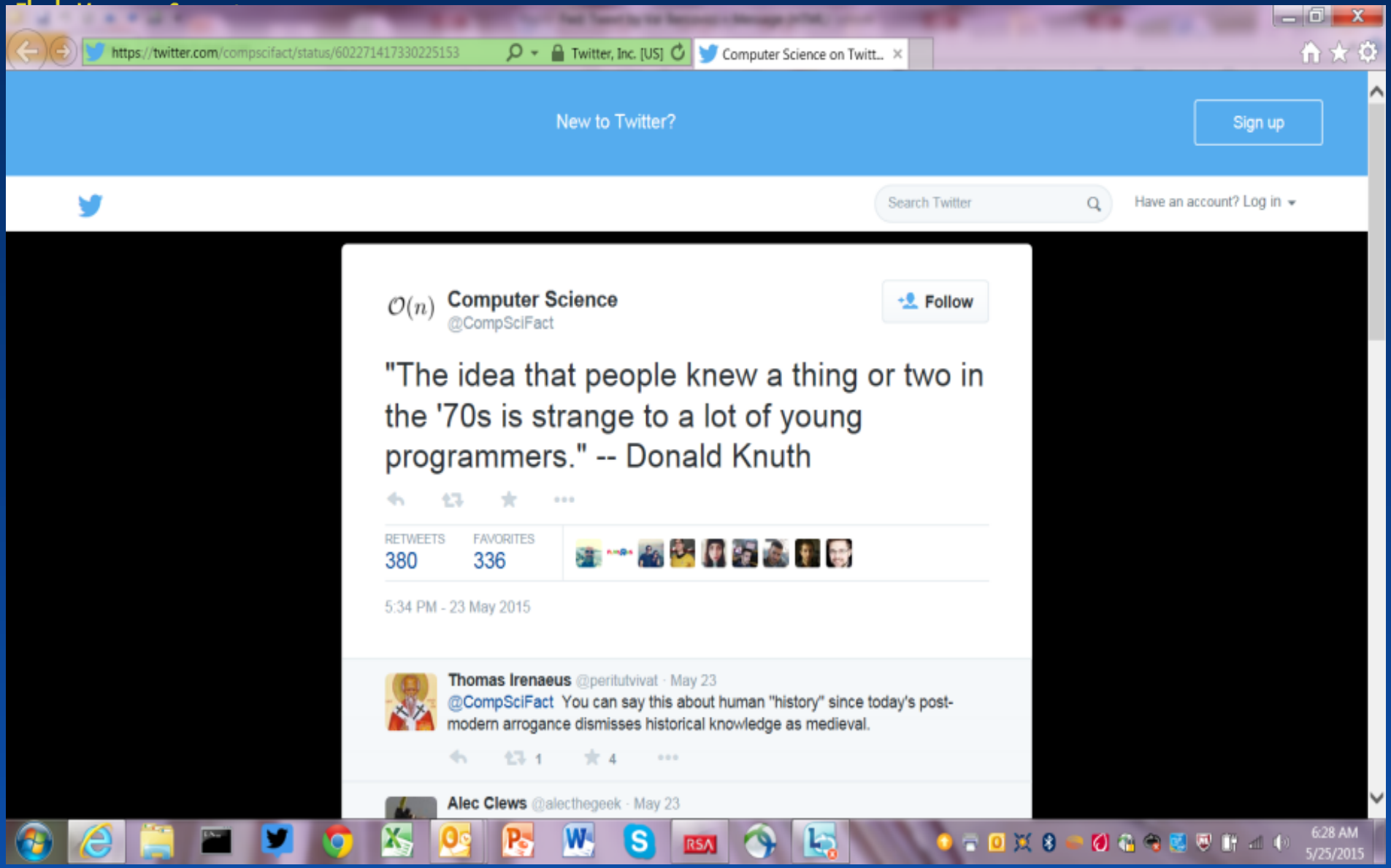
New Computational Approaches to Big Data

The Impact of Persistent Memory and Intelligent Data Encoding

Rob Peglar
SVP/CTO
Symbolic IO
rpeglar@symbolicio.com
@peglarr

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 @peritutvivat · May 23
@CompSciFact You can say this about human "history" since today's post-modern arrogance dismisses historical knowledge as medieval.
1
4
Alec Clews @alecthegeek · May 23
6:28 AM
5/25/2015

The Macro Trend - Back to the Future

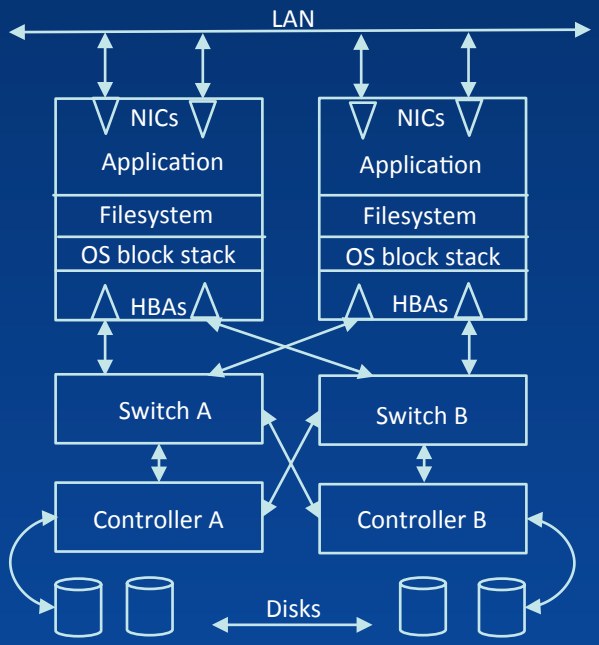
- In 1984 John Gage of Sun Microsystems said:
  - "The Network is the Computer"
    - Personal Note – I was working with Sun workstations in 1984
    - I appreciated what he said – he was right
- Back then we had compute & storage in a 'workstation'
  - Which we now know as a server
  - Everything was local – no SAN, no RDMA, no shared storage
- Sun's innovation was to put 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!

The Micro Trend
-The Start of the End of HDD

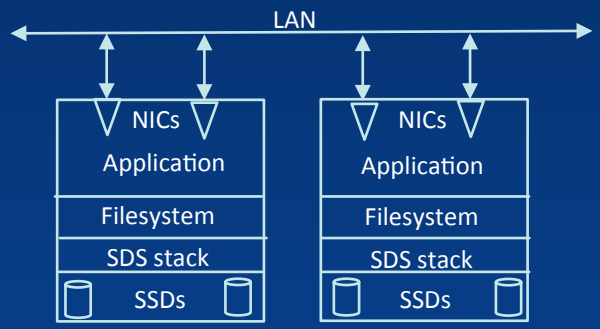
- The HDD has been with us since 1956
  - IBM RAMAC Model 305 (picture →)
  - 50 dual-side platters, 1,200 RPM, 100 Kb/sec
  - 5 million 6-bit characters (3MB)
- Today – the SATA HDD of 2017
  - 7 dual-side platters, 7,200 RPM, 100 MB/sec
  - 10 trillion 8-bit characters (10TB) in 3.5”
  - Over 2 million X denser and 10,000 X faster (throughput)
  - Problem is only 6X faster rotation speed – which means latency
- With 3D QLC NAND technology we get > 32 TB 2.5” SSDs
- Which means we’ve solved the capacity/density problem
  - Throughput & latency problem was already solved
  - Continues to improve by leaps and bounds (e.g. NVMe, NVMe-oF)
- Now we’re being hit with the memory bottleneck

