

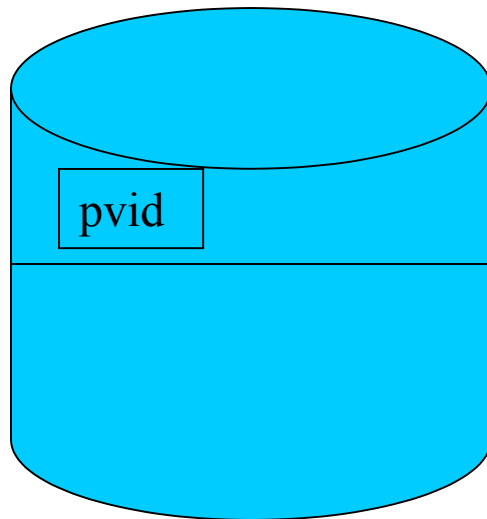
Tips & Tricks for Using LVM Effectively / Intro to VxVM

Renay Gaye
Hewlett-Packard
renay_gaye@hp.com

LVM Concepts

LVM Concepts-Physical Volume

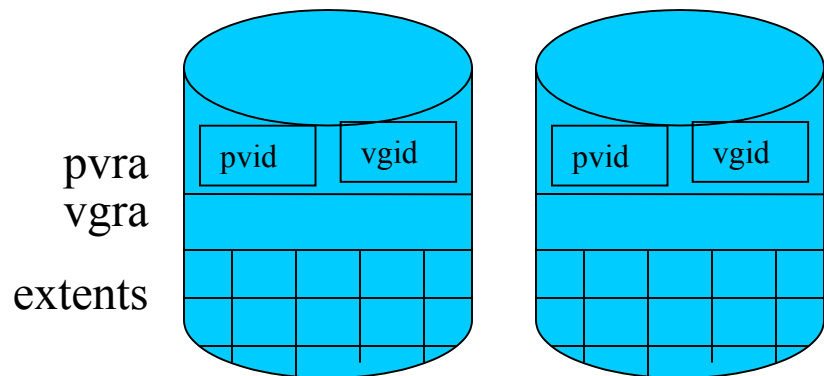
- LVM Managed Disk
- Each PV is assigned a unique PVID



`pvcreate /dev/rdisk/cntndn`

LVM Concepts-Volume Group

- One or more physical volumes
- pool of physical extents

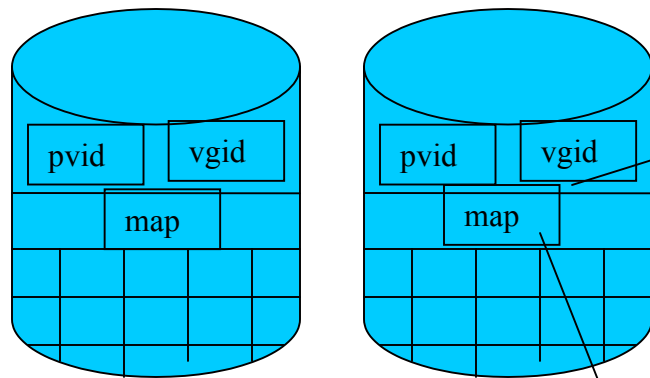


```
mkdir /dev/vgname
```

```
mknod /dev/vgname/group c 64  
0xnn0000
```

```
vgcreate vgname /dev/dsk/cntndn ...
```

Volume Group Map



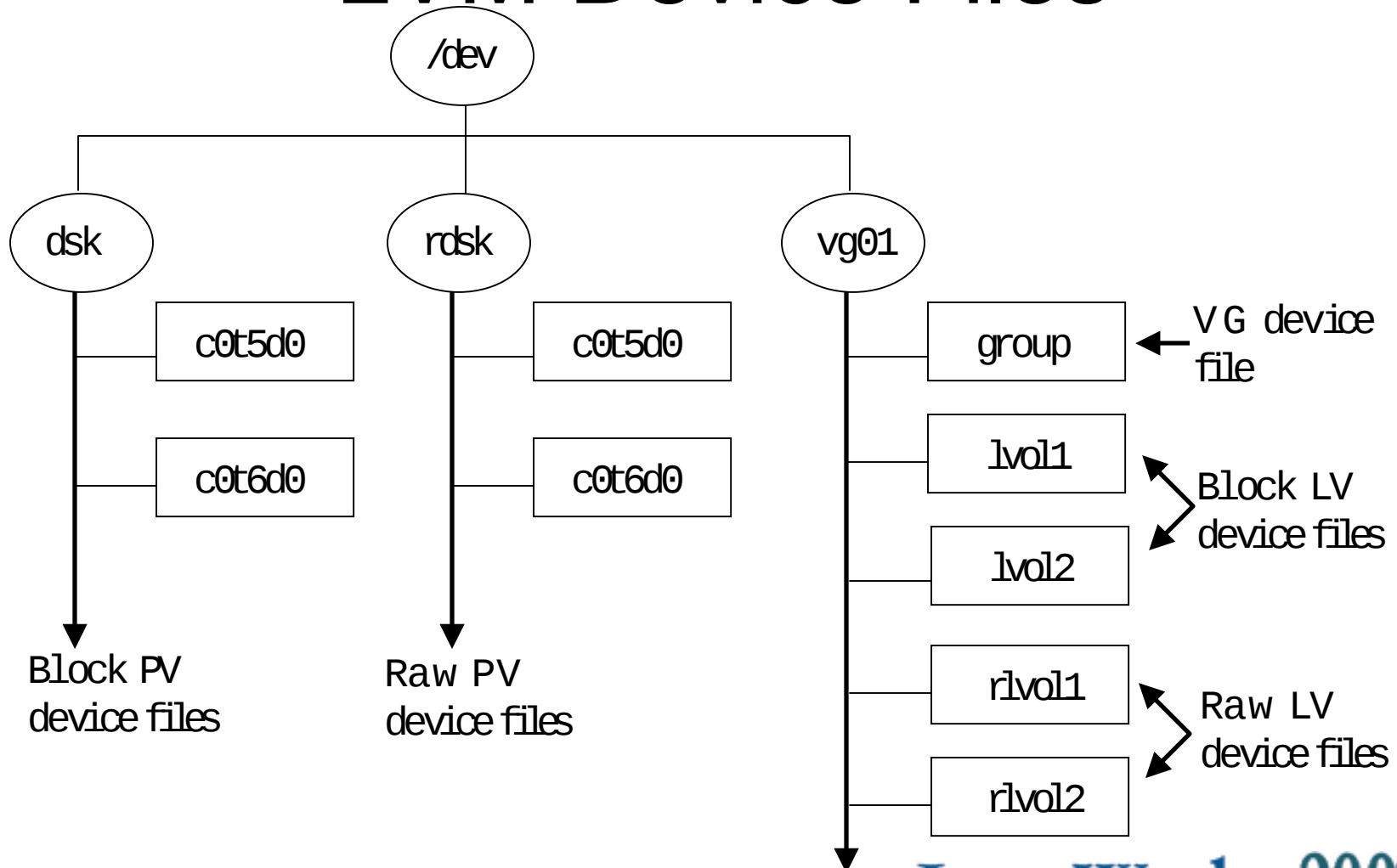
extent size (-s *n*)

* max extents/pv (-e *n*)

max. useable space/disk

	pe ₀	pe ₁	pe ₂	pe ₃	...			pe ₁₀₁₆
pv 0							??	
pv 1								
pv 2								
~								
~								
pv 15								

LVM Device Files



/etc/lvmtab

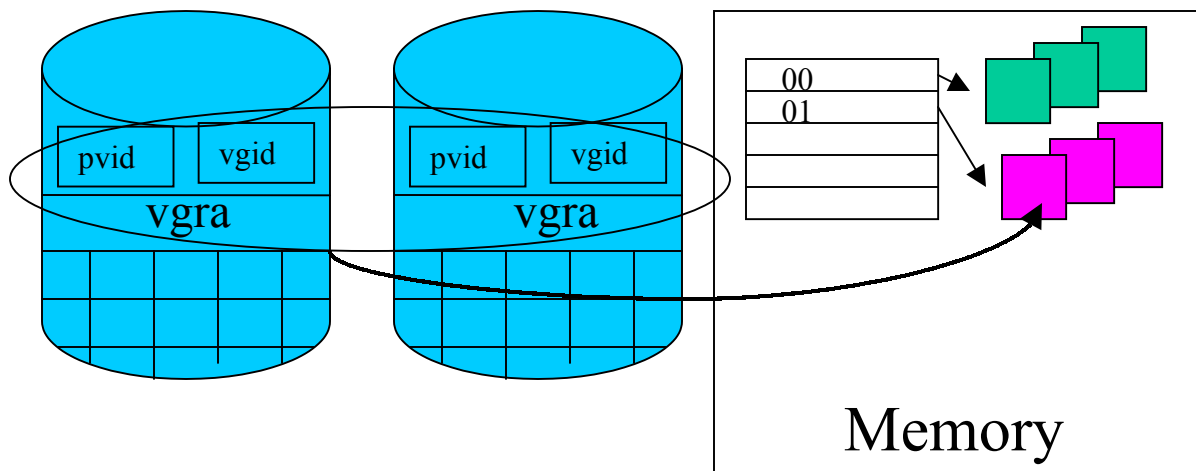
- Non-ascii file
- Records volume group/physical volume associations
- Used by many LVM commands for sanity checking

example:

```
/dev/vg00 (vgid)  
/dev/dsk/c0t6d0 (vgid,pvid)  
/dev/vg01 (vgid)  
/dev/dsk/c1t2d1 (vgid,pvid)  
/dev/dsk/c1t2d2 (vgid,pvid)
```

Volume Group Activation

- Done automatically when vg is created
- automatically at boot (/etc/lvmrc)
- required in order to access any lvol

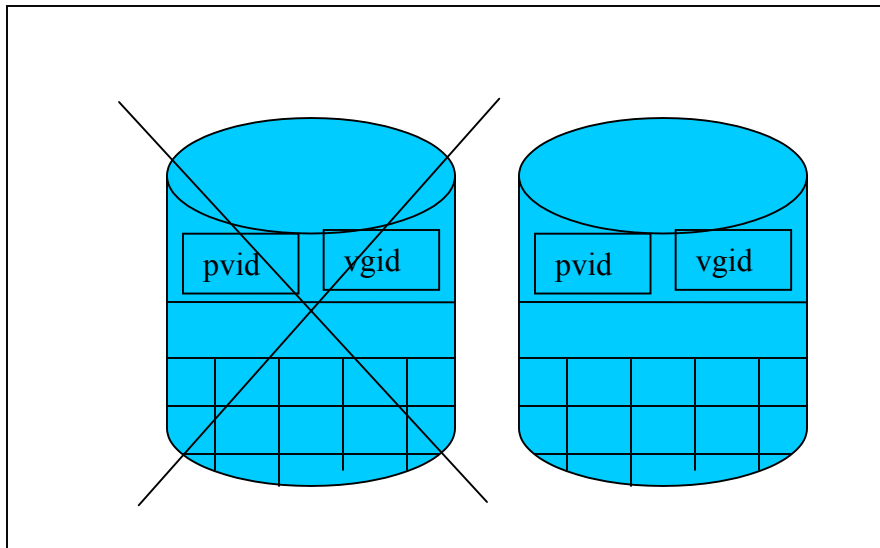


`vgchange -a y vname`

`vgchange -a r vname`

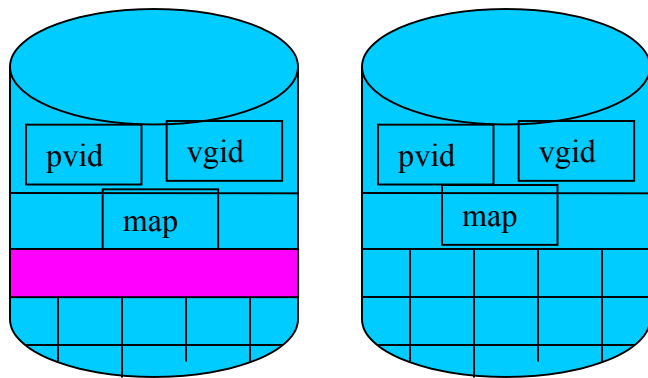
LVM Concepts-Quorum

vg01



- More than 1/2 disks in a vg required to activate the volume group
- Can override:
`vgchange -a y -q n vg01`
- Booting without quorum:
`ISL> hpux -lq`

LVM Concepts-Logical Volume



`lvcreate -L n vname`

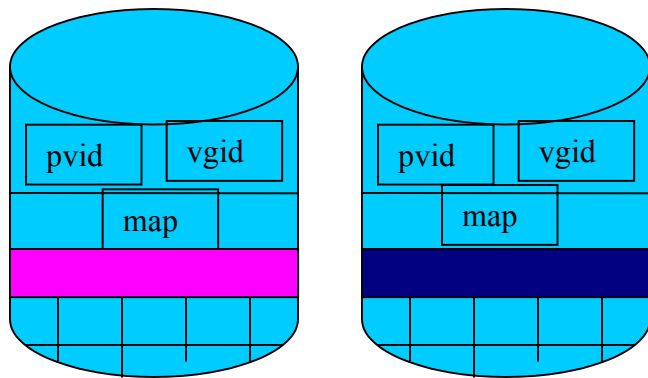
`lvcreate -L n -C y vname`

`lvcreate -l n vname`

`lvcreate -L n -n name vname`

	pe ₀	pe ₁	pe ₂	pe ₃	pe ₄	pe ₁₀₁₆	
pv 0	01	01	01	01	01	??	
pv 1							
pv 2							
pv 15							

Placing logical volumes



`lvcreate -n data1v vgdata`

`lvextend -L 500`
`/dev/vgdata/data1v`

`/dev/dsk/c4t2d0`

	pe 0	pe 1	pe 2	pe 3	pe 4	pe 1016		
pv 0	01	01	01	01	01	??	??	
pv 1	02	02	02	02	02	??	02	
pv 2								
~								
~								
pv 15								

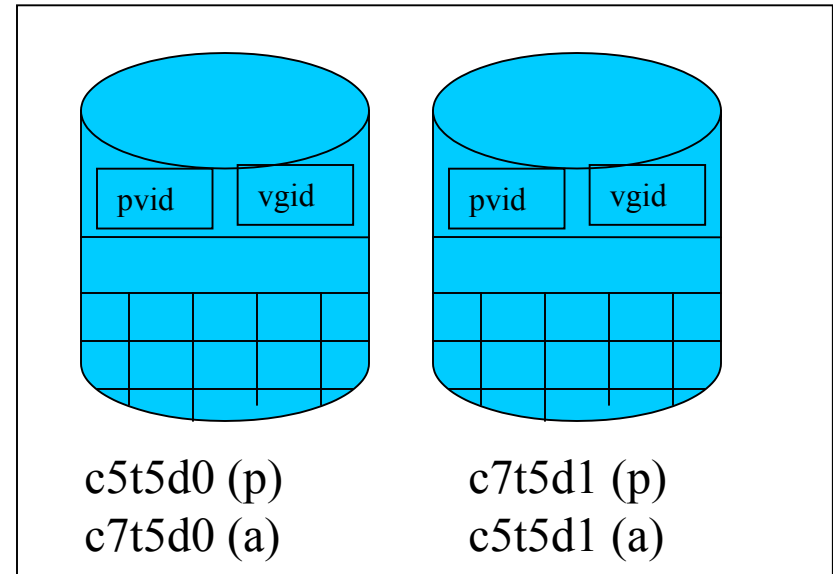
Multi-pathing solutions

Multipathing Solutions with LVM

- LVM pvlinks
 - active/passive
 - no additional software required
- Auto path
 - active/active, load balancing
 - supported for XP and VA arrays
 - additional cost product
- Powerpath
 - active/active, load balancing
 - supported for EMC arrays
 - additional cost product

PV links

```
pvcreate /dev/rdisk/c5t5d0
pvcreate /dev/rdisk/c5t5d1
mkdir /dev/vg01
mknod /dev/vg01/group c 64
0x010000
vgcreate vg01
/dev/dsk/c5t5d0 /dev/dsk/c7t5d0
/dev/dsk/c7t5d1 /dev/dsk/c5t5d1
```



PV links-switching the order

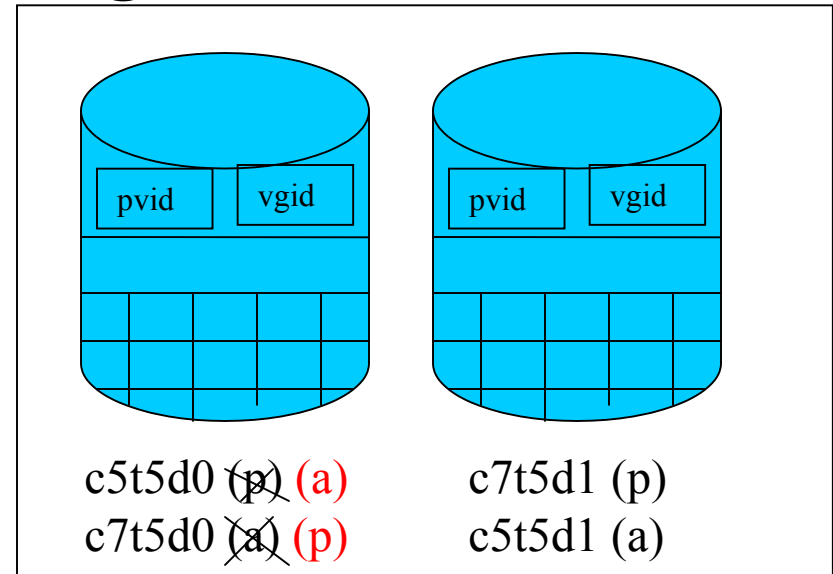
- Temporary Change

```
pvchange -s /dev/dsk/c7t5d0
```

