JIE LI

Research Assistant, Department of Computer Science, Texas Tech University

Email: jie.li@ttu.edu | Phone: (806)787-8258

Homepage: https://lijie.me

EDUCATION

- **Ph.D.**, Computer Science, Texas Tech University, Lubbock, Texas   Expected May 2024
- **M.S.**, Computer Science, Texas Tech University, Lubbock, Texas   2019
- **B.Arch.**, Architecture, Huaqiao University, Xiamen, China   2012

TECHNICAL STRENGTHS

Programming Languages:	C, C++, Python, Bash
Operating Systems:	UNIX/Linux, Mac OS
Databases:	MySQL, TimescaleDB (PostgreSQL), InfluxDB
Cloud Computing:	AWS & Azure Experience, Docker
Machine Learning Libraries:	Numpy, Pandas, Scikitlearn, Tensorflow

WORK EXPERIENCE

Data-Intensive Scalable Computing Laboratory (DISCL), TTU   **2019 - Present**
Research Assistant   *Lubbock, TX*

- **Research and Publication:** Conducted innovative research in High-Performance Computing, Computer Architecture, and Parallel and Distributed Computing. Authored and published research papers in reputable academic conferences and journals.
- **Professional Development and Networking:** Actively engaged in the academic community by attending conferences, workshops, and seminars. Presented research papers and posters at these events.
- **Mentorship and Education:** Mentored both graduate and undergraduate students in their independent research studies. Provided guidance on research topics, project development, and data analysis.
- **Software Development and Collaboration:** Played an integral role in developing and maintaining research software and tools. Wrote, tested, and documented code for various projects. Contributed to open-source software initiatives, fostering collaborative innovation.
- **Server Administration:** Managed two high-end servers (Hugo and Alita) hosted at the High-Performance Computing Center. Oversaw server configuration, maintenance, and software management. Ensured consistent server availability and reliability while troubleshooting issues as they arose.

Lawrence Berkeley National Laboratory (LBNL)   **2021, 2022, 2023 (Summer)**
Graduate Student Intern   *Berkeley, CA*

- **Data Integration and Analysis:** Integrated HPC monitoring data from diverse sources (LDMS, DCGM, Slurm, VictoriaMetrics) for comprehensive analysis of system-wide architectural efficiency, including CPU, GPU, DRAM, and HBM2 resource utilization. Identified critical trends and patterns within the data to drive insights into system performance, with a focus on NERSC's Cori and Perlmutter.
- **Machine Learning Expertise:** Conducted in-depth statistical analysis of job-level monitoring data. Applied a variety of machine learning models, including SVC, LinearSVC, Decision Tree, and Random Forests, to analyze jobs based on time-series features.

- **Innovative Data Processing:** Pioneered a novel approach by encoding time-series monitoring data as images. Achieved high accuracy through the training of a Convolutional Neural Network (CNN) to classify and predict job applications.
- **Simulation and System Design:** Designed and implemented a discrete event simulator to study resource management and job scheduling in HPC systems, with a specific focus on systems with disaggregated memory resources.

Teaching, Learning and Professional Development Center (TLPDC), TTU   **2018-2019**
Graduate Student Programmer   *Lubbock, TX*

- **Website Maintenance and Communication:** Maintained and updated TLPDC web pages, ensuring a fresh and relevant online presence. Facilitated communication with software application providers to meet product requirements efficiently.
- **Database Management and Security:** Managed the MySQL database with precision, safeguarding valuable data assets. Implemented robust backup strategies to protect against data loss. Proactively addressed and resolved database access issues to maintain uninterrupted operations.

SELECTED PROJECTS

High-Performance Computing System Health Monitoring & Performance Data Collection   **2021**
https://github.com/nsfcac/MonSter

- Explored mechanisms to acquire health status monitoring data from an HPC cluster via the Integrated Dell Remote Access Controller (iDRAC), enhancing cluster management and efficiency.
- Spearheaded the development of a suite of tools for automating iDRAC telemetry report configuration, metric analysis, and TimescaleDB table initialization. Efficiently handled diverse data sources and types, streamlining data processing and analysis.
- Designed and implemented a robust system monitoring infrastructure capable of asynchronous collection of health status data through the Redfish API and job accounting data via the Slurm REST API.
- The Slurm data collection code has been adopted and merged into Dell's Omnia project for broader industry utilization (Github link: https://github.com/dell/omnia).

