Inside MegaRAID

Dominik Mutterer, Field Application Engineer

Agenda

- Anwendungsgebiete & -beispiele
- Performance Betrachtungen
- Managementoberflächen
- Fragen & Antworten

Anwendungsbeispiele

Wann welcher Controller?

LSI MR SAS 9240, 9341

- 6 & 12Gb/s SAS
- über 700K IOPS
- PCI Express 2.0 & 3.0
- Bis zu 8 Ports
- RAID 0/1/5/6/10

Low Entry

LSI MR SAS 9250, 9260, 9280

- 6Gb/s SAS
- über 340k IOPS
- Bis zu 2.9GB/s Schreiben
- Bis zu 2.9GB/s Lesen
- 512MB DDR-IIII
- PCI Express 2.0
- Bis zu 24 Ports
- BBU

Value

LSI MR SAS 9266, 9271, 9286

- 6Gb/s SAS
- Über 670k IOPS
- Bis zu 3.0GB/s Schreiben
- Bis zu 4.2GB/s Lesen
- 1GB DDR-IIII
- PCI Express 3.0
- bis zu 8 Ports
- CacheVault und BBU

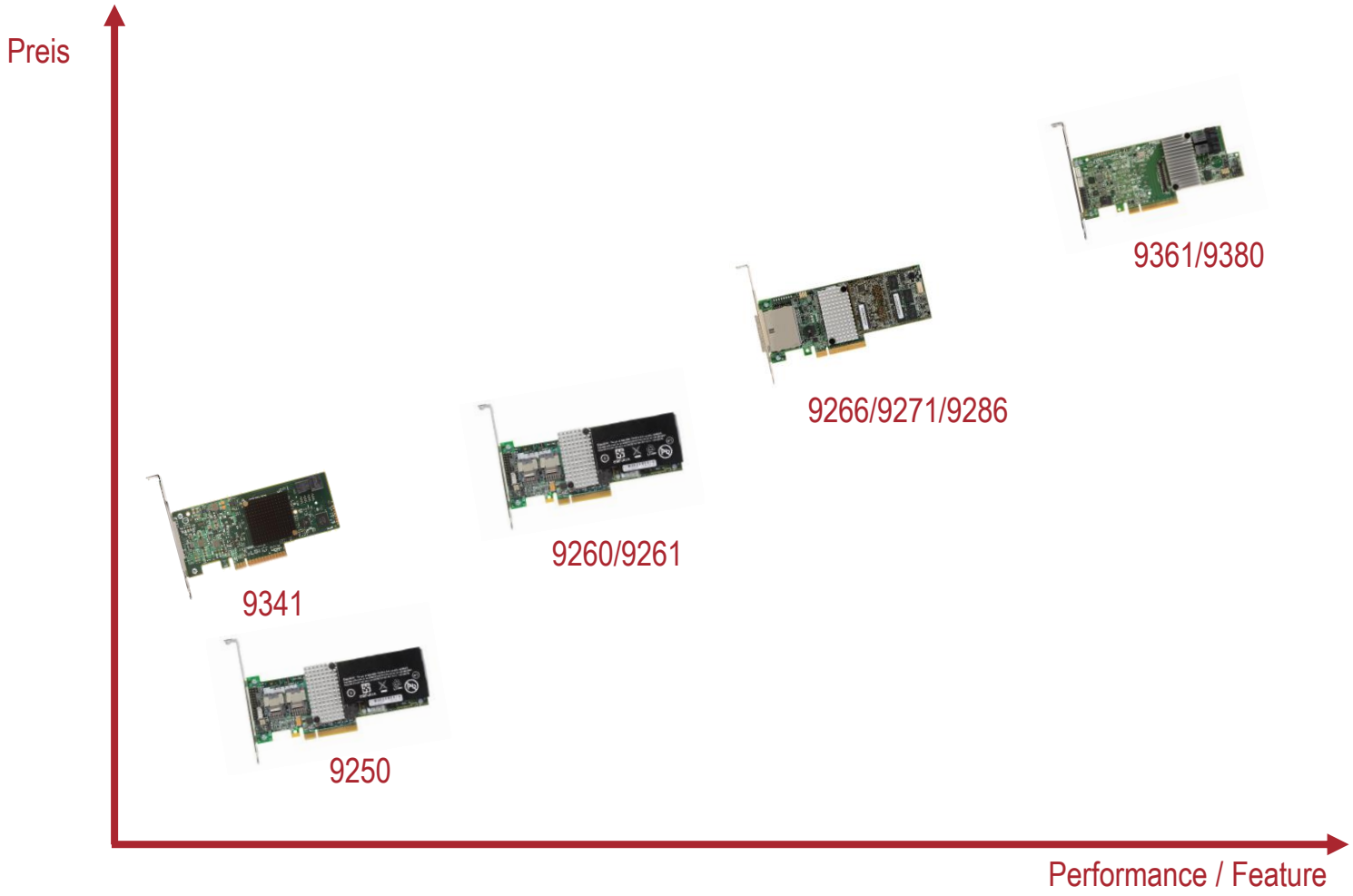
Performance

LSI MR SAS 9361, 9380

- 12Gb/s SAS
- über 700k IOPS
- Bis zu 6.5GB/s Schreiben
- Bis zu 6.0Gb/s Lesen
- 1GB DDR-III
- PCI Express 3.0
- Bis zu 16 Ports (NEU!!!)
- Nur CacheVault

Leadership

Preis-/Leistung



Einstiegssystem

Microserver MI106 [Ver.1.0]

_ Intel Celeron (max. 4 cores, up to 2GHz)
_ Upgradeable to 16GB RAM, (2 DIMMs)
_ 4x 3.5" HDD(s), configurable up to 32 TB (Gross)
_ Additional HDDs: 2 x 2,5" HDD(s) (fest installiert)
_ Expandable with 1 additional card(s), 1x PCI-E 2.0 (x1)
_ Compact, silent microserver for the office
_ Energy-efficient power supply 80plus Gold (> 87%)

12G SW RAID offloaded $$	6G HW RAID 0,1,10 Cache + BBU $	6G HW RAID Cache + BBU $$$
		
9341	9250	9260

Mittelklasse

2U Intel Dual-CPU SC825 Server [Ver.4.3]

_ Up to 2x Intel Xeon processor E5-2600v2 (4 - 12 cores, up to 3.5GHz)
_ Upgradeable to 512GB RAM, (16 DIMMs)
_ 8x 3.5" HDD(s) or 2.5" HDD(s), configurable up to 64 TB (Gross)
_ Expandable with 6 LP/ HL/ FL additional card(s), 3x PCI-E 3.0 (x16) 3x PCI-E 3.0 (x8)
_ Expandable to 4 on-board LAN ports (optional)
_ Energy-efficient power supply 80plus Platinum (> 94%)
_ Integrated IPMI (dedicated NIC)

Anwendung: mittlere Unternehmen; Web-, File-, Mail-, DB-Server; ggf. virtualisiert

6G HW RAID
Cache + BBU
$

9260

2. Gen 6G HW RAID
Cache + BBU/CV
$$

9266/9271

12G HW RAID
Cache + CV
$$$

9361/9380

AVAGO TECHNOLOGIES | STORAGE BY LSI™

Enterprise

2U Intel Dual-CPU RI8224M Server [Ver.1.0]

_ incl. CPU and RAM (optional variable) included in delivery
_ Up to 2x Intel Xeon E5-2600 (v2) Ivy Bridge CPUs (per node)
_ Upgradeable to 128GB DDR3 ECC Registered RAM, (8 DIMMs) (per node)
_ 6x 2.5" HDD(s), up to 5.4 TB configurable (per node)
_ Expandable with 1 LP additional card(s), 1x PCI-E 3.0 (x16) (per node)
_ Integrated IPMI (dedicated NIC) (per node)
_ Energy-efficient power supply 80plus Platinum (> 94%) red. NT

Anwendung: mittlere Unternehmen; virtualisiert, storageintensive Workloads, Datenbanken

2. Gen 6G HW RAID
Cache + BBU/CV
$$

9266/9271

12G HW RAID
Cache + CV
$$$

9361/9380

AVAGO TECHNOLOGIES | STORAGE BY LSI™

Advanced Software Optionen

SafeStore Key Management für SED Medien

- **Instant Secure Erase**

  Sofortiges löschen:

  - Bevor Laufwerke retourniert werden
  - Neue Verwendung für HDDs innerhalb eines Rechenzentrums

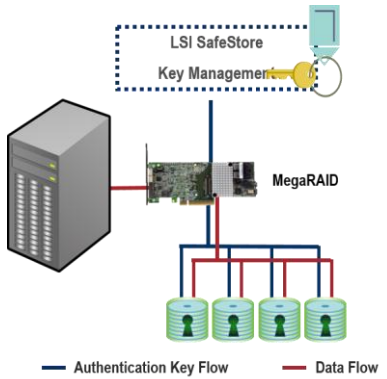
- **Auto-Locked Volumes**

  Simpel - Transparent

  - Automatisch gesichert, wenn Server heruntergefahren wird

  Sichert Daten, wenn ein Laufwerk

  - Ausfällt
  - gestohlen wird
  - wenn ein Server gestohlen wird



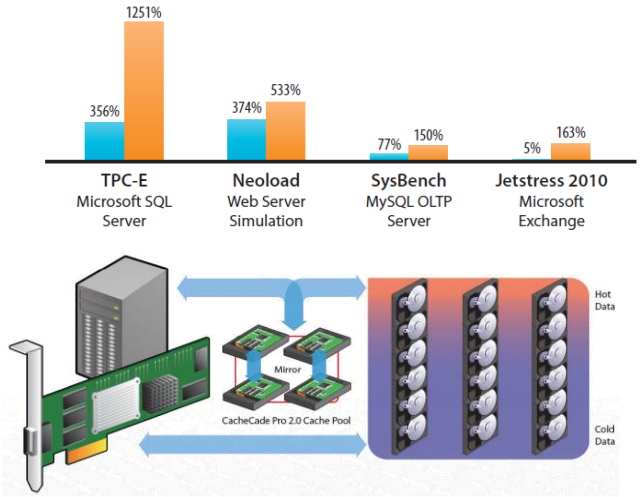
Einsatzgebiete

- HDD Austausch
- HDD Recycling
- Diebstahlschutz

Advanced Software Optionen

Cache Cade - SSD Hot Data Caching

- Beschleunigt was beschleunigt werden muss
  - Algorhytmus zur Erkennung von HotData
  - HotData werden auf einen Cache Pool aus SSDs verschoben
  - Beschleunigt Lese- & Schreibvorgänge



