MSL UR Program 2023

General Q&A

Q1 : Is it okay to submit multiple proposals for each topic/theme?
A1 : Yes, as long as it meets our requirements, you are welcome to submit multiple proposals. Also it's possible to combine two themes for one proposal.

Q2: Do you accept a joint proposal from two+ faculty members?
A2 : Yes. However, we will need separate budget plan from each faculty for two different research contracts.

ALL Project's requirements

- Up to 5 year program : Specific **Milestone** has to be proposed
- University PI has to describe **ultimate goal**, clear outcome and deliverables
- **Objectives and Key Result(OKRs)** has to be proposed
- **Top level architecture description** is required
- Research contract : Research result will be reviewed in June/October and there can be early termination if it doesn't meet MSL's requirements
- Minimum 2 Student researchers. Samsung summer internship is highly preferred
- Expecting high quality publications(Minimum 2) within three years
- Monthly Sync-up meeting and quarterly project status report and One year-end report.
- Fully justify all performance claims with thoughtful theoretical predictions and/or experimental outcome
- Include quantitative projections for performance improvement that are tied to representative values included in authoritative publications or in comparison to existing products
- Propose a tightly structured project which includes technical and project that demonstrates clear progress, are aggressive but achievable, and are quantitative
- Clearly define the merit of the proposed innovation compared to competing approaches and the anticipated outcome

<Template>

1. Statement of Problem & Significance of Research
    - Define the problem your proposal will solve
2. Project Plan & Suggested Solution (Problem aligned)
    - **Solution's guidelines** : How will you accomplish it?
    - Research Design & Methods
    - Research plan and technical Approach
    - Project Milestones : up to 5 years
        - Year 1 : 1Q~4Q Plan
        - Year 2 : 1Q~4Q Plan
        - Year 3~5 : High-level plan can be proposed. (Research agreement term will be renewed every two years with updated Milestones)
    - Potential Obstacles
3. Expected outcomes and result (Tangible / Intangible)
    - **Objectives and Key Results** (OKRs)
    - Solution's proposed application, use case, value, effect
4. Budget details : minimum 2 year($300k) ~ up to $1M for 5Year program
    - Including Overhead cost(Less than 55%), researcher's salary, benefit
    - Researchers should be invited for MSL summer intern program
    - Colocation server for testing Samsung Device can be supported MSL

Topic 1: Programming model for computational Storage

Abstract

Computational storage is defined as an architecture that increases the efficiency of the overall system by reducing the movement of data or accelerating functions by providing Computational Storage Functions (CSFs) attached to the storage, thereby reducing the burden on the host. Since the concept was introduced in the late 1990s in the Active Disk paper, system improvements have been evaluated through a variety of applications closely integrated with computational storage devices (CSDs). One of the main challenges of this concept is the lack of a programming model for developing applications universally such as OpenCL and CUDA in the GPU realm. Although standards bodies such as NVMe and SNIA are working to define instruction sets and APIs, they are still very early as a programming model. This Funding Opportunity Announcement (FOA) pursues new research and development in computational storage programming models. The technology must show how the proposed method will allow users to easily develop applications and how it can be integrated with the existing software such as databases, machine learning frameworks and storage. Applications and/or instructions can be found accompanying this announcement on the Samsung MSL webpage.

Requirements

All applications under this topic must:

- Propose a tightly structured project which includes technical and project that demonstrates clear progress, are aggressive but achievable, and are quantitative
- Clearly define the merit of the proposed innovation compared to competing approaches and the anticipated outcome.
- Be consistent with and have performance metrics (whenever possible) linked to published, authoritative analyses in the respective technology space.
- Include quantitative projections for performance improvement that are tied to representative values included in authoritative publications or in comparison to existing products
- Fully justify all performance claims with thoughtful theoretical predictions and/or experimental outcome.

Expected Outcome

- Publication describing the value proposition of research and possible real-life applications
- Working demo with Computational Storage / Smart SSD, final and intermediary at every quarter, meeting the requirements in listed in the above section
- Working simulated/emulated demo for real life use case

Question & Answer on proposal Theme

Q1: Is this topic for truly general-purpose programming model? Or if you are also interested in domain-specific ones as well?

A1: This topic is not about a programming model for only computational storage. This proposal is about what would be the ***best programming model*** for computational storage in the long term. Depending on the assumed architecture, a general purpose programming model or domain-specific one can be used. For example, it can be an extended version of the existing programming model such as SYCL, OpenMP, DPC++, or OpenCL. The proposed model can be used for other architectures like GPU, TPU, or DPU. Actually, it would be better. However, it must demonstrate that unique requirements of computational storage can be modeled well using the proposed programming model.

