



Coupling Facility Configuration Options

David Raften
Raften@us.ibm.com

Table of Contents

The Coupling Facility	5
Stand-alone Coupling Facilities.....	9
Logical Stand-alone Coupling Facilities.....	9
Message Time Ordering and STP	10
IEEE 1588 Precision Time Protocol (PTP).....	12
Coupling Facility LPAR on a Server.....	12
Dynamic Coupling Facility Dispatch.....	13
Dynamic CF Dispatch Performance	14
Recommendation: Make sure you have a XCF structures defined on each CF.....	15
Internal Coupling Facility.....	16
Performance of links.....	23
Dense Wavelength Division Multiplexer	24
Performance of Internal Coupling Facility (ICF)	26
Dedicated and Shared CPs.....	27
Performance of the Coupling Facilities.....	29
Number of CF Engines.....	29
Comparison of Coupling Efficiencies Across Coupling Technologies	30
Cost/Value Implications of Each CF Option	33
Standalone Coupling Facilities	33
System-managed CF Structure Duplexing	36
Asynchronous CF Structure Duplexing for Db2 Lock Structures.....	36
Structure Placement	39
Resource Sharing Environment.....	39
Resource Sharing: CF Structure Placement	42
Data Sharing Environments	43
Data Sharing: CF Structure Placement	43
Lock Structures	45
Cache Structures.....	45
Transaction Log Structures	46
Other CF Structures Requiring Failure Independence	47
Other CF Structures NOT Requiring Failure Independence	47
Failure Independence with System-managed CF Structure Duplexing	49
Combining an ICF with Standalone CF	49
Other Configuration Considerations	51
CFCC concurrent patch apply:	51
Single Server Parallel Sysplex:	52
Single ICF or Two ICFs:	53
Single z/OS Parallel Sysplex:	53
Other Considerations and Upgrade Paths	54

Coupling Facility Configuration Options

Pervasive Encryption Support	55
Summary	55
Appendix A. Parallel Sysplex Exploiters - Recovery Matrix	57
Appendix B. Resource Sharing Exploiters - Function and Value.....	60
Resource Sharing Exploiters.....	60
XCF, High Speed Signaling.....	61
System Logger, OPERLOG and LOGREC	61
z/OS GRS Star.....	62
z/OS Security Server (RACF), High-speed Access to Security Profiles	62
JES2, Checkpoint.....	62
SmartBatch, Cross System BatchPipes	63
VTAM GR (non LU6.2) for TSO	63
DFSMSHsm Common Recall Queue	63
Enhanced Catalog Sharing	63
WebSphere MQ Shared Message Queues	64
Workload Manager (WLM) Support for Multisystem Enclave Management	64
Workload Manager support for IRD	64
Recommended Configurations.....	65
Appendix C. Data Sharing Exploiters - Usage and Value	67
Db2 Data Sharing.....	67
VTAM, Generic Resources.....	67
VTAM, Multi-Node Persistent Session	68
CICS TS, CICS LOG Manager.....	68
VSAM RLS	69
IMS Shared Data	70
IMS Shared Message Queue.....	70
IMS Shared VSO.....	71
VSO DEDB.....	72
RRS, z/OS SynchPoint Manager.....	72
Recommended Configurations.....	73
Appendix D. CF Alternatives Questions and Answers	75

Coupling Facility Configuration Options

Coupling Facility Options

IBM's Parallel Sysplex® clustering technology (Parallel Sysplex) continues to evolve to provide increased configuration flexibility, improved performance and value to our customers.

With this increased flexibility in configuring the many Coupling Facility (CF) options, IBM has recognized the need to provide guidance in the selection of the Coupling Facility option that best meets a given customer's needs.

This paper examines the various *Coupling Facility* technology alternatives from several perspectives. It compares the characteristics of each CF option in terms of function, inherent availability, performance and value. It also looks at CF structure placement requirements based on an understanding of CF exploitation by z/OS® components and subsystems. Implementation of specific CF configurations is based on evaluation of Coupling Facilities and links in terms of:

- *Performance*
- *Business Value*
- *Cost*
- *Availability*

The objective of this white paper is to provide the customer with the information to make a knowledgeable decision. Recommendations of structure placements in relation to the above aspects are given but every customer needs to make the final decision that best suits their requirements and expectations.

More information on IBM Parallel Sysplex can be found at:

<https://www.ibm.com/products/zos/parallel-sysplex>

More information about the IBM Z® servers can be found at: <https://www.ibm.com/z>

Coupling Facility Configuration Options

The Coupling Facility

A Coupling Facility (CF) is the heart and soul of a Parallel Sysplex. From a hardware perspective, a CF is very similar to z/OS.

- It runs in an LPAR. The LPAR can use shared or dedicated CPs. The only difference is that a CF LPAR can also use a lower priced Internal Coupling Facility (ICF) specialty engine.
- The LPAR has memory. Where z/OS memory is divided into Private, Nucleus, CSA, SQA, etc., the CFCC uses pre-formatted “Structures”
- The LPAR has I/O capabilities. These are the CF Links to send requests in and out of the CF. The CF has no access to disk.
- It runs a licensed operating system, the Coupling System Control Code (CFCC). This is shipped and maintained as part of every IBM Z.
- The operating system has different releases (CFCC Levels)
- Operators can issue commands to the CFCC. They do so through the secure IBM Z Hardware Management Console (HMC) interfaces.

Despite the conceptual similarities to a z/OS LPAR, the CF has many key differences such as implications in the type of LPAR configured, where structures are placed, and what configuration options are chosen.

All IBM CF Control Code (CFCC) is Licensed Internal Code (LIC) that runs within a PR/SM™ Logical Partition (LPAR) regardless of whether the CF processors are dedicated or shared, or whether the servers also run z/OS or are “standalone” CFs. All CFs within the same family of servers can run the same CFCC code load and the CFCC code always runs within a PR/SM LPAR. Thus, when positioning certain CF configurations as fit for data sharing production such as a standalone model while characterizing other CF configurations as often more appropriate for test/migration (e.g., ICF or a logical partition on a server without System-managed CF Structure Duplexing) one must be careful not to erroneously assume such positioning is based on fundamental functional differences.

The inherent availability characteristics of the various CF offerings are, at their core, the same. However, IBM often recommends using one or more standalone CFs for maximum service availability. The essential reason IBM often makes such recommendations is that a Parallel Sysplex environment can deliver higher availability when the CFs are isolated from the processors on which software exploiters are executing from one another, to remove single points of failure from the deployment architecture. This paper explores many of the details behind IBM’s CF configuration and deployment recommendations and suggestions.

Coupling Facility Configuration Options

CFCC Levels

The following table highlights some of the CFCC functions in each release and on which servers they were made available on. All CF levels include the function of all previous CF levels. For more details on each function, please see “Coupling Facility Level (CFLEVEL) Considerations” linked off the Parallel Sysplex web page at:

<https://www.ibm.com/products/zos/parallel-sysplex>

CF improvements generally fall into one of several categories:

1. **Scalability**, to support larger Parallel Sysplex configurations
2. **Performance**, usually for data sharing environments since these use the Coupling Facility the most
3. **Functionality**
4. **RAS** benefits. These are in almost every new CF level, but not usually detailed.

In addition, a new server family and driver release often includes Parallel Sysplex support independent of the CFCC level such as for new links.

Coupling Facility Configuration Options

CF Level	Function	zEC12 zBC12	z13*	z14*	z15*	z16*	z17™
26	Nondisruptive Sys Managed copy process for lock structures Removal of support for CF images with VFM Removal of support for CF images on dedicated GPs						X
25	ICA SR coupling link protocol efficiency improvements Coupling Express2 LR coupling links CF Image scalability improvements Cache retry buffer support for IFCC recovery Lock record data reserved entries for structure full recovery Cache residency time metrics for directory/data entries Deprecate DYNDISP OFF/ON, keep only DYNDISP THIN option for shared-engine CF images Quantum-safe encryption for the CFCC firmware					X	
24	CFCC Fair Latch Manager improvements Message Path SYID Resiliency Enhancement Shared-Engine CF Default is changed to "DYNDISP=THIN" CF Monopolization Avoidance 384 Coupling CHPIDs per CPC				X		
23	Asynchronous cross-invalidate (XI) of CF cache structures Coupling Facility hang detect enhancements Coupling Facility ECR granular latching			X			
22	CE LR coupling link support (CL5) List notification enhancements CF performance scalability Support for CF Encryption Additional diagnostics		X				
21	Asynchronous CF Duplexing for Lock structures Up to 40 ICA SR links per CPC BCPii interface to collect CF information		X				
20	ICA SR coupling link support (CS5) Large cache structure performance CFCC processing scalability support 256 Coupling CHPIDs per CPC. A CF image is still limited to 128 coupling CHPIDs.		X				
19	Thin Interrupt support for shared CPs Flash memory exploitation for MQ shared queues	X					
18	RMF reporting improvements on CF links Non-disruptively dump enhanced diagnostic data Verification of local cache controls Db2® conditional writes Dynamic alter performance improvements	X					

Coupling Facility Configuration Options

	Storage class and castout class contention avoidance						
17	Non-disruptive dumps 2047 Structures per CF image 247 connectors to lock structures 127 connectors to (serialized) list structures 255 connectors to (unserialized) list structures						
16	SM Duplexing Perf. IMS™/MQ Msg Queue Perf						
15	RMF™ Reporting enh. 112 tasks						
14	CFCC dispatcher enh.						
13	Db2® Castout perf.						
12	IBM eServer™ zSeries® 990 (z990) compatibility SM Duplexing for zSeries 64-bit CFCC addressability Msg Time Ordering, Db2 Performance						
11	z990 Compatibility, SM Duplexing for 9672s						
10	z900 GA2 level						
9	IRD WLM Multi-System Enclaves WebSphere® MQ Shared Queues						
8	Dynamic ICF expansion into shared ICF SM Rebuild						
7	Shared ICF partitions on server Db2 Delete Name optimization						
6	ICB, ICF, TPF Support						
5	Db2 GBP duplexing Db2 castout performance Dynamic ICF expansion into shared CPs						
4	Dynamic CF Dispatch IMS SMQ extensions Internal Coupling Facility IMS and VSAM RLS performance						
3	IMS Shared Msg Queue						
2	Db2 performance VSAM RLS 255 cache connectors 1023 structures						
1	Dynamic Alter System Logger CICS® Temp Storage Queues						

More details on the CF levels can be found in the **Technical Guide** redbook for the current server at <https://www.redbooks.ibm.com/>, as well as the PR/SM Planning Guide at <https://www.ibm.com/docs/en/systems-hardware>.

Coupling Facility Configuration Options

For a list of the software levels that use the function and levels that can coexist with the CF, see the “Summary of CFLEVEL Functions” section of the z/OS MVS Setting Up a Sysplex document.

Stand-alone Coupling Facilities

The standalone CF provides the most robust CF capability, as the server is wholly dedicated to running the CFCC licensed internal code — all the processors, links and memory are for CF use only. A standalone CF models provide for maximum Parallel Sysplex connectivity since there are no other operating systems competing for the I/O connections. The maximum amount of links dependent solely on how many CPC drawers and fanout cards are installed.

A benefit of a standalone CF, which precludes running any operating system software, is that it is always failure-isolated from exploiting z/OS software and the server that z/OS is running on for environments without configuring System-Managed CF Structure Duplexing.

Standalone CFs can be obtained by ordering the ICF-only engine type in the standard server models.

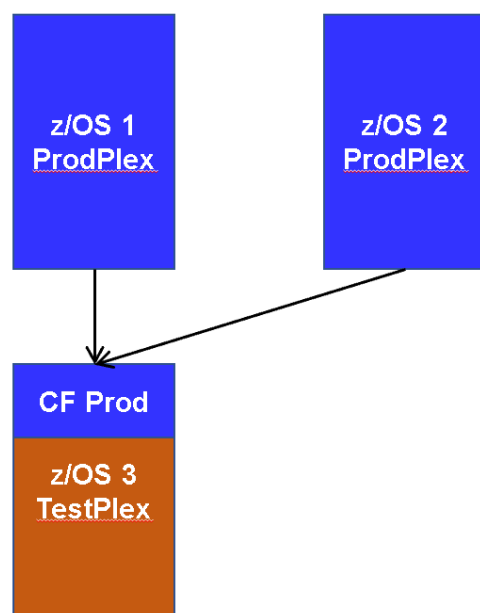
Logical Stand-alone Coupling Facilities

A standalone CF can also be a "logical" standalone. Here, the CF is isolated from the rest of the Parallel Sysplex LPARs by running on a different server. The difference is that other LPARs can be on the servers that are outside of the Parallel Sysplex environment. For example:

1. *A z/OS LPAR that is a member of a different Parallel Sysplex cluster*
2. *A z/OS LPAR that runs Db2 but connects to structures in a different CF*
3. *A network node (CMC)*
4. *A zTPF image*
5. *z/VM® or KVM for Z hypervisors with the images that they manage*
6. *A Linux® system*

In a logical standalone, one can get all the availability benefits of a physical standalone for the members of the Parallel Sysplex environment but without needing to dedicate a server for the CF. This is seen in the figure below:

Logical Stand-Alone CF

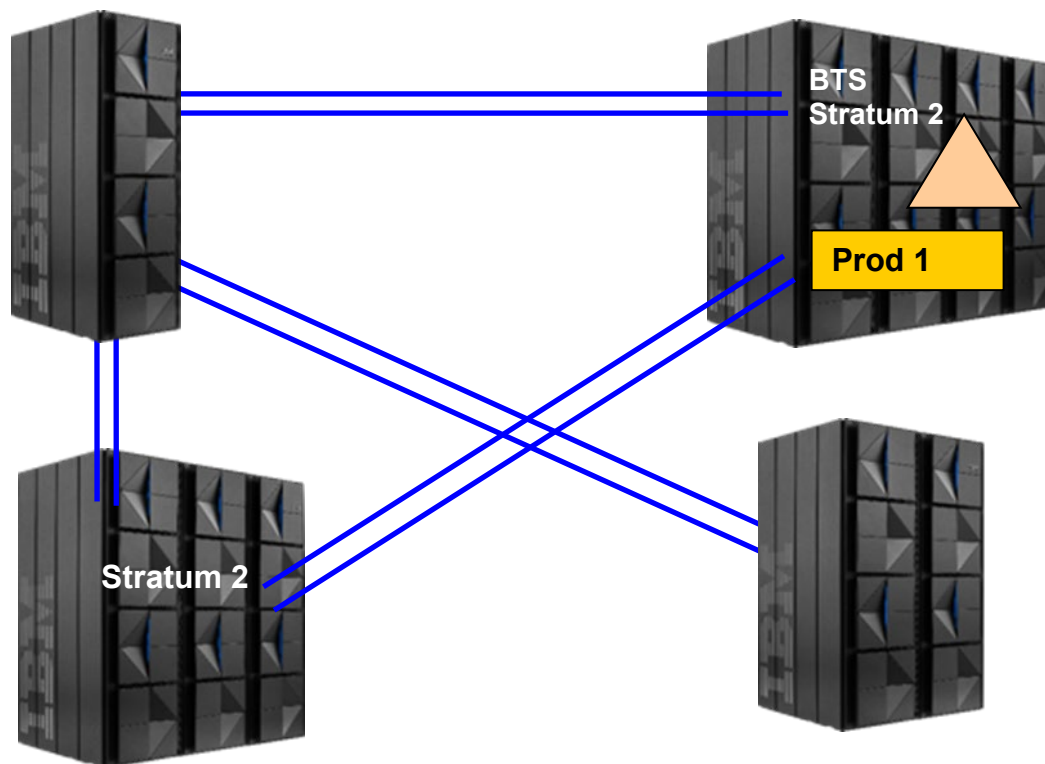


In this figure, CF Prod is a "Logical Stand-alone" CF. Although there are other z/OS images running on the same server, none of those are in the same Parallel Sysplex environment that CF Prod supports.

Message Time Ordering and STP

As processor and Coupling Facility link technologies have improved over the years, the synchronization tolerance between systems in a Parallel Sysplex cluster has become more rigorous. To help allow for any exchanges of time stamped information between systems in a sysplex involving the Coupling Facility observe the correct time ordering, time stamps are included in the message-transfer protocol between z/OS and the Coupling Facility. Therefore, the Coupling Facility requires connectivity to the same time source that the systems in its Parallel Sysplex cluster are using for time synchronization. If the Coupling Facility is on the same server as a member of its Parallel Sysplex cluster, no additional timer connectivity is required since the server already has connectivity to the time source. This is shown in the picture below. The Stand-alone Coupling Facility on the top left is the time source itself (Stratum 1). The CF on the top right has Server Time Protocol (STP) connectivity to the Stratum 1. If the Stratum 1 was the server on the top right (Prod 1 server), then the server on the top left would need STP connectivity.

Coupling Facility Configuration Options



All servers in the Sysplex need to be connected to the common Stratum 1 time source. Since STP uses coupling links and the Coupling Facility already has CF link connectivity to the servers in the Parallel Sysplex, the server containing the CF can be configured as the Stratum 1 for the timing network with minimal additional external links. That said, some chose to configure the Stratum 1 on a server with z/OS so STP messages can be monitored easier.

More information on STP can be found at

<https://www.redbooks.ibm.com/abstracts/sg247280.html?Open=>

Coupling Facility Configuration Options

IEEE 1588 Precision Time Protocol (PTP)

The STP Coordinated Timing Network (CTN) consists of a collection of servers and CFs that are time-synchronized to a time value called Coordinated Server Time (CST). The CTN is able to maintain synchronization to Universal Coordinated Time (UTC) through network connectivity and a time synchronization protocol to an external time reference receiving a time signal traceable to UTC. For IBM Z machines prior to IBM z15*, this external time reference would typically be Network Time Protocol (NTP). Starting with IBM z15, this external time reference could also be IEEE 1588 Precision Time Protocol (PTP). Prior to IBM z16*, the external time reference network connectivity occurred through the SE Service Word Interface and the internal service network.

