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)

Oracle Database 11g 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)

• 11.2.0.2.3 PSU (Grid - 12419353)
Automated Storage Management (ASM)

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

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...

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**) +

```
# 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)

```bash
# 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

```bash
multipath {
    wwid 3600507630bffc2ce00000000000001112
    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*",ID=="0.0.0300",OWNER="grid",GROUP="oinstall",MODE="0660",SYMLINK+="ASM0300"
KERNEL=="dasd*",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:

```bash
# umount tmpfs
# mount -t tmpfs shmfs -o size=1300m /dev/shm
# df -k /dev/shm
```

<table>
<thead>
<tr>
<th>Filesystem</th>
<th>1K-blocks</th>
<th>Used</th>
<th>Available</th>
<th>Use%</th>
<th>Mounted on</th>
</tr>
</thead>
<tbody>
<tr>
<td>shmfs</td>
<td>1331200</td>
<td>0</td>
<td>1331200</td>
<td>0%</td>
<td>/dev/shm</td>
</tr>
</tbody>
</table>

*** make permanent in the /etc/fstab file.***
Oracle asmcmd Error:

- ASMCD 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 /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 parameters that you can tweak
   based on workload characteristics.
   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 successfully 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 -l/opt/oracle/11gR2/perl/lib -l/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.
  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 txqueue len:1000 
    RX bytes:1414260847 (1.3 GiB) TX bytes:2735238017 (2.5 GiB)
New Features To Consider for 11gR2
Oracle RMAN Backup Compression

<table>
<thead>
<tr>
<th>Backup Compression</th>
<th>Backup Time</th>
<th>Compression Size</th>
<th>% Compression / Input MB/s</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘Basic’ 10gR2</td>
<td>02:48</td>
<td>278.95 MB</td>
<td>78.9 %</td>
</tr>
<tr>
<td>(BZIP2) Compression</td>
<td>(168 s)</td>
<td></td>
<td>7.89 MB/s</td>
</tr>
<tr>
<td>‘High’ 11gR2</td>
<td>08:41</td>
<td>224.82 MB</td>
<td>83.0 %</td>
</tr>
<tr>
<td>(BZIP2) Compression</td>
<td>(521 s)</td>
<td></td>
<td>2.54 MB/s</td>
</tr>
<tr>
<td>‘Medium’</td>
<td>01:08</td>
<td>295.53 MB</td>
<td>77.6 %</td>
</tr>
<tr>
<td>(ZLIB) Compression</td>
<td>(68 s)</td>
<td></td>
<td>19.46 MB/s</td>
</tr>
<tr>
<td>‘Low’</td>
<td>00:28</td>
<td>357.03 MB</td>
<td>73.0 %</td>
</tr>
<tr>
<td>(LZO) Compression</td>
<td>(28 s)</td>
<td></td>
<td>47.26 MB/s</td>
</tr>
</tbody>
</table>

- 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

<table>
<thead>
<tr>
<th>Proc</th>
<th>Memory</th>
<th>Swap</th>
<th>I/O</th>
<th>System</th>
<th>CPU</th>
</tr>
</thead>
<tbody>
<tr>
<td>r</td>
<td>b</td>
<td>swpd</td>
<td>free</td>
<td>buff</td>
<td>cache</td>
</tr>
<tr>
<td>0</td>
<td>5</td>
<td>186450688</td>
<td>25684</td>
<td>45113972</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>186108944</td>
<td>25684</td>
<td>45346640</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>5</td>
<td>185801168</td>
<td>25700</td>
<td>45549616</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>9</td>
<td>185453048</td>
<td>25700</td>
<td>45782904</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>185124276</td>
<td>25700</td>
<td>46000816</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>7</td>
<td>184783848</td>
<td>25708</td>
<td>46227136</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>10</td>
<td>184441684</td>
<td>25716</td>
<td>46449348</td>
<td>0</td>
</tr>
</tbody>
</table>

```
oracle@cnsiorap:/home/oracle> cat /proc/meminfo
MemTotal: 2560203484 kB
MemFree: 187199168 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: 17996 kB
PageTables: 28315696 kB
```
After an Hour….

<table>
<thead>
<tr>
<th>proc</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
</tr>
<tr>
<td>12</td>
</tr>
<tr>
<td>00</td>
</tr>
<tr>
<td>00</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>06</td>
</tr>
<tr>
<td>04</td>
</tr>
<tr>
<td>02</td>
</tr>
<tr>
<td>02</td>
</tr>
<tr>
<td>05</td>
</tr>
<tr>
<td>05</td>
</tr>
<tr>
<td>01</td>
</tr>
<tr>
<td>00</td>
</tr>
<tr>
<td>05</td>
</tr>
<tr>
<td>00</td>
</tr>
<tr>
<td>04</td>
</tr>
<tr>
<td>01</td>
</tr>
<tr>
<td>02</td>
</tr>
<tr>
<td>01</td>
</tr>
<tr>
<td>02</td>
</tr>
<tr>
<td>04</td>
</tr>
</tbody>
</table>

