

Linux on System z Failover-Szenario mit FCP

Christof Schmitt <christof.schmitt@de.ibm.com>



Übersicht

Für einen zuverlässigen Zugriff auf Daten sind Anbindungen der Plattensysteme über mehrere Pfade und das Spiegeln der Daten wichtige Voraussetzungen. In diesem Workshop wird an Hand eines Beispiels gezeigt, wie man diese Funktionen mit FCP unter Linux on System z einrichtet.



Agenda

- FCP mit Linux auf System z
- Multipathing
 - für Bandlaufwerke
 - für Plattensysteme
 - Beispiel-Szenario
- Spiegeln von Daten
 - LVM-Mirror
 - mdadm
 - Beispiele
- root-Dateisystem auf multipath-Gerät



FCP für Linux auf System z

- bessere Performance gegenüber FICON
- Asynchrones I/O-Modell
- Keine Einschränkungen der Plattengröße
- Bis zu 16 Partitionen auf SCSI-Platten
- Keine Limitierungen wie FICON-Topologie
- Flexible Konfiguration (keine IOCDs-Änderungen für Platten)
- Anbindung von Bandlaufwerken für Datensicherungen ins SAN



Anforderungen an die Storage-Umgebung

- FCP-Anbindung für ... ?
 - Bandlaufwerke
 - Platten für Daten
 - Platten für root-Dateisystem
- Sind Vorkehrungen notwendig gegen ... ?
 - Wartung oder Problemen in der SAN-Fabric
 - Plattensystem-Controller-Reset oder -Wartung
 - Ausfall von physikalischen Platten
 - Ausfall eines Speichersystems
 - nichtautorisierter Zugriff auf Speichergeräte im SAN



Mögliche Lösungen für die Anforderungen

- Im SAN
 - Zugriffsbeschränkungen über NPIV, Zoning, LUN-Maskierung
 - Redundante Fabrics
- Im Speichersystem
 - RAID zum Schutz gegen Ausfall physikalischer Platten
 - Spiegeln der Daten zur Remote-Lokation
- In Linux
 - Failover über mehrere Pfade bzw. SAN-Fabrics
 - Failover über gespiegelte Platten

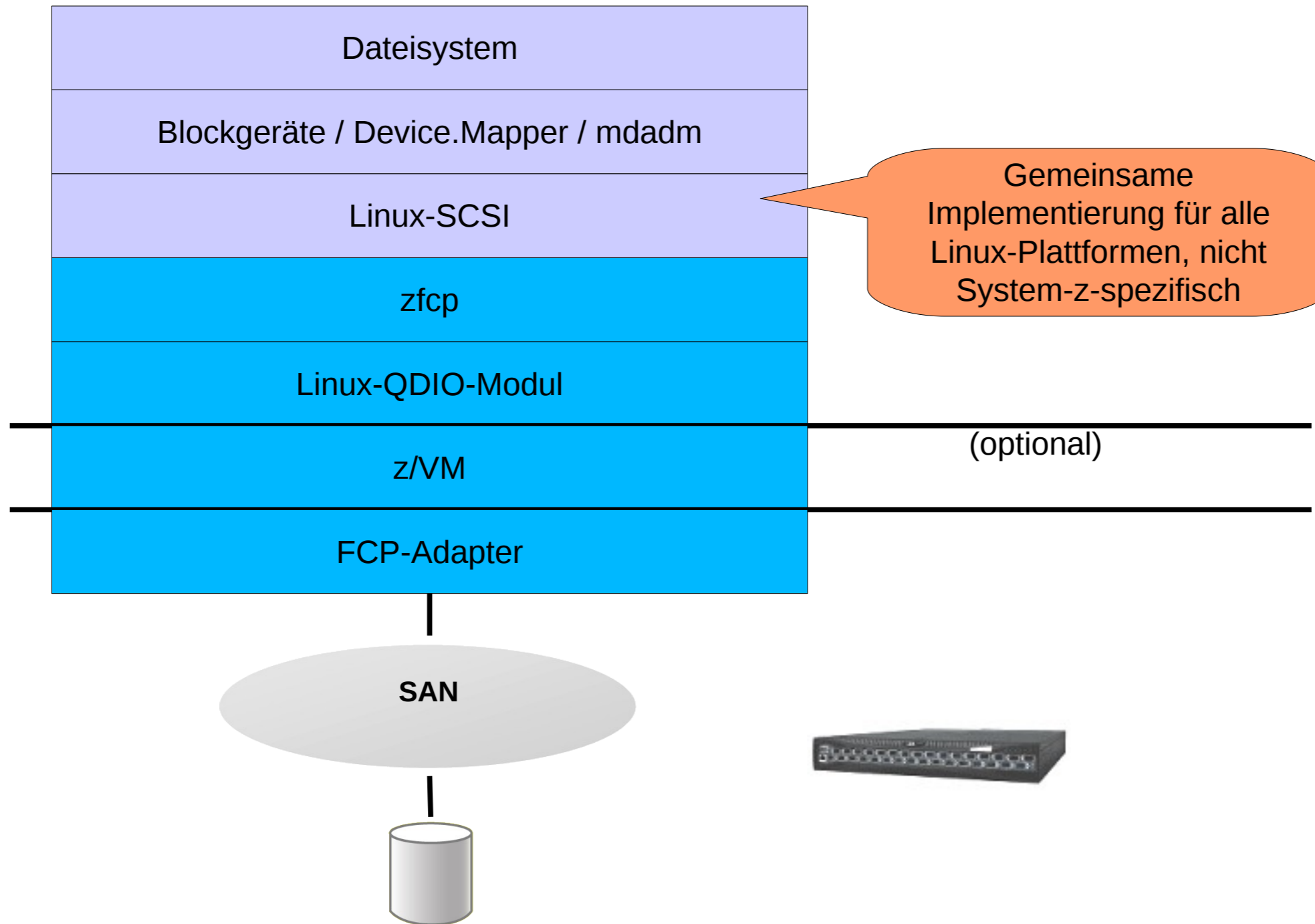


Generelle Empfehlungen

- Support-Matrix für Speichergeräte
 - basiert auf getesteten Umgebungen
 - empfohlene Softwarestände
 - RPQs
- Wartung von Linux
 - initiale Installation mit neuestem Service-Pack/Update
 - Wartungsstrategie



