FUJ00117489

V2.0

ORACLE

Root Cause Analysis

for Customer:

Fujitsu Services

on

SR No: 3-1610271271: Receiver is waiting for a latch dumping latch state for receiver (LCK0 issues)

Author: *Richard Watt*

Contributors:

John Cahill

Adrian Turner

Date: 5th July 2010

FUJ00117489

ORACLE

Contents

FUJ00117489

Executive Summary

Fujitsu Services are contracted with the Post Office to replace the current infrastructure used in 12000 branches across the UK. The project involves rolling out a (non Oracle) financial application using Oracle RAC database and tools products. Oracle Streams is used to provide a reporting database environment.

The migration onto the new platform started at the end of 2009 and is currently rolled out live to 600 branches.

Post Offices taking part in the pilot have been complaining about intermittent hangs which are having huge impact on over the counter services as the system hangs for 5 minutes or longer and users are unable to complete their financial transactions.

Fujitsu have been unable to reproduce the hangs in any test environment despite numerous attempts to inject the system with heavy load to simulate the production situation.
The focus of this escalation is to resolve the system wide hangs.

This document outlines the investigations which have been carried out by Oracle support and development to establish the root cause for these hangs and make recommendations on how to resolve the issue as per Service Request 3-1610271271.

FUJ00117489

ORACLE

Root Cause Analysis

The root cause of the hang was a defect in Oracle code that meant the LCK process did not clear the memory up efficiently.

The application's extensive usage of ddl to truncate database partitions has exposed the Oracle defect . Altering the ddl usage will drastically reduce the chance of the problem occurring. However, the Oracle defect would have still required resolution longer term.

This usage causes heavy growth of the database's shared memory pool. The hang is caused by lck inefficiently attempting to clear down this pool due to excessive growth in the "mvobj part des" region. The root cause of the lck process's behaviour is an Oracle defect. This has been resolved by an official patch.

FUJ00117489

Corrective Actions

The two defects (5618049 and 8528171) attribute growth in this region to large amounts of partition DDL taking place. These have been applied on Thursday 20th May.

Improved shared memory usage after patch application has been observed through analysis of data, this has shown the average free memory improve and the "mvobj part des" allocations reduced to 750mb compared to 2GB prior to patch application but lck issue still occurred on Saturday 22nd May.

Defect 7306915 resolves a memory leakage issue, this is proved to be occurring in Fujitsu's environment after analysis by support. This was applied in production on 30th May and proved to be successful.

Further application changes required to reduce ddl operations have been implemented on the 6th June by Fujitsu.
These are critical and a key action to resolving the critical business impact of this issue as they will reduce shared memory usage. It is not possible to completely remove all ddl operations and we will be focused on continuing to resolve issues in this area to ensure long term stability of the service.

Oracle have identified an issue with multi-versioned objects acquiring instance locks and are working on a fix to free this object in RAC rather than being queued up by LCK0. An official patch for 10.2.0.4.3 and 10.2.0.4.4 was provided on Tuesday 8th June in the form of a revised version of the previous fix for issue 8528171.

Oracle and Fujitsu are both able to reproduce the issue although not to the full extent of the live occurrences. This was sufficient to provide Oracle's Engineering team with a high level of confidence that the fixes will resolve the service hang issue. However due to the complexity of the environment Oracle are unable to 100% guarantee that no further issues will occur, based on experience with hundreds of other customers and the diagnostic information gathered, we believe this is the correct technical plan to lead to resolution of the current hangs being experienced by the Post Office.

Since installation in production environment no further issues have been reported. All corrective actions are now in place and post office roll-out is continuing as scheduled.

Formatted: Font: 11 pt, Font color: Auto

Formatted: Font: 11 pt, Font color: Auto

Formatted: Indent: Left: 0"

Formatted: Font: 11 pt, Font color: Auto, Not Superscript/ Subscript

Formatted: Superscript

FUJ00117489

Mitigation

The problem is shown to be related to the time the database instances are running. Regular re-cycling of the database instances is a proven method to prevent shared memory fragmentation.

An alternative approach of flushing shared memory has been recommended to Fujitsu as an alternative. This is documented in Service Request 3-1610271271.
This recommendation was reviewed and deemed not appropriate due to scheduling constraints.

FUJ00117489

Recommendations

By following the actions plan below we believe the issue will be significantly resolved. This is however, a complex issue and ongoing investigation may lead to other recommendations.

SR No: 3-1610271271: Receiver is waiting for a latch dumping latch state for receiver

1. Set _object_statistics=false. This is to enable the fix for defect 6001617 which is a common LCK0 CPU/load fix. Overall LCK0 load will be reduced with this parameter setting. (Completed - 14/05/2010)

2. Set event 14532 at level 1 to enable the fix for defect 5618049. This fix frees up "mvobj part des" heaps faster. (Completed – 20/05/2010)

3. Apply fix for defect 8528171 (provided in patch 9668554). This fix specifically addresses over allocations of "mvobj part des" heaps and reduces the number of library cache handles that LCK0 would need to free. (Completed - 20/05/2010)

4. Shared pool/library cache tuning activity, being led by Adrian Turner from Oracle consulting (scheduled for 6th June – Completed 06/06/2010, RA1005.6a – Partition Working Tables, RA1005.6b Aud$ Table and RA2605.2 Resource Manager)

Formatted: Font: 11 pt

5. Apply patch for defect 7306915. Delivered to customer on Friday 21st May as merge patch 9734573. This is aimed at further improving shared pool usage after reoccurrence of LCK issue on 22nd May.. (Completed - 30/05/2010)

This contains the following fixes:-
9668554 (applied 20th May, as discussed in point 3 above) + 7306915 (fix for 7445 [kglhdgn()+1101]).

