G2M RESEARCH

G2M Research Multi-Vendor Webinar: Edge Computing/Storage – Get (and Keep) Your Data Off of My Cloud

Tuesday September 15, 2020 (v1.0)

lightbits

ScaleFlux™

NGD systems

Webinar Agenda

9:00-9:05	Ground Rules and Webinar Topic Introduction (G2M Research)
9:06-9:35	Sponsoring Vendor presentations on topic (10 minute each)
9:36-9:36	Audience Survey 1 (1 minute)
9:37-9:43	Key Question 1 (1-minute question; 2 minutes response per vendor)
9:44-9:44	Audience Survey 2 (1 minute)
9:45-9:51	Key Question 2 (1-minute question; 2 minutes response per vendor)
9:52-9:52	Audience Survey 3 (1 minute)
9:53-9:59	Key Question 3 (1-minute question; 3 minutes response per vendor)
10:00-10:13	Audience Q&A (14 minutes)
10:14-10:15	Wrap-Up

G2M Research Introduction and Ground Rules

Mike Heumann

Managing Partner, G2M Research

Panelists

Josh Goldenhar
VP, Product Marketing
www.lightbitslabs.com

JB Baker
VP, Marketing
www.scaleflux.com

Scott Shadley
VP, Marketing
www.ngdsystems.com

Mike Heumann
Managing Partner
www.g2minc.com

The Edge is Full of Data

- 75% of enterprise data will be stored and processed at the edge by 2025 (Gartner)
- 90% of Industrial Enterprises will Utilize Edge Computing by 2022 (Frost and Sullivan)
- The Edge Computing market will grow by 19.9% by year between now and 2025
- Companies will spend $175B-$215B on IoT hardware by 2025

≠

Impacts of Moving All This Data

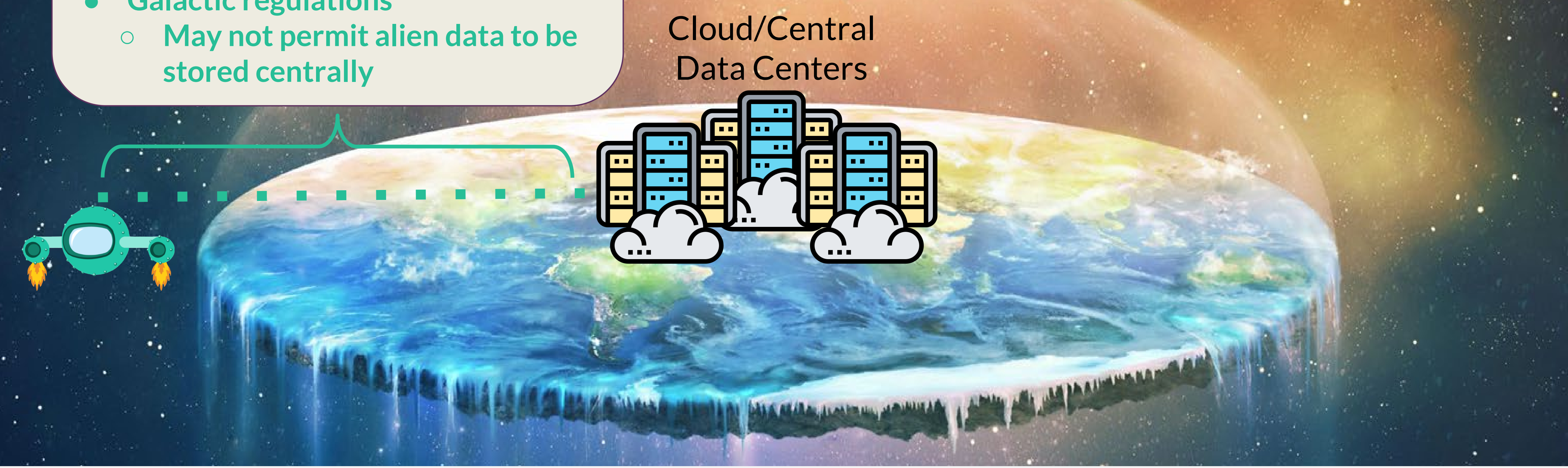
- In many cases, the data pipe to move large amounts of data from the edge doesn't exist
  - Think aircraft engine management systems, self-driving vehicle data gathering, or video surveillance
  - These systems can generate terabytes of data per hour
- What options are there to moving all the data?
  - Pre-processing the data and only moving the metadata
  - Sampling the data and sending samples only
  - Various data compression approaches
  - Combinations of these

Lightbits Labs

Josh Goldenhar
VP, Product Marketing
www.lightbitslabs.com

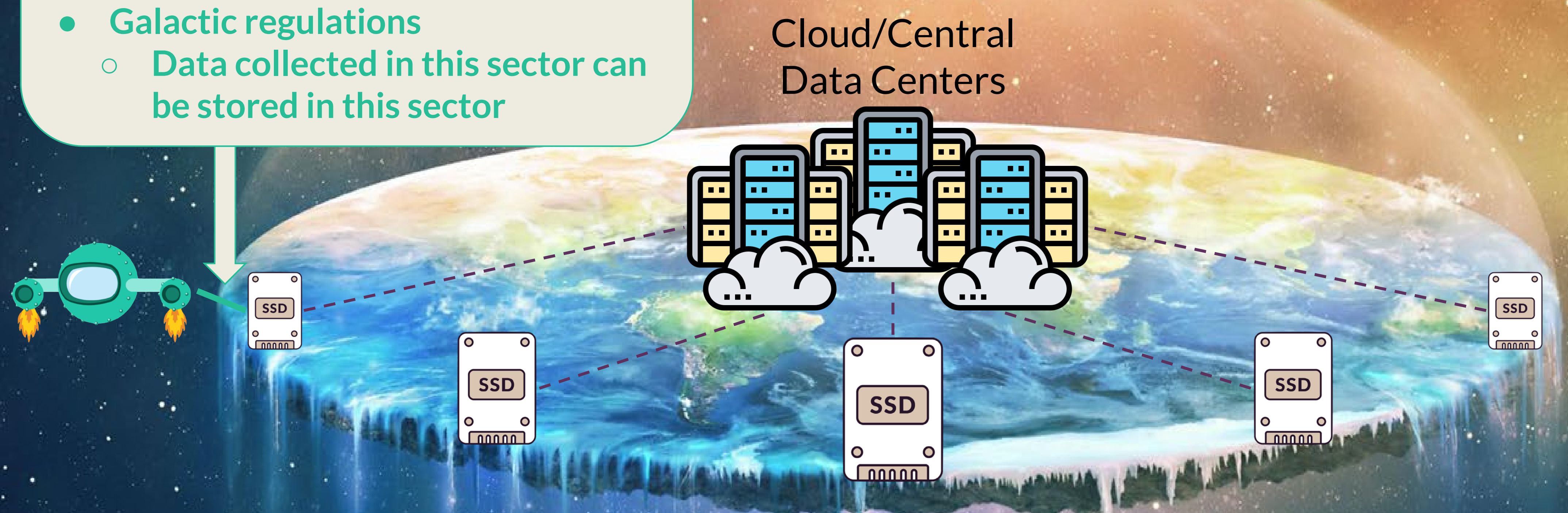
Data To/From Central Datacenters

- Latency high (long distance)
  - Hinders decision making
- Limited bandwidth
  - Can’t send all the data we might want to in the time we have
- Galactic regulations
  - May not permit alien data to be stored centrally



Storage at the Edge

- Latency low (short distance)
    - Decisions made quickly
- Local bandwidth
    - Bandwidth to edge storage may be higher than to central cloud
- Galactic regulations
    - Data collected in this sector can be stored in this sector

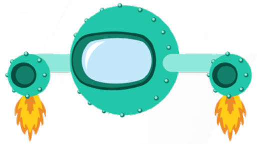


What's Driving Data to the Edge?