Monitoring Data Management and Query Performance Optimization   **2020**
https://github.com/nsfcac/MetricsBuilder

- Investigated and identified performance bottlenecks in InfluxDB. Optimized the database schema design, resulting in a remarkable **71.98%** reduction in data volume and a significant **1.76×** boost in query performance, enhancing data management efficiency.
- Designed and implemented a time series deduplication mechanism. It achieved an impressive average data volume reduction of **70.38%** and maintained data accuracy with a minimal error of only **0.74%** in reconstruction.
- Designed and developed MetricsBuilder, a data access accelerator. MetricsBuilder dramatically improved query performance by up to **25×** and reduced data transmission volume by **95%** compared to traditional SQL queries, streamlining data retrieval.
- Implemented an API using the OpenAPI specification. The API provided efficient data access services to data analysis consumers, including JavaScript data visualization applications and Grafana, ensuring seamless access to valuable insights.

PUBLICATIONS

[1] Jie Li, Rui Wang, Ghazanfar Ali, Tommy Dang, Alan Sill, and Yong Chen. Workload failure prediction for data centers. In *2023 IEEE 16th International Conference on Cloud Computing (CLOUD'23)*, pages 479–485, 2023c. doi: https://doi.org/10.1109/CLOUD60044.2023.00064

[2] Cristiano E. Caon, Jie Li, and Yong Chen. Effective management of time series data. In *2023 IEEE 16th International Conference on Cloud Computing (CLOUD'23)*, pages 408–414, 2023. doi: https://doi.org/10.1109/CLOUD60044.2023.00055

[3] Jie Li, George Michelogiannakis, Brandon Cook, Dulanya Cooray, and Yong Chen. Analyzing resource utilization in an hpc system: A case study of nersc's perlmutter. In *International Conference on High Performance Computing (ISC'23)*, pages 297–316. Springer, 2023b. doi: https://doi.org/10.1007/978-3-031-32041-5_16

[4] Jie Li, Brandon Cook, and Yong Chen. Arcode: Hpc application recognition through image-encoded monitoring data. *arXiv preprint arXiv:2301.08612*, 2023a. doi: https://doi.org/10.48550/arXiv.2301.08612

[5] Tommy Dang, Ngan VT Nguyen, Jie Li, Alan Sill, Jon Hass, and Yong Chen. Jobviewer: Graph-based visualization for monitoring high-performance computing system. In *2022 IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT'22)*, pages 110–119. IEEE, 2022. doi: https://doi.org/10.1109/BDCAT56447.2022.00021

[6] Tommy Dang, Ngan Nguyen, Jon Hass, Jie Li, Yong Chen, and Alan Sill. The gap between visualization research and visualization software in high-performance computing center. *The Gap between Visualization Research and Visualization Software (VisGap'21))*, 2021. doi: https://doi.org/10.2312/visgap.20211089

[7] Xi Wang, Antonino Tumeo, John D Leidel, Jie Li, and Yong Chen. Ham: Hotspot-aware manager for improving communications with 3d-stacked memory. *IEEE Transactions on Computers (IEEE Trans Comput)*, 70(6): 833–848, 2021. doi: https://doi.org/10.1109/TC.2021.3066982

[8] Jie Li, Ghazanfar Ali, Ngan Nguyen, Jon Hass, Alan Sill, Tommy Dang, and Yong Chen. Monster: an out-of-the-box monitoring tool for high performance computing systems. In *2020 IEEE International Conference on Cluster Computing (CLUSTER'20)*, pages 119–129. IEEE, 2020. doi: https://doi.org/10.1109/CLUSTER49012.2020.00022

[9] Ngan Nguyen, Jon Hass, Yong Chen, Jie Li, Alan Sill, and Tommy Dang. Radarviewer: visualizing the dynamics of multivariate data. In *Practice and Experience in Advanced Research Computing (PEARC'20)*, pages 555–556. 2020. doi: https://doi.org/10.1145/3311790.3404538

[10] Vung Pham, Ngan Nguyen, Jie Li, Jon Hass, Yong Chen, and Tommy Dang. Mtsad: Multivariate time series abnormality detection and visualization. In *2019 IEEE International Conference on Big Data (BigData'19)*, pages 3267–3276. IEEE, 2019. doi: https://doi.org/10.1109/BigData47090.2019.9006559

[11] Jie Li, Xi Wang, Antonino Tumeo, Brody Williams, John D Leidel, and Yong Chen. Pims: A lightweight processing-in-memory accelerator for stencil computations. In *Proceedings of the International Symposium on Memory Systems (MemSys'19)*, pages 41–52, 2019. doi: https://doi.org/10.1145/3357526.3357550