6. Oracle to deliver revised patch for defect 8528171 to free multi version objects. Completed 6th June.

FUJ00117489

Additional Recommendations:

Following the changes outlined above a period of close monitoring is recommended. For other application related recommendations, see onsite report provided by Adrian Turner.
See attached document below:-

Site Visit Report 20100514 FINAL (3).

Oracle would also recommend that further investment is made into the Load Testing environment, so that we can be confident that the architecture will support the additional load of further branches being rolled out. Oracle have offered to be involved in reviewing the testing strategy in further detail to ensure adequate load testing is carried out.

Use of Oracle Real Application Testing (RAT) should be considered – which combines a workload capture and replay feature with an SQL performance analyzer to help you test changes against real-life workloads, and then helps you fine-tune the changes before putting them into production.

The database replay feature would require upgrade of the DB to 11g. RAT also includes SQL Performance Analyser which is available on 10.2 and allows you to analyse the entire SQL workload for performance degradation in a fraction of the time it would take to do manually

To mitigate the high CPU run queues it is recommended that Fujitsu implement Database Resource Manager. Resource Manager only steps in when the server is 100% busy and throttles sessions, making them wait on "RESMGR: CPU QUANTUM" event. Under CPU starvation this will have the added benefit of preventing the database from impacting Clusterware and operating system processes. There is a known defect which must be reviewed prior to implementation (Oracle reference 4147766).

FUJ00117489

Appendix: Defect descriptions

The following gives technical detail on the underlying defects and what they are designed to fix:

Defect 8528171

PROBLEM DESCRIPTION:

Customer is doing thousands of add/exchange partition operations per day.
These operations cause too much shared pool memory to be allocated
for partition descriptor objects (mvobj part des). Because of this problem,
customer is having several issues on shared pool related operations.

These objects should be cleaned up more quickly when they are no longer
needed.
Typically in the AWR report, "mvobj part des" objects are
the largest portion of the shared pool.

WORKAROUND:

Restarting instance or flushing shared pool clears up the objects but these workarounds are not acceptable for the customer. Because they have a 24x7 system and flushing shared pool takes hours to complete.

FIX DESCRIPTION

This problem is addressed as follows:

1) Obsolete and invalid KGL objects are placed on the cold end of the KGH lru so they are freed faster.
2) Extent size for partition operations has been increased from 1k to 4k, this decreases fragmentation.
   3) Multi-versioned objects now do not acquire any instance locks, obsoletion is instead done manually in kkpox. This allows for these objects to be freed more quickly in RAC as they can be freed immediately rather than being queued up by LCK0.

Defect 6001617

PROBLEM DESCRIPTION:

LCK0 consumes CPU Spining on ksrbwait reading an ever increasing list of messages in "obj stat del channel" and holding "channel operations parent latch". ORA-4031's are also possible as messages are stored in the shared pool.

FIX DESCRIPTION:

This fix is activated by setting STATISTICS_LEVEL = BASIC
(or by setting "_object_statistics"=false) in the init.ora.
Setting this means the space advisory will not be available
for use as there will be no statistics for it to measure.

FUJ00117489

Defect 5618049

PROBLEM DESCRIPTION:

Db Leaking memory allocated for "mvobj part description" -- the versioned partition descriptor. It seems that this happens for both ADD PARTITION and DROP PARTITION (probably for other PMOs as well), and is not IOT specific.

kgh was unable to free memory because of a latching issue. In this case we had a number of versioned objects and were unable to free them.

The basic problem is that kgl is sometimes unable to free handles when called from kgh. This happens when we free one handle, realize this allows us to free a new object (since the ref count is now zero), but find this new object is protected by a second latch. Since we already hold a latch (to free the first handle) we are unable to free the second one.

FIX DESCRIPTION:

The fix tries to be proactive about freeing memory and will be enabled under an event 14532. The 10.2 fix tries to free memory proactively.

Defect 7306915
PROBLEM DESCRIPTION

ORA-4031 / KGL HANDLES MEMORY LEAK OF MVOBIND / MVOBTBL FROM PARTITION DDL
Repeated EXCHANGE PARTITION against a partitioned table with local partitioned indexes can cause a number of obsolete (OBS) unfreeable MVOBIND library cache handles in the library cache.
ie: ALTER SYSTEM FLUSH SHARED POOL will not flush the library cache entries. For high numbers of exchanges this can lead to shared pool pressure and subsequent ORA-4031 errors.

FIX DESCRIPTION
Merge patch 9668554 has been provided to customer to resolve leaks. Workaround is to reduce the number of exchange partition operations.