The Beginning of the End of Frames

- Choose one design…you pick



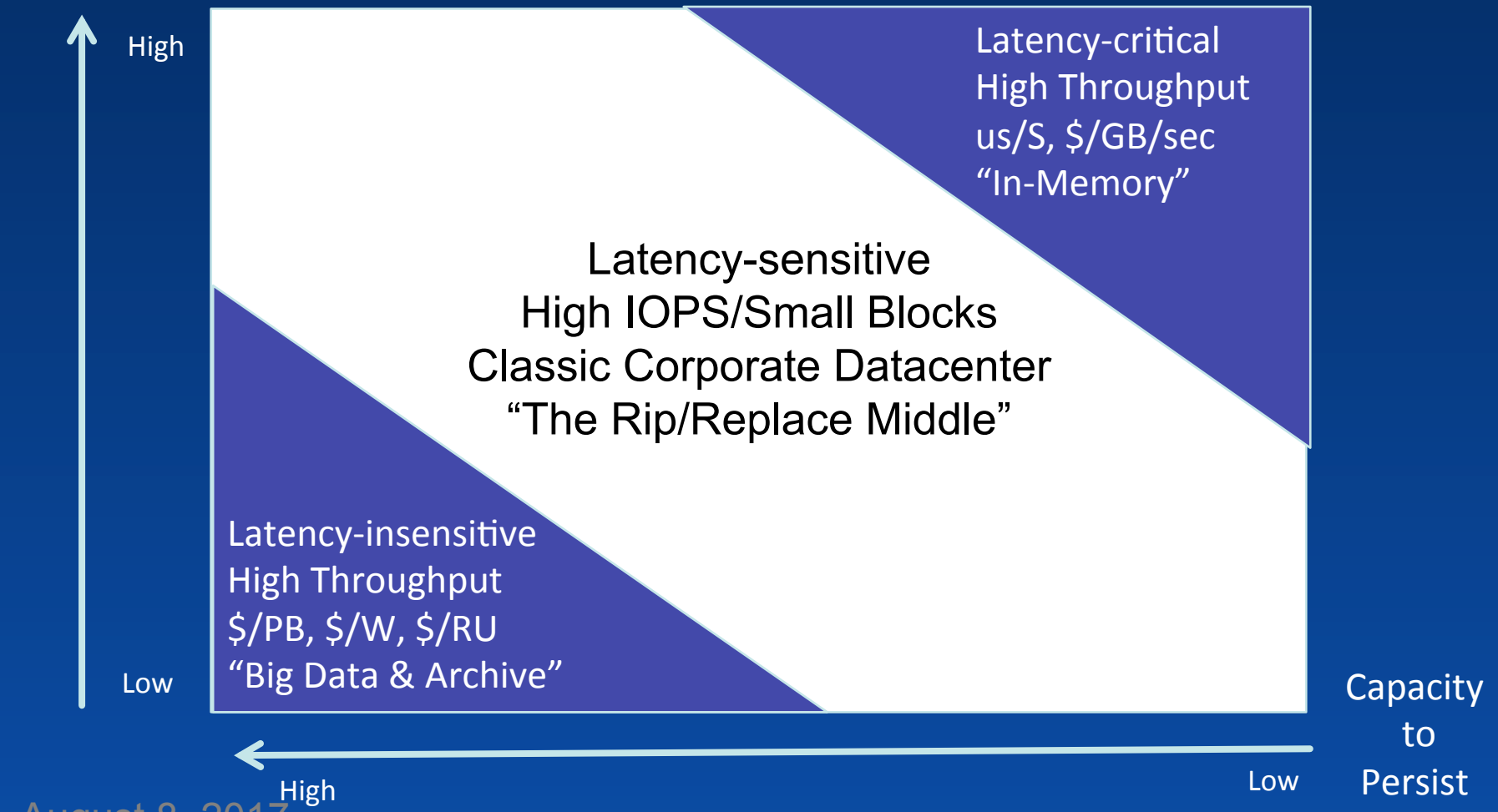
OR



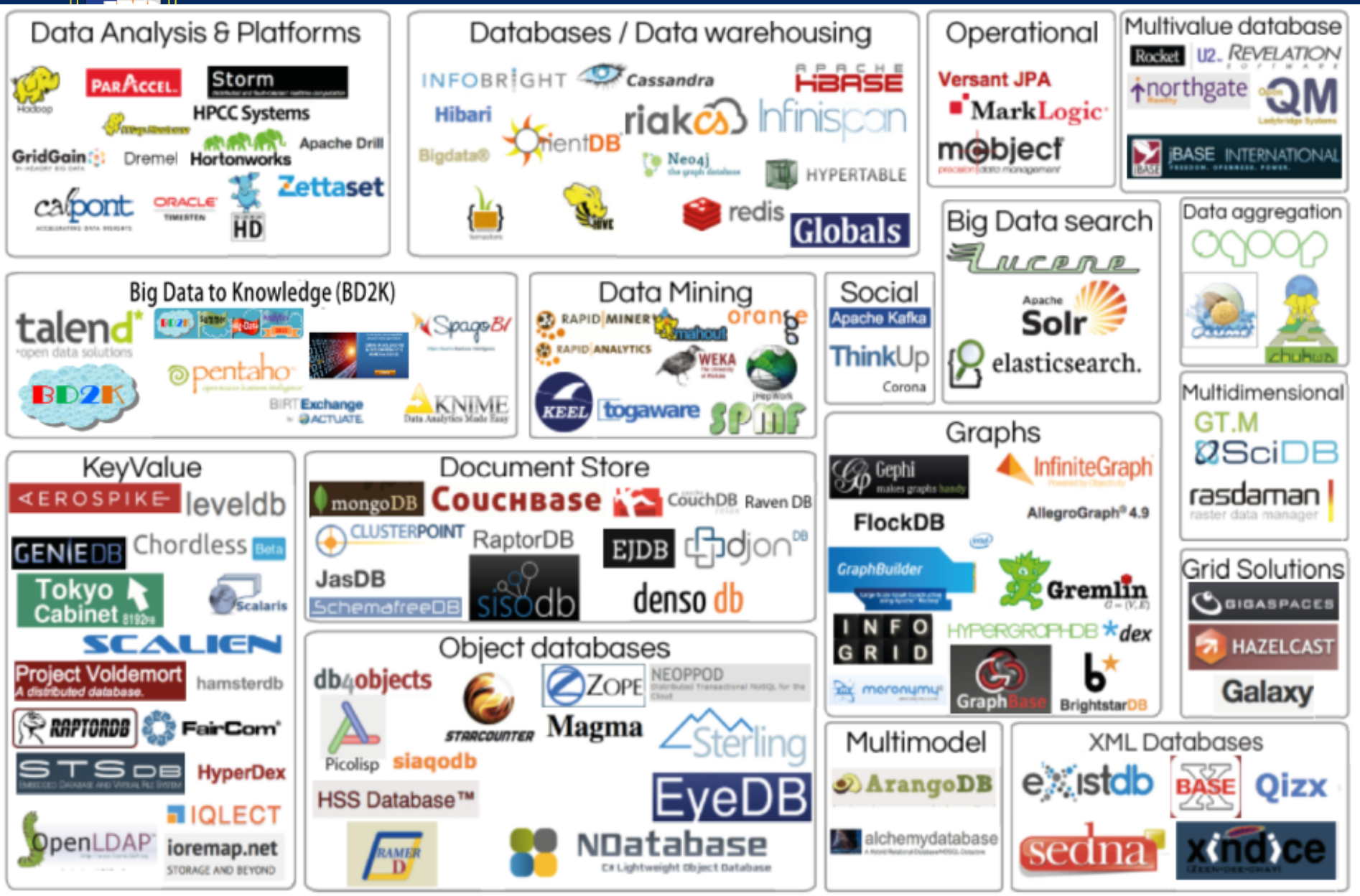
Which design is simpler?
Which design is more effective?
Which design is more scalable?

It's All About Workloads

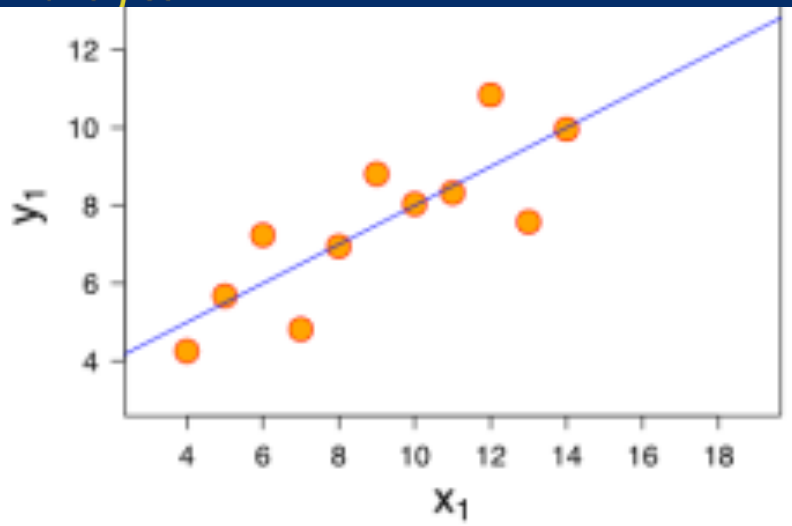
- Two interesting, strongly growing (high CAGR) areas
- One flat (or negative CAGR) large uninteresting area

