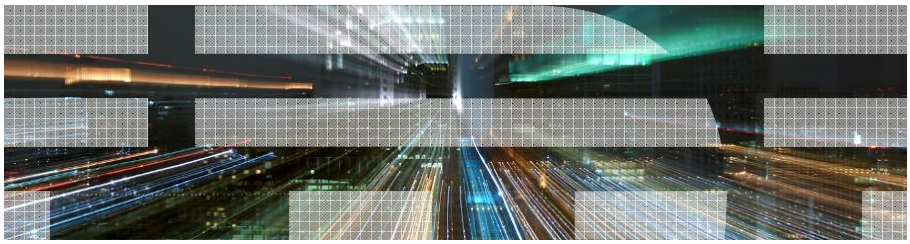
Mark A. Brooks mabrook@us.ibm.com
Nicole M. Fagen nfagen@us.ibm.com
October 2012

Parallel Sysplex® Resiliency



Trademarks

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

IBM®	MQSeries®	S/390®	z9®	IBM® (logo)
ibm.com®	MVS™	Service Request Manager®	z10™	AIX® BladeCenter®
CICS®	OS/390®	Sysplex Timer®	z/Architecture®	DataPower®
CICSPlex®	Parallel Sysplex®	System z®	zEnterprise™	DS4000®
DB2®	Processor Resource/Systems Manager™	System z9®	z/OS®	DS6000™
eServer™	PR/SM™	System z10®	z/VM®	DS8000®
ESCON®	RACF®	System/390®	z/VSE®	POWER7®
FICON®	Redbooks®	Tivoli®	zSeries®	ProtecTIER®
GDPS®	Resource Measurement Facility™	VTAM®		Rational®
HyperSwap®	RETAIN®	WebSphere®		System Storage®
IMS™				System x®
IMS/ESA®	Geographically Dispersed Parallel Sysplex™			XIV®

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.

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 & REPORT
- Best Practices for upgrading CFs
- Where is My Coupling Facility?
- Mirroring Couple Datasets
- Criticalpaging Function
- RRS Archive Logstream
- System Logger
- Automatic Restart Manager
- Root Causes of Sysplex Performance Problems
- Avoiding Planned Outages
- The Backup Plan
- Healthchecks
- Last Resort
  - SYNCHDEST

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***

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

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

Health

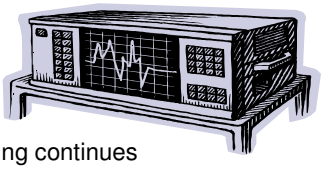
Failure Detection Interval

- 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)

Health

System Not Updating Status, System Not Sending Signals

- ISOLATETIME(x)
  - X seconds after the FDI exceeded fencing is initiated by all systems
    - •Fencing commands sent via the coupling facility to 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

System Not Sending Signals, System Updating Status

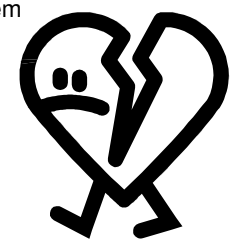
- System delays, performance issues, device/CF issues
  - Stalled I/O restarts, no buffer conditions, response times
  - SFM has nothing for these issues
  - Manual intervention to diagnose and repair
- Loss of signal connectivity
  - CONNFAIL(YES)
    - •SFM determines sets of systems that do have full 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)
Exception: CONNFAIL(NO) for GDPS environment

