



Cognizant

Presented in NYOUG Spring Conference on 04/14/2016

Exadata - Experiences

If we had to do it again...

Compute Nodes crashing - NetBackup Backups on OS

- Compute Node Servers started crashing left & right – in the middle of a business day
- Crash Dump Analysis showed NBU software interacting with Mega RAID driver for OS backup causing kernel panic

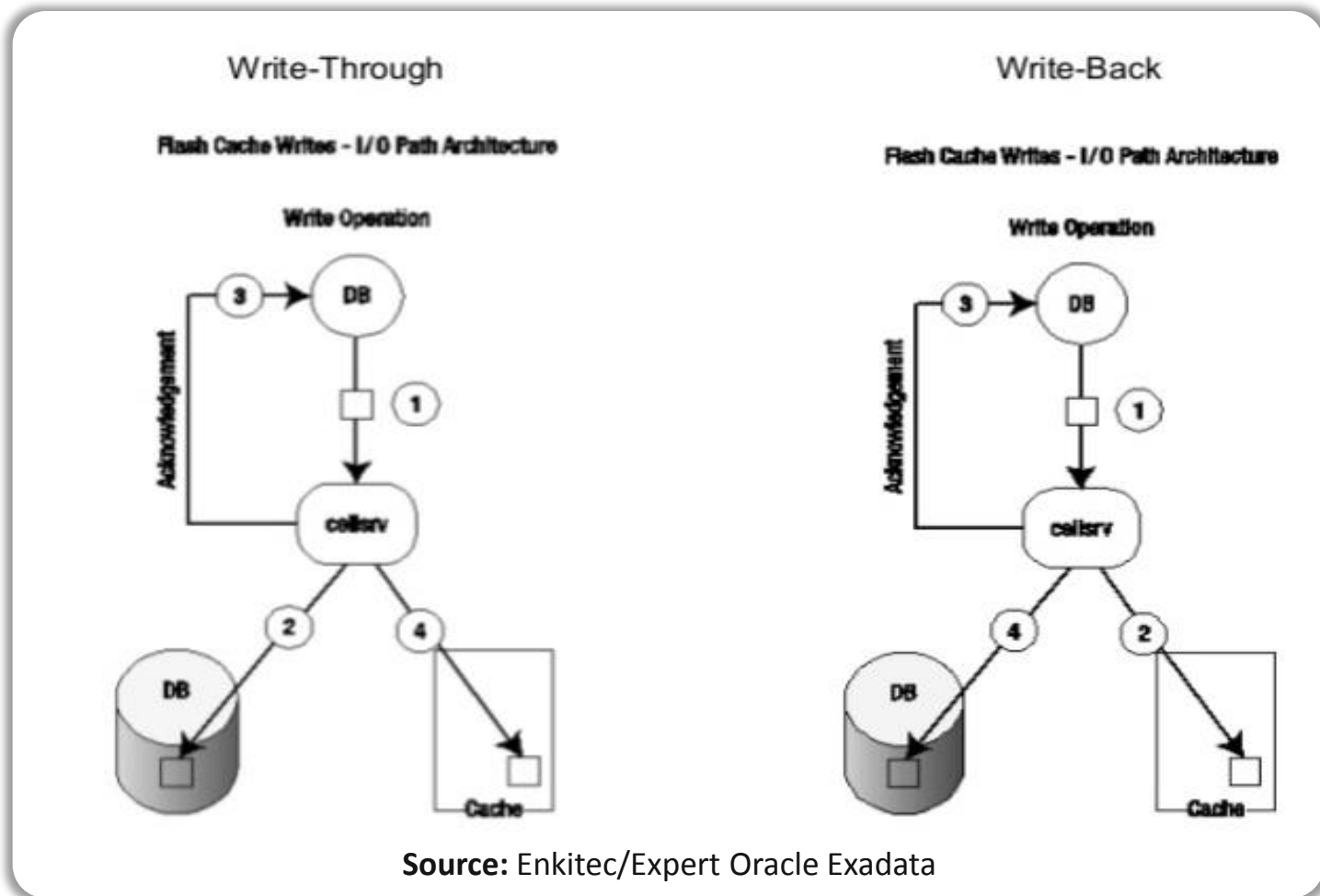
```
crash> log
```

```
Process bpbkar (pid: 23130, threadinfo ffff8112eff5c000, task ffff8112f88b2040)
Stack: 3934393637323935 ffff81187e233432 ffff8112eff5dea8 ffffffff8000d3ba
ffff81187e292d80 ffff81119dfda6e0 ffff8112eff5dea8 ffffffff8000a604
ffff81187e14b280 0000010100000000 0000000b3c11ba4c ffff81187bf0d480
Call Trace:
[<fffffff8000d3ba>] dput+0x2c/0x114
[<fffffff8000a604>] __link_path_walk+0xdf8/0xf42
[<fffffff8001a7cf>] vsnprintf+0x5e7/0x62f
[<fffffff8000cd07>] do_path_lookup+0x2ca/0x2f1
[<fffffff800477ae>] sprintf+0x51/0x59
[<fffffff880b7122>] :megaraid_sas:sysfs_max_sectors_read+0x0/0x25
[<fffffff800b68cf>] audit_filter_inodes+0xbe/0xf9
[<fffffff880b7140>] :megaraid_sas:sysfs_max_sectors_read+0x1e/0x25
```

- No vendor finger pointing !
 - Oracle developed new MEGA RAID driver rpm after they reproduced issue in-house
- Oracle also advised to go for native LVM Snapshot backups over to the NFS mounts

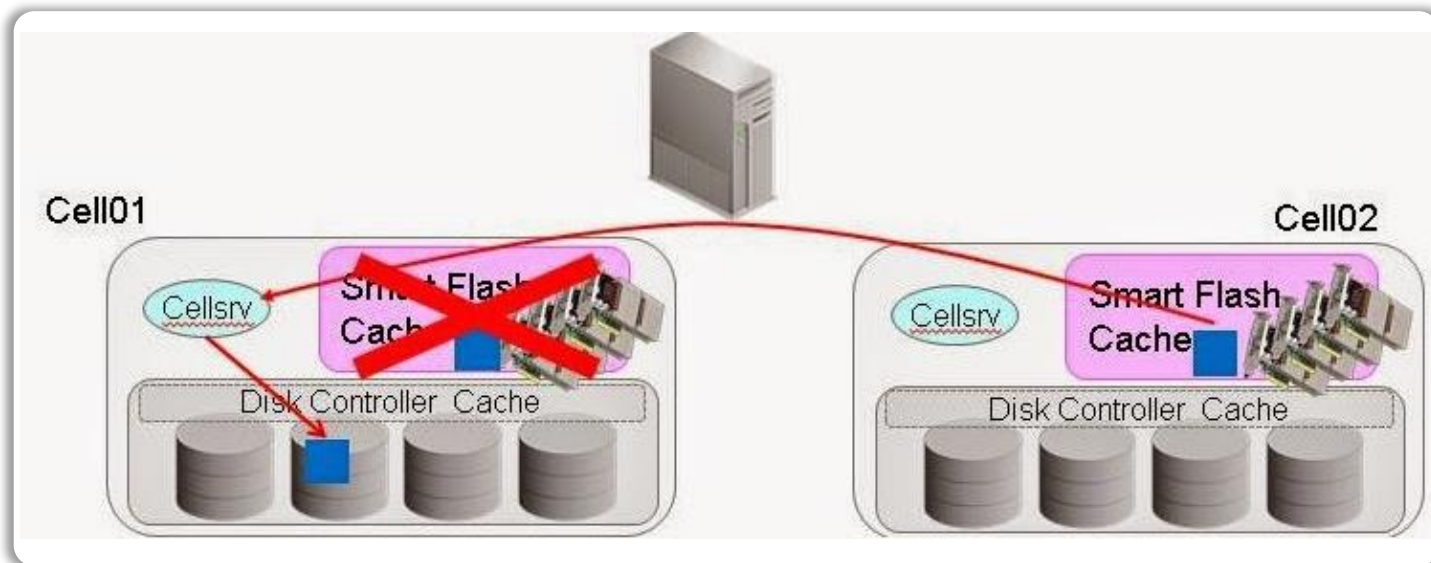
WBFC & NORMAL redundancy - Water and Oil

- WBFC = Write Back Flash Cache
- It works like this:



WBFC & NORMAL redundancy - Water and Oil (Contd...)

