IBM z Systems Masters Series

Solving Sysplex Questions with zCP3000

Gretchen Frye

Session ID: zCP3K_plex

June 2015, Gaithersburg, MD

Trademarks

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

AIX*	IBM i*	POWER7*	PureSystems	Tivoli*
BladeCenter*	IBM logo*	Power Systems	Storwize*	WebSphere*
DB2*	Informix*	PowerVM	System Storage*	zEnterprise*
Easy TIER*	PartnerWorld *	PureApplication	System x*	
IBM*	Power*	PureFlex	System z*	

* Registered trademarks of IBM Corporation

The following are trademarks or registered trademarks of other companies.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.
IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency which is now part of the Office of Government Commerce.
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.
Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.
Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.
Windows Server and the Windows logo are trademarks of the Microsoft group of countries.
ITIL is a registered trademark, and a registered community trademark of the Office of Government Commerce, and is registered in the U.S. Patent and Trademark Office.
UNIX is a registered trademark of The Open Group in the United States and other countries.
Java and all Java based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.
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.
Linear Tape-Open, LTO, the LTO Logo, Ultrium, and the Ultrium logo are trademarks of HP, IBM Corp. and Quantum in the U.S. and other countries.

* Other product and service names might be trademarks of IBM or other companies.

Notes:

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

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

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

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

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

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

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

Agenda

Use zCP3000 to look at sysplex questions:

- Collect the right data for a study
- Understand it, and identify current problem areas
- Model a new hardware configuration
- Maybe shared ICF and Thin Interrupt?
- z13 Link Migration Considerations
- Link consolidation on InfiniBand and ICA
- Sysplex Aggregation Pricing eligibility

This session is *not* covering:

- Sysplex concepts -> see the replay of Gene Sale's excellent session
- zCP3000 basic knowledge – please see the last page for reference urls

Collecting Model Input Data

Plan the Study

- Any current performance concerns?
- Which CFs matter, and which don't?
- Should the study include overview information?
- Are any partitions currently sharing channel paths?
- Are *all* of the mainframes (z/OS and CF) co-located? Is the distance changing?
- If there is a standalone CF, what type of machine is it on, what type of engines is the CF using and how many, and is it sharing?

Collecting Model Input Data

Capture the *right* period of time,

representative of what you want to do.

- Collect 1 week, maybe 2 if variability is not well understood.
- Assume 24x7 unless proven otherwise
- All members of “production” sysplexes
- If doing link consolidation, useful to include Test/Dev sysplexes.
- CPUMF (SMF 113) helps get the workload characterization right.
- All partitions that belong together as a production group
- Sysplex Aggregation Pricing reports require data for all sysplex members

Too Much Data!

2015 Techline testcase:

25 sysids @ 13.7MB average
342 MB of input data

- Use 64-bit CPSJAVA
- Collect one week, not two
- Shut down everything else.
- Increase max java heap size

  SET JAVA_OPTS="-Xms2G –Xmx6G“

  (in your C:\cpstools\zCP3000 64-bit dir, then cpstoolstart to start zCP3000)
- Once data loaded, delete I/O, then save 3PA
- Delete uninteresting intervals, then save 3PA
- Ask CPSTOOLS@US.IBM.COM for help

Agenda

Use zCP3000 to look at sysplex questions:

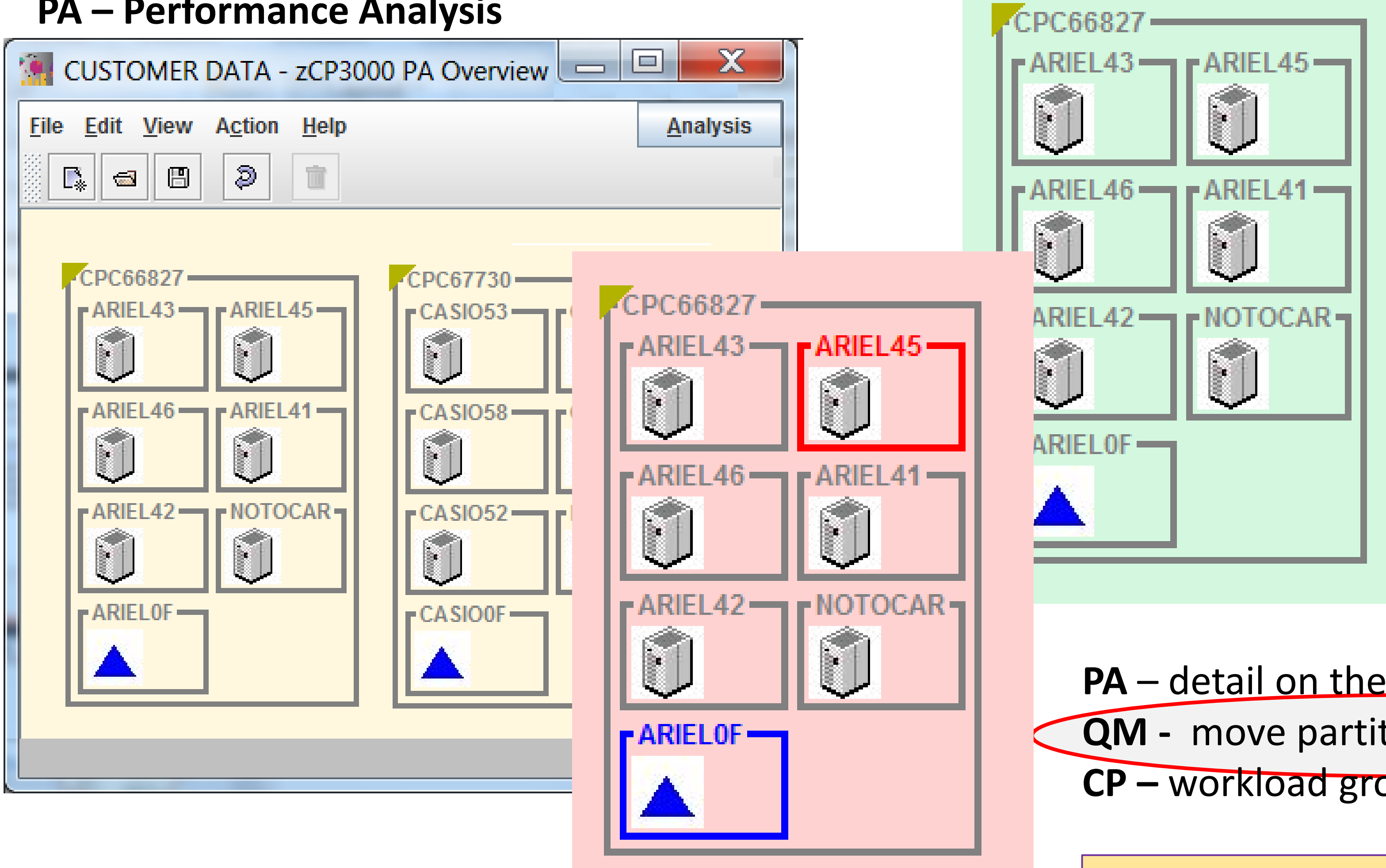
- Collect the right data for a study
- **Learn about the data, identify current problem areas**
- Modelling a new hardware configuration
- Using shared ICF and Thin Interrupt
- z13 Link Migration Considerations
- Link consolidation on InfiniBand and ICA
- Sysplex Aggregation Pricing eligibility

zCP3000 "modes"

IBM

PA – Performance Analysis

CP – Capacity Planning



QM – Quick Migration

PA – detail on the current, plus alt CPUs

QM - move partitions, change CPUs

CP – workload growth

As of Nov 2014, use QM mode, not CP mode for sysplex modelling.

Nov 2014 major changes in zCP3000

As of Nov 2014, use QM mode, not CP mode for sysplex

Why?

Heisenberg's uncertainty principle : you can measure this, or you can measure that, but not both at the same time.

Sysplex modelling function moves to QM mode, so that it can:

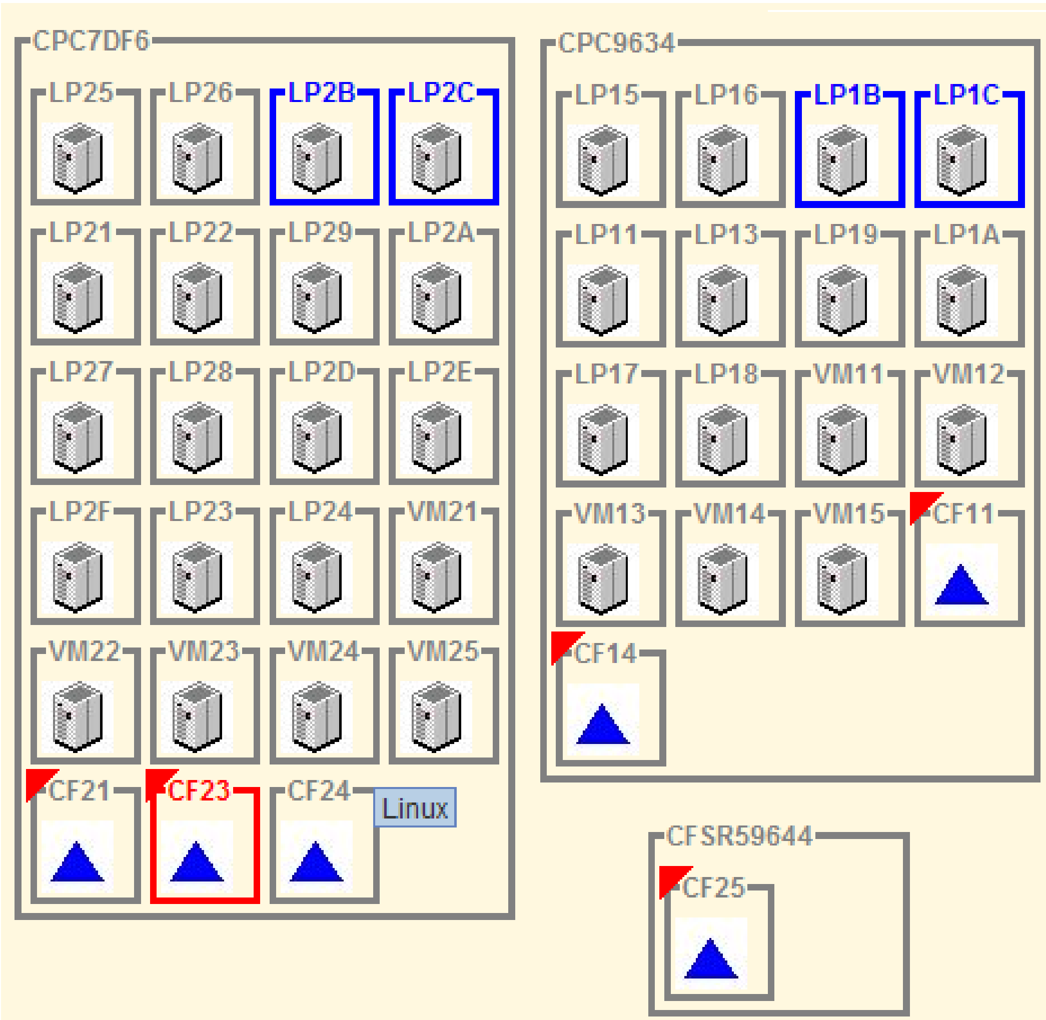
- Recalculate service time when the CF moves to a different speed engine.
- Change the dispatch mode to/from dedicated, dyndisp = Off/On/Thin
- **Can make multiple configuration changes and still see before vs after.**

Unchanged from CP mode:

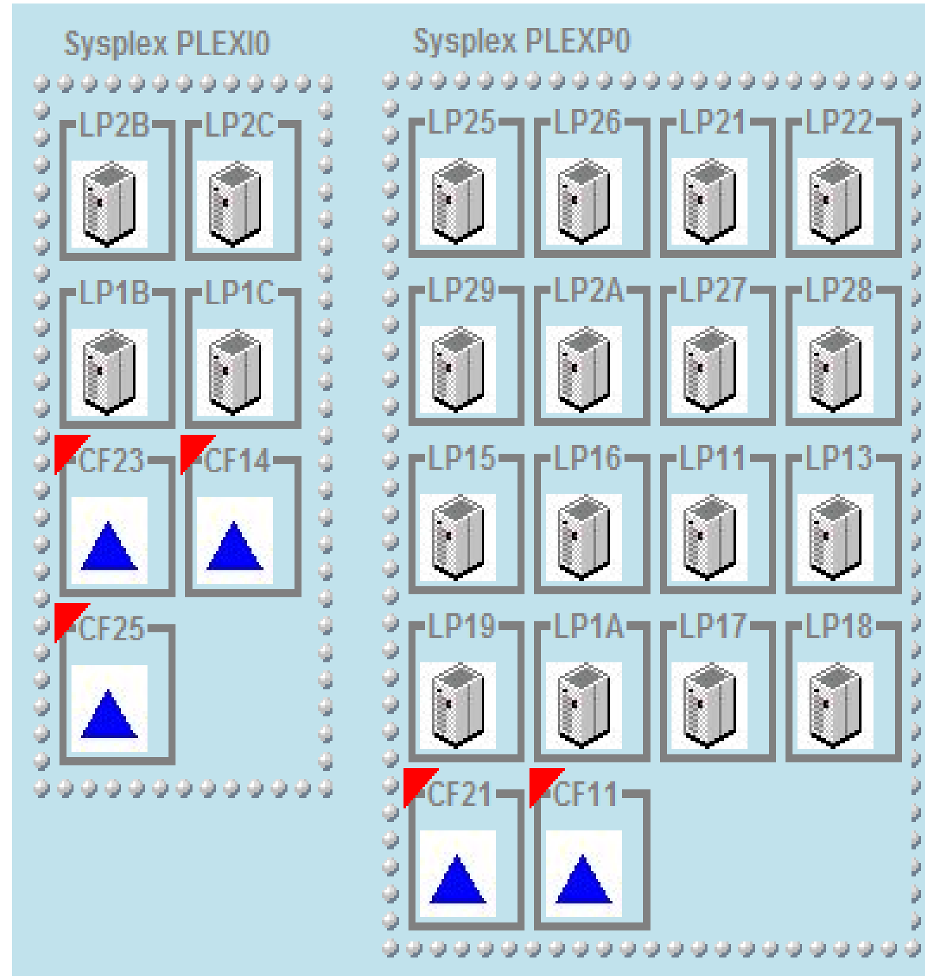
- Service time recalculation for a cf link type change or distance change.
- Projected service time when moving from dedicated to thin interrupt.

Understand the Collected Model Data

View->Physical

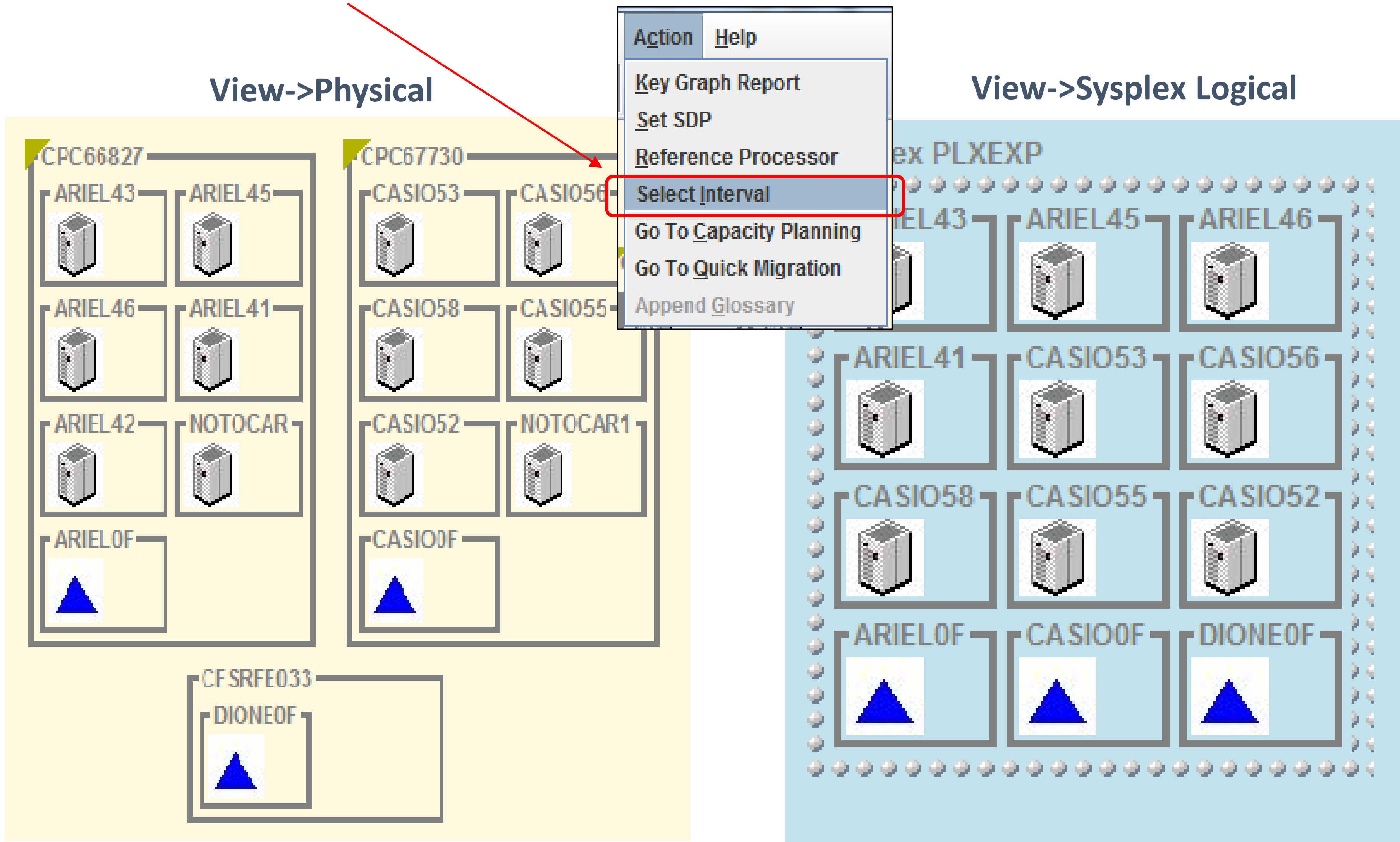


View->Sysplex Logical



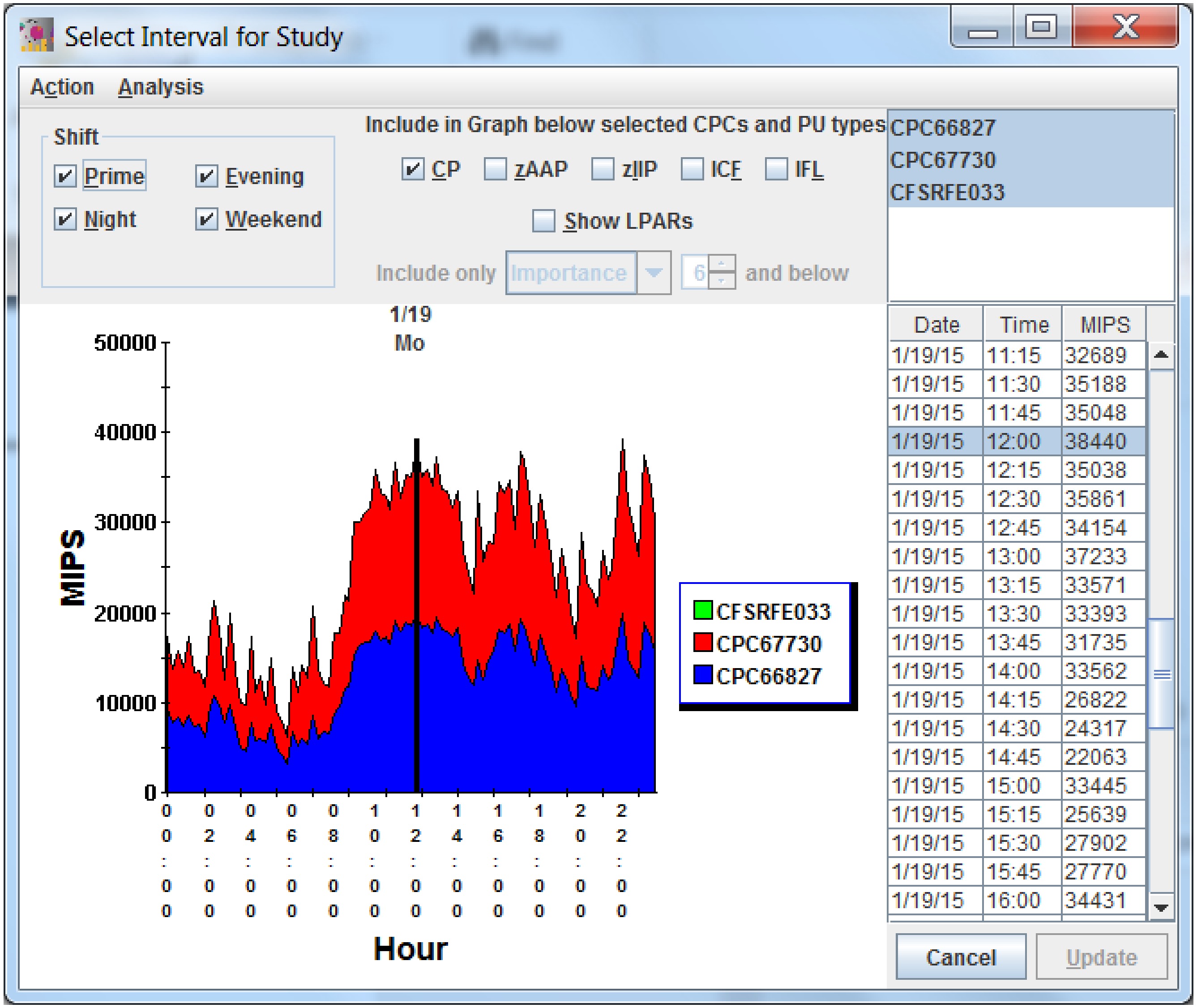
CF23 and its connections to PLEXIO sysplex members are highlighted at left.