Health

System Sending Signals, System Not Updating Status

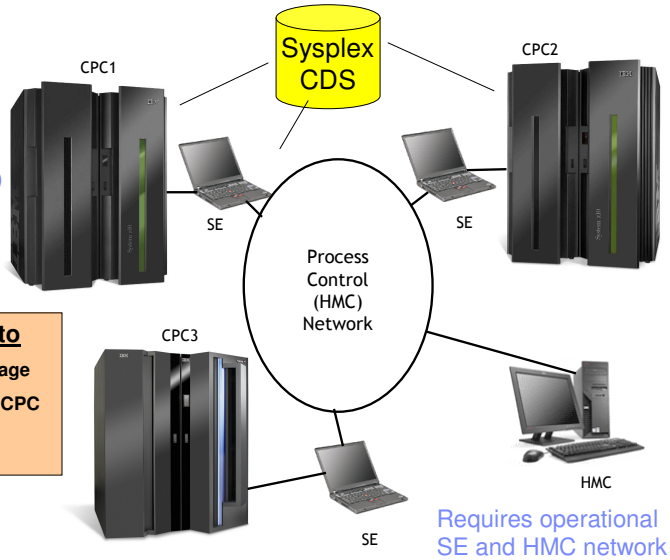
- SSUMLIMIT(x)
  - Indicates the length of time a system can remain in the state of not updating the heartbeat and sending signals, aka, the amount of time a system will remain in a "semi-sick" state.
  - Once the SSUMLIMIT has been reached the specified action will be initiated against the system
 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



Recommendation: SSUMLIMIT(900)

Health

SFM With BCPii



XCF uses BCPii to

- **Obtain identity of an image**
- **Query status of remote CPC and image**
- **Reset an image**

Requires operational SE and HMC network

BCPii, Why wait the FDI+ if the system is truly dead?

- BCPii allows XCF to query the state of other systems via authorized interfaces through the support element and HMC network
- Benefits:
  - XCF can detect and/or reset failed systems
  - Works in scenarios where fencing cannot work
    - CEC checkstop or powered down (via SE, not EPO)
    - Image reset, deactivated, or re-IPLed
    - No CF
  - Eliminates the need for manual intervention
  - Prevent human error that may lead to data corruption problems
  - **Reduction in sympathy sickness time**

- Requirements
  - z10 GA2, z196, or z114 with appropriate MCL's
  - Pair of systems at z/OS 1.11 or later
  - BCPii configured, installed available
  - XCF has security authorization to access BCPii FACILITY class resources
  - New version of sysplex CDS (toleration APAR OA26037 for z/OS 1.9 and 1.10)

Recommendation: Set this up. IT IS A CRITICAL COMPONENT OF RESILIENCY

MEMSTALLTIME

- Enable XCF to automatically take action when XCF signals are backing up to the point of adversely impacting other systems in the sysplex
- XCF action: terminate the stalled member that is consuming the highest quantity of buffer space related to the problem

Recommendation: MEMSTALLTIME(600-900)

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)

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

CFRM - MSGBASED

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

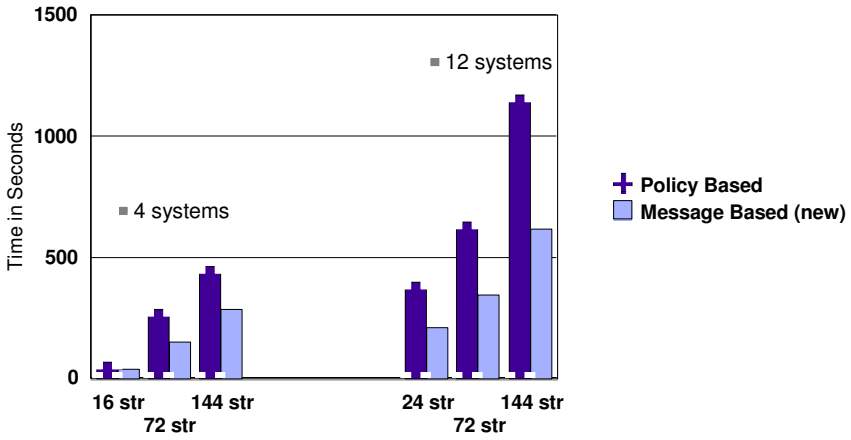
- Enable MSGBASED
  - Format CFRM CDS
    - ITEM NAME(MSGBASED) NUMBER(1)
  - SETXCF START,MSGBASED - switch occurs when there are no events outstanding for a structure
    - Events - connect, disconnect, rebuild events .. the reasons XES reaches out to connectors to a structure.

- The more systems in the sysplex connected to a structure the greater the improvement.

Recommendation: Leverage MSGBASED processing.