Einsatzgebiete

- Webserver
- Datenbanken
- Random-Workloads

Performancebetrachtungen

Housekeeping

			
Max. Bandbreite	3 Gb/s	6 Gb/s	12 Gb/s
Max. Bandbreite (4 Ports)	12 Gb/s	24 Gb/s	48 Gb/s
Max. Transferrate	300 MB/s	600 MB/s	1.200 MB/s
Max. Transferrate (4 Ports)	1.200 MB/s	2.400 MB/s	4.800 MB/s
Max. Transferrate (8 Ports)	2.400 MB/s	4.800 MB/s	9.600 MB/s

Housekeeping

	PCIe 1.0/1.1	PCIe 2.0/2.1	PCIe 3.0	PCIe 4.0
Erschienen	2003	2007	2012	~2016
Transfers/s (je Lane und Richtung)	2,5 GT/s	5,0 GT/s	8,0 GT/s	16,0 GT/s
Lanes (Breite)				
x1	250 MB/s	500 MB/s	985 MB/s	1969 MB/s
x2	500 MB/s	1000 MB/s	1969 MB/s	3938 MB/s
x4	1000 MB/s	2000 MB/s	3938 MB/s	7877 MB/s
x8	2000 MB/s	4000 MB/s	7877 MB/s	15754 MB/s
x16	4000 MB/s	8000 MB/s	15754 MB/s	31508 MB/s
(x32)	8000 MB/s	16000 MB/s	31508 MB/s	63015 MB/s

Medienlimitationen

Generation	Drive Type	Disk K IOPs	Sustained MB/s
Generation 2 (6 Gb/s)	SAS 2.5“	40 bis 250	80 bis 210
	SAS 3.5“	40 bis 250	90 bis 220
	SATA 2.5“	10 bis 70	40 bis 120
	SATA 3.5“	10 bis 70	80 bis 150
Generation 3 (12 Gb/s)	SAS HDD	40 bis 250	100 bis 220
	SATA HDD	10 bis 80	50 bis 150
	SAS SSD	10 bis 120	550
	SATA SSD	10 bis 100	550

Read Policies

- **Always Read Ahead** – Mit dieser Option nutzt der Controller Read-Ahead wenn die letzten beiden Disk zugriffen in sequenziellen Sektoren stattfanden. Wenn alle Reads in einem Random Muster stattfinden verwendet der Algorithmus kein Read Ahead, aber alle Anfragen werden dauerhaft auf mögliche sequenziellen Operationen überprüft.

- **No Read Ahead** – Nur die angeforderten Daten werden gelesen und der Controller tätigt kein Read Ahead von Daten

Write Policies

- **Write-Through –** mit dieser Caching Strategie werden Daten auf die Disks geschrieben bevor ein Abschluss Status ans Host OS gegeben wird – wir als sicherer gesehen, da bei einem Strom-Ausfall ungeschriebene Daten verloren gehen, wenn keine Cache-Absicherung vorhanden ist. Die Daten fließen direkt vom Host zu den Disks ohne die Daten vorher in den Cache kopiert zu haben was die Gesamt-Performance von Streaming Workloads mit Direct IO Mode erhöhen kann.

- **Write-Back –** mit dieser Caching Strategie wird bei Schreiboperationen ein Abschluss Status ans Host OS gegeben, sobald die Daten im RAID Cache geschrieben wurden. Die Daten werden auf die Disk geschrieben sobald dies vom Controller vorgegeben wird. Write-Back ist effizienter wenn temporäre und/oder die Größe der Anfragen niedriger sind als die Größe des Controller Caches. Write-Back ist sehr effizient in schreiblastigen Umgebungen. Cache Absicherung gegen Stromausfall wird empfohlen.

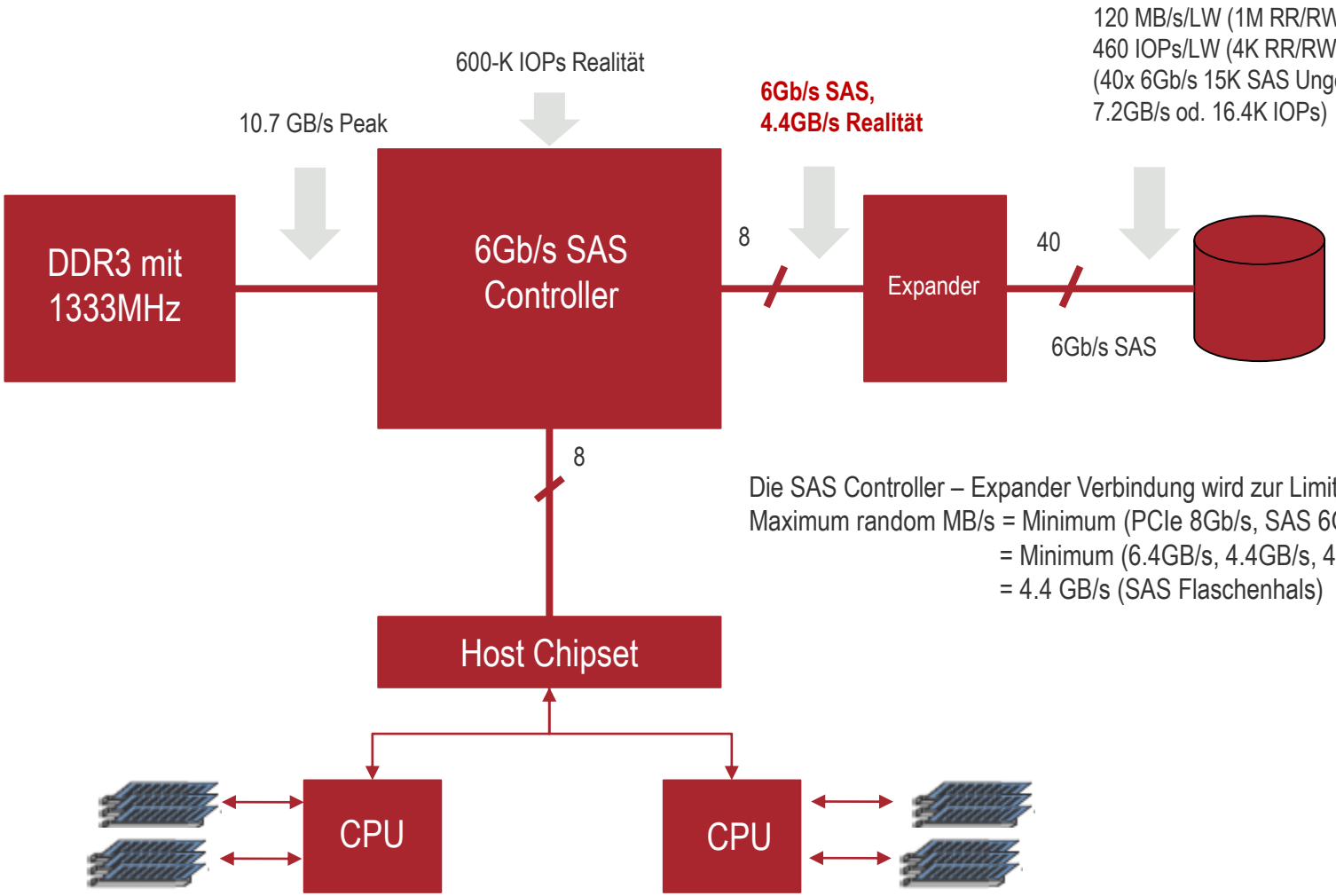
Data Placement Policies

- **Direct IO –** Alle Lese Daten umgehen den RAID Controller Cache und werden direkt in den Host Speicher transferiert. Alle Read Ahead Daten werden gecached – Alle Schreibvorgänge werden direkt vom Host Speicher unter Umgehung des RAID Controller Caches auf die Disks geschrieben, wenn Write-Through eingestellt ist. Empfehlung für alle Konfiguration.

- **Cached IO –** Alle Lese- und Schreibvorgänge gehen durch den Controller Cache auf dem Weg von oder zu dem Host (inklusive Schreibvorgänge in Write-Through Modus) – Nur für CacheCade 1.1 (Read Caching) ansonsten nicht zu empfehlen.

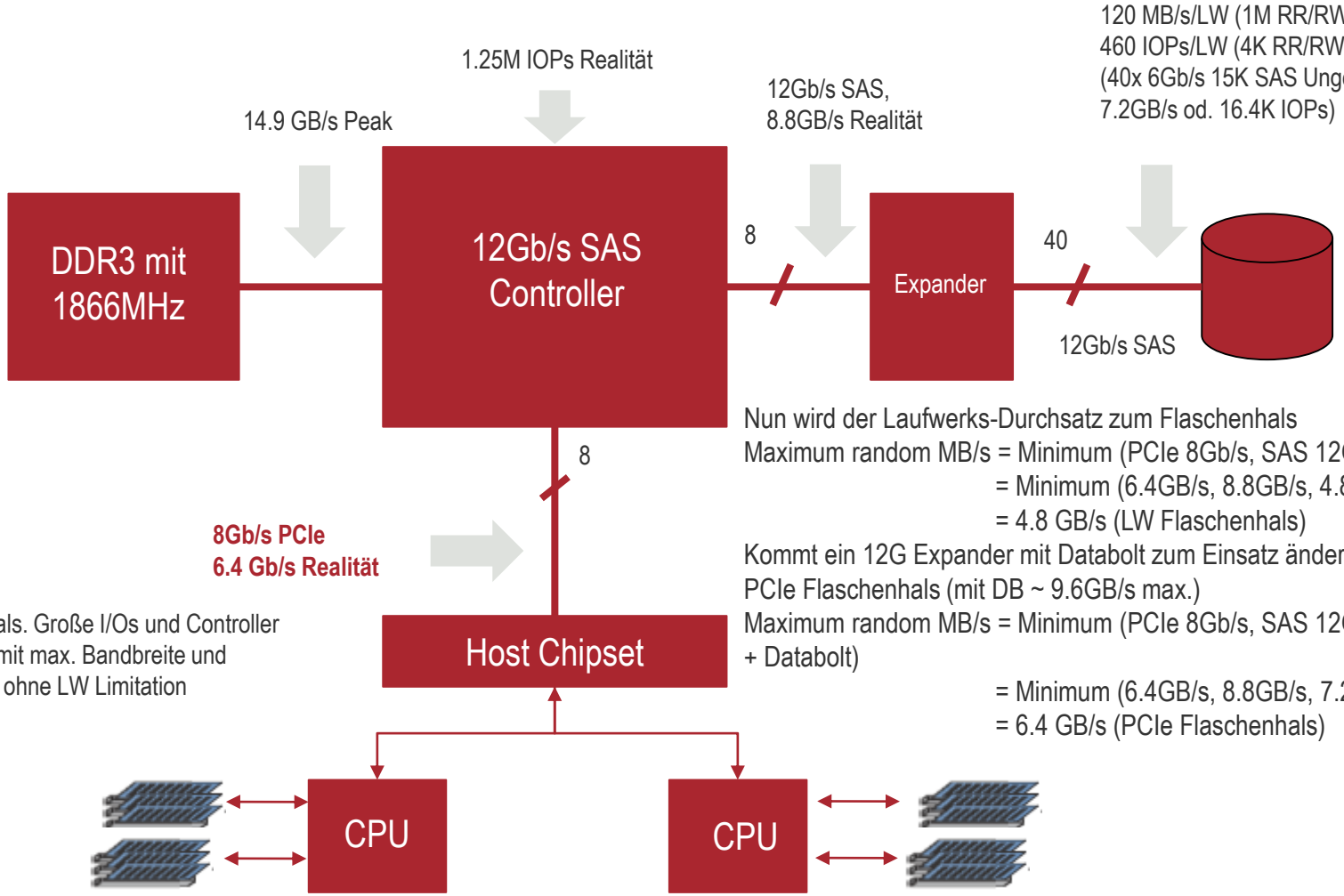
Limitations-Beispiele

6Gb/s SAS Controller Limitation



Die SAS Controller – Expander Verbindung wird zur Limitation:
Maximum random MB/s = Minimum (PCIe 8Gb/s, SAS 6Gb/s, 40x LW)
= Minimum (6.4GB/s, 4.4GB/s, 40x 120 MB/s)
= 4.4 GB/s (SAS Flaschenhals)

