# Tips - Implementing Oracle Solutions on Linux for IBM System z

Speakers: David Simpson & Tom Kennelly
Speaker’s Company: IBM Corporation

Date of Presentation: Wednesday, March 2, 2011: 3:00 PM-4:00 PM
Session Number: 8469
Room 203A (Anaheim Convention Center)

# Trademarks

The following are trademarks of the International Business Machines Corporation in the United States and/or other countries. For a complete list of IBM Trademarks, see www.ibm.com/legal/copytrade.shtml: IBM, IBM Logo, zSeries, MVS, OS/390, pSeries, RS/6000, S/390, System Storage, System z9, VM/ESA, VSE/ESA, WebSphere, xSeries, z/OS, z196, zEnterprise, z/VM.

The following are trademarks or registered trademarks of other companies

Java and all Java-related trademarks and logos are trademarks of Oracle Corporation, Inc., in the United States and other countries.
LINUX is a registered trademark of Linux Torvalds in the United States and other countries.
UNIX is a registered trademark of The Open Group in the United States and other countries.
Oracle, WebLogic and E-business Suite are registered trademarks of Oracle Corporation.
SET and Secure Electronic Transaction are trademarks owned by SET Secure Electronic Transaction LLC.
Intel is a registered trademark of Intel Corporation.
* All other products may be trademarks or registered trademarks of their respective companies.

NOTES:Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance ratios stated here.

IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.

All customer examples cited or described in this presentation are presented as illustrations of the manner in which some customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual customer configurations and conditions.

This publication was produced in the United States. IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. Consult your local IBM business contact for information on the product or services available in your area.

All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent Goals and objectives only.

Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the performance, compatibility, or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

Prices subject to change without notice. Contact your IBM representative or Business Partner for the most current pricing in your geography.

References in this document to IBM products or services do not imply that IBM intends to make them available in every country.

Any proposed use of claims in this presentation outside of the United States must be reviewed by local IBM country counsel prior to such use.

The information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

# Topics to Cover

- What's New – Oracle 10.2.0.5
- zEnterprise CPU Cache and Memory
- Oracle Databases, DASD, and FCP Disk Storage
- Oracle RAC Troubleshooting
- New Releases coming up

# What’s New – Oracle 10.2.0.5

# Oracle 10.2.0.5  Released

- **Jan. 2011 – Oracle 10.2.0.5 for System z Linux Released – Patch 8202632**
  - Patchset 10.2.0.5.2 delayed a bit as Patch cycles take 6 weeks, so should be out this week Patch **10248542**

- **Some reported issues with 10.2.0.5 base, so you'll want to patch up with the 10.2.0.5.2 patch set as soon as possible, applications may require some patches as well**

  With Banner Application we had to apply the following fix…

  ```
  update twgbparm set twgbparm_param_value = 'WWW2_USER' where twgbparm_param_name = 'WEBUSER';
  ```

  That took this query out of the top sql on the Automatic Workload Repository Reports (AWR):

  ```
  SELECT 'Y'
        FROM DUAL
       WHERE EXISTS
           (SELECT 'X'
              FROM GOVFGAC
             WHERE GOVFGAC_FGAC_USER_ID = lv_current_user
               AND (GOVFGAC_UPDATE_PREDICATE IS NOT NULL
                OR  GOVFGAC_SELECT_PREDICATE IS NOT NULL
                OR  GOVFGAC_DELETE_PREDICATE IS NOT NULL
                OR  GOVFGAC_INSERT_PREDICATE IS NOT NULL));
  ```

  - **SDO index ignored after 10.2.0.5 patchset -> Bug.10104555/9743250(96)SPATIAL QUERY PERFORMANCE DEGRADES AFTER APPLYING 10.2.0.5**

# Oracle 10.2.0.5

- The following Linux kernel parameters were updated:

```
net.ipv4.ip_local_port_range = 9000 65500
net.core.rmem_max = 2097152
net.core.wmem_max = 1048576
```

- **The following comes up on some of my systems…works OK when continuing…**

- **Checking for VERSION=2.6.16.21-0.8; found VERSION=2.6.16.46-0.12-default.**

# Oracle Support’s rpm Checker scripts

*** Prerequisite rpms to install AS10g(midtier) IBM: Linux on System z (s390x) [ID 1086769.1]**
*** Requirements for Installing Oracle RDBMS on SLES 11 on zLinux (s390x) [ID 1290360.1]**

```
# rpm -i ora-val-rpm-S10-DB-10.2.0.4-1.s390x.rpm

 error: Failed dependencies:
 openmotif-libs >= 2.2.4-21.12 is needed by ora-val-rpm-S10-DB-10.2.0.4-1.s390x
 compat-32bit >= 2006.1.25-11.2 is needed by ora-val-rpm-S10-DB-10.2.0.4-1.s390x
```

We found "**openmotif-libs-2.3.1-3.13.s390x.rpm**" on the SLES 11 SP1 SDK DVD, and
from the SLES10 SP3 CD "**compat-32bit-2006.1.25-11.2.s390x.rpm**“

**JUST RELEASED!!!**

S11 Database rpm checker (10.2.0.5)) (1.3 KB)
S11 Database rpm checker (11.2.0.2) (1.33 KB)

```
# rpm -i ora-val-rpm-S11-DB-11.2.0.2-1.s390x.rpm
   error: Failed dependencies:
   libstdc++43-devel-32bit >= 4.3.4_20091019-0.7.35 is needed by ora-val-rpm-S11-DB-11.2.0.2-1.s390x
```

**** I’m still getting this one on my Installs**

# Installing 10.2.0.5 Oracle on System z

- The 10.2.0.2 OUI does not recognize SLES-11 as a valid OS version. Therefore, you must invoke OUI as:

  **./runInstaller -ignoreSysPrereqs**

- During installation of 10.2.0.2, errors may be seen during the 'linking' phase of the base 10.2.0.2 software-only installation. If an error is encountered during the 'linking' phase:

  a) click continue (i.e. ignore any linking errors seen during the installation) and complete the installation.

  b) Continue to apply the required 10.2.0.5 patchset Patch:8202632 immediately, which will correct the problem.

- c) Apply the 10.2.0.5.2 patchset Patch:10248542

# PL/SQL Memory Leak – Patch - 5866410

## v$process_memory (for  PL/SQL Insert session only)

**BEFORE:**

```
CATEGORY         ALLOCATED            USED    MAX_ALLOCATED
---------- ---------------- ---------------- ----------------
SQL                60,080          23,072         438,560
PL/SQL             31,224          26,400          31,224
Freeable          131,072               0
Other             980,813                         980,813
```

**AFTER (Running PL/SQL FOR ALL insert)**

```
CATEGORY         ALLOCATED            USED    MAX_ALLOCATED
---------- ---------------- ---------------- ----------------
SQL                52,096          27,752         691,096
PL/SQL        977,385,024       4,434,584     977,389,224
Freeable        1,048,576               0
Other         121,035,317                     822,479,285
```

*************************************************** **AFTER APPLYING PGA PATCH 5866410**
*********************************************************************************

BEFORE:

```
CATEGORY         ALLOCATED            USED    MAX_ALLOCATED
---------- ---------------- ---------------- ----------------
SQL                60,080          23,072         323,080
PL/SQL             31,224          26,400          31,224
Freeable          196,608               0
Other             915,277                         915,277
```

AFTER (Running PL/SQL FOR ALL insert):

```
CATEGORY         ALLOCATED            USED    MAX_ALLOCATED
---------- ---------------- ---------------- ----------------
SQL                   37,968          15,184         329,568
PL/SQL             6,181,248       4,548,968       6,185,448
Freeable        1,048,576               0
Other                816,677                     702,739,117
```

