

IBM Storage Directions

**Common Romandie,
Genève, 7 Décembre 2018**

Dominique Salomon

IT Specialist Storage,

Middle East & Africa (MEA) Technical presales support.

IBM Client Center, Montpellier – France

Dominique.Salomon@fr.ibm.com

Last Updated: December 2018



Evolution générale



Three ways how IT uses data ... today

Procedural (if...then)

```
159 </tr>
160 </table>
161 </div>
162 </div>
163 </body>
164 <script type="text/javascript">
165 <!--
166 var currentImage = "bigImage1";
167 var pages = Math.ceil(photos.length / 9);
168 updatePages();
169 updateAllImages();
170 // document.getElementById("bigImage0").src = "images/wie
171 // document.getElementById("bigImage0").style.display = "
172 changePhotoDescription( "1" );
173
174 function updatePages() {
175     var j = 0;
176
177     var html = '<table style="width: 330px;" cellspacing="
178     if ( page != 0 ) {
179         html = html + '<a href="#" onclick="page=0; up
```

Statistical (big data)



Machine learning



Image: opendatascience.com

... and in 10 years

Statistical
(big data)

Machine learning

Procedural
(if...then)



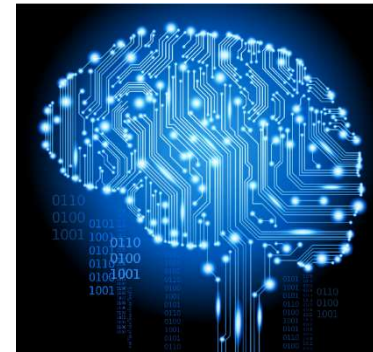
Image: Business over Broadway

Image: opendatascience.com

Cognitive Computing, Machine learning

- The simulation of human thought processes in a computerized model.
- Cognitive computing involves self-learning / machine learning systems that use data mining, pattern recognition and natural language processing to mimic the way the human brain works.

- Observe, Interpret, Evaluate, Decide
- Penetrate and analyze big data
- Solving “soft” problems
- Ever-changing data



Current examples



business as usual
classic / legacy IT



shopping, profiling,
fraud detection ...



autonomous driving,
image classification,
chatbots, gaming...



Image: Business over Broadway

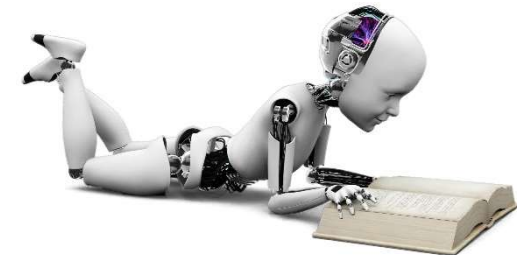


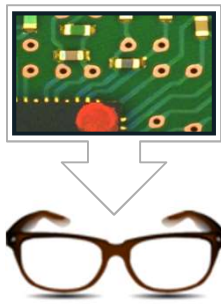
Image: opendatascience.com

Manual modelling

Accumulation of examples

Automatic modelling

How is data stored?

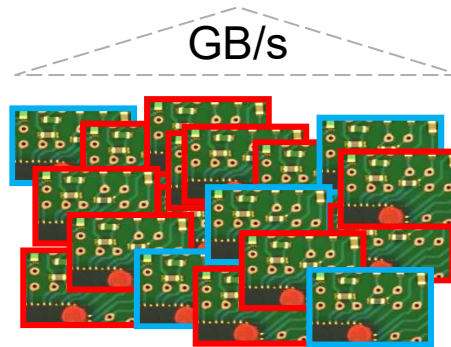


if...then...else

Procedural:
Archive test cases
for auditing



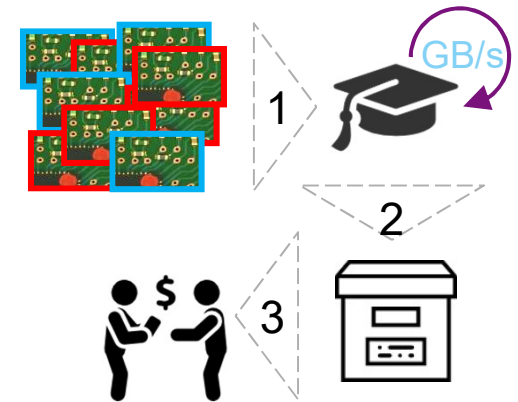
Image: Business over Broadway



Statistical:
Parallel processing of
many stored samples

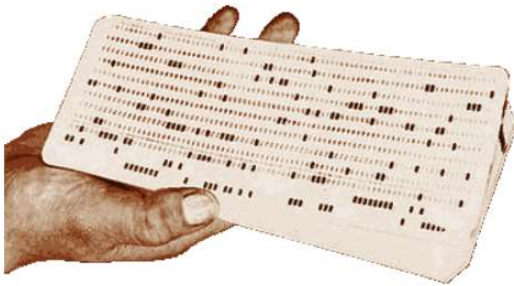


Image: opendatasience.com

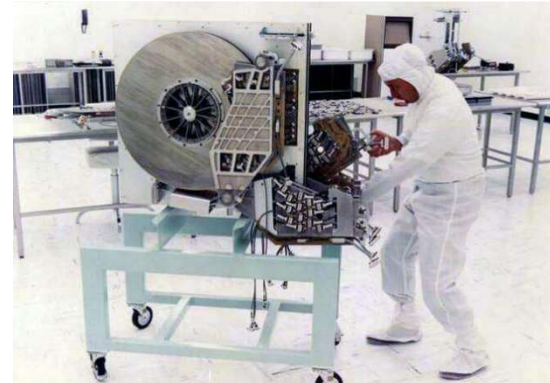


Machine Learning:
Train sample data, then
archive or trade data

Data storage : A bit of History ... and evolution to Flash Storage



© Copyright IBM Corporation 2018



Every generation has a defining industry



- Need To Re-Think How we are Doing Things in Today's World

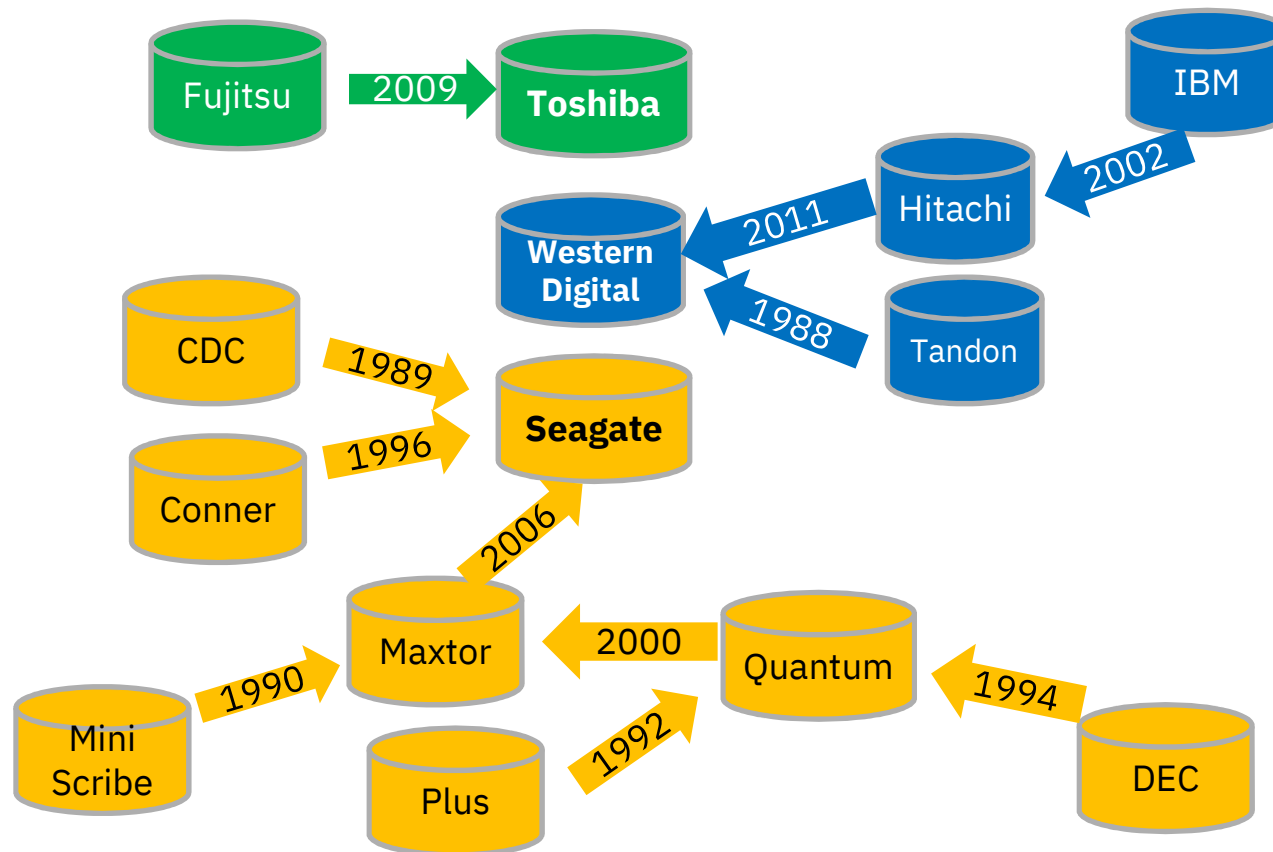
What Happens in an Internet Minute?



And Future Growth is Staggering



Consolidation of Disk Drive Vendors

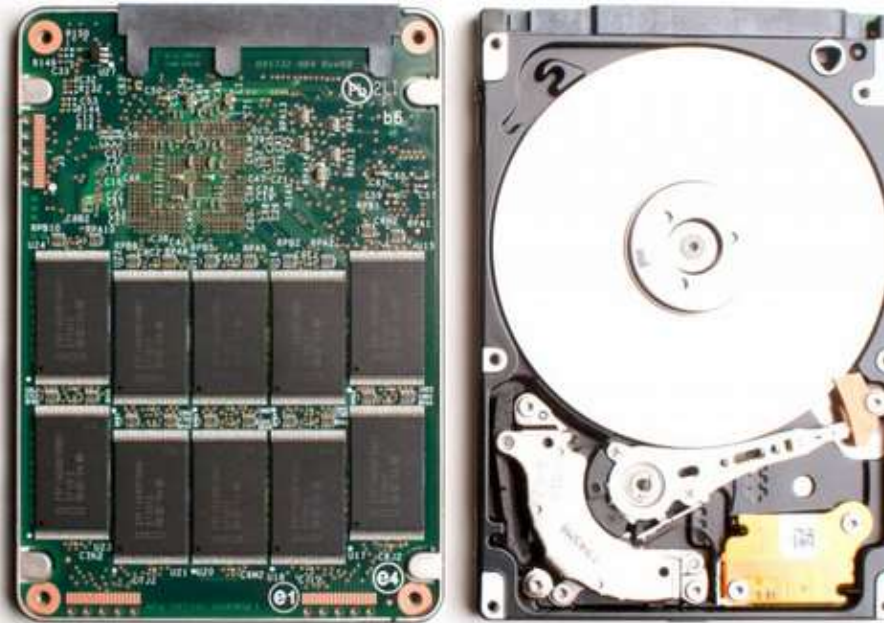


Evolution vers le Flash













SSD vs HDD

(Solid State Drive)

(Hard Disk Drive)



Seagate Offerings for Datacenter as of 26 APR

Exos™ 7E8	Exos 7E2	Exos 5E8	Exos 7E2000	Exos 10E2400	Exos 15E900
					
Formerly Enterprise Capacity 3.5, v 5.0	Formerly Enterprise Capacity 3.5, v 5.1	Formerly Archive HDD v3	Formerly Enterprise Capacity 2.5 HDD v3	Formerly Enterprise Performance 10K HDD v9	Formerly Enterprise Performance 15K HDD v6
Datasheet  Exos 7E8	Datasheet  Exos 7E2	Datasheet  Exos 5E8	Datasheet  Exos 7E2000	Datasheet  Exos 10E2400	Datasheet  Exos 15E900
Perfect for Mainstream bulk data storage	Perfect for Most economical enterprise drive per spindle	Perfect for Archival data	Perfect for Enterprise-class boot drives, blade and dense servers	Perfect for Mission-critical server and storage arrays	Perfect for Mission-critical server storage
Capacity 8TB, 6TB, 4TB, 2TB, 1TB	Capacity 2TB, 1TB	Capacity 8TB	Capacity 2000GB, 1000GB	Capacity 2400GB, 1200GB, 1800GB, 900GB, 600GB, 300GB	Capacity 900GB, 600GB, 300GB
Interface SAS, SATA	Interface SATA	Interface SATA	Interface SAS, SATA	Interface SAS	Interface SAS

Western Digital 14TB - Oct. 3, 2017

Ultrastar DC HC600 SMR Series

Purpose-Built Efficiency for
Sequential Write Workloads

Ultrastar 14TB

3.5-inch Helium Platform

7 200 Rpm

SATA 6Gb/s – SAS 12Gb/s

Enterprise **SMR** Hard Drive

SMR technology boosts capacity but data must be written sequentially so system customization is required; Supports random reads



<https://www.hgst.com/products/hard-drives/ultrastar-hs14>

Memories : a little history

DRAM (1966)

IBM Fellow Robert H. Dennard's crazy idea:

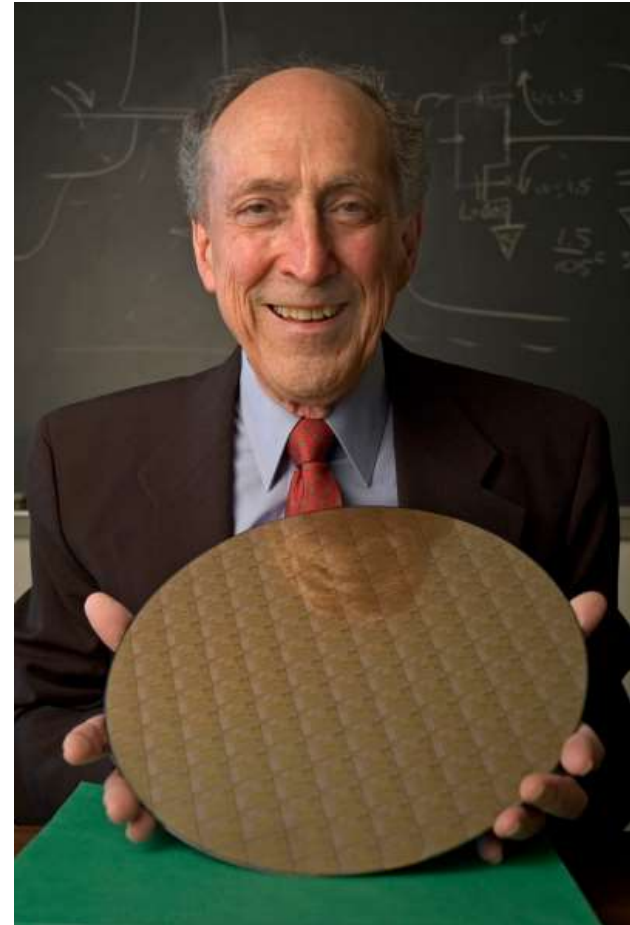
Capacitive cells on a chip that forget their content after a few seconds.

-but-

Periodically refresh the content before it fades. (DRAM = "Dynamic" RAM)

Patent [US3387286](#):

Field-effect transistor memory, 4. June 1968

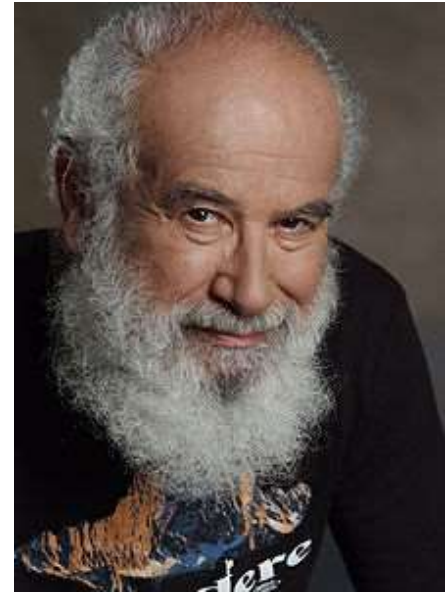


EPROM (1970)



Intel's Dov Frohman's idea:

Erasable programmable ROM - **EPROM**
*Trap charge so it can't escape, except
using UV or x-ray light.*

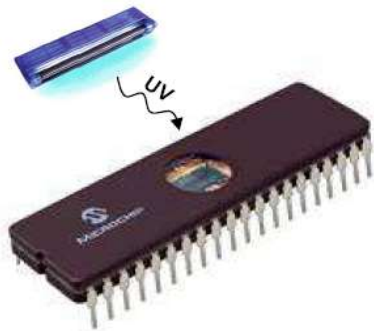


Dov Frohman, Intel

Flash technologie History

Modern Flash-Memory is built from EEPROM Blocks

EPROM (1970)



