



# Tru64 UNIX TruCluster Server LSM and EVA

Thomas Aussmann  
Consultant Proactive Services  
Hewlett-Packard GmbH  
thomas.aussmann@hp.com

© 2004 Hewlett-Packard Development Company, L.P.  
The information contained herein is subject to change without notice



4/20/2004



## Agenda

- Overview
- Best Practices
- Configuration
- Examples
- Resources

www.decus.de

2

4/20/2004




## Overview

- Logical Storage Manager - LSM
  - Software RAID solution
  - Located between device-driver and filesystem
  - Host based I/O
  - RAID-5 support
  - Spare disk support

www.decus.de 3

4/20/2004




## Overview

- Logical Storage Manager - LSM
  - Increased read performance
  - Double I/O on writes
  - More complex disaster recovery
- LSM within cluster
  - Root CFS support starts with Tru64 UNIX V5.1A
  - Not supported for member boot partition
  - Not supported for quorum disk
  - RAID 5 volumes not supported

www.decus.de 4

4/20/2004




## Overview

- Continuous Access - CA
  - No host based I/O
  - Data replication over dedicated SAN link(s)
  - Bi-directional data replication
  - Install latest firmware version

www.decus.de

5

4/20/2004

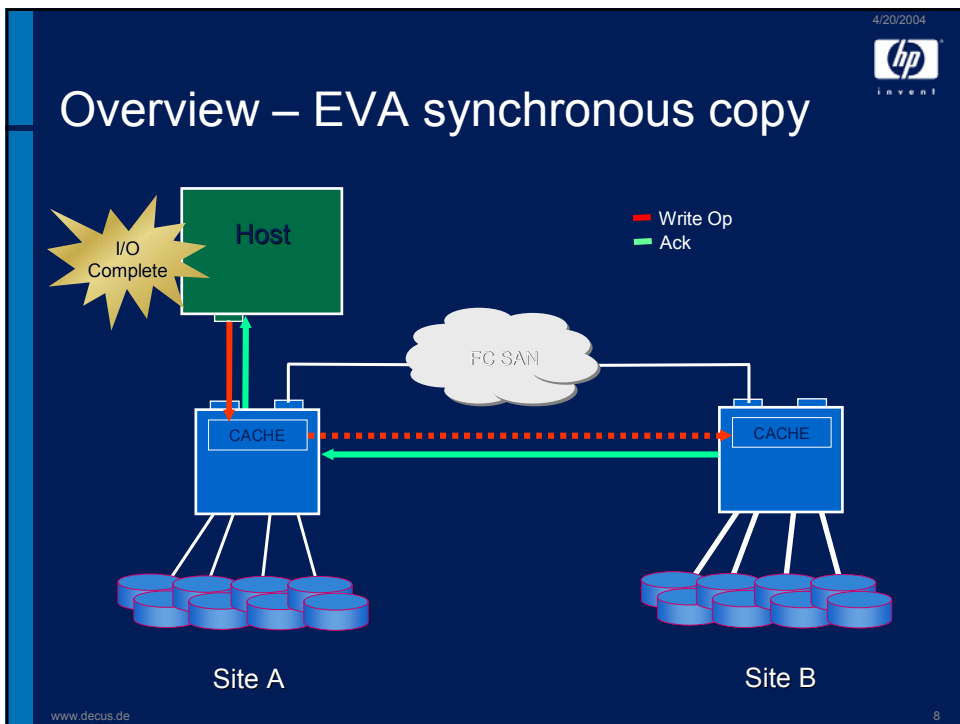
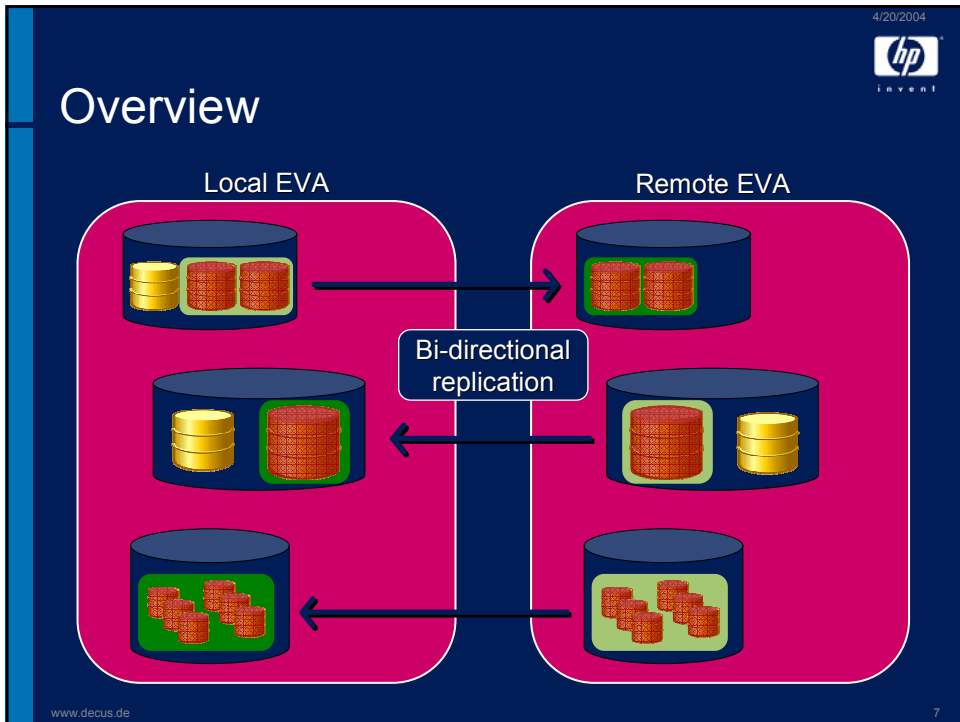



