-

Technical Notes
IBM Oracle International Competency Center (ICC)

August 15, 2023 ***Email address: ibmoracle@us.ibm.com***

Oracle 19c to 12c and 11.2.0.4 Database Performance Considerations with AIX on Power Systems including POWER10

Introduction

This document is intended to provide information, suggestions, and website links to assist with IBM Power Systems running the Oracle database. Suggestions for both the standalone database and Real Application Clusters (RAC) environments for Oracle 19c, 18c, 12c and 11.2.0.4 (terminal release) are included. POWER10 has been optimized for improved hybrid cloud performance and new hardware-enabled security capabilities. There is a required POWER10 AIX APAR which is listed in the AIX 7.X APAR table below. Issues on IBM Power Systems using AIX with the Oracle database can be minimized by following standard practices which apply to current IBM Power Systems generations.

Please note that Oracle has changed the database naming convention starting with Oracle 12.2. Oracle database 18c (year 2018) is the full release of 12.2.0.2. The recommended database product to target is 19c as it offers a greater duration of support by Oracle to March 2026 (see database release schedule).

The Oracle MOS link provides additional detail on understanding the new Oracle version numbers and the Oracle Database Release Schedule (Doc ID 742060.1):

Oracle version detail: https://docs.oracle.com/en/database/oracle/oracle-database/18/upgrd/oracle-database-release-numbers.html#GUID-1E2F3945-C0EE-4EB2-A933-8D1862D8ECE2

Oracle database release schedule - Doc ID 742060.1

https://support.oracle.com/epmos/faces/SearchDocDisplay?_adf.ctrl-state=z3pbgtsi9_266&_afrLoop=237916404979741

Table 1, Helpful reference links

Oracle and IBM Reference Information		
Oracle Database 19c & Oracle Database 19c RAC on IBM AIX - Tips	https://www.ibm.com/support/pages/node/6355761	Oracle 19c tips and considerations
Oracle Database on IBM Power Systems servers with Power10 Processors	https://www.ibm.com/support/pages/node/6515614	POWER10 tips and considerations
Oracle Database 11g and 12c on IBM Power Systems S924, S922 and S914 with POWER9 processors	https://www.ibm.com/support/pages/node/6355685	POWER9 tips and considerations on Techdocs
Current Oracle Database Release Schedule	For releases of Oracle for AIX, please refer to MOS note: Release Schedule of Current Database Releases (Doc ID 742060.1)	Oracle MOS document
Oracle 12c DB & RAC - IBM AIX: Tips and Considerations	https://www.ibm.com/support/pages/node/6355451	Tips, considerations and certification level information
Oracle Database on IBM Power with AIX Best Practices Presentation - Power Systems Virtual User Group	https://ibm.ent.box.com/v/PSVUG-OracleBestPracticesPart1 https://ibm.ent.box.com/v/PSVUG-OracleBestPractPart2	IBM Advanced Technology Group, ISV on Power - Oracle - AIX VUG call on 08/20/20 and 08/27/20. **Oracle Best Practices Part 1 (CPU & Memory)** **Oracle Best Practices Part 2 (I/O & Miscellaneous)**
Advanced Technology Group (ATG), ISV on Power - Oracle	https://w3.ibm.com/w3publisher/advanced-technical-group-oracle/collateral	Internal site - formerly Washington Systems Center Oracle IBM
Oracle Real Application Clusters on IBM AIX – Best practices in memory tuning and configuring for system stability	http://www.oracle.com/technetwork/database/clusterware/overview/rac-aix-system-stability-131022.pdf	RAC tuning and configuration Guide
Oracle Real Application Clusters (RAC) and Oracle Clusterware Interconnect Virtual Local Area Networks (VLANs) Deployment Considerations	http://www.oracle.com/technetwork/database/clusterware/overview/interconnect-vlan-06072012-1657506.pdf	RAC and VLAN deployment guide

Managing Raw Disks in AIX to use with Oracle Automatic Storage Management (ASM)	(MOS) Doc ID 1445870.1	My Oracle Support (MOS) https://support.oracle.com (Registration required)
IBM Service and Support Best Practices for Power Systems	https://www.ibm.com/support/pages/node/883882?mhsrc=ibmsearch_a&mhq=power%20systems%20best%20practices	Power best practices and recommendations
AIX Disk Queue Depth Tuning	https://www.ibm.com/docs/en/power8/8335-GCA?topic=HW4L4/p8ebj/p7ebjdrivequeuedepth.htm	IO queuing for disk
AIX Virtual Processor Folding is Misunderstood	https://www.ibm.com/support/pages/aix-virtual-processor-folding-misunderstood	
AIX Support Lifecycle Information	https://www.ibm.com/support/pages/aix-support-lifecycle-information	End of Service Reference Dates
IBM Spectrum Scale 5.1 is certified with Oracle RAC Database 19c on AIX 7.2 and AIX 7.3	https://www.ibm.com/support/pages/node/6575117	Certification with 19c version information (GPFS)
IBM Spectrum Scale (GPFS) Reference	http://www.ibm.com/support/knowledgecenter/STXKQY_4.2.2/com.ibm.spectrum.scale.v4r22.doc/bl1ins_oracle.htm	
Spectrum Scale Basic Installation Information with AIX on Oracle 12cR2 Oct 2019	https://www.ibm.com/support/pages/node/6355751	Installation notes and tips per certification with Oracle
IBM Techdocs Main Page - Technical Sales Library Review Recent docs, ICC Flashes, etc	http://www.ibm.com/support/techdocs/atsmastr.nsf/Web/Techdocs	Technical sales support database
Oracle RAC and DB Support Tools Bundle	MOS Doc ID 1459344.1	Get Proactive with Oracle Database Diagnostic Tools
Improving ASM Disk Discovery Time Best Practices	MOS Doc ID 1608549.1	Helpful ASM related information
Support Statement for Oracle DB Running on IBM Systems	https://www.ibm.com/support/pages/node/6449328	Reference the Oracle Roadmap for the corresponding links to Oracle documents on MOS

Power10 Performance Quick Start Guides	https://www.ibm.com/support/pages/system/files/inline-files/Power10_Performance_Quick_Start_Guides.pdf	Multiple QuickStart reference guides

Oracle DB 19c, 18c, 12c and 11.2.0.4 Standard Practices for IBM AIX

This document is a summary of standard practices for Oracle standalone and RAC database instances on 19c, 18c, 12c and 11.2.0.4 (terminal release) using IBM AIX. AIX 7.3 TL01 has been certified with Oracle Database version 19c. The minimum AIX version for POWER10 is AIX 7.2 TL5 SP3-2136 or AIX 7.2 TL4 SP5-2148. The minimum for POWER9 is AIX 7.2 TL2 SP2. AIX 7.2 TL05 or AIX 7.1 TL05 are suggested as a minimum strategy to maintain support.

The POWER certified version of AIX (minimum) with Oracle Database Release 19c is AIX 7.3 TL0 SP1, AIX 7.2 TL02 SP01 and AIX 7.1 TL05 SP01. Oracle Database Release 19c is the recommended long term support solution with AIX. The certified supported minimum AIX versions necessitate several IBM AIX APAR's be applied.

End of Service Pack Support

AIX 7.3 TL01 was released December 2022 with support projected to end of 2025. AIX 7.2 TL05 released Nov 2020 is current and has no projected end date listed. AIX 7.2 TL04 service pack support ended November 30, 2022. AIX 7.1 TL05 service pack support ends April 30, 2023. The terminal release of AIX 6.1 TL09 service pack support (EoSPS) ended April 2017.

AIX Support Lifecycle Information - https://www.ibm.com/support/pages/aix-support-lifecycle-information

Suggestions and Settings

The following pages discuss memory, CPU, I/O, network, and miscellaneous settings. In addition, some AIX APARs affecting Oracle 11.2.0.4, 12c, 18c and 19c are listed along with relevant Oracle patches for AIX 6.1, AIX 7.1, AIX 7.2 and AIX 7.3.

Memory

Specifications for kernel settings and Oracle large page usage are:

Kernel Settings

It is recommended to start with the default kernel settings and apply adjustments only as appropriately recommended. Adjusting the Restricted Tunables without proper guidance can have an undesirable impact on system stability. In most cases, AIX Support recommends the AIX 7 defaults for use with Oracle.

The general recommendation for most Oracle databases on AIX is to utilize 64KB page size and not 16MB page size for the SGA. If another application utilizes 16MB pages, then those

values would be non-zero, but the database would be configured not to use 16MB pages. Typically use of 16GB page size is not recommended.

A few AIX 7.2 features include:

- Security levels for the OS are enhanced and improved
- Starting with AIX 7.2 TL3 the default SMT setting has changed to SMT8 starting with POWER9 servers
- Up to 1,536 threads are supported in a single LPAR for POWER8 and POWER9 processors
- Maximum memory has been extended from 16TB to 32TB in a single LPAR for large applications

Some changes in AIX 6.1 and AIX 7.1 are:

- The elimination of the lru_file_repage tunable
- The default value for maxpin% increased from 80% to 90% in AIX 6.1
- New network option 'tcp_fastlo' introduced in AIX 7.1
- Change in LDR_CTRL setting for 'Shared Symtab'

LARGE PAGE USAGE WITH ORACLE

- IBM AIX supports three or four page sizes, depending on the hardware: 4KB, 64KB (medium), 16MB (large), and 16GB(huge). All four page sizes are supported by Power Systems from POWER5+ firmware level 240-202 onward.

- Page sizes set at 64KB and 16MB have been shown to benefit Oracle performance by reducing translation lookaside buffer processing to resolve virtual to physical addresses. Oracle 12c, 18c and 19c use 64KB pages for the SGA by default.

- Oracle Automatic Memory Management (AMM) uses the 64KB page size by default for the SGA and database (with the exception of the TNS Listener which is still 4KB).

PRE_PAGE_SGA

- The setting for 12.1 defaults to 'TRUE' which allocates all segments to the maximum. Prior to 12.1 the default was set to 'FALSE'.

- Pros:
  - With setting this to true, all segments are allocated to the MAXIMUM.

  - PRE_PAGE_SGA (at startup) will read "touching" all the memory pages. This can result in slower start up times but advantage is that all further requests to SGA memory are supposed to hit real physical memory and AIX will not need to do any additional allocations after startup.

- Cons:
  - It now takes more time for ANY ORACLE process to start as this "touching" of memory segments which is not done just during instance startup but also occurs for any new ORACLE process (i.e. a new database connection shadow process).

- The efficiency of this "touching" will depend on the page size used for the SGA. For example, an 80MB SGA using 64KB pages would need to "touch" 1250 pages whereas an SGA using 16MB pages would only need to "touch" 5 pages.

- To pin memory you set lock_sga to 'TRUE'. To use 16M pages one also needs to pin memory. If persistent memory usage issues are encountered overriding the default of pre_page_sga of 'TRUE' and setting it to 'FALSE' may be beneficial.

- If you are planning to use the "In-Memory" feature of Oracle database 12c, 18c or 19c it is recommended to set the pre_page_sga = TRUE (default)

LOCK_SGA = FALSE

- This is the default on all Oracle releases on AIX and means that the SGA is not pinned in memory.

- AIX Performance Support generally suggest not to pin the SGA.

- This is the suggested value, since it has been found that 64KB pages yield nearly the same performance benefit as 16MB pages, require no special management and minimize potential of negative impact of incorrectly configuring SGA size.

LOCK_SGA = TRUE

- To enable pinned SGA in Oracle set lock_sga to 'TRUE'

- To enable pinned shared memory (SGA) in AIX:
  vmo –p –o v_pinshm=1 (allows pinned memory—requires reboot)

Using 16MB Pages

- Generally, 16MB page size is not recommended as 64KB pages provide nearly the same performance and AIX is able to manage the distribution of memory between various requirements automatically. When using 16MB pages you need to set aside memory which can only be used for 16MB pages. For this reason, when using 16MB pages, you need to understand the dynamic memory requirements of your application and carefully size the amount of 16MB pages.

- If using 16MB pages you also need to pin memory (set lock_sga to 'TRUE')

- To set aside memory in AIX for 16MB pages you use the following command:
  - vmo –p –o lgpg_size=16777216 –o lgpg_regions=number_of_large_pages where number_of_large_pages=INT[SGA size -1)/16MB)]+1
  - Note that dynamically allocating 16MB pages ("-p flag") in a running LPAR can take a very long time and cannot be terminated. It is therefore strongly recommended to utilize the "-r" flag instead of the "-p" flag for the above vmo command. This tells AIX to allocate the 16MB pages during the next reboot and typically also results in a better distribution of allocated 16MB pages over all available processors.

- Capabilities required to allow Oracle to use 16MB large pages (implement as root): #chuser capabilities=CAP_BYPASS_RAC_VMM,CAP_PROPAGATE oracle
  See:http://www.ibm.com/developerworks/forums/servlet/JiveServlet/download/747-313057-14417535-361419/Oracle_Tuning_on_AIX.pdf
  and Oracle My Oracle Support note: 372157.1

- AMM CANNOT be used with pinned 16MB pages as it cannot be used together with "LOCK_SGA=TRUE. Setting large pages negates the need for AMM. Reference Oracle Metalink notes 443746.1 and 452512.1 explaining AMM.