# Review the Oracle “bdump” and “udump” files

**Found these messages:**

WARNING:io_submit failed due to kernel limitations MAXAIO for process=128 pending aio=127

WARNING:asynch I/O kernel limits is set at AIO-MAX-NR=65536 AIO-NR=20608

WARNING:Oracle process running out of OS kernel I/O resources (1)

WARNING:Oracle process running out of OS kernel I/O resources (1)

Add or upate entry in /**etc**/**sysctl.conf**

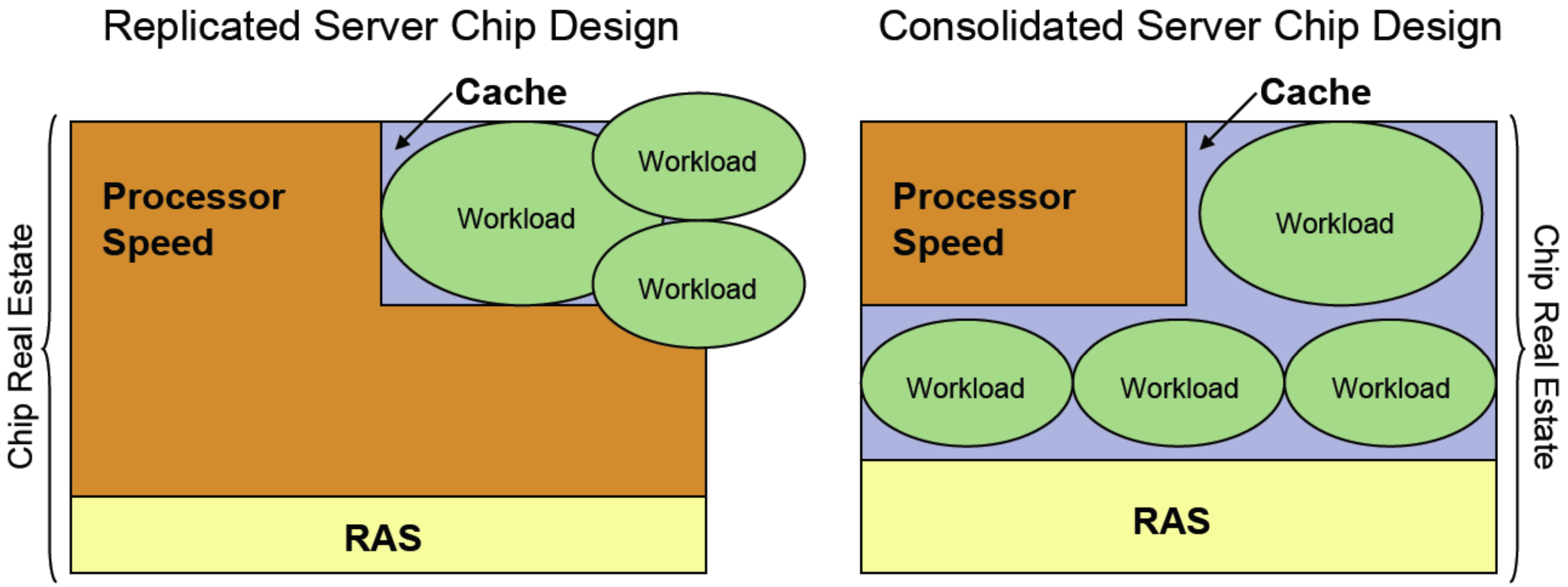
**fs.aio-max-nr = 1048576**

# Oracle 10.2.0.5 and 11gR2 Support

- Jan. 3, 2011 – Oracle 10.2.0.5 for System z Linux Released
  - Patch set 10.2.0.5.2 was delayed a bit as Patch cycles can take ~6 weeks, and the patches were released in mid January.
- 10.2.0.5 is the terminal release for 10gR2, there will be no more 10.2.0.**X** just 10.2.0.5.2
- **Patching end date for 10.2.0.4** has been pushed back from the 30-Apr-2011 date to **31-Jul-2011**
- Extended Support fees will be waived for 10gR2 Customers from August 2010 – July 2012 for Linux on System z customers.
- For 11.2 Customers Extended Support Ends Jan 2018 and Sustaining support Indefinitely.

# zEnterprise CPU Cache and Memory

# Chip Design Affects Virtualization Capabilities

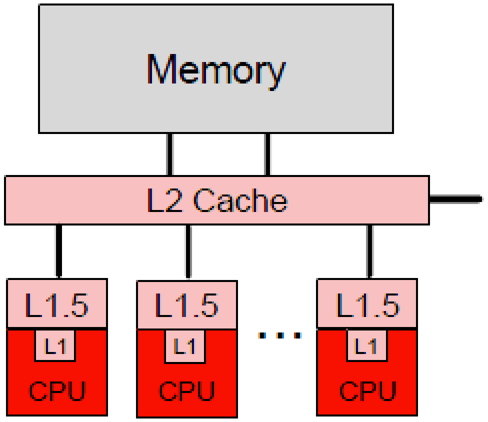


- Mixed workloads stress cache usage, requiring more context switches
- Working sets may be too large to fit in cache
- “Fast” processor speed is not fully realized due to cache misses

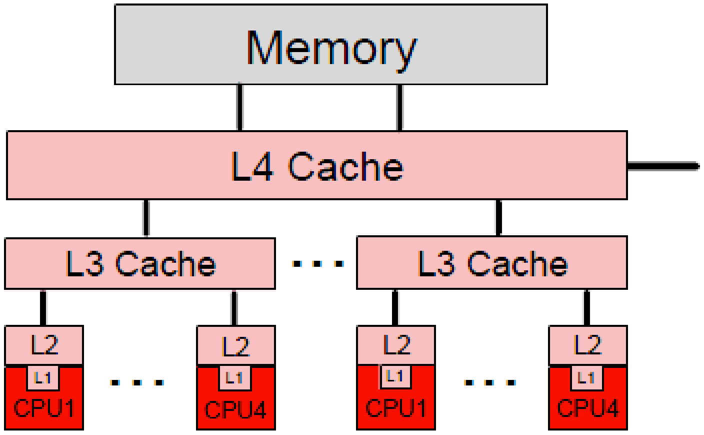
- System z cache is able to contain more working sets
- Processor speed is optimized by increased cache usage
- Additional RAS function is beneficial for mixed workloads

# zEnterprise 196 CPU Cache: Cache is King

- z10 EC
  - CPU
    - 4.4 Ghz
  - Caches
    - L1 private 64k i, 128k d
    - L1.5 private 3 MBs
    - L2 shared 48 MBs / book
    - book interconnect: star

