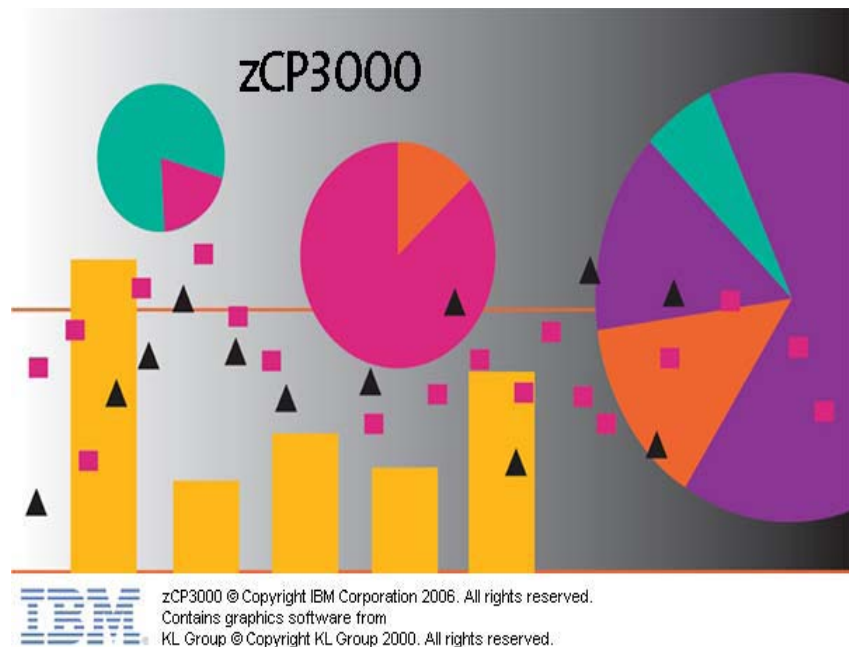# IBM Advanced Technical Skills
At the Washington Systems Center

Shared Coupling Facility Links in Sysplex

Gretchen Frye
frye@us.ibm.com

Version 2
3 December 2010

Shared CF Links in Sysplex

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. These and other IBM trademarked terms are marked on their first occurrence in this information with the appropriate symbol (® or ™), indicating US registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at http://www.ibm.com/legal/copytrade.shtml

The following terms are trademarks of the International Business Machines Corporation in the United States, other countries, or both:

CICSPlex®	MQSeries®	RMF™
CICS®	MVS™	System Storage®
DB2 Connect™	OMEGAMON®	System z10®
DB2®	OS/390®	System z®
DRDA®	Parallel Sysplex®	Tivoli®
DS8000®	PR/SM™	VTAM®
ESCON®	RAA®	WebSphere®
FICON®	RACF®	z/OS®
GDPS®	Redbooks®	z/VM®
HiperSockets™	Redbooks (logo) ®	z10™
IBM®	Resource Measurement	z9®
IMS™	Facility™	

The following terms are trademarks of other companies:

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

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

Other company, product, or service names may be trademarks or service marks of others

Introduction

This paper discusses a subset of issues having to do with the link configuration of a sysplex. There are many important considerations which this paper does NOT discuss, including

- ICF vs standalone CFs
- System managed duplexing
- Shared ICF engines
- Sysplex timer
- Structure placement

For information about these issues, and for an additional level of detail about coupling facility links which this paper does not provide, there are many excellent publications available, including, Coupling Facility Configuration Options[1] or the IBM RedBook "Getting Started with InfiniBand on System z10 and System z9" [2], or another IBM RedBook "System z Parallel Sysplex Best Practices"[3].

An Overview of Coupling Facility Links

The z/OS members and Coupling Facilities (CFs) of a Parallel Sysplex communicate over specialized, high-speed connections (links). Each z/OS image and coupling facility can belong to only one sysplex, and all members of a sysplex must be connected to all coupling facilities in that same sysplex. Generally, in a production sysplex, more than one coupling facility, and more than one channel path per CF-SYS connections, is configured. The channel paths between a z/OS system and a coupling facility are usually, but don't have to be, of the same type. For example, the connection between SYSA and CF1 could consist of 1 ICB-4 link and 1 ISC-3 link. zCP3000 cannot currently model links between a system-CF pair that are of more than one type.

Coupling Facility Link Types

Coupling Facility Links (CF links) are high speed fibre or copper interconnects used to communicate between CF partitions and z/OS partitions (images). CF links are also used for CF LPAR to CF LPAR communication for System Managed CF Structure Duplexing, There are a number of types of CF links, listed in order of speed (fastest first). Table 1 on page 2 lists the expected data transfer rate in a production environment.

- **Internal Coupling (IC)**, is a virtual link between a z/OS image and a CF residing on the same mainframe hardware. Using microcode and main memory to transfer data, it is the fastest type of communication link. All current IBM processors support IC links.

---

[1] ftp://public.dhe.ibm.com/common/ssi/ecm/en/zsw01971usen/ZSW01971USEN.PDF
[2] http://www.redbooks.ibm.com/redbooks/pdfs/sg247539.pdf
[3] http://www.redbooks.ibm.com/redpieces/pdfs/sg247817.pdf

- **Integrated Cluster Bus-4 (ICB-4)** links are high-speed copper based links between two co-located mainframes (up to 10 meters). ICB-4 is supported for z890, z990, z9, and z10, however it is not supported on the z196 system. ICB4 links are orderable in increments of 1.

- **InfiniBand-12x** links are high-speed fibre or copper based links between nearby mainframes (up to 150 meters). InfiniBand is supported for z9 architecture and above. InfiniBand connections between z9 and any other machine operate at Single Data Rate (SDR), whereas all other InfiniBand connections operate at Double Data Rate (DDR). InfiniBand links are ordered by adapter, each of which supports 2 ports and up to 16 chpids.

- **Integrated Cluster Bus-3 (ICB-3)** links are high-speed copper based links between two co-located mainframes (up to 10 meters). ICB-3 is supported for z800, z900, z890, z990, and z9. ICB-3 is not supported on z10 or z196. ICB3 links are orderable in increments of 1.

- **InfiniBand Long Reach (InfiniBand-1x)** links are fibre or copper fibre based links between two mainframes up to 10km apart (unrepeated), or up to 100km apart (repeated by DWDM). InfiniBand-1x is supported for z10 architecture and above. InfiniBand links are ordered by adapter, each of which supports 2 ports and up to 16 chpids.

- **InterSystem Coupling-3 (ISC-3)** links are fibre links between systems up to 20km apart (unrepeated), or up to 200km apart (repeated by DWDM). ISC3 links are orderable in increments of 1.

	Expected Data Transfer Rate (MB/sec)						
	IC	**IFB 12x**	**IFB 1x**	**ICB-4**	**ICB-3**	**ISC-3 <10km**	**ISC-3 >10km**
z800	1125				500	200	100
z900	1440				500	200	100
z890	3500			1500	500	200	100
z990	3500			1500	500	200	100
z9 BC	4000	600		1500	500	200	100
z9 EC	5000	600		1500	500	200	100
z10 BC	3200	1000	400	1500		200	100
z10 EC	7000	1000	400	1500		200	100
z196	10000	1000	400			200	100

Table 1. Expected Data Transfer Rate by CF Link Type.

What Is InfiniBand Architecture?

InfiniBand is an industry-standard specification developed by the InfiniBand Trade Association[TM4] that defines an input/output architecture used to interconnect servers, communications infrastructure equipment, storage, and embedded systems. It is a switched fabric communications link featuring high speed, low latency connections, and low cpu overhead. It is used on the System z platform to connect processor memory to I/O features and to provide coupling links to other servers in point-to-point configurations. It is also used on IBM System p and IBM BladeCenter platforms for high-speed cluster interconnect.

InfiniBand Host Channel Adapter (HCA) Types

Several different adapter are available on zEnterprise z196 processors and on System z z9 and z10 processors. They are referred to as InfiniBand-12X and the InfiniBand-1X "Long Reach". 12X and 1X refer to the number of physical lanes (send and receive pairs), each of which can support multiple transport services (in other words, lanes do not tie directly to channel paths). InfiniBand-12X refers both to the HCA1 adapter, which operates at Single Data Rate (SDR), and the HCA2 adapter, which operates at Single Data Rate (SDR) or Double Data Rate (DDR). The difference in speed between SDR and DDR can be seen in Table 1 on page 2 above.