ENVIRONMENTAL VARIABLES

LDR_CNTRL Settings

Oracle 11.2.0.4 (Single Instance)

- The use of 64KB page size for data, text, and stack regions is useful in environments with a large SGA and many online transaction processing (OLTP) users. For smaller Oracle instances, 4KB is sufficient for data, text, and stack. Use of 64KB for text, data and stack will increase the use of memory as memory is now allocated in 64KB chunks vs. 4KB in the past.
- 64KB page use for data, text, and stack is implemented separately from 64KB pages for the SGA and is done by means of an environment variable exported on behalf of the Oracle user. Please refer to the LDR_CNTRL Settings in MOS Note: USE OF AIX LDR_CNTRL ENVIRONMENTAL SETTINGS WITH ORACLE (Doc ID 2066837.1)
- Set for EACH database (oracle) user as well as for the Database TNS listener user.
- In the Oracle profile set and export:
  - LDR_CNTRL=DATAPSIZE=64K@TEXTPSIZE=64K@STACKPSIZE=64K <oracle user>
- In the Oracle Listener environment set and export:
  - LDR_CNTRL=DATAPSIZE=64K@TEXTPSIZE=64K@STACKPSIZE=64K <tnslistener user>
  - VMM_CNTRL=vmm_fork_policy=COR (add the 'Copy on Read' command)

- There is an option to ldedit the Oracle binaries so they use 64KB pages directly.

  Note that whenever a patch is applied or an Oracle relink is performed, this ledit will have to be performed again. # ldedit –btextpsize=64k –bdatapsize=64k –bstackpsize=64k $ORACLE_HOME/bin/oracle.
- When AME was enabled prior to Oracle 12.2, AIX used 4k page size instead of 64k page size for the Oracle database. Starting with Oracle 12.2 using 64k page size is enabled and recommended. Please see AME section below for more details.

Oracle 12.1 and later (Single Instance)

- Oracle 12.1 database and later releases by default will use 64k pages for text, data and stack. However, for the TNSLISTENER it still uses 4k pages for text, data and stack. To enable 64k page for the listener use the export command prior to starting the listener process. Note that running in an ASM based environment that the listener runs out of GRID_HOME and not ORACLE_HOME.
- The documentation for the "srvctl setenv" command changed in 12.1 or later releases. The -t or -T was removed in favor of -env or -envs
- In the Oracle Listener environment set and export:
  - LDR_CNTRL=DATAPSIZE=64K@TEXTPSIZE=64K@STACKPSIZE=64K <tnslistener user >
  - VMM_CNTRL=vmm_fork_policy=COR (add the 'Copy on Read' command)

Oracle 11g (RAC)

- As an Oracle user - srvctl setenv database -d <DB_NAME> -t "LDR_CNTRL=TEXTPSIZE=64K@DATAPSIZE=64K@STACKPSIZE=64K"
- As an Oracle Grid user - srvctl setenv listener -l LISTENER -t "LDR_CNTRL=TEXTPSIZE=64K@DATAPSIZE=64K@STACKPSIZE=64K"

Oracle 19c to 12c (RAC)

- As a Oracle Grid Infrastructure user:
- It is recommended to use 64KB pages with Database TNS listener (default name is LISTENER). This can be enabled by setting the LDR_CNTRL environment for the ora.LISTENER.lsnr resource with the following command.
  - Crsctl modify res ora.LISTENER.lsnr -attr "USR_ORA_ENV= LDR_CNTRL=DATAPSIZE=64K@TEXTPSIZE=64K@STACKPSIZE=64K VMM_CNTRL=vmm_fork_policy=COR" –unsupported
  - Additionally, do not try to set these using SRVCTL as the Java parser does not delimit the "@" sign
- Verification of LDR_CNTRL for RAC
  - crsctl stat res ora.LISTENER.lsnr -p|grep USR_ORA_ENV

    USR_ORA_ENV=LDR_CNTRL=DATAPSIZE=64K@TEXTPSIZE=64K@STACKPSIZE=64K VMM_CNTRL=vmm_fork_policy=COR

Stopping the listener then restarting it with 'srvctl start listener -n <nodename>' and then checking the BSS page size example:

root@rac201 # svmon -P 30802066|grep BSS

```
 888c4f        11 work text data BSS heap          m    118     0    0     118
 998ade        10 work text data BSS heap          m     15     0    0      15
```

.... This confirms the option took effect.

Without the above USR_ORA_ENV setting – note the "*sm*" in the output:
root@rac201 # svmon -P 30802092|grep BSS

```
 998b9e        11 work text data BSS heap          sm   1750     0    0    1750
 b30ef3        10 clnt text data BSS heap,          s    232     0    -       -
```

In the above two svmon commands the "-P" value is the PID of the "tnslsnr LISTENER" process.

SHARED SYMTAB

- The LDR_CNTRL=SHARED_SYMTAB=Y setting does not need to be specifically set in 11.2.0.4 or later releases. The compiler linker options take care of this setting and no longer needs to be specifically set. It is not recommended to have LDR_CNTRL=SHARED_SYMTAB=Y specifically set in 12c or later releases.
- Note: Having SHARED_SYMTAB=NO results in the shared symbol table run loader feature being disabled for any binaries running in that environment (including Oracle). As the shared symbol table feature is directly utilized to address the high work USLA heap usage reported in earlier bugs (where online patching feature was enabled) this resulted in the symptoms reported in this bug.
- In 11.2.0.4 (also 11.2.0.3) and 12c or later releases -- the reference to LDR_CNTRL=*SHARED_SYMTAB=Y* should be REMOVED if present
- LDR_CNTRL=*SHARED_SYMTAB=Y*@TEXTPSIZE=64K@DATAPSIZE=64K@STACKPSIZE=64K@<oracle user>

AIXTHREAD_SCOPE - Thread Tuning

- Both AIX 7.1 and AIX 6.1 specify the default environmental variable of AIXTHREAD_SCOPE=S (1:1).
- Oracle recommends system wide scope (AIXTHREAD_SCOPE=S) so this environmental variable is no longer required to be specifically set.
- Please refer to this link for additional thread tuning information:

https://www.ibm.com/support/knowledgecenter/en/ssw_aix_71/com.ibm.aix.performance/thread_tuning.htm

ACTIVE MEMORY EXPANSION (AME)

- Oracle 12cR2, including RAC, is certified for use with IBM's Active Memory Expansion (AME) using the following minimum product versions and levels of AIX 7.2 TL2 SP2 on POWER8 (native mode) with firmware FW860. Using the specified IBM product versions or newer results in AME using hardware compression which enables support of 64k pages with AME. Prior to these IBM product versions enabling AME disabled the use of 64k pages which in some cases can have a significant impact on Oracle database performance.

GLOBAL RUNQUE

- RT_GRQ (Global Runque) - This variable was recommended for pre-11.2.x releases on AIX. (Reference MOS - RAC and Oracle Clusterware Best Practices and Starter Kit (AIX) (Doc ID 811293.1). For 11.2.0.x and 12c or later database

versions, this is no longer required and should be removed. Improper use of this parameter can cause processor affinity issues and impact performance.

CPU

CPU specifications are:

- Symmetric multi-threading (SMT) mode: POWER10 - POWER8 support SMT8 while the POWER7 AIX supports up to SMT4. AIX/Oracle performance support encourage starting with the default and basing the SMT setting on which setting provides the best performance of your application.
- For AIX executing on POWER8 based systems, SMT4 mode may provide better performance than SMT8 mode, but it is workload dependent and both options should be tested.
- The SMT mode should be verified with the application to verify best performance. AIX executing on POWER10 based systems, SMT8 mode typically provides best performance.
- Virtual processor folding: This is a feature of Power Systems in which unused virtual processors are taken offline until the demand requires that they be activated. The default is to allow virtual processor folding, and this should not be altered without consulting AIX support. The minimum should be set to 2 (schedo parameter vpm_fold_policy=2) for both Standalone and RAC.
- For Oracle database environment set schedo parameter vpm_xvcpus to a value of 2 as we have seen AIX incorrectly folding too many processors if the parameter is left at default of 0.

- Please reference: 'AIX Virtual Processor Folding is Misunderstood' https://www.ibm.com/developerworks/community/blogs/aixpert/entry/aix_virtual_processor_folding_in_misunderstood110?lang=en for additional information.
- Specific to POWER7 SMT4 or POWER8 SMT4/8: Certain Oracle 11G or later parameters, including DB_WRITER_PROCESSES and PARALLEL_MAX_SERVERS, are partly derived from CPU_COUNT, and CPU_COUNT is equal by default to the number of logical CPUs. CPU_COUNT automatically adjusts to changes in virtual processor count and to SMT mode, up to three times the value on startup. Note that, when migrating from single-threaded platforms to Power Systems, or from POWER5 or POWER6 to POWER7 with SMT4, the value of CPU_COUNT will also increase, affecting DB_WRITER_PROCESSES, PARALLEL_MAX_SERVERS, and other dependent parameters. Queries that are sensitive to a degree of parallelism might change behavior as a result of migration to POWER7 or later generation. We suggest reviewing the PARALLEL_MAX_SERVERS parameter after migration, but to typically let DB_WRITER PROCESSES default, means not to set specifically.

RAC Minimum CPU

- schedo -p -o vpm_xvcpus=2

  Note that this is a critical setting in a RAC (also important for Standalone)

environment when using LPARs with processor folding enabled. If this setting is not adjusted, there is a high risk of RAC node evictions under light database workload conditions.

- This setting says that a minimum of 2 additional vp's will be online (e.g. not folded / disabled) at all times. With a shared processor systems using RAC, the minimum recommended value for vpm_vxcpus is 2, meaning there will be a minimum of 3 unfolded CPUs (the default 1 plus the 2 additional ones). This is recommended to avoid RAC reboot issues. A resource issue can be created when one Oracle process enters a tight loop polling on a fd and the Oracle process that is supposed to send to that fd does not get scheduled. Once that sending event occurs, things go back to normal and AIX housekeeping can run also.

  The default value of the vpm_xvcpus tunable parameter is 0, which signifies that folding is enabled. This means that the virtual processors are being managed. The schedo command is used to modify the vpm_xvcpus tunable parameter and is dynamic therefor not requiring a reboot.

For additional information on virtual processors, please refer to the 'Virtual processor management within a partition' section of the following AIX 7.1 document: https://www-01.ibm.com/support/knowledgecenter/ssw_aix_71/com.ibm.aix.performance/virtual_proc_mngmnt_part.htm?lang=en

I/O

I/O specifications are:

- If ASM is not used, max interpolicy striping (also known as *pp spreading* or *poor man's striping*), is suggested when logical volumes are created. To get the most benefit from spreading physical partitions across the LUNs, use a small physical partition size, for example, 32MB or 64MB. Note that the AIX default is not to spread PPs, means "INTER-POLICY = minimum". To utilize PP spreading, LVs have to be created first with "RANGE of physical volumes" set to "maximum" in smitty or the corresponding flag in mklv. After the LV was created that way you can then put a JFS2 file system on top of that LV.
- Async I/O is used even with Concurrent I/O (CIO)
  - With AIX 6.1 and 7.1, start with the asynchronous I/O defaults. With AIX 6.1, there is a new implementation of AIO. AIO kernel extensions are loaded at system boot (always loaded), AIO servers stay active as long as there are service requests, and the number of AIO servers is dynamically increased or reduced based on demand of the workload. The aio_server_inactivity parameter defines after how many seconds idle time an AIO server will exit. AIO tunables are now based on logical CPU count, and hence it is usually not necessary to tune minservers, maxservers, and maxreqs as in the past.
  - Oracle parameter (init.ora)

    disk_asynch_io = TRUE (this is the default value)
- Buffered file I/O on JFS2
  - The default filesystemio_options=ASYNCH

- In this case all table spaces, redo log file systems, and control file systems are using the kernel buffers rather than writing directly to disk.
- In this case, it does not matter whether redo log file systems and control file systems are 512Byte or 4k block size file systems.
- Oracle on AIX/POWER best performance is, however, usually achieved using CIO (though there are exceptions).
- Starting with Oracle 11.2.0.3+ There is a provision to create redo logs with 4k block size (init.ora parameter _disk_sector_size_override=TRUE needed). With this option set and redo log files created with 4k block size, redo related IO will be at minimum 4K in size. However, with a block size of 4K, there is increased redo wastage. In fact, the redo wastage, means additional IO written to disk in 4K blocks versus 512Byte blocks can be significant.

- Concurrent I/O (CIO) on JFS2
  - Set the Oracle parameter filesystemio_options=SETALL, or mount the filesystems (other than dump devices; may be required in older AIX /Oracle levels) with the CIO option. It is not necessary to both SETALL and mount filesystems with the CIO option. Using SETALL is the preferred method. For more information reference: *How To Check If CIO Is Being Used in 11.2.0.2 and above (MOS Doc ID 1478924.1)*
  - With Oracle 11.2.0.4 or later, set the Oracle parameter filesystemio_options=SETALL, AND DO NOT EXPLICITLY mount the filesystems with the CIO option. 11.2.0.4 or later versions implement CIO using AIX-internal O_CIOR option.
  - If using CIO with SETALL or with CIO/DIO mount ***you must*** create separate file systems for redo logs and control files (or a single filesystem for both), with an agblksize of 512Bytes rather than the default 4KB to prevent demoted IO and the accompanying significant performance degradation.
  - The ioo parameters aio_fsfastpath and posix_aio_fsfastpath accelerate CIO. They are enabled by default in AIX 6.1 and 7.1.
  - AIX 6.1, JFS2 with 11.2.0.2 and higher Bug notice:
    Bug 9310972 - ENHANCEMENT: INTRODUCING O_CIOR FLAG WHEN OPENING DATAFILES IN AIX 6.1. With AIX 6.1, IBM introduced a new open flag O_CIOR which is same as O_CIO, but this allows subsequent open calls without CIO. The advantage of this enhancement is that other applications like cp, dd, cpio, dbv can access database files in read only mode without having to open them with CIO.

    Note: Starting with Oracle 11.2.0.2 when AIX 6.1 is detected, Oracle will use O_CIOR option to open a file on JFS2.Therefore you should no longer mount the filesystems with mount option "-o cio".

- IBM mount advice for database files:
  - Data files: Use CIO filesystemio_options=SETALL, and default agblksize (4k); mount with no options.
  - Redo logs: Create with agblksize of 512Bytes and mount with no options. With filesystemio_options=SETALL, Oracle will use direct I/O for Redo logs.

