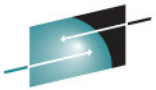
Installation Experiences & Recommendations for a Successful Install of Oracle 11gR2 on Linux on System z

Speaker Name: David Simpson
Speaker Company: IBM

Date : 1:30 PM, Monday, August 8, 2011
Session Number: 09881

Email: simpson.dave@us.ibm.com

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.
- Oracle is a Trademark of Oracle Corporation in the United States and other countries.

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

- 11gR2 Installation Changes
- Current Hot Topics with Oracle on System z Linux
- New Features to Consider for 11gR2
- Customer Experiences 11gR2 with Linux on System z

11gR2 Installation Changes

Oracle 11gR2 Documentation:

- Start with Oracle Support Notes (MOS) updated with the latest information:

1306465.1 - Getting Started 11gR2 on System z Linux

1290644.1 - Installing 11gR2 on SLES 11 on IBM: Linux on System z (s390x)

1308859.1 - Installing 11gR2 on SLES 10 SP3 on IBM: Linux on System z (s390x)

1306889.1 - 11gR2 RHEL 5 on System z Linux Requirements

- Two Types of Installs those Involving **Oracle Grid** (RAC – Real Application Clusters and Automated Storage Management) and those involving **Oracle Database Only**.

- For Oracle Grid Installs, the Oracle Grid Infrastructure Installation Guide 11g Release 2 (11.2) for Linux document **E17212-10** provides detailed information and has sections for System z Linux.

Oracle Software and Patches:

Link-> (not on E-Delivery)

http://www.oracle.com/technetwork/database/enterprise-edition/downloads

Oracle Database 11*g* Release 2 (11.2.0.2.0) for zLinux64

linux.zseries64_11gR2_database_1of2.zip (1,441,455,828 bytes)

linux.zseries64_11gR2_database_2of2.zip (1,009,427,871 bytes)

- For ASM or Oracle Grid (RAC):

Oracle Database 11g Release 2 Grid Infrastructure (11.2.0.2.0) for zLinux64

linux.zseries64_11gR2_grid.zip (756,155,780 bytes)

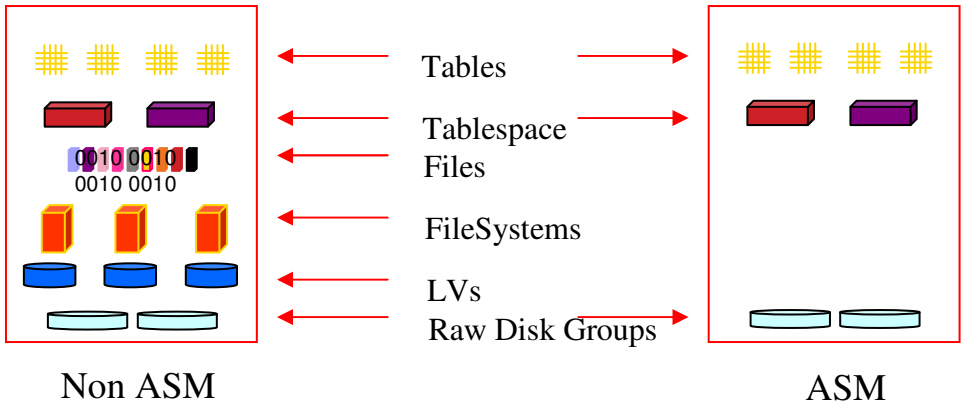
- 11.2.0.2.3 PSU (Database - 12419331)

https://updates.oracle.com/Orion/SimpleSearch/process_form?search_type=patch&patch_number=12419331&plat_lang=209P

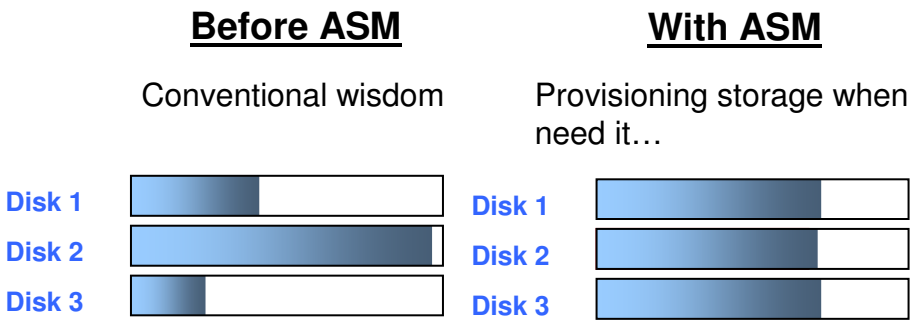
- 11.2.0.2.3 PSU (Grid - 12419353)

https://updates.oracle.com/Orion/Services/download/p12419353_112020_Linux-zSer.zip?aru=14001838&patch_file=p12419353_112020_Linux-zSer.zip

Automated Storage Management (ASM)



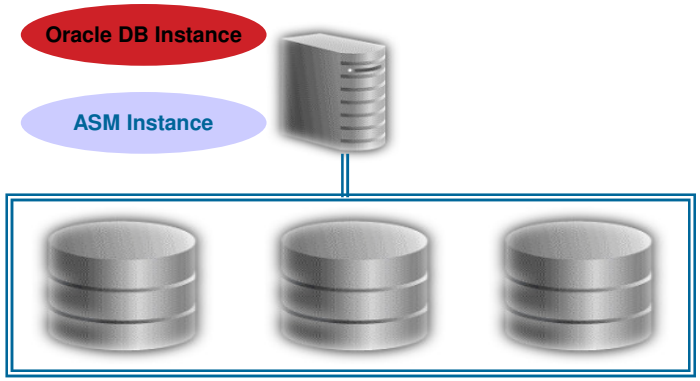
- **Eliminates need for conventional file system and volume manager**
- **ASM extends SAME (Stripe and Mirror Everything)**
- **Improved performance, scalability, and reliability**



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



Memory Sizing for 11gR2

- 11gR2 Oracle recommends **4.0 GB (4096 MB) of RAM** for all their Linux platforms.

- Testing with System z Linux has shown that **1GB** is too small (particularly if using Oracle grid's product), excessive Linux swapping occurring. **2GB** of virtual memory is the smallest we would recommend for an 11gR2 database.

- If upgrading from 10gR2 to 11gR2, we have seen an increase of approximately 200 mb with 11gR2. – **Customer Production Experience**

Disk Space:

i) Approximately **5.5 GB** of disk space is required for Oracle Grid Infrastructure (RAC) or a Single Instance Grid Cluster ASM Home. (1.8 GB 10gR2 for CRS before),

ii) Approximately **4.6 GB** of disk space is required for the database software. (2.1GB ASM, 2.5GB DB Home 10gR2)

iii) **1.0 GB** of disk space is recommended for the /**tmp** directory (or another temporary directory if environment variables **TMP** and **TEMP** are set to this directory) for Oracle to stage software for the install of executables.

Supported Kernel Versions for 11gR2

- **Red Hat 5.4+** -> Linux **2.6.18-238** or greater for Oracle RAC environments due to an incident of sporadic reboots with a lower kernel version and 10gR2 CRS
- Red Hat 6.0 is NOT Supported for any Linux Platform at this time
- SUSE 10 SP3 (or greater), Kernel -**2.6.16.60-0.54.5** or newer is required for an 11gR2 SUSE Installation.
- SUSE 11.0 SP1 (**2.6.32.12-0.7**) +

Result:

```
# cat /proc/version
Linux version 2.6.32.12-0.7-default (geeko@buildhost) (gcc version 4.3.4 [gcc-4_3-branch revision
152973] (SUSE Linux) ) #1 SMP 2010-05-20 11:14:20 +0200
```

Use the Linux rpm checker!

- Download the “rpm checker” from the bottom of My Oracle Support (MOS) Note **1306465.1**
- The rpm checker checks that the required rpms for Oracle Grid and Database installs. This prevents problems with the installation of Oracle.