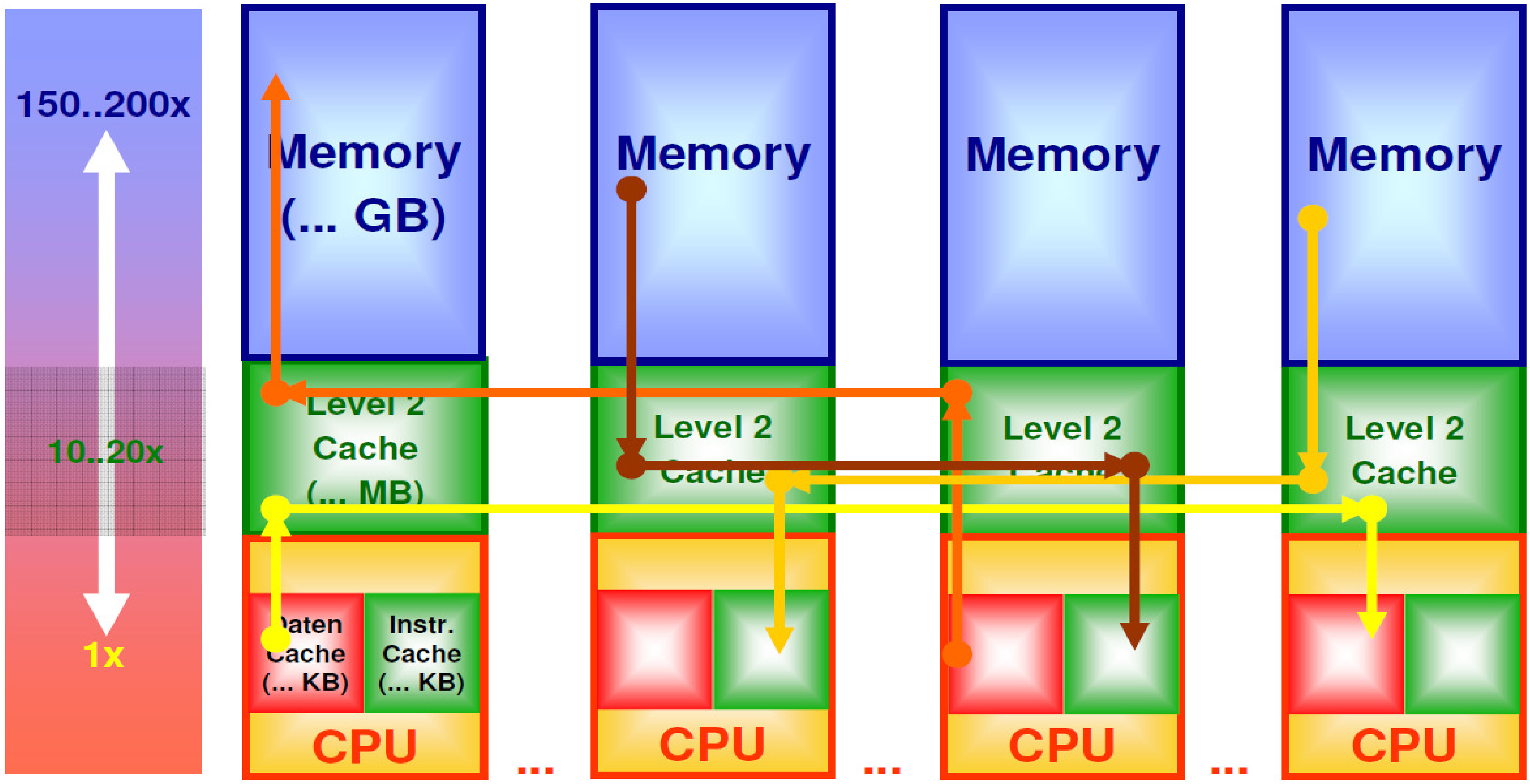


- z196
  - CPU
    - 5.2 Ghz
    - Out-Of-Order execution
  - Caches
    - L1 private 64k i, 128k d
    - L2 private 1.5 MBs
    - L3 shared 24 MBs / chip
    - L4 shared 192 MBs / book
    - book interconnect: star



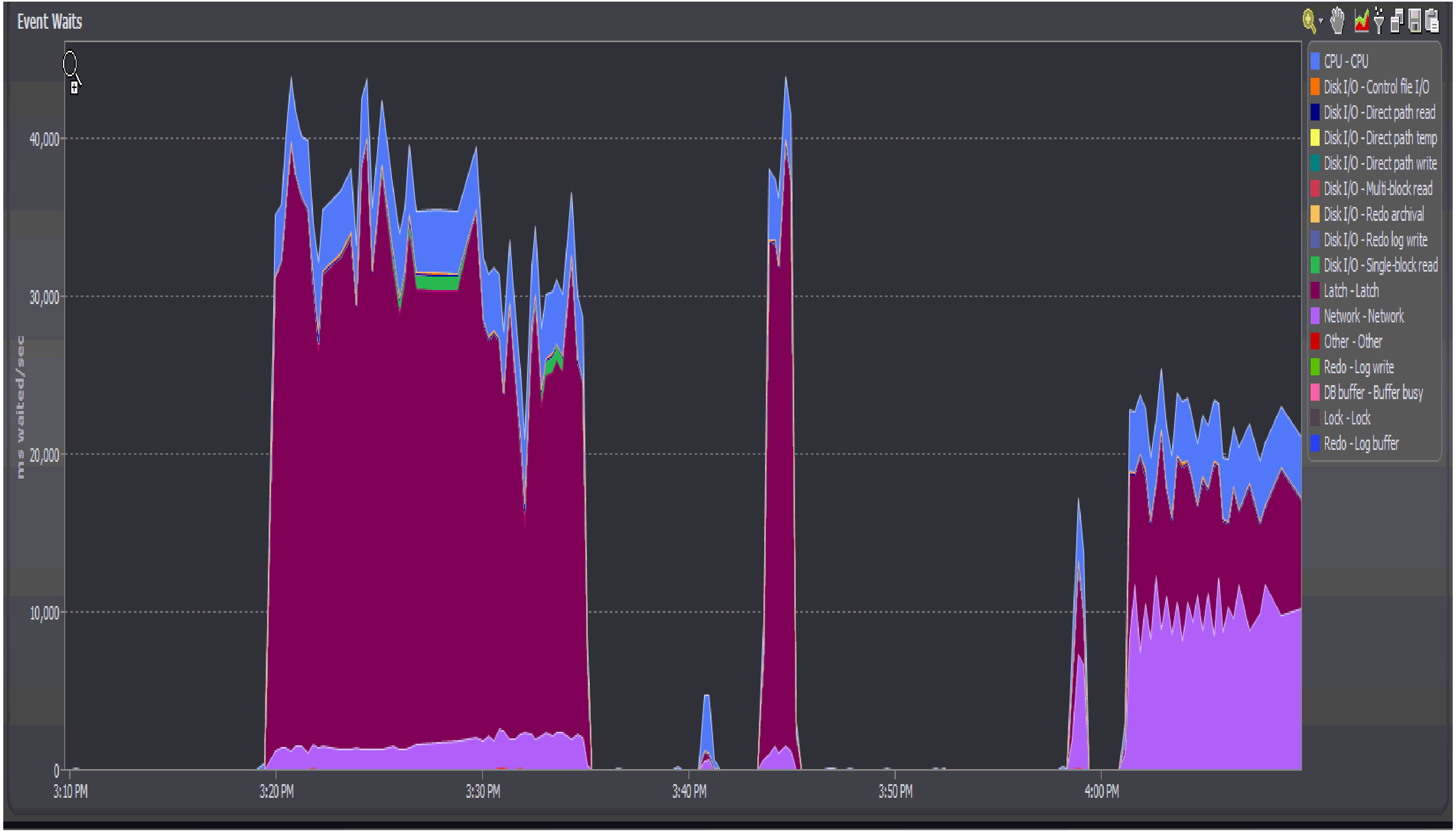
# Scalability: System-Structures optimized for data

