

RIDESHARING YOUR CLOUD DATA CENTER

Realize Better Resource Utilization with NVMe-oF™

CloudFest | March 13, 2018

Steve Fingerhut, Senior Vice President, General Manager, SSD and Cloud Software Business Unit

Toshiba Memory America

Machine Learning Team

100%

utilized 24/7



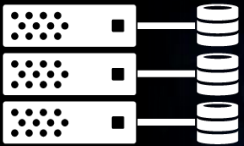
500%

idle Black Friday, year
Cyber Monday, etc.

Online Marketplace Team

What's wrong with my cloud data center?

DAS



direct-attached
flash

PROBLEM 1



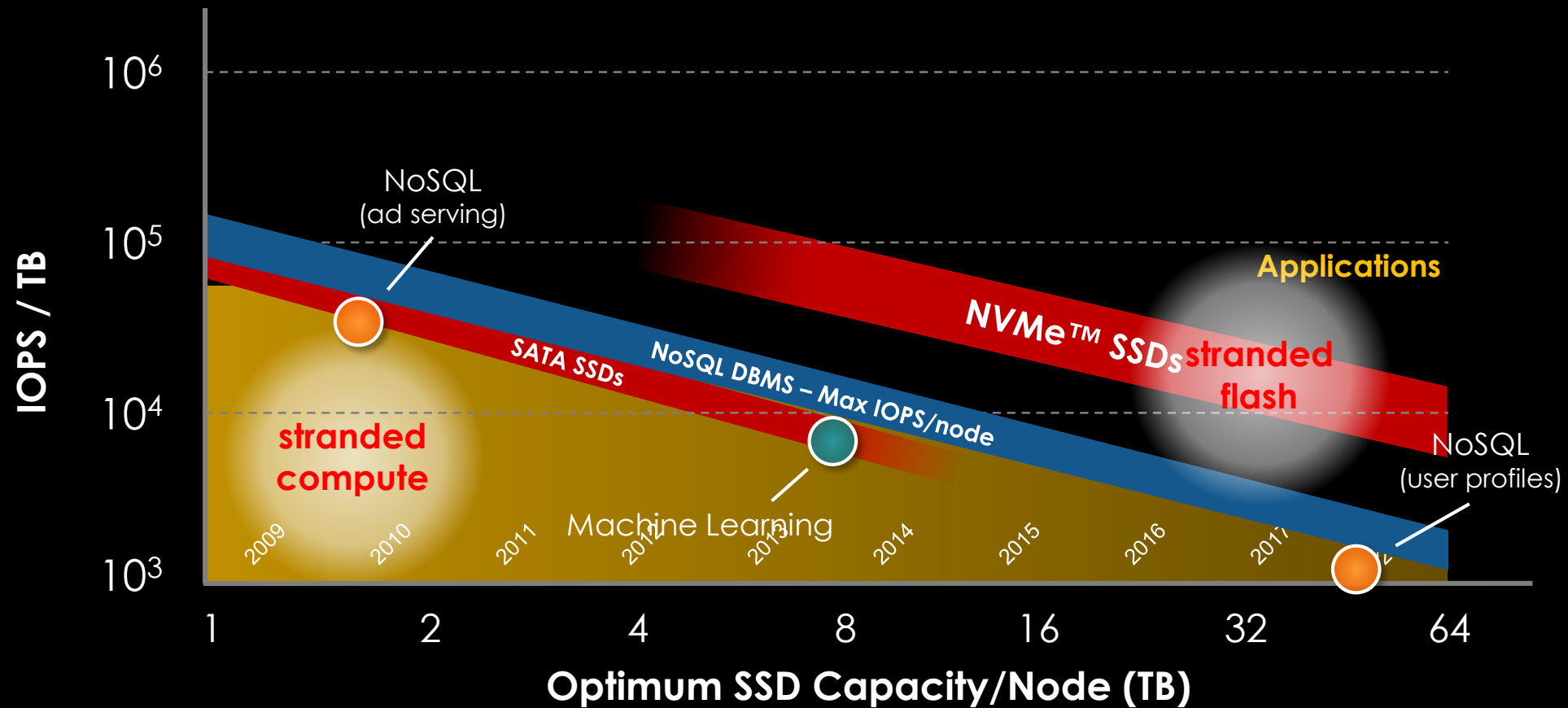
stranded flash & IOPS

PROBLEM 2



stranded compute

How much SSD capacity per node?



Too much wastes flash. Too little wastes CPU.