- Permanent Change

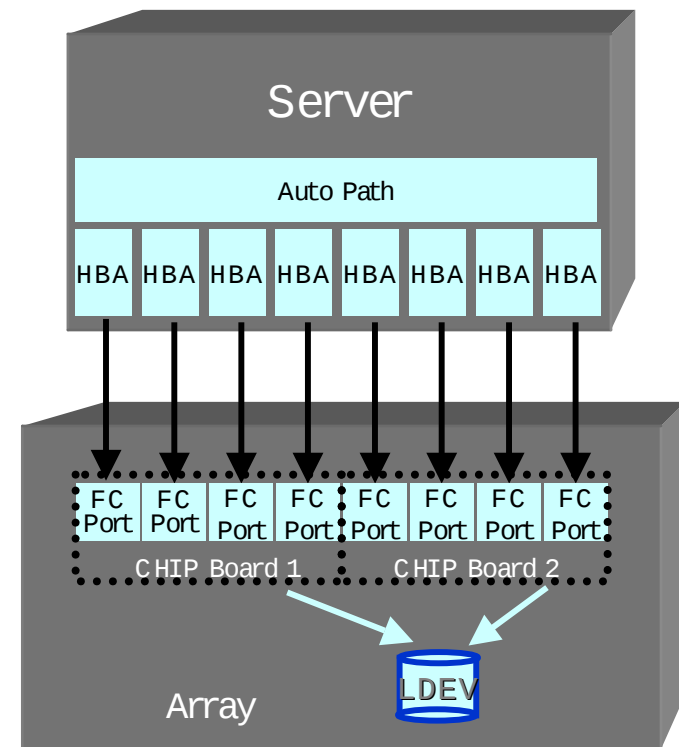
```
vgreduce vg01 /dev/dsk/c5t5d0
```

```
vgextend vg01 /dev/dsk/c5t5d0
```



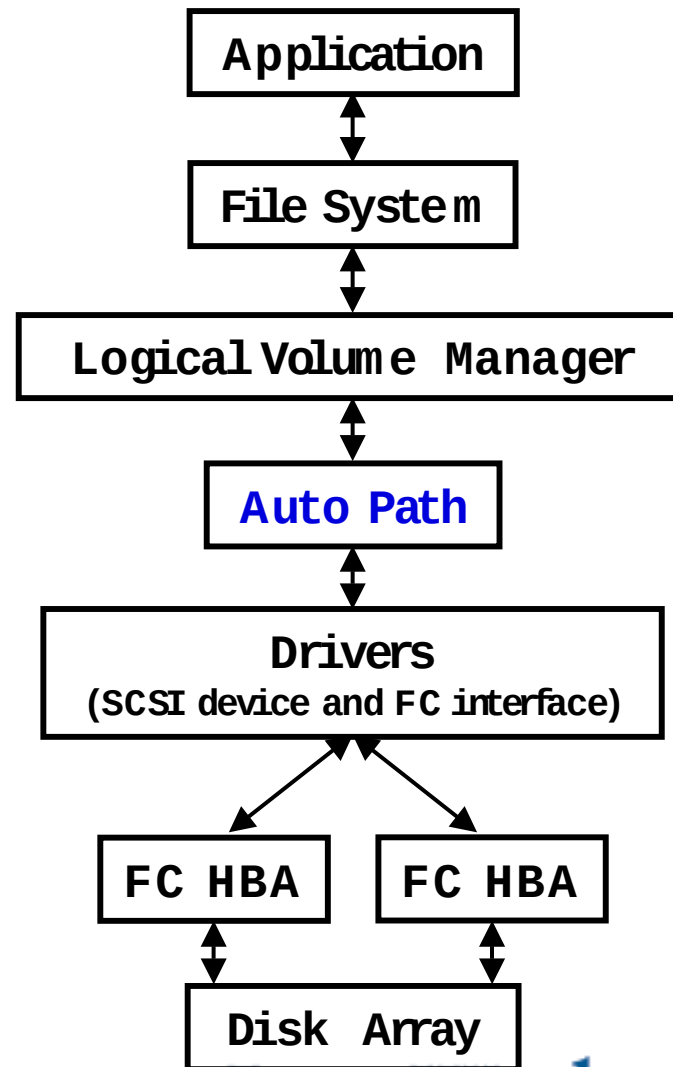
Auto Path-Dynamic load balancing

- Load balancing functionality
 - Supports up to 8 paths from a server to an end device
 - Provides dynamic load balancing across all paths to an end device
 - Choose from 4 load balancing policies, including “no load balancing”
 - Supports the XP and VA disk arrays
 - Load balancing supported in clustered environments



Auto Path driver

- A pseudo driver
- The Auto Path driver is layered between the LVM (Logical Volume Manager) and the SCSI device driver
- The driver provides the Command Line Interface



Moving Data

Moving Disks

Three Step Process

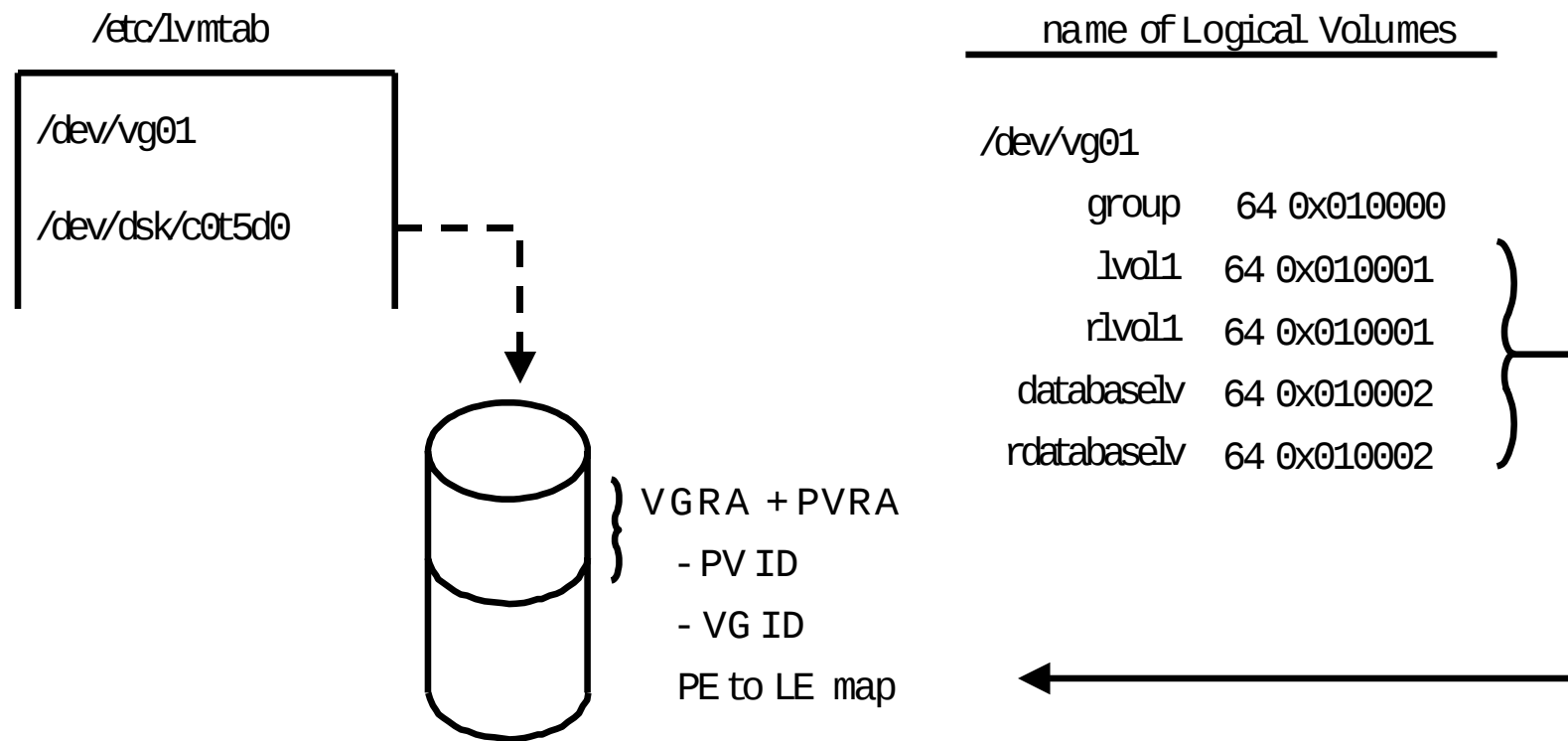
1. Remove definition of volume group
2. Move disk(s)
3. Add definition of volume group

Two commands

`vgexport (1m)`

`vgimport (1m)`

Volume Group Definition — Review



Exporting a Volume Group

Syntax:

vgexport [-p] [-v] [-m *file*] *VG*

- p Preview actions only
- v verbose
- m used to specify a map file for logical volume names
- Removes volume group definition from the system completely by updating `/etc/lvmtab` and kernel memory.
- The volume group must first be deactivated with `vgchange (1m)`.

Example:

```
vgchange -a n /dev/vg01
```

```
vgexport -v -m /etc/lvmconf/vg01.map /dev/vg01
```

Importing a Volume Group

Syntax:

```
vgimport [-p][-v][-m file] VG PV [PV...]
```

-p Preview actions only

-v verbose

-m used to specify a map file for logical volume names

Example:

```
mkdir /dev/vg01
```

```
mknod /dev/vg01/group c 64 0x010000
```

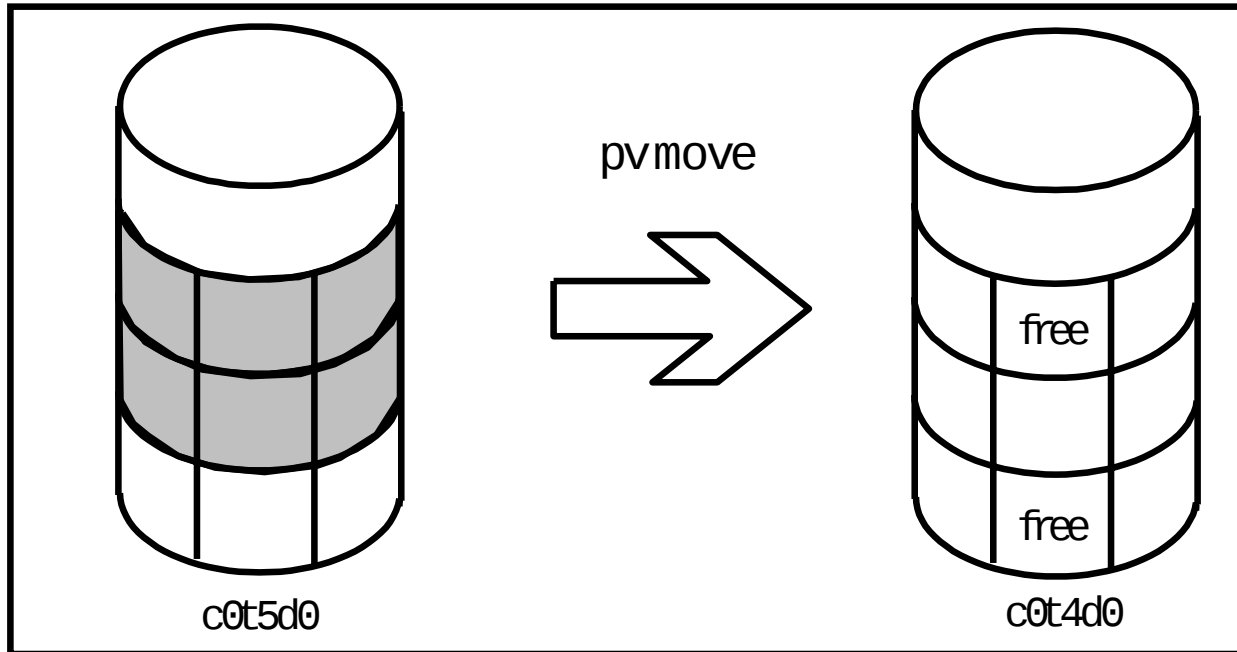
```
vgimport -v -m /etc/lvmconf/vg01.map /dev/vg01 /dev/dsk/c0t1d0
```

```
vgchange -a y /dev/vg01
```

```
vgcfgbackup vg01
```

Moving LVM Data

Volume group



Syntax:

```
pvmove [-n lv] from_PV [to_PV]
```

Example:

```
pvmove -n /dev/vg01/lvol1 /dev/dsk/c0t5d0 /dev/dsk/c0t4d0
```

Renaming LVM objects

Renaming Logical Volumes

- lvol names are not stored in the LVM maps or in /etc/lvmtab
- Simply rename device files and update /etc/fstab if necessary

Renaming Volume Groups

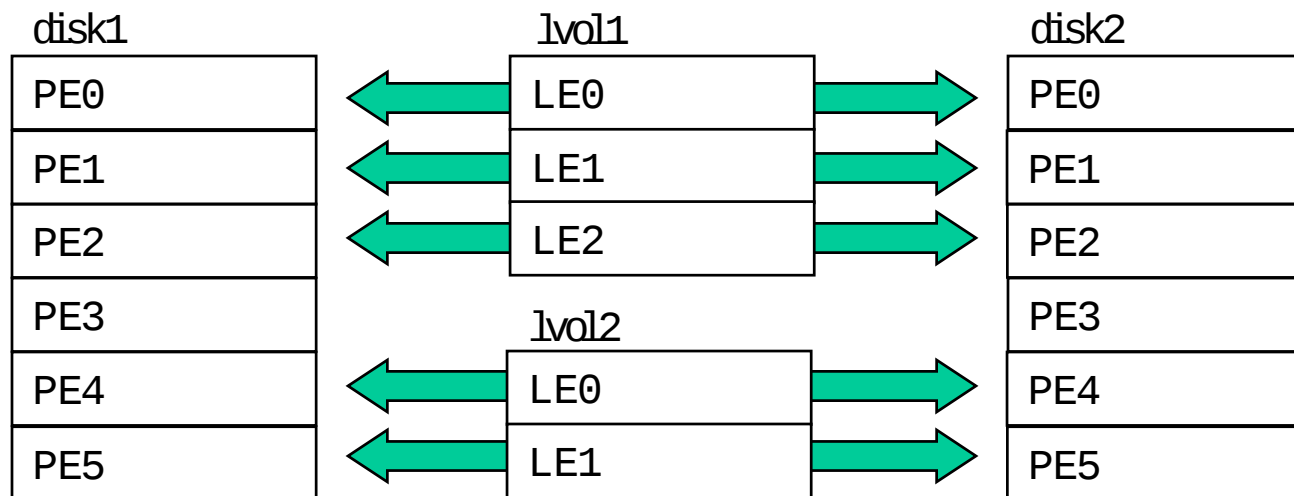
- Volume group name is kept in `/etc/lvmtab` and is used as the directory name to anchor the device files for the group
- Use `vgexport` and `vgimport` to rename volume group

LVM Mirroring

Mirrored Volumes

In a mirrored logical volume ...

- Each logical volume consists of one or more Logical Extents (LEs).
- Each LE maps to two or three Physical Extents (PEs) on disk.
- The logical volume remains accessible even when one of the disks is unavailable.



