Mission: AVAILABLE

This document can be found on the web, www.ibm.com/support/techdocs
Under the category of "White Papers."

August 2011-08-05

Nicole M. Fagen
David H. Surman

Table of Contents

Mission

Your mission, should you choose to accept it, is to configure your sysplex infrastructure for high availability so that the full benefits of the sysplex environment can be realized in your enterprise. Good luck, Sysprog!

This document identifies how to configure a parallel sysplex to maximize the high availability potential of your enterprise. The paper will leverage all of the latest z/OS functions and features, as well as new functions of the System z196 hardware platform. Once each item on the *Sysplex High Availability Checklist* has been reviewed and implemented in your environment, your mission will be complete.

GDPS (Geographically Dispersed Parallel Sysplex) Disclaimer

The high availability options detailed herein are for a non-GDPS parallel sysplex. The best practices and high availability configuration in a GDPS environment may differ. Please consult GDPS publications or support as needed to ensure the proper settings are selected for that environment.

Sysplex High Availability Checklist

- □ Coupling Facility (CF)
  - □ At least 2 CFs are defined and active in the sysplex, with physical connectivity to all systems in the sysplex
  - □ External CFs are preferred to internal CFs
  - □ Ensure there is enough coupling facility space for all the CF structures, and enough white space for non-duplexed structures to rebuild into other available coupling facilities should there be a CF outage.
  - □ Dedicated ICFs/CPs preferred. Dedicated processors are a must to successfully duplexing structures in a production environment.
  - □ Have at least two paths from each operating system image to the coupling facility. Additional paths may be required with heavy workloads.
  - □ Non-volatile and failure-isolated (standalone) CFs are preferred
  - □ CF to CF links available and being used
  - □ Coupling Facility upgrade procedures have been updated to use the latest features
- □ Coupling Facility structures are sized properly
- □ Coupling Facility structures have correct ALLOWAUTOALT setting
- □ CFRM formatted with MSGBASED
- □ CFRM formatted with SMDUPLEX
- □ CFRM formatted with SMREBLD
- □ Coupling Facility structures are duplexed to adhere to the high availability requirements of the application
- □ Sysplex Failure Management policy is active
  - □System names and weights specified
  - □ISOLATETIME(0)
  - □SSUMLIMIT(900)
  - □PROMPT is NOT specified
  - □MEMSTALLTIME(300)
  - □CFSTRHANGTME(900)
- □ SYSSTATDET is enabled
- □ AUTOIPL is enabled
- □ Failure Detection Interval is set to meet the needs of the enterprise, and has a correct relation to excessive spin parameters
- □ Signaling Configuration
  - □ XCF signaling structures sized properly
  - □ CTCs exist between systems in addition to XCF signaling structures
  - □ TCLASSes are defined appropriately for your workload and applications
  - □ MAXMSG is sufficiently large
- □ Couple data set mirroring pitfalls are understood and avoided
- □ CRITICALPAGING is enabled in DASD Swap environment
- □ All XCF and XES Healthchecks have been enabled to run on an ongoing basis, and any resulting warnings/exceptions have been investigated and addressed as needed

- Automation is established for critical sysplex messages
- Path busy interpretation is understood
- Understand the value of nondisruptive CF dumps and how the new dumps help ensure high availability

Coupling Facility Configuration

For a high availability environment, the following are recommended:

(1) Have at least 2 coupling facilities defined in the CFRM policy and physically available from all systems in the sysplex.
(2) External CFs are preferred to internal CFs.
(3) Ensure there is enough coupling facility space for all the CF structures and enough white space for non-duplexed structures to rebuild into other available coupling facilities should there be a CF outage.
(4) Use dedicated ICFs/CPs on the CFs whenever possible. Dedicated processors are a must to successfully duplex structures in a production environment.
(5) Have at least two paths from each operating system image to the coupling facility. Additional paths may be required with heavy workloads.
(6) Non-volatile and failure-isolated (standalone) CFs are preferred.
(7) CF to CF links
(8) Coupling Facility upgrade procedures have been updated to leverage the latest features

For additional information on each recommendation please refer to IBM Red Book: System z Parallel Sysplex Best Practices.

REALLOCATE

REALLOCATE is a process under which each allocated CF structure is evaluated and the CF location of the structure may be adjusted as deemed necessary by the system. In addition to the location possibly being adjusted, the simplex or duplex mode of the structure may be adjusted, and pending CFRM policy changes may be driven to take effect.

REALLOCATE evaluation processing considers the allocation criteria, the user-specified PREFLIST, and any pending changes to the CFRM policy to determine what actions to take for each structure. Should REALLOCATE processing determine that the structure is allocated in the most preferred coupling facility and no other structure-related changes need to take effect, then the structure will not be rebuilt. Note that the most preferred CF in the user-specified PREFLIST is not necessarily going to be the most preferred CF for REALLOCATE purposes, since CFRM also applies other criteria to making these structure placement/optimization decisions.

REALLOCATE processing is both simpler and more efficient than issuing SETXCF commands to start or stop duplexing, or initiating simplex rebuilds for structures when a CF needs to be emptied of all structures. Further, REALLOCATE processing can aid in

evacuating a coupling facility which is being taken down for maintenance or upgraded to a new machine. As stated above, REALLOCATE processing takes into consideration additional criteria for moving structures beyond the PREFLIST. The additional considerations ensure the structure is placed in the best possible location. The predecessor to REALLOCATE, POPULATECF(POPCF),only considers the PREFLIST. POPCF was designed to populate a CF will structures. Using POPCF, structures are moved into the target CF if the structure's PREFLIST had the target CF listed before the CF in which the structure currently resides. REALLOCATE, like POPCF, will move only one structure at a time, thereby causing significantly less burden on the z/OS systems and connectors to each structure, and minimizing the impact that the rebuild processing has on your ongoing sysplex workload.

As of z/OS 1.12 there are two new options associated with REALLOCATE processing. The first is a display to "test" or project the results of a new REALLOCATE command, if one were to be issued at the present time. The other is a display to "report" on the outcomes of the previous REALLOCATE that was performed in the sysplex.