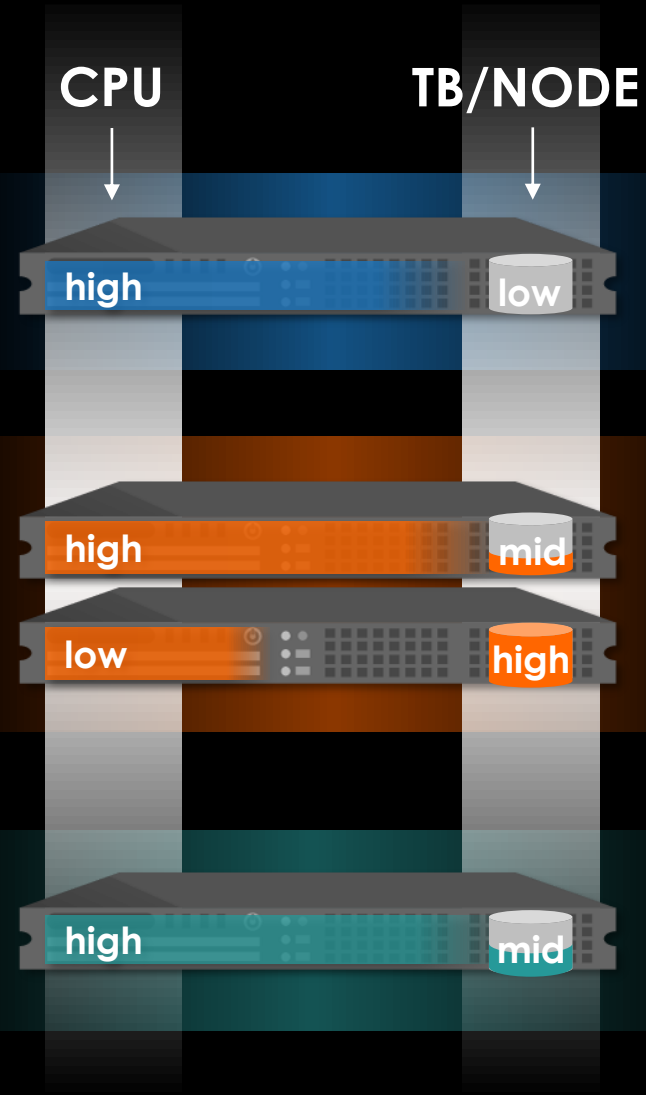
One size does not fit all.