RHEL5 - 11.2 Grid Infrastructure, SIHA, DB Install - Red Hat

S10 Grid Infrastructure/Database rpm checker - SLES 10

S11 Grid Infrastructure/Database rpm checker 11.2.0.2 -SLES 11

Running the Linux rpm Checker:

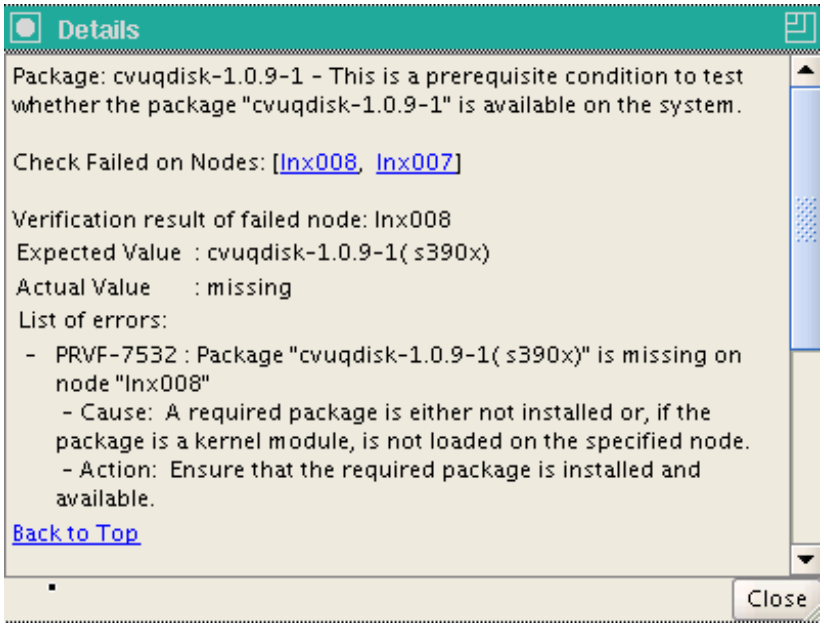
- Download the rpm checker, unzip then run rpm to install (the rpm checker does not actually install anything just checks the pre-reqs for you)

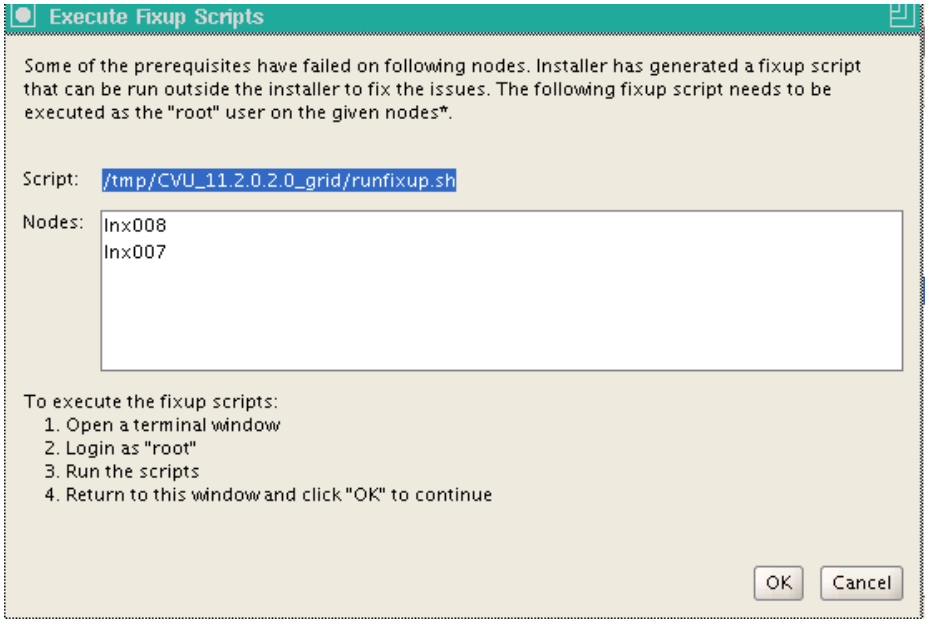
Result:

```
# rpm -ivh ora-val-rpm-EL5-DB-11.2.0.2-1.s390x.rpm
Preparing...                ########################################### [100%]
   1:ora-val-rpm-EL5-DB     ########################################### [100%]
********************************************************************************
*      Validation complete - please install any missing rpms                   *
*    The following output should display both (s390) - 31-bit and              *
*         (s390x) 64-bit rpms - Please provide the output to Oracle            *
*               Support if you are still encountering problems.                *
********************************************************************************
Found           glibc-dev (s390)
Found           glibc-dev (s390x)
Found           libaio (s390)
Found           libaio (s390x)
Found           compat-libstdc++-33 (s390)
Found           compat-libstdc++-33 (s390x)
Found           glibc (s390)
Found           glibc (s390x)
Found           libgcc (s390)
Found           libgcc (s390x)
Found           libstdc++ (s390)
Found           libstdc++ (s390x)
Found           libaio-devel (s390)
Found           libaio-devel (s390x)
```

Optional - Oracle Grid – cvudisk-1.0.9-1 rpm

- Oracle Grid install, you will need the cvudisk-1.0.9-1 rpm package from the Oracle 11gR2 distribution media.
- You can do this as part of a fix-up script or pre-install from the Oracle distribution.





NTP Time Check for Oracle Grid Installs

Red Hat:

modify /**etc/sysconfig/ntpd** add the -x flag
OPTIONS="-x -u ntp:ntp -p /var/run/ntpd.pid"

Restart the network time protocol daemon
/sbin/service ntpd restart

Ensure that the ntpd daemon is for system restart
chkconfig --level 35 ntpd on

SUSE:

modify /**etc/sysconfig/ntp** add the -x flag
NTPD_OPTIONS="-x -g -u ntp:ntp"

Restart the network time protocol daemon
/sbin/service ntp restart

Ensure that the ntpd daemon is for system restart
chkconfig --level 35 ntp on

Hardware Clock Synchronization Check

- With SLES 11 systems, you may encounter the following Warning, when Oracle runs the Oracle Grid System check.

PRVE-0029 : Hardware clock synchronization check could not run on node xxx"

- Not mandatory to fix, you can add the following lines to the " **/etc/init.d/halt.local** " file (**NOTE the # comment**)

```
CLOCKFLAGS="$CLOCKFLAGS --systohc"
#/sbin/hwclock --systohc
```

Oracle 11gR2 Installer – Many Improvements

- Easier to Install
- Improved De-Install process
- User Equivalency checker
- Automatically generated Fix Up scripts

Result:

```
root@lnx007 CVU_11.2.0.2.0_grid]# ./runfixup.sh
/usr/bin/id
Response file being used is :./fixup.response
Enable file being used is :./fixup.enable
Log file location: ./orarun.log
Installing Package /tmp/CVU_11.2.0.2.0_grid//cvuqdisk-1.0.9-1.rpm
Preparing...                ########################################### [100%]
   1:cvuqdisk                ########################################### [100%]
```

Multipath for FCP/SCSI Luns

```
multipath {
        wwid     3600507630bffc2ce0000000000001112
        alias    lun40
        path_grouping_policy    failover
        uid      501
        gid      501
        mode  660
    }
```

- **No longer require a disk partition for 11gR2! OS Vendors recommend this as well.**
- **Required for Device Persistence (tied to WWID)**
- **Required for Oracle grid user file permissions**
- **Use the /dev/mapper/<alias name> as the ASM Diskstring**

Linux UDEV Rules for Oracle

Create a **/etc/udev/rules.d/99-udev-oracle.rules** file to assign permissions for DASD devices.

```
vi /etc/udev/rules.d/99-udev-oracle.rules
```

Result:

```
KERNEL=="dasd*1",ID=="0.0.0300",OWNER="grid",GROUP="oinstall",MODE="0660",SYMLINK+="ASM0300"
KERNEL=="dasd*1",ID=="0.0.0305",OWNER="grid",GROUP="oinstall",MODE="0660",SYMLINK+="ASM0305"
```

Make an entry for each device you plan to use with Oracle ASM.