For clients requiring accuracy to UTC of 100 milliseconds or less, pulse per second (PPS) connectivity is also available.

IBM z16 introduces support for connecting the CEC directly to PTP or Network Time Protocol (NTP) external time references through the data center network. For example, bypassing the service word interface and internal service network. PTP or NTP traffic will now come into the CEC through ethernet connectivity to a new adapter card that is in turn connected directly to IBM z16 oscillators. This new adapter card features hardware timestamping capabilities, allowing IBM z16 to take full advantage of the accuracy of PTP. These enhancements will result in STP being capable of significantly enhanced accuracy to UTC compared to previous IBM Z machines and possibly eliminate the need for PPS connectivity.

Many sites are subject to UTC accuracy regulations, which the enhancements to the external time reference may help meet. These regulations include the European Securities and Markets Authority (ESMA) directive on markets in financial instruments (MiFID 2) Regulatory Technical Standard (RTS) 25. The regulations also include the US Financial Industry Regulatory Authority (FINRA) rules on financial transaction timing accuracy and the Securities and Exchange Commission (SEC) Rule 613 Consolidated Audit Trail (CAT).

Better UTC time accuracy against the external time reference also implies better UTC time accuracy compared to the distributed systems servers, getting tighter time synchronization between all systems involved in the financial transactions systems seeking a flexible and simple Coupling Facility (CF) configuration can enable a standard LPAR on their server with CF Control Code.

Coupling Facility LPAR on a Server

Customers seeking a flexible and simple Coupling Facility (CF) configuration can enable a standard LPAR on their server with CF Control Code. This might be particularly attractive to customers beginning their Parallel Sysplex implementation or those who have “spare” capacity on their server to support a CF. While

Coupling Facility Configuration Options

this approach does not offer the same availability as a standalone coupling solution since it does not offer the physical independence of a separate server, it is appropriate for Resource Sharing production and test application datasharing environments. Unlike the standalone CF and ICF approach which require additional hardware planning, this implementation offers the advantage of using existing MIPS to enable a Parallel Sysplex environment. Customers can start with a small amount of capacity and scale the CF up MIPS-wise, based on workload requirements. The [Dynamic CF Dispatch](#) feature (described below) is highly recommended for this environment.

Running the CFCC in a standard PR/SM LPAR on a general-purpose server is similar to running it on a standalone model. Distinctions are mostly in terms of price/performance, maximum configuration characteristics, and recovery characteristics. For example, a IBM z17™ can support up to 96 coupling links. Realistically, one would not do this if this server is also running a major z/OS workload as the links uses slots that would otherwise be used for FICON® and OSA channels. Also, running the CF on shared general purpose CPs can result in paying extra software license charts as the capacity will be associated with z/OS (as of z17, running the CF on dedicated general purpose CPs is not supported). The ability to run z/OS workloads side-by-side on the same physical server with the CF in the same Parallel Sysplex cluster can also introduce the potential for single points of failure for certain environments. These considerations will be explored later.

Customers seeking a coupling environment for intensive datasharing should closely evaluate the advantages of a standalone coupling solution or an Internal Coupling Facility (ICF) implementation.

Dynamic Coupling Facility Dispatch

The Dynamic CF Dispatch function provides an enhanced CFCC dispatcher algorithm for improved price/performance. Originally, the Coupling Facilities had a tight polling loop, looking for work arriving on their CF links. If it found work, the work was processed. If there were no new CF requests, it just looped around and looked again. Although this allowed for the fastest response times when the CF is running on dedicated CPs, it “did not play well with others” when sharing a processor with other LPARs on the server. PR/SM would dispatch the CF whenever it can, believing it was busy with productive work, and the CF can use resources that would otherwise go to z/OS. The LPAR could be “capped,” but that caused other tuning problems.

Dynamic CF Dispatch allows the strict tight polling loop algorithm to be replaced with a dispatching algorithm so the CF give up control of the physical processor voluntarily when there are no CF commands queued for execution, and be dispatched when work comes in. This behavior enables the processor to be fully used for customer workloads until CF processor resource is needed. CF CPU consumption is limited to the resource level required.

Coupling Facility Configuration Options

Dynamic CF Dispatch algorithms apply to all non-dedicated CF processor configurations, including standalone models.

There are three environments where DCFD is typically used:

1. Running multiple CF partitions that share a pool of CPs. This can help minimize the amount of CF engines that are configured. Many use this for small production environments where the CF capacity requirements are very small.
- 2 To support a System Programmer “sandbox” environment
- 3 To support a hot-standby CF. This can help eliminate the need to dedicate an entire processor for the second Coupling Facility use just in case of primary Coupling Facility failure.

When configured for use as a hot-standby, if there is a problem with the primary CF, there is a lot of XCF communication between z/OS images to rebuild structures to the next available CF in the PREFLIST. Unless CTCs are used to back up the XCF structures, you must remember to have an active set of XCF path structures in the hot-standby CF. XCF may see that using these structures results in higher response time than when going through the “primary” CF, so it will tend to favor the primary CF. Despite that, the backup CF will continue to see some minor XCF activity.

When the CF image is using shared general purpose processors that are from the same sharing pool as the z/OS shared processors, if a z/OS issues a CF command to a CF running with DCFD from the same server as that CF, then all CF commands are converted from synchronous mode to asynchronous to avoid possible processor deadlocks. CF commands destined to other CFs proceed as normal.

Dynamic CF Dispatch Performance

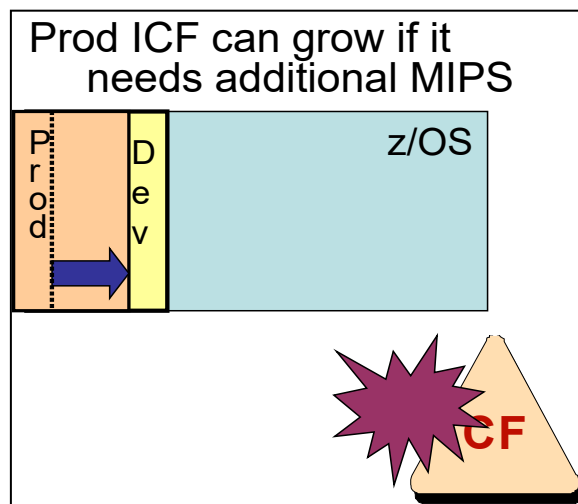
With the introduction of Dynamic Coupling Facility Dispatch (DYNDISP), additional value is provided for running shared CF LPARs. With DYNDISP=THIN, the percentage of the processor used in the CF logical partition is dependent on the request rate to the CF. Similar to z/OS, the CF would run when requests come in and then give up control back to PR/SM when there is no more work. N With Shared processor CF images the server will always use DYNDISP=THIN.

Note: Use DYNDISP=THIN is the only supported option when using shared CP

Coupling Facility Configuration Options

DYNDISP=THIN is the only supported option starting with the z16. Dedicated CPs (cores) are still recommended, but if you are using share CPs, use DYNDISP THIN. When running a Coupling Facility on a general processor, software charges apply to CFCC MIPS. However, DCFD eliminates the need to dedicate an entire core for Coupling Facility when using shared CPs.

IBM z16 is planned to be the last generation of IBM Z to support use of dedicated general-purpose processors (GPs) to provide processing capacity for Coupling Facility (CF) images. Use of dedicated ICF processors are much more cost-effective.



An example of using DYNDISP=THIN is if the production parallel sysplex (non-data sharing environment) is sharing a hot stand-by general purpose processor with a development CF (see figure above). Normally a stand-alone CF contains all of the CF structures (although the backup CF needs to also contain XCF structures). If the production CF is unavailable, the activity will move to the backup CF. Best performance can be ensured by giving this a high LPAR weight. This has an advantage by not requiring a second stand-alone CF image, but it charges its MSUs against z/OS on the rare occasions when it is active.

More information on DYNDISP can be found in the PR/SM Planning Guide on IBM documentation:

<https://www.ibm.com/docs/en/systems-hardware>

Recommendation: Make sure you have a XCF structures defined on each CF.

Coupling Facility Configuration Options

Coupling Thin Interrupts enable the CF to be used for more efficient processing with shared-engine CFs. When not doing productive work, the CF will go into an enabled wait state. Interrupts would then wake up the CF to process incoming messages. The CF proceeds to poll the CF links looking for work as normal. When it finds no more work to process, the CF will then put itself back to an enabled wait state. This enables faster CF response times and more efficient use of shared CF processors.

Coupling Thin Interrupts is also employed by z/OS to provide improved service times for asynchronous processes. A number of asynchronous coupling facility processes depended on polling by z/OS to detect their completion. With Coupling Thin Interrupts enabled on the z/OS side, an interrupt will be generated by

- The arrival of the response for an asynchronous CF request
- The arrival of a CF notification signal indicating the presence of work to be processed in a list structure such as processing incoming work in the form of XCF signals or messages queued to IMS or MQ shared queues.

This interrupt will cause the z/OS partition to be dispatched more promptly than the existing timer-interrupt-based algorithm would have, further reducing the latencies in processing the work and further improving response times.

Note: IBM recommends dedicated Coupling Facility CPs for best response times, but if you share CF CPs, Coupling Thin Interrupts is the default on z16 and the only option on z17.

Recommendation: Thin Interrupts support should always be used on the z/OS side, when the z/OS server supports it

Internal Coupling Facility

The chip design and technology on the Z servers provide the flexibility to configure the cores for different use. Cores can be used as Central Processor (CPs), System Assist Processors (SAPs), or specialty engines such as the Integrated Information Processors (zIIPs), Internal Coupling Facility processors (ICFs), and Integrated Facility for Linux processors (IFLs). For the general purpose processor model, the number of CPs are specified by the capacity setting, and the number of SAPs (provide I/O processing) are specified by the model number. This capability provides tremendous flexibility in establishing the best system for running applications. Cores can also be reserved as a spare.

Coupling Facility Configuration Options

An internal Coupling Facility is a CF image configured to run on ICF specialty engines. Special PR/SM microcode precludes the defined ICF processors from executing non-CFCC code. Changes to the number of ICF processors can be made dynamically through an On/Off Capacity Upgrade on Demand upgrade. Since the IBM Z servers can be configured as ICF-only, the maximum number of ICFs that can be defined for each server is the same as the maximum number of general purpose engines.

The ICF is a highly attractive CF option because, like the standalone models, can be attractively priced relative to the standalone CFs as a hardware feature. This is achievable given that the cost of the processor machine “nest” (frame, power and cooling units, etc.) is amortized over the server system as a whole and not just the burden of the CF component of the system, as is the case in a standalone Coupling Facility. In addition, an ICF does not require significantly more power and cooling for the server it is on, and does not require additional maintenance fees.

Running with an ICF can improve on software license charges. There are three possibilities for how software is charged:

- **Monthly License Charges (MLC)**

z/OS and associated software is charged based on the MSUs of the machine model it runs on. If z/OS requires nine core’s worth of capacity and the CF requires one core’s worth of capacity, then one can configure a 9-way z/OS server with one ICF engine and being charged based on the 9-way server, or one can configure a 10-way server, including one CP just being used to run the Coupling Facility. MCL charges here are based on the 10-way. Here, one can save on software by configuring a smaller server with an ICF

- **Sub-capacity Licensing**

z/OS runs the SubCapacity Reporting Tool (SCRT). This uses SMF70 records to report on its 4-hour rolling average. Since the Coupling Facility does not generate SMF70 records, its usage is not recorded by SCRT and does not affect the reported CPU usage by z/OS. From an IBM software point of view, running a CF on general purpose engines does not affect the software charges. However, ISVs may still be looking at the total general purpose MIPS on the floor and charge for that. Each client must have this discussion with their vendors.

Coupling Facility Configuration Options

- **One Time Charge**

Some countries have a one-time charge for z/OS. This is based on the total capacity of the server. In the example for MLC, there will be a one-time charge for the full capacity, a 10-way server, even though one CP is running the Coupling Facility. As with MLC, running an ICF would help reduce the charge. Any ISV software charges could also be based on the full server capacity. Each client must have this discussion with their vendors.

The ICF can be used to host a second Coupling Facility when a standalone CF is also being used, reducing the need for a second standalone Coupling Facility in a multisystem Parallel Sysplex configuration. The ICF can also serve as an active production Coupling Facility in many resource sharing environments or even in a data sharing environment with System-Managed CF Structure Duplexing.

Customers can use the ICF as a production Coupling Facility even in a single-server Parallel Sysplex configuration, running two or more z/OS logical partitions sharing data via the ICF. This configuration enables customers interested in getting into the Parallel Sysplex environment to take advantage of its continuous operations protection from software outages. Individual z/OS partitions can be taken down for maintenance or release upgrade, without suffering an application outage, through the datasharing provided by the remaining LPARs in the server. The ICF uses external Coupling Facility links or internal IC links. By putting the Coupling Facility in an ICF, both the benefits of reduced software license charges and reduced hardware costs can be realized.

Another typical use of an ICF in a single-server Parallel Sysplex environment is to use the capabilities of Intelligent Resource Director (IRD). While the z/OS Workload Manager (WLM) can manage CPU, Storage, and I/O resources for applications based on defined goals, the WLM scope is typically only within an LPAR. By WLM creating an IRD structure in a CF, the WLMs across multiple LPARs within the server can now communicate and adjust LPAR weights, the number of local CPs within the LPAR, I/O priority, and number of channels going to a disk subsystem. More information on WLM and IRD can be found in <http://www.ibm.com/systems/z/os/zos/features/wlm/>.

From a Total Cost of Ownership perspective, an ICF will generally be significantly cheaper than a standalone Coupling Facility solution. When performing the financial analysis, one must take into consideration the extra memory that may need to be configured and any memory increment sizes.

IBM z16 is planned to be the last generation of IBM Z to support use of dedicated general-purpose processors (GPs) to provide processing capacity for Coupling Facility (CF) images. Use of dedicated ICF processors are much more cost-effective.

Coupling Facility Configuration Options

While price can be a driving factor in purchasing decisions, configuration choices should be made based on availability, growth and performance requirements. Consider the following:

1. Cores can be configured as general-purpose CPs, IFLs, zIIPs, or ICFs.
2. When servers are upgraded to a new generation technology, the ICFs on that footprint are also affected by the server upgrade. Given that this upgrade requires a Coupling Facility outage along with a z/OS outage, customers should exercise their pre-planned reconfiguration policies and procedures. Standalone Coupling Facility upgrades are independent decisions from server upgrades.
3. When configuring an ICF, memory requirements are an important consideration.
4. ICF uses memory from the server pool of memory. Customers adding ICFs to their servers must determine whether additional memory is now required on the processor to support Coupling Facility functions.

Link Technologies

Coupling Short Distance 5 (CS5) is the fifth generation of short distance external coupling links and was available starting with the z13 family. The CS5 uses the IBM **Integrated Coupling Adapter (ICA SR)** connecting to the PCIe fanout. It provides up to 150 meters Coupling Facility connections

CS5 links support up to 7 subchannels, or 7 concurrent messages, per link. This allows better performance by reducing the bottleneck of the link during periods of high activity without requiring additionally physical links.

[Recommendation: Monitor subchannel usage via RMF MON III. Configure more subchannels and/or links if % Delay due to subchannel busy is 10% or greater.](#)

Using “Multiple Image Facility” (MIF), CS5 links can be used to connect to multiple z/OS Parallel Sysplexes at the “sender” end with multiple CFs at the “receiver” end. The z/OS’s each talking to a CF within its Parallel Sysplex.

The ICA SR1.1 links starting with the z15 are essentially the same as the ICA SR links, but with slightly different hardware. On the IBM z16 the ICA SR protocols were enhanced to provide up to 10% response time improvement for read requests and lock requests, and up to 25% improvement for write and duplexed write requests, compared to CF service times on the IBM z15 systems. This provided better coupling efficiencies. The CS5 links can connect to ICA SR, ICA SR1.1, and ICA SR2.0 links.

[IBM plans on discontinuing support for CL5 by z17+2](#)

Coupling Facility Configuration Options

The ICA SR 2.0 adapters come with the z17. The z17 will not support ICA SR or ICA SR 1.1, although they can connect to the ICA SR 1.1 on the z16 and the ICA SR 1.0 on the z15. They remain as link type CS5, but support:

- 8 primary send buffers / CHPID
- 4 logical CHPIDs / port
- Up to 48 adapters (96 ports) per CEC

Coupling Long Distance 5 (CL5) is the fifth generation of long distance external coupling links and is available starting with the z13 family. The CL5 uses the IBM **Coupling Express3 LR 10 Gb** (CE LR) links connecting to the I/O cage and uses the RDMA over Converged Enhanced Ethernet architecture. CE LR provides up to 10 kilometers base connectivity, up to 100 km with qualified DWDMs.

CL5 links support up to 32 subchannels, or 32 concurrent messages, per link. As with the CS5, this allows better performance by reducing the bottleneck of the link during periods of high activity without requiring additional physical links. The additional subchannels in the long-range environment provides additional throughput per CHPID by making more efficient use of the link bandwidth. Since each long-range request can take significantly longer than short range requests due to the latency increase from the distance involved and the speed of light over the fiber, the additional subchannels allowed more requests to be in-flight simultaneously, increasing the throughput.

[Recommendation: Monitor subchannel usage via RMF MON III. Configure more subchannels and/or links if % Delay due to subchannel busy is 10% or greater.](#)

Using “Multiple Image Facility” (MIF), CL5 links can also be used to connect to multiple z/OS Parallel Sysplexes at the “sender” end with multiple CFs at the “receiver” end. The z/OS’s each talking to a CF within its Parallel Sysplex.

The IBM z16 servers introduced the Coupling Express2 LR adapters. These provide a throughput improvement of up to 15% when connected to another CE2 LR adapter, compared to CE LR links on the IBM z15.