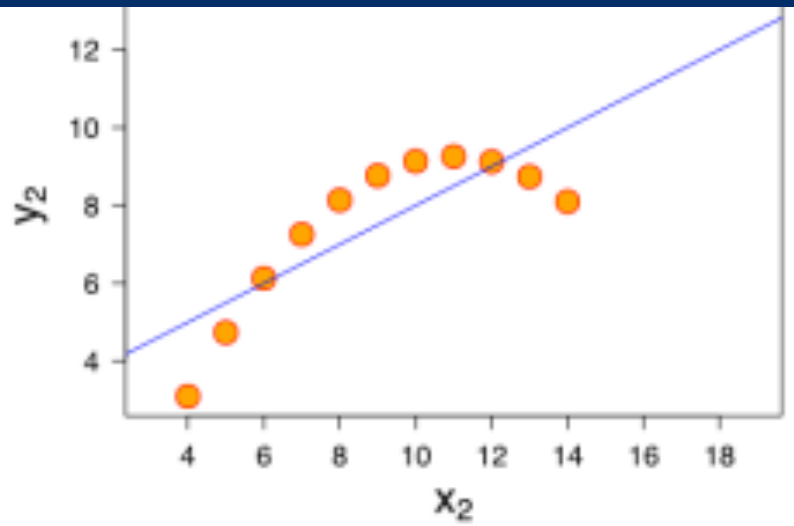


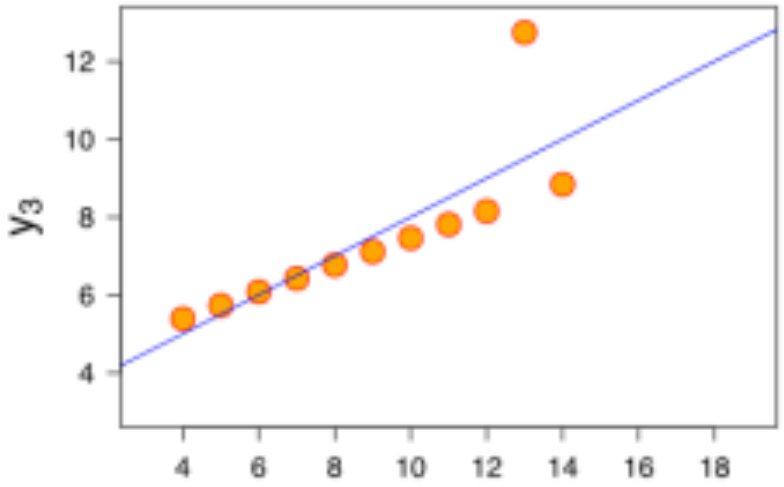
The Analytics Landscape

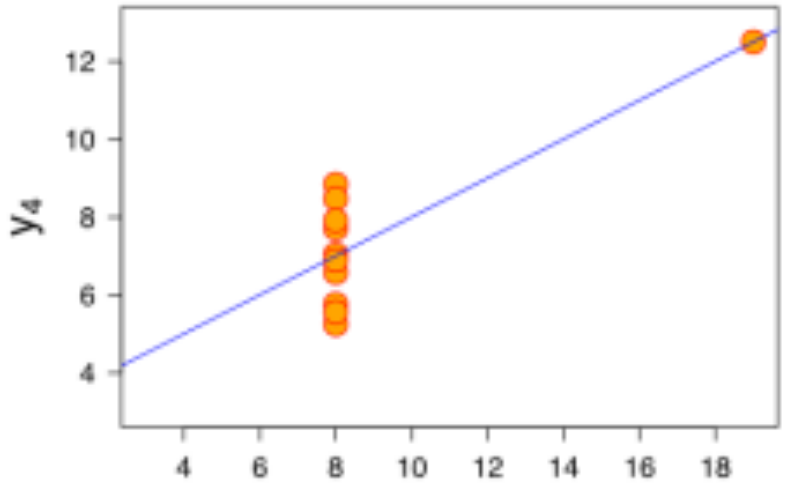


Data is Interesting – Anscombe's Quartet

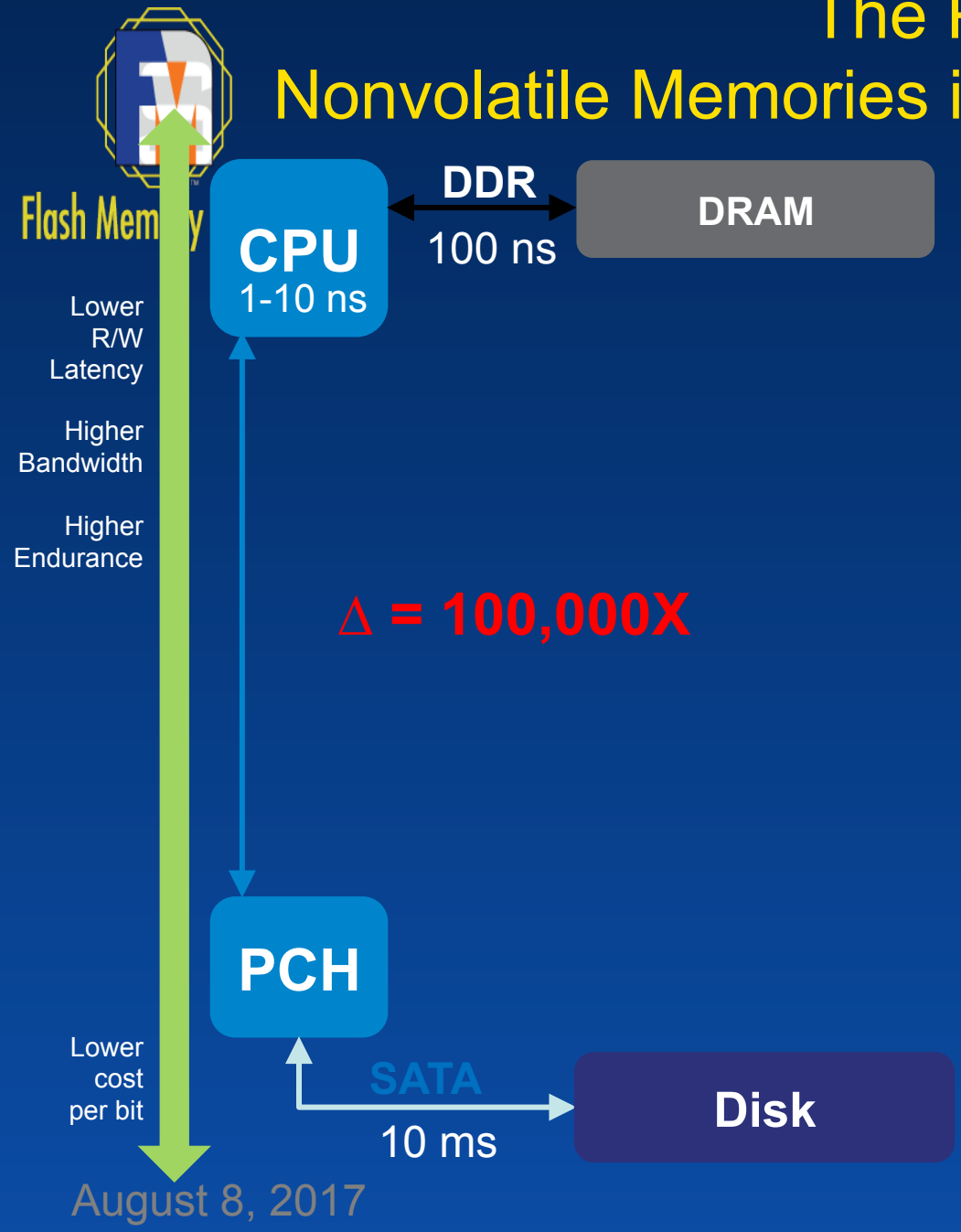






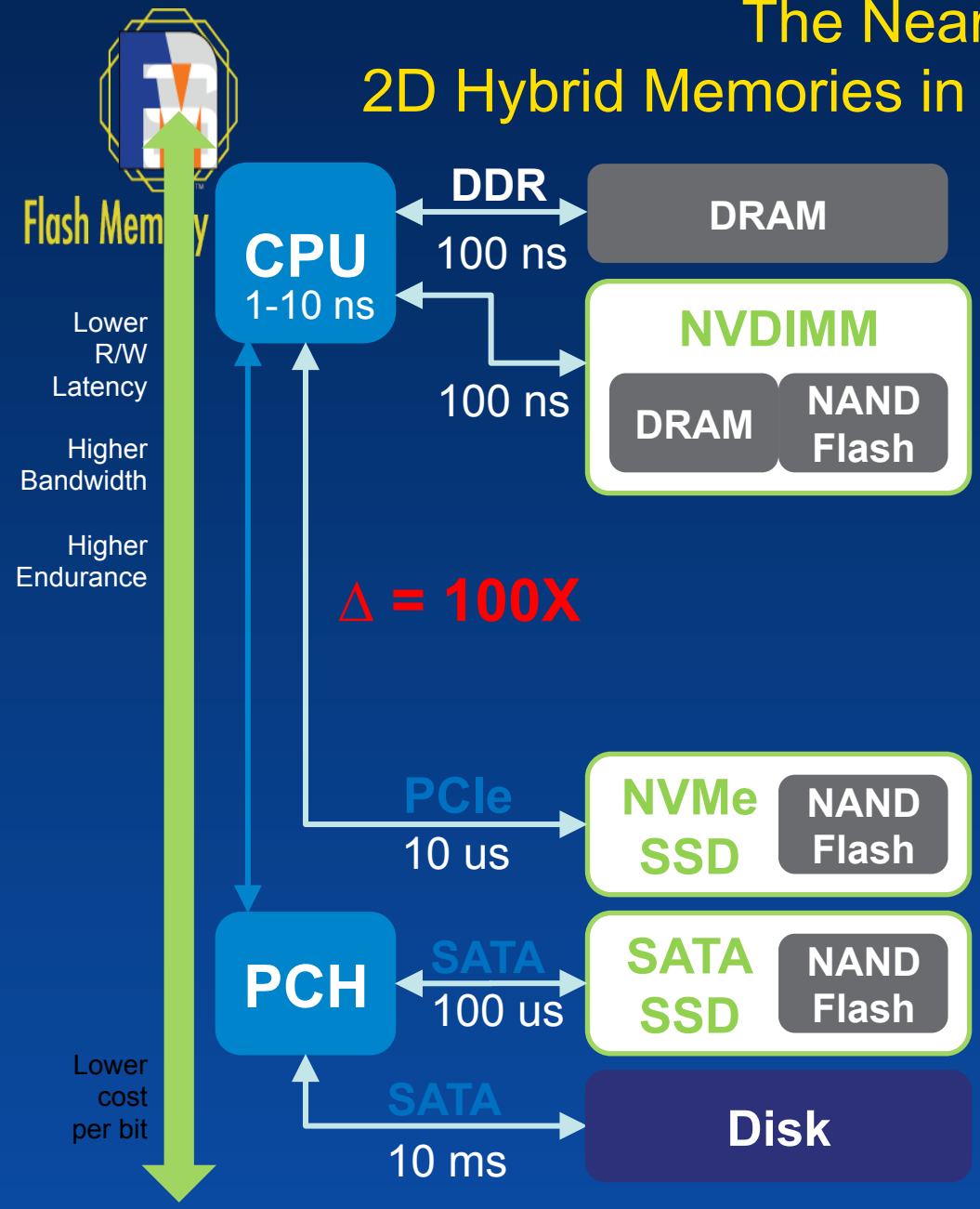


The Past: Nonvolatile Memories in Server Architectures



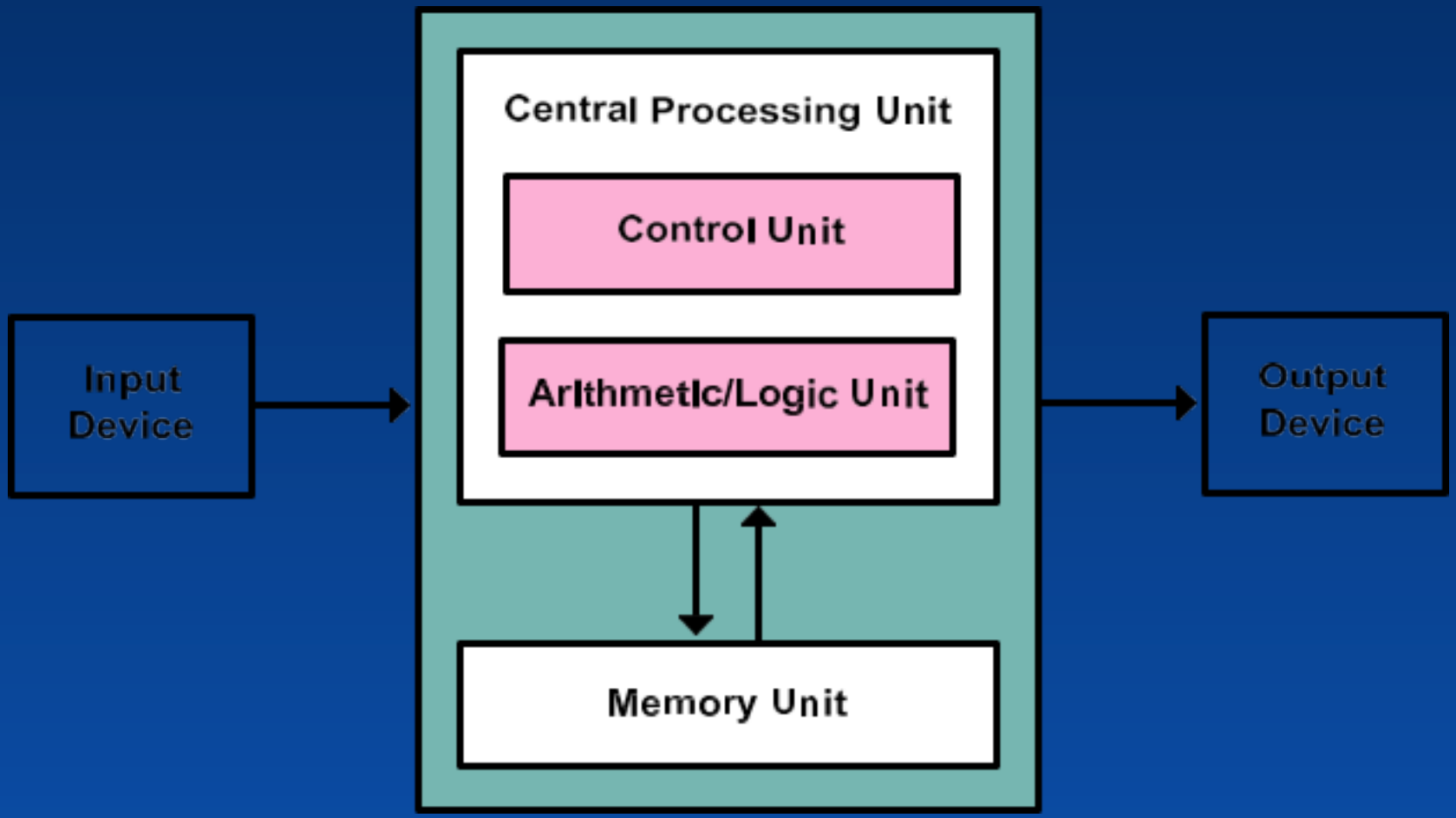
- For decades we've had two primary types of memories in computers: DRAM and Hard Disk Drive (HDD)
- DRAM was fast and volatile and HDDs were slower, but nonvolatile (aka persistent)