Workload Variability

STATELESS Web Server

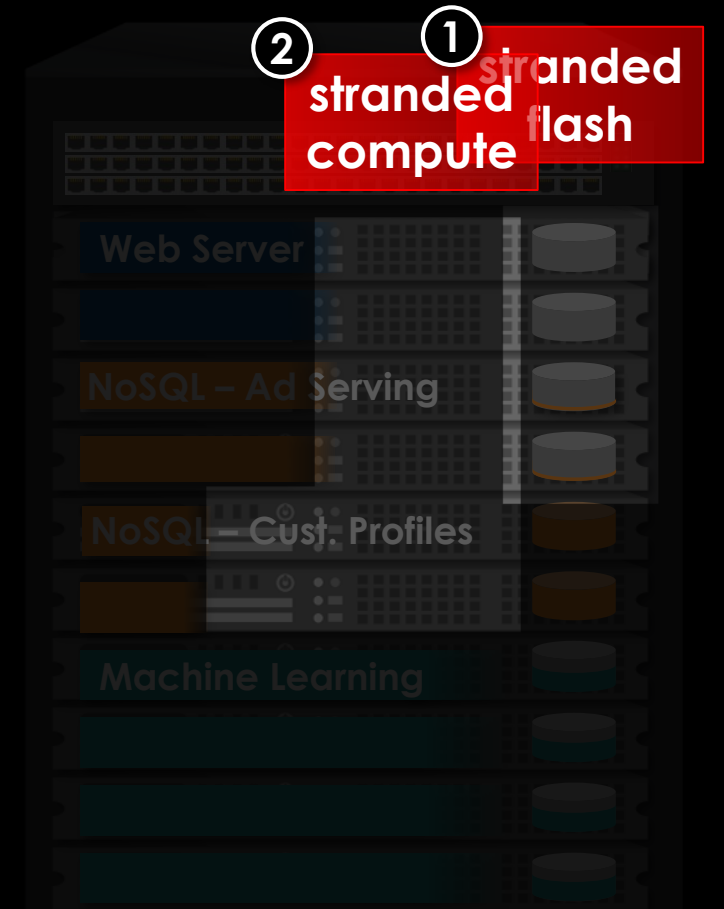
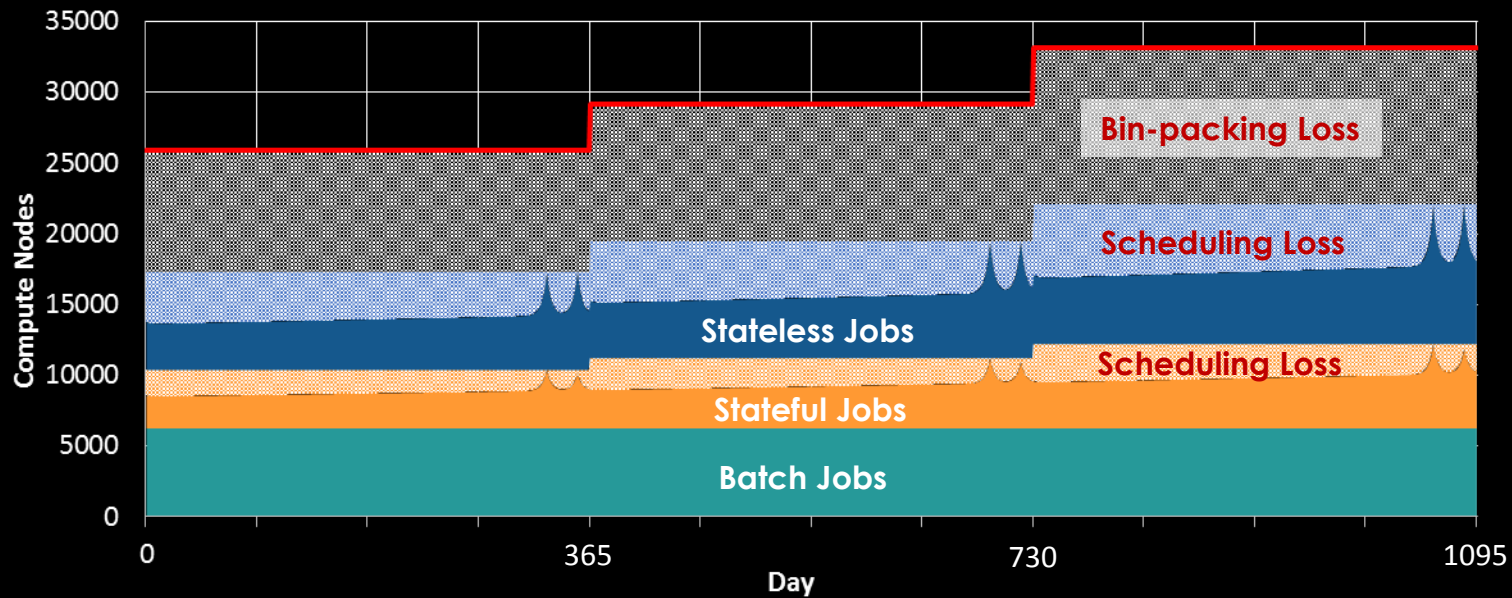
STATEFUL NoSQL DB (ad serving)
NoSQL DB (user profiles)

BATCH Machine Learning (shopping suggestions)