The CL5 links can connect only to other CL5 links.

Coupling Facility Configuration Options

Coupling Long Distance 6 (CL6) are available with the IBM z17 servers. As with the CL5, they support up to 10 km unrepeated, and up to 100 km with qualified DWDMs. CL5 and CL6 links connect to Coupling Express 3 LR 25Gb adapters.

- CL6 links can connect only to other CL6 links. This uses 25 Gb optics.
- For z17 communication back to previous servers, CL5 links must be installed. They can be on a Coupling Express3 LR adapter, running at 10 Gb optics, connected to CE LR or CE2 LR links on the z15 or z16.

CL5 and CL6 links support:

- 32 primary send buffers / CHPID (32 or 8 subchannels / device)
- Max 32 adapters (64 ports) per CEC

Extended Distance RPQ:

The RPQ titled, “**Ext Distance for IBM Z.**” is available for configuring over 10 km without DWDMs or going over 100 km with qualified DWDMs. The customer collects information such as the length of the fiber, the amount of bends in the fiber, the number of splices, the type of equipment, measurements of the light strength at both ends (“link loss budget”), any Dispersion Compensation Modules in the DWDMs, etc. This documentation is sent to IBM where it is run through a model. If IBM determines that this configuration will work, the RPQ is submitted and IBM will guarantee the configuration. The RPQ number varies depending on the server. More information can be found in “Coupling Facility Channel I/O Interface Physical Layer Document” (SA23-0395) as well as in the “IBM Z Connectivity Handbook” at <https://www.redbooks.ibm.com/abstracts/sg245444.html> .

Coupling Facility Configuration Options

Number of Books / Drawers	1	2	3	4	5
z14 12x IFB	8	16	24	32	
z14 1x IFB	16	32	48	64	
z14 CS5 (ICA SR)	20	40	60	80	
z14 ZR1 CS5	16				
z14 CL5 (CE LR)	64	64	64	64	
z14 ZR1 CL5	32				
z15 T01 ICA SR / ICA SR1.1	24	48	72	96	96
z15 T02 ICA	24	48			
z15 T01 CE LR	64	64	64	64	64
z15 T02 CE LR	64	64			
z16 ICA SR1.1	24	48	72	96	
z16 CE2 LR	64	64	64	64	
z17 ICA SR2.0	24	48	72	96	
z17 CE3 LR	64	64	64	64	

Note: Two I/O cages are needed to support the maximum amount of CE LR links.

The Internal Coupling channel (IC) is a microcoded “linkless” coupling channel between CF LPARs and z/OS LPARs on the same server. By using a memory-to-memory communication, it enables exceptional performance benefits. Additionally, LPARs using ICs to communicate internally within a server can simultaneously use external Coupling Links to communicate with CFs and z/OS systems external to the server. Please refer to the Performance section of this paper for details on the added value of these functions.

It is suggested that you define a minimum of internal coupling channels. For most customers, IBM suggests defining just one pair of ICP channel paths for each coupling facility logical partition (LP) in your configuration. For instance, if your general purpose configuration has several z/OS LPs and one CF LP, you would define one pair of connected ICP CHPIDs shared by all the LPs in your configuration. If your configuration has several z/OS LPs and two CF LPs, you still would only define one connected pair of ICP CHPIDs, but one ICP CHPID should be defined as shared by the z/OS LPs and one CF LP while the other ICP CHPID is defined as shared by the z/OS LPs and the other CF LP. Both of these examples best exploit the peer capabilities of these coupling channels by using the sending and receiving buffers of both channels.

Coupling Facility Configuration Options

Maximum number of ICP CHPIDs: Real CPU resources are used to implement the link function of connected ICP CHPIDs. Production environments should limit the maximum number of internal coupling links that are defined for a CPC to optimize the internal coupling link function utilization of CPU resources. This maximum number of internal coupling links is based on the number of available physical CPs on the server. This maximum number of internal coupling links can be calculated by taking the number of CPs in the CPC that are used for general-purpose CPs and for ICF CPs, and subtracting one from that total. For example: a server that consists of four general-purpose CPs and two ICF CPs would have a **recommended** maximum five $[(4 + 2 = 6) - 1 = 5]$ internal coupling links. This example represents a total of 10 ICP CHPIDs being defined.

Server	IC	12x IFB3	1x IFB	CS5 (ICA SR)	CL5 (CE LR)	CL6 (CE3 LR)	
z14	32	32	64	80	64		
z14 Model ZR1	32			16	32		
z15	64			96	64		
z16	64			96	64		
z17	64			96		64	

More information can be found in IBM Z Connectivity Handbook:

<http://www.redbooks.ibm.com/redbooks/pdfs/sg245444.pdf>

Performance of links

The Internal Coupling (IC) links provides a high-speed internal communication link between a z/OS LPAR and a CF LPAR on the same server. In comparison to the performance of external links, the IC channel can improve the coupling efficiency of the system by one to five percent. Since by definition an IC must be within the server, one cannot have a mix of different server types while using ICs. The capacity of an IC link is directly related to the cycle time of the server. IC links are memory-to-memory and do not require any additional hardware.

The link speed numbers shown in the table below reflect expected effective data transfer rates in a production environment.

Effective CF Link Speeds

Coupling Facility Configuration Options

Link Speed MB/sec	IC	CS5 (ICA SR)	12x IFB3	12x IFB	CL5 (CE LR)	CL6 (CE3 LR)
<u>z15 T02</u>	<u>7600</u>	6000 (0-70 m) 3700 (70-150m)	<u>NA</u>	<u>NA</u>	<u>700</u>	
<u>z15</u>	<u>8900</u>	6000 (0-70 m) 3700 (70-150m)	<u>NA</u>	<u>NA</u>	<u>700</u>	
<u>z16 A02</u>	<u>7300</u>	<u>5000</u>			<u>900</u>	
<u>z16</u>	<u>8100</u>	<u>5000</u>			<u>900</u>	
<u>z17</u>	<u>10,000</u>	5000 – up to 8K data size 3000 – for larger	<u>NA</u>	<u>NA</u>	<u>900</u>	<u>2000</u>

See section “Coupling Facility LPAR on a Server” for the maximum number of links on each server family.

Dense Wavelength Division Multiplexer

The Dense Wavelength Division Multiplexer (DWDM) is designed to use advanced dense wavelength multiplexing technology to transmit multiple channels over a single pair of fibers at remote distances by simultaneously using different frequencies. Depending on manufacturer, the technology supports at least 192 channels each running 100 Gbps or more over a distance of 100–300 km. This distance can be extended by using optical amplifiers. Within a Parallel Sysplex environment, Server Time Protocol (STP) can be used to extend a Parallel Sysplex environment greater than 100 km.

DWDM can provide significant benefits across a number of applications but is particularly valuable in that it enables the use of GDPS® for disaster avoidance and recovery. A GDPS/Metro configuration consists of a Parallel Sysplex cluster spread across two or more sites at distances of up to 100 km. It provides the ability to perform a controlled site switch for both planned and unplanned outages, with minimal or no data loss, while maintaining full data integrity. Distances of up to 200km are possible with qualified DWDMs and the RPQ titled, “**Ext Distance for IBM Z.**”

As the distances increase between the z/OS LPAR and the CF, and between the z/OS LPAR and the disk, service times increase for CF and I/O requests. Service time increases by approximately 10 microseconds

Coupling Facility Configuration Options

per km distance. Effect of transactions as part of **workloads**, however, may be elongated longer than the service time increases due to secondary effects of waiting for locks and other resources.

Performance Implications of Each CF Option

With the rich and diverse set of Coupling Facilities that are made available by IBM for Parallel Sysplex functions, it is both interesting and worthwhile to examine the performance characteristics of these alternative configurations. While each Coupling Facility has some unique features, the general performance characteristics of the CFs participating in a Parallel Sysplex cluster are similar. In the general discussion of coupled systems' performance, internal throughput rate (ITR), internal response time, and coupling efficiency are of interest. The internal throughput rate for online transaction processing (OLTP) may be denoted as the number of transactions completed per CPU second. The transaction's internal response time is a combination of the CPU busy time and the I/O time. The coupling efficiency of a given system is a measure of its productivity while participating in a Parallel Sysplex environment. Formally, the coupling efficiency of a Parallel Sysplex cluster may be defined as a ratio of the coupled systems' ITR to the aggregate ITR of the uncoupled systems.

A small overhead is associated with the enablement of a Parallel Sysplex cluster. This could be viewed as the cost to manage multiple z/OS images. Roughly, a 3% cost in total capacity is anticipated when a customer intends to migrate his applications from a single system to multi-systems. This includes overhead for shared disk, JES2 MAS, and (base) sysplex. This 3% cost is less if any of these options were already configured. On the other hand, the cost of datasharing in a Parallel Sysplex cluster depends on the characteristics of a customer's workloads. More specifically, it depends on the frequency and type of CF accesses made by the requesting systems. Even at a heavy production datasharing request rate (8–10 CF accesses per million instructions), the datasharing cost in a Parallel Sysplex cluster will generally be 11–12% of the total systems' capacity, resulting in a coupling efficiency of 88 - 89%.

The cost of a CF access is comprised of software and hardware components. The software component includes the subsystem (such as, Db2) and the operating system (z/OS) path lengths to setup for and complete a request, and in general is independent of the CF type. The hardware component is the CF service time, which includes the time spent in the requester hardware, in the requester's link adapter, in the coupling link, in the CF's link adapter, and in the Coupling Facility Control Code.

Coupling Facility Configuration Options

For CF read/write requests associated with data transfers, the link speed is of particular importance. The higher the effective link speed, the faster the data transfer. The “speed” refers not to how fast the links can carry a single bit from one point to the other, but rather how quickly it can get the bit into the system in the first place. IC “links” are fastest and depend upon the cycle time of the server. [See the table “CF Link Speeds.”](#)

The Coupling Facility Control Code processes the request and responds to the requesting system. The capacity of the CF to process requests depends on the speed and number of its processors in addition to the application workload characteristics. It is generally recommended that CFs run below 50% processor utilization. The reason for this is twofold:

- 1) as CF utilization grows above 50% (on a single engine CF), service time increases due to CPU queuing may noticeably degrade impacting coupling efficiency.
- 2) if one CF fails, there will be insufficient capacity in the remaining CF(s) to handle the load.

A requesting processor can issue CF commands in one of two modes: synchronous or asynchronous. A CF command is said to be executing synchronously when the requesting processor spins in an idle microcode loop from the time the request is sent over the link to the time it receives a response from the Coupling Facility. On the other hand, during the processing of an asynchronous CF command, the requesting CP does not wait for a response from the CF, but is free to do other pending work. However, software costs are incurred in performing the task switch and in the processing of the asynchronous CF command completion. z/OS will pick the mode that provides the best coupling efficiency. This tends to limit the coupling cost to approximately 18%, depending upon the activity rate..Standalone CF and General Purpose Server

A coupling facility always runs as an LPAR. From the CF’s point of view, it does not matter what else is running on that server. From a performance viewpoint, the only difference is that if z/OS was running on another LPAR, then some links can be the very fast Internal Coupling (IC) links instead of external links.

Performance of Internal Coupling Facility (ICF)

The Internal Coupling Facility can be used as a shared or dedicated CF partition with external or internal coupling links. The performance of an ICF with external coupling links is similar to that of a CF partition on a server, with the addition of a small degradation due to the multiprocessor effects of “turning on” an additional processor (note that these MP effects are much less than those associated with adding another processor to run additional z/OS work).

Coupling Facility Configuration Options

Dedicated and Shared CPs

Dedicated processors in a logical partition will always provide the highest coupling efficiency and best CF response times. Having dedicated CPs allows the CF LPAR to be dispatched by PR/SM without waiting. This helps improve CF response times and coupling efficiency. CF partitions with shared processors will have a lower coupling efficiency and CF response times will be longer. If shared CPs are used, Dynamic CF Dispatch with Thin Interrupts (DYNDISP=THIN) is the only supported option.

For production workloads running in a data sharing environment, the coupling facilities should be configured with dedicated processors. Due to the heavy usage of the coupling facilities, especially for lock structures, dedicated processors are needed to maintain good CPU performance and transaction response times. Resource Sharing-only environments have significantly lower CF access rates than Data Sharing environments. Even though dedicated CPs are still recommended for this situation, many sites have successfully run with shared CPs. Here, it is necessary to first understand which CF structures will be used, their access rates, and plans for future growth and exploitation.

For System-Programmer Test workloads, use of shared CPs can be used. These workloads are generally very small and response time is not an issue.

The use of shared processors for coupling facilities for Development depends upon response time and CPU requirements for the development workload, and how much CF activity is being used.

The performance of an Internal Coupling Facility partition with dedicated processors in a server is similar to that of a standalone CF because a dedicated LPAR requires only a little LPAR dispatching overhead. Defining the CF processors as ICFs can provide a less expensive alternative to standalone CFs, plus the ICF processors do not affect the capacity rating of the box so there is no increase in the software licensing fees. As was the case with standalone CFs, the performance of CF partitions with shared processors will result in lower coupling efficiencies than CF partitions with dedicated processors depending on the degree of processor resource sharing. This is due to the fact that sometimes when the CF partition has work to do, there will be a delay before the LPAR can dispatch it (when all physical processors are already busy).

Coupling Facility Configuration Options

An additional impact to CF performance occurs when a CF partition shares CPs with z/OS images in the same sysplex. Since processors are shared between the z/OS partitions sending the synchronous CF commands, and the CF partition that is executing them, a deadlock situation might occur in the process; the z/OS partition and the CF partition may simultaneously require the same processor resource. Hence, it is necessary for LPAR to “under the covers” treat all CF commands as asynchronous to the processor when those CF commands stem from z/OS partitions sharing the processors with the CF partition. Note that to the operating system, the CF command still appears to be processed as it was issued, either synchronously or asynchronously, thus, there is no additional software path length involved. However, the CF service times will be degraded resulting in a small loss in coupling efficiency.

Dedicating ICFs or general purpose CPs to a CF partition is just like dedicating CPs to z/OS partitions. For performance reasons, it is recommended that datasharing production environments run with only dedicated CF partitions.

Please see the discussion on [Dynamic Coupling Facility Dispatch](#) and [Coupling Thin Interrupts](#) earlier in this document.

Coupling Facility Configuration Options

Performance of the Coupling Facilities

The cost of a CF access is comprised of software and hardware components. The hardware cost can be minimized by using the most efficient links with faster engine speeds for the CFs. This reduces the time that z/OS is waiting for the response while the message is on the CF link and while the CF is busy processing the request. With this in mind, it becomes obvious that the best coupling efficiency is generally seen with the newest model CF.

Number of CF Engines

When determining if the Coupling Facility should be configured with one or more CPs (cores), two issues arise. The first one is capacity. The recommendation is to provide enough capacity for either Coupling Facility to support all the structures in a failover situation while staying under 50% busy for optimal response times. One would then need enough engines to make this so.

For typical z/OS workloads on a 1-CP server, the recommendation is usually to keep the highest priority workload <70% of the server capacity. With z/OS and Workload Manager, the total server utilization can be up to 100% busy, but the top priority workload should stay under 70% busy. This is because after 70% the response time of the top workload starts to rise. However, CF accesses rates do not follow a traditional random distribution. There tends to be a strong clustering of work coming in. For example, when a Db2 transaction wants to read a page, it generates a lock request and a GBP request close in time. When the transaction ends, it commits data, writing to the GBP and issuing (un)lock requests. Again, clustering the CF activity. Castout processing is another example. So with CFs, the response times tend to go up once the utilization hits 50% busy.

Do I need a Two-engine Coupling Facility?

There are two issues, availability and performance. For availability, assuming that one engine is sufficient to provide capacity for the Parallel Sysplex cluster, should an additional engine be configured for availability? This is not needed. Transparent CPU Sparing is designed to enable the hardware to activate a spare core to replace a failed core completely automatically, while preserving the application that was running at the time of the error. If one of the running CPUs fails and instruction retry is unsuccessful, a transparent sparing operation is performed. Using the support element, Dynamic CPU Sparing copies state information from the failed CPU into the spare CPU. To ensure that the action is transparent to the operating system, special hardware tags the spare with the CPU address of the failed CPU. The spare is ready to begin executing at precisely the instruction where the other CPU failed. Sparing is executed completely by hardware, with no operating system awareness. Applications continue to run without noticeable interruption.

Coupling Facility Configuration Options

With transparent CPU Sparing, even if there were a hardware hit, a second engine would not provide significant benefits.

Configuring a 2-CP Coupling Facility may improve response time performance. In addition to handling the basic Lock / List / Cache commands, the Coupling Facility will at times scan directories during castout processing and perform other administrative tasks. Some of these can take some time. In a 1-CP Coupling Facility, any Lock/List/Cache requests would queue up behind the administrative work until the CP is free. This elongates the response times and can affect coupling efficiency. A 2-CP CF can help alleviate this.

Comparison of Coupling Efficiencies Across Coupling Technologies

As discussed earlier, the coupling efficiency of a Parallel Sysplex cluster, particularly one that has heavy datasharing is sensitive to the performance of the operations to the Coupling Facility. As the IBM technology has advanced, many improvements in coupling technology have also been made to ensure good coupling efficiencies will be preserved. The following table estimates the “host effect” for a heavy data sharing production workload for various combinations of host processor and coupling technology. The values in the table represent the percentage of host capacity that is used to process operations to the coupling facility. For example, a value of 10% would indicate that approximately 10% of the host capacity (or host MIPS) is consumed by the subsystem, operating system and hardware functions associated with coupling facility activity. The table is based on a “coupling intensity” of 9 CF operations per million instructions (MI). Workloads that have lower coupling intensities will have lower “effects” with correspondingly smaller differentials between the options.