- There are 12 disks and 16 FMODs in each cell server
- DB/ASM layer don't know about caching in I/O subsystem
- I/O subsystem responsible to acknowledge write completion
- Depending on the ASM redundancy level, ASM writes to either 2 or 3 disks
- If we are absorbing those writes in WBFC, then, the writes go to 2 FMODs in different Fail groups



- Say, a FMOD faulted=> 4 Flash disks absorbing writes are gone
- The dirty buffers (that the DB thinks are on disk) on these are gone as well

WBFC & NORMAL redundancy - Water and Oil (Contd...)

- But, since ASM did 2 such writes (NORMAL redundancy), there's no loss of data and "resilvering" kicks in to save the day
- Say, another FMOD also faulted (Murphy?) that held the 2nd copy
- Instead of dismounting whole ASM DG, these blocks that are lost both primary and the copy are marked as BADFDATA in the DB(1557421.1)
- All DBs in cluster will have the BADFDATA if they had active writes and are caught in cross-fire
- You are in la-la land=> all DBs of the cluster are corrupted and need restore from RMAN
- DB Processing can still go on if important tables of your DB are not hit with corruption

WBFC & NORMAL redundancy - Water and Oil (Contd...)

Here's how, you can see how each disk has a partnered Flashdisk where the write is cached

```
[root@cellsvr10 ~]# cellcli -e list griddisk attributes name,status,cachedby,asmmode,status,asmdeactivationoutcome
DATA_CLUSTER_CD_00_cellsvr10 active "FD_00_cellsvr10, FD_10_cellsvr10, FD_11_cellsvr10, FD_09_cellsvr10" ONLINE Yes
DATA_CLUSTER_CD_01_cellsvr10 active "FD_15_cellsvr10, FD_13_cellsvr10, FD_12_cellsvr10, FD_14_cellsvr10" ONLINE Yes
DATA_CLUSTER_CD_02_cellsvr10 active "FD_15_cellsvr10, FD_13_cellsvr10, FD_12_cellsvr10, FD_14_cellsvr10" ONLINE Yes
DATA_CLUSTER_CD_03_cellsvr10 active "FD_05_cellsvr10, FD_06_cellsvr10, FD_04_cellsvr10, FD_07_cellsvr10" ONLINE Yes
DATA_CLUSTER_CD_04_cellsvr10 active "FD_05_cellsvr10, FD_06_cellsvr10, FD_04_cellsvr10, FD_07_cellsvr10" ONLINE Yes
DATA_CLUSTER_CD_05_cellsvr10 active "FD_00_cellsvr10, FD_03_cellsvr10, FD_01_cellsvr10" ONLINE Yes
DATA_CLUSTER_CD_06_cellsvr10 active "FD_15_cellsvr10, FD_13_cellsvr10, FD_12_cellsvr10, FD_14_cellsvr10" ONLINE Yes
DATA_CLUSTER_CD_07_cellsvr10 active "FD_00_cellsvr10, FD_10_cellsvr10, FD_11_cellsvr10, FD_09_cellsvr10" ONLINE Yes
DATA_CLUSTER_CD_08_cellsvr10 active "FD_00_cellsvr10, FD_10_cellsvr10, FD_11_cellsvr10, FD_09_cellsvr10" ONLINE Yes
DATA_CLUSTER_CD_09_cellsvr10 active "FD_00_cellsvr10, FD_03_cellsvr10, FD_01_cellsvr10" ONLINE Yes
DATA_CLUSTER_CD_10_cellsvr10 active "FD_00_cellsvr10, FD_10_cellsvr10, FD_11_cellsvr10, FD_09_cellsvr10" ONLINE Yes
DATA_CLUSTER_CD_11_cellsvr10 active "FD_05_cellsvr10, FD_06_cellsvr10, FD_04_cellsvr10, FD_07_cellsvr10" ONLINE Yes
DBFS_DG_CD_02_cellsvr10 active "FD_15_cellsvr10, FD_13_cellsvr10, FD_12_cellsvr10, FD_14_cellsvr10" ONLINE Yes
DBFS_DG_CD_03_cellsvr10 active "FD_05_cellsvr10, FD_06_cellsvr10, FD_04_cellsvr10, FD_07_cellsvr10" ONLINE Yes
DBFS_DG_CD_04_cellsvr10 active "FD_05_cellsvr10, FD_06_cellsvr10, FD_04_cellsvr10, FD_07_cellsvr10" ONLINE Yes
DBFS_DG_CD_05_cellsvr10 active "FD_00_cellsvr10, FD_03_cellsvr10, FD_01_cellsvr10" ONLINE Yes
DBFS_DG_CD_06_cellsvr10 active "FD_15_cellsvr10, FD_13_cellsvr10, FD_12_cellsvr10, FD_14_cellsvr10" ONLINE Yes
DBFS_DG_CD_07_cellsvr10 active "FD_00_cellsvr10, FD_10_cellsvr10, FD_11_cellsvr10, FD_09_cellsvr10" ONLINE Yes
DBFS_DG_CD_08_cellsvr10 active "FD_00_cellsvr10, FD_10_cellsvr10, FD_11_cellsvr10, FD_09_cellsvr10" ONLINE Yes
DBFS_DG_CD_09_cellsvr10 active "FD_00_cellsvr10, FD_03_cellsvr10, FD_01_cellsvr10" ONLINE Yes
DBFS_DG_CD_10_cellsvr10 active "FD_00_cellsvr10, FD_10_cellsvr10, FD_11_cellsvr10, FD_09_cellsvr10" ONLINE Yes
DBFS_DG_CD_11_cellsvr10 active "FD_05_cellsvr10, FD_06_cellsvr10, FD_04_cellsvr10, FD_07_cellsvr10" ONLINE Yes
RECO_CLUSTER_CD_00_cellsvr10 active "FD_00_cellsvr10, FD_10_cellsvr10, FD_11_cellsvr10, FD_09_cellsvr10" ONLINE Yes
RECO_CLUSTER_CD_01_cellsvr10 active "FD_15_cellsvr10, FD_13_cellsvr10, FD_12_cellsvr10, FD_14_cellsvr10" ONLINE Yes
RECO_CLUSTER_CD_02_cellsvr10 active "FD_15_cellsvr10, FD_13_cellsvr10, FD_12_cellsvr10, FD_14_cellsvr10" ONLINE Yes
RECO_CLUSTER_CD_03_cellsvr10 active "FD_05_cellsvr10, FD_06_cellsvr10, FD_04_cellsvr10, FD_07_cellsvr10" ONLINE Yes
RECO_CLUSTER_CD_04_cellsvr10 active "FD_05_cellsvr10, FD_06_cellsvr10, FD_04_cellsvr10, FD_07_cellsvr10" ONLINE Yes
RECO_CLUSTER_CD_05_cellsvr10 active "FD_00_cellsvr10, FD_03_cellsvr10, FD_01_cellsvr10" ONLINE Yes
RECO_CLUSTER_CD_06_cellsvr10 active "FD_15_cellsvr10, FD_13_cellsvr10, FD_12_cellsvr10, FD_14_cellsvr10" ONLINE Yes
RECO_CLUSTER_CD_07_cellsvr10 active "FD_00_cellsvr10, FD_10_cellsvr10, FD_11_cellsvr10, FD_09_cellsvr10" ONLINE Yes
RECO_CLUSTER_CD_08_cellsvr10 active "FD_00_cellsvr10, FD_10_cellsvr10, FD_11_cellsvr10, FD_09_cellsvr10" ONLINE Yes
RECO_CLUSTER_CD_09_cellsvr10 active "FD_00_cellsvr10, FD_03_cellsvr10, FD_01_cellsvr10" ONLINE Yes
RECO_CLUSTER_CD_10_cellsvr10 active "FD_00_cellsvr10, FD_10_cellsvr10, FD_11_cellsvr10, FD_09_cellsvr10" ONLINE Yes
RECO_CLUSTER_CD_11_cellsvr10 active "FD_05_cellsvr10, FD_06_cellsvr10, FD_04_cellsvr10, FD_07_cellsvr10" ONLINE Yes
```

WBFC & NORMAL redundancy - Water and Oil (Contd...)

```
select FILE_NAME, file_id, TABLESPACE_NAME from dba_data_files where FILE_ID in (select
distinct FILE# from gv$database_block_corruption);
alter database datafile 6 offline;

...

alter database datafile 15 offline;
run {
ALLOCATE CHANNEL 'T6' DEVICE TYPE 'SBT_TAPE' PARMS
'ENV=(NB_ORA_CLIENT=jc08bkp.mycompany.com
,NB_ORA_POLICY=nbu_db_policy,NB_ORA_SCHED=full), BLKSIZE=1048576';
restore datafile 6, 14, 2, 5, 3, 10, 15;
recover datafile 6, 14, 2, 5, 3, 10, 15;
}
alter database datafile 6 online;

...

alter database datafile 15 online;
```

