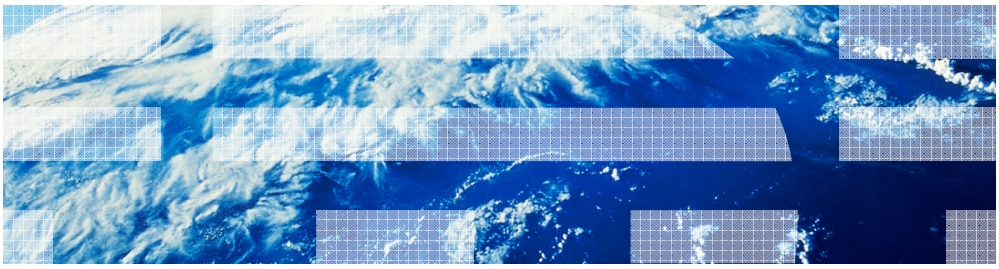
Mark Brooks – Parallel Sysplex Development mabrook@us.ibm.com
Nicole Fagen – Parallel Sysplex Support Team Lead nfagen@us.ibm.com
March 2011

Session 9028:
Parallel Sysplex Resiliency



A properly configured parallel sysplex can deliver near continuous availability. Often "properly configured" is equated with redundancy. Yes, redundancy is certainly a critical factor for enabling a sysplex to deliver on its promise of availability. But it is not sufficient. What one really needs is for the sysplex to be resilient. That is, the sysplex needs to be able to quickly resume normal operation after experiencing illness (such as sympathy sickness), change (such as reconfiguring of hardware or software), or misfortune (such as failures). In this presentation, we take redundancy as a given and explore ways to improve the resiliency of the sysplex. The topics were selected based on real world customer experiences. In particular we look at things like Sysplex Failure Management (SFM) parameters, the use of BCPii to detect failed systems, procedures for upgrading Coupling Facilities, Health Checks, and more.

Trademarks

The following are trademarks of the International Business Machines Corporation in the United States and/or other countries.

IBM*	FICON*	System x*
IBM (logo)*	IMS	System z*
ibm.com*	Parallel Sysplex®	System z9®
AIX*	POWER7	System z10
BladeCenter*	ProtecTIER*	Tivoli*
DataPower*	RACF*	WebSphere*
CICS*	Rational*	XIV*
DB2*	Redbooks®	zEnterprise
DS4000*	Sysplex Timer®	z/OS*
ESCON®	System Storage	z/VM*
* Registered trademarks of IBM Corporation		z9®

The following are trademarks or registered trademarks of other companies.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.
Cell Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license there from.
Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.
Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.
InfiniBand is a trademark and service mark of the InfiniBand Trade Association.
Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.
UNIX is a registered trademark of The Open Group in the United States and other countries.
Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.
ITIL is a registered trademark, and a registered community trademark of the Office of Government Commerce, and is registered in the U.S. Patent and Trademark Office.
IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency, which is now part of the Office of Government Commerce.

* All other products may be trademarks or registered trademarks of their respective companies.

Notes:

Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance ratios stated here.

IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.

All customer examples cited or described in this presentation are presented as illustrations of the manner in which some customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual customer configurations and conditions.

This publication was produced in the United States. IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. Consult your local IBM business contact for information on the product or services available in your area.

All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the performance, compatibility, or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

Prices subject to change without notice. Contact your IBM representative or Business Partner for the most current pricing in your geography.

For a complete list of IBM Trademarks, see www.ibm.com/legal/copytrade.shtml

Agenda

- Resiliency
- Sympathy Sickness
- SFM
  - FDI
  - ISOLATETIME
  - CONNFAIL
  - SSUMLIMIT
  - BCPii
  - MEMSTALLTIME
  - CRITICALMEMBER
  - CFSTRHANGTIME
  - Change to default action
- CFRM
  - MSGBASED
  - SMDUPLEX
  - SMREBUILD
- Coupling Facility Configurations
- Sizing Structures
- MAINTMODE and REALLOCATE
- REALLOCATE TEST
- REALLOCATE REPORT
- Best Practices for upgrading CFs
- Mirroring Couple Datasets
- Criticalpaging Function
- The Backup Plan
- Healthchecks
- Last Resort

re·sil·ient – adjective

- 1. springing back; rebounding.
- 2. returning to the original form or position after being bent, compressed, or stretched.
- 3. recovering readily from illness, depression, adversity, or the like; buoyant.

- **Resilient does not equal error free**. Single component failures will occur. Given this fact, our goal is to prevent a single component failure from becoming a sysplex impacting event.
- A resilient sysplex is one that is configured to achieve desired availability, is configured to scale to meet the needs of an enterprise, adheres to best practice operational procedures and leverages all available technology to recover from issues quickly.

Sympathy Sickness

- Sick systems don't play well with others
  - They don't respond when spoken to
  - They don't share their toys
- Hangs occur because others are:
  - Waiting for a response
  - Waiting to get an ENQ, latch, lock
- What can make a system sick?
  - Being dead
  - Loops (spin, SRB)
  - Low weighted LPAR
  - Loss of a coupling facility
- If a "sick" system does not recovery swiftly, or is not removed from the sysplex swiftly, other systems in the sysplex may be adversely impacted
  - ***In many cases, a long period of sympathy sickness has a greater negative impact on the sysplex than does the termination of an XCF group member, address space, structure connector, or even a system***

A resilient sysplex will take action to terminate a sick component, address space, application, structure connection or system when necessary to maximize the health of the entire sysplex.