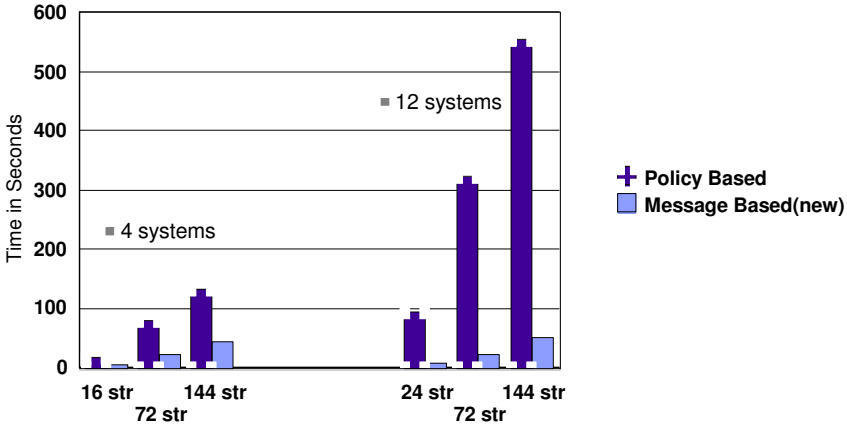
CFRM - MSGBASED

Structure Rebuild Improvements:



CFRM - MSGBASED

Duplexing Failover Improvements



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.

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.

User Managed Duplexing

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

Coupling Facility Configuration

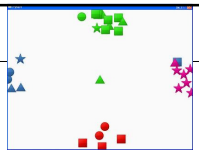
- Relative to coupling facilities, redundancy to a fair extent, permits resiliency
  - Have at least 2 Coupling Facilities defined in the CFRM policy and physically available.
  - Have at least two coupling links to / from each operating system to the coupling facility. Additional paths may be required with heavy workloads.
- 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.
- NonVolatile CFs are preferred.
- Provide 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.

Health

Sizing CF Structures

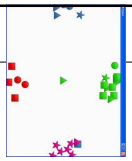
- IBM recommends using the CFSizer website or SIZER batch utility whenever the CFCC level is upgraded or there is a significant change in the workload using the structures
- CFSizer
  - http://www.ibm.com/systems/support/z/cfsizer/
- SIZER batch utility
  - http://www.ibm.com/systems/support/z/cfsizer/altsize.html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

Alter Processing



- Coupling Facility structures can be altered to meet the needs of exploiters.
  - Applications can issue IXLALTERs to change the entry to element ratio and the size of the structure.
  - Operators can use the SETXCF command to alter the size of the structure
  - ALLOWAUTOALT(YES) allows z/OS to initiate alters of the structure to align with in-use counts when the FULLTHRESHOLD is surpassed.
- When a structure is being altered and it is at its maximum size and near full, alter processing may cause undue burden on the coupling facility.

Alter Processing – OA34579



- OA34579 introduced a new SETXCF modify command to all for all types of alters to be disabled (or enabled) for a particular structure
- SETXCF MODIFY,STRNAME=structure_name,ALTER=DISABLED
- SETXCF MODIFY,STRNAME=structure_name,ALTER=ENABLED
- Verify the entry to element ratio is a “best” fit before disabling
  - Best fit for prime shift desired entry to element ratio
  - Best fit for the day
- Disabling alters is only recommended in environments where alter processing is know to have a negative performance impact.
- To display all structures with alter disabled or enabled:
  - DISPLAY XCF,STRUCTURE,ALTER={ENABLED|DISABLED}
  - Also, the display of a given structure will indicate if alter processing is disabled. D XCF,STR will contain START ALTER NOT PERMITTED

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.

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

XCF automatically issues REPORT display if any structure processed on the just completed REALLOCATE ends with return code greater than a warning.

Best Practices: Upgrading a Coupling Facility

UPGRADE

- Failing to follow proper procedures for upgrading a coupling facility may result in unexpected extended down time
  - Be sure you have the right CF !
- Recently published white paper documents best practices for upgrading a coupling facility
  - Push / Pull
  - Disruptive micro code upgrade
- Provides step by step process with short explanation for each step
- http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP101905
- White paper will be updated in the future as new options become available

Where is My Coupling Facility?