- Note on planning the Block Size of Redo Log Files: Unlike the database block size, which can be between 2K and 32K, redo log files always default to a block size equal to the physical sector size of the disk. Historically, this has typically been 512 bytes (512B). Some newer high-capacity disk drives offer 4K byte (4K) sector sizes for both increased ECC capability and improved format efficiency. Most Oracle Database platforms are able to detect this larger sector size. The database then automatically creates redo log files with a 4K block size on those disks. However, with a block size of 4K, there is increased redo wastage. In fact, the amount of redo wastage in 4K blocks versus 512B blocks can be significant.
- Control files: Create file system with agblksize of 512Bytes and mount with no options. With filesystemio_options=SETALL, Oracle will use direct I/O for control files.
- Archive logs: Mount -o rbrw . Do not use CIO; use the jfs2 rbrw option
- Dumps: Mount –o rbrw
- The file system mount option 'noatime', is suggested and applies to the JFS2 file system as an enhancement feature. It is suggested to use the 'noatime' mount option for JFS2 file systems hosting the Oracle binaries only.

- IOO tunables j2_nBufferPerPagerDevice and j2_dynamicBufferPreallocation:
  - For Oracle databases storing its data files in a JFS2 file system the defaults are typically too low. Evaluate if there is a shortage of buffers based on increasing values reported with: *vmstat –v | grep "external pager filesystem I/Os blocked with no fsbuf"*. If the reported value increases, first increase the setting for j2_dynamicBufferPreallocation from 16 (16k slabs) to 64; monitor. If increasing this parameter up to 256 does not help, then consider raising the value of j2nBufferPerPagerDevice which is the starting value for dynamic buffer allocation.
  - See help pages for information about these parameters. **Do not change AIX restricted tunables without the advice from IBM AIX support.**
- ASM considerations for standalone Oracle 11gR2:
  - For identifying, renaming, and securing ASM raw devices, see Managing Raw Disks in AIX to use with Oracle Automatic Storage Management (ASM).
  - ASM will use asynchronous I/O by default, so filesystemio_options=ASYNCH (default) is appropriate.
  - For clustered ASM (e.g. RAC) configurations, SCSI reservation must be disabled on all ASM hdisk and hdiskpower devices (e.g. reserve_policy=no_reserve). In standalone use (non-RAC) of ASM, hdisks and hdiskpower devices does not need to have SCSI) reservation disabled.
  - The following initialization parameters should be adjusted for ASM:
    - Add 16 to the value of processes
    - Add an additional 600KB to the value of large pool size
    - Add to shared pool size the aggregate of the values returned by these queries:
      - SELECT SUM(bytes)/(1024*1024*1024) FROM V$DATAFILE;

- SELECT SUM(bytes)/(1024*1024*1024) FROM V$LOGFILE a, V$LOG b WHERE a.group#=b.group#;
- SELECT SUM(bytes)/(1024*1024*1024) FROM V$TEMPFILE WHERE status='ONLINE';

- For disk groups using external redundancy, every 100GB of space needs 1 MB of extra shared pool, plus 2MB
- For disk groups using normal redundancy, every 50GB of space needs 1MB of extra shared pool, plus 4MB
- For disk groups using high redundancy, every 33GB of space needs 1MB of extra shared pool, plus 6MB
- Source: http://docs.oracle.com/cd/E18283_01/server.112/e16102/asminst.htm#CHDBBIBF
- ASM disk size mismatch issue with AIX. AIX shows that the size correctly while ASM shows it 8 times smaller. Pls see 11.2.0.4 Oracle Bug fixes.

Network

This section outlines the minimum values applicable to network configurations.

Kernel configurations

These values are generally suggested for Oracle, and can be considered as starting points (please note all udp settings are specific for RAC):

- sb_max >= 1MB (1048576) and must be greater than maximum tpc or udp send or recvspace (if you are using RAC and very large udp_recvspace, you might need to increase sb_max)
- tcp_sendspace = 262144
- tcp_recvspace = 262144
- udp_sendspace = db_block_size * db_file_multiblock_read_count+4
- udp_recvspace= 10 * (udp_sendspace)
- tcp_fastlo = 1. This is new in AIX 7.1 (no –p –o tcp_fastlo=1). The tcp_fastlo default setting is off or '0'. This can be extremely useful for improving performance folding or loop back (bequeath) connections and should be evaluated with application testing. Reference: AIX 6.1 TL9 (SP03/SP04) APAR IV67463 or AIX 7.1 TL3 (SP03/SP04) APAR IV66228 for a memory leak issue.
- rfc1323 = 1

  Ephemerals (non-defaults suggested for a large number of connecting hosts or a high degree of parallel query; also to avoid install-time warnings)

  - tcp_ephemeral_low=9000
  - tcp_ephemeral_high=65500
  - udp_ephemeral_low=9000
  - udp_ephemeral_high=65500

Jumbo frames are Ethernet frames larger than the standard maximum transmission unit (MTU) size of 1500 bytes. They can be up to 9000 bytes. They are used to reduce the number of frames

to transmit a given volume of network traffic, but they only work if enabled on every hop in the network infrastructure. Jumbo frames help to reduce network and CPU overheads. The use of Jumbo frames between the database RAC nodes in a cluster (RAC interconnect) is strongly recommended.

MTU adapter port specific settings will be overridden with setting 'mtu_bypass = ON'. This is complemented with 'tcp_pmtu_discover = 1' for MTU path discovery.

Miscellaneous Specifications

Live Partition Mobility (LPM) & SVC Upgrade

- Oracle RAC miscount timer - default of 30 seconds can be extended to 60 seconds if required to avoid node evictions and reset to 30 seconds after the LPM activity is completed.
- Use of Dedicated Adapter recommended instead of SEA
- Network bandwidth if minimal can cause restrictions
- IBM POWERVM Virtualization Introduction and Configuration SG 24-7940

Resource Limits

- ulimits (smit chuser or edit /etc/security/limits to create a stanza for Oracle/grid user and set -1 (unlimited) for everything except core. Example below

```
oracle:
  data = -1
  stack = -1
  fsize_hard = -1
  cpu_hard = -1
  data_hard = -1
  stack_hard = -1
  fsize = -1
  nofiles = -1
  cpu = -1
  rss = -1
```

- Maximum number of PROCESSES allowed per user (smit chgsys). Set this value to 16386 (16k)

Performance

TCP Setting of rfc1323=1

- This is a long-standing network tuning suggestion for Oracle on AIX, although the default remains 0 (in global "no" parameter list) in AIX 6.1 and 7.1. A network retransmission latency issue has recently been discovered when rfc1323=1 on AIX 6.1 TL6 and APAR IV00755 is also present. Our recommendation is to use rfc1323=1 (Oracle recommendation), and to ensure that "IV13121: TCP RETRANSMIT PROCESSING IS VERY SLOW 12/05/30 PTF PECHANGE" (or equivalent) is applied. Rfc1323 is default for the ISNO values set by the interface device driver, for VETH and 10Gbit it is enabled (and for NFS enabled by default).

- The default limit is 64K and the connection needs to be set on both sides. A setting greater than 64K is achieved with setting rfc1323=1 on both sides of the connection or it will revert to the default value of 64K.
- Reference: http://www-01.ibm.com/support/knowledgecenter/ssw_aix_71/com.ibm.aix.performance/rfc1323_tunable.htm?lang=en

Potential Oracle Database Performance Issue – Oracle Table'x$ksmsp'

Issue:
"If customers are using any query that accesses the table: 'x$ksmsp' they should disable such queries.

Oracle provides the use of x$ksmsp to allow customers to give a listing of the RAM heap to see how; free space is allocated within the shared pool, the sizes of available chunks on the freelist for the shared pool and RAM.

Unfortunately, this leads to a myriad of issues including; system hangs, heap issues (locking) etc in production systems and selecting from x$ksmsp on a production system is to be avoided.

This basically does the following for each subheap in the pool, it will:

1) Grab the shared pool latch for that subheap
2) Walk through ALL chunks of memory in that subheap calling back to ksmspc to extract information about the chunk of memory
3) Release the shared pool latch for that subheap

Even on a minimal sized shared pool this means you are holding the shared pool latch for a significant amount of time which then blocks anyone needing to perform any shared pool memory operation that requires that latch."

Precautions:
1) It is NOT recommended to run queries on X$KSMSP when the database instance is under load.

2) Performance of the database will be impacted, especially currently with very large SGAs.

3) Bug 14020215 was filed for ORA-600 errors and unplanned outages running queries directly on X$KSMSP.

4) There is a view, X$KSMSP_NWEX, in later versions of 11g that is safer to use for investigation of memory usage.

5) Oracle STRONGLY recommends you not run these queries unless specifically requested by Oracle Support to do so.

Recommend:
Suggest checking via a crontab entry running a query against a variety of X$ tables -

eliminating any query against the x$ksmsp resolves such latch contention issues.

Standalone Database and RAC Miscellaneous Tips and Reference

Database Upgrades

- **11g to 12c**

- Differences in RT priority setting from Oracle Database 11g to 12c (real time)
Power7: Oracle Database 11g R2 AIX 7.1 vs Power9: Oracle Database 12.1 AIX 7.2 - Oracle hidden parameters have changed from database 11g to 12c. After removing Logwriter setting, then database does not startup. It was observed that the LGWR (Log writer for redo) process used to be specified and placed in the real time class when started, along with others like key RAC processes via parameter _high_priority_processes . 12c now specifies all of these as the default for that parameter with the desire to remove the parameter, however LGWR is not being started in the RT class when removed.

This example is for a 19.8 RAC cluster:

```
# ps -efl |grep lg
  240001 A      grid 10355016          1    0   60  20 830ea6590 370716                Nov 16      -  0:04
asm_lgwr_+ASM1
  200001 A     root 13304066 27918764   0  60 20 809481480   236        16:50:35  pts/0  0:00 grep lg
  240001 A    oracle 26083592          1    0   39  20 819763590 386612              16:50:10      -  0:00
ora_lg01_rac1tst1
  240001 A    oracle 27132254          1    0   39  20 849129590 386612              16:50:10      -  0:00
ora_lg00_rac1tst1
  240001 A    oracle 33227068          1    0   39  20 87918f590 387764              16:50:10      -  0:00
ora_lgwr_rac1tst1
```

It seems in later versions 12c to 19c this parameter defines a ‘mask’ for various Oracle backgrounds. Suggestion is going with the default in 12x and above. It appears to be this way for some time now: Doc IDs 12951619.8 and 1523164.1 where LGWR is already added to high_priority_processes in 11.2.0.4 (with a backport available for 11.2.0.3.5) and so there will not be any need to adjust LGWR priority manually.

Troubleshooting Oracle RAC - Data Collection

- Collect and review ‘Last Gasp’ file http://www.oracle.com/technetwork/database/rac-aix-system-stability-131022.pdf (pg 7)
- Review CRS Logs for events with timestamps
- OSWatcher - network
- AWR Report
- RDA - collects Oracle logs
- Trace File Analyzer (TFA) and ORAchk (replaces RACcheck tool MOS Doc ID 1268927.2)

Performance

- Redo Logs – placement on separate file systems
- Bundle Patch or 11.2.0.4 (includes Bundle Patch)
- OCSSD Bin – Process & Thread Priority (see note in appendix)

Oracle Timers

- Network Heartbeat 1 per/sec with 30 sec default timeout
- CSS Voting Disk Heartbeat 1per/sec 27 sec default timeout

- AIX Driver read/write 30 Second timeout
- Oracle Miscount value – default is 30 sec, used 60 sec for LPM large lpar's successfully then changed back to 30 sec
- Oracle CRS OCSSD – default 200 seconds is normal timeout value; during reconfiguration this is temporarily reduced to 27s.

AIX fixes for Oracle 11.2.0.4 and 12c

Some of the common AIX fixes for Oracle 11g and 12c follow. The APAR number will be unique to the specific AIX TL level with the most current TL level providing a rollup of the earlier TL & SP levels. The APAR's listed may apply to either Oracle dbase version. If an APAR is not listed for a TL, then the fix is already included in that TL, or that TL is not vulnerable to the problem.

