

September 23-26, 2019 Santa Clara, CA

The NVRAM Standard, Bringing Coherence to the Crazy World of Persistent Memory

Bill Gervasi Principal Systems Architect Nantero, Inc.





Data processing is great

MANTERO

3

SD©

Data processing is great

Until something goes wrong





5

The Cost of Power Failure

According to Gartner, the average cost of IT downtime is **\$5,600** per minute. Because there are so many differences in how businesses operate, downtime, at the low end, can be as much as \$140,000 per hour, **\$300,000** per hour on average, and as much as \$540,000 per hour at the higher end. Jun 18, 2018

The 20 | The Cost of IT Downtime | The 20 https://www.the20.com/blog/the-cost-of-it-downtime/

Amazon.com Goes Down, Loses \$66,240 Per Minute





Checkpoint

🛗 November 12, 2015 🙎 Alexandr Omelchenko 🛛 🗲 Glossary

★★★★★ [Total: 21 Average: 4.2/5]

Checkpoint is a process that writes current in-memory dirty pages (modified pages) and transaction log records to physical disk. In SQL Server checkpoints are used to reduce the time required for recovery in the event of system failure. Checkpoint is regularly issued for each database. The following set of operations starts when checkpoint occurs:

- 1. Log records from log buffer (including the last log record) are written to the disk.
- 2. All dirty data file pages (pages that have been modified since the last checkpoint or since they were read from disk) are written into the data file from the buffer cache.
- 3. Checkpoint $\ensuremath{\mathsf{LSN}}$ is recorded in the database boot page.





oint

2019 Storage Developer Conference. © Nantero Inc. All Rights Reserved.

SD (®





8





9

System failure is a key factor in server software design

Data persistence is essential

Storage access time impacts transaction granularity









To reduce the penalties from checkpointing...

...move non-volatile storage closer to the CPU



SD[®]





13

September 23-26, 2019 Santa **The Search for**

THE HOLY GRAIL



SD[©]



When we no longer the 23 fear power failure...



DATA PERSISTENCE







When was the last time you read about a new volatile memory?

NRAM PCM

The non-volatile memory revolution is under way

2019 Storage Developer Conference. © Nantero Inc. All Rights Reserved.



17

ReRAN

SD[®]



From vacuum tubes

2019 Storage Developer Conference. © Nantero Inc. All Rights Reserved.



SD[®]

THIS is why the term "Persistent Memory" is insufficient

The industry must distinguish between deterministic and non-deterministic persistent memory

Only "Memory Class Storage" is fully deterministic AND persistent





Not all "persistence" is created equal





Flash Architecture	Layers of Cells	Bits per Cell	Number of Cell Voltage States	Cell Endurance ¹ (P/E Cycles)		
Planar SLC	1	1	2	~100,000		
Planar MLC	1	2	4	~3,000		
Planar eMLC/iMLC/pSLC	1	1	2	~20,000		
Planar TLC	1	3	8	<1,000		
Vertical SLC	Varies, 64 typical	1	2	TBD ²		
Vertical MLC	Varies, 64 typical	2	4	TBD ²		

	375GB Intel DC P4800X	1.6TB Intel DC P3700	1.6TB Intel DC P3608	2.4TB Micron 9100 Max	2.7TB Mangstor MX6300	
Endurance Per Usable GB	32.8 TB	27.35 TB	5.45 TB	2.73 TB	12.77 TB	
Usable Capacity	375GB	1.6TB	1.6TB	2.4TB	2.7TB 4TB	
Raw Capacity	448GB	2TB	2.3TB	4TB		
Spare Area / %	73GB / 16.3%	400GB / 20%	700GB / 30.4%	1600GB / 40%	1300GB / 32.5%	
Media Endurance Per Raw GB	27.4TB	21.9TB	3.8TB	1.6TB	8.63TB	

