IBM Education Assistance for z/OS V2R1

Item: Serial Rebuild with Structure Priority
Element/Component: BCP XCF

Material is current as of June 2013

Agenda

- Trademarks
- Presentation Objectives
- Overview
- Usage & Invocation
- Interactions & Dependencies
- Migration & Coexistence Considerations
- Installation
- Presentation Summary
- Appendix

Trademarks

- See url http://www.ibm.com/legal/copytrade.shtml for a list of trademarks.

Presentation Objectives

- Demonstrate the benefit of using Serial Rebuild Processing for Coupling Facility Events:
  - CF Failure / Loss of Connectivity Events
  - CF Gain Connectivity Events
  - Policy Initiated Start Duplexing Events

- Describe how to Enable/Disable Serial Rebuild Processing

Overview

- Problem Statement:
  - Unplanned events such as CF Failure and Loss of Connectivity will result in rebuilds and break-duplexing and reduplexing that are processed in parallel.
  - When many CF structures (up to 2048) are rebuilt at the same time, in parallel, performance problems are observed.

  - CF(s) overwhelmed updating new structures
  - Systems overwhelmed
    - XCF coordinating CF structure rebuild processing
    - Users managing CF structure rebuild processing

Overview

- Problem Statement:
  - Performance problems affect *availability* and result in other “side effects” for “hangs”, including ABEND026 for CFSTRHANGTIME
  - Applications will be unavailable during CF structure rebuild processing
    - Structures interferes with each other, causing system functions that depend on those structures to be unavailable for a long period of time.
    - Unimportant structures compete for CFRM attention with important ones.

Overview

- Solution:
  - Perform system initiated structure rebuilds in a more serial manner
    - Perform these actions in a policy controlled priority order which will insure the most important structures are rebuilt quickly
  - Perform policy initiated and CF Gain connectivity events in a strictly serial manner
- Benefit / Value:
  - Ensure the disruption from the CF Structure rebuilds and break duplexing activity is as small as possible.
  - Ensure that the most important structures in the installation's workload are available quickly.

Overview

- Background – CFRM Active Policy contention is generated when *many* CF structures are rebuilt at the same time. Couple data set data transfer (i.e. I/O) time can be significant which adds to the contention.
- CF structure rebuild processing has phases
  - Phase transitions generate *events. Confirmations* transition it to the next phase
  - Duplexing failover and user-managed rebuild each have multiple phases (Stop | Quiesce-Stop & Stop | Switch & Cleanup), (Quiesce, Connect, & Cleanup) resulting in multiple contention points, which multiply for hundreds of structures
- With current max of 2,048 structures and 255 connectors (per structure), active policy data may be around 192M:
  - 18 seconds to read the entire active policy
  - 45 seconds to write the entire active policy

Overview

- Background – Improvements
- CFRM Event and Confirmation Optimization
  - Improve performance through the *reduction* of CFRM active policy contention points
    - When servicing a request, also process other requests of the same type while continuously holding the CFRM active policy lock
    - Merge #N contention points for #N structures to 1 contention point for #N structures
- Message-Based Event and Confirmation Processing
  - Improve performance through the *elimination* of CFRM active policy contention points
    - Assign a single manager system to process acknowledgements and confirmations

Overview

- Background – Comparing Message-Based Event and Confirmation Processing vs. Policy Based Event Processing
- Test Environment - VICOM test
  - 4 z/OS systems, 2 coupling facilities, 101 structures
- Policy Based
  - 101 Structures need recovery at Time 0
  - 12 Structures completed duplex-failover after 19 seconds
  - 94 structures completed after 116 seconds
  - All complete after 127 seconds
- Message Based
  - 101 Structures need recovery at Time 0
  - 11 Structures completed duplex-failover after 42 seconds
  - 94 structures completed after 108 seconds
  - All complete after 120 seconds

Overview

- Background – Comparing Message-Based Event and Confirmation Processing vs. Policy Based Event Processing
- Total recovery time for message-based compared to policy-based was closer than expected.
  - Message-based still has the advantage of no contention, but it has a disadvantage that the acknowledgement I/O is still not optimized as it is when using policy-based.
- Duplexing failover takes longer when using message-based processing!
  - Simplex structures get to the connect phase much faster. The CFRM activity done when connecting to the rebuild new structure is interfering with duplexing failover.
- For each type of recovery (duplexing failover or rebuild), average recovery time is approximately equal to the total recovery time
- They all recover at about the same time!!