FCP-I/O-Stack für Linux auf System z

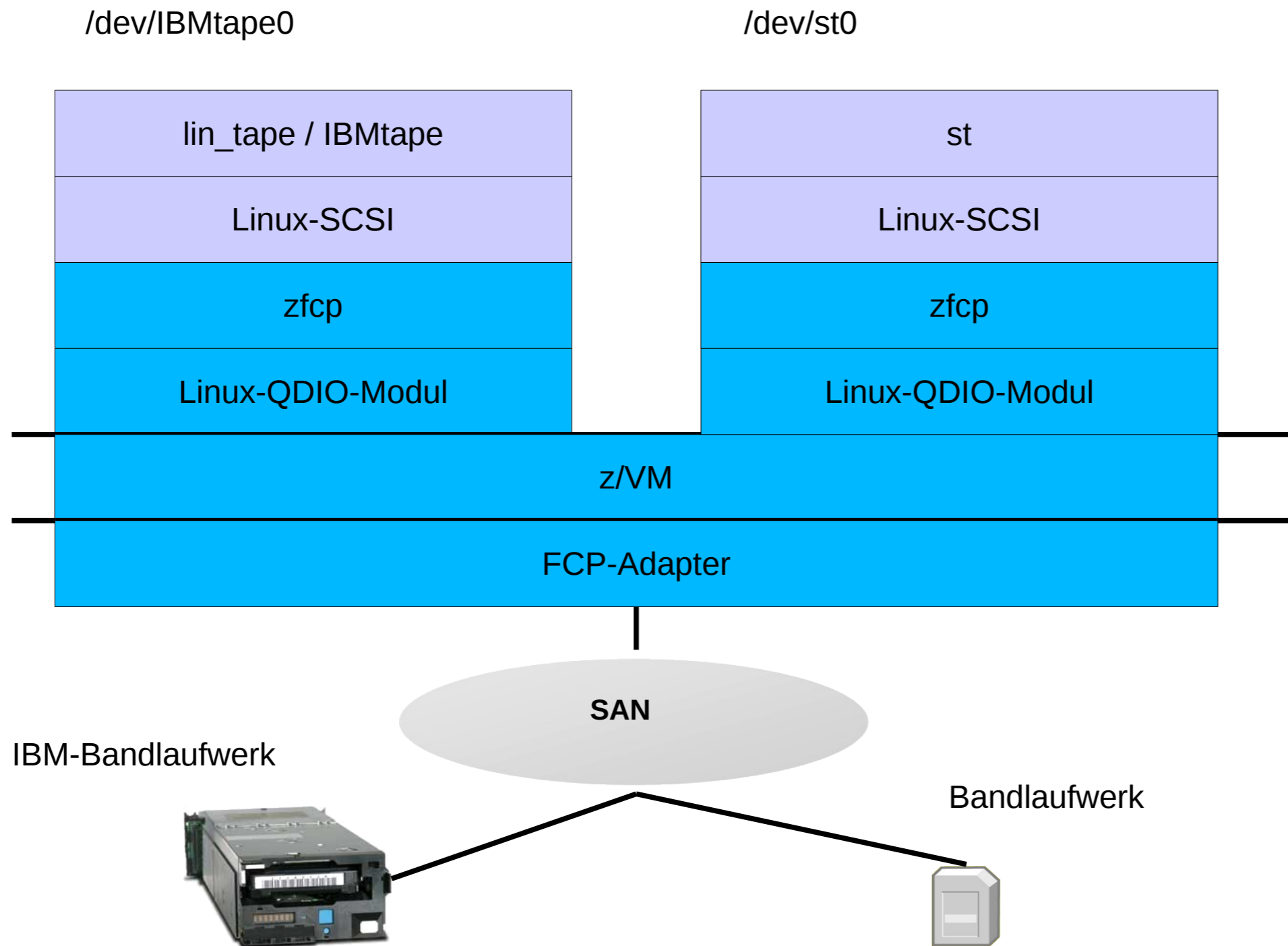


FCP-I/O-Stack für Linux auf System z

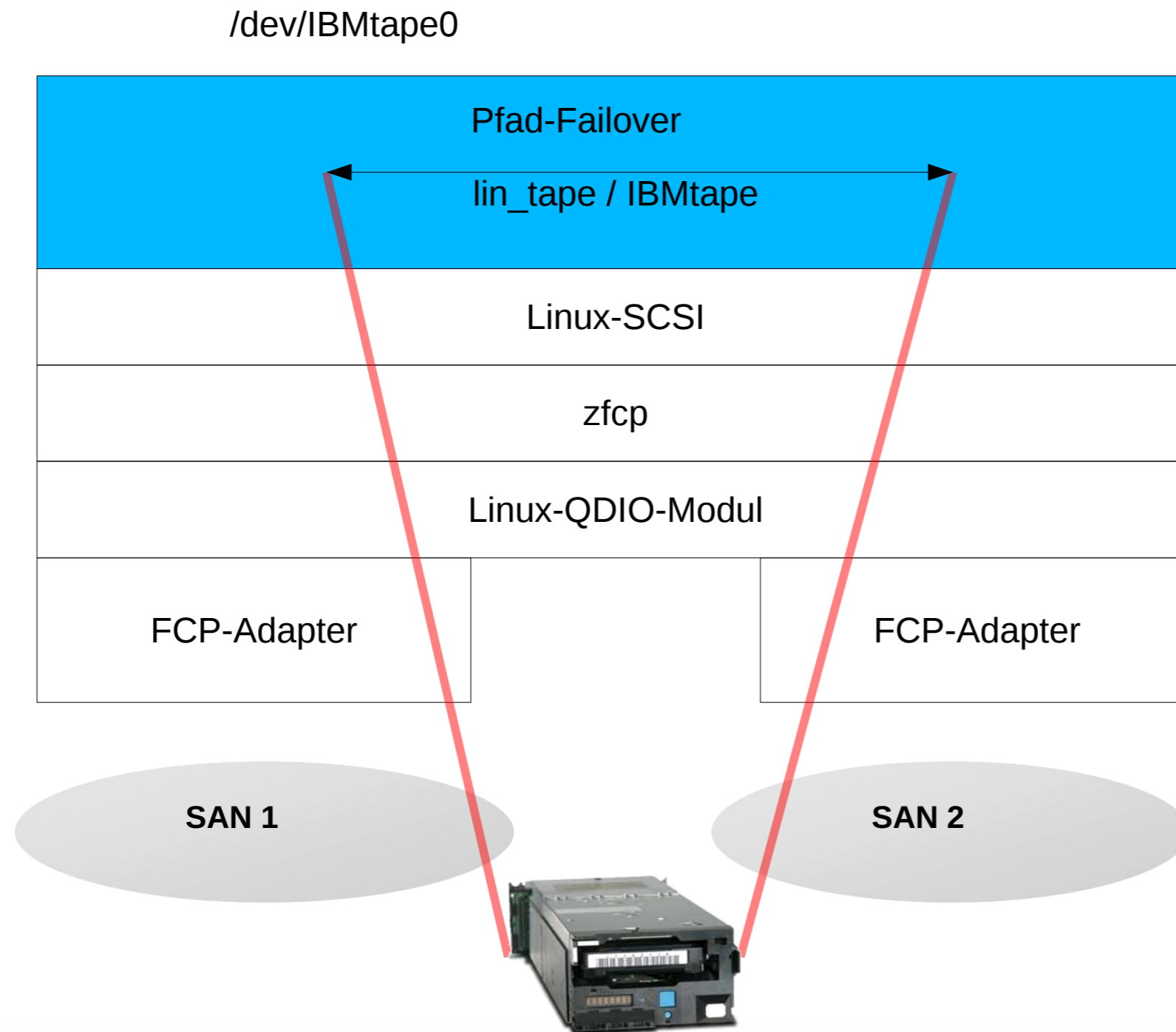
- zfcplib ist der System-z-spezifische “low-level driver”
- Obere Schichten sind für alle Linux-Plattformen gleich
- Gleicher Code und gleiches Setup für
 - redundante Pfade (“multipathing”)
 - gespiegelte Platten (“mirroring”)
 - Dateisysteme



FCP-Bandlaufwerke mit Linux auf System z



Redundante Pfade und Failover für IBM Bandlaufwerke



Redundante Pfade und Failover für IBM Bandlaufwerke

- Funktionalität von IBM-Tape-Treiber
 - lin_tape, alter Name IBMtape
- Service und Support zusammen mit IBM-Bandlaufwerk
- Einrichten
 - Modul-Parameter (z.B. in /etc/modprobe.conf.local)
 - options lin_tape alternate_pathing=1
 - Alle Pfade zum Tape-Drive einrichten
- Kein Failover zu anderen Bandlaufwerken oder Kopien der Daten
 - Aufgabe von Backup- oder Medienverwaltungs-Anwendung