The values in the table can be adjusted to reflect the coupling intensity for any workload. One can calculate the coupling intensity by simply summing the total req/sec of the CFs and dividing by the used MIPS of the attached systems (MIPS rating times CPU busy). Then, the values in the table would be linearly scaled. For example, if the workload was processing 4.5 CF operations per million instructions (or 4.5 CF ops/second/MIPS), then all the values in the table would be cut in half.

Coupling Facility Configuration Options

Host CF	z14 ZR1	z14	z15 T02	z15 T01	z16 A02	z16	17
z14 ZR1 CL5	18	18	18	18	18	18	
z14 ZR1 CS5	11	12	12	12	12	12	
z14 CL5	18	18	18	18	18	18	
z14 CS5	11	12	12	12	12	12	
z15 T02 CL5	18	18	18	18	18	18	
z15 T02 CS5	11	11	12	12	12	12	
z15 T01 CL5	18	18	18	18		18	
z15 T01 CS5	11	11	11	11	11	12	
z16 A02 CL5	18	18	18	18	18	18	
z16 A02 CS5	10	11	11	11	11	11	
z16 CL5	18	18	18	18	18	18	
z16 CS5	10	10	11	11	11	10	
z17 CL5			18	18	18	18	18
z17 CL6			n/a	n/a	n/a	n/a	18
z17 CS5			11%	11	10	10	10

Note 1: Assumes 9 CF requests / MI for production workload (9 CF operations / second per MIPS). If you are running less than 9 CF requests / MI, then the corresponding overhead would be proportionally less. For example, if you are running at 3 CF requests/MI (1/3 of “9”) then multiply all the numbers on the chart by 1/3.

Note 2: The table does not take into consideration any extended distance effects or system managed duplexing.

Note 3: For 9 CF requests/MI, host effect values in the table may be considered capped at approximately 18% due to z/OS 1.2 feature Synchronous to Asynchronous CF Message Conversion. Configurations where entries are approaching 18% will see more messages converted to asynchronous. As synchronous service times degrade relative to the speed of the host processor, the overhead % goes up. This could happen, for example, where the CF technology stays constant but you upgrade the host technology. This can be seen in the table by the % value increasing. z/OS converts synchronous messages to asynchronous messages when the synchronous service time relative to the speed of the host processor exceeds a breakeven threshold. At this point it is cheaper to go asynchronous. When all CF operations are asynchronous, the overhead will be about 18%. By the time you have reached $\geq 18\%$ in the table, z/OS would be converting almost every operation asynchronous.

Coupling Facility Configuration Options

Note 4: IC links scale with the speed of the host technology and would provide an 8 to 10% effect in each case.

The cap scales proportionally with the CF requests/MI activity rate. For example, If a configuration sees 4.5 CF requests/MI, then the cap would be at 9%.

To figure out how to scale the table, look at the RMF Subchannel Activity report, and identify the number of CF requests / second. In this case, there are about 41624 requests/second flowing from SYSA to CFP01.

```
COUPLING FACILITY NAME = CFP01
```

```
-----  
SUBCHANNEL ACTIVITY
```

```
-----  
# REQ
```

```
SYSTEM TOTAL -- CF LINKS -- PTH  
NAME  AVG/SEC TYPE GEN USE BUSY  
SYSA  37462K CBP  8 8  0  
41624 SUBCH 56 56
```

If SYSA was a System z14 at capacity setting 711 (rated at 16101 “LPAR MIPS” then this configuration has $41624/16101 = 2.59$ CF operations/MI. Since the numbers reflect 9 CF operations/MI, divide all the values by $(9/2.59) = 3.5$ to get a realistic estimate of the coupling overhead for each configuration option.

For example, If the z14 is connected to a z14 model ZR1 Coupling Facility, and you wish to estimate the effect of migrating from CS5 to CL5 links, the overhead would go from $11/3.5$ to $21/3.5$, or from approximately 3.1% to 6%.

IBM “MIPS” (Processor Capacity Index) values can be found at:

<https://www.ibm.com/support/pages/ibm-z-large-systems-performance-reference>

Note that the host effect is lower as the older generation servers are coupled to newer technology Coupling Facilities. Conversely, the host effect grows when newer generation servers are coupled to Coupling Facilities of the earlier technologies. The magnitude of change in host effect depends on the magnitude of change in the processor speed as well as the technology of the coupling link. Since the host effect is quite sensitive to the CF service time, the performance improvement from moving to new coupling technologies will be more dramatic for the higher speed host processors, particularly for

Coupling Facility Configuration Options

workloads with a high coupling intensity. This effect is clearly depicted in the table shown. Thus, it is of increasing importance to keep coupling technology reasonably current with the latest generation of processors. The table may be used to quantify the value of moving to a new coupling technology. However, it should be reiterated that the relative improvements between technologies are given for production workloads with heavy datasharing of 9 CF requests / MI. Workloads with lighter coupling intensities will have less relative difference between the coupling options. Customers desiring more Coupling Facility capacity can garner the processing power by upgrading the Coupling Facility model to a newer model or adding engines to existing models.

Since coupling efficiency is a function of the relative difference in processor speed between the server and CF, a good rule of thumb is to try to keep the CF on a comparable machine as the server while using the fastest link that can be configured. This can be done by keeping the Coupling Facility no more than one generation from the servers which they support

In addition to upgrading the CF server, coupling efficiency can also be improved by migrating to faster links if the configuration allows it.

Cost/Value Implications of Each CF Option

The Coupling Facility (CF) is at the heart of Parallel Sysplex technology. There are several options on how to configure the CF. They include:

- Standalone Coupling Facility
- Coupling Facility LPAR on a Server
- Internal Coupling Facility (ICF)
- System-managed CF Structure Duplexing

Standalone Coupling Facilities

When Parallel Sysplex clustering was first announced in April, 1994, the only option was to place the CFs on two standalone servers. As different options became available, the standalone CF still provided the configuration for the best availability. Over time, however, some of the differentiations are no longer as significant.

When ICFs became available, availability benefits of standalone CFs included:

1. **Microcode Upgrades:** Originally, a CF level upgrade required a POR of the entire server, affecting any z/OS LPARs that may be on that server. However, today Concurrent Patch Apply allows microcode patches to be dynamically installed on the coupling facility. To pick up a new CF level, the CF only needs to be “recycled” after dynamically moving structures to the other CF in the sysplex. Whereas in the past this was a significant differentiator, today it no longer is.

2. **Independent server maintenance and upgrades:** Today's servers are designed to support dynamic microcode upgrades and even version upgrades within the family. This is designed to be done in the middle of first shift without affecting production workload. Planning for the unexpected, many chose to perform microcode maintenance and upgrades off shift. However, many may choose to also (manually) move the CF on that server has its structures moved to the other CF in the Parallel Sysplex cluster. A standalone CF would not be affected by z/OS server upgrades. While the process of moving structures off a CF is not **automatic**, it can be **automated**. As confidence is built over time through experience that the dynamic microcode upgrades do not affect the stability of the server, and that structures can get dynamically rebuilt anyway, the process of moving structures should not be an issue.
3. **Standalone CF server upgrades:** A standalone CF server can be upgraded from one family to another without affecting the z/OS LPARs (if structures are first moved to the other CF). While this is a true statement, one should balance this with the fact that a server upgrade to support additional z/OS capacity and functionality that also contains ICFs would have the ICFs upgraded without an additional charge for the ICF upgrade if the ICF was already installed.
4. **Number of CF links:** In a channel constrained server, it is difficult to add another LPAR to hold a CF which itself requires channels for CF links; at least two CF links to each other external server running in the Parallel Sysplex cluster. Multiple Logical Channel Subsystems reduced this constraint.
5. **Failure Isolation:** If an IRLM Lock structure, for example, is on the same server as one of its connectors, and the server fails, then the structure is lost with an IRLM. Any IRLM on another server does not have enough information to be able to re-create the lost structure and proceeds to bring itself down to avoid data integrity problems. This brings down the Db2s and requires a group-wide restart. To avoid this situation, it is recommended that the IRLM structure and any others requiring failure isolation be situated on a CF not running z/OS in the Parallel Sysplex cluster.

This problem is no longer applicable with System-Managed CF Structure Duplexing of the structure. Although duplexing increases coupling overhead, much of this overhead was reduced with asynchronous system managed duplexing for lock structures. More information on duplexing and failure isolation can be found later in this document. An analysis of the performance impact of duplexing can be found in White Paper ZSW01975-USEN available on the Parallel Sysplex Web site: <https://www.ibm.com/products/zos/parallel-sysplex>

Coupling Facility Configuration Options

After analyzing the cost of System-Managed CF Duplexing, many data sharing customers conclude that better price-performance can be obtained for their environment by having an external CF to hold structures requiring failure isolation. If the standalone CF should have a problem, the structures can be rebuilt on to the other CF running as an ICF. z/OS LPARs usually are on that server as well. If there is a double failure and the server running the ICF and z/OS fail as well, then group-wide restart of Db2 will be required. To avoid this possible problem of a double failure, some sites chose to configure two standalone CFs. Again, asynchronous duplexing for Db2 lock structures significantly helps.

There is an increasing number of multi-site Parallel Sysplexes. A Parallel Sysplex environment that spans two sites up to 100 km apart from each other together with disk remote copy is a basis of disaster recovery solutions such as GDPS/Metro. If the site with the primary disk (Site 1) fails, then the entire workload is shifted to the site with the secondary disk (Site 2). To avoid failure isolation for either site as well as performance, standalone CFs are typically configured for both locations.

A balance of availability and performance can be obtained by configuring a single standalone CF containing structures that require failure isolation, with an ICF that contains structures not requiring failure isolation. System-Managed CF Structure Duplexing can then be enabled for those structures where the benefit outweighs the cost. This can typically be for those structures that have no fast method of recovery such as CICS named counters, CICS Temporary Storage, or WebSphere MQ non-persistent queues. The WebSphere MQ administrative structures should also be duplexed.

Coupling Facility Configuration Options

System-managed CF Structure Duplexing

System-managed CF Duplexing allows any enabled structure to be duplexed across multiple Coupling Facilities and maintains that duplexed copy during normal use of the structure by transparently replicating all updates to the structure in both copies. This provides a robust failure recovery capability through failover to the unaffected structure instance.

System-Managed CF Structure Duplexing provides:

- An easily exploited common framework for duplexing the structure data contained in any type of CF structure, with installation control over which structures are and are not duplexed.
- High availability in failure scenarios by providing a rapid failover to the unaffected structure instance of the duplexed pair with very little disruption to the ongoing execution of work by the exploiter and application.

With this support, many of the failure isolation issues surrounding data sharing structures are no longer relevant as there is no single point of failure and a standalone CF is no longer required. If, for example, an IRLM lock structure is duplexed between two ICFs on different servers, then a failure of a server containing one of the ICFs together with one of the IRLM instances does not stop the surviving IRLM(s) from still accessing the duplexed structure on the other ICF.

There is a performance cost to System-managed CF Structure Duplexing, so the benefits of being able to run with only ICFs have to be weighed against the added coupling overhead. More details about this can be found in White Paper System-managed CF Duplexing, ZSW01975USEN. It is linked off the Parallel Sysplex web page:

<https://www.ibm.com/products/zos/parallel-sysplex>

Asynchronous CF Structure Duplexing for Db2 Lock Structures

Asynchronous CF Structure Duplexing is designed to:

- Improve performance with cross-site duplexing of lock structures at distance or in same site.
- Maintain robust failure recovery capability through the redundancy of duplexing, even for non-standalone CFs.
- Reduce z/OS, CF, and link utilization overhead associated with synchronous duplexing of lock structures

The systems tries to recover quickly if a structure has a problem so service is not interrupted. Since data managers such as IRLM serialize on consistency of data, rapid recovery of a lock structure is important. Many years ago we introduced System Managed Duplexing that allows multiple copies of the same lock structure in two different CFs. The system failed over to either copy if one copy failed. This is very good for availability, but because of protocol exchanges, it affects coupling overhead.

Coupling Facility Configuration Options

With Asynchronous CF duplexing for Db2 lock structures, secondary CF lock structure updates are performed asynchronously with respect to primary updates in order to drive out cross-CF latencies due to “handshaking” protocols. This is intended to avoid the need for synchronous communication delays during the processing of every duplexed update operation while keeping the same redundancy of lock structure data and same recovery characteristics.

There is more performance benefit with distance since the cost of synchronous handshaking is greater with distance. All distance latencies associated with the round trip to and from the primary lock structure are not avoided by this support. That is unavoidable distance latency. This support is designed to avoid the additional latencies associated with the duplexing protocol itself.

Asynchronous CF Duplexing can provide service time similar to in a simplex environment. Although z/OSs need access to both the primary and secondary structures, normal communication is only to the primary structure. The primary CF lock structure is updated synchronously with respect to the transaction. The CF receives this update and acknowledges it, and the transaction continues, just as it would in simplex mode. Just like in simplex mode, the application (IRLM) does not talk directly to the secondary lock structure. After the primary structure is updated, while the transaction is continuing running, the CF sends this update to the secondary lock structure. The secondary structure will lagging behind. When a transaction does a commit, then the primary structure instance synchronizes it updates with the secondary. All committed transactions' updates are in the secondary, and the only locks that are not in the secondary already are for uncommitted transactions.

Given this configuration:

CECA	CECB
z/OSA	z/OSB
 Db2A	 Db2B
IRLMA	IRLMB
 ICFA	 ICFB
- LOCK (Prim)	- LOCK (Sec)

If ICFA holding the primary lock structure fails, then the Db2s and z/OSs talk to each other to identify the locks for in-flight units of work that were not updated at the secondary structure, and the secondary structure is updated. This can be done very quickly.

Coupling Facility Configuration Options

If CECA fails, then zOSA also fails. And all of its locks and changes for uncommitted transactions running on zOSA get backed out when Db2A restarts. There is no need to reconstruct them into the former secondary (now only) structure on ICFB.

If ICFB or CECB fails, the system just falls back to using the primary lock structure in ICFA which is fully up to date.

Asynchronous CF structure duplexing requires CFCC Level 21 with service level 02.16, CF to CF connectivity via coupling links, and exploitation, and PTF support for z/OS V2.2, Db2 V12, IRLM V2.3, and optionally z/VM 6.4 for guest support.

Recommendation: Always run with Asynchronous CF Duplexing for Db2 (IRLM) lock structures, even with not otherwise duplexing structures. This can only help response time and performance.

Heuristic CF Access Synchronous to Asynchronous Conversion: When a z/OS image sends a command to a Coupling Facility, it may execute either synchronously or asynchronously with respect to the z/OS task which initiated it. Often, synchronous execution of CF commands provides optimal performance and responsiveness by avoiding the back-end latencies and overhead that are associated with asynchronous execution of CF requests. These may include creating an SRB to send the CF request, dispatching the SRB, managing the I/O interrupt when the response comes back, re-dispatching the SRB, then redispatching the original task. However, under some conditions, it may actually be more efficient for the system to process CF requests asynchronously. Examples include:

- If the CF to which the commands are being sent is more than a few kilometers away (e.g. in a multi-site Parallel Sysplex cluster),
- If commands are being sent to a duplexed structure,
- If the CF is running on slower server technology than the z/OS server,
- If the CF is configured with shared processors, dynamic CF dispatching, or CFs on low weighted LPARs using shared processors,
- For some long running CF requests such as Db2 32K GroupBufferPool requests,
- If the responses are getting delayed due to overloaded CFs, overloaded links, etc.
- Anything else that may cause the service times to be long either of a permanent or transient nature may be converted.

z/OS can dynamically determine if starting an asynchronous CF request instead of a synchronous CF request would provide better overall performance and thus self-optimize the CF accesses. The system will

Coupling Facility Configuration Options

observe the actual performance of synchronous CF requests. Based on this real-time information, z/OS will heuristically determine how best to initiate new requests to the CF.

Sync/async optimizations are granular based on the type of CF request and the amount of data transfer on the request, so the algorithm factors in these considerations as well.

Structure Placement

The first half of this document provided a detailed look at coupling alternatives and their common characteristics. For example, all Coupling Facilities run in a Logical Partition (LPAR). This LPAR requires either shared or dedicated processors, storage, and high-speed channels known as coupling links. For test configurations, CF LPARs with Dynamic CF Dispatching can be used. Alternatively, ICF LPARs with Internal Coupling (IC) links provide the added flexibility of supporting multisystem configurations but without external links that need to be configured.

The standalone Coupling Facility was defined as the most robust production Coupling Facility configuration option, capable of delivering the highest levels of performance, connectivity and availability. Several Coupling Facility options exist on the servers. The Internal Coupling Facility (ICF) is generally the most attractive production CF option, given its superior price/performance. The ICF is comparable in performance to a standalone Coupling Facility. With proper configuration planning the ICF can support a highly available Parallel Sysplex environment, either used in combination with a standalone CF or used exclusively for “suitable” configurations or when used in conjunction with CF Duplexing. But what constitutes a “suitable” environment for exclusive use of ICFs? When is a standalone Coupling Facility required for availability reasons? Are there any production scenarios where a single CF is acceptable?

To answer these questions, it is important to have a clear understanding of two distinct Parallel Sysplex configuration environments: Resource Sharing and Data Sharing. These two Parallel Sysplex environments require different degrees of availability. Understanding what these terms mean and matching a customer’s business objectives to these basic environments will help simplify the Coupling Facility configuration process.

Following is an explanation of each environment, its corresponding benefits, and availability requirements.

Resource Sharing Environment

The Systems Enabled Parallel Sysplex environment represents customers that have established a functional Parallel Sysplex environment, but are not currently in production datasharing with IMS, Db2, or VSAM/RLS. Systems Enabled is also referred to as Resource Sharing. This environment is characterized as

Coupling Facility Configuration Options

one in which the Coupling Facility technology is exploited by *key z/OS components*. In the Parallel Sysplex Resource Sharing environment, CF exploitation can help deliver substantial benefits in the areas of simplified systems management, improved performance and increased scalability, and help eliminate redundant hardware.

