Martin L. Putra

CONTACT	Crerar 219, 5730 S Ellis Ave Website: https://www.martinputra.com
	Email: martinluttap@uchicago.edu Others: LinkedIn, GitHub

RESEARCH INTERESTS

Areas: Systems Support for Genomics, Cluster Resource Management, Operating Systems, Storage Systems
Focus: Design and build novel genomics data processing systems to achieve
1. Performance: data- & task- parallelism for increased throughput and cluster utilization
2. Efficiency: scheduling based on novel workload characterization for improved time- and cost-efficiency
3. Scalability: reliable execution on large-scale fleet of (possibly transient) hybrid cloud machines

EDUCATION

University of Chicago — Chicago, IL
Ph.D in Computer Science (*M.S. conferred in 2022*) — 2020 - Present
Advisors: Robert L. Grossman, Haryadi S. Gunawi
Bandung Institute of Technology (ITB) — Bandung, Indonesia
B.Eng. in Computer Science — 2015-2019

EXPERIENCES

Center for Translational Data Science (CTDS), Chicago, IL — 2020 - Present
Working to optimize Genomics Data Commons (GDC)'s pipeline automation system by:
1. *Characterizing Workload*: I analyzed a **large-scale, 2-years-long production trace** to characterize GDC's workload and **build prediction models** for GDC's genomics workflows execution time.
2. *Building a Hybrid Cloud Simulator*: I integrated results from (1) into a high fidelity simulator which supports investigation of novel **high-throughput & cost-efficient scheduling policies**.
3. *Migrating into State-of-Art Workflow Engine*: I built an **semi-automatic translator** to migrate existing CWL production workflows into Nextflow [C1]. Experimental results showed up to **72% speedup of exec. time**.
4. *Investigating Novel Workload Features*: My extensive communication with domain experts suggests that there is an **understudied workload feature(s)** that can lead to **tail latency**.

CONFERENCE PUBLICATIONS

Bibliometrics on Google Scholar and DBLP

BIBE'23   [C1] **Martin L. Putra**, In Kee Kim, Haryadi S. Gunawi, Robert L. Grossman. CNT: Semi-Automatic Translation from CWL to Nextflow for Genomic Workflows. In The 23rd IEEE International Conference on BioInformatics and BioEngineering (BIBE), 2023.
Artifact available upon request.

SOSP'21   [C2] Huaicheng Li, **Martin L. Putra**, Ronald Shi, Xing Lin, Gregory R. Ganger, Haryadi S. Gunawi. IODA: A Host/Device Co-Design for Strong Predictability Contract on Modern Flash Storage. In the Proceedings of the 28th Symposium on Operating Systems Principles (SOSP), 2021.
Pre-accepted sans discussion at the PC meeting; Artifact Available, Functional and Results Reproduced

JOURNAL PUBLICATIONS

TOS'23   [J1] Huaicheng Li, **Martin L. Putra**, Ronald Shi, Fadhil I. Kurnia, Xing Lin, Jaeyoung Do, Achmad I. Kistijantoro, Gre- gory R. Ganger, Haryadi S. Gunawi. Extending and Programming the NVMe I/O Determinism Interface for Flash Arrays. ACM Transactions on Storage (TOS), Volume 19, Issue 1, February 2023. [Extended version of C2]

TALKS   **CNT: Semi-Automatic Translation from CWL to Nextflow for Genomic Workflows.**
Conference Talk, BIBE'23, Online — 2023

TECHNICAL SKILLS	**Programming Languages**: Bash, C, C++, Python, JavaScript, Groovy
	Data Science Toolkit: NumPy, Pandas, Scikit-learn, Matplotlib, Gnuplot
	Testbed / Cloud Platform: OpenStack, ChameleonCloud, Emulab, AWS, Azure Cloud, GCP
	Systems Toolkit & Experiences: UNIX Tools, QEMU, FEMU, Linux Kernel (storage & md subsystems)
	Bioinformatics Toolkit: CWL, Nextflow, bwa, STAR, FastQC, Picard, GATK

SERVICE

Secondary / Sub-reviewer

ASPLOS'24: The 29th ACM Intl' Conf. on Architectural Support for PL and OS	2023
OSDI'24: The 18th USENIX Symposium on Operating Systems Design and Implementation	2023
FAST'24: The 22nd USENIX Conference on File and Storage Technologies	2023

Artefact Evaluation Committee (AEC)

ATC'23: The 2023 USENIX Annual Technical Conference	2023

Other Activities

UC Santa Cruz's Summer of Reproducibility Project Co-lead	2023
UChicago's Data Science Institute (DSI) Summer Lab Mentor	2022

Last update: December 21, 2023