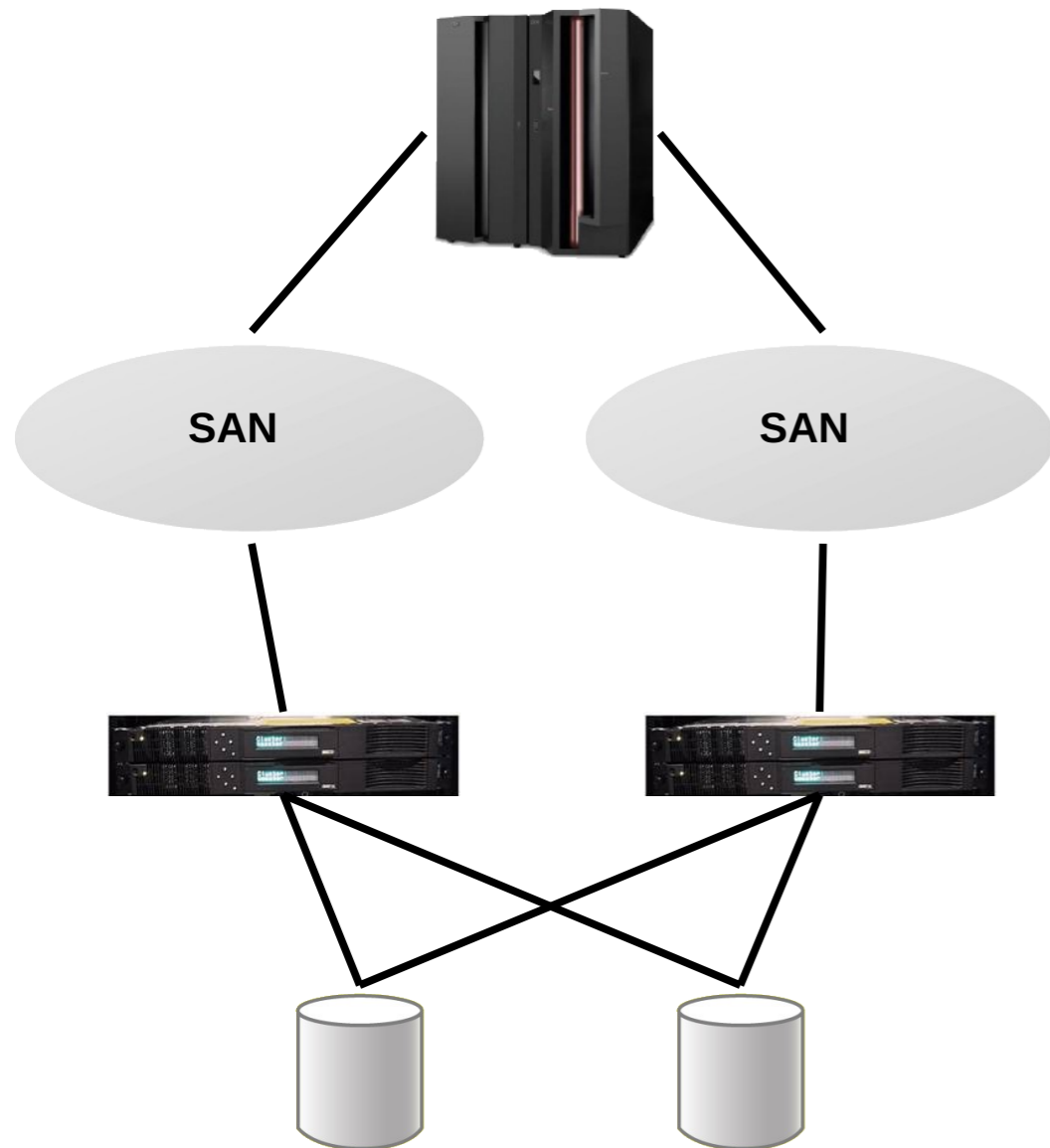
Redundante Pfade für Disks

■ Empfohlene Lösung

- Device-Mapper multipathing (dm-mpath)
 - Teil von Standard-Distributionen (SLES, RHEL, ...)
 - Wird in System-Tests und Device-Qualifikationen getestet
 - unterstützt mehr als 2 Pfade, Beispiele verwenden nur 2

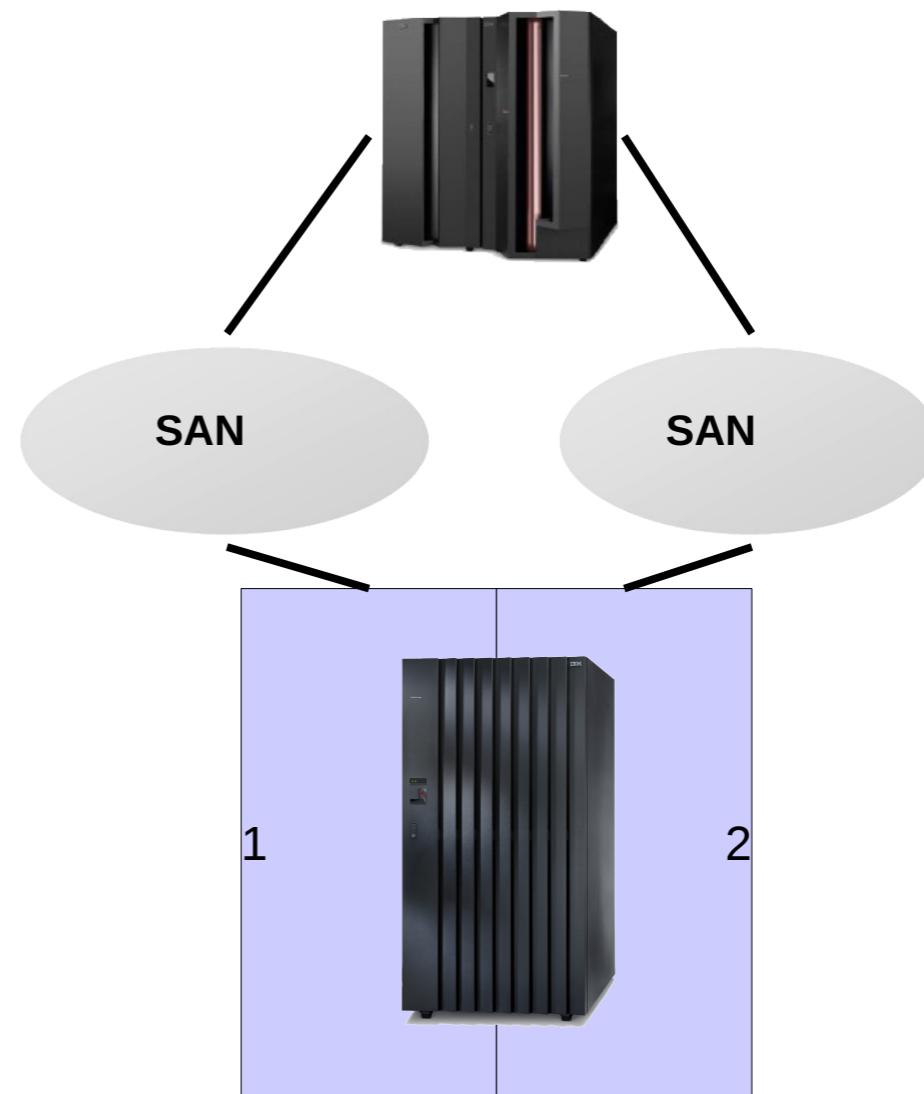


Gründe für "multipathing"



SVC

- 2 oder mehr Knoten
- Jeder Knoten kann offline sein



Storage server, z.B. DS8000

- 2 Controller
- Jeder Controller kann offline sein

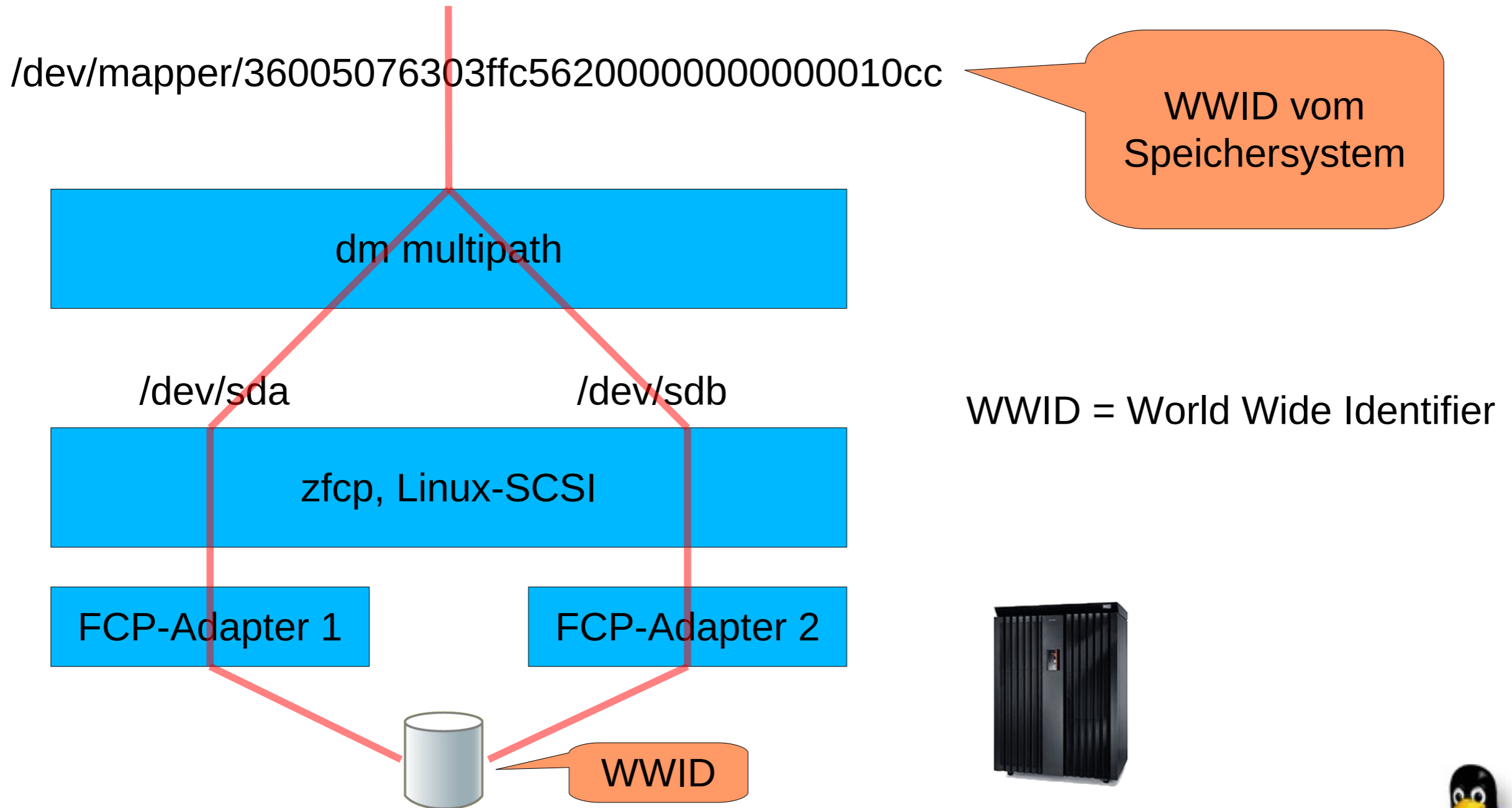


Gründe für “multipathing”