Perhaps the Systems Enabled Resource Sharing environment can be best understood through consideration of some examples.

First, simplified systems management can be achieved by using XCF Signaling Structures in the Coupling Facility versus configuration of dedicated Channel to Channel (CTC) connections between every pair of systems in the Parallel Sysplex cluster. Anyone that has defined ESCON® CTC Links between more than two systems can certainly appreciate the simplicity that XCF Signaling Structures can provide. Further, with the coupling technology improvements delivered with faster engines and CF links since the G5 servers, XCF communication through the CF could outperform CTC communication as well. More detail in XCF Performance can be found in the Washington System Center flash:

<ftp://ftp.software.ibm.com/software/mktsupport/techdocs/xcfflsh.pdf>

Improved performance and scalability is delivered with enablement of GRS Star. With GRS Star, the traditional ring-mode protocol for enqueue (ENQ) propagation is replaced with a star topology where the CF becomes the hub. By using the CF, ENQ service times are now measured in microseconds versus milliseconds. This ENQ response time improvement can translate to a measurable improvement in overall Parallel Sysplex performance. For example, a number of customers have reported significant reduction in batch window elapsed time attributable directly to execution in GRS Star mode in the Parallel Sysplex cluster. The GRS structure can also be used to dynamically assign tape drives across nodes in a Parallel Sysplex configuration. With this function enabled, tape drives can be reassigned on demand between LPARs without operator intervention.

RACF® is another example of an exploiter that can provide improved performance and scalability. RACF is capable of using the Coupling Facility to read and register interest as the RACF database is referenced. When the Coupling Facility is not used, updates to the RACF database will result in discarding the entire database cache working set that RACF has built in common storage within each system. If an installation enables RACF to use the Coupling Facility, RACF can selectively invalidate changed entries in the database cache working set(s) thus improving efficiency. Further, RACF will locally cache the results of certain command operations. When administrative changes occur, such commands need to be executed on each individual system. The need for this procedure can be eliminated in an XCF environment by leveraging command propagation. Refreshing each participating system's copy of the RACF database is still needed. However, this too is simplified when RACF operates within a Parallel Sysplex cluster by

Coupling Facility Configuration Options

leveraging command propagation. With RACF, you can take this whole process one step further by caching the RACF database in the Coupling Facility. RACF has implemented a store-through cache model. As a result, subsequent read I/Os of changed RACF pages can be satisfied from the Coupling Facility, eliminating the DASD I/O, providing high-speed access to security profiles across systems.

Other Resource Sharing CF exploitation examples include VTAM® Generic Resource for TSO session balancing, JES2 Checkpoint sharing for scalability, System Logger single-system-image support of sysplex-merged Logrec and Operlog, SmartBatch for improved batch performance, Shared Catalog, and many others.

It is easy to migrate to a Parallel Sysplex resource sharing environment. Generally, resource sharing exploiters are components shipped with z/OS. This avoids the need to possibly plan for data sharing exploitation concurrently on IMS, Db2, and CICS datasharing subsystems. Further, planning considerations associated with shared database performance tuning, handling of transaction affinities, etc., are avoided. Resource Sharing has high value, either as a straightforward migration step toward the full benefits of Application Enabled Data Sharing **or** as an end point Parallel Sysplex configuration goal.

Structure availability requirements differ significantly between the Resource Sharing and Data Sharing environments in terms of CF structure placement. For the installation that plans to only perform Resource Sharing, CF failure independence is generally not required.

Failure independence describes the need to isolate the CF LPAR from exploiting z/OS system images in order to avoid a single point of failure. This means that the CF and z/OS LPARs should **not** reside on the same physical server when failure independence is needed. Stated another way, ***failure independence is required if concurrent failure of the CF and one or more exploiting z/OS system images prevents the CF structure contents from being recovered, or causes a lengthy outage due to recovery from logs.***

Failure independence is a function of a given z/OS to CF relationship. For example, all connectors to a structure on a standalone CF are failure independent. However, with an ICF, all connections from z/OS images on the same footprint are failure dependent.

Note: The term *failure independence* and *failure isolation* are used interchangeably in Parallel Sysplex documentation. Do not to confuse the term “failure isolation” with “system isolation,” which occurs during sysplex partitioning recovery. More guidelines on failure-independence can be found in “Value of Resource Sharing,” GF22-5115.

Coupling Facility Configuration Options

Resource Sharing: CF Structure Placement

For Resource Sharing environments, servers hosting Internal Coupling Facility features along with exploiting z/OS images are acceptable. For example, if the CF is exploited by System Logger for Operlog or Logrec, and the CF should fail, no attempt at recovery of the lost data would be made. Hence, if the machine hosting both the CF and an exploiting z/OS system image should fail, no failure independence implications result. Another example is GRS operating in a Star configuration. GRS Star demands two Coupling Facilities. These Coupling Facilities could be Internal Coupling Facilities on two separate servers within the Parallel Sysplex cluster. A configuration such as this will not compromise GRS integrity. This is because in the event a z/OS image terminates, ENQs held by that image are discarded by GRS. So if both one of the CFs and a z/OS image should fail, CF Structure rebuild into the alternate CF would not involve recovery of the ENQs held by the system that died. The surviving GRS images would each repopulate the alternate CF structure with ENQs they held at the time of failure. Thus, failure independence is not required.

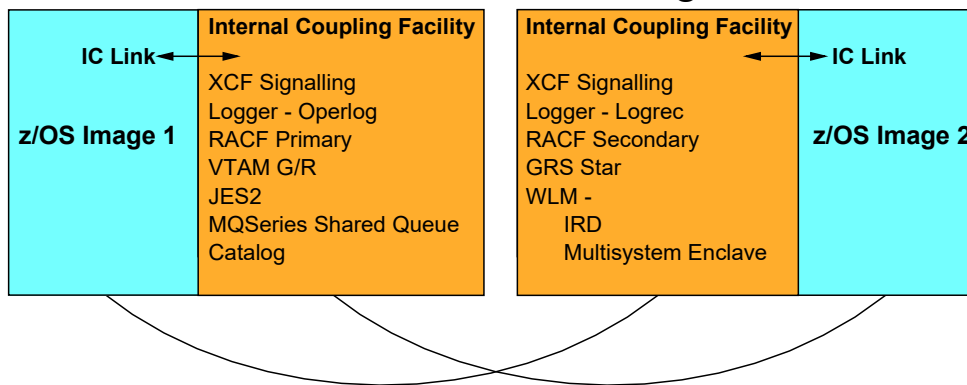
The above scenarios similarly apply to the other Resource Sharing environment CF exploiters. Thus, this environment does not demand configuration of a standalone CF (or more than one).

In the following diagram, the loss of a single server that hosts an Internal Coupling Facility will not result in a sysplex-wide outage. Actual recovery procedures are component dependent and are covered in great detail in Parallel Sysplex Test and Recovery Planning (GG66-3270). In general, it is safe to say that a production Resource Sharing environment configured with two Internal Coupling Facilities can provide cost effective high availability and systems integrity.

For Resource Sharing we stated that the GRS Star structure does not require failure independence. This is true. It does not mean, however, that there is no benefit from having at least one failure independent CF. Recovery for a failed CF will be faster if it is not first necessary to detect the failure of an exploiting z/OS image and perform Parallel Sysplex partitioning actions before the CF structure rebuild process can complete. For Resource Sharing exploiters other than GRS, the impact of a longer CF recovery time is not significant relative to the value gained in terms of reduced cost and complexity through leverage of ICFs. For GRS however, the placement of the GRS Star structure warrants consideration of the cost/complexity savings versus recovery time in a Data Sharing Environment.

Structure Placement

Resource Sharing



Data Sharing Environments

The data sharing Parallel Sysplex environment represents those customers that have established a functional Parallel Sysplex cluster and have implemented IMS, Db2 or VSAM/RLS data sharing. When data sharing enabled, the full benefits of the Parallel Sysplex technology are afforded, including dynamic workload balancing across systems with high performance, improved availability for both planned and unplanned outages, scalable workload growth both horizontally and vertically, etc. In this environment, customers are able to take full advantage of the ability to view their multiple-system environment as a single logical resource space able to dynamically assign physical system resources to meet workload goals across the Parallel Sysplex environment. These benefits are the cornerstone of the IBM Z Parallel Sysplex architecture.

Data Sharing: CF Structure Placement

The data sharing enabled environment has more stringent availability characteristics than the Resource Sharing environment. The data sharing configurations exploit the CF cache structures and lock structures in ways which involve sophisticated recovery mechanisms, and some CF structures do require failure independence to minimize recovery times and outage impact.

Two Coupling Facilities are required to support this environment. Without System-managed CF Structure Duplexing implemented for the data sharing structures, it is strongly recommended that this Coupling Facility configuration include at least one standalone CF. The second CF can run on a server as an Internal

Coupling Facility Configuration Options

Coupling Facility. The standalone CF will be used to contain the CF structures requiring Failure Independence. The ICF(s) will support those structures not requiring failure independence as well as serving as a backup in case of failure of the standalone CF. This should be viewed as a robust configuration for application enabled data sharing.

Two standalone CFs provide the highest degree of availability since there is always the possibility that an ICF or server might fail while the primary (standalone) CF is unavailable. This provides protection against a double failure.

Coupling Facility Configuration Options

This type of protection is also available with a logical standalone coupling facility, when the ICF is configured on a server that does not contain a connector to a structure that does not require failure isolation, or a z/OS instance that is exploiting the ICF on that server. The advantage of this is cost, in that an ICF can be configured instead of requiring a standalone coupling facility, while still maintaining a highly available Parallel Sysplex environment.

See section: [Standalone Coupling Facility](#) for more information.

With System-managed CF Structure Duplexing, two images of the structures exist; one on each of the two CFs. This eliminates the single point of failure when a data sharing structure is on the same server as one of its connectors. In this configuration, having two ICFs provides a high availability solution. A cost / benefit analysis of System-Managed Duplexing needs to be done for each exploiter to determine the value in each environment. More information can be found in System-managed CF Duplexing, ZSW01975USEN available off of the Parallel Sysplex web site at <https://www.ibm.com/products/zos/parallel-sysplex>

Lock Structures

Accounts that are in production data sharing must worry about preserving the contents of their **Lock structures** used by the lock managers such as IRLM or SMSVSAM. Should these structures be lost along with one or more of the z/OS images that were using them, the installation will be forced to recover from logs. This is a time consuming process that will vary depending on the specific customer environment. Recovery of the lock structures from logs can be avoided if the CF structure is isolated from exploiting lock managers, allowing rebuild of the lock structure contents from in-memory lock information on each of the connected systems. This can also be achieved using CF Duplexing for the relevant lock structures. See the section above discussing Asynchronous CF Duplexing for lock structures.

Although GRS Star uses a lock structure, it does not require failure independence.

Cache Structures

Cache structure recovery for the data sharing environments differs greatly depending on whether the cache structure exploiters support CF structure rebuild or not. Both IMS and VSAM/RLS support CF structure rebuild for their caches, and, given their caches do **not** contain changed data, can recover the loss of a cache structure without requiring recovery from logs. Failure of a data sharing IMS or VSAM/RLS system instance does not affect the ability to recover the cache structure contents and therefore, CF failure independence is not demanded for these cache structures.

Db2 Group Buffer Pools and IMS VSO DEDB caches contain changed data and do not support CF structure rebuild in the event of a CF failure. The IMS VSO DEDB cache support in IMS Version 6 supports multiple

Coupling Facility Configuration Options

copies of a data area. Each area has its own copy of the structure. This avoids a single point of failure. As such, it does not **demand** failure-isolation between the CF and any sharing IMS instance.

Db2 GBPs support user-managed CF structure duplexing. As Db2 writes buffer updates to a primary GBP, it simultaneously writes to the secondary, duplexed GBP. Registering interest and Reads are only done to the primary structure. Since both GBP instances are kept in synch with each other in terms of changed data, there are no concerns about the loss of the primary GBP as the data will still be available in the alternate. With Db2 GBP duplexing, there is no concern about using an ICF to hold the production data.

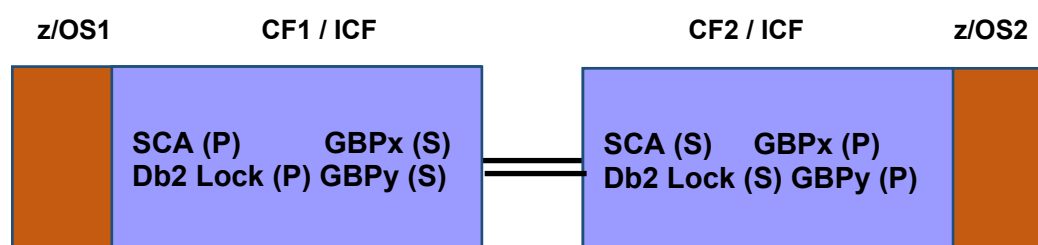
As exploiters take advantage of System-Managed CF Structure Duplexing, the possibility exists to eliminate the requirement for a standalone CF. If, for example, the Db2 SCA and Lock structures were duplexed, then with the primary and secondary structures on different ICFs, there are no availability issues. A cost / benefit analysis of System-Managed Duplexing needs to be done for each exploiter to determine the value in each environment. More information on System-managed CF Structure Duplexing can be found in ZSW01975USEN.

Transaction Log Structures

Several subsystems/components exploit the System Logger to provide logging for transactions and media recovery. These CF structures warrant failure independence to avoid a single point of failure for the CF log data. The System Logger component is able to rebuild these structures from local processor storage if the CF fails. But if the CF and a System Logger instance should fail, recovery is not possible, unless System Logger Staging Data Sets are in use — which impacts performance as it requires log records to be forced to disk before control can be returned for every log write request. Use of CF Duplexing can provide this failure-isolation requirement and can also allow the use of staging data sets with their accompanying performance implications to be avoided.

Coupling Facility Configuration Options

Subsystems/components which exploit the System Logger and warrant failure independence include: z/OS Resource Recovery Services, IMS Common Queue Server component for Shared Message Queues, and CICS system and forward recovery logs.



Other CF Structures Requiring Failure Independence

A number of other list structure exploiters provide CF failure recovery via CF rebuild from local processor memory contents on each connector. Without System-Managed duplexing, failure independence will require full reconstruction of the CF structure contents. A concurrent z/OS system image failure containing an exploiter of these structures would preclude successful CF structure rebuild. CF structures in this category include:

- VTAM Multi-Node Persistent Sessions
- VTAM Generic Resource structures for LU 6.2 Sync Level 2 sessions
- Db2 SCA
- IMS Resource Structure Added in IMS V8 to keep track of IMS resources and their status in an IMSplex environment.)

Other CF Structures NOT Requiring Failure Independence

There are a set of CF structures in support of data sharing which do not require failure independence, either because recovery does not need failure independence to recover the CF contents, or because recovery for CF structure loss is not possible in any event. CF structures in this category include:

- IMS Shared Message Queues
- VTAM Generic Resources for non-LU6.2 sync level 2
- CICS Shared Temp Store Queues

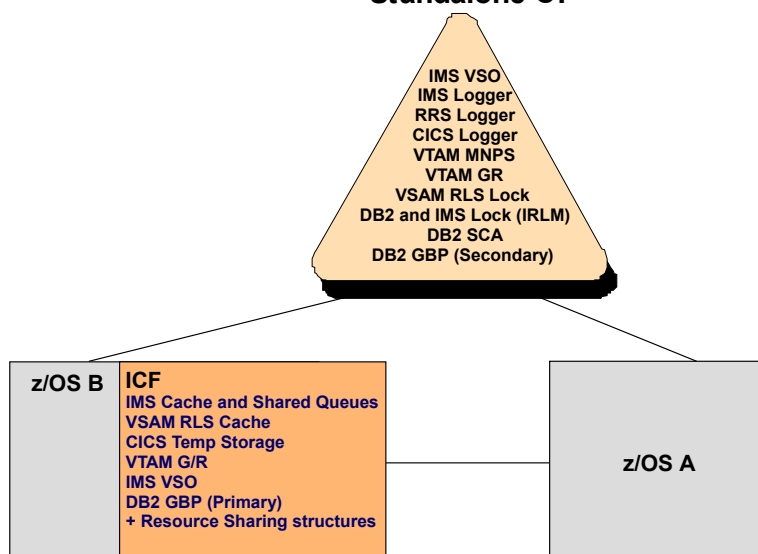
Coupling Facility Configuration Options

As one can see from the above discussion, a production datasharing environment has much more stringent availability requirements than a production Resource Sharing environment. Alternatively, these structures support System-Managed CF Structure Duplexing, providing full high availability support for these structures as well.

This means that careful thought must be given regarding structure placement in the event of a system image failure. A properly configured datasharing environment consisting of a single standalone Coupling Facility and one Internal Coupling Facility would have structures placed in the following way.

In the picture below, structures that do not require failure independence are placed in the ICF, while structures that do require failure independence are placed in the standalone CF.

Structure Placement - Data Sharing Standalone CF



Failure Independence with System-managed CF Structure Duplexing

For information on System-Managed CF Structure Duplexing, please see System-Managed CF Structure Duplexing (ZSW01975USEN) available on the Parallel Sysplex web site

www.ibm.com/products/zos/parallel-sysplex

Combining an ICF with Standalone CF

Many customers choose to take advantage of the availability provided by the standalone CF with the price benefits of an ICF. This is done by configuring two Coupling Facilities, one as a standalone and one as an ICF. The only caveat is that without duplexing, any application data sharing structures requiring failure-isolation should reside on the standalone model for highest availability. For example, this would include the Db2 SCA and IRLM's lock structure. Any duplexed structures (such as Db2 Group Buffer Pools or by using System-Managed CF Structure Duplexing) actually reside on both CFs, so there is no single point of failure. Any Resource Sharing structures could reside on the ICF without loss of availability.