- Sysplex Support Team #1 Callout
  - Cannot allocate ISGLOCK structure WAIT0A3-BC
- Environment: DR Site, most likely
- http://www-03.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/FLASH10786
- GRS New Function APAR OA40451

General Recommendations for all Couple Datasets

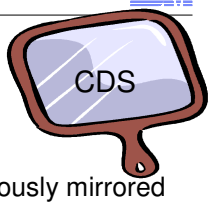
- Always run with a primary and alternate CDS
- Place CDSs on different volumes
- Place CDSs on different physical devices whenever possible
- Be prepared to deal with loss of a CDS by having:
  - A third one pre-formatted for use
  - Operational procedures or system automation to add it as an alternate, thus restoring redundancy

Health

Synchronous Mirroring of Couple Datasets

- Avoid synchronous mirroring of CDS
  - Risk of I/O delay or long busy conditions, which can:
    - Degrade timely access to the CDS by users
    - Lead to permanent I/O error and CDS being removed from service
- Especially Sysplex CDS and CFRM CDS
  - Removing both primary and alternate from service results in a **sysplex-wide outage**
- Possible Exception: LOGR CDS
  - If log-stream data is being mirrored to DR site, and data in the log stream is to be used at DR site, LOGR CDS must be mirrored as well
  - Need time consistent copies of all relevant data if the log stream is to be usable
    - Off-load data sets, staging data sets, LOGR CDS, MVS Catalogs

Using Copies of Couple Datasets at DR Site is Risky



- Customer Goal: Simplify DR configuration and minimize time to recovery by copying CDS to DR site
  - Allows all volumes on device to be copied or asynchronously mirrored without need for manual exclusion of volumes containing CDS
  - Need not run format utility to create CDS at DR site
  - Need not run policy utility to create policies in the CDS at DR site
- Problem: Requires great care to avoid sysplex outages
- Risks:
  - Sysplex outages
  - Data Integrity issues
  - CDS at primary site unexpectedly removed from service
  - 0A3 wait-state when GRS cannot allocate ISGLOCK structure
  - Residual data for inaccessible structures at other site
  - One, some, or all CF's ripped away from active sysplex

Recommendations if Using Copies of Couple Datasets at DR Site

- All CFs used by the DR site should be defined in the CFRM policy used by the primary site
- Do not allow DR site to gain access to CF's in primary site
- Do not allow DR site to gain access to DASD in primary site
- When configuration changes, be sure you maintain these conditions
  - Needs to be part of your change management procedures
- Reference: Hot Topics February 2011 Issue 24 p.69 "*Mirror, mirror, on the wall, should couple dataset be mirrored at all?"*
  - *http://publibfp.dhe.ibm.com/epubs/pdf/eoz2n1c0.pdf*

CRITICALPAGING

- Problem: Loss of system(s) during hyperswap (or other dasd swap) which were expected to survive
- Cause: Page fault in critical code path while DASD freeze/swap is in progress
- Solution: CRITICALPAGING Function
  - "Hardens" storage of critical address spaces
  - Reduces potential for page faults in address spaces that participate in the critical path:
    - 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
    www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/FLASH10733
  - White Paper
    www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP101800

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

RRS Archive Logstream

- Many performance problems reported associated with the ARCHIVE logstream
  - Mainly IXGLOGR data allocation issues
- Prior to z/OS 1.10, RRS had to be stopped and restarted to have RRS stop using the ARCHIVE logstream
- z/OS 1.10 and higher, installations can enable / disable use of the ARCHIVE logstream
- SETRRS ARCHIVELOGGING,ENABLE|DISABLE
- Impact of disabling archive logging
  - Archive logstream is not used by RRS for recovery purposes
  - Archive logstream data is not used by IBM support
  - Archive logstream data is only used by a given enterprise
    - Is the enterprise using it??