Hadoop® (big data)



Unique resource mixes: CPU intensity, flash capacity, IOPS, memory
Tough to optimize node configurations

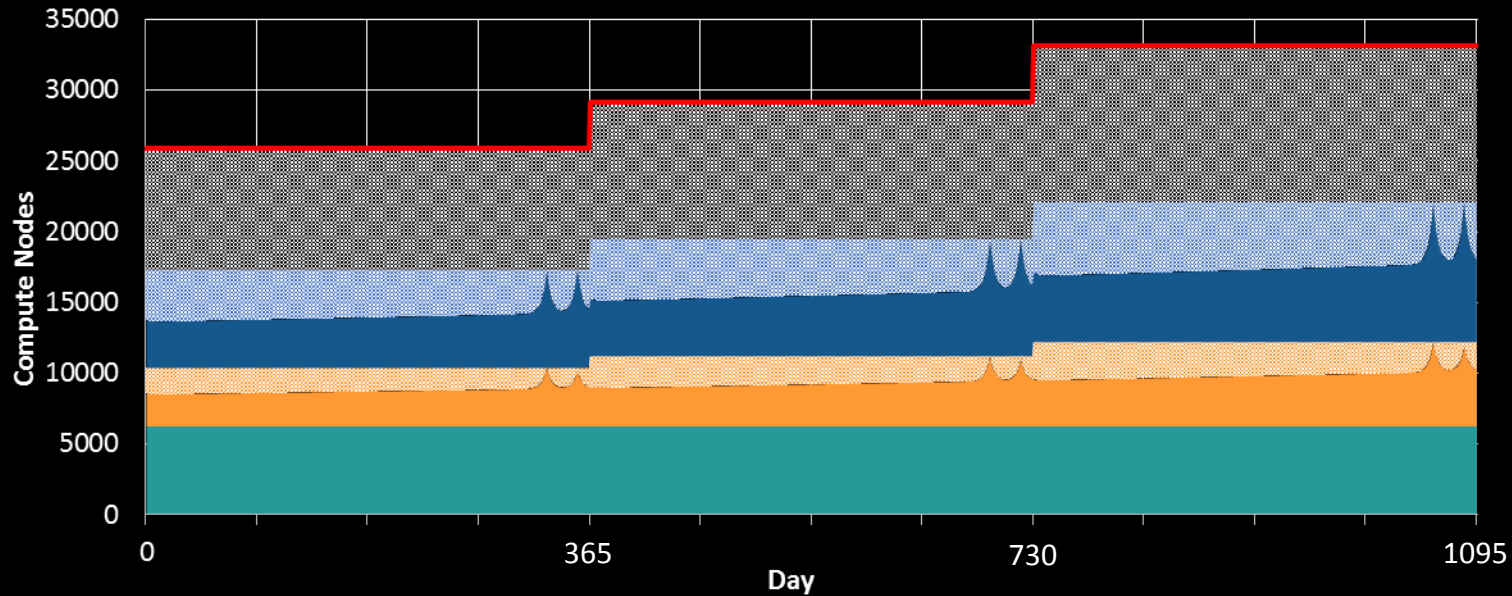
Application Silos with DAS Strands Resources



