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

Agenda
- Overview
- Best Practices
- Configuration
- Examples
- Resources
Overview

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

- 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
Overview

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

- 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
Overview

Bi-directional replication

Local EVA --> Remote EVA

Overview – EVA synchronous copy

I/O Complete

Write Op

Ack

Site A

Site B
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

Best Practices

• LSM
  − No aligning problem
  − No LSM spare disk(s) needed
  − One disk for more than one volume possible
  − LSM striping not recommended
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 dskn nconfig=0"
    - "voldisk moddb dskn nconfig=1"
    - Disable automatic load balancing
      - modify /sbin/lsmbstartup
      - "vold_opts=-k -x noloadbalance -x noautoconfig"
      - For rootdg, add disks with "voldctl add disk dskn"

- 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
    - ...
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

• 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
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

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
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

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
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

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
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>

Examples

```bash
root@tanjana:~# volmigrate cluster_root dsk3
volassist volrootmir root cluster_root 1640575dsk3 active
Addvol /dev/vol/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@tanjana:~# volmigrate -l 734002 dsk5
root@tanjana:~# volassist adddisk cluster_usrvol
root@tanjana:~# volmigrate -l 4718532 cluster_varvol
```
Examples
Examples

```
root@tanja:~# mount -t advfs
root@tanja:~# mount -t advfs/root
root@tanja:~# mount -t advfs/usr
root@tanja:~# mount -t advfs/var
root@tanja:~# mount -t advfs/temp
root@tanja:~# mount -t advfs/swap

root@tanja:~# df -t advfs
```

Resources

- Tru64 UNIX Best Practices documentation
- Tru64 UNIX V5.1B online documentation sets
- TruCluster Server V5.1B online documentation sets
- EVA Design Workshops