TEST

DISPLAY XCF,REALLOCATE,TEST

The output of the REALLOCATE TEST states the expected results if a REALLOCATE were to be issued in the current configuration – what would such a hypothetical REALLOCATE process actually do? The error/exceptions conditions, warning conditions, structures successfully REALLOCATEd, structures already in preferred CFs, summary of structures that would be allocated in each CF subsequent to the execution of the hypothetical REALLOCATE action, and a summary report of the REALLOCATE process, are all provided as output. The following is a sample of the results of a REALLOCATE TEST:

```
D XCF,REALLOCATE,TEST
IXC347I  21.05.03  DISPLAY XCF 551

COUPLING FACILITY STRUCTURE ANALYSIS PERFORMED FOR REALLOCATE TEST.
------------------------------------------------------------------------
STRUCTURE(S) WITH AN ERROR/EXCEPTION CONDITION

NONE
------------------------------------------------------------------------
STRUCTURE(S) WITH A WARNING CONDITION

NONE
------------------------------------------------------------------------
STRUCTURE(S) REALLOCATED SUCCESSFULLY

STRNAME: CI1_DFHLOG                                INDEX: 13
   SIMPLEX STRUCTURE ALLOCATED IN CF(S) NAMED: CF1A
   CFNAME     STATUS/FAILURE REASON
   --------   ------------------------------------------------------
   CFRP       PREFERRED CF 1
                                INFO110: 00000028 AC000800 00010011
   CF1A       PREFERRED CF ALREADY SELECTED
                                INFO110: 00000028 AC000800 0002000F

   1 REALLOCATE STEP(S): REBUILD
```

```
STRNAME: CI1D_DFHLOG1                                INDEX: 15
   DUPLEXED STRUCTURE ALLOCATED IN CF(S) NAMED: CF1A     CFRP
   CFNAME     STATUS/FAILURE REASON
   --------   ------------------------------------------------------
   CFRP       PREFERRED CF 1
                                 INFO110: 0000001E CC000B00 00010011
   CF1A       PREFERRED CF 2
                                 INFO110: 0000001E CC000B00 0002000F

   2 REALLOCATE STEP(S): KEEP=NEW, DUPLEX

… lines omitted …
-------------------------------------------------------------------------
STRUCTURE(S) ALREADY ALLOCATED IN PREFERRED CF(S)

STRNAME: ADSW_DFHJ01                                 INDEX: 54
   CFNAME     STATUS/FAILURE REASON
   --------   ------------------------------------------------------
   CF1A       PREFERRED CF 1
                                 INFO110: 0000003C CC000B00 0000000F
   CFRP       PREFERRED CF 2
                                 INFO110: 0000003C CC000B00 00000011

STRNAME: ADSW_DFHJ02                                 INDEX: 55
   CFNAME     STATUS/FAILURE REASON
   --------   ------------------------------------------------------
   CFRP       PREFERRED CF 1
                                 INFO110: 0000003C AC000800 00000011
   CF1A       PREFERRED CF ALREADY SELECTED
                                 INFO110: 0000003C AC000800 0000000F

… lines omitted …
COUPLING FACILITY STRUCTURE ANALYSIS OUTPUT FOR REALLOCATE TEST

CFNAME: CFRP
   COUPLING FACILITY     :   002817.IBM.02.00000002D4C6
                             PARTITION: 10   CPCID: 00
   CONNECTED SYSTEM(S):
   CSK      SA0      SB0      SC0      SD0      SE0      SF0
   SG0      SH0

   ACTIVE STRUCTURE(S):
   ADSW_DFHJ01(NEW)       ADSW_DFHJ02            ADSW_DFHJ04
   ADSW_DFHJ06            ADSW_DFHLGLOG1(NEW)    APPCLOG
   CI1_DFHLOG             CI1_DFHLOG1            CI1_DFHLOG2
   CI1_DFHSHUNT           CI1_DFHSHUNT1          CI1_DFHSHUNT2
   CI1D_DFHLOG(OLD)       CI1D_DFHLOG1(OLD)      CI1D_DFHLOG2(NEW)
   CI1D_DFHSHUNT(OLD)     CI1D_DFHSHUNT1(OLD)    CI1D_DFHSHUNT2(NEW)
   CM1D_DFHLOG(OLD)       CM1D_DFHSHUNT(OLD)     COUPLE_CKPT1(NEW)
   CQS_FF_LOGSTR(OLD)     CQS_FP_LOGSTR(NEW)     CRTWDB2_GBP0(NEW)
   CRTWDB2_GBP16K0(OLD)   CRTWDB2_GBP16K1(NEW)   CRTWDB2_GBP20(NEW)
   CRTWDB2_GBP21(NEW)     CRTWDB2_GBP32K(NEW)    CRTWDB2_GBP32K1(NEW)
   CRTWDB2_GBP8K0(OLD)    CRTWDB2_GBP8K1(NEW)    CRTWDB2_LOCK1(OLD)
   CRTWDB2_SCA(OLD)       CST4DB2_GBP0(NEW)      CST4DB2_GBP16K0(OLD)
   CST4DB2_GBP8K0(NEW)    CST4DB2_LOCK1          CST4DB2_SCA
   CST5DB2_GBP0(NEW)      CST5DB2_GBP8K0(OLD)    CST5DB2_LOCK1
   CST5DB2_SCA            CST6DB2_GBP0(NEW)      CST6DB2_GBP16K0(OLD)
   CST6DB2_GBP20(OLD)     CST6DB2_GBP21(NEW)     CST6DB2_GBP32K1(NEW)
   CST6DB2_GBP8K0(OLD)    CST6DB2_LOCK1(OLD)     CST6DB2_SCA(OLD)
   DFHCFLS_CI1D(OLD)      DFHNCLS_CI1D(OLD)      DFHXQLS_CI1D(NEW)
   DFHXQLS_CI1S           FFMSGQ_STR(OLD)        FFOVFLO_STR
   FPMSGQ_STR(NEW)        IGWLOCK00(NEW)         IRLMLOCK1(OLD)
   IRRXCF00_B002          IRRXCF00_P001          IRRXCF00_P003
   ISGLOCK                ISTGENERIC(NEW)        ISTMNPS(OLD)
   IXCPLEX_PATH1          IXCPLEX_PATH4          IXCPLEX_PATH5
   IXCPLEX_PATH7          IXCPLEX_PATH9          LOGGER_OPERLOG(OLD)
   RLSCACHE01             RRSLOG_RESTART         RRSLOG_RMDATA(OLD)
   SQ00APPL1(OLD)         SQ00APPL2(NEW)         SQ00CSQ_ADMIN(NEW)
   SYSARC_DFHSM_RCL(NEW)  SYSASFPBP01            SYSZWLM_D4C62817(NEW)
   SYSZWLM_WORKUNIT(OLD)  SYSZWLM_74562094(NEW)  TVS_IGWLOG
```

```
   VSAMCACHE1

CFNAME: CF1A
   COUPLING FACILITY     :   002094.IBM.02.000000057456
                             PARTITION: 21   CPCID: 00
   CONNECTED SYSTEM(S):
   CSK      SA0      SB0      SC0      SD0      SE0      SF0
   SG0      SH0

   ACTIVE STRUCTURE(S):
   ADSW_DFHJ01(OLD)       ADSW_DFHJ03            ADSW_DFHJ05
   ADSW_DFHJ07            ADSW_DFHLGLOG1(OLD)    CI1D_DFHLOG(NEW)
   CI1D_DFHLOG1(NEW)      CI1D_DFHLOG2(OLD)      CI1D_DFHSHUNT(NEW)
   CI1D_DFHSHUNT1(NEW)    CI1D_DFHSHUNT2(OLD)    CM1D_DFHLOG(NEW)
   CM1D_DFHSHUNT(NEW)     COUPLE_CKPT1(OLD)      CQS_FF_LOGSTR(NEW)
   CQS_FP_LOGSTR(OLD)     CRTWDB2_GBP0(OLD)      CRTWDB2_GBP16K0(NEW)
   CRTWDB2_GBP16K1(OLD)   CRTWDB2_GBP20(OLD)     CRTWDB2_GBP21(OLD)
   CRTWDB2_GBP32K(OLD)    CRTWDB2_GBP32K1(OLD)   CRTWDB2_GBP8K0(NEW)
   CRTWDB2_GBP8K1(OLD)    CRTWDB2_LOCK1(NEW)     CRTWDB2_SCA(NEW)
   CST4DB2_GBP0(OLD)      CST4DB2_GBP16K0(NEW)   CST4DB2_GBP8K0(OLD)
   CST5DB2_GBP0(OLD)      CST5DB2_GBP16K0        CST5DB2_GBP32K
   CST5DB2_GBP8K0(NEW)    CST6DB2_GBP0(OLD)      CST6DB2_GBP16K0(NEW)
   CST6DB2_GBP20(NEW)     CST6DB2_GBP21(OLD)     CST6DB2_GBP32K1(OLD)
   CST6DB2_GBP8K0(NEW)    CST6DB2_LOCK1(NEW)     CST6DB2_SCA(NEW)
   DFHCFLS_CI1D(NEW)      DFHCFLS_CI1S           DFHNCLS_CI1D(NEW)
   DFHNCLS_CI1S           DFHXQLS_CI1D(OLD)      FFMSGQ_STR(NEW)
   FPMSGQ_STR(OLD)        HSA_LOG                IGWLOCK00(OLD)
   IRLMLOCK1(NEW)         IRRXCF00_B001          IRRXCF00_B003
   IRRXCF00_P002          ISTGENERIC(OLD)        ISTMNPS(NEW)
   IXCPLEX_PATH10         IXCPLEX_PATH2          IXCPLEX_PATH3
   IXCPLEX_PATH6          IXCPLEX_PATH8          LOGGER_OPERLOG(NEW)
   OSAMCACHE1             RLSCACHE02             RRSLOG_DELAYED
   RRSLOG_MAIN            RRSLOG_RMDATA(NEW)     SQ00APPL1(NEW)
   SQ00APPL2(OLD)         SQ00CSQ_ADMIN(OLD)     SYSARC_DFHSM_RCL(OLD)
   SYSIGGCAS_ECS          SYSZWLM_D4C62817(OLD)  SYSZWLM_WORKUNIT(NEW)
   SYSZWLM_74562094(OLD)  TVS_IGWSHUNT
------------------------------------------------------------------------
REALLOCATE TEST RESULTED IN THE FOLLOWING:
         1  STRUCTURE(S) REALLOCATED - SIMPLEX
         4  STRUCTURE(S) REALLOCATED - DUPLEXED
         0  STRUCTURE(S) POLICY CHANGE MADE - SIMPLEX
         0  STRUCTURE(S) POLICY CHANGE MADE - DUPLEXED
        51  STRUCTURE(S) ALREADY ALLOCATED IN PREFERRED CF - SIMPLEX
        51  STRUCTURE(S) ALREADY ALLOCATED IN PREFERRED CF - DUPLEXED
         0  STRUCTURE(S) NOT PROCESSED
        30  STRUCTURE(S) NOT ALLOCATED
       118  STRUCTURE(S) NOT DEFINED
  --------
       255  TOTAL

         0   STRUCTURE(S) WITH AN ERROR/EXCEPTION CONDITION
```

REPORT

The output of the REALLOCATE REPORT states the results of the last REALLOCATE action that was actually performed in the sysplex. The start/end times for the REALLOCATE process, the error/exception conditions, warning conditions, structures successfully REALLOCATEd, structures already in preferred CFs, and the summary report of the REALLOCATE process, are all provided. The following is a sample of the results of a REALLOCATE REPORT:

```
D XCF,REALLOCATE,REPORT
IXC347I  21.38.39  DISPLAY XCF 847

THE REALLOCATE PROCESS STARTED ON 11/04/2010 AT 14:57:28.33.
THE REALLOCATE PROCESS ENDED ON 11/04/2010 AT 15:00:27.40.
------------------------------------------------------------------------
STRUCTURE(S) WITH AN ERROR/EXCEPTION CONDITION

NONE
------------------------------------------------------------------------
STRUCTURE(S) WITH A WARNING CONDITION

NONE
------------------------------------------------------------------------
STRUCTURE(S) REALLOCATED SUCCESSFULLY

STRNAME: ADSW_DFHJ01                              INDEX: 54
   2 REALLOCATE STEP(S): KEEP=NEW, DUPLEX
   COMPLETED ON SYSTEM SA0 ON 11/04/2010 AT 14:59:06.52.

STRNAME: ADSW_DFHJ02                              INDEX: 55
   1 REALLOCATE STEP(S): REBUILD
   COMPLETED ON SYSTEM SG0 ON 11/04/2010 AT 14:59:11.29.

STRNAME: ADSW_DFHJ03                              INDEX: 56
   1 REALLOCATE STEP(S): REBUILD
   COMPLETED ON SYSTEM SG0 ON 11/04/2010 AT 14:59:16.51.
… lines omitted…
------------------------------------------------------------------------
STRUCTURE(S) ALREADY ALLOCATED IN PREFERRED CF(S)

STRNAME: ADSW_DFHLGLOG1                           INDEX: 64
   EVALUATED ON SYSTEM SG0 ON 11/04/2010 AT 14:59:34.75.

STRNAME: COUPLE_CKPT1                             INDEX: 5
   EVALUATED ON SYSTEM SG0 ON 11/04/2010 AT 14:57:31.31.
… lines omitted …
------------------------------------------------------------------------
REALLOCATE PROCESSING RESULTED IN THE FOLLOWING:
        28  STRUCTURE(S) REALLOCATED - SIMPLEX
        22  STRUCTURE(S) REALLOCATED - DUPLEXED
         0  STRUCTURE(S) POLICY CHANGE MADE - SIMPLEX
         0  STRUCTURE(S) POLICY CHANGE MADE - DUPLEXED
        30  STRUCTURE(S) ALREADY ALLOCATED IN PREFERRED CF - SIMPLEX
        11  STRUCTURE(S) ALREADY ALLOCATED IN PREFERRED CF - DUPLEXED
         0  STRUCTURE(S) NOT PROCESSED
        46  STRUCTURE(S) NOT ALLOCATED
       118  STRUCTURE(S) NOT DEFINED
  --------
       255  TOTAL

         0  STRUCTURE(S) WITH AN ERROR/EXCEPTION CONDITION

         0  STRUCTURE(S) MISSING PREVIOUS REALLOCATE DATA
```

Notice, the system name, date and time of the rebuild event are noted in the report. If additional details pertaining to a particular rebuild are needed, the operlog can be reviewed for the given system, date and time.

CF Maintenance Mode (MAINTMODE)

A coupling facility can be placed in maintenance mode to ensure that no additional CF structure allocations can take place in the particular coupling facility, until such time as the CF is taken out of maintenance mode. MAINTMODE is also commonly used to indicate to REALLOCATE evaluation processing that a coupling facility is to be taken out of service for maintenance, and therefore, that REALLOCATE processing should move all structures out of that CF and into "more preferred" CFs that are available (and not in maintenance mode). The commands to place a coupling facility in MAINTMODE and to remove a coupling facility from MAINTMODE are:

```
SETXCF START,MAINTMODE,CFNAME=cfname

SETXCF STOP,MAINTMODE,CFNAME=cfname
```

MAINTMODE is a significant simplification function. Prior to MAINTMODE, performing planned CF reconfiguration actions, such as CF disruptive maintenance or upgrade, required the definition and activation of a new CFRM policy, one which did not include the coupling facility being taken down for service. The use of MAINTMODE eliminates the need to create, maintain, or activate any "alternative" CFRM policies during these sorts of CF maintenance actions.

Upgrading a Coupling Facility – Implement/Leverage Best Practices

IBM White Paper WP101905 contains the best practices for performing coupling facility upgrades. Leveraging these best practices will improve your sysplex availability while also minimizing risk and elapsed time for performing such upgrades.

Sizing Coupling Facility Structures

There are two IBM tools available to assist with the proper sizing of coupling facility structures.

First, the CFSizer website can be used to obtain accurate sizings for coupling facility structures which are being created for the first time, or when an application workload or sysplex infrastructure parameter (for example, number of systems in the sysplex) is being changed in a way that may affect proper CF structure sizing. The CF Sizer output can then be used to update the CFRM policy appropriately with new CF structure sizes, and these size changes can be implemented using the REALLOCATE function.

