Z Masters 2018

Solving Sysplex Problems with zCP3000

Session A26

Gretchen Frye
frye@us.ibm.com
IBM Washington Systems Center
z Systems Capacity Planning Tools - Developer

Agenda - Solving Sysplex Questions with zCP3000

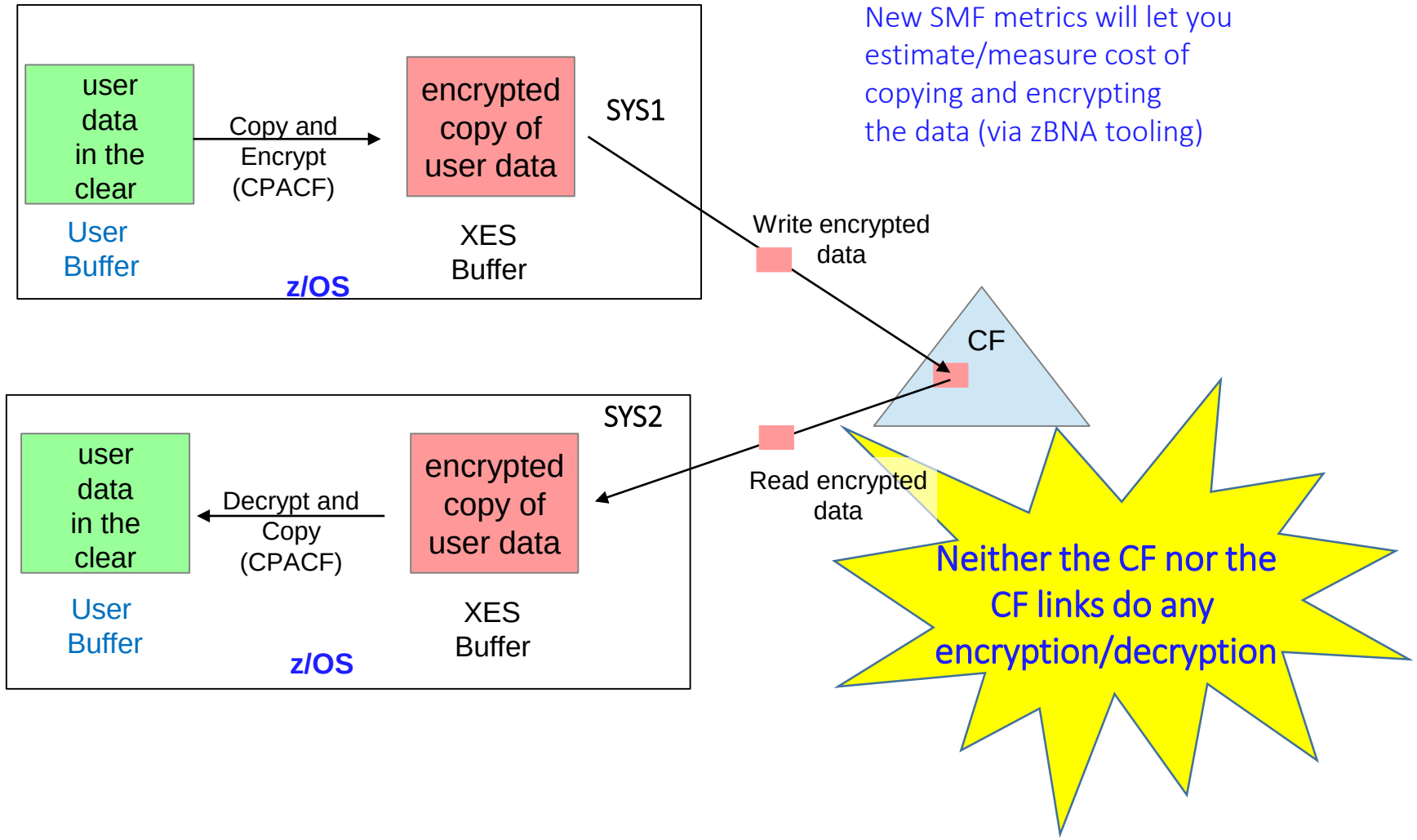
- **New Technology**
- Getting Started
- Understand problem areas and key resources
- Model new hardware or links or Thin Interrupt
- Coupling Link Considerations

Reference slides include links, graph recommendations, and references.