Topic 2: Impact of deep memory hierarchy on computer architecture and software system

Abstract

The traditional memory hierarchy consisting of CPU Cache, DRAM, and storage is showing inefficiencies with the advent of new applications such as Big Data and ML (Machine Learning). Various techniques have been proposed to respond to these needs. For example, Intel's latest CPUs have the option to integrate High Performance Memory (HBM); Compute Express Link (CXL) enables memory configurations with complex topologies with different media and latencies; and off-chip last-level caches overcome high latency of disaggregation and distribution. Such new technologies are making computer system memory hierarchies increasingly complex. Some of them have latency ranges not assumed by traditional CPUs, so latencies can increase from nanoseconds to microseconds. Some of these may require software changes to fully exploit the potential of the new technology. CPU innovations that support new memory hierarchy are also encouraged. This Funding Opportunity Announcement (FOA) pursues new research and development in computer architectures that can accommodate such deep memory hierarchy. The technique must show how memory hierarchies should be defined in order to demonstrate clear performance benefits at the application level. Applications and/or instructions can be found accompanying this announcement on the Samsung MSL webpage.

Requirements

All applications under this topic must:

- Propose a tightly structured project which includes technical and project that demonstrates clear progress, are aggressive but achievable, and are quantitative
- Clearly define the merit of the proposed innovation compared to competing approaches and the anticipated outcome.
- Be consistent with and have performance metrics (whenever possible) linked to published, authoritative analyses in the respective technology space.
- Include quantitative projections for performance improvement that are tied to representative values included in authoritative publications or in comparison to existing products
- Fully justify all performance claims with thoughtful theoretical predictions and/or experimental outcome.

Expected Outcome

- Publication describing the value proposition of research and possible real-life applications
- Working demo with Computational Storage / Smart SSD, final and intermediary at every quarter, meeting the requirements in listed in the above section
- Working simulated/emulated demo for real life use case

Topic 3: Rack-scale computing architecture for net zero carbon

Abstract

Data centers consume 1% of the world's electricity in 2020 and have remained nearly constant since 2010, thanks to a variety of technologies including virtualization, the cloud, and Moore's Law. However, the Paris Agreement requires us to achieve the challenging target of reducing electricity use by 53% by 2030. Since more than 60% of the power in the data center is being used by servers, new technologies are needed to use power more efficiently at the rack level. However, Dennard scaling is no longer working and Moore's Law is slowing down. Also, most data centers are already cloud-based and most servers are already virtualized. In these circumstances, further power efficiency improvements are challenging. This funding opportunity announcement (FOA) pursues new research and development into rack-scale architectures and computing models that can deliver better power efficiency. This technique must show that the proposed architecture uses less power than traditional architectures with real-world applications without compromising performance. In addition, the proposed computing model must be general enough to cover a wide range of applications. Applications and/or instructions can be found on the Samsung MSL webpage accompanying this announcement.

Requirements

All applications under this topic must:

- Propose a tightly structured project which includes technical and project that demonstrates clear progress, are aggressive but achievable, and are quantitative
- Clearly define the merit of the proposed innovation compared to competing approaches and the anticipated outcome.
- Be consistent with and have performance metrics (whenever possible) linked to published, authoritative analyses in the respective technology space.
- Include quantitative projections for performance improvement that are tied to representative values included in authoritative publications or in comparison to existing products
- Fully justify all performance claims with thoughtful theoretical predictions and/or experimental outcome.

Expected Outcome

- Publication describing the value proposition of research and possible real-life applications
- Working demo with Computational Storage / Smart SSD, final and intermediary at every quarter, meeting the requirements in listed in the above section
- Working simulated/emulated demo for real life use case

Topic 4: Applications of near storage compute for Information-Centric Network (ICN)

Abstract

With the rapid development and need of near real time data analytics at the source there is a growing need of processing and inference data using near storage compute. This paradigm requires computation, data and application services in close proximity to end users. Future storage networking technologies will result in a fast response time, low latency and scalable data growth. Information-centric networking (ICN) is once such newly proposed future Internet paradigm in which communication is based on content names irrespective of their locations. At the same time, ICN promises efficient content delivery, mobility support, scalability, and security for content. The near data processing and ICN provide an opportunity to reduce latency, support security, scalability and solution simplicity for applications requiring near real time data analytics. Here we solicit research proposals on how computational storage (Smart SSDs) can be used in conjunction with ICN for real life solutions, with data centric approach.