System Logger – Dynamic Logstream Definition Updates

- With a LOGR CDS format level of HBB7705 or higher many logstream attributes can be changed dynamically
- Removes the need for application outage to make a change
- IXCMIAPU update
- LOGR CDS formatted HBB7705 with SMDUPLEX
- Attributes updated when CF structure rebuilds
  - LS_SIZE, LS_DATACLAS, LS_MGMCLAS, LS-STORCLAS, OFFLOADRECALL, LOWOFFLOAD, HIGHOFFLOAD, STG-SIZE, STG-DATACLAS, STG-MGMTCLAS, STG-STORCLAS, STG-DUPLEX, DUPLEXMODE, LOGGERDUPLEX
- Attributes updated on dataset switch
  - RETPD, AUTODELETE, LS-SIZE, LS-DATACLAS, LS-MGMTCLAS, LS-STORCLAS, OFFLOADRECALL, LOWOFFLOAD, HIGHOFFLOAD
- Attributes updated on last disconnect / first connect
  - RETPD, AUTODELETE, LS-SIZE, LS-DATACLAS, LS-MGMTCLAS, LS-STORCLAS, OFFLOADRECALL, LOWOFFLOAD, HIGHOFFLOAD, STG-SIZE, STG-DATACLAS, STG-MGMTCLAS, STG-STORCLAS, STG-DUPLEX, DUPLEXMODE, LOGGERDUPLEX, MAXBUFSIZE

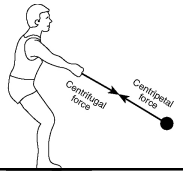
System Logger

- Every great once in a while, a logstream connection is left in a “confused” state. The connection is not fully connected and the connection is not completely disconnected. The connection is not fully functional.
  - Situation 1
    - D LOGGER,C,LSN=log-stream-name shows number of connectors on this system equal to 0.
    - D LOGGER,L,LSN=log-stream-name shows number of connectors to the log stream greater than 0.
  - Situation 2
    - D LOGGER,C,LSN=log-stream-name shows log stream status in "disconnect pending" state.
  - Situation 3
    - Offload dataset recall is taking a very long time, IXG310I, IXG311I and IXG312E, IXG281I may appear.
    - D LOGGER,ST,REC shows dataset recalls have been backed up for a long time

System Logger

- If any of the 3 situations persist for an extend period of time, SETLOGR FORCE may be needed / used to resolve the issue.

- SETLOGR FORCE,DISCONNECT,LSN=logstream_name
- SETLOGR FORCE,DELETE,LSN=logstream_name
- SETLOGR FORCE,NORECALL
  - Issue to stop logger from waiting for the recall of the current dataset being recalled



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)

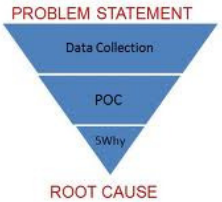
Exception: Only use "in place" for GDPS environment

Root Causes of Sysplex Performance Problems

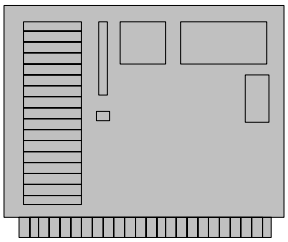
- **Real storage**
  - Change in workload
  - Defect
  - Poor configuration
- **CF slow downs**
  - Service time degradation
  - Not meeting expectations
  - More requests going ASYNC
  - New configuration
  - Increase workload
- **CPU**
  - Not enough CPU to handle workload
  - Capping LPARs
  - All LPARs running hot at the same time and there are vertical lows
- **Virtual storage shortages**
- **SRBs Looping**
- **Contention for global resources**
- **Spin loops**
- **IO delays**
  - CDS access delays
  - Logger offload hangs

You're not getting wet because it's raining…

You're getting wet because your umbrella has a hole in it.