Challenges:

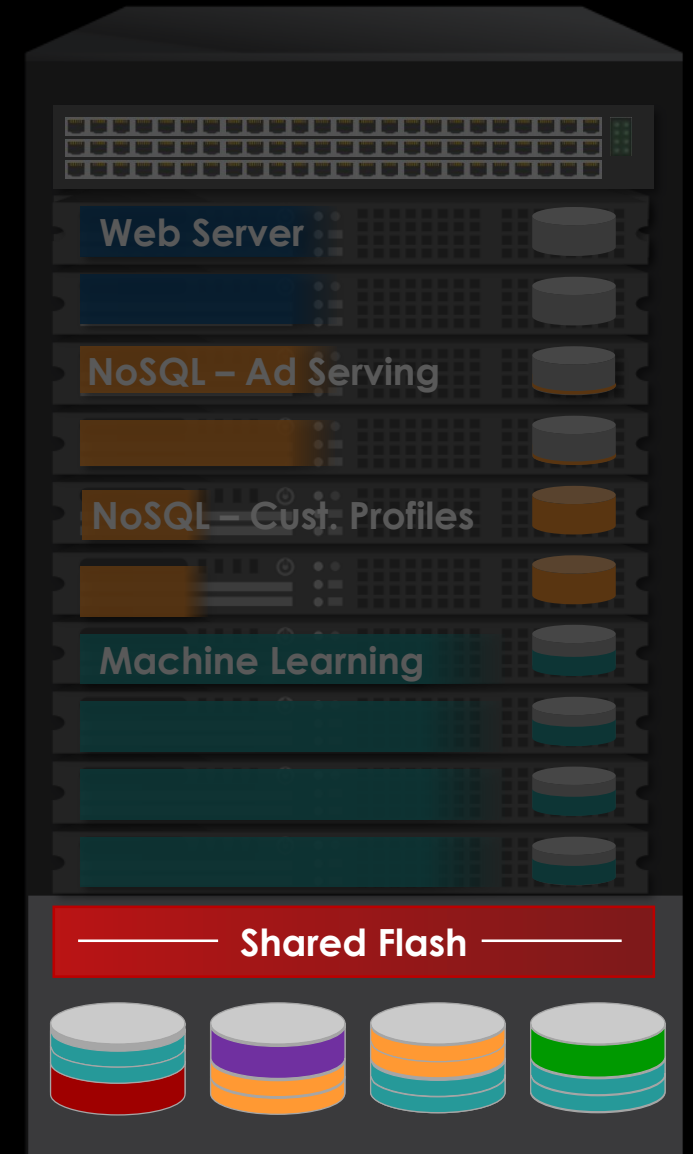
- ① Variable resource needs \Rightarrow **stranded flash**
- ② Reserve capacity for demand peaks \Rightarrow **stranded compute**

Disaggregation Optimizes Use of Flash



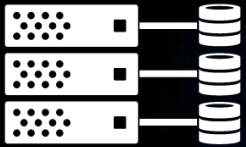
Benefits:

- Storage can now be provisioned optimally for every application
- Eliminates stranded flash
- Significant improvement in bin-packing loss (stranded compute)



What's wrong with my cloud data center?

