

# Wanyi Su

+1 778-319-9915 • wysujhu@gmail.com • Burnaby, BC, Canada

Personal Website: <https://wanyisu.netlify.app/> • GitHub: <https://github.com/suansuan0915>

## SKILLS

**Languages:** Python, SQL (SnowSQL, MySQL, PostgreSQL), Scala, NoSQL (Cassandra), MATLAB, R, bash scripting.

**BI and Other Software:** Tableau, Snowflake, Databricks, SVN, Git, Alteryx, Docker, Kubernetes, Control-M, Jupyter Notebook, IntelliJ, MS Office.

**Frameworks and Libraries:** PySpark, Spark Scala, Airflow, Kafka, Hadoop; pandas, NumPy, Sklearn, Matplotlib, Seaborn, PyTorch; Flask.

**Cloud Platform:** Azure (DevOps, Machine Learning, Databricks), AWS (EC2, S3, Redshift).

**Certificates:** Microsoft Certified: Azure Data Scientist Associate, CFA level 2.

## EDUCATION

**Simon Fraser University** | Burnaby, BC, Canada

Sep 2021 – Jan 2023

*M.S. in Computer Science - Big Data (GPA: 4.0/4.33)*

**Johns Hopkins University** | Baltimore, MD, United States

Jul 2016 – Aug 2017

*M.S. in Finance*

**Southwestern University of Finance and Economics** | Chengdu, China

Sep 2012 – Jun 2016

*Major in B.A. in Marketing; Minor in B.E. in Finance*

## WORK EXPERIENCE

**EnviroNics Analytics** | Toronto, ON, Canada

May 2023 – Dec 2023

*Data Engineer*

- Led weekly update and maintained the cloud based ETL pipeline for a real-time mobility data product using **Airflow**.
- Refactored and optimized an ETL pipeline for a geospatial product using tools (**Azure DevOps**, **Airflow**, **Alteryx**) and languages (**SnowSQL**, **Python**).
- Conducted testing and anomaly detection by machine learning and statistical modeling in **Databricks** and **Snowflake**.
- Automated and visualized data quality check process (hard & soft QC) and data flow monitoring by designing **Tableau** dashboard.
- Drafted technical documentation for two products, and cooperated with technical/non-technical stakeholders about product statuses/issues.

**SFU Computing Science School** | Burnaby, BC, Canada

Sep 2022 – May 2023

*Graduate Teaching Assistant (Part-Time)*

**Insurance Corporation of British Columbia (ICBC)** | North Vancouver, BC, Canada

May 2022 – Sep 2022

*Data Engineer Intern*

- Developed and maintained an ETL data pipeline in **Spark Scala** with 100% accuracy to facilitate data migration.
- Implemented data validation using **SQL** queries to extract large scale data from different databases (**Oracle**, **Drill**, **Hive**).
- Applied software engineering best practice during the pipeline building process, including unit testing, **shell scripting**, CI/CD automation using **Jenkins**, version control using **SVN**, etc.).
- Built **Tableau** dashboards to provide varying levels of visualization for stakeholders and automate data quality monitoring.
- Collaborated with internal and external stakeholders in **Agile** environment (**Scrum** meetings, **JIRA**, **Kanban** board).

**Morgan Stanley** | Baltimore, MD, United States

Jan 2018 – May 2018

*Operations Data Analyst*

- Analyzed 20+ derivatives trading transaction data daily. Created **Tableau** dashboards and presented to clients and directors.
- Queried from internal database using **MySQL**, quantified risk metrics and developed KRI reports using v-lookup, pivot table.
- Streamlined internal risk management processes and improved drafting time by 20%. Led weekly conference calls with Asian teams.

## PROJECTS

**Music Distributed Microservice Application Project**

Mar 2022

- Built a music containerized application, including 3 microservices, running within **Kubernetes** cluster on **AWS EKS**.
- Implemented **REST** requests between local machine and **DynamoDB** with functions of reading, creating, editing, deleting.
- Scaled and tested the application by load testing using **Gatling** simulation to observe accuracy, latency, system utilization, and health of system mesh of the microservices under small, medium, and large request load scenarios.
- Visualized and quantified the performance metrics using **Kiali**, **Grafana** dashboard, and **Prometheus**.
- Managed the project using **Scrum** and Kanban in a team with 3 teammates and applied version control using **GitHub**.
- Explored failure scenarios and tested with remediation to inspect multiple system evolving.