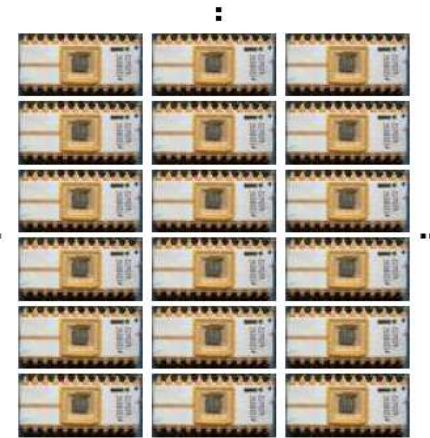
Delete with UV,
rewrite with 12 V

EEPROM



Delete with 12V,
rewrite with 5 V

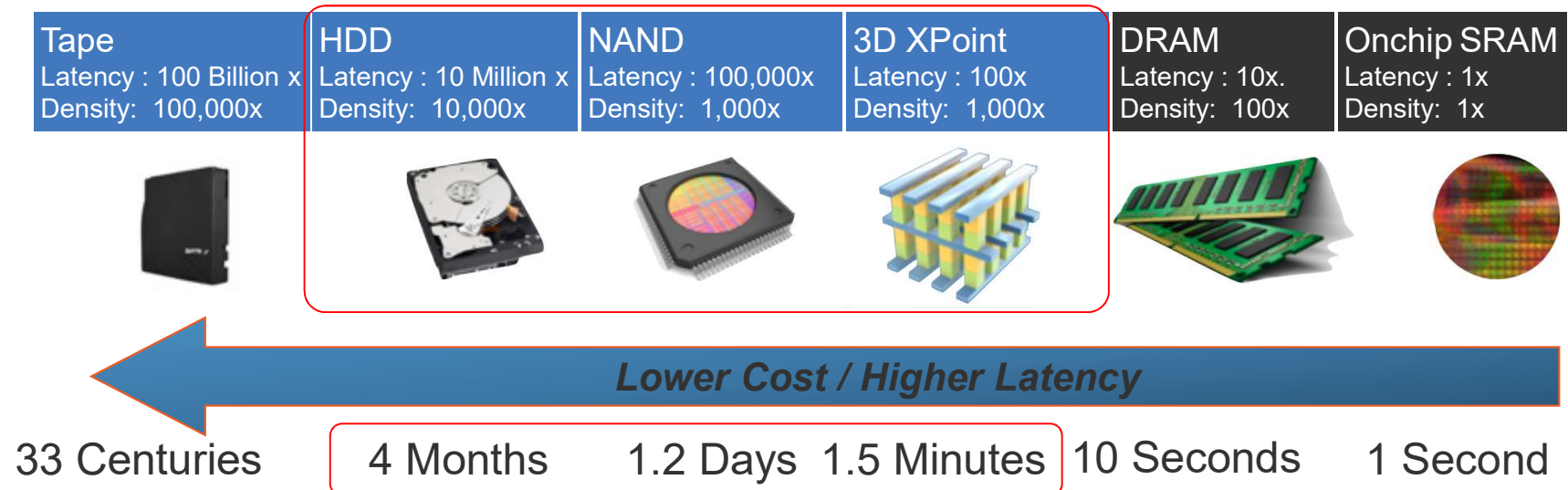
Flash Memory



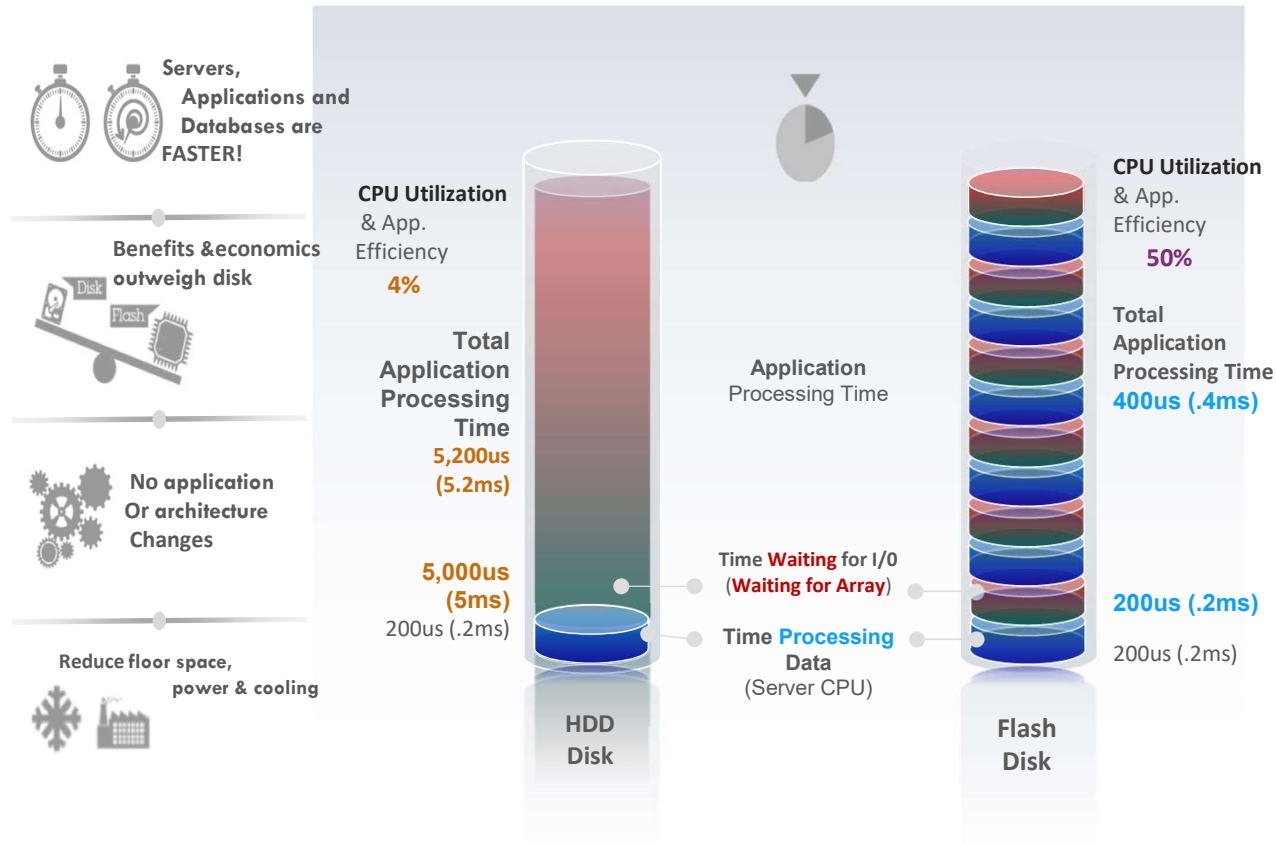
Block-wise
deletion
with 10V



The Storage Media Toolkit



Why it is so critical to reduce the **storage atency**



HDD vs SSD performances

	NL-SAS 7,200	SAS 10,000	SAS 15,000	SSD Consumer
Read IOPS	100 (150)	200 (320)	300 (500)	~5,000
Write IOPS	100 (120)	200 (275)	300 (400)	~1,000
Read MB/s	100 - 180	120 - 200	175 - 200	200 - 500
Write MB/s	100 - 180	120 - 200	150 - 200	200

http://www.theregister.co.uk/2016/01/19/toshibas_tweaked_endurance_wrings_out_low_write_ssd/

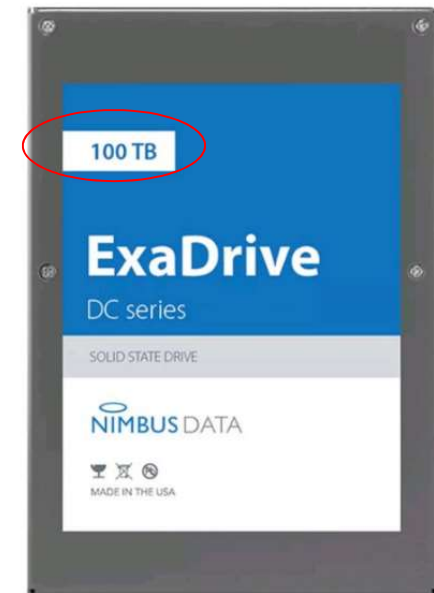
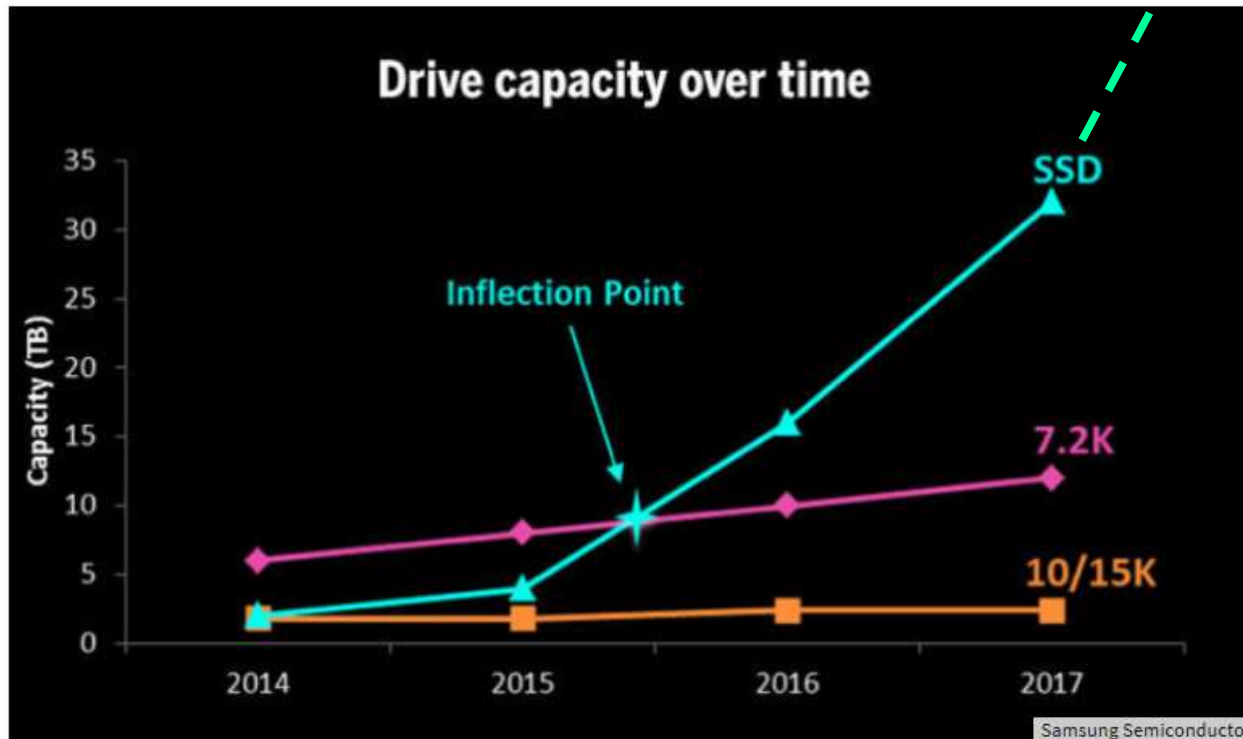
Toshiba Enterprise SSD disks

Product	Capacity	Random Read IOPS (4k)	Random Write IOPS (4k)	Seq Read	Seq Write	Endurance
PX04SHB	200/400/800GB - 1.6TB	270,000	125,000 (1.6GB 120,000)	1.9GiB/sec (1.6TB - 1.5GiB/sec)	850MiB/sec (1.6TB - 750MiB/sec)	25 DWPD
PX04SMB	400/800GB - 1.6/3.2TB	270,000	90,000 (3.2TB - 85,000)	1.9GiB/sec (3.2TB - 1.5GiB/sec)	850MiB/sec (3.2TB - 750MiB/sec)	10 DWPD
PX04SVB	480/960GB - 1.92/3.84TB	270,000	60,000 (3.84TB - 55,000)	1.9GiB/sec (3.84TB - 1.5GiB/sec)	850MiB/sec (3.84TB - 750MiB/sec)	3 DWPD
PX04SRB	480/960GB - 1.92/3.84TB	270,000	22,000	1.9GiB/sec (3.84TB - 1.5GiB/sec)	850MiB/sec (3.84TB - 750MiB/sec)	1 DWPD
PX04SLB	2TB, 4TB	270,000	19,000	1.9GiB/sec (4TB) 1.5GiB/sec (2TB)	850MiB/sec (4TB) 750MiB/sec (2TB)	0.5 DWPD

Solid state drives are now larger than hard disk drives

In 2015, the capacity of HDDs was surpassed for the first time by SSDs and because SSDs are scaling at a faster rate than HDDs, we will never look back

The largest SSD built in 2018 :



How Flash Gets Denser (and Cheaper) Flash Devices :

■ 1. Denser Lithography

- Lower endurance
- Limits to improvements

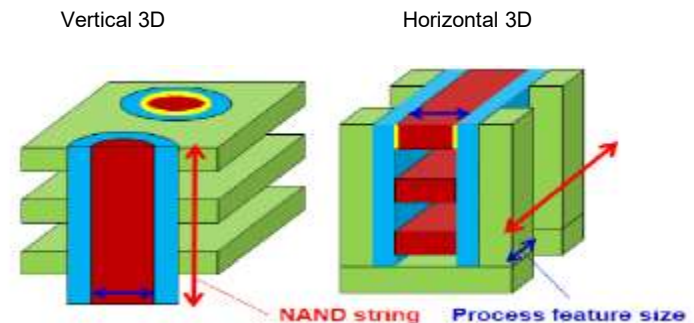
■ 2. More Bits per cell

- Signal processing to discriminate between 2, 4, 8 or more signal levels when a cell is read.
- Improvement in ~2X density when a step is taken (about every 3-4 years)
- Lower Endurance
- Now on 4 bits per cell (QLC)
- Not clear if we ever get 5 bits per cell for Enterprise

■ 3. Skyscraper (3D)

■ Who builds higher?

- Larger cells, higher endurance, MLC ok to use in 'healthy' 19~20 nm lithography
- Current best is 64 high. 200 or more high may be possible.

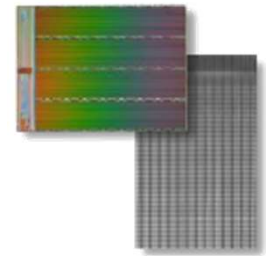


Source: Akira Goda, Micron: "Opportunities and Challenges of 3D NAND Scaling"

2D Planar NAND → 3D NAND

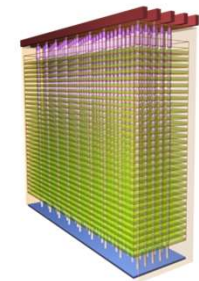
2D Planar NAND has limitations

- Implemented in a plan, XY dimensions
 - Dies are stacked to increase capacity at a package level, but this can only be extended so far
- Physical size of the NAND cell is at its limits (15nm current commercial limit)
 - Gate is small enough that minimal electrons, don't allow for significant separation in voltage levels, and can be greatly effected by leakage due to wear, which results in limit P/E cycling
 - Fine for SLC and MLC, but difficult for TLC, and likely not possible for QLC



3D NAND

- Implemented in XYZ dimensions
 - Can take advantage of die stacking to increase density
- Technology node size increased, which allows for more electrons on the gate of each cell
 - Increases P/E cycles, and enables larger delta between each voltage level per bit
 - Enables TLC endurance higher than planar MLC, and enables QLC!



2018 Flash News

64-layer **3D NAND Quad-Level Cells** (4 bits/cell)

256GB, 512GB and 1TB drives.

355,000 random write IOPS and 365,000 random read IOPS

3.18GB/sec reads and 2.96GB/sec writes

Five-year warranty and a 1.5 million hours MTBF rating.

and NVMe ...

https://www.theregister.co.uk/2018/07/23/toshiba_wd_96_layer_flash/



En résumé : la proposition de valeur du Flash

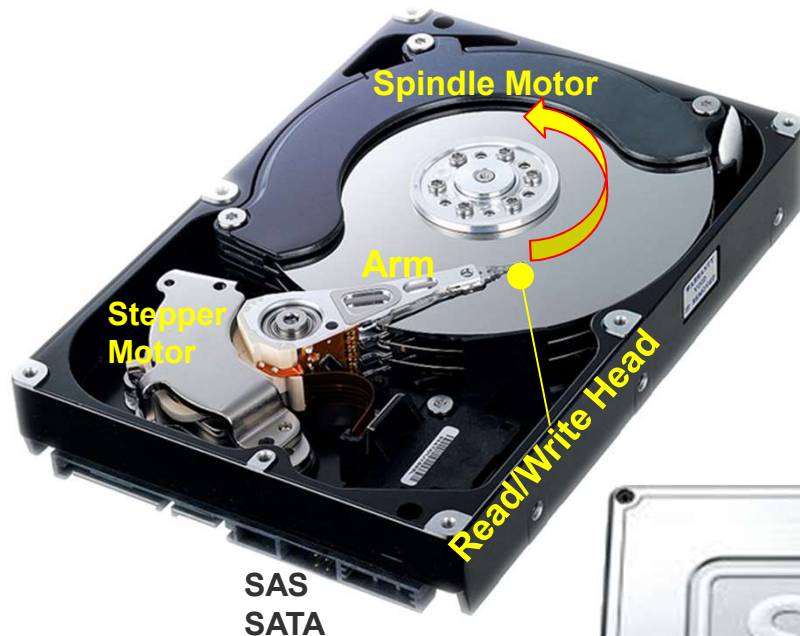
- **Latence** beaucoup plus faible que les technologies disques (intérêt pour OLTP, par exemple)
- **Densité d'I/Os** très importante (intérêt pour les batches, le VDI, par exemple)
- **Place au sol** et impact environnemental très faibles (intérêt pour IT hébergée ou DC contraint)

 <p>Réduction de 85% Temps d'exécution des batches</p>	 <p>Réduction de 90% Temps de réponse OLTP</p>	 <p>150-200 µs Latence</p>
 <p>Réduction de 75% de la place au sol Stockage d'un PB dans un seul rack. En activant la compression, stockage de deux fois plus</p>		 <p>Réduction de 80% Consommation d'énergie</p>

	Enterprise SSD (3,2TB)	SSD Haute Densité (15TB)	Disque HDD 15krpm
kIOPS Lectures aléatoires (4KB)	350	190	0,2 – 0,4
kIOPS Ecritures aléatoires (4KB)	300	30	0,2 – 0,4
MB/s Lectures séquentielles (64KB)	3000	1200	250
MB/s Ecritures séquentielles (64KB)	2000	900	250

NB: Données constructeur Toshiba pour les SSDs

!!! SSDs – Flash Pretending To Be Spinning Disk



When Reading or Writing to the HDD, instructions must be sent which tell the drive head where it needs to move with respect to Data Tracks, Cylinders and Sectors.

Same communication protocols are required

When Reading or Writing to an SSD, in spite of not having any moving parts, instructions must be sent which tell the non-existent drive head where it needs to move with respect to Data Tracks, Cylinders and Sectors.

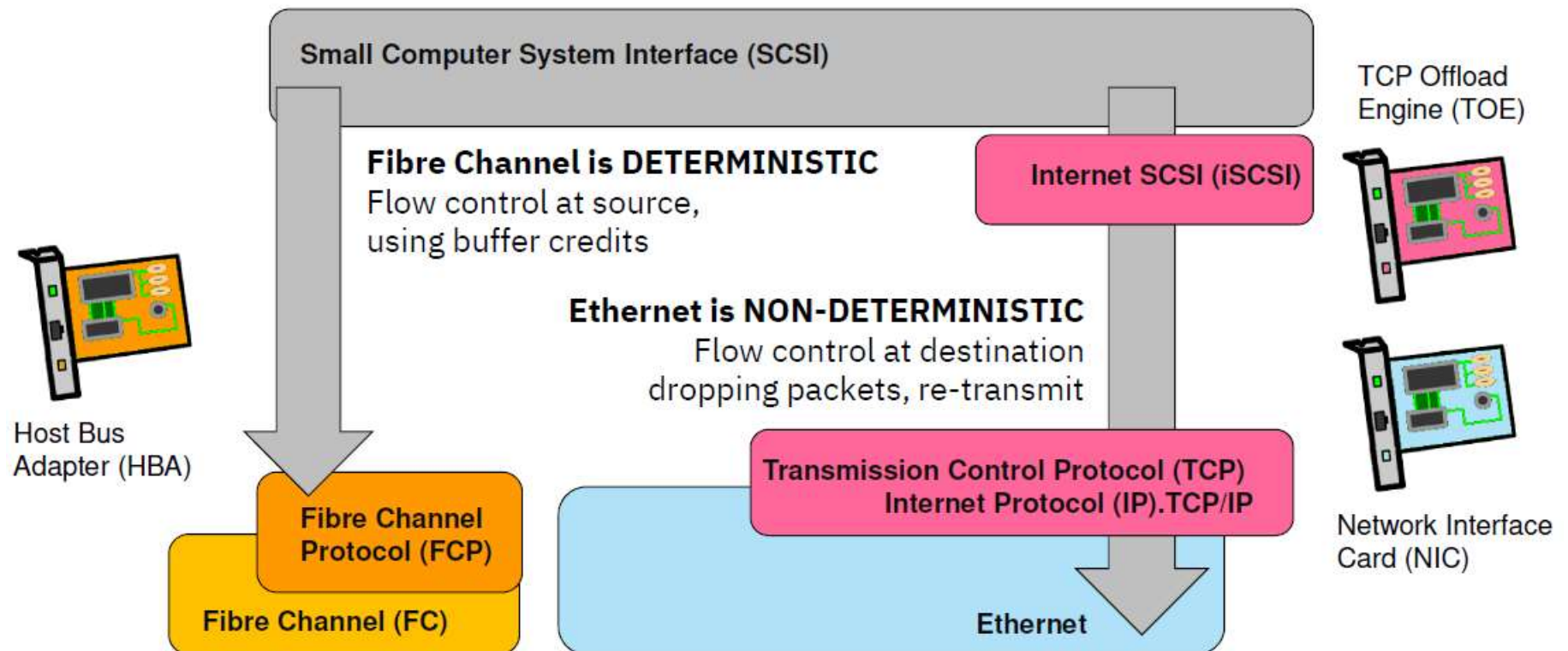
This is because it must follow the same communications protocol as the HDD it is mimicking.