## Overview

- Continuous Access - CA
  - Increased response time on writes
    - Remote data replication latency (cache only) added
    - Synchronous copy method on EVA
    - Synchronous/asynchronous copy method on XP
  - Decreased I/O performance after disaster recovery
    - Like LSM, (re)synch is needed

www.decus.de

6




4/20/2004  


## Overview

- Single Point Of Failure
  - LSM
    - Member down (loss of boot-disk)
    - Cluster down (loss of boot- and quorum disk)
    - Applications/data available if cluster survives
  - CA
    - Cluster down
    - Manual switch needed

www.decus.de 9


4/20/2004  


## Best Practices

- LSM
  - No aligning problem
  - No LSM spare disk(s) needed
  - One disk for more than one volume possible
  - LSM striping not recommended

www.decus.de 10

4/20/2004



## Best Practices

- LSM
  - Configuration database per diskgroup
    - Automatically managed number and location
    - May all be placed within one cabinet
    - Check location with „voldg list <diskgroup>“
  - Manually determine distribution of configurations
    - „voldisk moddb dsknnn nconfig=0“
    - „voldisk moddb dsknnn nconfig=1“
    - Disable automatic load balancing
      - modify /sbin/lsmbootstrap
      - „vold\_opts=-k -x noloadbalance -x noautoconfig“
      - For rootdg, add disks with „voldctl add disk dsknnn“

www.decus.de 11

4/20/2004




## Best Practices

- EVA
  - Build diskgroups with multiple of 8 physical disks
  - Build several diskgroups to distribute
    - System disks and application binaries
    - Database files
    - Redo and archiv files
    - .....

www.decus.de 12

4/20/2004




**Best Practices**

- EVA
  - Initialize disks before use
    - Use dd to zero disk
    - Avoids performance issue on first write
  - Within cluster
    - Distribute member boot disks on both EVA's
    - Create quorum disk on MSA1000 or HSG80
    - Duplicate boot disk for members

www.decus.de 13


4/20/2004



**Best Practices**

- Configuration
  - Keep initial installation disk
  - Configure LSM before cluster creation
  - Have LSM configuration information available
    - volprint, sys\_check
  - Save LSM configuration regularly
    - volsave
  - Use DRL
    - Speed up recovery times
    - Impact on performance


www.decus.de 14

4/20/2004 

## Configuration

- Increased size for log subdisks in cluster
  - 65 KB instead of 2 KB per GB
- Member share a common LSM configuration
  - Configuration can be managed from any member
  - Any member can handle LSM I/O directly
  - Symmetric I/O model
- No additional LSM I/O within cluster
- Private „in memory“ DRL per member

www.decus.de 15


4/20/2004 

## Configuration

- Increased cluster activity
  - Keep „in memory“ private structures consistent
  - Install latest patchkit (BL24)
    - Performance enhancements for CLSM
- Voldisk list can give different results
  - Only for disks not part of LSM
  - Typically limited to disabled disk groups
- Volstat statistics only refer to member executed on

www.decus.de 16




4/20/2004  


## Configuration

- different setups for system related CFS
  - One sliced disk for each volume
    - Easy access to AdvFS within own partition
  - Same disk but own simple partition per volume
    - Manipulate disklabel to skip private region
  - All volumes within one sliced or simple disk/partition
    - Even more complex, but default