- Data moves from the HDD to DRAM over a bus where it is the fed to the processor
- The processor writes the result in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

The Near Past:
2D Hybrid Memories in Server Architectures

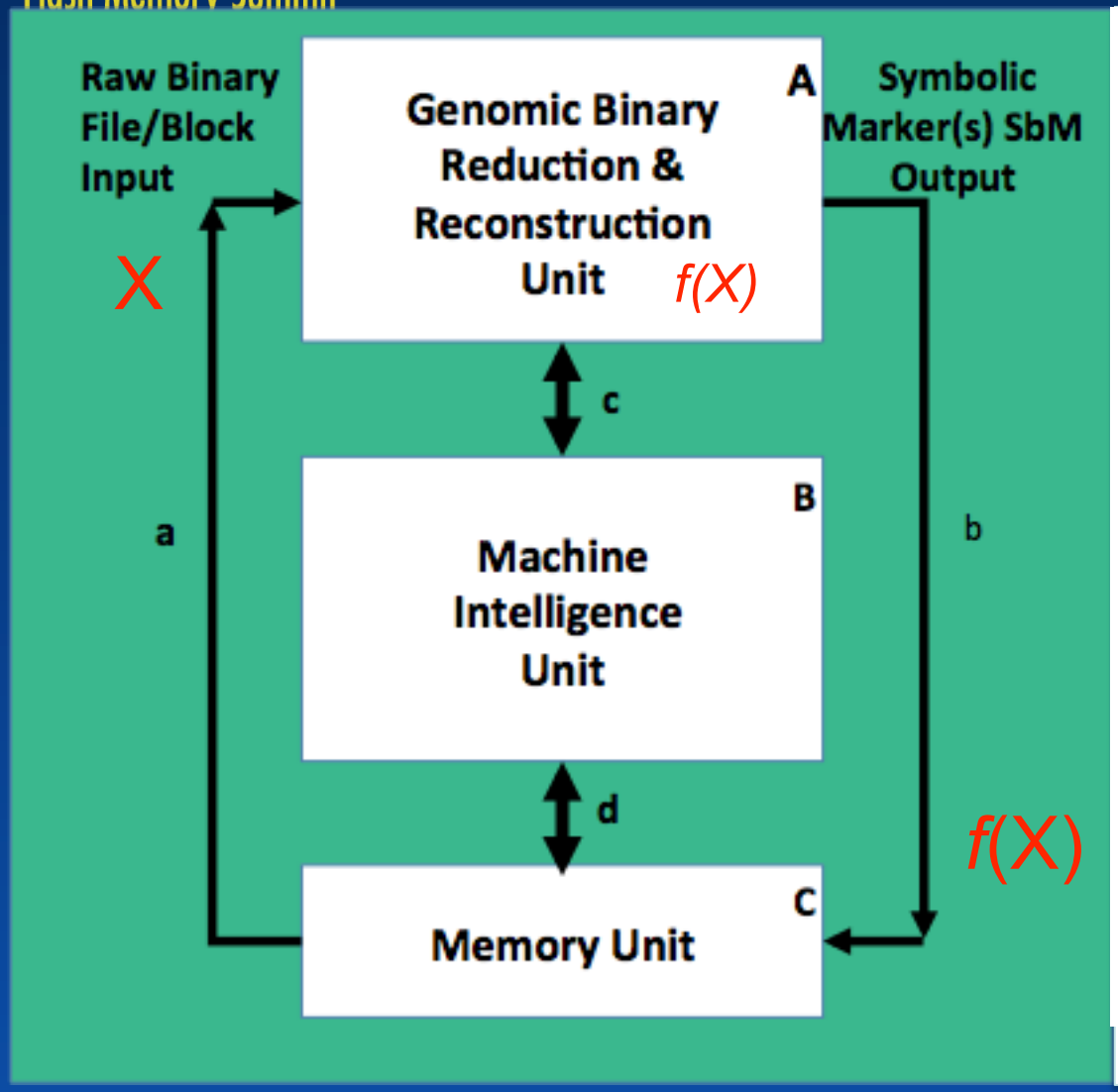


- System performance increased as the speed of both the interface and the memory accesses improved
- NAND Flash considerably improved the nonvolatile response time
- SATA and PCIe made further optimization to the storage interface
- NVDIMM provides super-capacitor-backed DRAM, operating at DRAM speeds and retains data when power is removed (-N, -P)