Change Course in time

- When an imminent crisis is detected
- Techniques to take a poor performing disk out of the equation quickly
 - You suspect a cell disk is having a high service time and causing I/O to be queued up
 - Or your internal monitoring scripts alert you to a poor performing disk, that the Exadata Software doesn't report (more on this later)
 - Impact will be cluster-wide for all the DBs, a very sluggish I/O
 - You need to move in quickly and before you get slammed with 4 different critical incidents
 - From cell:
 - `cellcli -e list griddisk attributes name, status, asmmodestatus, asmdeactivationoutcome`
 - `cellcli -e alter griddisk all inactive`
 - Restart the cell node (the disks from this cell are still out of equation)
 - ➡ Immediately performance should be back to good
 - `cellcli -e list griddisk attributes name, status, asmmodestatus, asmdeactivationoutcome`
- From cell:
 - Have the Oracle FE reach the DC, ASAP
 - Once, he is standing-by:
 - `cellcli -e alter griddisk all active`
 - `cellcli -e ALTER PHYSICALDISK 20:0 drop for replacement;`
 - Back in business

Corruption In Storage Cell disk

- Cell Machine goes down and doesn't come back up online
- From ILOM, you see:

```
INIT: version 2.86 booting
      Welcome to Oracle Linux Oracle Linux Server release 5.10
      Press 'I' to enter interactive startup.
Starting udev:                                [ OK ]
Loading default keymap (us):                  [ OK ]
Setting hostname [REDACTED]:                  [ OK ]
Setting up Logical Volume Management:  No volume groups found
                                           [ OK ]

Checking filesystems
CELLSYS contains a file system with errors, check forced.
CELLSYS:
Inode 491713 has imagic flag set.

CELLSYS: UNEXPECTED INCONSISTENCY; RUN fsck MANUALLY.
        (i.e., without -a or -p options)
                                           [ FAILED ]

*** An error occurred during the file system check.
*** Dropping you to a shell; the system will reboot
*** when you leave the shell.
Give root password for maintenance
(or type Control-D to continue):
```

- The cell machine goes on a recycle loop

Corruption In Storage Cell disk (Contd...)

- From ILOM console, you see:

From ILOM console:

```
+++++++
md: Autodetecting RAID arrays.
md: Scanned 0 and added 0 devices.
md: autorun ...
md: ... autorun DONE.
EXT3-fs (md4): error: unable to read superblock <===== That
EXT3-fs (md6): error: unable to read superblock
Kernel panic - not syncing: Attempted to kill init!
Pid: 1, comm: init Not tainted 2.6.39-400.128.17.el5uek #1
Call Trace:
[<ffffffff8106f3af>] panic+0xbf/0x1f0
[<ffffffff81188e2a>] ? free_vfsmnt+0x3a/0x50
[<ffffffff8105b39f>] ? account_entity_enqueue+0x8f/0xa0
[<ffffffff81071b49>] find_new_reaper+0xc9/0xd0
[<ffffffff81073e25>] forget_original_parent+0x45/0x1b0
[<ffffffff8105b072>] ? sched_move_task+0xb2/0x150
[<ffffffff81073fa6>] exit_notify+0x16/0x170
[<ffffffff81074383>] do_exit+0x283/0x460
[<ffffffff810745a1>] do_group_exit+0x41/0xb0
[<ffffffff81074627>] sys_exit_group+0x17/0x20
[<ffffffff8150f302>] system_call_fastpath+0x16/0x1b
Rebooting in 60 seconds..
```

- Copy a good diagnostic iso image from a good node: /opt/oracle.cellos/diag.iso
- Use the cell machine's ILOM GUI redirected CDRom as the input boot

Corruption In Storage Cell disk (Contd...)

- Using CLI make sure that the bios boot is set to CDROM as follows:
 - `cd /HOST`
-> `set boot_device=cdrom`
Set 'boot_device' to 'cdrom'
 - Power cycle
 - At boot prompt, get into "linux rescue" mode
 - At shell, corruption fixed via:
 - `e2fsck /dev/md6 -6`
 - `e2fsck /dev/md6 -y`

Corruption In Storage Cell disk (Contd...)

```
-sh-3.2# df -h
Filesystem                Size      Used Available Use% Mounted on
none                      11.8G    456.0k    11.8G   0% /dev
/dev/cdrom-sr0             71.4M    71.4M         0 100% /mnt/iso
/dev/md6                   9.8G     5.6G     3.7G  60% /mnt/cell

-sh-3.2# e2fsck /dev/md6 -6_ _ _
e2fsck 1.38 (30-Jun-2005)
/dev/md6 is mounted.

    WARNING!!!  Running e2fsck on a mounted filesystem may cause
    SEVERE filesystem damage.

Do you really want to continue (y/n)? y
yes

CELLSYS: recovering journal
CELLSYS contains a file system with errors, check forced.
Pass 1: Checking inodes, blocks, and sizes
Inode 491713 has imagic flag set. Clear<y>? yes
...

CELLSYS: ***** FILE SYSTEM WAS MODIFIED *****
CELLSYS: 77519/1310720 files (1.8% non-contiguous), 1496474/2621424 blocks
-sh-3.2#
-sh-3.2# e2fsck /dev/md6 -y
e2fsck 1.38 (30-Jun-2005)
/dev/md6 is mounted.

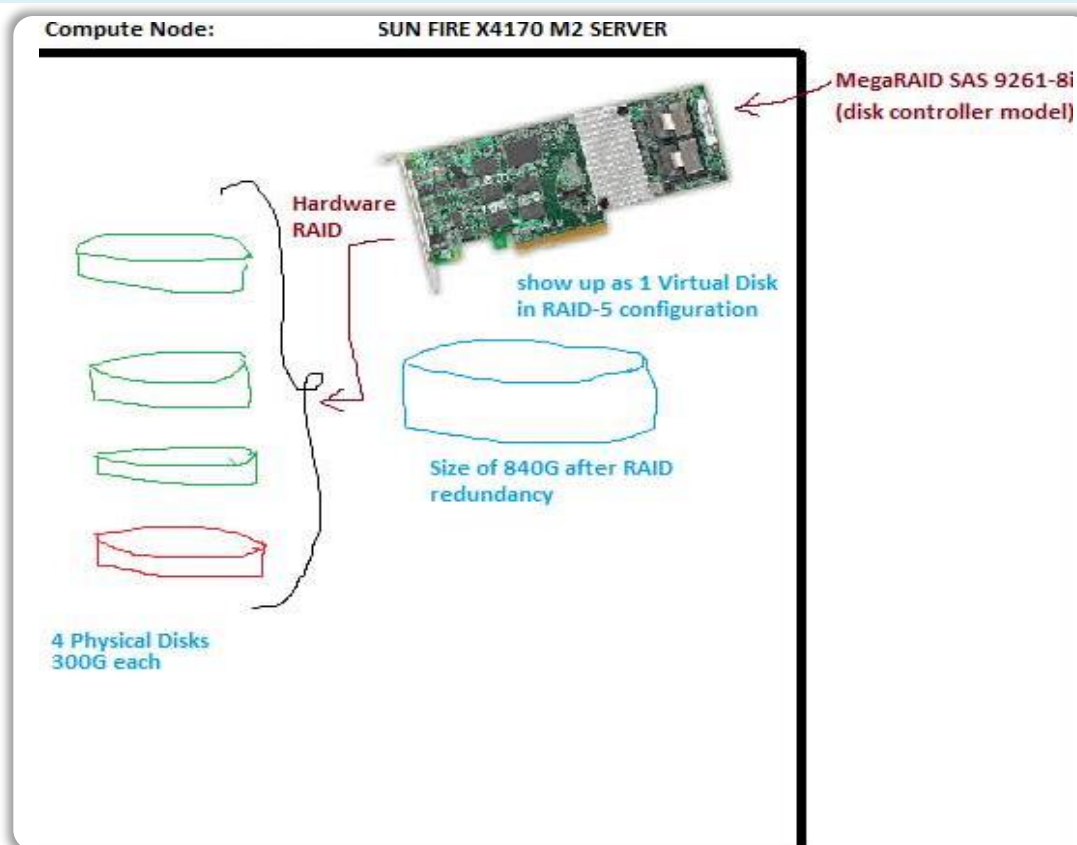
    WARNING!!!  Running e2fsck on a mounted filesystem may cause
    SEVERE filesystem damage.

Do you really want to continue (y/n)? y
yes

CELLSYS: clean, 77519/1310720 files, 1496474/2621424 blocks
```

When the home seems small for the whole family...