[12] Xi Wang, Antonino Tumeo, John D Leidel, Jie Li, and Yong Chen. Mac: Memory access coalescer for 3d-stacked memory. In *Proceedings of the 48th International Conference on Parallel Processing (ICPP'19)*, pages 1–10, 2019. doi: https://doi.org/10.1145/3337821.3337867

PRESENTATIONS

Conference Presentations

- Workload Failure Prediction for Data Centers, *CLOUD'23*   July 2023
- A Holistic View of Resource Utilization on Perlmutter (Poster), *SC'22*   Nov. 2022
- Advanced Visualization and Data Analysis of HPC Cluster and User Application Behavior, *SC'21*   Nov. 2021
- MonSTer: An Out-of-the-Box Monitoring Tool for HPC Systems , *CLUSTER'20*   Sept. 2020
- PIMS: A Lightweight Processing-In-Memory Accelerator for Stencil Computations, *MemSys'19*   Oct. 2019

Research Seminar Talks

- Towards Cycle-Accurate Simulation for xBGAS   Apr. 2023
- A Holistic View of Resource Utilization on Perlmutter   Aug. 2022
- SST and Cycle-accurate Simulation of xBGAS   May 2022
- Lightweight Checkpointing of Loop-Based Kernels Using Disaggregated Memory   Feb. 2022
- Deep Learning and Monitoring Metrics to Image Encoding for Detecting Applications in HPC systems Nov. 2021
- Detecting and Identifying Applications by Job Signatures   Sept. 2021
- Predicting Abnormal Workloads in HPC Systems   May 2021
- The IEEE Cluster2020 Experience, MonSTer Review and Future Work   Sept. 2020
- Monitoring Operating System Status on a Raspberry Pi cluster   July 2020
- Experiences of Storing and Querying Monitoring Data of Large-scale HPC Platforms Apr. 2020
- Collecting and Storing Telemetry Metrics from RedRaider Cluster   Apr. 2020
- Profiling Power Consumption of Jobs with SLURM   Jan. 2020
- Monitoring Power Usage of Jobs Running on Quanah Cluster   Sept. 2019

MENTORING EXPERIENCE

Undergraduate Students (including REU participants)

- Mentoring Yusheng Han and Zachary Kay on the topic "Running HPC Applications on the RedRaider Cluster and Analyzing Performance Behaviors". Independent Study (CS4000)   Spring 2022
- Mentoring Casey Root on the topic "Monitoring Queue Status via SLURM Rest API". Independent Study (CS4000)   Spring 2021

Graduate Students

- Mentoring Cristiano Caon on the topic "Investigating the Data Volume Reduction and Query Optimization in Time Series Databases". Outcomes include a conference publication in *CLOUD'23*. Independent Study (CS7000) and Master's Thesis   Fall 2022
- Mentoring Aniruddh Sanjaysinh Chavda and Huyen Nguyen on the topic "Usage Behavior Analysis with Clustering Job Accounting Data". Advanced Operating System (CS5379)   Spring 2021
- Mentoring Ruonan Wu on the topic "Job Accounting Data Analysis for Quanah Cluster ". Advanced Operating System (CS5379)   Spring 2021
- Mentoring Ashritha Puradamane Balachandra on the topic "Improving Query Performance of InfluxDB". Advanced Operating System (CS5379)   Spring 2020

SERVICES

Paper Reviewer

- The Journal of Supercomputing
- IEEE International Parallel and Distributed Processing Symposium (IPDPS'23)
- IEEE/ACM International Symposium on Cluster, Cloud and Internet Computing (CCGrid'22)
- IEEE International Conference on Distributed Computing Systems (ICDCS'22)
- The International Conference for High-Performance Computing, Networking, Storage, and Analysis (SC'22)
- International Parallel Data Systems Workshop (PDSW'22)
- IEEE International Conference on Big Data (BigData'20, BigData'21, BigData'22)
- IEEE International Conference on Smart Data Services (SDMS'20)

Volunteer

- Student volunteer of SC'21, St. Louis, Missouri   2021
- Student volunteer of SC'19, Denver, Colorado   2019
- Volunteer of Paul's Project - Grace Campus, Lubbock, Texas   2019

AWARDS AND HONORS

- Best Poster Award, NSF Cloud and Autonomic Computing Industry Advisory Board Conference   2022
- Summer Thesis/Dissertation Research Award ($2300), Lubbock, Texas   2019

AWARDS AND HONORS

- Best Poster Award, NSF Cloud and Autonomic Computing Industry Advisory Board Conference   2022
- Summer Thesis/Dissertation Research Award ($2300), Lubbock, Texas   2019