- **HCA1-O** is an optical InfiniBand connection on System z9. HCA1-O is used in combination with the 12X IB-SDR link rating to provide a link rate of up to 3 GBps. HCA1-O to HCA1-O communication (z9 to z9) is not supported.
- **HCA2-C** is a copper InfiniBand connection from a book to I/O cages on a System z10 and zEnterprise z196. It is used with the 12X IB-DDR link rating to provide a link rate of up to 6 GBps.
- **HCA2-O** is an optical InfiniBand connection for interconnecting zEnterprise z196s and System z10s, or for connecting zEnterprise z196s and System z10s to System z9. HCA2-O is used in combination with the 12X IB-DDR link rating to provide a link rate of up to 6 GBps between z196 or z10 servers and up to 3 GBps between z10 and z9 servers. When communicating with a System z9, HCA2-O negotiates down to the 12X IB-SDR link rating to provide a link rate of up to 3 GBps.
- **HCA2-O LR** provides an optical InfiniBand connection for long reach coupling links from System z10 to System z10. It is used with the 1X IB-DDR link rating to provide a link rate of up to 5 Gbps or automatically scales down to 2.5 Gbps (1X IB-SDR) depending on the capability of the attached equipment. The PSIFB Long Reach feature is not available on System z9.

---

[4] Of which IBM is a founding member.

CF Link Configurations

A CF Link is a channel path between one system and one coupling facility. Each CF Link channel path (chpid) consists of 7 hardware buffers which map to the 7 subchannels used by software to queue requests[5]. Seven requests can concurrently run to or from the coupling facility. An individual request travels over one and only one of the subchannels, and the speed with which an individual request travels to the coupling facility is considered the link speed. There can be, and often are, more than one channel path to a coupling facility. These channel paths from a z/OS system to a coupling facility are usually all of the same type, but it is possible to configure them to be of different types. For example, SYSA may have both an ICB4 channel path and also an ISC3 channel path defined to CF1.

Dedicated Links

A dedicated link means that the 7 subchannels from one LPAR are mapped to the 7 hardware buffers of the channel path (Figure 1. Dedicated CF Link Channel Path.). Channel path busy does not occur on a dedicated link[6], because z/OS will queue a request, rather than attempt to run it, if all subchannels are already busy. The utilization of this channel path is simply the sum of service times for all requests from SYSA to CF1, divided by the number of hardware buffers (in Figure 1, there are 7). This type of CF link can be configured as IC, ISC-3, ICB-4, or InfiniBand links.
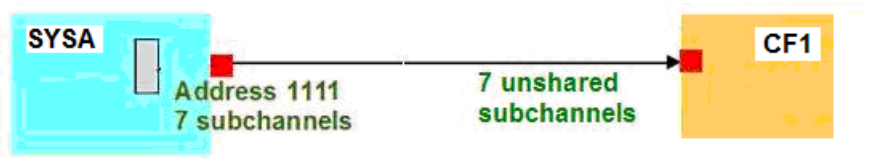

Figure 1. Dedicated CF Link Channel Path.

Shared Channel Paths

The link to a coupling facility can be shared by two or more sysplex members running in partitions on the same mainframe (Figure 2). A single channel path can only go to one coupling facility, thus the z/OS members and coupling facility all have to be in the same sysplex. A shared link means that the 14 subchannels from two LPARs (SYSA and SYSB) are mapped to the 7 hardware buffers of the channel path. The z/OS system will queue requests in a software queue if it knows the subchannels are busy, however it does not know about the requests on the link from other systems sharing the link. A Path Busy condition can occur when z/OS attempts to start an operation (having already obtained a subchannel) but finds that all the hardware buffers are busy with operations from this LPAR or the other LPAR sharing the link. The utilization of this channel path is simply the

---

[5] All modern links run in peer mode. Older machines and links could run in "compatibility" mode, with 2 sending subchannels and 2 receiving subchannels, but that is not considered in this paper.
[6] Well, ok, technically, it may occur, but very rarely, if a link buffer is stolen for a brief moment for some internal communication.

4

sum of service times for all requests from SYSA to CF1 and from SYSB to CF1, divided by the number of hardware buffers (7). This type of cf link can be configured as IC, ISC-3, ICB-4, or InfiniBand links.
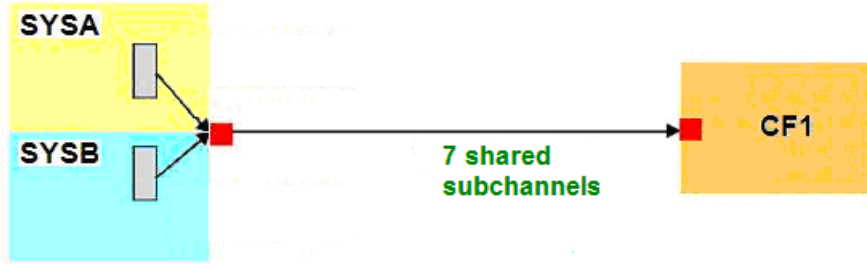

Figure 2. One channel path shared by two partitions.

Shared Host Channel Adapters

Each ISC or ICB type link can support one and only one channel path, however each InfiniBand adapter can support up to 16 chpids (a maximum of 8 is recommended), enabling the same underlying hardware to support multiple communication paths, to multiple coupling facilities, and to one or more sysplexes. For example, the 7 subchannels making up the InfiniBand link (chpid) from SYSA to CF1 and the 7 subchannels of the InfiniBand link from SYSA to CF2, as shown in Figure 3, can run independently on the same InfiniBand cable, whereas this would require 2 physical ICB4 cables.
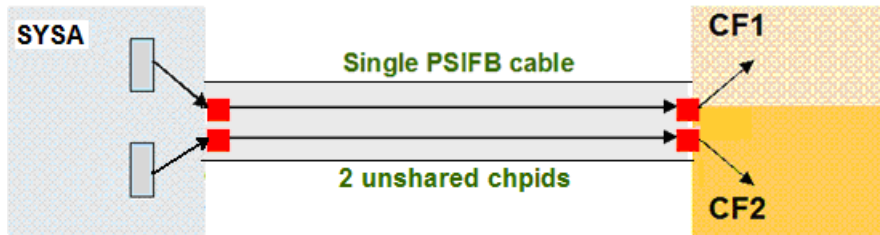

Figure 3. Multiple Chpids per InfiniBand Link.

Shared and Dedicated CHPIDs on a Shared Host Channel Adapter

In addition to the sharing of the physical link, channel paths can be logically shared, just as they can on ISC and ICB link types. So there can be a number of combinations of dedicated, logically shared, and physically shared links, with coupling facilities and z/OS member systems on either end of the InfiniBand link, as show in Figure 4.
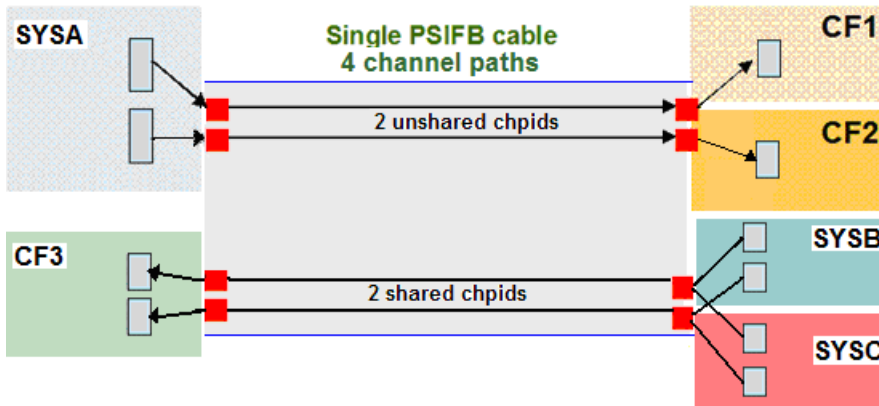
5

Figure 4. A mix of shared and unshared chpids.

So while ISC-3 and ICB-4 links can only be shared by partitions resident on the same CEC sharing the channel path, which can have only one destination coupling facility, InfiniBand links can support z/OS members on one machine talking to a CF on a second, and z/OS members on the second machine talking to a CF on the first, with or without sharing channel paths.

In zCP3000, the type of sharing involving shared channel paths is referred to as "logical sharing", and the type of sharing where multiple channel paths are sharing an InfiniBand link, is call "physical sharing".