Same Protocols !!

SAS
SATA



How SCSI commands get sent from Servers to Storage Devices



NVMe SSDs – Removing Bottlenecks



SCSI was created to deal with Hard disks HDDs.

This means that much of its code was designed to deal with rotating disk platters and physical disk head movement which can only **handle a single request at a time.**

This was ideal for disks but wholly **inefficient for dealing with solid-state devices.**



Regardless of their form factor, **NVMe (Non Volatile Memory – Express)** eliminate the **bottleneck of SCSI protocols and the old HDD interfaces** like SAS or SATA in order to deliver improved performance and lower latency with Flash Devices like SSDs

<http://nvmexpress.org/resources/specifications/>

Les protocoles SCSI et NVMe



Protocole SCSI

- 1982 (Small Computer System Interface)
- Unités logiques sur un bus périphérique
- Conçue pour des **HDDs**.
- 2005: SAS (Serial-Attached SCSI)
- Mono canal, mono file d'attente (Queue).
Traitements de commandes en série.
- 32 commandes
- Composants logiciels à tous les niveaux

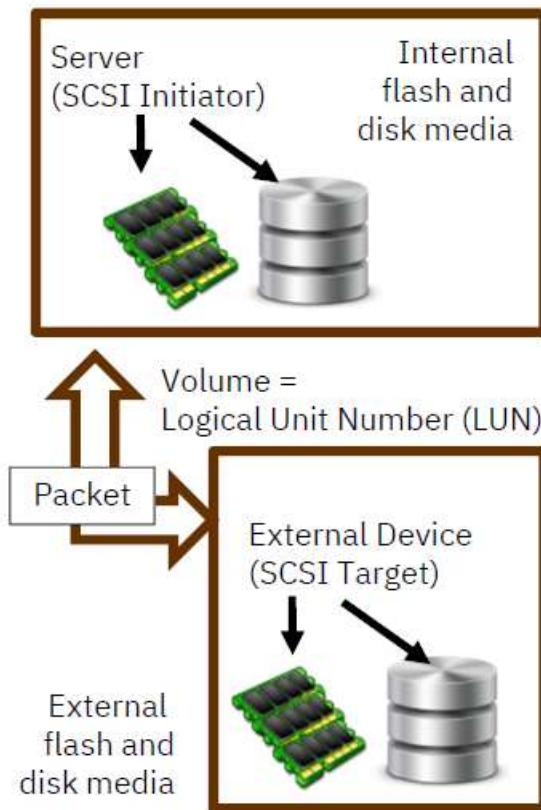


Protocole NVMe

- 2011 (**Non-Volatile Memory + PCI"e"**)
- Conçu comme une extension mémoire, pour des supports **Flash**.
- Traitements en **parallèle**, utilisation des CPUs multi cœurs.
- Accès direct aux supports Flash en PCIe (suppression des bus SAS ou SATA).
- **Jusqu'à 64 000 files d'attente** (Queues) et 64 000 commandes par file.
- Chaque cœur du contrôleur gère des files d'attente dédiées par SSD.

Le but de NVMe: réduire la latence et augmenter le débit en s'adaptant à la technologie Flash.

NVMe – An Alternative to SCSI command set

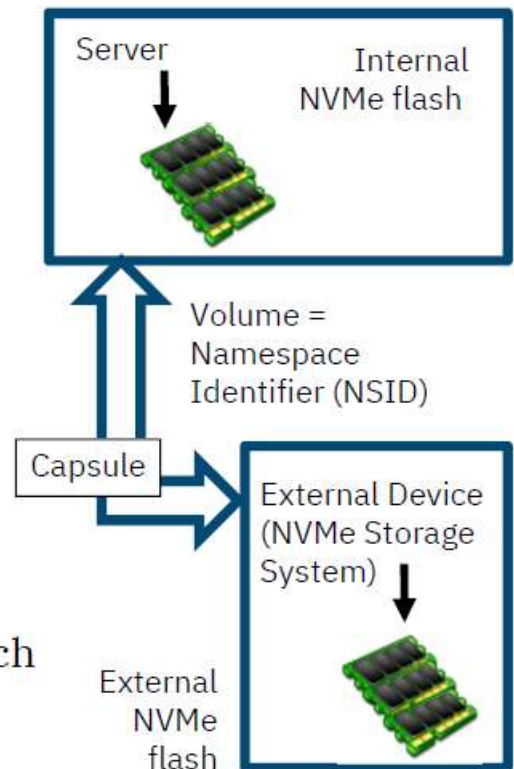


Small Computer System Interface (SCSI)

- 34 commands
- For both flash and disk
- Can be sent over network such as FCP, iSCSI, FCoE
- Terminology: LUN, Packet

Non-Volatile Memory Express (NVMe)

- 15 commands
- For flash only
- Can be sent over network such as FC-NVMe or NVMe-OF
- Terminology: NSID, Capsule



How NVMe reduces "gate latency": By **Parallelism** !



Serialized

vs.



Full parallel

NVMe extension on the Network : Clarifying Two Key Terms

- **NVM Express (NVMe™ - Non Volatile Memory Express)...**
 - ...Is the command protocol to address **non-volatile memory used with PCIe**

From the specification:

NVMe™ is a scalable host controller interface designed to address the needs of Enterprise and Client systems that **utilize PCIe-based SSDs**. The interface provides optimized command submission and completion paths. It includes support for **parallel operation** by supporting up to **65,535 I/O Queues** with up to 64K outstanding commands per I/O Queue. Support has been added for many Enterprise capabilities such as end-to-end data protection (compatible with SCSI Protection Information, commonly known as T10 DIF, and SNIA DIX standards), enhanced error reporting and virtualization.

- **NVMe over Fabrics™ (NVMe-oF™)...**
 - ...Is the command protocol to address **non-volatile memory using a non-Pcie network** which can be **FC, Ethernet or InfiniBand**

RDMA Explained* - Necessary For Speed



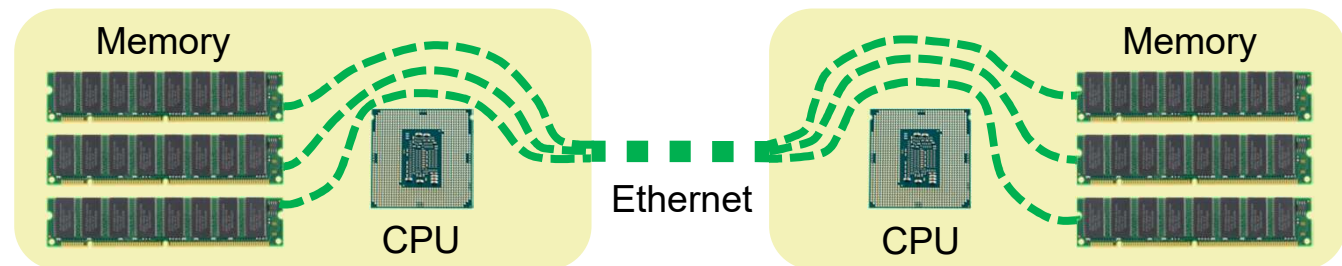
RDMA = Remote Direct Memory Access

- **Direct Memory Access (DMA)** is an ability of a device to access host memory directly, without the intervention of the CPU.
- **Remote Direct (or Dynamic) Memory Access (RDMA)** is the ability of accessing (read, write) memory on a remote machine without interrupting the processing of the CPU(s) on that system.

RDMA is a way of moving buffers between two apps across a network.

RDMA differs from traditional network interfaces because it bypasses the operating system. This allows programs that implement RDMA to have:

1. The absolute lowest latency
2. The highest throughput
3. Smallest CPU footprint



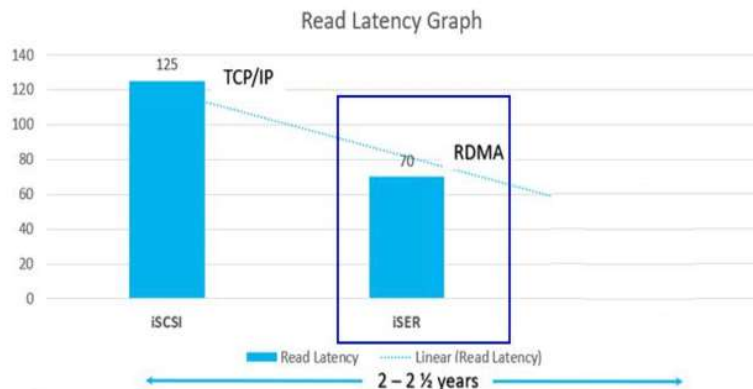
Benefits of NVMe-oF

- **NVMe-oF** - uses **NVMe** across your existing SAN Infrastructure **FC, Ethernet or InfiniBand**.
 - It has direct-attach performance with all the flexibility of being on a SAN
 - It has the potential for improved latency compared to SCSI right out of the box
 - It has even **better Latency & IOPS** when used with optimized parallel apps
- **!! Reduce your Application License costs**
 - Less server CPU core time spent on IO means **more CPU core time for your applications**
 - More CPU time for applications means **fewer 'per core' licenses to buy !!!**

Evolution des protocoles réseau: iSER sur Ethernet

SRP	iSER	iSER	iSCSI	SCSI FICON
		RoCE, iWARP		
RDMA		RDMA		FCP
Infiniband 40/56 Gbps (évolutions 100Gbps)		Ethernet 10Gbe (évolutions futures vers 25, 50, 100Gbe)		FC 16Gbps (évolutions 32Gbps)

- **iSER (iSCSI extension for RDMA) sur Ethernet** repose sur l'un des protocoles suivants:
 - **iWARP** (Internet Wide Area RDMA Protocol)
 - **RoCE** (RDMA over Converged Ethernet)



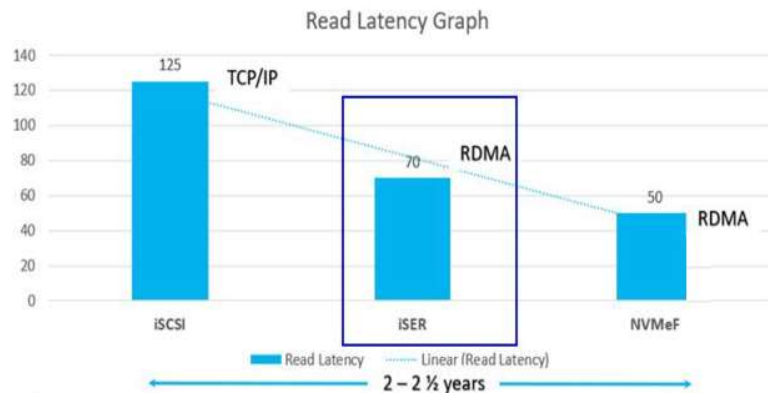
iSER utilise les protocoles **RDMA** mais pas ceux de NVMe

Il n'utilise pas la couche TCP/IP tout en assurant la compatibilité avec le protocole **iSCSI**

Il permet ainsi d'avoir sur **Ethernet** une très faible latence, un gros débit, et une diminution de l'utilisation des CPUs des contrôleurs (du fait du RDMA).

Evolution des protocoles réseau: NVMe-oF sur Ethernet

NVMe-oF	SRP	iSER	iSER	NVMe-oF		NVMe-oF	SCSI FICON
			RoCE, iWARP	NVMe-TCP	iSCSI	FC-NVMe	
RDMA			RDMA			FCP	
Infiniband 40/56 Gbps (évolutions 100Gbps)			Ethernet 10Gbe (évolutions futures vers 25, 50, 100Gbe)			FC 16Gbps (évolutions 32Gbps)	



NVMe-oF sur Ethernet n'utilise pas les commandes SCSI et permet d'avoir une **diminution supplémentaire de la latence** par rapport à iSER.

Evolution des protocoles réseau: NVMe-oF sur FC (FC-NVMe)

NVMe-oF	SRP	iSER	iSER	NVMe-oF		NVMe-oF	SCSI FICON
			RoCE, iWARP	NVMe-TCP	iSCSI	FC-NVMe	
RDMA			RDMA			FCP	
Infiniband 40/56 Gbps (évolutions 100Gbps)			Ethernet 10Gbe (évolutions futures vers 25, 50, 100Gbe)			FC 16Gbps (évolutions 32Gbps)	

FC-NVMe permet l'utilisation des reseaux **SAN Fibre Channel** et le **protocole FCP** mais avec une latence et une consommation CPU plus faibles.

The Evolution of NVMe?

- NVM Express Management Interface (NVMe-MI™) Specification officially released. Provides out-of-band management for NVMe™ components and systems and a common baseline mgmt. feature set across all NVMe™ devices and systems.

- NVM Express over Fabrics (NVMe-oF™) Specification published; extending NVMe™ onto fabrics such as Ethernet, Fibre Channel and InfiniBand®, providing access to individual NVMe™ devices & storage systems.

IBM Delivers NVMe
& NVMe-oF
Solutions

2018

1st Partial NVMe
solutions from
Competitors

- NVMe Spec 1.3 published. Addresses the needs of mobile devices, with their need for low power and other technical features, making it the only storage interface available for all platforms from mobile devices through data center storage systems.

- NVM Express Spec 1.0 published by industry leaders on March 1

2011

- NVMe Specification 1.1 released on Oct 11

2012

- Nothing Happened

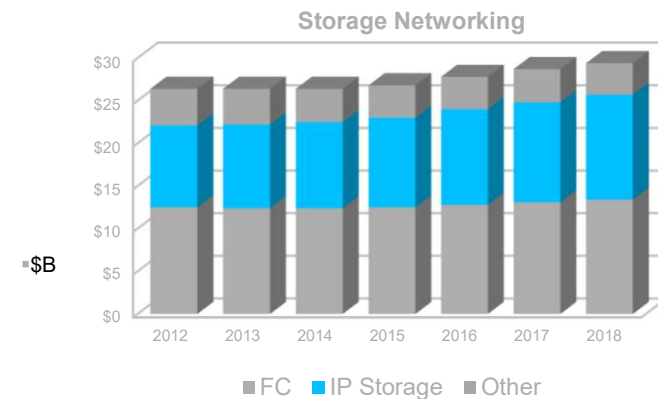
2013

- NVM Express Spec 1.2 released on November 3
- NVM Express Work Group was incorporated at NVMe Inc., the consortium responsible for the development of the NVM Express specification
- Work on the NVM Express over Fabrics (NVMe-oF™) Specification kicked-off

nvm
EXPRESS

Directions in Storage Networking

- **10GE to 100GE dominates the Cloud infrastructure**
 - CSPs adopt new Ethernet technology faster than Enterprise
 - Less constrained by legacy install base.
- **FC continues link speed generations - now on Gen 6 at 32Gbps**
 - Expect gradual decline in FC SAN share of storage attachment
 - Storage fabrics for new workloads, CSPs, Cold storage all favor IP storage attach – iSCSI, NAS, and REST Object Storage APIs.



IBM Storage Systems portfolio with focus on Flash & Virtualization



4 axes fondamentaux dans la stratégie d'IBM

Performances
maximales avec
la technologie
Flash



Optimisations
avec la
virtualisation du
stockage

IBM LEADER – > 100 000
Appliances SVC – Plusieurs
milliers de PB de stockage
virtualisé

Optimisations
grâce au Cloud



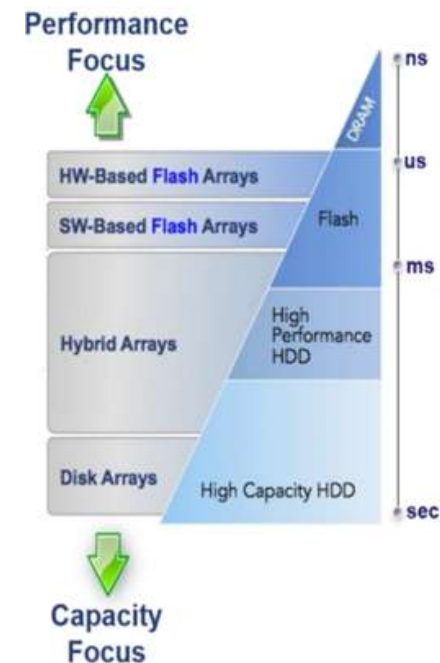
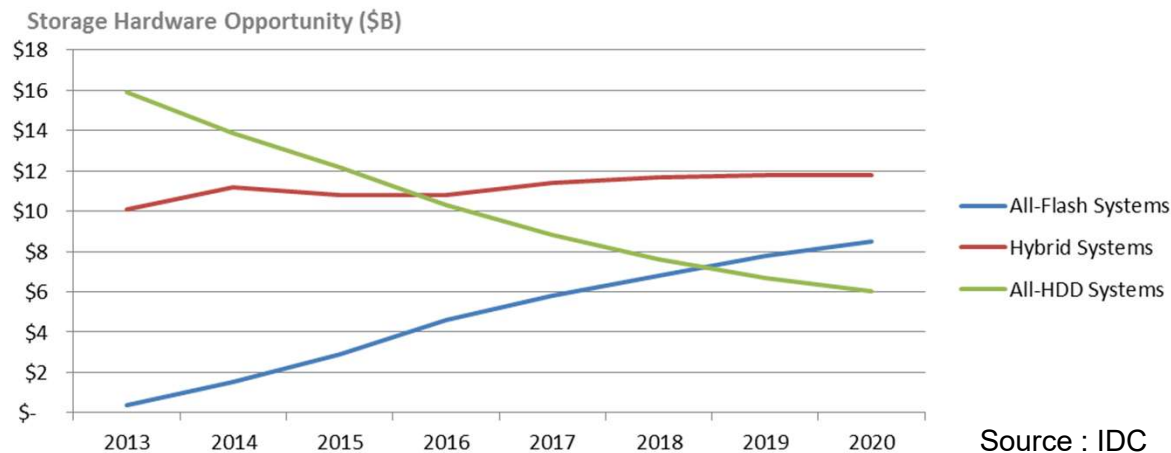
Software Defined
Storage

IBM LEADER
40% PDM



Le marché de la technologie Flash

Prévision des achats de technologies Flash (\$M)



Terminologie:

- **All Flash Array (AFA):** baie de stockage sur une base Flash ou SSD, **sans possibilité d'ajout de HDD** (SAS, NL-SAS) dans la baie.
 - **HW-Based All Flash Array (AFA):** Conception axée sur la performance, sans fonctions logicielles (snapshots/clones, compression, réplication, déduplication, etc.) « sur » le chemin des données.
 - **SW-Based All Flash Array (AFA):** Idem précédent, mais avec des fonctions logicielles.
- **Hybrid Arrays:** baie de stockage classique **intégrant des HDDs et des SSDs**.

