

SKILLS

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

Technologies – Apache Spark, Hadoop, Apache Storm, Redis, Microsoft Azure, Nodejs, Weka Machine Learning

EXPERIENCE

Yahoo Sports

Sunnyvale, CA

Daily Fantasy Backend Team – Software Engineering Intern

Jan - Apr 2018

- Backend Java API development to build out features in the Yahoo Sports Daily Fantasy platform (>1 million users).

SAP

Waterloo, ON

Big Data Tooling Team – Software Developer Intern

May - Aug 2017

- Full stack JavaScript/Node.js development of features for a customer facing Database Explorer web app.
- Built an advanced data preview interface leveraging SAP HANA's high speed querying capabilities to allow users to interactively create charts and better visualize data.

Communications Research Centre

Ottawa, ON

Network Applications Group – Software Engineering Intern

Sep - Dec 2016

- Prototyping and development of data engineering applications leveraging Apache Spark and Microsoft Azure and using Java and C# for a city-wide radio spectrum usage testbed system (>10 GB of data generated per day).
- Created an automated batch data analysis pipeline leveraging Apache Spark (in Java) and Azure Data Factory (with a C# custom activity), reducing execution time from about 5 hours to 30 minutes.

EDUCATION

University of Waterloo

Waterloo, ON

BASc Candidate – 3A Computer Engineering

Expected Apr 2020

- Cumulative Average: 93.42% | Overall GPA: 3.98/4 | 2B Class Rank: 1/164
- Awards – Engineering Faculty/Staff Upper Year Scholarship, Microsoft Tuition Scholarship, Richard & Elizabeth Madter Engineering Entrance Scholarship, President's Scholarship of Distinction & Research Award
- Undergraduate Research Assistantships – contributed to projects within research groups focusing on distributed computing systems, as well as on LDPC stochastic decoding algorithms.

PROJECTS

Full list and details available at linkedin.com/in/shencarl or csheng8.github.io/#projects

Hockey Analytics with Spark

2018

- Leveraging Apache Spark (including Spark Streaming and Spark MLlib) for a variety of hockey analytics related projects using data parsed from NHL APIs, including a Redzone-like streaming analytics algorithm, and a machine learning projected team points model.

Spatial Data Visualization Tool

2017

- Created an interactive data visualization tool for dynamically filtering and aggregating spatial data on a map based on zoom level and map bounds by leveraging a Geohash spatial index within an Apache Druid data store.

Parallelized User Behavior Prediction Model

2016

- Ported components of a novel user behavior prediction model to a MapReduce & Hadoop framework to improve speed of the model using parallelized cluster computing.
- Secondary author of publication in Cluster Computing 2017 Vol 20 Iss 2 (DOI 10.1007/s10586-017-0749-z)

Smart Bed Monitoring System

2015

- Built a system to recognize bed-related situations and monitor sleep patterns using a decision tree generated by applying a Weka machine learning algorithm on an experimentally gathered training data set.
- Received awards at the 2015 national Canada Wide Science Fair.