

Ankai Jie

M. ankaijie@gmail.com

P. 647-965-4305

W. ankaijie.com

L. ca.linkedin.com/in/ankaijie

Education

Candidate for B.CS

**Honors Computer Science, Major in
Data Science with Business Option**

University of Waterloo
3.9 GPA

Sept 2015 - Present

Skills Summary

Languages

Python, Java, SQL, Shell/Bash, R, Scala,
C++, Javascript

Database/Data Analytics

Hadoop, HBase, Hive, MySQL, SQL Server,
Postgres, MariaDB, Oracle, Redshift,
Apache Giraph, Spark

Cloud

AWS Configuration/Automation, EC2,
EMR, RDS, ECS, Boto3, Azure, Bluenote

Web Development

AngularJS, Backbone.js, J2EE, Node.js,
JQuery, D3, HTML, SASS, CSS, Bootstrap,
Spring, MVC, MongoDB, Grunt, AJAX

Other

Git, Maven, Docker, Jenkins, Ansible,
Linux

Awards

Winner, Hack the North, 2016

Top 10 hack from over 250 teams

Cloud Award, Microsoft, 2016

Best use of Azure and Machine Learning

University of Waterloo, 2011—2015

Deans Honour List, 2015, 2016

Honour Roll for Pascal, Cayley, Fermat,
Euclid math contests

Marsh Scholarship, Marsh LLC, 2015

Distinction in computer science and
mathematics

Extracurriculars

Microsoft Student Partner, 2017

Debate, Tennis, MMA

Work Experience

Data Engineer / The Meet Group

San Francisco, California

May 2017 - August 2017

- Developed Hive UDFs/UDAFs in Java to assist billion record distinct count approximations with sketchsets and statistical extrapolation, on Mapreduce/Tez
- Created scalable Airflow DAGs to execute custom Redshift queries for analysis of time series data and store results to HBase
- Integrated Python and Ansible to automate custom deployments of MapReduce jobs, ETL transactions, REST services, and NodeJS applications on AWS
- Developed Python library to manage Amazon EMR, ECS, EC2, RDS resources
- Built REST endpoints in Scala to process and deliver HBase and MariaDB data for visualizations and analysis on the web

Data Science Engineer / Scotiabank

Toronto, Ontario

May 2016 - August 2016

- Researched and developed process on Hadoop using Java, SQL, R to analyze billion node temporal graphs and predict consumer behavior
- Implemented all pairs shortest path in Apache Giraph for time series data aggregation and other graph processing algorithms in Java
- Automated data visualization and built ETL transactions with Python/Spark
- Created mechanism for graph visualizations on the web with NodeJS and D3.js

Web and J2EE Developer / Dawn Infotek

Toronto, Ontario

May 2015 - August 2015

- Built a robust J2EE data management system using AngularJS, Spring MVC
- Implemented RESTful web service interfaces using Java
- Created and automated test cases for end-to-end testing

Projects — github.com/AnkaiJie

Academic Fraudalyzer — Python • Celery • Flask • SQL

- Developed custom data pipeline in Python/SQL to ingest Scopus Academic data
- Theory-crafted graph processing algorithms with associate dean of Computer Science to highlight fraudulent authors in academia from citation data
- Visualizes million record datasets in graphs/tables through Flask web application

HeyKanye! — Python • SQL Server • Azure Cognitive Services

- Rap song generator that used Markov chains and parse trees to generate hip hop lyrics using lyrics of existing songs
- Created a custom edge detection algorithm to sync and sing lyrics to a beat
- Top 10 of 250 teams at Hack the North, Microsoft Award (Azure/Machine Learning)

Rogue 5 — C++ • NCurses

- Remake of traditional Rogue game with sentient AI and interactive combat