AIX 6.1 Service Pack Support ends Apr 2017 for AIX 6.1					**Symptoms**
APAR level:	**TL09**	**APAR #**			
UDP MULTICAST BUG MAKES ORACLE RAC UNSTABLE, INCLUDING NODE EVICTIONS	Base	IV35888			
THREAD_CPUTIME() RETURNS INCORRECT VALUES APPLIES TO AIX	Base	IV30219			
DISABLE MULTICAST LOOPBACK FOR MPING symptom: Dropping packet due to direction mismatch. Was expecting r but got s.	Base	IV36225			
THREAD_CPUTIME_FAST RETURNS INVALID VALUES IN POWER4/POWER5 – AWR does not generate	SP03	IV74746			AWR does not run in automatic mode
XMGC NOT TRAVERSING ALL KERNEL HEAPS. APPLIES TO AIX 6100-09	SP02	IV53582			
VLAN ADAPTER ON PCIE2 4-PORT ADAPTER(10GBE SFP+) DROPS PACKETS		IV76194			Node hangs
CANNOT UNMOUNT FILE SYSTEM AFTER USING SHARED SYMBOL TABLE.	Base	IV76409			6.1.8 - 6.1.9 If the shared symbol table feature is used (LDR_CNTRL=SHARED_SYMTAB=Y or -bshrsymtab linker flag)
UMOUNT FAILS WITH DEVICE BUSY ERROR EVEN WITHOUT ACTIVE PROCESS	Base	IV39905			Unable to unmount a file system unmount fails with "Device busy" error
IMPROVED DRIVER'S FAIRNESS LOGIC TO AVOID I/O STARVATION - large number of LUNs with heavy I/O stress, may observe poor I/O perform on some LUNs	TL09	IV75385			

PERFORMANCE REGRESSION WHEN USING OLSON TIMEZONE FORMAT. Default for AIX 6.1 and AIX 7 is Olson Time. POSIX was traditional format and previous older versions of AIX used POSIX	TL09	IV86773			HIPER- Poor perf, Node hangs/evictions. See appendix for details.
Memory Leak on File System Unmount: The system runs out of memory. Common indication is inability to create a new process with error message "0403-031 The fork function failed. There is not enough memory available	TL09	IV40315			Mem Growth
PROBLEMS CAN OCCUR WITH THREAD_CPUTIME AND THREAD_CPUTIME_FAST	TL09 SP08	IV93840			Get ORA-0600 errors – Is HIPER
MEMORY LEAK IN PTH_GET_TLS() - free_tls_mem() is not calling tls_free() if there is only tls entry for that pthread during pthread_exit(). Outcome of java profiler (MALLOCDEBUG) and RSS do not match.	TL09	IJ11198			Mem leak - outcome of java profiler (MALLOCDEBUG) and RSS do not match.
POTENTIAL DATA LOSS USING VIRTUAL FC WITH NUM_CMD_ELEMS > 256	TL09	IV91199			HIPER - DMA errors with multi virtual fc adapters if over 256
POWER9 VPM FOLD THRESHOLD	TL09	IJ11675			Number of unfolded cores is less than expected
ASSERT IN ADAPTER DRIVER DURING ERROR RECOVERY	TL09	IJ15458			Possible hang or sys crash during error recovery
Non Critical AIX 6.1 APAR level:	**TL9**	**APAR**			
DISABLE MULTICAST LOOPBACK FOR MPING symptom: Dropping packet due to direction mismatch. Was expecting r but got s.	Base	IV36225			
LPARSTAT -H AND -H WILL NOT SHOW HYPERVISOR STATISTICS APPLIES TO AIX 6100-09	Base	IV53394			
CAT /PROC/SYS/FS/JFS2/MEMORY_USAGE MAY RETURN INVALID ARGUMENT	SP03	IV54359			
IV46203: UMOUNT FAILS WITH DEVICE BUSY ERROR EVEN WITHOUT ACTIVE PROCESS APPLIES TO AIX 6100-08	Base	IV39905			
LOADING 5.3 TLS ENABLED LIBS BY 5.2 APPS CAUSED CORE DUMP IN 32B APPLIES	Base	IV30118			
A SPECIAL-PURPOSE LINKER FLAG WORKS INCORRECTLY.	Base	IV42840			
HANG UNDER SOME CIRCUMSTANCES WHEN A C++ DTOR UNLOADS LIBRARIES	Base	IV63322			
IV37549: ATTEMPT FAILED.OPENED STATE DOES NOT RETRY NPIV LOGIN, IF FIRST APPLIES TO AIX 6100-09	Base	IV37549			

THREAD_CPUTIME_FAST RETURNS INVALID VALUES IN POWER4/POWER5 – AWR does not generate	Base	IV74746			AWR does not run in automatic mode
ENABLING 'TCP_FASTLO' OPTION COULD LEAD TO MEMORY LEAK	SP03 & 04	IV67463			
VSCSI ADAPTER STOPS RESPONDING, CAUSING I/O TO HANG	Base	IV50780			
AIX 7.1 Service Pack Support for AIX 7.1 ends April 2023					
APAR level:	**TL03**	**APAR**	**TL04** released 12/4/15	**APAR**	
NETWORK PERFORMANCE DEGRADATION ON FC5899 (AUSTIN) ADAPTER APPLIES TO AIX 7100-03	SP01	IV58687			
XMGC NOT TRAVERSING ALL KERNEL HEAPS. APPLIES TO AIX 7100-03 14/04/17	SP01	IV53587		HIPER	
UDP MULTICAST: SHORT PACKET FOR SOME LISTENERS. APPLIES TO AIX 7100-03	SP01	IV33047			
DATA PTR INCORRECTLY INCREMENTED IN UDP RECEIVE. APPLIES TO AIX 7100-03	SP01	IV34454			
TATX/LOOKUPS/FILE OPENS APPLIES TO AIX 7100-03	SP01	IV44289			
POSSIBLE STARVATION OF LARGE I/OS UNDER HEAVY WORKLOAD APPLIES TO AIX 7100-03	SP01	IV44347			
IV60218/60052: PORT/DEVND FC5899 DRIVER HOG CPU WHEN ENTSTAT ON CLOSED	SP01	IV60218/ 60052			
UDP SEND PERFORMANCE ENHANCEMENTS APPLIES TO AIX 7100-03	SP01	IV54257			
DISABLE MULTICAST LOOPBACK FOR MPING symptom: Dropping packet due to direction mismatch. Was expecting r but got s.	Base	IV36204			
AWR does not run in automatic mode on P5	SP03	IV74746			
VLAN ADAPTER ON PCIE2 4-PORT ADAPTER (10GBE SFP+) DROPS PACKETS	Base	IV75642	Base	IV76558	Node hangs
CANNOT UNMOUNT FILE SYSTEM AFTER USING SHARED SYMBOL TABLE. APPLIES TO AIX 7100-03 & 04. If shared symbol table feature is used (LDR_CNTRL=SHARED_SYMTAB=Y or -bshrsymtab linker flag)	SP06	IV76410 See also IV40079	Base	IV76411	Not resolved on SP05, Fixed in SP06.
UMOUNT FAILS WITH DEVICE BUSY ERROR EVEN WITHOUT ACTIVE PROCESS	Base	IV40079 See also IV76410			Resolved in 7.1.3.6

PERFORMANCE REGRESSION WHEN USING OLSON TIMEZONE FORMAT. Default for AIX 6.1 and AIX 7 is Olson Time. POSIX was traditional format and previous older versions of AIX used POSIX	SP04	IV87788 HIPER	Base	IV86730	Poor perf, Node hangs/evictions. 7.1.3.4 up fixed in 7.1.3.7 - is HIPER. See appendix for details.
IMPROVED DRIVER'S FAIRNESS LOGIC TO AVOID I/O STARVATION - large number of LUNs with heavy I/O stress, may observe poor I/O perform on some LUNs	Base	IV78514	Base	IV75433	
Memory Leak on File System Unmount: The system runs out of memory. Common indication is inability to create a new process with error message "0403-031 The fork function failed. There is not enough memory available	Base	IV40336			Mem Growth
PROBLEMS CAN OCCUR WITH THREAD_CPUTIME AND THREAD_CPUTIME_FAST	TL03 SP08	IV93884	TL04 SP03	IV93845	Get ORA-0600 errors – Is HIPER
GETSOCKNAME RETURNS INVALID PATH NAME FOR AF_UNIX SOCKETS	TL05	IJ04311			TL5 and up
SYSTEM CRASH WITH "MACHINE CHECK TYPE - SLB ERROR"			TL4 SP2	IJ00747	Corruption seen and lack of dump
MEMORY LEAK IN PTH_GET_TLS() - free_tls_mem() is not calling tls_free() if there is only tls entry for that pthread during pthread_exit(). Outcome of java profiler (MALLOCDEBUG) and RSS do not match.	TL04	IJ11193	TL05	IJ11195	Mem Leak
POTENTIAL DATA LOSS USING VIRTUAL FC WITH NUM_CMD_ELEMS > 256	TL03 TL04	IV90142 IV90915	TL05	IV91202	HIPER - DMA errors with multi virtual fc adapters if over 256
GETSOCKNAME RETURNS INVALID PATH NAME FOR AF_UNIX SOCKETS	TL01 SP05	IJ04311			Invalid path name
POWER9 VPM FOLD THRESHOLD	TL04	IJ12054	TL05	IJ10664	Number of unfolded cores is less than expected
When using AIX Active Active PCM, disk open processing for ALUA supporting disks can take more than 1 sec. ASM startup delays encountered during SP2 to SP3 upgrade.			TL05	IJ11305	TL5 SP32 Causes delays with ASM
ASSERT IN ADAPTER DRIVER DURING ERROR RECOVERY			TL05	IJ16284	Possible hang or sys crash during error recovery
VPM INTELLIGENT FOLDING NOT SUPPORTED applies to TL5	TL05			IJ20661	xvcpu related - VPM may target a non-optimal core to fold rather than cores in the smaller srads.
CRASH ON P10 AIX LPAR TESTING ORACLE DATABASE	TL05 SP03/ newer	IJ41089			Customer may experience a kernel crash on p10 platforms.
Memory consumption increases when cached_routes no option is enabled.	TL05 SP09	IJ45294			HIPER

VIOS CRASH IN GIDS_OK(), IP_OUTPUT_POST_FW(), IP_OUTPUT() AND TCP_OUT - VIOS xxx	TL05.43 to TL5.47	IJ45295			HIPER
Non Critical AIX 7.1 APAR level:	**TL3**	**APAR**	**TL4**	**APAR**	
A SPECIAL-PURPOSE LINKER FLAG WORKS INCORRECTLY.	**Base**	IV42782			
CAT /PROC/SYS/F/JFS2/MEMORY_USAGE MAY RETURN INVALID ARGUMENT	SP3	IV55030			
IV44690: ADD ABILITY TO REORDER TOC SYMBOLS IN LIMITED CIRCUMSTANCES APPLIES TO AIX 7100-03	Base	IV44690			
IV63130: HANG UNDER SOME CIRCUMSTANCES WHEN A C++ DTOR UNLOADS LIBRARIES. APPLIES TO AIX 7100-03	Base	IV63130			
ATTEMPT FAILED. OPENED STATE DOES NOT RETRY NPIV LOGIN, IF FIRST APPLIES	Base	IV37484			
THREAD_CPUTIME_FAST RETURNS INVALID VALUES IN 4/POWER5 – AWR does not generate	Base	IV74746			
ENABLING 'TCP_FASTLO' OPTION COULD LEAD TO MEMORY LEAK	SP03/04	IV66228		IV67443	
AIX 7.2 (Dec 2015)					
APAR level:	**TL00**	**APAR**	**TL01**	**APAR**	
POSSIBLE SYSTEM CRASH USING PROCFS TO READ 32BIT PROCESS MAP FIL	SP1	IV79441			
AFTER LIVE UPDATE IFIX STATE MAY BE LEFT AS Q; REBOOT REQUIRED	SP1	IV79639			
MIRRORVG/SYNCVG ON MINIMAL AND MIGRATION INSTALL FAILS	SP1	IV79848			
SYSTEM CRASH AFTER APPLICATION SETS SIGNAL MASK	SP1	IV80412			IV80412m1a fixes conflict with IV79441
MULTIBOS MAY FAIL TO MOUNT OR REMOVE A STANDBY INSTANCE	SP1	IV81482			
CORE DUMP IN MANY COMMANDS WHEN USING NIS	SP1	IV82224			Network Information Services(NIS) auth
PERFORMANCE REGRESSION WHEN USING OLSON TIMEZONE FORMAT. Default for AIX 6.1 and AIX 7 is Olson Time. POSIX was traditional format and previous older versions of AIX used POSIX	SP00, SP01	IV86890, IV87789			Poor perf, Node hangs/evictions Is HIPER. See appendix for details.
SOME APPLICATIONS MAY FAIL (core dump) AFTER BEING LINKED ON AIX 7.2 TL1 APPLIES TO AIX 7200-01			TL01	IV94362	Updating to current AIX 7.2 TL1 causes Oracle core dump

					with EBS Is HIPER
PROBLEMS CAN OCCUR WITH THREAD_CPUTIME AND THREAD_CPUTIME_FAST	TL00 SP03	IV93883	TL01 SP01	IV93885	Get ORA-0600 errors – Is HIPER
GETSOCKNAME RETURNS INVALID PATH NAME FOR AF_UNIX SOCKETS	TL02 SP03	IJ04933			Invalid path name
SYSTEM CRASH WITH "MACHINE CHECK TYPE - SLB ERROR"			TL1	IJ01368	Corruption seen and lack of dump
MEMORY LEAK IN PTH_GET_TLS() - free_tls_mem() is not calling tls_free() if there is only tls entry for that pthread during pthread_exit(). Outcome of java profiler (MALLOCDEBUG) and RSS do not match.	TL01	IJ11194	TL02 TL03	IJ11196 IJ11197	Mem Leak – Java profiler mismatch
POTENTIAL DATA LOSS USING VIRTUAL FC WITH NUM_CMD_ELEMS > 256	Base TL01	IV92284 IV91511	TL02	IV91191	HIPER - DMA errors with multi virtual fc adapters if over 256
POWER9 VPM FOLD THRESHOLD	Base TL01	IJ10423 IJ12055	TL02 TL03	IJ10535 IJ10425	Number of unfolded cores is less than expected
When using AIX Active Active PCM, disk open processing for ALUA supporting disks can take more than 1 sec. ASM startup delays Encountered during SP2 to SP3 upgrade.			TL1 SP4	IJ11457	TL1 SP4 Causes delays with ASM process startup.
LDEDIT MISSING OPTION LIST IN THE USAGE OUTPUT Base 7.2.0 thru 7.2.4 (TL04)	TL03 SP02	IJ17766	TL04	IJ17973	Excessive Memory consumption. See appendix for work around
ASSERT IN ADAPTER DRIVER DURING ERROR RECOVERY	TL03	IJ17157			Possible hang or sys crash during error recovery
VPM INTELLIGENT FOLDING NOT SUPPORTED	7.2 Base	IJ21131	TL02 TL04	IJ23481 IJ21390	xvcpu related - VPM may target a non-optimal core to fold rather than cores in the smaller srads.
REPORT ALLOCATIONS ARE MISSING WITH MALLOCDEBUG=LOG SET	7.2 Base	IJ21356			
APPLICATIONS MAY FAIL UNEXPECTEDLY WITH PERMISSION ERRORS	TL05	IJ31049			Hiper - AIX 7200-05 Technology Level * with bos.mp64 below the 7.2.5.2 level.
SETREUID() INCORRECTLY SETS SAVE UID	TL05 SP01	IJ31207			After updating to AIX 7.2 TL5 SP1, setreuid() behaves differently, saving an incorrect value to suid. Permission errors may arise.