Select a study interval



Select a study interval

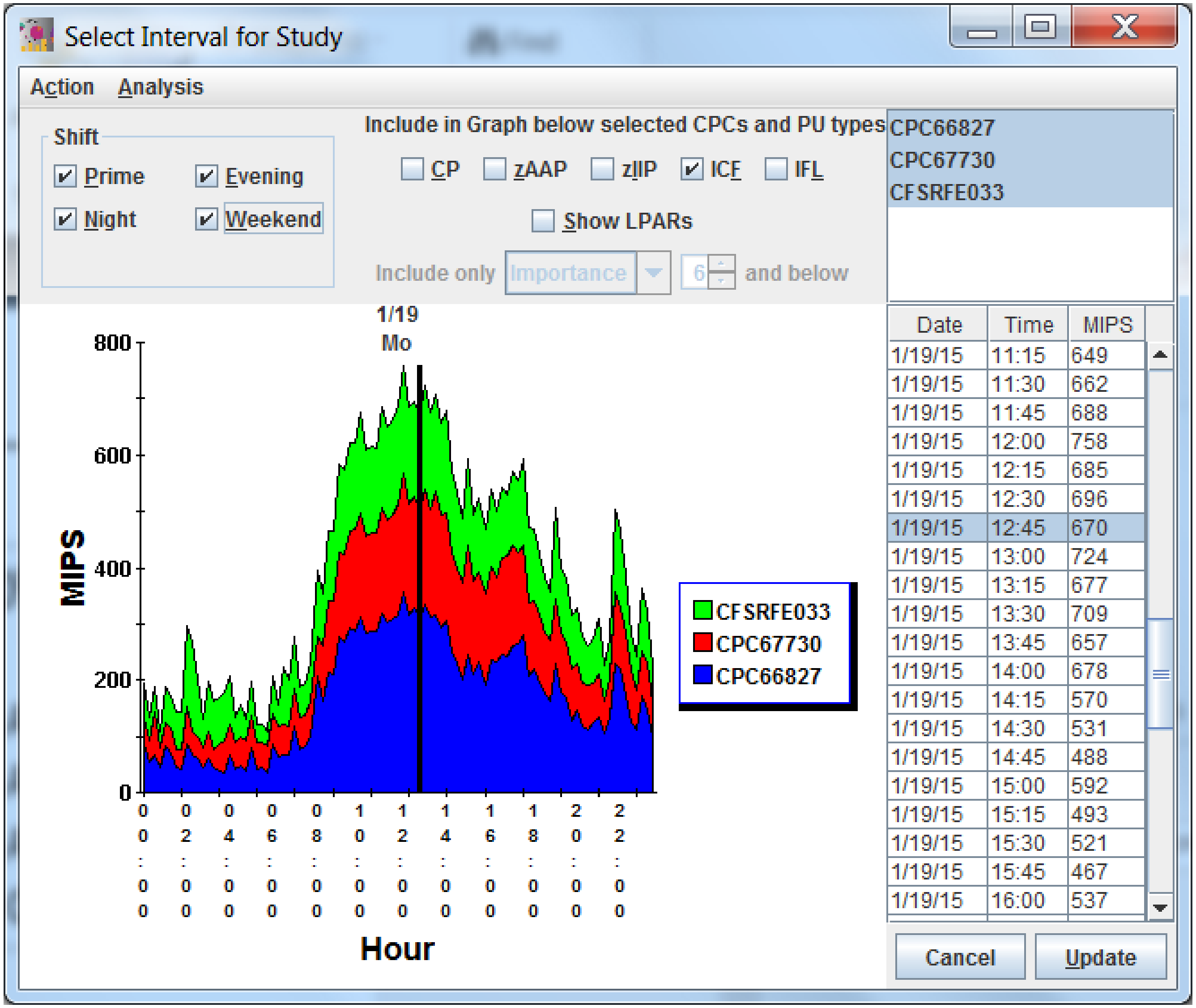
IBM



The default is 90^{th} % of

- All shifts
- All intervals
- All CPCs
- GCP used

Select a study interval



Keep

- All shifts
- All intervals
- All CPCs

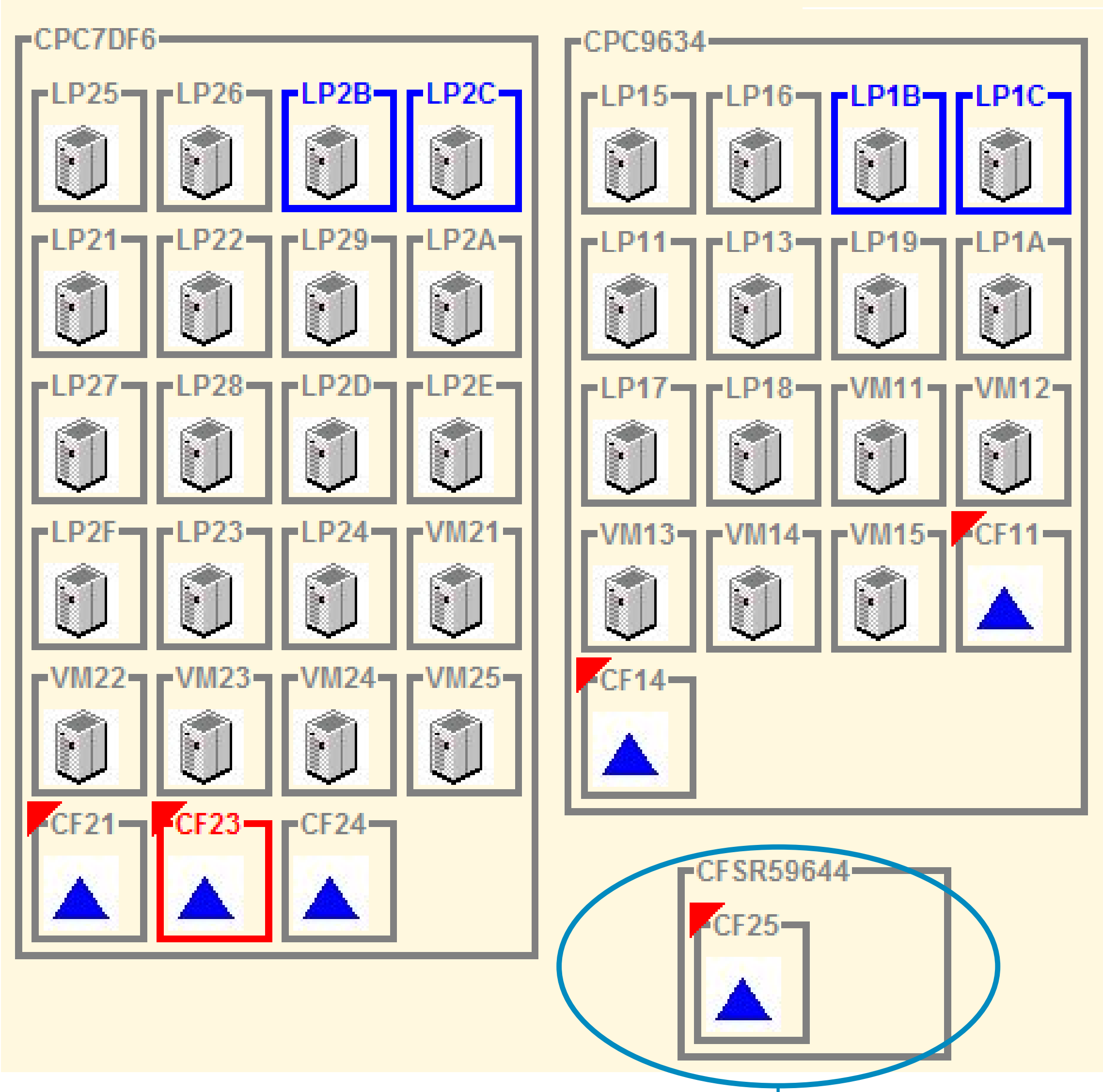
Change this:

- Uncheck CP
- Check ICF
- Reselect 90%

Look at Collected Model Data

Incomplete Configuration Information

- Red Triangles flag things that you need to fix.
- InfiniBand links are flagged until you refine the type.
- Standalone CFs will often need further definition.



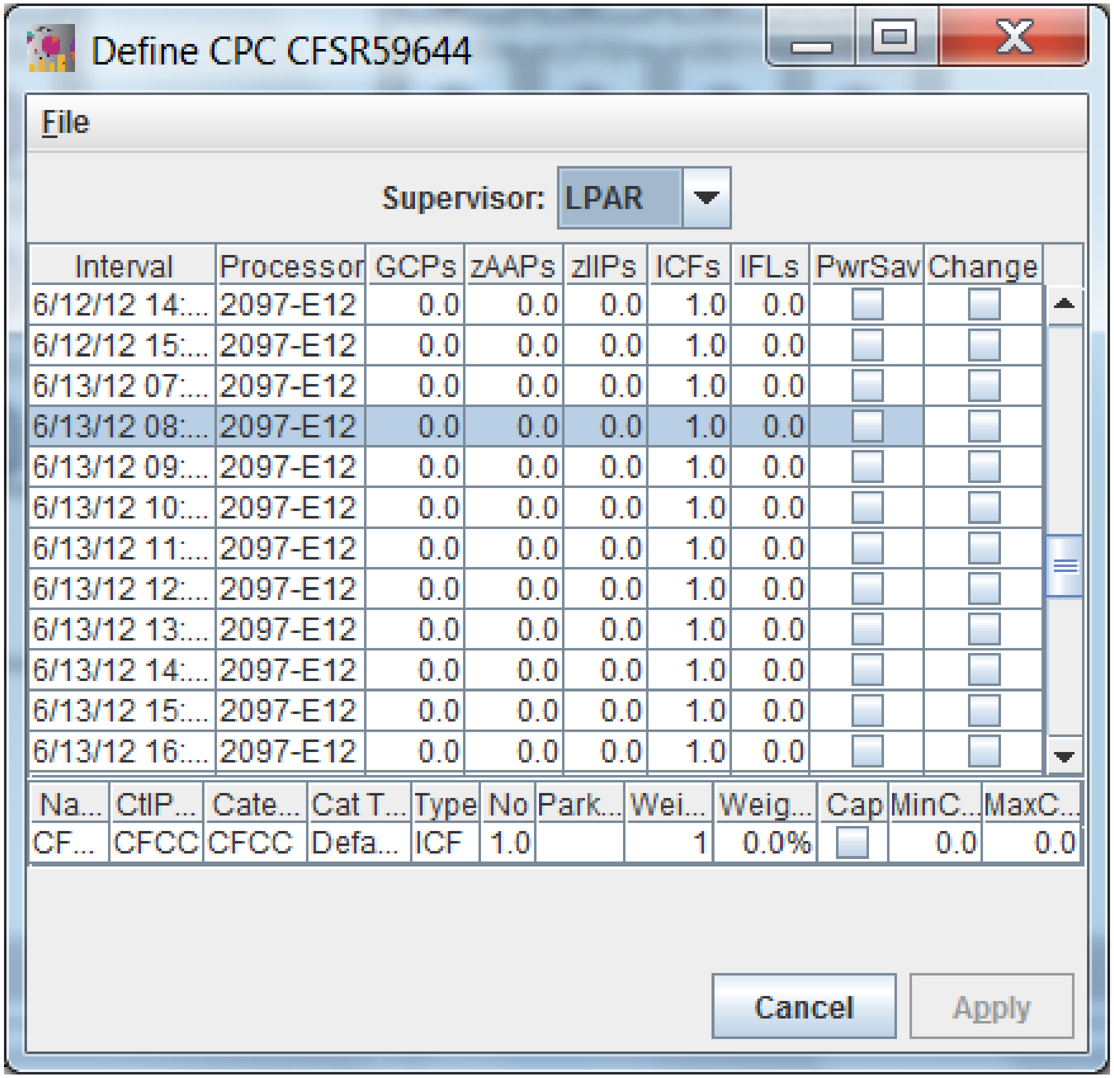
Looking at Collected Model Data

We know this config is not right.

(This config needs editing.)

We do not have SMF for this machine. This is all we know about CF25:

- **CFTYPE="2097-E12"**
- **CFNCPS="01"**
- **SR="59644"**
- **Plus, we know it must be sharing ICF/GCP resource with *someone*.**
- **But we don't even know this is ICF.**



Interval	Processor	GCPs	zAAPs	zIIPs	ICFs	IFLs	PwrSav	Change
6/12/12 14:...	2097-E12	0.0	0.0	0.0	1.0	0.0		
6/12/12 15:...	2097-E12	0.0	0.0	0.0	1.0	0.0		
6/13/12 07:...	2097-E12	0.0	0.0	0.0	1.0	0.0		
6/13/12 08:...	2097-E12	0.0	0.0	0.0	1.0	0.0		
6/13/12 09:...	2097-E12	0.0	0.0	0.0	1.0	0.0		
6/13/12 10:...	2097-E12	0.0	0.0	0.0	1.0	0.0		
6/13/12 11:...	2097-E12	0.0	0.0	0.0	1.0	0.0		
6/13/12 12:...	2097-E12	0.0	0.0	0.0	1.0	0.0		
6/13/12 13:...	2097-E12	0.0	0.0	0.0	1.0	0.0		
6/13/12 14:...	2097-E12	0.0	0.0	0.0	1.0	0.0		
6/13/12 15:...	2097-E12	0.0	0.0	0.0	1.0	0.0		
6/13/12 16:...	2097-E12	0.0	0.0	0.0	1.0	0.0		

Na...	CtlP...	Cate...	Cat T...	Type	No	Park...	Wei...	Weig...	Cap	MinC...	MaxC...
CF...	CFCC	CFCC	Defa...	ICF	1.0		1	0.0%		0.0	0.0

This could be an empty ICF partition on another CEC (check the SR#).

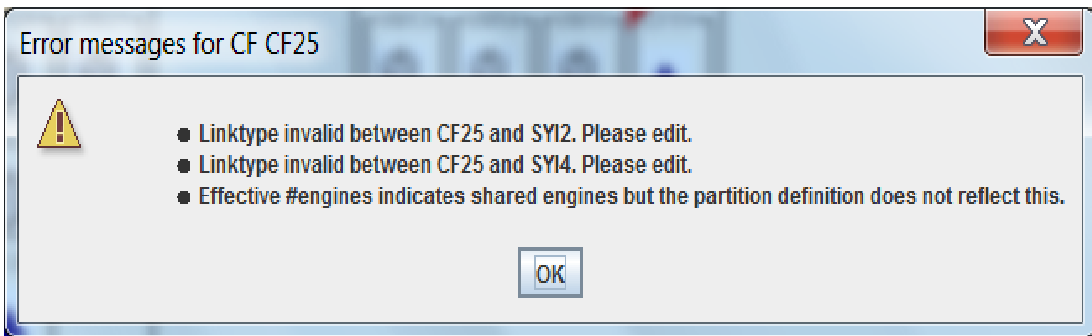
Or, this CEC contains another CF.

Or, this CF uses GCP, which it is sharing w/ z/OS sysplex members, or z/VM-Linux, or whatever.

zCP3000 CF Analysis

IBM

Error Messages Summary

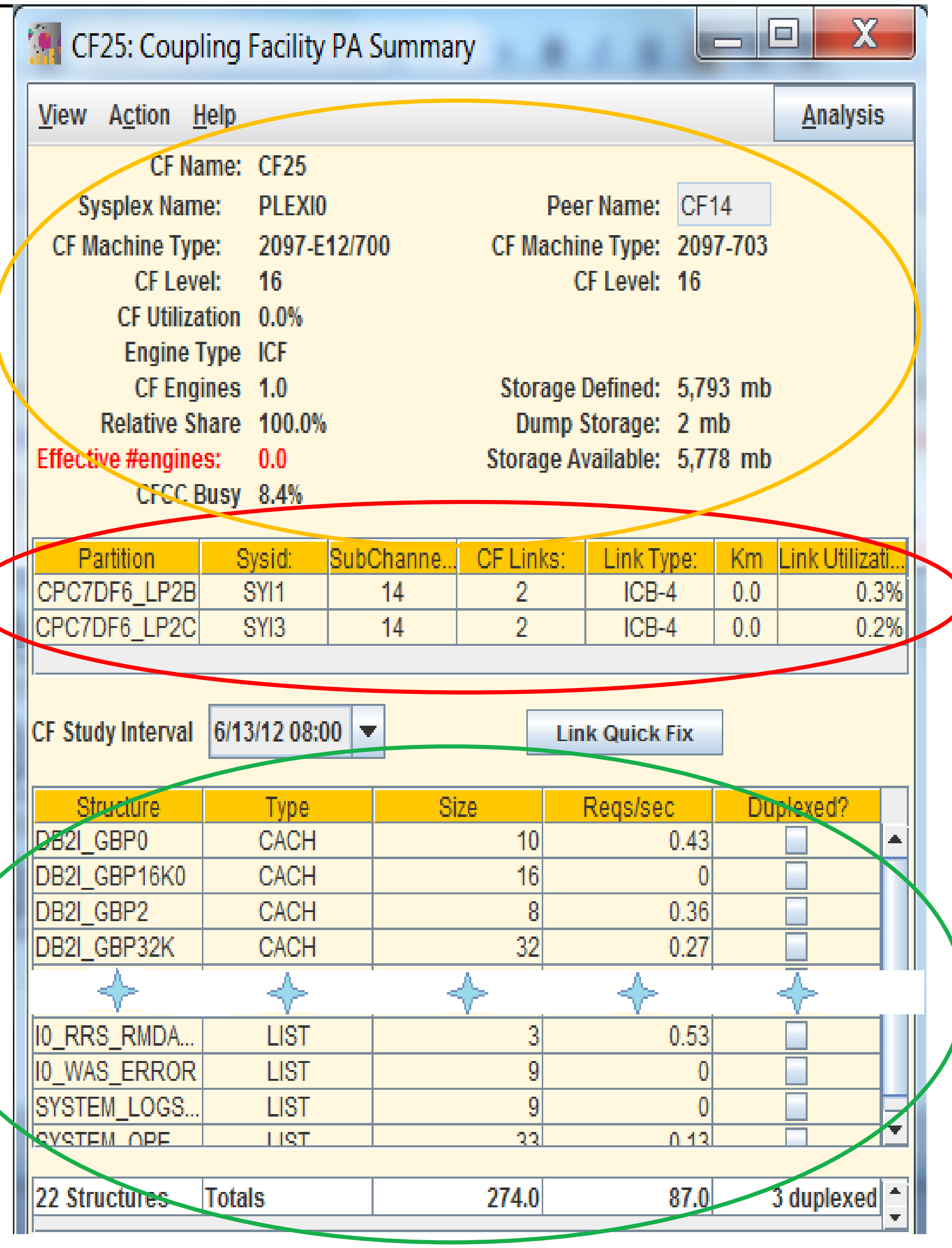


CF Window Summary:

CF Information

Link Information

Structure Information (sortable!)



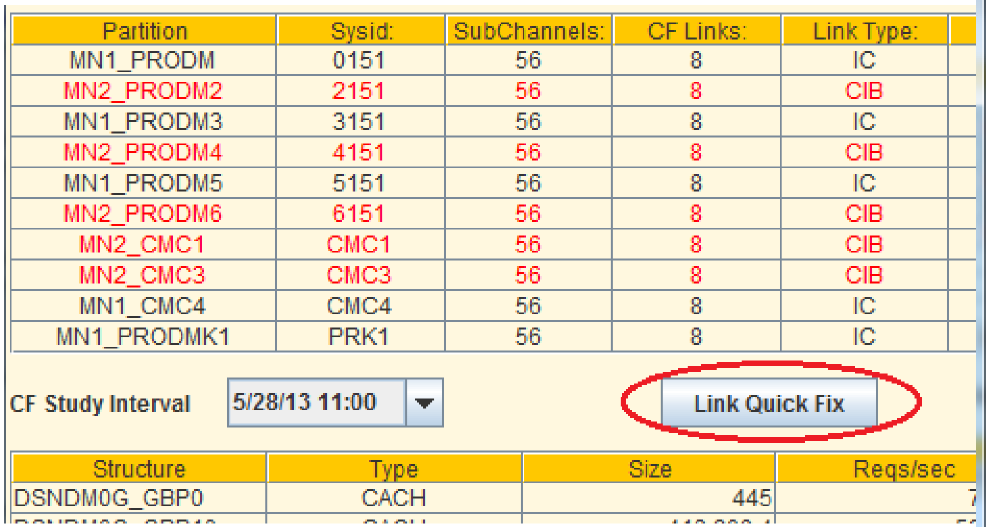
Partition	Sysid:	SubChanne...	CF Links:	Link Type:	Km	Link Utilizati...
CPC7DF6_LP2B	SYI1	14	2	ICB-4	0.0	0.3%
CPC7DF6_LP2C	SYI3	14	2	ICB-4	0.0	0.2%

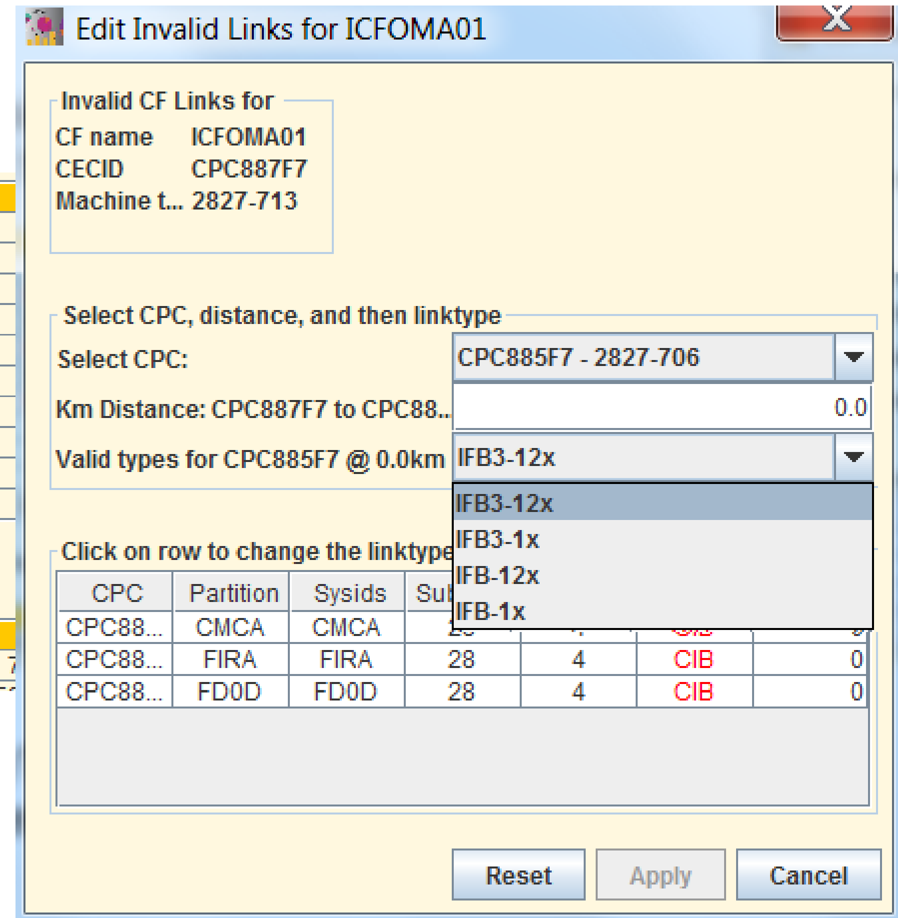
Structure	Type	Size	Reqs/sec	Duplexed?
DB2I_GBP0	CACH	10	0.43	
DB2I_GBP16K0	CACH	16	0	
DB2I_GBP2	CACH	8	0.36	
DB2I_GBP32K	CACH	32	0.27	
I0_RRS_RMDA...	LIST	3	0.53	
I0_WAS_ERROR	LIST	9	0	
SYSTEM_LOGS...	LIST	9	0	
SYSTEM_OPE...	LIST	33	0.13	
22 Structures	Totals	274.0	87.0	3 duplexed

Graph CF1014 on the CF Window

Fix the CF Link Types

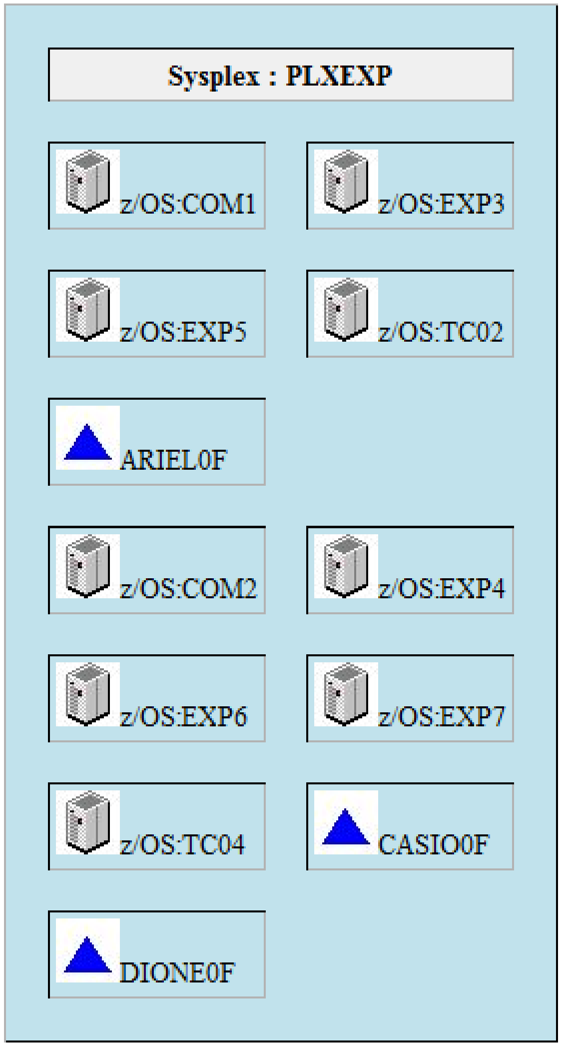
- InfiniBand links come in looking alike, and the distance is not known:





- Why? To model changes, you must understand what you are changing from.

Sysplex Topology Report



Graph PLEX1007 on the Logical Sysplex Window

The PLXEXP sysplex for this configuration is comprised of 9 z/OS systems and 3 coupling facilities, residing on 3 different processors.