Comprehensive, award-winning portfolio

Backup & Archive

Backup



IBM
Spectrum
Protect

VM Data
Availability



IBM
Spectrum
Protect Plus

Archive



IBM
Spectrum
Archive

Management

Monitoring
& Control



IBM
Spectrum
Control

Container
& VM APIs



IBM
Spectrum
Insights



IBM
Spectrum
Connect

Copy Data
Management



IBM
Spectrum
Copy Data
Management

Private
Cloud



IBM
Spectrum
Access
Blueprint

Hybrid Cloud
Disaster
Recovery



IBM
Spectrum
Virtualize for
Public Cloud

Cluster Virtualization for Servers

High-
Performance
Computing



IBM
Spectrum
LSF

High-
Performance
Analytics



IBM
Spectrum
Symphony

New-Gen
Workloads



IBM
Spectrum
Conductor

Infrastructure



IBM Cloud
Object Storage

Scale-Out
Object



Cloud Object
Storage System



IBM
Spectrum
Virtualize

Virtualized Multi-Vendor
Block

Storwize V5030F



Storwize V5000

Storwize V7000F



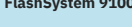
Storwize V7000

Available as

VersaStack™
Solution
by Cisco and IBM



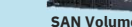
FlashSystem 9100



FlashSystem V9000



SAN Volume
Controller



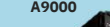
IBM
Spectrum
Accelerate

Scale-Out
Block

FlashSystem A9000



FlashSystem A9000



FlashSystem A9000R



FlashSystem A9000R



FlashSystem A9000R



FlashSystem A9000R



FlashSystem A9000R



FlashSystem A9000R



FlashSystem A9000R



FlashSystem A9000R



FlashSystem A9000R



FlashSystem A9000R



FlashSystem A9000R



IBM
Spectrum
Scale

Scale-Out
File



Elastic Storage
Server



IBM
Spectrum
NAS

Traditional
NAS/File

All-Flash
Acceleration



FlashSystem 900

High-End
Storage

DS8888F



DS8886F



DS8884F



DS8886



DS8884



Tape &
Virtual Tape

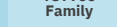
LT08 &
Tape Drives



TS7700
Family



TS2900
Autoloader



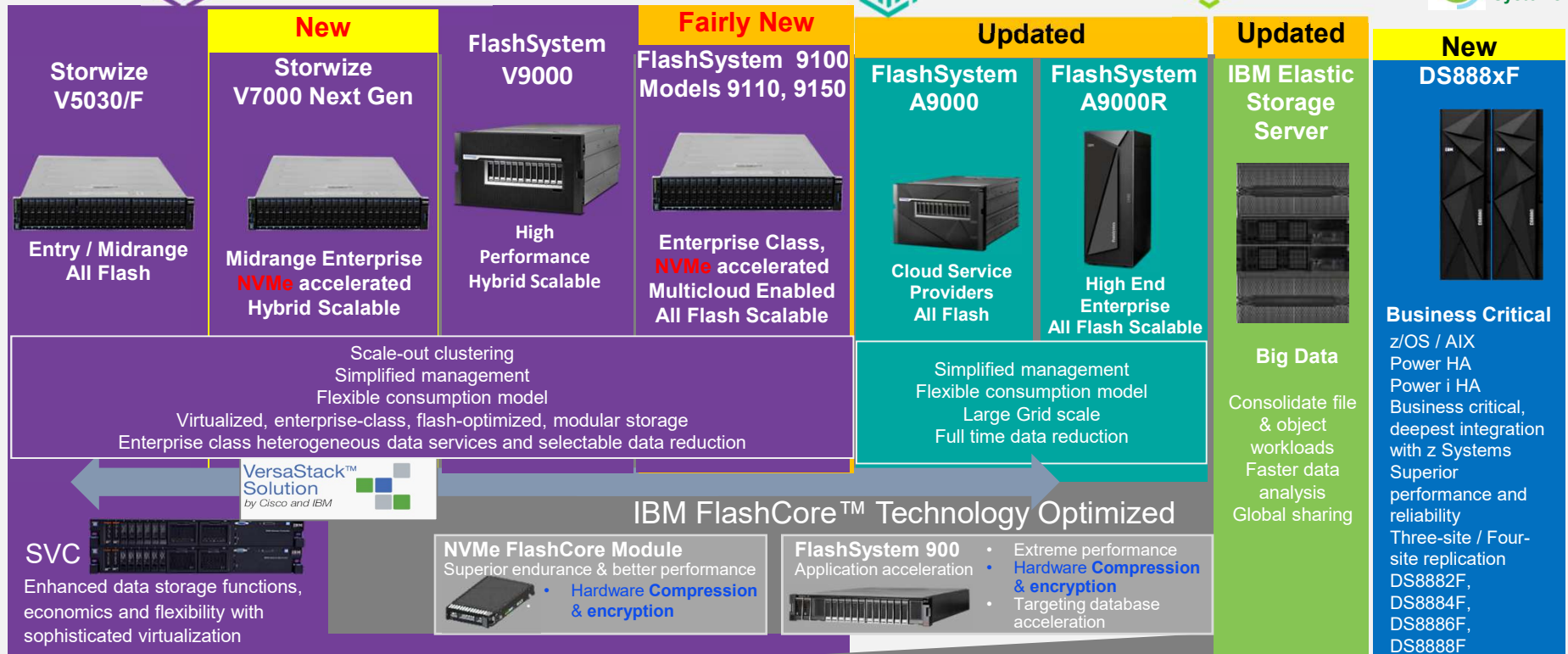
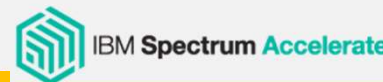
Tape
Libraries



Networking

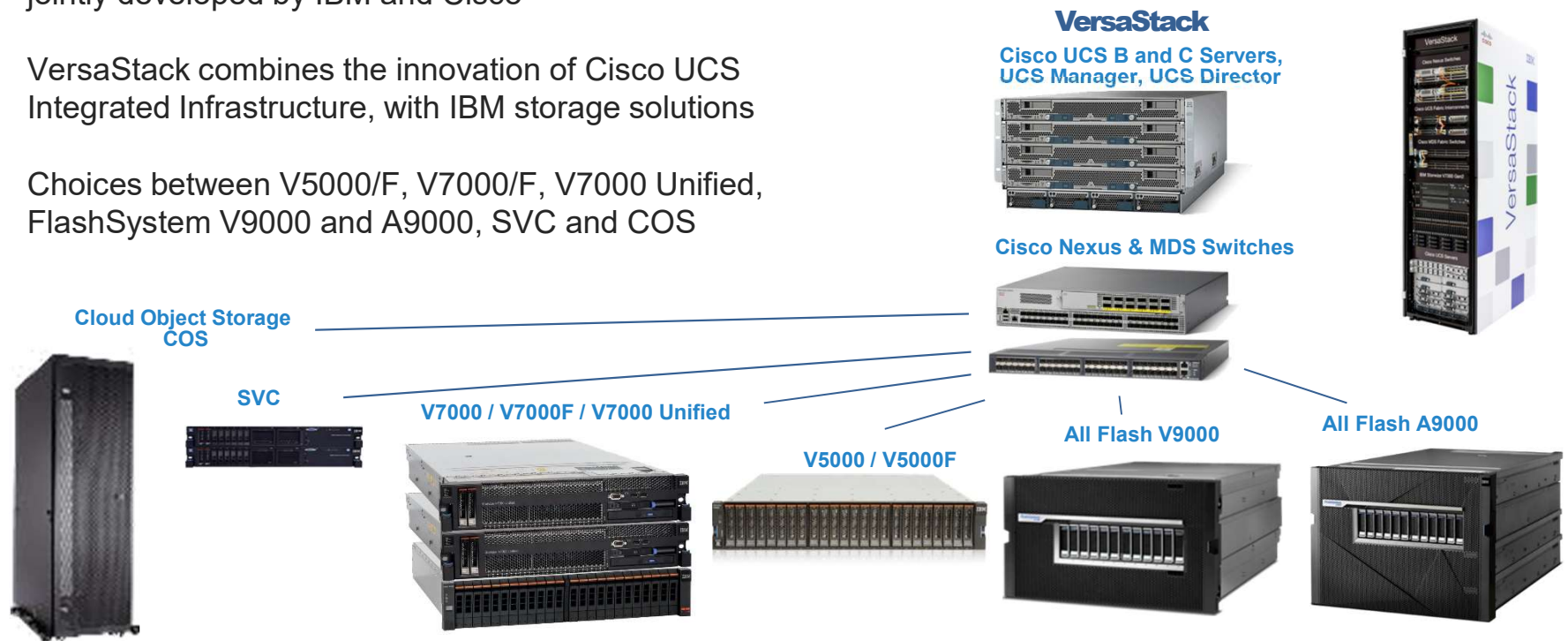


IBM Systems Flash/Hybrid Storage Offerings Portfolio



VersaStack™ Solutions from IBM and Cisco

- VersaStack, an integrated infrastructure solution jointly developed by IBM and Cisco™
- VersaStack combines the innovation of Cisco UCS Integrated Infrastructure, with IBM storage solutions
- Choices between V5000/F, V7000/F, V7000 Unified, FlashSystem V9000 and A9000, SVC and COS



The IBM storage systems built on Spectrum Virtualize

Comprehensive range of virtualized software defined storage systems

- One code base on all platforms
- One set of functions (selectively licensed)
- One client experience



Spectrum Virtualize

Storwize V5000
Storwize V5030F

Entry-Midrange Hybrid or All-Flash Block Storage

ECONOMY

More performance & scalability for traditional applications & analytics in mid-size businesses

Storwize V7000 Gen2+ & V7000F

Midrange Hybrid or All-Flash Block Storage

SCALABILITY

Highly scalable, high performance virtualization for analytics & mixed workloads in mid-size businesses

NEW (14/12/2018) Storwize V7000 Gen3

Midrange - Enterprise Hybrid

IBM FlashCore™ Technology Optimized

CONSOLIDATION

PERFORMANCE

NVMe-oF Ready
Lowest **end-to-end NVMe** starting price
Hybrid configurations
Highest performance fully-integrated storage virtualization solution

FlashSystem V9000

Enterprise Hybrid or All-Flash Block Storage

IBM FlashCore™ Technology Optimized

PERFORMANCE

Highest performance fully-integrated storage virtualization solution

FlashSystem 9100

FlashSystem 9110 & 9150 All Flash

Enterprise Class, **NVMe** accelerated Multi-Cloud Enabled

End-to-end NVMe
IBM FlashCore™ Highest performance

SAN Volume Controller

Enterprise Storage Virtualization of Flash & Hybrid

Enterprise Software-Only Block Storage

VERSATILITY

IBM Storwize Family

IBM FlashSystem

SDS

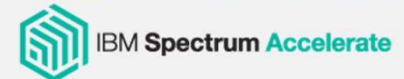
ENTRY

MIDRANGE

ENTERPRISE

IBM Systems Storage Flash/Hybrid & Virtualize positioning

IBM Spectrum Virtualize



SVC

Enterprise Class
Virtualization Appliance

Storwize V5000

Entry
Hybrid Solution

Storwize V5030/F

Entry / Midrange
Soft compression
All Flash

Storwize V7000/F

Midrange
Hard compression
All Flash

Storwize V7000 Next Gen

Midrange Enterprise
NVMe FlashCore & SSDs
Hard compression
Hybrid Solution

FlashSystem 9100

Enterprise Class,
NVMe FlashCore & SSDs
Hard compression
All Flash

FlashSystem A9000

Cloud Service Providers
All Flash

FlashSystem A9000R

High End Enterprise
All Flash

Simplified management
Flexible consumption model
Large Grid scale
Full time data reduction



Data Reduction :
Compression & Deduplication

NVMe FlashCore Module
Superior endurance & better performance
Hardware Compression & encryption

FlashSystem 900
Application acceleration
FlashCore technology
Extreme performance
Hardware Compression & encryption
All-Flash

IBM Storwize Family


IBM FlashSystem Family

IBM Systems Storage Flash/Hybrid & Virtualize positioning

IBM Spectrum Virtualize



SVC




Enterprise Class
Virtualization Appliance

Storwize V5000



Entry
Hybrid Solution

Storwize V5030/F



Entry / Midrange
Soft compression
All Flash

Storwize V7000/F



Midrange
Hard compression
All Flash

Storwize V7000 Next Gen




Midrange Enterprise
NVMe FlashCore & SSDs
Hard compression
Hybrid Solution

FlashSystem 9100




Enterprise Class,
NVMe FlashCore & SSDs
Hard compression
All Flash

FlashSystem A9000



Cloud Service Providers
All Flash

FlashSystem A9000R



High End Enterprise
All Flash

Simplified management
Flexible consumption model
Large Grid scale
Full time data reduction



IBM Spectrum Virtualize

Data Reduction :
Compression & Deduplication

NVMe FlashCore Module

Superior endurance
& better performance
Hardware
Compression
& encryption



FlashSystem 900
Application acceleration



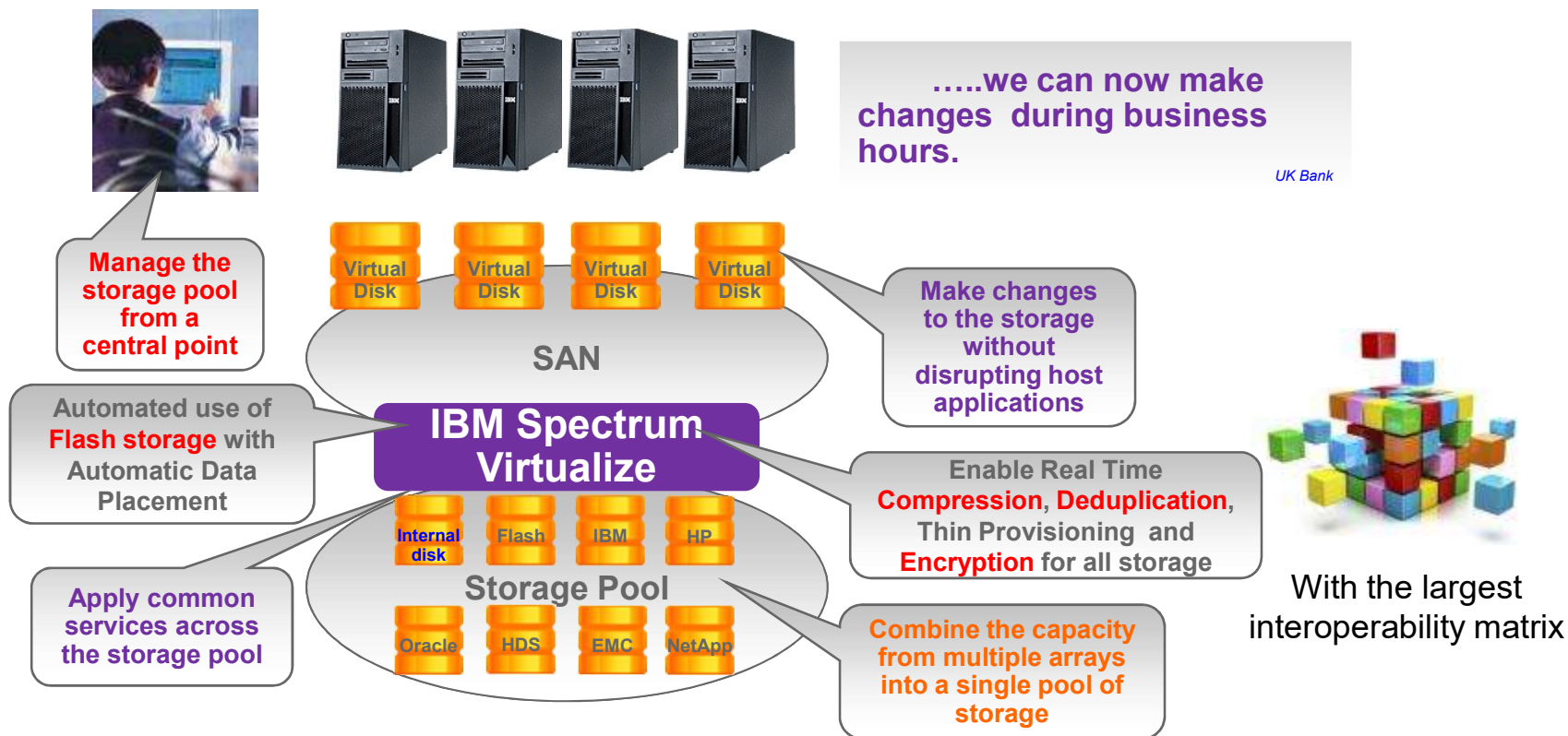
FlashCore technology
Extreme performance
Hardware Compression
& encryption
All-Flash