- Latency sensitive applications
  - These require a fast response time and cannot tolerate the high latencies associated with data traveling from the edge all the way to a core cloud
- Cloud egress bandwidth reduction
  - Frequently accessed and/or large data-sets/files stored at the edge (e.g. CDN), reduces outbound traffic and bandwidth requirements from central datacenters
- Cloud ingress bandwidth reduction
  - Data generated at edge devices that is pre-processed, reduced and stored at the edge and in turn reduced inbound central datacenter bandwidth requirements

What's Driving Data to the Edge?

- Limited connectivity
  - Allows edge devices to supply services at locations with limited or no cloud connectivity
- Security
  - Some data just simply can't be transmitted into centralized locations
- Regulatory requirements
  - Sometimes it's required to keep data close to the edge where it's generated to comply with the local municipal or national requirements on data movement and location



Give Me an Example

(There has to be a mention of AI/ML or 5G in nearly any presentation today…)

- AI decision making with edge devices
  - Machine Learning may take place on very large data set in the cloud or central data center
  - The resulting models are transferred on a regular basis (relatively small compared to the training data sets) to the edge
  - Sensors, devices (cars, traffic lights, wearables, phones, etc) send data to edge locations where it can be utilized quickly/locally to make a decision based on the model (inference)
  - Incoming data is small, numerous and needs to be acted on quickly requiring low latency storage and processing
  - 5G!

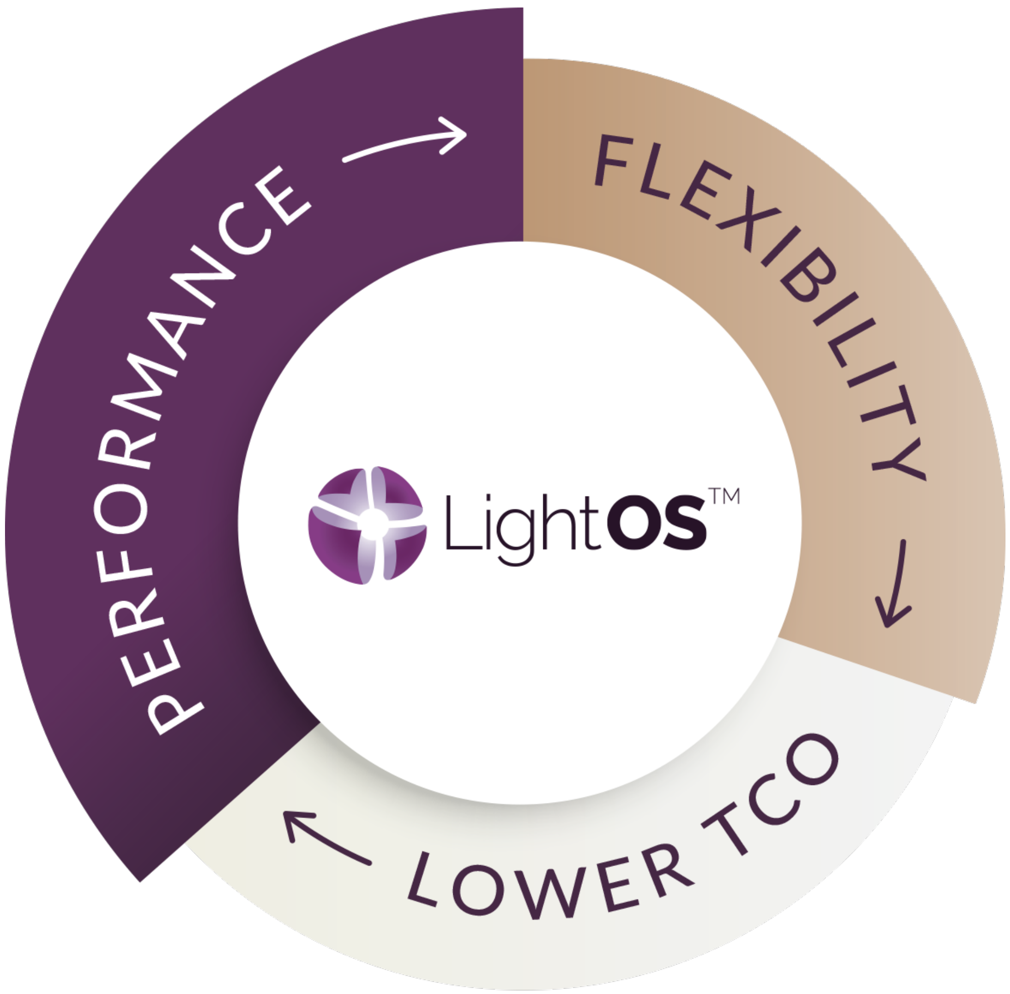
Issues with Edge Data Centers

- Often limited on:
  - Power
  - Cooling
  - Space
  - Serviceability

- Hardware and Networking
  - Hardware selection may be limited
  - Specialized configurations frowned upon
  - Making the most use of any resource is paramount

Any storage solution needs to be fast, multi-purpose, robust and flexible in implementation on standard hardware and networks - does such a solution exist????

LightOS Central Values



LightOS: Hyperscale Storage For All

Flexibility and Operational Efficiency

- Software-defined, highly available block storage, standard infrastructure
- Easy to deploy and allows for scaling storage & compute independently
- Rich data services

