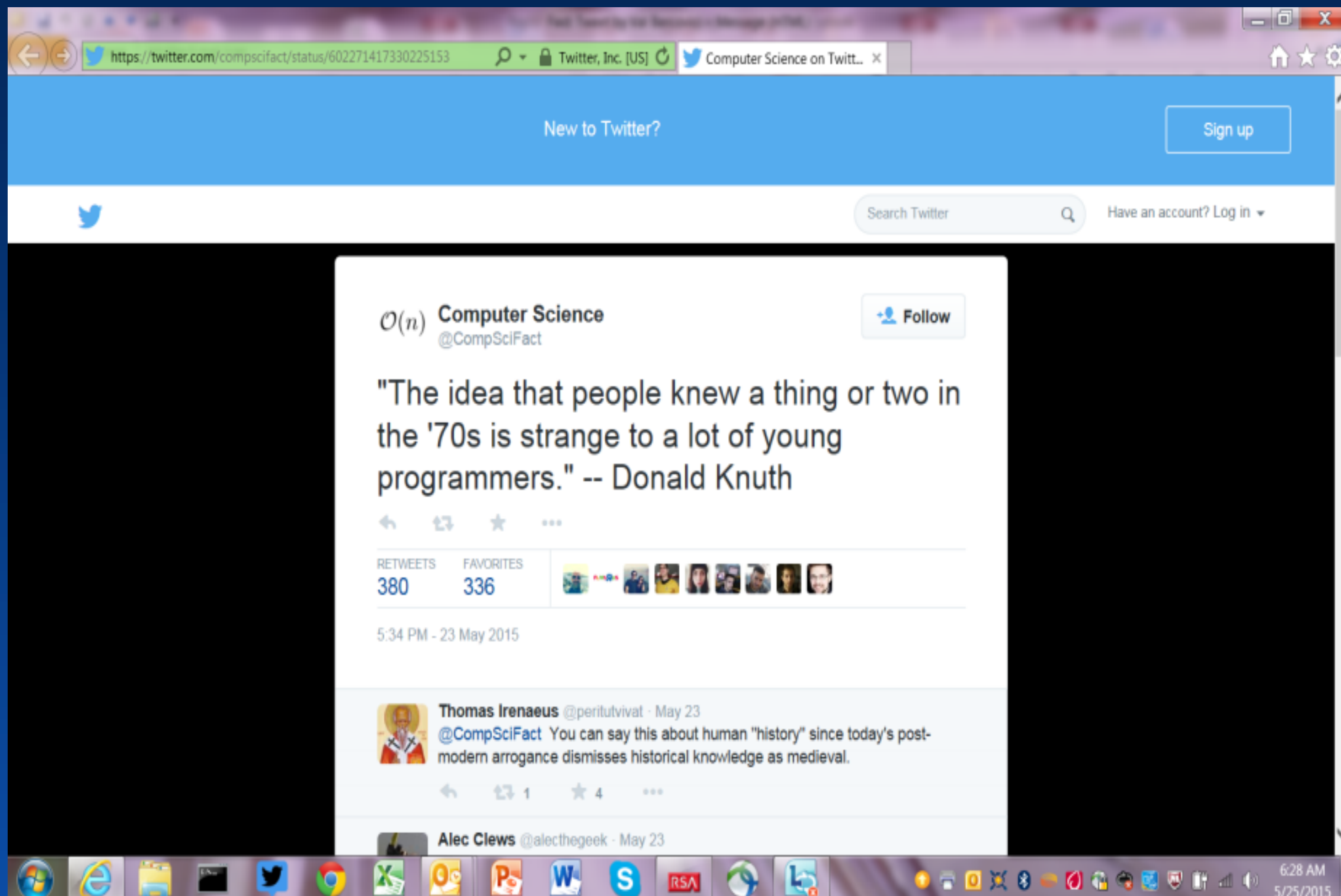
Opportunities from our Compute, Network, and Storage Inflection Points