Extending and Reducing Mirrors

Create a new, mirrored logical volume:

```
# lvcreate -m 1 -L 16 -n myfs1 vg01
```

Mirror an existing logical volume:

```
# lvextend -m 1 /dev/vg01/myfs1
```

Mirror an existing logical volume to a specific disk:

```
# lvextend -m 1 /dev/vg01/myfs1 /dev/dsk/c0t3d0
```

Add a second mirror:

```
# lvextend -m 2 /dev/vg01/myfs1
```

Remove a logical volume's mirrors:

```
# lvreduce -m 0 /dev/vg01/myfs1 /dev/dsk/c0t3d0
```

Check a mirrored logical volume's status:

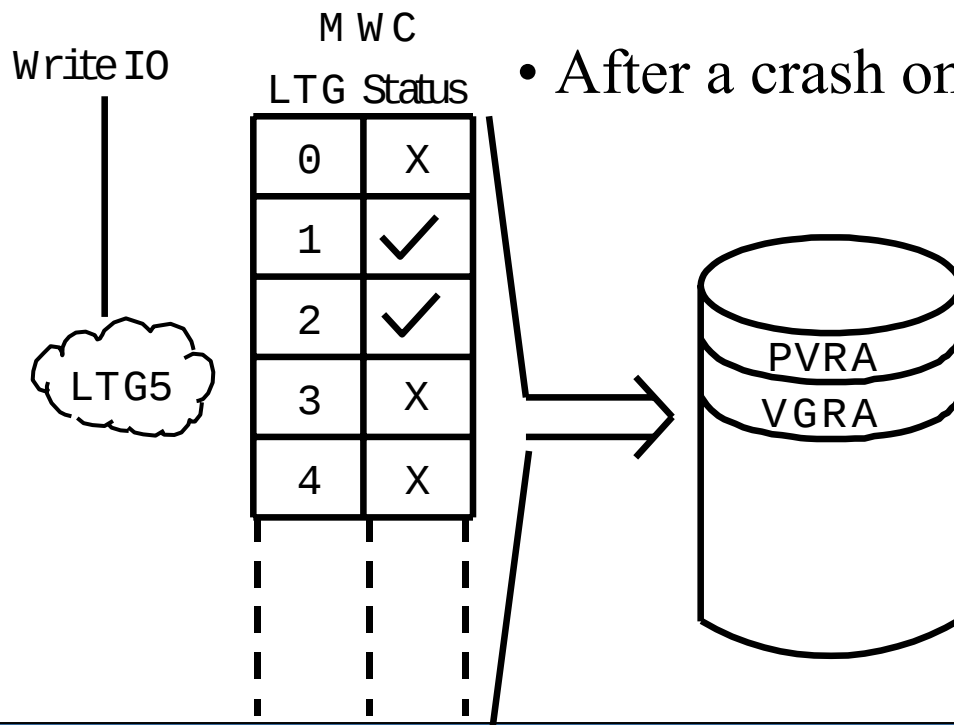
```
# lvdisplay -v /dev/vg01/myfs1
```

Mirrored I/O Scheduling

	Parallel (-d p)	Sequential (-d s)
Read	Access PV with lowest outstanding I/Os	Read in PV order
Write	Schedule writes simultaneously to all PVs	Schedule writes in PV order

MWC/MCR

- Writes are recorded in MWC in memory
- MCR record written to disk when a write is done to a logical track group not already recorded
- After a crash only "dirty" LTGs need be resynced

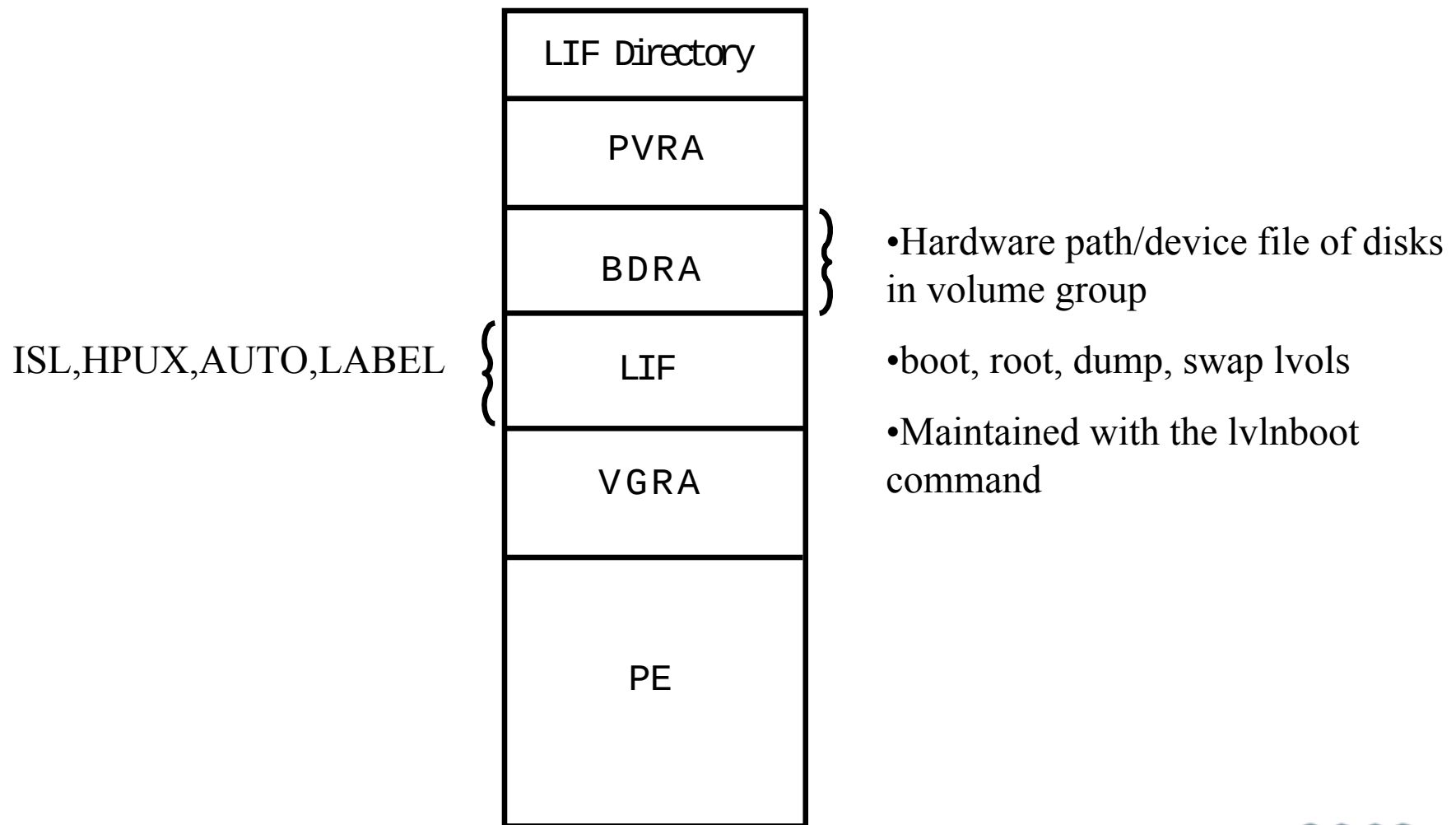


Mirror Consistency Recovery Options

		Advantage	Disadvantage
1	MWC/MCR -M y	Fast recovery on a system crash	Runtime overhead to write MCR record
2	NOMWC -M n -c y	No extra overhead at runtime	Slow recovery on a system crash
3	NONE -M n -c n	Application can do necessary recovery	No system managed recovery

LVM Boot Disks

LVM Boot volume structure



Mirroring the LVM Boot Disk

- Initialize disk for LVM. Leave room for BDRA/LIF

```
pvcreate -B /dev/rdisk/cntndn
```

- Add disk to root VG

```
vgextend vg00 /dev/rdisk/cntndn
```

- Install boot files in LIF area

```
mkboot /dev/rdisk/cntndn
```

- Change the auto file on both the primary and alternate boot disk

```
mkboot -a "hpux -lq" /dev/rdisk/cntndn
```

- Mirror each of the lvol's in the root vg

```
lvextend -m 1 /dev/vg00/lvoln /dev/dsk/cntndn
```

- Add the mirror disk definition to /stand/bootconf

Booting When BDRA is damaged

- Boot system into maintenance mode
`ISL> hpux -lm`
- Activate vg00
`vgchange -a y vg00`
- Use lvinboot to examine/repair BDRA
`lvinboot -v`
`lvinboot -b /dev/vg00/lvol1`
`lvinboot -r /dev/vg00/lvol3`
`lvinboot -s /dev/vg00/lvol2`
`lvinbtoo -d /dev/vg00/lvol2`
- Reboot the system
`reboot`

Moving the Boot Disk

- Problem: /etc/lvmtab contains the old device files for the boot disk
 - Solution: boot into maintenace mode, export and reimport volume group
- Problem: The BDRA and Label files contain the old device information
 - Solution: use lvinboot to fix

Cookbook for moving root disk

Boot from new device. Reply Y to interact with IPL

```
ISL> hpux -lm
```

```
# vgexport -v -m vg00.map vg00
```

```
# mkdir /dev/vg00
```

```
# mknod /dev/vg00/group c 64 0x0000000
```

```
# vgimport -v -m vg00.map vg00 /dev/dsk/new_device_file
```

```
# vgchange -a y vg00
```

```
# vgcfgbackup vg00
```

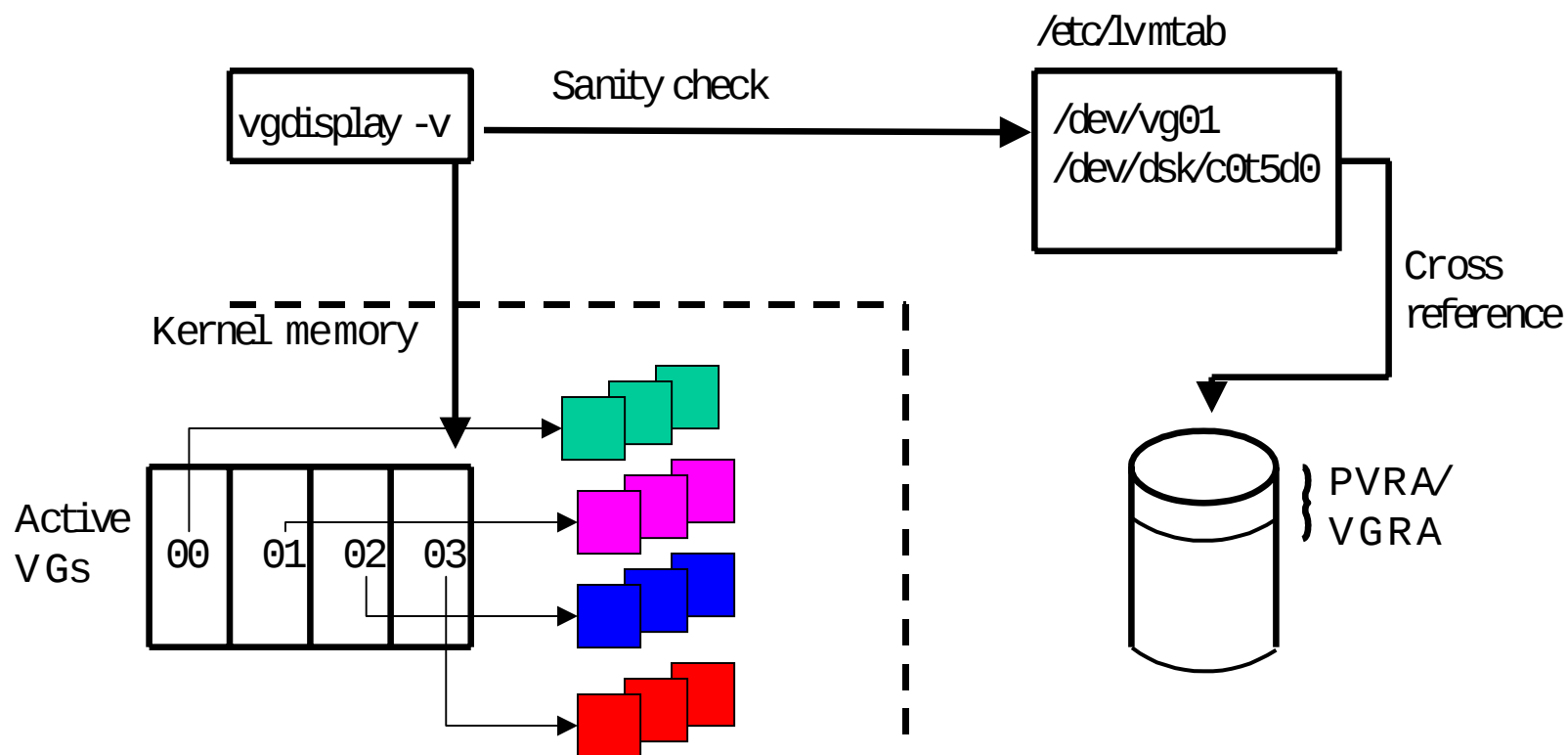
```
# lvinboot -R
```

```
# lvinboot -v
```

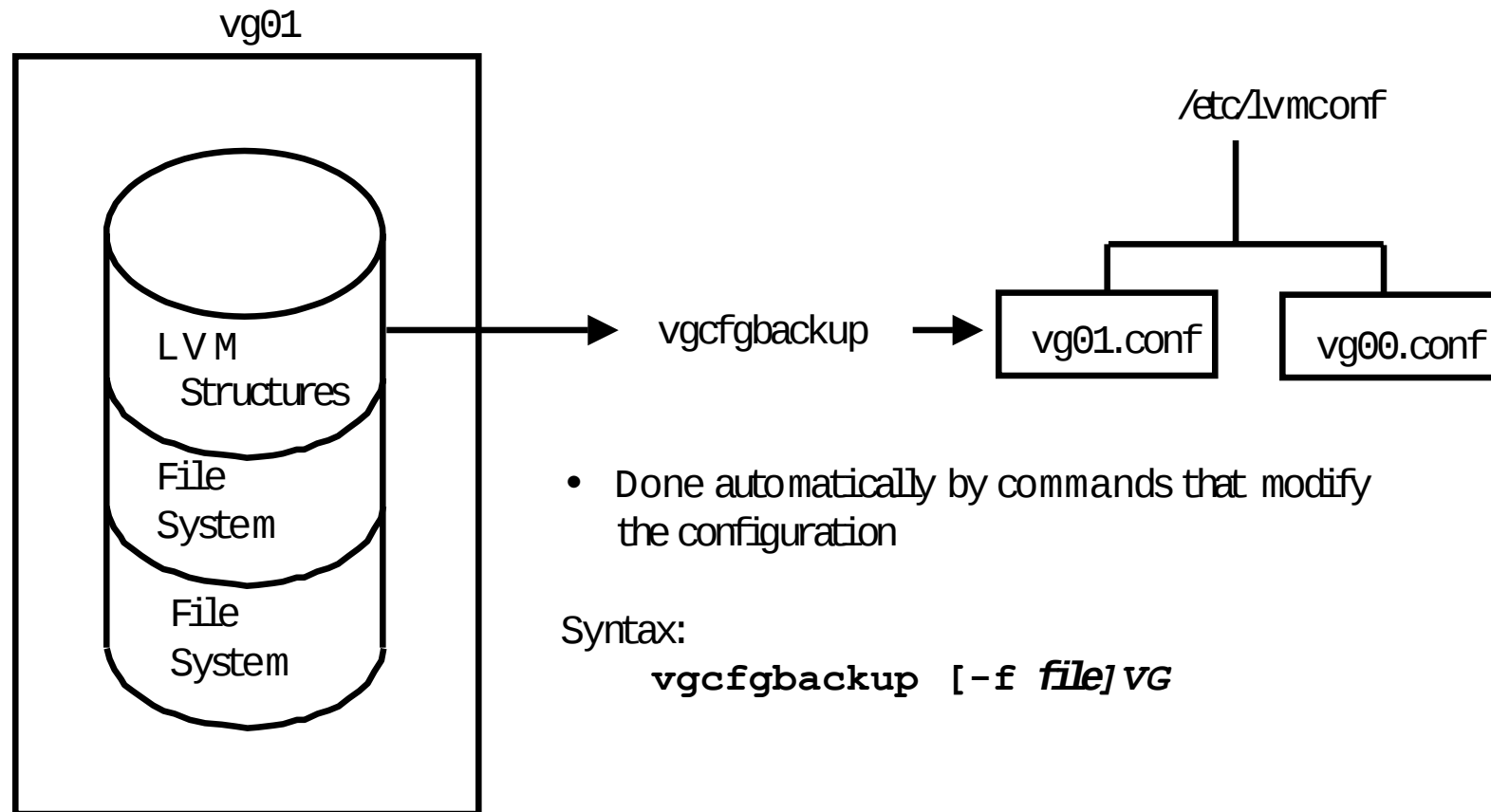
Recovering Corrupt LVM Information

Valuable Data?