CRASH ON P10 AIX LPAR TESTING ORACLE DATABASE	TL04	IJ41090	TL05 SP00 SP01 & 02	IJ41091	Customer may experience a kernel crash on p10 platforms.
CRASH ON P10 AIX LPAR TESTING ORACLE DATABASE	TL05 SP03/ 04	IJ41092	TL05 SP05 & newer	IJ41093	Customer may experience a kernel crash on p10 platforms.
Memory consumption increases when cached_routes no option is enabled or Systems running the VIOS 3.1.2.x. IJ45429 - MEMORY LEAK IS SEEN DUE TO CACHED_ROUTE NO OPTION	TL05 SPxx	IJ44534/ IJ45429			HIPER
MEMORY LEAK IS SEEN DUE TO CACHED_ROUTE NO OPTION	TL0	IJ45349	TL5.5 to TL5.7	IJ45429	HIPER - Systems running the VIOS xxx
MEMORY LEAK IS SEEN DUE TO CACHED_ROUTE NO OPTION - 7.2 TL5.100 to 201	TL5.100 to TL5.201	IJ45349			HIPER - Systems running the VIOS xxx
VIOS CRASH IN GIDS_OK(), IP_OUTPUT_POST_FW(), IP_OUTPUT() AND TCP_OUT - VIOS xxx	Base	IJ44047			HIPER
VIOS CRASH IN GIDS_OK(), IP_OUTPUT_POST_FW(), IP_OUTPUT() AND TCP_OUT - VIOS xxx	TL00	IJ45346	TL5.100 to TL5.201	IJ44786	HIPER
VIOS CRASH IN GIDS_OK(), IP_OUTPUT_POST_FW(), IP_OUTPUT() AND TCP_OUT - VIOS xxx	TL5.4 to TL5.6	IJ45427			HIPER
AIX 7.3					
APAR level:	**TL00**	**APAR**	**TL01**	**APAR**	
CRASH ON P10 AIX LPAR TESTING ORACLE DATABASE	TL00	IJ41094	TL01	IJ41088	Customer may experience a kernel crash on p10 platforms.
AIX 7.3 CHANGED SOME PERFORMANCE TUNING COMMANDS' DEFAULT OUTPUT	TL00	IJ38518 HIPER			scripts parsing -x output may fail
Memory consumption increases when cached_routes no option is enabled.	TL00	IJ44513/ IJ44047	TL01	IJ44761/ IJ44749	IJ44047 HIPER – see Misc VIOS APAR
VIOS CRASH IN GIDS_OK(), IP_OUTPUT_POST_FW(), IP_OUTPUT() AND TCP_OUT - VIOS xxx	TL00	IJ44512	TL1	IJ44749	HIPER
VIOS CRASH IN GIDS_OK(), IP_OUTPUT_POST_FW(), IP_OUTPUT() AND TCP_OUT - VIOS xxx	Base	IJ44543			HIPER

VIOS & Miscellaneous					
P8-E880 9119-MHE & MME Firmware MFG:BELL3 ADAPTER FAILING ON AIXERR AND STX APPLIES TO AIX 7100-03	7.1.3	IV77596			Ports on adapter (feature code EN27/EN28, CCIN 57D4) may become un-usable
VIOS CRASH IN GIDS_OK(), IP_OUTPUT_POST_FW(), IP_OUTPUT() AND TCP_OUT - VIOS 3.xxx – 3.xxx Applies to AIX 7.1.5, 7.2 & 7.3		IJ44047 IJ44512 IJ44543 IJ44749 IJ44786 IJ45295 IJ45346 IJ45427			HIPER System Crash

Oracle PSU, N-Apply Bundle Information and older version reference on MOS

Recommended N-Apply Bundle 2 for AIX 11.2.0.4 with critical fixes (Doc ID 2022567.1)
11.2.0.4. AIX bundle 2 can be downloaded here: Patch 20877677

Recommended N-Apply Bundle 1 for AIX 12.1.0.2 with critical fixes (Doc ID 2022559.1)
12.1.0.2. AIX Bundle 1 can be downloaded here: Patch: 20117253

Linked from:
Master Note - Recommended N-Apply Bundles for Oracle DB on AIX (Doc ID 1947965.1)

Quick Reference to Patch Numbers for Database PSU, SPU (CPU), Bundle Patches and Patchsets (Doc ID 1454618.1)

'My Oracle Support' (MOS) website Doc ID 1507249.1 - IBM POWER7 AIX and Oracle Database performance considerations (Oracle Database 10.2 & 11.2.0.3 focused).

Commands to verify Grid Infrastructue Patch Set Update levels (gipsu):

- $CRS_ORA_HOME/OPatch/opatch lsinventory -bugs_fixed |grep -i 'GI PSU'
- $ORACLE_HOME/OPatch/opatch lsinventory -bugs_fixed | grep -i 'DATABASE P'

AIX Related

Bug #		Description	Version affected	Fixed in	Ref. MOS Note id.
		Diagnostic			
17208793	BUG	KERNEL DUMP WHEN CRS NEEDS TO REBOOT A CLUSTER NODE – RAC Clusterware Patch creates sysdumpstart when a node reboots.	RAC 11.2 and 12.1	Included 11.2.0.4.6 & 12.1.0.2.3	Flag needs to be set to activate dump collection
		Merged 11.2.0.4 Bug Fixes			
		BUG 16901346 - OSYSMOND PROCESS TAKING ALMOST 5% OF CPU BECAUSE OF HIGH OPEN FDS COUNT			
		BUG 17733927 - CSS CLIENTS TIMEOUT UNDER HEAVY CONNECTIVITY LOADS ON AIX			
		OR			
		Merge patch is: BUG 19202720 - MERGE REQUEST ON TOP OF 11.2.0.4.0 FOR BUGS 16901346 17733927			
23088473	Bug	MERGE REQUEST ON TOP OF 11.2.0.4.0 FOR BUGS 21394872 13957757 16470836 17715578			
		11.2.0.4 Bug Fixes			
17294810		BUG 17294810 - PUTTING THREAD_CPU TIMER IMPLEMENTATION BACK TO 11.2.0.4			
Very 17501296		After upgrade to 11.2.0.4 unable to delete any rows from table with context index due to error: "PLS-00306	11.2.0.4		
18066615		Bug 18066615 : WRONG RESULTS AFTER UPGRADE TO 11.2.0.4	11.2.0.4		
19181582		DEADLOCK BETWEEN LG0N ON 'LGWR WORKER GROUP ORDERING'			
17551261		ORA-904 WHILE RUNNING SELECT QUERY which is in 11.2.0.4. and fixed in terminal release of 12.1.01	11.2.0.4		
14602788		HIGH CPU CONSUMPTION BY QMON PROCESS AFTER UPGRADE TO 11.2.0.4 fixed in 12.1 AQ:11GQ: Q00* PROCESS SPIN IN BOTH INSTANCES	11.2		

14275161		ORA-7445 [EVAOPN3()+135] INSERTING FROM PL/SQL AFTER UPGRADE TO 11.2.0.4 fixed in 12.1 - ORA-600 [RWOIRW: CHECK RET VAL] ON CTAS	11.2		
18828868	Patch	EXPLOSION IN AQ BACKGROUND SLAVES RESULTING IN ORA-18/ORA-20	11.2.0.3		
18194267		AIX-11.2.0.4-SIHA:ORAAGENT.BIN MEMORY LEAK	11.2.0.4		
18828868	Patch	EXPLOSION IN AQ BACKGROUND SLAVES RESULTING IN ORA-18/ORA-20	11.2.0.4		
18261183	Patch	GRID INFRASTRUCTURE ORAAGENT.BIN GENERATES <DEFUNCT> THREADS	11.2.0.4		
17802407		ORA-7445 [QERIXFETCHFASTFULLSCAN+8212] AFTER UPGRADE TO 11.2.0.4	11.2.0.4		
13807031	Patch	RMAN DUPLICATE DOES NOT FOLLOW SYMBOLIC LINKS FOR REDOLOG FILES IN 11.2 - fixed in 12.1	11.2.0.4		
19382066	RAC Enhance ment	IMPROVE RELIABILITY OF PROCESS PRIORITY SETTING	11.2.0.4, 12.1	12.2	Use caution if priorities are specially set
22085533	Patch	ACFS PERFORMANCE ISSUE ON AIX	11.2.0.4		Fixes approx. 10 Bugs
19987218	Bug	LOG FILE PARALLEL WRITE SLOWER THE HIGHER CPU_COUNT . Workaround: set _log_parallelism_max to 8 or lower	11.2.0.4		
25294369	Bug	ORACHK: VALIDATE OS LEVEL if OLSON TZ FORMAT IS ENABLED OR NOT	11.2.0.4		Perf & evictions. See appendix for details.
27434665	Bug - ASM	Lpar with virtual disks maped from VIOS as LV's. LV's on VIOS mapped to the lpar as virtual disks are showing 8X smaller size than the actual size while creating ASM disks from Oracle. (note: PSU is dated 4/2018)	11.2.0.4		Patch 28813878: GRID INFRASTRUCTURE PSU 11.2.0.4.190115 did not include fix
32109594	Bug	ASM IS UNABLE TO START WITH SMR CORRUPTED ERROR POST OCT 2020 PSU APPLY	11.2.0.4		See Appendix for more detail –> 11g Upgrades
32123313	Bug	SMR FILE KEEPS GETTING CORRUPTED AFTER APPLYING OCT 2020 DBPSU IN AIX 7 (Shared Memory Region errors)	11.0.2.4		See Appendix for more detail –> 11g Upgrades
		12.1 Bug Fixes			12.1.0.1 fix support to end Aug 31, 2016
20160490		CPU LEAK BY OCSSD PROCESS AFTER GRID UPGRADE FROM 12.1.0.1 TO 12.1.0.2	12.1.0.1 & 12.1.0.2		
21378561	Patch	SHARED POOL LATCH CONTENTION WHEN POPULATING INMEMORY STORE	12.1.0.2		
21095582	Bug	12.1 ALERT LOG DOES NOT SHOW LARGE PAGES USAGE IN AIX	12.1.0.1		Fixed in 12.1.0.2 nuisance Bug involving Alert Log accuracy

19145675	Base Bug	In Mem REPOPULATE READS HAVE TOO MANY SINGLE READS - BLOCKS BEING READ TWICE AFTER UPGRADE TO 12.1.0.2 (Duplicate Bug 20352562, 19787571)	12.1.0.2	12.1.0.2.2	Fixed in 12.1.0.2.2
13542050		Use of KGL mutexes might block on bogus mutex holder			
2062511.1	Doc	AWR IS NOT BEING GENERATED AFTER UPGRADE FROM 11203 TO 12102 ON POWER4, POWER 5 OR POWER5+ FRAME A (Ref: APAR IV74746)	12.1		
21378561	Bug	SHARED POOL LATCH CONTENTION WHEN POPULATING INMEMORY STORE.	12.1.0.2		
19881439	Bug	LATCH VIOLATION ISSUE ORA 600 [525]	12.1.0.2n etwork		
20400122	Bug	PROCESS MEMORY USAGE INCREASE BY 5 MB AFTER UPGRADE TO 12.1	12.1.0.2		
27846298 ~~21238475~~	Bug	HIGHER EXCLUSIVE MEMORY ALLOCATION AT OS LEVEL IN "WORK TEXT DATA BSS HEAP"	12.1.0.2		
19382066	RAC Enhance ment	IMPROVE RELIABILITY OF PROCESS PRIORITY SETTING	11.2.0.4, 12.1	12.2	Use caution if priorities are specially set
21915719	Bug	12C HANG: LGWR WAITING FOR 'LGWR ANY WORKER GROUP'	12.1.0.2		Crash, hang & user connection
23089357 ~~22832149~~	Bug	SHUTDOWN FAILS WITH ORA-449 "BACKGROUND PROCESS 'MMON' UNEXPECTEDLY TERMINATED with AIX 7.1.4, 7.2	12.1.0.2		Surfaced in AIX 7.1 TL4 and AIX 7.2 with Shutdown Immediate
23091948	Bug	PAGING ISSUE AFTER MIGRATE TO 12.1.0.2 VERSION. (memory consumption)	12.1		See appendix note for 12.1 memory consumption issues with AIX 7.1
23056988	Patch	ACFS PERFORMANCE SIGNIFICANTLY SLOWER THAN JFS2	12.1.0.2		Oracle 12.1.0.2.160719 for ACFS
21095582	Bug	12.1 ALERT LOG DOES NOT SHOW LARGE PAGES USAGE IN AIX	12.1.0.2		Large pgs configured but only 4K pages show in Alert Log
21915719	Bug	12C HANG: LGWR WAITING FOR 'LGWR ANY WORKER GROUP'	12.1.0.2		11203 to 12102 upgrade, frequent connectivity issues (DB went to hung state) and application users unable to connect to database
22992306	Bug	Host Metric CPU Utilization in Metric Group LPAR Performance on AIX too high with AIX 6.1	EM 12.1.0.5.0		Enterprise Manager Base - Host Metric CPU Utilization
21829392	Patch	M001 PROCESS SPINNING IN KEWFSU1_SEL_USAGE_1()	12.1.0.2		
21051872	Patch (Proactive	ORA-00600:[KRR_PROCESS_READ_ERROR_1] DIAGNOSTIC DUMP DEREFERENCES ZEROED FIELDS	12.1.0.2.4		Diag fix to dump correct diag infor for ORA-600 [krr_process_read_erro

