G2M RESEARCH

G2M Research Multi-Vendor Webinar: Advanced NVMe™ SSDs – Addressing the Blast Radius Problem

Tuesday March 31, 2020

Webinar Agenda

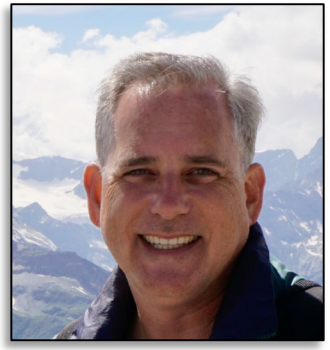
9:00-9:05	Ground Rules and Webinar Topic Introduction (G2M Research)
9:06-9:29	Sponsoring Vendor presentations on topic (8 minute each)
9:30-9:40	Key Question 1 (2-minute question; 3 minutes response per vendor)
9:41-9:42	Audience Survey 1 (2 minutes)
9:43-9:53	Key Question 2 (2-minute question; 3 minutes response per vendor)
9:54-9:55	Audience Survey 2 (2 minutes)
9:56-10:06	Key Question 3 (2-minute question; 3 minutes response per vendor)
10:07-10:18	Audience Q&A (12 minutes)
10:19-10:20	Wrap-Up

G2M Research Introduction and Ground Rules

Mike Heumann

Managing Partner, G2M Research

Panelists



Josh Goldenhar
VP Product Marketing
www.lightbitslabs.com

Joel Dedrick
VP/GM, Networked Storage SW
www.kioxia.com

Jonmichael Hands
Product Manager, SSDs
www.intel.com

Mike Heumann
Managing Partner
www.g2minc.com

What is the “SSD Blast Radius” Problem?

- Unlike processors, SSDs have continued to increase in capacity according to Moore’s Law
  - Enterprise U.2 SSD of 8TB or greater are typical
  - Max capacities of 16TB are here, 32TB are close
- At these capacities, a single SSD can easily support one or more servers
  - If the SSD fails, the impact of that failure on a server (or cluster) can be catastrophic
- Simply adding drives for redundancy can be very expensive
  - And using lower-capacity drives has its own issues