Requirements

All applications under this topic must:

- Propose a tightly structured project which includes technical and project that demonstrates clear progress, are aggressive but achievable, and are quantitative
- Clearly define the merit of the proposed innovation compared to competing approaches and the anticipated outcome.
- Be consistent with and have performance metrics (whenever possible) linked to published, authoritative analyses in the respective technology space.
- Include quantitative projections for performance improvement that are tied to representative values included in authoritative publications or in comparison to existing products
- Fully justify all performance claims with thoughtful theoretical predictions and/or experimental outcome.

Expected Outcome

- Publication describing the value proposition of research and possible real-life applications
- Working demo with Computational Storage / Smart SSD, final and intermediary at every quarter, meeting the requirements in listed in the above section
- Working simulated/emulated demo for real life use case

Topic 5: Mitigating data loading and caching bottlenecks for large scale AI training acceleration

Abstract

Artificial Intelligence (AI) / Machine Learning (ML), specifically Deep Neural Networks (DNNs), is stressing storage systems in new ways, moving the training bottleneck to the data ingestion phase, rather than the actual learning phase. Training these models is data-hungry, resource-intensive, and time-consuming. It uses all of the resources in a server; storage, DRAM, and CPU for fetching, caching, and pre-processing the dataset (collectively called the input data pipeline) to the GPUs that perform computation on the transformed data. Additionally, there is an output data pipeline that periodically checkpoints the model state to persistent storage. It is becoming increasingly important to have Data Storage and Ingestion (DSI) pipeline backed by performant and scalable storage backend requiring smaller data center footprint. Here we solicit research proposal to mitigate data loading and caching bottlenecks for large scale AI trainings specifically using object storage.

Requirements

All applications under this topic must:

- Propose a tightly structured project which includes technical and project that demonstrates clear progress, are aggressive but achievable, and are quantitative
- Clearly define the merit of the proposed innovation compared to competing approaches and the anticipated outcome.
- Be consistent with and have performance metrics (whenever possible) linked to published, authoritative analyses in the respective technology space.
- Include quantitative projections for performance improvement that are tied to representative values included in authoritative publications or in comparison to existing products
- Fully justify all performance claims with thoughtful theoretical predictions and/or experimental outcome.

Expected Outcome

- Publication describing the value proposition of research and possible real-life applications
- Working demo of proposed approach and solution utilizing Samsung's high capacity SSDs
- Desired, but not required, that proposed research to utilize DSS OSS: https://github.com/OpenMPDK/DSS

Topic 6: Federated AI – How near data processing can help core challenges of federated learning

Abstract

Federated Learning enables devices/servers to collaboratively learn a shared prediction model while keeping all the training data on the device, decoupling the ability to do machine learning from the need to store the data in the cloud. Applications requiring lesser storage space, lower generic compute and less power can be greatly benefitted by FL. While FL solves the issue of space in terms of communication it uses up a larger bandwidth through the communication of model updates back and forth. This can lead to bottlenecks in the bandwidth for communication and poses latency challenge for disturbed computing. By bringing compute and data closer together these challenges can be mitigated e.g. utilizing computation storage. Here we solicit research proposal on how computational storage (Smart SSDs) can be used for real life Federate AI use cases and applications such as accelerated training, reduce data management, decentralized systems, digital forensic.

Requirements

All applications under this topic must:

- Propose a tightly structured project which includes technical and project that demonstrates clear progress, are aggressive but achievable, and are quantitative
- Clearly define the merit of the proposed innovation compared to competing approaches and the anticipated outcome.
- Be consistent with and have performance metrics (whenever possible) linked to published, authoritative analyses in the respective technology space.
- Include quantitative projections for performance improvement that are tied to representative values included in authoritative publications or in comparison to existing products
- Fully justify all performance claims with thoughtful theoretical predictions and/or experimental outcome.

Expected Outcome

- Publication describing the value proposition of research and possible real-life applications
- Working demo with Computational Storage / Smart SSD, final and intermediary at every quarter, meeting the requirements in listed in the above section
- Working simulated/emulated demo for real life use case

Important Deadlines

Inquiries on Theme: 5/31 (Tue)

- Based on Inquiries collected, MSL will post Answers by 6/10

First Draft : 6/24(Fri) – *optional. This will allow PI to elaborate proposals based on MSL's requirement*

Final Submission : 7/22 (Fri)

For inquiries, please contact ssimsl@samsung.com