From Oracle we can then work with the new ASM Disk Device:

ALTER DISKGROUP DG2 add disk '/dev/ASM0305';
ALTER DISKGROUP DG2 rebalance power 2;

Current Hot Topics with Oracle 11gR2 on System z

Current Hot Topics

- **New 11gR2 Oracle VKTM process (Virtual Time Keeper)**
  - **VKTM** is responsible for providing a wall-clock time (updated every second) and reference-time counter (updated every 20 ms) **even when the database is idle for a long time (CPU Idle).** The VKTM timer service centralizes time tracking and offloads multiple timer calls from other clients.
  - **_disable_highres_ticks='true'** # disable high-res tick
    _timer_precision=2000 #VKTM timer precision in ms

    ****** Work with Oracle support to get approval to use in heavy memory 11gR2 over-commit environments.**
  - VM Q3 (which means it will never be swap out to release all it's memory). Have observed if we stop the database the Linux machine goes to Q1 (or Q2) releasing memory. Restart the database, the machine goes back to Q3.

Current Hot Topics

- ORA-600[KFDADD03] WHEN CREATING A DISKGROUP USING FCP/SCSI STORAGE
  - Bug 12346221 when creating ASM disk group
  - See note for long term ASMLib direction - Oracle ASMLib Software Update Policy for Red Hat Enterprise Linux Supported by Red Hat [ID 1089399.1] – no plans for RH6
  - Recommendation use UDEV rules opposed to ASMLib.

- When NLS_LANG and LANG values are set to different character set, DBCA can't be launched.
  - In Japan, DBCA related processes can't be terminated normally. Sending Ctrl-C doesn't work and processes remained as zombie. – open SR
  - No problems with French in Quebec or in Latin America – Mexico
  - Recommendation -> don’t set any Oracle ENVIRONMENT variables when installing per the release notes.

Oracle Automatic Memory – MEMORY_TARGET

- New memory management parameter **MEMORY_TARGET**(AMM – Automatic Memory management)
- Combines ASMM (Automatic Shared Memory Management) parameters SGA_TARGET and PGA_AGGREGATE_TARGET into one parameter.
- If you set **MEMORY_TARGET** too large …

  ORA-00845: MEMORY_TARGET not supported on this system

  The Oracle alert log shows:

  WARNING: You are trying to use the MEMORY_TARGET feature. This feature requires the /dev/shm file system to be mounted for at least 847249408 bytes.

- The error is really that the **MEMORY_TARGET** needs a larger **/dev/shm**

Run the following to resize tmpfs:

```
# umount tmpfs
# mount -t tmpfs shmfs -o size=1300m /dev/shm
# df -k /dev/shm
Filesystem           1K-blocks      Used Available Use% Mounted on
shmfs                1331200           0   1331200   0%  /dev/shm
```

***** make permanent in the /etc/fstab file.**

Oracle asmcmd Error:

- **ASMCMD is a command line interface that allows the DBA to look at Disk usage and files on raw disk volumes**

- **Some Systems may see an error when running the Oracle asmcmd command**

$ asmcmd

Can't load '/u01/grid/11.2/perl/lib/site_perl/5.10.0/s390x-linux-thread-multi/auto/XML/Parser/Expat/Expat.so' for module XML::Parser::Expat: libexpat.so.0: cannot open shared object file: No such file or directory at /u01/grid/11.2/perl/lib/5.10.0/s390x-linux-thread-multi/DynaLoader.pm line 203.

at /u01/grid/11.2/perl/lib/site_perl/5.10.0/s390x-linux-thread-multi/XML/Parser.pm line 14

Compilation failed in require at /u01/grid/11.2/perl/lib/site_perl/5.10.0/s390x-linux-thread-multi/XML/Parser.pm BEGIN failed--compilation aborted at /Compilation failed in require at /u01/grid/11.2/lib/asmcmddisk.pm line 133.

BEGIN failed--compilation aborted at /u01/grid/11.2/lib/asmcmddisk.pm line 133.

Compilation failed in require at /u01/grid/11.2/bin/asmcmdcore line 186. grid@cnsiorap:/home/grid> asmcmd

Can't load '/u01/grid/11.2/perl/lib/site_perl/5.10.0/s390x-linux-thread-multi/auto/XML/Parser/Expat/Expat.so' for module XML::Parser::Expat: libexpat.so.0: cannot open shared object file: No such file or directory at /u01/grid/11.2/perl/lib/5.10.0/s390x-linux-thread-multi/DynaLoader.pm line 203.

at /u01/grid/11.2/perl/lib/site_perl/5.10.0/s390x-linux-thread-multi/XML/Parser.pm line 14

Compilation failed in require at /u01/grid/11.2/perl/lib/site_perl/5.10.0/s390x-linux-thread-multi/XML/Parser.pm BEGIN failed--compilation aborted at /u01/grid/11.2/perl/lib/site_perl/5.10.0/s390x-linux-thread-multi/XML/Parser.pm line 18.

Compilation failed in require at /u01/grid/11.2/lib/asmcmddisk.pm line 133.

BEGIN failed--compilation aborted at /u01/grid/11.2/lib/asmcmddisk.pm line 133.

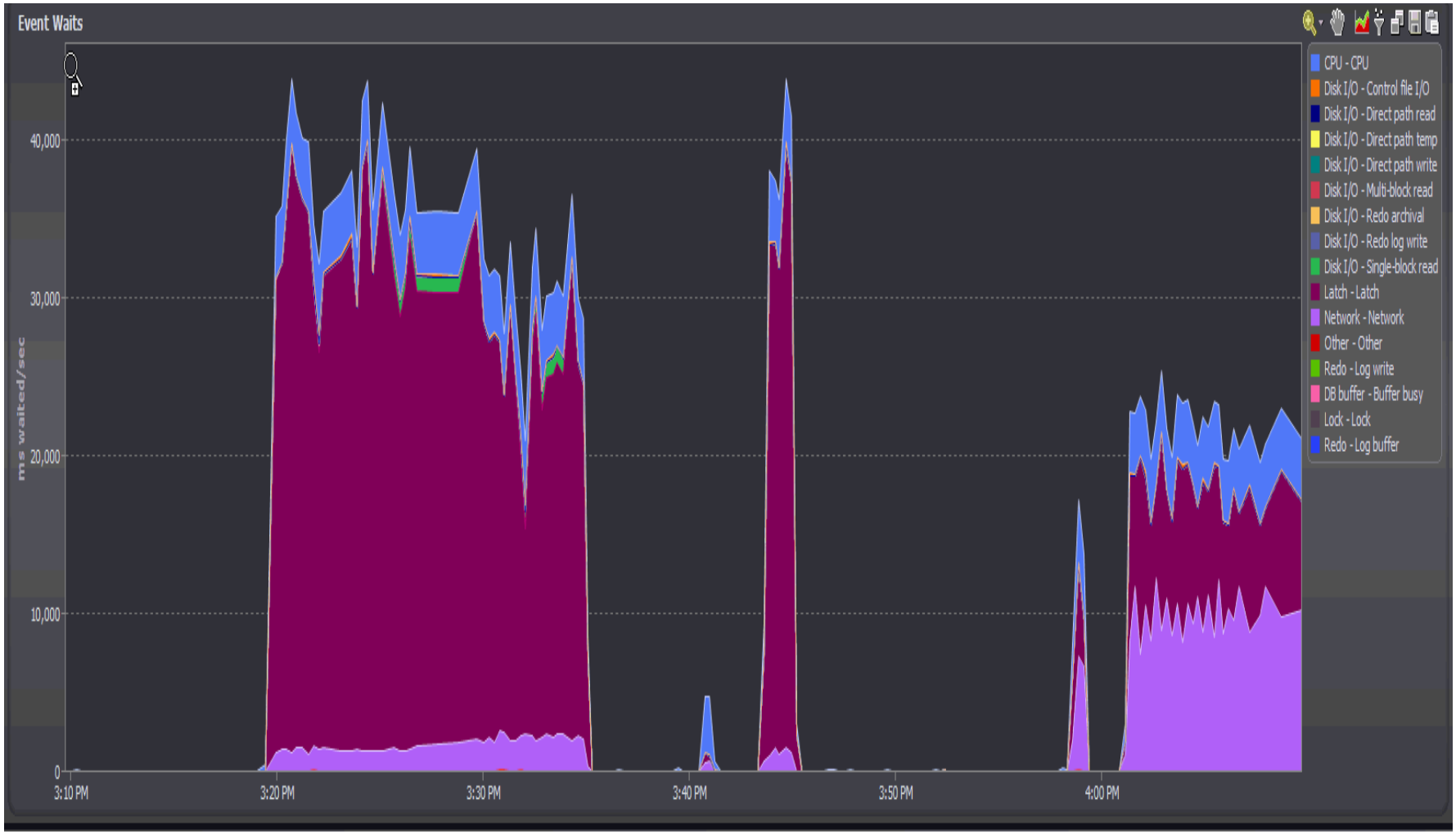
ASMCMD Error How to resolve

- [Cause] ASMCMD command calls libexpat.so.0 internally. With SLES 11 SP1(s390x) libexpat.so.0 is renamed to libexpat.so.1 (Also occurred with SLES 10 SP3 system)