Processor	Partition Type	ID	(#) Link Type
CPC66827 2827-723	CF	ARIEL0F	(5) IFB3-1x (4) IC
	SYS	EXP3	(4) IFB-1x (5) IFB3-1x (4) IC
	SYS	COM1	(4) IFB-1x (5) IFB3-1x (4) IC
	SYS	TC02	(4) IFB-1x (5) IFB3-1x (4) IC
	SYS	EXP5	(4) IFB-1x (5) IFB3-1x (4) IC
CFSRFE033 2098-E10	CF	DIONE0F	(4) IFB-1x
CPC67730 2827-722	CF	CASIO0F	(5) IFB3-1x (4) IC
	SYS	EXP4	(4) IFB-1x (5) IFB3-1x (4) IC
	SYS	COM2	(4) IFB-1x (5) IFB3-1x (4) IC
			(4) IFB-1x (5) IFB3-1x

CF Health Check

A single report giving a summary view of several key performance metrics.

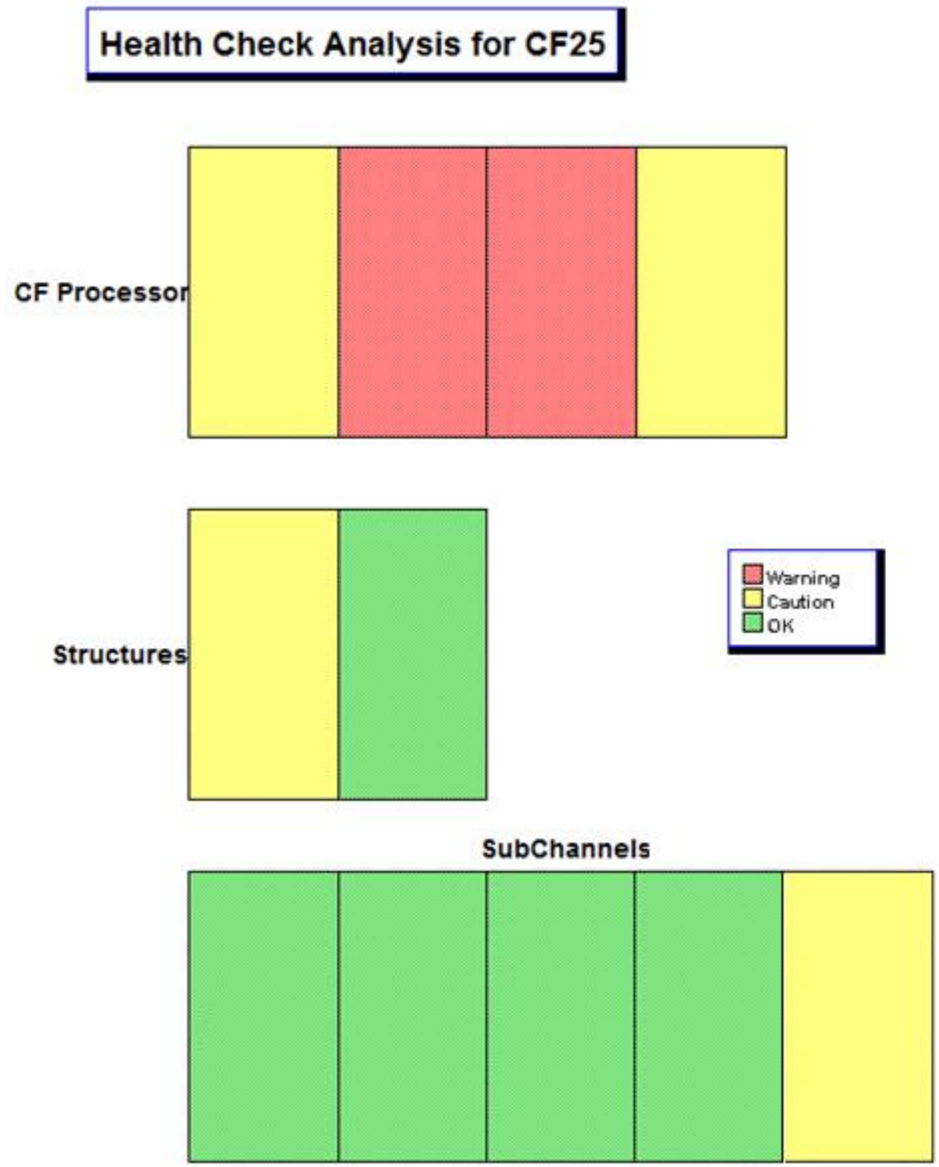
WARNING (RED) is a strong indication of a problem.

CAUTION (YELLOW) indicates a potential problem.

OK (GREEN) means that the analysis didn't find a problem.

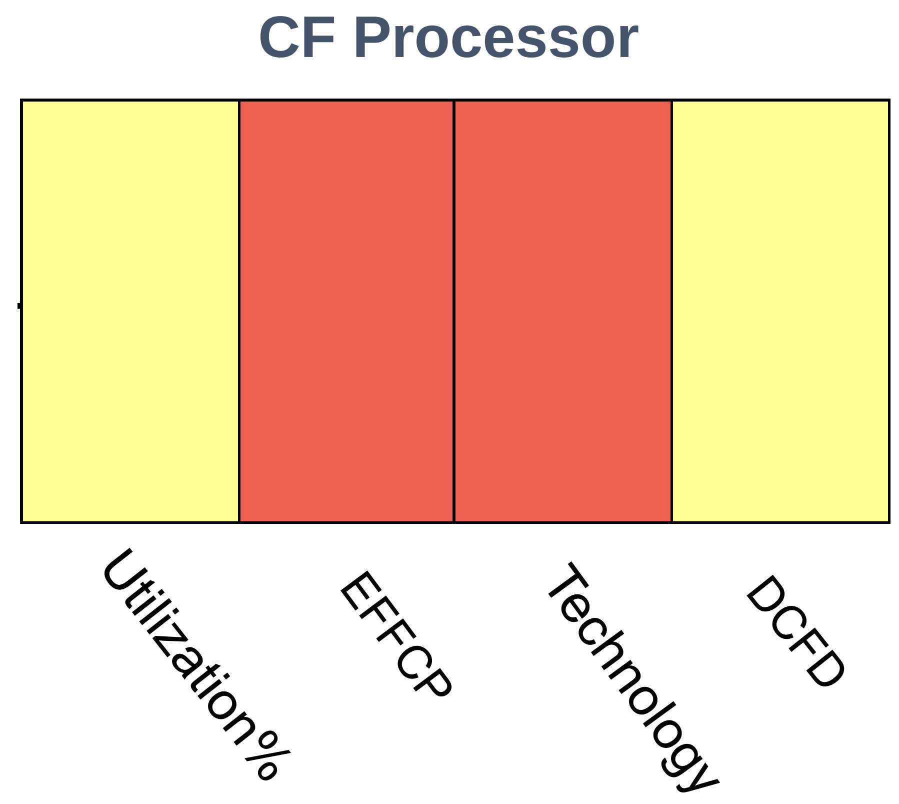
Red or Yellow means that we think you should look further at this area. It does not mean there *is* a problem.

Graph CF1000 on the CF Window



CF Health Check – CF Processor

- Partition Utilization
- Effective Engines >= 1*
- Current Technology?
- Dynamic CF Dispatch*



* Special logic for Thin Interrupt

Graph CF1000 on the CF Window

CF Health Check - Structures

Lock Contention

Lock Requests that had to wait

Fix in the Application.

False Lock Contention

2 locks hash to the same index

Increase the Structure size.

Graph CF1000 on the CF Window

CF Health Check Analysis : zCP3000 Graph and Text: CF1000

File Action

Structures

There were 157 structures in CF21. Looking at data from all intervals, the busiest structure was IXC2, a LIST type structure, with an average rate of 2,148.01 requests per second. Requests from all 157 structures totaled 7,125.75 per second, on average. Approximately 5% of all requests to IXC2 were from system SYE1. The table below shows some general information about activity to different structure types on the CF21 coupling facility.

	Structures	Average Req/sec
LOCK	11	2,745.1
LIST	90	3,841.4
CACH	56	539.2
All	157	**7,125.8**

Structures Rule 1 - Lock Contention - Caution

The table below shows information about lock structures. At least one lock structure had an interval where Lock Contention exceeded 2.0%. High Lock Contention can result in an increase in utilization and reduction in throughput. If the total request number is trivial, high Lock Contention percentage is not a problem. Otherwise, you may want to check the other applications that are running on the systems. In some cases, batch applications that share the databases with online applications hold locks for a much longer time. The time that the lock is held by the batch program can be shortened by taking more frequent checkpoints.

Structures	Type	Average Req/sec	Lock Contention	False Lock Contention
DB2P_LOCK1	LOCK	1,932.3	1.84%	0.21%
IGWLOCK00	LOCK	0.2	16.05%	0.54%
DB2R_LOCK1	LOCK	0.9	1.73%	0.35%
IRLMLOCK	LOCK	453.7	0.02%	0.00%
DLILOCKR	LOCK	0.0	0.00%	0.00%
DB2T_LOCK1	LOCK	82.5	0.39%	0.12%
DB2E_LOCK1	LOCK	3.6	0.55%	0.12%
ISGLOCK	LOCK	271.9	1.88%	0.48%
DB2V_LOCK1	LOCK	0.2	1.20%	0.22%
DLILOCKE	LOCK	0.0	0.00%	0.00%
DLILOCK	LOCK	0.0	0.00%	0.00%

CF Health Check - Subchannels

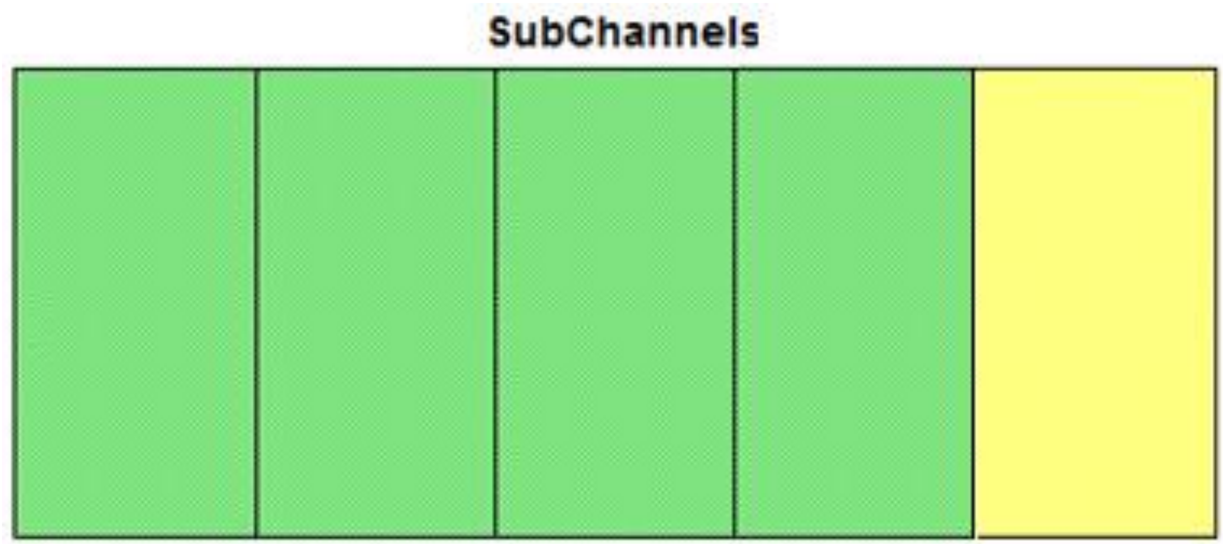
Subchannel Utilization
should be very low (<10%)

Sync -> Async Conversion
z/OS looked at the on-ramp and decided to go with Plan B.

Subchannels Offline
there is a sustained pattern of path busy.

Delayed Requests
more than 10.0% of requests were delayed

SubChannel Busy
Path busy at the hardware link level.



Graph CF1000 on the CF Window

Service Time

From when z/OS issues a CF request to when the return is recognized by z/OS.

- Host Hardware
- Link Latency
- Data Transfer
- CF Busy

Does not include queue time.

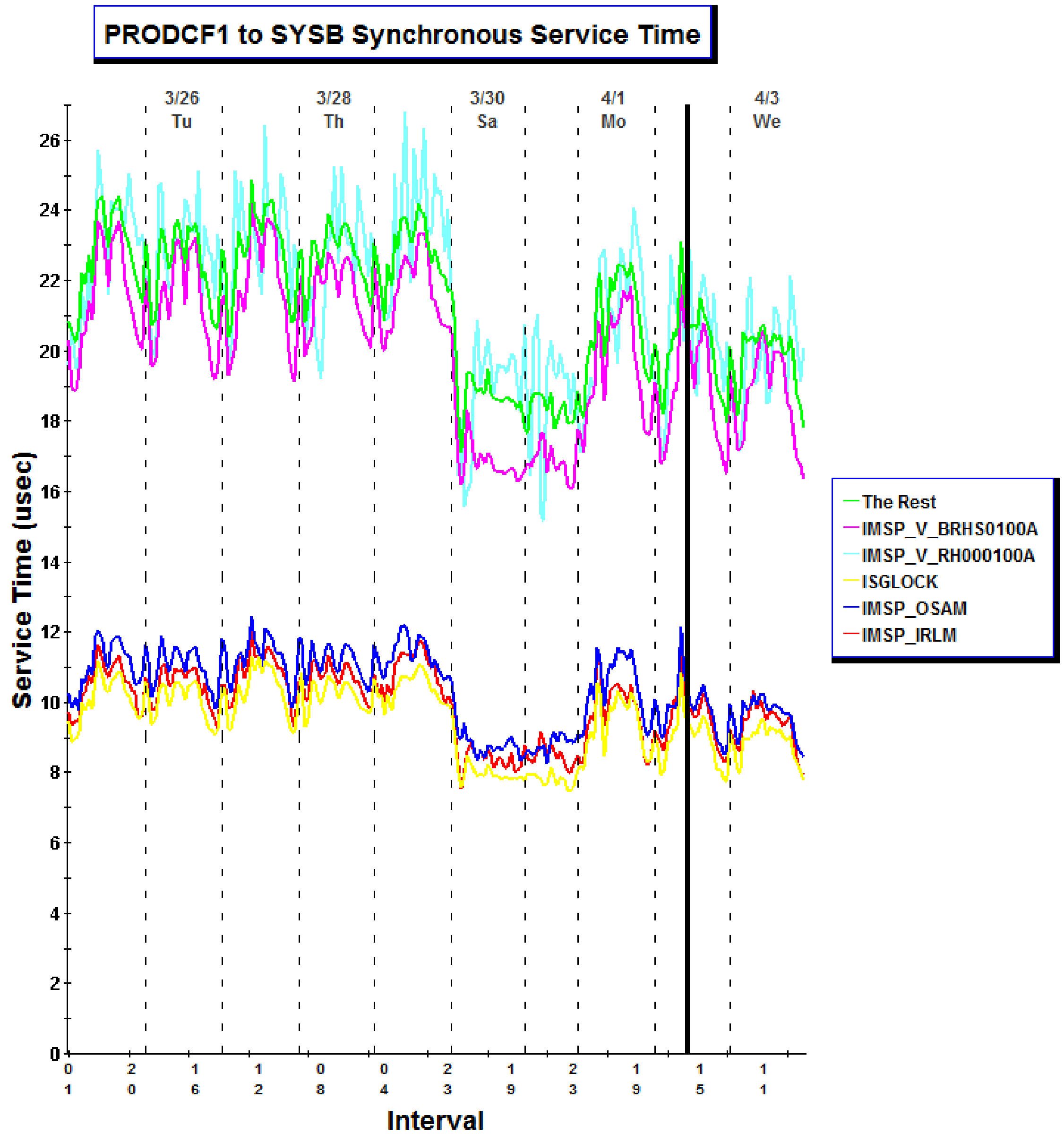
Service Time

• z/OS wants to run all requests synchronously (except for XCF).

• Insufficient bandwidth or slow service will cause it to convert to async.

• Therefore, except for IXC requests and distances>1km, all async requests should be viewed as a lack of resource.

• Service Time graphs available from the CF window, the Link window, and the Structure detail window.



Graph CFL012 on the CF Link Window

Identify key coupling links

Suppose there are 2 plexes, with 5 CFs and 20 z/OS members.

- Given the objectives of the study, which CFs are important?
- Sort the links by request rate (CF window, click on column header)
- Look at the CF Structures Report to see which links are busiest for the busiest (aka most important) structures.
- If anything shows up yellow in the Subchannels part of the CF Health Check, it’s worth looking at.

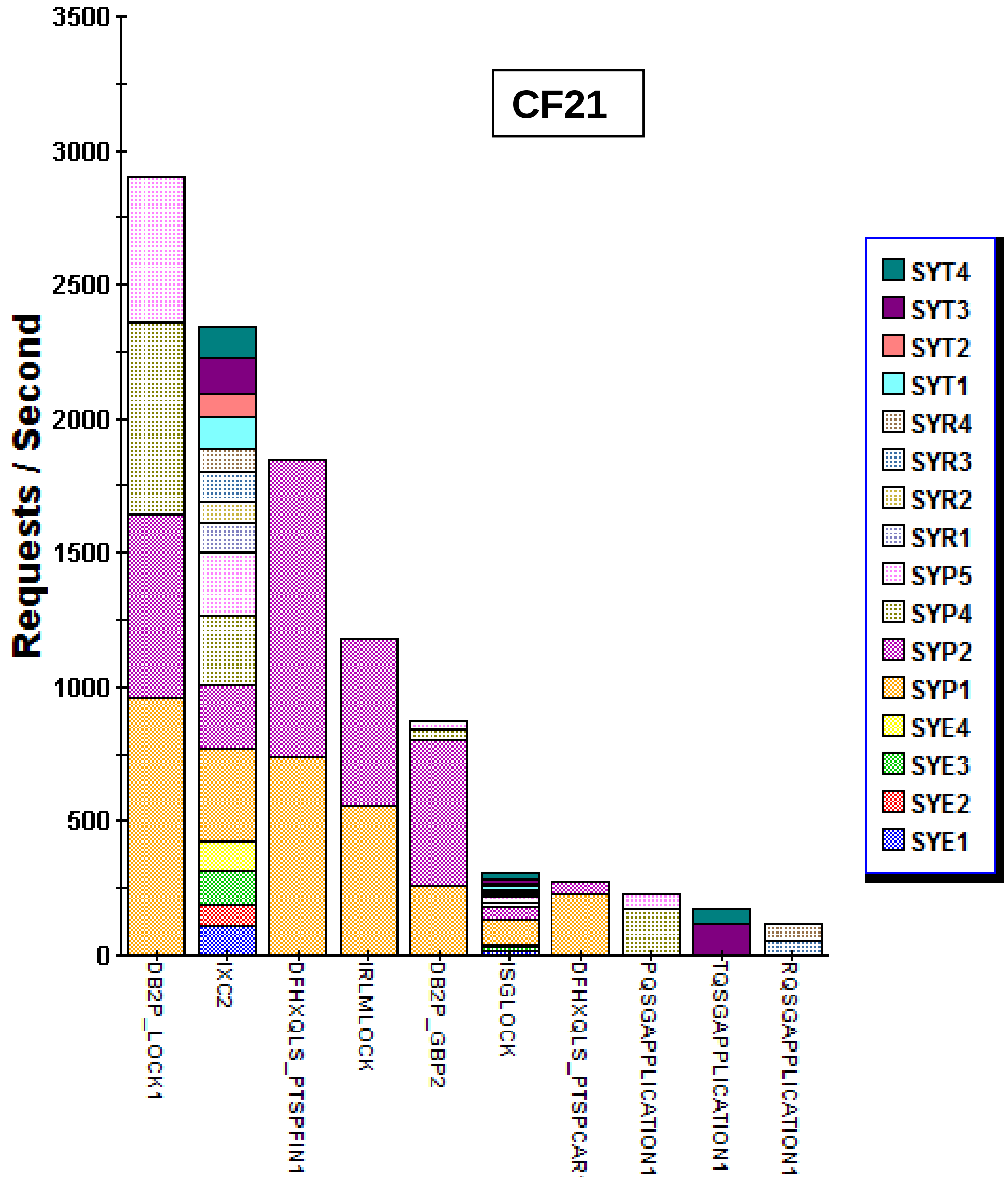
CF Structures Report

- Use to identify the busiest structures.
- Use to identify the key CF links for the busiest structures.

In your customer deliverable

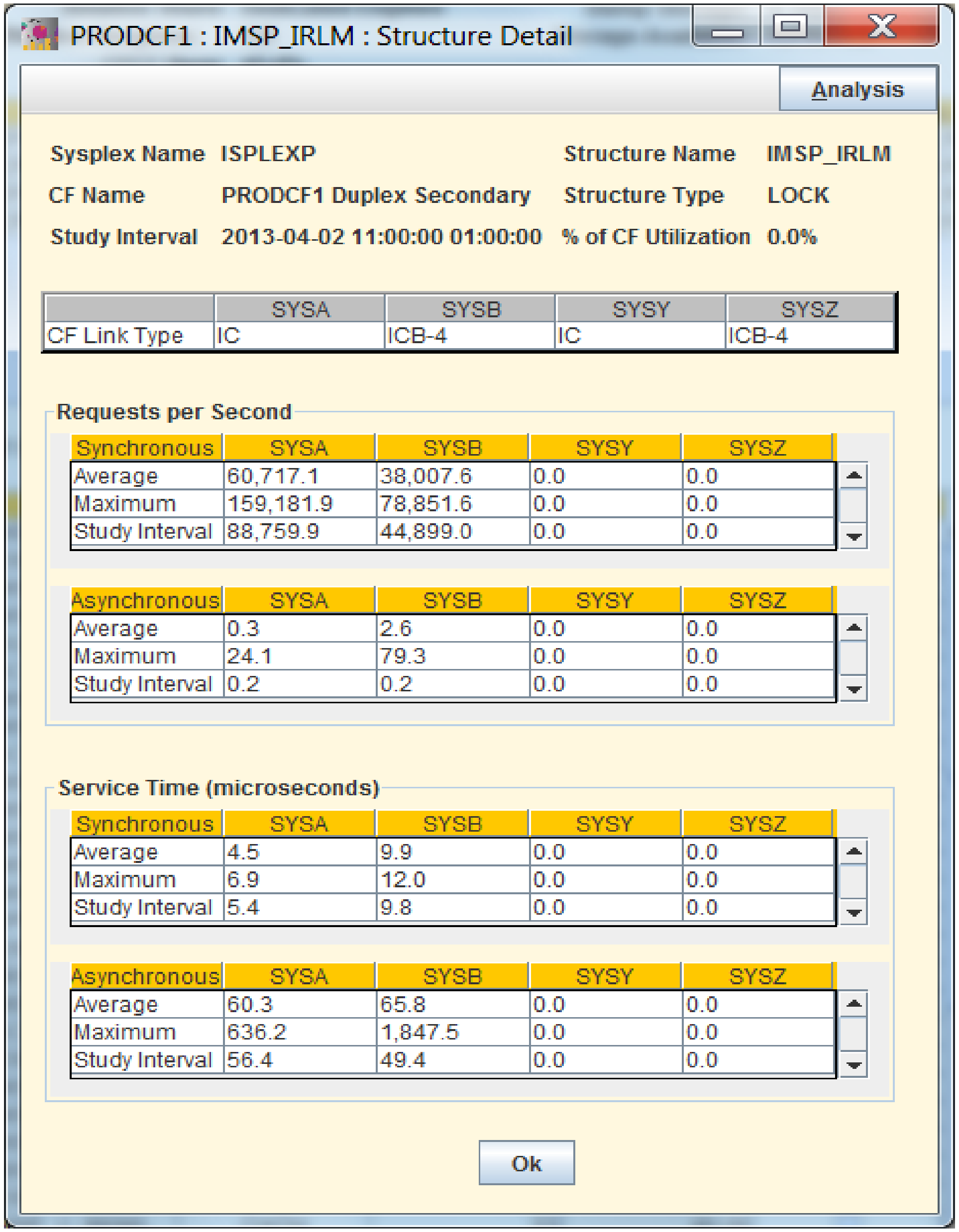
- focus on critical applications
- focus on identified problems
- otherwise, use this to identify important structures and links.