- Redundante Controller in Speichersystemen
- Annahme, dass Clients alle Controller verwenden
- Jeder Controller (oder Knoten) kann offline sein
 - Wartung an Hardware
 - Software-Updates im Controller
 - interne Resets
- Gleicher Ansatz auch für Lastverteilung
 - bevorzugter Controller in DS6000
 - Lastverteilung über alle verfügbaren Pfade



DM-multipathing unter Linux



Linux

DM-multipathing unter Linux

- basiert auf generischer Device-Mapper-Schnittstelle
- multipathd-Daemon für
 - Lesen der Konfiguration
 - Einrichten der Multipath-Geräte
 - Regelmäßiges Überprüfen aller Pfade
 - Wiederaktivieren eines Pfades
- verwendet udev zur automatischen Konfiguration
- multipath als CLI (Kommandozeile)
- kpartx-Modul für Partitionen auf multipath-Geräten
- Pfade werden automatisch erkannt und zugeordnet (WWID)



Einrichten von zfcplib für Multipathing

- Alle Pfade einrichten
- Keine gemeinsamen Komponenten in Pfaden
- Beispiel, besser: Setup von Distributionen verwenden

```
# cd /sys/bus/ccw/drivers/zfcplib/  
# echo 1 > 0.0.3c00/online  
# echo 1 > 0.0.3d00/online  
# echo 0x500507630313c562 > 0.0.3c00/port_add  
# echo 0x500507630303c562 > 0.0.3d00/port_add  
# echo 0x401040cc00000000 > 0.0.3c00/0x500507630313c562/unit_add  
# echo 0x401040cc00000000 > 0.0.3d00/0x500507630303c562/unit_add
```

üblicherweise gleiche
FCP-LUN (siehe
Speichersystem)

Unterschiedliche Adapter und Ports



Einrichten von zfcplib für multipathing

- In zfcplib und SCSI ist jeder Pfad als Device sichtbar
- multipathing ist darüber implementiert, im Blocklayer

```
# ls SCSI
```

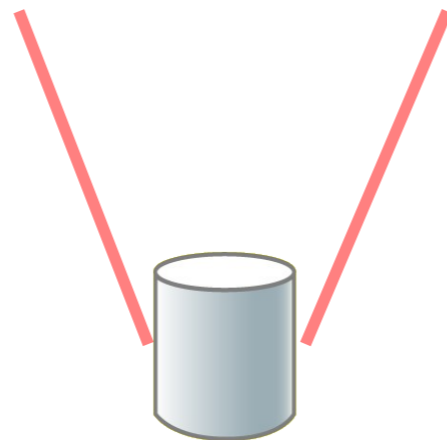
```
[0:0:0:1] disk IBM 2107900 2.27 /dev/sda
```

```
[1:0:0:1] disk IBM 2107900 2.27 /dev/sdb
```

```
# ls zfcplib -D
```

```
0.0.3c00/0x500507630313c562/0x401040cc00000000 0:0:0:1
```

```
0.0.3d00/0x500507630303c562/0x401040cc00000000 1:0:0:1
```



DM-Multipathing-Setup (SLES10)

- Im Normalfall keine Konfiguration notwendig
 - /etc/multipath.conf muss nicht existieren
- Scannen für multipath-Geräte und multipathd aktivieren
 - chkconfig –add boot.multipath
 - chkconfig –add multipathd
- SCSI-Devices für multipathing nicht mounten
- Neustart oder entsprechende Dienste jetzt starten
 - /etc/init.d/boot.multipath start
 - /etc/init.d/multipathd start
- Setup überprüfen
 - multipath -ll
 - ls -l /dev/mapper/



DM-Multipathing-Setup

■ Partitionen

- werden automatisch erkannt
- notwendig? oder besser logische Volumes über LVM?
- falls notwendig: Partitionen ohne Multipathing anlegen

■ multipath devices entfernen

- umount /data
- multipath -F



DM-Multipathing

WWID

```
# multipath -ll
```

```
36005076303ffc562000000000000010cf dm-0
```

```
IBM,2107900
```

```
[size=5.0G][features=1 queue_if_no_path][hwhandler=0]
```

```
\_ round-robin 0 [prio=2][active]
```

```
\_ 1:0:0:1087324176 sdb 8:16 [active][ready]
```

```
\_ 0:0:0:1087324176 sda 8:0 [active][ready]
```

pathgroup

- Keine Konfigurations-Datei notwendig
- Aber: Defaults können bei Bedarf überschrieben werden
- multipath -t zeigt aktuell verwendete Hardware-Tabelle
 - enthält defaults und Konfiguration



DM-Multipathing (RHEL5)

- alle Pfade konfigurieren
 - /etc/zfcp.conf
 - /sbin/zfcpconf.sh oder Neustart
- /etc/multipath.conf anpassen:

```
#blacklist {  
#     devnode "*"   
#}  
  
#defaults {  
#     user_friendly_names yes  
#}
```
- chkconfig --add multipathd
- /etc/init.d/multipathd start



DM multipathing setup (RHEL5)

- `user_friendly_names` und Aliase
 - `/dev/mapper/mpath0` anstelle von `/dev/mapper/36005076303ffc562000000000000010ce`
- Aber:
 - WWID ist eindeutig, Alias eventuell nicht
 - Zuordnung wird durch Konfiguration bestimmt
- Empfehlung: Eindeutige WWIDs verwenden

`/dev/mapper/36005076303ffc562000000000000010*`

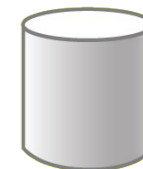
...cc



...cd



...ce



DM-Multipathing mit bevorzugtem Pfad

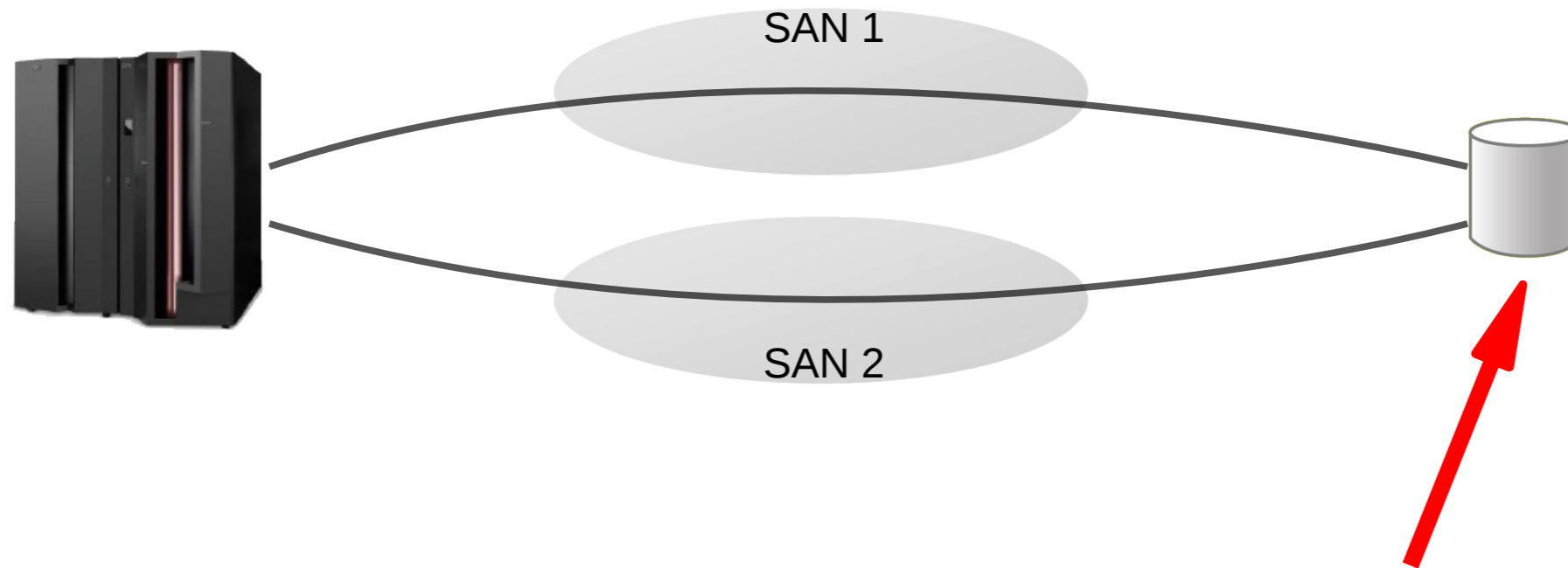


- active / passive controller (DS6000)
- Konfiguration für übliche Speichersysteme in Standard-Hardware-Tabelle
- 2 pathgroups
 - active/enabled
 - Information stammt vom Speichersystem (ALUA)

```
# multipath -ll
3600507630effca200000000000001229 dm-0 IBM,1750500
[size=3.0G][features=1 queue_if_no_path][hwhandler=0]
\_ round-robin 0 [prio=50][active]
  \_ 0:0:0:1076445202 sdaw 67:0 [active][ready]
\_ round-robin 0 [prio=10][enabled]
  \_ 1:0:0:1076445202 sdcb 68:240 [active][ready]
```



