YEASIR RAYHAN

Ph.D. Candidate in Computer Science, Purdue University
yrayhan@purdue.edu
yeasirrayhanprince.github.io · Google Scholar · LinkedIn · GitHub

RESEARCH SUMMARY

My research leverages machine learning and modern hardware to build fast and adaptive database systems across heterogeneous hardware and workloads. My work spans learned query optimization across row- and column-oriented DBMSs with CPU and GPU execution engines, learned hardware-aware scheduling on NUMA and chiplet-based servers, and buffer management for CXL-era tiered memory. More recently, I have been building database agents by designing harnesses that enable LLMs to autonomously interact with and optimize database systems.

EDUCATION

Purdue University - Graduate **August, 2021 – December, 2027 (expected)**
Ph.D. Candidate in Computer Science
- Tentative Title: Generative Database Agents
- Advisor: Walid G. Aref
- Collaborators: Ryan Marcus

Bangladesh University of Engineering and Technology (BUET) - Undergraduate **2014 – 2018**
B.Sc. in Computer Science and Engineering
- Advisor: Tanzima Hashem
- Thesis: Efficient Scheduling of Generalized Group Trips in Road Networks

SELECTED PUBLICATIONS

1. Gen-DBA: Generative Database Agents. **Yeasir Rayhan** and Walid G. Aref. *Preprint, ArXiv, 2026.*
2. P-MOSS: Scheduling Main-Memory Indexes Over NUMA Servers Using Next Token Prediction. **Yeasir Rayhan** and Walid G. Aref. *SIGMOD 2026.*
3. Virtual-Memory Assisted Buffer Management in Tiered Memory. **Yeasir Rayhan** and Walid G. Aref. *DaMoN Workshop, SIGMOD 2026.*
4. Exploring Next Token Prediction for Optimizing Databases. **Yeasir Rayhan** and Walid G. Aref. *aiDM Workshop, SIGMOD 2025.*
5. GTX: A Write-Optimized Latch-free Graph Data System with Transactional Support. Libin Zhou, Lu Xing, **Yeasir Rayhan** and Walid G. Aref. *SIGMOD 2025.*
6. SIMD-ified R-tree Query Processing and Optimization. **Yeasir Rayhan** and Walid G. Aref. *SIGSPATIAL 2023.*
7. An Efficient Approach for Indoor Facility Location Selection. **Yeasir Rayhan**, Tanzima Hashem, Muhammad Aamir Cheema, Hua Lu, and Mohammed Eunus Ali. *EDBT 2023.*
8. AIST: An Interpretable Attention-Based Deep Learning Model for Crime Prediction. **Yeasir Rayhan** and Tanzima Hashem. *ACM TSAS 2023.*

RESEARCH EXPERIENCE

Gen-DBA: Generative Database Agents 2026
Graduate Researcher, Purdue University *ArXiv 2026*

- Built Gen-DBA, a general-purpose database agent that pairs an LLM with a database harness to optimize across learning tasks, database engines, hardware platforms, data and query workloads, and optimization objectives.
- Post-trained an LLM and developed a database harness that enables the LLM to interact with DBMSs, observe and collect runtime statistics, and optimize query plans and physical database designs.
- Evaluated Gen-DBA across three database optimization tasks, achieving a 3.5× improvement in tail query latency on JOB, a 2.48× improvement in DSB workload latency for storage layout optimization, and a 54× improvement in tail cardinality-estimation Q-error on IMDb over PostgreSQL.

Buffer Pool Management for Tiered-Memory Database Systems 2025
Graduate Researcher, Purdue University *DaMoN 2026, ArXiv 2025*

- Built vmcache^n, an n-tier virtual-memory-assisted buffer pool spanning DRAM, CXL memory, and disk, keeping each page's virtual address fixed while migrating its physical frame across memory tiers.