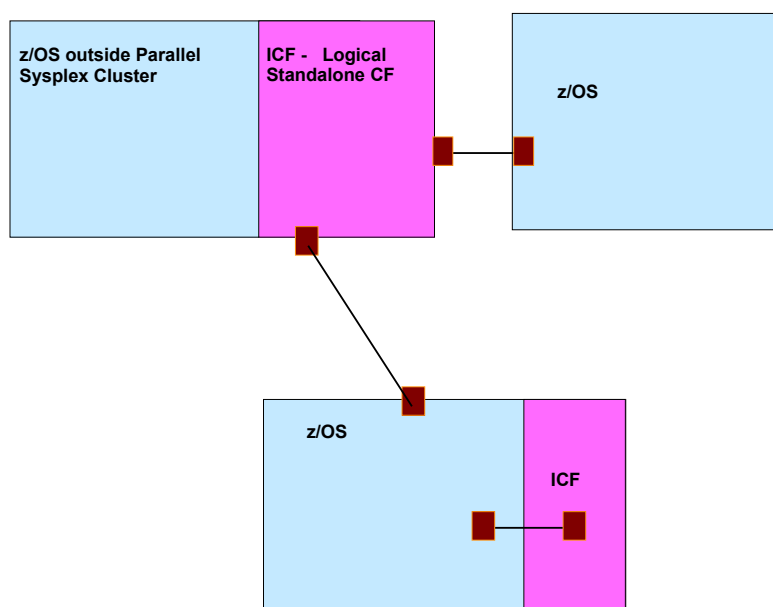
Coupling Facility Configuration Options

An example of how this may be configured is as follows:

Standalone CF	ICF
XCF Path 1	XCF Path 2
Db2 SCA	RACF®
IRLM Lock	Logger (OperLog)
Db2 GBPO (P)	Db2 GBP0 (S)
Db2 GBPI (S)	Db2 GBP1 (P)
Db2 GBP2 (P)	Db2 GBP2 (S)
	GRS Lock
	ECS (Catalog)
	JES2 Checkpoint

It should be noted that if an ICF is configured on a server whose logical partitions are not part of the Parallel Sysplex cluster that the ICF belongs to, then the ICF acts as a “logical” standalone Coupling Facility.

An example of this is shown in the figure below:



Coupling Facility Configuration Options

Other Configuration Considerations

CFCC concurrent patch apply:

CFCC code can be delivered as a new **release level** or as a **service level** upgrade, within a particular release level. Typically, a new **release level** will be delivered as part of an overall system level driver upgrade and will require a reactivate of the CFCC partition in order to utilize the new code. **Service level** upgrades are delivered as LIC, and are generally concurrent to apply and could be done non-disruptively while the CF is running. When applying concurrent CFCC LIC, the code will immediately get activated on all of the Coupling Facility images that are defined on the server.

To support migration from one CFCC release level to the next or the occasional disruptive patch, you have the ability to run different levels of the Coupling Facility code concurrently in different coupling facility partitions on the same server.

For example, if “CFProd” and “CFTest” both reside on the same server, once the new hardware level code is installed, the structures on the CFTest could be moved to a backup Coupling Facility in the Parallel Sysplex. Recycle the CFTest image. As the CF is restarted it picks up the new hardware code level. Move the structures back again once the new code has been activated. Once you are comfortable with this level, you can then recycle CFProd. This process significantly enhances the overall Sysplex availability characteristics of disruptive CFCC LIC by avoiding z/OS impact while allowing testing of the new hardware code level.

When migrating CF release levels, the lock, list, and cache structure sizes increase to support new functions. This adjustment can have an impact when the system allocates structures or copies structures from one coupling facility to another at different CFCC levels. For any CFCC release level upgrade, you should always run the CFSIZER tool that takes into account the amount of space needed for the current CFCC levels. The CFSIZER tool is available at <http://www.ibm.com/systems/support/z/cfsizer/>. Also consider using the SIZER utility at <http://www.ibm.com/systems/support/z/cfsizer/altsize.html>.

To support migration of new release or service levels that are marked as **disruptive**, you have the option to selectively activate the new LIC to one or more Coupling Facility images running on the server, while still running with previous level active on other Coupling Facility images.

The CF only needs to be recycled to pick up the new microcode after it is downloaded to the server. This provides the ability to:

- Selectively apply the new LIC to one of possibly several CFs running on the serve while still running with the previous level active on another Coupling Facility image. For example, if there is a CF that

Coupling Facility Configuration Options

supports a test Parallel Sysplex and a CF that supports a production Parallel Sysplex on the same server, you can install the new LIC to the server, but may only choose to apply this to the test CF without affecting the production CF. Once you are confident with the new code, you can then selectively apply it to other CF images on the server.

- Allow all other LPARs on the server where a CFCC patch will be applied to continue to run without being impacted by the application of the “disruptive” CFCC patch.

A given hardware driver level contains the driver for only a single CF level. While you can continue to run a CF at a given release level across non-disruptive hardware driver updates, as soon as you recycle that CF it will pick up the new code.

Single Server Parallel Sysplex:

Some companies that have a single server have seen the value of a Parallel Sysplex environment for z/OS availability. Sometimes called a “Sysplex-in-a-box,” by performing rolling IPLs they can keep the applications available across software upgrades. Although this does not provide for additional hardware redundancy, the hardware mean time to failure is so good that this is not perceived as a risk. A single server Parallel Sysplex can be configured with two z/OS images with data sharing connecting to one or two CFs using IC links.

There are several other use cases for a single server Parallel Sysplex. They include:

- A large corporation sometimes has an I/T department that provides computer services for the subsidiaries. At times they define the model architecture that are thoroughly tested and that the subsidiaries need to follow. This can include having a Parallel Sysplex with data sharing. Some subsidiaries may be large and implement this environment with two or more servers. Other subsidiaries however may not need as much computing power and chose to implement it on a single server.
- In a disaster recovery (D/R) environment, the primary, production site can be running in a normal multi-server Parallel Sysplex. The secondary, backup site has to run in the same *LOGICAL*, but not *PHYSICAL* configuration. For example, if Db2 was defined as data sharing in the production site, it has to restart in data sharing mode in the recovery site. But it can share the same server (or even z/OS image) with its data sharing partner. One of the Db2s will be started with “Restart Light” to come up, back out any in-flight units of work to release locks, then shut itself down to free up storage.
- A GDPS/Continuous Availability solution requires a Parallel Sysplex at both sites. Having a single server Parallel Sysplex is a way of reducing the amount of footprints while relying on the very rapid recovery capabilities of GDPS/Continuous Availability if there are any hardware related events

Coupling Facility Configuration Options

- It is IBM's recommendation that if the test environment reflect the production environment while being isolated from it as much as possible. If there is a Parallel Sysplex running in production, there should be a separate Parallel Sysplex in the test and development environments. These can be on a single server.

Single ICF or Two ICFs:

Since the CF(s) are on the same server as the z/OS image, if there is a major hardware event, everything on the server would be affected. In this specific case, the number of CFs would not make a difference. In addition, all links can be defined as ICF links that do not require any hardware, so again, no difference. There are however other things that should be considered

A single CF can help save money. IBM always recommends having dedicated ICF engines when doing production data sharing. Having a single CF would require only a single ICF engine to be purchased. Although the Dynamic CF Dispatching with Thin Interrupts enabled helps simplify the planning of this, sharing ICF engines is still not recommended for production environment. In addition, having a second CF increases the amount of memory needed on the server. Despite that, IBM does not recommend a CF.

Having two CFs will improve availability. There are some types of CF failures that can affect an entire CF image. Having a second CF allows the structures to be rebuilt while maintaining z/OS availability.

Recommendation: IBM recommends always having at least two CF images available at all times.

The IBM Z servers support dynamic code upgrades. This can sometimes include a CF upgrade. Having two CFs allows one to move all structures off of CF1 and recycle it. CF1 is now at the new level and CF2 continues to run at the n-1 level. Moving structures on to CF1 allows testing of the new CF code while keeping CF2 as a backup if there are issues. If there are no issues, then the structures from CF2 can be moved to CF1 and be recycled, bringing both CFs to the current level.

Recommendation: Configure two CFs, even with a one-server Parallel Sysplex

Single z/OS Parallel Sysplex:

There are some customers that have a single production z/OS instance, and they are currently satisfied with the overall availability and other characteristics of that single instance. However, they would like to implement certain z/OS services that require a Coupling Facility. An excellent example is VSAM Record Level Sharing (RLS) and/or Transactional VSAM (DFSMSStvs), technologies that are particularly helpful in supporting concurrent batch program and online transaction application access to common VSAM files. Many customers want to avoid taking their transaction applications offline when they run their overnight batch. They often want to continue providing transactions processing during nighttime and weekend hours because of the Internet and mobile user demands. The ability to run batch concurrently with online work can greatly enhance service availability, even with a single z/OS instance.

Coupling Facility Configuration Options

Other Considerations and Upgrade Paths

Single server and single z/OS Parallel Sysplex configurations with data sharing enable non-disruptive upgrades. Without data sharing, a server model upgrade needs requires an outage of several hours either during the MES or during the “push-pull” of servers. If this outage cannot be tolerated, then if a Parallel Sysplex with data sharing is already in place, it is straightforward to add a second machine with coupling links on a temporary or permanent basis, add it to the Parallel Sysplex, shift the workload to that second machine over some period of time, then either upgrade or retire the older machine model. This process can enable the hardware upgrade without suffering an extended outage.

Single server Parallel Sysplexes affects capacity planning. Ordinarily any clustering technology involving two or more nodes must provide enough capacity to support peak business critical operations when one node is offline. For example, if there are two nodes in the cluster and the peak utilization of the mission-critical workload is 10,000 MIPS, then either both nodes be configured to handle this peak load and run at under 50% busy, or else both servers be configured with Capacity Back-Up to be able to grow to handle the work if the other server is unavailable. In a single server Parallel Sysplex is different, all processing capacity is available to the entire cluster no matter how many nodes are active. The data sharing members are separate z/OS instances but not physically separated. If one member is offline the MIPS that it was using is now immediately available to the data sharing partner. Work that was being balanced between the two nodes are now directed to one node, using the same total MIPS. In both normal operations and abnormal operations (with one member offline) the utilization and capacity remain essentially level -- a uniquely efficient high availability solution.

Multi-server Parallel Sysplex customers that are concerned about loss of server capacity can implement two or more z/OS instances per machine so that an outage of a z/OS instance does not also remove processing capacity from the Sysplex. The surviving z/OS instance(s) on that machine can still tap into that server's full capacity. Operators can then take a z/OS instance offline with no material impact to the total capacity in the Sysplex. Customers also provision enough permanent and/or temporary capacity such as Capacity Backup and Capacity for Planned Events on every machine in the Parallel Sysplex so that they still have enough Sysplex-wide capacity in the rare event there's a whole machine outage. Typically they configure automation so that temporary capacity is instantly provisioned if a machine departs the Sysplex.

The sheer variety of CF configuration options provides great flexibility to customers since they can choose particular configuration options that meet their application availability and other service delivery requirements. When requirements change, they can typically reconfigure rapidly, with zero or near zero business impact and with no loss of investment. Single server Parallel Sysplex configurations are important, cost-effective, high availability options that are perfectly matched to many customer scenarios, and they represent unique IBM Z capabilities.

Coupling Facility Configuration Options

Pervasive Encryption Support

End-to-end encryption for both CF data in flight and data at rest in CF structures is supported. The encryption key for the CF data is kept within the CFRM Couple Data Sets. Prior to a z14 CF, if both CDSs are lost then all encrypted structures (as specified in the CFRM policy) need to be rebuilt. With a z14 CF, a copy of the key is also kept within the Coupling Facility.

Summary

This paper described various Coupling Facility configuration alternatives from several perspectives. Characteristics of each CF option was compared in terms of function, inherent availability, and price/performance. CF configuration alternatives was discussed in terms of availability requirements based on an understanding of CF exploitation by z/OS components and subsystems. Given the considerable flexibility IBM offers in selection of CF technology and configuration options and inherent tradeoffs between qualities of service that can be selected, there is some risk in trying to condense the content of this document into a few simple conclusions. However, we will do our best to do just that.

The various CF technology options have a great deal in common in terms of function and inherent availability characteristics, and that the fundamental decision points in selecting one option over another have more to do with price/performance and availability criteria dictated by customer choice of Parallel Sysplex environment (Resource Sharing versus Data Sharing).

The Internal Coupling Facility is a robust, price/performance competitive, and strategic CF option for many configurations, and a technology base built upon with System-Managed duplexing can be used as a high availability solution. The various CF technology options affects the performance characteristics and significant improvements in cost/performance value can be introduced through technology enhancements such as ICF, Dynamic ICF Expansion, and different coupling links.

From the perspective of configuring for availability, the configuration requirements are dictated by the recovery capabilities of specific CF exploiters. While there is much flexibility in the configuration choices based on cost versus value tradeoffs, a simple yet important distinction emerged regarding availability in Resource Sharing versus Data Sharing environments.

Those interested in the value of a Parallel Sysplex environment can obtain significant benefit from a Systems Enabled Resource Sharing environment with minimal migration cost. The value derived from system-level CF exploitation's (Resource Sharing) such as improved RACF and GRS performance, dynamic tape-sharing across the Parallel Sysplex cluster, simplified XCF Parallel Sysplex communications, can be easily introduced into existing configurations.

Coupling Facility Configuration Options

Further, Resource Sharing environments do **not** require a standalone CF in order to provide a highly-available Parallel Sysplex environment nor do they generally require use of CF Duplexing. For example, a configuration with two Internal CFs provides roughly the same robustness as a configuration with two standalone CFs for such an environment.

As one migrates forward to an environment supporting application data sharing to achieve all of the workload balancing, availability and scalability advantages offered by Parallel Sysplex technology, a standalone Coupling Facility becomes important from an availability perspective. This is due to the need for failure independence between the CF and exploiting z/OS system and subsystem components. System-Managed CF Structure Duplexing removes the requirement for a standalone Coupling Facility, but this has to be evaluated against the cost of duplexing.

A single standalone Coupling Facility configured with a second CF as an ICF or other server CF option can provide high availability at significant cost reduction in many customer configurations. With System-Managed CF Structure Duplexing, the configuration can be reduced down to two ICFs while still providing the same high availability configuration.

Finally, through exploitation of the significant CF technology alternatives and introduction of ever-evolving new CF options, IBM continues to demonstrate a strategic commitment to Parallel Sysplex clusters as a fundamental component of the IBM Z business value proposition.

Coupling Facility Configuration Options

Appendix A. Parallel Sysplex Exploiters - Recovery Matrix

This table assumes System-Managed CF Structure Duplexing is not configured. A table showing exploitation of duplexing can be found at the end of GM13-0103: System-managed CF Structure Duplexing technical paper, found at the Parallel Sysplex web site: ibm.com/systems/z/psa

Assumes single CF failure or 100% Loss of Connectivity failure

Exploiter	Function	Structure Type	Failure Independence Required?	Failure Impact
Db2	SCA	List	Yes	Data sharing continues
Db2 (IRLM)	Serialization	Lock	Yes	Data sharing continues
Db2 GBPs	Shared catalog/directory Shared user data	Cache	Yes	Shared data using the GBP is recovered from logs
Db2 GBPs (duplexing)	Shared catalog/directory Shared user data	Cache	Not applicable with duplexed structures	Data sharing continues
IMS (IRLM)	Serialization	Lock	Yes	Data sharing continues
IMS OSAM	Caching	Cache	No	Data sharing continues
IMS VSAM	Caching	Cache	No	Data sharing continues
IMS VSO DEDB	Caching	Cache	N/A	Data sharing continues / Duplexed structures
IMS CQS	Transaction balancing	List	No	Transaction balancing continues
z/OS System Logger IMS CQS Shared Message Queue	Merged log LSTRMSGQ	List	Yes (1)	Data sharing continues
z/OS System Logger IMS CQS Shared EMH	Merged log LSTREMH	List	Yes (1)	Data sharing continues
VTAM ISTGENERIC	SI to network generic resources	List	Yes (2)	Session balancing continues
VAM MNPS	Multi-node persistent sessions	List	Yes	MNPS continues
z/OS REesource Recovery Services (RRS)	Sync point management of merged logs	List	Yes (1)	High-speed Logging continues
z/OS System Logger – OPERLOG	Merged logs	List	Yes (2)	High-speed Logging continues
z/OS System Logger – LOGREC	Merged logs	List	Yes (2)	High-speed Logging continues

Coupling Facility Configuration Options

z/OS System Logger CICS – DFHLOG	CICS requirement. Emergency restart Backout processing	List	Yes (1)	Data sharing continues
z/OS System Logger CICS – DFHSHUNT	CICS requirement Forward recovery	List	Yes (1)	Data sharing continues
z/OS System Logger CICS	Forward Recovery	List	Yes (1)	Data sharing continues
CICS VSAM/RLS IGWLOCK00	Serialization	Lock	Yes	Data sharing continues
CICS VSAM/RLS	Caching of VSAM data	Cache	No	Data sharing continues
CICS Temp Storage queue	Tempstor affinities	List	No	Function terminates. Load balancing continues
CICS Named Counters	Unique index generation	List	No	User manage recovery needed
CICS data tables	Caching of VSAM data	Cache	No	“Scratch-pad” data lost
z/OS Allocation ISGLOCK	Shared tape	Lock	No	Shared tape continues
z/OS GRS ISGLOCK	ENQ/DEQ	Lock	No	ENQ/DEQ continues
z/OS XCF IXCxxxxxx	High-speed signaling	List	No	High-speed signaling continues
RACF	Cached security database	Cache	No	Normal RACF processing continues
JES2	Checkpoint	List	N/A	JES2 continues if duplexing
SmartBatch	Cross systems BatchPipes	List	N/A	Function not available. Rebuild supported for planned reconfiguration
WLM	Multi-system enclaves	List	No	Drop out of multi-system management
WLM	Intelligent Resource Director	List	No	Structure repopulated from WLM data over next 10s to 2m

- 1) For the System Logger, Failure Independence can be achieved through the use of Staging data sets or proper configuration planning. It is recommended to NOT use staging data sets due to the performance impact
- 2) Though failure independence is suggested, it is not necessary. Typical installations could live with the loss of data
- 3) Failure independence is required for LU6.2 SYNCVL2 only

Coupling Facility Configuration Options

For all exploiters, failure independence is the need for the CEC that is hosting the Coupling Facility to be physically separated from the z/OS image that is hosting the member of the multi-system applications.