- Typically, multiple ORACLE_HOMEs exist with multiple patch sets/patches/versions
- More common in a highly consolidated environment, as patching windows/bugs exposed vary from DB to DB
- Unearth the space, you knew (or didn't know) you had it all along



When the home seems small for the whole family (Contd...)

- LSI card is the controller for 3 active disks and is battery backed up for write back cache
- 4th disk in Compute nodes was “hot spare”
 - if any of the 3 disks in the RAID-5 config broke, the RAID config will intake the 4th disk and RAID rebuild will be initiated
- Since 11.2.3.2.1 (seems like medieval age now!), you can use up the 4th disk also in the active RAID config
 - To increase the available storage on the DB nodes

```
[root@jcd02 ~]# /opt/MegaRAID/MegaCli/MegaCli64 -pdlist -a0 | grep -iE "slot|firmware"
Slot Number: 0
Firmware state: Online, Spun Up
Device Firmware Level: 0B70
Slot Number: 1
Firmware state: Online, Spun Up
Device Firmware Level: 0B70
Slot Number: 2
Firmware state: Online, Spun Up
Device Firmware Level: 0B70
Slot Number: 3
Firmware state: Online, Spun Up
Device Firmware Level: 0B70
[root@xdhfd5cn02 ~]#
```

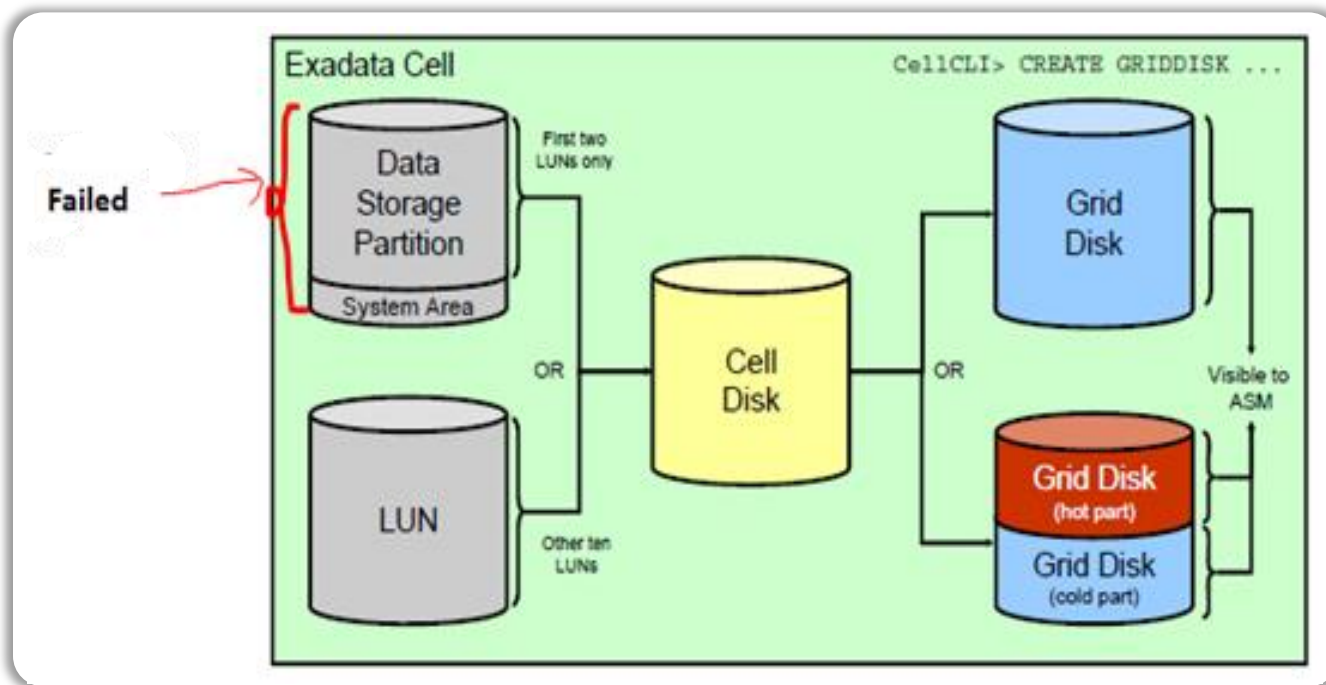
When the home seems small for the whole family (Contd...)

- If a compute node's disk faulted, the RAID-5 offers the protection, but we still need to replace the disk asap
 - **The RAID-5 is now said to be in a degraded mode**

```
[root@jcdb03 oracle]# /opt/MegaRAID/MegaCli/MegaCli64 -LdPdInfo -a0 | grep -iE "target|state|slot"
Virtual Drive: 0 (Target Id: 0)
State : Degraded
Slot Number: 0
Firmware state: Online, Spun Up
Foreign State: None
Slot Number: 1
Firmware state: Online, Spun Up
Foreign State: None
Slot Number: 2
Firmware state: Online, Spun Up
Foreign State: None
Slot Number: 3
Firmware state: Offline
Foreign State: None
```

When Cell System disk Crashes while you do the Cell imaging?

- It's PROD cell imaging weekend
- You type and wait for the prompt back:
 - `./patchmgr -cells /home/oracle/cell_group -patch`



(Source: Oracle Documentation)

When Cell System disk Crashes while you do the Cell imaging?

(Contd..)

- Of the 9000 minutes in a year, your cell disk chooses to crash during this time.
- Got the prompt back with failure message:
 - [failed] Patch or rollback failed as reported by /root/_patch_hctap_/_p_/install.sh - query state on the cell.
 - [ERROR] Patch or rollback failed. Please run cleanup before retrying.
- Of the 12 disks, the system disk, where you are going to image, crashes (Murphy?)
 - Any other disk, just a drop-rebalance, re-run failed imaging and replacement
- You hope and pray the disk crash alert is false alert, but it's real
- Bad thing: every single critical OLTP application is waiting for the window to be over (tough negotiated 11h window, when you are into heavy consolidation for ROI)
- Good thing: You haven't panicked and can actually formulate a POA during crisis 😊

When Cell System disk Crashes while you do the Cell imaging? (Contd..)

- Here's a simple sequence, ****AFTER**** the disk is replaced:

```
cellcli -e alter griddisk all inactive
cellcli -e alter cell shutdown services all
#./patchmgr -cells /root/cells.txt -cleanup <=====
where the file cells.txt should have just one word in it: the failed
cellnode, for example cellnode310
#./patchmgr -cells /root/cells.txt -patch_check_prereq
#./patchmgr -cells /root/cells.txt -patch
# cellcli -e alter cell startup services cellsrv
# cellcli -e alter griddisk all active
```

OEL5 to OEL6 Baggage (DB nodes)

- A lot, repeat, lot of things may need fixing when you go from a image (say 12.1.1.1.1) running OEL5 to an image, say 12.1.2.1.2 running OEL6
- Here's a sampling of what you may need to be prepared for:
 - ASR not working & Mutt Issues
 - DBFS mount was unable to mount

```
export ORACLE_HOME=/u01/app/home/version/11.2.0.3/sw
export LD_LIBRARY_PATH=$ORACLE_HOME/lib
export TNS_ADMIN=$ORACLE_HOME/network/admin
nohup $ORACLE_HOME/bin/dbfs_client /@dbfsdb -o failover -o
allow_other /mnt/dbfs >/tmp/dbfs_mount.out &
```

OEL5 to OEL6 Baggage (DB nodes) (Contd...)