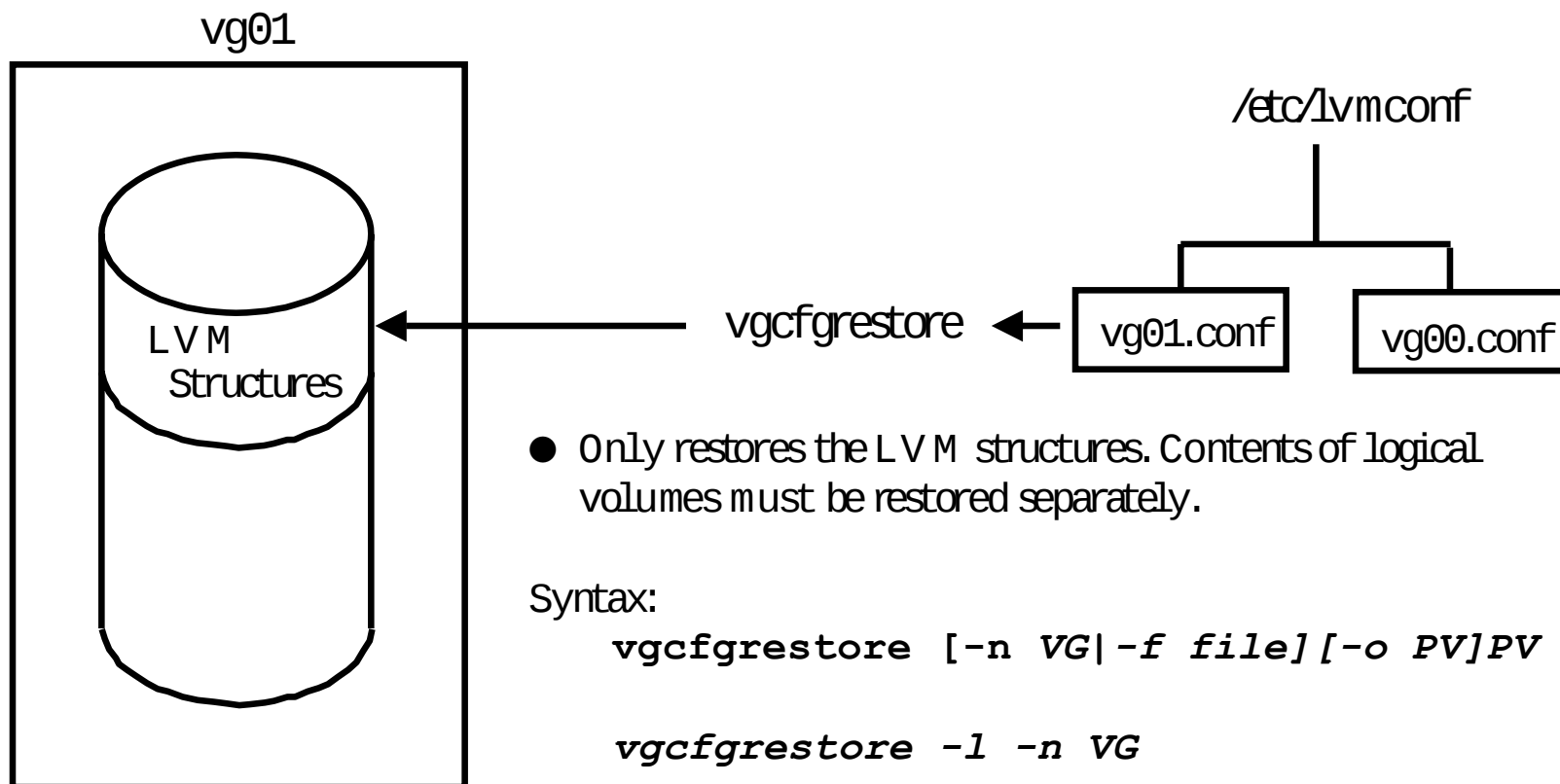
- The various lvm data structures are crucial to the continued availability of logical volumes



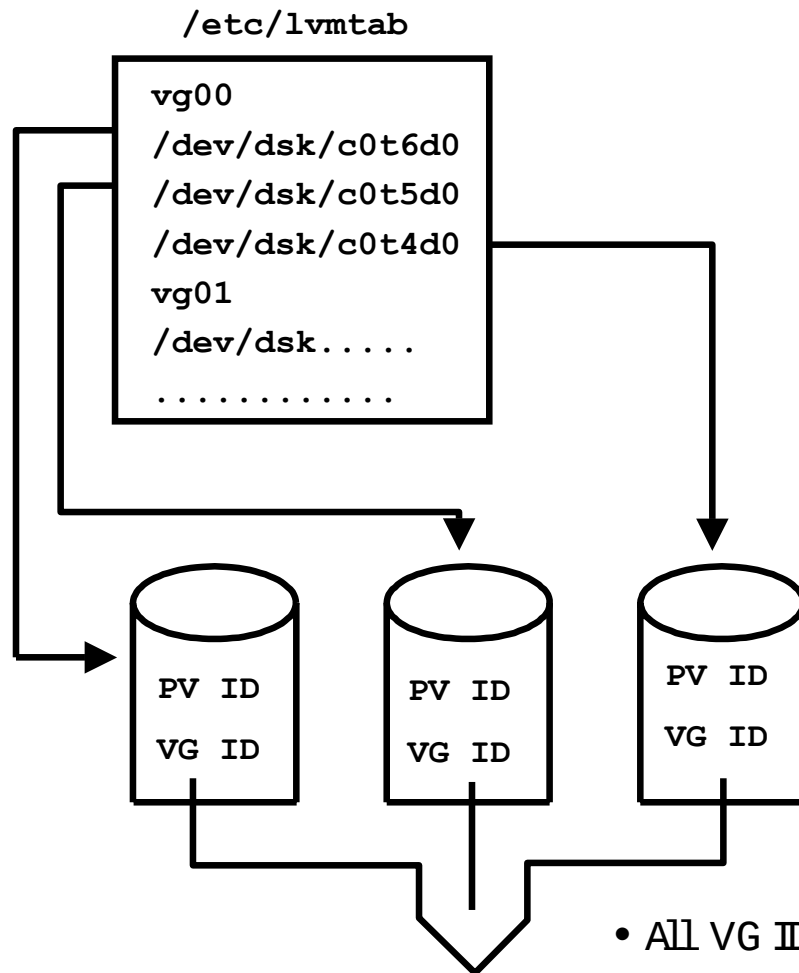
Backing up LVM Structures



Recovering LVM Structures



LVM Control File



- Not ascii data
 - can use strings (1) to read the ascii part
- Used primarily at boot up but also used to sanity check commands
- Can be rebuilt if lost

- All VG IDs the same - all disks in same vg

Recover or Repair `/etc/lvmtab`

Syntax:

```
vgscan [-v] [-p]
```

Semi-intelligent look at every disk to classify it

- imports those that are known to be on this system
- recommends the import of "new" disks
- rebuilds `/etc/lvmtab`

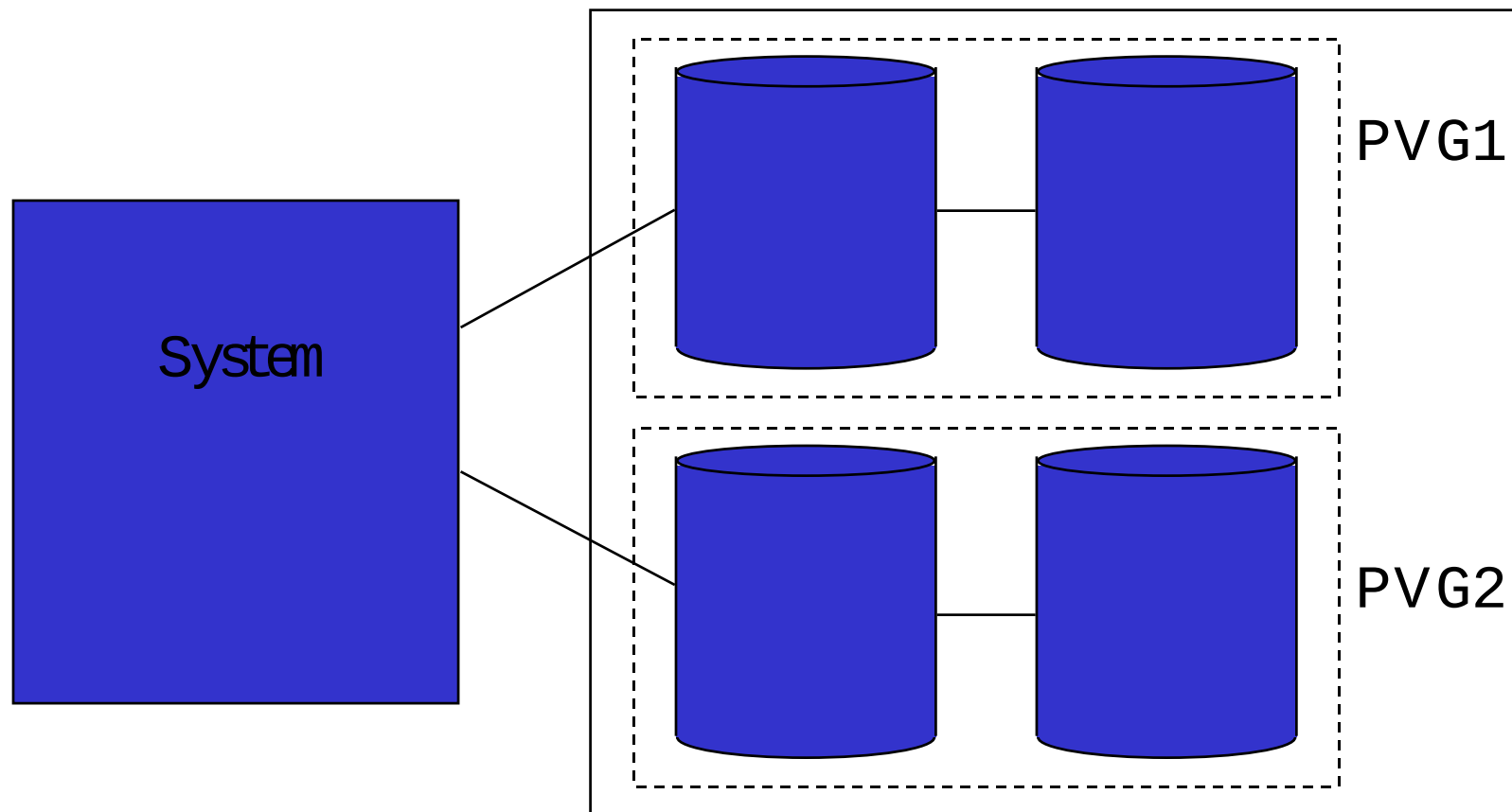
Ways to confuse `vgscan`

- redundant group files
- "old" LV M structures

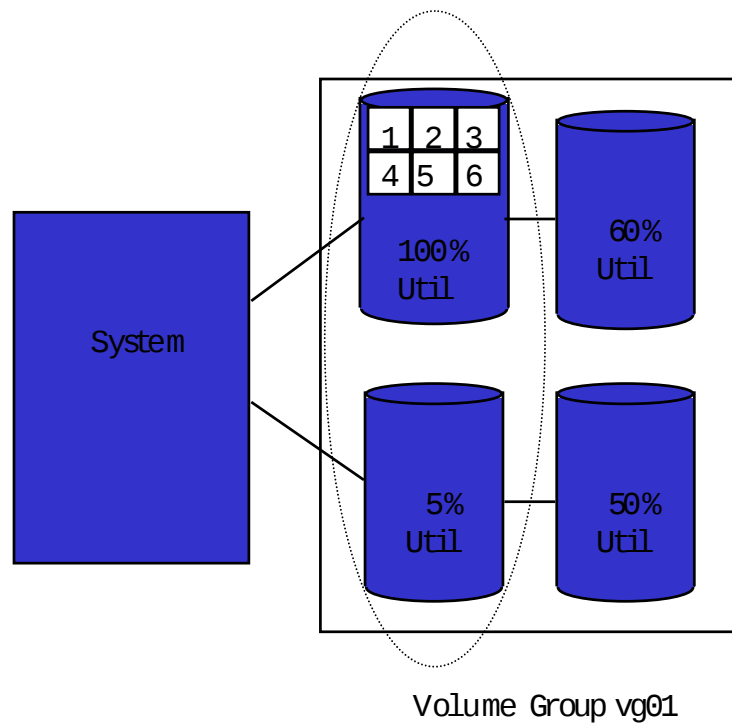
LVM Performance Tips

Use PVGs to Offload Busy Controllers

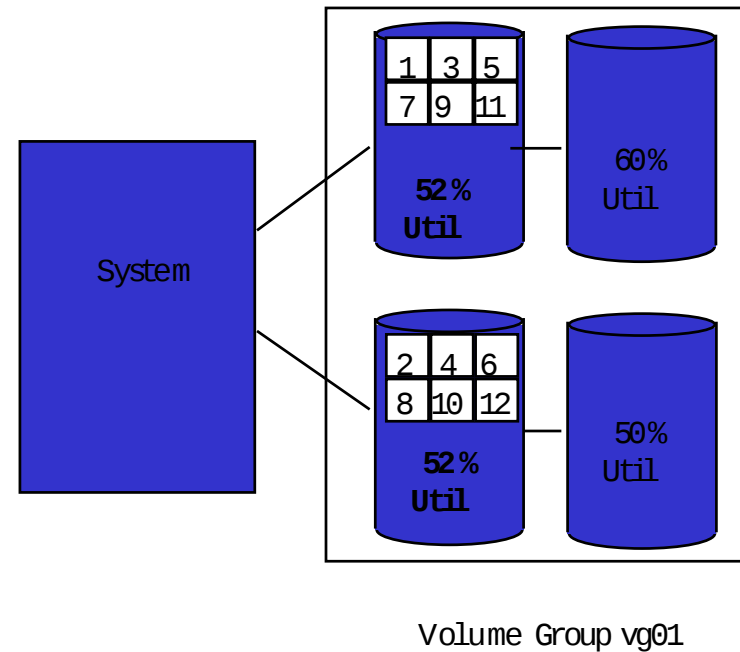
Volume Group vg01



Use Striping to Offload Busy Drives



Without Striping



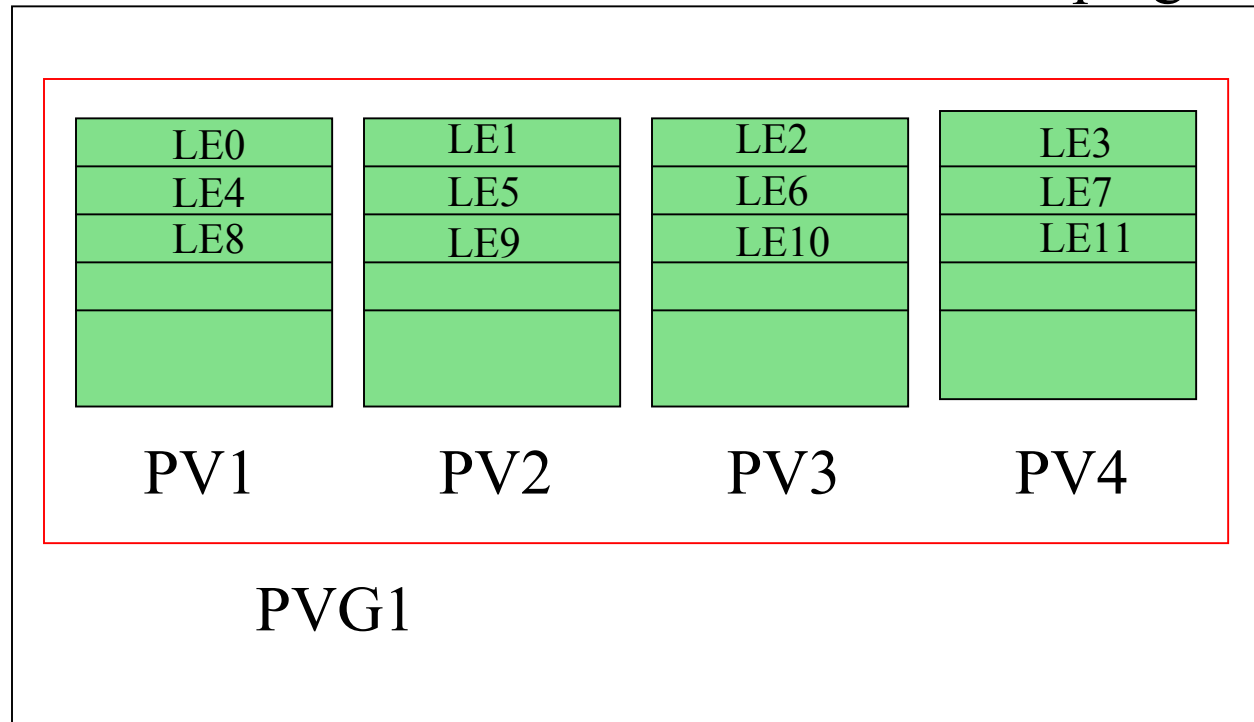
With Striping

LVM Distributed Allocation

Logical Volume

LE0
LE1
LE2
LE3
LE4
LE5
LE6
LE7
LE8
LE9
LE10
LE11

Volume Group vg01



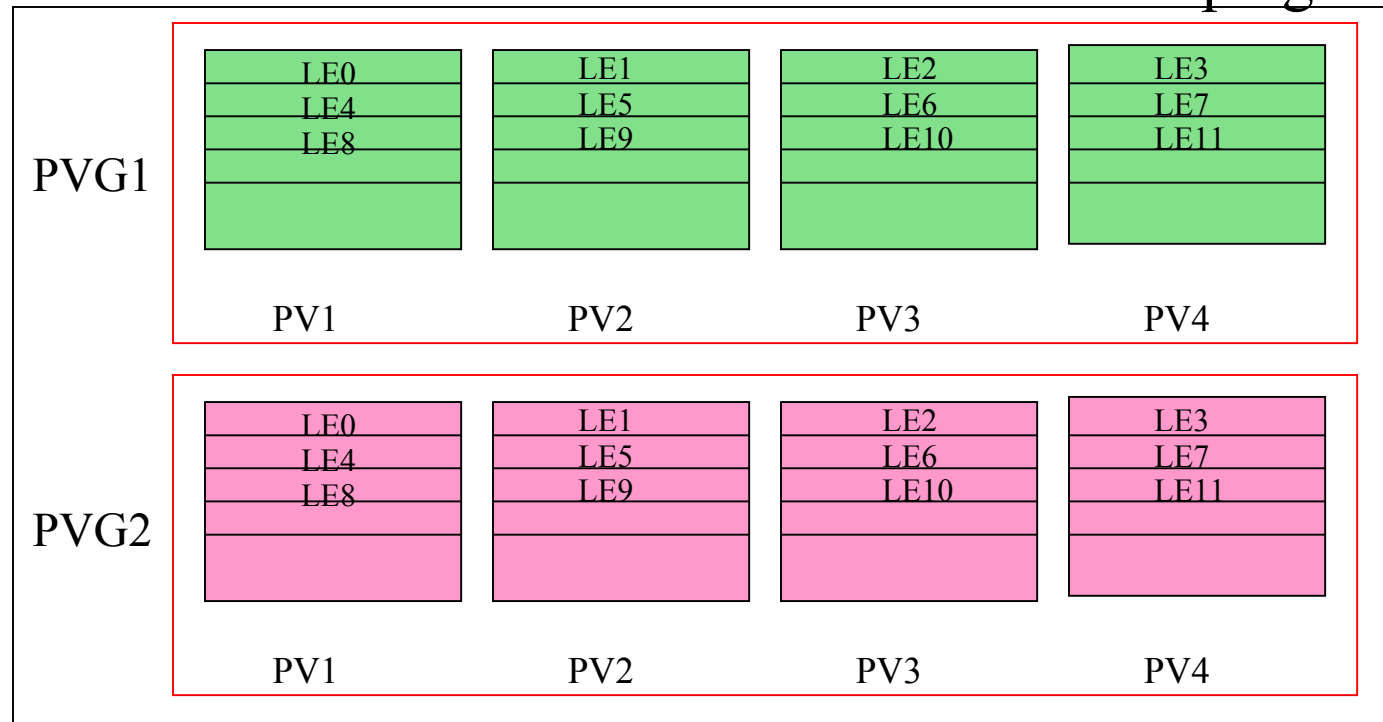
```
lvcreate -l 12 -s g -D y /dev/vg01
```