The Brave New Persistent World

Rob Peglar
Senior VP & CTO
Symbolic IO

Wisdom



https://twitter.com/compscifact/status/602271417330225153
Twitter, Inc. [US]
Computer Science on Twitt...
New to Twitter?
Sign up
Search Twitter
Have an account? Log in
Computer Science
@CompSciFact
Follow
"The idea that people knew a thing or two in the '70s is strange to a lot of young programmers." -- Donald Knuth
RETWEETS
380
FAVORITES
336
5:34 PM - 23 May 2015
Thomas Irenaeus @perituvivat · May 23
@CompSciFact You can say this about human "history" since today's post-modern arrogance dismisses historical knowledge as medieval.
Alec Clews @alecthegeek · May 23
6:28 AM
5/25/2015

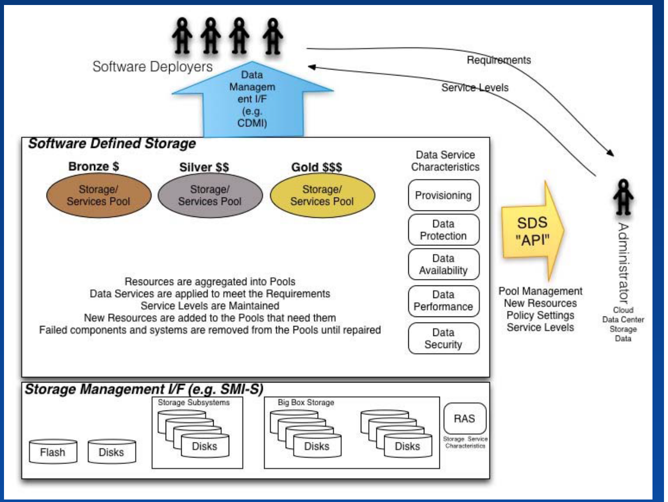
The Macro Trend – Back to the Future

- In 1984 John Gage of Sun Microsystems said:
  - “The Network is the Computer”
  - Personal Note – I was working with Sun workstations in 1984, and appreciated what he said – he was right
- We had compute & storage in a ‘workstation’
  - Which we now know as a server
  - Everything was local – no SAN, no RDMA, shared-nothing

- Sun’s innovation was built-in LAN interfaces
  - 10Mb/sec Ethernet – inside the workstation – as standard
- The entire design of SunOS revolved around LAN connections
- Today, “the Server is the Computer” – back to the future
- Hyperscale – Hyperconvergence – Hyperclustering – HyperHype?
  - No…it really is one of those once-in-a-lifetime inflection points!

Trend #1 – Software-Defined Everything

- Software-Defined is a trend to further abstract low layers of the 'stack' away
  - collapse all resources into host-based entities
- Simplifies deployment and allows high levels of automation
- RESTful APIs FTW – simple automation, scripting, networking
- SD Storage (SDS) is not new – been around since ~1999
  - Early players did local HDD and SAN array aggregation
  - Today local SSD, HDD and SAN/NAS/Object array abstraction
  - Leverages powerful compute cores & high-bandwidth LANs
  - Hypervisor-based and bare-metal based – both kinds exist
- SD Networking (SDN) is newer
  - Leverage host-based switching/routing s/w stacks
  - Use server-based NICs like switch-based ports
  - QoS, out-of-band, other operations much easier to abstract
  - Mostly hypervisor-based today (e.g. VMW NSX)