Sysplex Failure Management

- A Sysplex Failure Management (SFM) policy that implements best practices is a critical component of a resilient sysplex
- A good SFM policy enables automatic, timely, corrective action to be taken when applications or systems appear to be causing sympathy sickness
- SFM is your backstop that protects your sysplex when your operators and/or your automation are inattentive, unable, or incapable of resolving the problem
  - Every SFM parameter was created in response to actual incidents
  - You have full control over how quickly SFM reacts
  - It is vitally important to have the backstop in place

XCF_SFM_ACTIVE health check

Sysplex Failure Management

- Define an SFM policy to help meet your availability and recovery objectives
  - Applications or systems are not permitted to linger in an extremely sick state such that they adversely impact other systems in the sysplex
  - Applications or systems are not terminated prematurely
  - SFM settings may also vary depending on if there are operators continuously monitoring systems or if operators must be paged
- A suitable SFM policy is but a component of a resilient sysplex. You must still:
  - Ensure no hardware or software single points of failure
  - Have sufficient redundancy to allow for recovery
  - Sysplex enable workloads
  - Workload balancing

XCF_SFM_ACTIVE health check

Failure Detection Interval

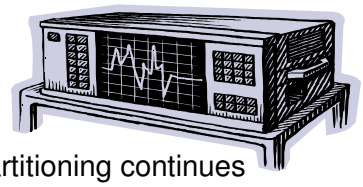
- Amount of time a system is permitted to appear unresponsive
  - Not updating heartbeat
  - Not sending signals
- FDI = MAX(User defined FDI, spin FDI)
  - User defined FDI is specified in COUPLExx or via SETXCF command
  - Spin FDI = (N+1)*spintime+5
    - N is the number of spin actions defined in the active EXSPATxx
    - Spintime is the spin interval defined in the active EXSPATxx
    - In a shared CP environment with default EXSPATxx parmlib settings, spin_FDI = 165 seconds
- FDI value
  - Too short -> unnecessary actions by SFM -false positives
  - Too long -> elongates sympathy sickness window -needless pain
- Just the right FDI
  - Best practice: spin FDI (do not specify user FDI)

IXC470I SYSTEM xxx EFFECTIVE VALUES: INTERVAL=165 OPNOTIFY=168
DEFAULT USER INTERVAL: 165
DERIVED SPIN INTERVAL: 165
DEFAULT USER OPNOTIFY: + 3
COMPUTED FOR: XCF INITIALIZATION

XCF_FDI health check

System Not Updating Status, System Not Sending Signals

- ISOLATETIME(x)
  - X seconds after the FDI exceeded fencing is initiated by all systems
    - •Commands are sent across the coupling facility to the target system
    - •I/O is isolated
      - •No new I/O is initiated
      - •Any ongoing I/O is terminated



- After fencing completes successfully, sysplex partitioning continues
  - Other systems in the sysplex clean up for system that was removed
  - Shared resources are released
- If fencing fails IXC102A is issued
  - Operator must reset the image and respond down to IXC102A

Recommendation: ISOLATETIME(0)

Health

XCF_SFM_SUM_ACTION health check

System Not Sending Signals, System Updating Status

- System delays, performance issues, device/CF issues
  - Stalled I/O restarts, no buffer conditions, response times
  - SFM has nothing for these issues
  - Manual intervention to diagnose and repair
- Loss of signal connectivity
  - CONNFAIL(YES)
    - •SFM determines sets of systems that do have full signal connectivity
    - •Selects a set with largest combined system weights
    - •Systems in that set survive, others are removed
    - •To ensure CONNFAIL makes the best decisions for the sysplex, ensure the weights assigned to each z/OS system adequately reflect the relative importance of the system
  - CONNFAIL(NO)
    - •Operator prompted with IXC409D to determine which system to terminate

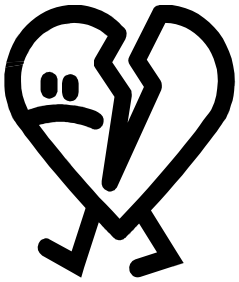
Recommendation: CONNFAIL(YES)

Health

XCF_SFM_CONNFAIL health check

System Sending Signals, System Not Updating Status

- SSUMLIMIT(x)
  - Indicates the length of time a system can remain in the state of not updating the heartbeat and sending signals, aka, the amount of time a system will remain in a "semi-sick" state.
  - Once the SSUMLIMIT has been reached the specified action will be initiated against the system
    - ISOLATETIME(0)



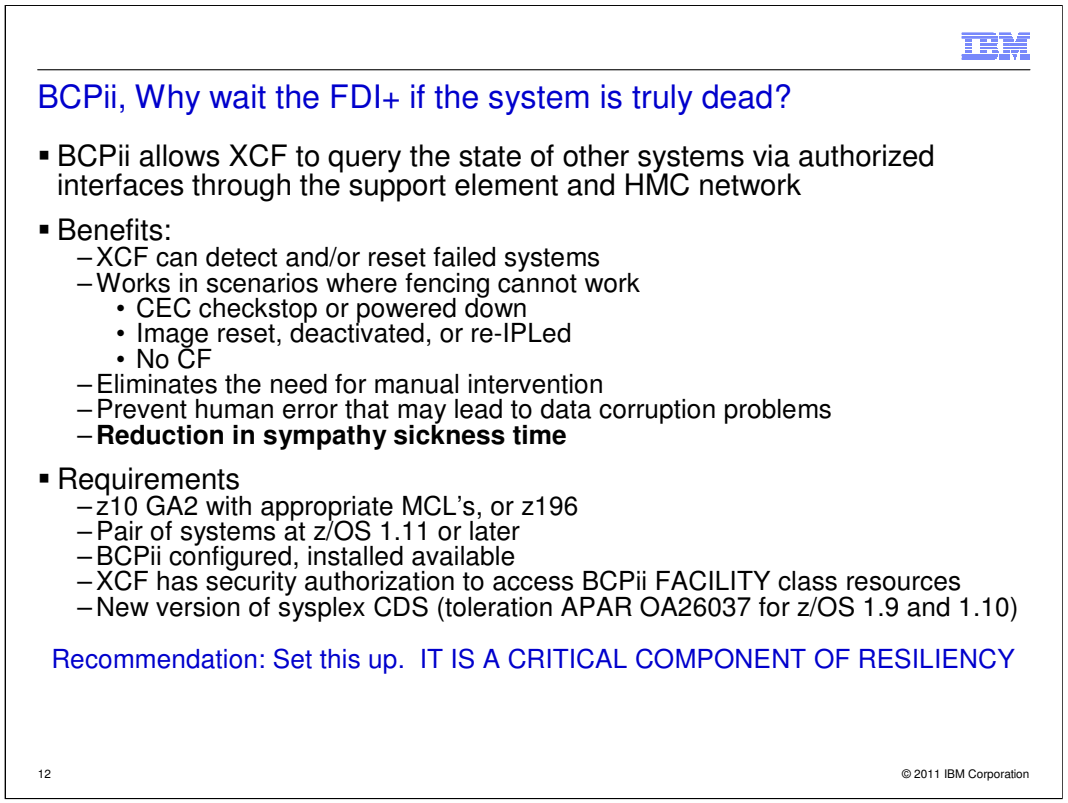
Recommendation: SSUMLIMIT(900)