Identifying Shared Resources

Partitions on the same system can share a channel path, so that what appear to be separate CF links from separate z/OS images actually run over the same physical link.

Unfortunately, it is not possible to know from the SMF records that shared channels are being used. You need to know from the customer which channel paths are shared. You either need the IOCDS or a statement from the customer.

Ask the Customer

The easiest way to figure this out is simply to ask the account team or the customer what chpids are shared, and how the InfiniBand links are configured. Ask something along the lines of "there are 2 ICB channels between CEC1 and CEC2; are they shared by the links running from SYSA, SYSB, and SYSC over to CF1 and CF2?" Something like that.

Identifying Shared CHPIDs from the IOCDS

If you have the IOCDS, shared channel paths (all types, not just CF links) can be identified in CHPID macro in the IOCDS. For example, Figure 5 shows that the

6

ICB link at chpid F8 is shared between the z/OS systems running in partitions LP31A and LP31B.

```
CHPID PATH=(CSS(0),F8),SHARED,
      PARTITION=((LP31A,LP31B),(=)),
      PCHID=280,TYPE=CBP
```

Figure 5. IOCDS example of a shared ICB-4 CF link

Here is the same CHPID defined as an InfiniBand link.

```
CHPID PATH=(CSS(0),F8),SHARED,
      PARTITION=((LP31A,LP31B),(=)),
      CPATH=(CSS(0),F1),
      CSYSTEM=CPCCF1,AID=00,PORT=1,TYPE=CIB
```

Figure 6. IOCDS example of a shared InfiniBand CF link.

Note that the IOCDS uses the partition name, so you have to figure out what the sysid is that this belongs to. Note also that, while this chpid is defined, it may not actually exist.

Identifying Shared Host Channel Adapters from the IOCDS

Using the IOCDS, the AID and PORT keywords on the CHPID macro specify the host channel adapter ID and port number of the CIB channel path. Different CHPIDs that are using the same HCA id are sharing the link adapter.

7

Performance Implications

There are many important considerations for good performance which this paper does NOT discuss, including the implementation of system managed duplexing, shared ICF engines, lock contention, and other issues. This paper is concerned only with giving a basic understanding of how changes to the coupling facility links can impact performance.

Changing the Link Type

Different link types will have different characteristics of
data transfer time and link latency, which will affect the service time for the request. Service time can affect the Host Sysplex Cost and whether a CF request is processed synchronously or asynchrously.   These topics are discussed further below.

Host Sysplex Cost

The cost of sysplex to the z/OS system is the additional time to complete an operation (ie, a transaction or a database operation) in a sysplex environment as compared with a non-sysplex environment. Host cost includes the time spent forming and sending the request, and for synchronous requests, the time spent waiting for the request to complete (the service time of the request). Additionally, for duplexed structures, the request is also sent to the peer CF, and some additional time is required to overlap the sending of the requests.

An increase in service time causes the sysplex   cost for the z/OS system to grow because it is now waiting longer for the same request to complete. Likewise, keeping the link type the same but increasing the underlying engine speed of the z/OS system can result in more cost because for the same service time the faster processor could have been executing more instructions. . Table 2 lists the host effect (cost) when the primary application is involved in data sharing. The numbers in the chart below are based on 9 CF ops/Mi - may be scaled linearly for other rates.

Table 2. Coupling Technology versus Host Processor Speed

Host CF	z890	z990	z9 BC	z9 EC	z10 BC	z10 EC	z196
z890 ISC	13%	15%	16%	17%	19%	21%	NA
z890 ICB	9%	10%	10%	11%	12%	13%	NA
z990 ISC	13%	14%	14%	15%	17%	19%	NA
z990 ICB	9%	9%	9%	10%	12%	13%	NA
z9 BC ISC	12%	13%	14%	15%	17%	19%	23%
z9 BC PSIFB 12X	NA	NA	NA	NA	13%	14%	16%
z9 BC ICB	8%	9%	9%	10%	11%	12%	NA
z9 EC ISC	12%	13%	13%	14%	16%	18%	22%
z9 EC PSIFB 12X	NA	NA	NA	NA	13%	14%	16%
z9 EC ICB	8%	8%	8%	9%	10%	11%	NA
z10 BC ISC	12%	13%	13%	14%	16%	18%	22%
z10 BC PSIFB 12X	NA	NA	11%	12%	13%	14%	15%
z10 BC ICB	8%	8%	8%	9%	10%	11%	NA
z10 EC ISC	11%	12%	12%	13%	15%	17%	22%
z10 EC PSIFB 12X	NA	NA	10%	11%	12%	13%	15%
z10 EC ICB	7%	7%	7%	8%	9%	10%	NA
z196 ISC	NA	NA	11%	12%	14%	16%	21%
z196 PSIFB 12X[7]	NA	NA	9%	10%	11%	12%	14%

For example, if you considered a 2 engine partition of a 2097-708 to be the equivalent of 1800 mips and that partition was doing heavy data sharing and sending 16,200 requests per second to the coupling facilities (16200/1800 = 9 CF ops per MIP), then the table estimates that the cost will be approximately 10% for z10 EC CF technology with ICB4 links. If this partition was only doing 8,100 requests per second (thus, 8100/1800 = 4.5 CF ops per MIP), then the 10% cost would be scaled down to 5%.   l. Resource sharing type workloads typically operate at a much lower request rate (for example 1 or 2 CF ops per MIP), and with less data being transported, so the cost for a resource sharing workload would typically be much less.

Link Utilization or Service Time?

While link utilization is useful to know, it should not be the only thing considered, or even the first, when considering a coupling facility link change or upgrade. The goal should not be to maximize utilization of the CF link, but rather to optimize the service time seen by your important applications. For example, if there were 3 z/OS systems sharing 2 ISC-3 links to the coupling facility, link utilization is the sum of the 3 utilizations. Let's say link utilization is 30%. Yes, you are getting good use out of that link, and it seems like you have plenty of capacity available. You do have more available bandwidth, but because some of SYSA's lock

---

[7] PSIFB 1X links would fall approximately halfway between PSIFB 12X and ISC links.

requests are converted to asynchronous due to encountering channel path busy, the service time to satisfy a lock request from SYSA's database application may degrade during peak hours.

Synchronous Conversion to Asynchronous

zOS keeps a table of CF service times for each structure it has access to, and depending on the estimated efficiency tradeoff, and the processor speed of the z/OS system, XES may convert any synchronous operation to asynchronous. This will happen most often to CF operations that
1) Are affected by a long distance effect
2) Are affected by the overhead of duplexing
3) Have large amounts of data transfer

And in later processors, the conversion threshold is lowered as the engine speed gets faster. This is because you don't want your z/OS running on a screaming fast z196 waiting (as z/OS does for synchronous requests) for slow links or old CF technology. This is also the reason for the Rule of Thumb to not let the CF technology be much slower than the z/OS technology, because even though the CF may be less than 50% utilized, delay at the CF due to slow engine speed is going to be paid for by more expensive cycles at the z/OS end of the link. The heuristic conversion algorithm limits this, but service time for a request executed asynchronously is always longer than for a request executed synchronously, and that longer service time is what the application sees.

Should shared ICB4 links be shared in InfiniBand?

You can replace ICB4 links one for one with InfiniBand links, and have the InfiniBand channel paths also shared, just as they were in ICB4, but this should be reconsidered, especially for links carrying production workload. ICB4 channel paths may have been configured to be shared at a time when the sysplex had just been set up, when the sysplex-exploiting workload was much less, and when fewer ICB4 adapters could be installed because of cost considerations. Over time, the assumptions justifying shared channel paths may have changed, and should be reviewed at this time. This is particularly true for the cost consideration. ICB4 links are come in billable increments of one, and each can carry only one channel path, but one billable InfiniBand unit (a 2 port adapter) can carry up to 16 channel paths, so there is less need to share.

Consider this example. Suppose there are two ICB4 links running between CEC1 and CEC2, and there are three z/OS systems on CEC1 sharing those two ICB4 channel paths going to the coupling facility on CEC2. This is a configuration like that shown in Figure 2. Each system thinks it has 2 channel paths (14 subchannels), but in fact, the subchannels are shared, and when there is contention, systems will start seeing path busy conditions.