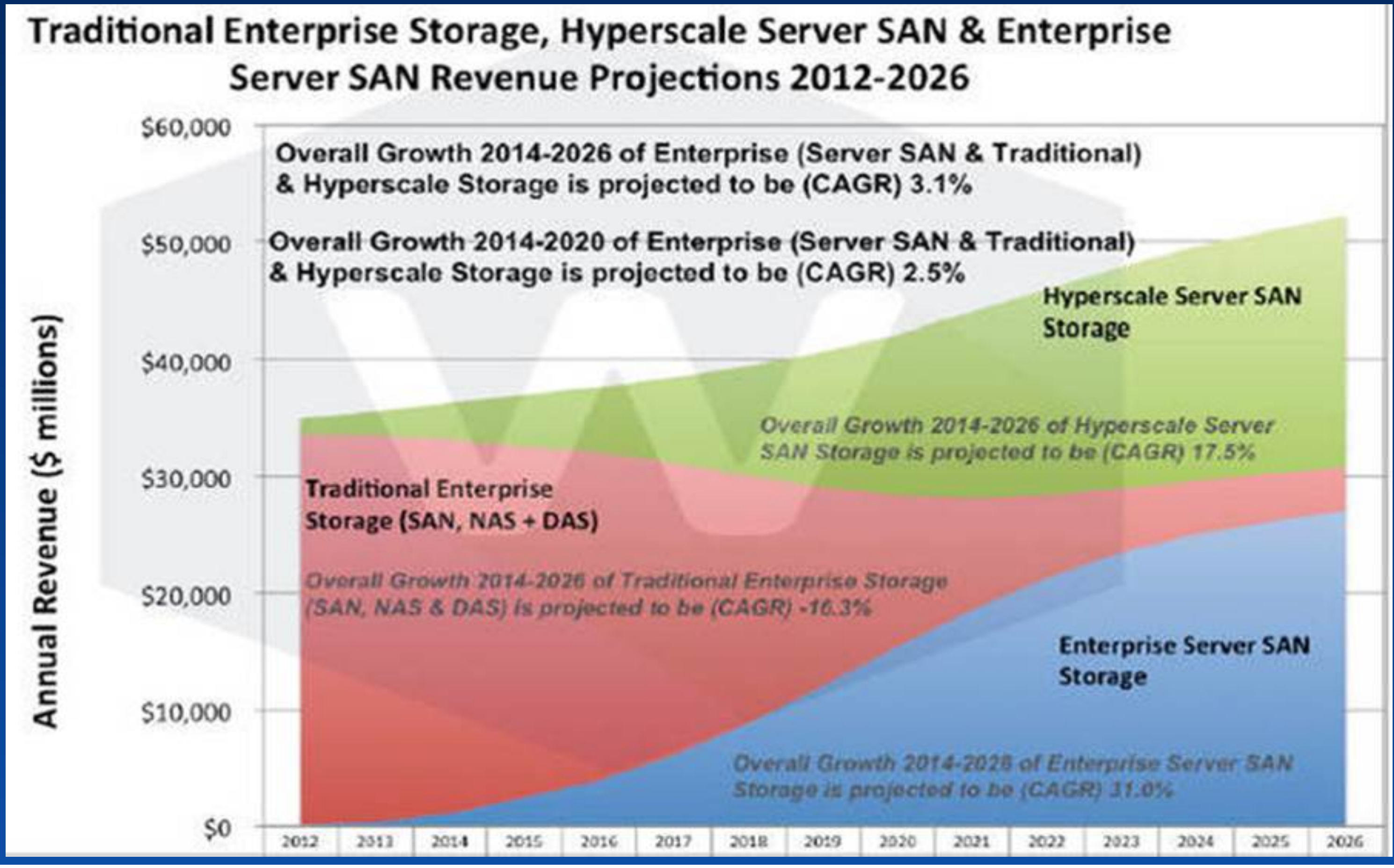


Courtesy SNIA

Trend #2 – The Start of the End of SAN

- Local (host-based) storage is once again rising
  - especially SSD
  - PCI-E/NVMe – high throughput/low latency storage
- SAN was useful for sharing and aggregating storage
  - provisioning/assigning/command & control paths
  - the logical extension of multi-initiator SCSI in the late ’80s
- Today SDS and host-based clustered file and object systems
  - perform the same functions w/o array controllers in the control & data paths
- Moving away from host-fabric-array configurations using channel technology (e.g. FC, SAS)
- Moving (back) towards host-host configurations using LAN technology (e.g. Ethernet, IB)
- Commoditize, standardize, virtualize, containerize, …