Health

IXC446I SYSTEM *sysname* IS IN MONITOR-DETECTED STOP STATUS BUT IS SENDING XCF SIGNALS. SFM WILL TAKE SSUM ACTION AT *actiontime* IF SYSTEM REMAINS IN THIS STATE.

IXC101I SYSPLEX PARTITIONING IN PROGRESS FOR ST37 REQUESTED BY XCFAS. REASON: SFM STARTED DUE TO STATUS UPDATE MISSING

XCF_SFM_SSUMLIMIT health check



SFM will automatically exploit BCPii and as soon as the required configuration is established. (a) Pairs of systems running z/OS 1.11 or higher (b) BCPii configured, installed, and available (c) XCF has security authorization to access BCPii defined FACILITY class resources (d) z10 GA2 with appropriate MCL's, or z196 (e) New version of the sypslex CDS is primary in the sysplex (f) toleration AAPR OA26037 for z/OS 1.9 & 1.10 (g) SYSSTATE DETECT function is not enabled.

•See topic "Assigning the RACF TRUSTED attribute" in *MVS Initialization and Tuning Reference* for information on using RACF to assign the TRUSTED attribute to the XCF address space.

•Refer to the "BCPii Setup and Installation" topic in *MVS Programming: Callable Services for High Level Languages* for information on installation and configuration steps and SAF authorization requirements to enable BCPii to invoke z/Series Hardware APIs.

•A system running on z/OS V1R11 and down-level hardware is only eligible to target other systems that are enabled to exploit the full functionality of the System Status Detection (SSD) partitioning protocol. A system not running on the requisite hardware can not be the target of SSD partitioning protocol functions.

•Install toleration PTFs for OA26037 on V1R10 and V1R9 systems in the sysplex to use the newly formatted sysplex couple data set required by the protocol.

•By default, the SYSSTATDETECT function is enabled in V1R11. The current setting of the SYSSTATDETECT function can be determined by issuing a DISPLAY XCF,COUPLE command. SYSSTATDETECT is the name of the XCF FUNCTIONS

MEMSTALLTIME

- Enable XCF to automatically take action when XCF signals are backing up to the point of adversely impacting other systems in the sysplex
- Action XCF will take: terminate the stalled member with the highest quantity of signals backed up

Recommendation: MEMSTALLTIME(600-900)

13

MEMSTALLTIME enables system to break out of an XCF signaling traffic jam. SFM will automatically start removing the largest build up. In the picture above, imagine all the blue cars were instantly removed.

IXC633I “member is impaired”
GROUP *gnme* MEMBER *mnme* JOB *jnme* ASID *asid*
{DEEMED | CONFIRMED} IMPAIRED AT *ipdate iptime* ID: *s#.r#*
LAST MSGX: *sgdate sgtime sgexit* STALLED *sgwork* PENDINGQ
LAST GRPX: *grdate grtime grexit* STALLED *grwork* PENDINGQ
LAST STAX: *stdate sttime stexit* STALLED

IXC634I GROUP *grpname* MEMBER *membername* JOB *jobname* ASID *asid* NO LONGER
IMPAIRED. *text* AT *ResumeDate ResumeTme* ID: *stall#*

IXC635E SYSTEM *sysname* HAS IMPAIRED XCF GROUP MEMBERS

IXC636I GROUP *grpname* MEMBER *membername* JOB *jobname* ASID *asid* IMPAIRED,
IMPACTING [CRITICAL] FUNCTION *function*

CFSTRHANGTIME

- Enable XES to automatically take action if a connector does not respond to a structure event in a timely fashion
- XES corrective actions:
  - Stop rebuild
  - Force user to disconnect
  - Terminate connector task, address space or system
  - RAS: ABEND026 dumps collected

Recommendation: CFSTRHANGTIME(900-1200)

14

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*

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

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

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.

CRITICAL MEMBER

- z/OS 1.12 GRS has declared itself to be a critical member
- In cooperation with XCF, GRS monitors its ability to perform its work (such as ENQ processing)
- If GRS cannot perform work for as long as the FDI, GRS is said to be “impaired”
- If GRS is impaired for more than N seconds, SFM will remove the system from the sysplex
  - N is determined by the SFM MEMSTALLTIME parameter
  - For MEMSTALLTIME(n), N=n seconds
  - For MEMSTALLTIME(NO), N=MAX(FDI, 120 seconds)

z/OS 1.11 Change to Default Partitioning Processing

- Prior to z/OS 1.11 the default action was PROMPT
- With z/OS 1.11 the default action is ISOLATETIME(0)
- D XCF,C indicates what is desired by the system
  - Both systems participating are at z/OS 1.11 then isolation will transpire
  - If either system participating is lower than z/OS 1.11 then PROMPT

16

As mentioned previously in this presentation, there are times when removing a system from the sysplex is the best way to ensure resiliency. As of z/OS 1.11 the sysplex default action ensures resiliency for systems which do not have SFM declarations.