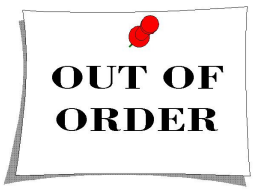
Common Reasons to IPL



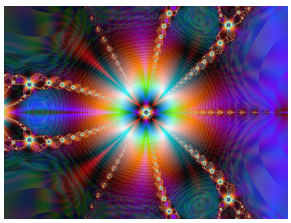
- Upgrade to a new release of z/OS
- Mass application of service
- Move to a new processor
- To harden dynamic changes
- To address storage creep / fragmentation
- Recover from non-reusable ASIDs or non-reusable LXs
- Operator training
- Maintain ability to do an IPL if you really need to do it
- Because it's the way it's always been done

Planned Outage Avoidance

- Dynamically change things on the system that used to require IPL
- Ability to change subsystems without having to restart them
- Capabilities to plan ahead to accommodate future non-disruptive growth
- Error Recovery Improvements
- Increase possibility of surviving longer between IPLs



Dynamic Changes



- APF List
- LNKLST
- LPA – add, delete, and sometimes even update
- Exits
- Subsystems (SETSSI ADD,S=ssn)
- System Symbols (IEASYMUP in z/OS 1.6)
- Number of Page Datasets (PAGEADD PAGEDEL)
- PPT (SCHEDxx – SET SCH=xx)
- RACF Started Class, CDT, database templates
- SVCs
- JES2, JES3
- **See Planned Outage Avoidance Redbook for details!**
  - **http://www.redbooks.ibm.com/redbooks/pdfs/sg247328.pdf**

Resiliency is Closely Related to Availability

- Configure the sysplex such that all of the high availability options are leveraged
- **Mission: Available** white paper
- http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP101966

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

Notify System Programmer

Message	Suggested Action
IXC518I	XCF not using CF xyz *Normal when a CF is being removed from a sysplex Action: D XCF,CF and D CF to determine which CFs are physically and logically available, recover as needed
IXC101I IXC105I	Partition has started for a system Partition has completed for a system *Normal when a system has been varied out of a sysplex Action: Collect a standalone dump if the system was removed unexpectedly by SFM

Notify System Programmer

Message	Suggested Action
IXL008I	Path to CF invalidated (links miscabled) 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

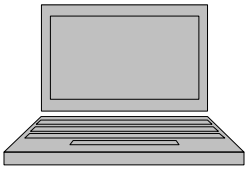
Notify System Programmer

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

SYNCHDEST

- Ensure CONSOLxx contains SYNCHDEST(groupname)
- Groupname is the name of a console group declared in CNGRPxx
- SYNCHDEST will try to present the SYNCH WTOR or SYNCH WTO to the physically attached MCS consoles on the system where the message was issued
- SYNCHDEST increases the likelihood of a system programmer noticing the message sooner during the hours when the systems are actively monitored
- The SYNCH message will go to the first declared console and stay there for 125 seconds then roll to the next console in the list. Eventually, the SYNCH message will go to the system console and stay there until replied to.
- Check the priority box and respond to SYNCH WTOR
- SYNCH WTOs will stay for 10 seconds and then system processing will resume.
- Refer to WSC FLASH10761 on synchronous WTOR processing
  - www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/FLASH10761

Last Resort



- System Console
  - Ensure access to the system console
  - z/OS 1.11 and above V CN(*),ACTIVATE not required to enter commands
- If MCS, SMCS and EMCS consoles are not responding, access the system via the System Console on the HMC
- Strongly consider enabling accessing of the HMC over the web
- Consider leveraging the Console Actions -> Monitor System Events to monitor for SYNC WTORs or SYNC WTOs, especially for times when system programmers are not actively monitoring the system

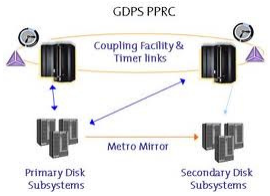
Healthchecks

- 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.

Health Checks in a GDPS Environment

- Some of the IBM defaults for certain health checks do not coincide with settings that are recommended in a GDPS environment. You may need to:
  - Update the check, or
  - Ignore certain results, or
  - Deactivate it entirely
- For example:
  - XCF_SFM_CONNFAIL should be no
  - XCF_TCLASS_HAS_UNDESIG ignore exception for GDPS TClass
  - Deactivate CATALOG_IMBED_REPLICATE check on k-sys
  - Deactivate RACF_GRS_RNL check on k-sys
  - XCF_FDI minimum FDI should be “spin FDI” (165 seconds)
- GDPS has its own checks as well



Questions?