- [Solution] This problem has been reported in an Oracle Bug. Workaround is to create a symbolic link:

```
cd /oracle/app/11.2.0/grid/lib  ($GRID_HOME/lib)
ln -s libexpat.so.1 libexpat.so.0
```

10gR2 High CPU, Latches –Shared Connections



Oracle 11gR2 – New Mutex locking

1) ORA-00600: internal error code, arguments: [kkspsc0: basehd]
applied patch

2) ORA-00600: internal error code, arguments: [kglLockOwnersListAppend-ovf]
applied patch

3) cursor: mutex S and library cache lock
1. Download and apply the 11.2.0.2.2 PSU Patch 11724916

2. Enable event 106001 to address Bug 10187168.

To enable the fix "**_cursor_features_enabled**" needs to be set to a value that depends on the patch level. Please note that the value for ***_cursor_features_enabled*** is different for each version

4) resmgr cpu:quantum wait event when not cpu bound
Advisory DEFAULT_MAINTENANCE_PLAN (Doc ID 786346.1)
- we disabled this and that helped

5) **Oracle 11.2.0.2 PSU (Patch Set Update)** includes a slew of parmeters that you can tweak based on workload charcteristics.
Note: **10411618 - Enhancement to add different "Mutex" wait schemes [ID 10411618.8]**

Do Not use NOARP for Oracle Grid Installs

- Oracle Grid Install when the network interfaces are set with NOARP you can encounter **BUG – 10173295** when running the root.sh script on the first node.

  Error:

  Did not succssfully configure and start ASM at /opt/oracle/11gR2/crs/install/crsconfig_lib.pm line 6470.

  /opt/oracle/11gR2/perl/bin/perl -I/opt/oracle/11gR2/perl/lib -I/opt/oracle/11gR2/crs/install/opt/oracle/11gR2/crs/install/rootcrs.pl execution failed

  CRS-1013:The OCR location in an ASM disk group is inaccessible. Details in /opt/oracle/11gR2/crs/log/dhsora1/client/clscfg.log

  Oracle Database 11g Clusterware Release 11.2.0.2.0 - Production Copyright 1996, 2010 Oracle. All rights reserved.

  2011-03-16 20:01:53.085: [CLSCFG][53553008]clscfg_main: Configuration type [4]

  ibctx: Failed to read the whole bootblock.

- Update the network interfaces to have ARP enabled (the following is incorrect)

  ifconfig –a

  eth0 Link encap:Ethernet HWaddr 02:00:02:00:00:A2
  inet addr:130.35.55.234 Bcast:130.35.55.255 Mask:255.255.252.0
  inet6 addr: fe80::200:200:100:a2/64 Scope:Link
  UP BROADCAST RUNNING **NOARP** MULTICAST MTU:1492 Metric:1
  RX packets:5749678 errors:0 dropped:0 overruns:0 frame:0
  TX packets:2799431 errors:0 dropped:0 overruns:0 carrier:0
  collisions:0 txqueuelen:1000
  RX bytes:1414260847 (1.3 GiB) TX bytes:2735238017 (2.5 GiB)

New Features To Consider for 11gR2

Oracle RMAN Backup Compression

Backup Compression	Backup Time	Compression Size Source DB - 1.29 GB	% Compression / Input MB/s
'Basic' 10gR2 (**BZIP2**) Compression	02:48 (168 s)	278.95 MB	78.9 % 7.89 MB/s
'High' 11gR2 (**BZIP2**) Compression	08:41 (521 s)	224.82 MB	83.0 % 2.54 MB/s
'Medium' (**ZLIB**) Compression	01:08 (68 s)	295.53 MB	77.6 % 19.46 MB/s
'Low' (**LZO**) Compression	00:28 (28 s)	357.03 MB	73.0 % 47.26 MB/s

- **RMAN Command -> CONFIGURE COMPRESSION ALGORITHM 'Low'**
- **Oracle Advanced Compression Feature required for Low, Medium, High**
- **Very High CPU observed with BZIP2**
- **Secure File LOBs can utilize this compression Technology**

Oracle HugePages Configuration:

- SLES 10 SP3+ (2mb), Red Hat 5 (2mb), SLES 11 SP 1(1 mb)
- Calculate nr_hugepages using script from MOS Note **401749.1**
  Then set kernel parameter:

  # **sysctl -w vm.nr_hugepages=<**value from above> …. then
  # **sysctl –p** (to load)

- Set the oracle memlock limit to be as the size of the Hugepages:
  - Set value (in KB) slightly smaller than Linux Guest size (No harm setting to greater than Oracle SGA requirements)

  cat /**etc/security/limits.conf** | grep memlock

  oracle soft memlock 3436560
  oracle hard memlock 3436560

- Set Oracle parameter - **use_large_pages="only"** to ensure instance will always start with large pages

Oracle HugePages – small 4K Page Example

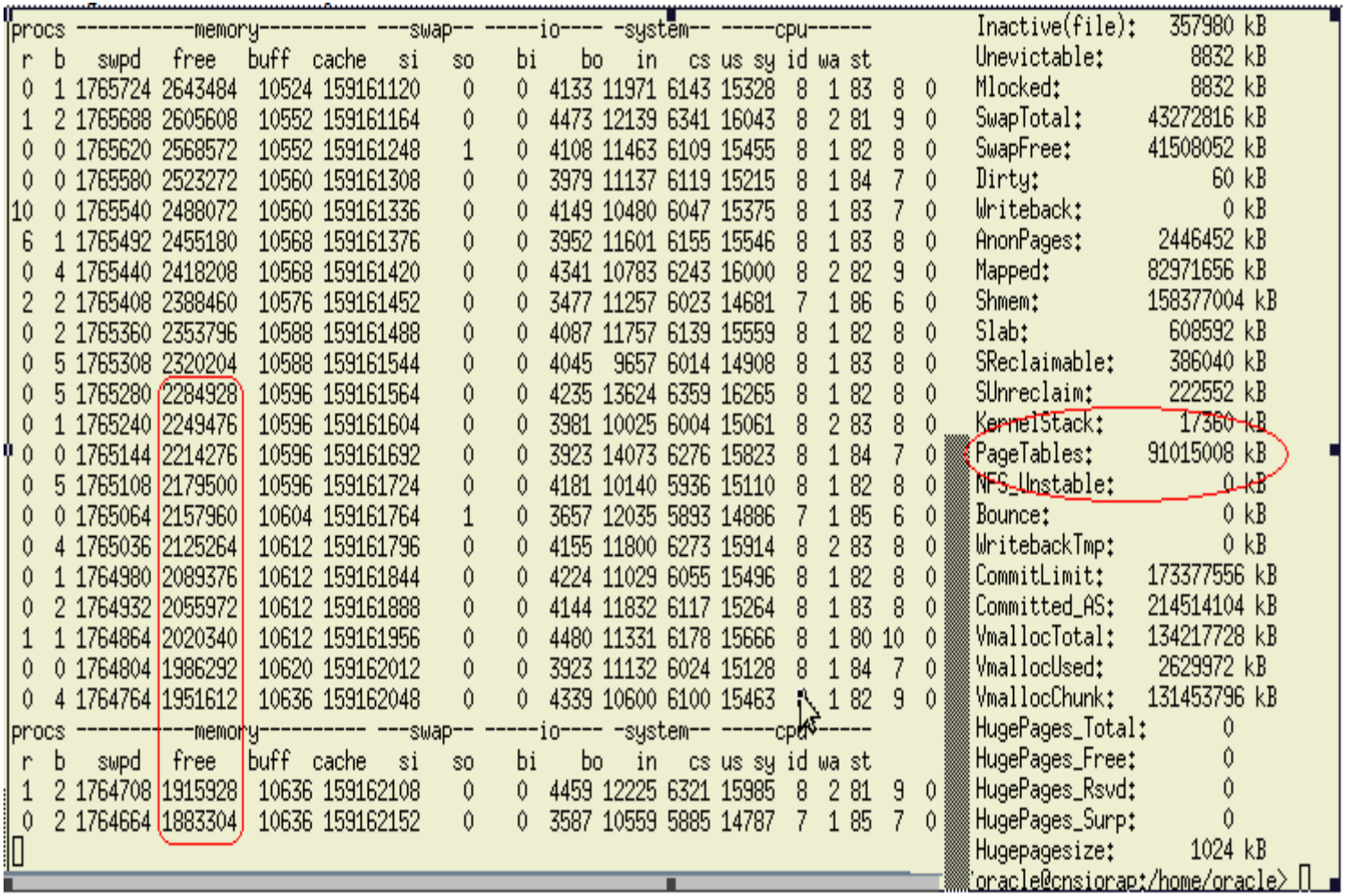
Starts out fine – 485 TPS and 58ms response time