If you replaced the two ICB4 adapters with two InfiniBand adapters, you would then have capacity available on those links to run up to 32 channel paths (16 recommended) between CEC1 and CEC2, so it may no longer be necessary to

10

share those two channel paths between the three systems. Each z/OS system may continue with two channel paths to the coupling facility, just as they had before, but now those would be unshared, so there would be 6 active channel paths carrying CF requests between CEC1 and CEC2, no path busy conditions, and you still won't have used all of the capacity of the 2 InfiniBand adapters.

Appendix A: Defining Shared Channels in zCP3000

Clarify the Current Link Configuration

It is sometimes the case that there are several key pieces of information about the sysplex link configuration which cannot be derived from SMF data, and it is important to add this information to the model so that comparisons can be made in CP mode.

- In some cases, the channel type is entirely missing from the SMF 74-4 records.
- Even when the basic CF link type is known to SMF (ie, ISC or ICB), sometimes the level cannot be discerned. For example, System z9 supports both ICB-3 and ICB-4 links, but in SMF, it is simply "ICB". You must update this yourself.
- Any sharing information, either for shared channel paths (chpids) or shared channel adapter is not known to SMF. You must define this relationship yourself based on information from the customer.

Important: the CF link type in PA mode describes what is currently installed. Changing the link type in PA mode *will not* change the service time characteristics of the work. To see the effect of a CF link type change, define the current environment and do comparisons in CP mode.

Shared Channel Support in zCP3000

zCP3000 support for shared CF channel paths both for non-InfiniBand (ISC and ICB type links) and for InfiniBand type links became available in October 2008. Support for InfiniBand shared host channel adapters became available in October 2010. It does not require a new level of CP2KEXTR, however it is important to note that ***shared channels and shared adapters are *not* automatically recognized and set up by zCP3000***. Shared channel paths and shared host channel adapters must be manually specified, because there is not sufficient information in the SMF records to determine the configuration. Once sharing had been defined in zCP3000, however, the sharing status is reflected in analysis reports, and the configuration of the shared channels is saved when the model is saved.

Locating Sharable CF Links in zCP3000

The customer needs to identify which links are shared. Once you know that, then you have to locate which links are being referred to. This may be stated in a narrative fashion, for example "SYSA and SYSB share links to CF1", but you have to figure out which partitions hold these three elements, and to confirm that they could share channel paths or adapters.

12

There are some rules to help you figure this out. Some of this is obvious, but I'll state it anyway.

- **Only 2 CECs can share a channel path**, so all z/OS sysplex members and coupling facilities using CF links on a shared channel must reside on the same pair of CECs. For example, SYSA on CEC1 cannot share a link with SYSB on CEC2 to CF1 which resides on the standalone CECCF1.
- **The channel path can be only one specific type (ie, ISC-3 or ICB-4)**, so the z/OS sysplex members and coupling facilities can share the link only if they all think it is the same type of link. If the link from SYSA on CEC1 to the CF1 coupling facility is an ICB-3 link, and the link from SYSB to CF1 is an ISC-3 link, then they are using two separate physical channels.
- **z/OS systems must talk to the same CF in the same Sysplex** in order to share a channel path. This is true for all link types.
- **InfiniBand type HCA2-O link adapters (but not channel paths) can be shared by SYSes and CFs in different sysplexes.** But it is still true that an InfiniBand channel path, like ISC-3 and ICB-4, can only be shared by z/OS systems going to the same CF.

Use both the zCP3000 physical view of the configuration and the logical view of the sysplexes, combined with input about sharing from the customer, to help you locate sysplex objects on CEC pairs that can potentially share a link.

Testcase Configuration

To illustrate how to define shared CF links in zCP3000, we'll use a testcase based on real data, where there are three sysplexes:

- A production sysplex WPLEX, with 6 z/OS members and 2 CFs.
- A production sysplex TPLEX, with 6 z/OS members and 2 CFs.
- A test/development sysplex DEVPLEX with 5 z/OS member and 2 CFs.

These 17 z/OS images and 6 coupling facilities run on 6 mainframes along with 8 other (non-sysplex) z/OS or z/VM images. EDF input files were created for 19 of the 25 active partitions (excluding CFs). Figure 7. zCP3000 Main View of the Physical Configuration below shows the zCP3000 physical view of the configuration. In zCP3000, connections can be shown by single clicking on a CF or SYS. In Figure 7, the WSPACF01 has been single clicked, making it red and

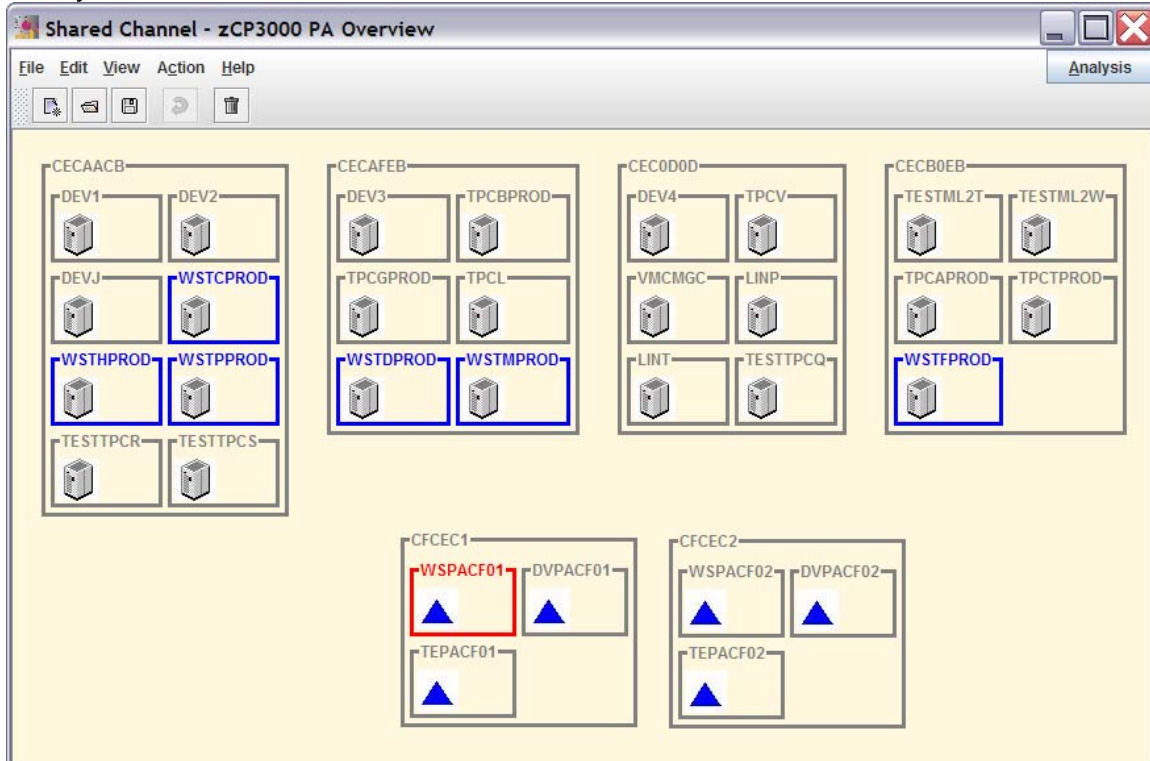the systems that it connects to, outlined in blue.

Figure 7. zCP3000 Main View of the Physical Configuration.