Analyst Data - Wikibon



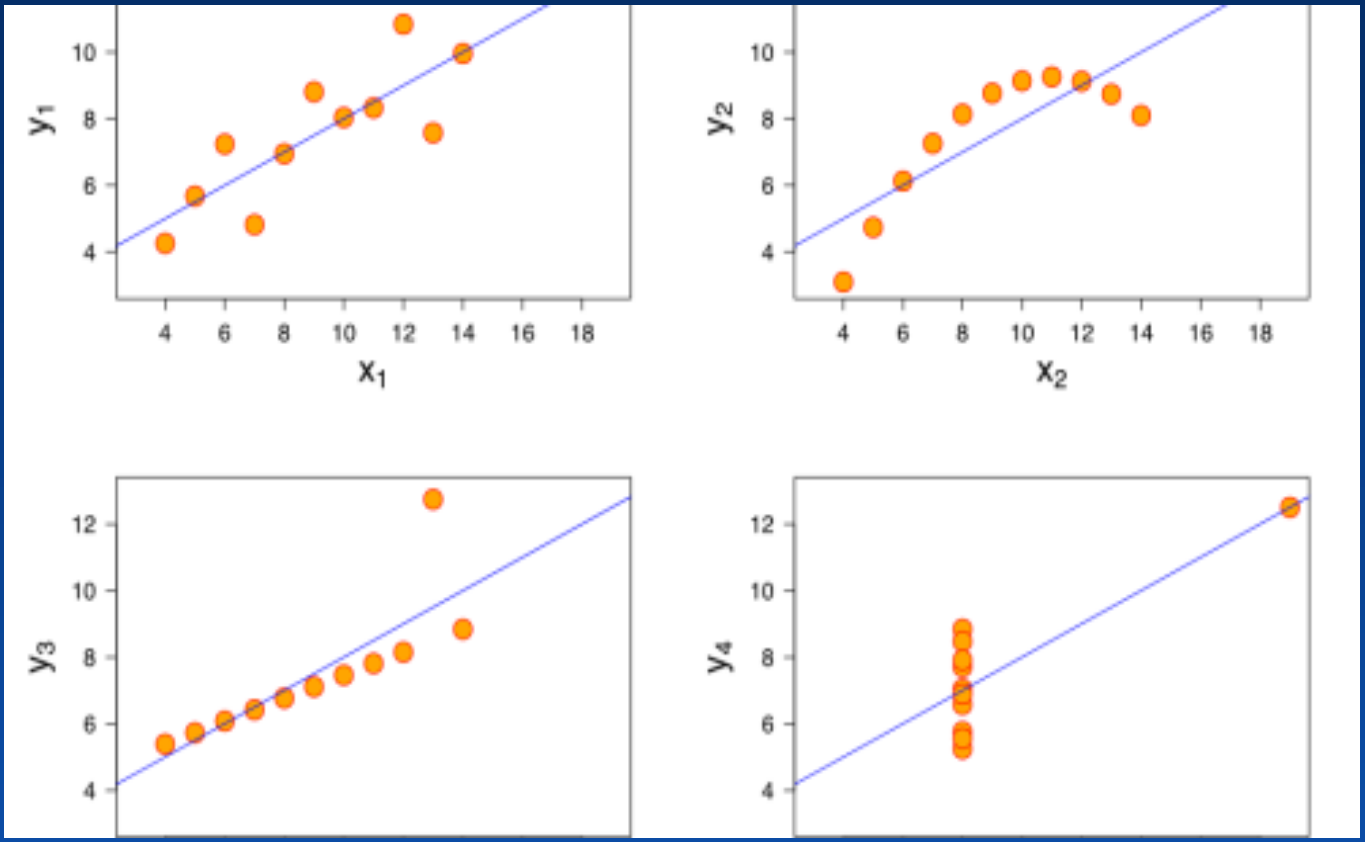
Trend #3 – The Start of the End of HDD

- The HDD has been with us since 1956
  - IBM RAMAC Model 305 (pictured)
  - 50 dual-sided platters, 1,200 RPM, 100 Kb/sec
  - 5 million 6-bit characters (3MB)
- Today – the SATA HDD of 2016
  - 7 dual-sided platters, 7,200 RPM, 100 MB/sec
  - 8 trillion 8-bit characters (8TB) in 3.5”
  - Over 2 million X denser and 10,000 X faster (throughput)
  - Problem is only 6X faster rotation speed – which means high latency
- With 3D TLC NAND technology we can easily build >10TB 2.5” SSDs
- We’ve solved the capacity/density problem - the throughput & latency problem was already solved
  - And continues to improve (e.g. NVM-E)
- On a $/TB basis SSDs are nearing price parity as deployed in servers, 2.5” U.2