Linux Page Tables at 27 GB

```
procs -----------memory---------- ---swap-- -----io---- -system-- -----cpu------
 r  b   swpd   free   buff  cache   si   so    bi    bo   in   cs us sy id wa st
 0  5      0 186450688  25684 45113972    0    0   746    68  166  355  1  1 95  3  0
 2  7      0 186108944  25684 45346640    0    0 76304  9432 6961 17518  9  3 63 25  0
 0  5      0 185801168  25700 45549616    0    0 67533 10097 6654 16722  8  2 66 24  0
 0  9      0 185453048  25700 45782904    0    0 76757 10541 7208 18075  9  3 62 26  0
 0  2      0 185124276  25700 46000816    0    0 71771  9892 6829 17447  9  3 61 27  0
 0  7      0 184783848  25708 46227136    0    0 74709 10148 6978 17523  9  3 64 25  0
 0 10      0 184441684  25716 46449348    0    0 73587  9983 7009 17791  9  3 61 27  0
```

```
oracle@cnsiorap:/home/oracle> cat /proc/meminfo
MemTotal:       260209484 kB
MemFree:        167199168 kB
Buffers:            25952 kB
Cached:         57855304 kB
SwapCached:             0 kB
Active:         60408136 kB
Inactive:         502528 kB
Active(anon):   60393244 kB
Inactive(anon):         0 kB
Active(file):     14892 kB
Inactive(file):   502528 kB
Unevictable:        7808 kB
Mlocked:            7808 kB
SwapTotal:      43272816 kB
SwapFree:       43272816 kB
Dirty:                104 kB
Writeback:              0 kB
AnonPages:        3037540 kB
Mapped:         55001156 kB
Shmem:          57360472 kB
Slab:             375452 kB
SReclaimable:     153232 kB
SUnreclaim:       222220 kB
KernelStack:       17296 kB
PageTables:     28315696 kB
```

After an Hour….



```
procs -----------memory---------- ---swap-- -----io---- -system-- -----cpu------
 r  b   swpd   free   buff  cache   si   so    bi    bo   in   cs us sy id wa st
 0  1 1765724 2643484  10524 159161120    0    0  4133 11971 6143 15328  8  1 83  8  0
 1  2 1765688 2605608  10552 159161164    0    0  4473 12139 6341 16043  8  2 81  9  0
 0  0 1765620 2568572  10552 159161248    1    0  4108 11463 6109 15455  8  1 82  8  0
 0  0 1765580 2523272  10560 159161308    0    0  3979 11137 6119 15215  8  1 84  7  0
10  0 1765540 2488072  10560 159161336    0    0  4149 10480 6047 15375  8  1 83  7  0
 6  1 1765492 2455180  10568 159161376    0    0  3952 11601 6155 15546  8  1 83  8  0
 0  4 1765440 2418208  10568 159161420    0    0  4341 10783 6243 16000  8  2 82  9  0
 2  2 1765408 2388460  10576 159161452    0    0  3477 11257 6023 14681  7  1 86  6  0
 0  2 1765360 2353796  10588 159161488    0    0  4087 11757 6139 15559  8  1 82  8  0
 0  5 1765308 2320204  10588 159161544    0    0  4045  9657 6014 14908  8  1 83  8  0
 0  5 1765280 2284928  10596 159161564    0    0  4235 13624 6359 16265  8  1 82  8  0
 0  1 1765240 2249476  10596 159161604    0    0  3981 10025 6004 15061  8  2 83  8  0
 0  0 1765144 2214276  10596 159161692    0    0  3923 14073 6276 15823  8  1 84  7  0
 0  5 1765108 2179500  10596 159161724    0    0  4181 10140 5936 15110  8  1 82  8  0
 0  0 1765064 2157960  10604 159161764    1    0  3657 12035 5893 14886  7  1 85  6  0
 0  4 1765036 2125264  10612 159161796    0    0  4155 11800 6273 15914  8  2 83  8  0
 0  1 1764980 2089376  10612 159161844    0    0  4224 11029 6055 15496  8  1 82  8  0
 0  2 1764932 2055972  10612 159161888    0    0  4144 11832 6117 15264  8  1 83  8  0
 1  1 1764864 2020340  10612 159161956    0    0  4480 11331 6178 15666  8  1 80 10  0
 0  0 1764804 1986292  10620 159162012    0    0  3923 11132 6024 15128  8  1 84  7  0
 0  4 1764764 1951612  10636 159162048    0    0  4339 10600 6100 15463  8  1 82  9  0
procs -----------memory---------- ---swap-- -----io---- -system-- -----cpu------
 r  b   swpd   free   buff  cache   si   so    bi    bo   in   cs us sy id wa st
 1  2 1764708 1915928  10636 159162108    0    0  4459 12225 6321 15985  8  2 81  9  0
 0  2 1764664 1883304  10636 159162152    0    0  3587 10559 5885 14787  7  1 85  7  0
```

```
Inactive(file):   357980 kB
Unevictable:        8832 kB
Mlocked:            8832 kB
SwapTotal:      43272816 kB
SwapFree:       41508052 kB
Dirty:                60 kB
Writeback:             0 kB
AnonPages:       2446452 kB
Mapped:         82971656 kB
Shmem:         158377004 kB
Slab:             608592 kB
SReclaimable:     386040 kB
SUnreclaim:       222552 kB
KernelStack:       17360 kB
PageTables:     91015008 kB
NFS_Unstable:          0 kB
Bounce:                0 kB
WritebackTmp:          0 kB
CommitLimit:    173377556 kB
Committed_AS:   214514104 kB
VmallocTotal:   134217728 kB
VmallocUsed:      2629972 kB
VmallocChunk:   131453796 kB
HugePages_Total:       0
HugePages_Free:        0
HugePages_Rsvd:        0
HugePages_Surp:        0
Hugepagesize:       1024 kB
oracle@cnsiorap:/home/oracle>
```