- Missing RPMs
 - Both Oracle
 - authconfig (x86_64), libXmu (x86_64), libXpm (x86_64), libXt (x86_64)
 - And 3rd Party
 - Example: lftp-3.7.11-7.el5, mutt-1.4.2.2-6.el5, tcpdump-3.9.4-15.el5, ttmkfsdir-3.0.9-23.el5
- OS Services not started with auto enabled option
 - CA-WAAE | acpid | altiris | netbackup | rscd | vasd | vxpbx_exchanged
 - Vintela issues
 - NFS issue (uid & group id showing 'nobody' for nfs filesystem)
 - /etc/cron.allow will not have anything and needs restore
 - etc

Compute Node Imaging – Gone Haywire

Sometimes, you sail smoothly into the perfect storm inside the SAME outage window

- Dbnodeupdate.sh on one of the nodes complains about RPMs
 - Higher version rpms or
 - rpms from 2 versions present or
 - missing rpms
 - “[ERROR] Unable to uninstall sas_snmp”
 - “ERROR: Given location (-l) does not exist or does not have the proper rpms available”
 - “ERROR: It is also possible the system already is on the same or higher release”
 - “[ERROR] Ramfs image file not found /boot/initrd-2.6.39-400.126.1.el5uek.img”
- Cell physical disk “not present”, thankfully, after the reboot of cell node imaging (Oracle replaced it)
- Reboot hung after dbnodeupdate.sh for one of the db nodes (recycle via ILOM)
- Missing IB Drivers load in kernel

Compute Node Imaging – Gone Haywire (Contd...)

- Missing IB Drivers load in kernel
- Way out of the mess, short of going back to the node reimaging
 - Manually load IB modules (% insmod mlx4_ib.ko, % insmod mlx4_core.ko)
 - Copy initrd from another node (“[ERROR] Ramfs image file not found /boot/initrd-2.6.39-400.126.1.el5uek.img”)
 - Reboot into later kernel, run the dbnodeupdate.sh

Lesson Learnt:

- Don’t troubleshoot, don’t be a hero... just rollback, the very first time dbnodeupdate.sh fails

Compute Node Imaging – Gone Haywire (Contd...)

```
[root@dbsvr201 ~]# ./dbnodeupdate.sh -r
#####
...
# - In case of any problem when filing an SR, upload the following:
#
#   - /var/log/cellos/dbnodeupdate.log
#   - /var/log/cellos/dbnodeupdate.<runid>.diag
#   - where <runid> is the id of the failing run.
#
#####
Continue ? [Y/n]
Y
(*) 2015-06-18 16:41:32: Unzipping helpers (/root/dbupdate-helpers.zip) to /opt/oracle.SupportTools/dbnodeupdate_helpers
(*) 2015-06-18 16:41:32: Initializing logfile /var/log/cellos/dbnodeupdate.log
(*) 2015-06-18 16:41:33: Collecting system configuration details. This may take a while...

Active Image version   : 11.2.3.3.0.131014.1
Active Kernel version  : 2.6.39-400.126.1.el5uek
Active LVM Name        : /dev/mapper/VGExaDb-LVDbSys1 <-----
Inactive Image version : 11.2.3.2.1.130302
Inactive LVM Name      : /dev/mapper/VGExaDb-LVDbSys2 <----- btw, when you extend /, ensure you extend both active & inactive LVMS to same size
Current user id        : root
Action                 : rollback <-----
Rolling back to        : 11.2.3.2.1.130302
Create a backup         : No (backups cannot be made for rollbacks)
Shutdown stack         : No (Currently stack is down)
Logfile                : /var/log/cellos/dbnodeupdate.log (runid: 180614164130)
Diagfile               : /var/log/cellos/dbnodeupdate.180614164130.diag
Server model           : SUN FIRE X4170 M2 SERVER
dbnodeupdate.sh rel.   : 3.26 (always check MOS 1553103.1 for the latest release of dbnodeupdate)

Continue ? [Y/n]
```

When proverbial stuff seem to hit the proverbial stuff

- If you start to get all kinds of Hardware failure at the same time, like:

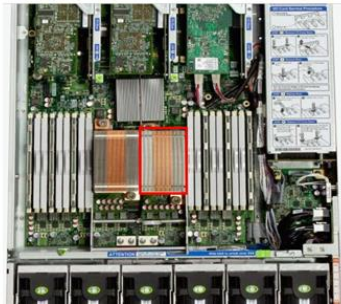
Description A processor component is suspected of causing a fault.

Component Name

/SYS/MB/P0

Fault class

fault.cpu.intel.internal



Description A processor component is suspected of causing a fault.

Component Name

/SYS/MB/P0

Fault class

fault.cpu.intel.quickpath.interconnect



Description A memory component is suspected of causing a fault.

Component Name

/SYS/MB/P0/D5

Fault class

fault.memory.intel.dimm_ue



Description A memory component is suspected of causing a fault.

Component Name

/SYS/MB/P1/D8

Fault class

fault.memory.intel.dimm_ue



When proverbial stuff seem to hit the proverbial stuff (contd...)

- The cell machine has gone for a reboot
- With so many hardware alerts, you have a few choices:
 - Do nothing, as the m/c came up albeit after multiple resets during BIOS initialization – ‘one-time-only-transient-failure’
 - all components came up “cleanly”, manual clearing of faults in ILOM
 - Replace the System Board as the common part
 - Replace the System Board and both CPUs
 - Replace the System Board, both CPUs, and 3 out of 6 Memory DIMMs.
- Problem of this nature not easy to get to the bottom of (from vendor)
- We took the route to replace the system board only
- Sent the replaced part to Oracle lab for RCA – confirmed it was the motherboard

When proverbial stuff seem to hit the proverbial stuff (contd...)

- This is very similar to when you hit multiple FMOD alerts at the same time:

Status	FAILED
Manufacturer	Sun
Model Number	Flash Accelerator F20 PCIe Card
Size	23GB
Serial Number	12334ASDFAS
Firmware	D21Y
Slot Number	PCI Slot 2; FDOM: 0
Cell Disk	FD_04_jccomstl3
Grid Disk	Not configured
Flash Cache	Present
Flash Log	Present



Status	FAILED
Manufacturer	Sun
Model Number	Flash Accelerator F20 PCIe Card
Size	23GB
Serial Number	12334ASDFAS
Firmware	D21Y
Slot Number	PCI Slot 2; FDOM: 2
Cell Disk	FD_06_jccomstl3
Grid Disk	Not configured
Flash Cache	Present
Flash Log	Present



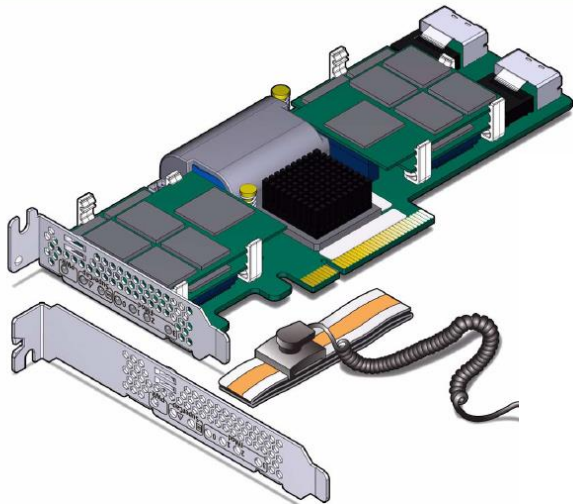
Status	FAILED
Manufacturer	Sun
Model Number	Flash Accelerator F20 PCIe Card
Size	23GB
Serial Number	12334ASDFAS
Firmware	D21YD
Slot Number	PCI Slot 2; FDOM: 1
Cell Disk	FD_05_jccomstl3
Grid Disk	Not configured
Flash Cache	Present
Flash Log	Present



