

Jack Wang
Software Engineering
github.com/jj22wang | jj22wang@uwaterloo.ca

PROFESSIONAL EXPERIENCE

AUGUST 2016 - PRESENT

Yelp

Data Quality Engineer | python, sklearn, Kafka, SQL

- Worked to source and display accurate data to customers
- Investigated and fixed Yelp business data errors
- Designed a service to utilize machine learning algorithms to improve business location accuracy
- Improved on user survey scoring algorithms to reduce false positives when calculating business categories
- Updated merge logic and added extensive test coverage to handle duplicate business removal

JANUARY 2016 - APRIL 2016

Capital One

Data Scientist | python, pySpark, Hive, JavaScript, sklearn

- Developed web applications for data visualization using Javascript and D3
- Adapted and optimized existing scikit-learn machine learning algorithms to service data applications
- Designed a parallelized batching pipeline using pySpark and Hive, to precalculate and cache expensive machine learning processes
- Migrated web applications to be servable from a Docker environment, reducing deployment complexity on AWS

MAY 2015 - AUGUST 2015

Loyalty One

R&D Engineer | SCALA, Spark, HDFS, Kafka, elasticsearch

- Developed a distributed computing API using the Spark framework, to eliminate boilerplate code for feature generation and data obfuscation
- Parallelized existing data analytics processes, improving process times by up to 100 times
- Implemented automated testing using Jenkins
- Executed large scale performance testing on over 80TB of testing data, to optimize Hadoop chunk sizes, data structures and transformation processes
- Prototyped an end to end data streaming pipeline using Spark Streaming, Kafka and elasticsearch
- Provided API support, debugging and code assistance for data science specialists

EDUCATION AND AWARDS

2014 – 2019 Candidate for B.SE.

The University of Waterloo

2010 – 2014 OSSD, Gifted

Nothern Secondary School

COMPUTER SKILLS

Languages python, SCALA, JavaScript, JAVA, C++

Frameworks Spark, pySpark, sklearn, SciPy, Hive, HDFS, SQL, elasticsearch, kafka, Gradle, D3

Environemnts Windows, Linux, OSX, Bash, AWS, Git, Docker

SIDE PROJECTS

Tagger | python, Discogs API

A python script to automatically tag song attributes using existing tags and filename

- Researched previous work and papers on this topic done by Sony
- Designed an algorithm to chunk and weight the different segments of a filename
- Created a module to query the Discogs Database API in order to scrape information from existing file information available

MeetMe | JavaScript, jQuery, Google Maps API

A Google Maps service to find the closest meeting point between a set of starting locations

- Aggregated distance matrixes from API calls to determine if optimal meeting point is found
- Designed an algorithm to repeated query different locations to balance travel times
- Utilized geometric distances and previous queries to update new search areas

MyoGuitar | JAVA, Lua, Android API

A Guitar playing app to utilize the Myo armband

- Utilized Android's developer API to implement multi-touch functionality
- Mapped contact point combinations to keys

FunctionSolver | HTML, CSS, JavaScript

A web application to interpret, reduce and graph mathematical equations

- Wrote a recursive chunking algorithm to parse and syntax check user input