



Developing a Server OS Drive for Big Data and Cloud Storage

C.C. Wu

EVP of Innodisk USA Corporation

The logo for Flash Memory Summit features a yellow sunburst icon above the text "Flash Memory" in black and "SUMMIT" in white on a blue rectangular background.

Flash Memory Summit Agenda

- Boot up storage background
- Various mechanical designs
- Reliability
- Product Lifespan
- Performance



Background

- VMware ESXi, Microsoft Hyper-V, CentOS - popular Server Operating Systems
- ESXi and Hyper-V require a minimum of 64GB for boot up storage
- CentOS requires 8GB for boot up storage
- Available form factors:
 - 2.5"
 - USB Drive
 - SD/microSD
 - SATADOM™
 - mSATA/Slim

The logo for Flash Memory Summit 2016, featuring a yellow sunburst icon above the text "Flash Memory" and "SUMMIT" in a blue box.

Flash Memory Summit Challenge

- Small footprint form factor
- PLP design
- Lifespan issues
- 24/7 server uptime
- Kernel panics with bit errors
- Continuous log files
- Reliable Performance



Boot Up Storage Considerations

- 2.5" SSD occupy precious drive bay slots
- USB drives are consumer focused and lower quality
- SD Cards are similar to USB drives
- SATADOM™ combines all these benefits together

Form factor comparison



2.5" SSD

SATA



MO-297

SATA



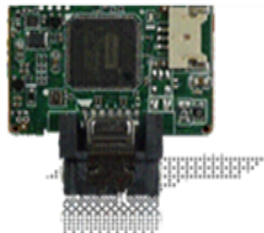
mSATA

SATA



M.2 2280

SATA, NVMe



SATADOM

SATA



USB Drive

USB



SD card

SDIO



Interface / Performance

	SD UHS-II	USB3.1	SATA III	PCIe Gen3(NVMe)
Bus Speed	312MB/s	1.21GB/s	750MB/s	1GB/s Gen3 x1 2GB/s Gen3 x2 4GB/s Gen3 x4 M.2
Number of Queues			1	65536
Commands per Queue			32	65536
Application	Camera Storage	Portable Drive	Operation System Drive	Operation System Drive

Flash Memory Summit 2016
Santa Clara, CA



SATA Flash Storage - Mainstream

- Slim/mSATA/SATADOM™
- Internal Server boot up storage
- Advantages
 - Performance
 - Stable protocol
 - Design for OS
 - No External DRAM buffer

CrystalDiskMark 5.1.2 x64

檔案(F) 設定(S) 佈景主題(T) 說明(H) 語言(Language)

All 5 1GiB D: 0% (0/238GiB)

	Read [MB/s]	Write [MB/s]
Seq Q32T1	534.3	215.6
4K Q32T1	127.8	118.4
Seq	432.9	213.9
4K	20.50	61.94



Reliability

- Operation System cannot accept any single bit corruption... Cause kernel panic
- PLP design and external DRAM issue
- PE Cycle for MLC, iSLC, and TLC?
- Nonstop server



Product Lifespan

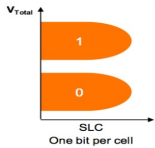
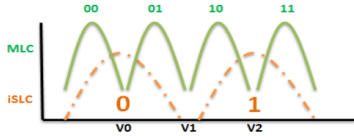
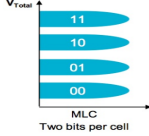
- iSMART tool
- LBA Write: TBW
- Flash Erase Count
 - MLC: 3,000.
 - iSLC:20,000
 - SLC:100,000
- iAnalyzer: Analyses OS's loading

The screenshot shows the iSMART software interface. The top navigation bar includes 'Device Information', 'iAnalyzer', 'Settings', 'S.M.A.R.T.', 'Alert', and 'System Information'. The 'S.M.A.R.T.' tab is active, displaying a table of attributes for an Innodisk SATA DOM device. The table has four columns: ID, Attribute Name, Item Value, and Raw Value. The data is as follows:

ID	Attribute Name	Item Value	Raw Value
A7	Avg Erase Count	0	020000010000000000000000
A9	Device Life	100	000064006400000000000000
AA	Spare Block Count	172	03006401AC00000000000000
AB	Program Fail Count	0	020000010000000000000000
AC	Erase Fail Count	0	020000010000000000000000
C0	Unexpected Power Loss Count	0	000000000000000000000000
C2	Temperature	30	00001E641E00000064021E
E5	Flash ID	0	02006401983C95937AD101
EB	Later Bad Block Read	0	020000000000000000000000
EB	Later Bad Block Write	0	020000000000000000000000
EB	Later Bad Block Erase	0	020000000000000000000000
F1	Total LBAs Written	143	020064018F00000000000000
F2	Total LBAs Read	65	020064014100000000000000

The interface also shows 'Model: Innodisk SATA DOM' and 'Device: 1 2'. The footer contains the Innodisk logo, copyright information for 2015, and the version number 'Ver 5.1.6(T)'.