High Performance

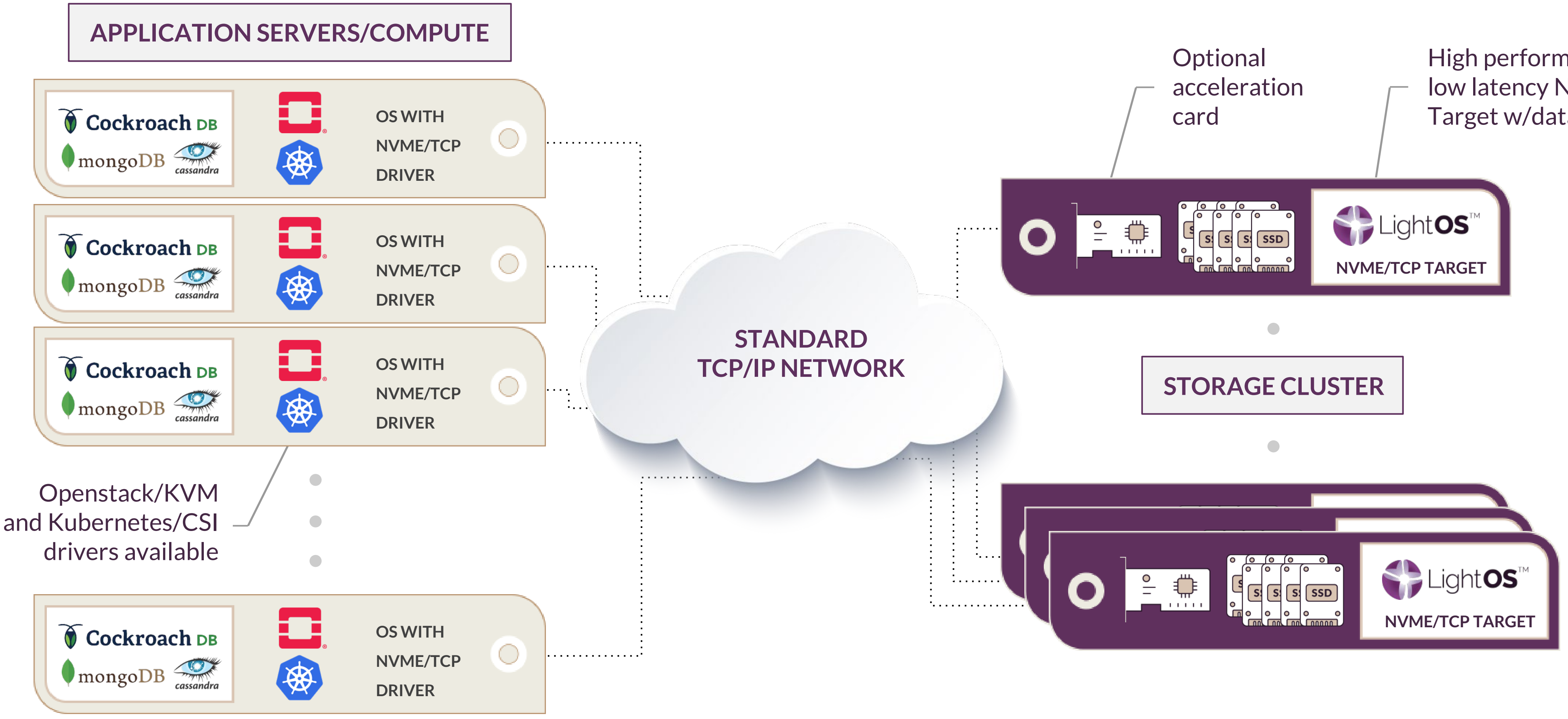
- Low latency (<200μs), high IOPs (Millions per target server), high bandwidth

Lower TCO

- Maximize utilization/ROI
- Increases flash endurance

High Performance Software Defined Storage

Standard servers, NICs and SSDs, optional hardware accelerator



Rich Data Services & Integrations

Local flash performance, array-like features

NVMe virtual block devices (NVMe/TCP):

- All volumes thin provisioned
- Line rate compression
- Data protection level per volume
- Volume expansion
- Consistent response time

Environment:

- Standard Linux distributions
- Kubernetes via CSI driver
- Openstack via Cinder driver
- Management via API and CLI
- Monitoring via Prometheus and Grafana (or choice)

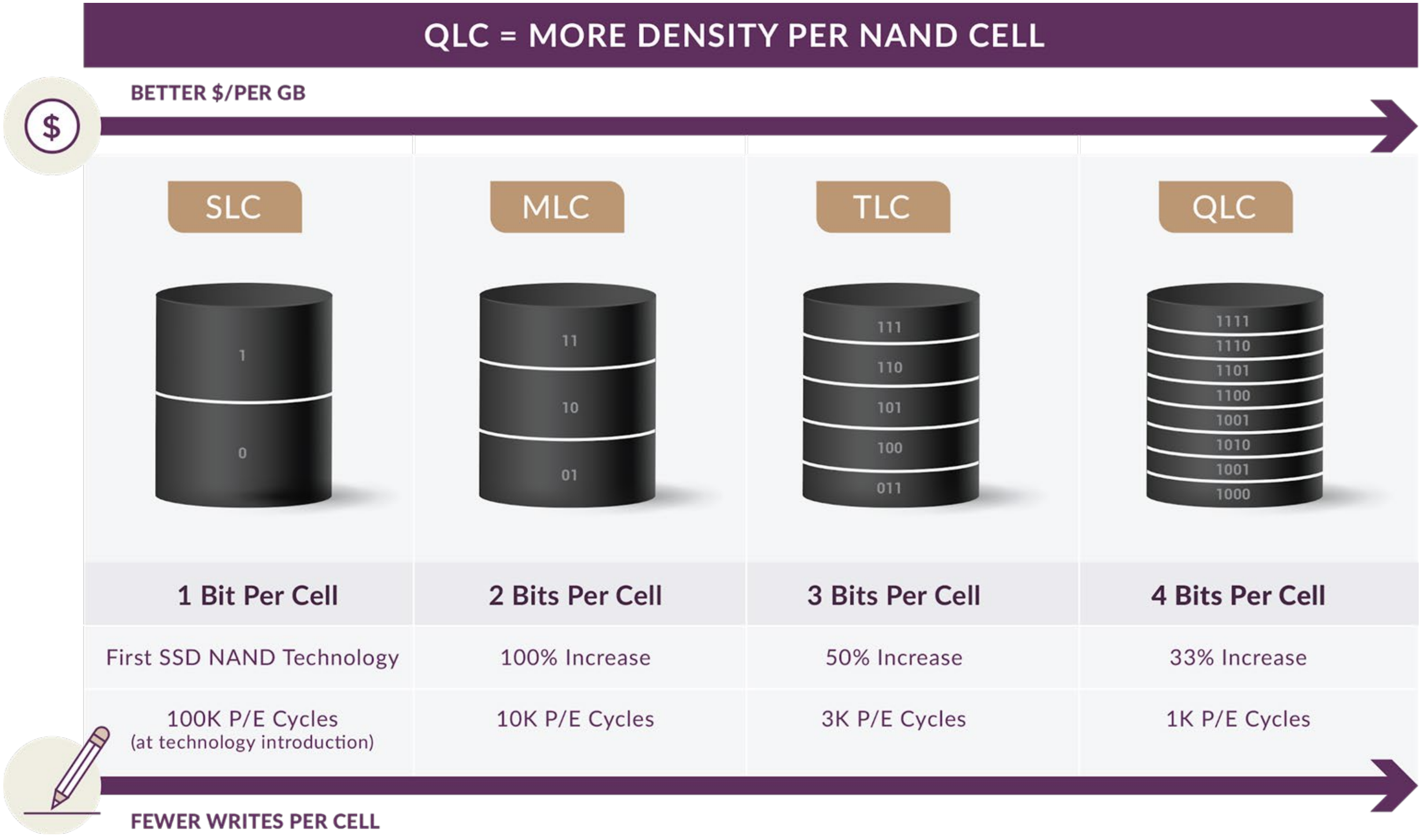
LightOS™

LightOS™

LightOS enables QLC Flash:

- Up to 5 times endurance
- Aggregates writes for higher performance
- Reduces drive over-provisioning
- Allows for high capacity, high density while preserving high performance

Increase Storage Density and Reduce TCO



Edge Storage Extraordinaire!