www.decus.de 17

4/20/2004  


## Configuration

- Different offsets for AdvFS depends on
  - Disklayout and partitions used
  - LSM sliced or simple disk
  - Single or multiple LSM volumes per media
- Known offsets
  - 16 blocks for disklabel and bootstrap info
  - 4096 blocks for LSM configuration data


www.decus.de 18

4/20/2004 

## Configuration

- LSM not yet configured
  - Run volsetup on initial/any cluster member
  - Run volsetup –s on existing other members only
  - New members are automatically configured
  - Use the same connectivity for LSM volumes
    - Disk group on same bus or member


www.decus.de 19

4/20/2004 

## Configuration

- With LSM already configured
  - Initial LSM configuration is propagated to cluster
  - All LSM volumes are available on new cluster
    - System-related volumes must be explicitly mounted
  - DRL will be disabled if logdisk is too small
    - Remove old log subdisk, add new one
  - New members are automatically configured


www.decus.de 20

4/20/2004  


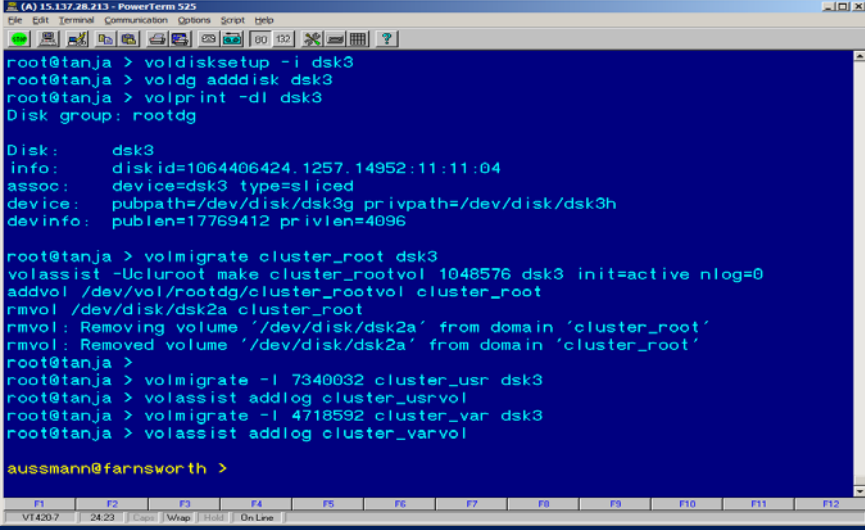
## Configuration

- Use volmigrate to convert AdvFS to LSM volumes
  - volmigrate must be used for cluster\_root
  - For cluster\_usr/\_var use either volmigrate or volencap
  - Never use „volrootmir“ within cluster
  - Use „volassist mirror <volume> <diskmedia>“
- Member specific
  - You may encapsulate primary swap
  - You may add LSM mirrored secondary swap
    - volume set start\_opts=norecov <swapvol>

www.decus.de21

4/20/2004  


## Examples



```


root@tanja > voldisksetup -i dsk3
root@tanja > voldg adddisk dsk3
root@tanja > volprint -dl dsk3
Disk group: rootdg

Disk:      dsk3
info:     diskid=1064406424.1257.14952:11:11:04
assoc:    device=dsk3 type=sliced
device:   pubpath=/dev/disk/dsk3g privpath=/dev/disk/dsk3h
devinfo:  publen=17769412 privlen=4096

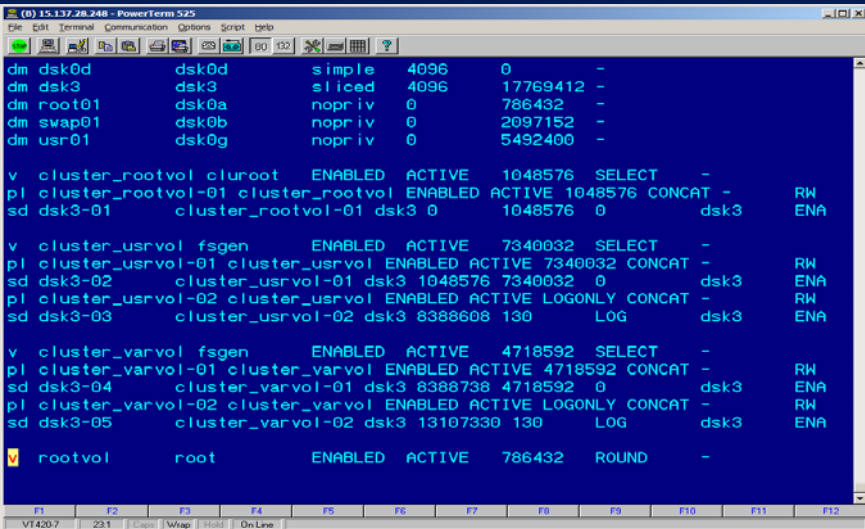
root@tanja > volmigrate cluster_root dsk3
volassist -Ucluroot make cluster_rootvol 1048576 dsk3 init=active nlog=0
addvol /dev/vol/rootdg/cluster_rootvol cluster_root
rmvol /dev/disk/dsk2a cluster_root
rmvol: Removing volume '/dev/disk/dsk2a' from domain 'cluster_root'
rmvol: Removed volume '/dev/disk/dsk2a' from domain 'cluster_root'
root@tanja >
root@tanja > volmigrate -l 7340032 cluster_usr dsk3
root@tanja > volassist addlog cluster_usrvol
root@tanja > volmigrate -l 4718592 cluster_var dsk3
root@tanja > volassist addlog cluster_varvol

aussmann@farnsworth >
    
```