The Classic Von Neumann Machine



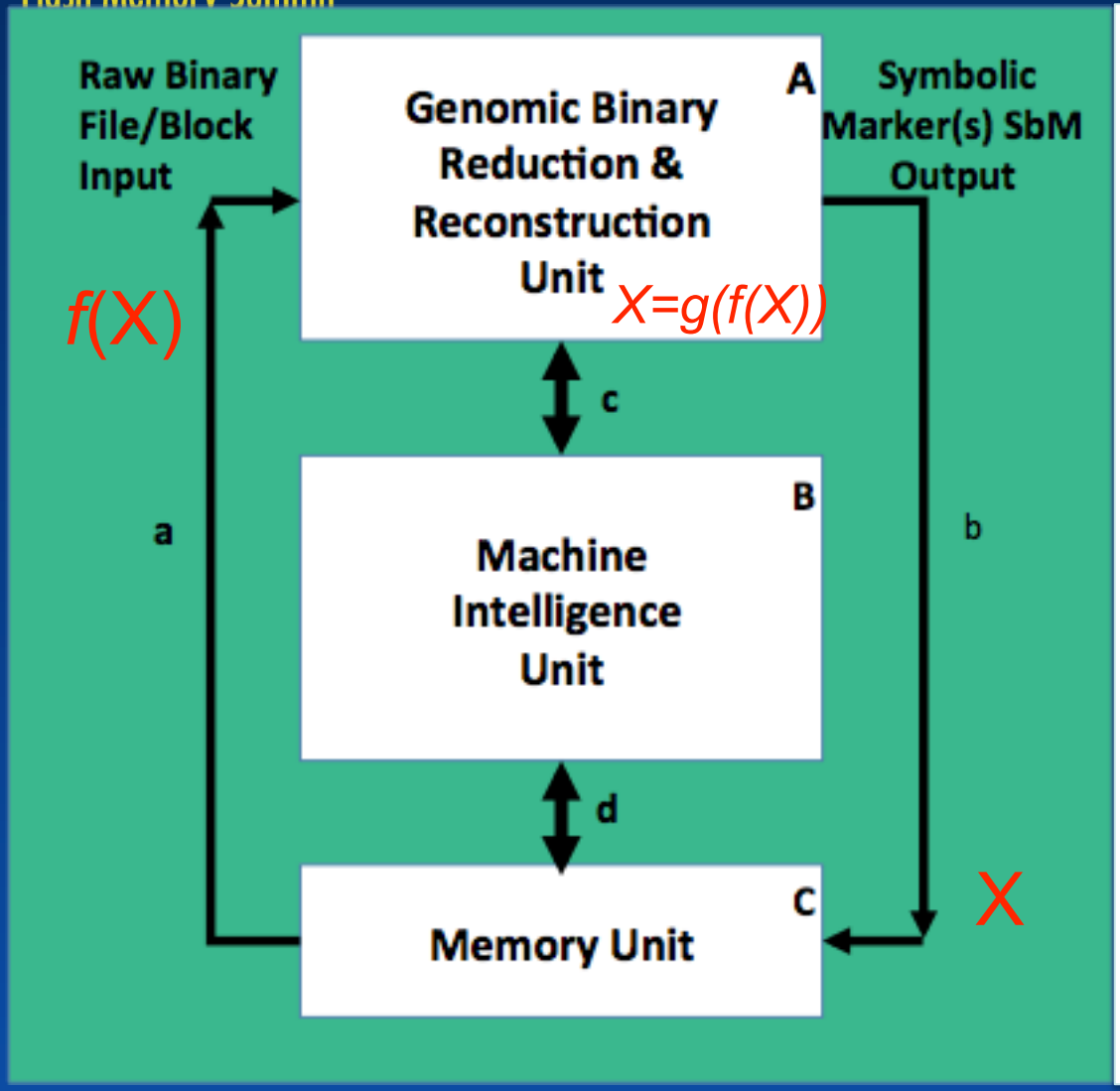
Symbolic IO IRIS i1 Memory Amplification – SymCE™ Data Deconstruct



1. Data Input (line a) into Genomic Binary Reduction/Reconstruct Unit (Box A) from Memory Unit (Box C)
2. GBR DE-constructs raw binary with assists for further deconstruction continuously with Machine Intelligence Unit (Box B) ('On the Fly' from initial data to all new/ modified data.
3. GBR outputs Symbolic Markers (SbM's) to Memory Unit (Box C) incrementally as most binary is similar.
4. File/Block data is now available for usage by Application, Network transfer, Visual Display, or to external Storage(line e)

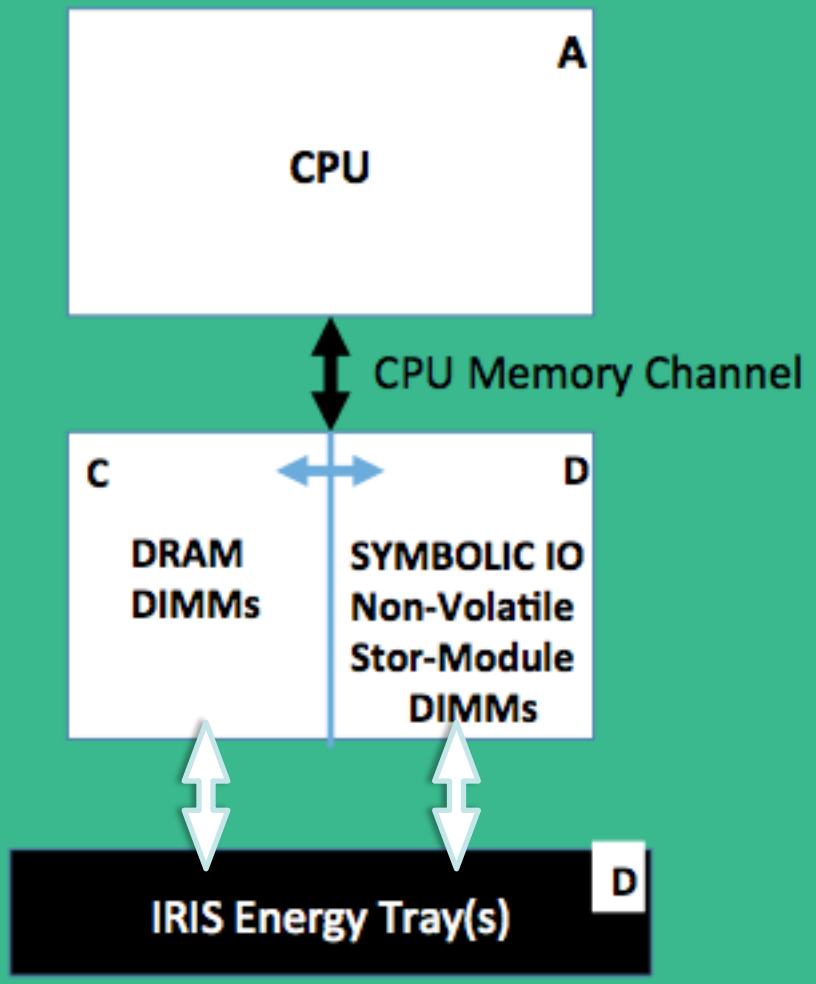
Symbolic IO IRIS i1 Memory Amplification – SymCE™ Data Reconstruct