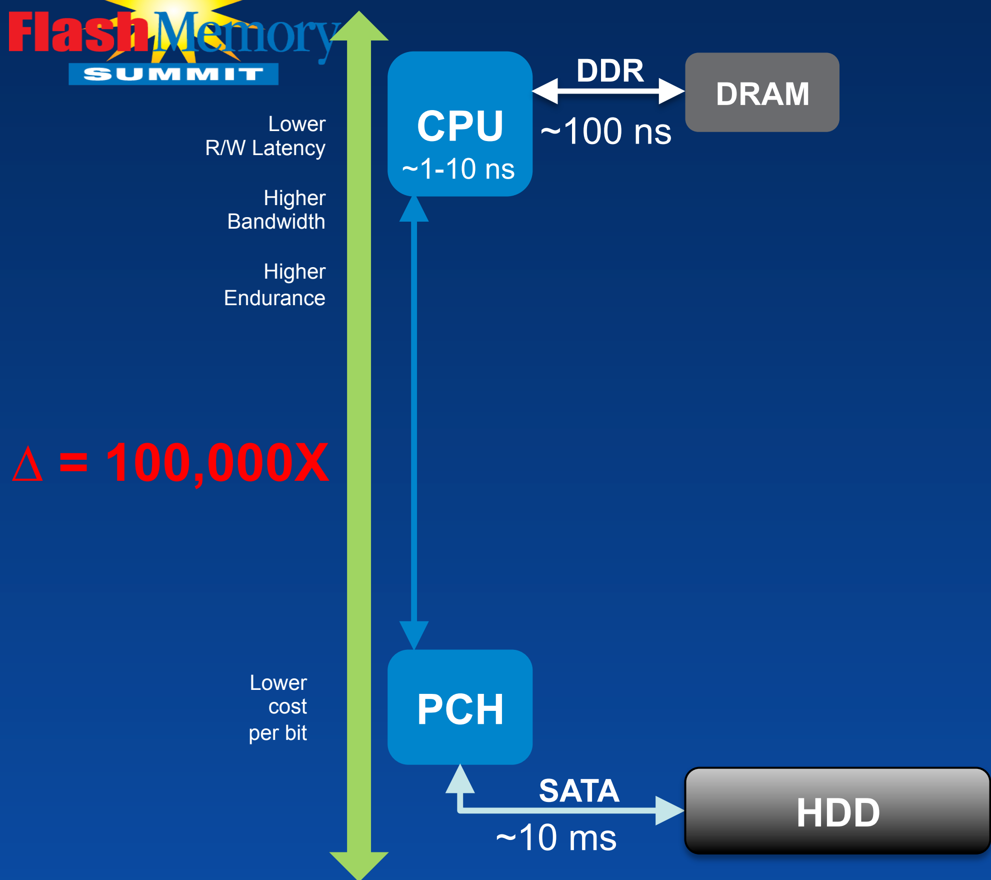
It's All About Workloads

- Two interesting, strongly growing (positive CAGR) areas
- One flat (or negative CAGR) large uninteresting area