**The key problem of current microprocessor-systems:**
**Memory access does not scale with CPU-cycletime !**



Source: **Klaus-Dieter Mueller (IBM)**

# High CPU, Latches – Database Shared Connections



# CMM enabled – Maybe NOT!

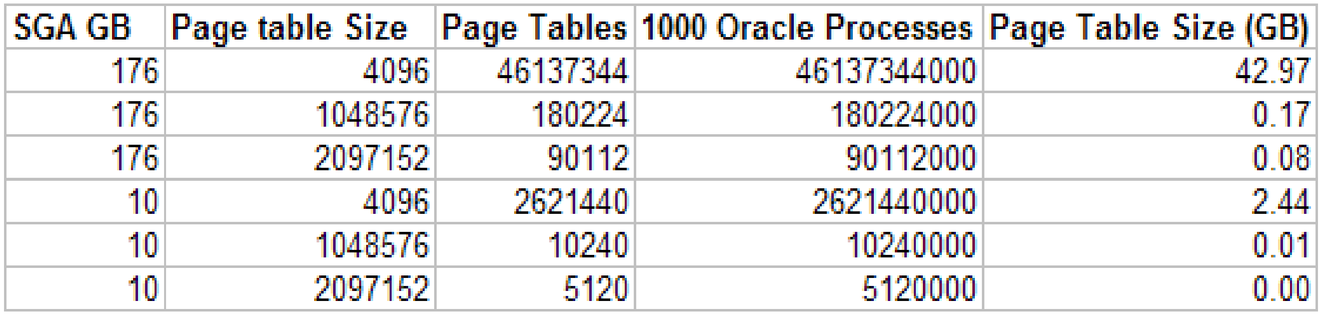
In /var/log/messages we encountered the following:

```
Jan 31 14:03:37 zlomsp10 kernel: kernel BUG at mm/page-discard.c:187!
Jan 31 14:03:37 zlomsp10 kernel: illegal operation: 0001 [#1]
Jan 31 14:03:37 zlomsp10 kernel: CPU: 0 Not tainted
Jan 31 14:03:37 zlomsp10 kernel: Process oracle (pid: 16449, task: 000000017cf5ac58, ksp: 000000010a44fa30)
Jan 31 14:03:37 zlomsp10 kernel: Krnl PSW : 0704000180000000 000000000021f032 (page_discard+0x9a/0x2d8)
Jan 31 14:03:37 zlomsp10 kernel: Krnl GPRS: 0000000000989680 000000010a44f838 0000000000000028 0000000000604998
Jan 31 14:03:37 zlomsp10 kernel: 000000000021f02e 0000000000000000 00000200054f0000 0000000000000001
Jan 31 14:03:37 zlomsp10 kernel: 000000000000001b 0000000000000007 00000001188a7ad8 0000000083a19f28
Jan 31 14:03:37 zlomsp10 kernel: 0000000083a19f28 00000000004c3138 000000000021f02e 000000010a44f8c0
Jan 31 14:03:37 zlomsp10 kernel: Krnl Code: e3 10 b0 06 00 90 a7 11 00 08 a7 84 00 0c c0 20 00 15 bf 27
Jan 31 14:03:37 zlomsp10 kernel: Call Trace:
```

Then 15 minutes later (900 seconds) Oracle throws this and crashes as it complete the archive operation.
**ORA-00494: enqueue [CF] held for too long (more than 900 seconds) by 'inst 1, osid 29091**

- CMM code is changed in SLES 11 (mm never accepted upstream)
- **Solution:** Either Move to SLES 11 and Oracle 10.2.0.5+ or remove **'cmma=yes'** from zipl.conf

# Oracle and HugePages

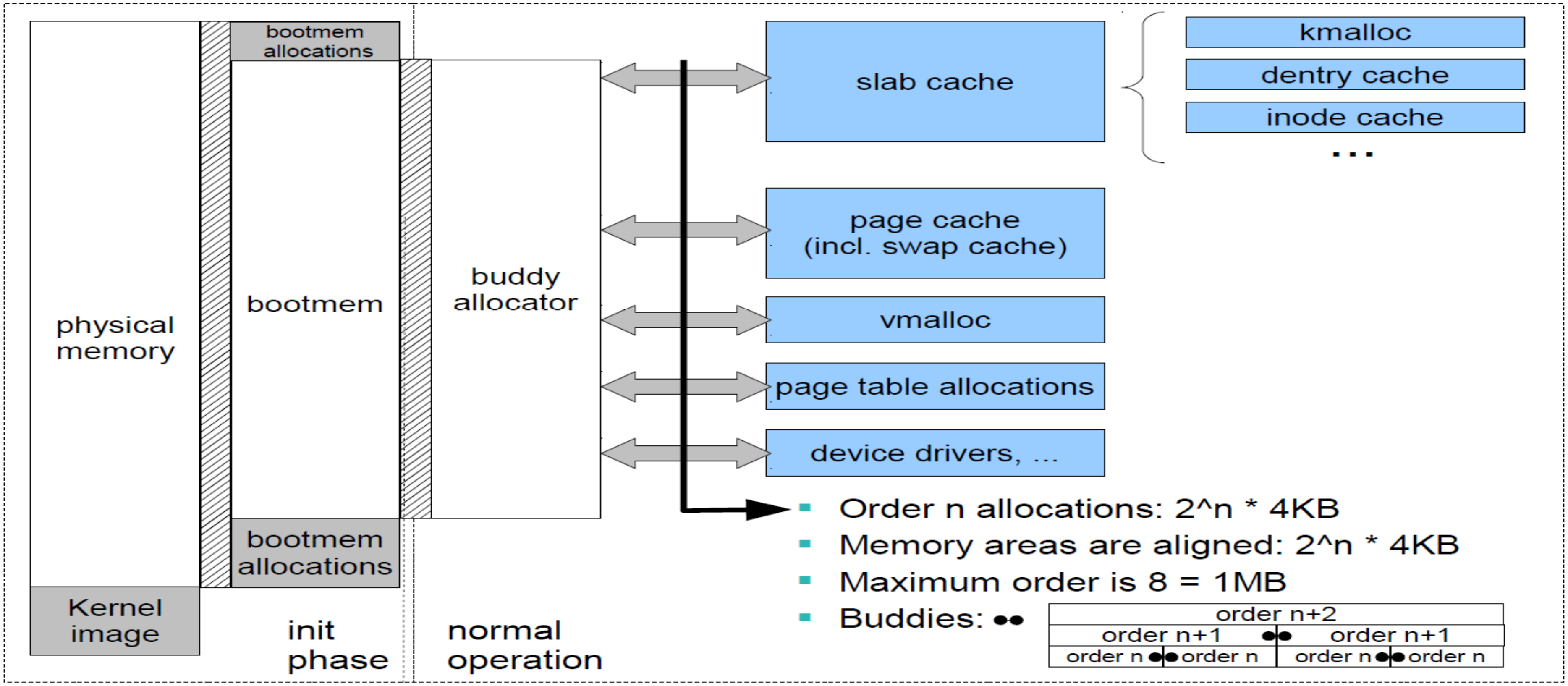


| SGA GB | Page table Size | Page Tables | 1000 Oracle Processes | Page Table Size (GB) |
|---:|---:|---:|---:|---:|
| 176 | 4096 | 46137344 | 46137344000 | 42.97 |
| 176 | 1048576 | 180224 | 180224000 | 0.17 |
| 176 | 2097152 | 90112 | 90112000 | 0.08 |
| 10 | 4096 | 2621440 | 2621440000 | 2.44 |
| 10 | 1048576 | 10240 | 10240000 | 0.01 |
| 10 | 2097152 | 5120 | 5120000 | 0.00 |

- Some debate will this even work with System z and Oracle -> Main benefit is the reduced Linux Page tables Size, we can then in turn use this memory for Oracle SGA
- If Running Native LPAR – z10+ has hardware acceleration – needs to be tested ~ 10% gain, most likely no performance gain other than free memory benefit for z/VM on current releases.
- Not recommended to use AMM (Automatic Memory Management) - Oracle Support Note – 361323.1
- Some Distributions have 1024 kB (SLES 11 SP1) pages tables and some 2048 kB (SLES 10 SP3, RHEL 5.3+)
- Not Currently supported for 10gR2 on System z, backport has been requested by Several Customers.
- Helps Reduce TLB misses and nested page table traversal
- Oracle SGA consists of a series of pointers, each Oracle process than connects needs to create a memory map.

