Tru64 UNIX TruCluster Server LSM and EVA

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

Agenda

- Overview
- Best Practices
- Configuration
- Examples
- Resources

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

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

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

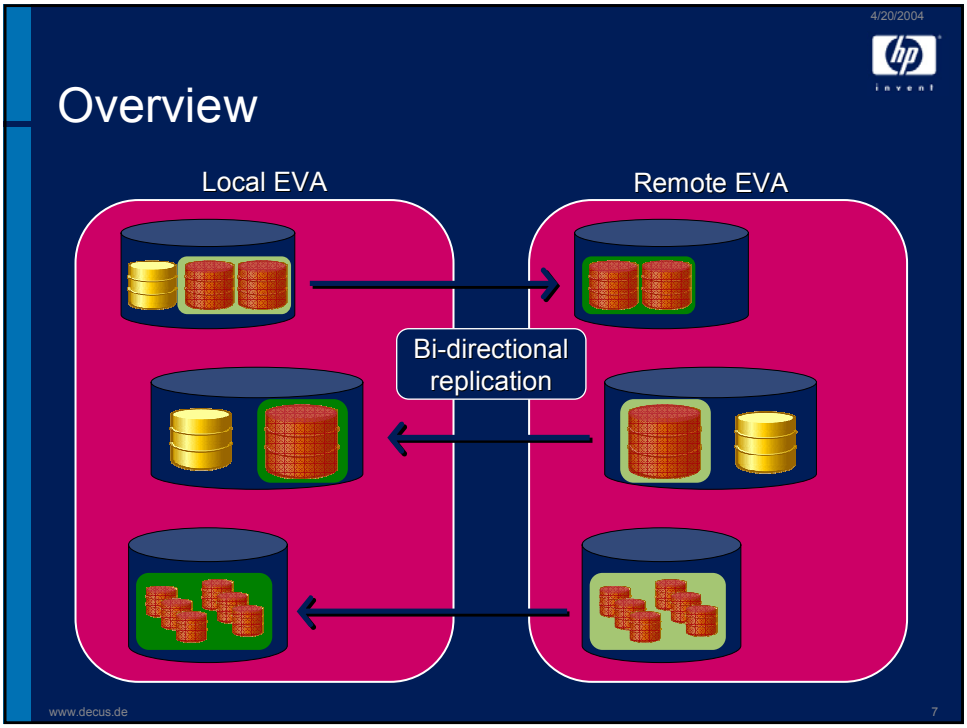
www.decus.de 5

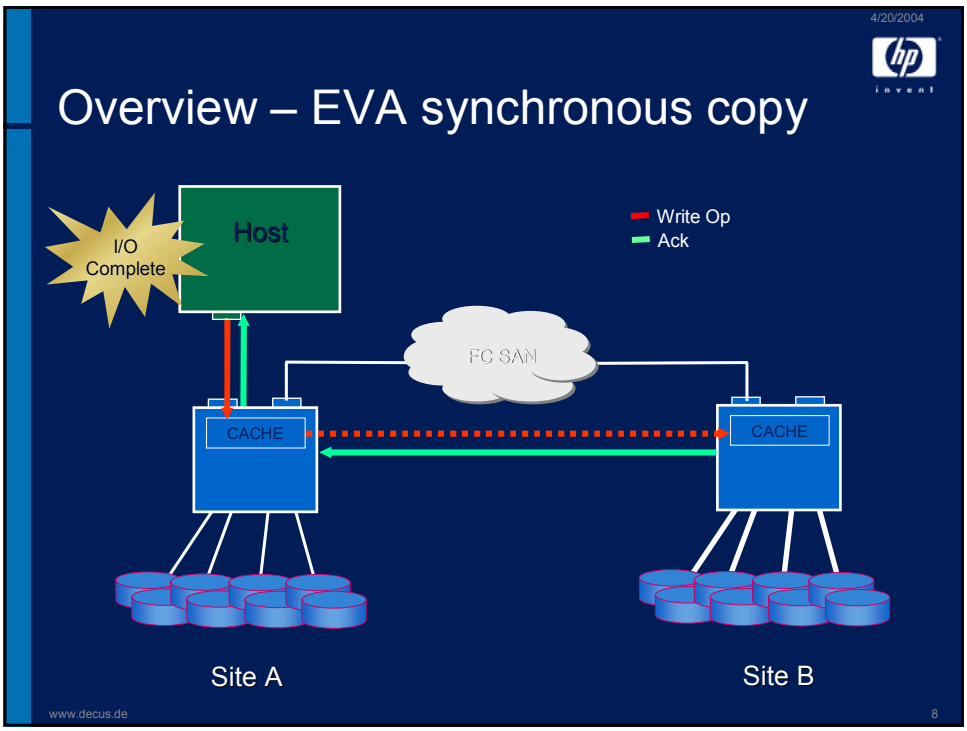
4/20/2004

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
Local EVA
Remote EVA
Bi-directional replication
www.decus.de
7