Graph CF1010 on the CF Window



Service Time for Important Structures

- Pick a busy structure.
- Note which links are busiest.
- Are most requests running synchronously?
- Are they getting good service time?
- The IC links to SYSA and the ICB-4 links to SYSB are adequate to the current demand.


PRODCF1 : IMSP_IRLM : Structure Detail

Analysis

Sysplex Name	ISPLEXP	Structure Name	IMSP_IRLM
CF Name	PRODCF1 Duplex Secondary	Structure Type	LOCK
Study Interval	2013-04-02 11:00:00 01:00:00	% of CF Utilization	0.0%

	SYSA	SYSB	SYSY	SYSZ
CF Link Type	IC	ICB-4	IC	ICB-4

Requests per Second

Synchronous	SYSA	SYSB	SYSY	SYSZ
Average	60,717.1	38,007.6	0.0	0.0
Maximum	159,181.9	78,851.6	0.0	0.0
Study Interval	88,759.9	44,899.0	0.0	0.0

Asynchronous	SYSA	SYSB	SYSY	SYSZ
Average	0.3	2.6	0.0	0.0
Maximum	24.1	79.3	0.0	0.0
Study Interval	0.2	0.2	0.0	0.0

Service Time (microseconds)

Synchronous	SYSA	SYSB	SYSY	SYSZ
Average	4.5	9.9	0.0	0.0
Maximum	6.9	12.0	0.0	0.0
Study Interval	5.4	9.8	0.0	0.0

Asynchronous	SYSA	SYSB	SYSY	SYSZ
Average	60.3	65.8	0.0	0.0
Maximum	636.2	1,847.5	0.0	0.0
Study Interval	56.4	49.4	0.0	0.0

Ok

Sysplex Graph Recommendations

Probably in every study

- zCP3000 main view : Topology
- Sysplex Logical window : Sysplex Topology, Link Summary
- For each CF : CF Summary, CF Health Check,
  CF Migration Summary report

PA Mode

PA and QM Mode

QM only

Look for areas of interest

- Thin Interrupt Effect
- CF Structures Table
- CF Link Topology
- CF Link Summary, Link Migration Comparison
- Service Time for Synchronous Structures
- Request Rate by System over Time
- Request Rate by Request Type and System

Drill down in areas of interest

- CF Logical Utilization over Time
- Advanced CF (shared ICF)
- Subchannel Busy
- Synchronous Service Time (Link)
- Lock Contention % of Requests
- Synchronous Intensity
- Delayed Requests
- Busiest Structures with Queue Time

Agenda

Use zCP3000 to look at sysplex questions:

- Collect the right data for a study
- Learn about the data, identify current problem areas
- **Modelling a new hardware configuration**
- Using shared ICF and Thin Interrupt
- z13 Link Migration Considerations
- Link consolidation on InfiniBand and ICA
- Sysplex Aggregation Pricing eligibility

Modeling configuration changes

- CF utilization changes w/ the speed of the CPC or alt.
- zCP3000 **calculates** new service times for config changes
  - For each structure, link, and sys
  - Additional adjustment based on current service time (the crappiness factor).
  - Recalc Sync/async on the new service time.
- Service time recalc is triggered in QM mode when:
  - Linktype or distance changes
  - CF machine type changes via drag and drop
  - CF Dispatch mode changes
- The current configuration must be understood:
  - Link types
  - Link distance
  - Shared channel paths
  - CF model, engine type, and LPAR weight

Coupling Links

Responsiveness depends on

- Type of Request
- Amount of data being carried*
- How request will be executed (sync or async)
- Simplex or Duplex
- Path busy or anticipated path busy (queued request)
- Link latency
- Link speed
- Link distance
- CF speed and responsiveness
- z/OS speed

Service time does not include queue time.

-> see CF Health Check, Subchannels, Delayed Requests

Sysplex Overhead is dependent on

- How request will be executed (sync or async)
- Service Time, for synchronous requests

* (not measured)

So lets model a new z13…

Collect data for all sysplex members (and maybe dev/test).

Fully define the sysplex in it's current state in PA mode.

- What *type* of IFB links: IFB3, IFB-12x, or IFB-1x?
- Link distance, if it will be changing
- Identify link any chpids shared by z/OS partitions.
- For standalone CFs : model, engine type, and LPAR weight
- Add hardware feature codes, if possible.
- Maybe generate some reports, like the Health Checks for key SYS and CF

Save then go to QM mode.

- Add a new z13
- Move partitions to the new CPC → service time recalc
- Define engines and lpars of the new z13 → service time recalc
- Upgrade link types → service time recalc
- Change dispatch mode → service time recalc

Look at the CF Migration Report →

CF Migration Summary Report

IBM

Before and after comparison of:

- CF configuration summary, CF link utilizations
- Configuration summary for all CFs and z/OS members of the sysplex

Sysplex PLXEXP Configuration After Migration							
	CECID	Model	Partition	SCP Level	Logical Engines	Relative Weight	Engine Type
DIONE0F	newZ13	2964-701	DIONE0F	16	1	Dedicated Engines	ICF
COM1	CPC66827	2827-723	ARIEL43	ZV011300	2	1.6%	GCP

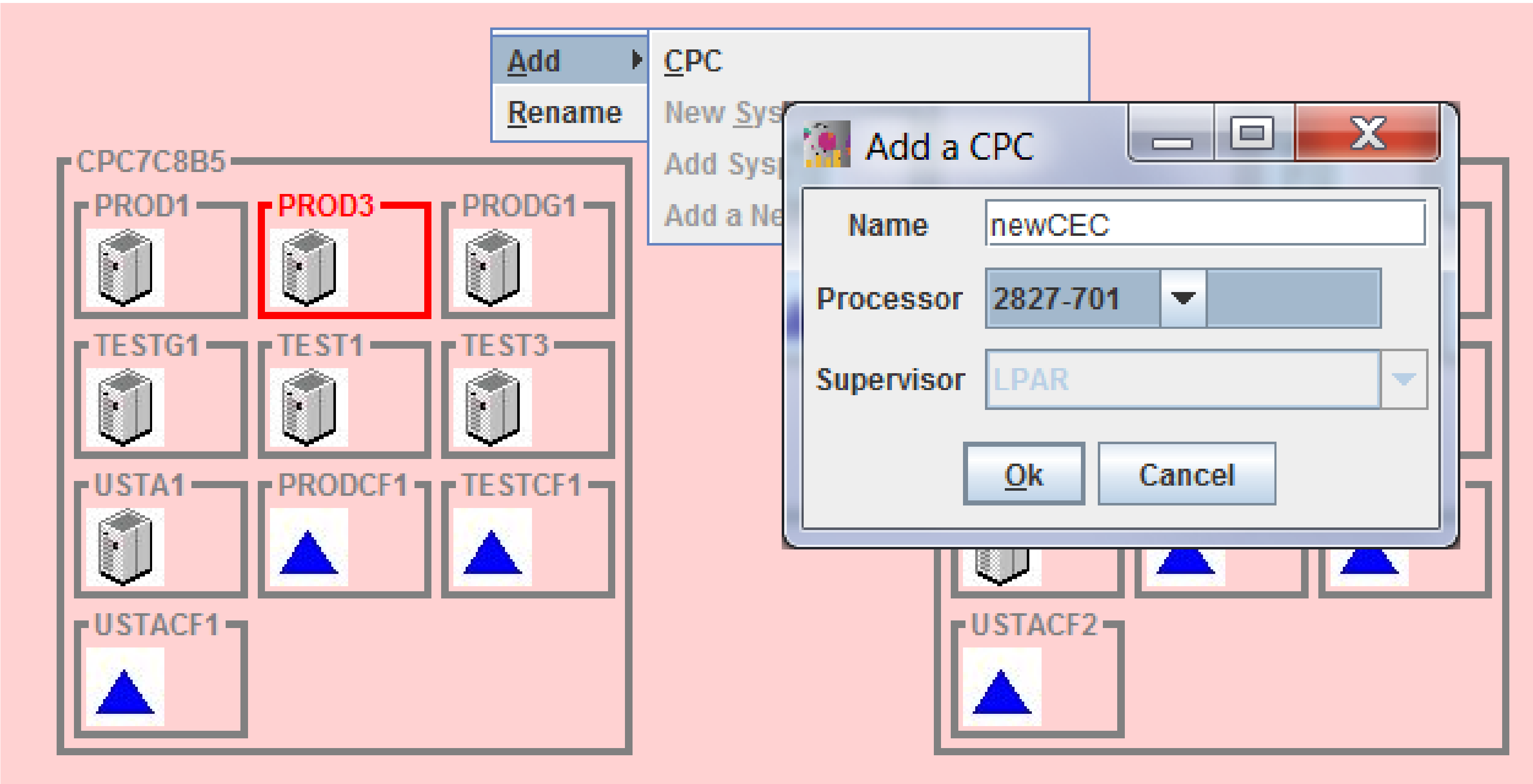
- Key Structures

Structures	Type	Duplexed?	Reqs/sec	Configuration Before Migration			Configuration After Migration		
				%Sync	Service Time		%Sync	Service Time	
					Sync	Async		Sync	Async
DB2DS0E_LOCK1	LOCK	No	27,508.9	96.1%	15.8	47.5	100.0%	14.2	□
CQSTMSGQ	LIST	No	13,318.3	88.7%	21.4	57.0	88.7%	24.6	44.8
ISGLOCK	LOCK	No	9,700.3	99.9%	15.3	56.1	100.0%	14.2	□
IXCLST01	LIST	No	5,071.4	0.0%	-	61.7	0.0%	-	56.5
DB2DS0S_LOCK1	LOCK	No	2,285.5	98.1%	15.1	44.3	100.0%	14.3	□

Graph CFQ100 on the CF Window (QM only)

Change the CF machine type

Step 1 – add the new machine



Change the CF machine type

Step 2 – move some partitions there

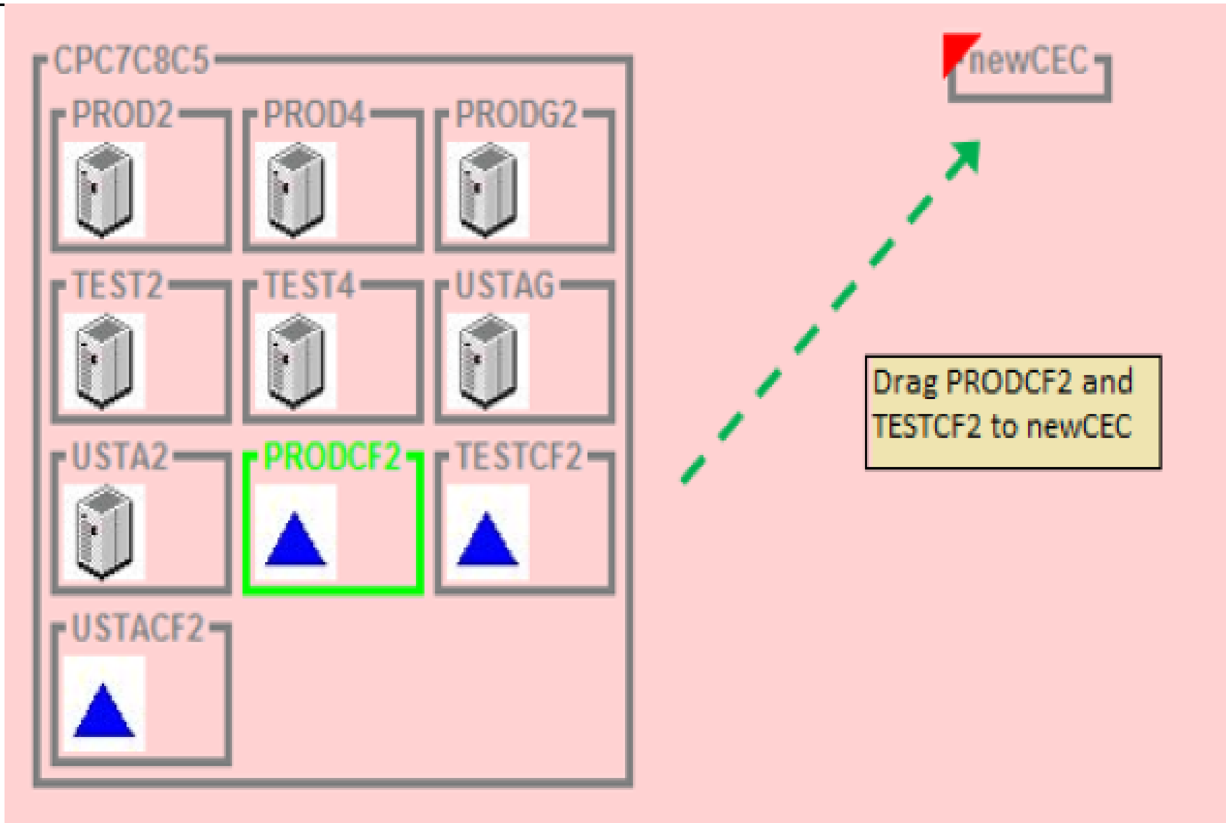


Figure 3. Error due to empty CPC

The new CPC will still be marked in error because the definition needs to change to accommodate the partitions that just moved there.

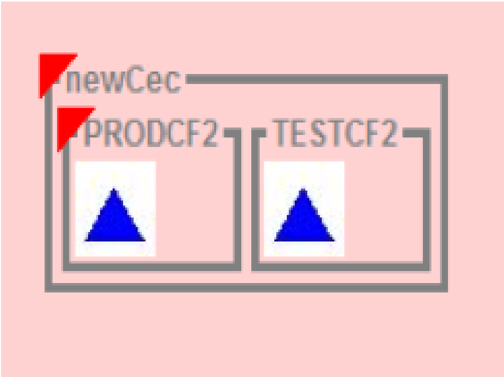
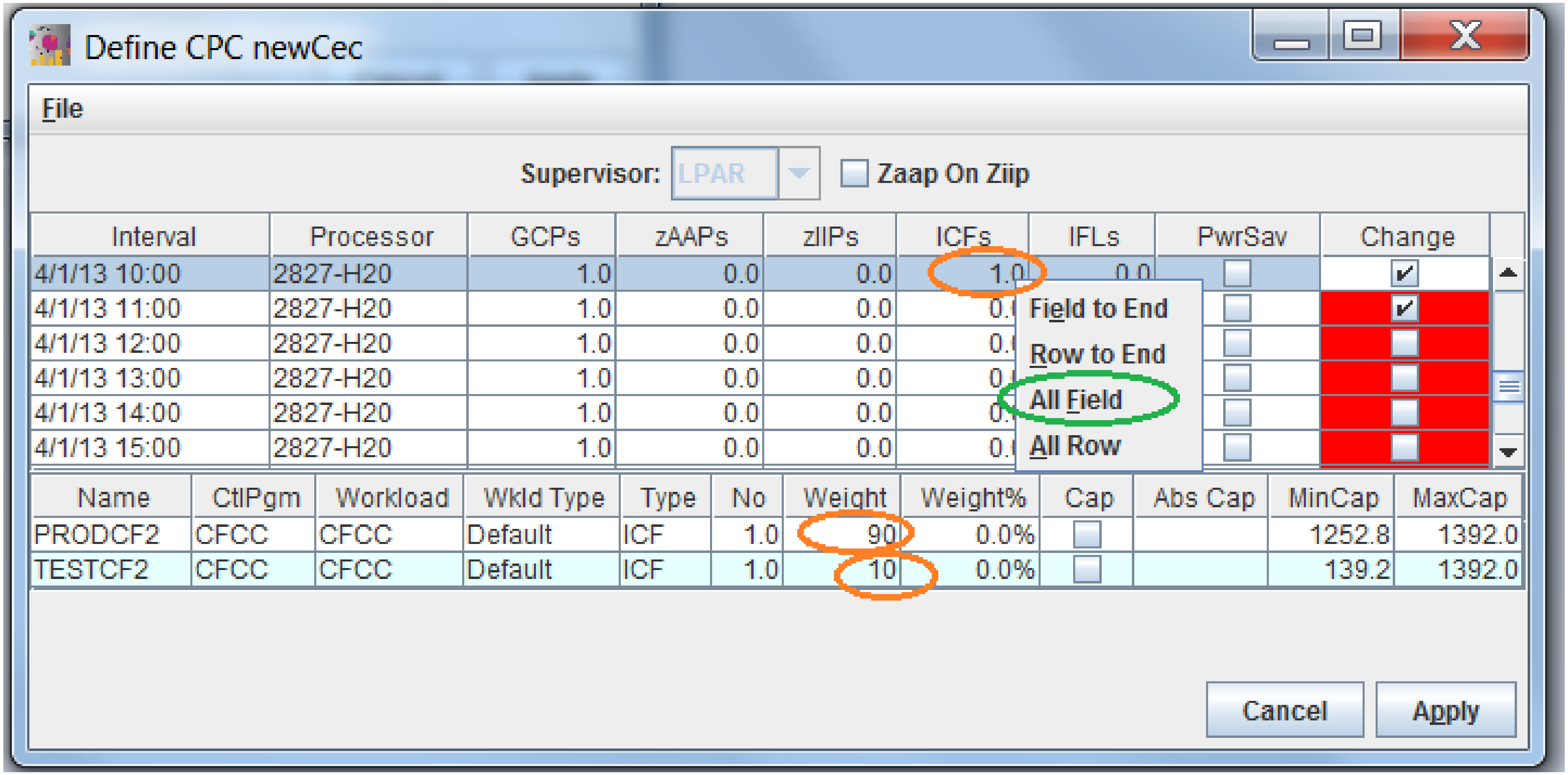


Figure 4. CEC Definition needs to be fixed on new CPC.

Configuration differences will cause CF request service times to be recalculated. Edit->Undo to reverse this change.

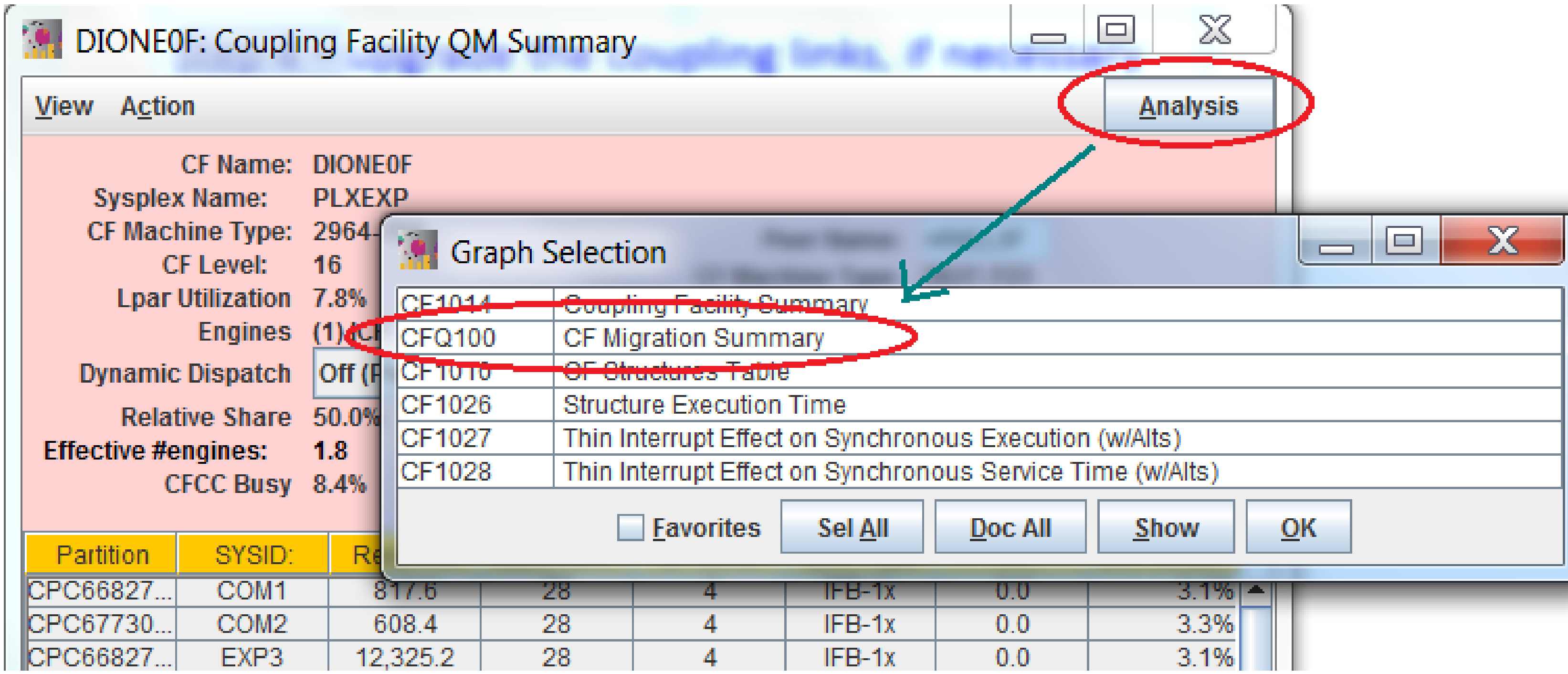
Change the CF machine type

Step 3 – define real engines and partitions



Change the CF machine type

If necessary, upgrade the coupling links.



Otherwise, you're done,
generate a CF Migration Summary for each major CF.