www.decus.de22

4/20/2004  


## Examples



```

dm dsk0d      dsk0d      simple  4096    0      -
dm dsk3       dsk3       sliced  4096    17769412 -
dm root01    dsk0a      nopriv  0        786432 -
dm swap01    dsk0b      nopriv  0        2097152 -
dm usr01     dsk0g      nopriv  0        5492400 -


v cluster_rootvol clurroot  ENABLED ACTIVE  1048576 SELECT -
pl cluster_rootvol-01 cluster_rootvol ENABLED ACTIVE 1048576 CONCAT - RW
sd dsk3-01     cluster_rootvol-01 dsk3 0      1048576 0      dsk3  ENA

v cluster_usrvol fsgen      ENABLED ACTIVE  7340032 SELECT -
pl cluster_usrvol-01 cluster_usrvol ENABLED ACTIVE 7340032 CONCAT - RW
sd dsk3-02     cluster_usrvol-01 dsk3 1048576 7340032 0      dsk3  ENA
pl cluster_usrvol-02 cluster_usrvol ENABLED ACTIVE LOGONLY CONCAT - RW
sd dsk3-03     cluster_usrvol-02 dsk3 8388608 130    LOG    dsk3  ENA

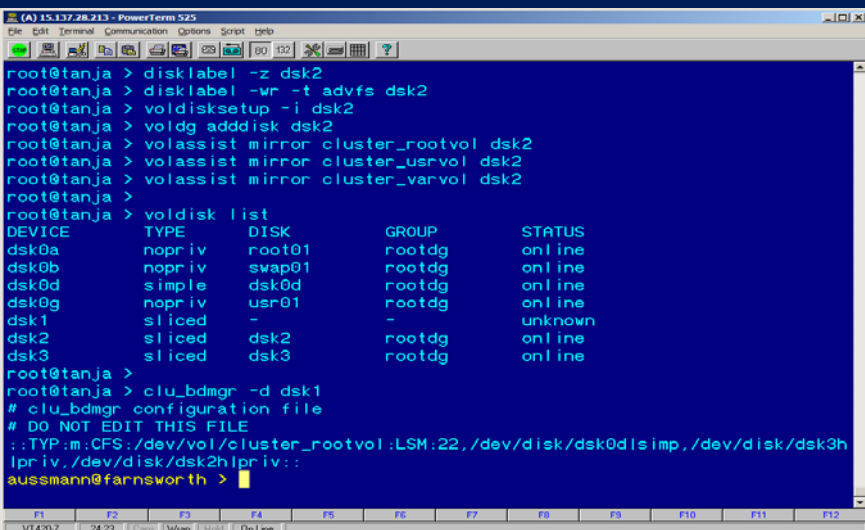
v cluster_varvol fsgen      ENABLED ACTIVE  4718592 SELECT -
pl cluster_varvol-01 cluster_varvol ENABLED ACTIVE 4718592 CONCAT - RW
sd dsk3-04     cluster_varvol-01 dsk3 8388738 4718592 0      dsk3  ENA
pl cluster_varvol-02 cluster_varvol ENABLED ACTIVE LOGONLY CONCAT - RW
sd dsk3-05     cluster_varvol-02 dsk3 13107330 130    LOG    dsk3  ENA

rootvol      root        ENABLED ACTIVE  786432  ROUND -
    
```