Data is Interesting - Anscombe's Quartet

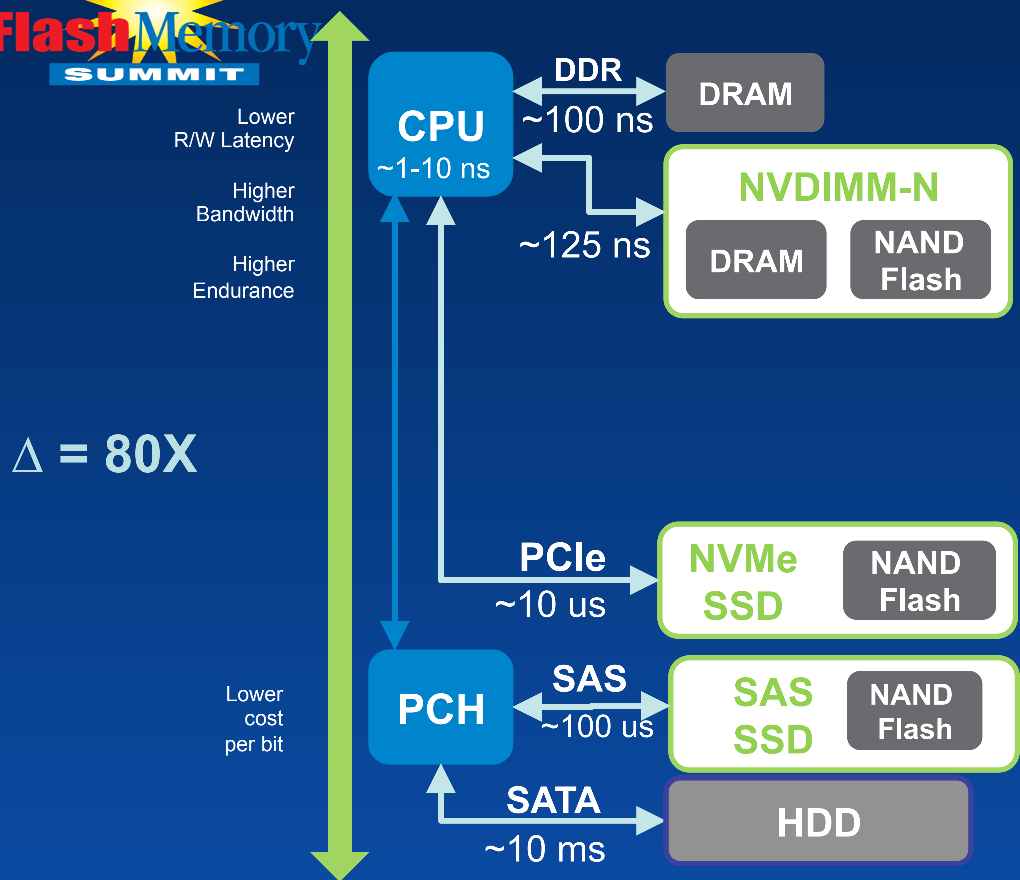


The Past: Nonvolatile Memories in Server Architectures



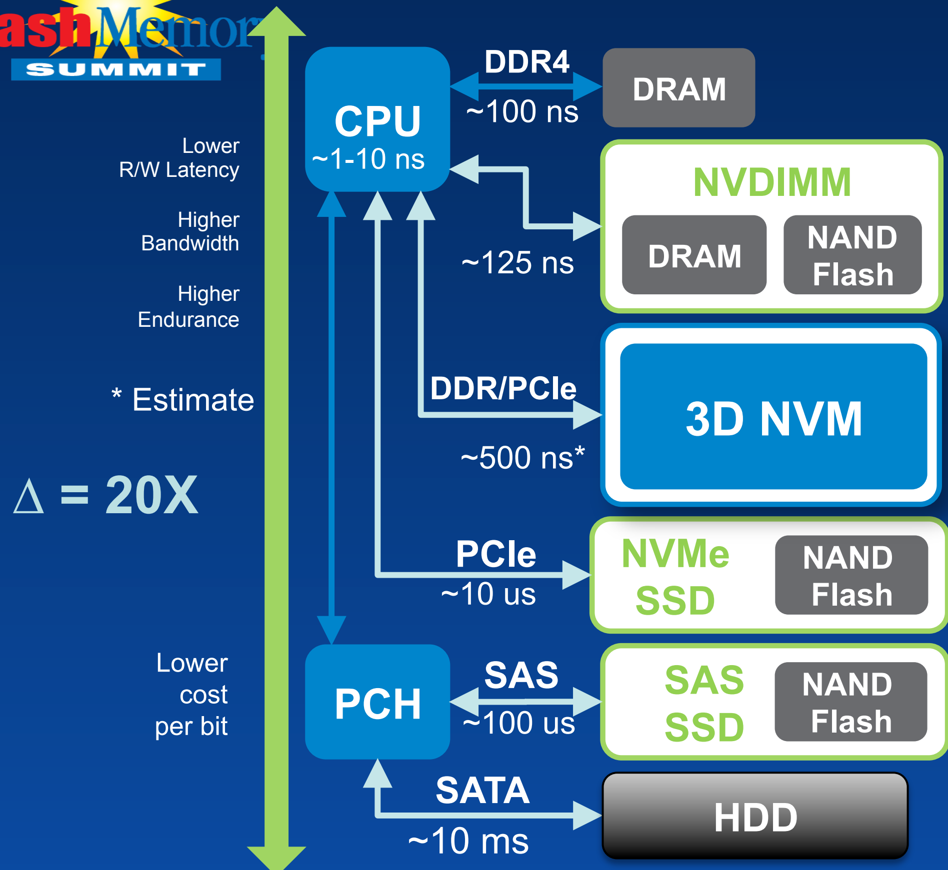
- For decades we've had two primary types of memories in computers: DRAM and Hard Disk Drive (HDD)
- DRAM was fast and volatile and HDDs were slower, but nonvolatile (aka persistent)
- Data moves from the HDD to DRAM over a bus where it is the fed to the processor
- The processor writes the result in DRAM and then it is stored back to disk to remain for future use
- SATA HDD is 100,000 times slower than DRAM (!!!)

