Research Training for UTA Undergraduate Students

Current Members

Project A: Optimizing Search Tree:

Ye, Dongqing

Ye, Dongchen

Kharel, Aabhas

Kale, Nirmik Milind

Project B: Optimizing Replacement Algorithms

Farahanipad, Farnaz

Vadgama, Siddharth Kiranbhai

Lamsal, Safal

Jain, Mehul

Hinkel, George L & Vivek Patel

Research Projects

- "SDC: A Software Defined Cache Supporting Flexible Key-value-style Data Caching" Lead: Fan Ni
- "*Wormhole*: A Fast Ordered Index for In-memory Data Management" Lead: Xingbo Wu
- "RASI: A Road-Aware Spatial Index for Intelligent Transportation Systems" Lead: Xingsheng Zhao
- "OC-Cache: An Open-Channel SSD Based Cache for Virtual Machine Systems" Lead: Haitao Wang

Data Storage Research: Opportunities

- Byte-addressable, high-capacity, and high-speed **non-volatile memory** (NVM), such as Intel and Micron's 3D XPoint (NVM), has been on production or is expected to be widely deployed in the next 5-10 years, reminiscent of adoption of flash-based SSDs in the last 15 years.
- Computing power and network capability are added into storage devices making **processing in storage devices** attached on the network (e.g., NVMe over Fabrics) a serious option for large-scale cost-effective distributed data processing.
- Storage devices are becoming increasingly diverse on storage media, performance characteristics, durability, and cost. This **heterogeneity** must be well managed at one server and a distributed system.

Data Storage Research: Challenges and Opportunities

- With NVM's expected much higher density than DRAM, the NVM at each server can grow very large and hit "**memory capacity wall**". A standalone NVM blade may be provided to provide a large pool of shared storage in a disaggregated datacenter architecture.
- With highly distributed storage and support of high-speed networks, it is a promising and challenging effort to build a **global address space** for high space utilization and easy programming.
- With a sea of data streaming in from numerous IoT devices, management cost, in particular, in terms of time and space costs of metadata, can be staggering. **O(1) management algorithms** must be developed.
- With ever-faster storage devices, performance bottlenecks are shifting to the software, specially on the **I/O stack in the OS** and distributed file or storage systems.

An Envisioned Storage Model

- Traditional **block address space** for performance-insensitive services, such as data archiving, on the disks;
- **Object-based key space** for easy uses by applications, such as convenient encapsulation and sharing of data and code, and development of scalable database systems;
- **Byte-addressable NVM memory space** to support immediate persistency of program data. Issues such as data placement, serialization, and consistency need to be carefully studied for a high-performance, cost-effective, and easy-to use storage system.

Summer Schedule

- Core members:
  - On-site meetings are required.
  - Commitment is expected.
  - Reasonable productivity is expected

- Participants:
  - Work at your own pace
  - No commitment is expected (don't need excuses for missing meetings or no progress)

- A meeting proposal will be announced at least one week before.
- Meeting schedule will accommodate constraints of core members.