# Memory and Linux Guest Size – Kernel Image

## Basic memory allocation



- Order n allocations: 2^n * 4KB
- Memory areas are aligned: 2^n * 4KB
- Maximum order is 8 = 1MB
- Buddies: ●●

# OOM killer – when the kernel runs out of memory

- “Nobody likes the out-of-memory (OOM) killer. Its job is to lurk out of sight until that unfortunate day when the system runs out of memory and cannot get work done; the OOM killer must then choose a process to sacrifice in the name of continued operation. It's a distasteful job, one which many think should not be necessary. But, despite the OOM killer's lack of popularity, we still keep it around; think of it as the kernel equivalent of lawyers, tax collectors, or Best Buy clerks. Every now and then, they are useful. “ (John Corbet, Linux Weekly News)
- Kicks in when kernel is out of memory and no swap space is available, or when kernel is out of memory and swap device is stuck
- Any OOM situation without 100% swap usage is a Kernel BUG
- Swap token algorithm (Sles 10), multiple parallel direct reclaimers (Sles 11)
- Frequent improvements (Sles 11 over Sles 10), better OOM heuristics: what process to kill?

# Oracle Databases and Using DASD and FCP Disk Storage

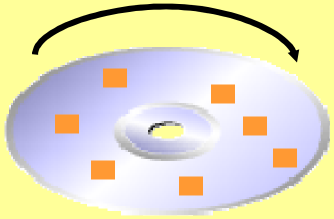
# I/O Metrics for Databases

## Disk bandwidth

**Metric = IOPS and latency**

**Best with High RPM**

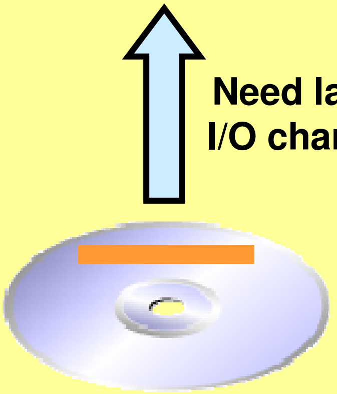
**and fast seek time**



**OLTP**
**(Small random I/O)**

## Channel bandwidth

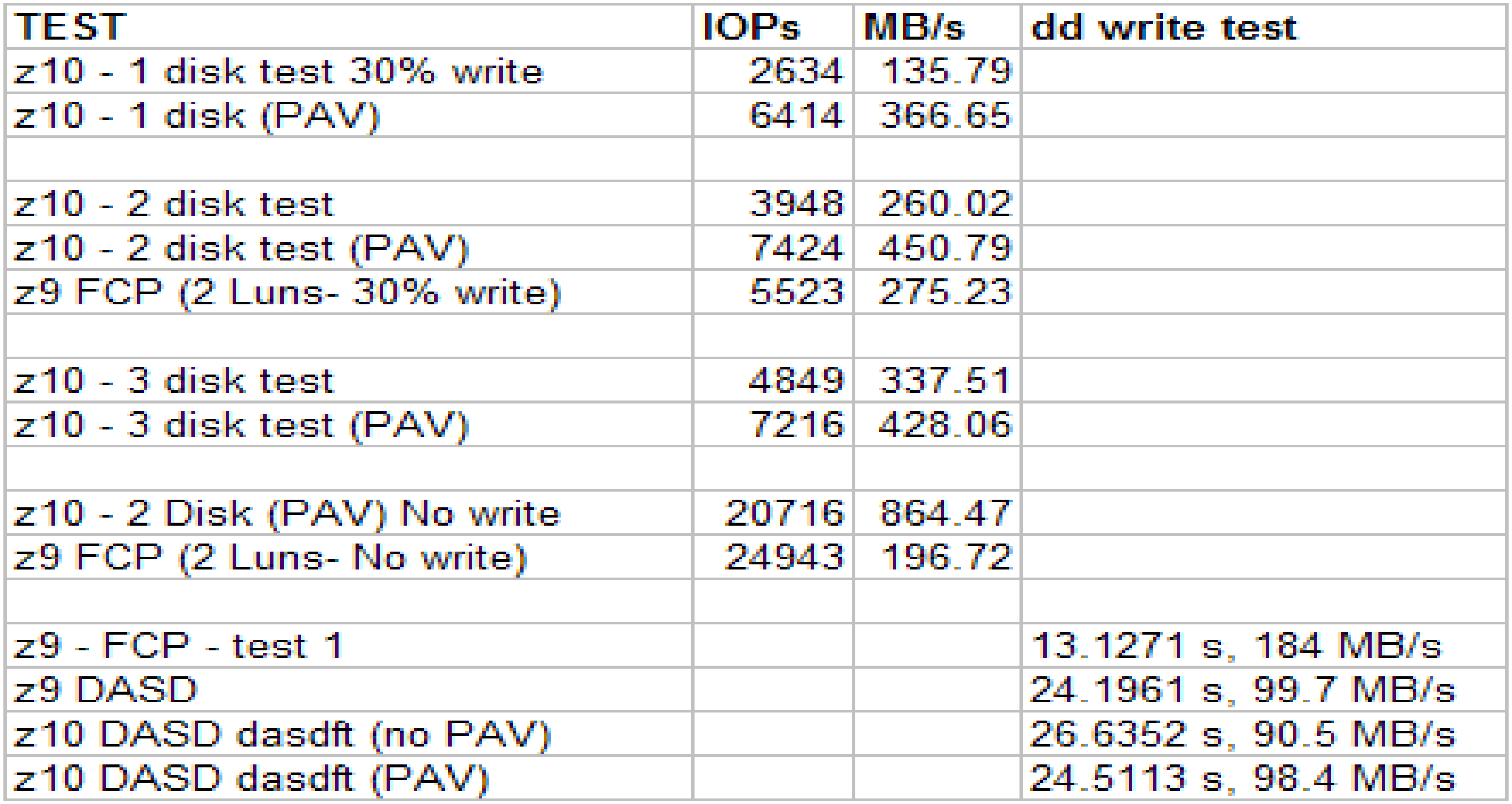
**Metric = MBPS**



**Need large I/O channel**

**DW/OLAP**
**(Large sequential I/O)**

# Working with DASD PAV Storage



| TEST | IOPs | MB/s | dd write test |
|---|---|---|---|
| z10 - 1 disk test 30% write | 2634 | 135.79 | |
| z10 - 1 disk (PAV) | 6414 | 366.65 | |
| | | | |
| z10 - 2 disk test | 3948 | 260.02 | |
| z10 - 2 disk test (PAV) | 7424 | 450.79 | |
| z9 FCP (2 Luns- 30% write) | 5523 | 275.23 | |
| | | | |
| z10 - 3 disk test | 4849 | 337.51 | |
| z10 - 3 disk test (PAV) | 7216 | 428.06 | |
| | | | |
| z10 - 2 Disk (PAV) No write | 20716 | 864.47 | |
| z9 FCP (2 Luns- No write) | 24943 | 196.72 | |
| | | | |
| z9 - FCP - test 1 | | | 13.1271 s, 184 MB/s |
| z9 DASD | | | 24.1961 s, 99.7 MB/s |
| z10 DASD dasdft (no PAV) | | | 26.6352 s, 90.5 MB/s |
| z10 DASD dasdft (PAV) | | | 24.5113 s, 98.4 MB/s |