12Gb/s SAS Controller PCIe Limitation



Nun wird der Laufwerks-Durchsatz zum Flaschenhals
Maximum random MB/s = Minimum (PCIe 8Gb/s, SAS 12Gb/s, 40x LW)
= Minimum (6.4GB/s, 8.8GB/s, 4.8 GB/s)
= 4.8 GB/s (LW Flaschenhals)

Kommt ein 12G Expander mit Databolt zum Einsatz ändert sich das zum PCIe Flaschenhals (mit DB ~ 9.6GB/s max.)
Maximum random MB/s = Minimum (PCIe 8Gb/s, SAS 12Gb/s, 40x LW + Databolt)
= Minimum (6.4GB/s, 8.8GB/s, 7.2 GB/s)
= 6.4 GB/s (PCIe Flaschenhals)

PCIe Flaschenhals. Große I/Os und Controller Interface laufen mit max. Bandbreite und Geschwindigkeit ohne LW Limitation

10.66 GB/s (DDR3-1333 SDRAMs 8-byte primary data bus)

10.66 GB/s (DDR3-1333 SDRAMs 8-byte primary data bus)

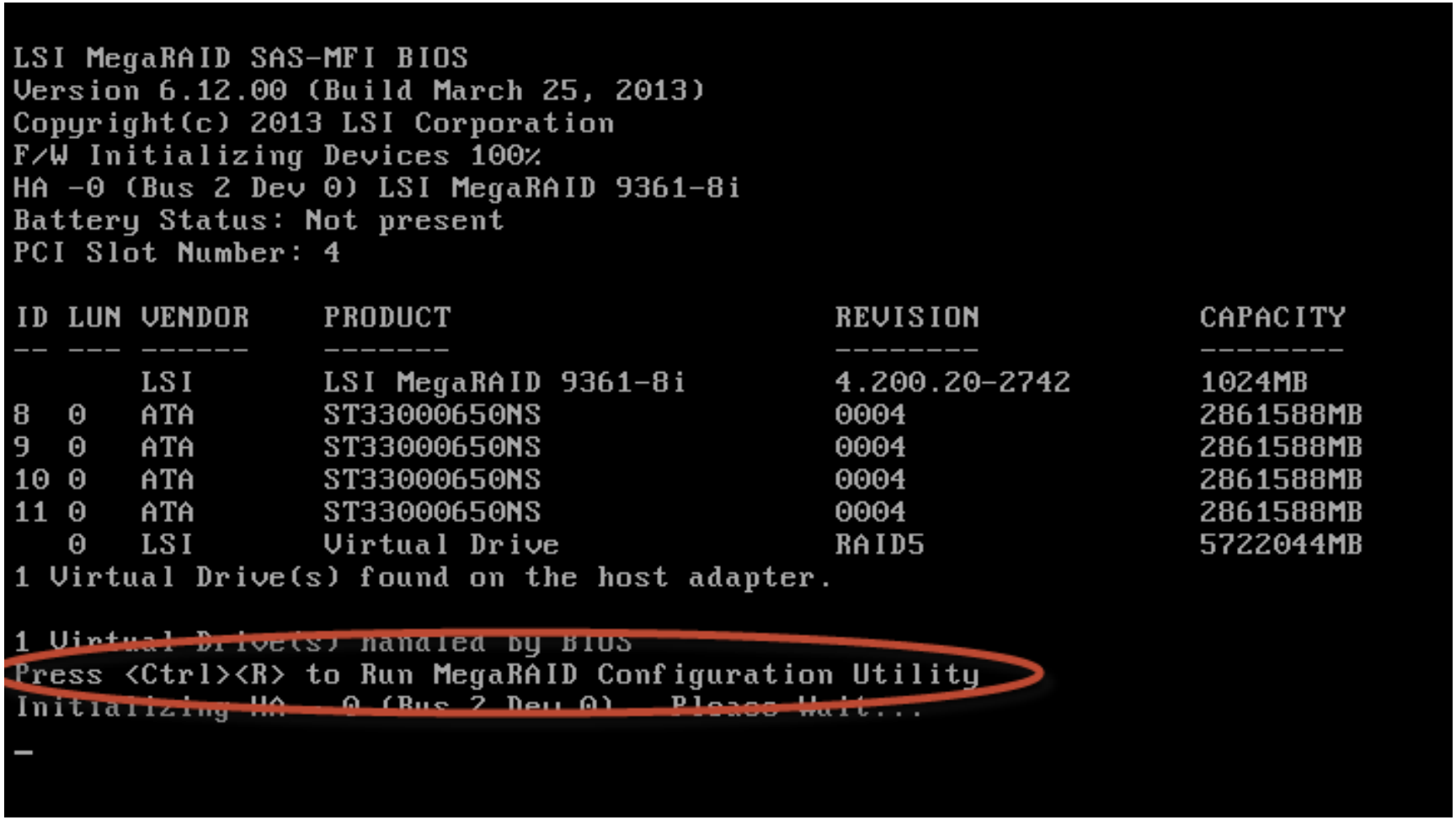
Wer viel misst, misst Mist…

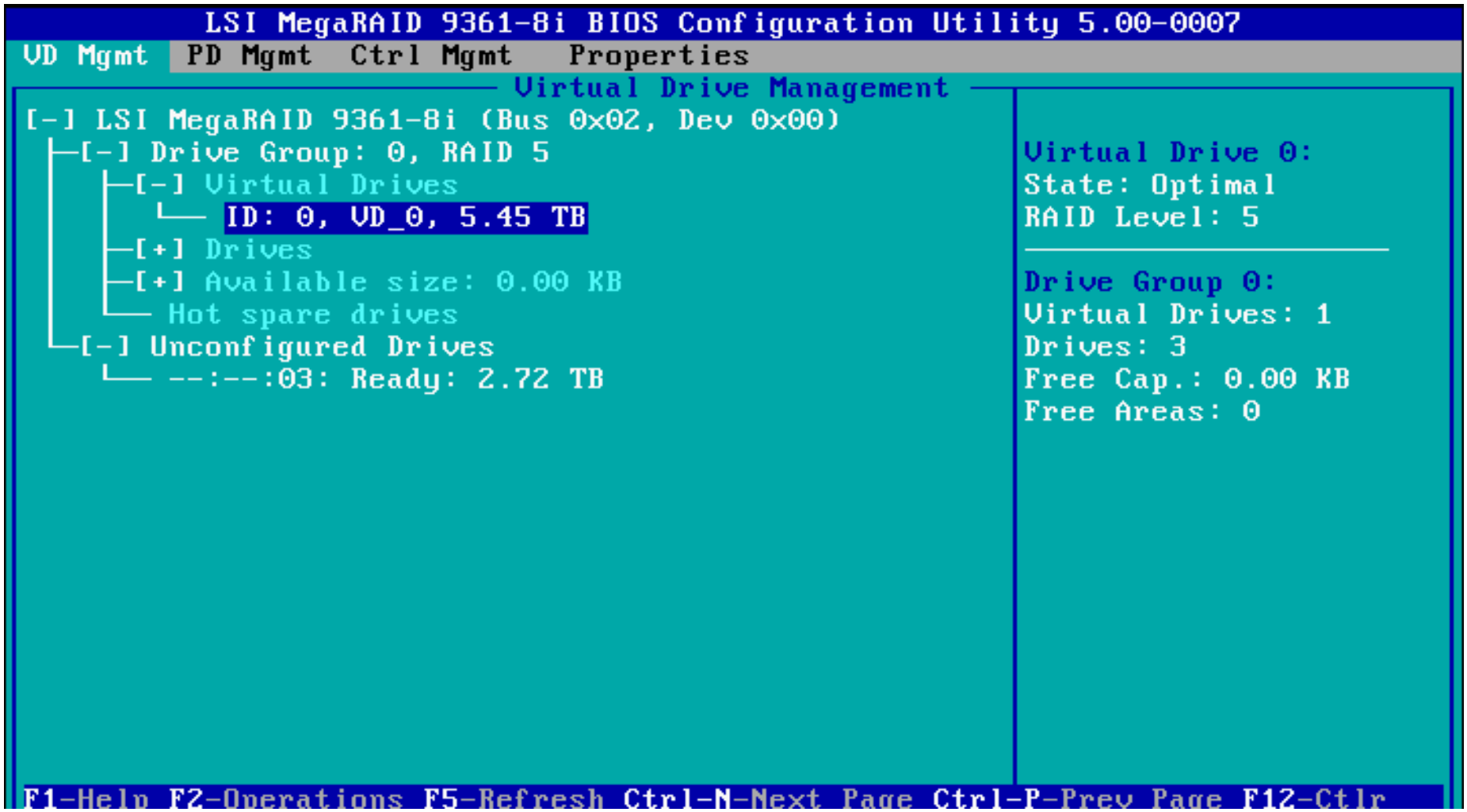
- Traue keinem Benchmark, den Du nicht selbst getunet hast
- Beachte die gegebenen Limitationen
- Formuliere, was gebenchmarked werden soll
- Stelle die Erwartungen korrekt auf
- Verwende ein empfohlenes, aussagekräftiges Benchmark-Tool

