** to gather data, and perform APK analysis for comprehensive app information retrieval.
- Applied **statistical methods** using **NumPy** and **pandas** to analyze datasets.
- Conducted **taint analysis** on applications using **FlowDroid**, collecting datasets for detailed security evaluation.

Education

2023–2025 **University of Waterloo**,
Software Engineering, Master of Applied Science.

2022-2023 **Concordia University**,
Computer Science, Master of Applied Science.

2016-2022 **York University**,
Computer Science, Bachelor of Science.