Vincent Le

https://github.com/vincent-tn-le

+1 (647) 936 0155
vincent.tn.le@gmail.com
https://vincent-tn-le.github.io/website

WORK EXPERIENCE

Egress Software, Data Engineer   *Toronto,* **Aug 2021 – June 2023**

- Orchestrated and scaled the streaming pipelines using Apache Spark and Azure (Blob Storage, Synapse) to support reporting, internal insights, and product R&D for 10+ thousand client organizations
- Automated the preprocessing pipelines for unstructured streaming data using PySpark and Azure to detect outliers, pseudonymize identifiable information and compute aggregate statistics
- Deployed and scheduled ETL batch pipelines using Azure Databricks to process unstructured client data ingested by event hubs into Delta Tables based on medallion architecture
- Developed and implemented the test suite using Databricks and Python to ensure data integrity across geographies
- Trained new team members on how to use the technology stack and guided them towards outstanding performance evaluations

Wealthsimple, Operations Specialist   *Toronto,* **Feb 2021 – Aug 2021**

- Developed automated scripts to input end of day financial transfers information into legacy database using Python
- Built data visualizations to track internal metrics and KPIs for institutional transfers team using AWS Redshift

PwC, Data Analyst Intern   *Toronto,* **May 2019 – Dec 2019**

- Created real-time dashboards supporting account migrations for banking client using Tableau ensuring 100+ thousand accounts onboard correctly
- Conducted time series analysis using Python to identify outliers and approve entries for purchase deal processes remediating $2 million of inventory

RBC, Data Scientist Intern   *Toronto,* **May 2018 – Sep 2018**

- Trained autoencoder models in Python with Tensorflow to detect outliers in event data
- Built and scheduled the ETL pipeline using PySpark to preprocess internal network access data by user required for misuse detection strategies

PROJECTS

Price Aggregator Pipeline Dashboard (A group project) • *Javascript, React, Python, SQL*   *Ongoing*

- Built a web app to display art supply infographics based off a database of products and prices over time
- Collaborated with a team of colleagues to incorporate Python with SQL database, Heroku integration, React UI components to provide features and a simple modern user interface

TECHNICAL SKILLS

- **Programming Languages:** Python, SQL, R, Scala, C, JavaScript
- **Web Development:** HTML, CSS, React.js
- **Tools:** Git, Azure (Data Lake, Event Hubs, Blob storage), Apache (Spark, Kafka, Cassandra), Databricks, Tableau, AWS

EDUCATION

Bachelor of Mathematics, Statistics Major – University of Waterloo Class of 2020

ADDITIONAL INFORMATION

Certifications   AWS Cloud Practitioner, Databricks Certified Data Engineer Associate, Google Data Analytics

Interests   Home Cooking, Badminton, Visual Arts, UW StarterHacks