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

IXC109I FENCE OF SYSTEM SA0 SUCCESSFUL.

IXC101I and IXC105I are issued as they have been in the past.

CFRM - MSGBASED

- Minimize serialized writes to the CFRM CDS by enabling one system to be the manager to coordinate structure recovery / rebuild protocols

- Enable MSGBASED
  - Format CFRM CDS
    - ITEM NAME(MSGBASED) NUMBER(1)
  - SETXCF START,MSGBASED - switch occurs when there are no events outstanding for a structure
    - Events - connect, disconnect, rebuild events .. the reasons XES reaches out to connectors to a structure.

Recommendation: Leverage MSGBASED processing.

Health

There is a switch to revert back to non-MSGBASED processing.

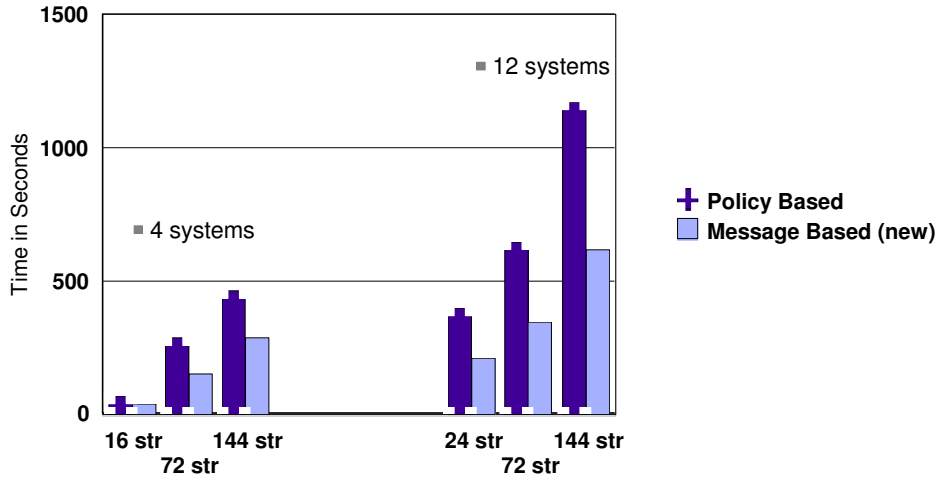
SETXCF STOP,MSGBASED - signal sent to all other systems in the sysplex. When each system processes the notification it will discard MSGBASED requests. Systems will then read the policy to obtain the next event to process. Pending signals will also be discarded, the system will not wait for a pending signal to complete.

This function has been in the field since z/OS 1.8.

XCF_CFRM_MSGBASED health check

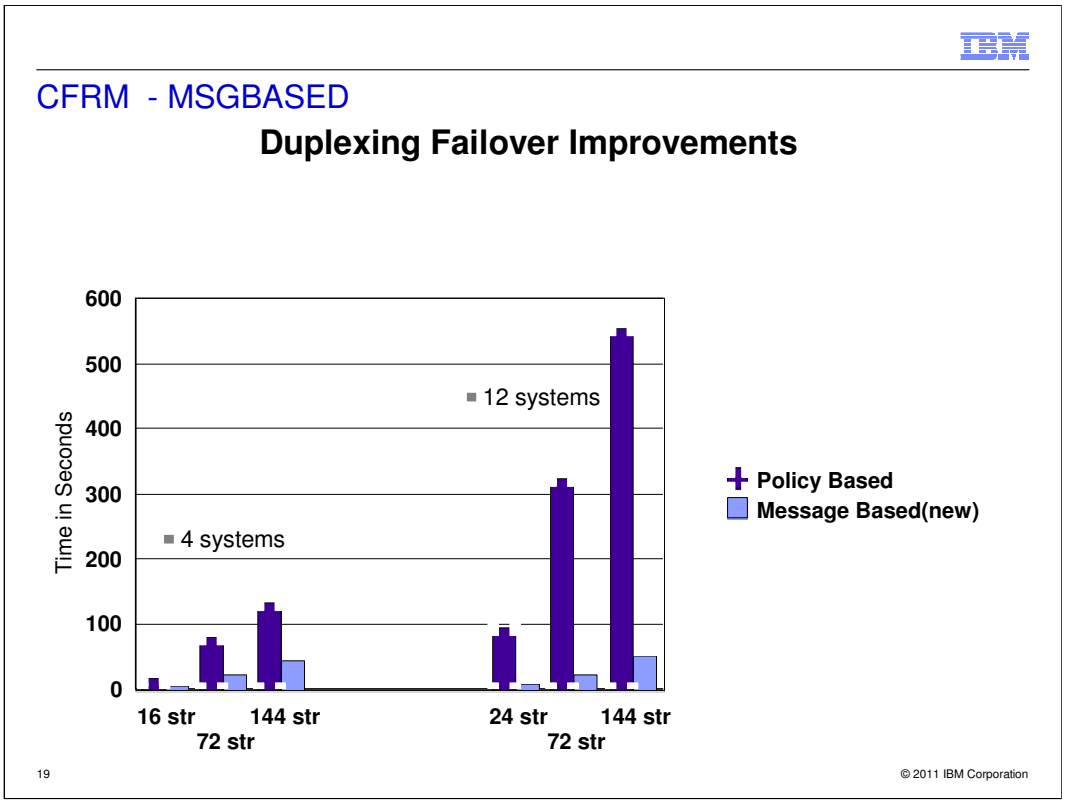
CFRM - MSGBASED

Structure Rebuild Improvements:



18

With 12 systems, 144 active structures, MSGBASED usage reduced rebuild time by almost 50%. 1175 seconds down to 620 seconds.