New Technology Support

- Support for the latest hardware
- Integrated Coupling Adapter (ICA)
- Coupling Express Long Range (CE LR)
- Coupling Facility (CF) Processor Scalability
  - Improved exploitation of nway
- CF Encryption
  - zCP3000 - CEC1064 CF Encryption Estimation
  - CEC1065 CF Encryption w/ alts
  - zBNA support to estimate impact from encryption

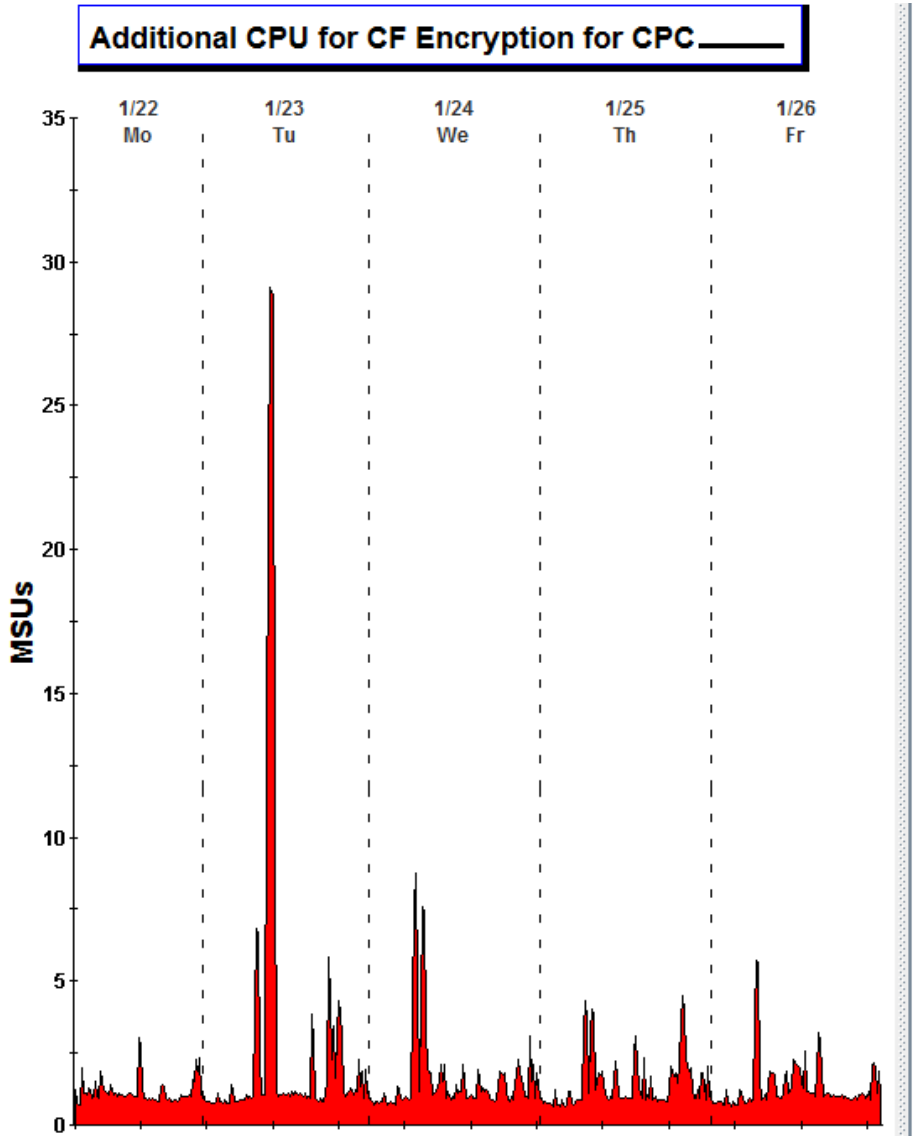
CF Encryption



New SMF metrics will let you estimate/measure cost of copying and encrypting the data (via zBNA tooling)