4K Page Tables after 70 minutes

Linux Swap and PageTables using 87.7 GB of Memory!

```
procs -----------memory---------- ---swap-- -----io---- -system-- -----cpu------
 r  b   swpd   free   buff  cache   si   so    bi    bo   in   cs us sy id wa st
338  8 1766820 1096980   1200 158901132    1  467 11419   721 2140 2724  1 93  0  0  7
125 13 1767088 1096700   1316 158896948    8  135  7199  1092 2227 4262  2 91  0  0  7
420  4 1767396 1073704   1416 158891792   17  137 18407 25048 5875 11215  6 80  4  5  1
302  5 1767588 1089200   1424 158876220    3  172  1256   329 1705 1483  0 93  0  0  6
227  7 1767652 1088700   1448 158870652    9   97  4889   361 1987 1926  1 92  0  0  7
165 16 1767796 1093696   1444 158858216    0  129  3617   605 2205 2874  2 91  0  0  7
452 16 1768980 1074352   1480 158858772   35  453 11801 14244 4667 8128  5 85  2  2  6
257 14 1769204 1096292   1276 158828368    5   84  1320   505 2066 2657  2 91  0  0  7
177  6 1769172 1098028   1320 158821092    0   20  1647   447 1761 1984  2 91  0  0  7
217 16 1769600 1095124   1364 158816144   19  224  2167  1055 2029 2703  2 91  0  0  7
144 17 1770068 1088160   1256 158814320   12  239  1760   659 1884 2295  2 91  0  0  7
122 11 1771576 1082412   1276 158810608   11  561  1817   868 1862 2049  2 92  0  0  7
219 10 1772768 1073684   1260 158807908   29  408  2385   863 2200 2916  2 91  0  0  7
315  3 2033292 1076748   1152 158561024  100 86901 21179 87940 45540 33283  0 93  0  0

SReclaimable:     386028 kB
SUnreclaim:       222484 kB
KernelStack:       16880 kB
PageTables:     91964268 kB
NFS_Unstable:          0 kB
Bounce:                0 kB
WritebackTmp:          0 kB
CommitLimit:   173377556 kB
Committed_AS:  214527304 kB
VmallocTotal:  134217728 kB
VmallocUsed:     2629972 kB
VmallocChunk:  131453796 kB
HugePages_Total:       0
HugePages_Free:        0
HugePages_Rsvd:        0
HugePages_Surp:        0
Hugepagesize:       1024 kB
oracle@cnsiorap:/home/oracle>
```

A little while Later….

74 ms response time and 0 TPS

Configuring HugePages for Oracle

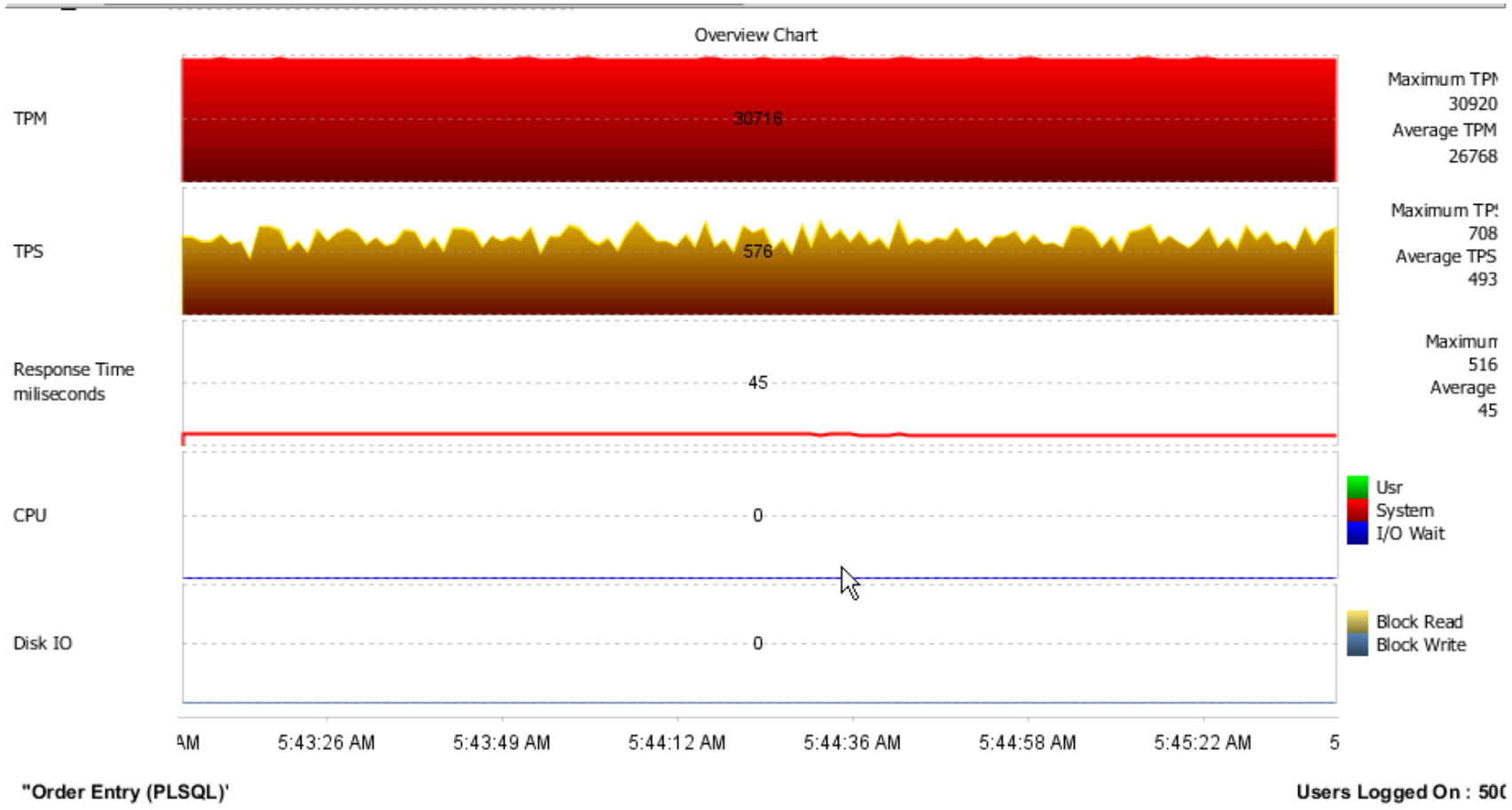
```
MemTotal:       260209484 kB
MemFree:        10667956 kB
Buffers:           52416 kB
Cached:          3909240 kB
SwapCached:            0 kB
Active:          3350056 kB
Inactive:        3008640 kB
Active(anon):    3324948 kB
Inactive(anon):        0 kB
Active(file):      25108 kB
Inactive(file):  3008640 kB
Unevictable:        7872 kB
Mlocked:            7872 kB
SwapTotal:      43272816 kB
SwapFree:       43272816 kB
Dirty:               168 kB
Writeback:             0 kB
AnonPages:       3030216 kB
Mapped:           128528 kB
Shmem:            299404 kB
Slab:             252440 kB
SReclaimable:      30804 kB
SUnreclaim:       221636 kB
KernelStack:       17184 kB
PageTables:       383412 kB
NFS_Unstable:          0 kB
Bounce:                0 kB
WritebackTmp:          0 kB
CommitLimit:    53774356 kB
Committed_AS:    5645992 kB
VmallocTotal:   134217728 kB
VmallocUsed:      2629972 kB
VmallocChunk:   131433316 kB
HugePages_Total:   233600
HugePages_Free:     77314
HugePages_Rsvd:     48003
HugePages_Surp:         0
Hugepagesize:        1024 kB
```

```
 0  2      0 10915880  52240 3939944    0    0  4309  2339 6287 15651  8  1 83  7  0
 0  2      0 10898188  52240 3939944    0    0  4197  2227 5992 14832  8  1 83  7  0
 0  0      0 10880420  52248 3939944    0    0  4269  2199 6094 15135  8  1 83  7  0
 1  0      0 10868204  52248 3939944    0    0  4099  2249 5814 14416  8  1 84  7  0
 0  0      0 10857216  52256 3939944    0    0  4379  2267 5910 14829  9  1 83  8  0
 0  0      0 10850168  52256 3939944    0    0  3859  2111 5474 13579  8  1 85  6  0
 0  2      0 10840936  52264 3939944    0    0  4232  2376 5884 14668  9  1 82  8  0
 7  1      0 10830716  52264 3939944    0    0  3984  2268 5654 14243  9  1 83  7  0
 5  1      0 10824900  52272 3939944    0    0  3696  1815 5358 12752  7  1 86  6  0
 2  4      0 10817872  52280 3939944    0    0  4341  2327 5592 14119  9  1 83  7  0
 0  1      0 10804156  52280 3939944    0    0  4203  2225 5565 13930  8  1 83  7  0
 0  1      0 10795620  52280 3939976    0    0  4352  2193 5569 13812  8  1 83  7  0
procs -----------memory---------- ---swap-- -----io---- -system-- -----cpu------
 r  b   swpd   free   buff  cache   si   so    bi    bo   in   cs us sy id wa st
 2  0      0 10785656  52284 3939988    0    0  4557  2335 5637 14175  9  1 82  8  0
 0  1      0 10779592  52284 3939984    0    0  3867  2095 5441 13446  8  1 85  7  0
 0  1      0 10774076  52300 3939984    0    0  4197  2287 5608 14003  9  1 83  7  0
 0  1      0 10770368  52300 3939988    0    0  4200  2151 5431 13439  8  1 84  7  0
```

Page Tables are now at 0.365 GB vs 88.1 GB before!!!