IBM Storwize Family

IBM FlashSystem Family

IBM Spectrum Virtualize

is the foundation of an **Efficient SAN Storage Infrastructure**



Data Reduction

Thin Provisioning

Capacity is allocated on demand as storage is first written to

Compression

Data is compressed before being written to storage

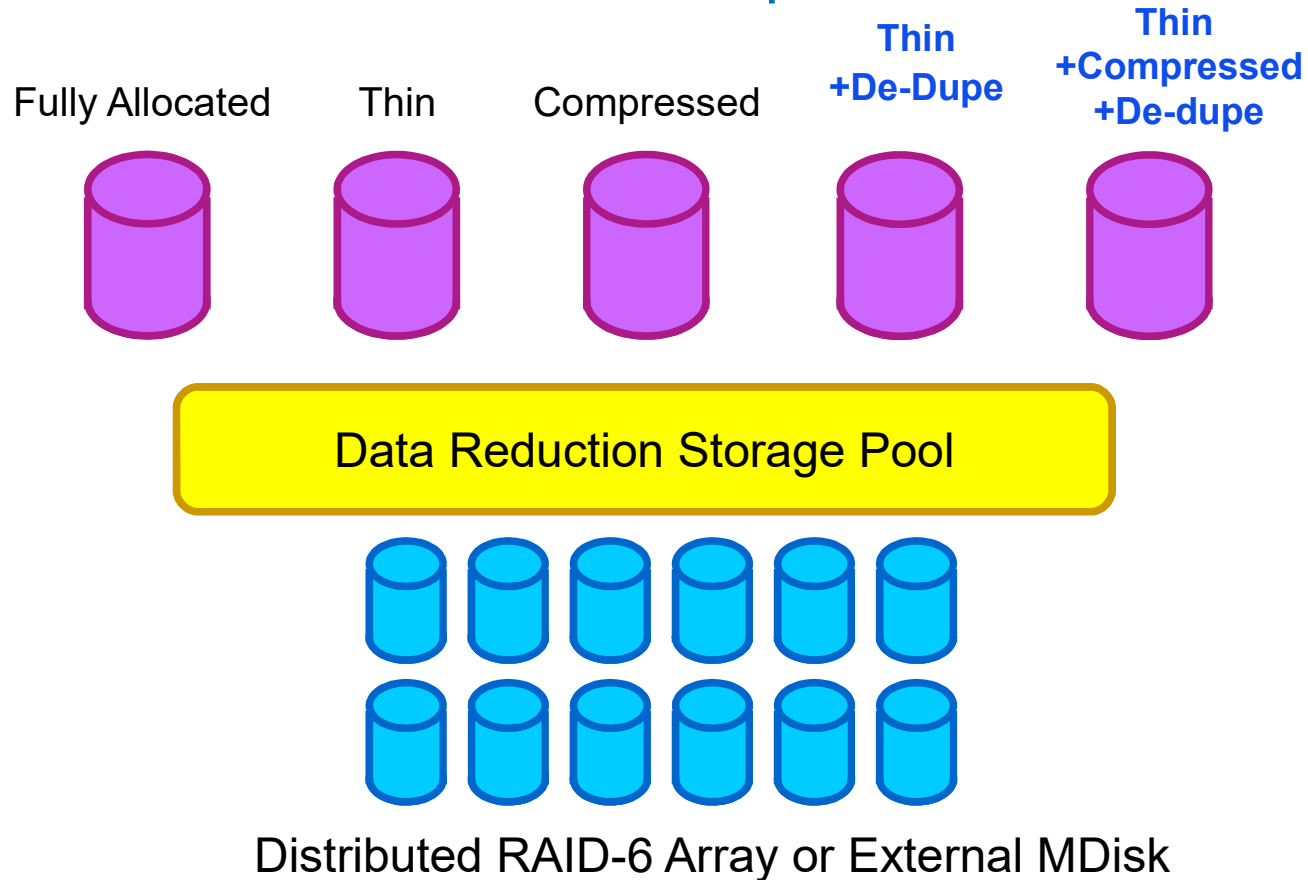
De-Duplication

Duplicates of data are detected and are replaced with references to first copy



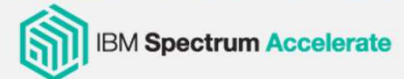
IBM Spectrum Virtualize Code 8.1

New Data Reduction Pools & Deduplication




IBM Systems Storage Flash/Hybrid & Virtualize positioning

IBM Spectrum Virtualize




SVC




Enterprise Class
Virtualization
Appliance

Storwize V5000




Entry
Hybrid Solution

Storwize V5030/F



Entry / Midrange
Soft compression
All Flash

Storwize V7000/F



Midrange
Hard compression
All Flash

Storwize V7000 Next Gen




Midrange Enterprise
NVMe FlashCore & SSDs
Hard compression
Hybrid Solution

FlashSystem 9100




Enterprise Class,
NVMe FlashCore & SSDs
Hard compression
All Flash

FlashSystem A9000



Cloud Service Providers
All Flash

FlashSystem A9000R



High End Enterprise
All Flash

Simplified management
Flexible consumption model
Large Grid scale
Full time data reduction



Data Reduction :
Compression & Deduplication

NVMe FlashCore Module
Superior endurance
& better performance
Hardware
Compression
& encryption



FlashSystem 900
Application acceleration



FlashCore technology
Extreme performance
Hardware Compression
& encryption
All-Flash

IBM Storwize Family

IBM FlashSystem Family

IBM FlashSystem 900

Premier all flash array with hardware compression/encryption and 3D TLC

IBM Storage and SDI

Capacity and Density

- Up to 180TB usable capacity with **IBM enhanced 3D TLC** in 2U
- IBM-enhanced 3D TLC IBM FlashCore®
- **440TB maximum effective capacity** for compressible workloads in **2U**

Inline Hardware Technology

- **Inline Hardware Compression/Encryption with no performance penalty**
- Supports average 2:1 compressibility on all usable capacity Modules

World Class Performance even with Compression

- **Consistent response time; latency as low as 85 microseconds**
- Response time for real workloads **improves** with compressible data
- High-speed **16Gb FC-NVMe** interface



Unique & Valuable

- [illegible]



No one else.

IBM is the only vendor to deliver Tier 0 ultra-dense purpose-built TLC flash modules with inline high-performance compression!

FlashCore: Hardware Accelerated I/O

NEW Inline Hardware Compression

Unique & Valuable

FlashSystem data **compression/decompression** algorithm is a Modified Dynamic GZIP algorithm



- **Implemented completely in hardware**, no processor intervention
- This technology **originated with Mainframe - Z** and has been adapted to work in an IBM FlashCore™ flash controller
- Improved economics with no performance impact
 - Compression is performed as the first step in the inbound data path, and decompression is the last step in the outbound data path. This minimizes the amount of data written to flash
 - Data protection (ECC) is implemented on top of compressed data
 - Compression and decompression are completely transparent above the FlashCore module except for management of space
- FlashCore modules are also **self-encrypted** (hardware)

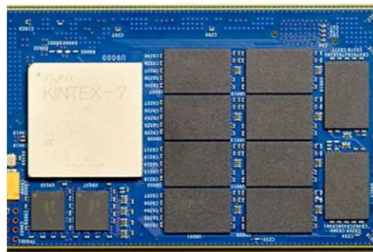
Who has it?

IBM is the **only** vendor to deliver these purpose-built TLC flash modules with inline high-performance compression!

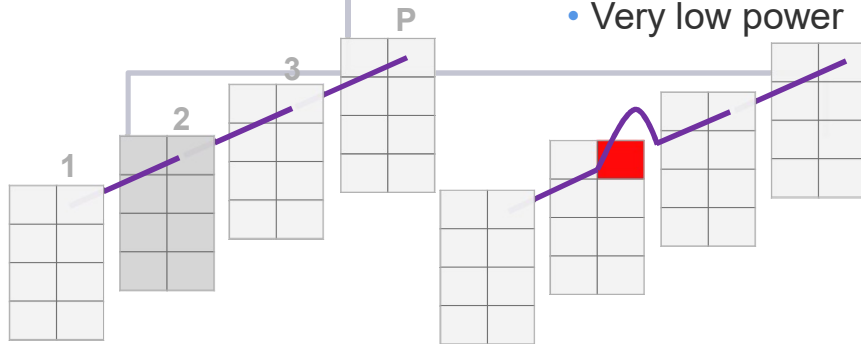
FlashCore Uses IBM Patented Variable Stripe RAID™

Unique & Valuable

Add Variable Stripe RAID™



- Maximum level of flash module data protection
 - Maximum wear life
 - Fast writes
 - Scalable
 - Fast at reads
 - Non-volatile
 - Very low power



Advantages:

- Protects data from a chip failure
- Dynamically re-stripes data at a sub-chip level
- Preserves life, protection and performance

IBM VSR™ protected flash modules:

Now we can build our system using far more advanced technology than SSDs



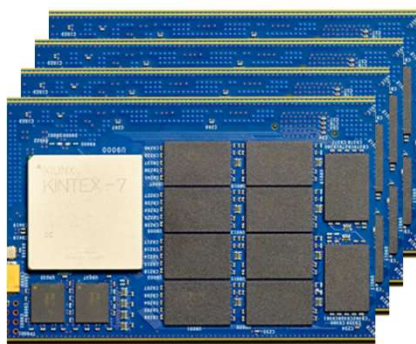
Who has it?

Only the IBM FlashSystem



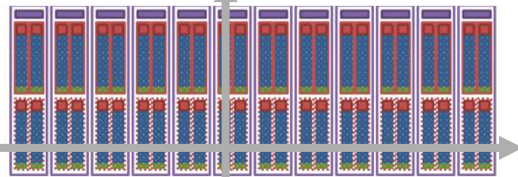
FlashCore Two-Dimensional (2D) Flash RAID

Employs a 2nd Dimension RAID of Protection



- Maximum level of system protection
- Maximum level of flash module data protection
 - Maximum wear life
 - Fast writes
 - Scalable
 - Fast at reads
 - Non-volatile
- Very low power

2D RAID
across
modules



VSR within modules

Advantages:

- VSR protects from flash chip or sub-chip issues
- System-level RAID protects against abrupt module failure and controller failure



Who has it?

Only the IBM FlashSystem



Custom Designed Flash Media With Field Proven Architecture

The Right Systems For The Right Environments

All IBM FlashSystem Products Share A Consistent Core Architecture

- **IBM MicroLatency™ FlashCore Modules (FCM)** delivering industry leading performance
- IBM FlashCore™ Technology with IBM Enhanced 3DTLC Flash
 - Non-blocking Crossbar Switch
 - Hardware Only Data Path
 - Concurrent Code Load and Concurrent Maintenance
- High Availability Flash Enclosure with dual canister design
- **Inline, high-speed compression/decompression**
- Integrated Data Integrity Checking
- Two-Dimensional RAID



IBM FlashSystem 900

Speed and compression to optimize business—critical application performance and data center efficiency

IBM Storage and SDI

- **New High-speed 16Gb Fibre Channel NVMe over Fabrics (FC-NVMe)** interface
 - Read latency reduced by 15% to **85 microseconds**
 - Supports both **Fibre Channel** and **FC-NVMe**
 - 4 x 16 Gb FC Adapter
- New “extra large” **18TB module** offers **44TB effective capacity** with inline data compression
 - **2X** module capacity for only **15%** higher price (**40% better \$/TB**)
 - **Store up to 440TB** with inline data compression to achieve benefits like smaller data center footprint, lower power costs, and lower cooling costs
- **Annoucement date: 10/23/2018**
- **GA date: 12/7/2018**



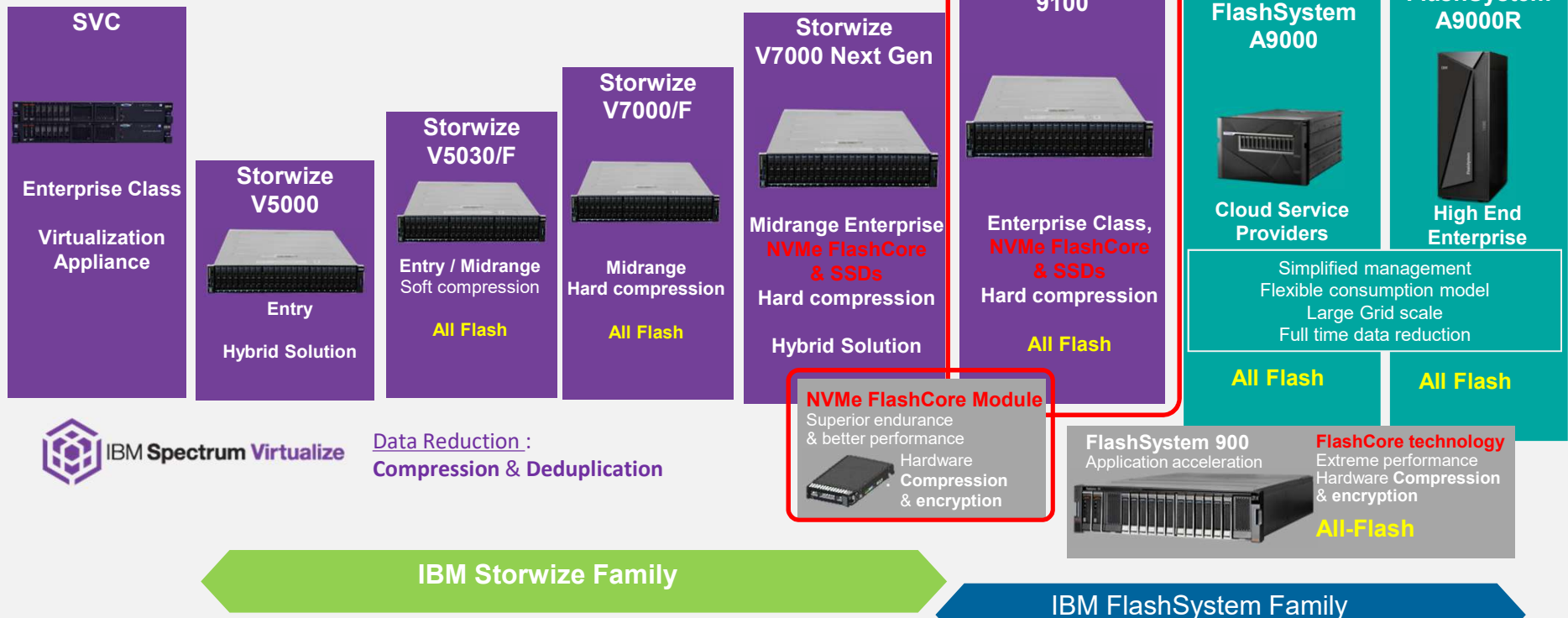
FlashSystem 900 AE3 Configuration Matrix

Module	Small (3.6TBu)				Medium (8.5TBu)			Large (18TBu)			XLarge (18TBu)		
	Min			Max	Min		Max	Min		Max	Min		Max
Qty	6	8	10	12	8	10	12	8	10	12	8	10	12
Usable	14.4	21.6	28.8	36.1	51.3	68.4	85.5	108	144	180	108	144	180
Expected Effective Cap. (2:1 typical)	28.8	43.3	57.7	72.2	102.6	136.8	171	NA	NA	NA	216	288	360
Max Effective	43.9	65.9	87.9	109.9	131.9	175.9	219.9	131.9	175.9	219.9	263.8	351.8	439.8

- Additionally:
 - No mixing of MicroLatency modules of varying capacity
 - FlashSystem 900 chassis does not support mixing MLC & TLC flash media
 - Systems cannot be upgraded in the field to new flash media (MLC → TLC)
 - Large capacity modules have the same effective capacity as the Medium modules. These are designed for system that need additional capacity but have mixed workloads that may not be able to take full advantage of compression.

IBM Systems Storage Flash/Hybrid & Virtualize positioning

IBM Spectrum Virtualize



New IBM FlashSystem 9100

Simple, Agile, Ready for the Future

Powerful NVMe End to end storage system for enterprise applications

Predictive analytics with **AI-cloud based** support

6X better density with **460TB** raw capacity in a **2U** enclosure

5X data reduction with data efficiency capabilities **compression** and **deduplication** for up to **2PB in 2u**

Consistent response time : latency as low as **85 microseconds**

100% data availability for systems using **HyperSwap**

3X faster access to valuable analytics driving better business results with **NVMe** technology – **up to 2.5M IOPS in 2U**



© Copyright IBM Corporation 2018



IBM FlashCore™ technologies

+ IBM Spectrum Virtualize

FlashSystem 9100 : A **end-to-end NVMe** Solution

A modern storage system designed to support today's performance and growth requirements in order to easily address the needs of multi-cloud environments and expanding data centers.

Up to 24 NVMe Drives in Control enclosure :

- **4.8TB, 9.6TB FCM** and **19.2TB FCMs** with hardware compression
- 1.92TB, 3.84TB, 7.68TB & 15.26TB **NVMe SSDs**

Up to 20 SAS 12Gb/s Expansion enclosure :

- Only for 2,5" SAS Flash disks (SSDs)
- 1,92TB, 3,84 TB, 7,68 TB & 15,36 TB SAS SSDs

EasyTier:

- Use **FCMs** or **NVMe SSDs** as a high-speed tier for **SAS SSDs** or for other **storage that is being virtualized**

NVMe-oF:

- **NVMe-oF** “ready” host interfaces

FlashCore Modules - FCM

- **FlashCore** technology for improved endurance
- High performance
- **with hardware compression**
- 4.8TB – 19.2TB



NVMe SSDs

- Dual Ported for performance
- Industry standard
- 1.92 – 15.36TB



FlashSystem 9100



NVMe-oF ready Host Interfaces



- 16Gb Fibre Channel
- 25Gb Ethernet (iWarp)
- 25Gb Ethernet (RoCE)

FlashSystem 9100 Features/Capabilities

- **Storage Virtualization (IBM Spectrum Virtualize embedded)**
- Virtual Disk Mirroring
- HyperSwap
- Metro Mirror - synchronous remote replication
- Global Mirror - asynchronous remote replication
- Global Mirror over IP - remote replication over the Internet
- Importing and exporting existing LUNs via Image Mode
- Unmap
- Vvols
- Public Cloud offerings
- IBM Cloud Private
- Container support
- Thin Provisioning
- Fast-write cache
- Auto tiering (Easy Tier)
- Distributed RAID 5 & 6
- Transparent Cloud Tiering
- Cloud Snapshot
- **Compression**
- **Deduplication (New : code 8.2)**
- FlashCopy (Snapshot)
- FlashCopy rollback
- Encryption of data at rest
- Data migration and pooling

IBM FlashSystem 9100

FlashSystem 9110 and FlashSystem 9150

IBM Storage and SDI

Dual Active-Active Array Controllers w/**NVMe** to Flash Media
Dual-ported 2.5" NVMe Flash bays (24)

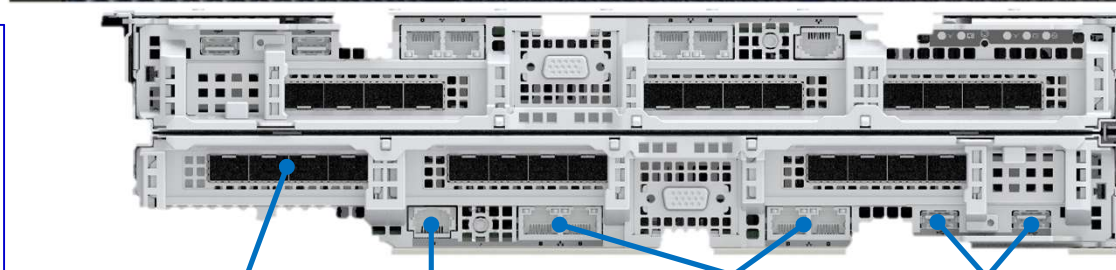
Flash Core Modules FCM redesigned to 2.5" industry
standard form factor
(Includes: HW Compression, HW Encryption and 64 Layer 3DTLC)