Flash Memory Summit



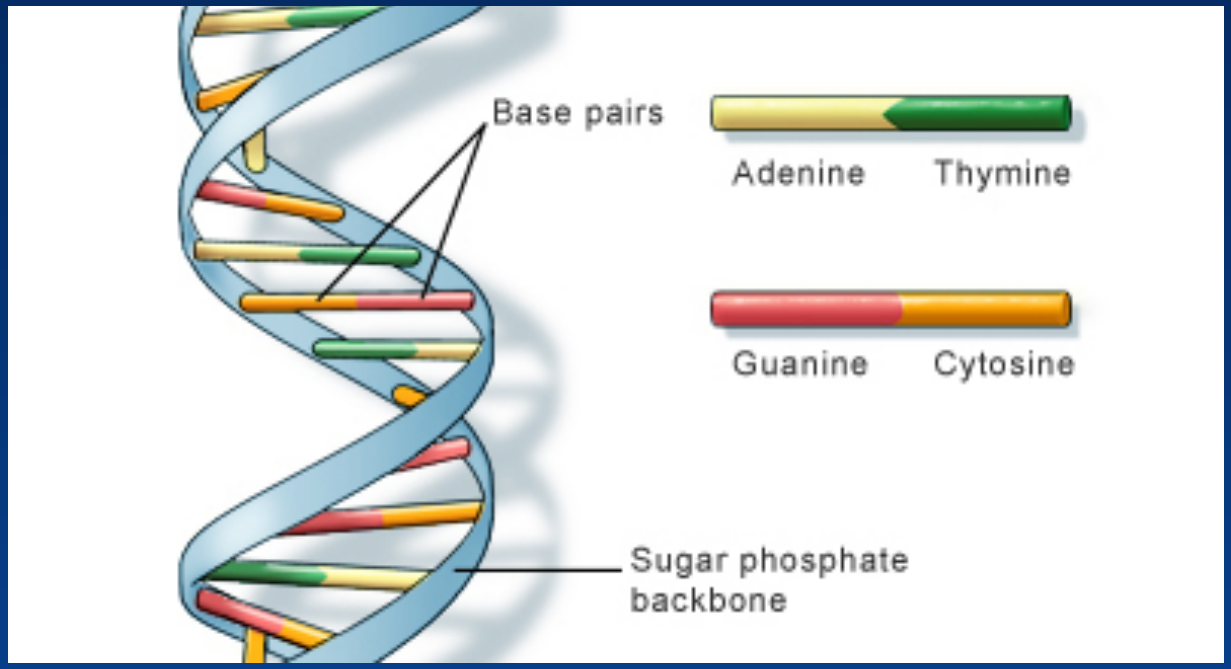
1. Data Input (line a) into Genomic Binary Reduction/Reconstruct Unit (Box A) from Memory Unit (Box C)
2. GBR DE-constructs raw binary with assists for further deconstruction continuously with Machine Intelligence Unit (Box B) ('On the Fly' from initial data to all new/ modified data.
3. GBR outputs Symbolic Markers (SbM's) to Memory Unit (Box C) incrementally as most binary is similar.
4. File/Block data is now available for usage by Application, Network transfer, Visual Display, or to external Storage(line e)

Symbolic IO IRIS i1 SymCE™ - Energy Mgmt.



- Symbolic IO fuses State of the Art Storage Array Technology by utilizing Super Capacitors, called IRIS Energy Tray(s) (Box D) to power the Symbolic IO Server when power is lost or coming back up.
- On all memory channels attached to CPUs within any Symbolic IO Server, DRAM is inserted along with Patented Non-Volatile StorModules that are used to De-Stage RAM when power is lost, and then to Stage (restore) RAM contents when power is again restored to the Server.

What is this?



- DNA is, effectively, a binary encoding (AT=0, CG=1)
- Humans have roughly 3 billion base pairs (genome)
- Very small form factor (human genome = a thimble-full of material)
- DNA can replicate itself
- Genes consist of DNA (few hundred bases->few million)
- Most genes are identical between people (~25,000 genes per person)
- Order matters – genes are markers, markers interact based on order
- Very small space needed – results in large organism when expressed

The Present: Amplified PM in Server Architectures

Lower R/W Latency

Higher Bandwidth

Higher Endurance

Lower cost per bit

CPU 1-10 ns

DDR 100 ns

DRAM +

NVDIMM

DRAM | **NAND Flash**

Amp Capacity

O(10 TB)

DDR 500 ns *

PCIe 5 us *

3D PM

O(100 TB)

Δ = 2-20X

PCIe 10 us

NVMe SSD | **NAND Flash**

O(1 PB)

PCH

SATA 100 us

SATA SSD | **NAND Flash**

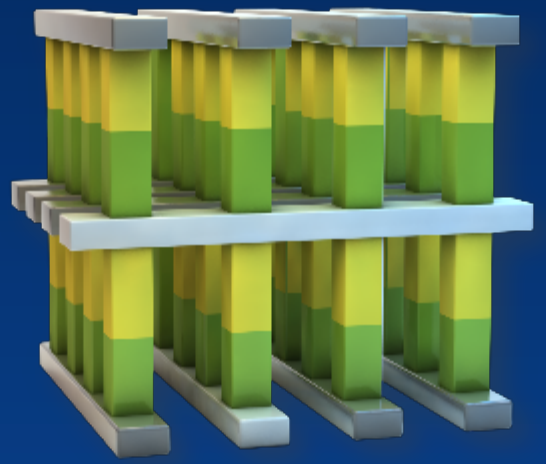
O(zero)

SATA 10 ms

Disk

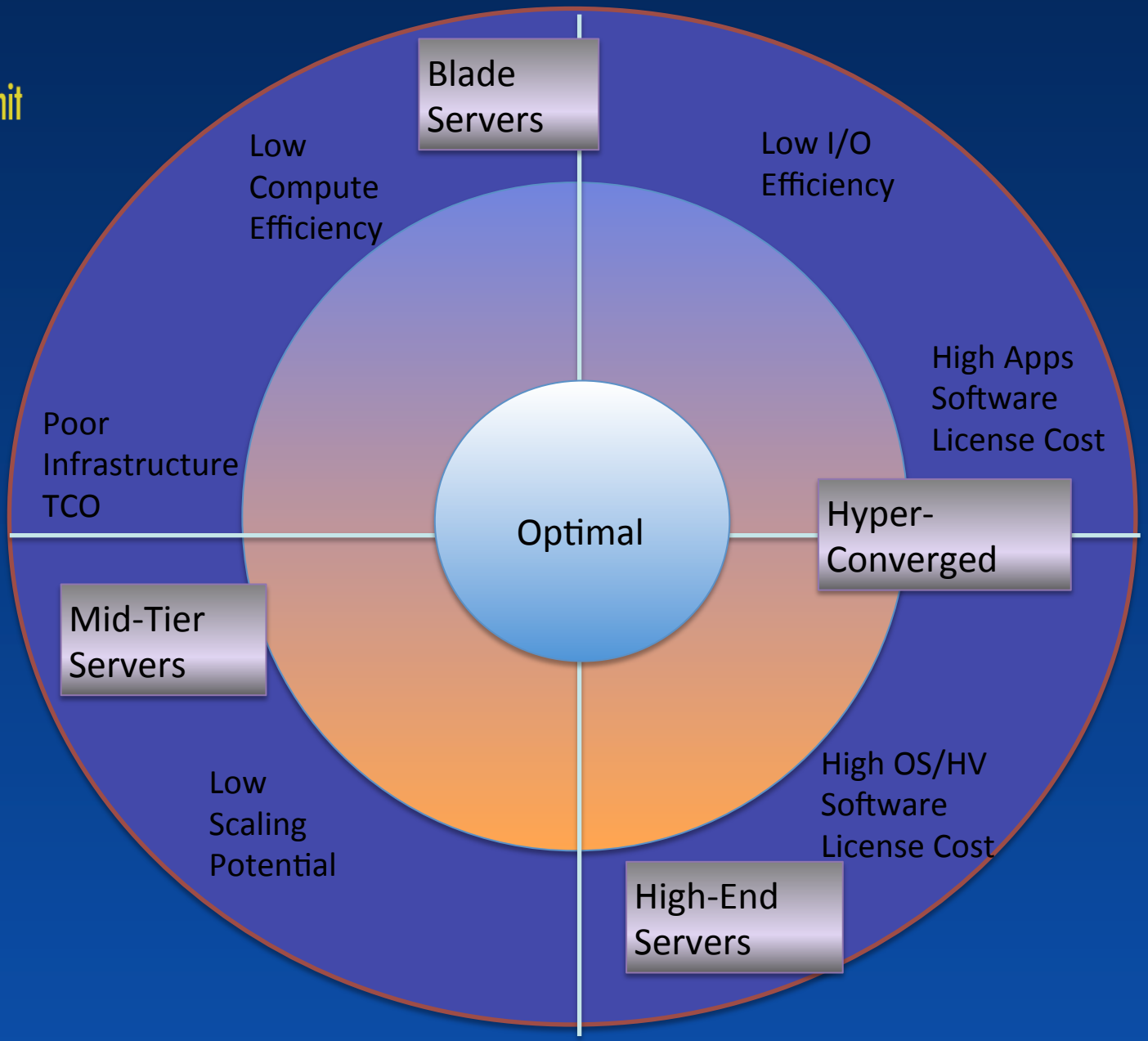
O(zero)