Mirroring and Striping

Logical Volume

LE0
LE1
LE2
LE3
LE4
LE5
LE6
LE7
LE8
LE9
LE10
LE11

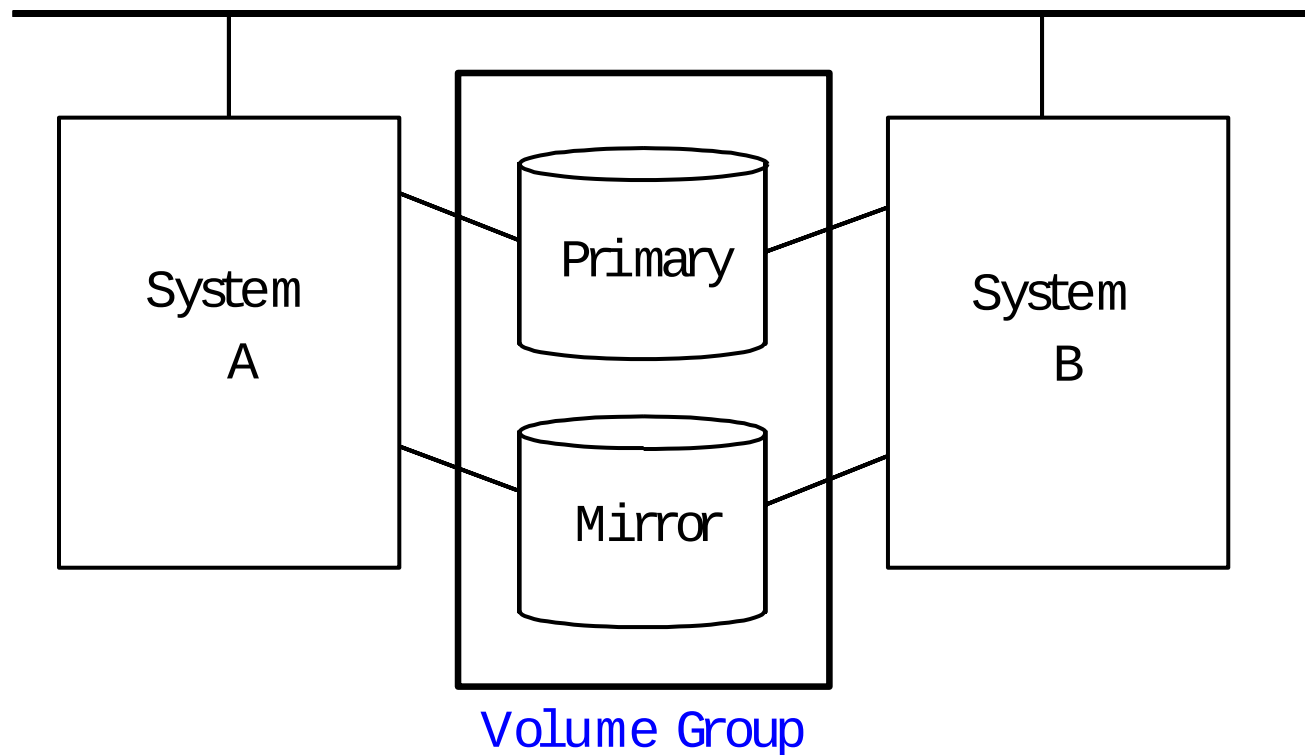
Volume Group vg01



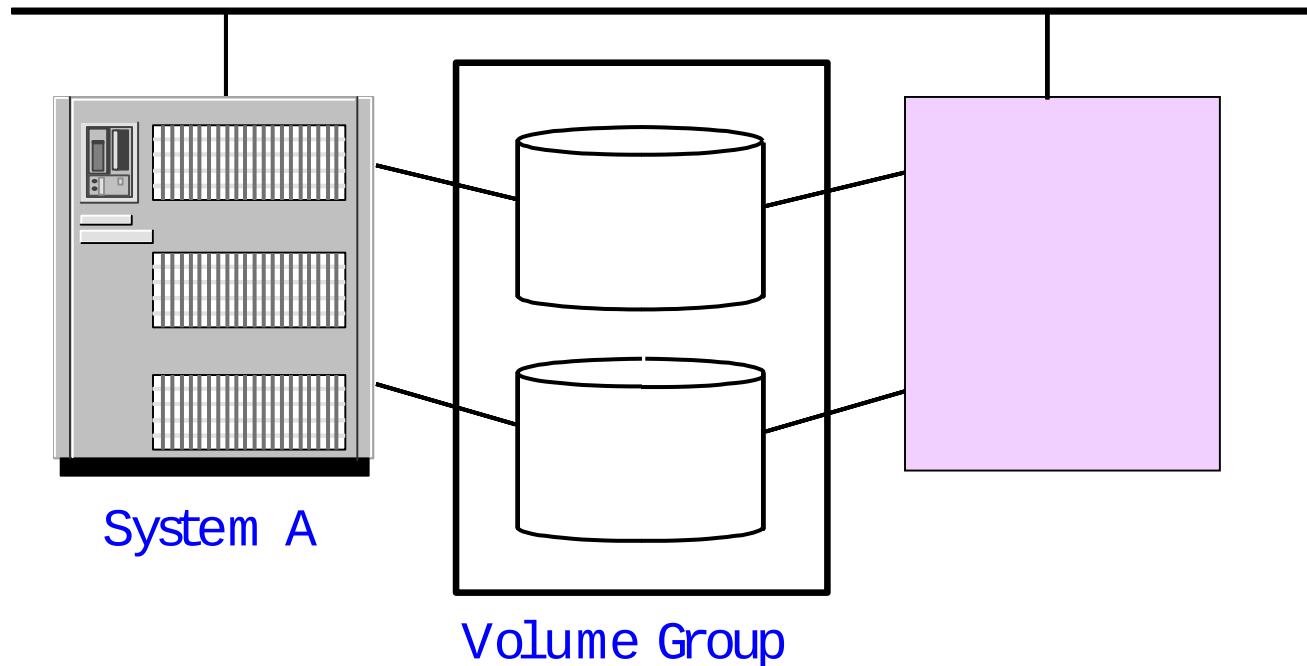
```
lvcreate -l 12 -m 1 -s g -D y /dev/vg01
```

LVM in an MC/ServiceGuard Environment

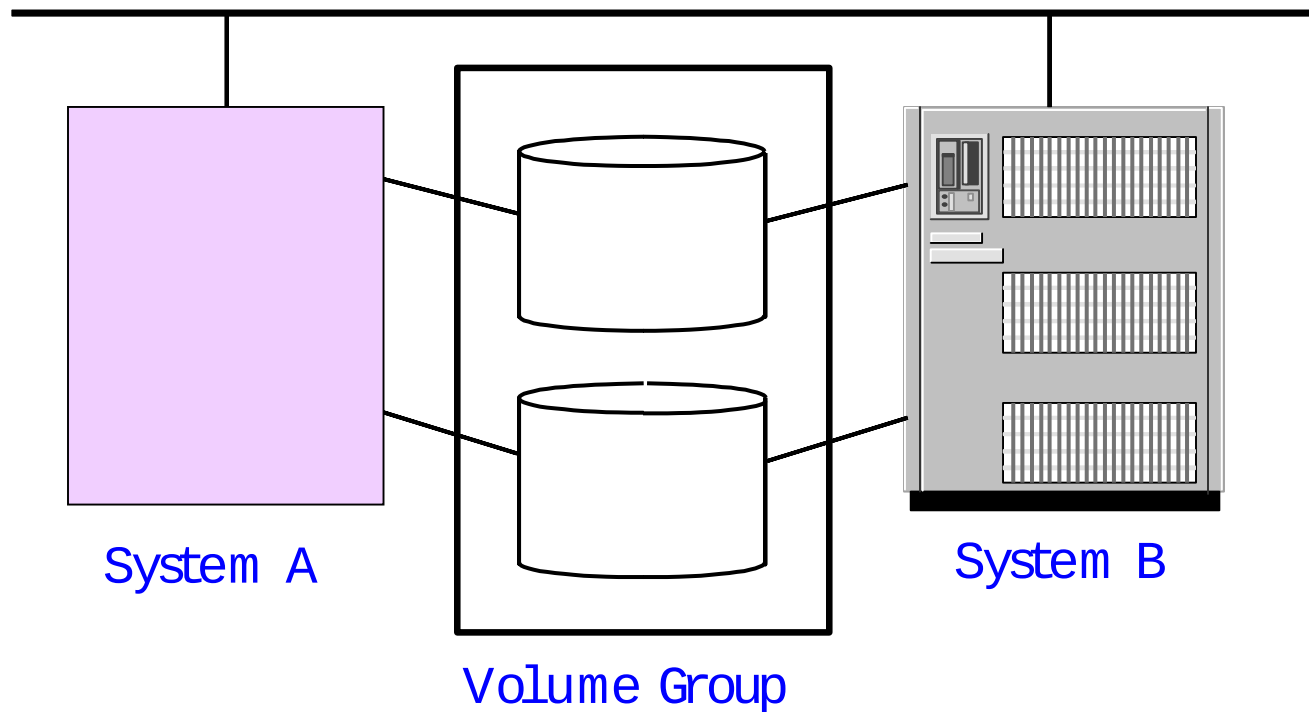
MC/ServiceGuard Configuration



Step 1: Configure Volume Group on First System



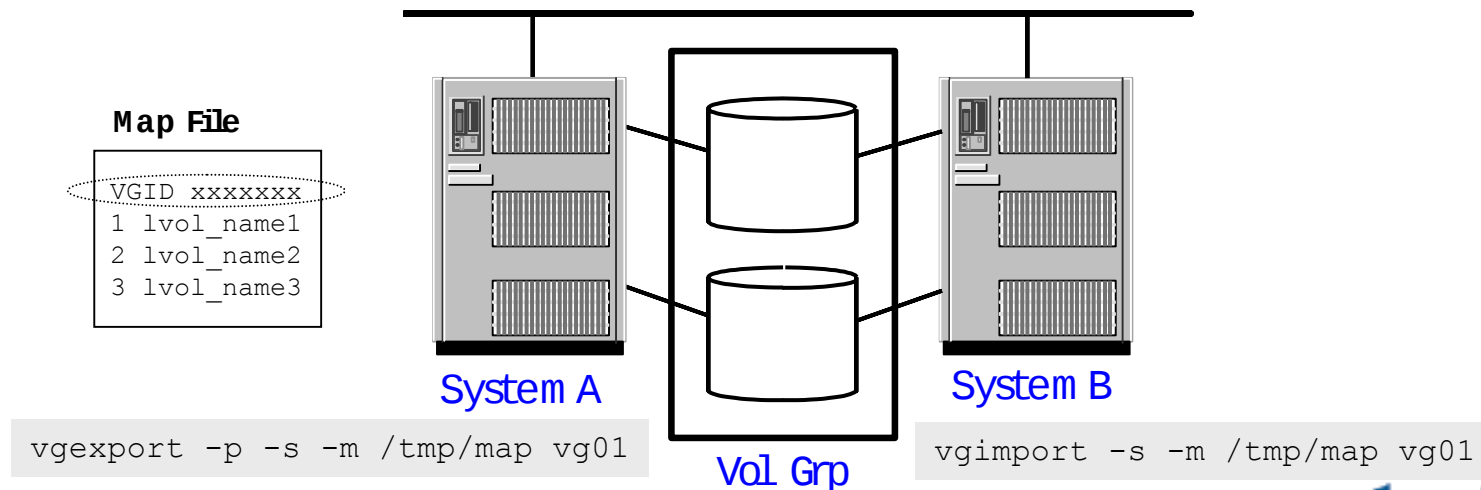
Step 2: Import Volume Group to Second System



VG Export and Import -s Option

vgexport -p -s -m <mapfilename> <vg_name>
create *mapfile* without removing the VG,
and save the VGID for the *vgimport* command

vgimport -s -m <mapfilename> <vg_name>
scan for disks that have the same *VG ID* as in the *mapfile*



LVM Definitions-Both Nodes

Node 1

/dev/vg01

group 64 0x010000

lvol1 64 0x010001

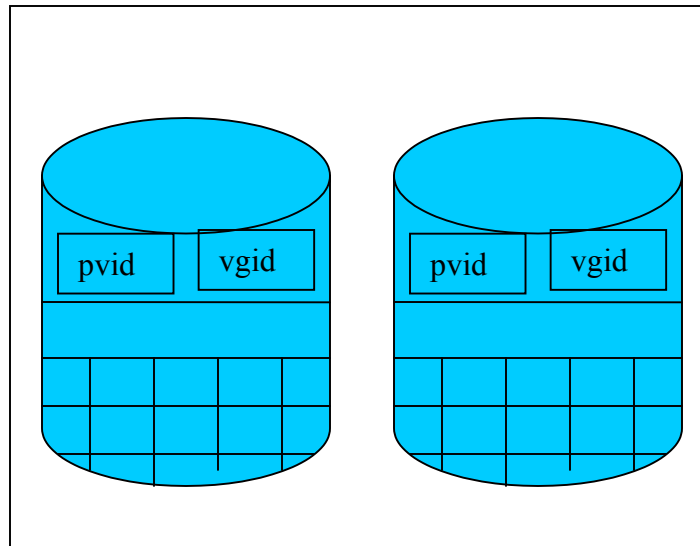
rlvol1 64 0x010001

database1v 64 0x010002

rdatabase1v 64 0x010002

/etc/lvmtab

```
/dev/vg01
/dev/dsk/c5t5d0
/dev/dsk/c7t5d0
/dev/dsk/c7t5d1
/dev/dsk/c5t5d1
```



