SULTAN MAHMUD SAJAL

+1-814-380-3595 • smsajal116@gmail.com • LinkedIn Profile • Google Scholar • smsajal.github.io

BRIEF BIOGRAPHY

PhD graduate in Computer Science with experience at NetApp, Meta and Microsoft Research, specializing in cloud infrastructure, large scale systems performance, and power sufficiency and power budget optimization. Proven record of impactful research (EuroSys '24 Runner-Up Paper Award, OSDI 2023 publication) and industry contributions resulting in $100M+ annual savings.

WORK EXPERIENCE

NetApp, Inc. (San Jose, CA) - Performance Engineer 05/01/2026 - Present

- Working on optimizing performance of AIDE (AI Data Engine), a distributed platform for ingesting, cataloging, and serving enterprise storage data for AI workloads.

Meta Platforms, Inc. (Menlo Park, CA) - Performance and Capacity Engineer 05/06/2024 - 02/01/2026

- Optimized rack power budgets for multiple AI inference hardware programs, resulting in $100M/year in data center OpEx savings and establishing a standardized budget optimization framework across Meta's AI infrastructure hardware programs.
- Drove Cross-functional collaboration to increase power limit of GPUs used by recommendation services, leading to 7% reduction in GPU capacity required without any increase in power budget, resulting in $400k/year benefit.
- Collaborated with Cross-functional teams to deliver power budget for Flash storage racks delivering 7.3EB storage to support GenAI training.
- Collaborated with Cross-functional teams to deliver power budget for Flash storage racks delivering 9.9EB storage to support MySQL and other services.
- Redesigned the Power Sufficiency Simulator pipeline, increasing input accuracy and reducing Thrift service dependencies by 77%.
- Created a methodology for machine-level power estimation method for power capped machines, improving power capping efficiency by 11% or reducing overage risk by 0.15%, depending on the scenario.
- Developed a technique to estimate power usage at the power device level from host-level metrics, achieved 99%+ accuracy in eliminating 0.9%–6% estimation error across power device types.
- Improved idle host power usage estimation accuracy by up to 31%, significantly enhancing estimates in rack-level budget planning.

Microsoft Research, Microsoft Corporation (Redmond, WA) - Research Intern 05/31/2022 - 08/26/2022

- Designed and implemented novel admission control techniques for Microsoft Azure to maximize resource utilization while meeting strict SLA requirements.
- Achieved over 95% resource utilization and maintained 99.9%+ availability in a production cloud environment.

Gray Systems Lab, Microsoft Corporation (Redmond, WA) - Research Intern 05/24/2021 - 08/20/2021

- Developed realistic benchmarks for Azure HDInsight by reconstructing Spark workloads from real-world query traces.
- Generated synthetic datasets and workloads from user workload statistics while preserving privacy guarantees.

The Pennsylvania State University (State College, PA) - Graduate Assistant 08/20/2018 - 05/05/2024

- Contributed to research projects as a Research Assistant, resulting in publications at EuroSys 2024 and 2021.
- Served as a Teaching Assistant for multiple undergraduate courses, leading recitations, holding office hours, and grading assignments and projects.

REVE Systems (Dhaka, Bangladesh) - Junior Software Engineer 10/02/2017 - 07/10/2018

- Built a prototype machine translation system supporting Bangla and 25 additional languages.
- Developed a Bangla NLP platform prototype hosting applications such as a machine translator, spell checker, and more.

EDUCATION

Ph.D., Computer Science and Engineering 08/21/2018 - 08/10/2024
The Pennsylvania State University, PA, USA

B.Sc., Computer Science and Engineering 02/16/2013 - 09/13/2017
Bangladesh University of Engineering and Technology, Dhaka, Bangladesh

TECHNICAL SKILLS & SELECTED COURSES

Languages: Python, Java, C++, Scala, R
Frameworks/Tools: Pandas, Apache Spark, Docker, Kubernetes, Git
Cloud Platforms: AWS, Azure
Web & Devops: Nginx, Varnish, Memcached
Databases: MySQL, PostgreSQL, Presto
Scripting/Markup: Bash, LaTeX, HTML/CSS
Specialized Areas: Distributed Systems, Systems Benchmarking, Power Budget Optimization, Performance Analysis