* estimated

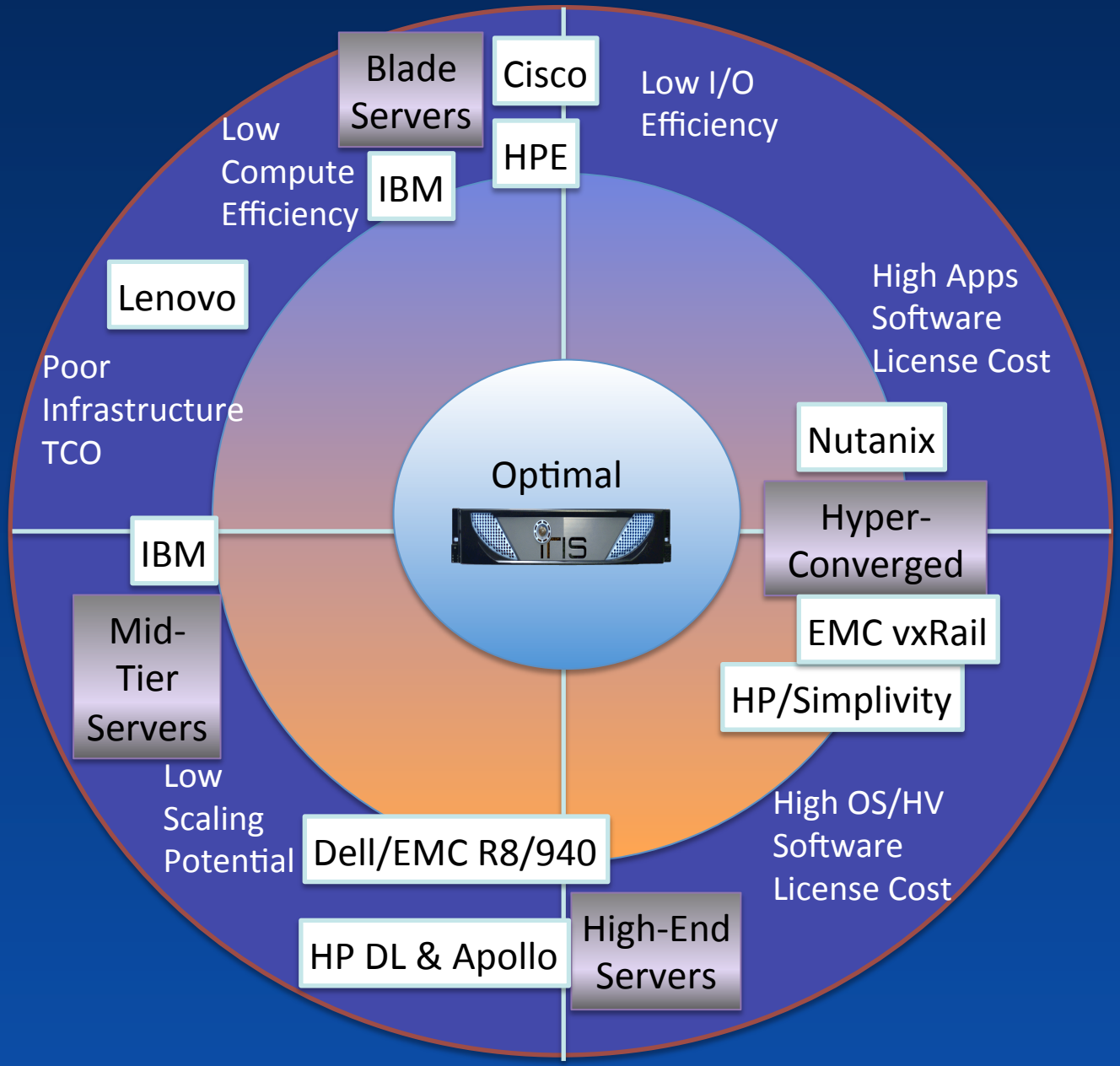


- PM technology provides the benefit “in the middle”
- It is considerably lower latency than NAND Flash
- Performance can be realized on PCIe or DDR buses
- Lower cost per bit than DRAM while being considerably more dense

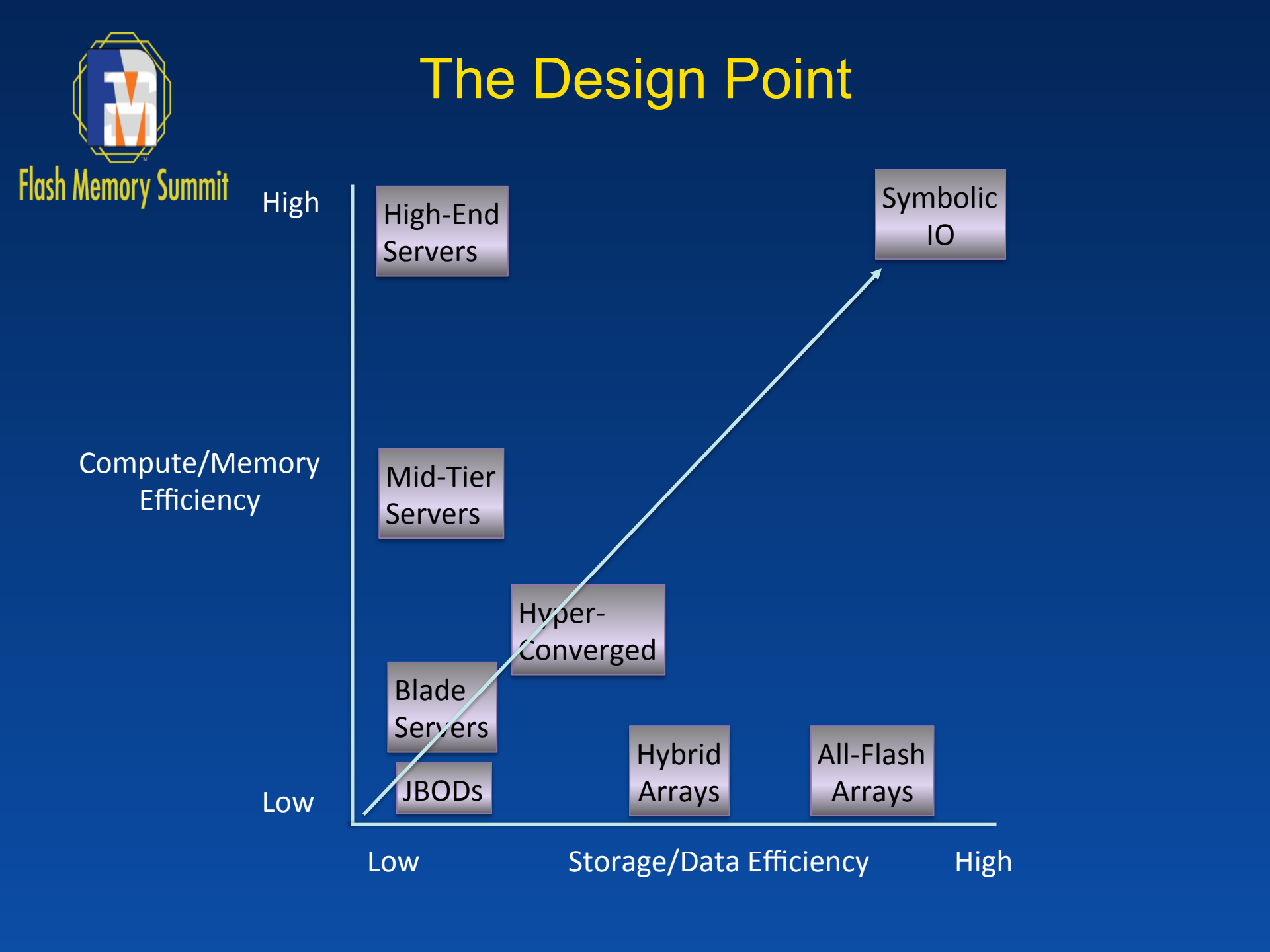
The Server Problem



The Server Problem




ELECTRIC LIGHT DID NOT COME FROM THE CONTINUOUS
IMPROVEMENT OF CANDLES


Flash Memory Summit
The Design Point
High
Compute/Memory Efficiency
Low
High-End Servers
Symbolic IO
Mid-Tier Servers
Hyper-Converged
Blade Servers
JBODs
Hybrid Arrays
All-Flash Arrays
Low
Storage/Data Efficiency
High