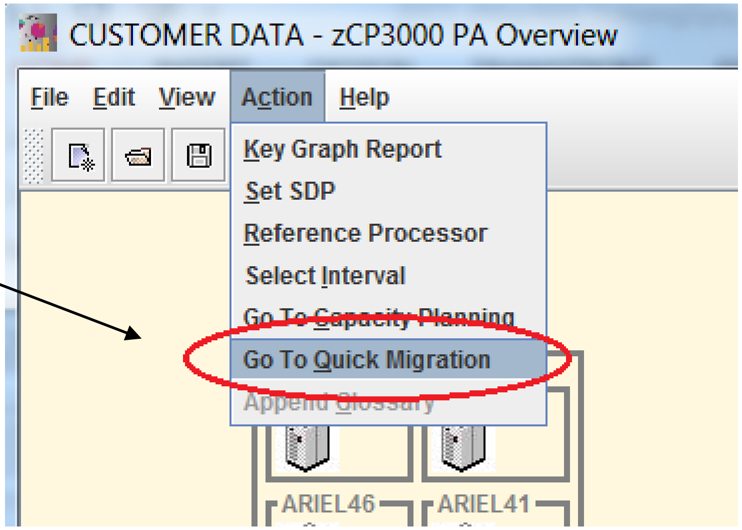
CF Coupling – Relative Link Speed in zCP3000

zCP3000 Name	Link type	Link latency (mics)	Link speed (Mbs/sec)	Supported on
ISC3	ISC3	12	200	z13,zBC12,zEC12,z114,z196,z10,z9,z990,z890,z900,z800
IFB-1x	IFB 1X	10	400	z13,zBC12,zEC12,z114,z196,z10
IFB-12x	IFB 12X – z9	8	600	z9
IFB-12x	IFB 12X	8	1000	z13,zBC12,zEC12,z114,z196,z10
ICB4	ICB4	4	1500	z10,z9,z990,z890
IFB3-12x	IFB3 12X – z114	3.5	2500	z114
IC	IC-z10 BC	1	3200	z10-BC
IFB3-12x	IFB3 12x – zBC12	3.5	4000	zBC12
IC	IC-z9 BC	1	4000	z9
IFB3-12x	IFB3 12X	3.5	5000	z13zEC12,z196
ICA	CS5	3.5	5000	z13
IC	IC-z9 EC	1	5000	z9
IC	IC-z114	1	6500	z114
IC	IC-zBC12	1	7100	zBC12
IC	IC-z10 EC	1	7500	z10
IC	IC-z13	1	8500	z13
IC	IC-z196	1	8900	z196
IC	IC-zEC12	1	9400	zEC12

To model a link type change:

(estimate new service time)

- Must be in QM Quick Migration Mode
  (CP Mode is being phased out)



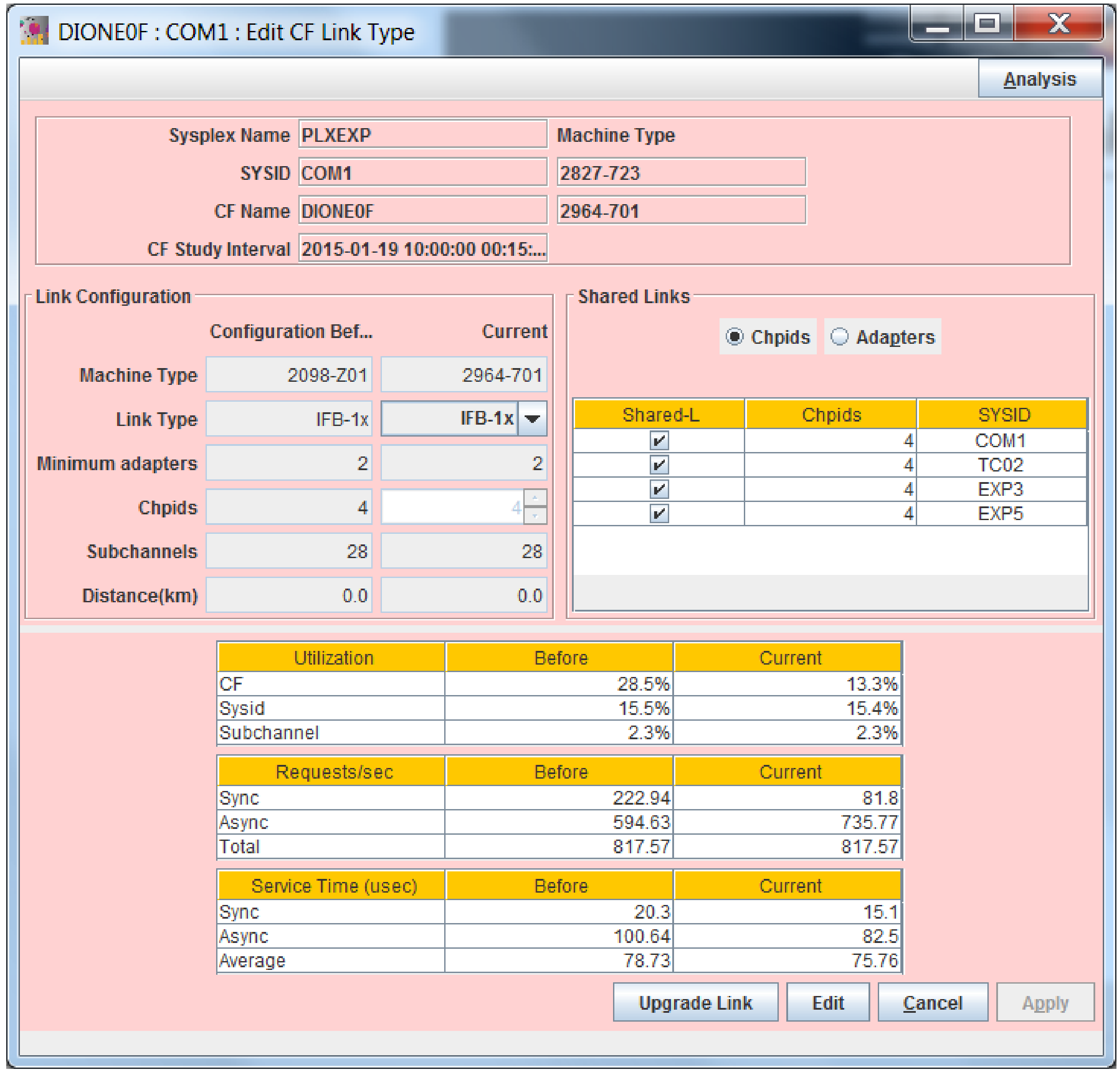
- One CF-SYS set of links at a time

- If CEC model change also, do that first.

Change the Coupling Links

Double click a link on the CF window to look at the link detail.

Use the “Upgrade” button to model a new link type.



Shared-L	Chpids	SYSID
✔	4	COM1
✔	4	TC02
✔	4	EXP3
✔	4	EXP5

Utilization	Before	Current
CF	28.5%	13.3%
Sysid	15.5%	15.4%
Subchannel	2.3%	2.3%

Requests/sec	Before	Current
Sync	222.94	81.8
Async	594.63	735.77
Total	817.57	817.57

Service Time (usec)	Before	Current
Sync	20.3	15.1
Async	100.64	82.5
Average	78.73	75.76

Change the Coupling Links

Link Configuration

	Configuration Bef...	Current
Machine Type	2098-Z01	2964-701
Link Type	IFB-1x	IFB-1x
Minimum adapters	2	2
Chpids	4	4
Subchannels	28	28
Distance(km)	0.0	0.0

Upgrade Link

Upgrade
2964-701
IFB3-12x
2
4
28
0.0

Utilization	Before	Current	Estimated
CF	28.5%	13.3%	13.3%
Sysid	15.5%	15.4%	15.5%
Subchannel	2.3%	2.3%	0.2%

Requests/sec	Before	Current	Estimated
Sync	222.94	81.8	225.03
Async	594.63	735.77	592.54
Total	817.57	817.57	817.57

Service Time (usec)	Before	Current	Estimated
Sync	20.3	15.1	8.49
Async	100.64	82.5	79.24
Average	78.73	75.76	59.77

Changing the Coupling Links

We already changed the CF machine type.

Now we're changing the link type.

Utilization	Before	Current	Estimated
CF	28.5%	13.3%	13.3%
Sysid	15.5%	15.4%	15.5%
Subchannel	2.3%	2.3%	0.2%

Requests/sec	Before	Current	Estimated
Sync	222.94	81.8	225.03
Async	594.63	735.77	592.54
Total	817.57	817.57	817.57

Service Time (usec)	Before	Current	Estimated
Sync	20.3	15.1	8.49
Async	100.64	82.5	79.24
Average	78.73	75.76	59.77

CF Link Migration Comparison

Service Time (usec)	Before	Current	Estimated
Sync	20.3	15.1	8.49
Async	100.64	82.5	79.24
Average	78.73	75.76	59.77

Duplex		Structure Type	Requests/sec	%Synchronous		Service Time (usec)			
						Synchronous		Asynchronous	
				Before	After	Before	After	Before	After
IXCLST01		LIST	592.5	0.0%	0.0%	0.0	0.0	100.5	79.2
ISTGENERIC		LIST	143.0	98.6%	100.0%	22.0	9.2	156.9	29.4
ISGLOCK		LOCK	78.1	100.0%	100.0%	17.1	7.2	0.0	27.5
IRRXCF00_P001		CACH	3.7	96.7%	100.0%	23.9	9.8	122.4	30.1
OPERLOG		LIST	0.2	95.0%	100.0%	29.1	9.2	68.2	29.4
Total/Weighted Average			817.6	27.3%	27.5%	20.3	8.5	100.6	79.2

- The Total/Weighted Average will agree w/ the Request Activity Summary on the Link Window.
- All structures have better sync service time, and better yet, more requests will run synchronously.
- Ie, currently 3.3% of IRRXCF00_P001 requests run asynchronously w/ a service time of 122.4 microseconds. After the z13 and IFB3 migration, those 3.3% will run synchronously w/ a service time of 9.8 mics.

Graph CFQL001 on the CF Link Window (QM only)

Heuristic Sync->Async Conversion

- zOS keeps a table of observed CF service times for each structure.
- Depending on the estimated efficiency tradeoff, it may convert any synchronous operation to asynchronous.
- Conversion is done *before* the request is ever launched.
- Requests converted this way are async from the start. They do not count as a “changed” request.
- With 10 mics added per km distance, anything over 2km will certainly be async.

Approximate Threshold for Conversion

z/OS Host		Microseconds
z9-EC	2094-701	36.92
z9-BC	2096-Z01	44.65
z10-EC	2097-701	26.00
z10-BC	2098-Z01	31.27
z196	2817-701	26.00
z114	2818-Z01	26.45
zEC-12	2827-701	26.00
zBC-12	2828-Z01	26.00
z13	2964-701	26.00

Table 1. The meaning of “slow”

Note: This is **COMPLETELY DIFFERENT** from subchannel busy sync->async conversions.

Agenda

Use zCP3000 to look at sysplex questions:

- Collect the right data for a study
- Learn about the data, identify current problem areas
- Modelling a new hardware configuration
- **Using shared ICF and Thin Interrupt**
- z13 Link Migration Considerations
- Link consolidation on InfiniBand and ICA
- Sysplex Aggregation Pricing eligibility

Effective number of engines

Calculated as:

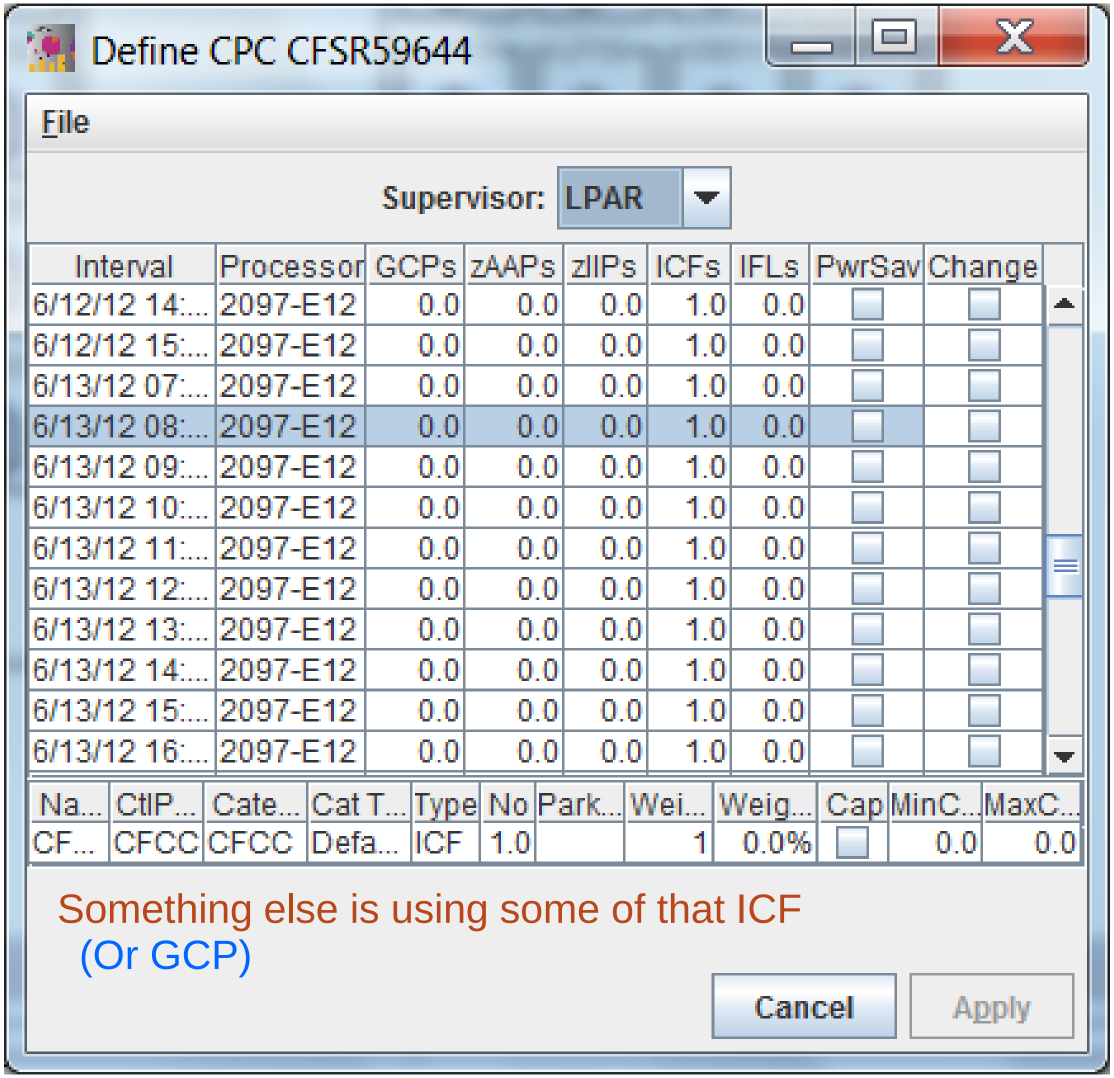
$$\frac{\text{(CF Busy + CF wait)}}{\text{Interval time}}$$

When EFFCP < Defined CPs

➡ **Shared Engines***

***Probably.**

If effcp>90%, it could be LP mgmt.



Interval	Processor	GCPs	zAAPs	zIIPs	ICFs	IFLs	PwrSav	Change
6/12/12 14:...	2097-E12	0.0	0.0	0.0	1.0	0.0	☐	☐
6/12/12 15:...	2097-E12	0.0	0.0	0.0	1.0	0.0	☐	☐
6/13/12 07:...	2097-E12	0.0	0.0	0.0	1.0	0.0	☐	☐
6/13/12 08:...	2097-E12	0.0	0.0	0.0	1.0	0.0	☐	☐
6/13/12 09:...	2097-E12	0.0	0.0	0.0	1.0	0.0	☐	☐
6/13/12 10:...	2097-E12	0.0	0.0	0.0	1.0	0.0	☐	☐
6/13/12 11:...	2097-E12	0.0	0.0	0.0	1.0	0.0	☐	☐
6/13/12 12:...	2097-E12	0.0	0.0	0.0	1.0	0.0	☐	☐
6/13/12 13:...	2097-E12	0.0	0.0	0.0	1.0	0.0	☐	☐
6/13/12 14:...	2097-E12	0.0	0.0	0.0	1.0	0.0	☐	☐
6/13/12 15:...	2097-E12	0.0	0.0	0.0	1.0	0.0	☐	☐
6/13/12 16:...	2097-E12	0.0	0.0	0.0	1.0	0.0	☐	☐

Na...	CtlP...	Cate...	Cat T...	Type	No	Park...	Wei...	Weig...	Cap	MinC...	MaxC...
CF...	CFCC	CFCC	Defa...	ICF	1.0		1	0.0%	☐	0.0	0.0

Something else is using some of that ICF

(Or GCP)

Shared vs Dedicated CF Engines

Why is this still so important?

- Typically low CF utilization
- Newer engines are usually faster
- Need to stay within one generation of z/OS hardware

→ hard to justify dedicated engines

Enhanced Methods for sharing CF engines

Dynamic ICF Expansion **z10 and below only**

allows a CF partition to have *both* shared and dedicated ICF engines.

Dynamic CF Dispatch **(for the test/dev CF)**

CF gives up control of the processor when there is no more work to do.

Thin Interrupt **(for test/dev and maybe some prod)**

link hardware generates an interrupt when something arrives in the link buffer

PR/SM can quickly dispatch a shared engine CF with new work,

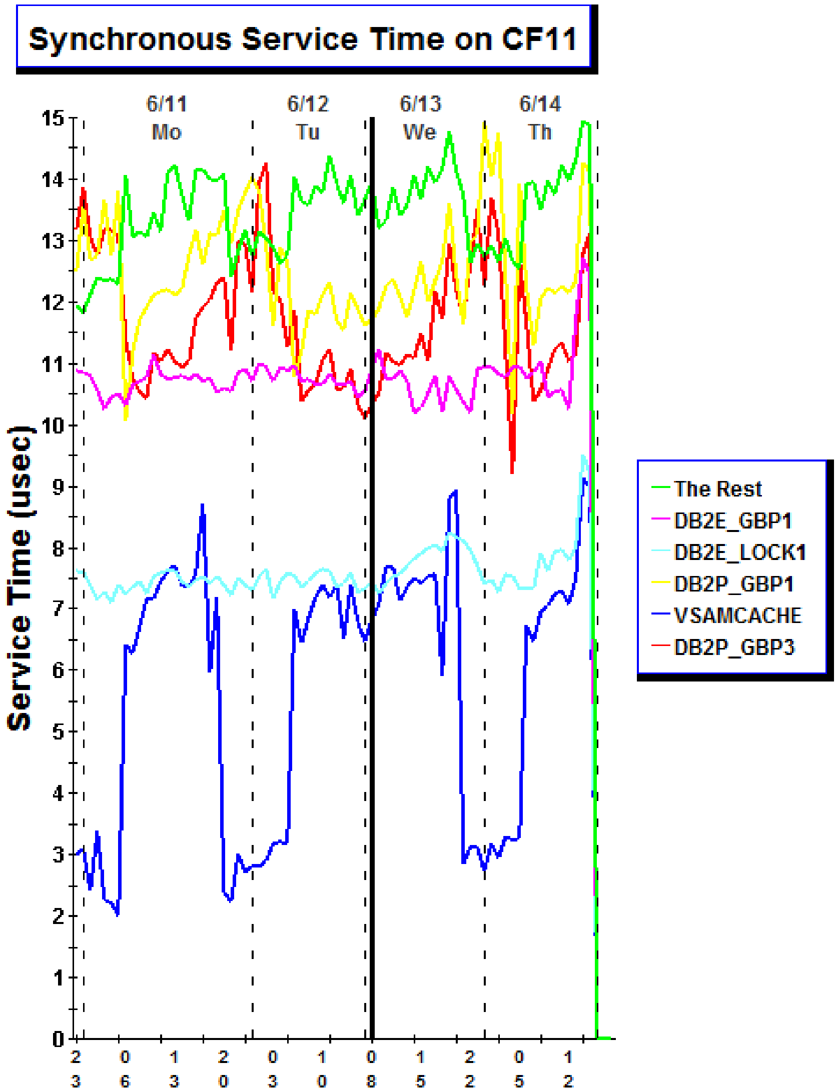
and the CF will drop when work is done.

Testcase : the effect of losing Dynamic ICF Expansion

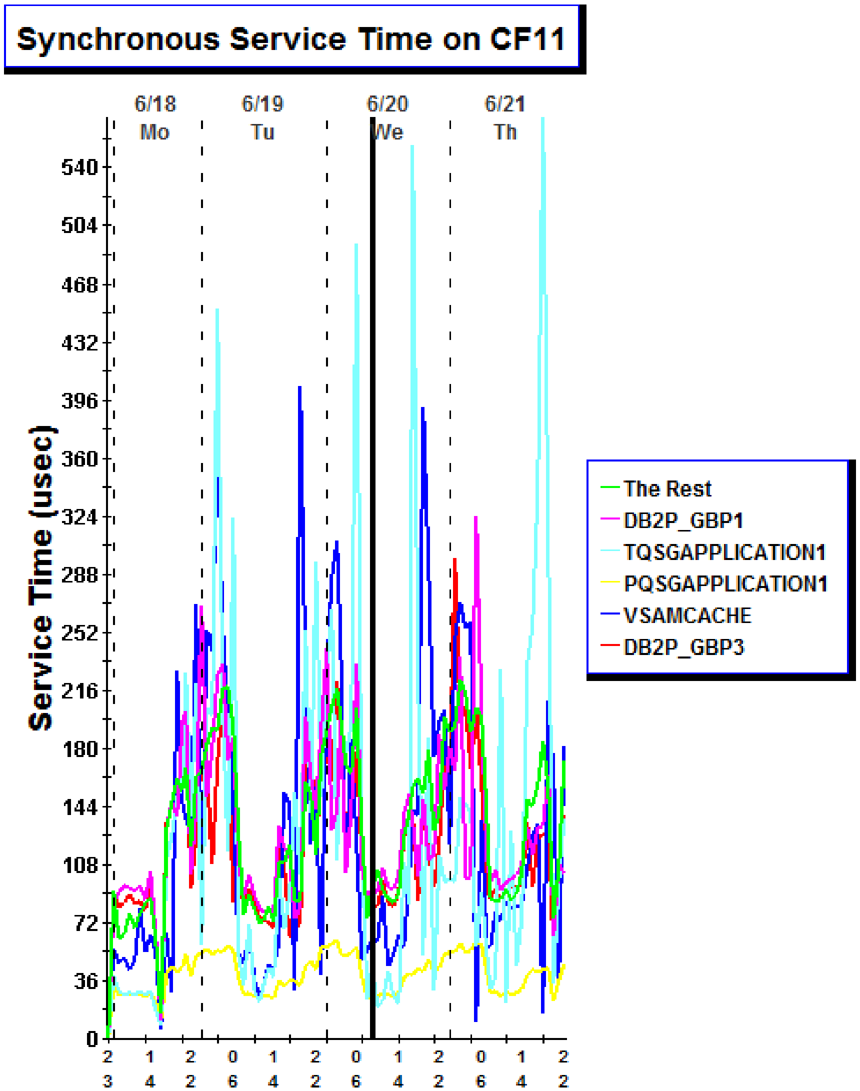
before : 1 dedicated and one share ICF in production, one shared ICF in non production on z10

after : 2 shared ICFs in for production, one shared ICF for non-production on z196

The effect of sharing ICF engines



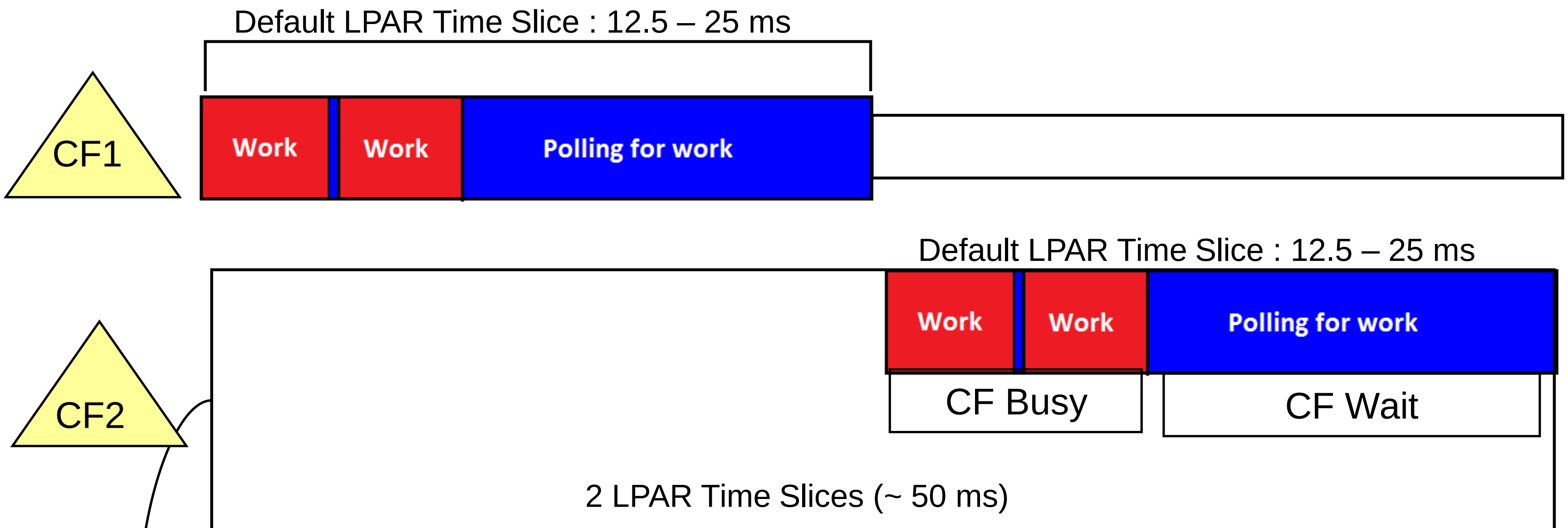
z10 + ICB4 + Dyn_ICF_Expansion



z196 + IFB12x + shared engines

“Regular” Sharing

Both CF1 and CF2 have DYNDISP=OFF (default setting)



Difference in Scale

microsec (sysplexOp) vs millisec(LPAR slice)

$$\text{EFFCP} = \frac{(\text{CF Busy} + \text{CF wait})}{\text{Interval time}}$$