					r_1]
22226365	Bug	THREADED_EXECUTION=TRUE - SCMN PROCESS RES MEMORY INCREASE Ref: Oracle Multi-Threaded Oracle Model- 12c Doc ID 1958348.1. Note: Multi-Threaded mode is False by default in AIX and Oracle 12.1 & 12.2.	12.1.0.2		Grid Infra - threaded execution processes consume a high amount of memory causing servers to be bounced to release the memory
24796183	Bug	RAC NODE DOESN'T GET CRASHED BY KILLING OCSSD AFTER CRASHDUMP ENABLED with no vendor clusterware	12.1.0.2		Clusterware trying to collect crashdump - not working (no crashdump, and node not rebooted)
25164540	Bug	RAC NODE DOESN'T GET CRASHED AFTER KILLING OCSSD ON GI/VCS VENDOR CLUSTER	12.1.0.2		This second one impacts AIX when vendor clusterware is in use
18775971	Patch	ORA-04031: UNABLE TO ALLOCATE ("SHARED POOL Default Calculation Of Shared Pool Subpools Is Vulnerable Leading To ORA-4031 (Doc ID 2225248.1)	12.1		Shared pool allocation in Oracle 12c with > 32 LPs. Recommend to upgrade to 12.2 or apply patch.
20480452	Merge Patch	Fixes 2 Bugs: Bug 18485835 WRONG RESULT ON QUERY WITH IN/ANY SEMIJOIN SUBQUERY IF OFE LESS THAN 11 AND Bug 19475484 APPSPERF_DBIM: 12.1.0.2 DATA DICTIONARY SQL PLAN REGRESSION ISSUE	12.1.0.2.0		Note: If issues with patch – Request new build of the Merge Patch to overlay on top of the installed PSU.
24916212	Backport Bug Fix	IMPROVE PERFORMANCE OF LNXADD, LNXSUB, LNXMUL, LNXSUM ON AIX – SIMD with PL/SQL related	12.1.0.2		Fixed in base 12.1 & 18.1, backported to 12.1
		12.2 Bug Fixes			
23089357 ~~22832149~~	Bug	APPSST12102:SHUTDOWN FAILS WITH ORA-449 "BACKGRND PROCESS ''MMNL' UNEXPECTEDLY TERMINATED with AIX 7.1.4, 7.2. Back out 22832149 and install 23089357 for complete resolution.	12.1.0.2		EBS RI may be affected. Surfaced in AIX 7.1 TL4 & 7.2. Replacement Bug.
29306513	Bug	12.2.0.1 JAN-2019 GIRU FAILS TO BE APPLIED USING GRIDSETUP.SH '-APPLYPSU'	12.2.0.1		Workaround available – MOS Doc 2504773.1
24916212	Backport Bug Fix	IMPROVE PERFORMANCE OF LNXADD, LNXSUB, LNXMUL, LNXSUM ON AIX – SIMD with PL/SQL related	12.2.0.2		Fixed in base 12.2 & 18.1, backported to 12.1
25416731	Patch	TABLESPACE IO STATISTICS MISSING FROM AWR REPORT	12.2.0.1, 18.1		Seen in 18.4
29135891	Bug	SCM0 PROCESS USING 100% CPU - CPU capacity of 5 cores was significantly used up. Also seems to be related in significant increase in control file reads in the DB.	12.2		RAC perf issue. The RAC internal bug is Bug 24590018 - DLM Statistics Collection and Management Slave

		RAC Node Eviction/Hang			
20235486	Bug	This lowers the priority of osysmond from 0 to 20 so sched can run			
		schedo -p -o vpm_xvcpus=2 Currently set to 0, this sets to 2 always gives us a second cpu to finish what osysmond.bin is waiting for. Min is quantity of 2.			Frequently not set properly.
17208793	**Diag Bug**	KERNEL DUMP WHEN CRS NEEDS TO REBOOT A CLUSTER NODE – RAC Patch creates sysdumpstart when a node reboots. RAC 11.2 and 12.1 Included 11.2.0.4.6 & 12.1.0.2.3. Flag needs to be set to activate dump collection			
13940331	Bug	OCSSD threads are not set to the correct priority (Doc ID 13940331.8)			
16901346	Bug	This reduces the CPU utilization of osysmond			
IV53587		Shipped 7.1.3.3 & 6.1.9.3 XMGC NOT TRAVERSING ALL KERNEL HEAPS	HIPER		
Data Collection		Collect CSSD logs & Last Gasp file data			
IV87788	APAR	PERFORMANCE REGRESSION WHEN USING OLSON TIMEZONE FORMAT. Default timezone format for AIX 6.1 and AIX 7 is Olson Time. POSIX was traditional format and previous older versions of AIX used POSIX	HIPER		AIX 7.1.3, 7.1.4, 7.2 Poor perf, node hang/evictions. See appendix for details.
		18c & 19c Bug Fixes			
24916212	Backport Bug Fix	IMPROVE PERFORMANCE OF LNXADD, LNXSUB, LNXMUL, LNXSUM ON AIX – SIMD with PL/SQL related	18.1		Fixed in base 12.1 & 18.1, backported to 12.1
30159782	Generic Bug	Remote Copy Fails if using openssh-7.2. 7.4, etc.	12.2.0.1, 18.1 & 19.1		Oracle GRID install issue on system with the newer version(s) of ssh. See notes below.
29340414	20.1 Bug Fix	GIPC RANK ISSUE - with remote interfaces metrics for HAIP failover situations.			Backportable to 19.5 & 19.6. Potential HA failover issues when interface does down.
25416731	Patch	TABLESPACE IO STATISTICS MISSING FROM AWR REPORT	12.2.0.1, 18.1		Seen in 18.4
32367954	Bug	NODES EVICTED WITH ERROR CLSB:8997: [INFO] ORACLE CLUSTERWARE INFRASTRUCTURE ERROR IN OCSSD: FATAL	19.1		fix backs out interconnect monitoring

		SIGNAL 6 HAS			introduced in 19.8 – no more system statx calls.
32588318	Bug	ACFS SLOW PERFORMANCE ON IBM-AIX 7.2 19CCOMPARED TO JFS/LINUX	12.1 - 19c		
30227028	Bug	GRID only - Incorrect Node Aliveness Check Causes Cluster Reconfiguration Hang	19.10		See MOS Doc ID 2767299.1 for 3node or larger cluster info
32550751	Bug	AIX CHANGE MGA TO USE SHM* API'S TO IMPROVE SCALABILITY	19 & 18c		PDB performance – decreased overall cpu usage
33748144	Merge Patch	Merge Patch for RU 19.10.0.0.0 OF 33557617 31844249 33608329. The following are the bugs fixed by this patch: 33380871: HIGH CPU ON KSLWT_UPDATE_STATS_ELEM 33529654: DIAGNOSTIC PATCH FOR BUG 32869560 ", 31844249: MAX NUMBER OF CPUS LIMIT FOR SERVERS WITH MORE THAN 1031 CPUS 33608329: DIAGNOSTIC PATCH FOR BUG 33474100	19.10		AWR top 10 processes identifies KXSGETRUNTIMELOCK AND SSKGSLCAS. KXSGETRUNTIMELOCK gathers stats about executions of shared cursors.
31844249	Patch	MAX NUMBER OF CPUS LIMIT FOR SERVERS WITH MORE THAN 1031 CPUS	19.xx		
32869560	Patch	HIGH CPU ON KXSGETRUNTIMELOCK AND SSKGSLCAS. The fix is included in 23c but after 19.15 setting the flag from default of true to false disappeared and still needs to be included	19.10		Fixed in 19.10.
33380871	Patch	HIGH CPU ON KSLWT_UPDATE_STATS_ELEM Included starting with 19.15 + newer and in 23c. Port needed for 19.11 -19.14	19.10		Fixed in 19.10.
33608329	Patch	kslgess Diagnostic Patch for BUG 33474100. Bug 33474100 which is still listed 'being worked' by Oracle Dev. Diagnostic patch was for 19.10 can be requested via SR to port to newer version.	19.10		Fixed in 19.10.
2829438.1	Fix	Memory Expansion option active on IBM AIX, impact on DB Instance Shutdown. Lengthy database shutdown, about 20 minutes when SGA_TARGET is set to 1250 GB on an IBM AIX on POWER Systems (64-bit) 7.2 platform.	19.11 & newer		MOS doc

Additional Information

This Technical Note was authored by IBM's Wolfgang Tertel and Wayne Martin. For more information on this Technical Note, please send your questions to the IBM Oracle International Competency Center at **ibmoracle@us.ibm.com**

Appendix

Some of the following reference information is in original formatting to not alter the data represented.

Oracle 12.2.0.2 Database Release Naming Convention Changes

Beginning with release 12.2.0.2, new releases will be annual. The version will be the last two digits of the release year. Please more information please see MOS Note 742060.1 The release originally planned as 12.2.0.2 will now be release 18 (for 2018), and the release originally planned as 12.2.0.3 will be release 19. Releases 18 and 19 will be treated as under the umbrella of 12.2 for Lifetime Support purposes.

Instead of Patch Sets, Patch Set Updates, and Database Bundle Patches, the new releases will be maintained with Release Updates (RU) and Release Update Revisions (RUR). For more information see Release Update and Release Update Revisions for Database Proactive Patch Program (Doc ID 2285040.1)

The new feature releases of the database product will be provided annually and patchsets will no longer be released. To support both security-related fixes and high-priority non-security fixes to each annual release, quarterly Release Updates (RUs) will be provided each January, April, July and October. To allow customers to keep current on just security-related fixes once their feature environment becomes stable, quarterly Release Updates Revisions (RURs) will be provided each January, April, July and October for the RUs from the past 6 months.

• RUs are proactive, highly tested bundles of critical fixes which enable customers to avoid known issues. They are specific to a particular annual release.
• RURs contain security and regression fixes to a RU that extend the RU's lifetime up to two quarters. They are specific to a particular RU.
• The legacy terms 'Patchset', 'Patchset Update', and 'Database Bundle Patch' will no longer be meaningful for 12.2 database software.

OCSSD Bin – Process & Thread Priority

Oracle RAC ocssd.bin, cssdagent, cssdmonitor, and osysmond.bin all (and should) have priority (PRI) of '0' and a scheduling policy (SCH) of '-'as well as all threads for ocssd.bin, cssdagent, and cssdmonitor also have a PRI '0'.

The threads are often overlooked and forgotten. The main process maybe correct but the threads associated with those processes also need to run (inherit) at those elevated priorities.

Known 11.2.0.3 OCSSD Bug 13940331
Priority should be 0 and not 60 (see PRI column in data).
Please ensure the ocssd threads are also running at "0" and "—"?

ps -p ps -ef | grep ocssd.bin | grep -v grep |awk '{print $2}'` -mo THREAD

Bug 13940331 - AIX: OCSSD threads are not set to the correct priority (Doc ID 13940331.8)

On AIX has found that the threads created for OCSSD were not inheriting the correct priority this may cause evictions when the machine is running under high workload.
The processes show the functions "ntevque->netevpque->poll" in their call stack and may hang.

Log Example:

```
svp0090bdc:root# ps -p `ps -ef | grep ocssd.bin | grep -v grep | awk '{ print
$2 }'` -mo THREAD
    USER      PID     PPID        TID ST  CP PRI SC    WCHAN        F     TT
BND COMMAND
  oracle  8454212 11141280          - A    0   0 35        * 10240103      -
   - /opt/GRID/app/grid/11.2.0.3/bin
       -        -        -   53412031 S    0   0  1 f1000f0a10032f40
8410400      -   - -
       -        -        -   67174423 S    0  60  1 f1000f0a10040140
8410400      -   - -
       -        -        -   69402727 S    0  60  1 f1000f0a10042340
8410400      -   - -
       -        -        -   69730365 S    0  60  1        -   418400      -
   - -
       -        -        -   70189183 S    0  60  1 f1000f0a10042f40
8410400      -   - -
       -        -        -   71958629 Z    0  60  1        -   c00001      -
   - -
       -        -        -   74317843 S    0  60  1 f1000f0a10046e40
8410400      -   - -
       -        -        -   76742815 S    0  60  1 f1000f0a10049340
8410400      -   - -
       -        -        -   82182299 S    0   0  1 f1000f0a1004e640
8410400      -   - -
       -        -        -   82313403 Z    0  60  1        -   c00001      -
   - -
       -        -        -   85000239 Z    0  60  1        -   c00001      -
   - -
       -        -        -   92143809 Z    0  60  1        -   c00001      -
   - -
       -        -        -  104005671 Z    0   0  1        -   c00001      -
   - -
       -        -        -   40894877 S    0  60  1 f1000f0a100a7040
8410400      -   - -
       -        -        -   41156961 S    0  60  1 f1000f0a100a7440
8410400      -   - -
       -        -        -   42992033 S    0  60  1        -   418400      -
   - -
       -        -        -   43188529 Z    0  60  1        -   c00001      -
   - -
       -        -        -   43254061 S    0  60  1 f1000f0a100a9440
8410400      -   - -
       -        -        -   43319609 Z    0  60  1        -   c00001      -
   - -
```

Database and Database Upgrade Notes & Issues

11g to 12c Upgrade Issues

- **Paging**

In a few cases, heavy paging has been observed after upgrading from Oracle Database 11.2.0.3.4 and AIX 6100 TL09 to DB 12.1.0.2 and AIX 7100 TL03. Both Oracle and IBM Support agree that based on the individual server environment, some additional memory may be warranted (based on the increased 12.1 memory footprint and increasing page size from 4K to 64K) or having the workload rebalanced to reduce paging which impacts system performance. The use of Perfpmr output has been helpful in determining and evaluating memory utilization.

- **Differences in RT priority settings from 11g to 12c (real time)**

- P7: 11g R2 AIX 7.1 vs P9: 12.1 AIX 7.2 Oracle hidden parameters have changed from 11g to 12c. Cust removed Logwriter setting and then is does not startup. Cust has observed that the LGWR (Log writer for redo) process used to be specified to be placed in the real time class when started, along with others like key RAC processes via parameter _high_priority_processes. 12c now specifies all of these as the default for that parameter so we would like to remove the parameter, however LGWR is not being started in the RT class when we remove it.