Overview

- Solution – CFRM Serial Rebuild Processing
  - Perform system initiated structure rebuilds for CF Failure/ Loss of connectivity in a more serial 'pipeline' manner
    - Perform these actions in a policy controlled priority order which will insure the most important structures are rebuilt quickly
    - Provide relief for “floods” of CF structure rebuild and start-duplexing activity
  - Perform policy initiated and CF Gain connectivity events in a strictly serial manner
  - Improve average performance by reducing *interference* among CF structure users

Overview

- Enhancements to the following CF structure rebuild processing triggers will be made:
  - CF Failure / Loss of Connectivity Events
    - System loss of connectivity to a coupling facility (which includes a CF failure)
  - CF Gain Connectivity Events
  - Policy Initiated Start Duplexing Events
    - Policy initiated CF structure duplexing actions (which includes duplexing a DUPLEX(ENABLED) structure after duplexing failover for a CF failure)

- This new support is collectively called CFLossConnRecoveryManagement or CFLCRMGMT

Usage & Invocation

- Message-based processing is enhanced with the ability to perform CFRM LOSSCONN recovery management

- With CFRM LOSSCONN recovery management, the message-based manager system is responsible for orchestrating and initiating processing

  – to recover from a loss of connectivity to a CF and all that processing is initiated in a more serial manner.
  – to process a gain of connectivity to a CF or start duplexing and all that processing is initiated in a serial manner.

Usage & Invocation

- When this function is enabled on all active systems in the sysplex, the message based manager system may enable CFRM LOSSCONN Recovery Management.
  - Message-based processing enablement (available on z/OS V1.8) is required to enable CFRM LossConn Recovery Management.

- CFLCRMGMT Function switch control enablement/disablement
  - SETXCF FUNCTIONS, ENABLE= CFLCRMGMT
  - SETXCF FUNCTIONS, DISABLE= CFLCRMGMT

  - COUPLExx parmlib member
    - FUNCTIONS ENABLE(CFLCRMGMT)

Interactions & Dependencies

- CFRM Serial Rebuild Processing
- When a loss of connectivity to a CF , CFRM will enter a new "state" of LOSSCONN RECOVERY IN PROGRESS.
- In this state, the system will "suspend" any in-progress REALLOCATE or POPULATECF processing and defer policy-initiated start/stop duplexing actions.
  - CFRM will defer the “less important” processing until the “more important” processing completes
  - The message-based manager system will keep track of structure state in order to take responsibility for "scheduling" recovery processing.
    - For each structure the manager system will use the state to decide the action that needs to be taken.
    - Using the state and desired action for each structure, the manager system will decide on the next structure to process.

Interactions & Dependencies

- CFRM Serial Rebuild Processing
- CFRM will perform recovery actions for structures in priority order.
- Within that ordering, structures will be processed in the following order of importance in order to achieve “closest to recovering” ordering:
  - Rebuild and duplexing failover (stop a duplexing rebuild)
  - Recommend disconnect (for REBUILDPERCENT with SFM)
  - Rebuild (start a user-managed rebuild)
- Within each recovery action, lock structures will be processed before list and cache structure.

Usage & Invocation

- CFRM Policy Information – defining the priority
  - The STRUCTURE definition statements have been updated to include a optional RECPRTY statement

  - This statement specifies the rebuild priority to be given to the structure for CFLCRMGMT when
    - The system loses connectivity to a coupling facility
    - 1= high priority 4 = low priority (1 digit decimal value).

  - When RECPRTY is not specified, the system takes a default:
    - RECPRTY(3) medium - All structures
    - Lock structures will be processed before list and cache structure.
    - RECPRTY does not apply to XCF signalling structures.
    - RECPRTY takes effect immediately with policy activation.