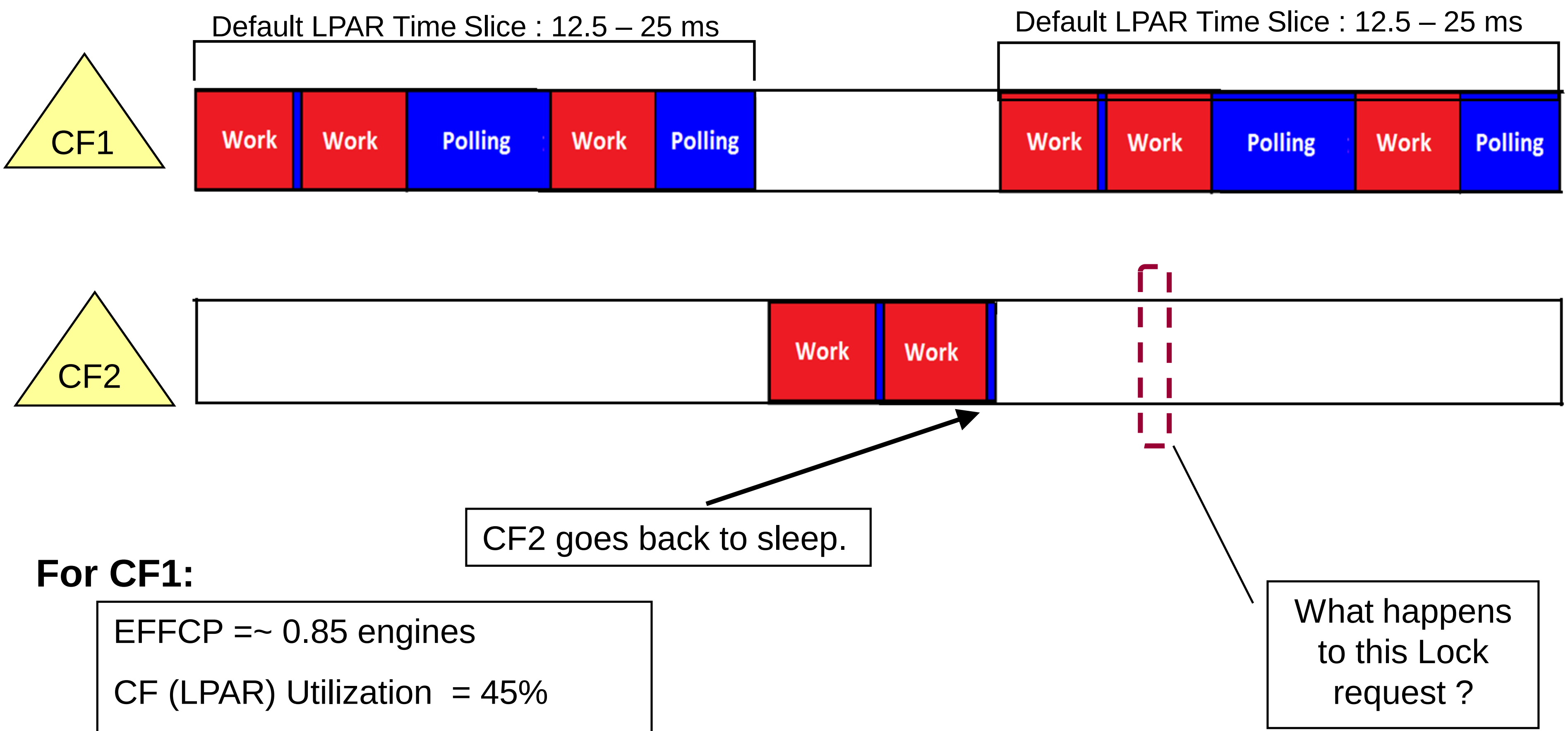
EFFCP =~ 0.5 engines

CF (LPAR) Utilization = 17.5%

CFCC Busy = 35%

Dynamic CF Dispatch

CF1 has DYNDISP=OFF & CF2 has DYNDISP=ON



For CF1:

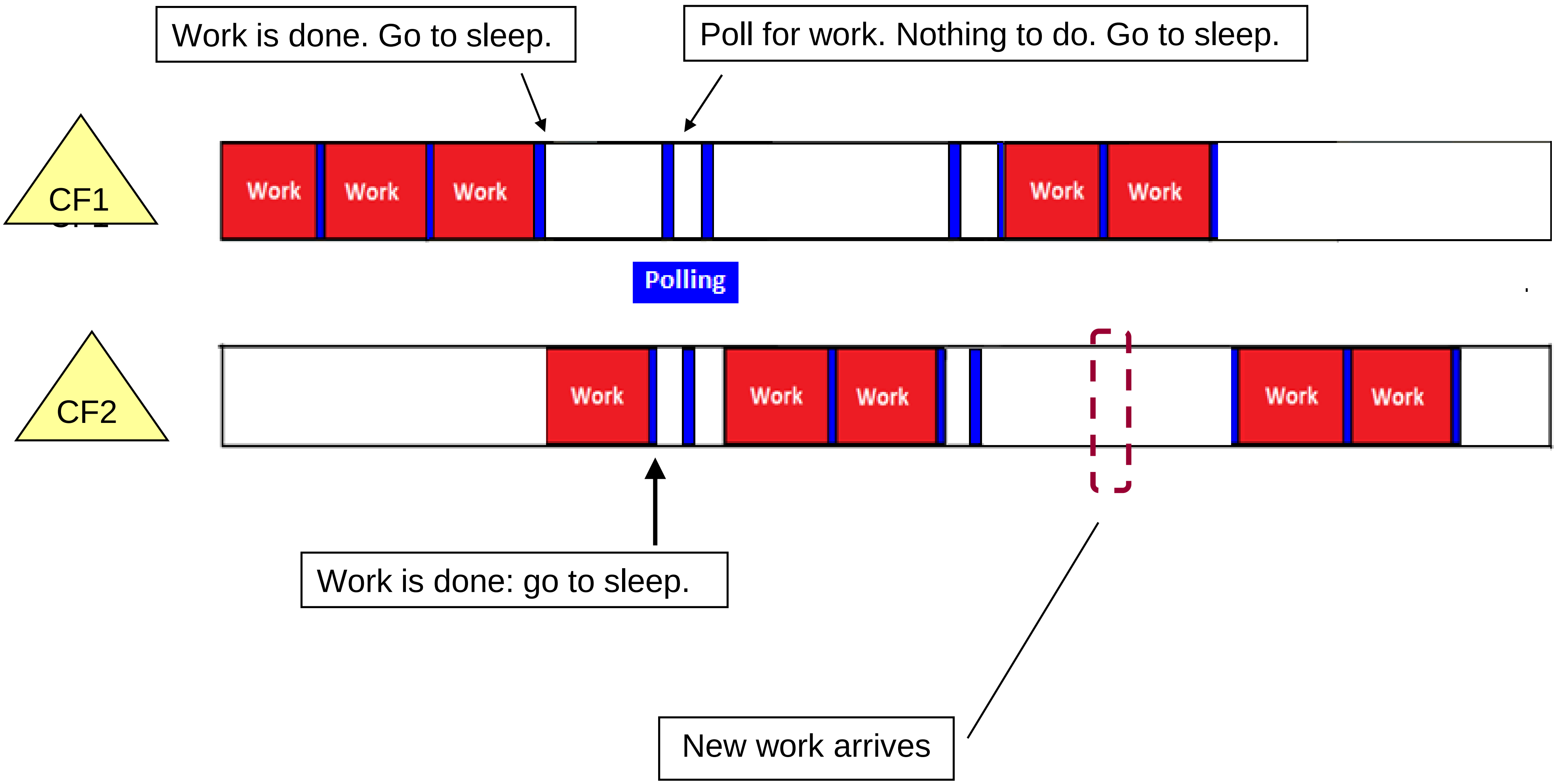
EFFCP =~ 0.85 engines

CF (LPAR) Utilization = 45%

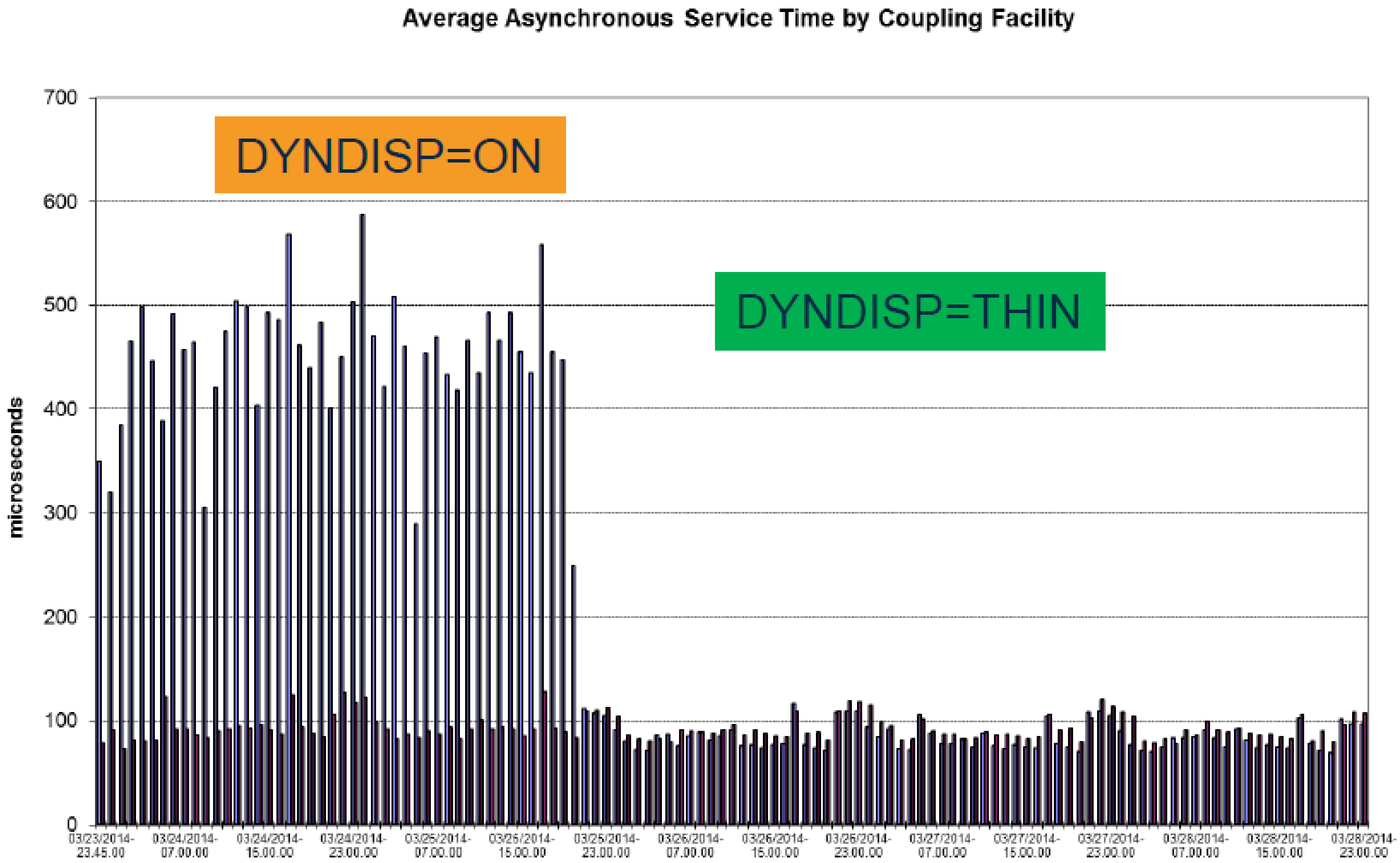
CFCC Busy = 56%

Thin Interrupt

Both CF1 and CF2 have DYNDISP=THININTERRUPT



Impact of Thin Interrupt



* Frank Kyne, Share Session 15602: The Skinny on Coupling Thin Interrupts

Shared vs Dedicated CF Engines

The Bottom Line is

Dedicated engines are still recommended

for production workloads for best responsiveness, reliability, and availability.

If the CF is always dispatched, it's always looking for new work.
Interrupts would not provide any response time benefit .

If ICF resource will be shared,

- All sharing CFs should set DYNDISP=THININTERRUPT
- LPAR weights & engines can be used to give more to the needy.

So how can zCP3000 help?

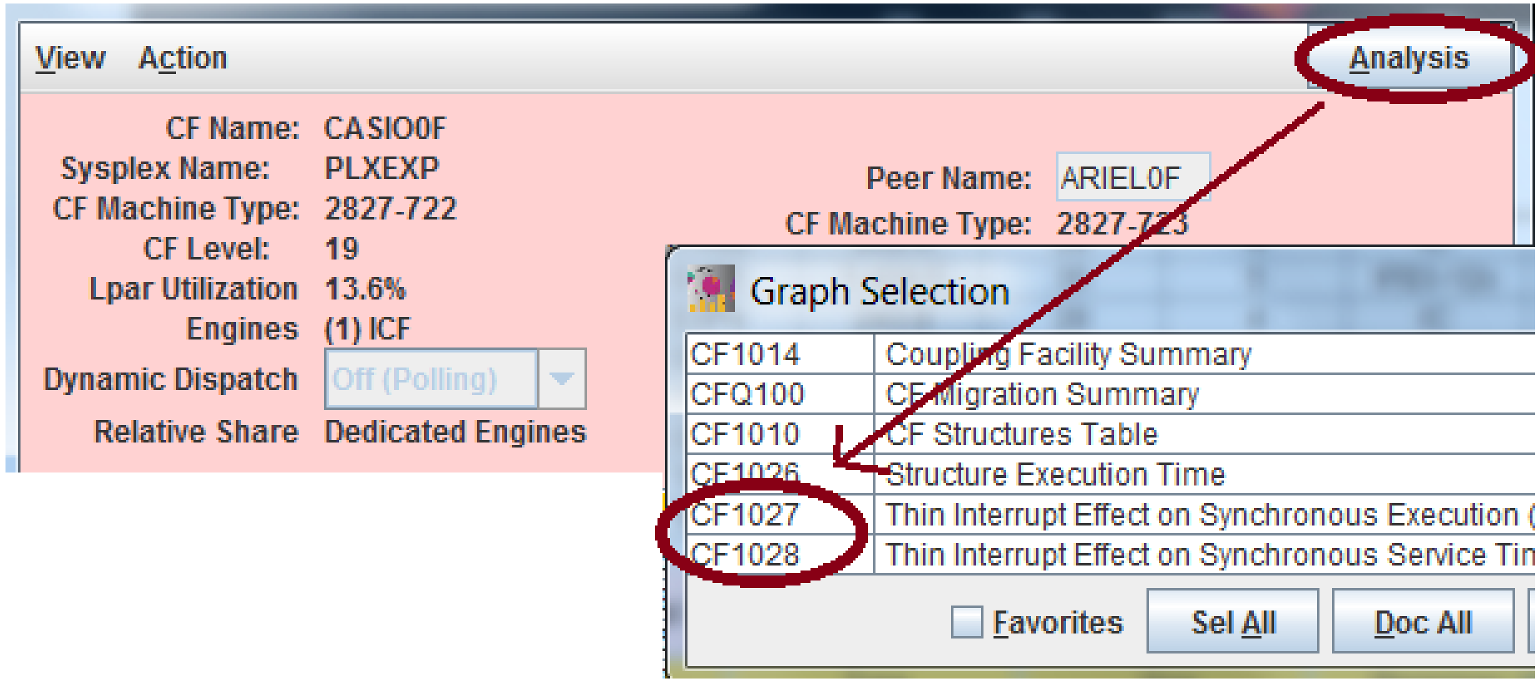
- Can identify service issues due to
  - shared CF engine in current configuration
  - old technology, or too great a difference

- Can identify issues that more/faster ICF won't fix
  - Lock contention
  - Subchannel busy
  - Poor service time due to slow links

- CF1027 report to compare service time dedicated vs Thin Interrupt.

- More changes zCP300 in 2015 to
  - CF Dispatch modelling (not just dedicated -> Thin)
  - z/OS effect of Thin Interrupt

Dedicated -> Thin Interrupt



Thin Interrupt performance for sync will be similar, a little bit worse for dedicated.

CF1027 Effect of Thin Interrupt

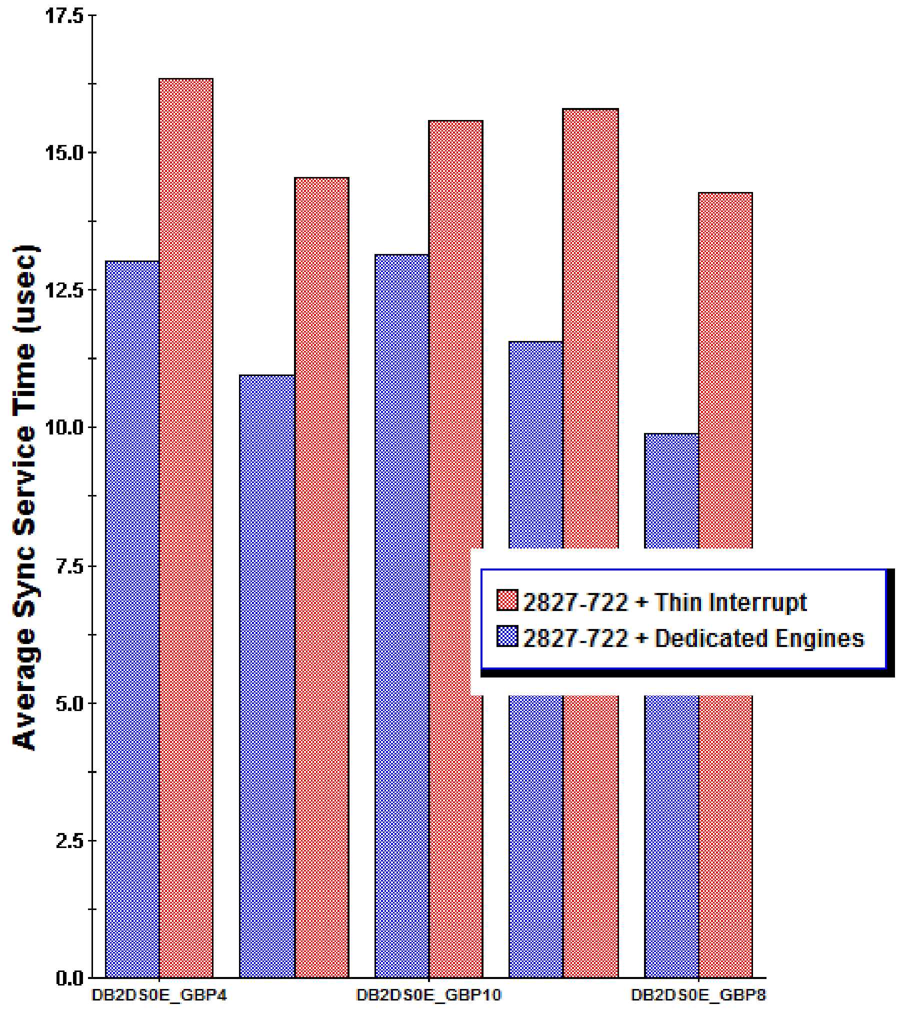


Table 1. Synchronous Service Time (microseconds)

Structure	Duplex	Current Configuration	Thin Interrupt
		2827-722	2827-722
DB2DS0E_GBP4	*	13.0	16.4
DB2DS0E_GBP2	*	11.0	14.6
DB2DS0E_GBP10	*	13.2	15.6
DB2DS0E_GBP49	*	11.6	15.8
DB2DS0E_GBP8	*	9.9	14.3

Agenda

Use zCP3000 to look at sysplex questions:

- Collect the right data for a study
- Learn about the data, identify current problem areas
- Modelling a new hardware configuration
- Using shared ICF and Thin Interrupt
- **z13 Link Migration Considerations**
- Link consolidation on InfiniBand and ICA
- Sysplex Aggregation Pricing eligibility

Coupling Link Connectivity

	Max Supported Coupling Links						
	z13	zEC12	zBC12*	z196	z114*	z10	z9
IC	32	32	32	32	32	32	32
ISC-3	n/a	48	32	48	48	48	48
ICA SR	32	n/a	n/a	n/a	n/a	n/a	n/a
HCA3-O LR (1x)	64	64	32	48	32	n/a	n/a
HCA3-O (12x)	32	32	16	32	16	n/a	n/a
HCA2-O LR (1x)	n/a	32	12	32	12	32	n/a
HCA2-O (12x)	n/a	32	16	32	16	32	n/a
HCA1-O (12x)	n/a	n/a	n/a	n/a	n/a	n/a	16
Max ext Links	*	104	72	104	72	64	64
Max chpids	256	128	128	128	128	64	64

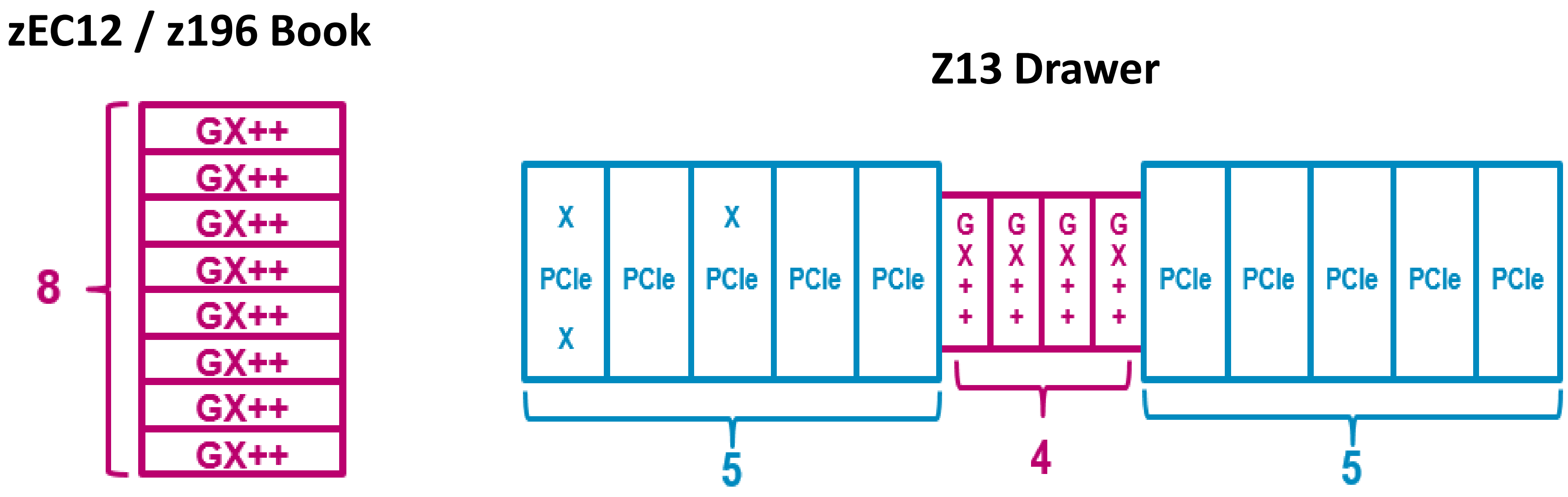
Link = Port ,   1, 2 or 4 ports per adapter

Coupling Link Configuration

	Fanout	Max Adapters (Features)	Ports per Adapter	Recommended chpids/adapter
ICA SR (z13)	PCIe	16	2	8
HCA3-O LR (1x)	GX++	16	4	8
HCA3-O (12x)	GX++	16	2	8
HCA2-O LR (1x)	GX++	16	2	8
HCA2-O (12x)	GX++	16	2	8
ISC-3 (zEC12)		48	1	1

In zCP3000, link = CHPID

z13 Link Migration problem



Configuring the same number of IFB coupling links on z13 as on z196/zEC12 may require additional CPC drawers

Is this an actual problem?

Use eConfig to determine the current installed adapters.
Define a new z13, then configure adapters.
zCP3000 will tell you how many books on the new machine.

Hardware Feature Codes

CPC66827

FC#	Adapter	LinkType	4 Books 2827-H89	Chpids Normal	Chpids Max
0172	ICA	ICA	0	0	0
0171	HCA3-O	IFB-12x	2	16	32
0170	HCA3-O LR	IFB-1x	2	16	32
0163	HCA2-O	IFB-12x	0	0	0
0168	HCA2-O LR	IFB-1x	0	0	0
0162	HCA2-C	Ficon	0	0	0
0167	HCA1-O	IFB-12x	0	0	0
3393	ICB-4	ICB-4	0	0	0
0217	ISC-3	ISC-3	0	0	0
0218	ISC-3	ISC-3	0	0	0

Cancel Apply

Hardware Feature Codes

Analysis

newZ13

FC#	Adapter	LinkType	2 Drawers 2964-N63	Chpids Normal	Chpids Max
0172	ICA	ICA	0	0	0
0171	HCA3-O	IFB-12x	3	24	48
0170	HCA3-O LR	IFB-1x	2	16	32
0163	HCA2-O	IFB-12x	0	0	0
0168	HCA2-O LR	IFB-1x	0	0	0
0162	HCA2-C	Ficon	0	0	0
0167	HCA1-O	IFB-12x	0	0	0
3393	ICB-4	ICB-4	0	0	0
0217	ISC-3	ISC-3	0	0	0
0218	ISC-3	ISC-3	0	0	0

Cancel Apply

Pull an eConfig report for the existing system to find the installed feature cod...