RESEARCH EXPERIENCE

The Pennsylvania State University (University Park, PA) - Graduate Assistant 08/21/2018 - 07/31/2024

- **Ph.D. Thesis:** Improving the Fidelity of Trace-Driven Experiments in Cloud Computing Systems.
- **Project:** Facilitate Isolated Experimentation for Reproducible Results in Cloud Systems
  - Motivate the need for isolated experimentation for realistic and reproducible experiments through realistic experiments.
  - Develop methodology and tool for orchestrating isolated experiments.
- **Project:** Upscale Workloads from Cloud Infrastructure and Large Datacenters
  - Developed novel upscaling techniques for real workloads to enable faithful systems experimentation under varying loads.
  - Winner of Best Runner Up paper in EuroSys 2024.
- **Project:** Downscale Workloads from Cloud Infrastructure and Large Datacenters
  - Developed novel techniques to downscale cloud workloads while preserving important characteristics such as arrival process and performance to facilitate realistic systems research and industry prototyping.
  - Published in EuroSys 2021.

Microsoft Research, Microsoft Corporation (Redmond, WA) - Research Intern 05/31/2022 - 08/26/2022

- **Project:** Efficient and Scalable Cloud Admission Control in Azure
  - Developed novel admissions control system, Kerveros, that guarantees SLAs for both allocated and reserved resources while maximizing resource efficiency.
  - After deployement in Azure, Kerveros achieved 99.9%+ availability while achieving 95%+ resource utilization.
  - Deployed in Microsoft Azure, and published as a paper in OSDI 2023.

Gray Systems Lab, Microsoft Corporation (Redmond, WA) - Research Intern 05/24/2021 - 08/20/2021

- **Project:** Development of Flight Simulator for Spark Jobs
  - Created realistic benchmarks for Spark workloads from query traces and generate synthetic representative datasets for the benchmark.
  - Developed workload analyzer and synthetic data generator in Apache Spark.

PUBLICATIONS

- **Sultan Mahmud Sajal**, Md Salman Estyak, Rubaba Hasan, T Zhu, B Urgaonkar, S Sen. "**TraceScaler: A Framework for Scaling Load in Real-World Traces for System Evaluation.**". ACM Transactions on Computer Systems. *(Invited Paper, Under Review.)*
- **Sultan Mahmud Sajal**. "**Improving the Fidelity of Trace-Driven Experiments in Cloud Computing Systems.**". PhD Dissertation, The Pennsylvania State University, (2024).
- **Sultan Mahmud Sajal**, T Zhu, B Urgaonkar, S Sen."**TraceUpscaler: Upscaling Traces to Evaluate Systems at High Load.**". Nineteenth European Conference on Computer Systems, (Eurosys 2024).
- **Sultan Mahmud Sajal**, L Marshall, B Li, S Zhou, A Pan, K Mellou, D Narayanan, T Zhu, D Dion, T Moscibroda, I Menache. "**Kerveros: Efficient and Scalable Cloud Admission Control.**". 17th USENIX Symposium on Operating Systems Design and Implementation, (OSDI 2023).
- **Sultan Mahmud Sajal***, R Hasan*, T Zhu, B Urgaonkar, S Sen. "**TraceSplitter: A New Paradigm for Downscaling Traces.**". Proceedings of the Sixteenth European Conference on Computer Systems, (EuroSys 2021).**Equal Contribution*

HONORS AND AWARDS

- **Runner Up Paper Award**, European Conference on Computer Systems, (EuroSys) 2024
- **Student Grant**, European Conference on Computer Systems, (EuroSys) 2024
- **Student Grant**, USENIX Symposium on Operating Systems Design and Implementation, (OSDI) 2023
- **Registration Grant**, ACM SIGMETRICS 2021
- **Student Grant**, USENIX Symposium on Operating Systems Design and Implementation, (OSDI) 2020
- **Student Travel Grant**, ACM Symposium on Cloud Computing (SoCC) 2019
- **Student Travel Grant**, ACM Symposium on Cloud Computing (SoCC) 2018
- **Technical Scholarship**, Bangladesh University of Engineering and Technology 2018

REFERENCES

Available upon request.