Interactions & Dependencies

- CF Serial Rebuild Challenges
- We want to allow structures that may “make progress” to “move ahead” of those that may not.
  - But allowing too much “move ahead” just gets us back to processing in parallel.
  - Allowing not enough “move ahead” causes recovery time to suffer for these dependencies.
  - Processing structures in a *pipeline* fashion allows some parallelism, with some optimization in phase transition
- What is the magic number.
  - Processing 2 structures at a time yields better average/total times than when processing 1 structure at a time.
  - But processing all structures at the same time has worse results.

Interactions & Dependencies

- CFRM Serial Rebuild Processing

- Policy initiated duplexing actions include the following:
  - Start a duplexing rebuild for structures with a policy specification of DUPLEX(ENABLED) as a result of starting a policy or gaining CF connectivity.
  - Stop a duplexing rebuild in response to a CFRM policy change. For example, changing the DUPLEX policy specification to DUPLEX(DISABLED).

- Policy initiated actions to start duplexing will occur in an order determined by the system, in alphabetical order. Actions to stop duplexing will be taken before actions to start duplexing so that CF space used by a structure will not prevent duplexing for another structure.

Interactions & Dependencies

- CFRM Serial Rebuild Challenges
- Processing in a serial manner loses the *advantages* of processing in parallel, including advantages of previous I/O and contention optimization
  - Some internal optimization to generalize and process CFRM requests as a set has also been added
    - New optimization on CDS requests allows CFRM to process I/O functions as a set with unlock after the end of each set.
    - For example (lock/read,write/unlock,lock/read,write/unlock) … becomes...
      (lock/read,write,read,write,unlock) …or, if lucky...
      (lock/read,write,write,unlock)

Interactions & Dependencies

- Comparing Message-Based Event and Confirmation Processing vs. CFRM LossConnRecovery MGMT
- Testing results with 101 Structures - Vicom Environment
- Message-based
  - 101 Structures need recovery at Time 0
  - 11 Structures completed duplex-failover after 42 seconds
  - 94 structures completed after 108 seconds
  - All complete after 120 seconds
- CFLCR MGMT
  - 101 Structures need recovery at Time 0
  - 11 Structures completed duplex-failover after 11 seconds
  - All complete after 67 seconds

Interactions & Dependencies

- Comparing Message-Based Event and Confirmation Processing vs. CFRM LossConnRecovery MGMT
- More Testing
- System Hardware Test
  - 2006 Structures
  - 5 system-managed duplexed structures
- CFLCRMGMT enabled
  - The recovery processing took about 16.5 minutes.
- CFLCFMGMT disabled
  - The recovery processing took just under 14 minutes.

Interactions & Dependencies

- Comparing Message-Based Event and Confirmation Processing vs. CFRM LossConnRecovery MGMT
- CFLCRMGMT enabled - about 16.5 minutes.
  - 2006 structures rebuilt
  - 1.5 minutes to rebuild ISGLOCK
  - 5 Duplexing Failover structures
  - No hangs were detected.
- CFLCFMGMT disabled - under 14 minutes.
  - 1008 structures rebuilt
  - 2.5 minutes to rebuild ISGLOCK.
  - 2 Structures duplexing failover .
  - Over half of the structures that rebuild was started for experienced hangs and CFSTRHANGTIME terminated almost all of those connectors/structures.
    - The termination allowed the "recovery" to be much faster ... since connector termination is much faster than rebuild.

Interactions & Dependencies

- Software Dependencies
  - Message Based Event & Confirmation Processing is a pre-requisite to enabling Serial Rebuild
  - All systems in the sysplex should enable CFRM LossConn Recovery management (CFLCRMGMT) before CFRM can properly process the structures in a serial manner
- Hardware Dependencies
  - None

Migration & Coexistence Considerations

- No toleration/coexistence APARs/PTFs
- No Rolldown to lower level z/OS releases.
- The function can be enabled one system at a time
  - The structures may not all be processed serially until all systems have enabled CFLCRMGMT
  - CFLCRMGMT should remain disabled until all systems are at V2.1
- In order for all systems to benefit from CFRM Serial Rebuild Processing
  - Enable the function only when all systems are at V2.1 or above

Installation

- Miscellaneous:
  - No APARs or PTFs needed
  - In order to enable Message-Based Event Processing a new version of the CFRM CDS needs to be formatted.
  - No hardware configuration updates required
  - PARMLIB FUNCTION statements in COUPLEXX members may be needed
  - CFRM Policy updates to STRUCTURE definition are optionally required
  - Planning CF Structure rebuild priority may be needed if RECPRTY will be used for STRUCTURES.
  - No special web deliverables are needed