Second, the Sizer batch utility can be used when upgrading from one CFCC level (CFLEVEL) to the next level. In order for this batch utility to carry out its sizing function, the uplevel coupling facility must be available for use in your sysplex, but must

have not yet been populated with CF structures. When the uplevel coupling facility is first made available in your sysplex, run the SIZER batch job to obtain new sizings for all of the active structures that are currently still allocated in the downlevel CF. Next, update the CFRM policy appropriately with the new structure sizes determined by the SIZER utility. Finally, make the uplevel coupling facility available for use and allow structures to rebuild (as appropriate) into the new coupling facility using the REALLOCATE function.

Note that the SIZER batch utility is only useful if the structure sizes on the downlevel coupling facility are currently believed to be appropriate and correct for the applications using the structures. If in fact the allocated structures on the downlevel coupling facility are inadequate, then the sizes determined by SIZER will also be similarly inadequate on the uplevel coupling facility. The SIZER batch utility is trying to determine "equivalence" between the structures allocated in a downlevel and uplevel CF; it is not trying to determine correctness or sufficiency of the structure sizes.

Recommendations:

(1) Do not make your CF structures too small. Initial allocations may fail when the connectors reject the attributes of a too-small structure, or in some cases it may not be possible to allocate an under-sized structure in the CF at all. If a very small structure is allocated, even if it is accepted for use by the connector, it is likely to encounter performance or availability problems. Generally speaking, erring on the side of "oversizing" CF structures is far less likely to lead to problems than "undersizing" them.

(2) The ratio between the initial allocation size of the structure (INITSIZE) and the maximum possible size of the structure (SIZE) should not exceed 1:2. The greater the discrepancy between the structure's INITSIZE and maximum SIZE, the greater the amount of "fixed overhead" space required to manage the potential future growth in the structure size – which takes away usable space from the structure given its initial allocation size! While it is desirable to allocate CF structures in such a way that they allow room for future expansion via structure alter, this should not be over-done.

AutoAlter Support - ALLOWAUTOALT

ALLOWAUTOALT(YES) specified by the installation in the CFRM policy, and IXLCONN ALLOWAUTO(YES) in the CF structure exploiter's support, allows XCF/XES to dynamically expand, contract and reapportion a coupling facility structure. This processing in commonly referred to as AutoAlter processing.

With AutoAlter, XCF/XES will expand and reapportion the entries and elements when the FULLTHRESHOLD is reached. XCF/XES will contract a structure when the CF space becomes 90% full.

Some applications rely heavily on ALLOWAUTOALT to ensure the best use is made of the structure. For example, DB2 GBPs rely on ALLOWAUTOALT to alter the topology of the group buffer pool structures. On the other hand, some structures issue their own IXLALTER requests to change the structure topology regardless of whether AutoAlter has been enabled for use via the CFRM policy. For example, System Logger issues IXLALTER to change the entry/element ratio of the structure to suit its needs. System logger relies heavily on the counts in the structure to perform offload processing and as such does not react well to a structure being contracted.

Because of these different structure exploitation designs, there are some CF structures for which the use of AutoAlter should be enabled, and other CF structures for which the use of AutoAlter should be avoided; care must be taken to specify the ALLOWAUTOALT(YES|NO) specification appropriately. To ensure the highest availability of applications, be sure to configure ALLOWAUTOALT to match the recommendation of the exploiting application.

Coupling Facility Structure Duplexing

Structure duplexing can be leveraged to ensure high availability of the data in the structure for exploiting applications. Whether a structure needs to be duplexed or not depends largely on the importance of the data in the structure. If the data is mission critical and persistent, or cannot easily be recovered any other way if lost, then duplexing is likely desirable. Similarly, if the data is needed to successfully recover an application duplexing may be appropriate. However, if the data is highly transient, not needed for recovery and not deemed mission critical, or the application already has another back up scheme, duplexing may not be necessary.

To ensure the availability of mission critical data, be sure to configure structure duplexing (either system-managed or user-managed) to match the recommendation of the exploiting application. Note that system-managed duplexing may have significant performance implications which must be considered prior to implementation for any given CF structure. For additional information please see System - Managed CF Structure Duplexing

CFRM Options

Ensure that the CFRM couple data set (CDS) formatting parameters, and the CFRM policy, leverage the following optional items:

MSGBASED, SMDUPLEX, SMREBLD

The output of D XCF,COUPLE,TYPE=CFRM will identify which options are currently available with the couple data sets your sysplex is currently using. Note that the options for the primary couple data set actually control which optional functions can be active in the sysplex (the alternate CDS may have additional optional features defined for it, but these cannot be used until the alternate CDS is promoted to become the primary CDS).

```
IXC358I  22.20.01  DISPLAY XCF
CFRM COUPLE DATA SETS
PRIMARY    DSN: SYS1.CFRM.CDS10
           VOLSER: CDSCFP     DEVN: 3B29
           FORMAT TOD         MAXSYSTEM
           11/01/2010 21:28:20       16
           ADDITIONAL INFORMATION:
            FORMAT DATA
             POLICY(8) CF(10) STR(1029) CONNECT(129)
             SMREBLD(1) SMDUPLEX(1) MSGBASED(1)
ALTERNATE  DSN: SYS1.CFRM.CDS11
           VOLSER: CDSCFA     DEVN: 5931
           FORMAT TOD         MAXSYSTEM
           11/01/2010 21:41:30       16
           ADDITIONAL INFORMATION:
            FORMAT DATA
             POLICY(8) CF(10) STR(1029) CONNECT(129)
             SMREBLD(1) SMDUPLEX(1) MSGBASED(1)
CRFM IN USE BY ALL SYSTEMS
```

Message-Based CFRM (MSGBASED)

Contention for the CFRM CDS and CFRM policy serialization, and I/O delays to the CFRM CDS, are known to cause recovery time issues in some configurations, when the older non-MSGBASED (policy-based) protocols are being used by CFRM.

In policy-based CFRM, the protocol for use of the CFRM CDS requires serialized writes to the CFRM policy by each connector, for processing various kinds of structure events. In message-based CFRM, the protocol minimizes the I/O and contention for the CFRM CDS and policy by performing much of this cross-system and cross-connector coordination for structure events via the exchange of XCF signals. XCF signals are much preferred for I/O and contention avoidance in this context.

Enabling MSGBASED processing ensures that systems will use XCF signals to communicate the stages of rebuild processing and recovery, as well as for coordinating other types of structure event processes. The more active connectors there are to a given structure, the greater the performance benefit of MSGBASED processing will be.

To use MSGBASED processing and minimize CFRM-related delays during processing of structure events, including rebuild times, the CFRM CDS must be formatted with:

ITEM NAME(MSGBASED) NUMBER(1)

The output of D XCF,STR,STRNAME=structure_name indicates if the structure is using policy based or message based protocols.

```
IXC360I  22.28.49  DISPLAY XCF
STRNAME: ISGLOCK
 STATUS: ALLOCATED
 EVENT MANAGEMENT: MESSAGE-BASED
 TYPE: LOCK
 POLICY INFORMATION:
  POLICY SIZE    : 74 M
  POLICY INITSIZE: 74 M
  POLICY MINSIZE : 0 K
  FULLTHRESHOLD  : 90
  ALLOWAUTOALT   : NO
  REBUILD PERCENT: 1
  DUPLEX         : DISABLED
  ALLOWREALLOCATE: YES
  PREFERENCE LIST: CF2      CF3      CF4      CF1
  ENFORCEORDER   : NO
  EXCLUSION LIST IS EMPTY

 ACTIVE STRUCTURE
 ----------------
 ALLOCATION TIME: 06/17/2011 13:19:36
 CFNAME         : CF3
 COUPLING FACILITY: 002817.IBM.02.000000094E15
                    PARTITION: 04   CPCID: 00
 ACTUAL SIZE    : 65 M
 STORAGE INCREMENT SIZE: 1 M
 USAGE INFO      TOTAL     CHANGED    %
  LOCKS:       8388608
 PHYSICAL VERSION: C7EFBDFB F5EFEC8F
 LOGICAL  VERSION: C7EFBDFB F5EFEC8F
 SYSTEM-MANAGED PROCESS LEVEL: 8
 XCF GRPNAME    : IXCLO0B0
 DISPOSITION    : DELETE
 ACCESS TIME    : 0
 MAX CONNECTIONS: 32
 # CONNECTIONS  : 9

CONNECTION NAME  ID VERSION  SYSNAME  JOBNAME  ASID STATE
---------------- -- -------- -------- -------- ---- ----------------
 ISGLOCK#JA0      02 00020274 JA0      GRS      0007 ACTIVE
 ISGLOCK#JB0      01 00010287 JB0      GRS      0007 ACTIVE
 ISGLOCK#JC0      04 00040253 JC0      GRS      0007 ACTIVE
 ISGLOCK#JE0      06 00060254 JE0      GRS      0007 ACTIVE
 ISGLOCK#JF0      09 00090240 JF0      GRS      0007 ACTIVE
 ISGLOCK#J80      03 0003026A J80      GRS      0007 ACTIVE
 ISGLOCK#J90      07 00070242 J90      GRS      0007 ACTIVE
 ISGLOCK#TPN      08 00080244 TPN      GRS      0007 ACTIVE
 ISGLOCK#Z0       05 0005023C Z0       GRS      0007 ACTIVE

 DIAGNOSTIC INFORMATION:  STRNUM: 000000B0 STRSEQ: 00000002
                          MANAGER SYSTEM ID:  06003B42
  NAME/MGR  #QUEUED   1STQESN   LASTQESN  CMPESN    NOTIFYESN
  J80       00000000  00000000  00000000  000000A5  000000A5
  MGR SYS   00000000  00000000  00000000  000000A5  00000000

EVENT MANAGEMENT: MESSAGE-BASED       MANAGER SYSTEM NAME:  J80
```

In z/OS 1.12 XCF supplied a new health check, XCF_CFRM_MSGBASED, to check for the recommended use of MSGBASED protocol.

A CFRM CDS formatted for MSGBASED is also automatically formatted for SMREBLD and SMDUPLEX.

System-Managed Duplexing (SMDUPLEX)

With SMDUPLEX, the system-managed duplexing rebuild process is supported. There are applications using structures that require manual operator intervention, log-based recovery processing, or long recovery times, should a simplex instance of the structure need to be rebuilt. SMDUPLEX allows for such structures to more quickly and more transparently recover from a structure failure, CF failure, or loss of CF connectivity which affects such structures.

To use SMDUPLEX, the CFRM CDS must be formatted with:

ITEM NAME(SMDUPLEX) NUMBER(1)

In addition to having a CFRM CDS formatted to use SMDUPLEX, structure definitions need to be updated to utilize the duplexing function. CFRM policy definition of DUPLEX(ALLOWED) allows manual initiation or termination of structure duplexing using SETXCF commands or the IXLALTER programming interface. The CFRM policy definition of DUPLEX(ENABLED) results in the system trying to establish duplexing for the structure automatically whenever two suitable coupling facilities, named in the PREFLIST, are available for use. Moreover, exploiters of structures must permit the duplexing by specifying ALLOWAUTO(YES) on the IXLCONN macro used to connect to the structure.

Note: The SMDUPLEX documented here is distinct from the DB2 GBP cache structure duplexing (also called user-managed duplexing). DB2 GBP duplexing does not depend on specification of SMDUPLEX, and is also highly recommended for availability.