G2M RESEARCH

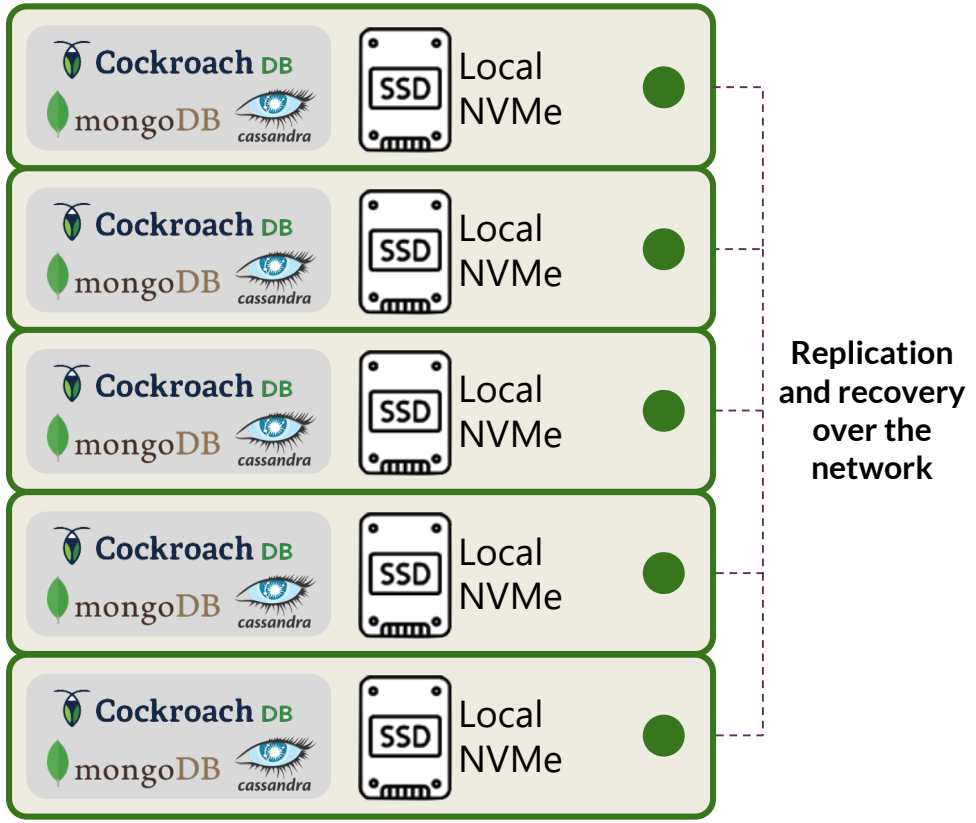
lightbits

Lightbits Labs

- Josh Goldenhar
- VP, Outbound Marketing
- www.lightbitslabs.com


lightbits
Hyperscale Storage for All

Cloud Native Applications: Scale-out, Distributed



- Racks of application servers
- Local flash for data storage
- Limited server SKUs
- Applications perform data replication
- Drive or server failures necessitate full rebuilds
- More than you can eat IOPs

Severe Under-Utilization

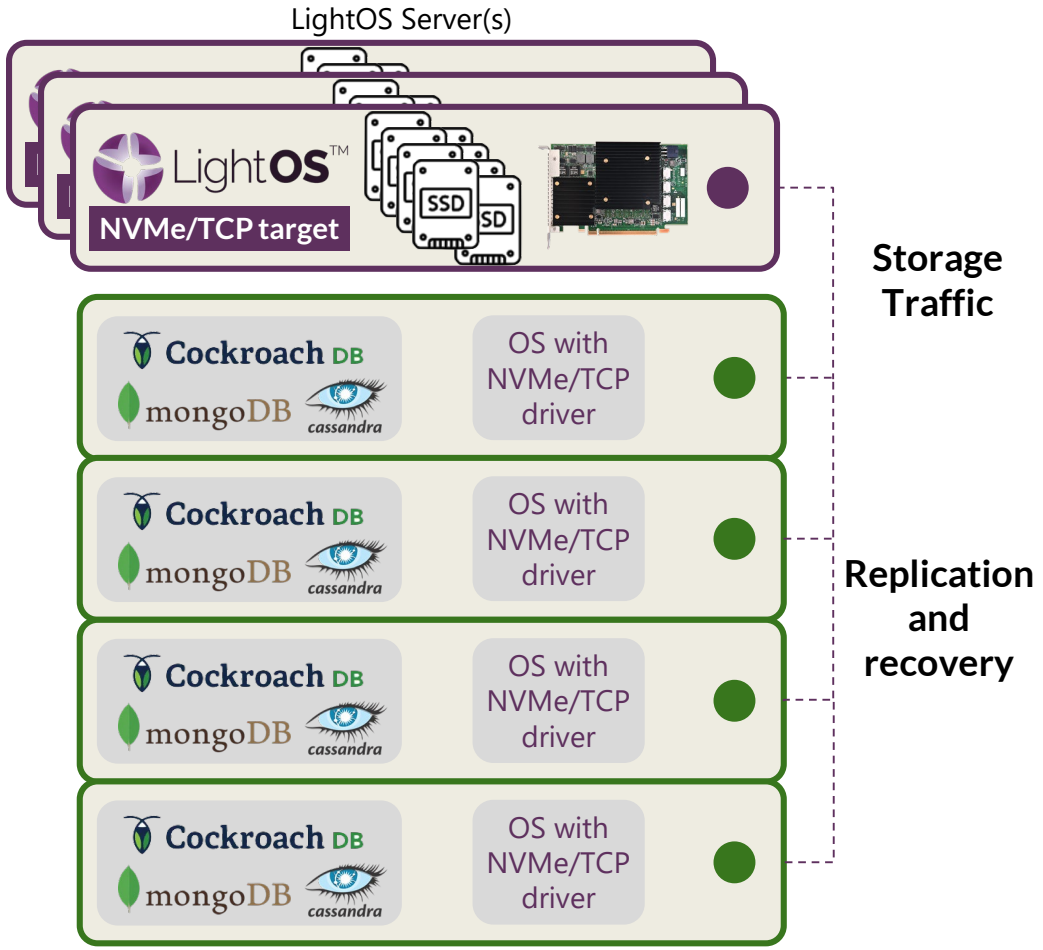
15-25% capacity utilization

50% performance utilization

50-85% of flash expense is wasted!