Node 2

/dev/vg01

group 64 0x010000

lvol1 64 0x010001

rlvol1 64 0x010001

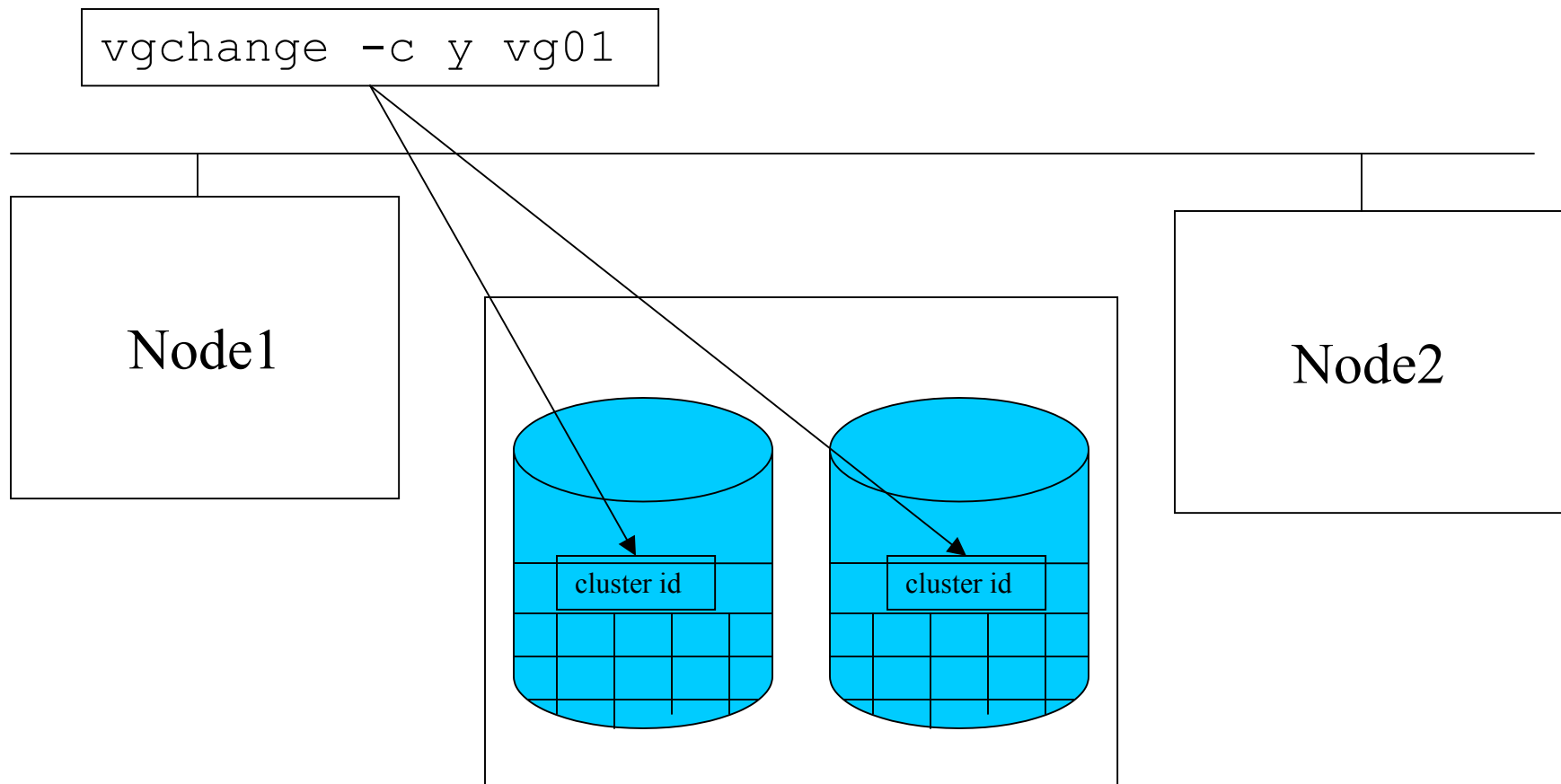
database1v 64 0x010002

rdatabase1v 64 0x010002

/etc/lvmtab

```
/dev/vg01
/dev/dsk/c5t5d0
/dev/dsk/c7t5d0
/dev/dsk/c7t5d1
/dev/dsk/c5t5d1
```

Cluster Volume Group

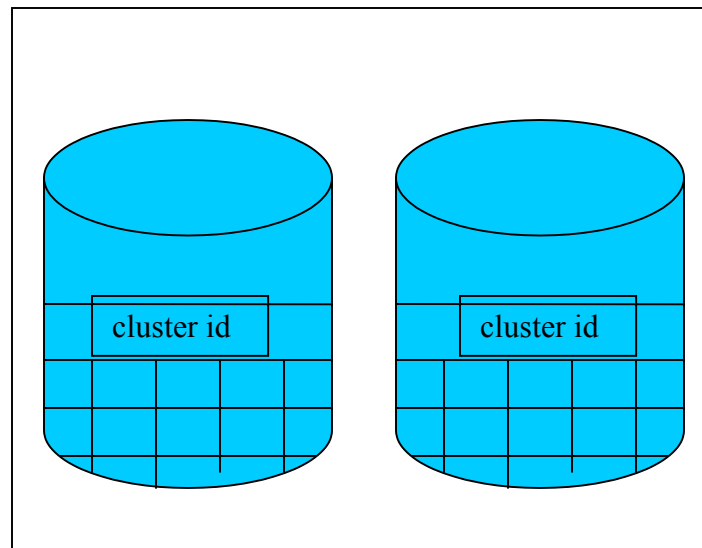


Volume Group Exclusive Activation

```
vgchange -a e vg01
```

Q. Do you
have the
volume group
activated?

Node1
cmlvmd



A. No I
do not

Node2
cmlvmd

Marking Volume Groups as M C/ServiceGuard Volume Groups

Marking Volume Group
for MC/ServiceGuard

vgchange -c y VGName

Marking Volume Group
as non-M C/ServiceGuard

vgchange -c n VGName

Standard Volume Group
Activation

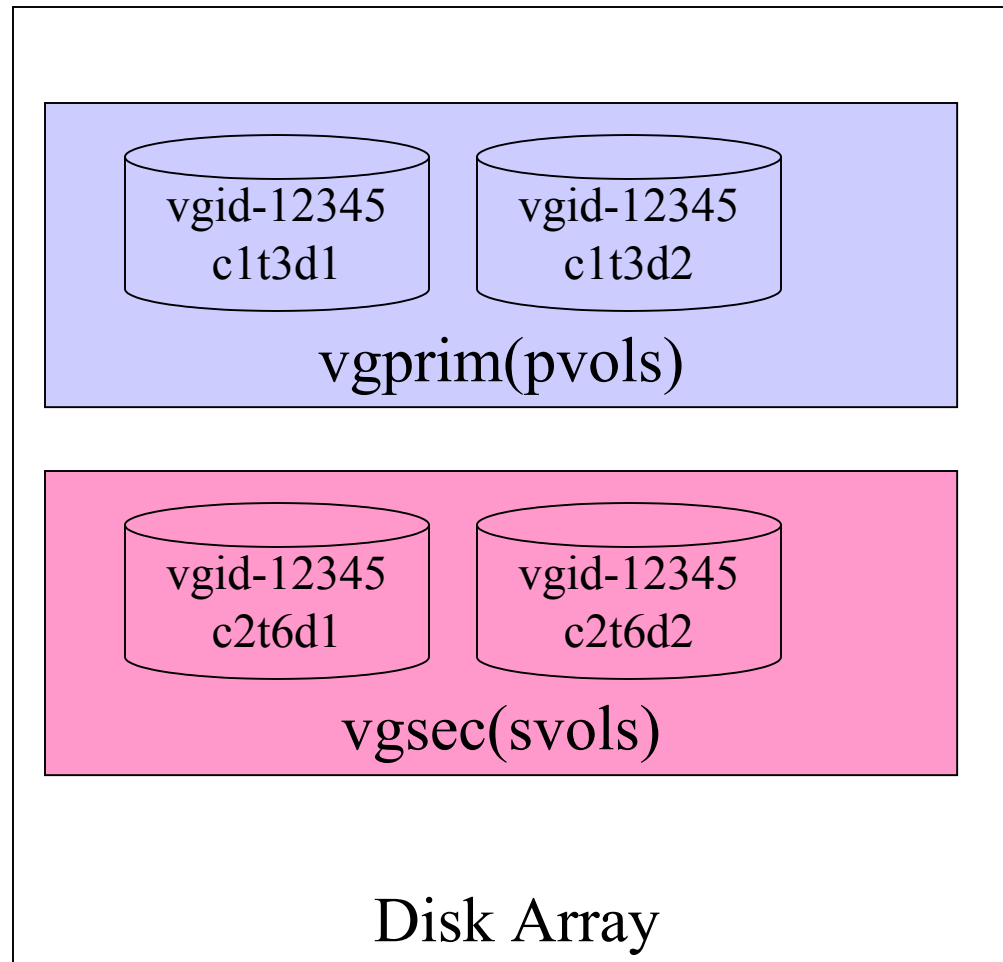
vgchange -a y VGName

Exclusive Volume Group
Activation

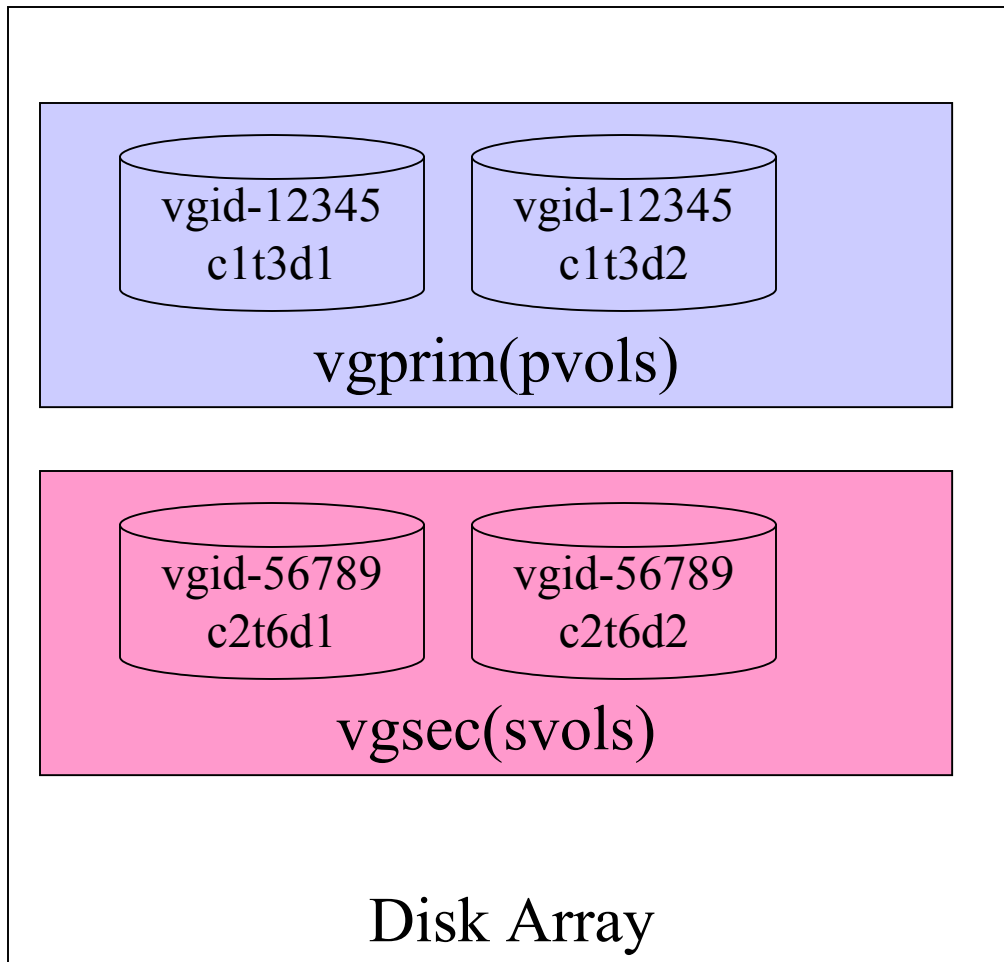
vgchange -a e VGName

Disk Array Data Replication Issues

Problem-Duplicate vgid

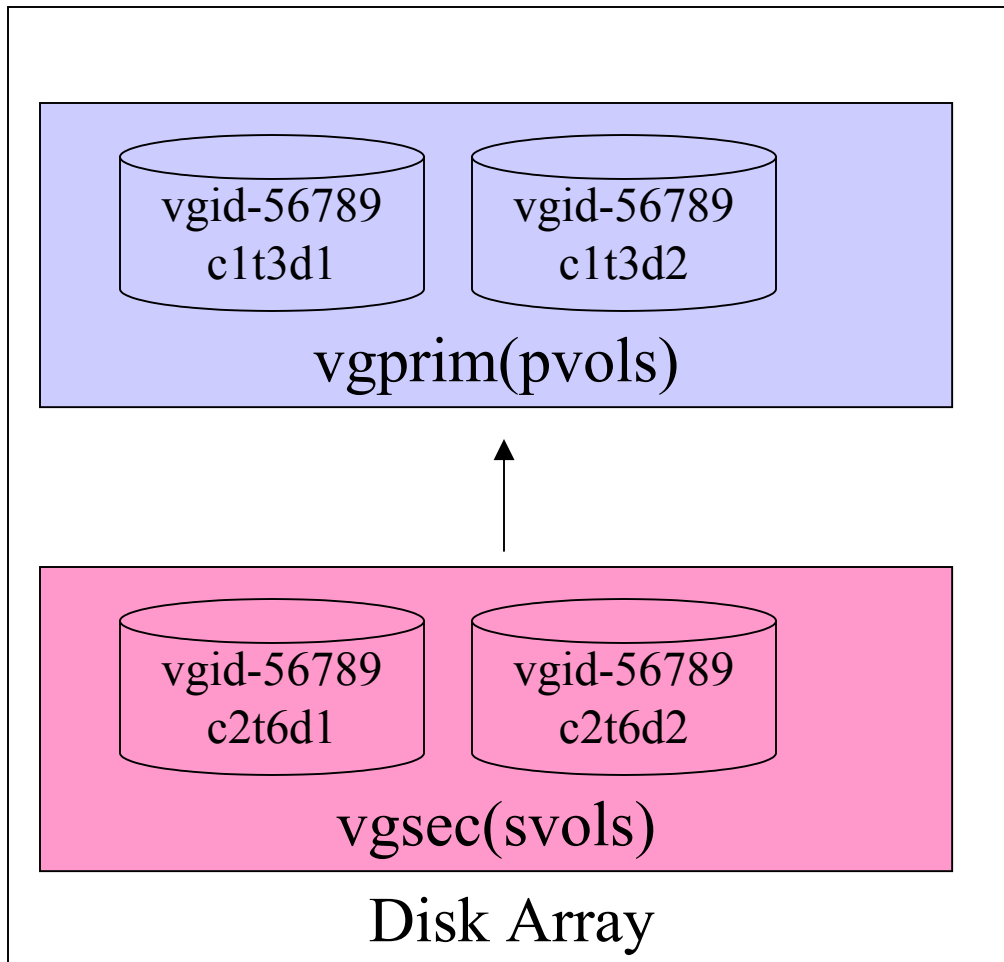


Solution-vgchgid



```
vgchgid /dev/rdisk/c2t6d1  
/dev/rdisk/c2t6d2  
  
mkdir /dev/vgsec  
  
mknod /dev/vgsec/group c 64  
0x020000  
  
vgexport -pm mapfile vgprim  
  
vgimport -m mapfile vgsec  
/dev/dsk/c2t6d1  
/dev/dsk/c2t6d2  
  
vgchange -a y vgsec  
  
vgcfgbackup vgsec
```

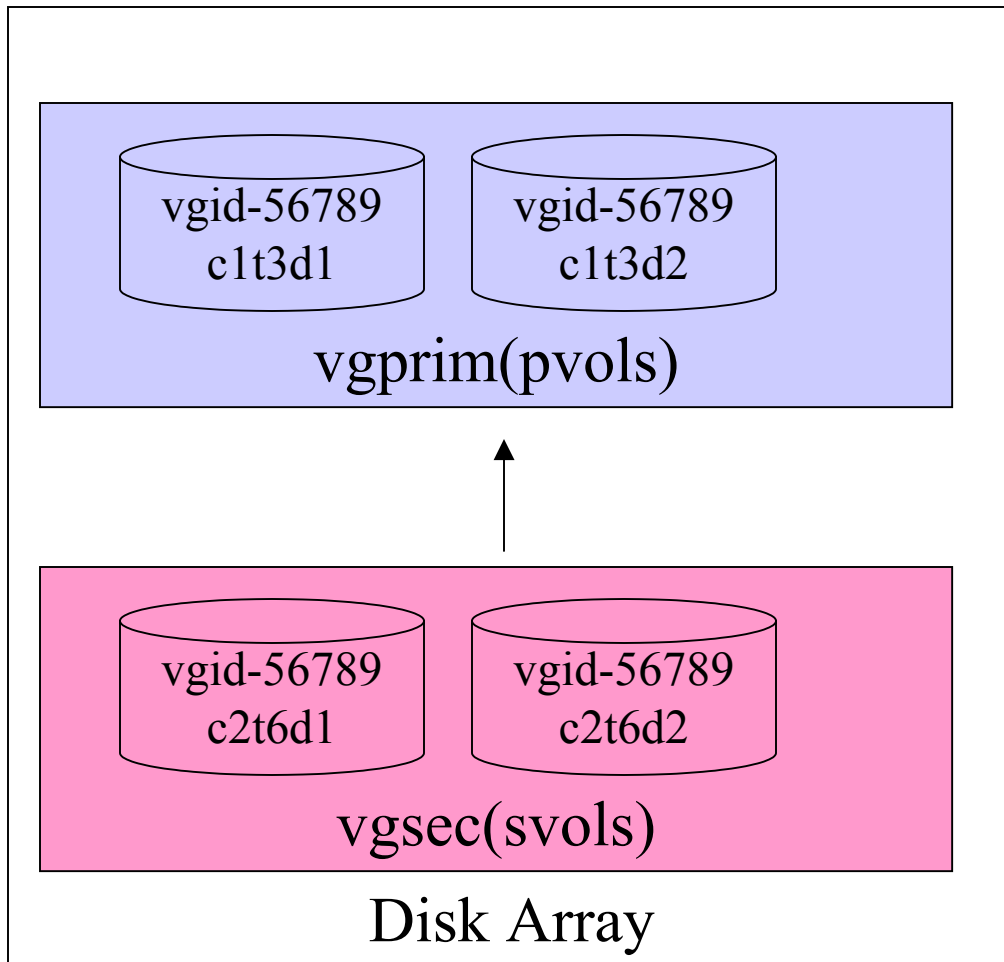
Problem after a reverse sync



/etc/lvmtab

```
/dev/vgpriv (vgid:12345)  
/dev/dsk/c1t3d1  
/dev/dsk/c1t3d2
```

