

Ceph Enhancements to Handle Computational Storage using IBM's Flash Core Modules

Presenters: Tim Fisher (fisher@us.ibm.com), Mohit Kapur (mohitk@us.ibm.com)

Contributors: Mohit Kapur, Martin Ohmacht, Ralph Bellofatto, Chris Ward, Tim Fisher, Bharat Sukhwani, Sameh Asaad

Thomas J Watson Research Center, IBM

FlashCore Module



Flash Memory Summit

FCM3 Offering →

FCM4 Coming Soon...

*2.5" Dual ported
NVMe SSD
U.2 Form Factor*

*Encryption with
FIPS 140-3 L2
Certification in
progress*

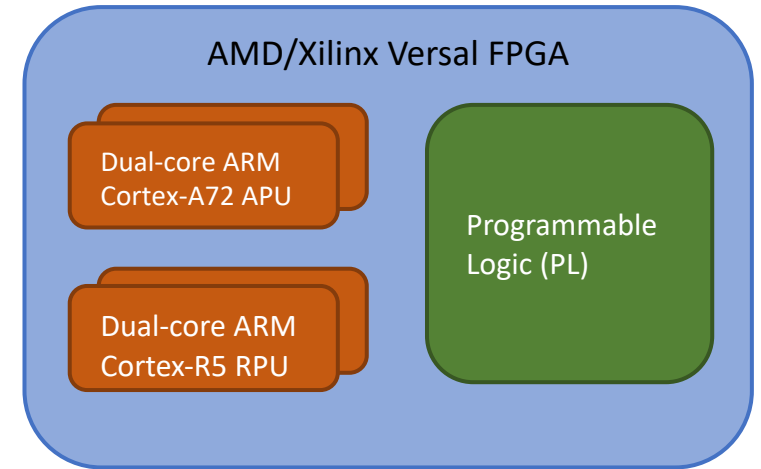
*Used exclusively
In IBM Storage
Appliances*



FCM Computational Storage

- Cutting Edge AMD / Xilinx FPGAs with embedded, powerful multi-core ARM processors
- Inline compression provides *supreme* system processing offload and system memory offload
- TCG Opal, AES-256 Encryption
- FTL processing and metadata contained completely inside FCM.
 - **Zero** system resources needed!
- Application Process Units (APUs) and Programmable Logic collect data for computation
- Real-time Processing Units (RPU) analyze collected

*FCM is positioned to lead and shape the evolution of computational storage in All Flash Arrays and **beyond**...*