Presentation Summary

- Without CFRM LOSSCONN recovery management, systems independently initiate rebuild processing to recover from a loss of connectivity to a CF and all that processing is initiated simultaneously.
- With CFRM LOSSCONN recovery management, the message-based manager system is responsible for orchestrating and initiating processing
  - to recover from a loss of connectivity to a CF and all that processing is initiated in a more serial manner.
  - to process a gain of connectivity to a CF or start duplexing and all that processing is initiated in a serial manner.
- The serial nature of the recovery processing will help reduce interference and reduce average recovery time.

Presentation Summary

- A new COUPLXX and SETXCF FUNCTIONS keyword will be introduced:
  - CFLCRMGMT: CFRM LOSSCONN Recovery Management

- Message-based processing is enhanced with the ability to perform CFRM LOSSCONN recovery management.
  - When the optional CFLCRMGMT function is enabled on all active systems in the sysplex, the message-based manager system may enable CFRM LOSSCONN recovery management.

- CFRM LOSSCONN recovery management may enhance average CF LOSSCONN recovery time by processing CF structures serially, rather than in parallel.

Appendix: Message Based Event Processing

- The Message-based event processing protocol available with z/OS V1R8 defines a single system as "manager" and all other systems as participants in the recovery process.
- The manager system is responsible for coordinating events and confirmations with the participant systems and is responsible also for updating the CFRM active policy, thus reducing I/O to the CFRM couple data set to a single, central control point. Communication between the manager system and the participant systems, on which applications or subsystems are running, is through XCF signaling.
- Must run with MSGBASED-formatted CFRM couple data sets to switch from Policy-Based to Message-Based as well as enablement on each system.

Appendix: Message Based Event Processing

- D XCF, STRUCTURE
- RECPRTY will be displayed for each structure on the display XCF, STRUCTURE operator command output message IXC360I if the RECPRTY has been specified in the CFRM Policy.

  - D XCF, STR, STRNAME=*
  - STRNAME: strname
  - .....
  - SYSTEM RECPRTY : 3
  - . . .
  - POLICY INFORMATION:
  - POLICY SIZE : policysize
  - POLICY INITSIZE: policyinitsize
  - POLICY MINSIZE : policyminsize
  - ...
  - [POLICY RECPRTY : recprty]

Appendix: Message Based Event Processing

- D XCF, STRUCTURE
- The DISPLAY XCF command will be enhanced with several status filter options to provide an operator with the ability to determine what "serial" processing is not yet complete and why it is still pending.

- DISPLAY XCF,
- [,{STRUCTURE|STR}]

[,{STRNAME|STRNM}={(strname[,strname]...)|ALL}]

[,{CONNAME|CONNM}={(conname[,conname]...)|ALL}]

[,{STATUS|STAT}=

([,DUPMISMATCH][,LOSSCONN] [,RBPROC][,RBPEND][,DUPENAB] [,DUPALLOW]

Appendix: DISPLAY STRUCTURE STATUS Filters

- DUPMISMATCH
  - Allocated but DUPLEXED state does not match policy – start or stop duplexing pending
- LOSSCONN
  - A connector has lost connectivity to the structure
- RBPROC
  - Structure in rebuild processing (other than duplex established)
- RBPEND
  - POPCF or REALLOCATE evaluation pending
- DUPENAB/DUPALLOW
  - Structure with policy DUPLEX specification of ENABLED or ALLOWED, respectively

Appendix: Enable CFLCRMGMT

- Use system command

```
RO *ALL,SETXCF FUNCTIONS,ENABLE=CFLCRMGMT
IXC373I XCF / XES OPTIONAL FUNCTIONS ENABLED:
        CFLCRMGMT
IXC548I CFRM EVENT MANAGEMENT ENVIRONMENT UPDATED
EVENT MANAGEMENT PROTOCOL: MESSAGE-BASED
REASON FOR CHANGE: MANAGEMENT LEVEL
TRANSITION SEQUENCE NUMBER: 00000003
TRANSITION TIME: 08/08/2012 15:17:15.109343
MANAGER SYSTEM NAME: S1
MANAGER SYSTEM NUMBER: 0100000B
MANAGEMENT LEVEL: 01052010
```