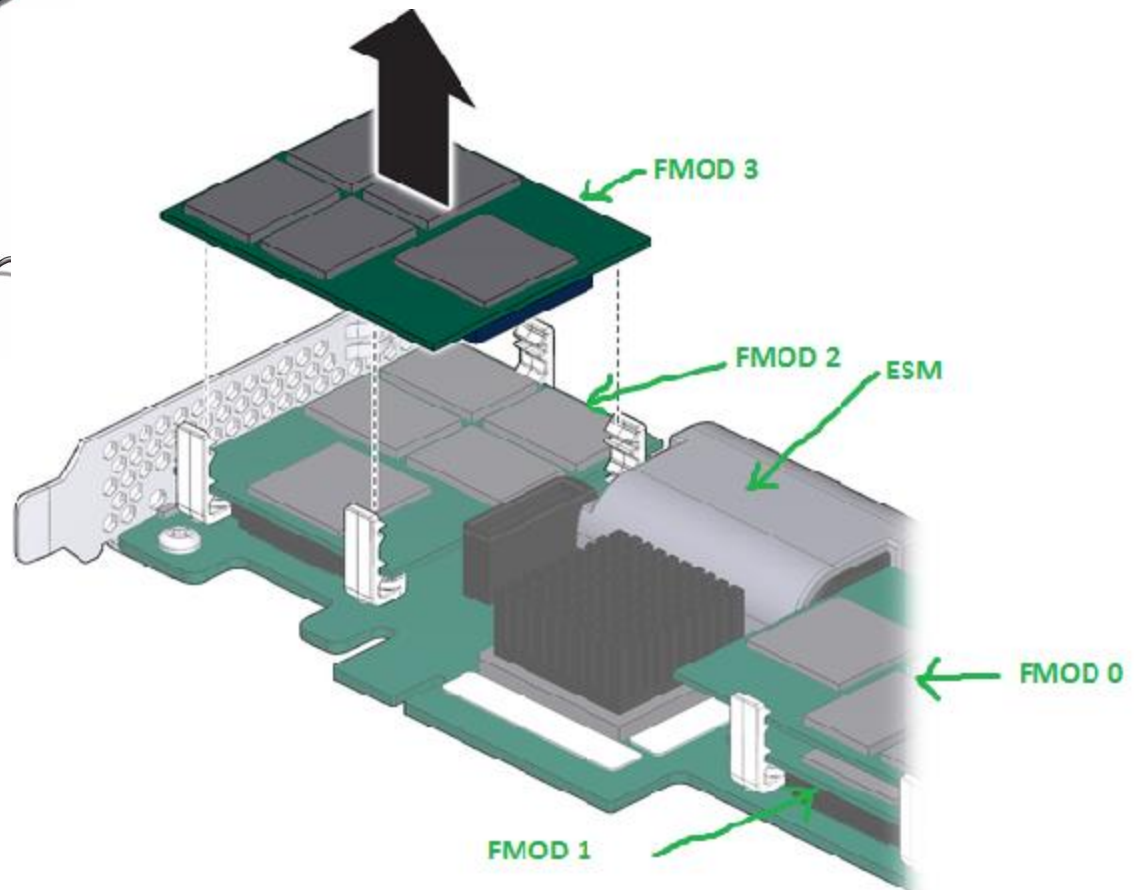
Status	FAILED
Manufacturer	Sun
Model Number	Flash Accelerator F20 PCIe Card
Size	23GB
Serial Number	1225M0CTCJ
Firmware	D21Y
Slot Number	PCI Slot 2; FDOM: 3
Cell Disk	FD_07_jccomstl3
Grid Disk	Not configured
Flash Cache	Present
Flash Log	Present



When proverbial stuff seem to hit the proverbial stuff (contd...)



Just replace the
whole card



Following sections touch briefly a few costly bugs that fell on our laps

- May not be entirely applicable to your environment
- Still, but, nevertheless, these have caused major heartaches either inside a planned outage or caused an unplanned outage

Compute Node won't come up after cell imaging

After cell imaging LVM was found to be corrupt in the initrd

Bug 13596277 - LVM.CONF IN INITRD INCORRECT AFTER APPLYING 11.2.2.4.2 DB MIN PACK

Copy a proper lvm.conf from the original initrd image and rebuild it as:

1. extract the initrds

```
# cd /boot
```

```
# mkdir x y
```

```
# cd x
```

```
# zcat ../initrd-2.6.18-238*.img | cpio -idmv
```

```
# cd ../y
```

```
# zcat ../initrd-2.6.18-194*.img | cpio -idmv
```

```
# cd ..
```

2. replace the lvm.conf

```
# mv x/etc/lvm/lvm.conf{,.orig}
```

```
# cp y/etc/lvm/lvm.conf x/etc/lvm/
```

3. rebuild a new initrd

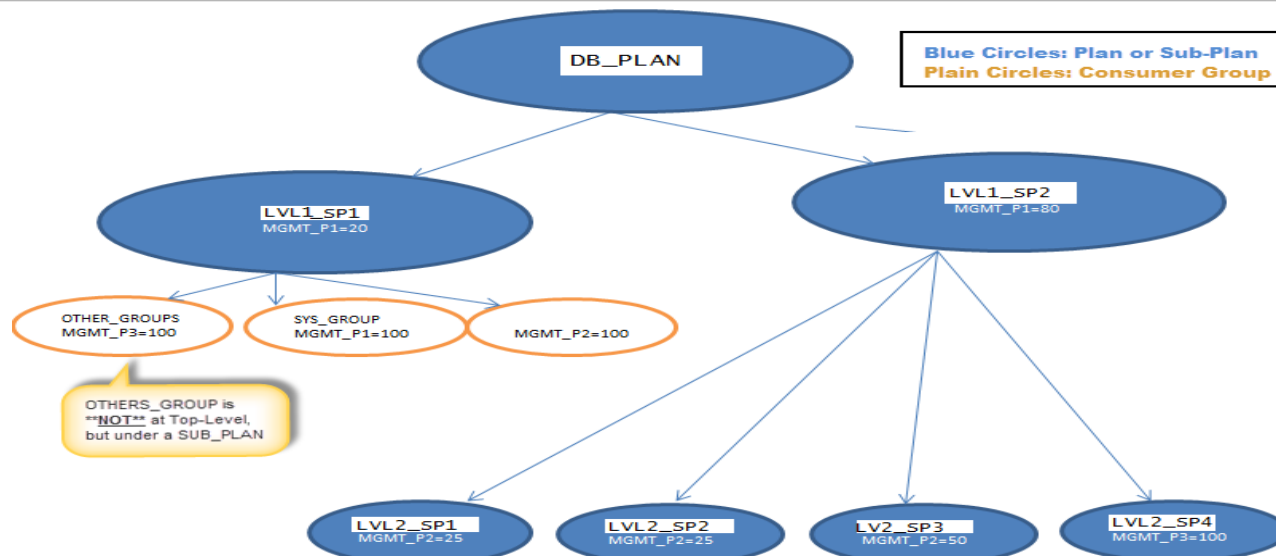
```
# cd x
```

```
# find ./ | cpio -H newc -o > /boot/initrd-new.img
```