A CFRM CDS formatted for SMDUPLEX is also automatically formatted for SMREBLD.

System-Managed Rebuild (SMREBLD)

With SMREBLD, system managed rebuild processing is supported. System-managed rebuild processing makes it possible to rebuild (and REALLOCATE) certain types of CF structures which would not otherwise support rebuild processing. Furthermore, for many types of structures that do support rebuild, the use of SMREBLD makes it possible to rebuild these structures even at a time when there are no active connectors to the structures operating at the time. Enablement for SMREBLD is strongly recommended to broaden the scope of CF structure rebuild processing for planned reconfiguration and REALLOCATE purposes.

To use SMREBLD, the CFRM CDS must be formatted with:

ITEM NAME(SMREBLD) NUMBER(1)

Failure Detection Interval (FDI)

The failure detection interval is the length of time a system can be in a "status update missing" condition – basically, not updating its sysplex couple data set "heartbeat" information that makes the system appear active to other systems – before other systems in the sysplex will consider a true failure to have been detected for that system. See Table A for a summary of the pros and cons associated with setting various FDI lengths.

	Pro	**Con**
Long FDI	Gives the system lots of time to try to recover from whatever is preventing it from updating its "heartbeat" and resume normal operation.	"Sick" system remains active in the sysplex for a long time with no action being taken to remove it, possibly causing sysplex sympathy sickness for the other systems. "False negative" – a system that is not going to recover from a problem is not quickly removed from the sysplex
Short FDI	Sick system is removed quickly from the sysplex, allowing the rest of the system to remain fully functional and minimizing the occurrence of any sysplex sympathy sickness conditions.	Perhaps the system was about to recover but was removed just before it could recover. "False positive" – a system that WAS going to recover from a problem is removed from the sysplex so quickly that it is not able to recover in time.

Table A. Pros and cons to both a long and short Failure Detection Interval value.

The "sweet spot" of the FDI is whatever amount of time is deemed to be long enough in your environment to allow all system recovery actions declared in the EXSPATxx parmlib member to execute in an attempt to break the system out of the problem situation which may be preventing the system from updating its heartbeat (and perhaps doing other more useful work as well) – and yet, not so long that it unnecessarily delays recovery actions beyond that time.

Beginning with release z/OS 1.11, the FDI is set to the larger of the user defined INTERVAL in COUPLExx and the calculated SPIN_FDI that is derived from the excessive spin actions, as follows:

SPIN_FDI = SPIN_TIME * (n+1) +5 where n is the number of spin actions declared in EXSPATxx.

The default SPIN_TIME is 40 seconds and the default number of spin actions is three. Thus, the default SPIN_FDI is 165. D XCF,COUPLE output contains the SPIN_FDI, shown as the DERIVED SPIN INTERVAL, the USER INTERVAL (as either defaulted to or specified in COUPLExx), and the effective resultant FDI, which is shown as the INTERVAL value in the following display.

```
IXC357I  22.48.40  DISPLAY XCF 260
SYSTEM J80 DATA
   INTERVAL   OPNOTIFY     MAXMSG    CLEANUP      RETRY   CLASSLEN
        165        168       2000         15        100        956

   SSUM ACTION  SSUM INTERVAL  SSUM LIMIT     WEIGHT  MEMSTALLTIME
       ISOLATE              0         900         10           900

   CFSTRHANGTIME
             900

   DEFAULT USER INTERVAL:    165
   DERIVED SPIN INTERVAL:    165
   DEFAULT USER OPNOTIFY: +    3

   MAX SUPPORTED CFLEVEL: 17

   MAX SUPPORTED SYSTEM-MANAGED PROCESS LEVEL: 17

   SIMPLEX SYNC/ASYNC THRESHOLD:               27
   DUPLEX SYNC/ASYNC THRESHOLD:                30
   SIMPLEX LOCK SYNC/ASYNC THRESHOLD:          27
   DUPLEX LOCK SYNC/ASYNC THRESHOLD:           36

   CF REQUEST TIME ORDERING FUNCTION: INSTALLED

   SYSTEM STATUS DETECTION PARTITIONING PROTOCOL ELIGIBILITY:
     SYSTEM CAN TARGET OTHER SYSTEMS.
     SYSTEM IS ELIGIBLE TO BE TARGETED BY OTHER SYSTEMS.

   SYSTEM NODE DESCRIPTOR: 002817.IBM.02.0000000xxxxx
                           PARTITION: 07    CPCID: 00

   SYSTEM IDENTIFIER: 4E152817 07003B42

   NETWORK ADDRESS: IBM390PS.R91

   PARTITION IMAGE NAME: J80

   IPL TOKEN: C7F65200 F05D49A1

   COUPLEXX PARMLIB MEMBER USED AT IPL:  COUPLE00

   OPTIONAL FUNCTION STATUS:
     FUNCTION NAME                   STATUS      DEFAULT
     DUPLEXCF16                      ENABLED     DISABLED
     SYSSTATDETECT                   ENABLED     ENABLED
     USERINTERVAL                    DISABLED    DISABLED
     CRITICALPAGING                  ENABLED     DISABLED
     DUPLEXCFDIAG                    DISABLED    DISABLED
```

Recommendation: Use the defaults for EXSPATxx and also allow the FDI to default to the calculated SPIN FDI based on those excessive spin defaults.

The z/OS 1.11 IBM HealthCheck XCF_FDI will check your system(s) to ensure the FDI is set properly based on the contents of the system EXSPATxx parmlib member.

Sysplex Failure Management (SFM)

The sysplex's SFM policy dictates what automatic actions are to be taken for XCF signaling connectivity failures, system status update missing conditions, reconfiguring systems in a PR/SM environment, handling general signaling sympathy sickness, handling critical member signaling sympathy sickness, and handling coupling facility structure event response hangs. High availability configurations leverage SFM options appropriately to remove "sick" applications, middleware instances, or systems from a sysplex swiftly and automatically. Avoiding manual operator intervention in dealing with such situations is imperative to avoiding extended periods of sysplex sympathy sickness…

Sysplex Sympathy Sickness

A well-thought-out SFM policy is necessary to ensure that systems which are not fully functioning are removed from the sysplex in a timely manner. If a "sick" system is not removed from the sysplex swiftly, then other systems in the sysplex may endure sysplex sympathy sickness. Sysplex sympathy sickness surfaces when one system in the sysplex in not able to fully participate in shared sysplex activities such as: failing to update its heartbeat in the sysplex CDS, inability to send and/or receive XCF messages for one or more XCF groups, failure to release sysplex-scope resources (including global ENQs and various forms of data sharing database and file locks) in a timely fashion, or lack of XCF signaling connectivity to other systems in the sysplex.

In these cases, the other systems in the sysplex suffer because they cannot fulfill their objectives and sysplex-scope "obligations" as a result of what is happening over on the "sick" system. For example, perhaps a system cannot obtain a global ENQ resource which is held by the "sick" system, which it needs in order to perform some important function. Perhaps a system is backing up with XCF signaling buffers because the "sick" system is not receiving the messages. When the "sick" system either restores full functionality by resuming normal operation, or when the "sick" system is removed from the sysplex (and resources held by that system are released for use by the rest of the sysplex), the sympathy sickness issues are resolved.

In many cases, the sysplex sympathy sickness time causes a greater impact to an enterprise than does the "hard failure" loss of a system. Thus, a sysplex failure management policy which does not allow "sick" systems to linger in a sysplex indefinitely, causing sympathy sickness elsewhere in the sysplex, is highly desirable to maximize availability.

Status Update Missing

The most commonly reported symptom of sysplex sympathy sickness is the status update missing condition. Each system in the sysplex updates its heartbeat in the sysplex CDS every 3 seconds. The heartbeat is merely a timestamp. Each system also reads in the heartbeats of all of the other systems in the sysplex every 3 seconds. If a reading system detects that another system has not updated its heartbeat for the length of the failure detection interval (FDI), the reading system will declare the system that is failing to update its heartbeat as "status update missing."

To minimize the length of time a system is status update missing or causing sysplex sympathy sickness, an appropriate SFM policy is required. The table below notes each SFM option and the high availability configuration recommendations around that option.

Parameter	**Brief Description** **High Availability Recommendation**
CONNFAIL(YES|NO)	SFM will use the weights of the systems to determine which systems should remain in the sysplex when faced with inter-system XCF signaling connectivity problems. CONNFAIL(NO) would result in an operator being prompted to make such XCF signaling connectivity related sysplex reconfiguration decisions, which can be a very complicated process to deal with manually, and which will likely lead to sysplex sympathy sickness until an operator responds.
	Recommendation: CONNFAIL(YES)
SYSTEM NAME(*sysname*) WEIGHT(*weight*)	The *weight* of a system is a value between 1 and 9999. The *weight* should be set to reflect the relative importance of each system. *SFM will seek to "save" the system with the higher weight, or the set of systems with the higher aggregate weight*, should various kinds of failure occur.
	Recommendations: Declare *sysname* to match the MVS system name, or use NAME(*) to define default system weights for otherwise unspecified systems. A low priority system (such as a test or development system) should have a low weight, while the most important production system should have the highest weight.

ISOLATETIME(interval)	The number of seconds to wait (after the system is in a status update missing condition) to isolate a system using the fencing services through the coupling facility for a system that is not sending XCF signals and has entered status update missing condition. The isolation interval is a number of seconds between 0 and 86400.
	Recommendation: set the isolation *interval* to zero. If a system is not sending XCF signals and has not updated its heartbeat in the sysplex CDS for the length of time specified for the FDI, then the system needs to be removed from the sysplex without further delays.
SSUMLIMIT(*ssumlimit*)	Number of seconds a system can be in a status update missing condition (thus looking unresponsive) and yet still sending XCF signals (thus looking responsive) before SFM removes the system from the sysplex. The *ssumlimit* is a number of seconds between 0 and 86400.
	Recommendation: 900. This allows system programmers up to 15 minutes to successfully resolve this ambiguous sort of status update missing condition before SFM removes the system from the sysplex automatically.
PROMPT	Indicates that an operator should be prompted to reply to a WTOR if a system enters a status update missing condition.
	Recommendation: Do not use the PROMPT option if it is truly the case that an operator must physically respond to the WTOR. The failure to respond to the critical XCF WTORs in a timely manner is a prime source for sysplex sympathy sickness situations. Use ISOLATETIME instead.
MEMSTALLTIME(seconds)	Number of *seconds* after which XCF is to take action to resolve XCF signaling stall conditions. These usually reflect a problem in the XCF member (middleware or application) that is not able to receive messages from other members of its XCF group as a result. The number of *seconds* can range from 0 to 86400. Please see the MEMSTALLTIME section of this paper for additional details.
	Recommendation: 600-900
CFSTRHANGTIME(seconds)	Specifies the number of seconds that a coupling facility structure connector can remain unresponsive to a CF structure event which requires a response, before the system takes action to relieve the hang. 0 up to 1800 seconds is the valid range.

	Please see the CFSTRHANGTIME section of this paper for additional details.
	Recommendation: 900

For more information about SFM definitions please see z/OS V1R12.0 MVS Setting Up a Sysplex.