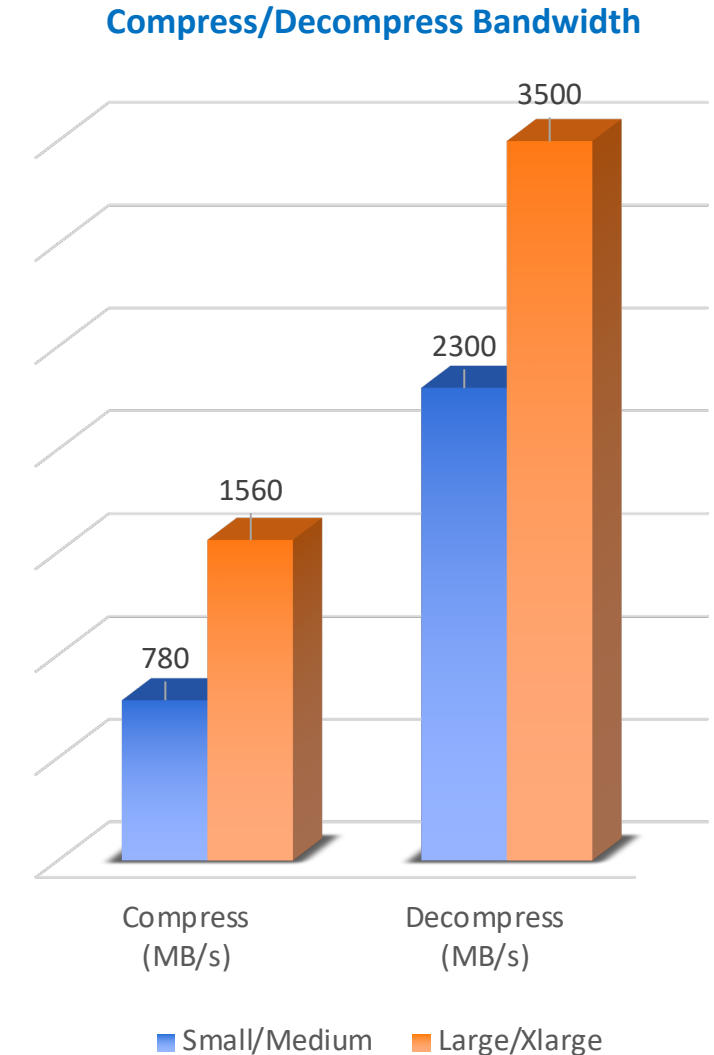
Hardware Compression



Flash Memory Summit

- Inline, At Speed, Hardware Compression
 - Shipping since 2017 in FlashSystems products
- Data is decompressed along side compression and checked bit by bit. Never any corruption!
- Takes advantage of already existing LSA mapping
- Decompression done in line with hardly any extra latency
- Data protection (ECC) is implemented on top of compressed data.
 - Allows garbage collection and other background data transactions to operate on compressed data
- Significant increase in appliance performance
- Significant reduction in appliance DRAM
- Compression and Decompression completely transparent above the Flash module except for management of space

Hardware compression inside the SSD is still one of the best examples of Computational Storage!



Ceph Enhancements to Handle Computational Storage using IBM's Flash Core Modules

Presenter: Mohit Kapur

Contributors: Mohit Kapur, Martin Ohmacht, Ralph Bellofatto, Chris Ward, Tim Fisher
Bharat Sukhwani, Sameh Asaad

Thomas J Watson Research Center, IBM

Climbing Mt. Computational offload



Flash Memory Summit

Higher System Performance

Data Privacy and Security

Reduce Latency

Real Time Analytics

Free CPU Cycles

Open Source Software Stack

Reconfigurable Hardware

Use case: **Compression**, Encryption, DB Query offload

CEPH 101



Flash Memory Summit

- Scalable
- Distributed Architecture
- Unified – Block, File, Object
- Self Healing, Resilient
- High Performance
- Open Source
- Cost Effective
- Integration and Ecosystem