- Use COUPLExx parmlib member

```
FUNCTIONS ENABLE(CFLCRMGMT)
```

Appendix: IXCYQUAA

```
...
QUACFSTCFLCRMGMT EQU X'01' CF LossConn recovery management is in        *
                           progress for the CF


...
QUASTRSTCFLCRMGMT EQU X'01' CF LossConn recovery management is in       *
                           progress for the structure
...
QUASTRMSGBASEDLEVEL DS F Level of message-based event processing        *
                           currently being used by CFRM. Valid when     *
                           QuaStrMsgBasedEventMgmt is on.
...
QUASTRRECPRTY DS H       RECPRTY for structure as specified in CFRM     *
                           active policy. Value of zero implies not     *
                           specified
QUASTRSYSRECPRTY DS H    RECPRTY for structure determined by the        *
                           system. Value of zero implies RECPRTY is not *
                           supported for the structure and it will not  *
                           participate in LOSSCONN RECOVERY management.
...
```

Appendix: DISPLAY COUPLE

```
D XCF,C
IXC357I  15.10.32  DISPLAY XCF
SYSTEM S1 DATA
...
    OPTIONAL FUNCTION STATUS:
      FUNCTION NAME                    STATUS      DEFAULT
      DUPLEXCF16                       ENABLED     DISABLED
      SYSSTATDETECT                    ENABLED     ENABLED
      USERINTERVAL                     DISABLED    DISABLED
      CRITICALPAGING                   DISABLED    DISABLED
      DUPLEXCFDIAG                     DISABLED    DISABLED
      CFLCRMGMT                        ENABLED     DISABLED

SYSPLEX COUPLE DATA SETS
...
```

Appendix: Recovery Management Messages – CF Failure Recognition

```
S1        IXC518I SYSTEM S1 NOT USING
                  COUPLING FACILITY  SIMDEV.IBM.EN.ND0100000000
                                     PARTITION: 00    CPCID: 00
                  NAMED LF01
                 REASON: CONNECTIVITY LOST.
                 REASON FLAG: 13300001.
S2        IXC518I SYSTEM S2 NOT USING
                  COUPLING FACILITY  SIMDEV.IBM.EN.ND0100000000
                                     PARTITION: 00    CPCID: 00
                  NAMED LF01
                 REASON: CONNECTIVITY LOST.
                 REASON FLAG: 13300001.
S1        IXC568I CF LOSSCONN RECOVERY PROCESSING INITIATED.
            MANAGER SYSTEM NAME      : S1
            MANAGER SYSTEM NUMBER    : 0100000D
            MANAGEMENT LEVEL         : 01052010
S2        IXC568I CF LOSSCONN RECOVERY PROCESSING REQUESTED.
            MANAGER SYSTEM NAME      : S1
            MANAGER SYSTEM NUMBER    : 0100000D
            MANAGEMENT LEVEL         : 01052010
```

Appendix: Recovery Management Messages - Successful

```
IXC518I SYSTEM S1 NOT USING
        COUPLING FACILITY  SIMDEV.IBM.EN.ND0100000000
                           PARTITION: 00    CPCID: 00
        NAMED LF01
       REASON: CONNECTIVITY LOST.
       REASON FLAG: 13300001.
IXC568I CF LOSSCONN RECOVERY PROCESSING INITIATED.  ← system that lost connectivity
  MANAGER SYSTEM NAME      : S1
  MANAGER SYSTEM NUMBER    : 0100000B
  MANAGEMENT LEVEL         : 01052010
IXC568I CF LOSSCONN RECOVERY MANAGEMENT STARTED.    ← manager system

... rebuild and duplexing failover ...

IXC568I CF LOSSCONN RECOVERY MANAGEMENT SUCCESSFUL. ← manager system
IXC568I CF LOSSCONN RECOVERY PROCESSING COMPLETED.  ← system that lost connectivity
  EVENT MANAGEMENT PROTOCOL: MESSAGE-BASED
  MANAGER SYSTEM NAME      : S1
  MANAGER SYSTEM NUMBER    : 0100000B
  MANAGEMENT LEVEL         : 01052010
```