XCF Member Stall Time (MEMSTALLTIME)

XCF Signaling can schedule multiple SRBs to run in a single XCF group member's message exit at one time. When XCF group members fail to pull in their messages in a timely manner, XCF's internal message buffers start to fill, which is called a "signaling stall" condition. Eventually, most/all XCF message buffers may become consumed and the XCF signaling paths start to back up, causing sysplex sympathy sickness issues. When XCF signaling buffers are a shared resource that is common to different XCF groups, a "signaling stall" may impact the middleware or applications associated with all of the XCF groups that are sharing the message buffer space.

When an XCF group stalls, IXC431I and IXC430E will be issued. Sample messages:

```
IXC431I GROUP xxxxxxxx MEMBER yyyy JOB zzzzzzzz ASID 0219
   STALLED AT 04/28/2010 02:31:46.676092 ID: 0.1
   LAST MSGX: 04/28/2010 02:31:42.046513  18 STALLED     14 PENDINGQ
   LAST GRPX: 04/28/2010 01:40:45.922416   0 STALLED      0 PENDINGQ
```

IXC431I is issued after 3 minutes of being stalled.

```
IXC430E SYSTEM ssss HAS STALLED XCF GROUP MEMBERS
```

IXC430E is issued after 5 minutes of being stalled.

```
IXC631I GROUP xxxxxxxx MEMBER yyyy JOB zzzzzzzz ASID 0219
   STALLED, IMPACTING SYSTEM ssss

IXC640E STALLED XCF GROUP MEMBERS ON SYSTEM ssss IMPACTING SYSPLEX
```

IXC631I and IXC640E are issued when other system(s) in the sysplex are being impacted by stalls. All of the messages noted above are issued on the system that has the stalled member(s).

We recommend setting the MEMSTALLTIME to 600-900 to allow XCF to take automatic actions to resolve matters when an XCF stall condition reaches the point of impacting other systems in the sysplex and the MEMSTALLTIME interval expires. When the MEMSTALLTIME expires, XCF will request a dump of the member consuming the most buffers to gather serviceability information, and then start to terminate the member. XCF will terminate the join task, jobstep task then ASCBXTCB. If the stall condition is still not relieved by these actions, the ASID will be MEMTERMed. If the group consuming the most XCF buffers is SYSGRS, SYSMCS, SYSXCF, or * XCF then the system on which the stalled members is running will be

removed from the sysplex, because a system cannot function without those members operating normally.

When MEMSTALLTIME results in an action being taken, message IXC615I is issued.

```
IXC615I GROUP grpname MEMBER membername JOB jobname ASID asid
SFM TERMINATING what to RELIEVE SYMPATHY SICKNESS

What =  JOIN TASK
   JOBSTEP TASK
   ADDRESS SPACE
   SYSTEM

IXC615I GROUP grpname MEMBER membername JOB jobname ASID asid
SFM MEMTERMING ADDRESS SPACE TO FORCE COMPLETION
```

Once the stall condition is relieved the following messages will appear:

```
IXC432I GROUP grpname MEMBER membername JOB jobname ASID asid text AT
ResumeDate ResumeTme ID: stall#

IXC632I GROUP grpname MEMBER membername JOB jobname ASID asid
NO LONGER IMPACTING SYSTEM sysname
```

CF Structure Hang Time (CFSTRHANGTIME)

When a connector to a structure fails to respond to a structure-related event (for example, requesting its participation in a structure rebuild or other recovery process) in a timely manner, it is possible that all of the connectors to the structure will be impacted, and for the application or middleware that is relying on the use of the CF structure to also be impacted. The most common example is a connector failing to respond to a phase of a rebuild (or other sysplex-wide process). The rebuild cannot progress unless all connectors respond to each phase of the rebuild. When one or more responses are missing, the structure may remain unusable for an extended period of time. The impact of this varies depending on the type of application or middleware product that is using the structure.

Recommendation: CFSTRHANGTIME 900, which will provide operators with 15 minutes to resolve the hang. If the hang is not resolved in 15 minutes then XES will start to take automatic actions to attempt to resolve the hang condition.

When the CFSTRHANGTIME is reached, the following corrective actions may be taken, depending on conditions at the time:

(1) Stop the rebuild
(2) Stop signaling path (XCF structures only)
(3) Force a disconnect (XCF signaling structures only)
(4) Terminate the connector's task
(5) Terminate the connector's address space
(6) Partition the connector system

For GRS, the system will be partitioned out of the sysplex immediately because GRS specifies TERMLEVEL=SYSTEM when it connects to the ISGLOCK structure via IXLCONN.

IXL040E or IXL041E will be issued to report a connector failing to respond to a structure event. Prior to z/OS 1.12, IXL042I and IXL043I were issued to indicate the response was received or is no longer needed.

```
IXL040E   CONNECTOR NAME: connector-name, JOBNAME: jobname, ASID: asid HAS
text. process FOR STRUCTURE structure-name CANNOT CONTINUE.
MONITORING FOR RESPONSE STARTED: mondate montime. DIAG: x

IXL041E   CONNECTOR NAME: connector-name, JOBNAME: jobname, ASID: asid HAS
NOT RESPONDED TO THE event FOR SUBJECT CONNECTION:
subject-connector-name. process FOR STRUCTURE structure-name
CANNOT CONTINUE. MONITORING FOR RESPONSE STARTED: mondate
montime.  DIAG: x

IXL042I   CONNECTOR NAME: connector-name, JOBNAME: jobname, ASID: asid  HAS
action. THE REQUIRED RESPONSE event FOR STRUCTURE structure-name
IS NO LONGER EXPECTED.

IXL043I   CONNECTOR NAME: connector-name JOBNAME: jobname ASID: asid HAS
action.  THE REQUIRED RESPONSE FOR THE event FOR SUBJECT
CONNECTION subject-connector-name, STRUCTURE structure-name IS
NO LONGER EXPECTED.
```

z/OS 1.12 introduced new messages IXL047I, IXL048I, IXL049E and IXL050I. The messages indicate a response is no longer required, action to be taken or an action was taken to resolve the hang.

```
IXL047I   THE RESPONSE REQUIRED FROM CONNECTOR NAME: conname TO STRUCTURE
strname, JOBNAME: jobname, ASID: asid  responsetype IS NO LONGER
EXPECTED.  REASON: reason

IXL048I   THE RESPONSE REQUIRED FROM CONNECTOR NAME: conname TO STRUCTURE
strname, JOBNAME: jobname, ASID: asid FOR THE event  FOR SUBJECT
CONNECTION subjectconnname IS NO LONGER EXPECTED. REASON: reason

IXL049E   HANG RESOLUTION ACTION FOR CONNECTOR NAME: conname TO STRUCTURE
strname, JOBNAME: jobname, ASID: asid: actiontext

IXL050I   CONNECTOR NAME: conname TO STRUCTURE strname, JOBNAME: jobname,
ASID: asid  HAS NOT PROVIDED A REQUIRED RESPONSE AFTER
noresponsetime SECONDS. TERMINATING termtarget TO RELIEVE THE
HANG.
```

It is better to lose one connector to a structure than to have all connectors to a structure hang. It is possible for the connector to reconnect to the structure after having been terminated, and hopefully resume normal operation. Meanwhile, the remaining connectors to the structure will be better able to fulfill their obligations without the unresponsive connector.

To emphasize the importance of CFSTRHANGTIME, consider the enhanced catalog structure, ECS, being hung during a rebuild process. When the ECS structure is being rebuilt, data set allocations cannot occur because catalog functions cease temporarily. Given this, system logger will not be able to allocated new offload datasets to offload data into. System logger will not be able to allocate or browse older offload datasets.

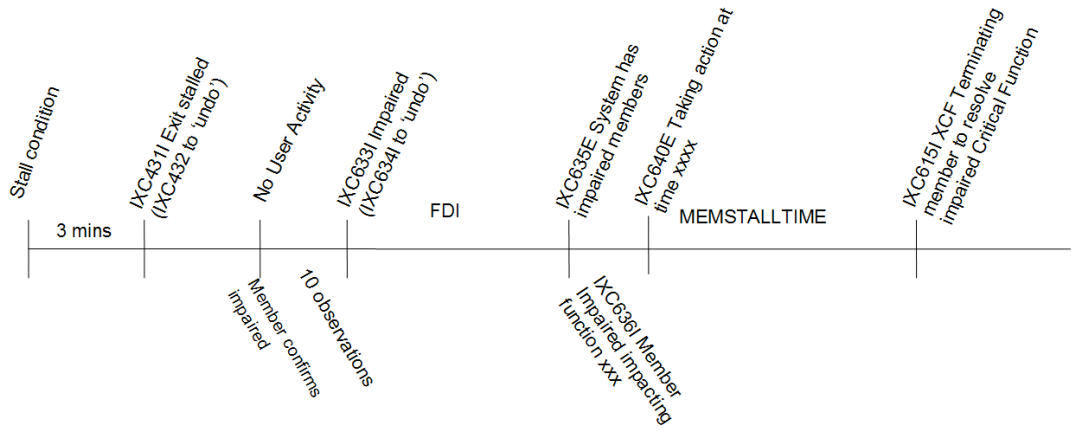
TSO users will not be able to log on to systems to do problem determination. Many different applications on all systems in the sysplex will suffer until the ECS structure rebuild is able to complete. CFSTRHANGTIME will take action and terminate the connector who appears unresponsive when the specified time is reached. After terminating the hung ECS structure connector the sysplex can continue operating.

CRITICAL MEMBERs

An application which joins an XCF group may specify CRITICAL=YES on the IXCJOIN macro, and thereby designate itself as a "critical member" to XCF. Members can update their member status via a member status exit. Along with CRITICAL=YES, the member specifies a TERMLEVEL. TERMLEVEL indicates what action should be taken when the member fails to update its status. TERMLEVEL options include MEMASSOC, ADDRSPACE and SYSTEM. MEMASSOC indicates the member should be canceled, ADDRSPACE indicates the address space associated with the member should be canceled and SYSTEM indicates the system should be removed from the sysplex. If TERMLEVEL SYSTEM is declared and the critical member goes "member status update missing," then the system on which the member was running will be terminated.

In z/OS 1.12, the GRS component joins as a critical XCF member. GRS specifies TERMLEVEL SYSTEM. If GRS is not functioning on one system, the entire sysplex can be impacted. Again, the goal of critical members support is to detect and avoid the possibility of sysplex sympathy sickness as much as possible.

The timeline and new messages associated with critical member support:



Stall condition = At least 1 exit stalled for 30 seconds or work item on head of queue for 30 seconds

No User Activity = All scheduled user exits stalled for at least 30 seconds or No user exits scheduled

System Status Dectection via BCPii (SYSSTATDETECT)

There are two major aspects to removing a system from a sysplex. First, the system must become "I/O isolated" so that it can no longer access nor modify any shared data in the sysplex environment; and second, the system must be logically removed from the sysplex so that sysplex-scope serialization and other sysplex-scope resources can be taken away from that system and made available to other systems in the sysplex.