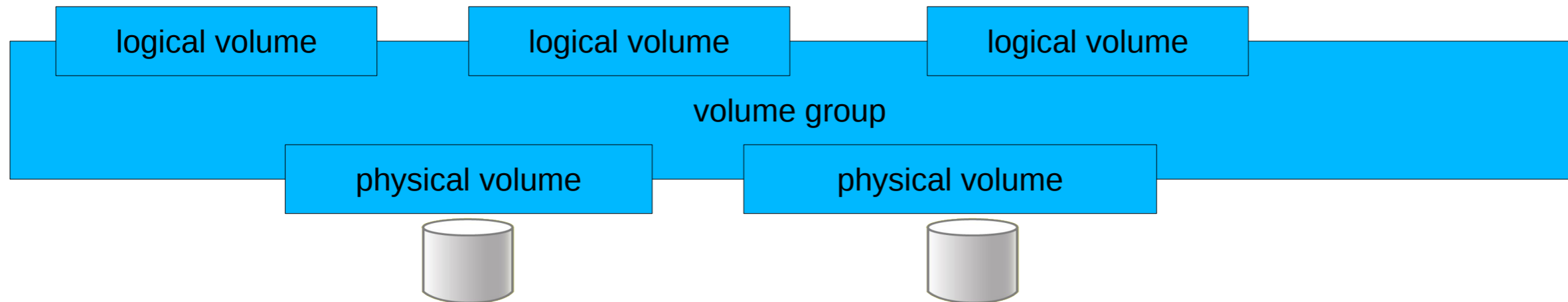
Spiegeln von Daten (“mirroring”)



- Ausfall eines Plattenspeichersystems?
- LVM / device-mapper mirroring
- Spiegel with mdadm

LVM / device mapper mirroring

- LVM (Logical Volume Manager)
 - Generisches System, aufbauend auf Device-Mapper

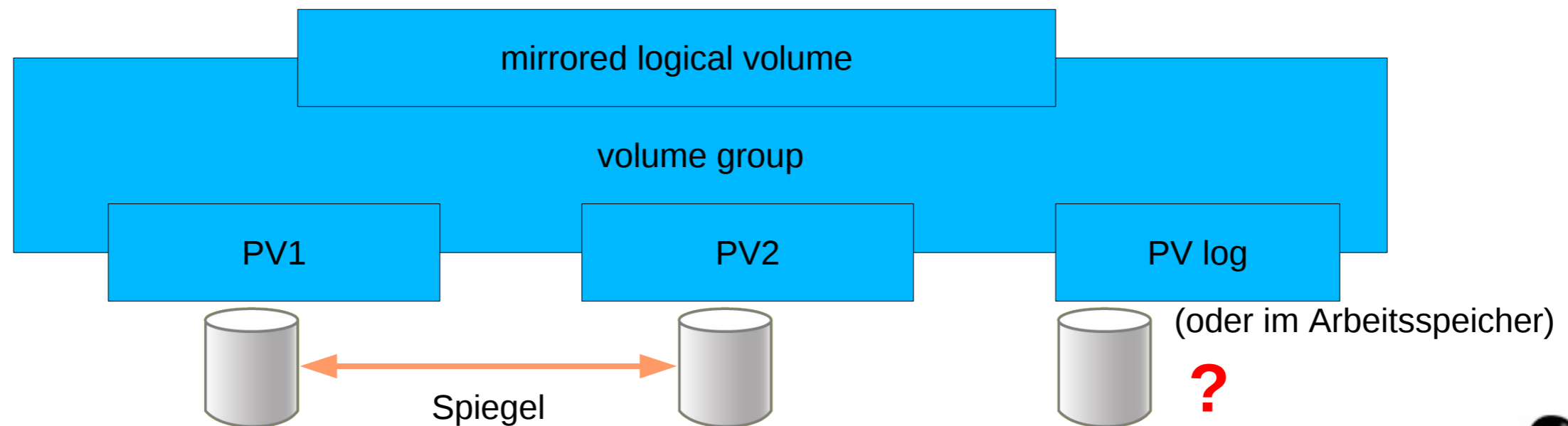


- “volume group” fasst “physical volumes” zusammen
- “logical volumes” verwenden Speicherplatz aus “volume group”
- Möglichkeiten: striping, mirroring, snapshots, ...



LVM / device mapper mirroring

- Spiegel mit LVM verwendet ein externes Log
 - Log enthält Status von Spiegel
 - --corelog hält Log im Speicher
 - erfordert kompletten resync nach Neustart
 - Alternative: zusätzliche Platte für Log
 - “single point of failure”



LVM / device mapper mirroring

```
# pvcreate /dev/mapper/36005076303ffc56200000000000010cd  
Physical volume "/dev/mapper/36005076303ffc56200000000000010cd"  
successfully created
```

```
# pvcreate /dev/mapper/36005076303ffc56200000000000010ce  
Physical volume "/dev/mapper/36005076303ffc56200000000000010ce"  
successfully created
```

```
# vgcreate vg0 /dev/mapper/36005076303ffc56200000000000010cd /dev/mapper/  
36005076303ffc56200000000000010ce  
Volume group "vg0" successfully created
```

```
# lvcreate vg0 -L1G -m1 --corelog  
Logical volume "lvol0" created
```

Log im Arbeitsspeicher

```
# mke2fs -j /dev/vg0/lvol0
```



LVM / device mapper mirroring

- `--corelog` erfordert `resync` nach jedem Neustart

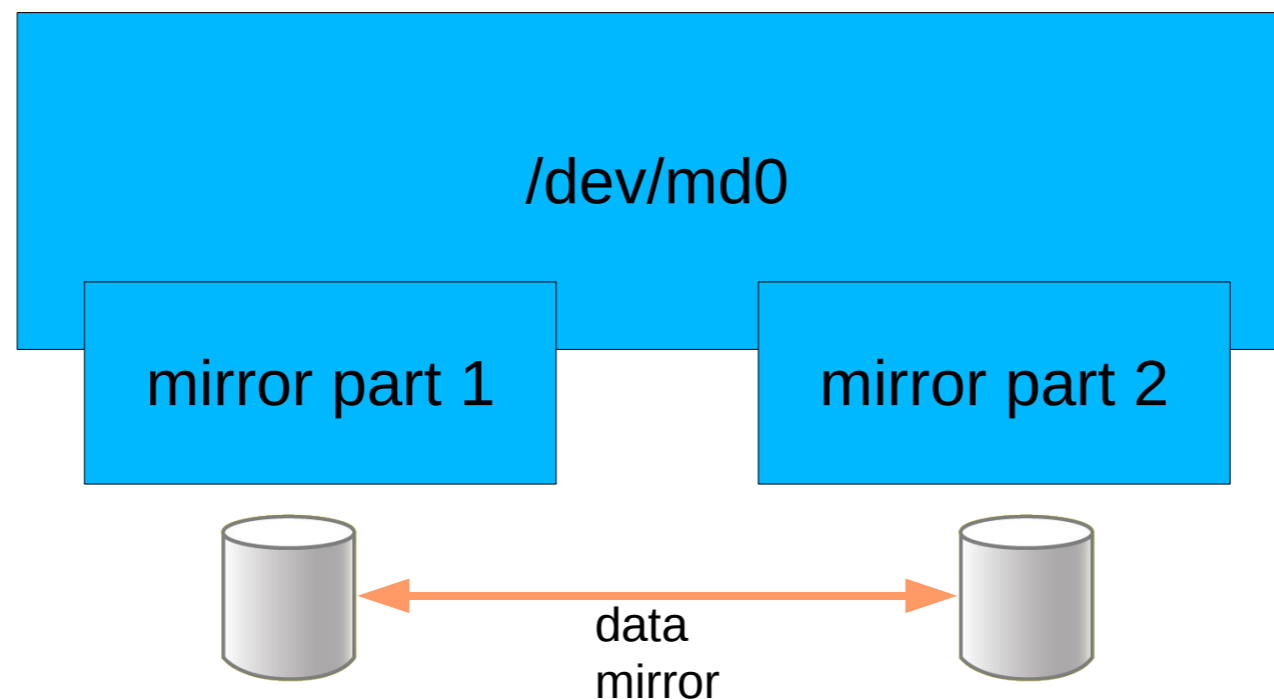
```
# lvs
LV      VG      Attr      LSize Origin Snap%   Move Log Copy%
lvol0  vg0     mwi-a-    1.00G                73.83
```

```
# vmstat 1
procs -----memory----- ---swap-- -----io----- -system-- -----cpu-----
 r  b    swpd    free    buff    cache    si    so     bi     bo    in    cs  us  sy  id  wa  st
 0  0        0 309180    7008    85356     0     0    2067   1874   526   197   3   2  65  30   0
 0  0        0 309180    7008    85356     0     0    1280   1280   119   264   0   0 100   0   0
 0  0        0 309180    7008    85356     0     0         0         0    94   230   0   0 100   0   0
 1  0        0 309180    7008    85356     0     0    1536   1536   113   272   0   0  99   0   0
 0  0        0 309180    7008    85356     0     0    1536   1536   108   265   0   0 100   0   0
 0  0        0 309180    7016    85348     0     0    6400   6420   186   438   0   0 100   0   0
 1  0        0 309180    7016    85348     0     0    5632   5632  2081   407   0   1  99   0   0
```