# PAV and FCP - Considerations

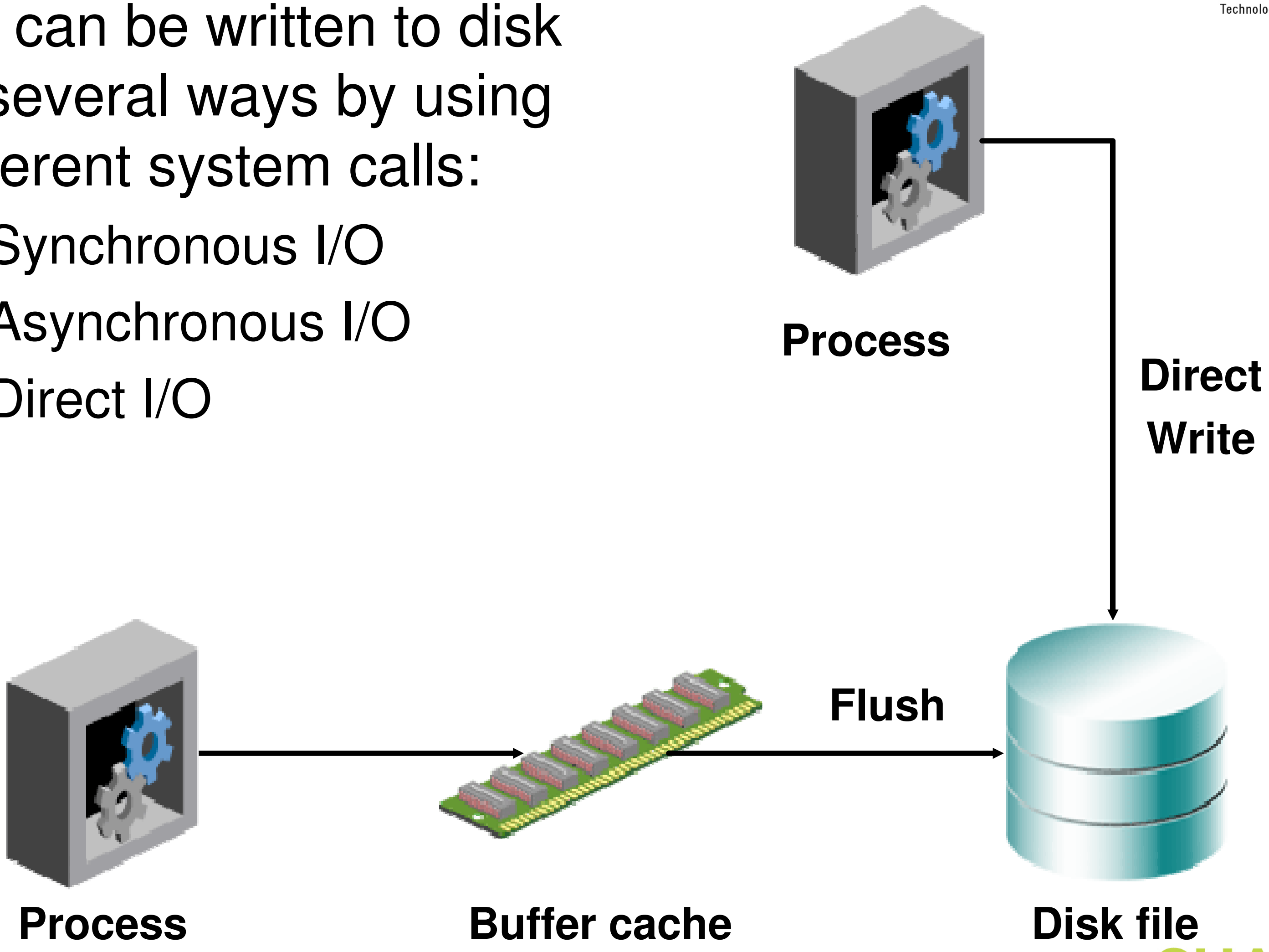
```
# lsdasd -u
Bus-ID    Name      UID
==============================================================================
0.0.0200  dasdk     IBM.75000000066342.1837.05.00000191000001f40000000000000000
0.0.0301  dasdl     IBM.75000000066342.1837.09.0000000100000d0a0000000000000000
0.0.0400  dasda     IBM.75000000066342.1838.04.000000000000ffef0000000000000000
0.0.38cd  alias     IBM.75000000066342.1838.04.000000000000ffef0000000000000000
0.0.38ce  alias     IBM.75000000066342.1838.04.000000000000ffef0000000000000000
0.0.38cf  alias     IBM.75000000066342.1838.04.000000000000ffef0000000000000000
0.0.38d0  alias     IBM.75000000066342.1838.04.000000000000ffef0000000000000000
0.0.38d1  alias     IBM.75000000066342.1838.04.000000000000ffef0000000000000000
0.0.38d2  alias     IBM.75000000066342.1838.04.000000000000ffef0000000000000000
0.0.38d3  alias     IBM.75000000066342.1838.04.000000000000ffef0000000000000000
0.0.38d4  alias     IBM.75000000066342.1838.04.000000000000ffef0000000000000000
0.0.0401  dasdb     IBM.75000000066342.1839.00.000000000000ffef0000000000000000
0.0.3905  alias     IBM.75000000066342.1839.00.000000000000ffef0000000000000000
0.0.3906  alias     IBM.75000000066342.1839.00.000000000000ffef0000000000000000
0.0.3907  alias     IBM.75000000066342.1839.00.000000000000ffef0000000000000000
0.0.3908  alias     IBM.75000000066342.1839.00.000000000000ffef0000000000000000
0.0.3909  alias     IBM.75000000066342.1839.00.000000000000ffef0000000000000000
0.0.390a  alias     IBM.75000000066342.1839.00.000000000000ffef0000000000000000
0.0.390b  alias     IBM.75000000066342.1839.00.000000000000ffef0000000000000000
0.0.390c  alias     IBM.75000000066342.1839.00.000000000000ffef0000000000000000
```

**For FCP Queue Depth (XIV):**

```
cat /sys/module/zfcp/parameters/queue_depth
echo 128 > /sys/bus/scsi/devices/0:0:0:1073955055/queue_depth
```

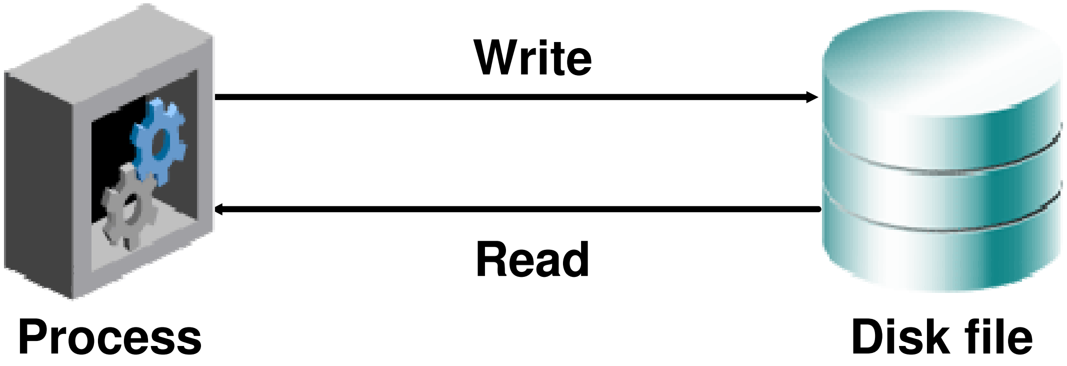
# I/O Modes

- I/O can be written to disk in several ways by using different system calls:
  - Synchronous I/O
  - Asynchronous I/O
  - Direct I/O