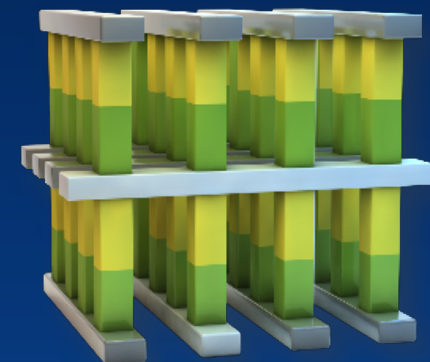
The Present: 2D Hybrid Memory Server Architectures



- System performance increased as the speed of both the interface and the memory accesses improved
- NAND Flash considerably improved the nonvolatile response time
- SAS and PCIe made further optimizations to the storage interface
- NVDIMM provides battery- or ultra-capacitor-backed DRAM, operating at near-top speeds (typ. 1600) and retains data when power is removed
- NVMe transport provides efficient use of PCI-Express bus (queues, etc.)

The Future: 3D Nonvolatile Memories in Server Architectures





Courtesy Micron

- NVM technology provides the benefit in 'the middle' – reduces the gap
- Significantly faster than NAND Flash with much higher endurance
- Performance can be realized on PCIe or DDR buses – storage or memory
- Lower cost per bit than DRAM while being considerably more dense
  - Software-enabled via PMEM & others

Opportunities From The Inflection Point

- There is no question whatsoever that persistent memory changes compute
- But does it change storage?
  - Is persistent memory just faster storage for what we have?
- Should I just throw persistent memory ‘at the problem’?
- This technique is currently being used in SSDs
  - 3DX included

- Throw NVMe at the problem – faster transport, less overhead, more queues, etc.
- Throw dense 3D NAND flash at the problem – 512TB in 3U – save W,BTU,RU
- That’s all well and good – BUT …

Solve the Weiji

- We have a weiji on our hands 危機, translated, 'critical point'
- Instead of treating data like we have for ~60 years now – blocks – look at bits
  - Like DNA – order matters – only two base pairs (A+T, C+G) – adapts over time - genomics
- Translate (encode) raw data into a better (space efficient, compute efficient, secure) form
- Use only persistent memory to hold metadata and translation aids (e.g. bit markers)
  - No disk necessary of any kind – SSD or otherwise – leverage byte-addressable methods
  - Takes only 4GB of DRAM to hold all possible combinations of 32-bit entities (2^32)
- It takes 112 bits (13 8-bit letters and a blank) to represent 'critical point' in ASCII
- It takes 26 bits (2 13-bit symbols) to represent 危機
- Don't store raw data – compute the representation of it – store the markers/instructors
- CHANGE THE GAME – not the rules – compute in-memory using 10X the raw surface

Opportunities from our Compute, Network, and Storage Inflection Points

The Brave New Persistent World

Rob Peglar
Senior VP & CTO
Symbolic IO