Cloud Native Applications with Lightbits' NVMe/TCP

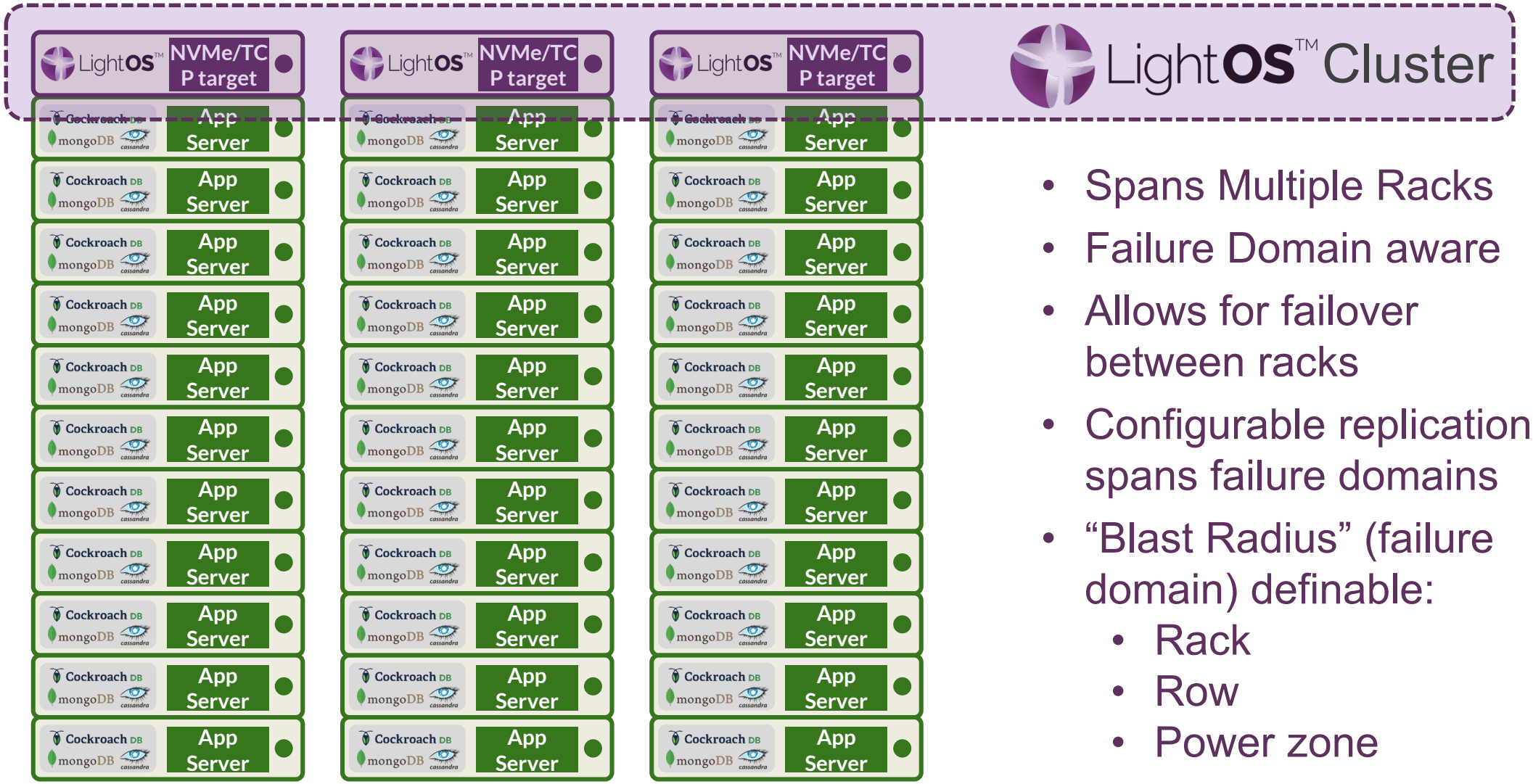
Higher uptime and utilization, faster recovery = lower TCO



Benefits:

- Local flash performance
- Drive failures do not affect applications
- Volumes can be any size and thin provisioned
- Server failure:
  - Instances can be launched anywhere
  - Recovery takes seconds to minutes
  - Higher uptime, increased SLAs

Multi-Rack, Scale-Out Approach



- Spans Multiple Racks
- Failure Domain aware
- Allows for failover between racks
- Configurable replication spans failure domains
- “Blast Radius” (failure domain) definable:
  - Rack
  - Row
  - Power zone

LightOS: Hyperscale Storage For All

- High performance
  - Low latency, high IOPs, high bandwidth
- Increased Operational Efficiency
  - Scale storage & compute independently
  - Software-defined, standard infrastructure
  - Rich data services
- Reduce TCO
  - Maximize utilization/ROI
  - Increase flash endurance

Disaggregate without increasing your Blast Radius!

