Ankai Jie

W: ankaijie.com **P**: 647-965-4305

E: ankaijie@gmail.com

EDUCATION

UNIVERSITY OF WATERLOO
Sept 2015 — Present
Year 4

Bachelor of Computer Science, Major in Data Science with Business Option
Deans Honours List, 3.9 GPA
Marsh Memorial Scholarship, President's Scholarship, Research Scholarship

SKILLS

LANGUAGES

Python, Scala, Java, C++, SQL, Bash, R, Javascript, Scheme, HTML, CSS

DATA ANALYTICS

Hadoop, Hive, HBase, Spark, MapReduce, Kafka, Tez, Yarn, Apache Giraph, Redshift, Mysql, SQL Server, PostgreSQL

INFRASTRUCTURE

AWS — EC2, EMR, RDS, ECS, Azure, Docker, Ansible, Jenkins, Airflow

OTHER

SBT, Node.js, JQuery, D3.js AngularJS, Backbone.js, Git, Bootstrap, MongoDB, SASS, J2EE, Maven, Spring

EXPERIENCE

Jan 2018 — April 2018

SOFTWARE ENGINEERING
Coursera
Mountain View, CA, USA

- Created Scala job for major degree partners to dynamically regrade learners for updated assignments
- Developed REST resources and database logic in Scala for team-based assessment system used in Master and Undergraduate level degree courses
- Integrated Coursera platform with slack for learners using Kafka

May 2017 — Aug 2017

DATA ENGINEERING
The Meet Group
San Francisco, CA, USA

- Developed MapReduce jobs to approximate billion record distinct counts using KMV Sketchsets
- Created an Airflow data pipeline on Hadoop for time series data analysis on HBase and Redshift
- Integrated Python/Ansible to automate MapReduce jobs, ETLs, and application deployment on AWS

May 2016 — Aug 2016

DATA SCIENCE ENGINEERING
Scotiabank
Toronto, ON, Canada

- Developed Hadoop data pipeline for behavioural analysis on billion node temporal graphs
- Implemented all-pairs shortest paths in Java for time series data aggregation using Apache Giraph
- Automated data visualizations, ETLs with Python/Spark, and graph visualizations with D3.js

May 2015 — Aug 2015

WEB DEVELOPER
Dawn Infotek
Toronto, ON, Canada

- Built a robust data management system using AngularJS and J2EE with Spring on the backend

PROJECTS

JOB MARKET ANALYSIS
Python, R, Tableau

- Used generalized linear models and geo-data to identify key disparities in the American job market
- Best Visualization Award at Waterloo Datafest 2018 using Indeed.com data

CITATION ANALYZER
Research Project
Python, Celery, Flask, SQL

- Theory-crafted graph processing algorithms with graduate studies professor to identify influential and fraudulent researchers
- Developed pipeline in Python to ingest Scopus data and visualize million record datasets on the web

HEY KANYE!
Python, SQL Server, Azure Cognitive Services

- Rap song generator that uses Markov chains and parse trees to generate custom hip-hop lyrics
- Created an edge detection algorithm to sync and sing lyrics to a beat via text to speech
- Winner of Hack the North 2016, Microsoft Azure award

MISCALLANEOUS

AWARDS
- Waterloo Datafest 2018 Winner
- Hack the North 2016 Winner
- Microsoft Azure Cloud Award 2016
- Oracle Java SE 8 Certification
- Euclid Math contest Honour Roll

COURSEWORK
- Data Structures (Enriched)
- Computational Statistics and Data Analysis
- Logic and Computation

ACTVITIES
- Microsoft Student Partner, 2016 — Present
- Debate, Tennis, MMA