With 12 systems, 144 active structures, MSGBASED usage reduced failover time in a duplex environment by up to 90%! 555 seconds down to 52 seconds.

CFRM - SMREBUILD

- System managed rebuild process
- Structures which do not support user managed rebuild can be relocated using system managed rebuild
- Enable SMREBUILD
  - Format CFRM CDS
    - ITEM NAME(SMREBUILD) NUMBER(1)
  - All connectors must specify
    - IXLCONN ALLOWAUTO=YES
- Must have access to the “old” structure throughout the rebuild, so SM rebuild does not provide a CF failure or CF lossconn recovery strategy

Recommendation: Leverage to simplify CF disruptive changes, balancing of workloads.

If SM rebuild is used for a particular structure and there is a CF failure or loss of connectivity to the CF (or structure) an alternate recovery mechanism will be required.

CFRM - SMDUPLEX

- System managed duplexing allows applications to transparently recover from failures automatically
  - Structure failure, CF failure, loss of connectivity to CF
  - Critical for applications which do not support user rebuilds
- But not without cost
  - Service times for duplexed requests are longer than simplex
  - Need links between CFs (and a pair of CFs)
- Setup Required
  - Format CFRM CDS
    - ITEM NAME(SMDUPLEX) NUMBER(1)
  - CFRM policy updates for relevant structures
    - DUPLEX(ALLOWED) – manual control of when to start duplexing
    - DUPLEX(ENABLED) – system seeks to maintain duplexing when feasible

Recommendation: Consider leveraging SMDUPLEX processing.

Example – CICS named counter structure

System Managed Duplexing White Paper
www.ibm.com/systems/z/advantages/pso/whitepaper.html

User Managed Duplexing

- For system managed duplexing, the system maintains both instances of the structure
- User managed duplexing, the exploiter determines which requests to duplex
- Refer to application recommendations for best practices
  - Example, DB2 Group Buffer best practice is to utilize user managed duplexing

Coupling Facility Configuration

- Relative to coupling facilities, redundancy to a fair extent, permits resiliency
- Have at least 2 Coupling Facilities defined in the CFRM policy and physically available.
- External CFs are preferred to internal CFs.
  - External CF in use by z/OS systems does not reside on CEC with any z/OS image using the CEC
  - Internal CF resides on a CEC with at least one z/OS image using it
  - Certain structures become unrecoverable if they reside in a non-failure-isolated CF Use dedicated CPs on the CFs whenever possible.
- Have at least two coupling links to / from each operating system to the coupling facility. Additional paths may be required with heavy workloads.
- NonVolatile CFs are preferred.
- Ensure there is enough space for all the structures and enough white space for structures on the other coupling facilities to rebuild into this coupling facility should there be a CF outage.

XCF_CF suite of health checks

XCF_SIG health checks

System Managed Duplexing White Paper

www.ibm.com/systems/z/advantages/pso/whitepaper.html

Sizing CF Structures

- IBM recommends using the CFSizer website or SIZER batch utility whenever the CFCC level is upgraded or there is a significant change in the workload using the structures
- CFSizer
  - http://www-947.ibm.com/systems/support/z/cfsizer/
- SIZER batch utility
  - http://www-947.ibm.com/systems/support/z/cfsizer/altsize.html
- IBM suggests that the INITSIZE to SIZE ratio not exceed 1:2

If structures are incorrectly sized applications may be impacted as a result of thruput delays. Structure connections may fail. If structures are significantly undersized then a new allocation, following deallocation, of the structure may fail.

MAINTMODE and REALLOCATE

- MAINTMODE
  - Structures must be moved off of a coupling facility prior to taking the coupling facility out of service
  - Place a CF in maintenance mode to ensure XCF does not allocate any new instances of structures on the CF
  - Commands
    - SETXCF START,MAINTMODE,CFNAME=cfhame
    - SETXCF STOP,MAINTMODE,CFNAME=cfname
- REALLOCATE
  - Command to initiate XCF / XES evaluation of structure location and movement of structures to the most desired location
    - optimize structure placement in normal use
    - move structures around in CF maintenance scenarios
    - Restore structures to their rightful place after a CF failure

Recommendation: Use MAINTMODE and REALLOCATE for CF maintenance and structure placement.

z/OS 1.12 REALLOCATE TEST

- D XCF,REALLOCATE,TEST
- Proactive look at the results of reallocate processing
  - Are some production structures going to move?
  - Is now the best time to move production structures?
- Analysis of TEST REALLOCATE provided in IXC347I
  - List of structures with an error / exception condition
  - List of structures with a warning condition
  - List of structures which will reallocate successfully and details pertaining to where structure will reside after reallocated
  - List of structures which reside in preferred CF
  - Coupling facility summary of where structure instances will reside after reallocate completes
  - Summary of reallocate processing analogous to IXC545I

Recommendation: Leverage REALLOCATE TEST to proactively plan for results of REALLOCATE processing.

Summary at the bottom of IXC347I

```
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
```

z/OS 1.12 REALLOCATE REPORT

- D XCF,REALLOCATE,REPORT
- Summarize the results of the previous REALLOCATE
- Operators can take corrective actions as needed
- Analysis of REALLOCATE REPORT provided in IXC347I
  - Start and stop time of last reallocate
  - List of structures with an error / exception condition
  - List of structures with a warning condition
  - List of structures which will relocated successfully
  - List of structures which resided in preferred CF prior to initiating last reallocate
  - Summary of reallocate processing analogous to IXC545I

Recommendation: Automate the display of the reallocate report after each IXC545I, review the results and take action as needed..