Mirroring mit mdadm

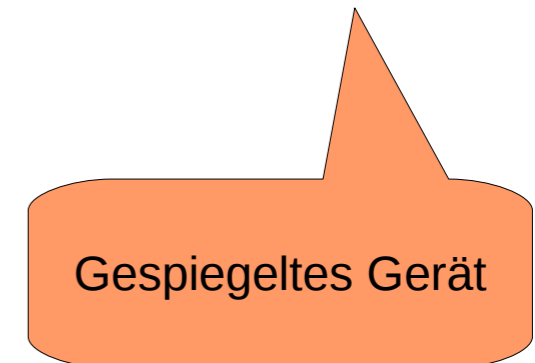
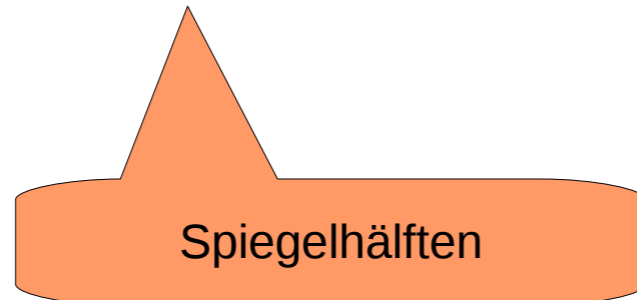
- Spiegelt Block-Devices, aber verwendet nicht den Device-Mapper
- Device-Namen: /dev/md0 anstelle von /dev/mapper/*
- kein externes Log
- ist generell gegenüber dem LVM-Mirror zu bevorzugen



mdadm device mirroring

■ initial setup

```
# mdadm --create --level=1 --raid-devices=2 --spare-devices=1 /dev/md0  
/dev/mapper/36005076303ffc56200000000000010c*
```



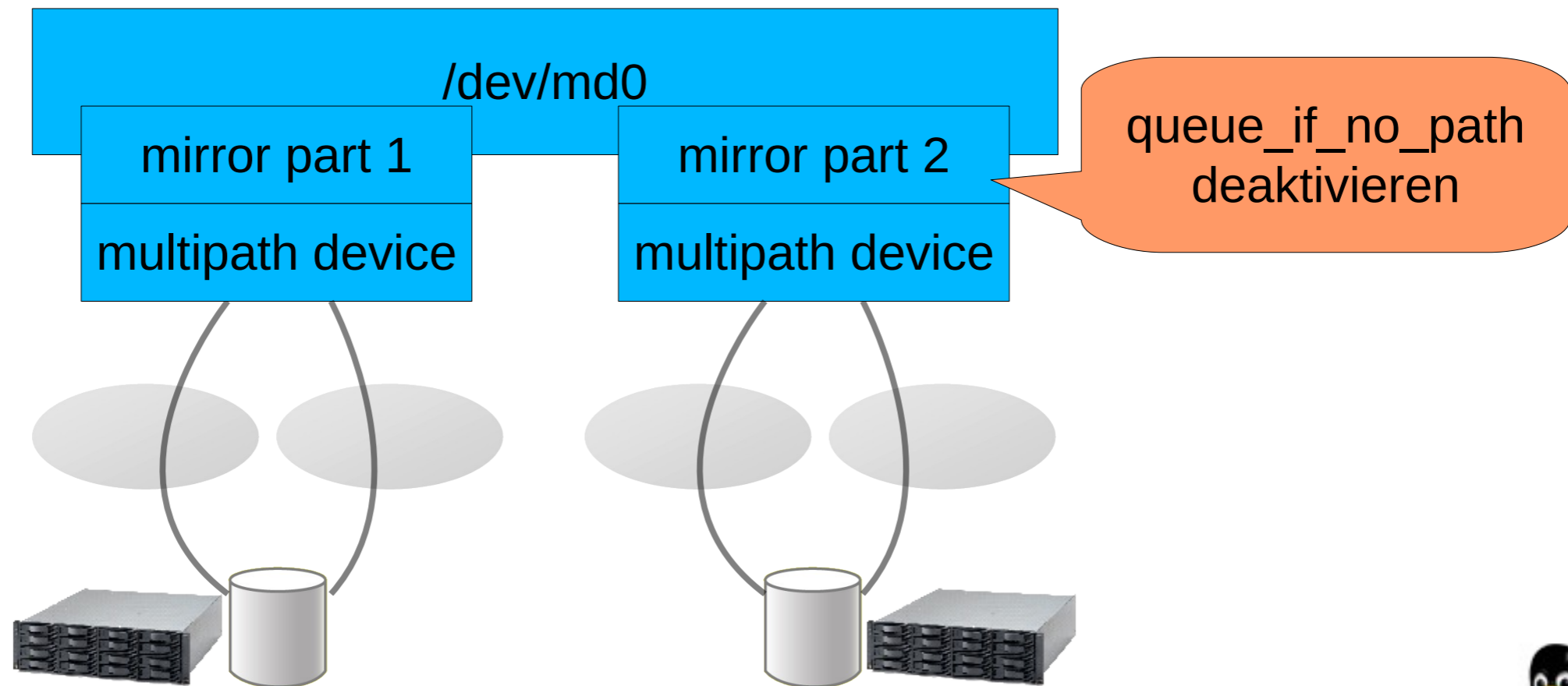
■ resync wird automatisch bei Bedarf gestartet

```
# cat /proc/mdstat  
Personalities : [raid1]  
md0 : active raid1 dm-2[2](S) dm-1[1] dm-0[0]  
5242816 blocks [2/2] [UU]  
[=====>.....] resync = 33.1% (1737024/5242816) finish=1.3min  
speed=44816K/sec
```



MD-mirroring auf DM-multipath

- Auch mit Datenspiegel wird Multipathing benötigt
- `queue_if_no_path` muss deaktiviert werden, damit Geräte-Fehler für den Datenspiegel sichtbar werden



MD-Mirroring auf DM-Multipath

■ queue_if_no_path deaktivieren

```
# multipath -ll
36005076303ffc56200000000000010cd dm-0 IBM,2107900
[size=5.0G][features=1 queue_if_no_path][hwhandler=0]
\_ round-robin 0 [prio=2][active]
  \_ 1:0:0:1087193104 sdd  8:48 [active][ready]
  \_ 0:0:0:1087193104 sdc  8:32 [active][ready]
```