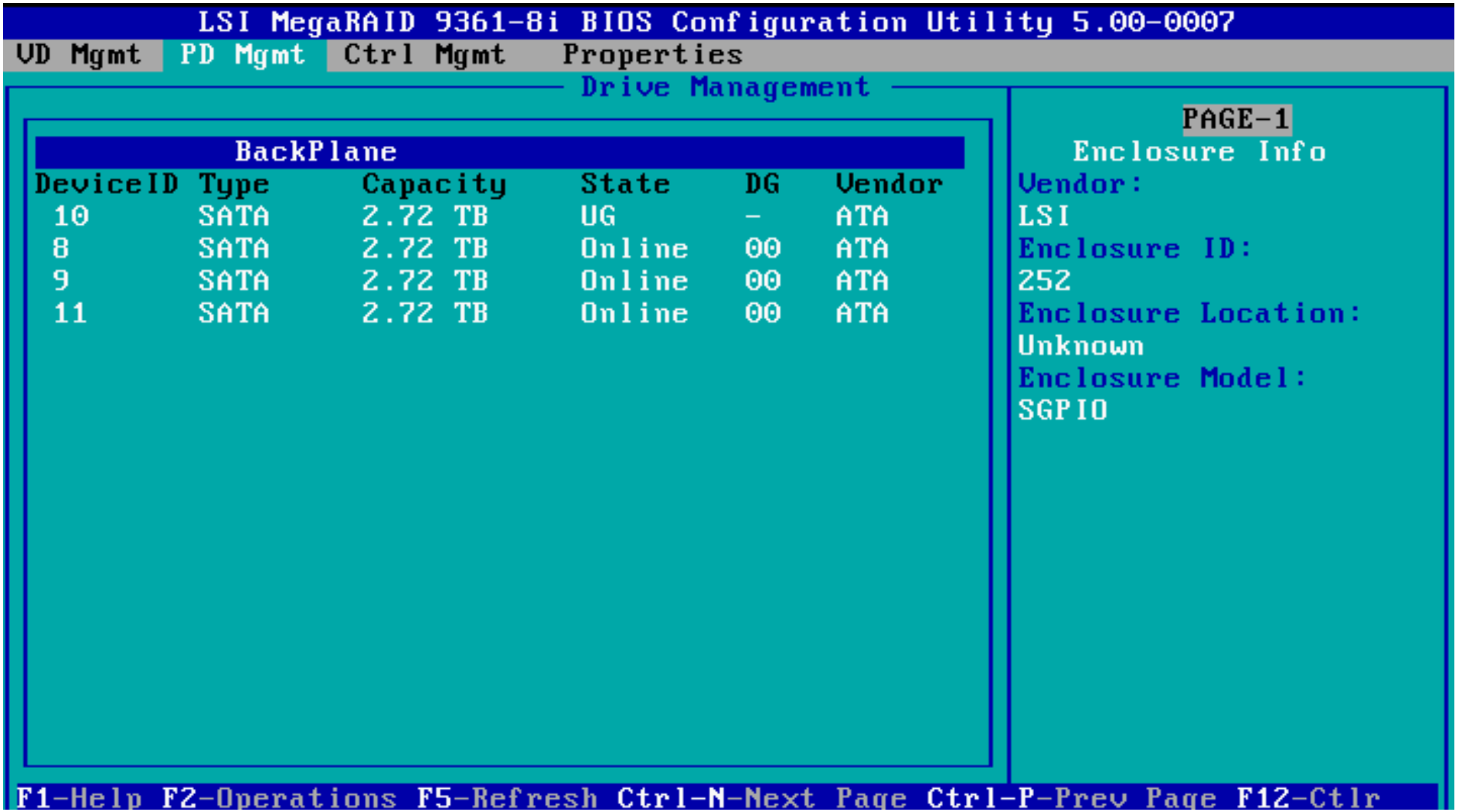
Managementoberflächen

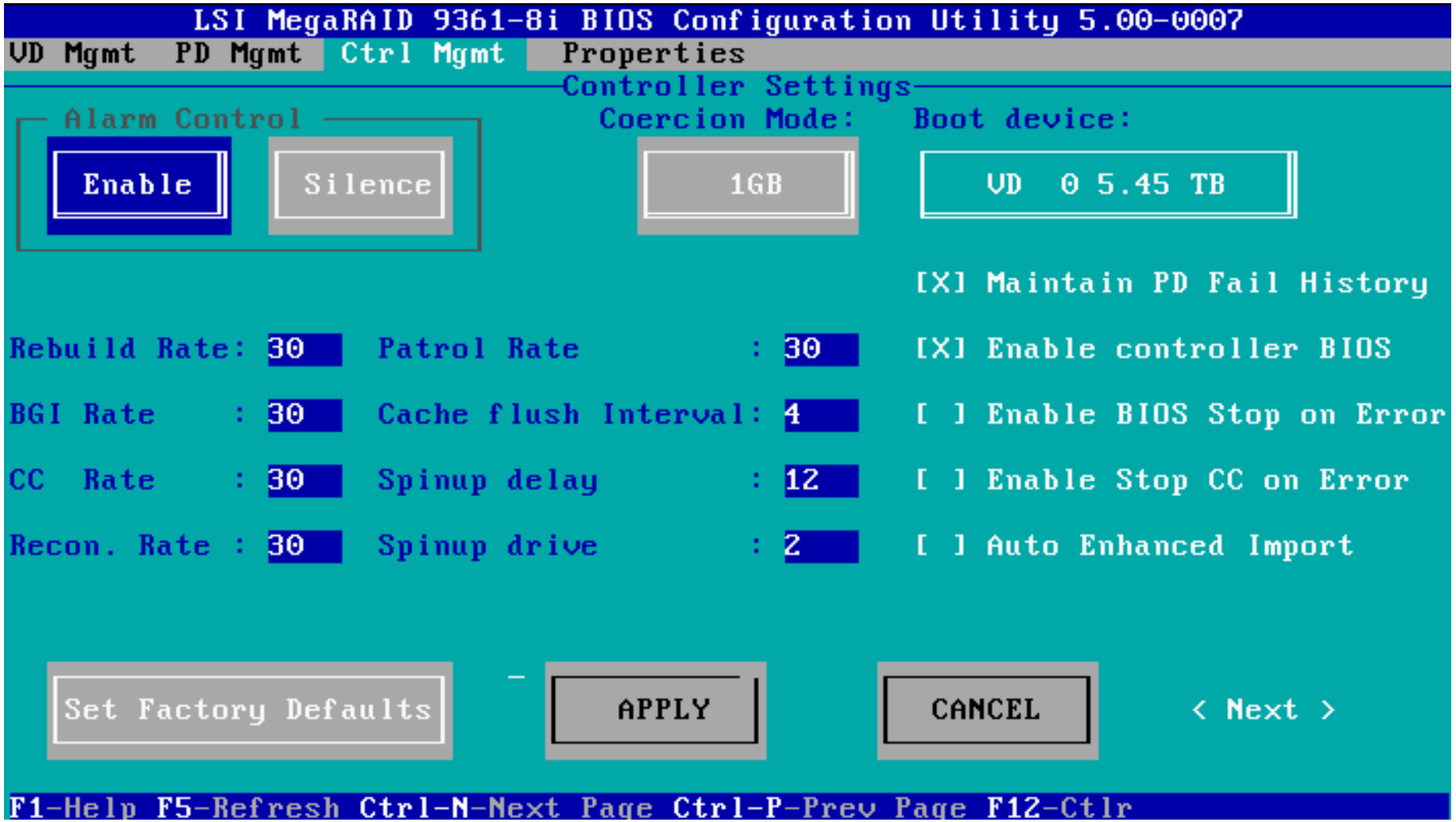
Configuration Utility (Pre-Boot)