Same Test with 1MB Oracle HugePages

1 MB Huge Pages 576 vs 4K Pages 485 TPS and 45ms vs 58 ms Response
(When running Good)



Two Hours later still Running strong…

```
[oracle@cnsiorap:/home/oracle>
MemTotal:       260209484 kB
MemFree:         10453276 kB
Buffers:            60092 kB
Cached:           3911428 kB
SwapCached:             0 kB
Active:           3540228 kB
Inactive:         3016924 kB
Active(anon):     3513552 kB
Inactive(anon):         0 kB
Active(file):       26676 kB
Inactive(file):   3016924 kB
Unevictable:         7872 kB
Mlocked:             7872 kB
SwapTotal:       43272816 kB
SwapFree:        43272816 kB
Dirty:                 20 kB
Writeback:              0 kB
AnonPages:        3218740 kB
Mapped:            129036 kB
Shmem:             299404 kB
Slab:              254420 kB
SReclaimable:       31316 kB
SUnreclaim:        223104 kB
KernelStack:        17360 kB
PageTables:        389768 kB
NFS_Unstable:           0 kB
Bounce:                 0 kB
WritebackTmp:           0 kB
CommitLimit:     53774356 kB
Committed_AS:     5860460 kB
VmallocTotal:   134217728 kB
VmallocUsed:       2629972 kB
VmallocChunk:   131433316 kB
HugePages_Total:   233600
HugePages_Free:     77309
HugePages_Rsvd:     47998
HugePages_Surp:         0
Hugepagesize:        1024 kB
```

```
 0  1      0 10458852  59956 3942708    0    0  3971 14633 6235 15858  9  1 80 10  0
 0  5      0 10458228  59964 3942700    0    0  4139 11295 6041 15235  9  1 80 10  0
 1  4      0 10458132  59964 3942712    0    0  3765 14725 6246 15765  8  1 81  9  0
 0  2      0 10457500  59972 3942704    0    0  3957 10909 5922 15024  9  1 80 10  0
 0  3      0 10457416  59972 3942712    0    0  3491 15451 6199 15713  8  1 82  8  0
 1  1      0 10457376  59980 3942712    0    0  3931  9919 5842 14763  9  1 81  9  0
 0  3      0 10448640  59980 3942728    0    0  3525 14445 6229 15595  8  1 82  9  0
 1  2      0 10449476  59980 3942744    0    0  4037 10345 5895 15057  9  1 80 10  0
 0  3      0 10448076  59988 3942736    0    0  3667 15081 6271 15912  9  1 81  9  0
 1  1      0 10451036  59988 3942728    0    0  3715 10472 5935 14777  8  1 82  8  0
 0  2      0 10451892  59988 3942728    0    0  3915 14064 6265 15958  9  1 80 10  0
 0  1      0 10450884  60004 3942720    0    0  3256 12665 5917 14908  8  1 84  7  0
 8  1      0 10449416  60004 3942724    0    0  4080 11723 6165 15571  9  1 80 10  0
15  0      0 10451084  60012 3942720    0    0  3715 10644 5862 15008  8  1 82  8  0
 5  3      0 10450120  60012 3942724    0    0  3664 14157 6289 15810  8  1 82  8  0
 1  2      0 10449976  60012 3942724    0    0  4152 10243 5955 15021  9  1 80 10  0
 1  5      0 10456096  60020 3942724    0    0  3333 13889 6190 15468  8  1 82  8  0
 2  4      0 10458024  60020 3942720    0    0  4176  9695 5917 14960  9  1 79 11  0
 0  2      0 10457412  60028 3942716    0    0  3661 14596 6273 15871  8  1 81  9  0
procs -----------memory---------- ---swap-- -----io---- -system-- -----cpu------
 r  b   swpd   free   buff  cache   si   so    bi    bo   in   cs us sy id wa st
 0  3      0 10458256  60028 3942720    0    0  3965 10697 5980 15029  9  1 81  8  0
 0  5      0 10458264  60028 3942720    0    0  3704 14120 6231 15743  8  1 82  9  0
 1  3      0 10459016  60036 3942708    0    0  3880 10425 5931 14894  9  1 81  9  0
```

Page tables: 0.371 GB and No Swap

HugePages for Large DB's with Many Connections

1 MB HugePages 510 TPS vs 4K Pages 488 TPS and 37ms vs 74 ms Response

HugePage Considerations

- Can not use Oracle Automatic Memory Management with Huge Pages. Set memory regions manually (**db_cache_size**, **shared_pool_size**)
- Not swappable: Huge Pages are not swappable. Therefore there is no page-in/page-out mechanism. Huge Pages are universally regarded as pinned.
- General guideline consider when combined Oracle SGA's are greater than **8 GB** (particularly if a lots of connections)
- Decreased page table overhead; more memory can be freed up for other uses. For example more Oracle SGA memory, and less physical I/O's (See also Document **361468.1**)
- Cat /proc/cpuinfo look for the "edat" feature to see if HW large page support is enabled as well.

Oracle 11gR2 new features

- RAC
  - ASM and clusterware consolidated
    - Grid infrastructure installation
  - OCR and Voting disks can be now in ASM
    - Auto backup of voting disk into OCR
  - Enhanced Cluster Verification Utility
    - Simplified installation
  - Enhanced RAC de-install utility
  - No more reboot of the nodes
  - **SCAN** (**S**ingle **C**lient **A**ccess **N**ame)
    - During cluster installation, SCAN is configured which is a domain / host name and resolves up to three ip addresses
    - A SCAN listener is created for each of the SCAN ip addresses
    - SCAN listeners provide the load balancing
    - client uses SCAN name to connect, no need to specify vip name
    - when new node is added, no need to edit tnsnames.ora