Solution



- export volume group
- reimport volume group

/etc/lvmtab

```
/dev/vgpriv (vgid:56789)
```

```
/dev/dsk/c1t3d1
```

```
/dev/dsk/c1t3d2
```

VxVM Concepts

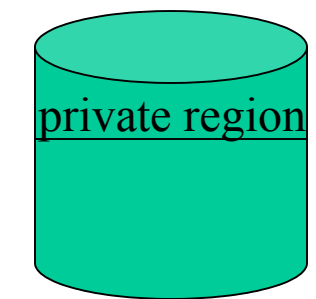
VxVM Intro

- Another Volume Manager Supported on HP-UX platform
- Can co-exist with LVM
- Allows flexible disk space management
- includes vmsa gui as well as cli
- Available from HP beginning with HP-UX 11i. Available from Veritas for other releases
- No rootability until 11i Release 1.5 (11.20)
- Lite & Full versions

VxVM Objects - Disks

Physical Disk:

Disk initialized for VxVM
use



cntndn or
enclosure based
name

Free disk pool

or

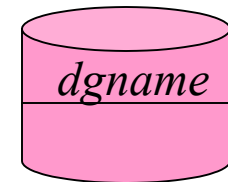
Disk Group - *dgroup*



dgroup01



dgroup02

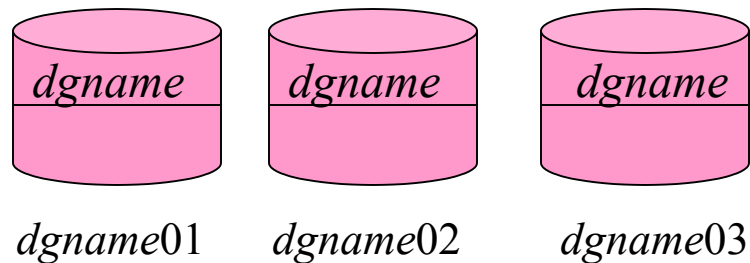


dgroup03

VMdisks

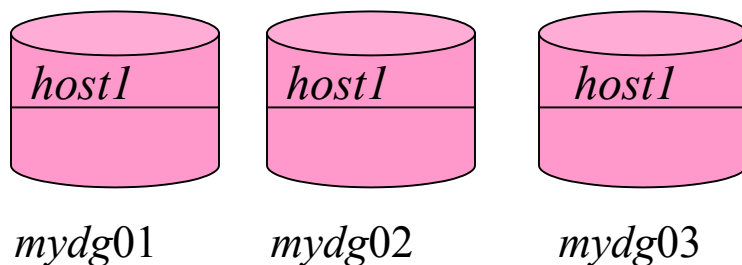
VxVM Concepts-Disk Group

- Similar to an LVM Volume Group
- One or more VMdisks
- Pool of space that can then be "partitioned" into volumes
- Entire group can be moved between systems using import/deport



VxVM Concepts-Disk Group "activation"

- There is no `/etc/lvmtab` file
- `hostid` in private region is compared to `hostid` on system
- `autoimport` flag indicates group should be imported at start up
- `deport` removes `hostid` from private region but leaves disks in tact

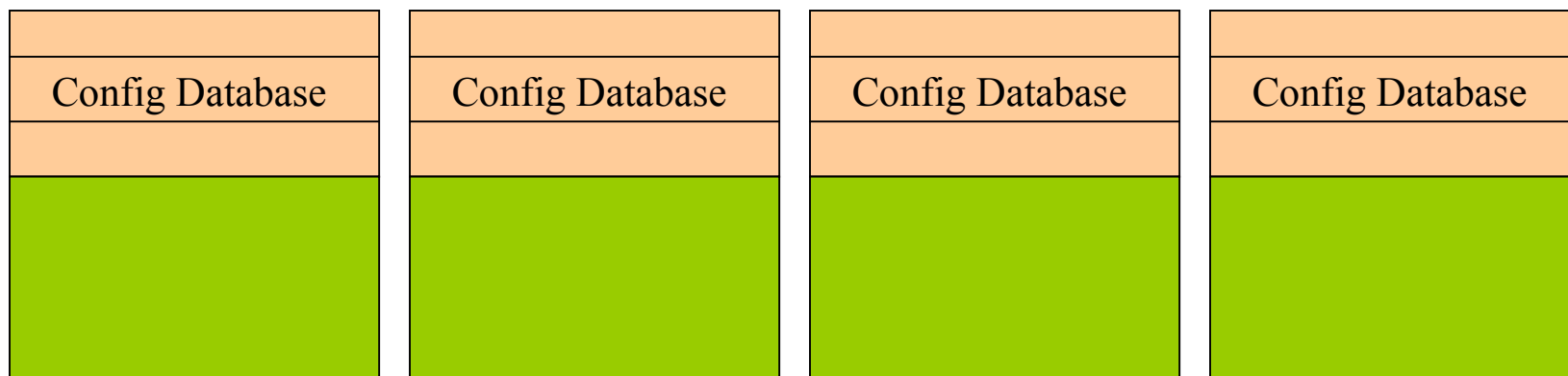


`/etc/vx/volboot`

`host1`

VxVM Concepts-Configuration Database

- Describes all objects contained in the disk group
- Replicated on multiple disks in the disk group (4 by default)



VxVM rootdg

- Required for VxVM to run
- Does not necessarily contain the root file system
 - 11i: will not include the root file system
 - 11i 1.5: will include the root file system
 - future releases will have choice
- rootdg cannot be moved between systems
- discourage including customer data

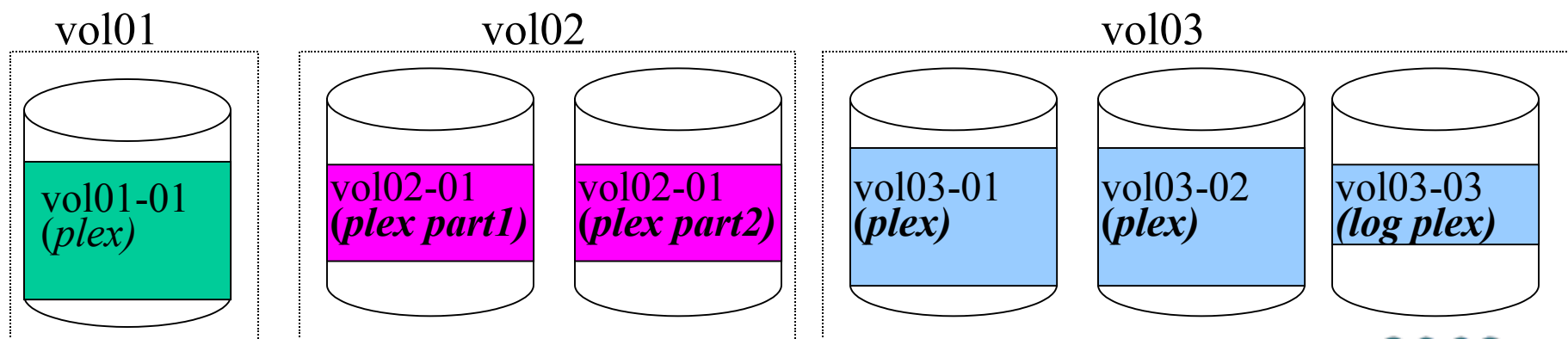
VxVM Concepts-Volume

- Similar to an LVM logical volume
- Analogous to a "partition"
- Can contain a file system, a swap area or raw data
- Represented by device files:

/dev/dsk/dgname/volname /dev/rdisk/dgname/volname

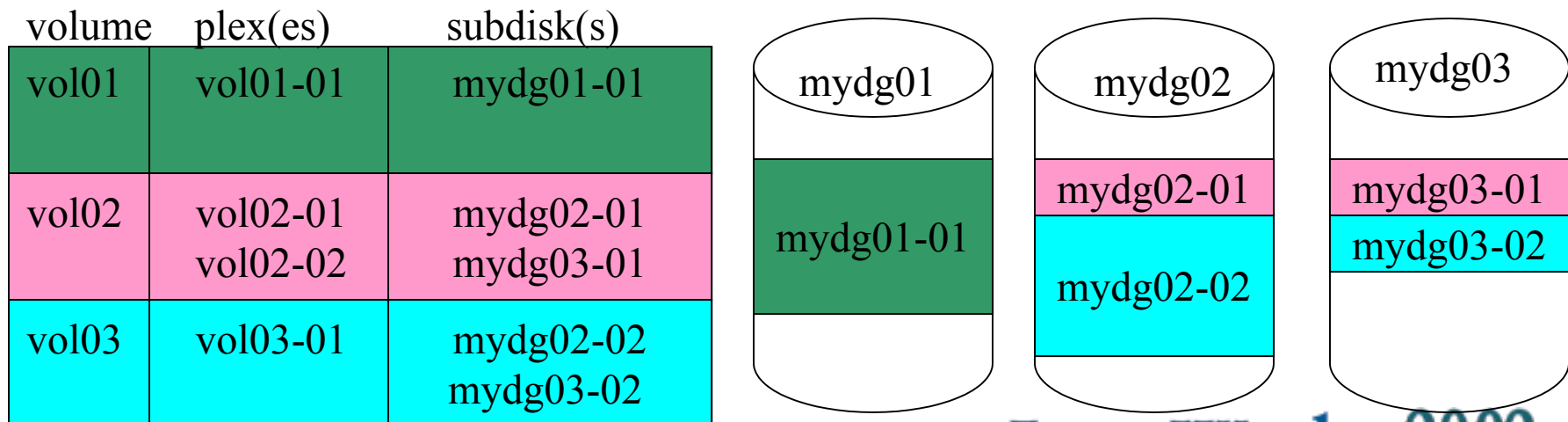
VxVM Concepts-plex

- Data Plex: A copy of the data
- Non-mirrored volumes will have one data plex, mirrored volumes will have multiple data plexes
- a plex can span disks
- log plex: mirrored and raid5 volumes can have a log plex

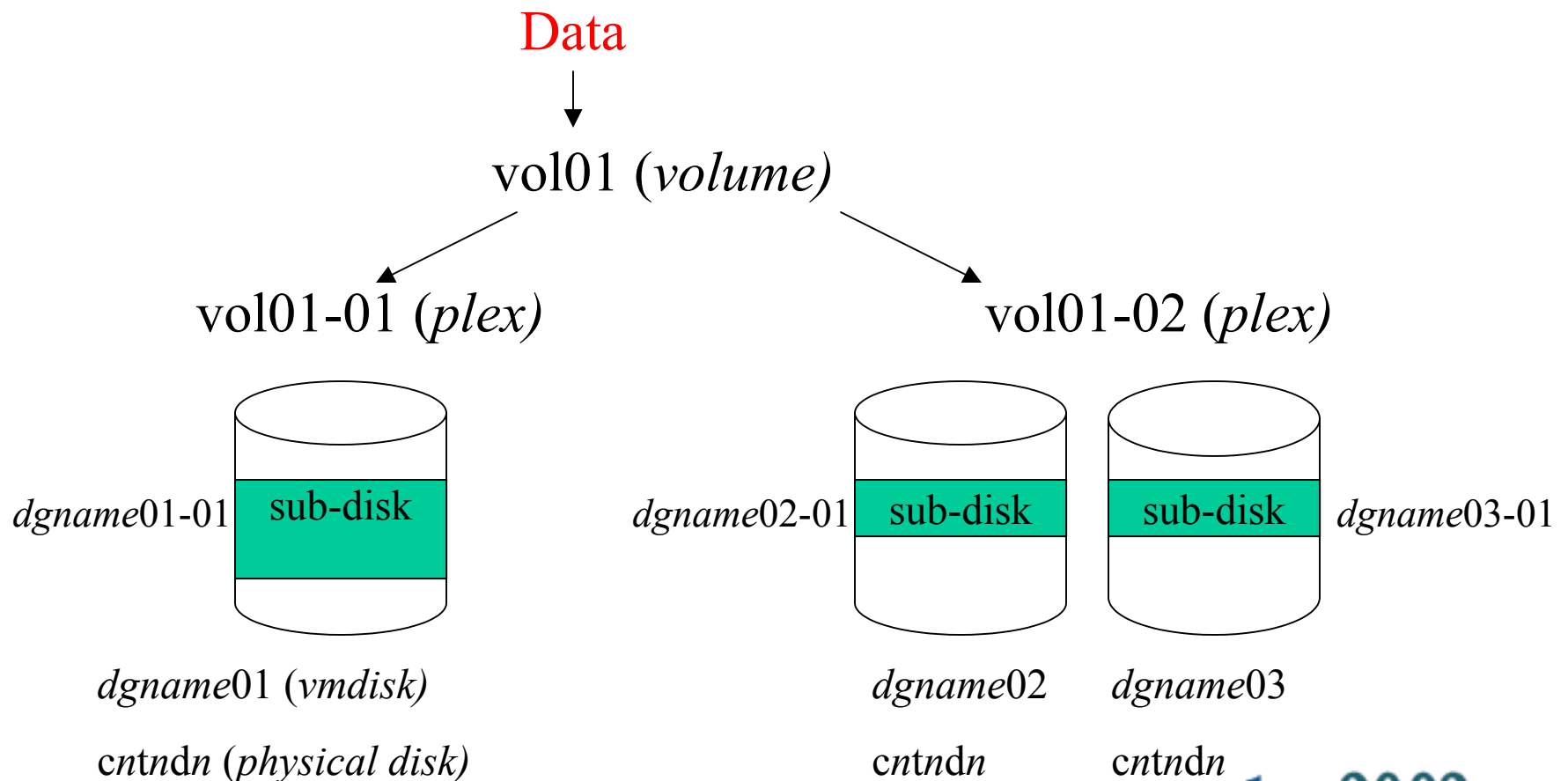


VxVM Concepts-subdisk

- Contiguous space on a single vmdisk
- entire subdisk belongs to a single plex
- usually created when volume is created