G2M RESEARCH

KIOXIA

Kioxia

- Joel Dedrick
- VP/GM, Networked Storage Software
- www.kioxia.com

KUMOSCALE™

SSDs & Blast Radius

Joel Dedrick

VP/GM Networked Storage Solutions

KIOXIA

Caveats

The following content applies to medium/large scale data centers (KumoScale target market)

Statements should not be construed as a guarantee, express or implied, regarding Kioxia products

New Problems with Modern SSDs

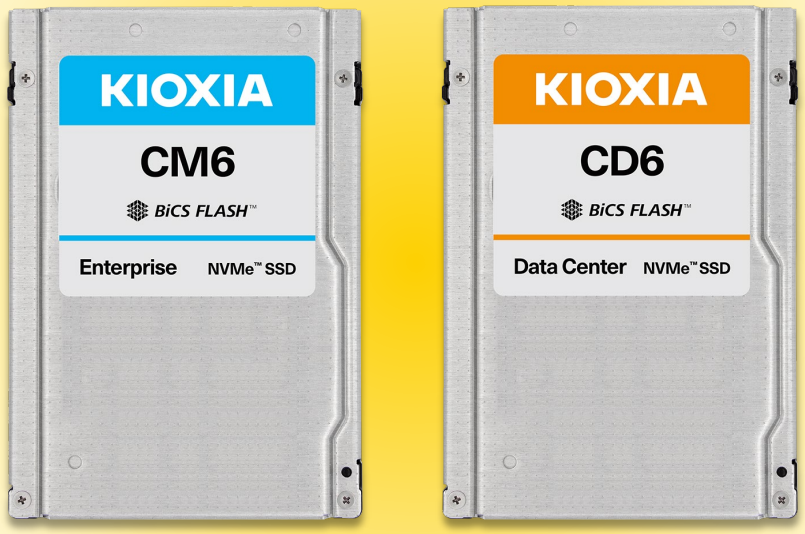
KUMOSCALE™

They’re too big

They’re too fast

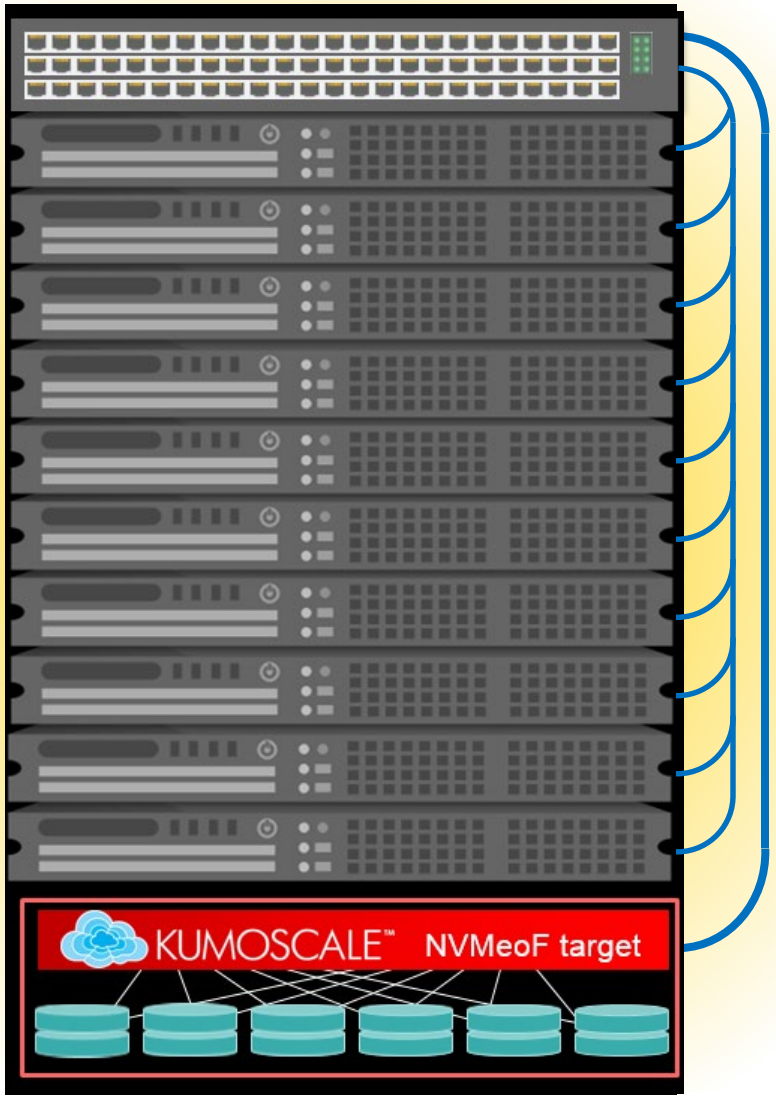
They’re too expensive

~~The amount of irreplaceable data you lose when one fails is too large.~~