FS9110
Dual 8-core
processors per
controller

Up to 1.5TB
Cache per
system

FS9150
Dual 14-core
processors
per controller

Up to 3 host
adapters per
controller



Interface Card Slot (3)

Tech Port

10GbE Ports (4)

USB Ports (2)

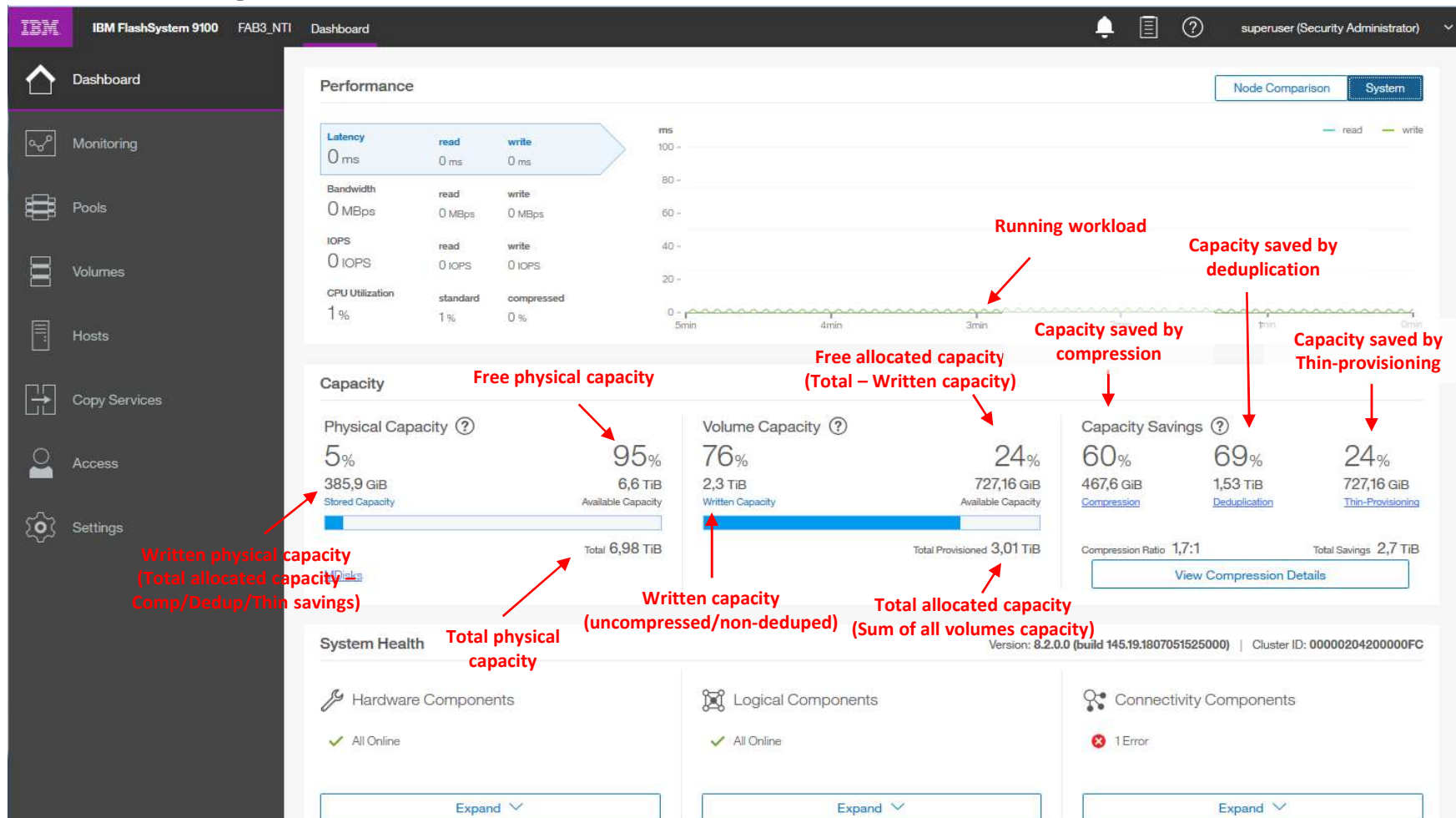
2KW PSU (2)

Unprecedented Storage Capacity In 2U Of Rack Space



Flash Media	Capacity per Drive with Inline Compression (max ratio varies 2:1 – max)	Capacity per Drive with Data Reduction Pools (2:1 – 5:1)	Max System Capacity in 2U with Inline Compression (max ratio varies 2:1 – max)	Max System Capacity in 2U with Data Reduction Pools (2:1 – 5:1)
FCM 4.8TB	9.6TB – 22TB	9.6TB – 24TB	230.4TB – 552TB	230.4TB – 576TB
FCM 9.6TB	19.2TB – 22TB	19.2TB – 48TB	460.8TB – 552TB	460.8TB – 1.1PB
FCM 19.2TB	38.4TB – 44TB	38.4TB – 96TB	921.6TB – 1PB	921.6TB – 2.3PB
Flash Media	Capacity per Drive with Data Reduction Pools (2:1 – 5:1)		Max Capacity in just 2U with Data Reduction Pools (2:1 – 5:1)	
NVMe 1.92TB	3.84TB – 9.6TB		92.6TB – 230.4TB	
NVMe 3.84TB	7.68TB – 19.2TB		184.32TB – 460TB	
NVMe 7.68TB	15.36TB – 38.4TB		368.64TB – 921.6TB	
NVMe 15.36TB	30.72TB – 76.8TB		737.28TB – 1.8PB	

FlashSystem 9100 : GUI

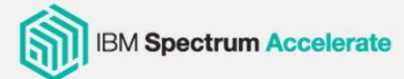


FlashSystem 9100 – Future Proofed, Flexible & Fast

- **Front to back NVMe technologies**
 - Supports unique world class IBM FCM drives with inline compression
 - Supports industry standard **NVMe SSD drives**
 - Storage Class Memory capable (SCM-capable)
 - NVMe-oF ready – simply requires a software update later in the year
- 24 drives in a 2U enclosure (***six times the density of FlashSystem V9000***)
- FlashSystem 9150 delivers **2.5M Cache Hit IOPS / 2U** and scales out for **up to 10 Million IOPS in 8U**
- FlashSystem 9100 delivers up to **34GB/s** in just 2U
- **Latency as low as 100µs**
- Internal bandwidth (CPU to drives) is significantly increased
- Host bandwidth also increased through more available ports
- Move from traditional RAID5 to DRAID6 for improved resiliency
- **Has IBM Spectrum Virtualize embedded**
- Can cluster up to four systems for tremendous scaling and performance

IBM Systems Storage Flash/Hybrid & Virtualize positioning

IBM Spectrum Virtualize



SVC

Enterprise Class
Virtualization Appliance

Storwize V5000

Entry
Hybrid Solution

Storwize V5030/F

Entry / Midrange
Soft compression
All Flash

Storwize V7000/F

Midrange
Hard compression
All Flash

Storwize V7000 Next Gen

Midrange Enterprise
NVMe FlashCore & SSDs
Hard compression
Hybrid Solution

FlashSystem 9100

Enterprise Class,
NVMe FlashCore & SSDs
Hard compression
All Flash

FlashSystem A9000

Cloud Service Providers
All Flash

FlashSystem A9000R

High End Enterprise
All Flash

Simplified management
Flexible consumption model
Large Grid scale
Full time data reduction



IBM Spectrum Virtualize

Data Reduction :
Compression & Deduplication

NVMe FlashCore Module
Superior endurance & better performance
Hardware Compression & encryption

FlashSystem 900
Application acceleration

FlashCore technology
Extreme performance
Hardware Compression & encryption
All-Flash

IBM Storwize Family

IBM FlashSystem Family

IBM Storage and SDI

Storwize V7000 Next Gen *(V7000 Gen 3)*



IBM Storwize V7000 Gen3 (GA : 14/12/2018)

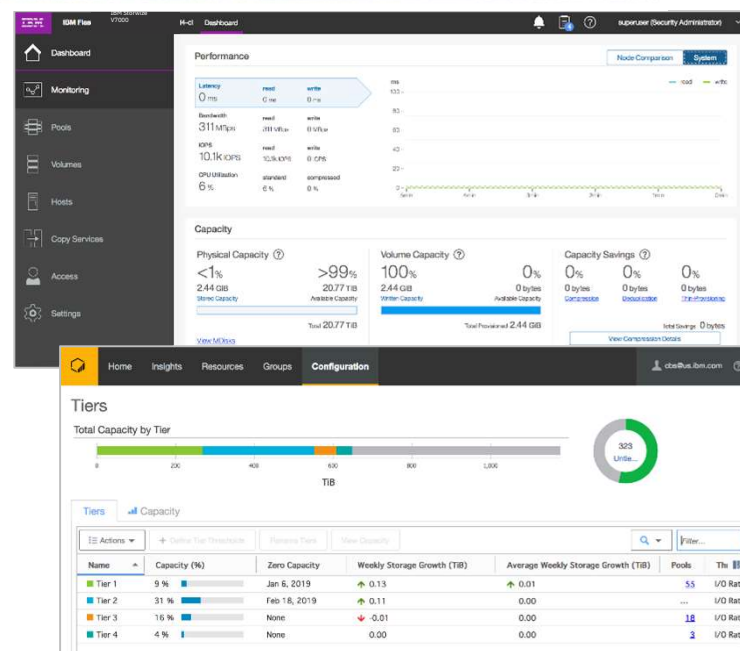
IBM Storage & SDI

Accelerate business execution

- **NVMe**-optimized IBM **FlashCore technology**
- Extensive AI-based storage resource management
- Optimize storage using AI-based data placement with Easy Tier

Modernize your IT infrastructure

- Leverage your existing SSD and HDD infrastructure
- Increase ROI with **IBM Spectrum Virtualize** to extend a rich set of data services across all your storage
- Predictive analytics and enhanced support with IBM Storage Insights



Storwize V7000 Gen3 – a new Control Enclosure

- **Two Models – V7000 Gen3, M/T: 2076**
 - **724** - NVMe Control enclosure for **hybrid** configurations
 - **U7B** - NVMe Control enclosure for **utility** configurations
- **Drives** – same set of NVMe and FCM drives as FS9100 – 800GB -> 19.2TB
- **Expansions** – **existing V7000 (2076) Gen2 expansions and drives, 12GbSAS, same drive VPD as V7000**
 - **Allow re-use but no data in-place migration from Gen2/2+ to Gen3**
 - Same attachment limits as current V7000 – **up to 20 2U Expansions, up to 8 5U Expansions**, combos up to 760 drives max
 - Hybrid: existing models **V7000 2076-12F/24F/92F**
 - Utility: existing models **V7000 2076-12F/24F/92F**
- **Software**
 - Existing software Control enclosure PID (5639-CB8) with optional advanced functions and full bundle
 - Existing software Expansion enclosure PID (5639-XB8) with optional advanced functions and full bundle
 - Existing software PID (5639-EB8) for External Virtualization
- **Clustering** – **up to 4way (8 node) clusters**
 - with **V7000 724/624/AF6/U7A** and 524

Storwize V7000 Gen3 – a new Control Enclosure

- Adapters – [two slots for host adapters per canister](#).
 - On-board 4 ports 10GbE and 1GbE Tech port per canister
 - 4port 16GbFC **FC-NVMe**
 - 2port 25GbE **NVMeoF ready** for **iWARP** or **RoCE** with **iSER support**,
- CPUs – [dual Intel Skylake 8 core 1.7GHz](#) processors per canister
- Cache Memory (system) – max 48 DIMM slots (24 per canister)
 - [Base includes 128GB](#) (8x 16GB)
 - Options for:
 - 256GB – adds 8x 16GB DIMMs
 - 384GB – adds additional 8x 16GB DIMMs and
 - [1152GB](#) – adds additional 24x 32GB DIMMs per system
- Encryption – same process as used for model 624. Encryption feature code, DFSA activation code
- At planned Availability, will ship with [8.2.0.x software](#) code level installed. Upgrade to 8.2.1.0 for new capabilities.

IBM Storwize V7000 Next Gen

The first Storwize system with NVMe

IBM Storage and SDI

V7000 Next Gen

Dual 8-core
processors per
controller.

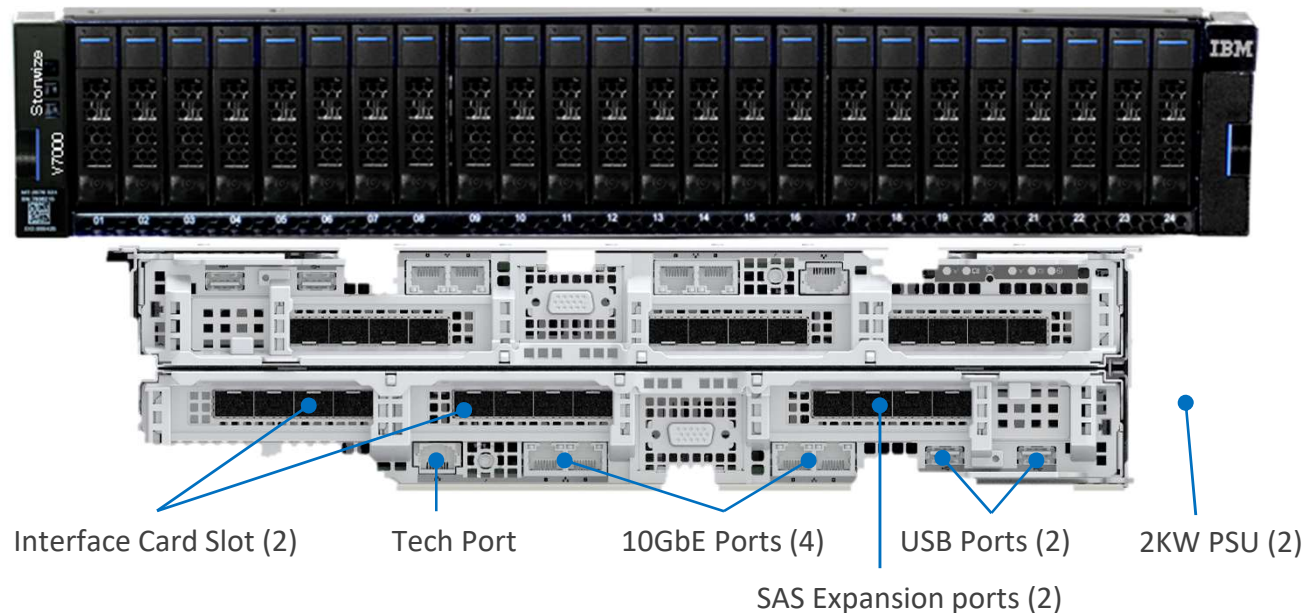
Four 8-core total

Up to 1.1TB
Cache per System

Up to 2 host
adapters per
controller

Dual Active-Active Array Controllers w/NVMe to Flash Media
Dual-ported 2.5" NVMe Flash bays (24)

Flash Core Modules - FCM redesigned to 2.5" industry standard form factor
(Includes: HW Compression, HW Encryption and 64 Layer 3DTLC)



New Storwize V7000 – Future Proofed, Flexible & Fast


- **End to end NVMe technologies**
 - Supports unique world class IBM Flash Core Modules - FCM drives with inline compression
 - Supports industry standard NVMe drives
 - Storage Class Memory capable (SCM-capable)
 - **Supports NVMe-oF**: **FC-NVMe** available now on **FC** and **Ethernet** implementation simply requires a software update in 2019.
- 24 drives in a 2U enclosure (*double the density of Storwize V7000 Gen2/2+ due to FCMs*)
- New Storwize V7000 provides up to **750,000 Cache Hit IOPS / 2U & scales out** for up to **3 Million IOPS** in 8u
- New Storwize V7000 provides up to **26 GB/s bandwidth in 2U & scales out** for up to **104 GB/s** in 8u
- Latency between **200µs to 250µs** with NVMe-based media
- Internal bandwidth (CPU to media devices) is significantly increased over V7000 G2+
- Host bandwidth also increased over V7000 G1 & G2
- Supports SAS SSD & HDD expansion, thus providing flexibility & performance via EasyTier

Figures Subject
To Change

IBM Systems Storage Flash/Hybrid & Virtualize positioning


IBM Spectrum Virtualize

SVC




Enterprise Class
Virtualization
Appliance

Storwize V5000




Entry
Hybrid Solution

Storwize V5030/F



Entry / Midrange
Soft compression
All Flash

Storwize V7000/F



Midrange
Hard compression
All Flash

Storwize V7000 Next Gen



Midrange Enterprise
NVMe FlashCore & SSDs
Hard compression
Hybrid Solution


FlashSystem 9100



Enterprise Class,
NVMe FlashCore & SSDs
Hard compression
All Flash


IBM Spectrum Accelerate

FlashSystem A9000



Cloud Service
Providers
All Flash

FlashSystem A9000R



High End
Enterprise
All Flash

Simplified management
Flexible consumption model
Large Grid scale
Full time data reduction



IBM Spectrum Virtualize

Data Reduction :
Compression & Deduplication

NVMe FlashCore Module

Superior endurance
& better performance
Hardware
Compression
& encryption



FlashSystem 900
Application acceleration



FlashCore technology
Extreme performance
Hardware Compression
& encryption
All-Flash

IBM Storwize Family

IBM FlashSystem Family

IBM FlashSystem A9000/A9000R

Enterprise class performance for critical applications

EFFICIENT

- With **IBM Spectrum Accelerate technology (XIV)**
- Lower TCO with Grid-based data reduction including **pattern removal, deduplication and compression**.
- **Cloud-ready** and tuning free for mixed workloads & **Multi-Tenancy features**
- Rich set of copy services including remote copy, mirroring, hyper-scale migration and RoW snapshots.