Figure 8 shows the zCP3000 Sysplex Logical view of the same configuration. In the sysplex logical view, each sysplex is shown with the z/OS systems and coupling facilities that belong to it. You can see the logical view by selecting the menu option View->Logical Sysplex.
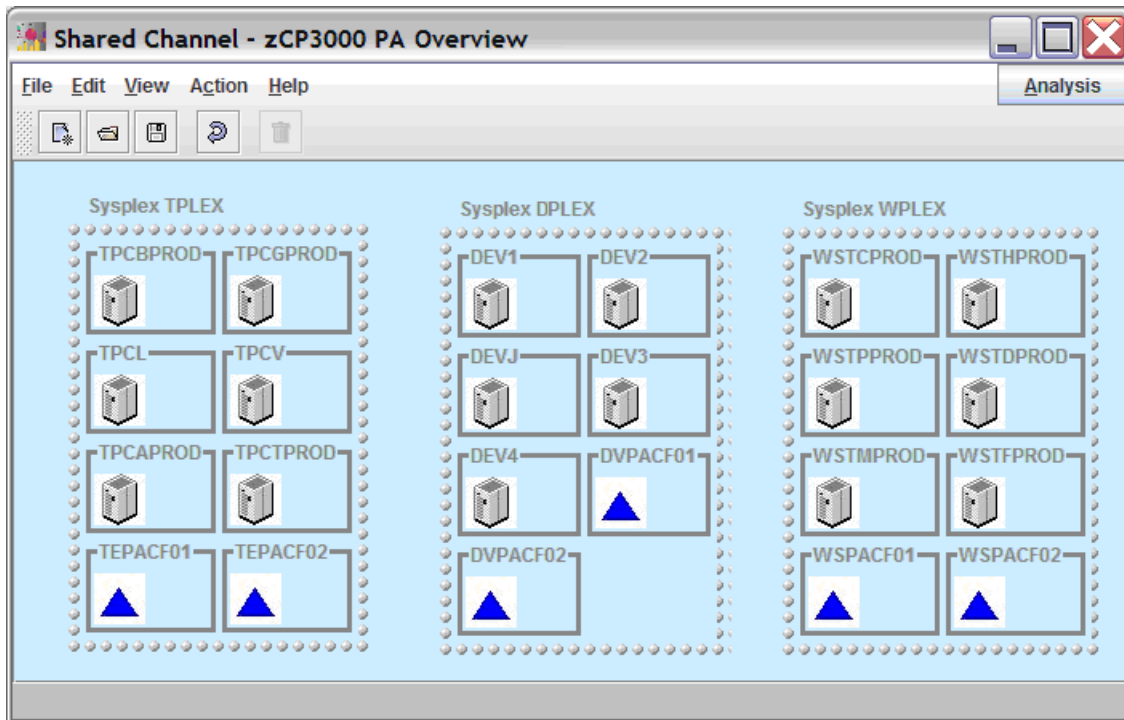
Figure 8. Logical Sysplex Configuration of 2 production + 1 test sysplexes.

The Logical Sysplex view of the configuration is useful in that it lets you see that there are 3 sysplexes in this complex, and that shared InfiniBand channels could be of benefit to this customer in a way that ISC or ICB links could not, but you will need to refer to the physical configuration to determine which physical channel paths can be shared.

IOCDS Information

IOCDS decks were collected from the 4 of the 6 mainframe machines[8]. From the IOCDS deck from CECAACB, it was determined that

- 4 ICB channels are defined as shared between WSTC, WSTH, and WSTP
- 4 ISC channels are defined as shared between DEV1, DEV2, and DEVJ.
- 4 ICB channels are defined as shared between DEV1, DEV2, and DEVJ.
- 4 ISC channels are defined as shared between TESTTPCR and TESTTPCS, but information was not collected from those z/OS systems so nothing further is known about them.

---

[8] IOCDS Statements are listed in Appendix B.

15

Scenario 1 : Share WPLEX Channel Paths from CECAACB

From Figure 8, it can be seen that the WPLEX sysplex consists of 6 z/OS members and 2 coupling facilities. You will need to edit each CF that has shared paths coming to it.

On the zCP3000 physical view (View->Physical), single-click WSPACF01 to see its connections. There are 3 z/OS images on CECAACB that are connected to WSPACF01. These three z/OS images are listed by their partition names, and the IOCDS tells us that they can share up to 4 ICB type cf link(s). These 4 ICB links may all go to WSPACF01 or some may go to other coupling facilities.

Double-click WSPACF01 to look at it in more detail. As shown in Figure 9, this window lists the z/OS systems that had an active link to this coupling facility and which also had EDF data loaded into the model.

The link table lists the CEC id, the partition name, the sysid, how many links there are, and what type. The partition name will match the names seen in the zCP3000 physical view. From the IOCDS, we know that the partitions for WSTC, WSTH, and WSTP share 4 ICB type channel paths, and it is probable[9] that the red-circled links listed in the table are 2 of those 4 shared channel paths, so we are going to change the link definition to recognize that. Note that you can see from the table that WSTC, WSTH, and WSTP are not currently defined with shared channels because the utilization is different for all three. The link utilization will be the sum of the 3 when they are defined as shared.

---

[9] There may be more ICB channel paths defined than are actually used. If that is true, you will not be able determine from the collected data which channel paths were actually used.

Figure 9. Detail View of WSPACF01 coupling facility.

Double-click the WSTC row (or any of the 3 that are going to be defined as shared, it doesn't matter which), to edit the link from WSPACF01 to WSTC, as shown in Figure 10.

Figure 10. Detail View of the WSPACF01-WSTC link.

The Shared Links section, in the middle-right of the CF Link window, lists the systems that could possibly share these channel paths[10], and whether it is currently defined as shared. Click the checkboxes for WSTP and WSTH. As you do so, you will notice the subchannel utilization increasing, as it reflects the summed utilizations for all z/OS systems using these channel paths. Hit the apply button to save these changes in the model. The sharing changes will automatically be applied to the other two CF links. On the CF window, the utilizations for the WSTC, WSTP, and WSTH systems will now be the summed utilizations of the shared channel paths, as shown in Figure 11.

---

[10] ICB links can only share channel paths, as indicated by the "Logical" sharing radio button. The "Physical" sharing radio button is enabled only for InfiniBand links.

Figure 11. WSPACF01 with shared channel paths.

Scenario 2: Upgrade from ICB-4 to InfiniBand

In this scenario, we will upgrade some shared ICB-4 channel paths to InfiniBand. As seen in Figure 12, the two shared ICB-4 links from WSTD and WSTM system on the CECAFEB machine to the WSPACF01 coupling facility on the CFCEC1 machine will be upgraded to InfiniBand links, and we will consider whether they should be shared again or not.
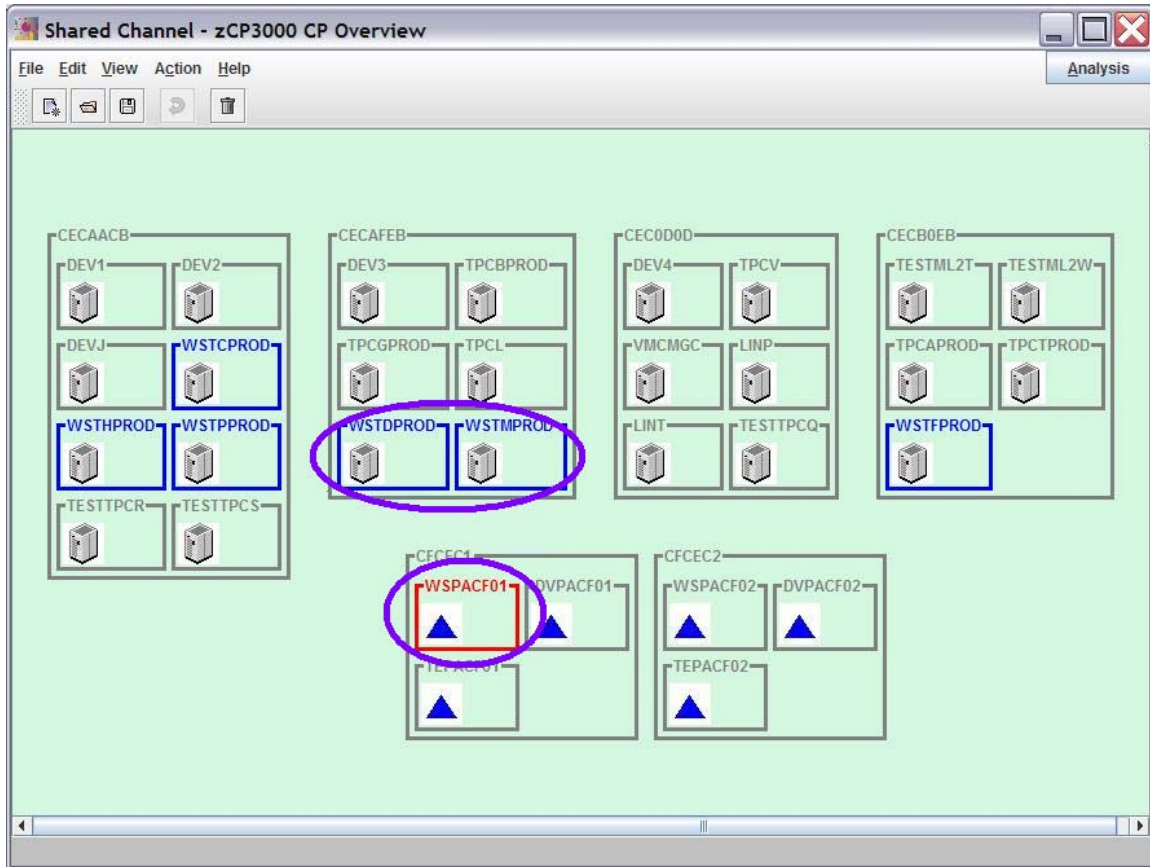

Figure 12. zCP3000 Capacity Planning Mode View

Changing a link type causes shared links to become unshared, because the link type changes are done one at a time, and as soon as a link has become InfiniBand, it cannot share the chpid with an ICB-4 type link.