"Write endurance" determines HOW persistent

Wear leveling needed if writes are limited

2019 Storage Developer Conference. © Nantero Inc. All Rights Reserved.



SD (®

Temperature sensitivity impacts long term retention

								Chern										
Application	Workload	Active Use	Retention	ention Functional UBER Jse Failure ver off) Rqmt (FFR)	UBER		55	1	1	2	2	3	5	8				
Class		(power on)	(nower off)		Failure Romt (FER)	Panure Romt (EER)	Ramt (FFR)	Ramt (FFR)	E FR)	<u> </u>	50	2	2	3	4	6	9	15
		,	(power on)			Power Off emperatu	45	4	4	5	7	10	17	27				
Client	Client	40°C 30°C 8 hrs/day 1 year	30°C	≤3%	≤10-15		40	7	8	10	14	20	31	52				
			1 year				35	14	16	20	26	38	61	101				
					- 6	30	28	32	39	52	76	120	199					
Forte and deal	Enternation	E FOC	1000	<20/	<10-16		25	58	65	79	105	155	244	404				
Enterprise	Enterprise	24hrs/dav	3 months	≤ 3 %	≤ 10 ⁻¹⁰			25	30	35	40	45	50	55				
			0					Active temp										
								Weeks of Data Retention										



Client

SD[®]





SD 🕑

Memory Class Storage

Full DRAM Speed

No endurance limits

Fully deterministic





is a

Memory Class Storage

25

SD®





26

Storage Class Memory ls **NOT** a **Memory Class Storage**







28



DRAM speed amber 23-26, 2019 State Class, CA Non-volatility

Unlimited write endurance

Wide temperature range

Scalable beyond DRAM

Flexible fabrication

Low power

Low cost



SD[®]

Low





Drop in replacement for DRAM NVRAM Memory Class Storage

Fully Deterministic

Permanently persistent

Always available

Host System













Comparing DRAM & NVRAM

No refresh is required

"Self refresh" can be power OFF

Some timing differences (but deterministic!)

Data persistence definitions

Greater per-die capacity



SD®



SD @

DDR5 NVRAM Specification brings coherence

NANTERO

34





















SD[®]





DDR5 DRAM is limited to 32Gb per die

September 23-26, 2019 Santa Clara, CA

DDR5 NVRAM enables up to 128Tb per die

SD©







Row Extension adds up to 12 more bits of addressing

Backward compatible with DDR5 – Acts like REXT = 0 until needed



SD (®



"ROW" includes bank group & bank...





43

SD[®]









45



NRAM: Carbon Nanotube Cell Memory Crosspoint tiles translated to DDRx





September 23-26, 2019 Santa Clera, CA

SD @

Full DRAM speed: DDR4, DDR5 Non-volatility: 12,000 year data retention **Unlimited write endurance** Wide temperature range: -55 °C to +300 °C tested Beyond DRAM: 512Gb/die in DDR5 window Flexible fabrication: Logic or memory Low power: 15% lower power than DRAM

Low cost: cheaper to build than DRAM



47

SD @

mber 23-26, 2019 Clara, CA

Or Revolution?







48

It's About the Software



2019 Storage Developer Conference. © Nantero Inc. All Rights Reserved.

49

ANTERO

SD @





50









DATA PERSISTENCE

Registers Cache Memory $\sqrt{}$ Storage Class Memory Class √NVMe

√ Tape, \sqrt{SSD} $\sqrt{\text{Hard Drive}}$



SD[®]

2019 Storage Developer Conference. © Nantero Inc. All Rights Reserved.

Storage





54

SD[®]





Power failure drives systems architectures

Checkpointing is a costly way to deal with system failure

SD©

Memory tiers balance safety and performance

Summary

Applications evolving to exploit MCS

DDR5 NVRAM standard brings coherence Persistence is moving closer to the CPU



September 23-26, 2019 Santa Clara, CA



Thank you for your time

Bill Gervasi bilge@nantero.com