Flash PE Cycle

Item	SLC Single Level Cell	iSLC SLC mode	MLC Multi Level Cell
Architecture	<p>SLC Flash has only two states: erased (empty) or programmed (full).</p> 	<p>Enhance iSLC, algorithm & Enhance ECC</p> 	<p>MLC Flash has four states: erased (empty), 1/3, 2/3, and programmed (full).</p> 
Performance	Best	Faster	Slower
Power Consumption	Lowest	Lower	Higher
Endurance (P/E Cycles)	100K/60K	20K	3K
Initial Data Retention	10 Years	10 Years	10 Years
Density	MLC > iSLC > SLC		
Cost	★★★★	★★★	★★
Application	<ol style="list-style-type: none"> 1. IPC, embedded, automation 2. Mission-critical applications 	<ol style="list-style-type: none"> 1. IPC/Kiosk/POS system 2. Embedded System 3. Server MB 4. Write intended application. 	<ol style="list-style-type: none"> 1. POS, Kiosk system 2. Commercial application



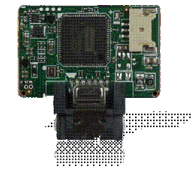
WAI issue

- Virtual Operation System has more loading with 4K random write.
- How to implement a FTL with low WA of 4K random write Without DRAM buffer
- Increase SRAM buffer
- 4K Mapping Algorithm
- Small Management Unit to reduce WAI



SMART Tool

- Install Microsoft Hyper-V Server 2012 R2
- 4.5 GB Write (143*32MB)



88% Random Write

ID	Attribute Name	Item Value	Raw Value
A7	Avg Erase Count	0	020000010000000000000000
A9	Device Life	100	000064006400000000000000
AA	Spare Block Count	172	03006401AC00000000000000
AB	Program Fail Count	0	020000010000000000000000
AC	Erase Fail Count	0	020000010000000000000000
C0	Unexpected Power Loss Count	0	000000000000000000000000
C2	Temperature	30	00001E641E00000064021E
E5	Flash ID	0	02006401983C95937AD101
EB	Later Bad Block Read	0	020000000000000000000000
EB	Later Bad Block Write	0	020000000000000000000000
EB	Later Bad Block Erase	0	020000000000000000000000
F1	Total LBAs Written	143	020064018F00000000000000
F2	Total LBAs Read	65	020064014100000000000000

ismart iAnalyzer

Navigation: Device Information | **iAnalyzer** | Settings | S.M.A.R.T. | Alert | System Information

Read Statistics:

- Read: (7%) Sequential, (93%) Random
- Sequential Read: (2%) 8MB, (5%) 4MB, (25%) 1MB, (11%) 128KB, (22%) 64KB, (35%) 32KB
- Random Read: (4%) 64KB, (31%) 32KB, (18%) 16KB, (9%) 8KB, (38%) 4KB

Write Statistics:

- Write: (12%) Sequential, (88%) Random
- Sequential Write: (2%) 8MB, (5%) 4MB, (84%) 1MB, (2%) 128KB, (2%) 64KB, (5%) 32KB
- Random Write: (17%) 64KB, (5%) 32KB, (7%) 16KB, (9%) 8KB, (62%) 4KB

innodisk Copyright© 2015 Innodisk Corporation. All Rights Reserved. Ver 5.1.6(T)