With the SYSSTATDETECT function enabled for use, z/OS 1.11 or higher will utilize BCPii functions to determine if a system whose status update is missing is actually in a known state where it cannot possibly resume processing without a re-IPL (perhaps in a nonrestartable disabled wait state, or on an image that has been reset or deactivated). If a system can not resume processing, and other systems in the sysplex can discover this through BCPii, then z/OS can automatically partition the failed system out of the sysplex without waiting for the failure detection interval (FDI) to expire – it can be removed at once!

The benefit of the SYSSTATDETECT function is that truly "dead" systems will be logically removed from the sysplex in a much more timely fashion, which will minimize the scope and duration of sysplex sympathy sickness.

Enablement of the SYSSTATDETECT function is highly recommended to maximize the availability of the sysplex. After BCPii is configured, you will need to format a sysplex CDS with:

ITEM NAME(SYSSTATDET) NUMBER(1)

Please see z/OS Setting Up a Sysplex "Using the System Status Detection Partitioning Protocol and BCPii for Availability and Recovery."

AUTOIPL

To minimize down time and bring a failed system back into the sysplex as quickly as possible, and/or maximize first failure data capture for a problem that leads to a z/OS disabled wait state, the AUTOIPL function can be leveraged. Using AUTOIPL, a standalone dump can be taken automatically for a pre-defined set of non-restartable disabled WAIT STATE codes. Alternatively, if the goal is to bring the z/OS system back into the sysplex as quickly as possible, AUTOIPL can be structured accordingly. Both functions can also be requested – AUTOIPL can initiate both a standalone dump and a re-IPL of the failed z/OS system once the standalone dump processing completes. Regardless, the goal of AUTOIPL is to eliminate manual intervention in dealing with failure situations that lead to a z/OS non-restartable disabled wait state. Please see z/OS V1R12.0 MVS Planning: Operations 6.4 Exploiting the automatic IPL function.

Default Partitioning Process Change

As stated above, we strongly recommend creating a proper SFM policy as a means to achieving sysplex high availability. If an installation does not have an SFM policy in use at all, then as of z/OS 1.11 XCF will nevertheless proactively seek to reduce sysplex sympathy sickness, as follows.

At z/OS 1.11 and above, XCF will attempt to automatically partition a system when it enters a status update missing situation, after the failure detection interval has expired. This behavior is identical to having an SFM policy with ISOLATETME(0) specified. This change was made to minimize the scope and duration of sysplex sympathy sickness even for sysplexes where an SFM policy has not been implemented.

In addition, when an operator initiates a VARY XCF to manually remove a system from the sysplex on zOS 1.11 and above, XCF will automatically fence (I/O isolate) the system if there is proper coupling connectivity established to enable XCF to do so, and XCF will then complete the partitioning action. This change was made to ensure that systems leaving the sysplex are automatically isolated if possible. Prior to this change, without SFM, the operator would have been prompted to manually RESET the system and then respond to the IXC102A prompt. Now, when XCF is able to automatically isolate a system, the IXC102A message will not be issued, and manual intervention in the partitioning process will not be needed.

The message pattern when manually partitioning a system from a sysplex, using the VARY XCF command at zOS 1.11 and above, will be similar to the following.

Operator initiates the VARY XCF offline for a system:

```
V XCF,SA0,OFFLINE
*0377 IXC371D CONFIRM REQUEST TO VARY SYSTEM SA0 OFFLINE. REPLY
 SYSNAME=SA0 TO REMOVE SA0 OR C TO CANCEL.
```

Operator confirms the removal of the system:

```
R 377,SYSNAME=SA0
IEE600I REPLY TO 0377 IS;SYSNAME=SA0
```

XCF initates sysplex partitioning in response to the VARY XCF request:

```
IXC101I SYSPLEX PARTITIONING IN PROGRESS FOR SA0 REQUESTED BY
*MASTER*. REASON: OPERATOR VARY REQUEST
```

Communication with SA0 stops swiftly and the system trying to partion SA0 tries to restart paths as logical sysplex partitioning has not completed:

```
IXC467I RESTARTING PATHIN DEVICE F710
        USED TO COMMUNICATE WITH SYSTEM SA0
        RSN: INTERVENTION REQUIRED
IXC467I RESTARTING PATHIN DEVICE F711
        USED TO COMMUNICATE WITH SYSTEM SA0
        RSN: INTERVENTION REQUIRED
```

```
IXC467I RESTARTING PATHIN DEVICE F712
        USED TO COMMUNICATE WITH SYSTEM SA0
        RSN: INTERVENTION REQUIRED
```

Clean up of lock structures:

```
IXL030I CONNECTOR STATISTICS FOR LOCK STRUCTURE IGWLOCK00,
CONNECTOR SD0:
      00030532 00000000 00000010 00F0000F 0064

      00000000 00000000 00000013 00000000
      00000120 000003F1 00000000 0000523C
      00000000 00000000 00000000 00000000
      00000000 00000000 00000000 00000000

      00000001 00000000 00000004 00000000
      00000001 00000002 00000000 0002066A
      00000000 00000000 00000000 00000000
      00000000 00000000 00000000 00000000

      00000002 00000000 00000008 00000000
      00000021 0000010F 00000000 00021178
      00000000 00000000 00000000 00000000
      00000000 00000000 00000000 00000000
IXL020I CLEANUP FOR LOCK STRUCTURE IGWLOCK00,
CONNECTION ID 04, STARTED BY CONNECTOR SD0
INFO: 0001 000404AB 00000044
IXL021I GLOBAL CLEANUP FOR LOCK STRUCTURE IGWLOCK00,
CONNECTION ID 04, BY CONNECTOR SD0
HAS COMPLETED.
INFO: 00000000 00000000 00000000 00000000 00000000 00000000
      00000000
IXL022I LOCAL CLEANUP FOR LOCK STRUCTURE IGWLOCK00,
CONNECTION ID 04, BY CONNECTOR SD0
HAS COMPLETED.
INFO: 0000000C 00000000 00000000 00000241 00000000 00000000
      00000000
IXL023I CLEANUP FOR LOCK STRUCTURE IGWLOCK00,
CONNECTION ID 04, BY CONNECTOR SD0
HAS COMPLETED.
INFO: 00000034 0000004B 00000000 00000000 00000000 00000000
      01000000 000000C6 00000048 00000000 00000000 000000C5
      00000047 00000000 00000000 00000000 00000000
```

Fencing of system automatically initated by XCF:

```
IXC108I SYSPLEX PARTITIONING INITIATING FENCE
  SYSTEM NAME:          SA0
  SYSTEM NUMBER:        02001F35
  SYSTEM IDENTIFIER:    4D852097 01001F35
```

Assuming fencing was successful:

```
IXC109I FENCE OF SYSTEM SA0 SUCCESSFUL.
```

BCPII connection for the partitioned system is released:

```
IXC113I BCPII CONNECTION TO SYSTEM SA0 RELEASED
  DISCONNECT REASON:    SYSTEM REMOVED FROM SYSPLEX
  IMAGE NAME:           SA0
  NETWORK ADDRESS:      IBM390PS.H131
  SYSTEM NUMBER:        02001F35
  IPL TOKEN:            C5EEBD05 023B9D98
```

XCF stopping paths to system that was removed:

```
IXC467I STOPPING PATHOUT STRUCTURE IXCPLEX_PATH3 LIST 67
        USED TO COMMUNICATE WITH SYSTEM SA0
        RSN: SYSPLEX PARTITIONING OF REMOTE SYSTEM
IWM051I STRUCTURE(SYSZWLM_4D852097), FOR SYSTEM SA0 CLEANED UP
IXC467I STOPPING PATHOUT STRUCTURE IXCPLEX_PATH2 LIST 54
        USED TO COMMUNICATE WITH SYSTEM SA0
        RSN: SYSPLEX PARTITIONING OF REMOTE SYSTEM
IXC467I STOPPING PATHOUT STRUCTURE IXCPLEX_PATH1 LIST 49
        USED TO COMMUNICATE WITH SYSTEM SA0
        RSN: SYSPLEX PARTITIONING OF REMOTE SYSTEM
IXC467I STOPPING PATHOUT STRUCTURE IXCPLEX_PATH6 LIST 67
        USED TO COMMUNICATE WITH SYSTEM SA0
        RSN: SYSPLEX PARTITIONING OF REMOTE SYSTEM
```

XCF completed partitioning of system:

```
IXC105I SYSPLEX PARTITIONING HAS COMPLETED FOR SA0
- PRIMARY REASON: OPERATOR VARY REQUEST
- REASON FLAGS: 000004
```

XCF Signaling Configuration

XCF signaling is used by applications to send messages from one member of a group to another member of a group. The various members may reside on the same system or on different systems in the sysplex. To ensure communication flow is uninterrupted, a high availability XCF signaling configuration is desired.

The following points contribute to a high availability XCF signaling configuration

(1) Multiple XCF Signaling Structures, appropriately sized, and allocated across multiple CFs that each have redundant coupling link connectivity to all systems in the sysplex.

(2) Multiple CTCs between systems (if CTCs are to be used). For the highest sysplex availability, it is recommended that BOTH XCF signaling structures in the CF, and CTC links between systems, be used. Having signaling paths using different transport technologies provides the most failure-isolation and redundancy.

(3) TCLASS definitions – adequate separation of message traffic by size, with - multiple paths assigned to each TCLASS

(4) MAXMSG allotment – enough space to send and receive peak message traffic volumes without encountering a significant number of "buffer full" or "reject" conditions

Please see IBM Redbook System z Parallel Sysplex Best Practices "XCF Signaling Paths" and "Transport classes" for further details.

Mirroring Couple Datasets

Simply put, the only couple data set that should ever be synchronously mirrored via disk replication technologies is the LOGR CDS. However, asynchronous mirroring of CDSs of any type is permitted.

Extreme caution should be used when IPLing systems using CDSs images which are mirrored copies of couple data sets from other sysplexes in order to avoid several different types of potentially high-impact outages. z/OS Hot Topics Newsletter #24 article "Mirror, mirror, on the wall, should couple datasets be mirrored at all?" documents the known pitfalls associated with mirroring CDSs. Please review these major pitfalls and establish procedures to ensure each issue is avoided.

CRITICALPAGING

If the system uses DASD Swap technologies (such as IBM HyperSwap) then CRITICALPAGING enablement is strongly recommended to achieve a high availability environment. IBM DASD Swap technologies are Basic HyperSwap and GDPS HyperSwap Manager; other vendors also provide similar DASD Swap capabilities. Please reference IBM Washington System Center FLASH10733.

The CRITICALPAGING function serves to "harden" critical z/OS address spaces and storage areas against paging, as paging I/O to DASD devices cannot occur at certain times during the DASD Swap processing.

System Console (HMC)

There are a few (very rare) problems which may require the use of the hardware system console on the HMC to resolve; for example, there are situations in which it may be necessary to use the operating system messages area of the HMC as a "console of last resort" for managing z/OS images. Due to this fact, it is imperative that the operations staff be able to access the HMC in an extremely timely fashion and that all CECs in the configuration are defined to use an HMC system console.

z/OS HealthChecks

XCF and XES (and a number of other sysplex-enabled functions and components) have been at the forefront in providing z/OS HealthChecks since the inception of this capability. We recommend enabling all delivered HealthChecks to perform their appointed checking, and then investigating all warnings/exceptions reported by those HealthChecks and resolving them whenever possible.