LightOS Meets the Challenges of Edge Storage

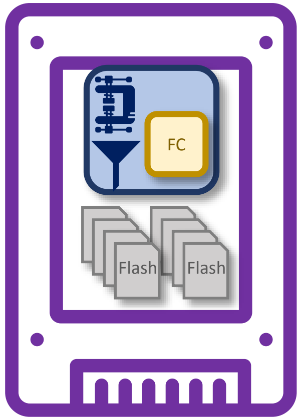
- Low latency and high bandwidth, with high performance density (IOPs and GBs per rack unit)
- Works on x86 hardware and any Ethernet NIC
- Supports high capacity with TLC and QLC NVMe SSDs
  - Supports fail-in-place for SSDs ElasticRAID
- Does not require special Ethernet switch configuration
- Easy to deploy and manage

Scaleflux

- JB Baker
  VP, Marketing
  www.scaleflux.com

Computational Storage At the Edge

- Computational Storage Drives (CSDs) add compute function(s) directly into the SSD
  - Enables distributed & parallel execution of tasks
  - Improves overall efficiency
- Example functions:
  - Pre-processing Data
    - Reduce Data Movement
    - Reduce Network Burden
  - Compression
    - Improve Edge Storage Economics
    - Improve Edge Compute Economics



Single FPGA combines Compute and SSD Functions

G2M RESEARCH

Pre-Processing Example: Data Filtering

Ordinary Storage:

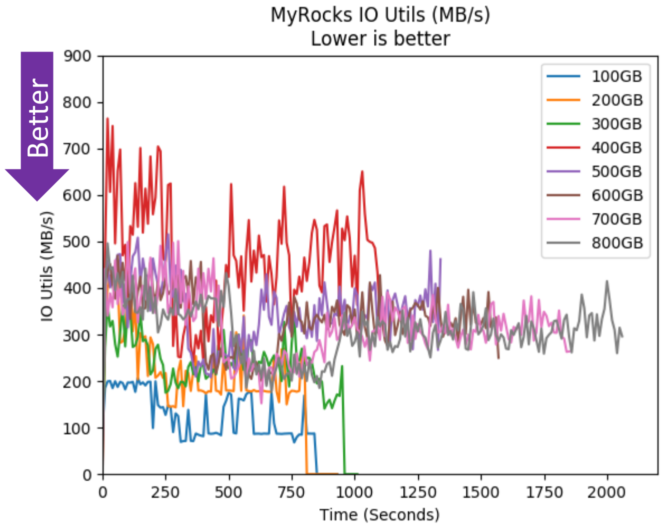
Data Filtering CSD:

- Data Movement, CPU Utilization, & Query Completion Time
  - 8 instances of RocksDB
  - 100GB per instance

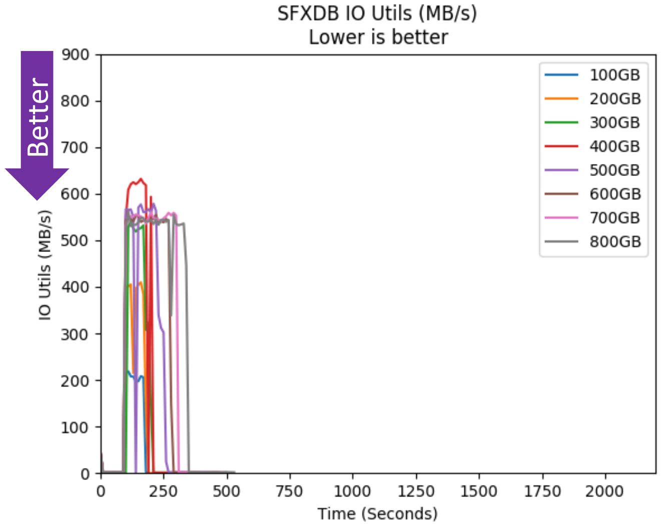
G2M RESEARCH

Pre-Processing Example: Data Filtering

Ordinary Storage:
- Massive total data movement
- Long completion times



Data Filtering CSD:
- Minimized data movement
- Faster completion

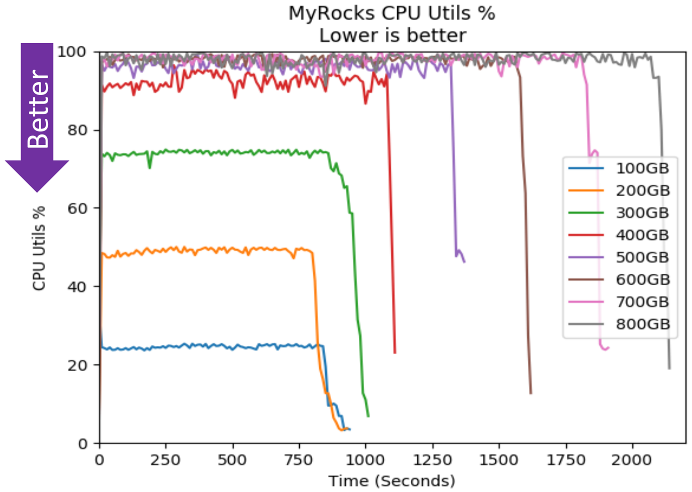


- Data Movement, CPU Utilization, & Query Completion Time
  - 8 instances of RocksDB
  - 100GB per instance

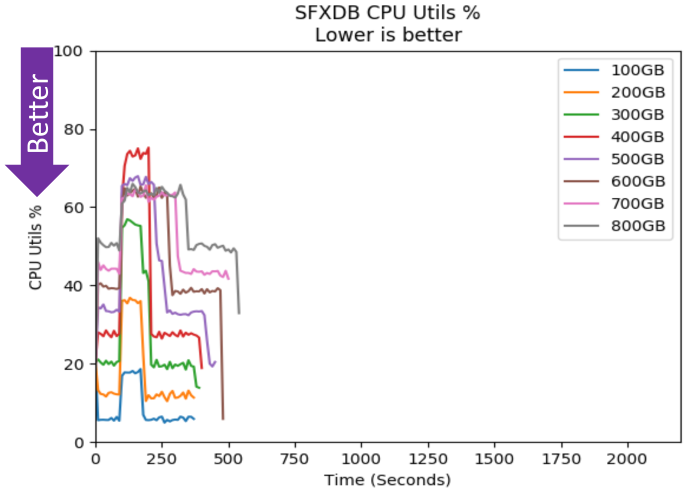
G2M RESEARCH

Pre-Processing Example: Data Filtering

Ordinary Storage:
- CPU pegged @ 4 processes
- Long completion times



Data Filtering CSD:
- No CPU bottleneck through 8 processes
- Minimized query times



- Data Movement, CPU Utilization, & Query Completion Time
  - 8 instances of RocksDB
  - 100GB per instance

Use CSDs to:

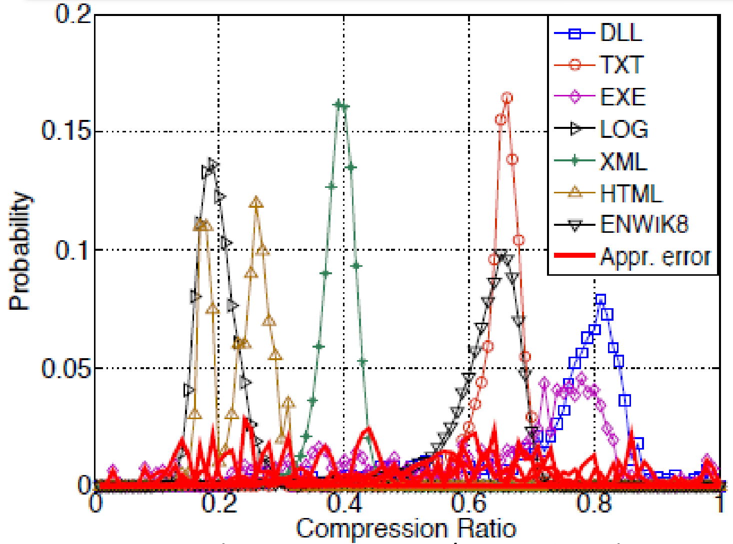
✓ Run analytics faster & with less CPU

✓ Transfer *information* not raw data

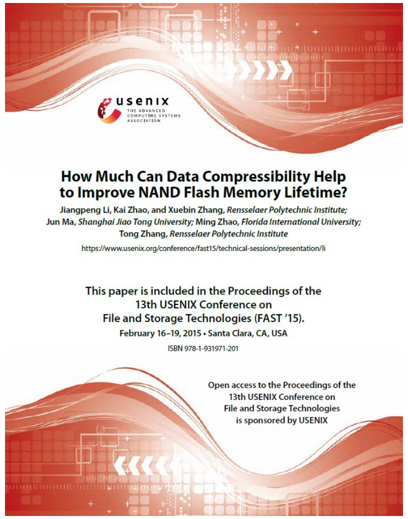
G2M RESEARCH

Compression: Storage Economics

File Types	**Compressibility** (Original Size:Compressed Size)
Images, Video, Encrypted	<1.2:1
Binaries, DLL, EXE	1.2:1
XML	2:1
HTML, Logs, Database	>2:1