DAS



direct-attached
flash

PROBLEM 1



stranded flash & IOPS

PROBLEM 2



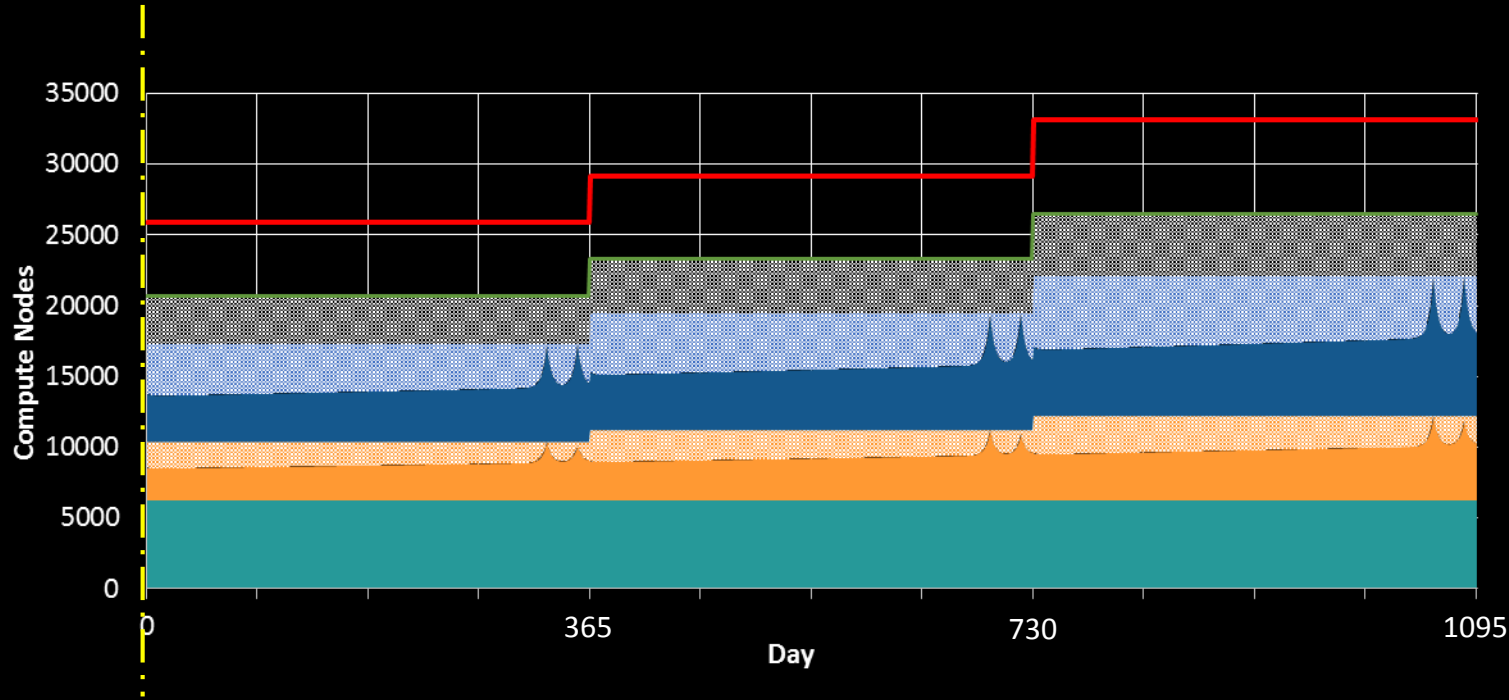
stranded compute

PROBLEM 3



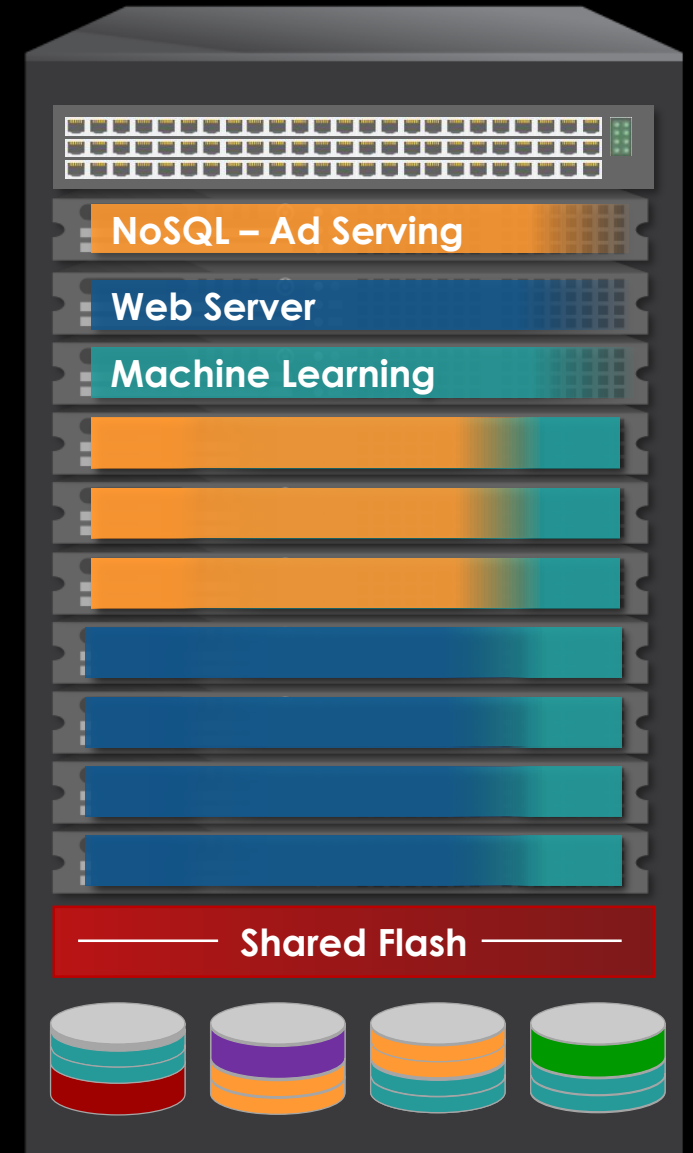
lost operational
agility & revenue

Disaggregate and Orchestrate for Agility



All applications can now be scaled out quickly – even stateful ones

- “Borrow” compute resources from lower priority applications during demand peaks
- Possible *only if any application can run on any node* – unrestricted by data locality



What Flash-enabled Agility Looks Like in Your Data Center



Provides optimal platforms for more applications with fewer node types

30-60%
fewer compute nodes
for the same workload

Eurosys 2015: "Large-scale cluster management at Google with Borg"