Installations should periodically review both the HealthChecks that have been disabled, and those where the default checking parameters have been overridden in some way by a PARM specification for a check, and ensure that the reason for overriding the check's default behavior is understood, and remains valid. In other words, the installation should try to maximize the benefits of the HealthChecker function by minimizing the amount of checking that is being suppressed or overridden in some way.

The current list of XCF/XES HealthChecks is given below. Please see IBM Health Checker for z/OS User's Guide for additional insight about each check.

XCF_CDS_MAXSYSTEM
XCF_CDS_SEPARATION
XCF_CDS_SPOF
XCF_CF_ALLOCATION_PERMITTED
XCF_CF_CONNECTIVITY
XCF_CF_MEMORY_UTILIZATION
XCF_CF_PROCESSORS
XCF_CF_STR_AVAILABILITY
XCF_CF_STR_DUPLEX
XCF_CF_STR_EXCLLIST
XCF_CF_STR_NONVOLATILE
XCF_CF_STR_POLICYSIZE
XCF_CF_STR_PREFLIST
XCF_CF_SYSPLEX_CONNECTIVITY
XCF_CFRM_MSGBASED
XCF_CLEANUP_VALUE
XCF_DEFAULT_MAXMSG
XCF_FDI
XCF_MAXMSG_NUMBUF_RATIO
XCF_SFM_ACTIVE
XCF_SFM_CFSTRHANGTIME
XCF_SFM_CONNFAIL
XCF_SFM_SSUMLIMIT
XCF_SFM_SUM_ACTION
XCF_SIG_CONNECTIVITY
XCF_SIG_PATH_SEPARATION

XCF_SIG_STR_SIZE
XCF_SYSPLEX_CDS_CAPACITY
XCF_SYSSTATDET_PARTITIONING
XCF_TCLASS_CLASSLEN
XCF_TCLASS_CONNECTIVITY
XCF_TCLASS_HAS_UNDESIG

Automation for Critical Sysplex Messages

To ensure that critical system/sysplex messages which are important for availability reasons are noticed, investigated and reacted to as swiftly as possible, it is highly recommended that certain sysplex messages be automated via your installation's automation product. Minimally, the automation should provide alerting and notification to the operations staff so that the messages are observed, understood, and responded to expeditiously. Even better, it may be possible to design automation scripts which can actually respond to some of these messages (WTORs) without human involvement. The following table identifies XES, XCF and GRS messages that should be automated upon. The table also states the suggested automation action.

Message	**Suggested Action**
IXC102A	Respond down after system has been reset. Action: Verify that the indicated system has been reset (or cause it to be reset) and respond DOWN immediately.
IXC409D	XCF Signaling paths between systems were lost Action: Respond with the name of the system to remove from the sysplex. Note that in most cases, SFM processing for CONNFAIL(YES) can address issues such as these.
IXC426D	System is sending signals but not updating its heartbeat. Action: Investigate the status of the indicated system and and react before sysplex sympathy sickness ensues. Respond with the system to take down if unable to understand why the system is only partially responsive, or to resolve that issue. If SSUMLIMIT is specified in the SFM policy, and the time limit is exceeded, XCF will take automatic action in this case.
IXC430E IXC431I	System has stalled XCF members, i.e. XCF members who are not receiving and processing their inbound messages in a timely fashion. Action: Look for IXC432I indicating the situation has been resolved. If there is no IXC432I begin problem determination. Assess if there is a resource constraint causing the member to be unable to process messages.

	Possible constraints include: resource contention, lack of processor capacity, auxiliary storage shortage, real storage shortage, message exit SRBs looping, etc. If deeper investigation is required the following slip can be used to collect a dump of XCFAS and the stalled member when the IXC431I is issued. SLIP SET,COMP=00C,REASON=020F0006, JOBLIST=(XCFAS,stalled_job),DSPNAME=('XCFAS'.*), SDATA=(COUPLE,ALLNUC,LPA,LSQA,PSA,SWA,RGN,SQA,TRT,CSA,GRSQ, XESDATA,SUM),END
IXC585E	FULLTHRESHOLD has been exceeded for the entries or elements of a structure. If ALLOWAUTOALT(YES) is defined for the structure XES may start to alter the structure attributes in order to attempt to relieve the structure full condition. Action(s): Check for IXC530I alter started, IXC588I alter started and targets, IXC590I altered ended and results of alter, and IXC586I indicating alter completed. Assess if the workload, consider the possibility that the workload has increased or changed significantly. If the alters are pervasive, disruptive to ongoing workload processing, or extremely long in duration, revisit the CFSizer or investigate the workload for the structure
IXC631I IXC633I IXC635E IXC636I IXC640E	Stalled XCF groups are impacting other systems in the sysplex. Stalled XCF groups are impacting the sysplex. Action(s): Check for relief messages IXC632I and IXC634I If MEMSTALLTIME is specified in the SFM policy, and the time limit is exceeded, XCF will collect documentation and take action to resolve the stall.
IXL008I	Path to CF has been invalidated. Action: D CF to determine if corrective action for the CF paths needs to be taken. The path may have been miscabled/misconfigured or incorrectly defined.
IXL040E IXL041E	Connector has not responded to a structure event. An ABEND026 RSN08118001 dump will be taken. Report a problem to the application that is failing to respond. Actions: Check to see if IXL042I or IXL043I, or, IXL049E or IXL050E has been issued, indicating the situation has been resolved. If the situation is still occurring, check to see if the address space is waiting on a resource

	(ENQ, latch, aux storage shortage, real storage shortage, processor, etc.). If CFSTRHANGTIME is specified in the SFM policy and the time limit is exceeded, the corrective action will be taken by the system. If CFSTRHANGTIME is not set, take action to resolve the hang, possibly by terminating the connection.
IXL044I	Persistent IFCCs (interface control checks) for a coupling facility were detected. Action(s): Consider collecting a nondisruptive dump of the CF while the problem is occurring, see Non-Disruptive CF Dump section of this paper. Also consider activating SYSXES ctrace collecting dumps on all systems in the sysplex. The slip below must be set on all system in the sysplex. When the slip is triggered a dump of XCFAS and its dataspaces will be captured on all systems in the sysplex. Also, when the slip is triggered the corresponding slips will be disabled on all the other systems in the sysplex to avoid taking multiple sets of sysplex dumps. Contact the IBM Hardware Support Center. TRACE CT,2M,COMP=SYSXES, R XX,OPTIONS=(HWLAYER,REQUEST,LOCKMGR),END Note: SUB=(GLOBAL) and SUB=(structurename) can be added if the problem structure is known. SLIP SET,ACTION=SVCD,MSGID=IXL044I,ID=IXL,IDGROUP=IXL4, JOBLIST=(XCFAS),DSPNAME=('XCFAS'.*), SDATA=(ALLNUC,CSA,PSA,LPA,LSQA,NUC,RGN,SQA,SUM,SWA,TRT,XESDATA,COUPLE), REMOTE=(DSPNAME,SDATA,JOBLIST),END
IXL045E	XES SRBs are encountering delays. Action(s): Determine if the system is overburdened and resolve the system processing or throughput bottleneck. Consider taking a dump while the condition is occurring and contact the IBM Software Support Center (compid 5752SCIXL). Console dump: DUMP COMM=(IXL045E) JOBNAME=(XCFAS,impacted_job),DSPNAME=('XCFAS'.*), SDATA=(ALLNUC,CSA,PSA,LPA,LSQA,NUC,RGN,SQA,SUM,SWA,TRT,XESDATA,COUPLE), REMOTE=(SYSLIST=*('XCFAS','impacted_job'),DSPNAME,SDATA),END Slip to capture dump upon recreate:

	SLIP SET,ACTION=SVCD,MSGID=IXL045E, JOBLIST=(XCFAS),DSPNAME=('XCFAS'.*), SDATA=(ALLNUC,CSA,PSA,LPA,LSQA,NUC,RGN,SQA,SUM,SWA,TRT,XESDATA,COUPLE), REMOTE=(DSPNAME,SDATA,JOBLIST),END
IXL158I / IXL157I	Path to CF has become not operational and path operational messages. Paths to coupling facilities should come online and stay online until VARYd or CONFIGUREd offline. Dropping or toggling of links is unexpected and should be investigated. Action: Verify the desired configuration for that path, configure the links online as needed..

Path Busy Conditions

When evaluating the occurrence of path busy conditions, generally no remedial action is required unless the percentage of requests which experienced one or more path busy conditions (count of path busies / total number of CF requests) is greater than 10%.

The actions that may be considered by the installation to try to reduce the impact of excessive path busy conditions include the following:

- Add more shared coupling CHPIDs to the CF, to increase the total number of link buffers that are available to the sharing z/OS LPARs.

- Decrease the "extent of sharing" of the coupling CHPIDs. Consider sharing the same number of coupling CHPIDs across fewer z/OS LPARs, or even dedicating coupling CHPIDs to just one z/OS LPAR.

- Attempt to eliminate one or more of the configuration or workload conditions described above which may be exacerbating the occurrence of path busy conditions.

The Path Busy recommendation pertains to environments where the PTFs for OA35117 are applied to all systems. OA35117 implemented accounting changes required to remove the sensitivity to machine speeds which may cause over-reporting when accumulating path busy counts. Code changes are required to ensure the accumulation is consistent across different machine generations. For more details please see OA35117.

Non-disruptive Coupling Facility Dump

System z parallel sysplex continuously strives to achieve 99.9999% availability. To attain this goal any issues on the platform must be resolved swiftly. First failure data capture is paramount to resolving issues swiftly. In 2010, IBM introduced the nondisruptive CF dumping capability. The nondisruptive dumps make it possible to obtain dumps of the coupling facility without having to terminate the CF. The nondisruptive dump enables IBM support to identify root cause of CFCC defects quickly.

There exists a class of problems which impact exploiters of coupling facility structures for which the root cause is an issue on the coupling facility. For example, sometimes a break duplexing event is triggered or a loss of a path to a coupling facility occurs as a result of a delay or temporary congestion in the coupling facility.