(Compressed Size/Original Size)


usenix THE ADVANCED COMPUTING SYSTEMS ASSOCIATION

How Much Can Data Compressibility Help to Improve NAND Flash Memory Lifetime?

Jiangpeng Li, Kai Zhao, and Xuebin Zhang, *Rensselaer Polytechnic Institute;* Jun Ma, *Shanghai Jiao Tong University;* Ming Zhao, *Florida International University;* Tong Zhang, *Rensselaer Polytechnic Institute*

https://www.usenix.org/conference/fast15/technical-sessions/presentation/li

This paper is included in the Proceedings of the 13th USENIX Conference on File and Storage Technologies (FAST '15).
February 16–19, 2015 • Santa Clara, CA, USA

ISBN 978-1-931971-201

Open access to the Proceedings of the 13th USENIX Conference on File and Storage Technologies is sponsored by USENIX

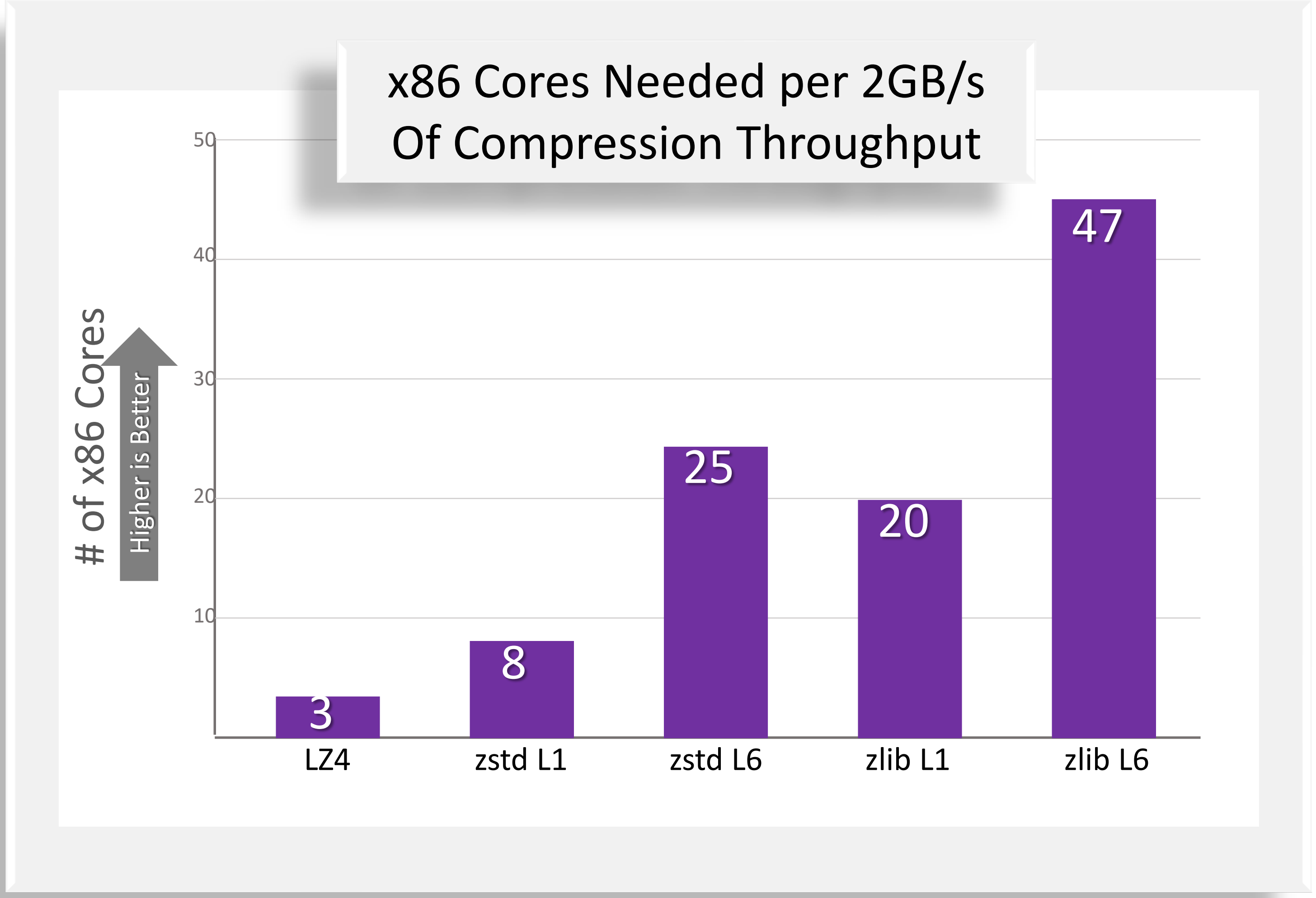
https://www.usenix.org/system/files/conference/fast15/fast15-paper-li.pdf

Double (or more) effective edge storage capacity

- ✓ Buy less physical capacity
- ✓ Use fewer drives
- ✓ Increase density per server
- ✓ Extend back up / offload windows

G2M RESEARCH

Compression: Compute Economics



CPU-based compression

- High tax on cores
- Adds latency
- Throughput doesn't scale with storage

CSD-based compression

- ✓ Frees up cores for applications
- ✓ Reduces the cost & power
- ✓ Saves on storage costs without sacrificing performance

- Test Set Up:
  - Test server CPU is Intel(R) Xeon(R) CPU E5-2667 v4 @ 3.20GHz
  - Measured compression throughput for 1 core with various compression algorithms
  - Source data was the Corpus Canterbury files for all runs
  - Graph metric is "how many cores are needed to achieve the compression throughput of 1 CSD"

G2M RESEARCH

NGD Systems

- Scott Shadley
  VP, Marketing
  www.ngdsystems.com

Computational Storage Saves the Edge

Who? Needs Compute?

What? Is Computational Storage?

When? Is it Needed?

Where? Is it Deployable?

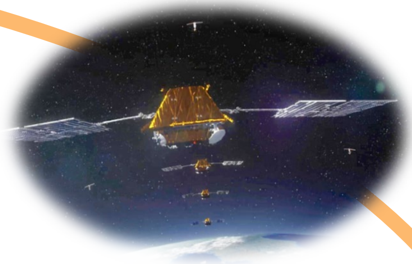
Why? Should YOU care?

Who Needs Computational Storage?

