

# SIMON WONG

[✉ /simon.wong@uwaterloo.ca](mailto:/simon.wong@uwaterloo.ca) · [🐙 /simonwongwong](https://github.com/simonwongwong) · [🌐 /simonwongwong.io](https://www.linkedin.com/company/simonwongwong) · [in/simonwongwong](https://www.linkedin.com/company/simonwongwong)

## SUMMARY

---

- **Languages:** Python · Scala · C/C++ · Java · Javascript · MATLAB
- **Tools:** Apache Spark · AWS (EC2, EMR, S3) · pandas · Hadoop · flask · SQL · Airflow · Docker · React

## EXPERIENCE

---

### Stripe

*Software Engineering Intern - Data Libraries Team*

January 2020 - May 2020

*San Francisco / Remote*

- Responsible for team's effort to rewrite and optimize critical Scalding pipelines by using Apache Spark
- Reduced compute costs for data pipelines by over \$140,000 per year through Spark migration
- Improved availability of over 2000 pipelines through upstream and direct optimizations
- Parallelized daily validation and verification jobs to reduce execution times by up to 94% for large jobs
- Discovered and rectified a complex logical error that exposed millions of pieces of sensitive PII

### Capital One

*Data Science Engineering*

May 2019 - August 2019

*Toronto*

- Parallelized a data visualization pipeline using python's multiprocessing library, improving runtime by up to 93%
- Scaled up data validation tools by utilizing PySpark, reducing runtime by 85% for DataFrames with > 25M rows
- Significantly increased testing suite code coverage by upgrading to *Hypothesis* testing
- Created and maintained a variety of tools to aid Data Scientists with creating Machine Learning models

### exactEarth Ltd

*Software Engineering Intern*

September 2018 - December 2018

*Cambridge*

- Increased efficiency of Spark jobs by designing and running experiments to improve resource allocation
- Designed and wrote a stateful Spark Streaming service to rate limit data streams from Kafka
- Created a queue API and microservice to manage Spark jobs using flask to optimize compute and data output

## PROJECTS

---

### Facebook Messenger Statistics

*Data science project*

[git.io/vpQno](https://github.com/simonwongwong/facebook-messenger-statistics)

- Wrote a Jupyter Notebook for exploratory analysis on Facebook Messenger usage, utilizing pandas
- Created a script to generate an interactive HTML report using plotly for interactive data visualization
- Aggregates and reports analytics such as most messaged users, most used words, most active times, etc.

### Bluetooth Bicycle Turn Signals

*Hardware project*

[git.io/vpQnd](https://github.com/simonwongwong/bluetooth-bicycle-turn-signals)

- Custom designed and built a wireless LED turn indicator for bicycles using ATMEGA328P chips
- Switches embedded into handlebars communicate with LED lights attached to bicycle using Bluetooth
- Designed a wireless power switch controlled by manipulating the flux path of magnets and a reed switch
- Developed an accompanying Python program for generating custom turn signal animations

## EDUCATION

---

### University of Waterloo

*B.ASc in Mechatronics Engineering - Cumulative average of 86%*

2018 - present

*Waterloo, ON*

- Relevant Courses: Algorithms and Data Structures, Computer Structures & Real-Time Systems, Microprocessors and Digital Logic, Microprocessor Systems and Interfacing