```
procs -----------memory--------- ---swap--- -----io----- -system- ----cpu------
```

```
Inactive(file); 357980 kB
Unevictable; 8832 kB
Mlocked; 8832 kB
SwapTotal; 43279615 kB
SwapFree; 41508052 kB
Dirty; 60 kB
Writeback; 0 kB
AnonPages; 2446452 kB
Nmapped; 82371665 kB
Shmem; 158377004 kB
Slab; 608592 kB
PageTables; 91015908 kB
NFS_Unstable; 0 kB
KernelStack; 17364 kB
WritebackTmp; 0 kB
CommitLimit; 173377596 kB
Committed_AS; 21451404 kB
VmallocTotal; 134217728 kB
Vmalloc_used; 2629372 kB
VmallocChunk; 131457396 kB
```

```
procs -----------memory--------- ---swap--- -----io----- -system- ----cpu------
```

```
HugePages_Total; 0
HugePages_Free; 0
HugePages_Rsvd; 0
HugePages_Rsvpt; 0
Hugepagesize; 1024 kB
```
4K Page Tables after 70 minutes

Linux Swap and PageTables using **87.7 GB** of Memory!

<table>
<thead>
<tr>
<th>proc</th>
<th>mem</th>
<th>swap</th>
<th>io</th>
<th>system</th>
<th>cpu</th>
</tr>
</thead>
<tbody>
<tr>
<td>338</td>
<td>8</td>
<td>1766820</td>
<td>1096980</td>
<td>1200</td>
<td>158901132</td>
</tr>
<tr>
<td>125</td>
<td>13</td>
<td>1767088</td>
<td>1096700</td>
<td>1316</td>
<td>158896348</td>
</tr>
<tr>
<td>420</td>
<td>4</td>
<td>1767396</td>
<td>1073704</td>
<td>1416</td>
<td>158891792</td>
</tr>
<tr>
<td>302</td>
<td>5</td>
<td>1767588</td>
<td>1089200</td>
<td>1424</td>
<td>158876220</td>
</tr>
<tr>
<td>227</td>
<td>7</td>
<td>1767652</td>
<td>1088700</td>
<td>1448</td>
<td>158870652</td>
</tr>
<tr>
<td>165</td>
<td>16</td>
<td>1767796</td>
<td>1093696</td>
<td>1444</td>
<td>158858216</td>
</tr>
<tr>
<td>452</td>
<td>16</td>
<td>1768980</td>
<td>1074352</td>
<td>1480</td>
<td>158858772</td>
</tr>
<tr>
<td>257</td>
<td>14</td>
<td>1769204</td>
<td>1096292</td>
<td>1276</td>
<td>158828368</td>
</tr>
<tr>
<td>177</td>
<td>6</td>
<td>1769172</td>
<td>1090828</td>
<td>1320</td>
<td>158821092</td>
</tr>
<tr>
<td>217</td>
<td>16</td>
<td>1769600</td>
<td>1095124</td>
<td>1364</td>
<td>158816144</td>
</tr>
<tr>
<td>144</td>
<td>17</td>
<td>1770068</td>
<td>1088160</td>
<td>1256</td>
<td>158841430</td>
</tr>
<tr>
<td>122</td>
<td>11</td>
<td>1771576</td>
<td>1082412</td>
<td>1276</td>
<td>158810608</td>
</tr>
<tr>
<td>219</td>
<td>10</td>
<td>1772768</td>
<td>1073684</td>
<td>1260</td>
<td>158807908</td>
</tr>
<tr>
<td>315</td>
<td>3</td>
<td>2033292</td>
<td>1076748</td>
<td>1152</td>
<td>158561024</td>
</tr>
</tbody>
</table>

```
[Image of a terminal output showing memory statistics]
```

**SHARE in Orlando 2011**
A little while Later….

74 ms response time and 0 TPS
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...

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 (Single Client Access Name)
  • 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
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

View Workload Replay: replay1_jfv

Status: In Progress

▼Summary
- Replay Name: replay1_jfv
- Capture Name: capturejfv1
- Duration (hh:mm:ss): 00:02:30
- Start Time: Jul 10, 2007 9:46:32 PM GMT -07:00
- End Time: Jul 10, 2007 9:48:12 PM GMT -07:00

Workload Profile
- Network Time (hh:mm:ss): 00:08:00
- Think Time (hh:mm:ss): 00:06:12
- Clients: 2
- Clients Finished: 2

Assessing the Replay
The Elapsed Time Comparison chart shows how much time the replayed workload has taken to accomplish the same amount of work as captured.

- When the Replay bar is shorter than the Capture bar, the replay environment is processing the workload faster than the capture environment.
- The divergence table gives information about both the data and error discrepancies between the replay and capture environments, which can be used as a measure of the replay quality.
Customer Experiences 11gR2 with Linux on System z
Oracle 10gR2 & zEnterprise Performance

Oracle RAC - Comparison z196 versus z10

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