Disaggregating Flash

KUMOSCALE™



Scale out flash approach enabled by NVMe-oF™, 100GE, KumoScale

Disaggregate

- Move SSDs out of compute nodes

Virtualize

- Volumes, not raw drives

Share

But wait, now the blast radius is *REALLY* big!

KUMOSCALE™

Data Protection – What Really Matters

Prime Directive:

KUMOSCALE™

Don’t lose *any* critical data – *ever*

Losing 500GB of irreplaceable, critical data is not obviously better than losing 16TB of it.

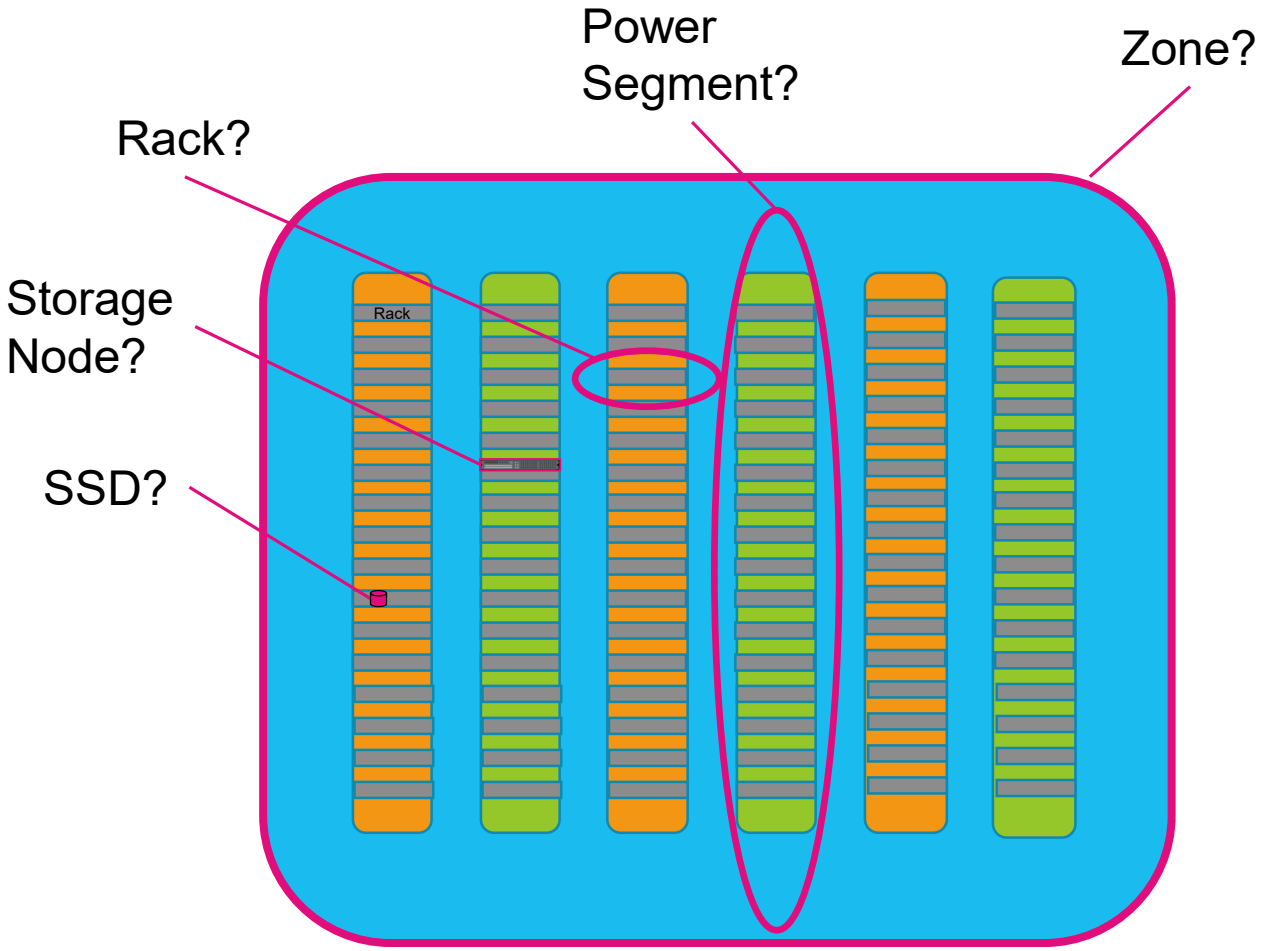
- **In terms of Data Protection, SSD capacity doesn’t matter – *and never has***