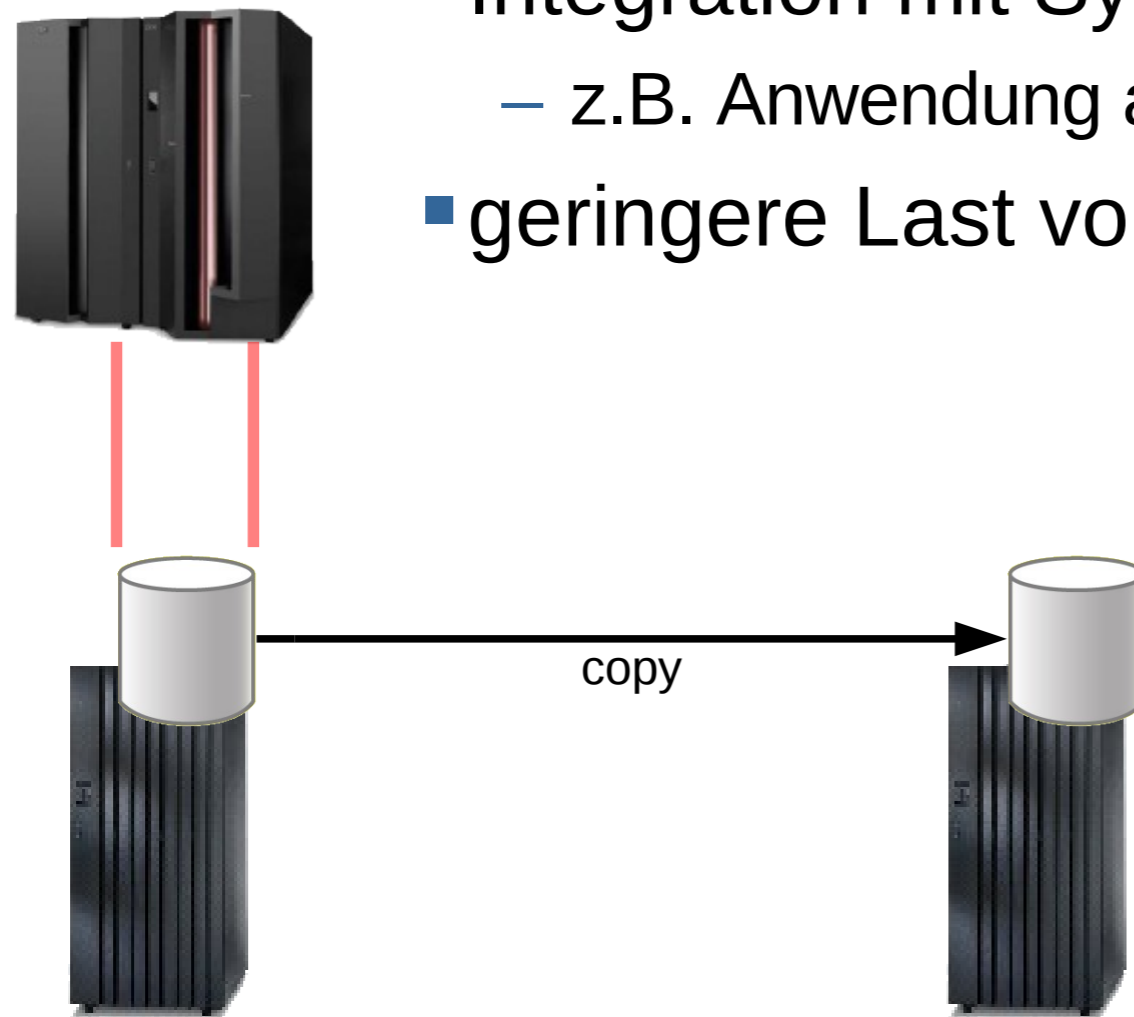
```
/etc/multipath.conf
defaults {
    no_path_retry fail
}
```

```
/etc/init.d/multipathd restart
# multipath -ll
36005076303ffc56200000000000010cd dm-0 IBM,2107900
[size=5.0G][features=0][hwhandler=0]
\_ round-robin 0 [prio=2][active]
  \_ 1:0:0:1087193104 sdd  8:48 [active][ready]
  \_ 0:0:0:1087193104 sdc  8:32 [active][ready]
```



Alternativen zu Mirroring in Linux

- “Remote copy” im Plattenspeichersystem
- Kopiert Daten unabhängig von Linux
- Integration mit System und Anwendungen
 - z.B. Anwendung anhalten, unmount, sync, ...
- geringere Last von Linux-Systemen



root-Dateisystem auf Multipath-Gerät



- Voraussetzung für root-Dateisystem auf SCSI
- Multipathing wird schon in initrd zum Booten benötigt
- Probleme:
 - Keine Option in Installations-Programmen, um root auf multipath-Gerät zu installieren
 - Installation auf single-path und späteres Anpassen
 - zipl unterstützt keine multipath-Geräte
 - Zusätzliche Platte für /boot (SCSI oder DASD)
 - Booten in root auf single-path für mkinitrd und zipl (?)



Installation von SLES10 auf multipath-Gerät

```

cp q loaddev
PORTNAME 50050763 0310C562 LUN 401040CD 00000000 BOOTPROG 0
BR_LBA 00000000 00000000

cp ip1 3c00
00: HCPLDI2816I Acquiring the machine loader from the processor controller.
00: HCPLDI2817I Load completed from the processor controller.
00: HCPLDI2817I Now starting the machine loader.
01: HCPGSP2630I The virtual machine is placed in CP mode due to a SIGP stop and
store status from CPU 00.
00: MLDEVLO12I: Machine loader up and running (version 0.18).
00: MLOPDM003I: Machine loader finished, moving data to final storage location.
Linux version 2.6.16.60-0.9-default (geeko@buildhost) (gcc version 4.1.2 2007011
5 (SUSE Linux)) #1 SMP Mon Mar 17 17:16:31 UTC 2008
We are running under VM (64 bit mode)
Detected 2 CPU's
Boot cpu address 0
Built 1 zonelists
Kernel command line: root=/dev/mapper/36005076303ffc56200000000000010cc-part1
TERM=dumb

```

/boot, wird von zipl verwendet

root-Dateisystem auf multipath-Gerät

```

Setup multipath devices: ok.
Waiting for device /dev/mapper/36005076303ffc56200000000000010cc-part1 to appear
: ok
rootfs: major=253 minor=1 devn=64769
fsck 1.38 (30-Jun-2005)
[/bin/fsck.ext3 (1) -- /] fsck.ext3 -a /dev/mapper/36005076303ffc562000000000000
10cc-part1
/dev/mapper/36005076303ffc56200000000000010cc-part1: clean, 91021/525888 files,
550917/1050241 blocks
fsck succeeded. Mounting root device read-write.
Mounting root /dev/mapper/36005076303ffc56200000000000010cc-part1

```



Zusammenfassung

- Failover-Anforderungen unterscheiden sich von System zu System
 - genaue Analyse und Planungen notwendig
- Multipathing ist eine wichtige Voraussetzung für sicheren Einsatz von FCP-Storage
- Datenspiegeln ist sowohl in Linux, als auch in Plattenspeichersystem möglich
- root-Dateisystem auf Multipath-Gerät erfordert zusätzlichen Aufwand



Weitere Informationen

- Device-mapper Resource Page (link to Multipath bug tracking)
<http://sources.redhat.com/dm/>
- Device-mapper and LVM2 Wiki <http://sources.redhat.com/lvm2/wiki/MultipathUsageGuide>
- The Software-RAID HOWTO (mdadm) <http://tldp.org/HOWTO/Software-RAID-HOWTO.html>
- How to setup / use multipathing on SLES
http://support.novell.com/techcenter/sdb/en/2005/04/sles_multipathing.html
- Redhat: LVM Administrator's Guide
http://www.redhat.com/docs/manuals/enterprise/RHEL-5-manual/Cluster_Logical_Volume_Manager/
- Redhat: Using Device-Mapper Multipath
http://www.redhat.com/docs/manuals/enterprise/RHEL-5-manual/en-US/RHEL510/DM_Multipath/
- multipath tools FAQ <http://git.kernel.org/?p=linux/storage/multipath-tools/.git;a=blob;f=FAQ>
- Enabling root-on-multipath for SLES9 on zSeries
<http://linuxvm.org/Info/HOWTOs/root-on-multipath.html>
- How to use FC-attached SCSI devices with Linux on System z – SC33-8291-03
http://www.ibm.com/developerworks/linux/linux390/october2005_documentation.html
- LVM Howto <http://tldp.org/HOWTO/LVM-HOWTO/>
- IBMtape/lin_tape driver and documentation <ftp://ftp.software.ibm.com/storage/devdvr/>



Weitere Informationen

- IBM DS6000 Interoperability Matrix
<http://www.ibm.com/systems/storage/disk/ds6000/pdf/interop.pdf>
- IBM DS8000 Interoperability Matrix
<http://www.ibm.com/systems/storage/disk/ds8000/interop.pdf>
- IBM SAN Volume Controller Interoperability
<http://www.ibm.com/systems/storage/software/virtualization/svc/interop.html>
- System Storage Interoperation Center (SSIC)
<http://www.ibm.com/systems/support/storage/config/ssic/index.jsp>



Vielen Dank!



