CARL SHEN

github.com/cshen98 | linkedin.com/in/cshen98 | cshen98.github.io | +1 (226) 700-3906 | carl.shen@uwaterloo.ca

SKILLS

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

Technologies – *Distributed Computing:* Hadoop, Spark; *Stream Processing:* Kafka, Storm; *NoSQL Datastores*: Redis, Druid; *ETL/Data Pipelines*: Gobblin, Data Factory; *Machine Learning:* Weka, Spark MLlib; *Web Backend*: Node.js, Jersey

EXPERIENCE

Facebook – Software Engineering Intern — Menlo Park, *CA* | May – Aug *2019*

LinkedIn – ETL/Apache Gobblin Team – Software Engineering Intern — Mountain View, *CA* | Sep – Dec *2018*

- Enhanced a *MapReduce*-based *Gobblin* Distcp flow to allow for splitting of files into block level granularity work units, alleviating a mapper skew bottleneck by increasing parallelization of large files; resulted in up to *~4x* performance gain in data movement for various pipelines that supported various LinkedIn-Microsoft product integrations.

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

- Implemented a heuristic-based projected points algorithm using a *Storm* topology and leveraging a *Redis* datastore, powering a set of new live projections features within the live contest details page of the Daily Fantasy platform; resulted in increased user engagement (*~5%* more page views) during live contests.
- Built and architected *MySQL/Datanucleus JDO* schemas and *Java JAX-RS/Jersey* APIs to develop components of the backend infrastructure for a new product, Yahoo Fantasy Slate, which had *~25k* users within weeks of launching.

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

- Created an automated batch processing pipeline for a city wide radio spectrum monitoring system, which involved porting and integrating *Spark* data analytics scripts to be invoked by a *Data Factory* workflow via calls to a *Livy* REST interface; reduced execution time compared to existing process by up to *~7x*.

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 Faculty 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
- Extracurriculars – IEEE CSCWD Conference (2016) Organizing Team, London Science Fair Judging & Organizing Roles

PROJECTS & PUBLICATIONS

Spark Hockey Analytics — *2018*

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

WiFi-Based Indoor Localization System (IEEE CSCWD 2017 | DOI 19.1109/CSCWD.2017.806671Z) — *2017*

- Implemented a localization approach based on WiFi signal strengths info relative to mobile beacons only, reducing setup costs compared to a traditional fingerprint approach while achieving a similar localization accuracy.

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

- Developed components of a novel smart home user behavior prediction model using *MapReduce* which provided improved execution speeds by a factor of up to *~5x*.

Smart Bed Monitoring System — *2016*

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