- Developed `move_pages2`, a workload-aware Linux kernel system call that gives the DBMS control over page migration through partial-page migration, configurable migration modes, and batch sizes.
- Achieved up to 4× higher TPC-C throughput than vmcache, a state-of-the-art two-tier buffer pool. The `move_pages2` system call achieved up to 2.3× higher query throughput and 2.6× higher page-migration throughput than Linux's native `move_pages` interface.

P-MOSS: Learned Scheduling of Main-Memory Indexes on NUMA and Chiplet Servers 2024–2025
Graduate Researcher, Purdue University *SIGMOD 2026, aiDM 2025*

- Built P-MOSS, a learned spatial query scheduler that places index partitions across NUMA nodes and assigns queries to CPU cores to maximize throughput on NUMA and chiplet-based servers.
- Formulated scheduling as next-token prediction using hardware PMU statistics. Trained a Decision Transformer with offline RL across hardware and workloads, followed by post-training on the target server.
- Evaluated P-MOSS on six hardware configurations from Intel, AMD, NVIDIA (ARM), and IBM using YCSB and OSM workloads, achieving up to 6× higher query throughput than state-of-the-art scheduling policies.

SIMD-ified R-tree Query Processing and Optimization 2022–2023
Graduate Researcher, Purdue University *SIGSPATIAL 2023*

- Built a SIMD-accelerated R-tree to accelerate spatial range selection and spatial joins using AVX-512 SIMD instructions.
- Designed two SIMD-friendly R-tree node layouts and introduced five query-processing optimizations, including replacing recursive tree traversal with a SIMD queue, software prefetching to hide cache misses, and vectorized binary search to prune bounding-box comparisons during spatial joins.
- Evaluated the SIMD-accelerated R-tree on spatial range selection and join workloads, achieving up to 4× lower range-selection latency and 9× lower spatial-join latency than scalar implementations.

AI-Native Database Operators May 2023 – June 2023
Database Software Engineer Intern, ITTIADB *Bellevue, WA*

- Prototyped user-defined functions (UDFs) in ITTIA DB that invoke PyTorch and ONNX models directly from SQL queries, enabling in-database machine learning inference without moving data out of the DBMS.

AIST: Interpretable Spatio-Temporal Crime Prediction January 2021 – August 2021
Research Assistant, BUET, Dhaka, Bangladesh (Advisor: Tanzima Hashem) *ACM TSAS 2023*

- Developed AIST, an interpretable attention-based deep learning model for spatio-temporal crime prediction that learns from past crimes, traffic flow, points of interest, and recurring crime trends.
- Modeled spatial and temporal dependencies using graph neural networks and attention mechanisms.
- Evaluated AIST on the Chicago 2019 crime dataset, reducing MAE by 4.1% and MSE by 7.45% over prior methods.
- Co-authored the proposal "Interpretable Spatio-Temporal Deep Learning Models for Safe Cities," funded by the ICT Division, Bangladesh (2020).

TEACHING

Purdue University August 2023 – Present
Graduate Teaching Assistant *West Lafayette, IN*

- CS 34800: Information Systems; CS 44800: Introduction to Relational Database Systems.

SKILLS

Programming and OS Kernels: C/C++, Python, Java, SQL; Linux kernel development, Linux system calls, Virtual memory, NUMA page migration.
Database Systems: PostgreSQL, DuckDB, Sirius, PostgreSQL extensions, Query optimization, Database indexing, Multi-tier buffer pool management, Query scheduling.
Hardware and Performance: NUMA, SIMD, Performance Monitoring Unit (PMU), CXL tiered memory, Heterogeneous CPU/GPU execution.
LLM and ML Systems: Hugging Face Transformers, vLLM, SGLang; SFT, GRPO, LoRA/rsLoRA, On-policy distillation, Constrained decoding, Tool-using agents, PyTorch, PyTorch Geometric.

AWARDS AND HONORS

Purdue Graduate Teaching Award: Purdue University, 2025.
CS Graduate Student Association Travel Grant: Purdue University, for SIGMOD 2025.
ACM SIGSPATIAL Travel Award: SIGSPATIAL 2023.
The Global Undergraduate Awards: Regional Winner (Asia), Computer Science, 2019.