FLEXIBLE

- Easily manage more than 100 FlashSystem A9000s &/or FlashSystem A9000Rs from a single pane of glass with HyperScale Manager.
- Non-disruptive data movement between A9000 and A9000R with Hyper-Scale Migration
- **VMware, Hyper-V** and **VersaStack** integrated



ULTRA FAST

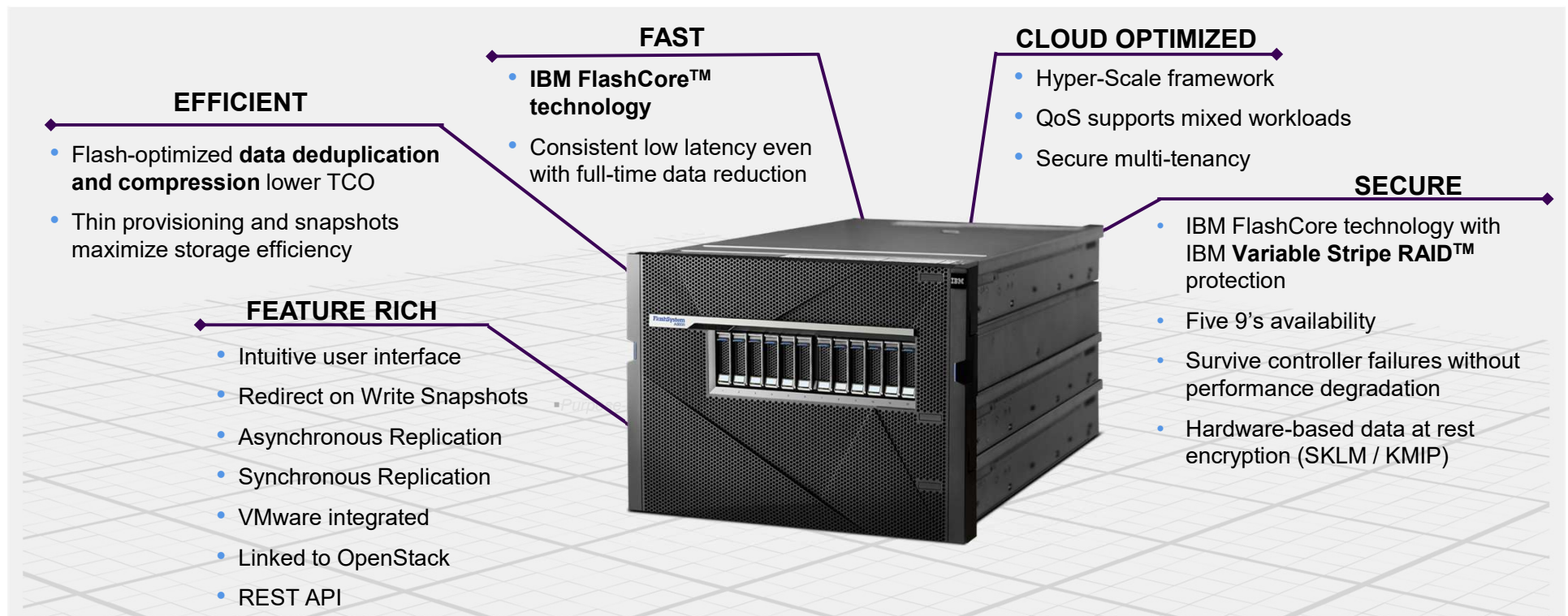
- Flash-optimized array for virtualizing the tiered enterprise data center
- **IBM FlashCore technology** combining high-performance and ultra-low latency,
- **NEW tier 0 3.6 TB, 8.5TB and 18TB** MicroLatency modules.
- Up to **4M IOPS** and **3.6PB** of capacity

DEPENDABLE

- IBM FlashCore technology with **IBM Variable Stripe RAID™** protection
- Protect data from disclosure with encryption for storage and support for redundant encryption key servers
- Get highest levels of data protection and availability with remote site replication options, and **IBM HyperSwap**.
- Enterprise Class Support with Remote support featuring Spectrum Control Storage Insights Foundation

Introducing IBM FlashSystem A9000

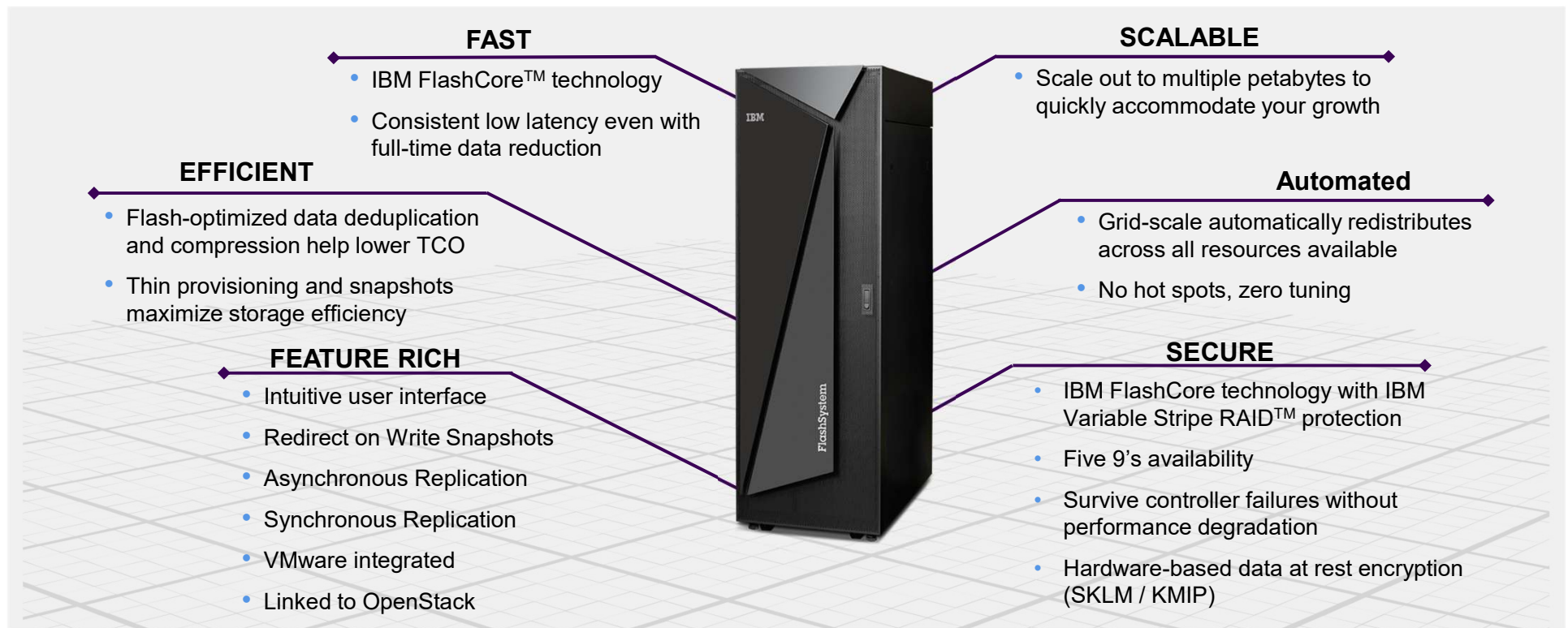
A Highly-Parallel All-Flash Platform For Cloud-Scale Business



Introducing IBM FlashSystem A9000R

IBM Storage & SDI

A Grid-Scale, All-Flash Platform Designed To Drive Your Business Into The Cognitive Era





FlashSystem A9000 and A9000R

At a Glance: What's new with v12.3

IBM Storage & SDI

Announcement:
October 9, 2018

FlashSystem A9000 & A9000R

Multi-Site High-Availability and Disaster Recovery

Supreme data protection with HyperSwap between two primary sites, and disaster recovery via simultaneous Asynchronous replication from the primary sites to a third site.



FlashSystem FlashSystem A9000R model 425 & U25

New and smaller entry point configuration

Single flash enclosure entry-point, vs 2 before, as low as 180TB

Start at
-50%
Capacity



Management Enhancements

System-wide pool A single pool can utilize the full system capacity

Monitor System level Internal & External latency Clear indication of system (vs env.) performance health

FlashSystem A9000R Model 425/U25

With new **Grid-Starter** entry-points

IBM Storage & SDI
Announcement: October 9, 2018



MicroLatency Module Size

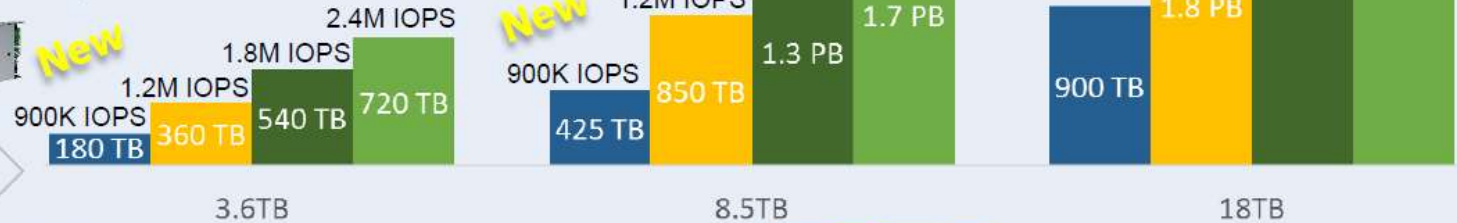
The lowest A9000R entry point is now at **180 TB**

The lowest **capacity optimized** entry point is now at **900TB**

Performance Optimized

Balanced

Capacity Optimized



* IOPS results represent random IO, 80/20% read/write
* Capacity information is Effective, assuming 5:1 data reduction

Number of Grid-Elements



IBM Block Portfolio Support for FC-NVMe & NVMe-oF

On February 20, 2018 IBM announced **new NVMe-oF hardware adapters** supporting **iWARP** or **RoCE** for Spectrum Virtualize and a Statement of Direction (SoD) covering FS900, A9000/A9000R Model 425 and IBM FlashSystem V9000 Model AC3, IBM Storwize V7000 models 624 and AF6, and IBM SAN Volume Controller Engine Model SV1

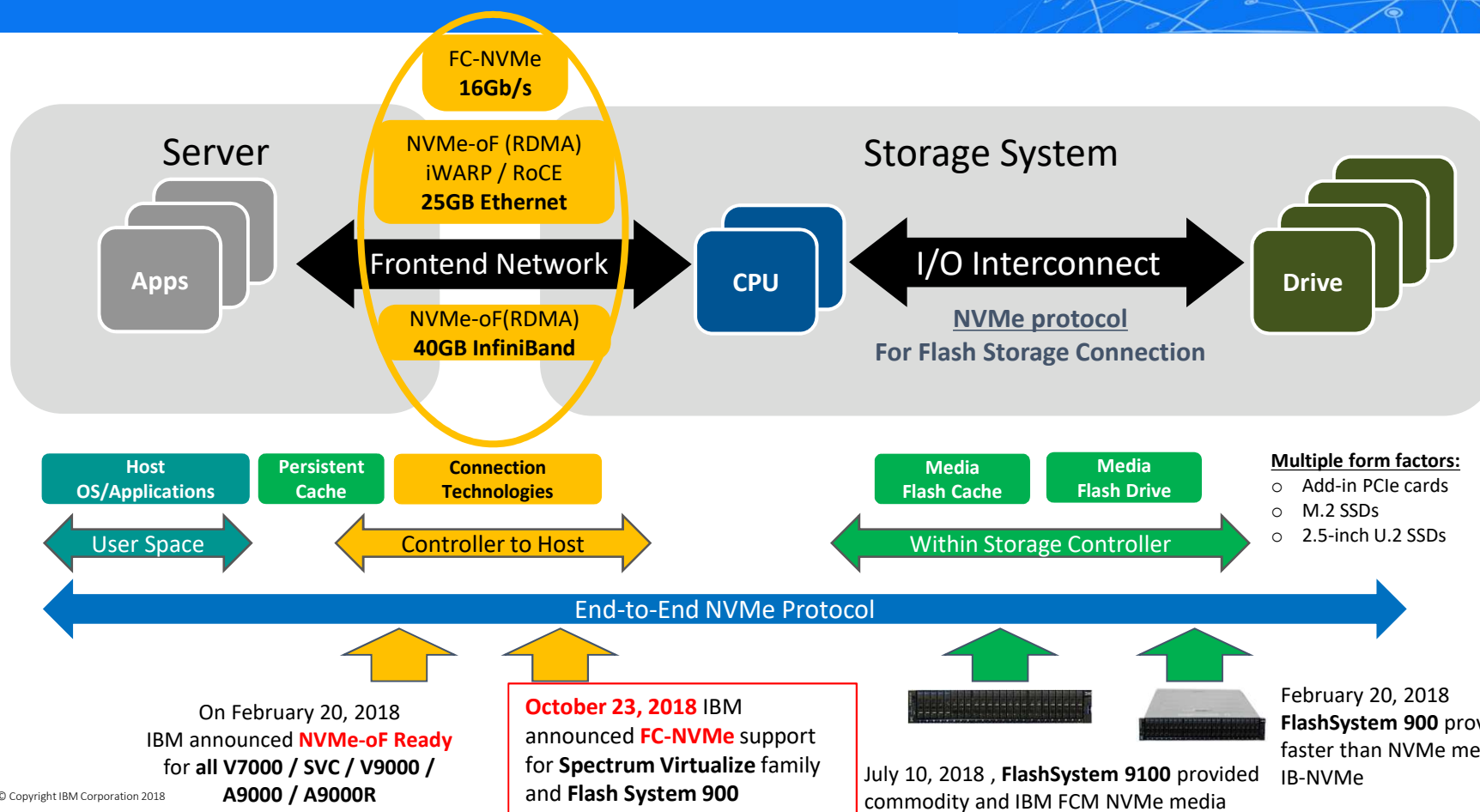
On February 27, 2018 IBM announced **FC-NVMe** hardware ready adapters for the A9000/A9000R Model 425/U25

On October 23, 2018 IBM announced **FC-NVMe firmware support for Spectrum Virtualize family and FS900**

IBM Block Systems	NVMe-oF Hardware		Software
	Type	Availability	Availability
Spectrum Virtualize Storwize V7000 / SVC / V9000 / FS9100*	25Gb iWARP/RoCE	Now	SOD
	16Gb FC-NVMe	Now	Now
FlashSystem 900*	IB NVMe-oF	Now	Now
	16Gb FC-NVMe	Now	Now
Spectrum Accelerate FlashSystem A9000* / A9000R*	16Gb FC-NVMe	Now	SOD

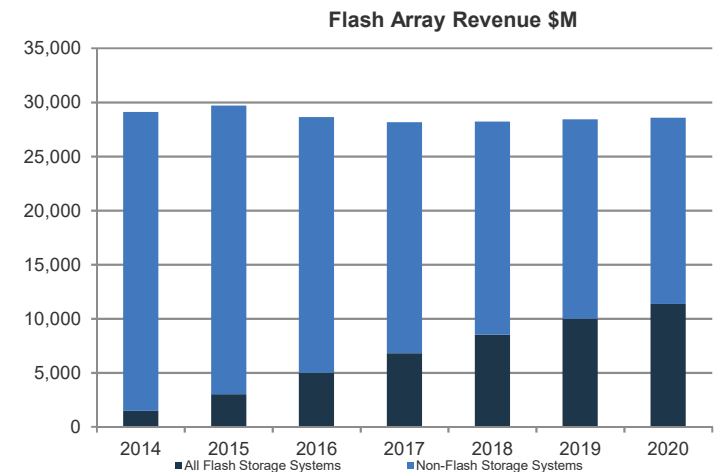
End-To-End NVMe Technologies - IBM Block Storage Offerings

IBM Storage and SDI



IBM Storage Systems Directions

- **Flash optimized systems** grow with innovations in Flash storage use
 - Storage systems designed as software on storage-rich servers with standard hardware
 - **NVMe, Compression deduplication, automated tiers**
- **Object Stores** grow in use
 - For Enterprise Clouds, and Cloud Storage Services
 - For geographically shared storage and fixed content
- **Software Defined Storage**
 - Storage services deployed as software on standard hardware with standard APIs for Block, File, and Objects
- **Storage architectures for Analytics, AI, and HPC**
- **Storage services for Cloud and Hybrid Cloud models**
 - Expect Container model to enable Hybrid Cloud





IBM Storage Systems Roadmap

- Continue Storage systems évolution to NVMe
- 32Gb Fibre Channel support
- NVMeoF support on 25Gb Ethernet (iSER-iWARP/RoCE)
- IBM Flash Core Modules (FCM) using 3D-QLC technology
- Multi Tenancy extension (Cloud)

BACKUP SLIDES

NVMe & NVMe-oF Technology...

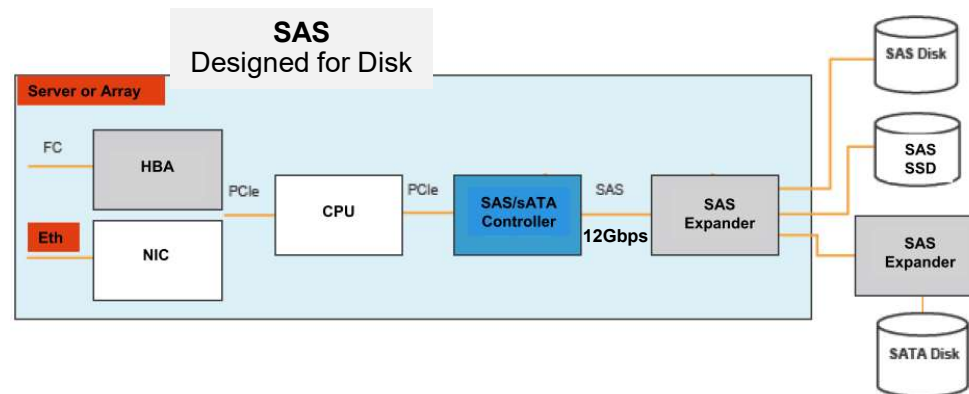
A little detail, but good to know

Where does NVMe performance come from ?