27

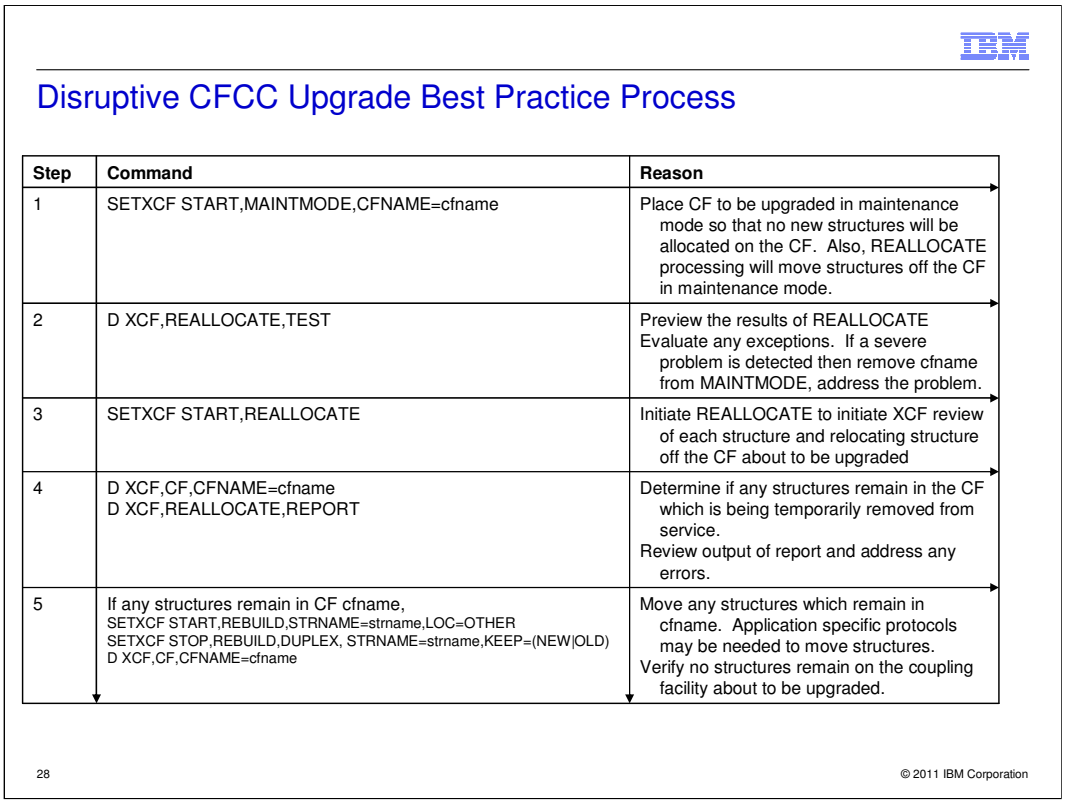
Summary at the bottom of IXC347I

```
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
```

Note: system automates the display if there is an exception



Disruptive CFCC Upgrade

The apply of the CFCC MCL requires the CF to be reactivated

CEC with CF is being taken out of service temporarily for other maintenance

All structures must be removed from the CF for some reason. The same physical coupling facility is being brought back into service once the disruptive changes are complete.

Disruptive CFCC Upgrade Best Practice Process

Step	Command	Reason
6	On each system, VARY PATH(CFname,xx,CFname,yy,etc),OFFLINE,UNCOND	Vary all paths to CF logically offline
7	RO* ALL,D CF,CFNAME=cfname D XCF,CF,CFNAME=cfname	Verify no system has active LOGICAL path to CF. Verify no system has connectivity to CF.
8	Complete the HW upgrade	IXL158I will be issued indicating paths are not operational when the CF is taken down. IXL157I will be issued indicating the paths are available when CF connectivity is restored.
9	On each system, VARY PATH(CFname,xx,CFname,yy,etc),ONLINE	Configure paths online.
10	RO *ALL,D CF,CFNAME=cfname	Verify all systems have connectivity to the CF and the paths are ONLINE. Verity CF to CF links are available.
11	SETXCF STOP,MAINTMODE,CFNAME=cfname	Permit XCF to allocate structures in the CF by taking the CF out of maintenance mode.
12	SETXCF START,REALLOCATE	Relocate structures to the desired CFs.
13	D XCF,REALLOCATE,REPORT	Verify all structure reside in desired CFs. Correct any errors noted by the report.

Push / Pull CF Upgrade Best Practice Process

Step	Command	Reason
0	Create new CFRM policy distinct from the currently active policy with the new CF definition and updated structure definitions based on CFSizer or SIZER.	This step can be done ahead of time to minimize net down time.
1	SETXCF START,MAINTMODE,CFNAME=cfname	Place CF to be upgraded in maintenance mode so that no new structures will be allocated on the CF. Also, REALLOCATE processing will move structures off the CF in maintenance mode.
2	D XCF,REALLOCATE,TEST	Preview the results of REALLOCATE Evaluate any exceptions. If a severe problem is detected then remove cfname from MAINTMODE, address the problem
3	SETXCF START,REALLOCATE	Initiate REALLOCATE to initiate XCF review of each structure and relocating structure off the CF about to be upgraded
4	D XCF,CF,CFNAME=cfname D XCF,REALLOCATE,REPORT	Determine if any structures remain in the CF which is being temporarily removed from service. Review output of report and address any errors.

Push / Pull upgrade implies the physical CEC on which the CF resides is being replaced (upgraded to a new machine).

If the CEC being replaced is the stratum 1 for the STP timing network, please move it to another CEC prior to removing the machine.