Configuration Utility

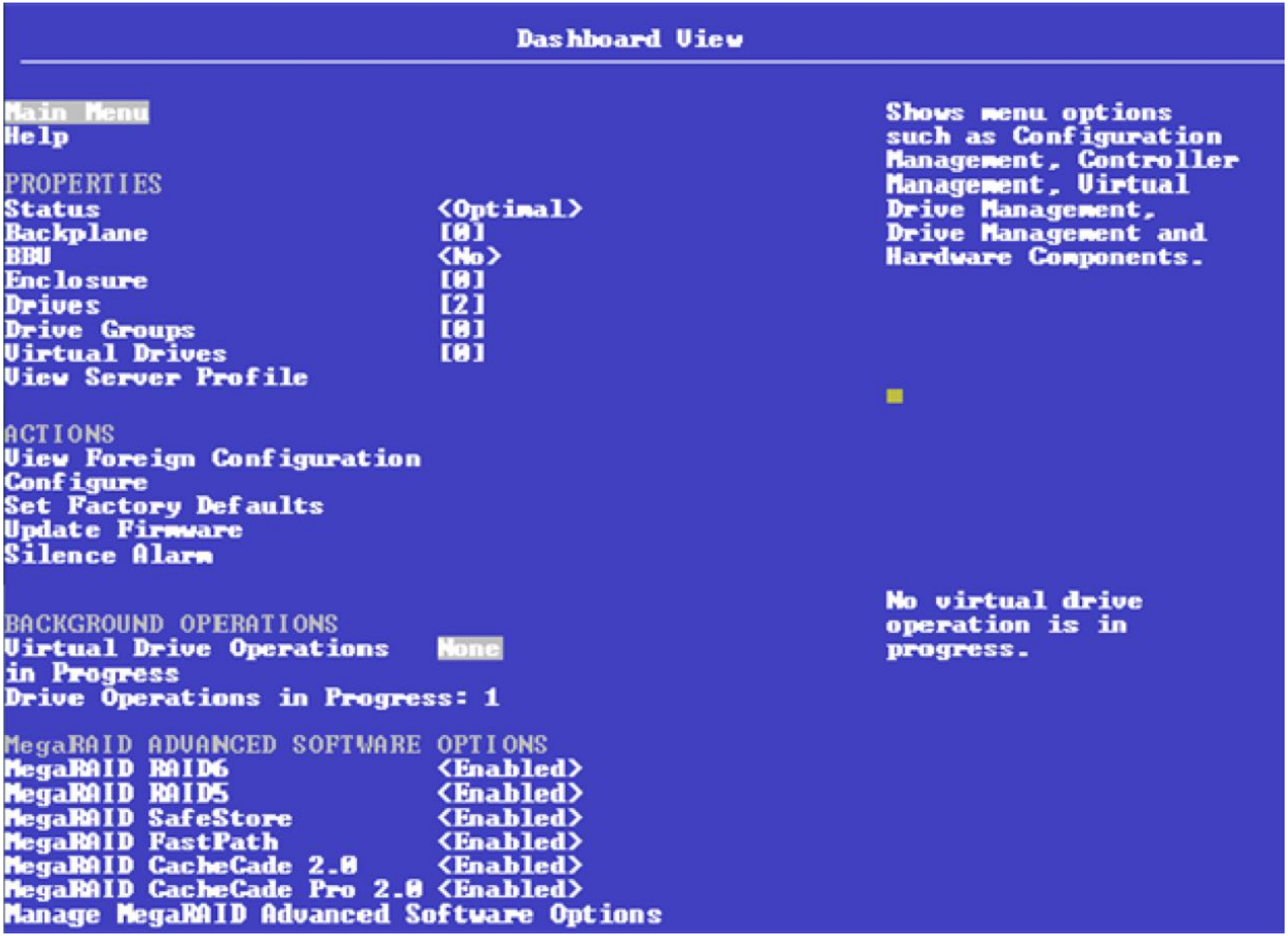


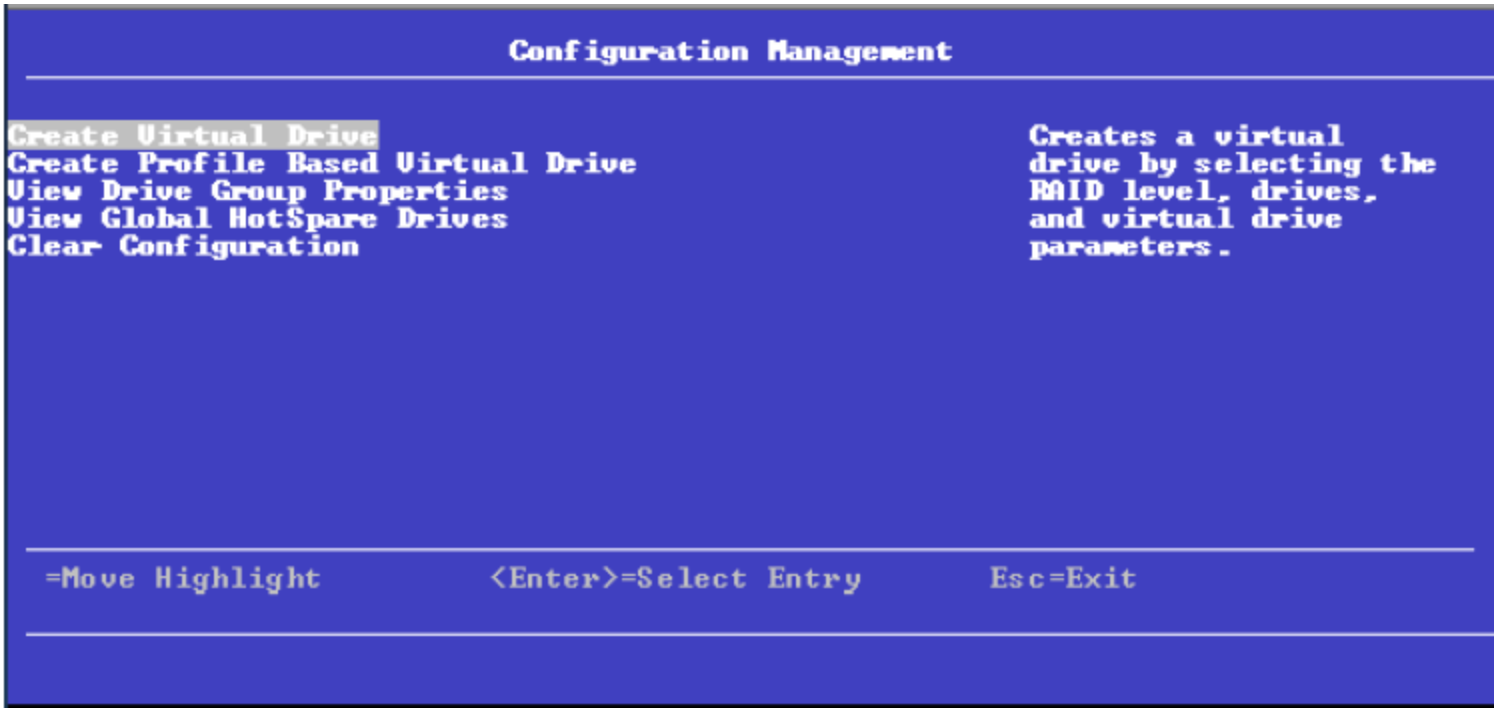

LSI MegaRAID 9361-8i BIOS Configuration Utility 5.00-0007
VD Mgmt
PD Mgmt
Ctrl Mgmt
Properties
Virtual Drive Management
[-] LSI MegaRAID 9361-8i (Bus 0x02, Dev 0x00)
[-] Drive Group: 0, RAID 5
[-] Virtual Drives
ID: 0, VD_0, 5.45 TB
[+] Drives
[+] Available size: 0.00 KB
Hot spare drives
[-] Unconfigured Drives
--:--:03: Ready: 2.72 TB
Virtual Drive 0:
State: Optimal
RAID Level: 5
Drive Group 0:
Virtual Drives: 1
Drives: 3
Free Cap.: 0.00 KB
Free Areas: 0
F1-Help F2-Operations F5-Refresh Ctrl-N-Next Page Ctrl-P-Prev Page F12-Ctlr