NOTE that this is for a 19.8 RAC cluster -- you want to check what your default values are before modifying alter system set
"_high_priority_processes"='LMS*|LM1*|LM2*|LM3*|LM4*|LM5*|LM6*|LM7*|LM8*|LM9*|LM*|LCK0|CKPT|DBRM|RMS0|*LGWR*|CR*|RS0*|RS1*|RS2*' scope=spfile;
... and after restart of instance:

```
# ps -efl |grep lg
  240001 A     grid 10355016       1   0  60 20 830ea6590 370716          Nov 16      -  0:04 asm_lgwr_+ASM1
  200001 A     root 13304066 27918764   0  60 20 809481480   236          16:50:35  pts/0  0:00 grep lg
  240001 A   oracle 26083592       1   0  39 20 819763590 386612          16:50:10      -  0:00 ora_lg01_rac1tst1
  240001 A   oracle 27132254       1   0  39 20 849129590 386612          16:50:10      -  0:00 ora_lg00_rac1tst1
  240001 A   oracle 33227068       1   0  39 20 87918f590 387764          16:50:10      -  0:00 ora_lgwr_rac1tst1
```

11g October 2020 Oracle PSU Upgrade Potential Issue

- **Oct PSU Upgrade -11.2.0.4 + AIX 7.1**

Oracle processes not starting after July & Oct 2020 PSU update. Experienced process Pmon errors after PSU update and could not startup the new instance after upgrading to 11.2.0.4 PSU versions July & Oct . Messages: Process Pmon can't attach to SMR (Shared Memory Region errors). Backing out the PSU for either July or Oct allowed the instance to start up normally. SMon can't attach to shared mem region.
Note: PMON and SMON are two required background processes. PMON is the Process Monitor which is responsible for recovering processes when the user process fails. PMON does the process cleanup. SMON is the System Monitor which is responsible for recovering the system after a failure.
11/17 October 2020 PSU. In particular, in 11.2.0.4 we are seeing SMR is corrupted. Shut down and restart the instance to recreate it. Errors in the alert.log. There are two Oracle Bug Fixes that appear to help resolve the issue (similar.

Bug 32109594 : ASM IS UNABLE TO START WITH SMR CORRUPTED ERROR POST OCT 2020 PSU APPLY
Bug 32123313 : SMR FILE KEEPS GETTING CORRUPTED AFTER APPLYING OCT 2020 DBPSU IN AIX 7 (Shared Memory Region errors)

19c Installation

- The minimum 19c database PSU level for installation of Oracle with AIX is version 19.3 which require PSU upgrades from the base version to get to current supported level

Memory – Oracle 11g

Some background on why Oracle 11g may require more memory than the previous 10g version. 11g uses "dynamic" SGA/PGA while previously10g used “static” SGA/PGA. In 11g the SGA/PGA is managed as one memory area. Even in an attempt to an "apples-to-apples" test of setting the 11g memory target to the same size (as combined SGA+PGA as a 10g install) would not really be applicable as 11g SGA dynamically sizes SGA and PGA allocations as needed within the overall memory target value.

In 10g the PGA and SGA both can be sized UP. However the SGA in 10g while it can be sized UP, it cannot be sized DOWN and the memory the SGA allocates will stay with SGA forever (or until the next database restart). Therefore, increasing the SGA memory with 11g is often done to accommodate this dynamic change in the function of 11g's SGA.

12c Adaptive Features

It has been found helpful to test turning this feature off to eliminate it as a cause of performance related issues. Try setting: OPTIMIZER_ ADAPTIVE_FEATURES to FALSE.

12.1 Increased Memory Becomes Consumed / Paging

System is paging after adding more memory for 12.1. The common environment is AIX 7.1, 12.1.0.2 & RAC. Significant idle connections are being seen, (ie, 2K for 11.2.0.4 vs 6K for 12.1.0.2) and the available memory is consumed.

We are seeing with 12c a high number of idle connections taking up memory and leading to paging with 12.1 and AIX. What is happening is Oracle DCD in 12c swamps the LPAR with open connections consuming memory and (thereby) taking down instances on the LPAR. To correct this condition, one would set these 2 parameters in the sqlnet.ora file:

```
SQLNET.EXPIRE_TIME=10
USE_NS_PROBES_FOR_DCD=TRUE
```

The latter 'NS_PROBES' command changes the DCD - Dead Connection Detection - behavior to pre-12c functionality by using NS probes instead of TCP/IP probes.

AIX ‘Shutdown Immediate’ command errors with ‘ORA-449’

This error is seen with AIX 7.1 TL04 and AIX 7.2 and 12.1.0.2.5. Oracle is providing a fix via Oracle Bug 23089357 (replaced 22832149). This error was not seen in AIX 7.1 TL03 and 12.1.0.2.5 and resolves a timing issue between the AIX and Oracle database processes. The error log files fill with informational errors that do not impact system performance.

This Oracle Bug fix may also resolve these errors:
ORA-27144: attempt to kill process failed
ORA-27300: OS system dependent operation:getthrds64 failed with status:4
ORA-27301: OS failure message: Interrupted system call
ORA-27302: failure occurred at skgpstime3

I/O Issues

Use IOSTAT -D to look for "service queue full" (qfull) conditions and high counts of Adapter Overflow errors with FCSTAT.

IBM Spectrum Scale Setting Info with Oracle RDBMS RAC

In addition to the performance tuning suggestions within the chapter titled *"Configuring and Tuning your system for GPFS"* of the *IBM Spectrum Scale: Administration Guide*, the following recommendations are provided.
Spectrum Scale, previously known as General Parallel File System or GPFS, is a high-performance clustered file system that is a complimentary solution when deploying Oracle Database Real Application Cluster (RAC) configurations. GPFS has the following certified uses:

- ORACLE_HOME directory for shared Oracle RAC database installation
- Database files for tablespaces and other general database object containers
- Oracle Clusterware registry and membership files including Oracle Cluster Registry (OCR) and Vote Disks, as well as the Grid Infrastructure Management Repository (GIMR)
- ORACLE_BASE for a common repository of alert logs and diag traces for the RAC cluster

After initial GPFS installation, the following configuration considerations and tuning parameters are suggested:

IBM Spectrum Scale Performance Tuning

Oracle databases open and access IBM Spectrum Scale files in the correct manner by default. Do not use any special mount options (eg. DIO) for GPFS. The Oracle database instance parameter "filesystemio_options" should remain at the default value of SETALL.

When configuring Network Shared Disk (NSD) devices, there will be a one-to-one relation of a storage LUN for each GPFS NSD. One or more LUNs/NSDs can be used for a single GPFS file system. Storage LUNs for a single file system should use the same RAID type (eg. RAID-5 or RAID-10). It is not recommended to mix RAID types within the *same* GPFS file system, but different file systems may use different RAID types. As an example, one might use RAID-10 arrays for Oracle REDO logs which need good sequential write performance and RAID-5 for data and index tablespaces which may be accessed in a random manner.

Storage LUNS / GPFS NSDs can be created from different arrays (different HDDs, controllers etc) within the storage subsystem. In this manner, when the GPFS file systems are created, multiple NSDs would be used to produces the desirable effect of spreading I/O across the various controllers and cache regions in storage subsystem. This method achieves the general objective of the commonly-used (S)tripe (A)nd (M)irror (E)verything... the so-called SAME strategy.

IBM Spectrum Scale provides the option to set block size for each file system individually and the Oracle-specific recommendations are:

- 512KB is generally suggested.
- 1MB is suggested for file systems that are 100TB or larger.

The mount options to suppress atime (-S) and mtime (-E) on data file systems may be helpful in reducing overhead for the file system management and increasing performance. One should understand whether any operating system utilities like backup software uses file modification time and therefore should not be suppressed.

To suppress atime and mtime (either or both), the parameters are set as follows:

- mmchfs <device> -E no Disable exact mtime tracking
- mmchfs <device> -S yes Suppress atime tracking

For IBM Spectrum Scale versions earlier than 4.2.3, I/O thread tuning parameters are recommended to be initially set as follows:

- prefetchThreads=150
- worker1Threads=450

For IBM Spectrum Scale versions 4.2.3, 5.0 and later, the I/O thread tuning is controlled by a single parameter, workerThreads. The recommended initial value should be set as follows:

- workerThreads=512 (or 1024)

IBM Spectrum Scale Resilience and Availability

Quorum:
Availability of the IBM Spectrum Scale cluster is paramount for production or mission-critical databases. As such, the parameter 'minQuorumNodes' may be set to decrease the possibility of losing cluster quorum and incurring unplanned downtime. Quorum loss or loss of connectivity occurs if a node goes down or becomes isolated from its peers by a network failure. Quorum is typically defined as one + half of the explicitly defined quorum nodes in the Spectrum Scale cluster.

In small clusters it may be desirable to have the IBM Spectrum Scale cluster remain online with only one surviving node. *In that case, tiebreaker disks must be used*. The following parameter values are an example of this configuration option (the names of the tiebreaker disks will be different of course):

- minQuorumNodes=1
- tiebreakerDisks="tiebreakerdisk1;tiebreakerdisk2;tiebreakerdisk3"

GPFS administration and file manager network:
As previously stated, availability of the IBM Spectrum Scale cluster network is an important consideration for production or mission-critical environments. As such, the GPFS cluster network may be protected using link aggregation methods such as IEEE 802.3ad or Etherchannel.

The GPFS network has modest bandwidth requirements as it does not transfer large sets of data from node to node. Although not a hard requirement, the administration and file manager network may be dedicated as it is in certification tests.

Storage failure detection:
IBM Spectrum Scale makes use of storage subsystems that employ SCSI-3 persistent reservations to control multi-node access to the shared storage. Failover times can be significantly reduced when this parameter is enabled in the file system cluster. IBM tests and certifies storage subsystems for use of this feature. Please see the IBM Spectrum Scale FAQ 4.1 "What disk hardware has IBM Spectrum Scale been tested with?" to confirm the currently supported storage subsystems and further considerations for implementation. To enable this feature, the following parameters should be set:

- usePersistentReserve=yes
- failureDetectionTime=10

For a device to properly offer SCSI-3 Persistent Reservation support for GPFS, it must support SCSI-3 PERSISTENT RESERVE IN with a service action of REPORT CAPABILITIES. The REPORT CAPABILITIES must indicate support for a reservation type of Write Exclusive All Registrants. Contact the disk system vendor to verify these capabilities are provided.

Notes:

- Only a subset of releases are certified for use in Oracle environments. To confirm the certified versions login to Oracle support (https://support.oracle.com/) and search on the certify tab for Product IBM Spectrum Scale and note target version to be used.
- For AIX go to IBM Spectrum Scale and Oracle RDBMS RAC (Doc ID 2587696.1)
- For Linux go to, http://www.oracle.com/technetwork/database/clustering/tech-generic-linux-new-086754.html

- Oracle certification is for storing RDBMS files in the IBM Spectrum Scale direct access model. Configuring an Oracle database to access through Protocol Nodes (NFS, SMB) is not certified.
- There is no plan to certify Oracle DB versions prior to 19c on IBM Spectrum Scale 5.1 as those versions are out of support.
- There are no currently supported levels of IBM Spectrum Scale qualified with Linux on Power.
- Oracle has not been certified with IBM Spectrum Scale on Linux on Intel and there are no current plans to do so.
- For the list of virtualization and partitioning technologies supported by Oracle, go to http://www.oracle.com/technetwork/database/virtualizationmatrix-172995.html

12.1 with AIX Requirements (Oracle MOS Doc id 1587357.1)

(Oracle Database (RDBMS) on Unix AIX,HP-UX, Linux, Solaris and MS Windows Operating Systems Installation and Configuration Requirements Quick Reference (12.1) (Doc ID 1587357.1)

AIX

OS Version	Patches/Packages	Kernel settings
AIX 6.1 Technology Level 7 Service Pack 3 ("6100-07-03-1207") or later, 64-bit kernel **Note:** You can install on AIX 6.1 Technology Level 7, but Oracle recommends that you install on AIX 6.1 Technology Level 9 or later. AIX 7.1 Technology Level 1 Service Pack 3 ("7100-01-03-1207") or later, 64-bit kernel **Note:** You can install on AIX 7.1 Technology Level 1, but Oracle recommends that you install on AIX 7.1 Technology Level 3 or later.	bos.adt.base bos.adt.lib bos.adt.libm bos.perf.libperfstat bos.perf.perfstat bos.perf.proctools xlC.aix61.rte:11.1.0.4 or later xlC.rte.11.1.0.4 or later If you are using the minimum operating system TL level for AIX 6.1 listed above, then install all the following AIX APAR fixes: IV16716, IV20880, IV21128, IV28319, IV30712, IV33433, IV34685, IV39104, IV45072, IV45073 If you are using the minimum operating system TL level for AIX 7.1 listed above, then install all the following AIX APAR fixes: IV16737, IV21116, IV21235, IV28925, IV34869, IV35057, IV39136, IV41415, IV45072, IV45073 **Note: APAR number may vary according to TL / SP level, please contact IBM for exact APAR number for your OS TL/SP version.** **For Oracle C++,** **Oracle C++ Call Interface,** **Pro*C/C++,** **Oracle XML Developer's Kit (XDK) –** Install IBM XL C/C++ Enterprise Edition for AIX, V11.1 (11.1.0.9) January 2012 PTF. IBM XL C++ Runtime for AIX, V11.1 (11.1.0.4) November 2011. **For Pro*COBOL** - IBM COBOL for AIX Version 4.1.1 (March 2012 PTF), Micro Focus Server Express 5.1 **For Pro*FORTRAN** - IBM XL Fortran Runtime for AIX, Version 13.1, January 2012 PTF **For ADA** - OC Systems PowerAda 5.5	Set the port range high enough to avoid reserved ports for any applications you may intend to use. If the lower value of the range you have is greater than 9000, and the range is large enough for your anticipated workload, then you can ignore OUI warnings regarding the ephemeral port range. tcp_ephemeral_low = 32768 tcp_ephemeral_high = 65535 udp_ephemeral_low = 32768 udp_ephemeral_high = 65535 **Shell limits** Soft File Descriptors at least 1024 KB Hard File Descriptors at least 65536 KB Soft FILE size -1 (Unlimited) Soft CPU time -1 (Unlimited) Soft DATA segment -1 (Unlimited) Soft STACK size -1 (Unlimited) Soft Real Memory size -1 (Unlimited) Processes (per user) -1 (Unlimited) Note: This limit is available only in AIX 6.1 or later. maxuprocs 16384 ncargs 128 **IOCP :** On IBM AIX on POWER Systems (64-Bit), enable I/O completion ports (IOCP) to ensure successful database and grid infrastructure installation. Use **smitty iocp** to change the characteristics to "**Available**". After modifying this parameter **reboot** the system.