What if it is a problem?

- Adopt the new ICA coupling link for z13-z13 links.
- Prioritize GX++ slots for 1x distance coupling
- Eliminate FICON Express8 (also uses GX++)
- Eliminate or consolidate Infiniband links

  4 chpids per port for 12x,
  8 chpids across 4 ports

Agenda

Use zCP3000 to look at sysplex questions:

- Collect the right data for a study
- Learn about the data, identify current problem areas
- Modelling a new hardware configuration
- Using shared ICF and Thin Interrupt
- z13 Link Migration Considerations
- **Link consolidation on InfiniBand and ICA**
- Sysplex Aggregation Pricing eligibility

CF Link Consolidation

Prior to InfiniBand and ICA

- One link = one chpid = one feature
- A link could be shared because the chpid could be shared, between members of the same sysplex on the same mainframe, going to the same CF

With InfiniBand and ICA

- The physical link can carry multiple chpids
- Chpids can be from the same sysplex or a different one
- Each chpid can still be shared by partitions on the same mainframe going to the same destination CF

zCP3000 Terminology

- Link = channel path (chpid)
- Adapter = the physical path

IFB-1x : 8-16 chpids on 4 ports on HCA2-O LR or HCA3-O LR

IFB-12x : 8-16 chpids across 2 ports on HCA2-O or HCA3-O

IFB3 : max 4 chpids per port(2) on HCA3-O

CF Link Consolidation

- Typically multiple physical ICB4 or ISC3 links/adapters/chpids
- Often chpids are shared between 2-3 partitions
- Meaning you really have (chpids/partitions) number of ICB4 links.

- Minimum two HCA2-O or HCA3-O adapters
- Meaning that each machine will have 16-32 channels paths
- IFB3 is a protocol, not an adapter, requiring
  - HCA3-O adapter types on both ends
  - Maximum 4 chpids per port

IBM recommends

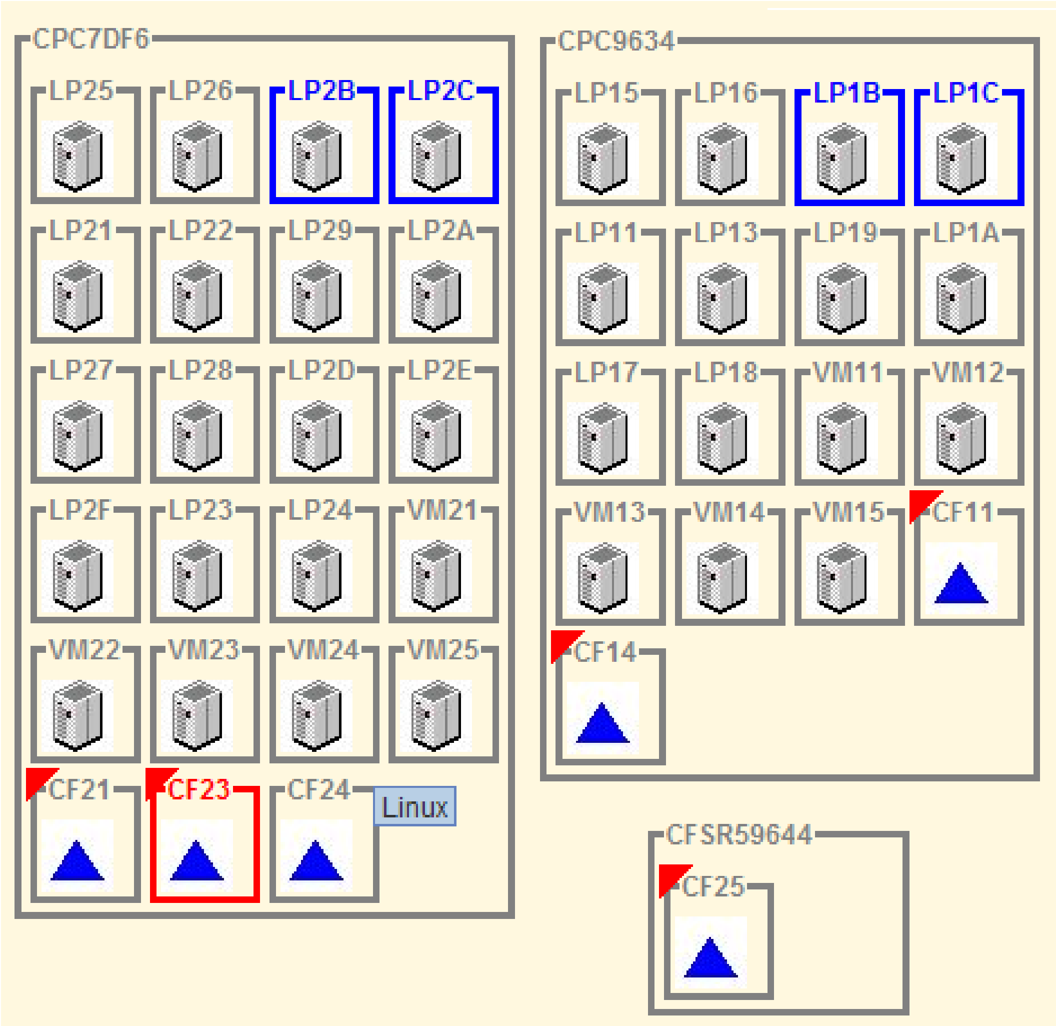
- One for one replacement of non-IFB chpids with IFB chpids
- Many to one replacement of non-IFB adapters with IFB adapters

Incomplete Configuration Information

Red Triangles mean something is not fully understood.

- Currently, all InfiniBand chpids come in with channel type “CIB”.
- Correct the current link definitions before proceeding:
  - Link type
  - Shared CHPIDs?
  - Distance, for remote CFs
- With CFCC 18, RMF will be able to differentiate various IFB link types and Detect if CIB link running "degraded“

* Not yet supported in zCP3000 *



Sysplex CF Link Summary Report

Summary of Links by CEC

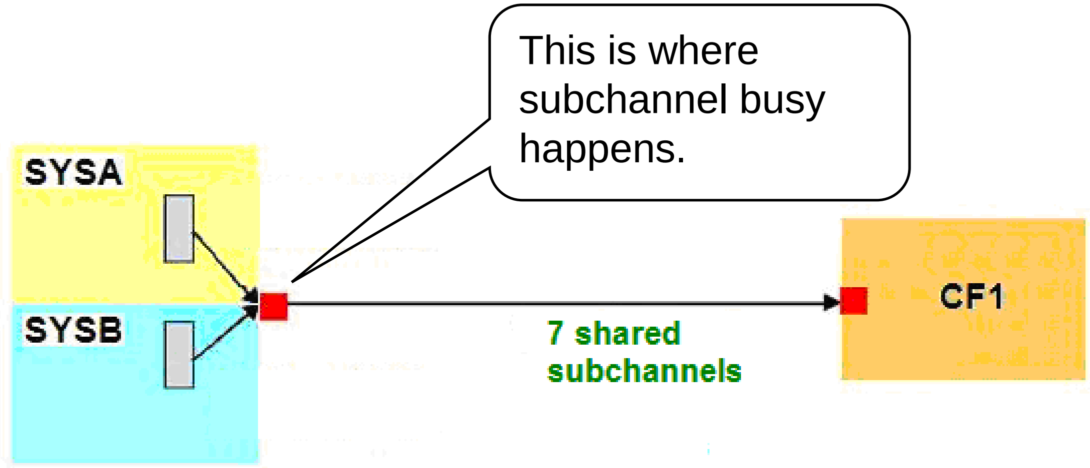
Processor ID	Link Type:	Defined Links	Configured CHPIDS	Minimum Adapters	Minimum Recommended Adapters
CPC7C8B5	IC	4	8	n/a	n/a
	ICB-4	8	8	8	8
CPC7C8C5	IC	4	8	n/a	n/a
	ICB-4	8	8	8	8

CEC Summary

Processor CPC7C8B5 SR# 7C8B5				
Partition	**CF/SYSID**	**Sysplex**	**OpSys**	**CF Links**
PROD1	SYSA	ISPLEXP	z/OS 1.11.0	(2) IC (2) ICB-4
PRODCF1	PRODCF1	ISPLEXP	CFCC Level 16	(4) IC (4) ICB-4
PROD3	SYSY	ISPLEXP	z/OS 1.11.0	(2) IC (2) ICB-4

Graph PLEX1008 on the Logical Sysplex Window

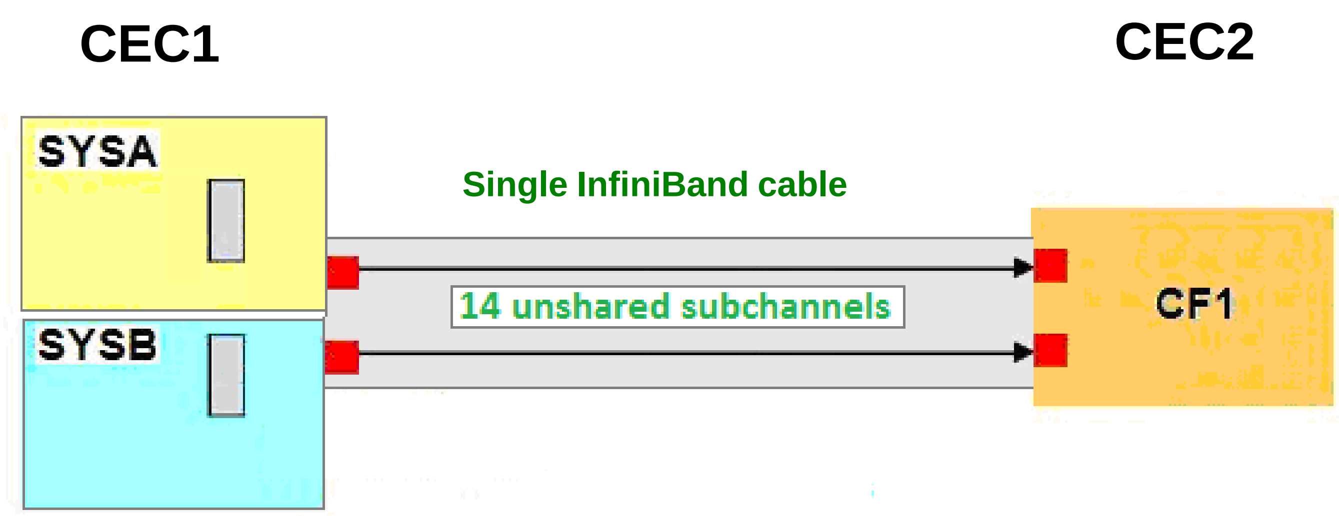
Shared Channel Paths



IOCP Definition of a shared cf link:

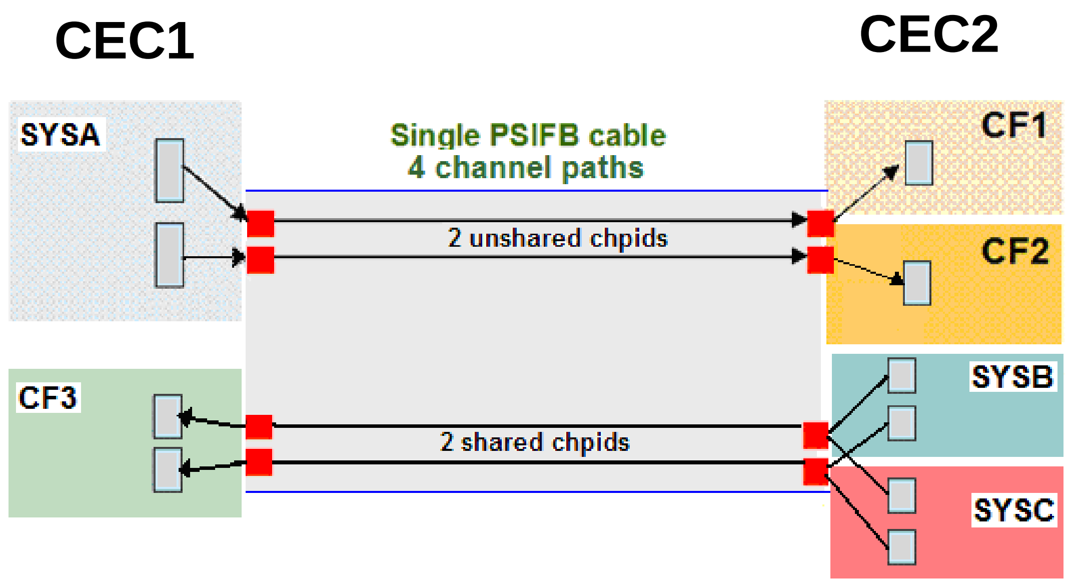
```
CHPID PATH=(CSS(0),F4),
      SHARED,
      PARTITION=((SYSA,SYSB),(=)),
      PCHID=101,
      TYPE=CFP
```

Multiple CHPIDs on a single InfiniBand Link



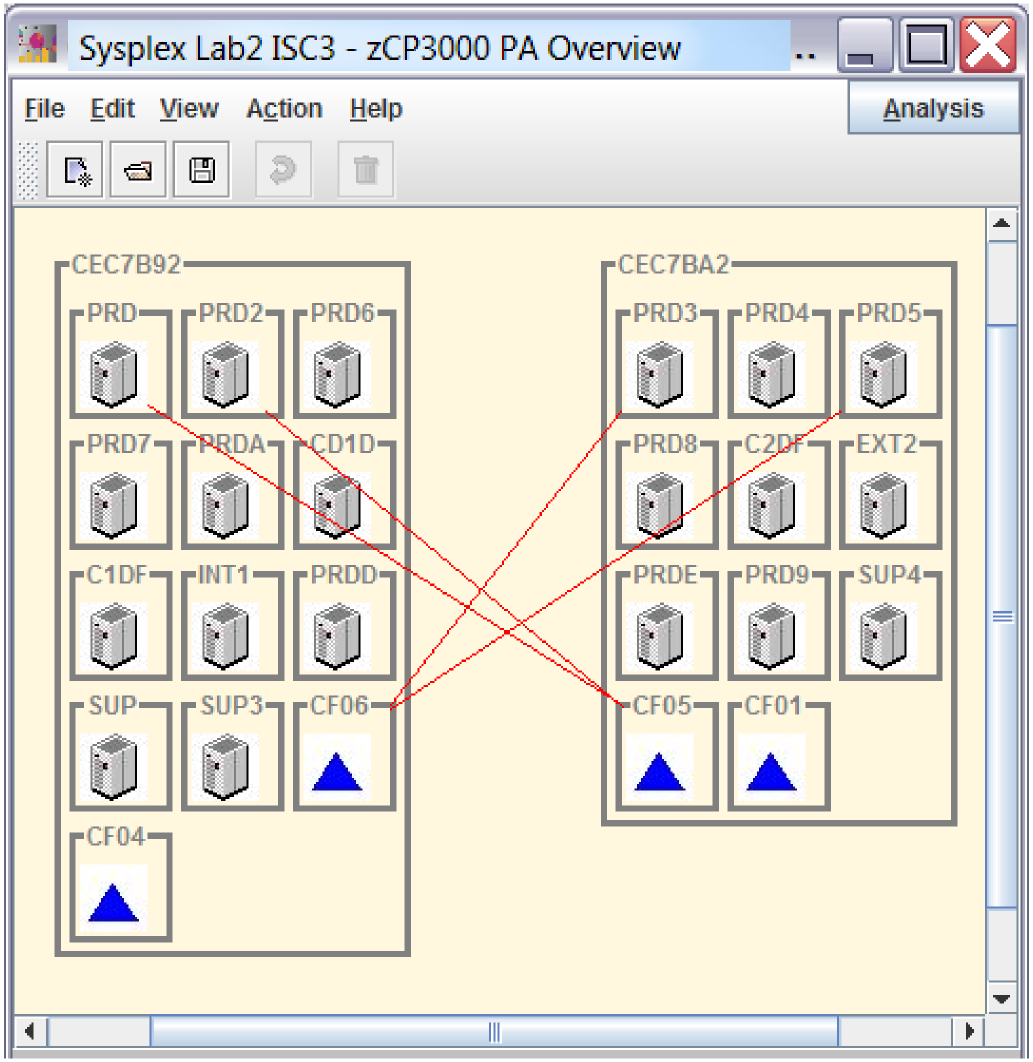
- This InfiniBand link carries 2 unshared chpids.
- Because the chpids are unshared, you should not see subchannel busy here.
- It connects a port on an HCA2-O adapter on CEC1 to a port on an HCA2-O adapter on CEC2. There is room for up to 6 more chpids on this cable.

Shared Chpids and Shared HCA3-O Adapters



- This InfiniBand link carries 4 chpids. You still have room to define up to 4 more on this port. There is also a 2nd port for each HCA2-O adapter, not shown.
- SYSA has one chpid each to CF1 and CF2; SYSA partition is not sharing these chpids with any other z/OS system on CEC1. The CF1 partition sees 1 chpid, and the CF2 partition sees 1 chpid.
- SYSB and SYSC each have one chpid defined to CF3. SYSB and SYSC each have 2 chpids to CF3, but they are shared. CF3 sees 2 chpids.

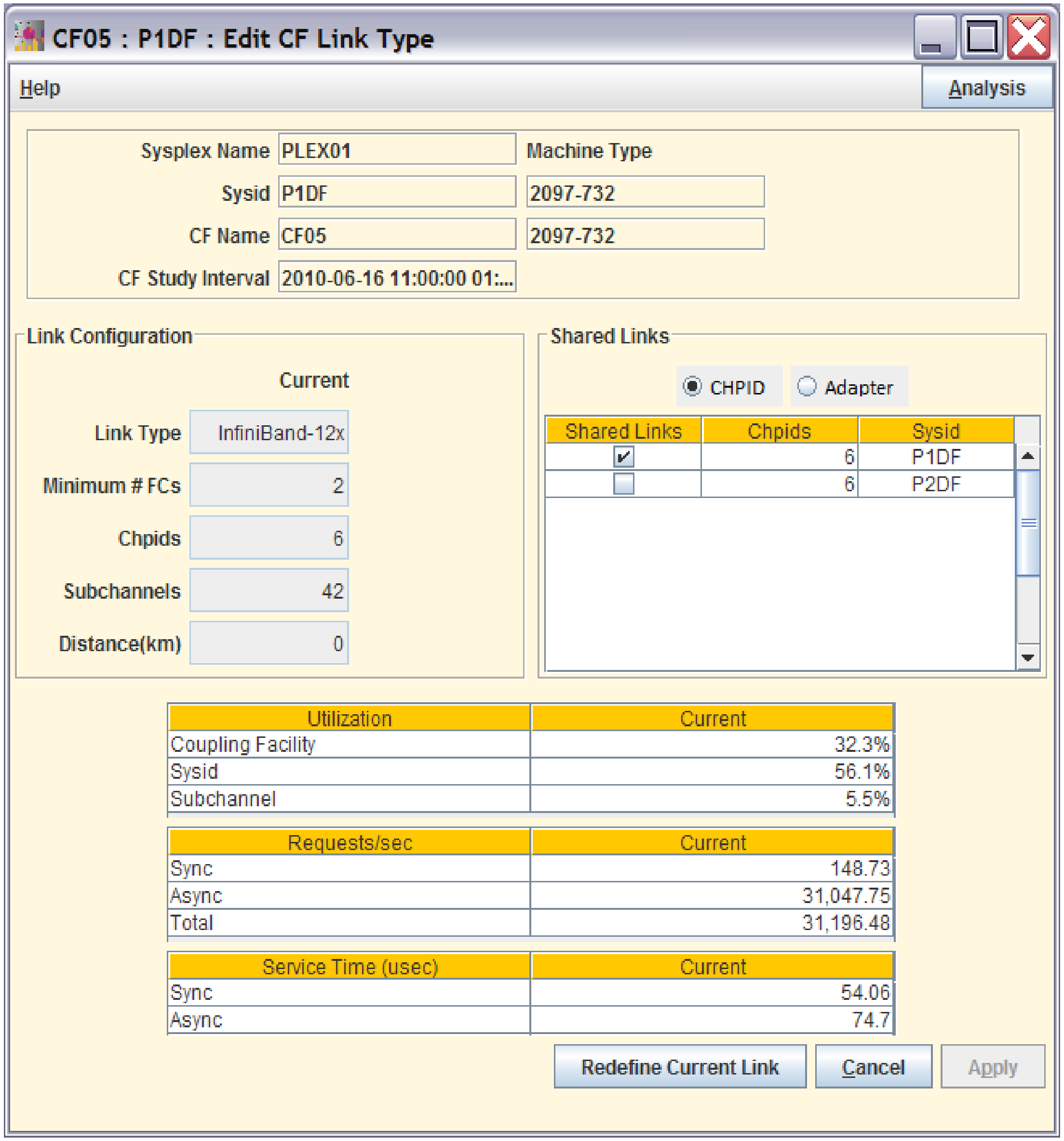
CF Link Consolidation – Shared CHPID



- 2 or more partitions on the same CEC
- talking to the same CF on another CEC
- same sysplex
- using the same chpid
- same type (ISC3, ICB4, IFB)

