Carl Shen

github.com/**csheng8**
linkedin.com/in/**csheng8**
csheng8.github.io
carl.shen@uwaterloo.ca

Skills

Languages – Java, C++, C#, JavaScript, Python, SQL

Technologies – ***Distributed Computing:*** Spark, Hadoop; ***Data Pipelines:*** Apache Storm, Kafka, Gobblin; ***Cloud:*** Azure; ***Data Stores:*** Redis, Apache Druid; ***Machine Learning:*** Weka, Spark MLlib, Photon ML; ***Web Backend:*** Jersey, Node.js

Experience

LinkedIn – Gobblin Team – Systems and Infrastructure Engineering Intern — Mountain View, *CA* | Sep – Dec *2018*

- Implemented emission of stateful ingestion statistics messages within the ***Gobblin-Kafka*** source and extractor constructs for the purpose of providing metrics for a profiling and monitoring dashboard.
- Built an end to end suite of block aware components within the ***Gobblin-DistCp*** module and ported existing file aware implementations of data publishing jobs to the new pipeline, improving performance of a major bottleneck.

Yahoo Sports – Daily Fantasy Backend Team – Software Engineering Intern — Sunnyvale, *CA* | Jan - Apr *2018*

- Owned a projected points feature that enhanced the Daily Fantasy platform's live contest user experience by building a ***Apache Storm*** topology and leveraging a ***Redis*** data store to implement a heuristic based algorithm; Increased visits to contest details page during live contests by ***~10%.***
- Implemented an automated contest set up flow that replaced a daily hour-long process by adding a set of APIs within a ***Jersey*** and ***Datanucleus*** backend framework.

Communications Research Centre Canada – Software Engineering Intern — Ottawa, *ON* | Sep - Dec *2016*

- Built a data engineering pipeline for a radio spectrum monitoring system, integrating and running ***Spark*** data analytics scripts using an automated ***Azure Data Factory***; Reduced execution time by ***~10x***.
- Led the transfer of knowledge for migrating from Matlab to Spark as a data analytics framework, which involved porting existing algorithms and deploying resources in Azure, as well as running weekly Spark tutorial sessions.

Education

University of Waterloo – BSE Candidate – 3A Software Engineering — Waterloo, *ON* | Expected Apr *2021*

- Overall GPA – ***3.98/4*** | Cumulative Average – ***93.04%***
- Awards – First in Class/Upper Year Scholarship ($500 | 2018), Microsoft Tuition Scholarship ($2000 | 2017), President's Research Award ($1500 | 2017), Madter Engineering Faculty Entrance Scholarship ($5500 | 2015)
- Undergraduate Research Assistantships – Distributed Systems Lab, Stochastic Decoding Group

Projects

Spark Hockey Analytics *2018*

- Using ***Kafka*** and ***Spark*** to process data from NHL APIs for machine learning related hockey analytics projects, including a Redzone-like streaming script for predicting goal probabilities of each currently live game.

Interactive Spatial Data Visualization Tool *2017*

- Created an interactive data visualization tool for dynamically filtering and aggregating spatial data on a map view by leveraging a Geohash geocoding algorithm as a spatial index within an ***Apache Druid*** data store.

User Behavior Prediction Model (Cluster Computing 2017 Vol 20 Iss 2 | DOI 10.1007/s10586-017-0749-z) *2016*

- Implemented components of a novel smart home user behavior prediction model using a ***MapReduce*** and ***Hadoop*** framework to parallelize the algorithms, improving execution speed by up to ***5x.***

Smart Bed Monitoring System *2015*

- Built a system to recognize and monitor bed and sleep related activity by implementing a decision tree generated using the ***Weka machine learning*** library on an experimentally gathered data set.