– Google –

Flexible, fast response to changing business conditions



Shared Accelerated Storage Solves DAS Problems

DAS



direct-attached
flash

PROBLEM 1



stranded flash & IOPS

PROBLEM 2



stranded compute

PROBLEM 3

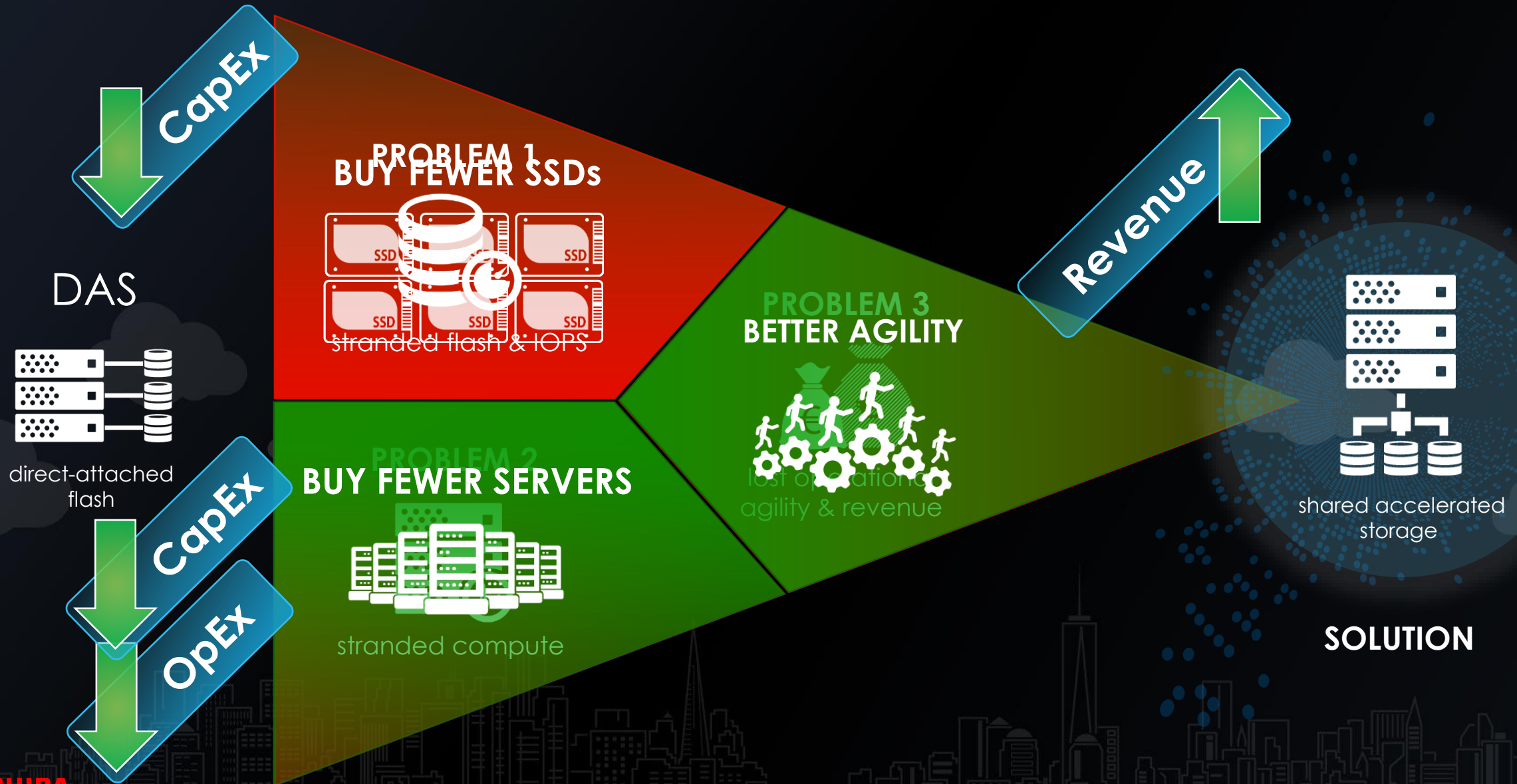


lost operational
agility & revenue

shared accelerated
storage

SOLUTION

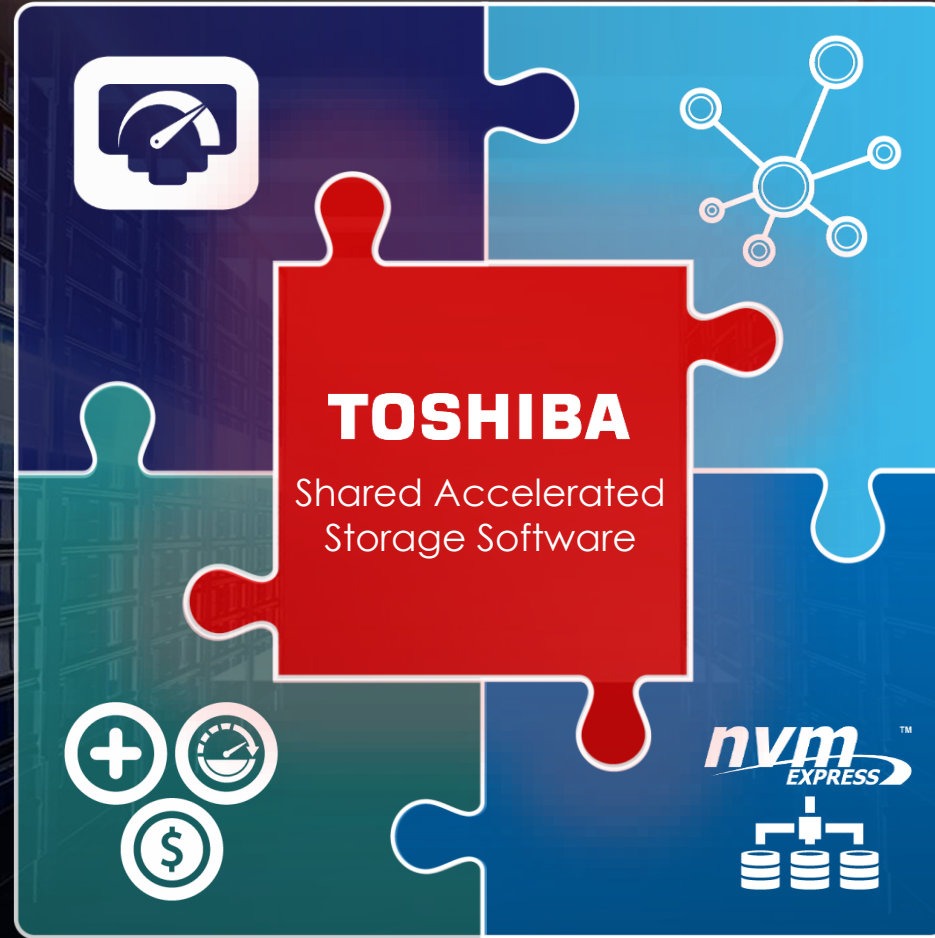
Shared Accelerated Storage Solves DAS Problems



