Runhan Yu

(781)786-1586 • runhanyu90@gmail.com
runhanyu.com •linkedin.com/in/runhanyu • github.com/ziyunch

SKILLS

Programming languages	Python, SQL, Java, R, Matlab, Lisp, Javascript, Bash
AWS technologies	IAM, EC2, Redshift, Glue, Athena, EMR, Lambda, S3, RDS, DynamoDB
Distributed computing tools	Spark, Airflow, Kafka, Hive, Hadoop, MapReduce
Others	Linux, git, Django, Emacs, Tableau, Vim, UNIX, CSS, HPC

EXPERIENCE

Rescale San Francisco, CA
Data Engineer *Oct. 2019 - Present*

- Developed and enhanced new features, APIs and services for platform (Python, Django, Java).
- Built data metrics collection system from scratch and manage data warehouse (Redshift, Glue).
- Prepared Dataclips to aggregate platform usage data for reporting and decision (SQL, Athena).

Insight Data Science New York, NY
Data Engineer Fellow –rxminer.net *Jan. 2019 – Sep. 2019*

- Developed an ETL pipeline to extract, integrate and transform prescription data from multiple providers in AWS cloud computing to enable nation-wide queries on prescription drug usage (Python, AWS, Airflow).
- Validated and combined public available Medicaid and Medicare datasets with NIH, FDA and NPPES sources into a SQL queryable databases in Redshift, visualized in website (SQL, JavaScript, CSS).
- Implemented custom connector to Redshift/PostgreSQL with 20 times more efficiency (Python).
- Built a real-time monitoring pipeline of IoT sensor data and latencies for data center management to handle 10,000+ events per second (Spark streaming, Kafka).

Brandeis University Waltham, MA
Graduate Research Assistant *Oct. 2012 – Dec. 2018*

- Developed programs to parse massive experiment data into structured analysis and visualization (R).
- Built models to tackle metrics and simulate enzyme kinetic mechanisms (Matlab).
- Developed a python-based pipeline that extracted 100,000 gene sequences encoding protein of interest from 200 million gene sequences in the 114 GB GenBank database from RESTful API (Python).
- Managed 24 graduate and undergraduate students in biochemistry research lab and chemistry teaching lab.
- Collaborated with 3 teams with multi-disciplines and multi-cultures.

PROJECTS

Ledger-like Plain Accounting App convertor

- Provide smooth transition from accounting app to ledger-like plain text accounting tools.
- Convert .csv file exported from Sui accounting app to a Beancount file (Python).
- Wrote a lisp function to quickly add transactions in Emacs (Lisp).

Multithread Web Crawler

- Write a class to handle multithreading website crawling inside the given domain.
- Feature a breath-first search algorithm and a multithread pool to visit all links asynchronously.
- Handle various status code, time-out and exceptions in a structured manner (Python).

Stomach Cancer Gene Variation - bit.ly/2sBSHTW

- Extracted cancer research data from the Cancer Genome Atlas Network (R).
- Mapped gene list with biological annotations with functional annotation tool DAVID (Python).
- Summarized and visualized stomach cancer-related gene candidates with R package MAFtools (R).

Finalist for 4th Tufts New England Case Competition

- Provided business case solutions to reduce readmission rates and costs and increase profits for the client.
- Developed a collaborative comprehensive care package for acute myocardial infarction for Boston Scientific.

EDUCATION

Brandeis University Waltham, MA
Ph.D. in Chemistry, with specialization in Quantitative Biology *May. 2019*

Nankai University Tianjin, China
B.S. in Materials Chemistry *Jun. 2012*
B.S. in Finance *Jun. 2012*