The Design Point of Symbolic IO

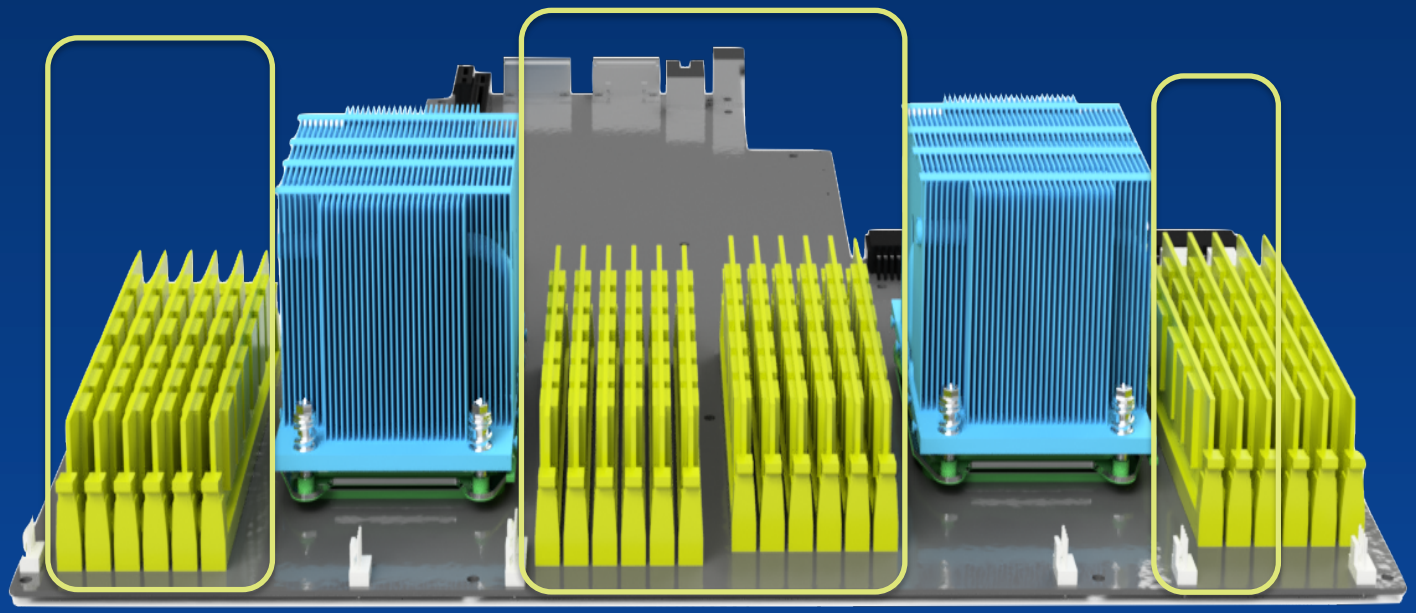
- All systems are not only the sum of their parts, they are the **blend of their interactions**
  - Compute and memory are tied together - literally (CPUs and RAM)
  - Storage and (persisted) data, similarly (channels/busses and media)
  - The design point for most systems is to optimize one or the other
    - via incremental improvement of components - faster, denser, cheaper
  - Until now, impossible to 'fuse' the two design points together
    - 'Hyperconverged' not a fusion – it is merely components in one box
  - Symbolic IO took a different approach…

- Symbolic IO goes beyond incremental improvement on one axis only
  - The use of persistence in the memory channel (StorModules) fuses the two
  - The software enables the hardware to operate as one entity
  - This reaches the most efficient design point possible

Storage becomes compute, compute becomes storage

StorModules™ - Primary Storage

Symbolic IO StorModule™
Performance Tier 1 & Primary Storage
DDR4 Speeds @ 68GB/s for IRIS-i1



- Persistence in the memory channel
- DDR-4 based DIMMs

- IRIS-i1 can use up to 21StorModules™
- 3D Xpoint™ DIMMs - tier 1.5
- (not available yet from Intel - Apache Pass)

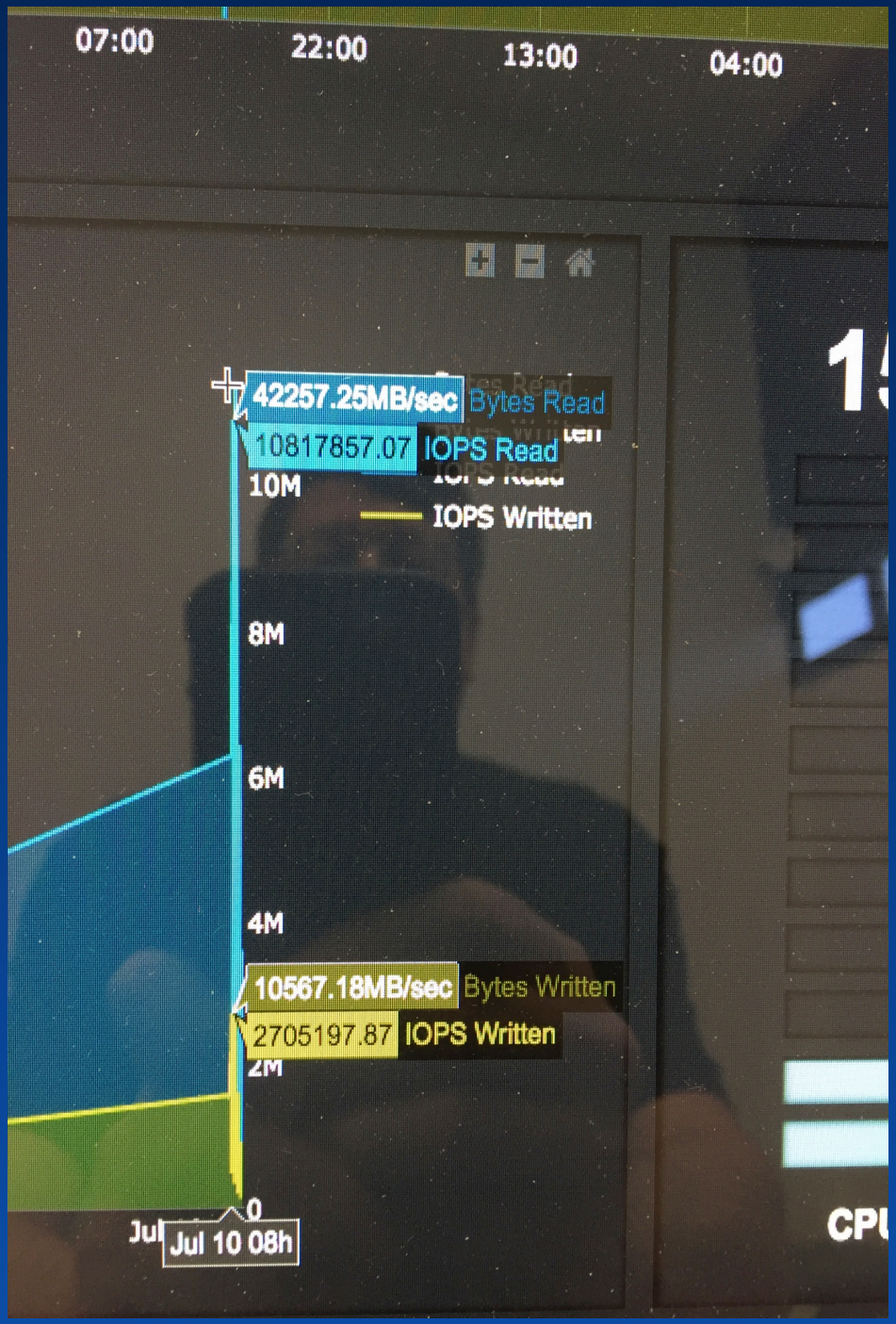
SymCE™ OS

- **Translates application I/O (read,write) into real-time computation**
  - no code change required - unlike other persistent memory approaches (e.g. pmem.io)
- **Infused hypervisor supporting persistent memory; fully orchestrated VDI**
  - 100's of VMs/virtual desktops per IRIS i1 – storage at memory speed
  - Simply migrate VMs from other platforms into IRIS i1 with SymCE tools

- State Independent Snapshot (SIS™) and Clone (SIC™) technology
- Patent Pending BLINK™ software
  - Allows for a complete infrastructure to be preserved, restored, or scaled up in minutes
  - Optional removable BLINK card (persistent media)
- BLINK™ contains machines, applications, security settings, configurations and data - the complete machine state
- BLINKs are completely application-consistent
- BLINKs are 100% secure and only contain Symbolic bit Markers and proprietary metadata
- Clone an IRIS via BLINK in minutes
- Selective & partial BLINKs - granular
  - Use with infused hypervisor - selectively blink VMs, volumes within VMs

POC #1 – Data Analytics Company

Challenges	**Symbolic Solution**	**Symbolic Advantages**
• High cost to maintain on-premises infrastructure • Limited virtualization capability on-premises (25 VMs per server) • High OPEX from significant use of AWS for compute	• Implement IRIS i1 to replace current HP on-premises infrastructure • Bring back all AWS workload to on-premises on IRIS i1 • Leverage SymCE OS for storage and virtualization	• Replace HP servers at 21:1 ratio (21 HP blades to one IRIS) • Run ~400 VMs per IRIS instead of 25 per HP • Save $450,000 for on-premises • Save $1,000,000 for every 400 VMs over 5 years by re-homing AWS


07:00
22:00
13:00
04:00
42257.25MB/sec Bytes Read
10817857.07 IOPS Read
10M
IOPS Written
8M
6M
4M
10567.18MB/sec Bytes Written
2705197.87 IOPS Written
2M
0
Jul 10 08h

New Computational Approaches to Big Data

The Impact of Persistent Memory and Intelligent Data Encoding

Rob Peglar
SVP/CTO
Symbolic IO
rpeglar@symbolicio.com
@peglarr