Shared Accelerated Storage for Cloud is Now Possible

High bandwidth,
low latency
networks

Mature
orchestration
frameworks

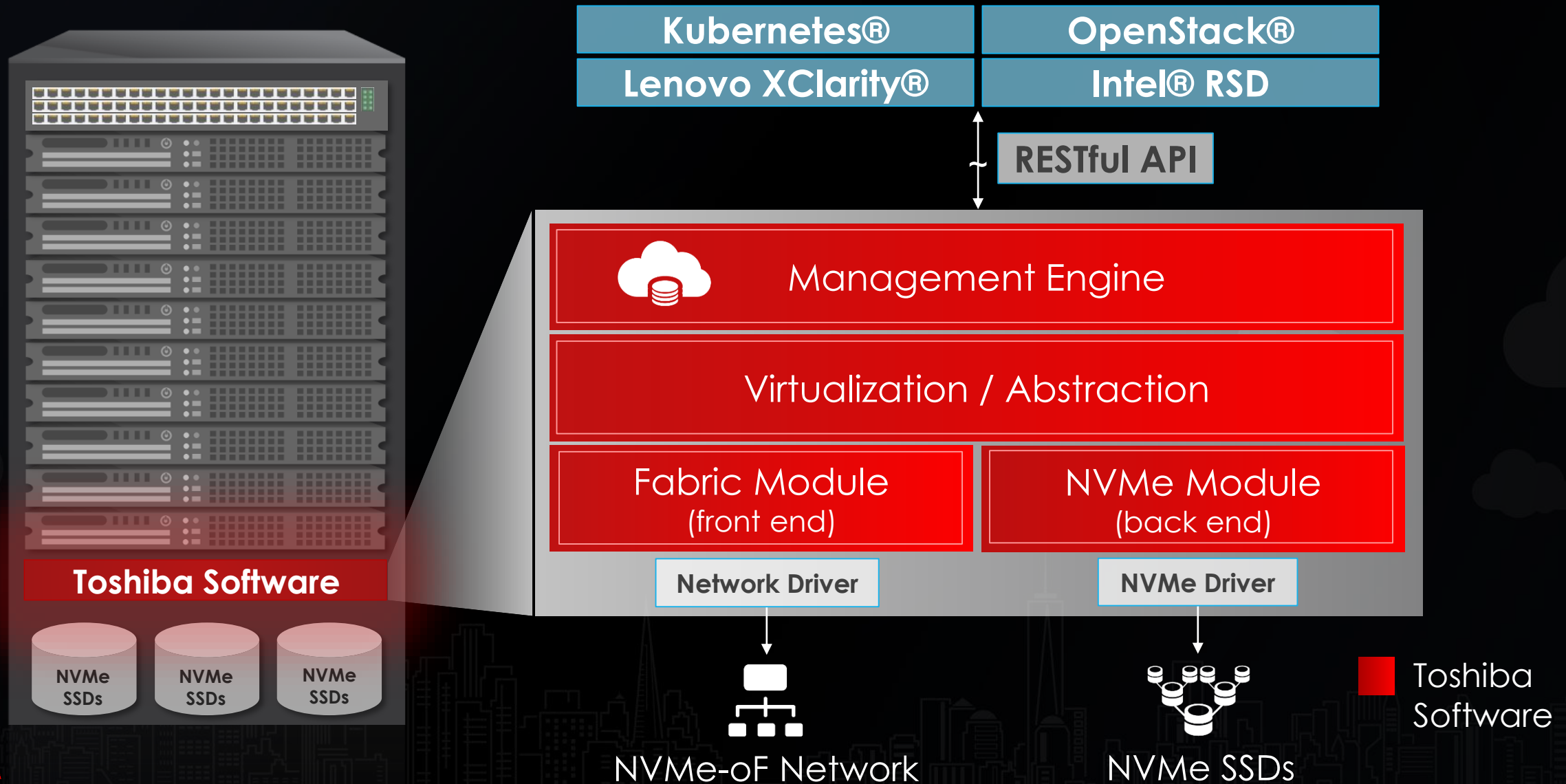


Bigger, faster &
more cost-effective
NVMe™ SSDs

NVMe-oF™
protocol

maximum data center efficiency

Toshiba Shared Accelerated Storage Software



Toshiba Software Key Benefits



Fast

<20 μ S
latency adder

8,000,000+
4K RR IOPS



Flexible

4,000+
namespaces

384 TB
SSD capacity



Future-Proof

Popular
cloud frameworks

Broad
ecosystem support



Optimized

100%
elastic

Zero
stranded capacity

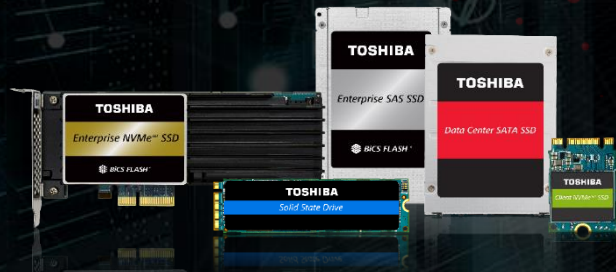


Inventor of **NAND flash memory**

Cutting-edge 64/96 layer
3D flash technology

Broad **solid state drive portfolio**

NVMe, SAS and SATA



TOSHIBA

World's largest **flash fab**

#2 in flash

Decades of storage experience

Major supplier to 10 of top 10
server & storage vendors

TOSHIBA



Thank You!

Visit Toshiba at Booth #C04

TOSHIBA
Leading Innovation >>>