4/20/2004
Overview – EVA synchronous copy
I/O Complete
Host
Write Op
Ack
FC SAN
CACHE
CACHE
Site A
Site B
www.decus.de
8

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 dsk*nnn* nconfig=0“
    - „voldisk moddb dsk*nnn* nconfig=1“
    - Disable automatic load balancing
      - modify /sbin/lsmbstartup
      - „vold_opts=-k -x noloadbalance -x noautoconfig“
      - For rootdg, add disks with „voldctl add disk dsk*nnn*“

www.decus.de 11

4/20/2004

Best Practices

- EVA
  - Build diskgoups 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 perfomance 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

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

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

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

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

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

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

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 to small
    - Remove old log subdisk, add new one
  - New members are automatically configured

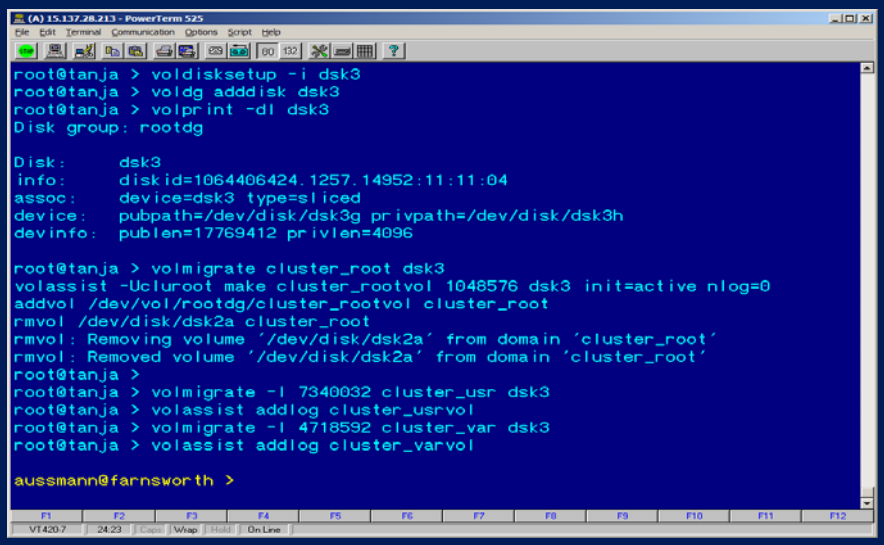
4/20/2004

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>

4/20/2004

Examples



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 cluroot       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

v  rootvol        root           ENABLED  ACTIVE   786432   ROUND     -
```

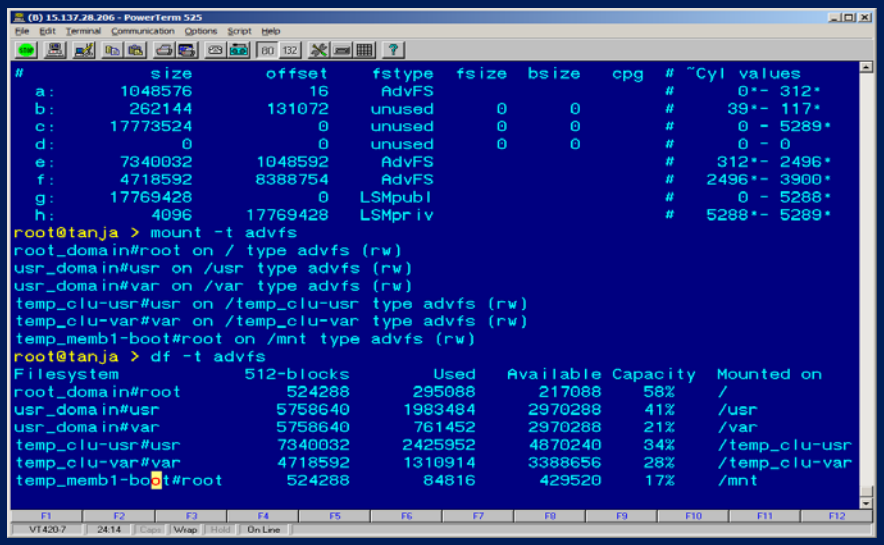

4/20/2004

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/dsk0dlsimp,/dev/disk/dsk3h
lpriv,/dev/disk/dsk2hlpriv::
aussmann@farnsworth >
```

4/20/2004

Examples



www.decus.de 25

4/20/2004

Resources

- Tru64 UNIX Best Practices documentation
  - 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
- TruCluster Server V5.1B online documentation sets
  - http://h30097.www3.hp.com/docs/pub_page/cluster_list.html
- EVA Design Workshops

www.decus.de 26


invent