Appendix: DISPLAY XCF,STRUCTURE

```
D XCF,STR,STRNM=*
IXC360I  15.28.30  DISPLAY XCF
CF LOSSCONN RECOVERY MANAGEMENT IS IN PROGRESS.
STRNAME: THRLSTMQ
...
 SYSTEM RECPRTY  : 3

 & AMPERSAND DENOTES CONNECTOR WHO LOST CONNECTIVITY TO STRUCTURE
 CONNECTION NAME  ID VERSION  SYSNAME  JOBNAME  ASID STATE
 ---------------- -- -------- -------- -------- ---- ----------------
 IXCLO06B0001     01 0001000C S1       LISTFILL 002C ACTIVE &
...
EVENT MANAGEMENT: MESSAGE-BASED       MANAGER SYSTEM NAME:  S1
                                      MANAGEMENT LEVEL   :  01052010
D XCF,STR
IXC359I  15.35.36  DISPLAY XCF
CF LOSSCONN RECOVERY MANAGEMENT IS IN PROGRESS.
STRNAME            ALLOCATION TIME   STATUS                    TYPE
...
```

Appendix: DISPLAY XCF,CF

```
D XCF,CF,CFNM=*
IXC362I  15.28.30  DISPLAY XCF
CF LOSSCONN RECOVERY MANAGEMENT IS IN PROGRESS.
CFNAME: A
...
D XCF,CF
IXC361I  15.34.56  DISPLAY XCF
CF LOSSCONN RECOVERY MANAGEMENT IS IN PROGRESS.
CFNAME     COUPLING FACILITY                SITE
...
```

Appendix: Serial Policy-Initiated Duplexing Action Deferred Message

```
IXC538I DUPLEXING REBUILD OF STRUCTURE DUPENABLED01
WAS NOT INITIATED BY MVS. REASON:
DEFERRED UNTIL PROCESS COMPLETION

IXC538I DUPLEXING REBUILD OF STRUCTURE DUPENABLED02
WAS NOT STOPPED BY MVS. REASON:
DEFERRED UNTIL PROCESS COMPLETION
```

Appendix: DISPLAY STRUCTURE STATUS Filters – Usage

- Did z/OS duplex all my DUPLEX(ENABLED) structures?
  - D XCF,STR,STAT=DUPENAB
  - If not, maybe delayed for more important work, rebuild processing, or stop duplex
    - D XCF,STR,STAT=(DUPMISMATCH,RBPROC,RBPEND)

      CF LOSSCONN RECOVERY MANAGEMENT IS IN PROGRESS.

      THE REALLOCATE PROCESS IS IN PROGRESS.

      POPULATECF REBUILD PENDING

      REBUILD IN PROGRESS
- Did z/OS resolve all duplexing mismatches?
  - D XCF,STR,STAT=DUPMISMATCH
  - If not, maybe delayed for more important work or rebuild processing
    - D XCF,STR,STAT=(RBPROC,RBPEND)

      CF LOSSCONN RECOVERY MANAGEMENT IS IN PROGRESS.

      THE REALLOCATE PROCESS IS IN PROGRESS.

      POPULATECF REBUILD PENDING

      REBUILD IN PROGRESS

Appendix: DISPLAY STRUCTURE STATUS Filters – Usage

- Did REALLOCATE (or POPCF) complete?
  - D XCF,STR,STAT=RBPEND

    THE REALLOCATE PROCESS IS IN PROGRESS.

    POPULATECF REBUILD PENDING

    POPULATECF REBUILD IN PROGRESS
  - If not, maybe delayed for more important work

    CF LOSSCONN RECOVERY MANAGEMENT IS IN PROGRESS.
- Did CF LOSSCONN RECOVERY complete?
  - D XCF,STR,STAT=LOSSCONN,STRNM=*

    CF LOSSCONN RECOVERY MANAGEMENT IS IN PROGRESS.

Appendix: Publications

- z/OS V2R1 MVS z/OS Setting up a Sysplex
- z/OS V2R1 MVS z/OS Sysplex Services Guide
- z/OS V2R1MVS z/OS Sysplex Services Reference
- z/OS V2R1 MVS System Messages Vol 10 (IXC - IZP)