That means that zCP3000 will not generate before and after reports showing both a link type change *AND* showing the effects of sharing. You will have to show the effect of sharing *AFTER* all of the links have been changed to InfiniBand.

Go to Capacity Planning (CP) Mode

Performance Analysis (PA) mode of zCP3000 only reflects the sysplex configuration that is currently installed. To project the effect of changes to a CF link, you must be in CP mode, which selects a single study interval and projects

20

growth and other changes based on that one interval. To convert to CP mode, on the zCP3000 physical view, select Action->Go To Capacity Planning Mode. The growth rate applies only to CPUs, so you can specify that or not, it doesn't matter for the CF link upgrade.

Change the WSPACF01-WSTD Link Type

When in CP mode, double click WSPACF01 to see the CF detail, and then double-click the WSTD row to edit the link from WSPACF01 to WSTD. WSTD and WSTM share 2 ICB-4 channel paths (chpids) to the WSPACF01, so in the CF window, WSTD and WSTM have the same link utilization (which is 2.8%).

In CP mode, the edit window gives you the choice to "Upgrade Link" or "Redefine Current Link".
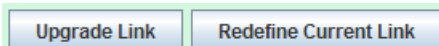
These are explained below.

Upgrade Link Button vs Redefine Current Link Button

Upgrade Link means that the link is correct as it is currently defined (ie, it really is an ICB4 link) and that you want to upgrade to something else. Redefine Current Link says that it isn't really ICB4 right now, and that you want to correct the definition. *The difference is that service time and host cost *will not change* if you redefine the link, but they will change for a link upgrade.*

Figure 13 shows the appearance of the window after the "Upgrade Link" button has been pressed and the new linktype has been selected. If the ICB4 channel path was shared with WSTM, then the Shared Links table would have shown this before you hit the "Upgrade Link" button. After hitting the "Upgrade Link" button, any shared links are deleted from the sharing table, the link utilization for the WSTD-WSPACF01 link will change to show the unshared utilization and the shared links table is disabled for updates.[11] This change will be reflected in reports generated from this window via the "Analysis" button, and will be made permanent in the data model if you hit the apply button.

---

[11] New function in the zCP3000 release on Dec 1st, 2010

21

Figure 13. Upgrade WSPACF01-WSTD from ICB4 to InfiniBand.

When changing from ICB4 to InfiniBand, the CF utilization will not change as a result of the link type change, however the sysid utilization may change slightly if there is a change in how many requests are serviced synchronously. The subchannel utilization will change both from the link type change and from not being a shared link any more.

Hit the Apply button to save this change. When you change the CF link type, the sysplex cost on the z/OS may change slightly, depending on how many synchronous requests become asynchronous or visa versa. When you apply the link type change, you will be asked what to do with the change in z/OS overhead. This dialog is shown in Figure 14.

Applying Host Cost Changes

The sysplex cost is the time spent on a z/OS to actually issue a CF request, and the time spent waiting for a synchronous request to complete. If the request is

22

asynchronous, the path length to issue the instruction is slightly longer, but the z/OS system will not wait for the request to complete, so it tends to result in slightly lower sysplex cost. If the request was synchronous and remains synchronous, then if the service time increases, overhead will increase because z/OS is waiting longer for the synchronous request to complete. This tends to be a very small number for all but significantly busy data sharing systems. Select the workloads that are most closely related to the sysplex workload. These workloads will have a slight change (an increase or decrease) to their cpu time used.
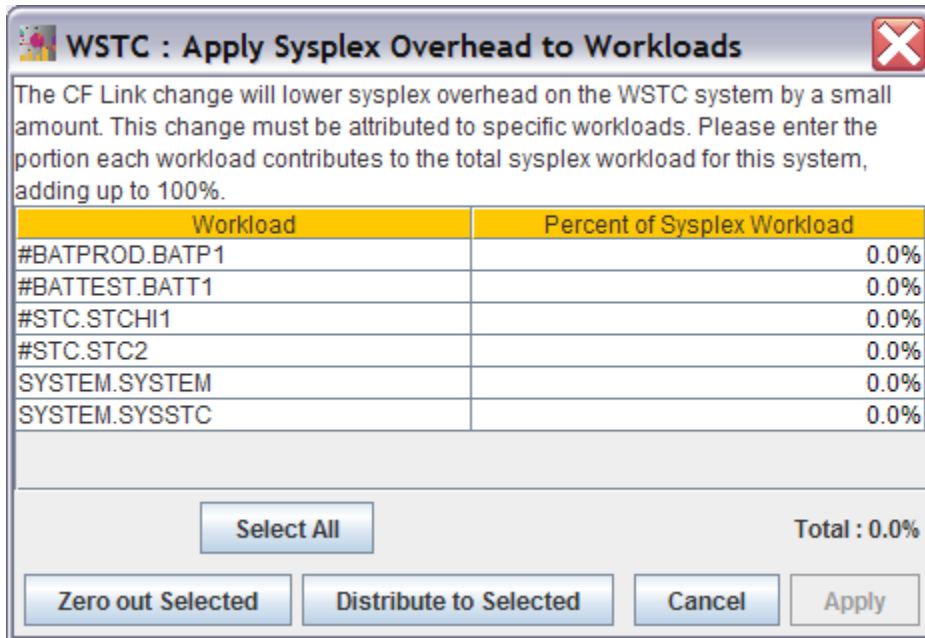

Figure 14. Dialog to Adjust Sysplex Overhead.

Please note that the sysid utilization row of the CF link window shows the currently logical utilization of the WSTD system, *which already includes sysplex cost since it is already running requests over the ICB-4 links*. The sysplex overhead referred to in Figure 14 refers only to the *change* in sysplex cost due to the link type change. It does not refer to the total sysplex cost from the current sysplex workload. By comparison, Table 2. Coupling Technology versus Host Processor Speed" on page 2 shows the estimated cost for a busy data sharing system.

23

Scenario 3: Sharing the InfiniBand Host Channel Adapter

Each InfiniBand adapter can support up to 16 chpids (a maximum of 8 is recommended), enabling the same underlying hardware to support multiple communication paths, to multiple coupling facilities, and to one or more. This capability is only supported by InfiniBand links, so for this scenario, assume that all of the ICB4 channel paths in all three sysplexes (WPLEX, PPLEX, and TPLEX) have been replaced, one for one, by InfiniBand chpids, and that we are going to configure the InfiniBand links so that the links but not the channel paths are shared.

