Wang Chen

wang1117chen@gmail.com | +1 206 504 8440 | 15835 14th NE Rd, Bellevue, WA, 98008

EDUCATION

University of Washington Seattle, WA

- Master of Science in Information Management, specialized in Data Science & Engineering
- Selected Coursework: Machine Learning, Interactive Information Visualization (**D3.js**), Software Architecture (**TypeScript**), Full-stack Software Development (**Node.js, React, AngularJS, MongoDB**), Advanced Database Internals and Design (**SQL, NoSQL**)
- Teaching Assistant of INFX562 (PG) Interactive Information Visualization (**D3.js, Tableau**)

09/2016 – 06/2018

SUMMARY OF SKILLS

- Programming languages: **Python, JavaScript, SQL**, Java, Scala, HTML5, CSS3, TypeScript
- Frameworks and tools: **Hadoop, Spark, AWS Certified Solutions Architect, Hive**, Druid, Pig, Node.js, Kafka, Cassandra (NoSQL)

INDUSTRY EXPERIENCES

Discovery, Inc Seattle, WA

Big Data Engineer *01/2020 – present*

- Designed and deployed large-scale, high performance data platforms with **AWS** cloud infrastructure to drive business decisions and improve customer engagement (**Terraform, Jenkins**)
- Built automatic and scalable streaming/batch data ETL pipelines (**Kafka, Spark, Docker**) to collect and process data sets
- Implemented **A/B testing** pipelines and platforms which enables fast feedback loops and more data-informed decision-making process
- Built and maintained **recommendation engine** with streaming/batch data processing (**Flink, Spark**) that increased engagement by 50%
- Integrated alternative open-sourced real-time/historical analytics database (**Apache Druid**) and data visualization tool (**Superset**) that realized high-performance and flexible data explorations
- Increased efficiency by 40% by designing and building self-serve tools (**AWS**) that enable non-tech users to deploy light ETL tasks

Ernst & Young Seattle, WA

Data Engineering Consultant (Clients: Microsoft, Nike, T-Mobile) *06/2018 – 01/2020*

- Created, maintained and documented a set of web/mobile analytics metrics and automated ETL data pipelines that better captured day-to-day user behaviors (**Streaming/Batch**) and helped make short/long-term business decisions
- Designed and implemented scalable data warehouses with Cloud infrastructure (**AWS, Azure**) that supports analytical solutions over product performance (**Redshift, S3, EC2, T-SQL, Python, Hive**)
- Built Data Analytics/Machine Learning platform (**Spark**) and developed a web-based data exploratory interface (**Node.js, D3.js**) that allows clients to explore data freely and make data-driven decisions more efficiently
- Tell data stories that describe analytical to diverse audiences (C-level communications)

Ernst & Young Seattle, WA

Technology Consulting Intern, Data Engineering & Analytics (Client: PSE) *06/2017 – 08/2017*

- Designed data models and created reusable automated ETL tools using **Python and Excel (VBA)** to collect data from multiple sources
- Developed and maintained an interactive web-based reporting dashboard (**React, D3.js**) that presented workflows and key indicators
- Communicated with end-users, gathered business requirements, and delivered data analysis reports in professional presentations

Amazon Lab126 Shenzhen, China

Business Analyst Intern *05/2016 – 09/2016*

- Used **MySQL** to develop an internal database for HR department to support their daily operations including recruiting and training
- Built reporting dashboards and generated weekly performance reports and monthly newsletter for HR department with **Tableau**
- Maintained and automated reports generation from database by writing **SQL** and **VBA** scripts, increasing efficiency by almost 50%

PROJECT EXPERIENCES

Broadband Data Analysis and Visualization Capstone – Data Visualization Seattle, WA

Explored broadband usage data and built interactive visualization dashboards to help government make decisions *12/2017 – 06/2018*

- Designed and developed a scalable data warehouse using star schema that successfully handled petabyte-level data
- Automated data ingestion flows (**Python, Spark**) and loaded aggregated data into HDFS for exploratory analysis and reports (**Hive**)
- Visualized data in an interactive web-based portal and created interactive infographics with **Node.js, D3.js and Tableau**

Social Media Data Mining Capstone – Data Science & Engineering Seattle, WA

Developed ingestion pipelines and data processing platform for real-time social media sentiment analysis *01/2017 – 06/2017*

- Built a high-performance data infrastructure using **Apache Kafka**, **Spark**, and **Cassandra** to prepare training dataset
- Used **Python** packages (**sklearn**) to build **Naïve Bayes Classifier** based on sentiment-labeled data that reaches 97% accuracy
- Conducted **regression analysis** between social media sentiment scores and stock performances across different industries
- Developed a web-based dashboard that visualizes real-time streaming data and model fitting results with **Node.js** and **D3.js**