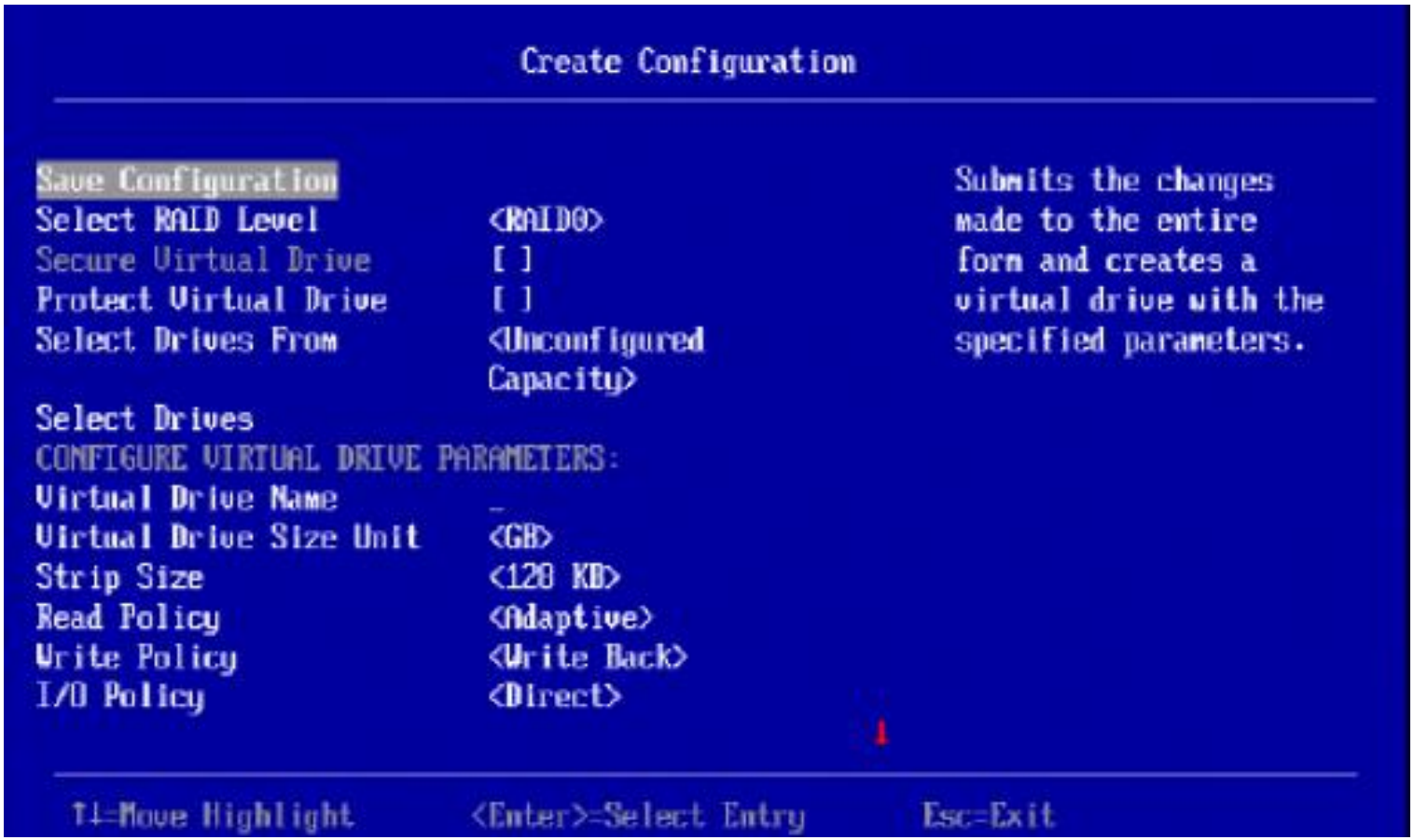


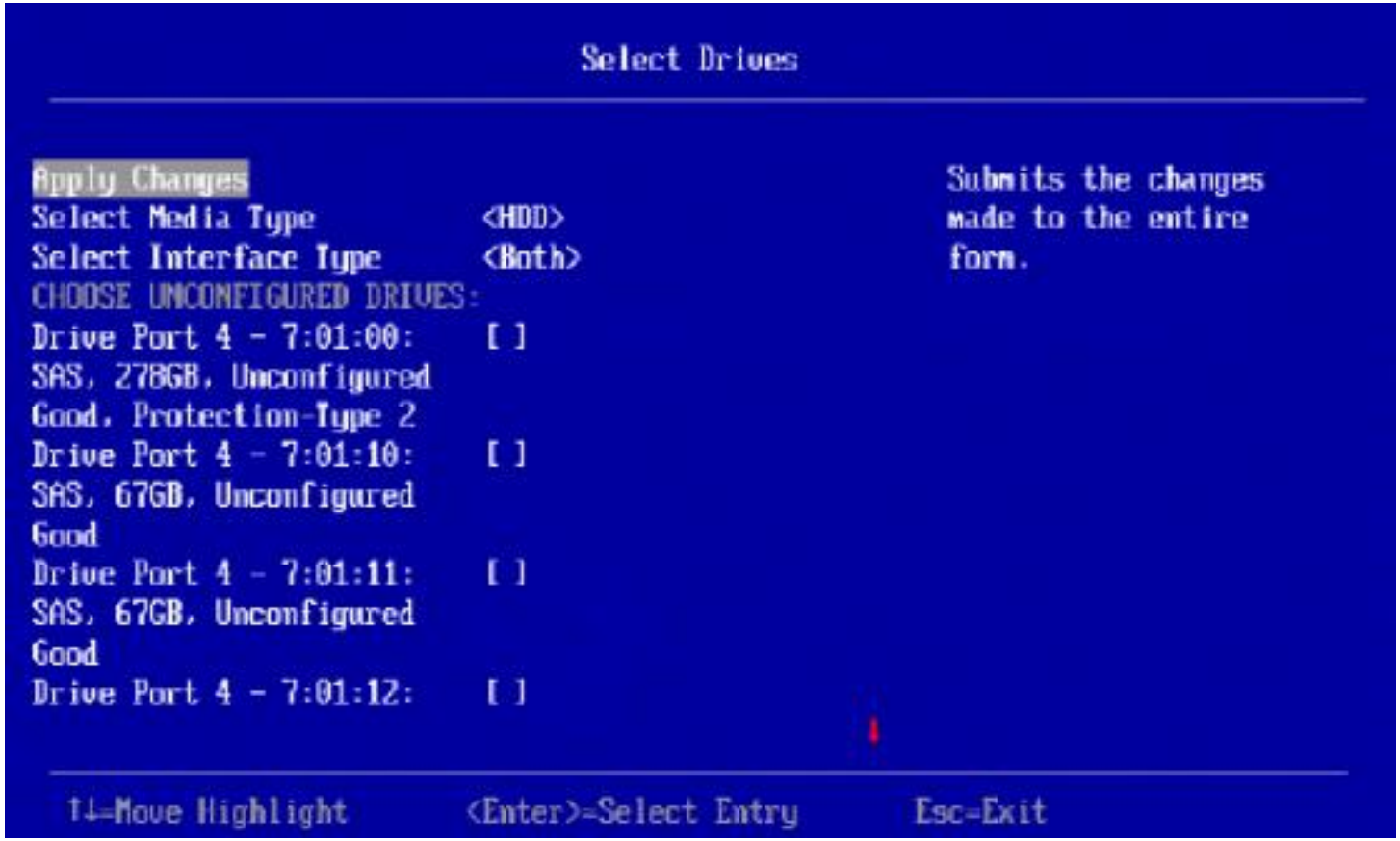


HII – Human Interface Infrastructure (EFI)