zCP3000 CF Encryption Report

CPC window
->Analysis
->CEC1064



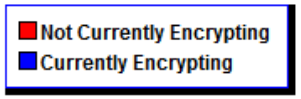
This graph represents the additional CPU utilization that will be required on a 2964-728 with 28 GCPs to encrypt and decrypt all data going to and from any attached Coupling Facility. It is based on new fields in the RMF 74 subtype 4 records which record the length of the data which was passed to the coupling facility. Be aware that for data which is currently encrypting that MSU consumption is already taking place. For data which is not currently encrypting, that additional MSU consumption will be consumed when encryption is activated.

The model contains data from 6 coupling facilities and 84 structures.

The top 5 structures and their required utilizations is illustrated in the table below:

Coupling Facility	Structure	Type	Encrypted Now	Average	Maximum
CF03N3I1	IXC_LIST_01	LIST	No	0	3
CF03N3I1	IXC_LIST_03	LIST	No	0	21
CF04N3I1	IXC_LIST_04	LIST	No	0	1
CF04N3I1	IXC_LIST_05	LIST	No	0	1
CF04N3I1	IXC_LIST_06	LIST	No	0	9

None of these structures are currently encrypted.



CF Encryption Cost Estimation

(for both zBNA and zCP3000)

Requires new SMF 74-4 fields.

z/OS support to generate new fields

- Will be included in z/OS 2.3 base
- Is rolled down to V2.2 (APAR OA51879)

Coupling Link Relative Speed

Internal Coupling Links			
Link type	Sub-chan	Latency (mics)	Speed (Mbs/sec)
IC-zEC12	7	1	9400
IC-z14, z196	**7**	**1**	**8900**
IC-z13	7	1	8500
IC-z10 EC	7	1	7500
IC-z13s	7	1	7300
IC-zBC12	7	1	7100
IC-z114	7	1	6500
IC-z9 EC	7	1	5000
IC-z9 BC	7	1	4000
IC-z10 BC	7	1	3200

External Coupling Links			
Link type	Sub-chan	Latency (mics)	Speed (Mbs/sec)
ICA (CS5) @ 70m	8	2	6000
IFB3	7	3	5000
IFB3 - zBC12, z13s	7	3	4000
ICA (CS5) @ 150m	8	2	3700
IFB3 – z114	7	3	2500
ICB4	7	4	1500
IFB 12X	7	8	1000
CE LR	8	13	800
IFB 12X – z9	7	8	600
IFB 1X	7/32	10	400
ISC3	7	12	200

z14 will be the last machine to support InfiniBand links.
z13 was the first to support ICA and CE-LR

Coupling Thin Interrupt

the arrival of an unsolicited signal on a coupling link generates an interrupt.

(starting with zBC12 and zEC12-GA2)

- Both z/OS and the Coupling Facility can exploit this.
- CF and z/OS exploitation are independent of each other.
- z/OS support mainly benefits low weight partitions
- For Coupling Facilities, it benefits CFs sharing CPU resource
- It does not benefit CFs with dedicated engines.

IBM recommends:

- Use dedicated engines for production CFs
- Otherwise enable Coupling Thin Interrupt for all CFs

Agenda - Solving Sysplex Questions with zCP3000

- New Technology
- **Getting Started**
- Understand problem areas and key resources
- Model new hardware or links or Thin Interrupt
- Coupling Link Considerations

Reference slides include links, graph recommendations, and references.

Getting Started

- **Plan the Study**
- Current performance concerns?
- Which sysplexes/CFs matter, and which don't?
- Should the study include overview information?

Always pull down the latest version of CP3KEXTR and zCP3000.

You may also need

- Partition names and lpar weights for CFs
- Current IOCP decks for all machines.

