Compute Everywhere: How Storage and Networking Expand the Compute Continuum

A SNIA Webcast Discussion Panel
LIVE: November 17, 2020 – 10:00 am PST
On Demand – snia.org/educational-library
Today’s Speakers

Moderator:
Jim Fister
Director, CMSI Applications Enabling
Principal, The Decision Place

Presenter:
Eli Tiomkin
NGD Systems
Chair, CMSI Computational Storage
Special Interest Group

Presenter:
Steve Adams
Intel

Presenter:
Chipalo Street
Principal Program Manager Lead
Microsoft, Azure IoT Edge
SNIA Legal Notice

The material contained in this presentation is copyrighted by the SNIA unless otherwise noted.

Member companies and individual members may use this material in presentations and literature under the following conditions:

- Any slide or slides used must be reproduced in their entirety without modification
- The SNIA must be acknowledged as the source of any material used in the body of any document containing material from these presentations.

This presentation is a project of the SNIA.

Neither the author nor the presenter is an attorney and nothing in this presentation is intended to be, or should be, construed as legal advice or an opinion of counsel. If you need legal advice or a legal opinion please contact your attorney.

The information presented herein represents the author’s personal opinion and current understanding of the relevant issues involved. The author, the presenter, and the SNIA do not assume any responsibility or liability for damages arising out of any reliance on or use of this information.

NO WARRANTIES, EXPRESS OR IMPLIED. USE AT YOUR OWN RISK.
SNIA-at-a-Glance

185 industry leading organizations

2,000 active contributing members

50,000 IT end users & storage pros worldwide
Traditional Enterprise Data States

Move/Transfer

Analyze/Decide

Store/Archive
Forward-Looking Enterprise Data States

- Compute everywhere to maximize data value
- Requires meaningful access to both data and data flow
Compute pendulum is swinging towards the middle

Latency, Cost, Sovereignty, Resilience, Capacity

Machine Time Decisions
Real-time Analysis
Local

Human Time Decisions
Postmortem Analysis
Global

Distributed
Centralized

Things
Premises
Metro/Network
Zones/Regions
Hyperscale DC

Physical & Mechanical world
Digital World

Physical & Mechanical world

Latency, Cost, Sovereignty, Resilience, Capacity
“Edge Compute” is the next sharing economy battleground

Distributed

Centralized

Hyperscale DC

Premises

Zones/Regions

Shared Skilled Labor

Shared Skilled Labor AND Shared Compute & Storage Resources

Inference

Inference & Grooming

Training

Metro/Network
The rise of the EDN* … the next big platform
The rise of the EDN* … the next big platform
The next decade: Composable E2E Platforms

- Scripts: invoke systems
- Queries: invoke insights
- Resources compose and data converges whenever, wherever, and for as long as needed
The next decade: Composable E2E Platforms

- Scripts: invoke systems
- Queries: invoke insights
- Resources compose and data converges whenever, wherever, and for as long as needed
The next decade: Composable E2E Platforms

- Scripts: invoke systems
- Queries: invoke insights
- Resources compose and data converges whenever, wherever, and for as long as needed
Azure meets the customer where they are
Azure meets the customer where they are
Azure meets the customer where they are

- Azure Stack HCI
- Azure Stack Hub
- Azure Stack Edge
- Metro / Zones
- Cloud

- Edge Servers
- Azure Native Edge Infrastructure
- Cloud
Azure meets the customer where they are

On prem edge  Metro / Zones  Cloud

Azure IoT Edge  Azure Stack HCI  Azure Stack Hub  Azure Stack Edge  Azure

Azure Stack Hub
Azure meets the customer where they are

On prem things
- Azure Sphere
- Azure RTOS
- Azure IoT Device SDK

On prem edge
- Azure IoT
- Azure IoT Edge

Metro / Zones
- Azure Stack HCI
- Azure Stack Edge

Cloud
- Azure Stack Hub
- Azure
Scaling requirements are not met with existing solutions
One CPU to many storage devices creates bottlenecks
These bottlenecks exist, we currently just shift where they reside

Technologies that ‘compose’ these elements just exacerbate the bottleneck

A way to augment and support without wholesale change is needed
Computational Storage View

- Computational Storage Function (CSF)
  - Send compute request to the drive
  - Allow drive to reduce data
  - Only return the results
  - Can be local or fabric attached
  - Reduces fabric and DDR BW consumption
  - Costs Saving
    - Reduced transfers
    - Reduced power
    - Free up host cycles
    - Potential for server removal
  - Potential for massively parallel compute

Computational Storage Systems
Computational Storage Devices

- **Computational Storage Drive (CSD):**
  A storage element that provides Computational Storage Function and persistent data storage.

- **Computational Storage Processor (CSP):**
  A component that provides Computational Storage Functions for an associated storage system without providing persistent data storage.

- **Computational Storage Array (CSA):**
  A collection of Computational Storage Devices, control software, and optional storage devices.
  (Many options here)
Using Computational Storage

Benefits

✔ Distributed Processing
✔ Faster Results
✔ Lower Power
✔ Smaller Footprint

Reduced data transfers
Reduced fabric provisioning

Scaling compute resources with storage provides access to results faster

Computational Storage resources ‘offload’ work from the overtasked CPU

Seamless architectures create new ‘servers’ with each storage device added

Additional CPU resources without added rack space
Problem Statement

- Databases growing at exponential rates

- Load and Search time key blocks in getting results

Computational Storage Solution

- Determine best way to increase performance
- Load Time Reductions due to CSD Offload of AI code

Results are Proven:

- Load Time Reduced > 95%
- Search Time Reduced > 60%
- Power Savings of > 60%

Technical paper to be published in the ACM journal on Computational Storage
Low latency actuation

High latency actuation

Microsoft Azure Edge OS running the application
With NGD’s Computational Storage Devices
Thanks for Watching Our Webcast

▪ Please rate this webcast and provide us with feedback

▪ A link to this webcast and the PDF of the slides are posted to the SNIA Compute Memory and Storage Initiative website at https://www.snia.org/forums/cmsi/knowledge/articles-presentations

▪ You can also find this webcast and many other videos and presentations on today’s topics in the SNIA Educational Library

▪ A Q&A from this webcast will be posted to the SNIA Compute, Memory, and Storage Blog
Where To Find Out More About Compute Everywhere and Computational Storage

- Website resources
  - www.snia.org/CMSI
- Twitter
  - @sniacomputational
- Blog
  - SNIAComputeMemory&Storage
- Videos
  - https://www.youtube.com/user/SNIAVideo/playlists
- Educational materials
  - https://www.snia.org/educational-library
- Joining SNIA and the Compute, Memory, and Storage Initiative
  - https://www.snia.org/member_com/join-SNIA