Push / Pull CF Upgrade Best Practice Process

Step	Command	Reason
5	If any structures remain in CF cfname, SETXCF START,REBUILD,STRNAME=strname,LOC=OTHER SETXCF STOP,REBUILD,DUPLEX, STRNAME=strname,KEEP=(NEW|OLD) D XCF,CF,CFNAME=cfname	Move any structures which remain in cfname. Application specific protocols may be needed to move structures. Verify no structures remain on the coupling facility about to be upgraded.
6	CONFIG CHP(xx,yy,zz,aa),OFFLINE,UNCOND R xx,CONTINUE	Configure the paths to the CF offline. Reply continue to IXL126I to remove the last path to the CF.
7	RO* ALL,D CF,CFNAME=cfname D XCF,CF,CFNAME=cfname	Verify no system has active path to CF. Re-verify no structures in the CF
8	Remove the “old” machine and bring in the “new” machine.	
9	ACTIVATE parms,SOFT=VALIDATE	If making changes to CF elements (CF control units or CF channel paths) in the I/O configuration, ensure that SOFT=VALIDATE is specified on the ACTIVATE system command. SOFT=VALIDATE is a requirement in all N-1 partitions when changes to CF elements are made.

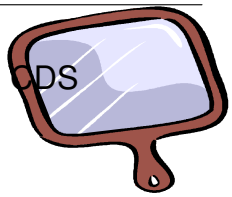
Push / Pull CF Upgrade Best Practice Process

Step	Command	Reason
10	SETXCF START,POLICY,TYPE=CFRM,POLNAME=new_policy name	Activate the new CFRM policy with the definition for the new CF. Update structure sizes per CFSizer or SIZER output.
11	CONFIG CHP(xx,yy,zz,aa),ONLINE	Configure paths online.
12	D XCF,CF,CFNAME=cfname RO *ALL,D CF,CFNAME=cfname	Verify XCF has the proper definition, including serial number for the new CF. D XCF,CF,CFNAME=cfname indicates the CF definition XCF logically knows about. D CF,CFNAME=cfname contains the physical information for the CF the image is connected to. The serial number must match. Verify all systems have connectivity to the new CF and all the paths are ONLINE. Verity CF to CF links are available.
13	SETXCF START,REALLOCATE	Relocate structures to the desired CFs.
14	D XCF,REALLOCATE,REPORT	Verify all structuresreside in desired CFs. Correct any errors noted by the report.

Couple Datasets

- IBM General CDS Recommendations
  - Always run with a primary and alternate
  - Place CDSs on different volumes
- Automatic Restart Manager (ARM)
  - Use ARM policy (or automation) to quickly restart failed elements "in place" when they fail on a running system
  - Use ARM policy or automation to quickly restart failed elements "cross systems" when a system fails.
    - Extremely important for subsystems
    - Process the logs and release the retained locks (DB2)

Mirroring Couple Datasets



- Goal: Customers desire to simplify DR configuration and minimize time to recovery.
- Problem: Sysplex outages or severe performance problems may result from mirroring CDSs.
  - Here mirroring means physical hardware level replication of DASD
  - Does not refer to software mirroring between primary and alternate CDS
- Recommendations:
  - The only CDS that should ever be considered for synchronous mirroring is LOGR CDS
    - Mirroring SYSPLEX CDS may result in sysplex slowdowns
    - Mirroring CDSs may result in IO errors against the CDS
      - IO error against the primary CDS and alternate CDS results in a loss of functions associated with the CDS
  - If an asynchronously mirrored CFRM is to be used at the DR site, the DR site CFs must be in the active policy.
  - Never IPL another system using a copied or mirrored SYSPLEX or CFRM CDS if the IPLing system has access to the primary dasd or CFs which are currently in-use
    - Steal CDSs from active sysplex
    - Steal CFs from active sysplex
  - Anytime a configuration changes ensure the mirroring or copying of CDSs is considered
- Reference: Hot Topics February 2011 Issue 24 p69 Mirror, mirror, on the wall, should couple dataset be mirrored at all?

Hardware level mirroring includes PPRC, XCR, metro mirror, global mirror, etc.

Software mirroring, always have a primary and alternate CDS is a MUST!!

CRITICALPAGING

- Problem Statement: Loss of system(s) during hyperswap (or other dasd swap) which were expected to survive due to critical code path encountering a page fault while DASD freeze / swap is in progress
- Solution: CRITICALPAGING Function, minimizes the likelihood of systems failing to survive a hyperswap (or other dasd swap) due to encountering a page-fault on a critical code path by “hardening” storage of critical address spaces.
  - Critical system address spaces
    - RASP (RSM), GRS, CONSOLE, XCFAS, address spaces associated with Basic HyperSwap in base (HSIB), Basic HyperSwap API (HSIBAPI), and GDPS HyperSwap Communication Task (often jobname GEOXCFST)

CRITICALPAGING

- Real storage assessment needed prior to enabling CRITICALPAGING to ensure application performance is not impeded
- If the system "never" pages perhaps no real storage needs to be added
- If the system pages often, to maintain current performance, a simple guideline
  - PLPA+EPLPA and CSA+ECSA
- References:
  - WSC Flash
    - http://www-03.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/FLASH10733
  - White Paper
    - http://www-03.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP101800

Recommendation: Perform real storage assessment and enable CRITICALPAGING function in DASD swap environments

The Back Up Plan – Notify System Programmer

- If the resiliency options are not implemented then operators must be engaged to assess the situation and determine which actions to take
- Call for help as quickly as possible

The Back Up Plan – Notify System Programmer

Message	System Programmer Action
IXC102A	Reset system and respond DOWN immediately.
IXC402D	Reset system and respond DOWN immediately.
IXC409D	Assess status of systems Respond with the name of the system to be removed
IXC426D	System is sending signals but not updating its heartbeat. Investigate swiftly and react before sysplex sympathy sickness ensues. Respond with the system to take down if unable to resolve immediately.
IXC631I IXC633I IXC635E IXC636I IXC640E	Investigate stalled members, pursue recovery options which include termination of stalled members.