CHPID PATH=(CSS(0),F4),SHARED,
PARTITION=((DEVJ,DEV1,DEV2),(=)),
PCHID=101,TYPE=CFP

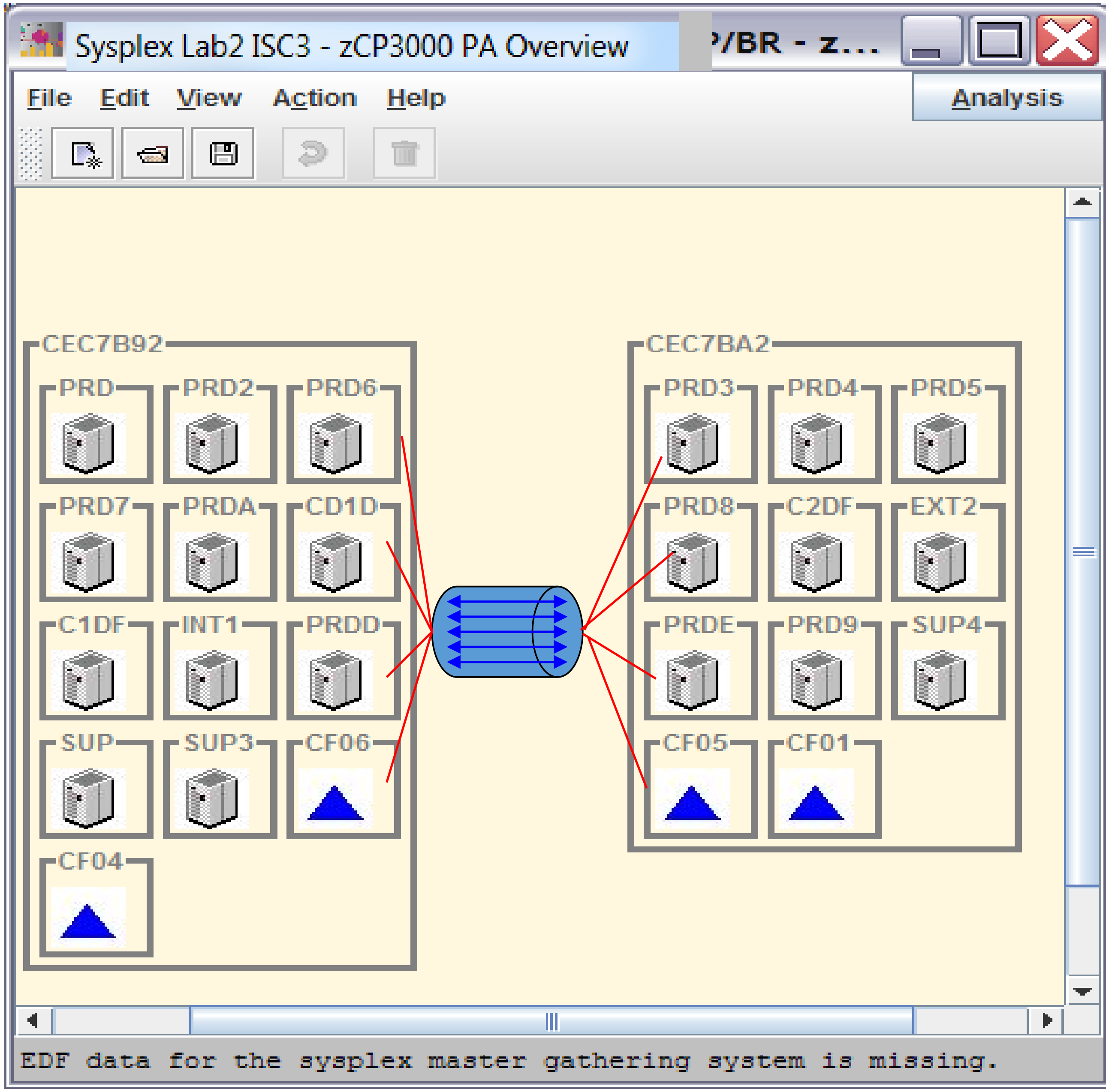
Defining Shared Channel Paths



You have to do this yourself

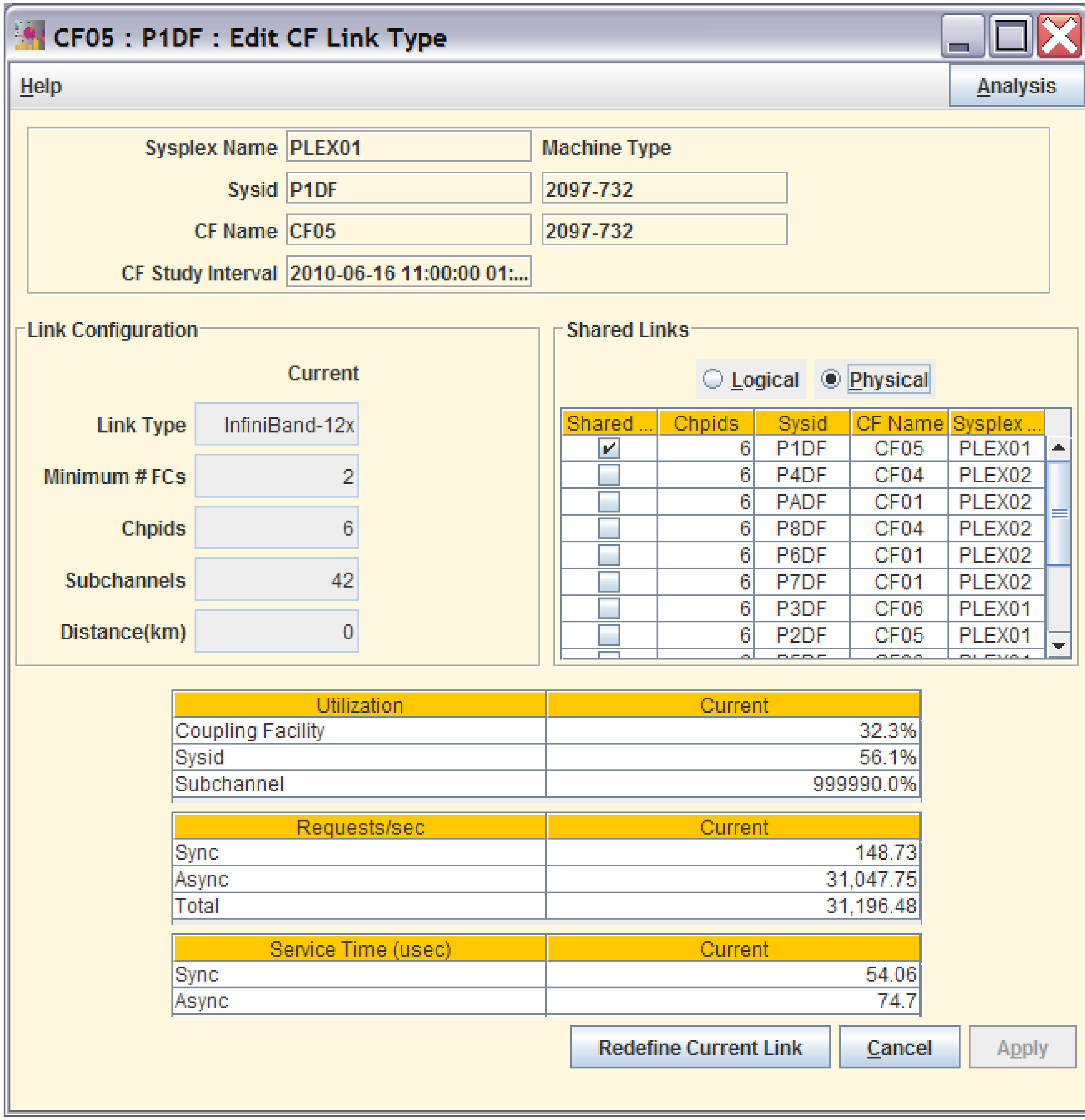
- CP3000 does not know if a link is shared.
- You *may* be able to tell from the CF HealthCheck.
- Shared Links table lists the partitions that could possibly share these 6 chpids.
- CF linktype must be fully defined and must match exactly.

Defining Shared Adapters



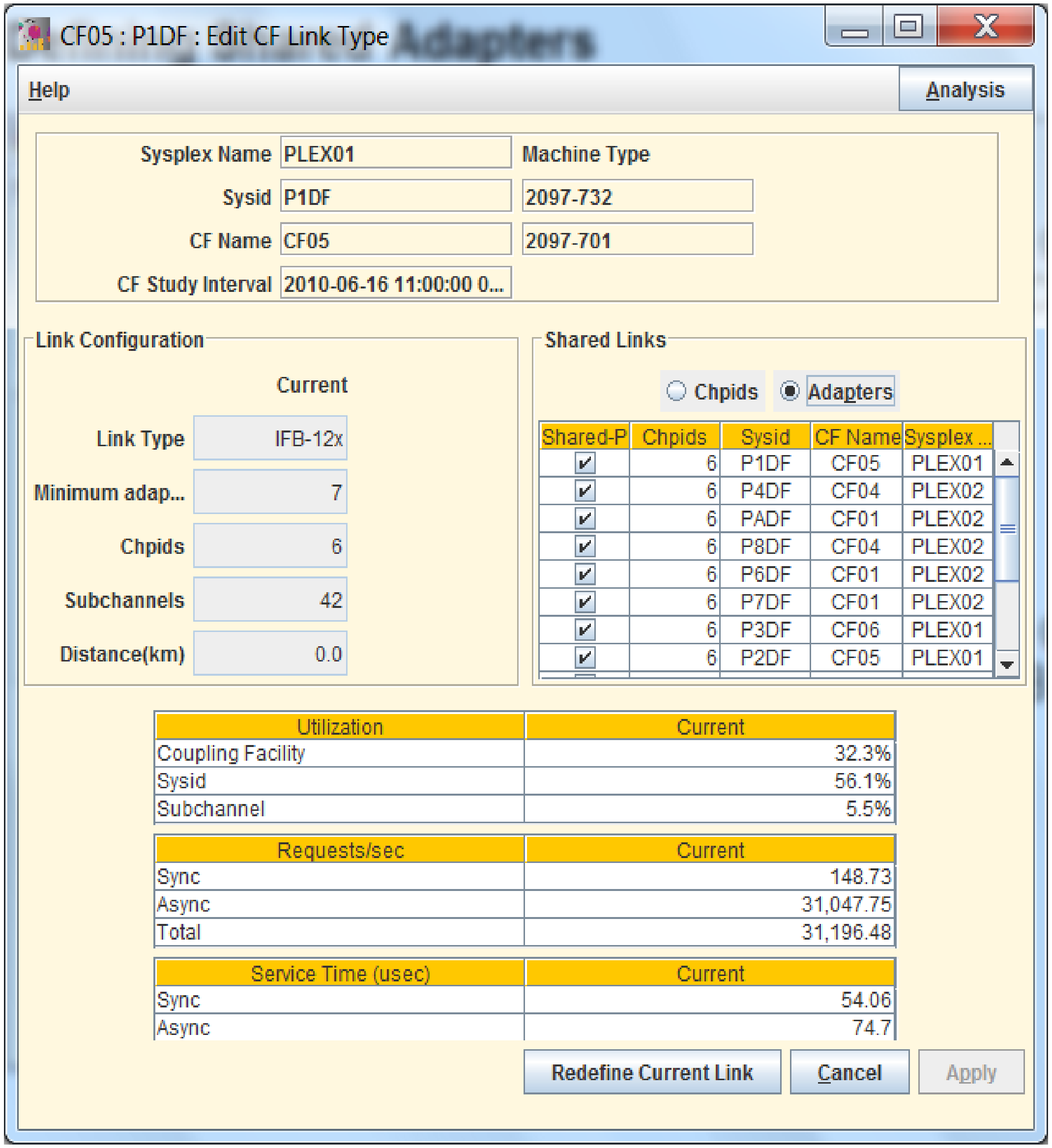
- 2 or more partitions on the same *pair* of CECs
- talking to any CF on either CEC
- same sysplex, or not
- using different chpids
- which run through the same IFB pipeline.

Defining Shared Adapters



- InfiniBand only
- Shared Links table lists the sysplex members and CFs on these 2 CECs that could run on the same physical InfiniBand feature.
- *NO* performance implications.
- P2DF could be *both* logical share and physical share.

Defining Shared Adapters



- 54 chpids running on 54 ISC3 links…

Replace with

- 54 chpids running on 7 IFB adapters.

Defining Shared Adapters

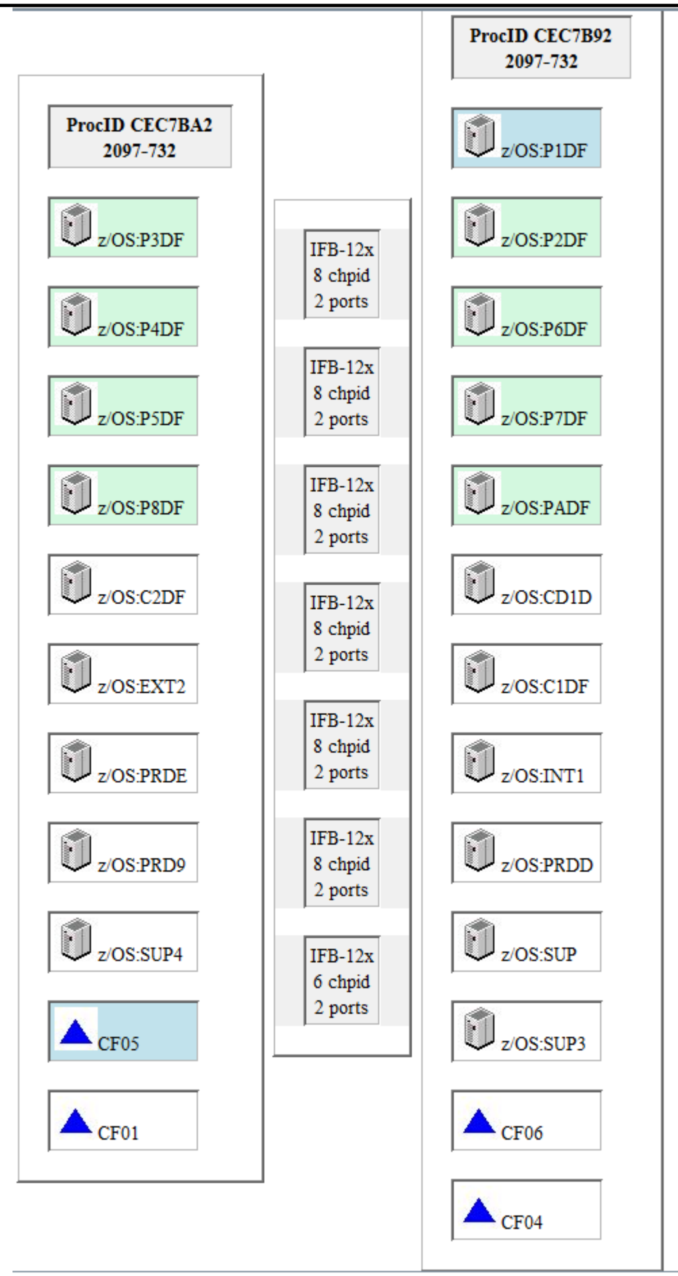
CF Link Topology Report
PLEX01:CF05:P1DF

CHPID Sharing

The P1DF system on CEC7B92 shares 6 chpids to the CF05 coupling facility on CEC7BA2 with 0 other partitions.

Adapter Sharing

54 chpids share 14 ports on 7 InfiniBand HCA2-O adapters.



Agenda

Use zCP3000 to look at sysplex questions:

- Collect the right data for a study
- Learn about the data, identify current problem areas
- Modelling a new hardware configuration
- Using shared ICF and Thin Interrupt
- z13 Link Migration Considerations
- Link consolidation on InfiniBand and ICA
- **Sysplex Aggregation Pricing eligibility**

Sysplex Aggregation Pricing

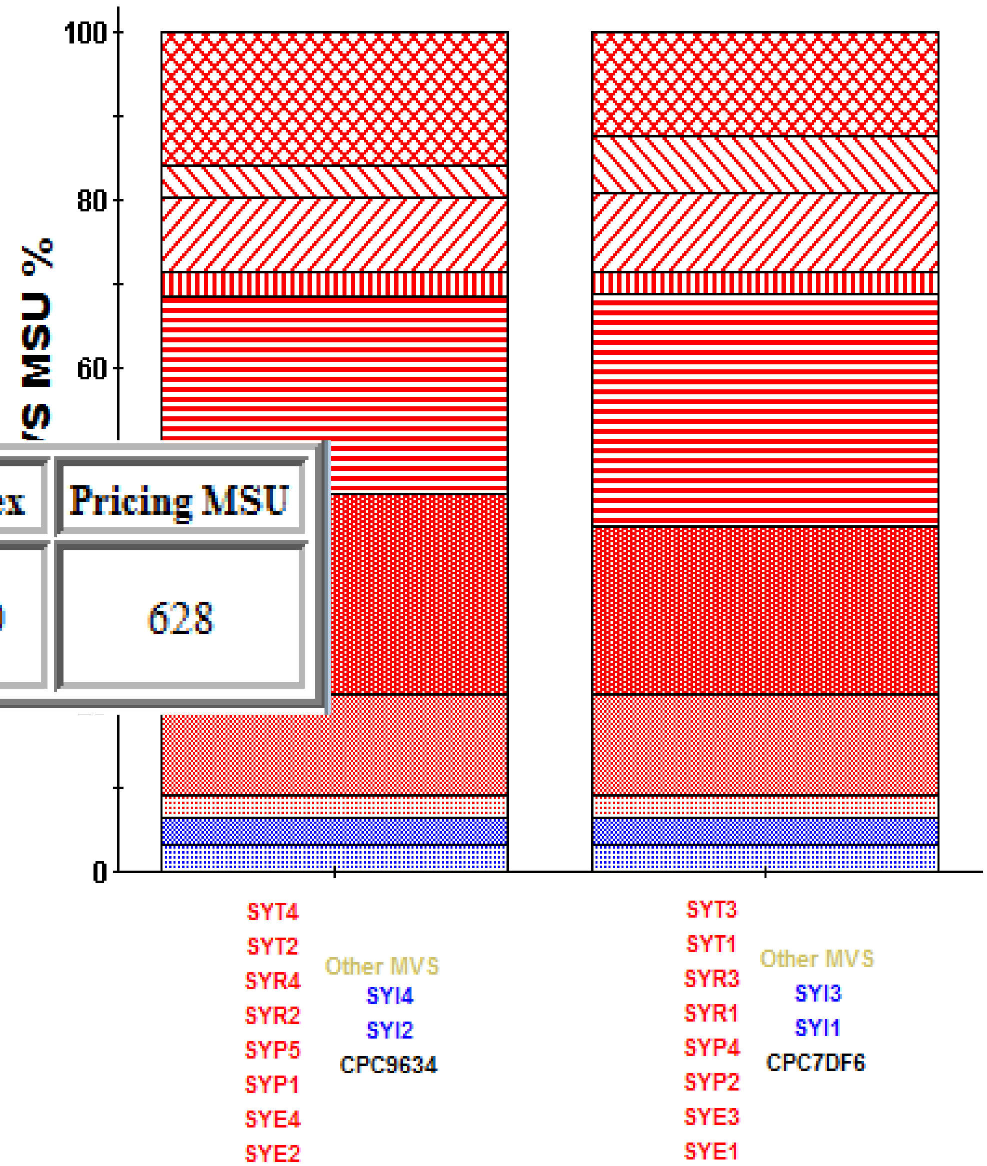
Be a hero

Show your customer how to save money on software licensing fees.

Sysplex Aggregation Pricing

Pay for software as if it was executing on a single large CPC.

CPC	CPU Model	Capacity (MSU)	PricingPlex	Pricing MSU
CPC9634	2097-703	312	PLEXP0	628
CPC7DF6	2097-703	316		



z/OS sysplex members use more than 50% of the machine.

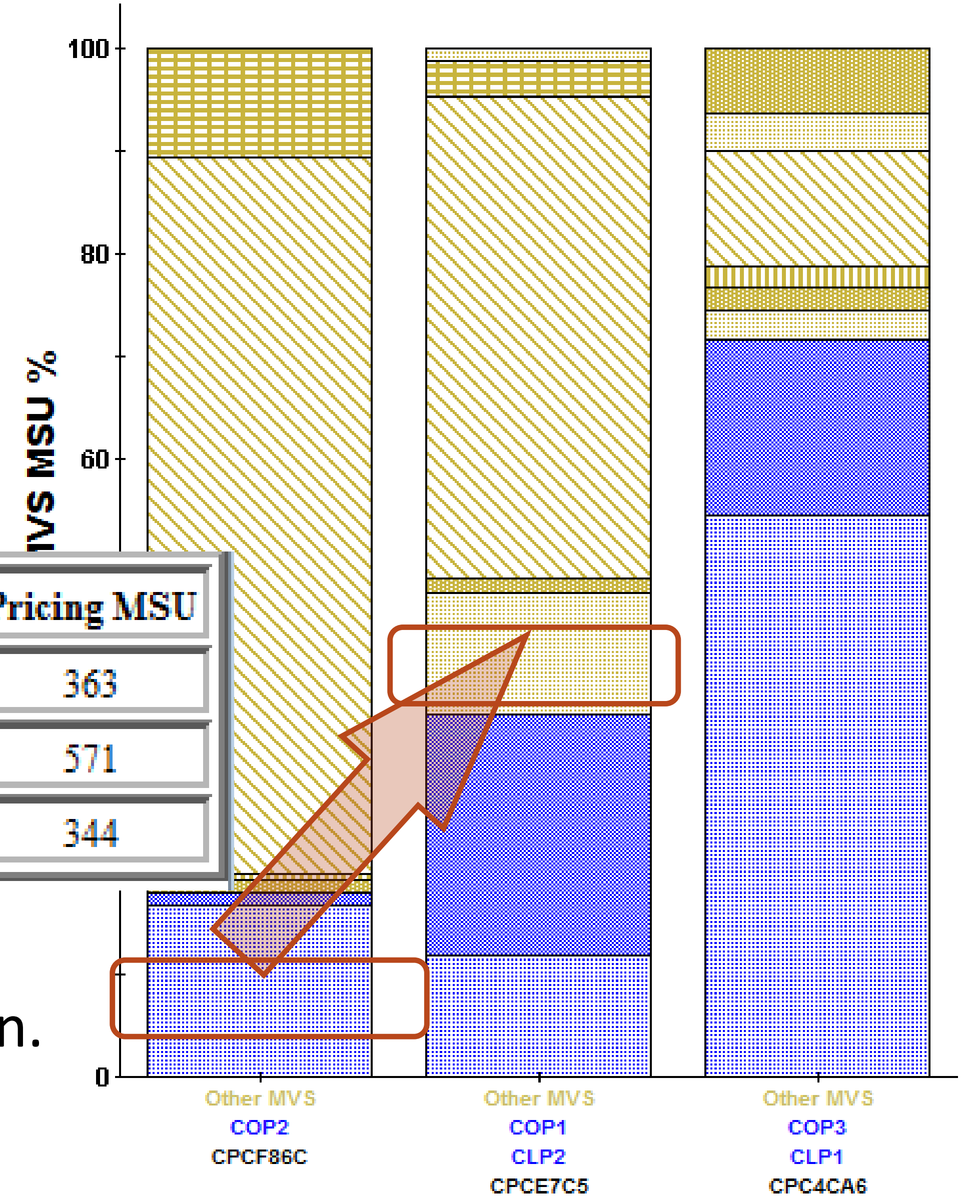
Graph PLEX1005 on the Logical Sysplex Window

Sysplex Aggregation Pricing

Sysplex members must use more than 50%.

Not currently eligible, but what if the z/OS in partition COP2 moved to CPCE7C5?

CPC	CPU Model	Capacity (MSU)	PricingPlex	Pricing MSU
CPCF86C	2094-705	363	Not Eligible	363
CPCE7C5	2097-706	571	Not Eligible	571
CPC4CA6	2817-604	344	Not Eligible	344



Use PlexCalc tool to verify qualification.

Graph PLEX1005 on the Logical Sysplex Window

Reference Slides

Sysplex Graph Recommendations

Probably in every study

- zCP3000 main view : Topology
- Sysplex Logical window : Sysplex Topology, Link Summary
- For each CF : CF Summary, CF Health Check, CF Migration Summary report

PA Mode

PA and QM Mode

QM only

Look for areas of interest

- Thin Interrupt Effect
- CF Structures Table
- CF Link Topology
- CF Link Summary, Link Migration Comparison
- Service Time for Synchronous Structures
- Request Rate by System over Time
- Request Rate by Request Type and System

Drill down in areas of interest

- CF Logical Utilization over Time
- Advanced CF (shared ICF)
- Subchannel Busy
- Delayed Requests
- Synchronous Service Time (Link)
- Lock Contention % of Requests
- Synchronous Intensity
- Busiest Structures with Queue Time

Index to recommended Graphs

zCP3000 main view

- INV1012 Topology

zCP3000 sysplex view

- PLEX1007 Sysplex Topology
- PLEX1008 CF Link Summary
- PLEX1005 Sysplex Aggregation

CF Window

- CF1014 CF Summary
- CF1000 Health Check Analysis
- CFQ100 CF Migration Summary report
- CF1027 Thin Interrupt Effect
- CF1010 Structures Table
- CF1001 Logical Utilization over Time
- CF1020 Advanced CF Utilization
- CF1017 Subchannel Busy
- CF1015 Delayed Requests
- CF1012 Service Time for Synchronous Structures
- CF1009 Request Rate by System over Time
- CF1004 Request Rate by Request Type and System
- CF1016 Busiest Structures with Queue Time

Structure Window

- STR1015 Lock Contention %
- STR1017 Synchronous Intensity

CF Link Window

- CFL005 CF Link Topology
- CFL001 CF Link Summary
- CFQL001 Link Migration Comparison
- CFL012 Synchronous Service Time

Need help?

IBM

Engage Global Techline

Techline System Design & Configuration

Geography

North America | Go!

Request Submission

Technical Sales Request Generator

eMail Techline

CPS Tools for System z Capacity Planning

Capacity Planning Support (CPS) tools are used
IBM System z processors. **These tools are fre**
Groups to discuss issues, ask questions, and c

The **Sizing and TCO Analysis Tools** page lists several restricted access
discern the total costs of ownership. The **Fit For Purpose Assessment**
that it will be running.

To engage help to run any of these tools, contact Global Techline

Program WebPage	Download	User's Guide
zSOFTCAP v4.3d	Download 07/23/2013	Download 07/23/2013
zPCR v8.4a	Download 09/23/2013	Download 07/23/2013

zMasters slide decks and replays:

Will be posted to TechDocs in late June.

15602: The Skinny on Coupling Thin Interrupts

https://share.confex.com/share/123/webprogram/Session15602.html

CPSTools Education & Downloads available at:

IBM - http://w3.ibm.com/support/americas/wsc/cpsproducts.html

BPs - **https://www-304.ibm.com/partnerworld/wps/servlet/ContentHandler/tech_PRS1762**

Contact: cpstools@us.ibm.com

Thank you!