NGD systems

SMART CITIES & AUTOMATION

EXTREME EDGE



Linux

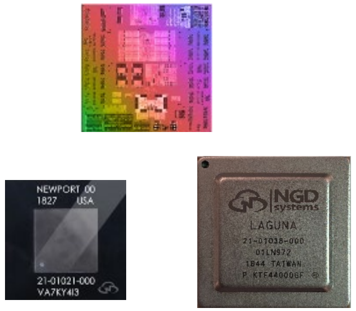
docker

DATACENTERS & CDN DELIVERY

AUTONOMOUS VEHICLES

What is Computational Storage?

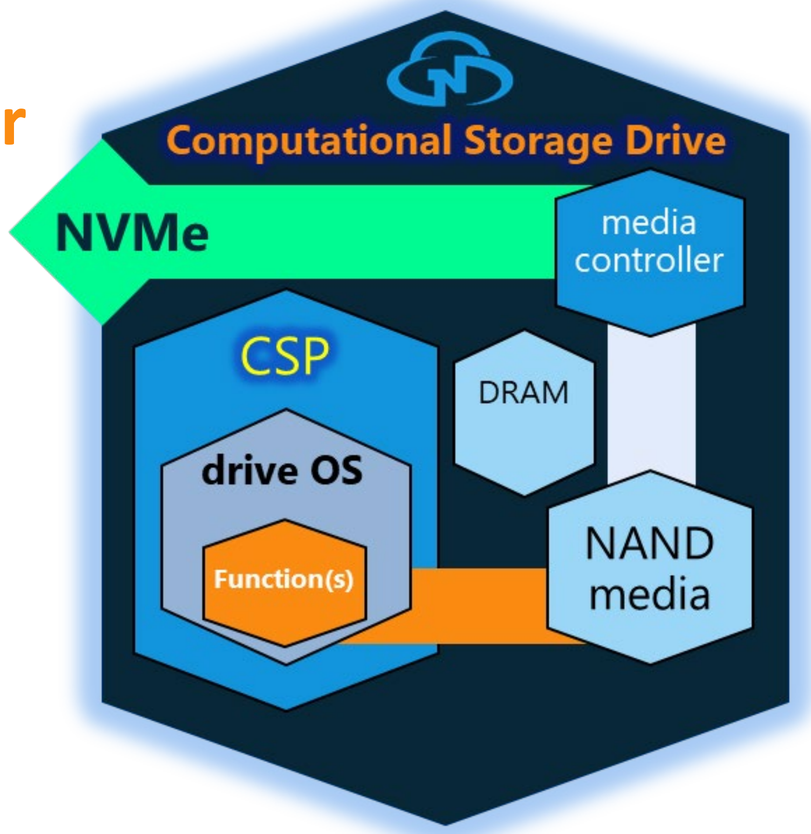
ASIC-Based Controller



Management & Firmware

Modular firmware

Flash Vendor & Type agnostic
SLC/MLC/TLC/QLC



Designed & Manufactured in USA

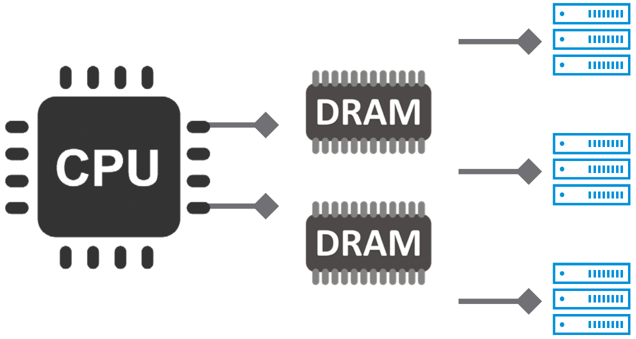
In-Situ Processing

- ✓ Full fledged on drive OS
- ✓ Quad-core 64-bit app processor
- ✓ Hardware acceleration

When is Computational Storage Needed?

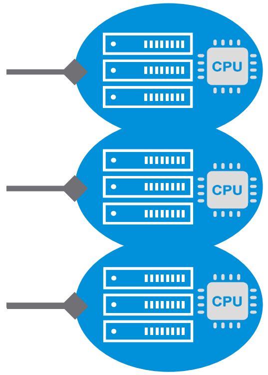
Pain Points - Before

- ❌ Physical Space
- ❌ Available Power
- ❌ Scaling Mismatch
- ❌ Bottleneck Shuffle