CLIENT nodes run
Block Device over
libRADOS

SERVER nodes run
Object Storage
Daemon

**Reliable Autonomous Distributed
Object Store**



Handling Hardware Compression in Ceph

Compressible data

Logical space:
22TB

Physical
space: 9.6TB

Incompressible data

Logical
space: 22TB

Physical space:
9.6TB



Ceph plugin to handle hardware compression

Merged with Main Branch
17.2.4

Testing

Add Plugin to
Kernel Devices

Right
Capability Set

New Device
Interface on top of
VDO

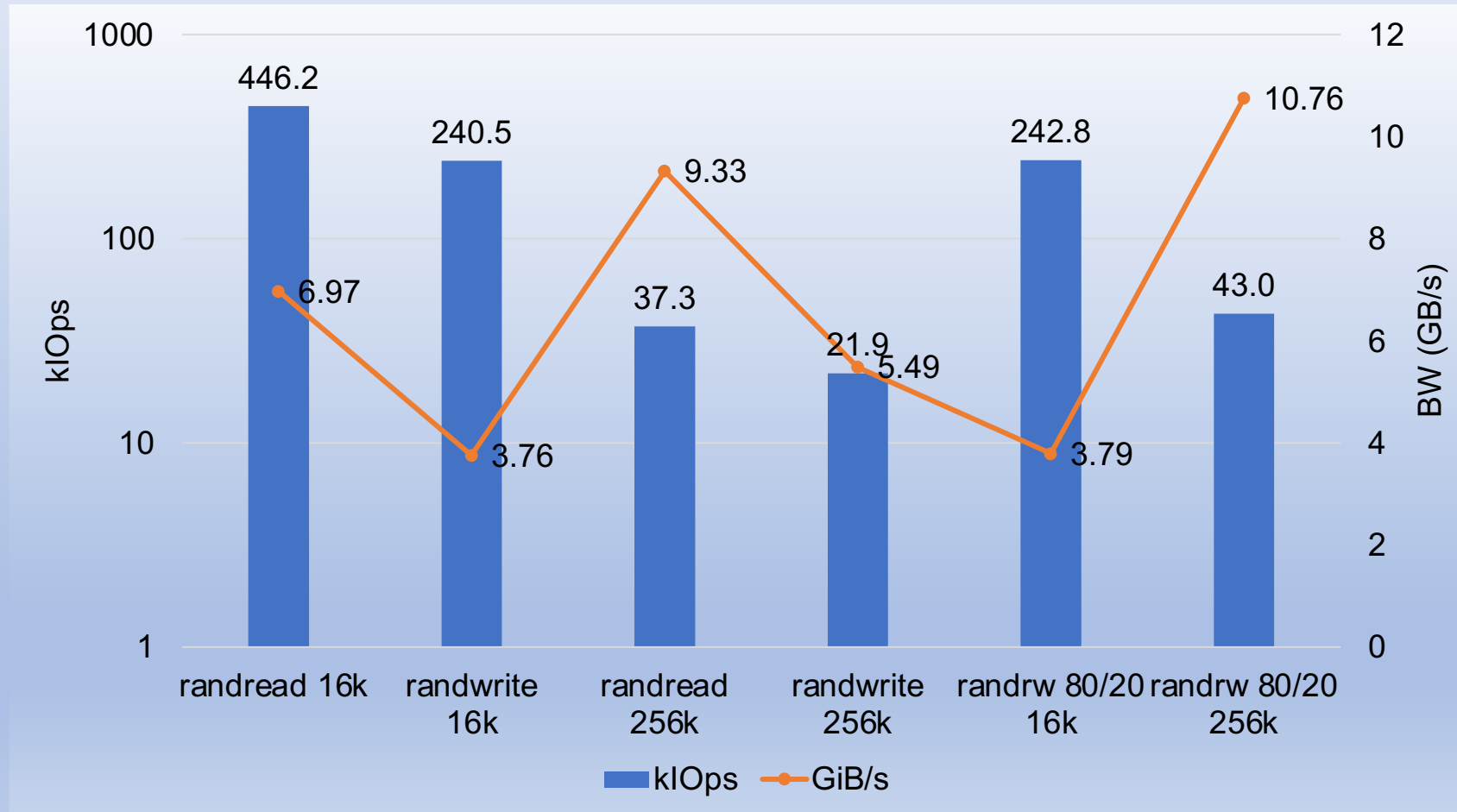
New Plugin

~ 15 files got
modified across the
stack



IOPs, Bandwidth across block sizes

6 Storage Nodes, 5 FCMs/Node, 2 Clients, 32 FIO threads reading from 32 Volumes





Flash Memory Summit

Thank You!

Questions?

Presenters: Tim Fisher (fisher@us.ibm.com), Mohit Kapur (mohitk@us.ibm.com)

Contributors: Mohit Kapur, Martin Ohmacht, Ralph Bellofatto, Chris Ward, Tim Fisher, Bharat Sukhwani, Sameh Asaad

Thomas J Watson Research Center, IBM

When you interact with IBM, this serves as your authorization to Flash Memory Summit or its vendor to provide your contact information to IBM in order for IBM to follow up on your interaction.

IBM's use of your contact information is governed by the IBM Privacy Policy.