PCI Express (PCIe)

Serial Attached SCSI (SAS) Protocol

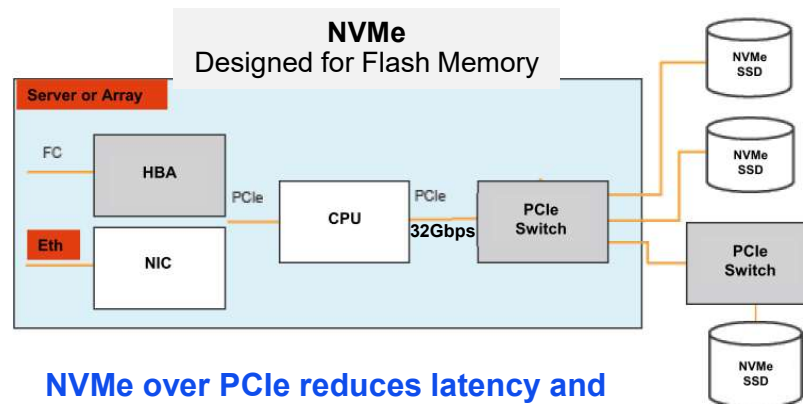
One Channel
One Queue
32 Commands
Software stack at each layer



Non-Volatile Memory (NVMe) Protocol

Up to 64,000 Queues
64,000 Commands
per Queue

Each Core has
dedicated queues
per SSD



NVMe over PCIe reduces latency and increases throughput

nvm
EXPRESS

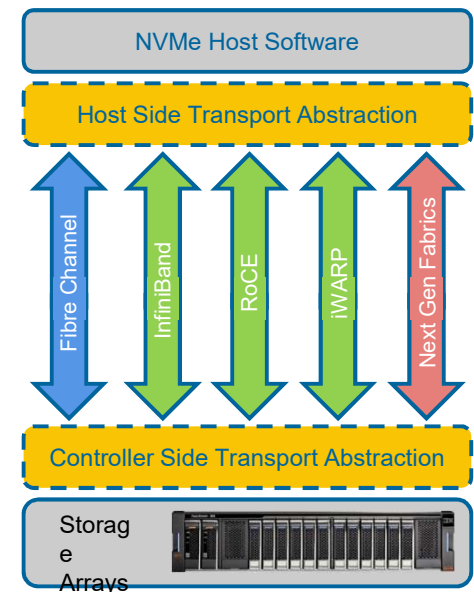
What is NVMe over Fabrics (NVMeF / NVMe-oF)

NVMe is inside the server or storage array whereas NVMe over Fabrics is across the network

- Direct Attached SSD (PCIe based) doesn't scale
- Networked storage is a must for large customers
- Only 13% of storage capacity shipped is DAS (inside the server), 87% of the total storage capacity shipped is external storage

Two types of fabric transports for NVMe currently part of the standard :

- NVMe over Fabrics using Remote Direct Memory Access (RDMA)
 - InfiniBand, RoCE / iWARP
- NVMe over Fabrics using the Fibre Channel Protocol (FCP)
 - FC-NVMe



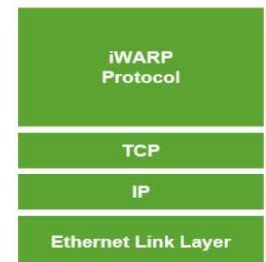
Goal of NVMe over Fabrics is to provide distance connectivity to NVMe devices with no more than 10 microseconds (μ s) of additional latency over a native NVMe device inside a server

RDMA-enabled Ethernet Technologies for NVMe over Fabrics

Connection
Technologies

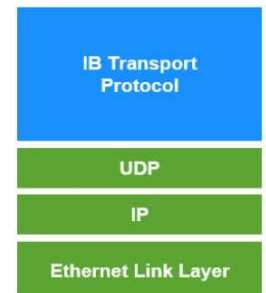
iWARP - Internet Wide Area RDMA Protocol

- Implements RDMA over Internet Protocol networks
- Layered on a mix of layers including DDP (Direct Data Placement), MPA (Marker PDU Aligned framing), RDMA protocol (RDMA over TCP/IP)



RoCE - RDMA over Converged Ethernet

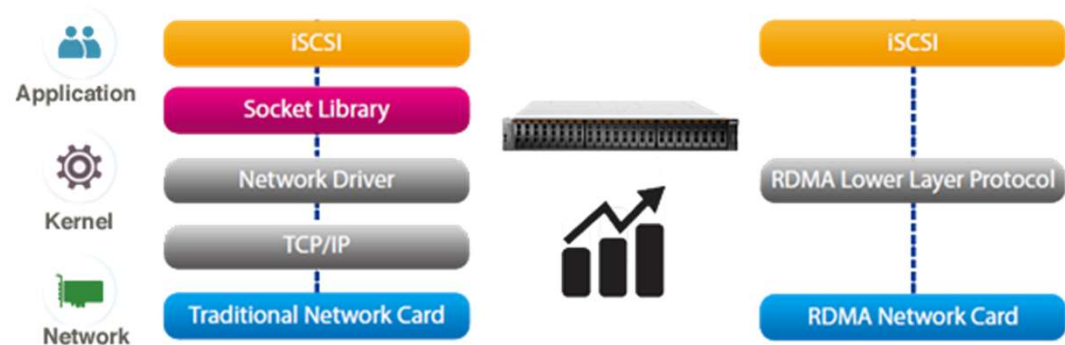
- RoCE is based on InfiniBand transport over Ethernet
- RoCEv2 enhances RoCE with a UDP header and Internet routability
 - Uses IP but not TCP
 - Uses InfiniBand transport on top of Ethernet
 - Requires a lossless DCB fabric



iSER Explained* - Fast, But Not NVMe-oF

iSER = iSCSI Extensions for RDMA

- iSER is an RDMA transport for iSCSI, the link protocol could be either Ethernet or InfiniBand at any supported speed (10, 40, 56, 100Gb/s).



What are the iSER advantages?

- iSER uses the **RDMA** protocol suite to supply **higher bandwidth** for block storage transfers (zero time copy behavior). To that fact, it eliminates the TCP/IP processing overhead while preserving the compatibility with iSCSI protocol.
- Enjoys the stability & benefits of the iSCSI protocol like security, high availability & so on.
- Faster than iSCSI, FC, FCoE
- Lower latency than iSCSI, FC, FCoE

iWARP Explained*



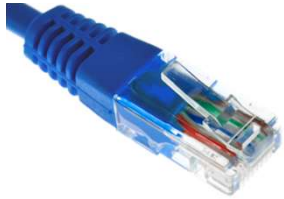
iWARP = Internet Wide Area RDMA Protocol.

- iWARP) is an update of the RDMA Consortium's RDMA over Transmission Control Protocol (TCP) standard.

Key Features:

- **Kernel Bypass** – With standard Ethernet the Processor busily aligns incoming packets into queues and aligns them for transmission. The processor gets interrupted when a queue is full and a transmission needs to occur. This all adds overhead driving up CPU utilization and slowing down network performance. However, with Kernel Bypass lets iWARP capable adaptors manage the RDMA network connection and avoiding the system software as much as possible. iWARP manages the source and destination on its own by stabling queue pairs and this bypasses the host processor. This means No Queues and no network related interrupts which means lower latency, greater bandwidth and lower CPU utilization. The advantages increase with the size of the messages. This means iWARP shines when large storage blocks are involved. Virtual Machine Migration is a perfect example of a prime activity for iWARP. One last item is that iWARP's communications are secure which protects data in flight.
- **Direct Data Placement** – Enables data placement directly into the User Queue and avoids interruptions

In all, **iWARP is the basics of InfiniBand applied to Ethernet**. This gives it tremendous capabilities for both legacy software and next generation applications.



RoCE Explained*



RoCE = RDMA Over Converged Ethernet.

- In order to promote the advancement of RDMA over Converged Ethernet (RoCE), the InfiniBand Trade Association (IBTA) launched the RoCE Initiative.

RDMA over Converged Ethernet or RoCE (pronounced “rocky”) is driving an advanced data center architecture that eliminates dedicated storage area networks and converges compute, networking and storage onto a single fabric. Leveraging the latest advances in reliable Ethernet and DataCenter Bridging (DCB), efficient RDMA mechanisms in RoCE provide lower CPU overhead and increase mainstream data center application performance.

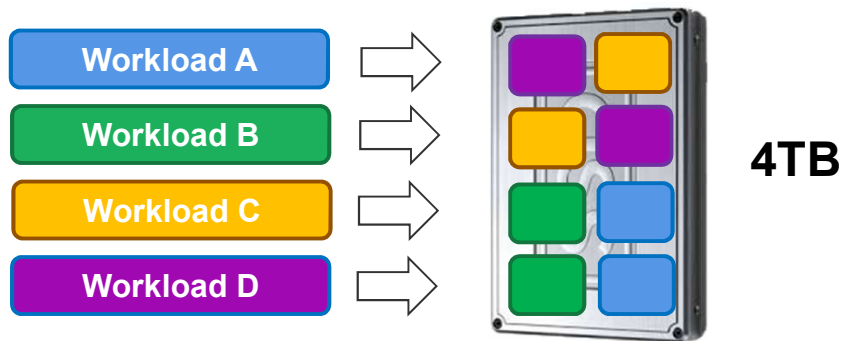
RoCE adopters can make use of RDMA’s capabilities without leaving the familiar transport and network management system of Ethernet. In this way, adopters are able to upgrade their application performance without investing in alternative switching interconnect technologies.

RoCE, now in its second version, is a routable protocol that takes RDMA and puts it on Ethernet networks. It’s not just a supercomputing technology anymore, now it is for the corporate data center too.

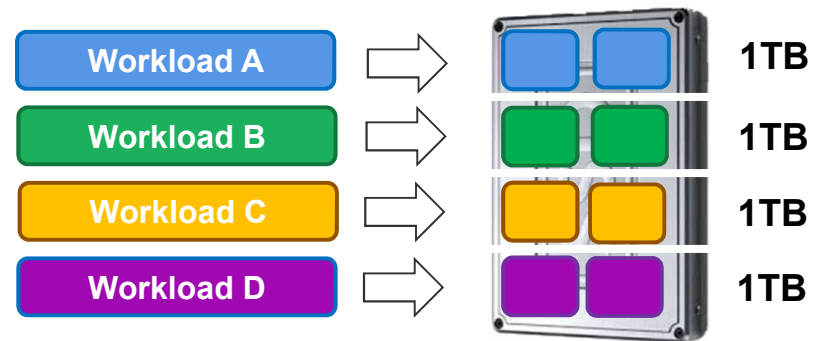
What is NVMe I/O Determinism?

- Service isolation region
- Increase Read IOPs and reduce max latency
- Provides strict QoS profile
- Significantly improves P99 and P9999 for a well-behaved host
 - Helps 99% of I/Os happen within a target latency

No I/O Determinism

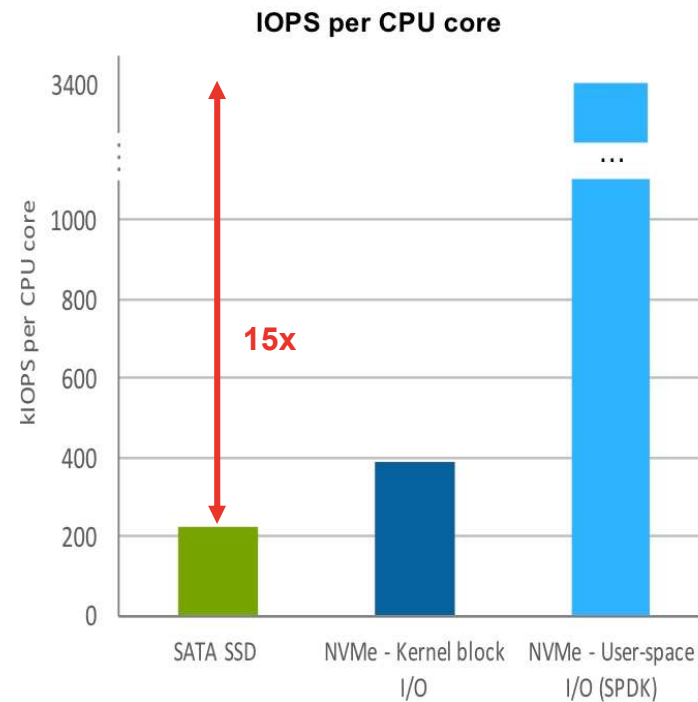


With I/O Determinism



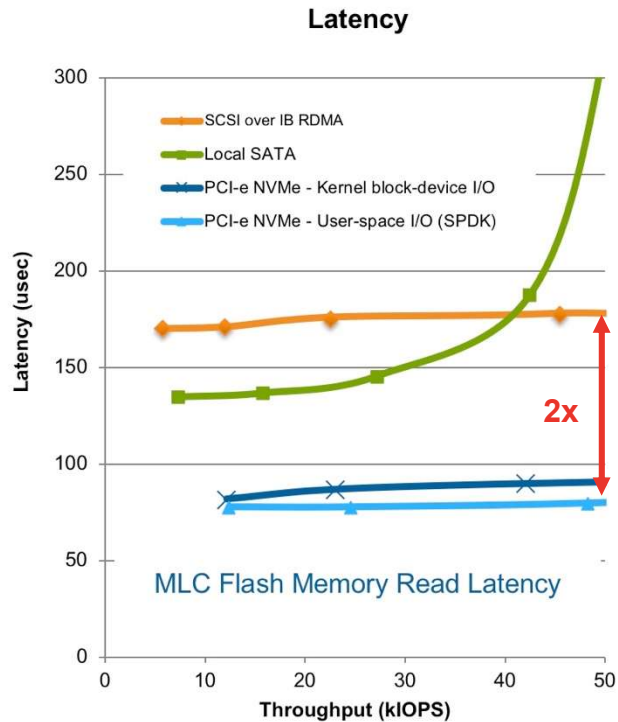
NVMe Micro-Benchmarks: CPU Utilization

**4 KiB Random
Reads achieved
using 1 CPU core
dedicated for raw I/O**

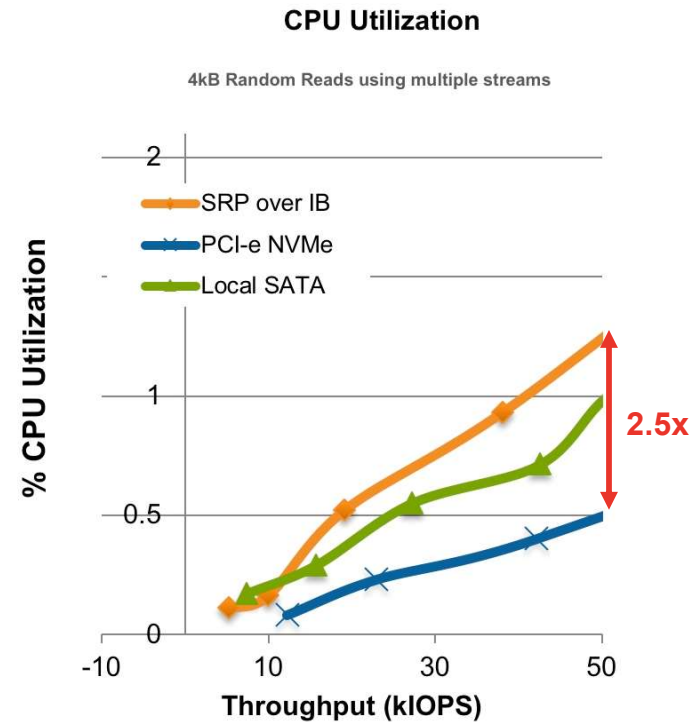


The NVMe-optimized I/O stack results in 15x efficiency improvement!

NVMe Performance



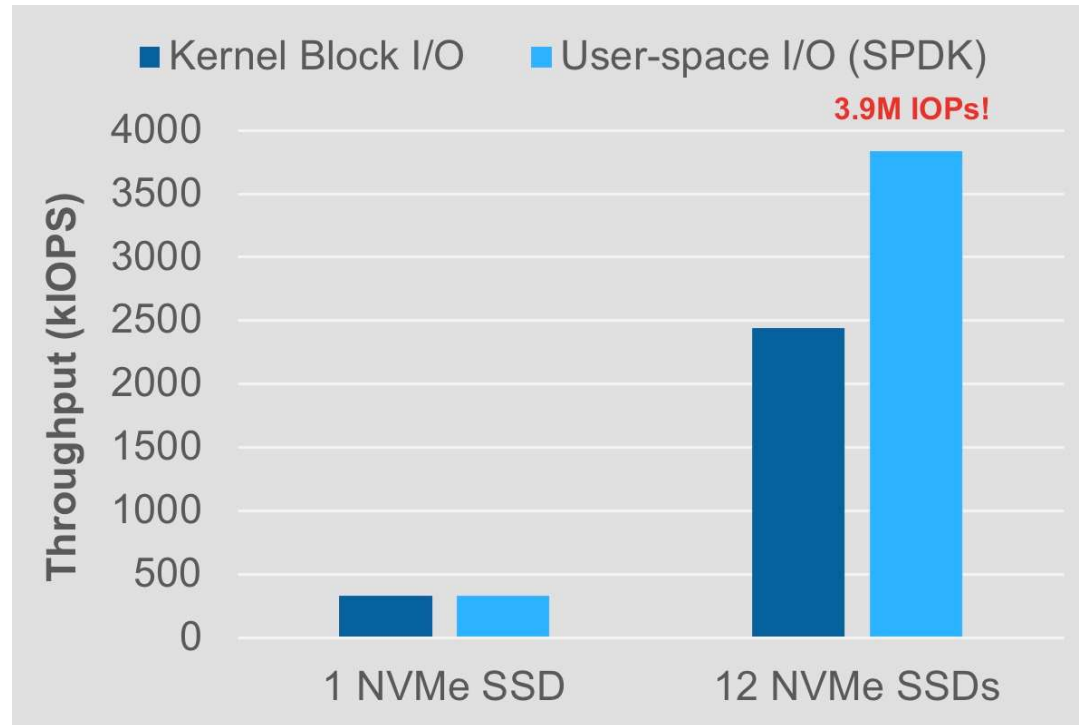
NVMe { Cuts the access latency by half.
Doubles the CPU efficiency.



NVMe latency improvement will be dramatic
with future NVMe enhancements.

NVMe Micro-Benchmarks: IOPS scalability

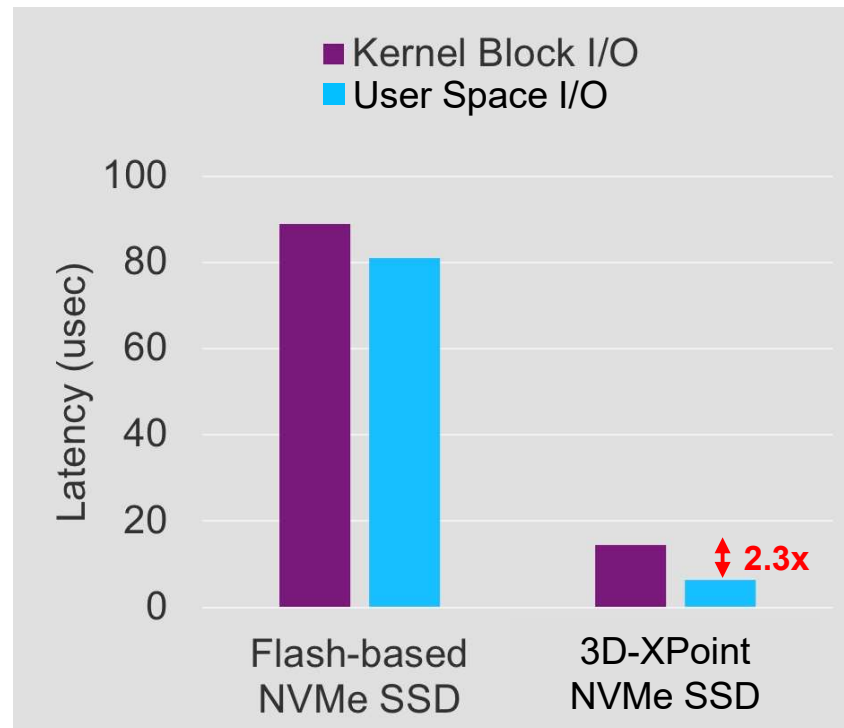
4 KiB random
read workload



The NVMe-optimized I/O stack results in extreme IOPS scalability!

NVMe Micro-Benchmarks: Latency

**Median latency
measured for 4 KiB
random read requests
at a queue depth of 1**



The NVMe-optimized I/O stack results in **15x** efficiency improvement!

The Future of NVMe

NVMe™ 1.4

- IO Determinism
- Persistent Controller Mem Buffer and Event Log
- Multipathing

NVMe-MI™ 1.1

- SCSI Enclosure Services (SES)
- NVMe-MI™ In-band
- Native Enclosure Management

NVMe-oF™ 1.1

- Enhanced Discovery
- TCP Transport Binding