Coupling Facility Configuration Options

Appendix B. Resource Sharing Exploiters - Function and Value

Systems Enabled represents those customers that have established a functional Parallel Sysplex environment using a Coupling Facility, but are **not** currently implementing IMS, Db2 or VSAM/RLS datasharing.

The following exploiters fall into this category.

Exploiter	Function	Benefit
z/OS XCF	High Speed Signaling	Simplified System Definition
z/OS System Logger	OPERLOG and LOGREC logstream	Improved Systems Management
z/OS Allocation	Shared Tape	Resource sharing
z/OS GRS/GRS STAR	Resource Serialization	Improved ENQ/DEQ performance, availability and scalability
Security Server (RACF)	High speed access to security profiles	Performance
JES2	Checkpoint	Systems Management
SmartBatch	Cross Systems BatchPipes	Load Balancing
VTAM GR (non LU6.2)	Generic Resource for TSO	Session Balancing/availability
z/OS RRS	Logger	Coordinate changes across multiple database managers
Catalog	Extended Catalog Sharing	Performance
MQSeries®	Shared message queues	System Management, capacity, availability, application flexibility
WLM	MultiSystem Enclaves	System Management
WLM	Intelligent Resource Director	System Management, Performance

Resource Sharing Exploiters

Efficient recovery in a Systems Enabled environment requires operations to take the necessary actions to complete the partitioning of failed images out of the Parallel Sysplex environment. This must be done as quickly as possible. Doing so will allow rebuild processing to continue (this includes proper COUPLExx interval specification, response to XCF's failure detection/partitioning messages and use of SFM). Operator training on failure recognition (i.e., responding to the IXC102A message) and the proper use of the XCF command is key and should not be overlooked. Installations that run with an active SFM policy may automate the partitioning process. Further automation could be applied to improve responsiveness. For additional details on availability items like CF volatility, isolate value, interval value and a sample recovery time line, please see WSC Flash 9829, "Parallel Sysplex Configuration Planning for Availability."

Coupling Facility Configuration Options

A brief description of each exploiter follows along with reasons to consider using them.

XCF, High Speed Signaling

Signaling is the mechanism through which XCF group members communicate in a sysplex. In a Parallel Sysplex environment, signaling can be achieved through channel to channel connections, a Coupling Facility, or both. Implementing XCF signaling through Coupling Facility list structures provides significant advantages in the areas of systems management and recovery. A signaling path defined through CTC connections must be exclusively defined as either outbound or inbound, a Coupling Facility list structure can be used as both. z/OS automatically establishes the paths. If more than one signaling structure is provided, the recovery from the loss of a signaling structure is automatically done by Cross System Extended Services (XES).

It is highly recommended to have redundant XCF signaling paths. A combination of CTCs and list structures may be used, two sets of CTCs, or two sets of structures. **Both** signaling structures should not be in the same CF. This situation can be monitored for with automation routines. WSC Flash: **Parallel Sysplex Performance: XCF Performance Considerations (Version 2)** contains more information on configuring and tuning XCF.

System Logger, OPERLOG and LOGREC

The system logger is a set of services which allows an installation to manage log data across systems in a Parallel Sysplex environment. The log data is in a log stream which resides on both a Coupling Facility structure and DASD data sets. System logger also provides for the use of staging data sets to enable the log data to be protected from a single point of failure. A single point of failure can exist because of the way the Parallel Sysplex environment is configured or because of dynamic changes in the configuration. Using system logger services, you can merge log data across a Parallel Sysplex environment from the following sources:

- A Logrec logstream. This will allow an installation to maintain a Parallel Sysplex view of logrec data.
- An Operations logstream. This will allow an installation to maintain a Parallel Sysplex view of SYSLOG data.
- CICS Transaction Server utilizing VSAM/RLS to provide the installation with VSAM datasharing in a Parallel Sysplex environment (discussed later).
- IMS Shared Message Queue providing support for transaction load balancing (discussed later).
- RRS merged logs.

Coupling Facility Configuration Options

z/OS GRS Star

A GRS Star configuration of Global Resource Serialization replaces the previous ring mode protocol and potential ring disruptions. The star configuration is built around a Coupling Facility, which is where the global resource serialization lock structure resides. By using the Coupling Facility, ENQ/DEQ service times could be measured in microseconds, not milliseconds. This has significant performance improvement. Installations that currently use GRS ring-mode or a third party alternative and host a fair number of z/OS images should consider migrating to GRS Star for the improved performance, scalability, and availability it provides. GRS Star is required for the Automatic Tape Switching (ATS Star) support.

z/OS Security Server (RACF), High-speed Access to Security Profiles

RACF uses the Coupling Facility to improve performance by using a cache structure to keep frequently used information located in the RACF database. The cache structure is also used by RACF to perform very efficient, high-speed cross invalidates of changed pages within the RACF database. Doing so preserves the working set that RACF maintains in common storage.

RACF is another example of an exploiter that can provide improved performance and scalability. RACF is capable of using the CF to read and register interest as the RACF database is referenced. When the CF is not used, updates to the RACF database will result in discarding the entire database cache working set that RACF has built in common storage within each system. If an installation enables RACF to use the CF, RACF can now selectively invalidate only changed entries in the database cache working set(s) thus improving efficiency. Further, RACF will locally cache the results of certain command operations. When administrative changes occur, such commands need to be executed on each individual system.

JES2, Checkpoint

JES2 supports placing its checkpoint in the Coupling Facility. When this option is selected, a Coupling Facility list structure is used for the primary checkpoint data set. The alternate checkpoint data set could reside either on DASD or in the other Coupling Facility. For recovery, JES2 provides its own recovery mechanism called the JES2 Reconfiguration Dialog. The dialog would be entered when a Coupling Facility fails that is hosting the checkpoint structure. Rebuild of the JES2 structure is supported. The benefits of having the JES2 checkpoint in the Coupling Facility include equitable access to the checkpoint lock across all members within the MAS complex and elimination of the DASD bottleneck on the checkpoint data set. In addition, members of the MAS are now capable of identifying who owns the lock in the event of a JES failure.

Coupling Facility Configuration Options

SmartBatch, Cross System BatchPipes

SmartBatch uses a list structure in a Coupling Facility to enable cross-system BatchPipes®. Cross-system BatchPipes is also known as pipeplex. In order to establish a pipeplex, the BatchPipes subsystems on each of the z/OS images in the Parallel Sysplex cluster must have the same name and connectivity to a common Coupling Facility list structure. It is possible to have multiple pipeplexes within a single Parallel Sysplex cluster by using different subsystem names. Each pipeplex must have its own list structure.

VTAM GR (non LU6.2) for TSO

VTAM provides the ability to assign a generic resource name to a group of active application programs that provide the same function. The generic resource name is assigned to multiple programs simultaneously, and VTAM automatically distributes sessions among these application programs rather than assigning all sessions to a single resource. Thus, session workloads in the network are balanced. Session distribution is transparent to the end user.

DFSMSHsm Common Recall Queue

The common recall queue (CRQ) is a single recall queue that is shared by multiple DFSMSHsm™ hosts. The CRQ enables the recall workload to be balanced across each of these hosts. This queue is implemented through the use of a Coupling Facility (CF) list structure. Prior to this enhancement, recall requests were processed only by the host on which they were initiated.

Enhanced Catalog Sharing

In a resource sharing Parallel Sysplex cluster, each image must have read/write access to both the Master Catalog and User Catalogs. The state of the catalog(s) must be accessible from all systems in the cluster to allow for data integrity and data accessibility. As the number of systems in the cluster grows, serialized access to catalogs could have negative sysplex-wide impact. z/OS contains enhancements that will improve the performance of shared catalogs in a Parallel Sysplex environment. With Enhanced Catalog Sharing (ECS), the catalog control record describing each catalog is copied into the Coupling Facility instead of DASD. This reduces the number of I/Os to the shared catalogs and improves overall system performance.

Coupling Facility Configuration Options

WebSphere MQ Shared Message Queues

WebSphere MQ supports shared queues for persistent and non-persistent messages stored in Coupling Facility list structures. Applications running on multiple queue managers in the same queue-sharing group anywhere in the Parallel Sysplex cluster can access the same shared queues for:

- High availability
- High capacity
- Application flexibility
- Workload balancing

You can set up different queue sharing groups to access different sets of shared queues.

All the shared queues stored in the same list structure can be accessed by applications running on any queue manager in the same queue-sharing group anywhere in the Parallel Sysplex cluster. Each queue manager can be a member of a single queue-sharing group only.

The MQ structure supports System-managed rebuild so it supports rebuild for planned and unplanned reconfigurations. Failure isolation is not required since persistent messages are recovered from the logs and non-persistent messages have no application impact if the queue is lost. The MQ structures also support System-Managed CF duplexing for high availability as well.

Workload Manager (WLM) Support for Multisystem Enclave Management

The z/OS Workload Manager supports Multisystem Enclaves. An enclave can be thought of as the “anchor point” for a transaction spread across multiple dispatchable units executing in multiple address spaces. The resources used to process the work can now be accounted to the work itself (the enclave), as opposed to the various address spaces where the parts of the transaction may be running. This can now be managed as a unit across multiple systems. Multisystem Enclaves support would provide system management and performance benefits.

Workload Manager support for IRD

Intelligent Resource Director consists of the functions of LPAR CPU Management, Dynamic Channel Path Management, and Channel Subsystem Priority Queuing. The scope of management is the set of all LPARs in the same server in the same Parallel Sysplex cluster. The first two functions require a CF structure.

CPU Management provides z/OS WLM with the ability to dynamically alter the logical partitions (LP) weights for each of the logical partitions within the LPAR cluster, thereby influencing PR/SM to allow for each logical partition to receive the percentage of shared CPU resources necessary to meet its workload goals.

Coupling Facility Configuration Options

Dynamic Channel Path Management (DCM) allows z/OS WLM to dynamically reconfigure channel path resources between control units within the cluster in order to enable business critical application's I/O processing requirements, necessary to meet its workload goals, are not being delayed due to the unavailability or over utilization of channel path resources.

The Workload Manager Multisystem Enclaves and IRD structures support System-managed Rebuild to move structures and System-Managed duplexing for high availability. Additionally, for both of these structures, there are no failure isolation requirements.

- If the Multisystem Enclave structure cannot be rebuilt, the system falls back to the pre-Multisystem Enclave management way of managing the enclaves: Each enclave is managed on a per-LPAR basis.
- If the IRD structure is unavailable, it gets reallocated on the other CF as soon as its loss is detected. For LPAR CPU Management, the structure is refreshed within 10 seconds (the WLM interval). For DCM, the history of CHPID usage is lost, and the system needs to relearn this for optimal performance. This is done in two minutes. During that time, the system runs no worse than in pre-DCM mode.

Recommended Configurations

Data Sharing Enabled represents those customers that have established a functional Parallel Sysplex environment doing database datasharing. Please reference the "Target Environments" section to understand the overall benefits when running in application enablement mode.

Coupling Facility Configuration Options

The following exploiters fall into this category.

Data Sharing Enabled Exploiters

Exploiter	Function	Benefit
DB2 V4/V5	Shared Data	Increased Capacity beyond single CEC/Availability
DB2 V6	Shared Data	Increased Capacity Availability, DB2 Cache Duplexing
DB2 IRLM	Serialization (Multisystem database locking)	Integrity/Serialization of Database datasharing
VTAM	Generic Resource	Availability/Load Balancing
VTAM (Multi-Node Persistent Sessions)	MNPS	Availability
CICS TS R1.0	CICSLOG Manager	Faster Emergency Restart, simplified systems management
DFSMS™ R1.3	Shared VSAM/OSAM	Increased Capacity beyond single CEC/Availability
IMS V5/V6		Increased Capacity beyond single CEC/Availability
IMS V5/V6 IRLM	Serialization (Multisystem database locking)	Integrity/Serialization of Database datasharing
IMS V6 SMQ	Transaction Balancing	Increased Capacity beyond single CEC/Availability
IMS V6 Shared VSO	Shared Data	Increased Capacity beyond single CEC/Availability
IMS V8 Resource Structure	Shared resource definition and status	Consistent IMSplex-wide resource definition and status
RRS	z/OS Synch point manager	Multisystem distributed synch point coordination

Appendix C. Data Sharing Exploiters - Usage and Value

Db2 Data Sharing

A Db2 data sharing group is a collection of one or more Db2 subsystems accessing shared Db2 data. Each Db2 subsystem in a datasharing group is referred to as a member of that group. All members of the group use the same shared catalog and directory and must reside in the same Parallel Sysplex cluster.

- The Coupling Facility list structure contains the Db2 Shared Communication Area (SCA).
- Each Db2 member uses the SCA to pass control information throughout the datasharing group. The SCA contains all database exception status conditions and other information necessary for recovery of the data sharing group.
- The Coupling Facility lock structure (IRLM) is used to serialize on resources such as table spaces and pages.
- It protects shared Db2 resources and allows concurrency.
- One or more Coupling Facility cache structures, which are group buffer pools that can cache these shared data pages.

For an in-depth discussion on Db2 Data Sharing recovery, consider reviewing Db2 on MVS™ Platform: Data Sharing Recovery (SG24-2218).

It is not required to have an SFM policy active in order to rebuild SCA/Lock when a loss of CF connectivity occurs although, as stated earlier, we strongly recommend running with an active SFM policy.

User-managed duplexing of the group buffer pools is a major improvement in Parallel Sysplex continuous availability characteristics. It is designed to eliminate a single point of failure for the Db2 GBP caches in the CF via Db2 managed duplexing of cache structures. This eliminates the need to perform lengthy log recovery (reduce from hours down to minutes) in the event of a CF failure containing the Db2 GBP cache structure. With Db2 (user-managed) GBP duplexing, at the time of a failure of the cache structure, Db2 will automatically switch over to use the duplexed structure in the other Coupling Facility. This enhancement makes it possible to safely configure your GBPs in the ICF.

VTAM, Generic Resources

VTAM provides the ability to assign a generic resource name to a group of active application programs that provide the same function. The generic resource name is assigned to multiple programs simultaneously, and VTAM automatically distributes sessions among these application programs rather than assigning all sessions to a single resource. Thus, session workloads in the network are balanced. Session distribution is transparent to the end user.

Coupling Facility Configuration Options

IMS support for VTAM Generic Resources to allow session load balancing across multiple IMS Transaction Managers. It provides a single system image for the end user. It allows the end user to logon to a generic IMS Transaction Manager name and VTAM to perform session load balancing by establishing the session with one of the registered IMS TMs. As the workload grows, IMS customers need the power that distributing the data and applications across the Parallel Sysplex cluster provides. This growth can be accomplished while maintaining the single system image for their end users. End users are able to access applications and data transparently, regardless of which IMS system in the Parallel Sysplex cluster is processing that request. Generic Resource support allows the end user to log on to a single generic IMS name that represents multiple systems. This provides for easy end-user interaction and improves IMS availability. It also improves systems management with enhanced dynamic routing and session balancing.

VTAM, Multi-Node Persistent Session

SNA network attachment availability is significantly improved in a Parallel Sysplex configuration through VTAM Multi-Node Persistent Session. With MNPS, client session state information is preserved in the Coupling Facility across a system failure, avoiding the need to reestablish sessions when clients relogon to host systems within the Parallel Sysplex cluster. This significantly reduces the outage time associated with getting clients reconnected to the server systems.

CICS TS, CICS LOG Manager

CICS TS exploit the System Logger for CICS backout and VSAM forward recovery logging as well as auto-journaling and user journaling. Exploiting the System Logger for the CICS backout log provides faster emergency restart. Exploiting the System Logger for VSAM forward recovery logs and for journaling allows merging of the logged data across the Parallel Sysplex cluster in real-time. This simplifies forward recovery procedures and reduces the time to restore the VSAM files in the event of a media failure. CICS TS enhancements include sign-on retention for persistent sessions, automatic restart of CICS data sharing servers, and system-managed rebuild of Coupling Facility structures. These include the Named Counters, Data Tables, and Temporary Storage structures.

For further details see the CICS Release Guide, GC33-1570 and the CICS Recovery and Restart Guide, SC33-1698.

Coupling Facility Configuration Options

VSAM RLS

CICS TS provides support for VSAM Record Level Sharing (RLS). This function supports full VSAM datasharing, that is the ability to read and write VSAM data with integrity, from multiple CICS systems in the Parallel Sysplex environment. This provides increased capacity, improved availability and simplified systems management for CICS VSAM applications. In addition, the CICS RLS support provides partial datasharing between CICS systems and batch programs. For further details see the CICS Release Guide, GC33-1570.

VSAM RLS requires CF structures. A CF lock structure with name IGWLOCK00 must be defined. In addition, CF cache structures must be defined. The number and names of the CF cache structures are defined by the customer installation. Refer to publication DFSMS/MVS™ DFSMSdfp™ Storage Administration Reference, SC26-4920-03, for details of how to specify the CF cache structure names in the SMS Base Configuration.

VSAM RLS provides nondisruptive rebuild for its CF lock and cache structures. Rebuild may be used for planned CF structure size change or reconfiguration. In the event of CF failure or loss of connectivity to CFs containing RLS structures, RLS internally rebuilds the structures and normal operations continue. However, the lock rebuild requires failure independence. The lock structure may be System-Managed duplexed for high availability.