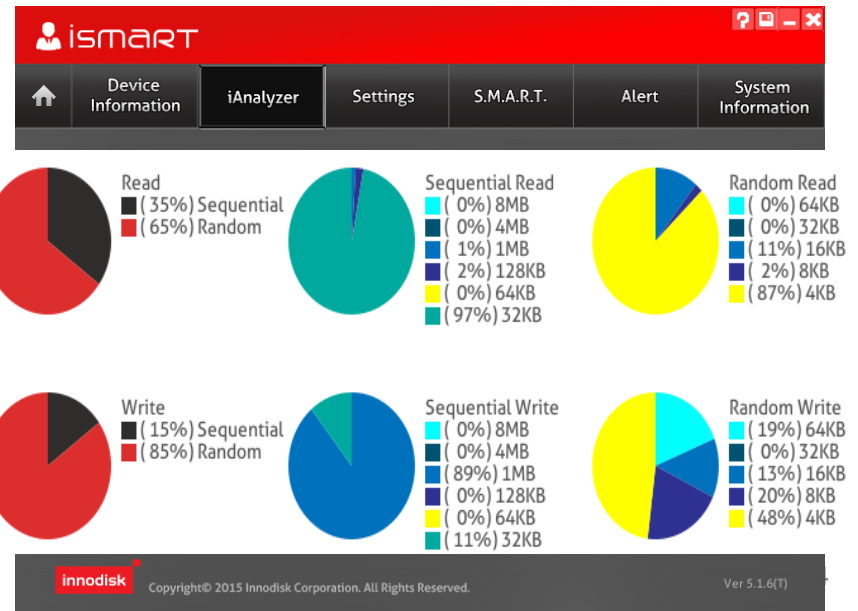


SMART Tool

- Install VM ESXi 6.0.0.0
- 512MB Write

85% Random Write

ID	Attribute Name	Item Value	Raw Value
A7	Avg Erase Count	0	020000010000000000000000
A9	Device Life	100	000064006400000000000000
AA	Spare Block Count	207	03006401CF00000000000000
AB	Program Fail Count	0	020000010000000000000000
AC	Erase Fail Count	0	020000010000000000000000
C0	Unexpected Power Loss Count	0	000000000000000000000000
C2	Temperature	30	00001E641E000000064021E
E5	Flash ID	0	02006401983C95937AD101
EB	Later Bad Block Read	0	020000000000000000000000
EB	Later Bad Block Write	0	020000000000000000000000
EB	Later Bad Block Erase	0	020000000000000000000000
F1	Total LBAs Written	16	020064011000000000000000
F2	Total LBAs Read	2	020064010200000000000000





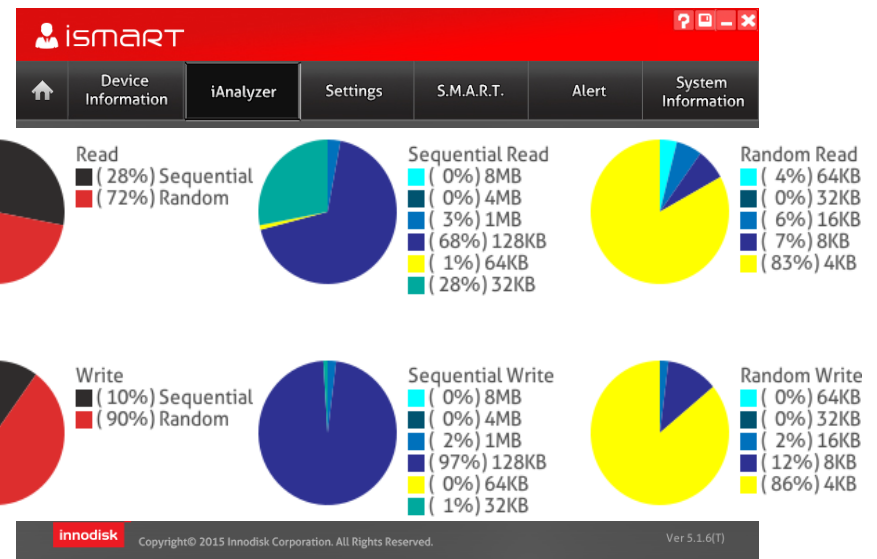
SMART Tool

- Running over 12 Hours
- 544MB Write

90% Random Write

ID	Attribute Name	Item Value	Raw Value
A7	Avg Erase Count	0	020000010000000000000000
A9	Device Life	100	000064006400000000000000
AA	Spare Block Count	207	03006401CF00000000000000
AB	Program Fail Count	0	020000010000000000000000
AC	Erase Fail Count	0	020000010000000000000000
C0	Unexpected Power Loss Count	0	000000000000000000000000
C2	Temperature	30	00001E641E00000064021E
E5	Flash ID	0	02006401983C95937AD101
EB	Later Bad Block Read	0	020000000000000000000000
EB	Later Bad Block Write	0	020000000000000000000000
EB	Later Bad Block Erase	0	020000000000000000000000
F1	Total LBAs Written	17	020064011100000000000000
F2	Total LBAs Read	76	020064014C00000000000000

Flash
Santa Clara, CA





NVMe Driver

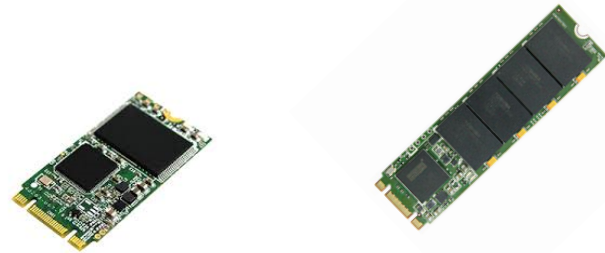


- Microsoft Hyper-V:
2008 R2 /2012/2012 R2 (Driver support)
- CentOS 7.0/ Kernel 3.3 :
Support NVMe
- VMware:
ESXi 5.5/6.0 (Driver support)



M.2 NVMe solution

- M.2 2242, 2280 SSD
- NO External DRAM
- PCIe Gen 3
- NVMe protocol
- 4K Mapping FTL
- Up to 1GB/s by Gen3 x1



CrystalDiskMark 5.1.2 x64

檔案(F) 設定(S) 佈景主題(T) 說明(H) 語言(Language)

All 5 1GiB D: 0% (0/119GiB)

	Read [MB/s]	Write [MB/s]
Seq Q32T1	667.9	192.0
4K Q32T1	133.7	114.2
Seq	619.7	189.8
4K	16.69	73.93

- PCIe Gen 3 x1
- NVMe interface



Conclusion

- PCIe (NVMe) becomes the next generation mainstream server boot up drive
- SATA III currently on the market
- Small form factor is very popular
- 4K Mapping FTL is a good solution



Thank You!

C. C. Wu, EVP Innodisk USA

cc_wu@Innodisk.com

www.Innodisk.com