As with shared channel paths, each CF involved has to be edited. As shown in Figure 15, the current level of channel path sharing is shown when the link from WSPACF01 to the WSTC system is double clicked. The minimum number of recommended InfiniBand adapters (2 for RAS reasons, otherwise 1 per 8 chpids) is shown in the "Minimum # FCs" field. Although three systems (WSTC, WSTP, and WSTH) had shared the ICB4 channel paths in the previous scenario, the process of changing the links to InfiniBand caused them to become unshared in zCP3000, so they are shown in Figure 15 as logically unshared. Logical sharing refers to shared channel paths, while physical sharing refers to shared adapters.
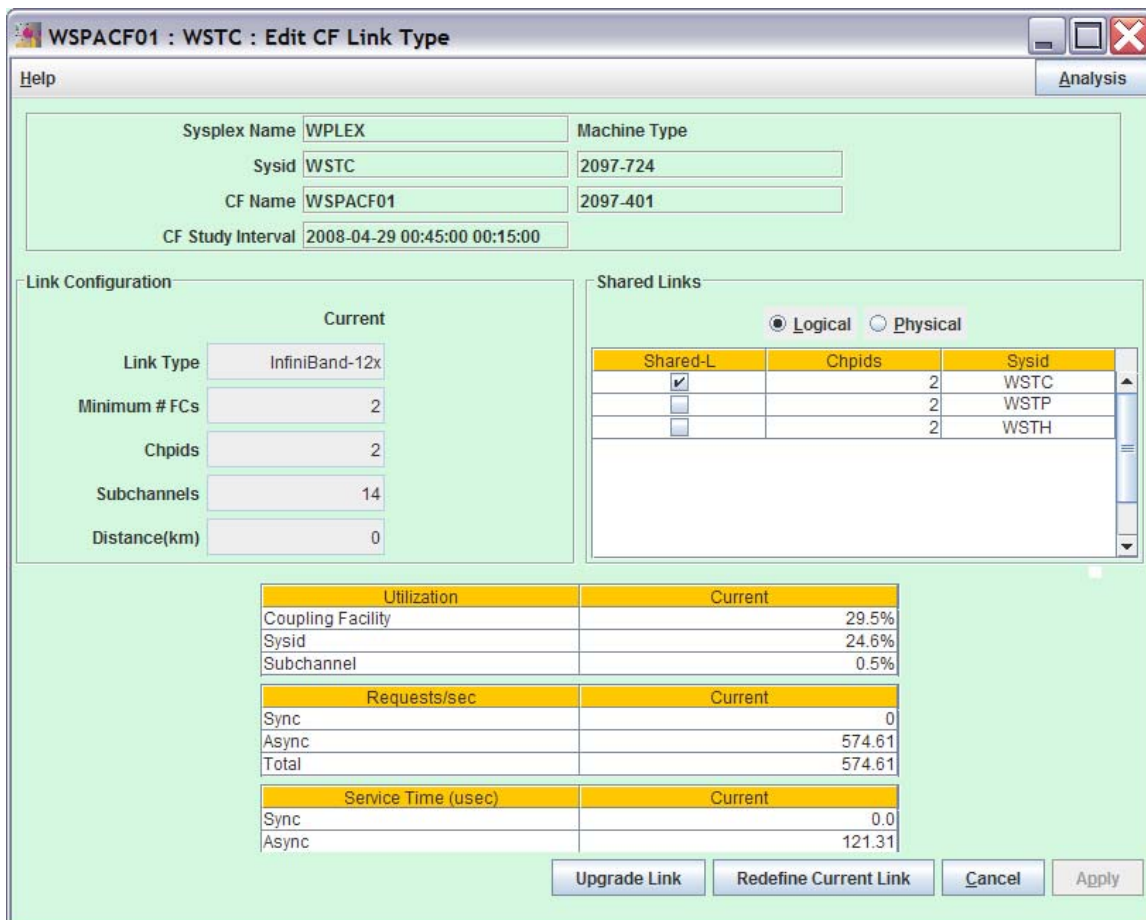
Figure 15 InfiniBand Link from WSPACF01 to WSTC

The "Physical" radio button control in the Shared Links box of the CF Link window is enabled only for InfiniBand links. Selecting the "Physical" radio button changes the view so that it is looking at shared adapters instead of shared links. As shown in Figure 16, a larger list of links is presented in the shared links box. The additional rows are channel paths that could potentially share the InfiniBand adapter with the 2 channel paths running from WSPACF01 to WSTC. The WSTP and WSTH systems could potentially also share channel paths, as shown in Figure 15, but this would increase the link utilization and potentially result in increased occurrences of channel path busy, so since there is additional capacity available on the InfiniBand adapter, we can avoid having to share the WSTC channel paths with WSTP and WSTH. As additional links are selected to share the InfiniBand adapter, a count is kept of how many channel paths have been configured. If the chpid count exceeded 16, then the minimum number of FCs (feature codes) would be increased to 3.

Figure 16 Shared Adapter selection for WSPACF01-WSTC

The report shown in Figure 17 shows the link we are editing and any systems sharing the channel paths in blue, and any channel paths sharing the InfiniBand adapter in green. If the WSCH and WSTP had also been checked as logically shared (ie, they shared two channel paths with WSTC), then they would have been shaded blue in this report.
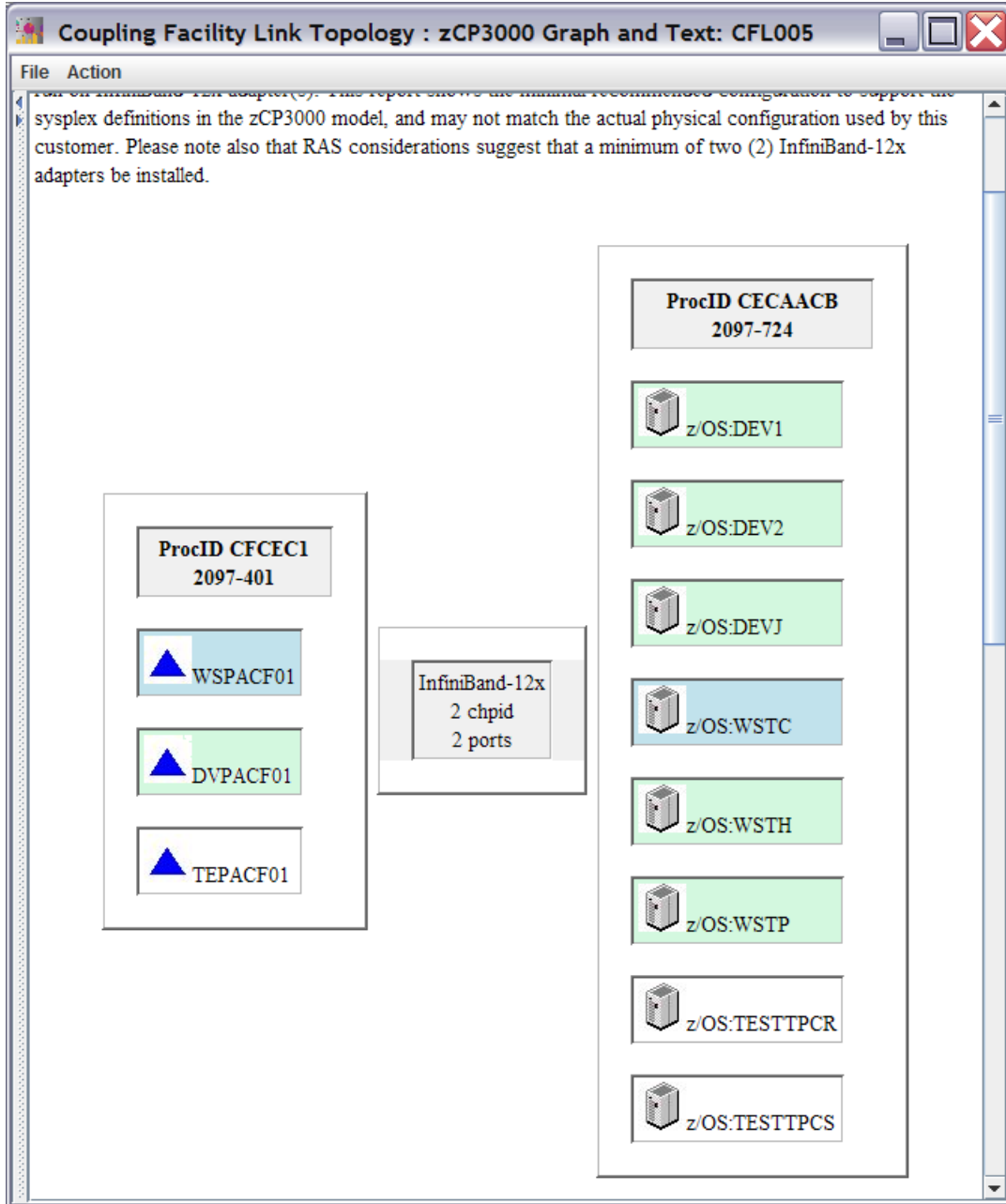

Figure 17 Shared InfiniBand Adapter on the CF Link Topology Report.