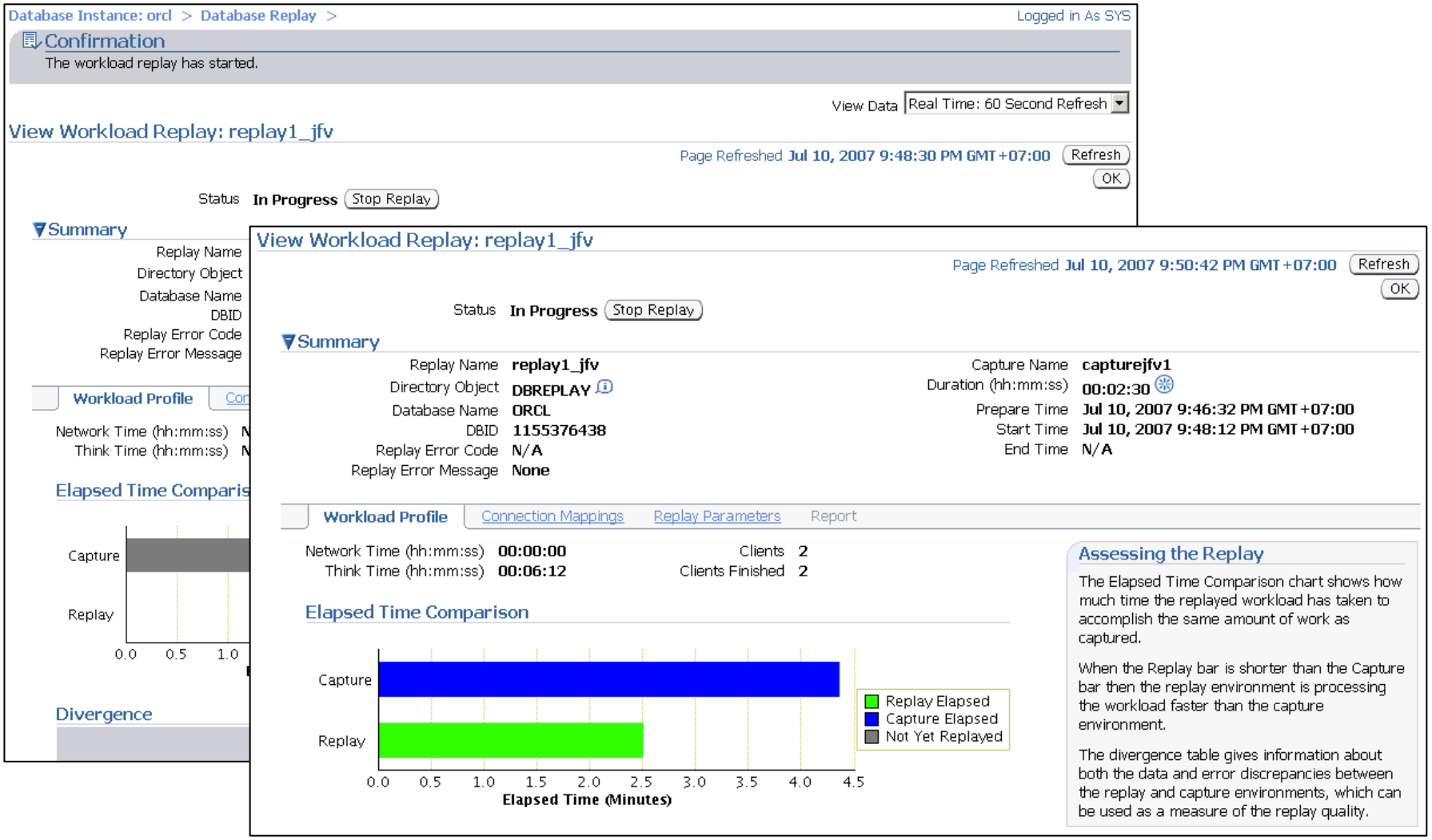
Oracle Database Replay

- Changes to system environments are a common occurrence:
  - Database upgrades
  - OS upgrades / changes
  - Platform changes
  - Storage changes (ECKD to FCP)
  - Single instance to RAC
  - Filesystem to ASM
  - DB configuration parameter changes
- In the past, realistic testing of Production workload is time consuming and rarely simulates production.

Database Replay

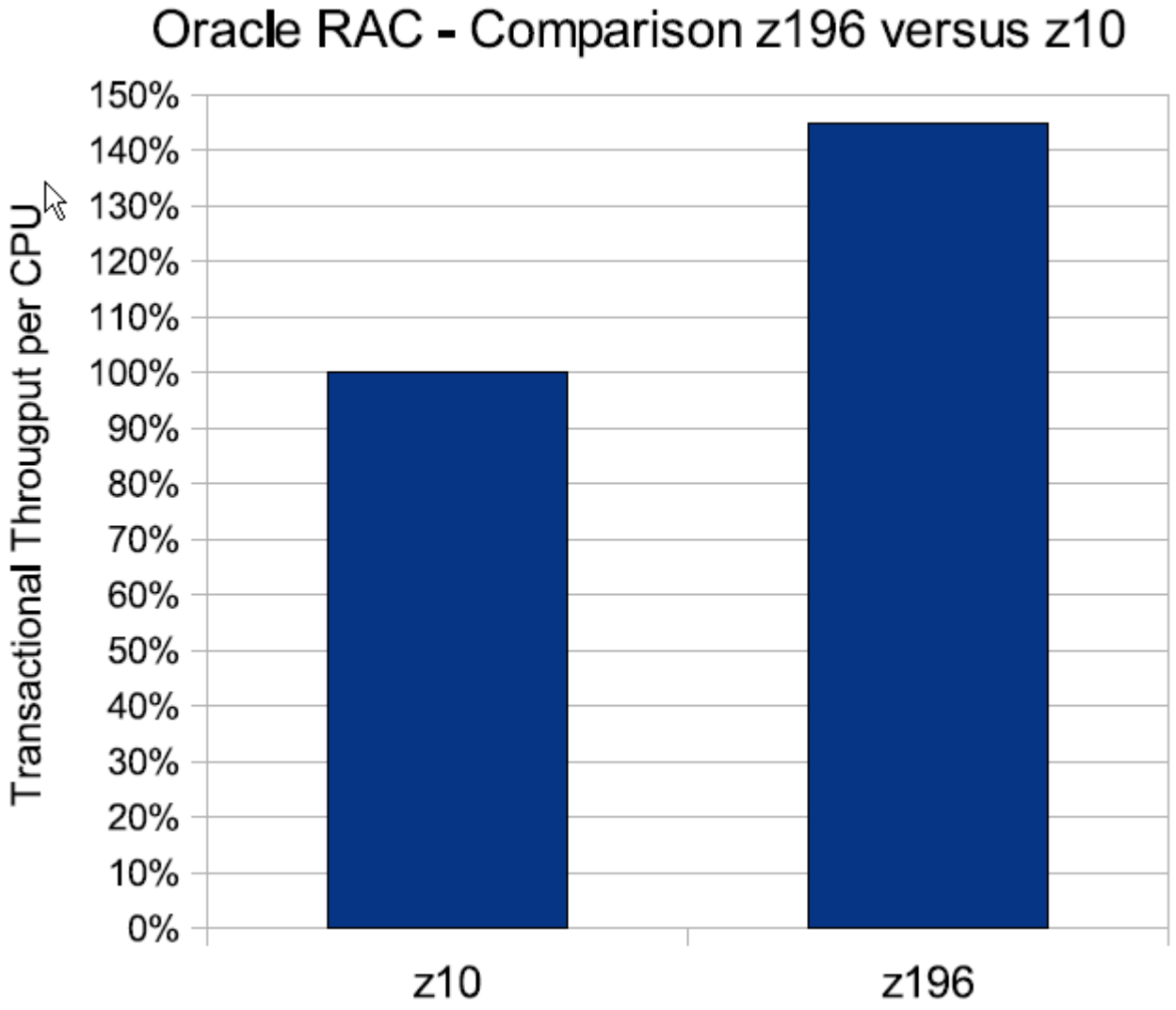
- Re-create actual production database workload in a test environment.
- Identify and analyze potential instabilities before making changes to production.
- Capture workload in production:
  - Capture full production workload with real load & concurrency
  - Move the captured workload to test system
- Replay workload in test:
  - Make the desired changes in test system
  - Replay workload with production load & concurrency
  - Honor commit ordering
- Analyze and report:
  - Errors
  - Data divergence
  - Performance divergence

Viewing Workload Replay Statistics



Customer Experiences 11gR2 with Linux on System z

Oracle 10gR2 & zEnterprise Performance



Source: http://public.dhe.ibm.com/software/dw/linux390/perf/ZSW03185-USEN-02.PDF

Starting a POC with Oracle on System z Linux

- **System Sizing**
  - Sizing Virtual Memory, Number of Virtuals CPs/IFLs to Physicals
  - Network requirements (Linux network parameters)
- **Disk Options**
  - HyperPAV, PAV, FCP/SCSI, LVM with striping
  - Oracle Orion
- **OS requirements**
  - DASD, Swap, FCP/SCSI
  - Use the rpm checker
  - ulimits, system timer (no more hang check timer)
- **Installing Oracle**
- **Loading the database**
  - Transportable Database, Data Pump, Migration Factory
- **Generating a test load**
  - Database Replay
- **Monitoring**
  - Enterprise manager, ADDM, ASH reports for Oracle
  - Linux vmstat, sar, nmon (steal, swap, run queue), iostat
  - Velocity, Performance Tool Kit


Thank
You