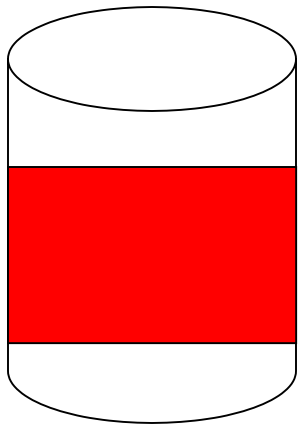
VxVM Objects - Review



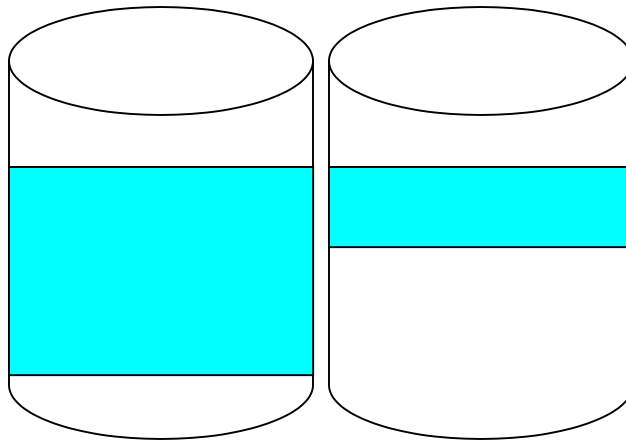
Data Plex Layouts

- Concatenated
- Striped
- raid5

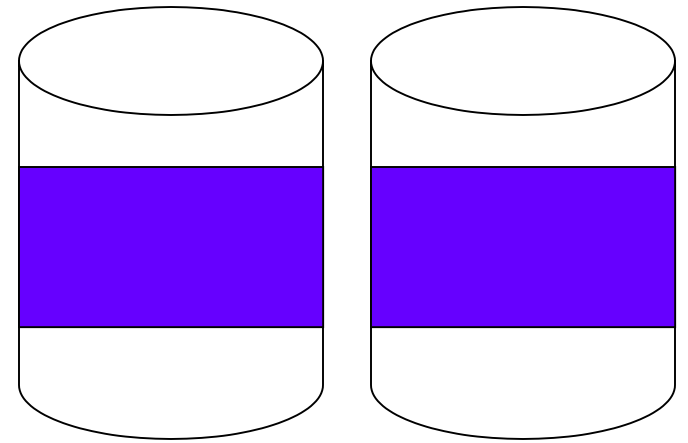
Concatenated Plex Examples



- Simple volume
- One plex
- One subdisk

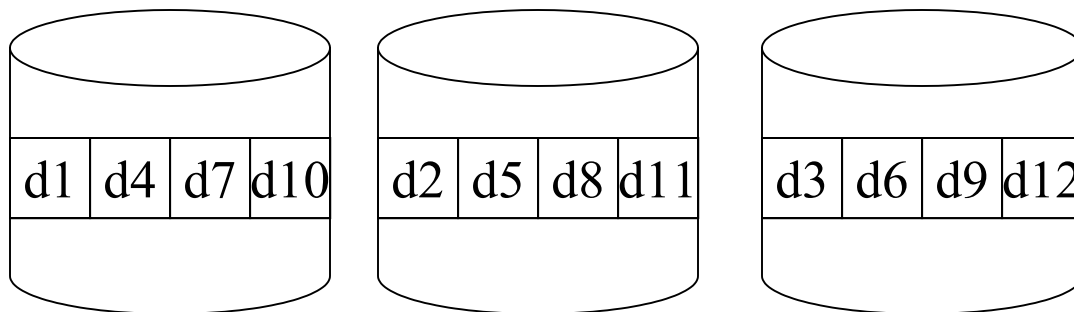


- Spanned Volume
- One plex
- Two subdisks

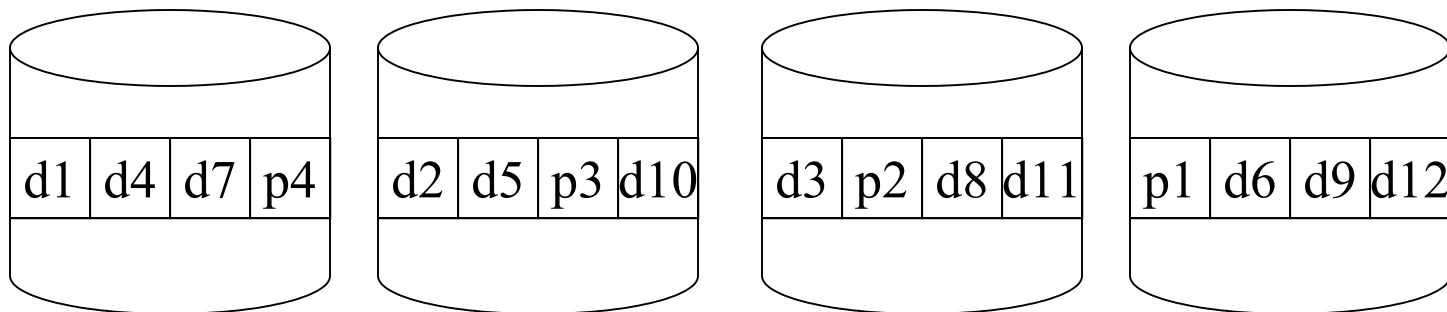


- Mirrored Volume
- Two plexes
- Two subdisks

Striped Plex Layout



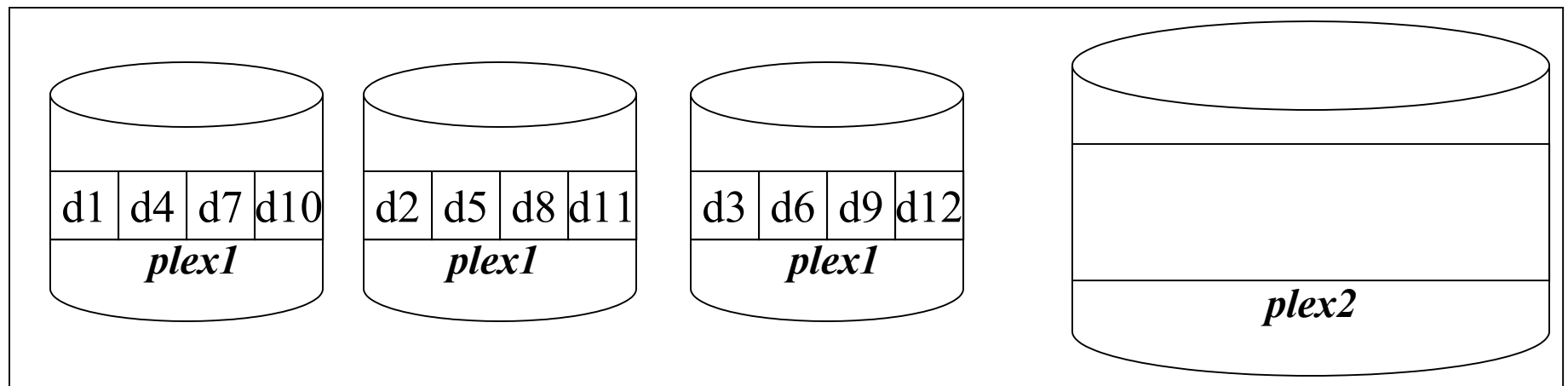
raid5 plex Layout



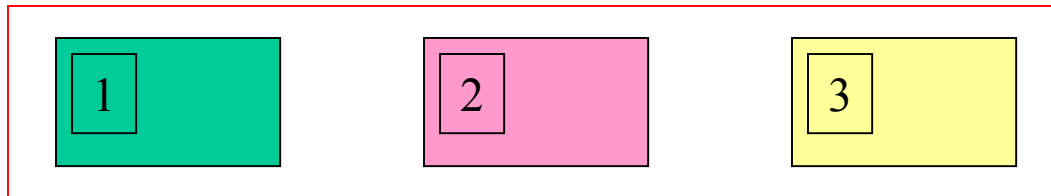
Mixed volumes

- You can combine plex layouts in a single volume
- By default the striped plex will be preferred over the concatenated plex
- You can not mirror raid5 plex layout

mixed volume

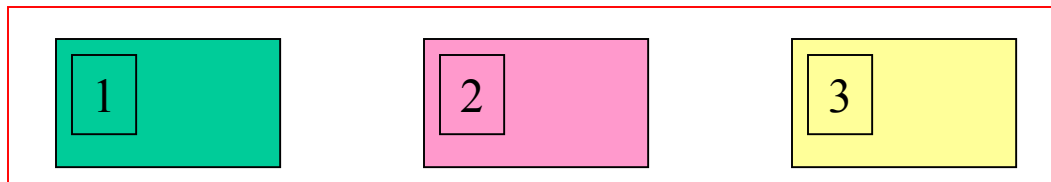


Layered Volumes-Stripe-Mirror

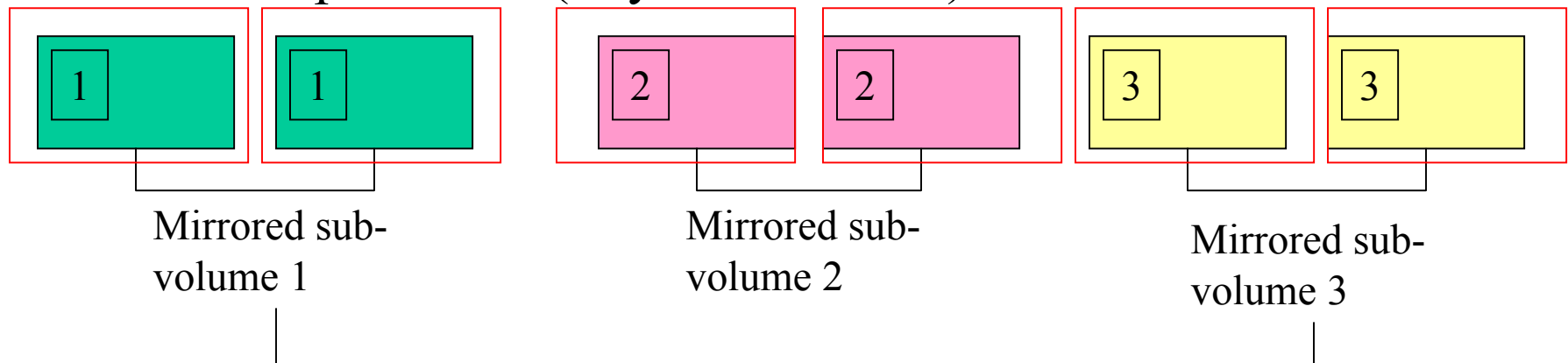


Raid 1+0

Mirror-Stripe



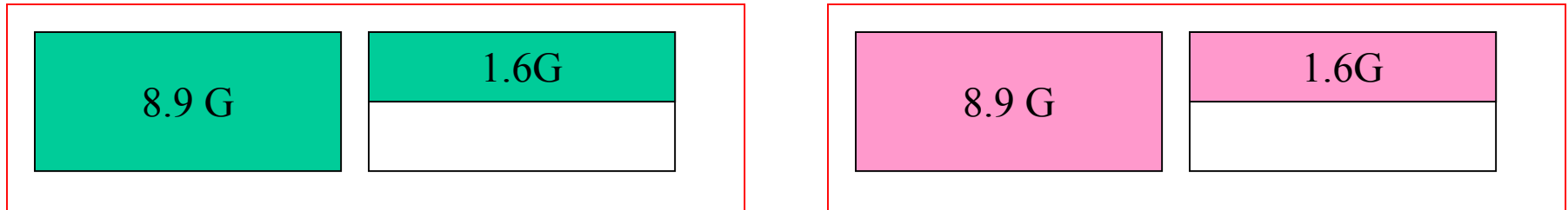
Raid 0+1 Stripe-Mirror (Layered Volume)



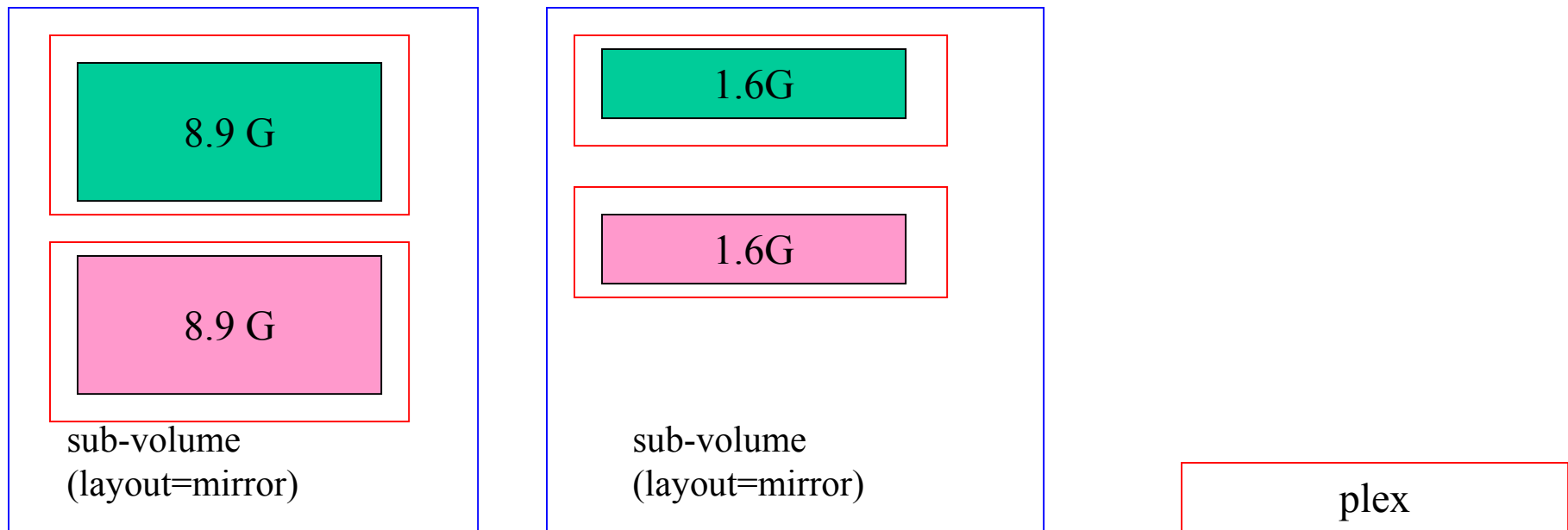
Striped-Volume

Layered Volumes-Concat-pro

Normal Mirror (spanning multiple disks):



Layered Volume (spanning multiple disks):

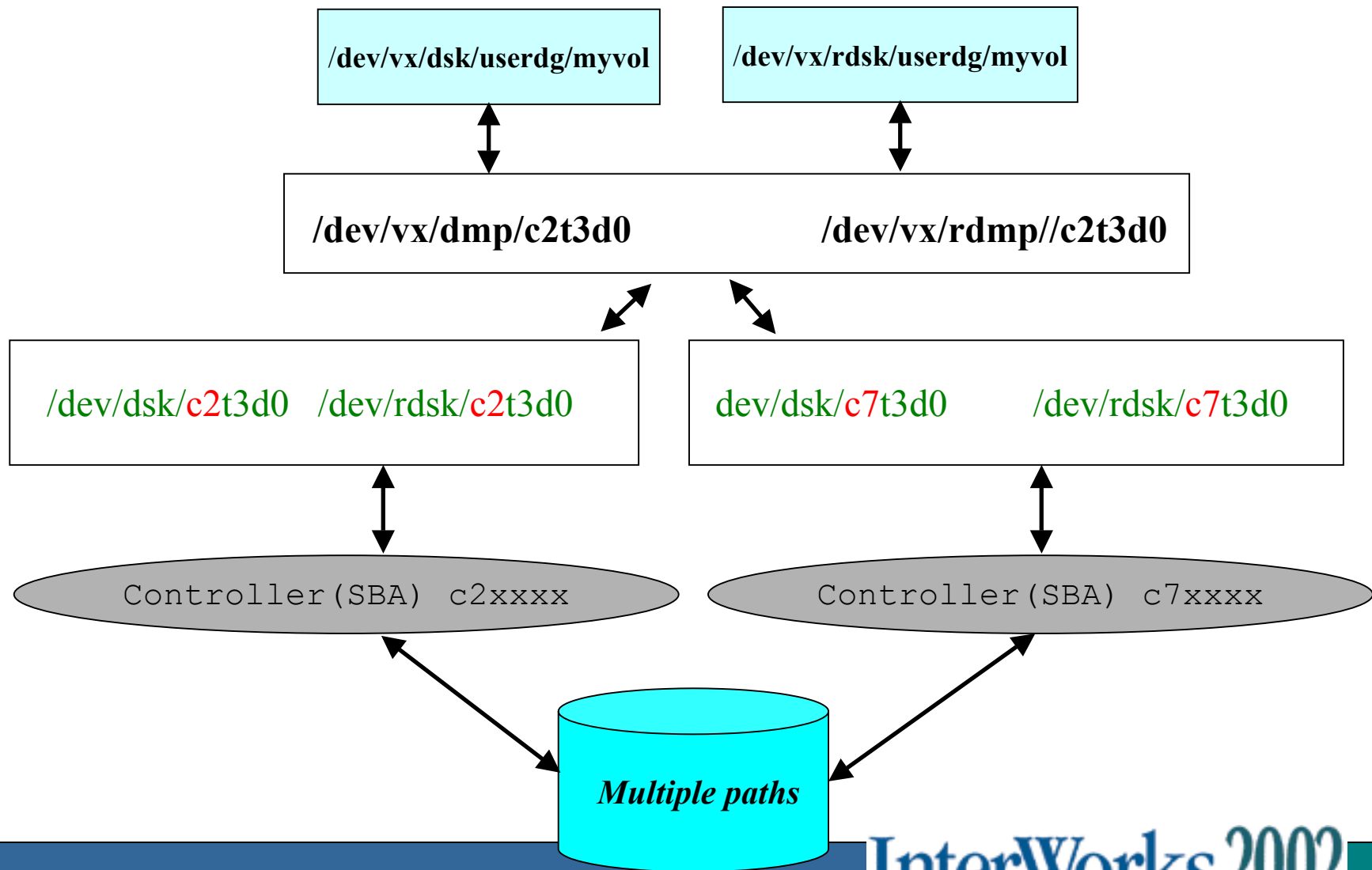


volume with layout=concat

Creating Volumes

- vxassist (cli) or vmsa (gui)
 - you create volume
 - plex, sub-disks automatically created
- bottom up
 - manually create sub-disk(s)
 - specify sub-disk(s) to form plex
 - specify plex(es) to form volume

Dynamic Multi-Pathing



VxVM-LVM Comparisons

LVM

Physical Volume

Physical Volume in a VG

Volume Group

Logical Volume

group of contig. extents

single mirror copy

MWC/MCR

pvra/vgra

extents

VxVM

Physical Disk

VMdisk

Disk Group

Volume

subdisk

plex

dirty region log

private region

public region

Information Sources

- H6285S Hands-On with LVM course
- H6487S Hands-On with MC/ServiceGuard course
- H5875S HP-UX System and Network Admin for Experienced UNIX System Administrators
- U1592AAE Intro to VxVM - HP Education Virtual Classroom Seminar

www.hp.com/education

- Various white papers and cookbooks from the Response Center