Phi Bya

hungbya@gmail.com | phibya.com | linkedin.com/in/phibya | github.com/phibya | 470-708-2689

SUMMARY: Software Engineer with 6+ years of experience building scalable data-intensive applications. Proficient in full-stack development and performance optimization. Proven track record of delivering production systems handling large number of users and terabytes of data.

TECHNICAL SKILLS

- Languages: JavaScript/TypeScript, Python, R, C++, SQL, HTML/CSS
- Backend: Node.js, REST APIs, GraphQL, Microservices
- Frontend: React.js, D3.js, Cytoscape.js, Echarts, Zustand, WebSockets
- Databases: MongoDB, PostgreSQL, MySQL, Elasticsearch
- Cloud & DevOps: Docker, Singularity, Google Cloud Platform, JetBrains TeamCity CI/CD, Coder, Terraform
- Frameworks: Meteor.js, Astro.js, PyTorch, Pandas, Polars, scikit-learn, dplyr

PROFESSIONAL EXPERIENCE

Research Assistant Professor | Auburn University | Auburn, AL May 2024 – Present

- Lead data science and engineering initiatives for 5 PhD students and cross-functional research teams
- Architected and deployed HPC infrastructure (768 CPU cores, 16 GPUs, 6TB RAM) processing 50TB+ genomic data
- Reduced researcher onboarding time by 80% (1 week to 2 days) through automated environment setup and comprehensive documentation

Graduate Research Assistant | Auburn University | Auburn, AL August 2023 – May 2024

- Built end-to-end data analysis and machine learning pipeline in Python and R for single-cell genomic analysis processing 1M+ cells
- Optimized dimension reduction algorithms (PCA, t-SNE, UMAP) for single-cell data with millions of cells
- Developed interactive visualization dashboard using React.js and D3.js serving 100+ concurrent users
- Deploy deep learning models for cell classification and trajectory inference with 95% accuracy

Graduate Research/Teaching Assistant | University of Nevada, Reno | Reno, NV January 2018 – August 2023

- Taught R programming and statistical methods to 50+ students across diverse backgrounds (biology, biochemistry & computer science).
- Designed, implemented and maintained protein knowledge databases with 3M+ records using MongoDB.
- Built NLP pipeline using Python to crawl and index 200M+ scientific articles in Elasticsearch
- Developed entity recognition models for detecting gene-disease-organism relationships.

Software Engineer | YouthDev Co., Ltd | Ho Chi Minh City, Vietnam May 2016 – December 2017

- Engineered real-time trending detection system processing 40M+ daily content items
- Implemented high-performance scoring algorithm for trending detection reducing latency by 60%
- Built recommendation engine using collaborative filtering serving 100K+ users

Research and Development Engineer | YouNet Group | Ho Chi Minh City, Vietnam April 2014 – May 2016

- Developed sentiment analysis models and APIs using Naive Bayes achieving 87% accuracy
- Built user attribute prediction system using Random Forest, Collaborative Filtering, and Matrix Factorization for 50M+ social media profiles with 82% precision

EDUCATION

Ph.D., Computer Science, Auburn University — May 2024
M.Sc., Computer Science, University of Nevada, Reno — May 2022
B.Eng., Information Systems, UIT, Vietnam National University — May 2016

KEY PROJECTS

CareerScan: A minimalist and distraction-free job search engine https://careerscan.io

- Built web crawling system collecting real-time job postings from LinkedIn and multiple job platforms
- Implemented Elasticsearch-powered search engine indexing 35M+ job postings with advanced filtering and full-text search
- Created data pipeline for deduplication and normalization of job listings
- Tech stack: React.js, Echarts.js, Meteor.js, Node.js, Elasticsearch, MongoDB, Docker

Analyze and visualize large-scale single-cell data https://cytoanalyst.com

- Architected end-to-end data processing pipeline handling millions of data points with distributed computing
- Built real-time interactive dashboard with WebGL-accelerated for smooth rendering of large datasets
- Optimized data streaming and chunking for visualization of datasets exceeding available memory
- Implemented WebSocket-based collaboration system supporting concurrent multi-user analysis sessions
- Tech stack: R, Python, React.js, Echart.js, D3.js, Meteor.js, WebSockets, Node.js, MongoDB, Docker

ThYme: A comprehensive database of the Thioester-active enzYmes https://thyme.tinnguyen-lab.com

- Designed and implemented database system managing 3M+ enzyme records and 3.2B+ biological entities with sub-second query performance
- Built automated data ingestion pipeline with microservices architecture for database updating
- Developed high-performance APIs (REST & GraphQL) with caching layer for optimal response times
- Implemented full-text search capabilities and complex query optimization for scientific data retrieval
- Tech stack: React.js, Meteor.js, Apollo GraphQL, MongoDB, Docker, Redis

Intelligent platform for systems-level analysis https://ipsa.tinnguyen-lab.com

- Developed high-performance analysis engine with LLM integration for automated data interpretation and insights generation supporting 500+ organisms with 1000+ monthly analyses
- Optimized algorithm performance reducing computation time from hours to minutes
- Tech stack: R, Python, React.js, Echart.js, D3.js, Meteor.js, WebSockets, Node.js, MongoDB, Docker

FAT-PTM: A database for post-translational modification data https://fat-ptm.tinnguyen-lab.com

- Built curated databases for protein post-translational modification managing 600+ pathway networks and 49K+ protein modifications with complex relational queries
- Developed interactive graph visualization interface for real-time biological network editing and analysis
- Tech stack: React.js, Cytoscape.js, D3.js, Meteor.js, Node.js, MongoDB, Docker

DeconBenchmark – Cellular deconvolution framework https://github.com/tinnlab/DeconBenchmark

- Standardized and containerized 50+ algorithms into a unified framework.
- Tech stack: R, Python, Matlab, Docker

Cancer subtype discovery using multi-modal data integration https://cran.r-project.org/package=PINSPlus

- Developed novel clustering algorithms for data integration and patient stratification
- Published R package (PINSPlus) with 30,000+ downloads on CRAN

PUBLICATIONS

- Published 30+ peer-reviewed papers on method developments, data analysis, and software systems
- Google Scholar: https://scholar.google.com/citations?user=yw6kjSYAAAAJ