www.decus.de 23

4/20/2004  


## Examples




```

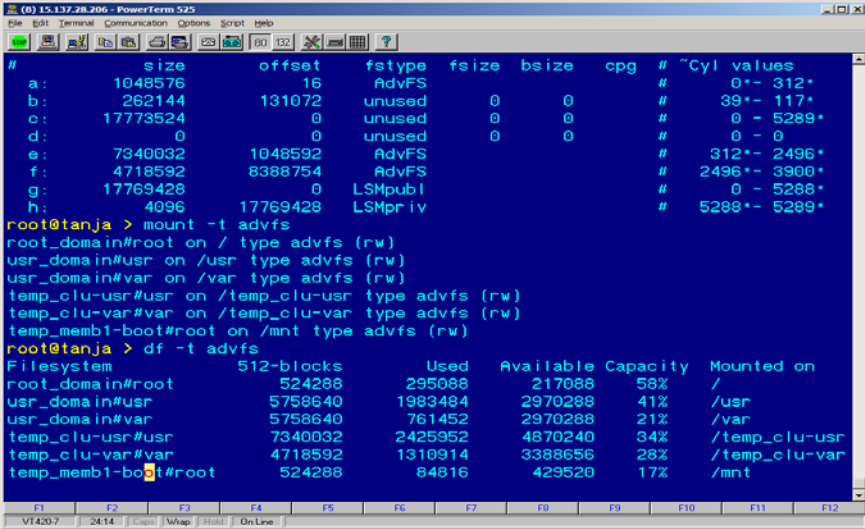
root@tanja > disklabel -z dsk2
root@tanja > disklabel -wr -t advfs dsk2
root@tanja > voldisksetup -i dsk2
root@tanja > voldg adddisk dsk2
root@tanja > volassist mirror cluster_rootvol dsk2
root@tanja > volassist mirror cluster_usrvol dsk2
root@tanja > volassist mirror cluster_varvol dsk2
root@tanja >
root@tanja > voldisk list
DEVICE      TYPE      DISK          GROUP        STATUS
dsk0a      nopriv   root01       rootdg       online
dsk0b      nopriv   swap01       rootdg       online
dsk0d      simple   dsk0d        rootdg       online
dsk0g      nopriv   usr01        rootdg       online
dsk1       sliced   -            -            unknown
dsk2       sliced   dsk2         rootdg       online
dsk3       sliced   dsk3         rootdg       online

root@tanja >
root@tanja > clu_bdmgr -d dsk1
# clu_bdmgr configuration file
# DO NOT EDIT THIS FILE
:: TYP:m;CFS:/dev/vol/cluster_rootvol:LSM:22,/dev/disk/dsk0dls imp,/dev/disk/dsk3h
lpr iv,/dev/disk/dsk2h lpr iv::
aussmann@farnsworth >
    
```

www.decus.de 24

4/20/2004 

## Examples



```


#          size      offset  fstype  fsize  bsize  cpg  # ~Cyl values
a:      1048576         16   AdvFS
b:      262144      131072  unused    0     0   #      39* - 117*
c:     17773524         0  unused    0     0   #      0 - 5289*
d:         0         0  unused    0     0   #      0 - 0
e:     7340032     1048592  AdvFS
f:     4718592     8388754  AdvFS
g:     17769428         0  LSMpubl
h:         4096     17769428  LSMpriv
#      5288* - 5289*

root@tanja > mount -t advfs
root_domain#root on / type advfs (rw)
usr_domain#usr on /usr type advfs (rw)
usr_domain#var on /var type advfs (rw)
temp_clu-usr#usr on /temp_clu-usr type advfs (rw)
temp_clu-var#var on /temp_clu-var type advfs (rw)
temp_memb1-boot#root on /mnt type advfs (rw)

root@tanja > df -t advfs
Filesystem          512-blocks      Used    Available Capacity    Mounted on
root_domain#root          524288      295088      217088      58%      /
usr_domain#usr           5758640     1983484     2970288      41%      /usr
usr_domain#var           5758640       761452     2970288      21%      /var
temp_clu-usr#usr         7340032     2425952     4870240      34%      /temp_clu-usr
temp_clu-var#var         4718592     1310914     3388656      28%      /temp_clu-var
temp_memb1-boott#root     524288       84816      429520       17%      /mnt

```

www.decus.de 25

4/20/2004 

## Resources

- Tru64 UNIX Best Practices documentation
  - [http://h30097.www3.hp.com/docs/best\\_practices](http://h30097.www3.hp.com/docs/best_practices)
- Tru64 UNIX V5.1B online documentation sets
  - [http://h30097.www3.hp.com/docs/pub\\_page/doc\\_list.html](http://h30097.www3.hp.com/docs/pub_page/doc_list.html)
- TruCluster Server V5.1B online documentation sets
  - [http://h30097.www3.hp.com/docs/pub\\_page/cluster\\_list.html](http://h30097.www3.hp.com/docs/pub_page/cluster_list.html)
- EVA Design Workshops

www.decus.de 26