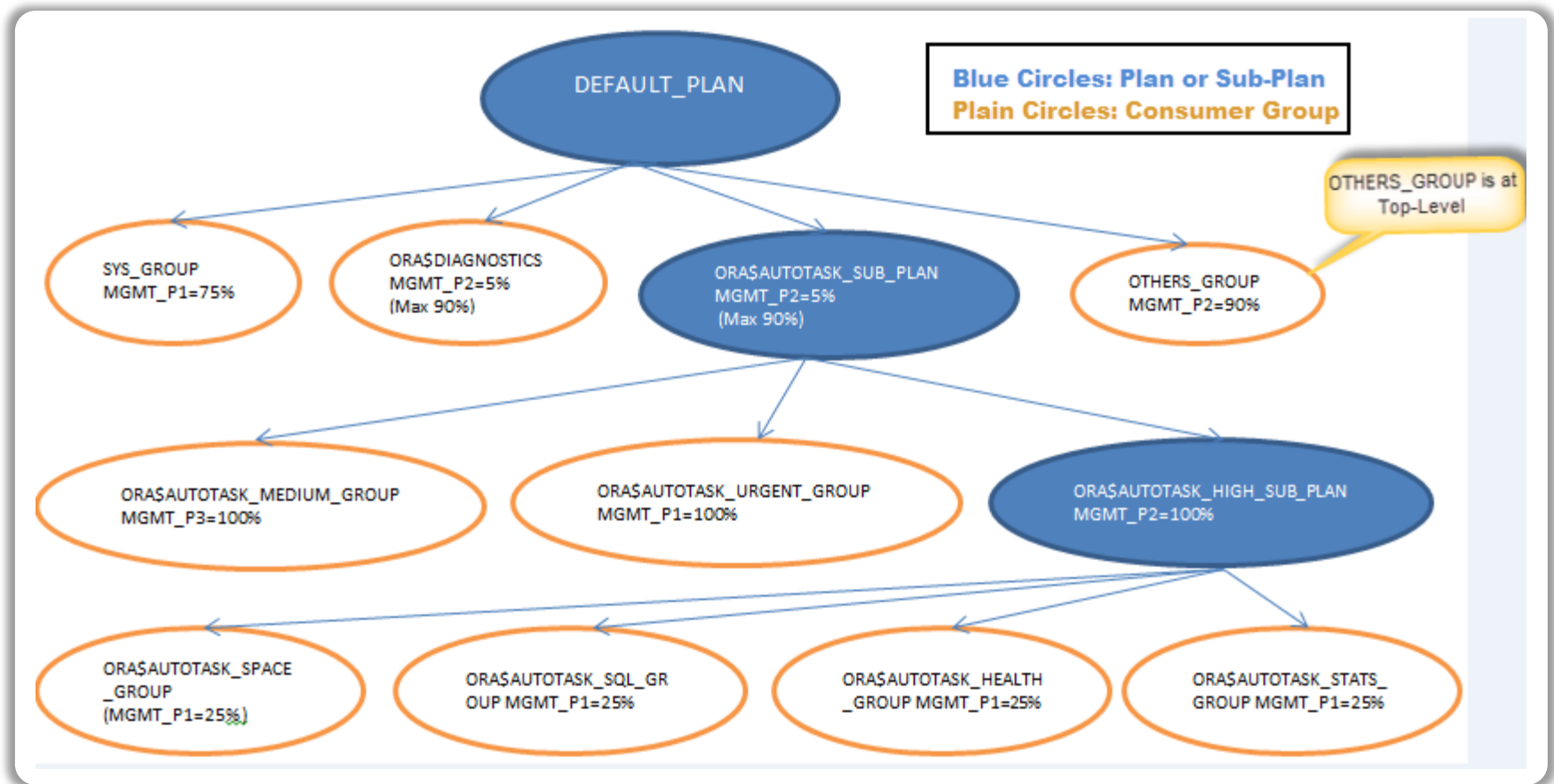
4. modify the grub setting

RM Bug causing ASM DG dismount

- Right after a tough negotiated outage window was over, (remember consolidation?), the databases were thrown open to users for business (Cell Image to 12.1.1.1.1)
- Within a few mins all the DBs in the cluster, came crashing down
- Bug 20408010 : EXADATA:CELL:ORA-00600 [DISKIOSCHED::GETCATINDEX:2] [4294967295]
- Culprit/exposing condition for the bug was a DBRM plan on one of the DBs where OTHERS_GROUP was NOT at the top level
- When cellsrv received sql processing via iDB, it also gets the RM directives
 - OTHERS_GROUP not at the top level exposed the bug in the 12.1.1.1.1 code and caused the cellsrv to crash~~ it's like a cell machine crashing
 - When another cellsrv in another cell machine also crashed, that was it
 - ASM normal redundancy can tolerate only a single failgroup down



RM Bug causing ASM DG dismount (contd...)



RM Bug causing ASM DG dismount (contd...)

- When cellsrv received sql processing via iDB, it also gets the RM directives
- Cell system I/O processing always has the iorm active, for internal latency processing
- It doesn't honor the DBRM directives in this mode, though it's 'active'
- You can't turn it 'off', as in 'off'

```
CellCLI> list iormplan detail
```

```
name:          jccomcell5_IORMPLAN
catPlan:
dbPlan:
objective:     basic
status:        active
```

```
CellCLI> alter iormplan inactive
```

```
IORMPLAN status cannot be set to inactive
```

```
IORMPLAN successfully altered
```

- OTHERS_GROUP not at the top level exposed the bug in the 12.1.1.1.1 code and caused the cellsrv to crash~~ it's like a cell machine crashing
- When another cellsrv in another cell machine also crashed, that was it
- ASM normal redundancy can tolerate only a single failgroup down

Chip Paused – our most painful issue (well one of our many!)

- Bug 17021128 : NIWOT "CHIP PAUSED" CAUSES HIGH SERVICE TIME ON ALL DRIVES
- Cell m/c disk controller fails to realize disks having poor performance AND user IO keeps hitting to that disk without knowing its health state
- User IO queued continuously to the faulted/poor performing disk can bring down entire IO performance of the entire cell due to IO contention
- Consequentially, the entire cluster !
- Expected behavior by design:
 - faulted disk or a disk having poor performance will be isolated by disk controller
 - Exadata Storage Server Software then marks the disk from normal status to poor performance status
 - Once disk is marked as Poor Performance by Exadata software, associated grid disks will be dropped from ASM to avoid poor IO performance caused by the problem disk

Chip Paused – our most painful issue (well one of our many!) (contd...)

- What this bug does:
 - Disk controller and disk get into a pause loop allowing I/O issue to happen because the controller never realizes the disk is having a performance issue
 - You cannot see these Pause loops using your usual go-to MegaCLI parser tool
 - You must review the FwTermLog directly to confirm a poor performing disk
- Permanent fix was to upgrade the firmware
 - Via a cell-one off 17878822
 - Or Cell image upgrade to 11.2.3.3.0 or higher (we chose this, as the one-offs are not as thoroughly regression tested as a major release)
- This taught us one of the techniques to take a poor performing disk out of the equation quickly

- VPN goes out during Cell Imaging
 - Rollback and image again, necessarily
- Monitor ASR trap files

```
[root@dbn07 compmon]# cat /opt/oracle.cellos/common/traps.state
1 ; Fri Mar 21 07:06:13 2016 ; 766e1bb1-d007-49cd-b374-9395066c68ac ; Physicaldisk 252:1 Make
Model: HITACHI H103030SCSUN300G is at status predictive failure. Raised fault id: HALRT-02008 ;
Physical disk should be replaced. Exadata Compute Server: Disk Serial Number: 3345GBV14E
```

- Choose your battles, not every alert needs to be actionable

Configuration check discovered the following problems:

[WARNING] Ethernet interface eth0 is not at 1000Mbps. It is at 100 Mbps. Check cables and switches.

- When there's a sudden spike in DB node hard disk devices, check the BBU charge
 - it may be doing its regular discharge/charge (learn cycle)

Other Tidbits (contd...)

- When there's a sudden spike in DB node hard disk devices, check the BBU charge – it may be doing its regular discharge/charge (learn cycle)

Subject: EM Event: Critical:dbnode02.cust.com - Disk Device sda is 95.58% busy.

Target name=dbnode02.cust.com

Target type=**Host**

Host=dbnode02.cust.com

Event reported time=**Apr 20, 2016 10:19:15 PM EDT**

Message=Disk Device sda is 95.58% busy.

Metric=Disk Device Busy (%)

Metric value=95.58

Disk Device=

Severity=**Critical**

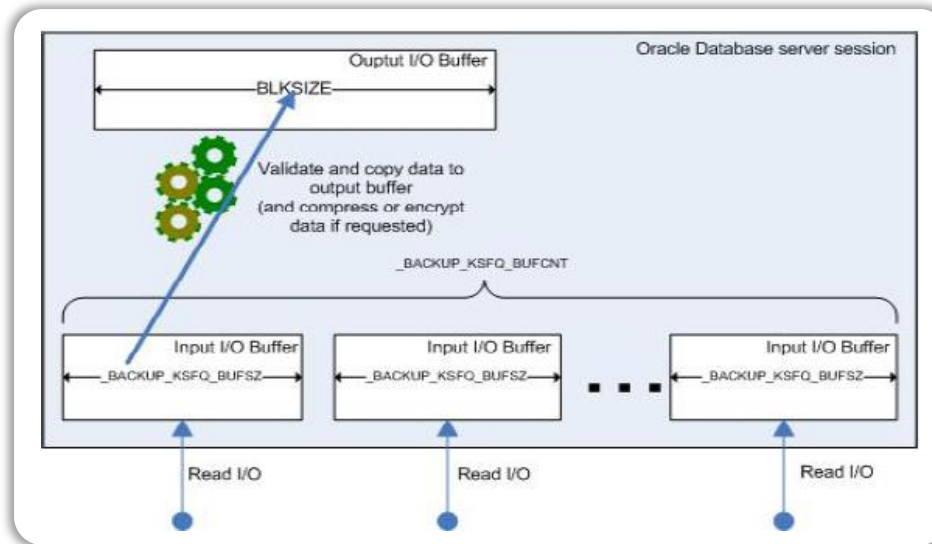
- At least, Monthly once, review all the Exachks, at least for PROD clusters
- Again, once a month at least, review the Metalink Critical Patch Review note 1270094.1, so you know if your system is vulnerable to a industry-wide hot issue (we have been the trail blazers to the world on a few bugs there !)

Non-Exadata Specific

- What follows are experiences that were faced in our Exadata environment
 - But they are not Exadata specific issues, you can face them in non-Exadata world as well

Active Clone running slow after DB upgrade?

- Business critical nightly clone job took a longer time after source DB upgrade
- Took a month's grueling investigation with Support/DEV
- Only manual workaround available (DEV working on fix):
 - Increase “_backup_ksfq_bufcnt”=64
 - Increase “_backup_ksfq_bufsz”=4



Active Clone running slow after DB upgrade? (Contd...)