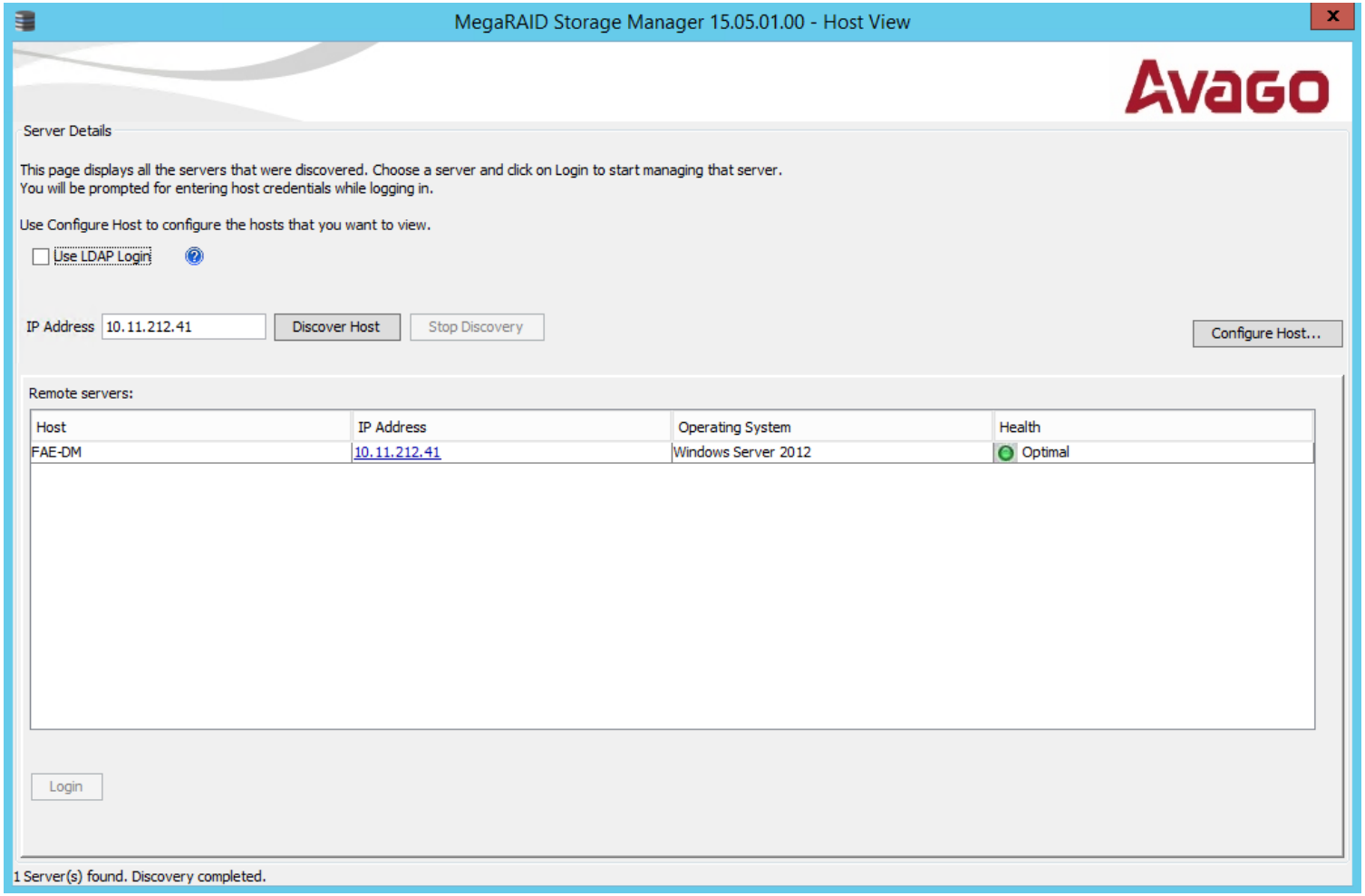


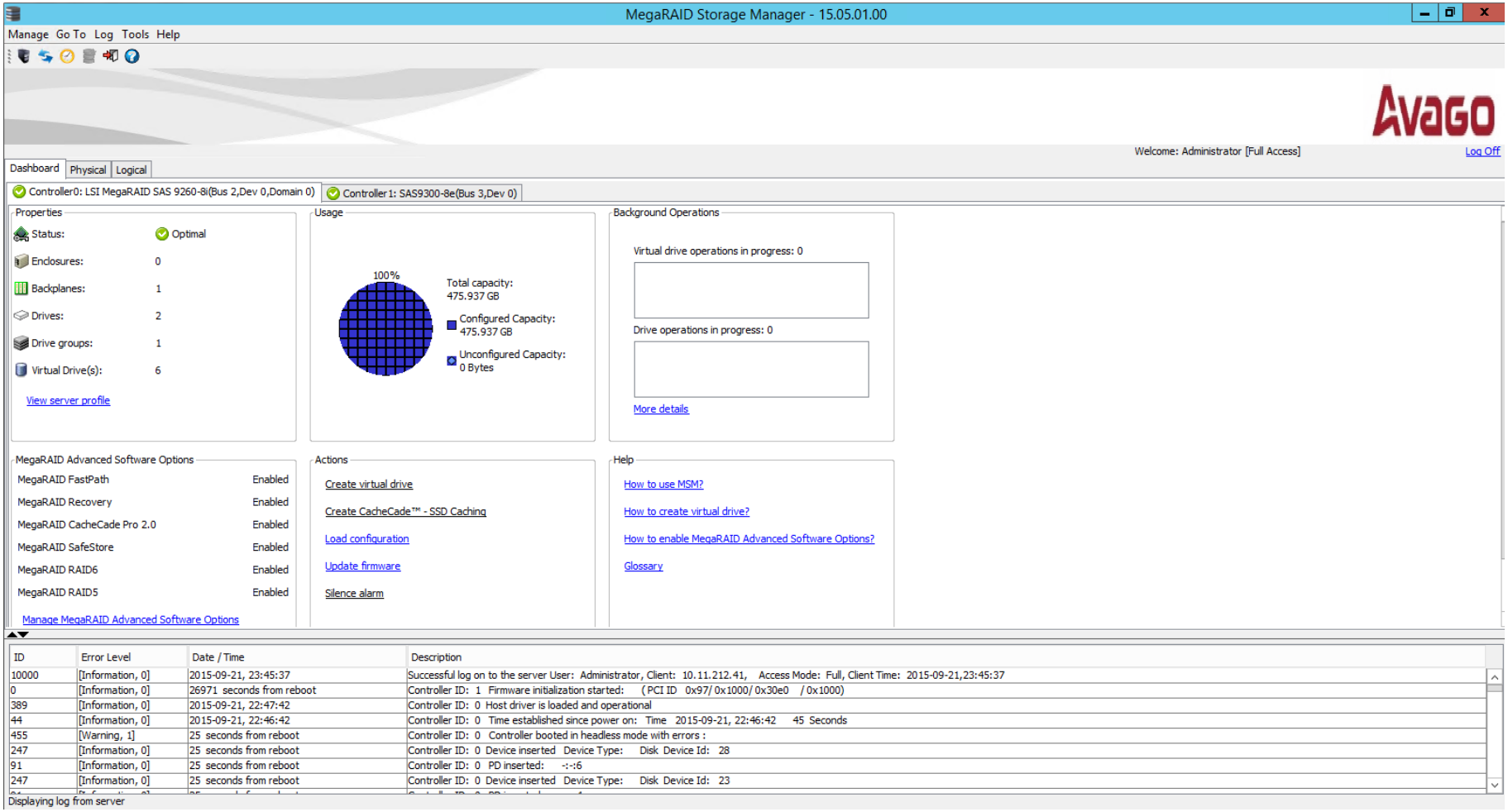


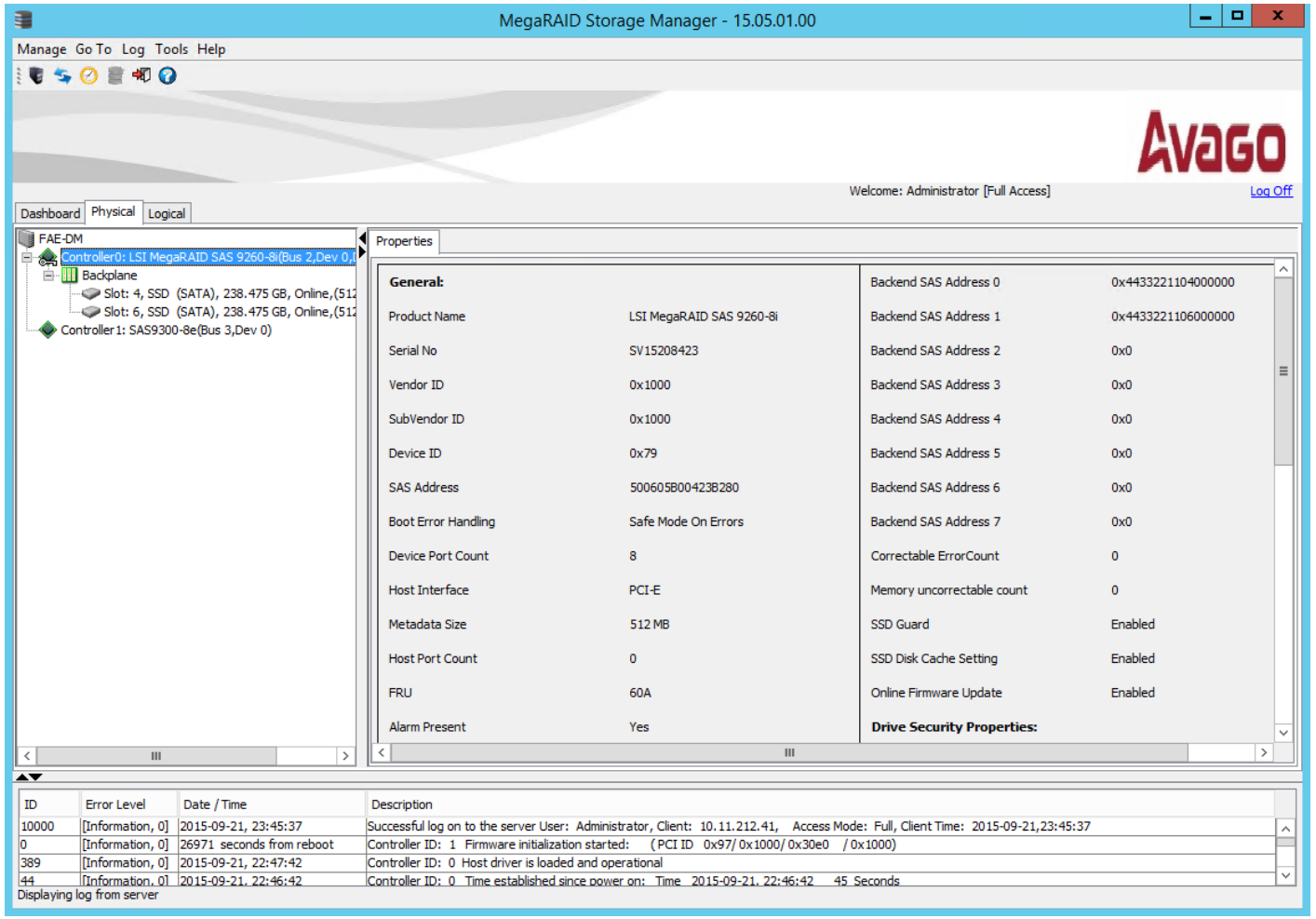


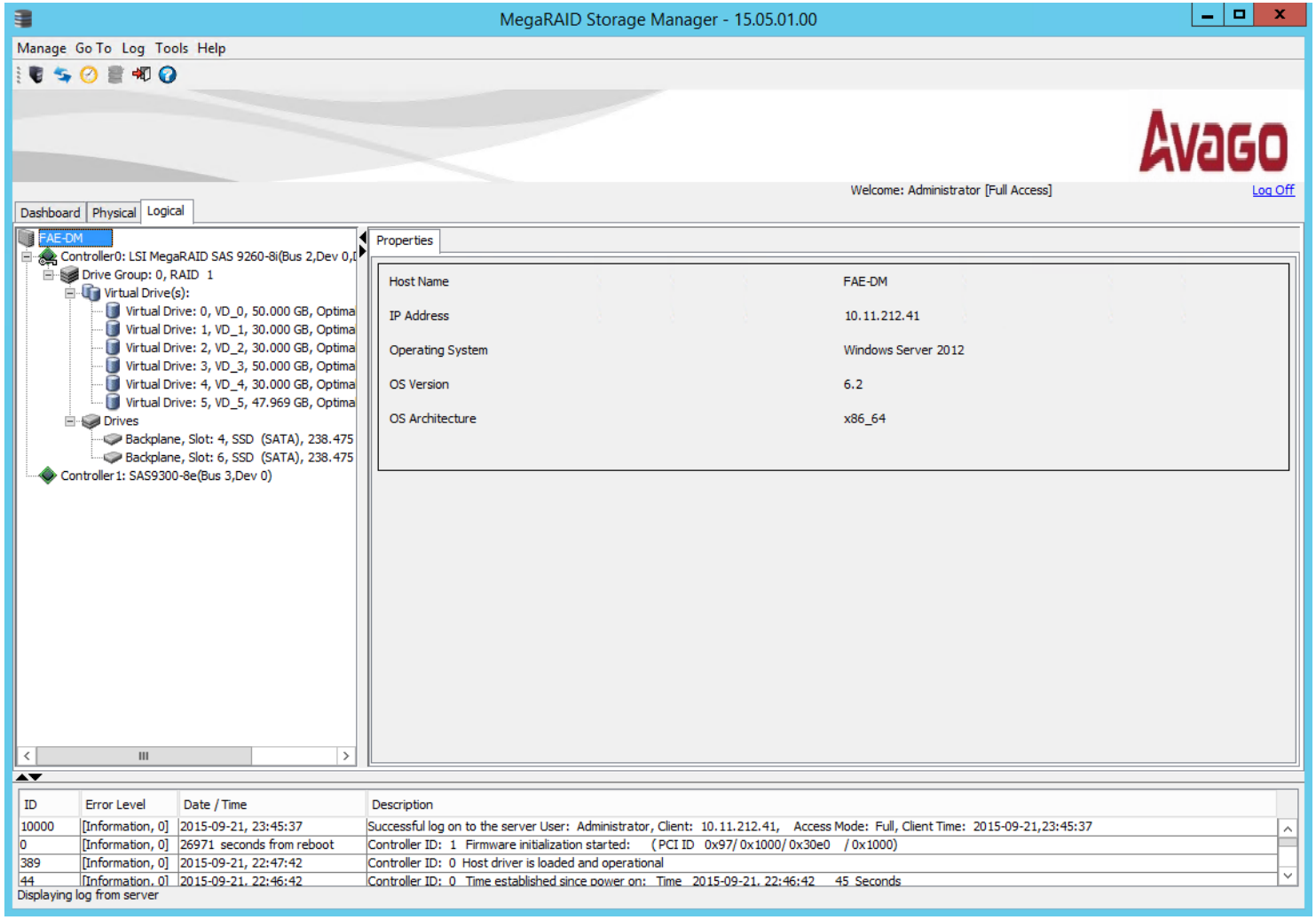

Select Drives
Apply Changes
Select Media Type <HDD>
Select Interface Type <Both>
CHOOSE UNCONFIGURED DRIVES:
Drive Port 4 - 7:01:00: [ ]
SAS, 278GB, Unconfigured
Good, Protection-Type 2
Drive Port 4 - 7:01:10: [ ]
SAS, 67GB, Unconfigured
Good
Drive Port 4 - 7:01:11: [ ]
SAS, 67GB, Unconfigured
Good
Drive Port 4 - 7:01:12: [ ]
Submits the changes made to the entire form.
↑↓=Move Highlight
<Enter>=Select Entry
Esc=Exit

MegaRAID Storage Manager


MegaRAID Storage Manager 15.05.01.00 - Host View
AVAGO
Server Details
This page displays all the servers that were discovered. Choose a server and click on Login to start managing that server.
You will be prompted for entering host credentials while logging in.
Use Configure Host to configure the hosts that you want to view.
Use LDAP Login
IP Address 10.11.212.41
Discover Host
Stop Discovery
Configure Host...
Remote servers:
Host
IP Address
Operating System
Health
FAE-DM
10.11.212.41
Windows Server 2012
Optimal
Login
1 Server(s) found. Discovery completed.