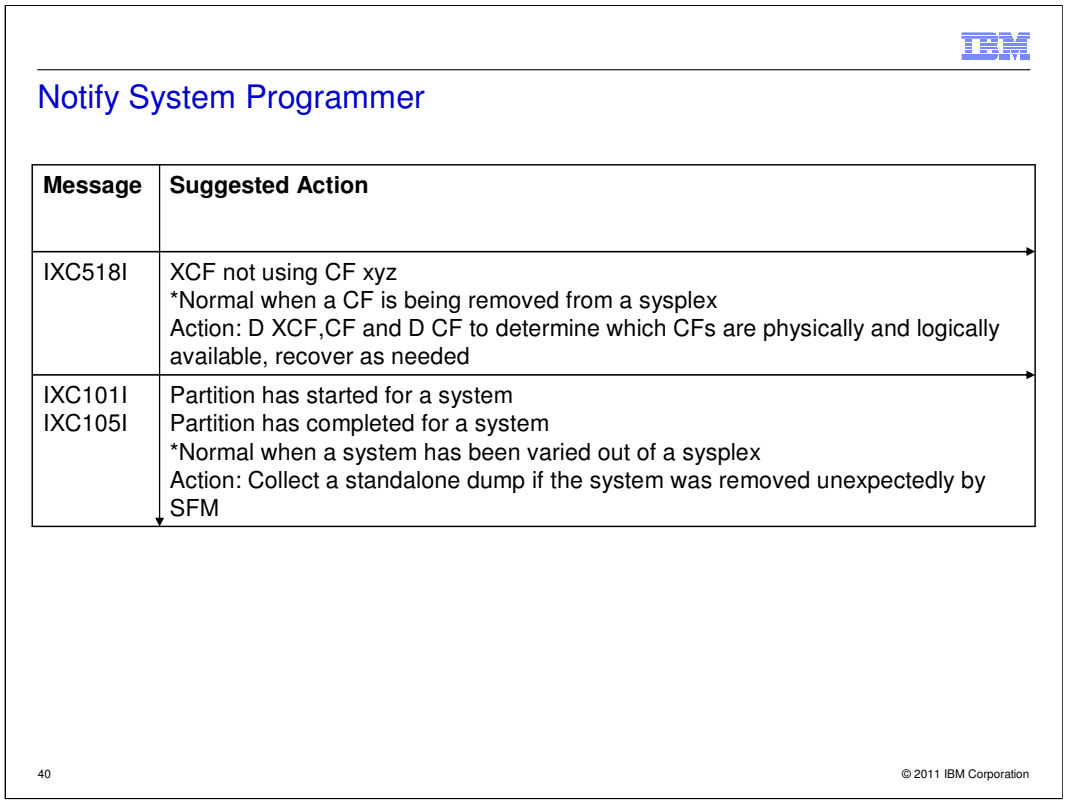
Recommendation: Leverage resiliency options, ISOLATETIME, SSUMLIMIT, CONNFAIL and MEMSTALLTIME.

The Back Up Plan - Notify System Programmer

Message	Suggested Action
IXL040E IXL041E	Determine why connector has not responded. Consider terminating the connector. If the hang exceed 2 minutes ABEND026 RSN08118001 dump will be taken. Open a PMR to the application failing to respond.

Recommendation: Leverage CFSTRHANGTIME

.



The above messages occur when an unexpected, infrequently encountered, error situation occurs. System programmers should be notified.

IBM is pursing leveraging the automatic generation CF non-disruptive serialized dumps in situations to enhance serviceability.

Notify System Programmer

Message	Suggested Action
IXL008I	Path to CF invalidated Action: D CF to determine if corrective action for the CF paths needs to be taken.
IXL044I	IFCCs for a coupling facility were detected. Action(s): Consider collecting a nondisruptive dump of the CF while the problem is occurring. Also consider collecting dumps on all systems in the sysplex. Contact the IBM Hardware Support Center. SLIP SET,ACTION=SVCD,MSGID=IXL044I, JOBLIST=(XCFAS),DSPNAME=('XCFAS'.*), SDATA=(ALLNUC,CSA,PSA,LPA,LSQA,NUC,RGN,SQA,SUM,SWA,TRT,XESDATA,COUPLE), REMOTE=(DSPNAME,SDATA,JOBLIST),END

The above messages occur when an unexpected, infrequently encountered, error situation occurs. System programmers should be notified.

IBM is pursing leveraging the automatic generation CF non-disruptive serialized dumps in situations to enhance serviceability.

Notify System Programmer

Message	Suggested Action
IXL045E	XES SRBs encountering delays. Action(s): Determine if the system is overburdened and resolve the bottleneck. Consider taking a dump while the condition is occurring and contact the IBM Software Support Center (compid 5752SCIXL). 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	Path to CF not operational Action: Verify the desired configuration for that path, take action as needed. Consider collecting a nondisruptive dump of the CF while the problem is occurring. Also consider collecting dumps on all systems in the sysplex. Contact the IBM Hardware Support Center. SLIP SET,ACTION=SVCD,MSGID=IXL158I, JOBLIST=(XCFAS),DSPNAME=('XCFAS'.*), SDATA=(ALLNUC,CSA,PSA,LPA,LSQA,NUC,RGN,SQA,SUM,SWA,TRT,XESDATA,COUPLE), REMOTE=(DSPNAME,SDATA,JOBLIST),END

The above messages occur when an unexpected, infrequently encountered situation, occurs. System programmers should be notified.

IBM is pursing leveraging the automatic generation CF non-disruptive serialized dumps in situations to enhance serviceability.

Last Resort

- System Console
  - Ensure access to the system console
  - z/OS 1.11 and above V CN(*),ACTIVATE not required to enter commands
- Strongly consider enabling accessing of the HMC over the web

Consider running consoles in distributed mode. Allows 99 active consoles per system, 250 consoles can be defined per system. Reduces the number of times SYSZMCS#MCS resources is required for console definition changes.

Healthchecks

- 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_STR_POLICY_SIZE
- 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

Recommendation: Investigate exceptions and take action as appropriate.

Questions?