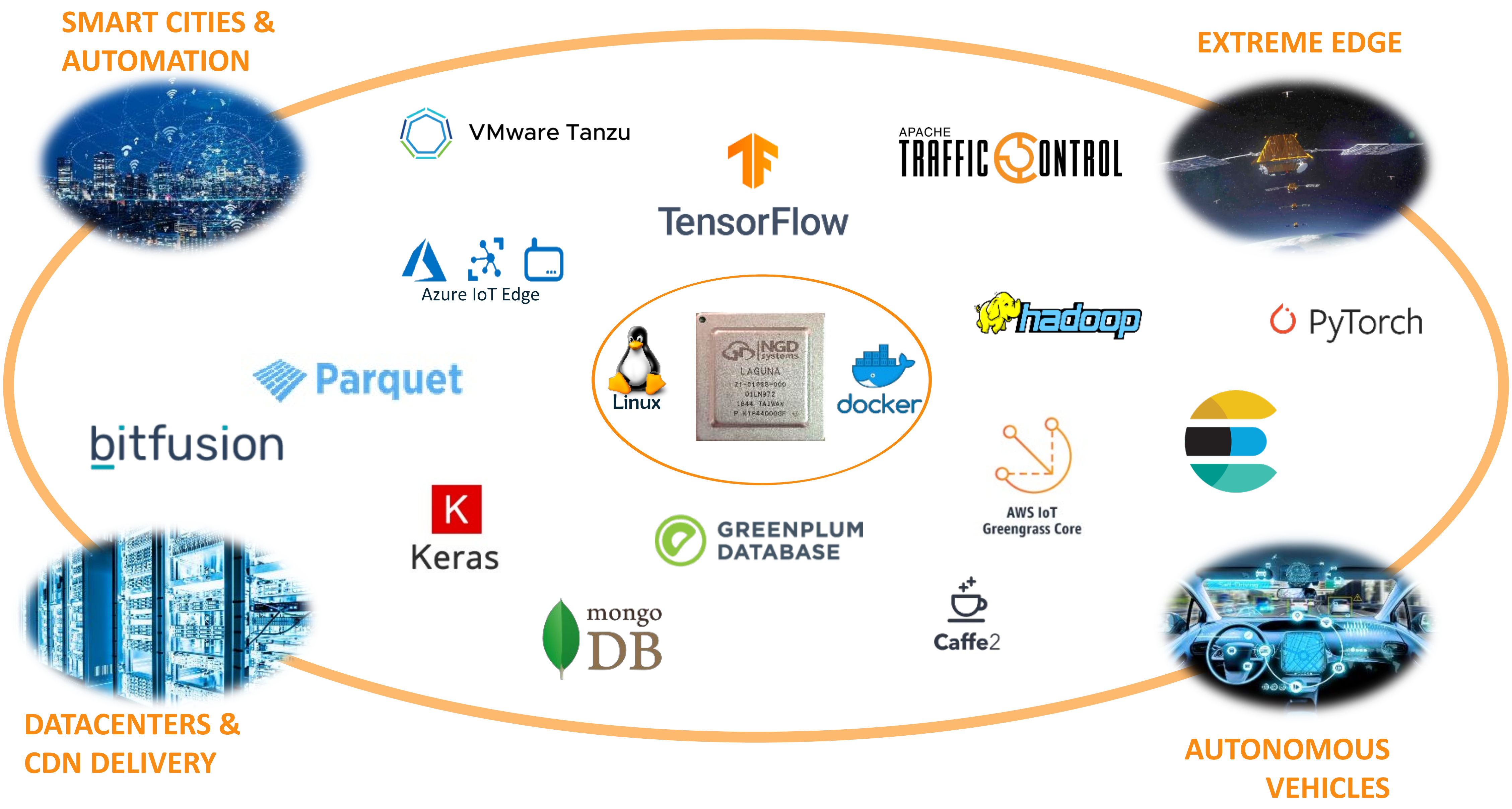


Computational Storage Benefits

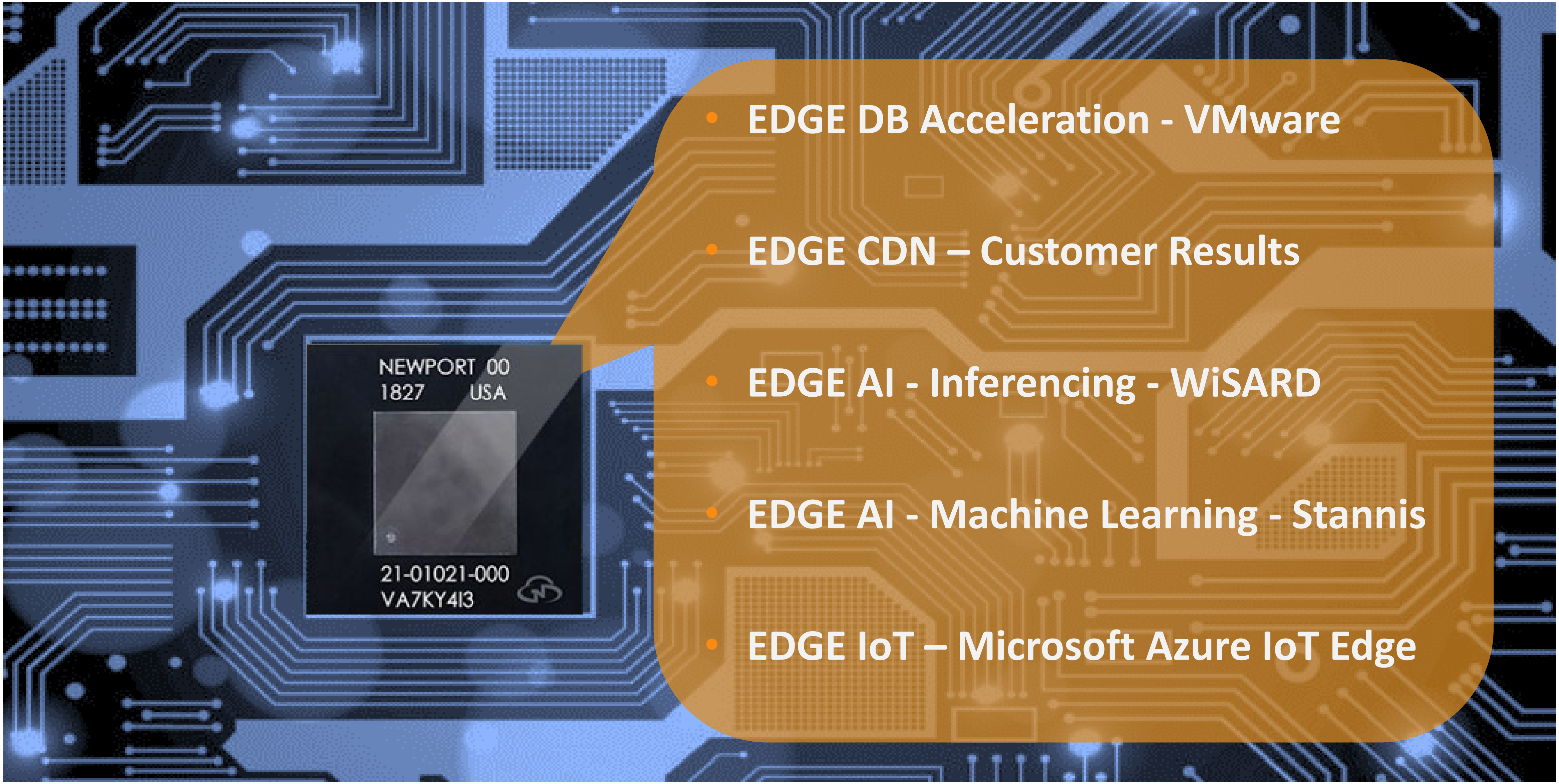
- ✓ Smaller Footprint
- ✓ Lower Power
- ✓ Distributed Processing
- ✓ Faster Results



Where is Computational Storage Deployed?



Why You Should Care!



- EDGE DB Acceleration - VMware
- EDGE CDN – Customer Results
- EDGE AI - Inferencing - WiSARD
- EDGE AI - Machine Learning - Stannis
- EDGE IoT – Microsoft Azure IoT Edge

Edge Analytics –VMware – xLabs Innovation Center.

Computational Storage allows it to be drive level.

Reducing footprint, server cost, while still offering full fault tolerance

vmware

DELL EMC

bitfusion

GREENPLUM DATABASE

1 Segment Host/Data Node PER DRIVE
Database is Mirrored and Fault Tolerant

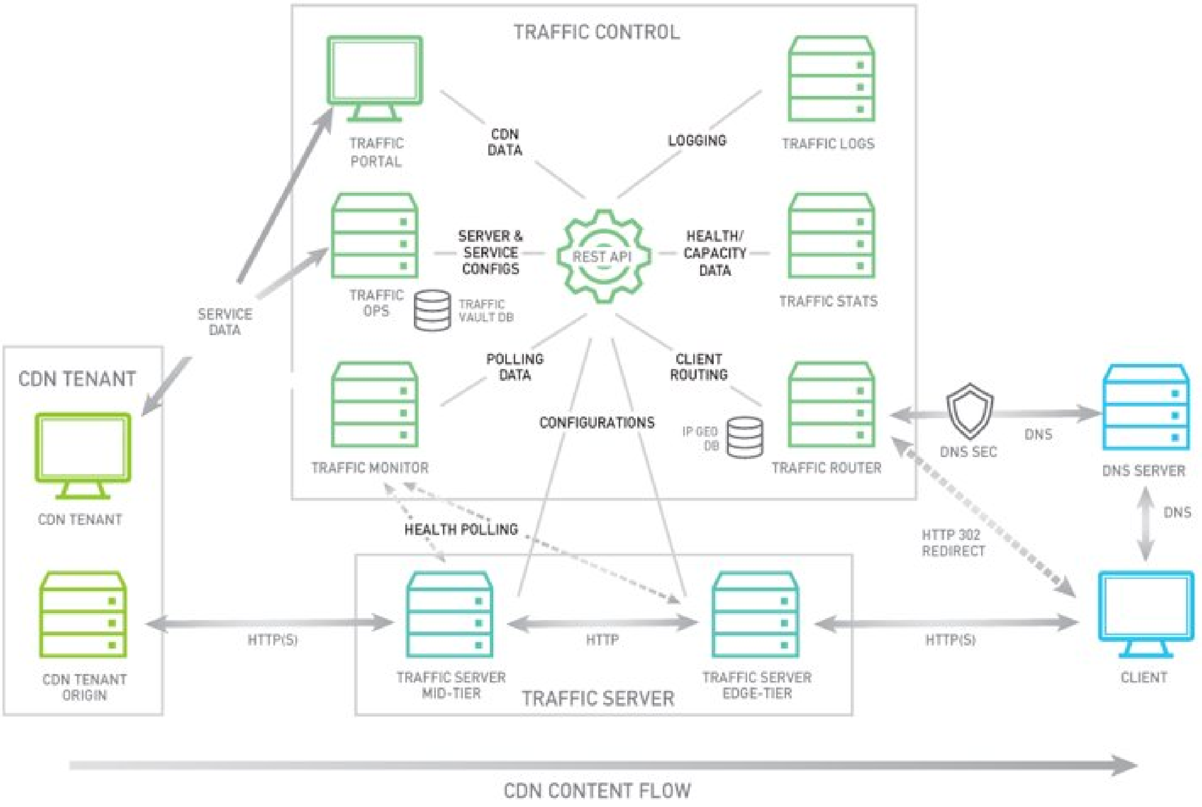
VMware Tanzu

Full management including full resiliency and data loss protection at server level