Blast Radius

KUMOSCALE™

Which “blast radius” is the one that matters?

The largest one that could conceivably fail.



Desideratum # 2

KUMOSCALE™

Don’t waste CAPEX $$ protecting noncritical data

Much in-process data is ephemeral or inherently replaceable. Not worth the performance cost to protect it.

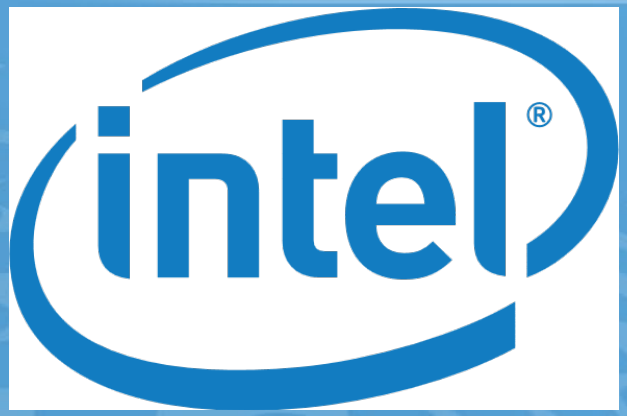
➢ ***At scale, “one size fits all” is prohibitively expensive***

Conclusions

Data protection needs to span the largest failure domain. SSD Boundaries are irrelevant.

Protection shouldn’t be mandatory. Pay only where you need it.

G2M RESEARCH



Intel

- Jonmichael Hands
- Product Manager, SSDs
- www.intel.com

Blast Radius of SSD – a distraction

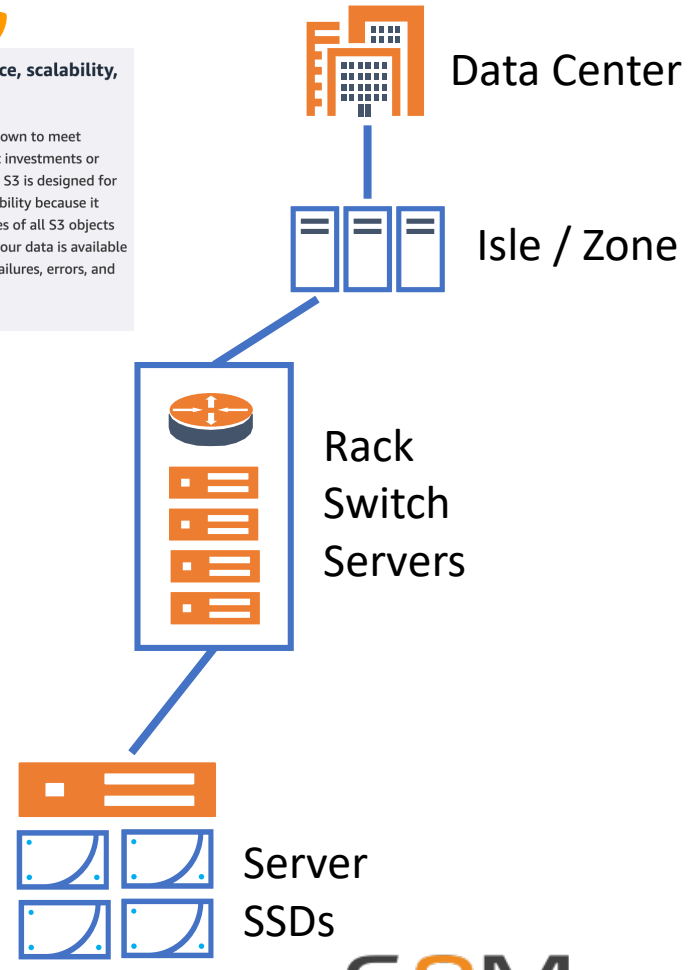
- Top cloud vendors don't talk about blast radius. They talk about durability, availability, and TCO
- Durability and availability requirements vary drastically by deployment scale
- SDS / HCI distribute data (e.g. Ceph CRUSH, VMware vSAN)
- Blast radius a function of SSD bandwidth, network bandwidth, and replication schema (RAID & EC)
- Small deployments (server, AFA, storage array) rely on rebuild time