RAC Tips

BUG 17208793 - KERNEL DUMP WHEN CRS NEEDS TO REBOOT A CLUSTER NODE.

Node Evictions during LPAR Migration and SVC Upgrade

Oracle RAC Support typically will suggest the css_miscount not be changed from the default of 30 sec (the Oracle documented setting). As in the case of evictions during an LPAR migration or SVC upgrade, if the timeout value exceeds 30 seconds an eviction will occur. Setting the miscount timeout value to 60 seconds temporarily can help overcome this by adjusting the timeout value of the process to temporarily

increase it and set it back to default of 30 seconds. In one situation, the LPAR migration took 35 seconds vs the timeout of 30 seconds and this helped overcome node evictions and permit LPAR's to be migrated.

This solution is as follows:
1) Before performing the activity set the miscount to 60
2) Perform the action this
3) After the activity is complete -- reset the miscount back to 30

(Note: Default in Oracle Exadata for the miscount timer is 60 seconds)

Reference MOS Doc: CSS Timeout Computation in Oracle Clusterware (Doc ID 294430.1)

In general, Oracle Clusterware accesses a Voting File (on disk) every second for read and write with less than a kilobyte of data read or written. An acknowledgement of the write I/O needs to be received in 200 seconds under normal operations
(long disk timeout) and 27 seconds during a reconfiguration in the cluster (short disk timeout).

The cluster interconnect (network) always needs to respond in 27 seconds.

The Short Disk Timeout only comes into play when there is a RAC node reconfiguration going on. A reconfiguration means that either a node is leaving or joining the cluster. If there was a network timeout causing a reconfiguration then the short disk timeout would come into play and that could cause a storage timeout..... But without a reconfiguration the storage timeout is 200 seconds.

Oracle RAC General Reference:

Oracle Clusterware (CRS) operates on two heartbeat mechanisms

Network heartbeat - allows communication between the RAC nodes (via cluster interconnect) and to maintain cluster membership

Voting Disk heartbeat - Both of these heartbeats have an associated timeout value. Network heartbeat has the timeout known as Miscount while Voting Disk heartbeat timeout is represented by DiskTimeOut.

The values of these timeouts vary by versions and platforms and are well documented under Metalink Note 294430.1. These timeouts govern the maximum time (in seconds) which the network heartbeat can be missed or I/O to the voting disks has to complete.

There is also a third timeout parameter, Short DiskTimeOut (SDTO), which is only enforced during cluster formation or reconfiguration (node leaving or joining the cluster).

What is NETWORK and DISK HEARTBEAT and how it registers in VOTING DISKS/FILES

1. All nodes in the RAC cluster register their heartbeat information in the voting disks/files. RAC heartbeat is the polling mechanism that is sent over the cluster interconnect to ensure all RAC
a. Nodes are available.
b. Voting disks/files are just like attendance register where you have nodes mark their attendance (heartbeats).

2. CSSD process on every node makes entries in the voting disk to ascertain the membership of the node. While marking their own presence, all the nodes also register the information about their communicability with other nodes in the voting disk. This is called NETWORK HEARTBEAT.

3. CSSD process in each RAC maintains the heart beat in a block of size 1 OS block in the hot block of voting disk at a specific offset. The written block has a header area with the node name. The heartbeat counter increments every second on every write call. Thus heartbeat of various nodes is recorded at different offsets in the voting disk. This process is called DISK HEARTBEAT.

4. In addition of maintaining its own disk block, CSSD processes also monitors the disk block maintained by the CSSD processes of other nodes in cluster. Healthy nodes will have continuous network & disk heartbeats exchanged between the nodes. Break in heartbeats indicates a possible error scenario.

5. If the disk is not updated in a short timeout period, the node is considered unhealthy and may be rebooted to protect the database. In this case, a message to this effect is written in the KILL BLOCK of node. Each nodes reads its KILL BLOCK once per second, if the kill block is not overwritten, node commits suicide.

6. During reconfig (leaving or joining), CSSD monitors all nodes heartbeat information and determines whether the nodes has a disk heartbeat including those with no network heartbeat. If no disk heartbeat is detected, then node is considered as dead.

CSS NETWORK HEARTBEAT

There are 3 threads in CSS that deal with CSS heartbeats:

SendingThread - This thread periodically wakes up and sends appropriate packets (based on the join state) to other nodes so that other members know we are still alive.

ClusterListener Thread - This thread listens for incoming packets from other nodes by calling clscselect and dispatches them for appropriate handling. When the ClusterListener receives a packet, it calls an internal function to process the packet. The packet info is put in memory to let it get seen by the polling thread so it can decide if any action is needed.

Polling Thread - Periodically wakes up and scans to see who is active and has been checking in regularly by reading from the node db in memory. If the last packet time in memory reaches 1/2 of miscount you will see this famous message:

2016-12-11 22:12:55.954:
[cssd(3277434)]CRS-1612:Network communication with node systemnode01 (1) missing for 50% of timeout interval. Removal of this node from cluster in 14.111 seconds

VIOS

Interconnect:

- A dedicated 10G Ethernet Adapter connection is recommended as minimum to provide sufficient bandwidth for cluster timing sensitive traffic

- RAC cluster traffic - interconnect traffic should be dedicated and not shared. Sharing of interconnect can cause timing delays leading to node hang/eviction issues

Virtual Ethernet:

If you choose to use virtual ethernet for the RAC cluster interconnect for its convenience or to allow use of Live Partition Mobility (LPM) please include the following configuration on the VIO client RAC nodes.

When using virtual ethernet for the Oracle RAC public network or cluster interconnect the attribute poll_uplink for the AIX ethernet device (ent) on the client LPAR should be set to "yes". This allows the status of the physical link backing the virtual ethernet to be reflected in the IFF_DEVHEALTH status for the virtual ethernet and allows Oracle VIP and HAIP addresses to failover quicker in the case of a failure of the physical link. The option can be changed with the AIX chdev command, as shown in the following example. The attribute should be set with the "-P" option of chdev to make the change permanent.

Example for change poll_uplink for ent2:

If the device is not busy you can change the attribute dynamically using the following command. If the device is busy the change will take effect after setting the option with the use of the "-P" option and rebooting.

- chdev -l ent2 -a poll_uplink=yes

The following command should be used to set the attribute permanently.

- chdev -l ent2 -P -a poll_uplink=yes

Olson Timezone

The default timezone format for AIX 6.1 and AIX 7 is Olson Time. POSIX was the traditional format and previous older versions of AIX used POSIX and evolved to using the Olson format which is the current standard.

A significant performance regression was introduced in the time related APIs (IV87788 starting in AIX 7.1.3.4) when the Olson timezone format is used (ex: "TZ=America/Sau_Paulo"). Calls to time functions such as tzset(), mktime(), and localtime() may take significantly longer to complete, which can compound and have an impact on application performance, depending on how the application uses these function calls. Investigating and suspecting this may be causing a performance issue leading to a node hang and also node evictions.

Summary:
Timezone function tzset() was calling stat() on every call to check for updates to the timezone rules file. This stat() call is costly and was impacting the performance of all timezone functions. Added a limit to only call stat() a maximum of once every minute. Workaround is using POSIX timezone format.

ORA-0600 Errors / Core Dump

IBM HIPER APAR - ORA 600 ERRORS AND ORACLE CORE DUMPS AFTER AIX SP UPGRADE

Reported only on 11gR2 - 11.2.0.4

Oracle Alert (MOS):
ALERT: Database Corruption ORA-600 ORA-7445 errors after applying AIX SP patches - AIX 6.1.9.7 to SP08 or AIX 7.1.4.2 to SP03(Doc ID 2237498.1)

IBM Notice:
http://www-01.ibm.com/support/docview.wss?uid=isg1SSRVPOAIX71HIPER170303-1247

PROBLEM SUMMARY:
The thread_cputime or thread_cputime_fast interfaces can cause invalid data in the FP/VMX/VSX registers if the thread page faults in this function. APAR's IV93840, IV93884, IV93845, IV93883, IV93885

AFFECTED LEVELS and FIXES

Affected AIX Levels	Fixed In	iFix / APAR (ftp://aix.software.ibm.com/aix/ifixes)/
6100-09-08	6100-09-09	IV93840
7100-03-08	7100-03-09	IV93884
7100-04-03	7100-04-04	IV93845
7200-00-03	7200-00-04	IV93883
7200-01-01	7200-01-02	IV93885

DataPump Suggestions

Data Pump technology fully uses all available resources to maximize throughput and minimize elapsed job time. For this to happen, a system must be well-balanced across CPU, memory, and I/O. In addition, standard performance tuning principles apply. For example, for maximum performance you should ensure that the files that are members of a dump file set reside on separate disks, because the dump files will be written and read in parallel. Also, the disks should not be the same ones on which the source or target tablespaces reside. Any performance tuning activity involves making trade-offs between performance and resource consumption.
If your database is running in a shared CPU LPAR with CPU folding active, the default, ensure that schedo parameter vpm_xvcpus is set to a value of 2 as otherwise AIX could be too aggressive folding VPs.

Memory Consumption with AIX 7.2

APAR IJ17766 – LDEDIT MISSING OPTION LIST IN THE USAGE OUTPUT
https://www-01.ibm.com/support/docview.wss?uid=isg1IJ17766

A potential new issue that may affect memory usage identified starting with AIX 7200-03-02. The linker has an option "shrsymtab" which Oracle uses so that all Oracle processes use a shared symbol table. This significantly reduces the memory footprint used by the Oracle processes. The Oracle binary is linked at install time, and the make file used to link Oracle uses output from the 'ldedit' command to determine if the system Oracle is being installed on supports the shrsymtbl option. Specifically the Oracle install process looks for the "shrsymtab" string in the output (see output listed below). At some point AIX changed the output returned by `ldedit` to not include that string, even when shrsymtbl option is supported, leading to Oracle processes no longer using the shared symbol table.

The ldedit command doesn't show the option list in the usage output # ldedit ldedit:
Usage: ldedit -b option [-V] File while with previous levels the output was # ldedit ldedit: Usage: ldedit -b option [-V] File ldedit: Options are: [no]lpdata maxdata[/dsa] maxstack [no]rwexec textpsize datapsize [no]shrsymtab stackpsize [no]forkpolicy:[cow|cor].

This output is used by Oracle to determine if shrsymtab feature was available, and since AIX 7.2 TL3 SP2 doesn't show it, Oracle will not use it, resulting in more memory being used. **APAR IJ17766** has been opened to address it and have ldedit return the "old" output.

Interim Workaround:
In the meanwhile you can do the following to workaround the issue:
1. Get the dump output of Oracle.
dump -X 64 -Hov <oracle_exec>
2. The flags section should have:
Flags
0x00/0x8000
Flags=(SHRSYMTAB)
3. If not set, then set it as follows:
ldedit -b shrsymtab <oracle_exec>
ldedit: File oracle updated.
4. restart the executable

APAR IJ17973 applies to AIX 7.2 TL4 LDEDIT MISSING OPTION LIST IN THE USAGE OUTPUT
Note: You may check with Oracle support to know which are the executables that use shrsymtab feature in order to know which ones have to be changed so that they'll use it.

Openssh GRID Install Issue – Oracle MOS Doc ID 30159782.8
There is a potential issue with a Oracle GRID install issue when installing on a system with the newer version(s) of ssh. This issue is documented in MOS Oracle Bug 30159782 (Generic) - Remote Copy Fails if using openssh - 7.2. 7.4, etc.

Description: During a GI installation a remote copy may fails with OpenSSH 7.2, 7.4 etc.

The crsconfig log file shows that cluutil command fails:
2019-08-05 12:29:16: oracle.ops.mgmt.cluster.ClusterException: protocol error:
filename does not match request: failed

The error can be reproduced including a path with both double and single quotes (i.e., "'<path>'"):
$ /opt/openssh8_0_p1/bin/scp -p "<MACHINE>:'/tmp/t'" "/tmp/tt"

Workaround: Downgrade the openssh package

The following version of ssh, and newer have exposed this problem:

```
Version shown by `ssh -V` command:
   oracle@rac154:/home/oracle=> ssh -V
   OpenSSH_7.5p1, OpenSSL 1.0.2p  14 Aug 2018

AIX fileset versions:
     openssh.base.client   7.5.102.1801  COMMITTED  Open Secure Shell
   Commands
     openssh.base.server   7.5.102.1801  COMMITTED  Open Secure Shell Server
     openssh.license       7.5.102.1801  COMMITTED  Open Secure Shell License
  openssh.man.en_US     7.5.102.1801  COMMITTED  Open Secure Shell

Oracle versions confirmed as being affected:
    19.1.0
    18.1.0
    12.2.0.1 (Base Release)

The fix for 30159782 is first included in Oracle:
    20.1.0
    19.6.0.0.200114 (Jan 2020) OCW Release Update Revision(OCW RU)
    OCW Release Update 18.9.0.0.200114 (Jan 2020)
    OCW Release Update 12.2.0.1.200114 (Jan 2020)
```

Additional information in MOS note for details and work around: INS-06006 GI RunInstaller Fails If OpenSSH Is Upgraded to 8.x (Doc ID 2555697.1)