Showcased at Vmworld 2020 - Session ID [OCTO478] –

Computational Storage, Tanzu Greenplum, vSphere Bitfusion

Apache Traffic Control Customer Lab Results.



Problem Statement

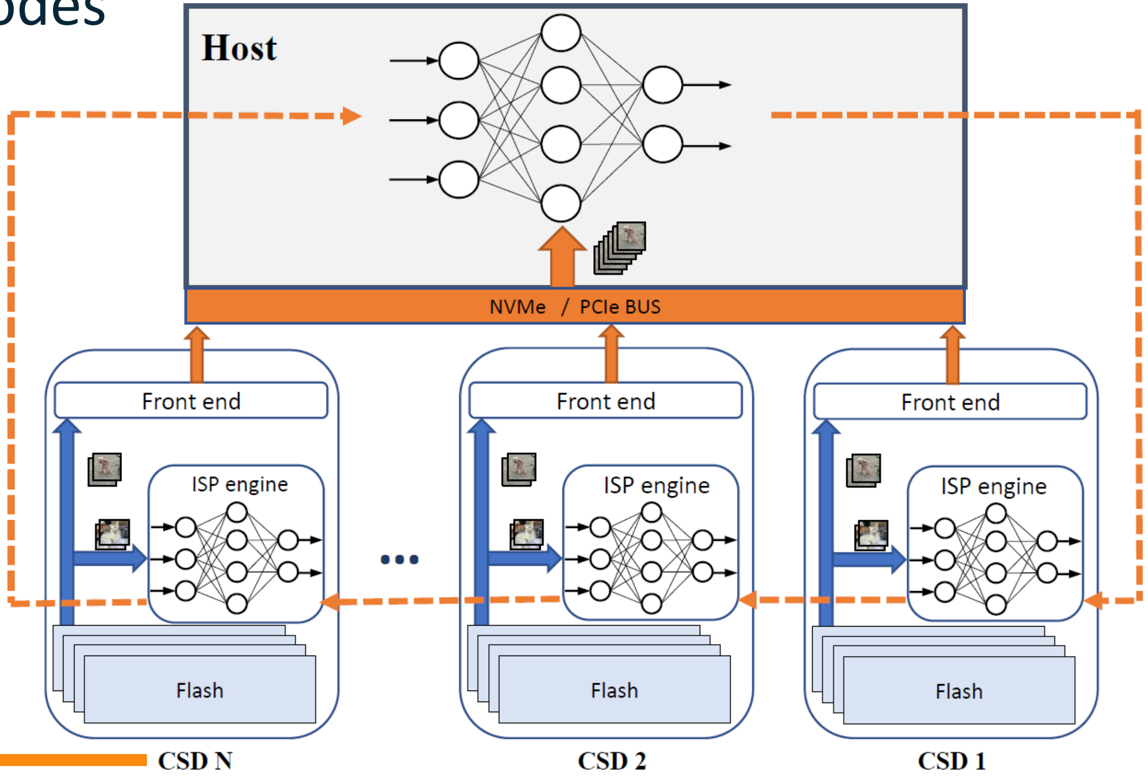
- Open source CDN – Traffic Control
- Focus on Time to First Frame (TtFf)
- Complex System to be pulled apart to find single step for impact

Process

- Identified a Single Storage Instance
- Allocated Storage and processing to NGD Computational Storage
- Performance impact to whole system
  - **50% faster step performance**
  - **10% overall system improvement**

Machine Learning Using the STANNIS Framework

- System for Training of Neural Networks In Storage
- Matches the processing workload of all nodes
- Determines batch size on each node
- Load balancing on size of input data



THE Computational Storage Edge Products You Need.

- Large breadth of **SSD** solutions and capacity options
- Leading **W/TB** Energy Efficiency
- Industry's only 16-Channel M.2
- Largest capacity NVMe U.2

Form Factor	Availability	Raw Capacity TLC (TB)	QLC Coming CQ'4
M.2 22110	NOW	up to 8	Stay
U.2 15mm	NOW	up to 32	Tuned
EDSFF E1.S	NOW	up to 12	For
E3.S	Q3	up to 32	More

M.2

U.2

EDSFF

NGD's Computational Storage device is powered by Azure Iot Edge

https://www.youtube.com/watch?v=D7Ab8zIi3kw

The ONLY *Edge-Ready*
Computational Storage Drive on the Market
Learn More Here: www.NGDSystems.com

Panel Questions and Audience Surveys

Audience Survey Question #1

- How big of a problem will edge data be for your organization over the next five years? (select one):
  - It is already a critical concern for our organization: 3%
  - We expect it to be a critical problem for our organization w/i the next 5 years: 18%
  - We expect it to be one of our significant data problems w/i the next 5 years: 26%
  - It is a concern, but not one of our high priorities w/i the next 5 years: 9%
  - It is not a concern for us within the next 5 years: 9%
  - Unknown how much of a problem it will be for us: 35%

G2M RESEARCH

Panel Question # 1

- What vertical markets and/or use cases do you see as the having the biggest issues with edge data over the next 5 years?
  - Lightbits Labs
  - ScaleFlux
  - NGD Systems

Audience Survey Question #2

- For those of you who believe that edge data will be an issue for your organization, what options are you looking at? (select all that apply):
  - Pre-processing of data as it is ingested at the edge: 31%
  - Sampling the data and transmitting samples only: 15%
  - Data compression approaches: 27%
  - In-situ processing of the data at the edge: 35%
  - New networking techniques/capabilities: 27%
  - A combination of these approaches: 50%
  - Other: 8%

G2M RESEARCH

Panel Question # 2

- Current thinking regarding edge data and edge computing is that the amount of data will always exceed the ability to transmit it, either to the cloud or to other datacenters. Does your organization believe that this will continue to be true?
  - ScaleFlux
  - NGD Systems
  - Lightbits Labs

Audience Survey Question #3

- Who do you expect can help your organization solve your edge data issues? (select one):
  - Our server vendors: 12%
  - Our data storage/data networking vendors: 8%
  - Edge computing resource vendors: 20%
  - Consultants specializing in edge computing: 8%
  - Cloud computing/storage vendors: 48%
  - Others: 4%

G2M RESEARCH

Panel Question #3

- Assuming that transmission will continue to be an issue, what are some of the most promising approaches to reduce the volume of data generated at the edge?
  - NGD Systems
  - Lightbits Labs
  - ScaleFlux

Audience Q&A

Thank You For Attending

G2M

RESEARCH