aws

Industry-leading performance, scalability, availability, and durability

Scale your storage resources up and down to meet fluctuating demands, without upfront investments or resource procurement cycles. Amazon S3 is designed for 99.999999999% (11 9's) of data durability because it automatically creates and stores copies of all S3 objects across multiple systems. This means your data is available when needed and protected against failures, errors, and threats.

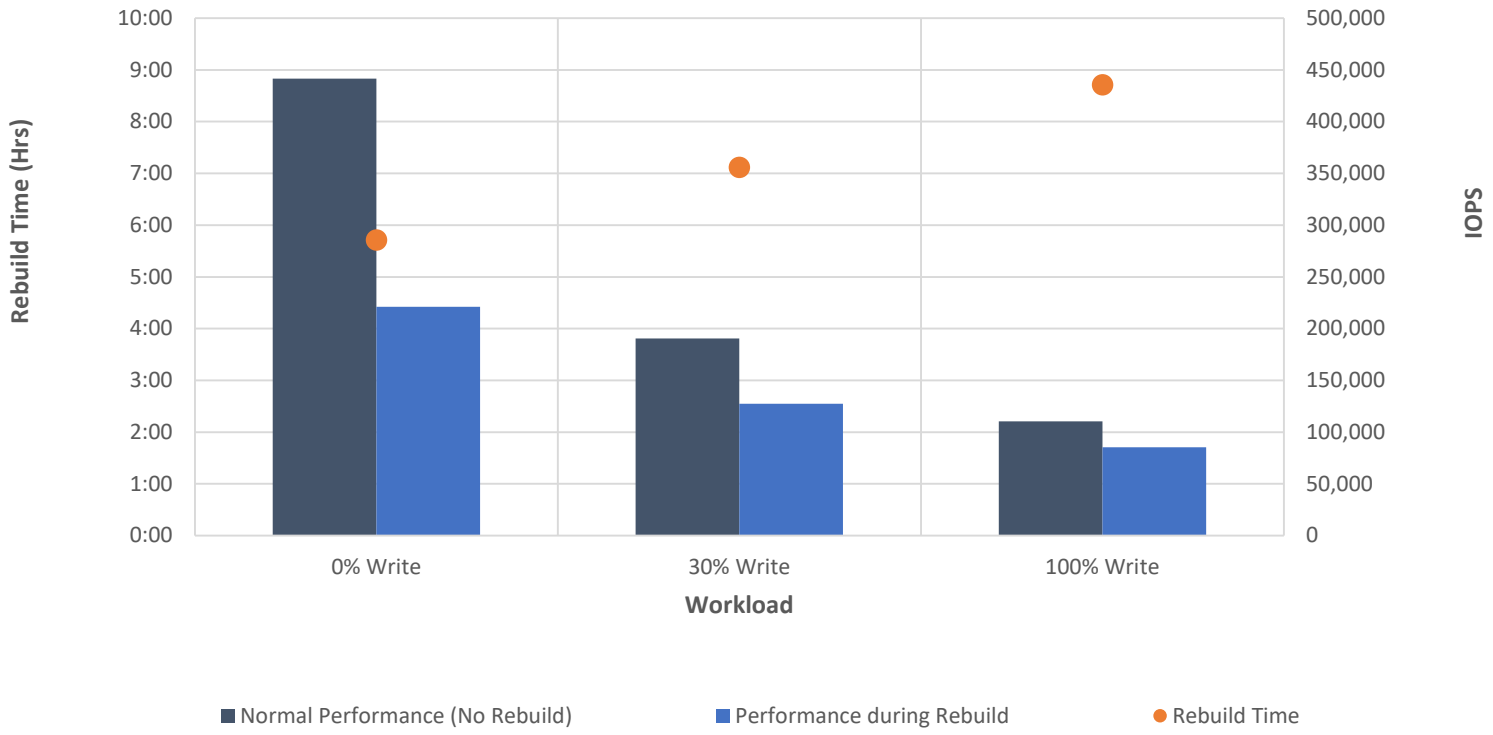
Learn about S3 data durability »