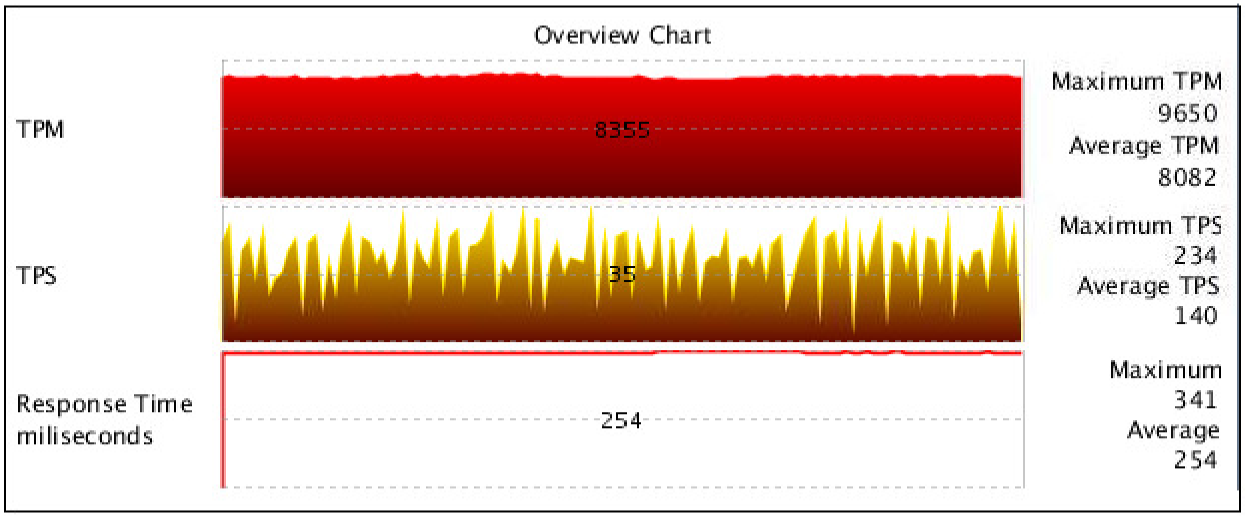
# Direct I/O

- Direct I/O is considered to be the high-performance solution.
  - Direct reads and writes do not use the OS buffer
  - Direct reads and writes can move larger buffers than file system I/Os.
  - Set `filesystemio_options=setall for LVM file systems`

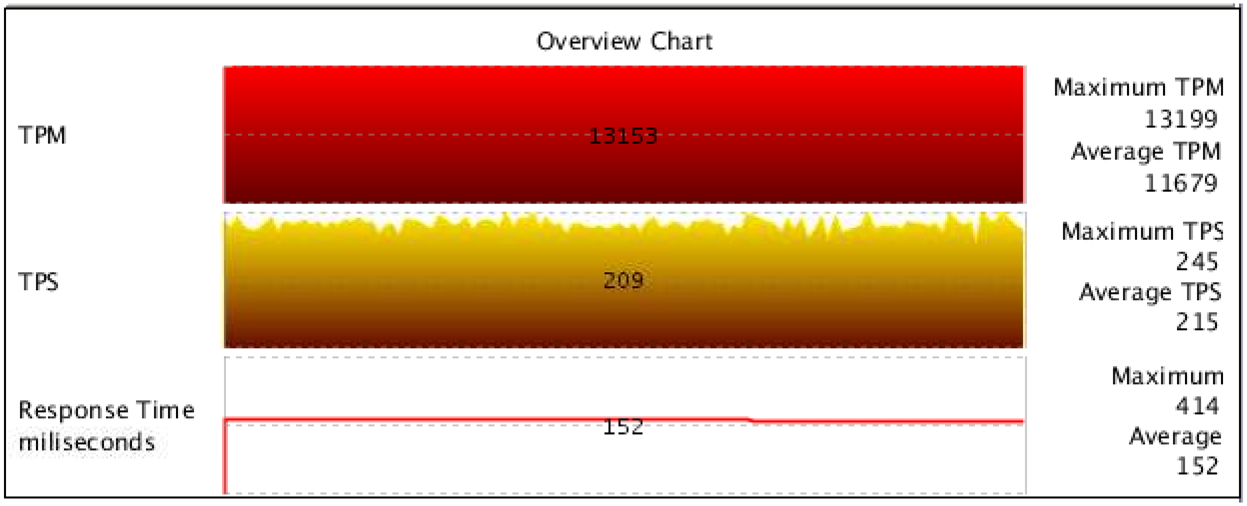


# filesystemio_options Oracle parameter Swingbench Test

filesystemio_options=none, SGA Undersized (Linux cache) – 10GB Linux



filesystemio_options=setall, SGA rightsized – 10GB Linux -

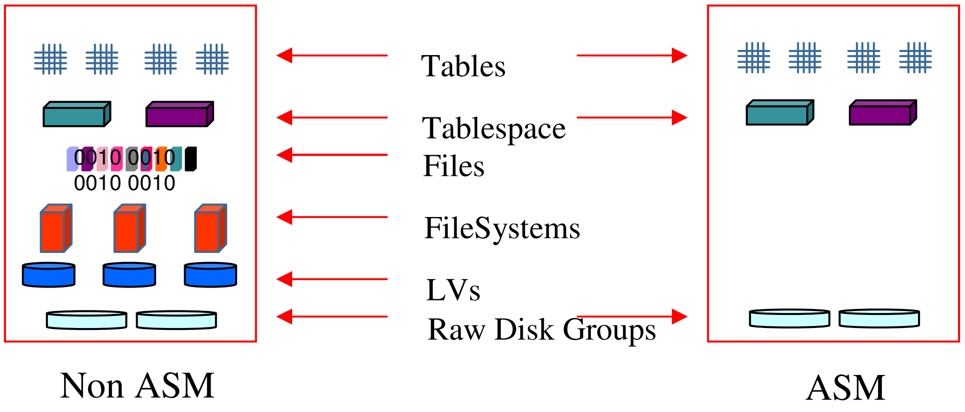


SHARE in Anaheim 2011

# Oracle I/O – Direct and ASYNCH I/O

- *FILESYSTEMIO_OPTIONS* = { *none* | **setall** | directIO | asynch }.
- Most of the time use **filesystemio_options = setall** (provides asynch + directIO) as we tend to size to what's needed with virtualization.
- Some DSS Applications work better with asynch (particularly lots of TEMP writes).
- Size the Oracle buffer cache if changing from **none** or if migrating to ASM to account for any compensation of the Linux file system buffer cache.
- Linux will eventually use up all the free memory available (feature) for such things as file caching so understanding the data your looking at before changing.
- If using Oracle ASM not as important to set.