Coupling Facility Configuration Options

Transactional VSAM (DFSMSStvs) is available as a priced feature on z/OS. Based on VSAM/RLS, it allows full read/write shared access to VSAM data between batch jobs and CICS online transactions. In addition to the RLS requirements, DFSMSStvs uses RLS, which requires the z/OS System Logger.

IMS Shared Data

IMS n-way data sharing is IMS/DB full function support, which allows for more than 2-way data sharing.

Each z/OS image involved in IMS n-way Parallel Sysplex data sharing must have at least one IRLM. Each IRLM is connected to the IRLM structure in a Coupling Facility.

With IMS data sharing the concept of a data sharing group applies. A data sharing group is defined using the CFNAMES control statement. The IMS subsystems are members of this group.

Three Coupling Facility structure names must be specified for Parallel Sysplex data sharing. The names are for the:

- IRLM structure (the IRLM lock table name)
- OSAM structure
- VSAM structure

IMS Parallel Sysplex data sharing uses the OSAM and VSAM structures for buffer invalidation only.

Like the VSAM/RLS and Db2 counterparts, IMS data sharing can provide increased application availability and virtually unlimited horizontal growth.

Support is provided to utilize the Coupling Facility for store through caching (storing changes) and store clean data (storing unaltered data) for OSAM databases. IMS allows OSAM database buffers to be selectively cached in a central Coupling Facility, for increased performance.

IMS Shared Message Queue

The IMS Shared Message Queue (SMQ) solution allows IMS/ESA® customers to take advantage of the Coupling Facility to store and share the IMS message queues among multiple IMS systems. Incoming messages from an IMS on one Central Processor Complex (CPC) in the Parallel Sysplex cluster could be placed on a shared queue in the Coupling Facility by the IMS Common Queue Server (CQS) for processing by an IMS on another CPC in the Parallel Sysplex cluster. This can provide increased capacity and availability for their IMS systems. This SMQ solution takes advantage of the MVS function Cross System Extended Services (XES) to access the Coupling Facility, and the Automatic Restart Manager (ARM) to improve availability by automatically restarting the components in place or on another CPC. With shared

Coupling Facility Configuration Options

queues, a message can be processed by any IMS sharing the queues. This can result in automatic workload distribution as the messages can be processed by whichever IMS subsystem has the capacity to handle the message. An IMS customer can use shared queues to distribute work in the Parallel Sysplex cluster. In a shared queues environment, all transaction-related messages and message switches are put out on shared queues, which permits immediate access by all the IMSes in a shared queues group. This provides an alternative to IMS transfer of messages across MSC links within a shared queues group. The IMS SMQ solution allows IMS customers to take advantage of the Parallel Sysplex environment, providing increased capacity, incremental horizontal growth, availability and workload balancing.

Fast Path Shared Expedited Message Handler (EMH): This provides the capability to process Fast Path transactions in a Parallel Sysplex environment using the Coupling Facility. This enables workload balancing of Fast Path transactions in a Parallel Sysplex environment and hence provides increased capacity. This capability can also improve the system and data availability to the end user since multiple IMS subsystems can have access to the shared EMH queues in the Coupling Facility. This provides for increased throughput in Fast Path Message Processing.

IMS Shared VSO

IMS V6 Fastpath support provides block level data sharing of IMS Fast Path Data Entry Data Base (DEDB) Virtual Storage Option (VSO) databases, using the Coupling Facility. The VSO structure can be duplexed. This provides for improved workload balancing and offers increased capacity and growth in moving into Parallel Sysplex environments.

For information on handling the recovery from these situations, refer to IMS Operations Guide, SC26-8741, or the ITSO Redbook S/390® MVS Parallel Sysplex Continuous Availability SE Guide, SG24-4503.

Coupling Facility Configuration Options

VSO DEDB

There is no structure rebuild support for VSO DEDB Areas, either manual or automatic.

If two structures are defined for a specific area, then all data is duplexed to each structure, performed by IMS. This enables each structure to act as a backup for the other in the case of any connectivity failure from one or all IMSes, structure failure, or CF failure. In any of these cases, all IMS subsystems are informed and processing continues from the remaining structure.

The decision whether to define two structures for any specific area is installation-dependent. As each structure is a store-in cache, there is always the possibility that at any given point in time, data has not been hardened to DASD. Thus, any loss of a single structure will inevitably require DB recovery action, which could be time consuming. This problem is alleviated by the System-Managed CF duplexing support by these structures.

RRS, z/OS SynchPoint Manager

Many computer resources are so critical to a company's work that the integrity of these resources must be protected. If changes to the data in the resources are corrupted by a hardware or software failure, human error, or a catastrophe, the computer must be able to restore the data. These critical resources are called protected resources or, sometimes, recoverable resources. Resource recovery is the protection of the resources. Resource recovery consists of the protocols, often called two-phase commit protocols, and program interfaces that allow an application program to make consistent changes to multiple protected resources.

In the OS/390® Release 3 (with additional functions in following releases) IBM provided as an integral part of the system, commit coordination or synchpoint management as part of z/OS Recoverable Resource Management Services (RRMS). These services enabled transactional capabilities in recoverable resource managers. Additionally, these services provided Protected Communication Resource Managers distributed transactional capabilities in many z/OS application environments.

The Recoverable Resource Management Services consist of three parts for managing transactional work in z/OS: 1) Context Services for the identification and management of work requests, 2) Registration Services for identifying a Resource Manager to the system, and 3) Resource Recovery Services or RRS to provide services and commit coordination for all the protected resource managers participating in a work request's transactional scope in many of z/OS application environments. Through RRMS capabilities, extended by communications resource managers, multiple client/server platforms can have distributed transactional access to z/OS Transaction and Database servers.

Coupling Facility Configuration Options

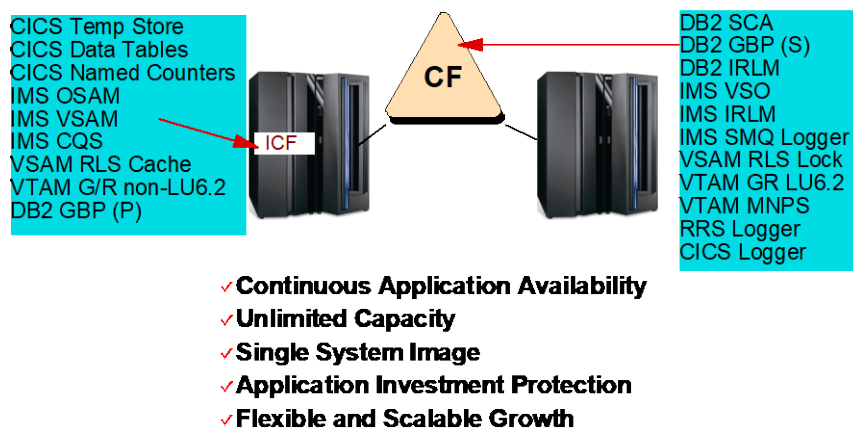
RRS uses five log streams that are shared by the systems in a sysplex. Every z/OS image with RRS running needs access to the Coupling Facility and the DASD on which the system logger log streams are defined.

Note that you may define RRS log streams as Coupling Facility log streams or as DASD-only log streams.

The RRS images on different systems in a sysplex run independently but share logstreams to keep track of the work. If a system fails, RRS on a different system in the sysplex can use the shared logs to take over the failed system's work.

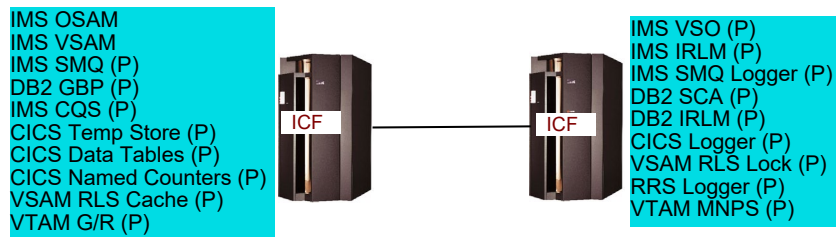
Recommended Configurations

Sample Data Sharing Configuration Without SM Duplexing



For simplicity, only the Primary duplexed structures are shown. Although IMS OSAM and VSAM structures do support System-Managed CF Structure Duplexing, they are not shown as duplexed since they obtain little value from this.

Sample Data Sharing Configuration With SM Duplexing



- ✓ Continuous Application Availability
- ✓ Unlimited Capacity
- ✓ Single System Image
- ✓ Application Investment Protection
- ✓ Flexible and Scalable Growth

Appendix D. CF Alternatives Questions and Answers

Typical questions and their associated answer relative to IBM's Coupling Facility and its usage.

Do internal CFs require an LPAR?	Yes. The Coupling Facility Control Code (CFCC) always runs in a Logical Partition – regardless of how the machine is configured
How is ICF selected from the HMC?	On machines with the ICF feature installed, see the Image Profile – Processor Page (on the HMC). You will notice there is a new option that will allow you to select ICF as a CP source. You must be in LPAR mode to see the option.
Are ICFs sized the same as standalone CFs?	Yes, ICF should still be sized in the traditional fashion using formulas and techniques that are subsystem dependent. ICF provides the MIPS part of the sizing equation – Links and Storage are still needed. Links are not needed in when using Internal Coupling channels (IC).
Should internal CFs be used in all coupling configurations?	No, what is very important is to first understand the failure independence that is required.. When doing application data sharing without all the data sharing structures duplexed, at least one standalone CF is strongly recommended. But an internal CF can be an excellent backup for this standalone CF, or can be used exclusively in system-enabled configurations. One can also use System-Managed duplexing to remove failure isolation issues after analyzing the host CPU cost vs. benefits.
Is adding an processor to the CF subject to the same MP effects as traditional z/OS engines?	Yes. It is similar to adding another z/OS engine in all ways except in what is being run on it. The ICF will vie for resources (storage hierarchy, busses within the system, etc.) with whatever is running on the other engines.
If we choose to leverage ICF on an IBM machine, will this affect software pricing?	No, due to the fact that the CPs required to support the ICF are dedicated to the coupling function, no IBM software license charges apply to the partition.
Which is better: ICFs or standalone CFs?	Performance: Since ICFs can share the same footprint as the z/OSs that they connect to, efficient IC links can be employed. This yields the best performance (with pricing advantages). For a Resource Sharing-only environment, this is the recommended configuration. In a Data Sharing environment, many structures require failure-isolation. Since there is a performance hit to System-Managed CF Structure Duplexing, configuring a standalone CF as the primary location to hold the structures requiring failure-isolation would provide the best performance.

Availability: System-Managed CF Structure Duplexing removes availability concerns of an ICF-only environment for those structures being duplexed. Generally, all the structures will NOT be duplexed due to performance issues. In addition, some microcode upgrades needed for the Coupling Facility are disruptive and the CEC would be needed to be re-IMLed. This affects the z/OS workload. For these reasons, standalone CFs provide better availability.

System Management: Processor family upgrades are sometimes needed to support capacity requirements. If an ICF was configured on the same footprint, then the ICF would automatically be upgraded at the same time, certainly keeping it "within one generation" of the z/OS workload. In addition, less footprints to manage, install, power, etc. makes ICFs easier to manage than standalone CFs.

Software Pricing: Since neither standalone CFs or ICFs can run z/OS, the capacity used for both of these are not calculated as part of the software licensing charge calculations. If running a "normal LPAR" as a CF using general purpose CPs, then the capacity used for this coupling facility is could have software pricing implications. Contact your IBM software sales representative.

Overall: In a Resource Sharing-only environment, two ICFs are recommended. Once Data Sharing is introduced, at least one standalone CF should be considered.

Under which circumstances should I consider enabling Dynamic Dispatch on a coupling facility? Dynamic Coupling Facility Dispatch (DCFD) is an enhancement to CF LPARs that consist of shared CP or ICF processors. With Dynamic CF Dispatch enabled, the CFCC LPAR voluntarily gives up use of the processor when there is no work to do. This uses very little CP resource, leaving the general purpose CPs to process the normal production work.

The tradeoff for a decreased usage of production CP or ICF processors is an increase in response time from the Coupling Facility with Dynamic Dispatch enabled. Any request that arrives at the coupling facility when it has given up its use of the processor will wait. So enabling Dynamic Dispatch should be limited to environments where responsive performance is not a primary objective, such as a test CF image. It could also be used as a backup CF. If the primary CF should fail, the backup CF in the production system would sense an increasing queue of CF work and begin employing the general purpose CPs to execute the work. As the activity rate to this backup CF increases, it's response time improves at the expensive of increased usage of the CP.

Coupling Facility Configuration Options

	DCFD=THIN is the only option supported
	DCFD=THIN is the only option with shared processors on z17
Can we consider the Internal Coupling Facility (ICF) as an option for a low cost entry CF for migration and/or to support a development environment?	Yes, the Internal Coupling Facility (ICF) engine can host a Coupling Facility without increasing software licensing costs. ICF provides a low cost CF which can be used as an entry level, test/development or single system production CF.
Why would I want a Parallel Sysplex environment when I have no plans to move to an application data sharing environment?	The IT industry is moving as rapidly as possible to clustered computer architectures. The z/OS Parallel Sysplex environment is a highly advanced, state of the art clustered computer architecture because of its advanced data sharing technology. An additional benefit is the (PSLC) pricing feature. Prior to database data sharing customers receive value through the resource sharing capabilities of exploiters like VTAM, GRS, JES2, System Logger, Catalog, IRD, etc. This allows fail-over capability and single point of control for systems management.
Is the Coupling Facility a good vehicle for XCF signaling or should we still use CTCs?	XCF is recommended over CTCs. XCF signaling through CF performance is better than CTCs while reducing the need for additional hardware. Use of multiple XCF signaling structures in multiple CFs is essential for high availability. More information is available in WSC flash: ibm.com/support/docview.wss?uid=tss1flash10011 Parallel Sysplex Performance: XCF Performance Considerations.
Should I use System-Managed CF Structure Duplexing for all of my data sharing structures?	System-Managed Duplexing has additional performance costs over "simplex" mode. This includes the additional CF request to the duplexed structure for each update, time for the duplexed structure to coordinate the request, and z/OS receiving the additional response and reconciling them. This has to be balanced with the benefits including eliminating the requirement for a standalone CF.
What other benefits does System-Managed Duplexing provide?	The primary purpose of System-Managed Duplexing is to provide a rapid recovery in a loss of connectivity or failure situation. This is especially true in the case of a structure that would otherwise not support recovery such as CICS Temporary Storage, or WebSphere MQ Shared Queues.
Where can I get more information on configuring System-Managed Duplexing?	The technical paper "System-Managed CF Structure Duplexing," ZSW0-1975USEN, contains detailed information on what it is, how it works, and the environments where it would provide the most benefit

Coupling Facility Configuration Options

How does XCF choose a path for signaling?	For XCF, the path with the best response time is preferred. XCF considers all of the paths available to it, the CTCs and CF structure paths, and the one that provides the best response time is selected.
How are paths chosen for CF messages?	Path selection is generally done by the server firmware based upon different classes of links, classified by performance. We use from the highest-performance class that has available buffers for use.
How can I test the effectiveness of new links?	Add the new links, then vary the old links offline for the duration of the testing
What is the maximum number of CHPIDs to a single CF?	Each z/OS is limited to 8 CHPIDs to a given CF Each z/OS can use only one CU to a given CF

Coupling Facility Configuration Options



©Copyright IBM Corporation 2025

IBM Corporation

New Orchard Road

Armonk, NY 10504

U.S.A.

Produced in the United States of America,

04/2025

IBM, ibm.com, the IBM logo, IBM Z, IBM eServer, BatchPipes, CICS, Db2, DFSMS, DFSMSdfp, DFSMSHsm, DFSMS/MVS, ESCON, FICON, GDPS, IMS, IMS/ESA, InfiniBand, MQSeries, MVS, OS/390, Parallel Sysplex, PR/SM, RACF, Resource Link, RMF, S/390 Sysplex Timer, System z, System z9, System z10, VTAM, WebSphere, z14, z15, z16, z17, z/OS, z Systems and z/VM are trademarks or registered trademarks of the International Business Machines Corporation.

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 therefrom.

InfiniBand and InfiniBand Trade Association are registered trademarks 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.

Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

The registered trademark Linux® is used pursuant to a sublicense from the Linux Foundation, the exclusive licensee of Linus Torvalds, owner of the mark on a worldwide basis.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

OpenStack is a trademark of OpenStack LLC. The OpenStack trademark policy is available on the [OpenStack website](#).

Red Hat®, JBoss®, OpenShift®, Fedora®, Hibernate®, Ansible®, CloudForms®, RHCA®, RHCE®, RHCSA®, Ceph®, and Gluster® are trademarks or registered trademarks of Red Hat, Inc. or its subsidiaries in the United States and other countries.

TEALEAF is a registered trademark of Tealeaf, an IBM Company.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Worklight is a trademark or registered trademark of Worklight, an IBM Company.

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

The information contained in this documentation is provided for informational purposes only. While efforts were made to verify the completeness and accuracy of the information contained in this documentation, it is provided "as is" without warranty of any kind, express or implied. In addition, this information is based on IBM's current product plans and strategy, which are subject to change by IBM without notice. IBM shall not be responsible for any damages arising out of the use of, or otherwise related to, this documentation or any other documentation. Nothing contained in this documentation is intended to, nor shall have the effect of, creating any warranties or representations from IBM (or its suppliers or licensors), or altering the terms and conditions of the applicable license agreement governing the use of IBM software.

References in these materials to IBM products, programs, or services do not imply that they will be available in all countries in which IBM operates.

Product release dates and/or capabilities referenced in these materials may change at any time at IBM's sole discretion based on market opportunities or other factors and are not intended to be a commitment to future product or feature availability in any way.

ZSW01971-USEN-29