Nguyen Thanh Toan

Phone: 0869118227
Email: thanhtoan1742@gmail.com
Github: github.com/thanhtoan1742

ABOUT ME

Highly motivated Computer Science graduate with strong expertise in backend development, algorithm design, and high-performance computing using C++, CUDA, Rust, Go, and Python. Proven track record in developing specialized databases, AI/ML pipelines, and optimizing complex bioinformatic algorithms. Eager to leverage technical depth to solve challenging problems.

EDUCATION

- Ho Chi Minh City University of Technology (Bach Khoa TP.HCM).
- Computer Science Major.

ACADEMIC ACHIEVEMENTS

- 3rd prize in VietNam national Informatics contest (2018).
- Participated in several national and regional ACM-ICPC contests (national 2019, regional 2020, 2021, 2022).
- Codeforces max-rating 1900 (candidate master).
- TOEIC L&R 960/990, TOEIC S&W 320/400.
- Ho Chi Minh City University of Technology, 2019-2023, GPA: 8.01/10

EXPERIENCE

- Programming Languages: C++, C, CUDA, Rust, Go, Python
- HPC & Data Science: GPU Optimization, PyTorch, Numerical Methods, Statistical Analysis
- Backend & Systems: REST, gRPC, SQL
- DevOps & Tools: Docker (including CUDA support), Containerization, Git, CI/CD Streamlining

PROJECTS

Bioturing - Rdstream *(Sep-2025 - Now)*

- Rdstream is a Rust-Python library that converts RDS files to HDF5/H5AD files.
- Single cells analysis is done by many scientist in R and stored as RDS file (R's storage file format). RDS can only easily read and work on by R. This library convert RDS to a more standard HDF5 format.
- Architected and optimized data pipeline by implementing a streaming conversion process (RDS to HDF5/H5AD) in Rust-Python, successfully controlling peak memory usage for large bioinformatic datasets.
- Convert Seurat project to Anndata H5AD file format.

Bioturing - Exen *(June-2025 - Now)*

• Exen is a specialized single-cell database in C++ and CUDA capable of processing queries on 500M+ cells.
• The database is implemented and optimized in C++ and CUDA.
• Implemented many statistical methods such as differential expressed genes (T-Test and FDR correction), effect-size calculation (Cohen's D).
• Reduced numerical error in statistical calculations (mean/variance) crucial for accurate differential gene expression analysis.
• Implemented common expression optimization to reduce time on marker genes queries.

Bioturing - HIFED *(Dec-2024 - Feb-2025)*

• HIFED is a train and inference pipeline for histology image features extraction and dowstream processing.
• The pipeline is optimized for maximum GPU utilization and reasonable memory usage. Implemented in pytorch.
• Implemented the first iteration which includes building data pipeline, testing models and optimize pre-processing, inference and post-processing steps.

Bioturing - New-Luin *(May-2024 - Now)*

• New-luin is a Python library implemented many algorithms used in analyzing single cell and spatial data (similar to scanpy, cuml, cugraph).
• Algorithms are implemented in C++ and CUDA for maximum performance.
• Implemented polygon matching algorithms, fixed radius near neighbors on 2D space, linkage clustering algorithm.
• Streamline build and deployment process for multiple CUDA versions and Python platform.

OPSWAT - Metadefender ICAP Server *(Sept 2023 - May-2024)*

• Metadefender ICAP Server is a server that receive files and help do malware detection, data loss prevention and many cybersecurity related task to the files.
• Developed the integration with OPSWAT online scan service Metadefender Cloud.
• Developed custom block page solution for infected content.
• Resolved many technical debts and problems related to database interaction.

Teko - Tekone Inventory Service *(May 2022 - Oct 2022)*

• Inventory Service is a microservice responsible for managing what is available and where they are.
• Developed incoming warehouse feature to handle incoming order for omni-channel retail platform.
• Developed many watch-dog to clean up inventory data (clean up faulty stock, clean up unused warehouse) for omni-channel retail platform.
• Resolved many technical debts presented in the in-house developed reliable asynchronous messaging system.