Beispiel: SLES10-Installation auf multipath-root-Dateisystem

- 2 Platten für / and /boot

```

schmichr@schmichrtp:~
File Edit View Terminal Tabs Help
YaST @ 9.152.37.202 Press F1 for Help

Configured ZFCP Disks
In this dialog, manage
ZFCP disks on your
system.

To configure a new
ZFCP disk, click Add.

To remove a configured
ZFCP disk, select it
and click Delete.

Warning

When accessing a ZFCP
device READ/WRITE,
make sure that this
access is exclusive.
Otherwise there is a
potential risk of data

Configured ZFCP Disks
Minimum Channel      Maximum Channel
0x0000               0xffff [Filter]

Channel Number |          WPN          | zfcpl-LUN
0.0.3c00 | 0x500507630310c562 | 0x401040cc00000000
0.0.3c00 | 0x500507630310c562 | 0x401040cd00000000

[Add] [Delete]
[Cancel] [Next]

```



Beispiel: SLES10-Installation auf multipath-root-Dateisystem

- /, /boot and swap

```

schmichr@schmichrtp:~
File Edit View Terminal Tabs Help
YaST @ 9.152.37.202 Press F1 for Help

Expert Partitioner

For a root file system
on SCSI disks, add a
/boot partition on
DASD to use for IPL.

The table to the right
shows the current
partitions on all your
hard disks. Nothing
will be written to
your hard disk until
you confirm the entire
installation in the
last installation
dialog. Until that
point, you can safely
abort the
installation.

Hard disks are
designated like this:

Device  ID  Size  F  Type  Mount  Mount By
-----  --  ---  -  ---  ---    ---
/dev/sda  ()  5.0 GB  F  IBM-2107900
/dev/sda1  4.0 GB  F  Linux native (Ext3)  /  I
/dev/sda2  1011.9 MB  F  Linux swap  swap  I
/dev/sdb  ()  5.0 GB  F  IBM-2107900
/dev/sdb1  4.9 GB  F  Linux native (Ext3)  /boot  I

[ Create ] [ Edit ] [ Delete ] [ dasdfmt ]
[ LVM... ] [ EVMS... ] [ RAID...v ] [ NFS... ] [ Expert..v ]
[ Back ] [ Abort ] [ Finish ]

```



Beispiel: SLES10-Installation auf multipath-root-Dateisystem

■ Initialer IPL über /boot-Platte

```

x3270-4 t6360008
File Options
00:
00: CP SET LOADDEV PORTNAME 50050763 0310C562 LUN 401040CD 00000000
00:
00: CP IPL 3C00
00: HCPLDI2816I Acquiring the machine loader from the processor controller.
00: HCPLDI2817I Load completed from the processor controller.
00: HCPLDI2817I Now starting the machine loader.
01: HCPGSP2630I The virtual machine is placed in CP mode due to a SIGP stop and
store status from CPU 00.
00: MLOEVL012I: Machine loader up and running (version 0,18).
00: MLOPDM003I: Machine loader finished, moving data to final storage location.
Linux version 2.6.16.60-0.9-default (geeko@buildhost) (gcc version 4.1.2 2007011
5 (SUSE Linux)) #1 SMP Mon Mar 17 17:16:31 UTC 2008
We are running under VM (64 bit mode)
Detected 2 CPU's
Boot cpu address 0
Built 1 zonelists
Kernel command line: root=/dev/disk/by-id/scsi-36005076303ffc562000000000000010cc
-part1 TERM=dumb

```

set loaddev: Port, LUN

IPL: FCP adapter



Beispiel: SLES10-Installation auf multipath-root-Dateisystem

- Nach Installation ist root-Dateisystem nur single-path
- dedizierte Platte für /boot

```
# mount
```

```
/dev/sda1 on / type ext3 (rw,acl,user_xattr)
```

```
/dev/sdb1 on /boot type ext3 (rw,acl,user_xattr)
```

```
[...]
```

```
# lsscsi
```

```
[0:0:0:1087127568]disk  IBM    2107900    2.27 /dev/sda
```

```
[0:0:0:1087193104]disk  IBM    2107900    2.27 /dev/sdb
```

```
# lszfcp -D
```

```
0.0.3c00/0x500507630310c562/0x401040cc00000000 0:0:0:1087127568
```

```
0.0.3c00/0x500507630310c562/0x401040cd00000000 0:0:0:1087193104
```



Beispiel: SLES10-Installation auf multipath-root-Dateisystem

- Zusätzlicher Pfad für root-Dateisystem

/etc/sysconfig/hardware/hwcfg-zfcp-bus-ccw-0.0.3d00 anlegen:
(oder zweiten Pfad über YaST anlegen)

```
[...]  
ZFCP_LUNS="  
0x500507630310c562:0x401040cc00000000"
```

Zweiten Pfad über hwup Skripte aktivieren (oder reboot)

```
# chccwdev -d 3d00  
Setting device 0.0.3d00 offline  
Done  
# modprobe vmcp  
# vmcp det 3d00  
FCP 3D00 DETACHED  
# vmcp att 3d00 \  
FCP 3D00 ATTACHED TO T6360008 3D00
```



Beispiel: SLES10-Installation auf multipath-root-Dateisystem

- multipathing for root, aber nicht für /boot

```
# lszfcp -D
0.0.3c00/0x500507630310c562/0x401040cc00000000 0:0:0:1087127568
0.0.3c00/0x500507630310c562/0x401040cd00000000 0:0:0:1087193104
0.0.3d00/0x500507630310c562/0x401040cc00000000 2:0:0:1087127568
```

```
# lsscsi
[0:0:0:1087127568]disk  IBM      2107900      2.27 /dev/sda
[0:0:0:1087193104]disk  IBM      2107900      2.27 /dev/sdb
[2:0:0:1087127568]disk  IBM      2107900      2.27 /dev/sdc
```

Query wwid for /boot device

```
# scsi_id -g -u -s /block/sdb
36005076303ffc562000000000000010cd
```

blacklist device for /boot in /etc/multipath.conf

```
blacklist {
    wwid 36005076303ffc562000000000000010cd
}
```



Beispiel: SLES10-Installation auf multipath-root-Dateisystem

- Multipathing für nächsten Neustart aktivieren

```
# chkconfig --add boot.multipath
boot.multipath      0:off 1:off 2:off 3:off 4:off 5:off 6:off B:on
# chkconfig --add multipathd
multipathd          0:off 1:off 2:off 3:on  4:off 5:on  6:off
```

- System auf multipath-Geräte umstellen

root und swap in /etc/fstab (aber /boot nicht verändern!)

```
/dev/mapper/36005076303ffc56200000000000010cc-part1 /           [...]
/dev/mapper/36005076303ffc56200000000000010cc-part2 swap        [...]
/dev/disk/by-id/scsi-36005076303ffc56200000000000010cd-part1 /boot  [...]
```

Kernel-Parameter in /etc/zipl.conf

```
parameters = "root=/dev/mapper/36005076303ffc56200000000000010cc-part1
TERM=dumb"
[...]
parameters = "root=/dev/mapper/36005076303ffc56200000000000010cc-part1
TERM=dumb 3"
```



Beispiel: SLES10-Installation auf multipath-root-Dateisystem

- Boot-Vorgang auf multipath-Geräte umstellen

initrd mit multipath-tools neu anlegen

```
# mkinitrd -f mpath
```

```
# zipl
```

```
# shutdown -r now
```

```
# multipath -ll
```

```
36005076303ffc56200000000000010cc dm-0 IBM,2107900
```

```
[size=5.0G][features=1 queue_if_no_path][hwhandler=0]
```

```
\_ round-robin 0 [prio=2][active]
```

```
\_ 1:0:0:1087127568 sdc 8:32 [active][ready]
```

```
\_ 0:0:0:1087127568 sda 8:0 [active][ready]
```

```
# mount
```

```
/dev/mapper/36005076303ffc56200000000000010cc-part1 on / type ext3
```

```
(rw,acl,user_xattr)
```

```
[...]
```



Trademarks

The following are trademarks of the International Business Machines Corporation in the United States, other countries, or both.

Not all common law marks used by IBM are listed on this page. Failure of a mark to appear does not mean that IBM does not use the mark nor does it mean that the product is not actively marketed or is not significant within its relevant market.

Those trademarks followed by ® are registered trademarks of IBM in the United States; all others are trademarks or common law marks of IBM in the United States.

For a complete list of IBM Trademarks, see www.ibm.com/legal/copytrade.shtml:

*, AS/400®, e business (logo)®, DBE, ESCO, eServer, FICON, IBM®, IBM (logo)®, iSeries®, MVS, OS/390®, pSeries®, RS/6000®, S/30, VM/ESA®, VSE/ESA, WebSphere®, xSeries®, z/OS®, zSeries®, z/VM®, System i, System i5, System p, System p5, System x, System z, System z9®, BladeCenter®

The following are trademarks or registered trademarks of other companies.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.

Cell Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license therefrom.

Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

ITIL is a registered trademark, and a registered community trademark of the Office of Government Commerce, and is registered in the U.S. Patent and Trademark Office.

IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency, which is now part of the Office of Government Commerce.

* All other products may be trademarks or registered trademarks of their respective companies.

Notes:

Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance ratios stated here.

IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.

All customer examples cited or described in this presentation are presented as illustrations of the manner in which some customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual customer configurations and conditions.

This publication was produced in the United States. IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. Consult your local IBM business contact for information on the product or services available in your area.

All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the performance, compatibility, or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

Prices subject to change without notice. Contact your IBM representative or Business Partner for the most current pricing in your geography.