Appendix B: Sample IOCDS Statements

CECAACB CF Link CHPIDs

```
CHPID PATH=(CSS(0),EC),SHARED,PARTITION=((TESTTPCR,TESTTPCS),(=)),PCHID=380,TYPE=CFP
CHPID PATH=(CSS(0),ED),SHARED,PARTITION=((TESTTPCR,TESTTPCS),(=)),PCHID=388,TYPE=CFP
CHPID PATH=(CSS(0),EE),SHARED,PARTITION=((TESTTPCR,TESTTPCS),(=)),PCHID=381,TYPE=CFP
CHPID PATH=(CSS(0),EF),SHARED,PARTITION=((TESTTPCR,TESTTPCS),(=)),PCHID=389,TYPE=CFP
CHPID PATH=(CSS(0),F0),SHARED,PARTITION=((TPCW),(=)),PCHID=100,TYPE=CFP
CHPID PATH=(CSS(0),F1),SHARED,PARTITION=((TPCW),(=)),PCHID=108,TYPE=CFP
CHPID PATH=(CSS(0),F2),SHARED,PARTITION=((TPCW),(=)),PCHID=190,TYPE=CFP
CHPID PATH=(CSS(0),F3),SHARED,PARTITION=((TPCW),(=)),PCHID=198,TYPE=CFP
CHPID PATH=(CSS(0),F4),SHARED,PARTITION=((DEVJ,DEV1,DEV2),(=)),PCHID=101,TYPE=CFP
CHPID PATH=(CSS(0),F5),SHARED,PARTITION=((DEVJ,DEV1,DEV2),(=)),PCHID=109,TYPE=CFP
CHPID PATH=(CSS(0),F6),SHARED,PARTITION=((DEVJ,DEV1,DEV2),(=)),PCHID=191,TYPE=CFP
CHPID PATH=(CSS(0),F7),SHARED,PARTITION=((DEVJ,DEV1,DEV2),(=)),PCHID=199,TYPE=CFP
CHPID PATH=(CSS(0),F8),SHARED,PARTITION=((DEVJ,DEV1,DEV2),(=)),PCHID=280,TYPE=CBP
CHPID PATH=(CSS(0),F9),SHARED,PARTITION=((DEVJ,DEV1,DEV2),(=)),PCHID=290,TYPE=CBP
CHPID PATH=(CSS(0),FA),SHARED,PARTITION=((DEVJ,DEV1,DEV2),(=)),PCHID=2A0,TYPE=CBP
CHPID PATH=(CSS(0),FB),SHARED,PARTITION=((DEVJ,DEV1,DEV2),(=)),PCHID=2B0,TYPE=CBP
CHPID PATH=(CSS(0),FC),SHARED,PARTITION=((WSTCPROD,WSTHPROD,WSTPPROD),(=)),PCHID=281,TYPE=CBP
CHPID PATH=(CSS(0),FD),SHARED,PARTITION=((WSTCPROD,WSTHPROD,WSTPPROD),(=)),PCHID=291,TYPE=CBP
CHPID PATH=(CSS(0),FE),SHARED,PARTITION=((WSTCPROD,WSTHPROD,WSTPPROD),(=)),PCHID=2A1,TYPE=CBP
CHPID PATH=(CSS(0),FF),SHARED,PARTITION=((WSTCPROD,WSTHPROD,WSTPPROD),(=)),PCHID=2B1,TYPE=CBP
```

CECAFEB CF Link CHPIDs

```
CHPID PATH=(CSS(0),F0),SHARED,PARTITION=((DEV3),(=)),PCHID=110,TYPE=CFP
CHPID PATH=(CSS(0),F1),SHARED,PARTITION=((DEV3),(=)),PCHID=118,TYPE=CFP
CHPID PATH=(CSS(0),F2),SHARED,PARTITION=((DEV3),(=)),PCHID=180,TYPE=CFP
CHPID PATH=(CSS(0),F3),SHARED,PARTITION=((DEV3),(=)),PCHID=188,TYPE=CFP
CHPID PATH=(CSS(0),F4),SHARED,PARTITION=((DEV3),(=)),PCHID=111,TYPE=CFP
CHPID PATH=(CSS(0),F5),SHARED,PARTITION=((DEV3),(=)),PCHID=119,TYPE=CFP
CHPID PATH=(CSS(0),F6),SHARED,PARTITION=((DEV3),(=)),PCHID=181,TYPE=CFP
CHPID PATH=(CSS(0),F7),SHARED,PARTITION=((DEV3),(=)),PCHID=189,TYPE=CFP
CHPID PATH=(CSS(0),F8),SHARED,PARTITION=((TPCBPROD,TPCGPROD,TPCL),(=)),PCHID=280,TYPE=CBP
CHPID PATH=(CSS(0),F9),SHARED,PARTITION=((TPCBPROD,TPCGPROD,TPCL),(=)),PCHID=290,TYPE=CBP
CHPID PATH=(CSS(0),FA),SHARED,PARTITION=((TPCBPROD,TPCGPROD,TPCL),(=)),PCHID=2A0,TYPE=CBP
CHPID PATH=(CSS(0),FB),SHARED,PARTITION=((TPCBPROD,TPCGPROD,TPCL),(=)),PCHID=2B0,TYPE=CBP
CHPID PATH=(CSS(0),FC),SHARED,PARTITION=((WSTDPROD,WSTMPROD),(=)),PCHID=281,TYPE=CBP
CHPID PATH=(CSS(0),FD),SHARED,PARTITION=((WSTDPROD,WSTMPROD),(=)),PCHID=291,TYPE=CBP
CHPID PATH=(CSS(0),FE),SHARED,PARTITION=((WSTDPROD,WSTMPROD),(=)),PCHID=2A1,TYPE=CBP
CHPID PATH=(CSS(0),FF),SHARED,PARTITION=((WSTDPROD,WSTMPROD),(=)),PCHID=2B1,TYPE=CBP
```

CEC0D0D CF Link CHPIDs

```
CHPID PATH=(CSS(0),EC),SHARED,PARTITION=((TESTTPCQ),(=)),PCHID=118,TYPE=CFP
CHPID PATH=(CSS(0),ED),SHARED,PARTITION=((TESTTPCQ),(=)),PCHID=188,TYPE=CFP
CHPID PATH=(CSS(0),EE),SHARED,PARTITION=((TESTTPCQ),(=)),PCHID=308,TYPE=CFP
CHPID PATH=(CSS(0),EF),SHARED,PARTITION=((TESTTPCQ),(=)),PCHID=398,TYPE=CFP
CHPID PATH=(CSS(0),F0),SHARED,PARTITION=((DEV4),(=)),PCHID=110,TYPE=CFP
CHPID PATH=(CSS(0),F1),SHARED,PARTITION=((DEV4),(=)),PCHID=180,TYPE=CFP
CHPID PATH=(CSS(0),F2),SHARED,PARTITION=((DEV4),(=)),PCHID=300,TYPE=CFP
CHPID PATH=(CSS(0),F3),SHARED,PARTITION=((DEV4),(=)),PCHID=390,TYPE=CFP
CHPID PATH=(CSS(0),F8),SHARED,PARTITION=((TPCV),(=)),PCHID=2A0,TYPE=CBP
CHPID PATH=(CSS(0),F9),SHARED,PARTITION=((TPCV),(=)),PCHID=2B0,TYPE=CBP
CHPID PATH=(CSS(0),FA),SHARED,PARTITION=((TPCV),(=)),PCHID=2A1,TYPE=CBP
CHPID PATH=(CSS(0),FB),SHARED,PARTITION=((TPCV),(=)),PCHID=2B1,TYPE=CBP
```

CECB0EB CF Link CHPIDs

```
CHPID PATH=(CSS(0),F8),SHARED,PARTITION=((TPCAPROD,TPCTPROD),(=)),PCHID=280,TYPE=CBP
CHPID PATH=(CSS(0),F9),SHARED,PARTITION=((TPCAPROD,TPCTPROD),(=)),PCHID=290,TYPE=CBP
CHPID PATH=(CSS(0),FA),SHARED,PARTITION=((TPCAPROD,TPCTPROD),(=)),PCHID=2A0,TYPE=CBP
CHPID PATH=(CSS(0),FB),SHARED,PARTITION=((TPCAPROD,TPCTPROD),(=)),PCHID=2B0,TYPE=CBP
CHPID PATH=(CSS(0),FC),SHARED,PARTITION=((WSTFPROD),(=)),PCHID=281,TYPE=CBP
CHPID PATH=(CSS(0),FD),SHARED,PARTITION=((WSTFPROD),(=)),PCHID=291,TYPE=CBP
CHPID PATH=(CSS(0),FE),SHARED,PARTITION=((WSTFPROD),(=)),PCHID=2A1,TYPE=CBP
CHPID PATH=(CSS(0),FF),SHARED,PARTITION=((WSTFPROD),(=)),PCHID=2B1,TYPE=CBP
```