StorCLI

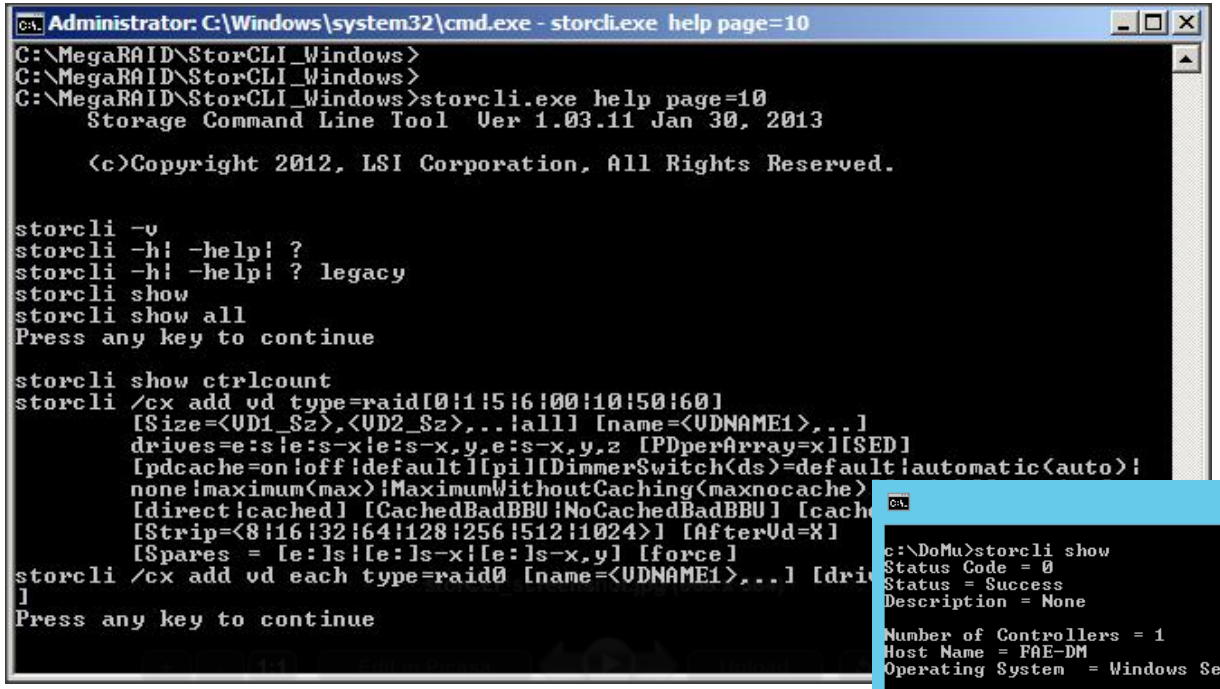
OS Unterstützung

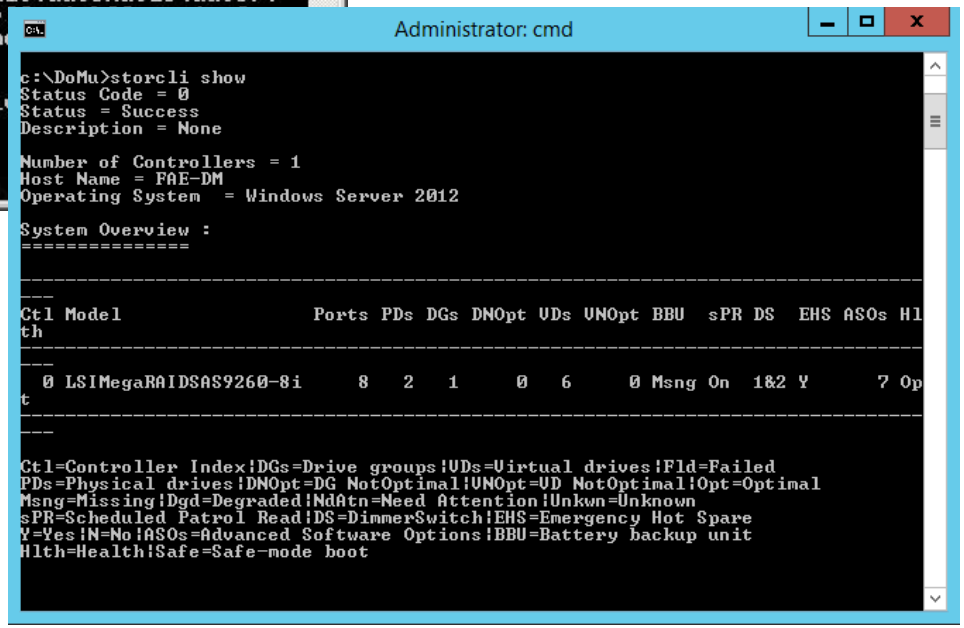
- **Windows**
- **Linux**
  - RHEL
  - SLES
  - Fedora
  - Ubuntu
- **Solaris**
- **VMware**
- **FreeBSD**
- **UEFI**
- **Pre-Boot (limitiert)**

StorCLI Syntax

<[object identifier]> <verb> <[adverb | attributes | properties] > <[key=value]>

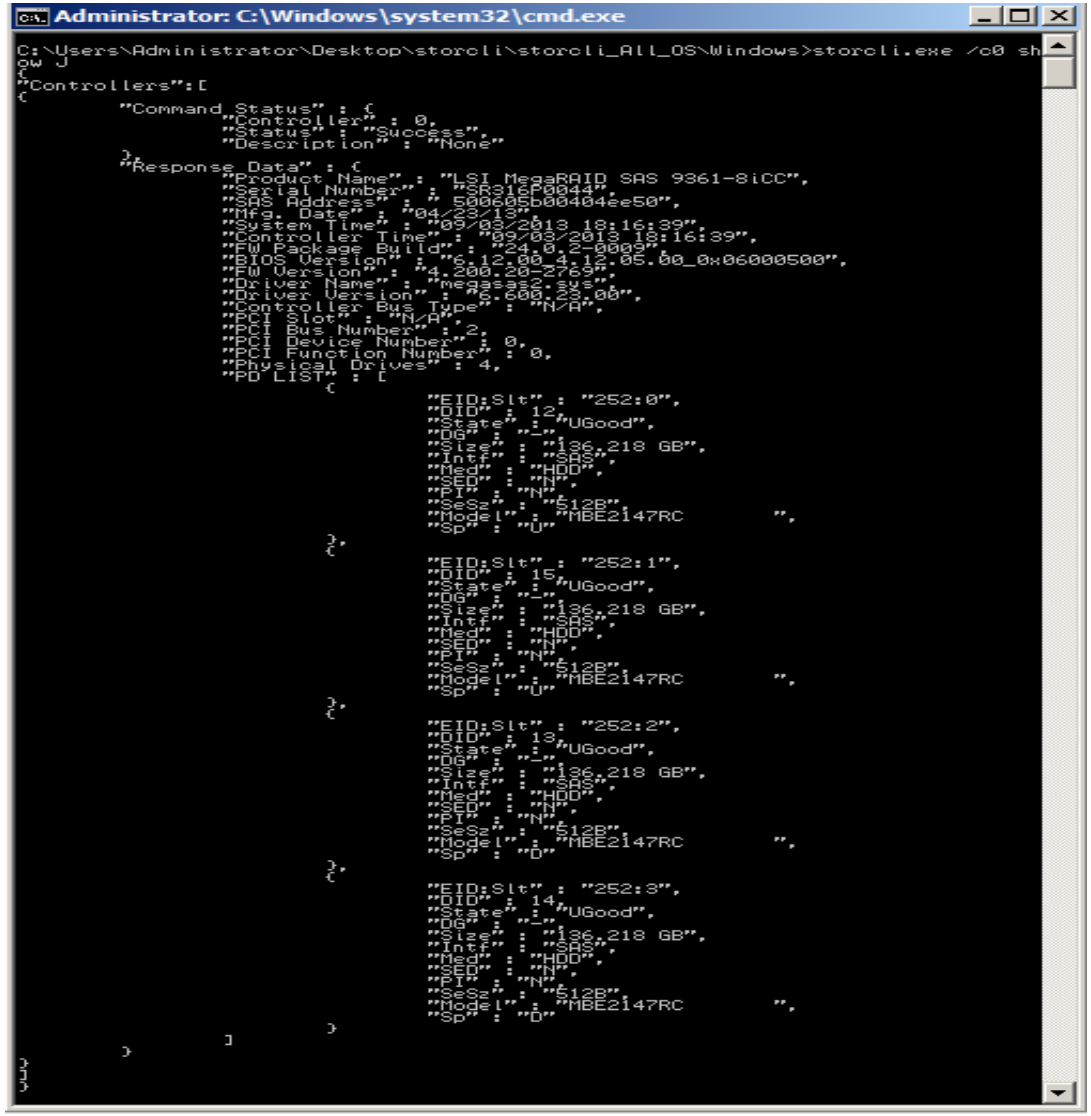
Object Identifier	Description
No object identifier specified	If there is no object identifier, the command is a system command.
/cx	This object identifier is for controller *x* .
/cx/vx	This object identifier is for a virtual drive *x* on controller *x* .
/cx/vall	This object identifier is for all virtual drives on controller *x*.
/cx/ex	This object identifier is for an enclosure *x* on controller *x*.
/cx/eall	This object identifier is for all enclosures on controller *x*.
/cx/fx	This object identifier is for a foreign configuration *x* on controller *x*.
/cx/fall	This object identifier is for all foreign configurations on controller *x*.
/cx/ex/sx	This object identifier is for the drive is slot *x* on enclosure *x* on controller *x*.
/cx/ex/sall	This object identifier is for all the drives on enclosure *x* on controller *x*.
/cx/ex/dx	This object identifier is for the drive group *x* on enclosure *x* on controller *x*.
/cx/ex/dall	This object identifier is for the all drive groups on enclosure *x* on controller *x*.



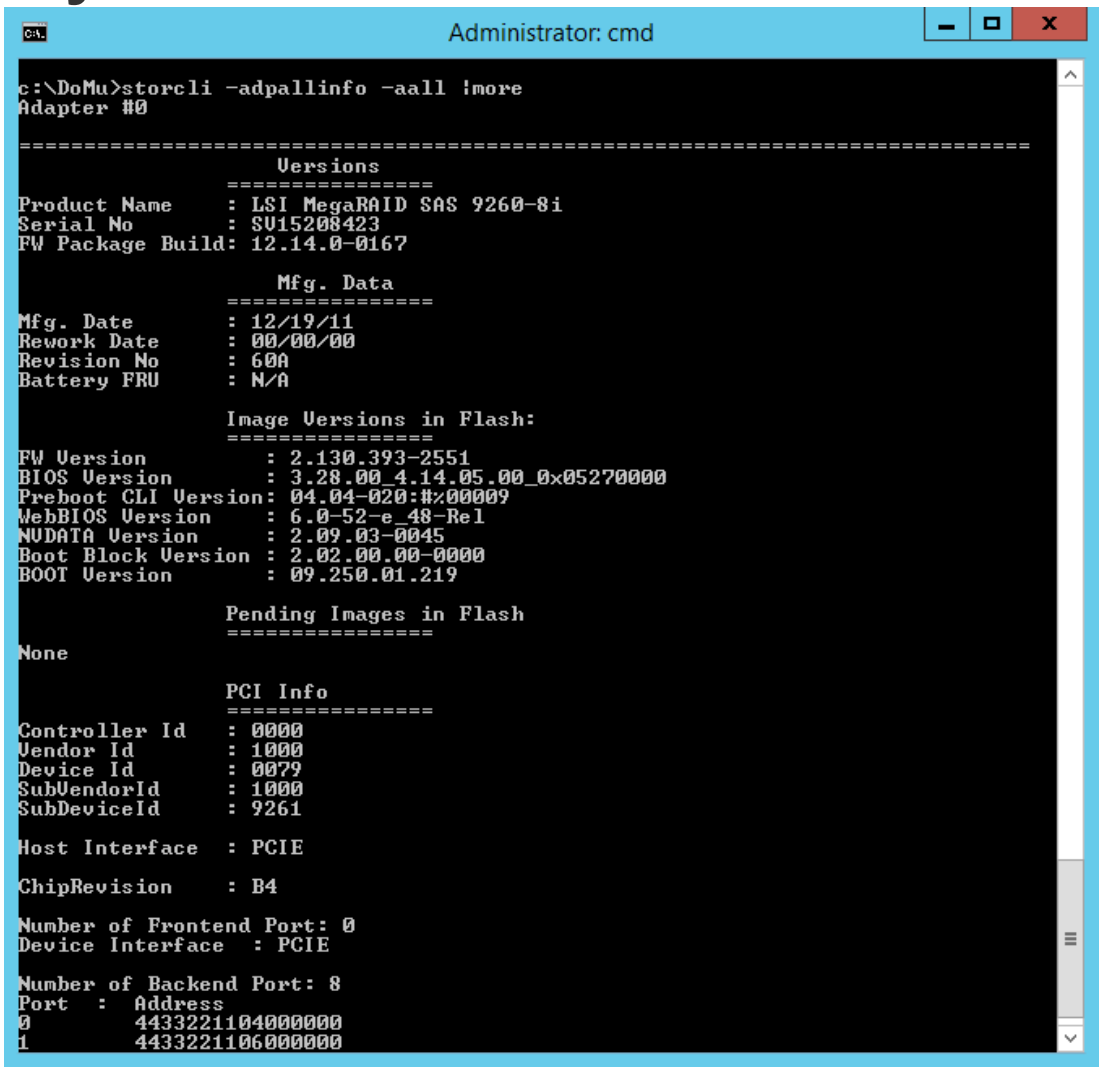


/c0 show

/c0 show in JSON Format



MegaCLI Syntax enthalten



Fragen & Antworten


AVAGO
TECHNOLOGIES
STORAGE
BY LSI™