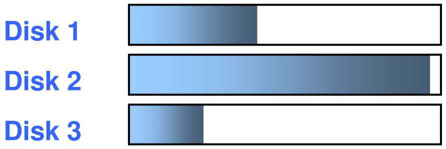
# Automated Storage Management ( ASM )



- Eliminates need for conventional file system and volume manager
- ASM extends SAME (Stripe and Mirror Everything)
- Improved performance, scalability, and reliability

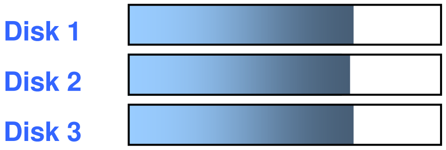
**Before ASM**

Conventional wisdom



**With ASM**

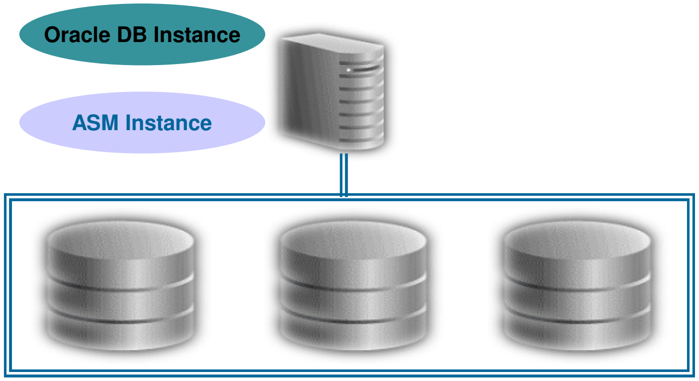
Provisioning storage when you need it…



F

## ASM is Oracle's integrated clusterware

- Capacity on demand
  - Add/drop disks online
  - **Automatic I/O load balancing**
  - Stripes data across disks to balance load
  - Best I/O throughput
  - Automatic mirroring and stripping
- **Easy to manage**
- Can only host datafiles, not binaries



**Disk group**

# LVM vs ASM for Single Instance Databases

- With LVM you do not have to Install, Learn and Manage an ASM instance.
- Ensure that the LVM is stripped (especially for random I/O)
- When adding storage you should add same number of disks in the stripe plex. i.e. if you stripe across 4 disks, you'll need to add another 4 disks.
- ASM is handy for dynamically resizing and migrating disk while the system is running.
- ASM is a logical stepping stone to other Oracle MAA strategies – RAC
- ASM has a daemon processes that runs called oclsd.bin that needs to be watched to ensure under utilized Linux Guests drop from queue in order to release memory.
- If using ASMLib (a utility for ASM) needs to make sure you upgrade with the Kernel level when patching the Linux operating system.
- Memory constrained systems – the ASM instance processes are sometimes the first to go. (OOM – Out of Memory Manager)

# Oracle RAC Troubleshooting

# Reasons For Oracle Node Evictions:

- **Hang Check Timer** -
  replaced by **oprocd** process in 10.2.0.4 and tuned with the CRS **diagwait** parameter:
  - Oracle still recommends the Linux hang checker to be set in 10g & 11g though:

    **/sbin/insmod hangcheck-timer hangcheck_tick=1 hangcheck_margin=10 hangcheck_reboot=1**

  - Oracle oprocd process sets a timer, then sleeps. When oprocd wakes up again and gets scheduled onto the cpu if it sees that a longer time has passed than the acceptable margin, oprocd will reboot the node.

**Linux Script - /var/log/messages**

```
Apr 17 20:36:23 orainst040 logger: WARNING Fri Apr 17 20:36:23 CDT 2009 The date loop took longer than 2 seconds  from 1240018575 to
     1240018583
Apr 17 20:42:51 orainst040 logger: WARNING Fri Apr 17 20:42:51 CDT 2009 The date loop took longer than 2 seconds  from 1240018963 to
     1240018971
```

**/etc/oracle/oprocd/<nodename>.oprocd.lgl**

```
Apr 24 05:20:03.665 | INF | TrackHistoricalTrends: added first sample 3242554777 in 10 to 50 percentile
Apr 26 05:26:23.593 | INF | TrackHistoricalTrends: added first sample 2642327278 in 10 to 50 percentile
```

2 entries where the delay in scheduling was approximately 3.24 and 2.64 seconds respectively

- **I/O** - Inability for a Node to communicate with the Voting disk – after 200 seconds will evict.

- **Network** Interconnect cannot ping other node for more than 60 seconds
  - z/VM uses OSA Cards with redundancy connected with 10GB Switch

- **Lack of Memory** - causing Linux to do emergency page scans and kill processes
  - Increase swap space to provide a "**safety net**" for unexpected workloads

- **VM Page Reordering: For every 8GB ~1 second of delay.**
  Turn off for Really Large Oracle Guests. http://www.vm.ibm.com/perf/tips/reorder.html

# Trouble Shooting – Oracle RAC Evictions

- **Scenario ->** Oracle RAC in Production for Several years, and started having sporadic node evictions after Linux Kernel upgrade.

- The "**init.cssd fatal**" and the "**init.cssd daemon**" processes were somehow terminating.

- When the Linux kernel was rolled back the problem went away (just the kernel rpm only)

# Oracle RAC Node Evictions – cont'd

What we were seeing could be reproduced with the init.cssd daemon script being killed

```
root@orausr07 bin]# ps -ef | grep init.cssd
root     1998    1  0 17:47 ?      00:00:02 /bin/sh /etc/init.d/init.cssd fatal
root     2704  1998  0 17:47 ?      00:00:00 /bin/sh /etc/init.d/init.cssd oprocd
root     2722  1998  0 17:47 ?      00:00:00 /bin/sh /etc/init.d/init.cssd oclsomon
root     3043  1998  0 17:47 ?      00:00:00 /bin/sh /etc/init.d/init.cssd daemon

[root@orausr07 bin]# kill -11 1998
[root@orausr07 bin]# ps -ef | grep init.cssd | grep -v grep
root     2704    1  0 17:47 ?      00:00:00 /bin/sh /etc/init.d/init.cssd oprocd
root     2722    1  0 17:47 ?      00:00:00 /bin/sh /etc/init.d/init.cssd oclsomon
root     3043    1  0 17:47 ?      00:00:00 /bin/sh /etc/init.d/init.cssd daemon
root    14435    1  0 18:15 ?      00:00:00 /bin/sh /etc/init.d/init.cssd fatal
```

**We then ran kill -11 3043 (daemon process) and the following appeared in /var/log/messages (NO - Oracle CRS failure. Rebooting for cluster integrity. messages )**

**/var/log/messages**

```
Oct 25 18:15:38 orausr07 logger: Cluster Ready Services completed waiting on dependencies.
Oct 25 18:15:38 orausr07 logger: Oracle CSS Family monitor restarting.
Oct 25 18:16:20 orausr07 logger: Oracle CSS restart. 0, 1
```

Box goes down..So what it looked like to to us was init.cssd fatal was respawning with this particular kernel level. What was causing this respawing at this kernel level was the BIG question.

# Oracle RAC Node – Eviction Conclusions

- Extra tracing  in ocssd.bin - CLSMON_Args=-trace. This throws a print every 500 msec (so you'll need lots of disk)

- We ran sh –x in the inittab.cssd (this will increase the logging substantially and disk!

- We ran strace on the Oracle Linux process that were running

- Collaboration with Oracle, Linux and IBM support teams with everyone on daily status calls.

# New Release Coming Up! – Q1 2011

# Oracle 11gR2 (Generic) – New Features

- SecureFile LoBs
- Database Replay (Real Application Testing)
- 11.2.0.1 vs 11.2.0.2 -> Performance and Bug Fixes
- Come to the Session on Thursday – 203B – 1:30pm

SHARE
Technology • Connections • Results

SHARE
in Anaheim
2011