G2M RESEARCH

Intel® VROC RAID 5 Rebuild

RHEL* 7.4 with Intel® SSD DC P4510 Series



Rebuild Time with No Workload- 1:35

CPU Utilization during Rebuild:

Workload	CPU Usage (no Rebuild)	CPU Usage (during Rebuild)
0% Write	0.7 cores	2.9 cores
30% Write	0.9 cores	1.9 cores
100% Write	1.1 cores	1.6 cores

52 Physical cores on this 2 socket, Intel® Xeon® Platinum 8170 Processor-based system

Configuration Summary

RAID Volume: 4TB / SSD, 4 Disk RAID 5

- 12 TB Storage Capacity
- 4 TB Parity Overhead

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit www.intel.com/benchmarks. 1. **System configuration**: Intel® Server Board S2600WFT family, Intel® Xeon® 8170 Series Processors, 26cores@ 2.1GHz, RAM 192GB , BIOS Release 7/09/2018, BIOS Version: SE5C620.86B.00.01.0014.070920180847
OS: RedHat* Linux 7.4, kernel- 3.10.0-693.33.1.el7.x86_64, mdadm - v4.0 - 2018-01-26 Intel build: RSTe_5.4_WW4.5, Intel ® VROC Pre-OS version 5.3.0.1039, 4x Intel® SSD DC P4510 Series 4TB drive firmware: VDV10131, Retimer
BIOS setting: Hyper-threading enabled, Package C-State set to C6(non retention state) and Processor C6 set to enabled, P-States set to default and SpeedStep and Turbo are enabled
Workload Generator: FIO 3.3, RANDOM: Workers-8, IOdepth- 8, No Filesystem, CPU Affinitized
Performance results are based on testing as of November 9, 2018 and may not reflect all publicly available security updates. See configuration disclosure for details. No product can be absolutely secure.

G2M RESEARCH

Blast Radius within SSD

- Average SSD capacity increasing, QLC NVMe SSDs offering lower TCO in “warm” storage (TCO$/TB/rack, IOPS/TB)
- First order principle - reduce device failure rate
  - We know how SSDs fail (hint, nothing like HDDs)
  - firmware errors, media errors, hardware failure, endurance (rare)
  - Implement new recovery mechanisms, make firmware more resilient
- Zoned Namespaces - Reduce size of failure domain
  - Previously drive included protection from media failures and better UBER, but better handled by upper layer software in distributed systems
  - Failure domain now can be zone (around the size of NAND EB)

G2M RESEARCH

Panel Questions and Audience Surveys

Panel Question # 1

- With the “blast radius” of individual SSDs growing beyond the needs of a single server (in many cases), how important do you think “networked SSDs”, SSD virtualization, and/or scale-out flash storage software will be to enabling the use of high-capacity SSDs?
  - Lightbits Labs
  - Kioxia
  - Intel

Audience Survey Question #1

- To what extent is the SSD “blast radius” problem an issue for your organization? (check one):
  - It is a significant problem across most of our mission-critical workloads today: 23%
  - It is a problem for a number of our mission-critical workloads today: 23%
  - It is a problem for a couple of our workloads today: 13%
  - It is not a problem today, but we expect it to be in the next 2-3 years: 26%
  - We do not see it as an issue for our workloads in the next 2-3 years: 15%

Panel Question #2

- Concepts such as computational storage and “smart SSDs” are seen as an important solution to the blast radius issue. How useful are these approaches to standard enterprise workloads?
  - Kioxia
  - Intel
  - Lightbits Labs

Audience Survey Question #2

- When looking at solutions for the blast radius problem, which of these approaches has your organization explored? (check all that apply):
  - Scale-Out Flash Storage (SOFS) software solutions: 42%
  - Distributed File Systems: 39%
  - Networked SSDs (Ethernet, NVMe-oF, etc.): 37%
  - Composable Infrastructure: 16%
  - Centralized storage arrays: 26%
  - Other: 11%

Panel Question # 3

- Most of today's solutions to the blast radius issue are "external" to the SSD. Are there options for new technologies "internal" to the SSD that can help solve the blast radius issue?
  - Intel
  - Lightbits Labs
  - Kioxia

Audience Q&A

G2M RESEARCH

Thank You For Attending

G2M

RESEARCH