- This parameter has no impact or meaning outside the context of RMAN and will only influence the IO done via RMAN.
- This parameter only impacts the IO buffers used by RMAN and does not impact any other IO.
- The way RMAN works is it uses IO buffers in each RMAN channel to read from datafile and write to destination, called I/P Buffers and O/P Buffers, respectively.
- The write destination of the o/p buffers could be backup media like dedup appliance in the case of backup or the auxiliary instance in the case of RMAN clone.
- By setting this buffer count parameter to 64 and default buffer size of 4MB, the channel uses 128 buffers for both i/p and o/p, thereby speeding up the clone process

Mystery Shutdown (was not at all fun)

- A database, let's call it AAAA, was going down cleanly & mysteriously
- No human, no script, no job that we were aware of did that !
- We were losing hair fast, as this one DB kept going for a shutdown, in random intervals
- Finally, we nabbed the culprit, in its act, red-handed: Was our DBUA !

[Thread-48] [2016-01-14 22:52:25.335 EST] [Database.checkClusterDB:3408] Stop RAC db (inplace upgrade need db to be open in mount mode first)

- AAAA was a clustered DB of 4 instances, but 3 instances shutdown and disabled
- Two conditions
 - If DBUA was started from one of the 3 nodes above and
 - if the engineer did not select the correct DB to upgrade in the 1st screen (click Next twice, as DBUA loads slower) and AAAA being alphabetically the 1st DB selected by default in dbua
- then this AAAA db was shutdown with above message in DBUA log !

Scan Listener relocation with MTS

#1. TNS-12518/TNS-12541 TNS:no listener :

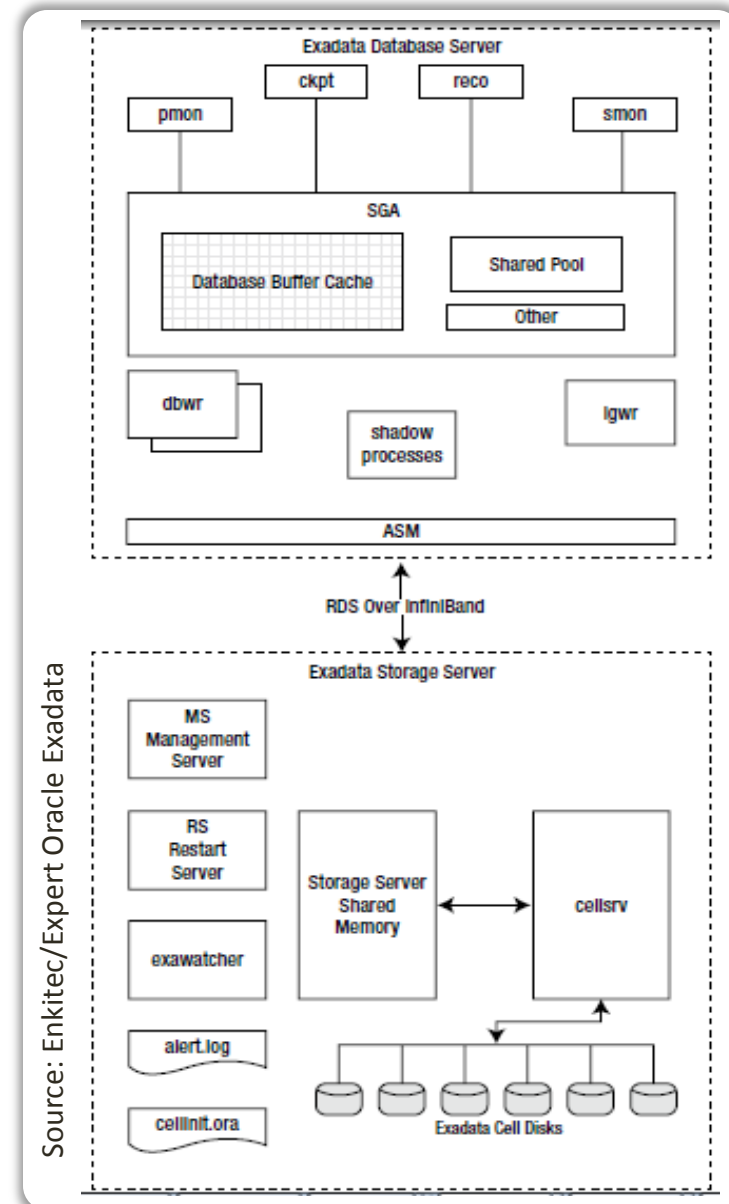
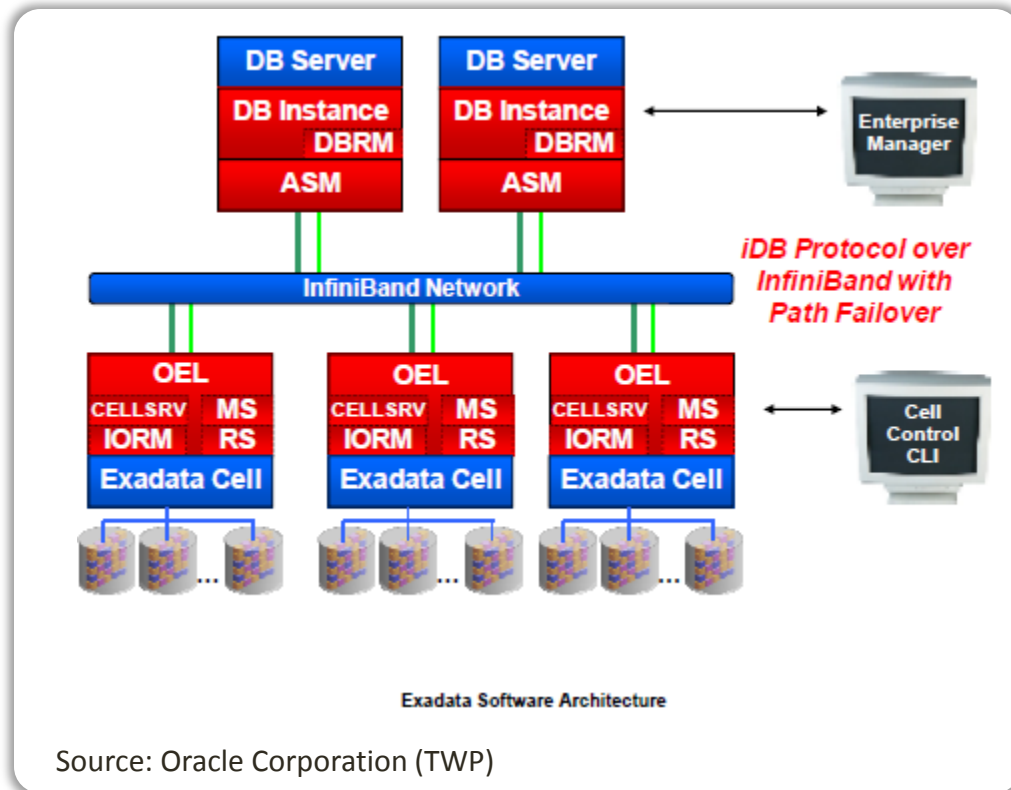
- Configured dispatchers using scan-name and not using node VIPs.
- Relocated scan from node#1 to node#4 to node#3
- Any shared server connection through node#4's/node#1's local listener resulted into TNS-12541 TNS:no listener
- TNS-12541 indicates that listener is not able to contact designated dispatcher.

#2 TNS-12518/TNS-12540 : TNS: internal limit restriction exceeded.

- TNS-12540 is an after effect of several TNS-12541 errors
- Listener maintains a pool of connection slots.
- Size of this pool is 1024
- These connection slots are occupied by each connection that is with the listener
- If all connection slots are occupied, listener throws TNS-12540 indicating resource limit reached
- Once listener hands-off connection to a dispatcher/shared-server or to a dedicated shadow server process, it's slot is reclaimed
- However, in case of TNS-12541 this reclamation doesn't happen
- Slots for connections resulting into TNS-12541 are not reclaimed

Solution : In order to avoid TNS-12541 & TNS-12540 errors, should configure dispatchers using node VIPs

Appendix A: Exadata Architecture



KEEP CHALLENGING™

Thank You

Jecca Chandrasekaran Narasimhan

Exadata Architect: Oracle Solution Practice - Engineered Systems

Cognizant Technology Solutions US

jcnars.wordpress.com

@jcnars