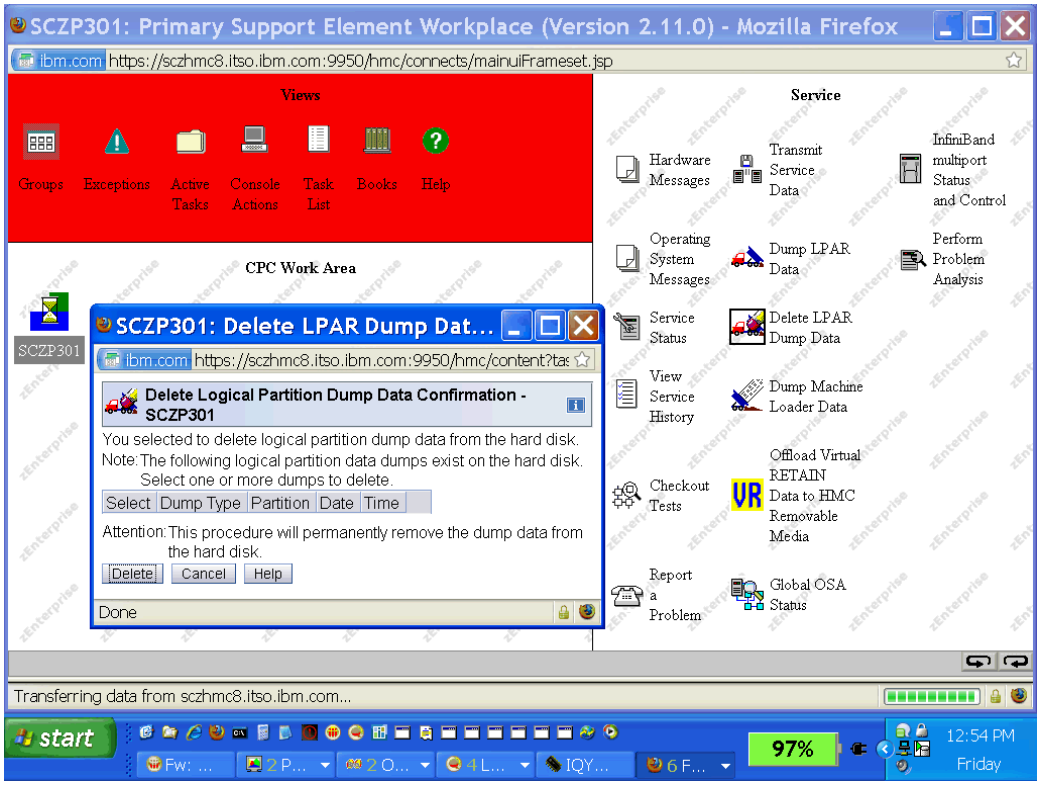
Historically, SYSXES CTRACE (component trace) was collected to observe the flow of commands going to the coupling facility and the responses from the coupling facility, from the outside. Unfortunately, observing the "to and from" flow of commands does not always permit IBM support to clearly identify precisely what was happening inside the coupling facility at the time of the error. In order to identify the cause of the error, a dump of the coupling facility itself may be required.

Prior to z196 processors, a dump of a coupling facility was a disruptive action that resulted in a coupling facility terminating. As a result, all of the structures had to rebuild to another coupling facility. Further, if the problem involved a duplexed structure then disruptive dumps of BOTH coupling facilities at the same time were required to reach root cause. For sysplexes with only two coupling facilities, taking two disruptive dumps for the two CFs resulted in a sysplex-wide outage! So, for improved serviceability and problem determination on z196, the ability to take a nondisruptive dump of a coupling facility was provided. The ability to take a nondisruptive dump was subsequently rolled back to the z10 via CFCC 16 Service Level 4.01. The z10 nondisruptive CFCC dumps phone home automatically. z196 nondisruptive dumps phone home automatically at DR86 SYSTEM EC N29802 MCL301 and above.

With the new non-disruptive CF dumping support, when a break duplex event is encountered, non-disruptive dumps may be taken on both CFs involved in the break. If the break occurs, z/OS ABEND026 dumps may also be taken on all z/OS LPARs physically connected to the coupling facility. The z/OS dumps will be taken on the z/OS LPARs regardless of whether they are actually using the affected CF or have active connectors to the impacted structure.

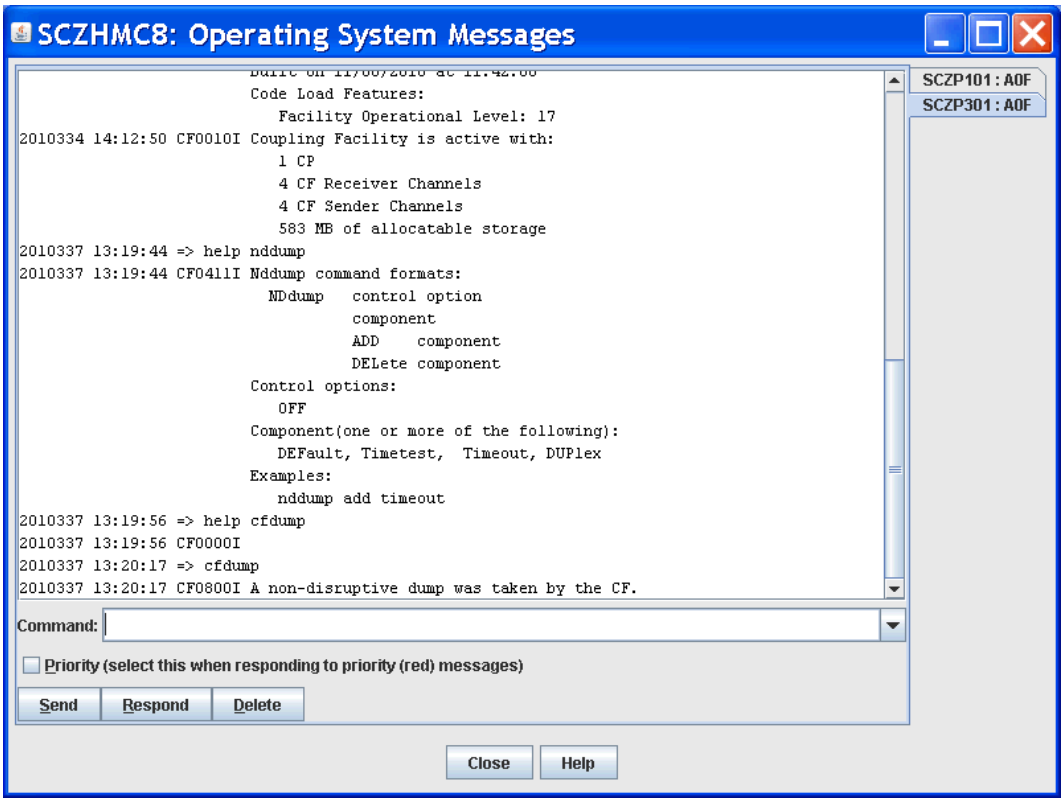
For other situations where the coupling facility is believed to be involved, system programmers or the CE can manually initiate a nondisruptive dump of the coupling facility. The steps to initiate a nondisruptive dump are:

(1) Delete existing LPAR dumps using the Service Task on the SE.

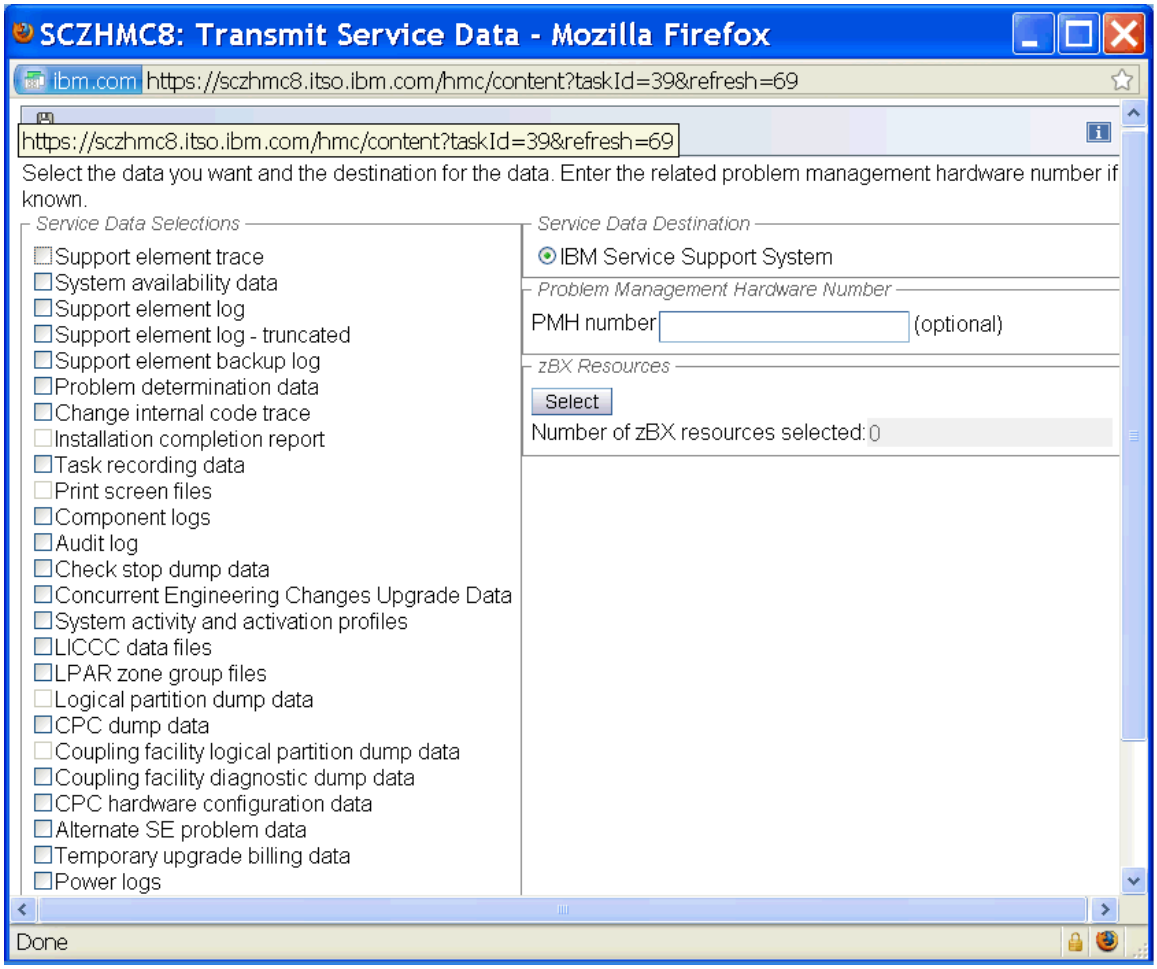


(2) On the OPERMSG console for the CFCC image, you will use the CFDUMP command to initiate the dump.

(3) Message CF0800I will appear indicating the dump has occurred. Please be aware that might take a few minutes once you enter the command. BE PATIENT!



(4) Transmit the support element log, system availability data, problem determination data, coupling facility logical partition dump data and coupling facility diagnostic dump data.



When a nondisruptive dump is initiated by the coupling facility, IXL051E will be issued by z/OS.

```
IXL051E   dumptype DUMP OF COUPLING FACILITY cfname
type.mfg.plant.sequence PARTITION: partition side CPCID: cpcid
INITIATED BY THE requestor [FOR STRUCTURE strname SID sid] [DIAG
DATA: diagdata]
```

With the PTFs for OA35342 applied a z/OS operator command can be issued to dump a coupling facility. Please refer to the hold data for OA35342 once it is generally available.

We recommend that the nondisruptive dumping procedures be tested in a non-production environment to establish familiarity with the protocol in the event that such a dump is required to diagnose a problem in the future. This is analogous to the "best practice" to ensure that standalone dump procedures are updated for each new release of z/OS.

Conclusions

If you have analyzed and implemented all of the recommendations contained herein, then your enterprise is configured to maximize availability. Congratulations on the successful completion of your mission!

Trademarks

A full list of U.S. trademarks owned by IBM may be found at: http://www.ibm.com/legal/copytrade.shtml .

Feedback

Please send comments or suggestions for changes to surman@us.ibm.com and nfagen@us.ibm.com .

Acknowledgements

The authors would like to thank all the many contributors and reviewers of Mission: Available, the IBM Parallel Sysplex Development and Support teams. A special thanks to Mark Brooks, William Neiman and Daniel Rinck.