Collecting Model Input Data

Capture the *right* period of time, (representative of what you want to do)

- Collect 1 week of data.
- Assume 24x7 unless proven otherwise
- All members of “production” sysplexes
- **If you are looking at link consolidation, include all members, all sysplexes, including dev/test.**
- CPUMF (SMF 113) helps get the workload characterization right.
- All partitions that belong together as a production group

Too Much Data!

2018 Testcase:

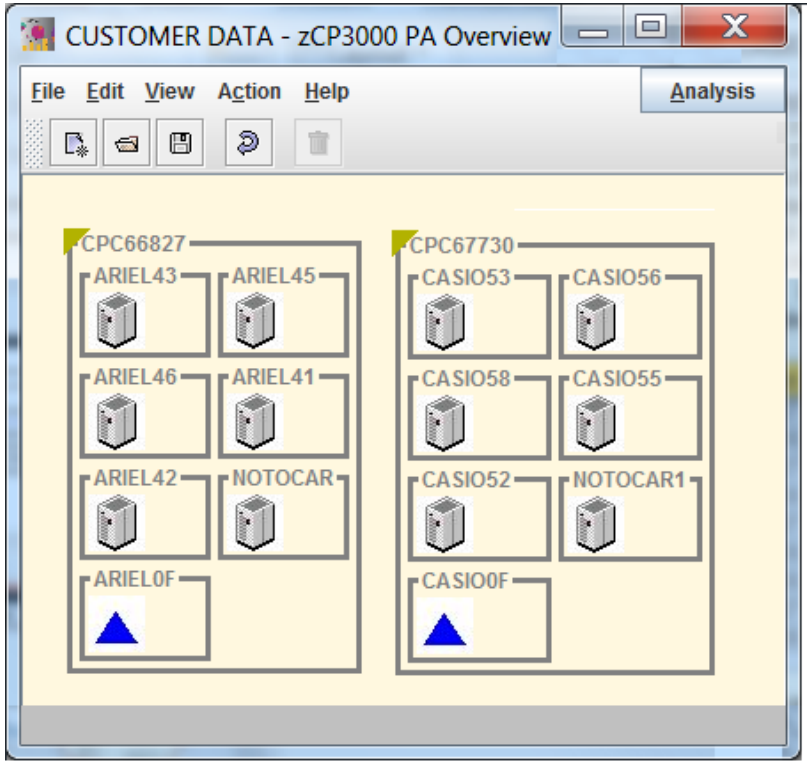
73 EDFs @ 3.8MB average

273 MB of input data

- Collect one week, not two
- Shut down everything else.
- File->Trim option to delete I/O (BCU, BCUT, ACT, PATH) from EDF*
- After load, delete uninteresting shifts (weekend, night), then save 3PA
- After load, delete uninteresting intervals, then save 3PA
- Ask CPSTOOLS@US.IBM.COM for help

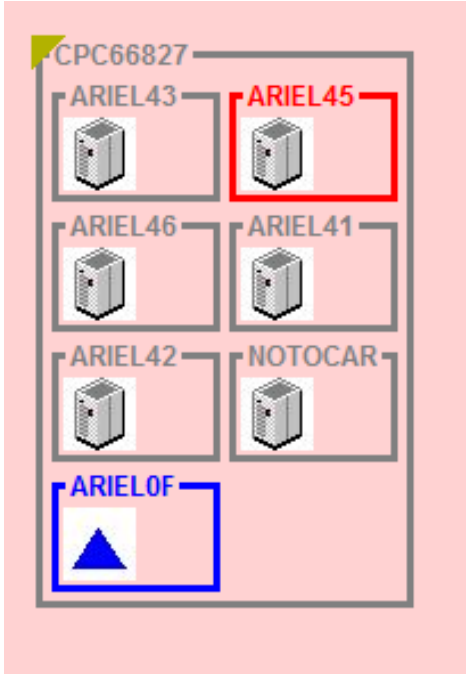
zCP3000 “modes”

PA – Performance Analysis



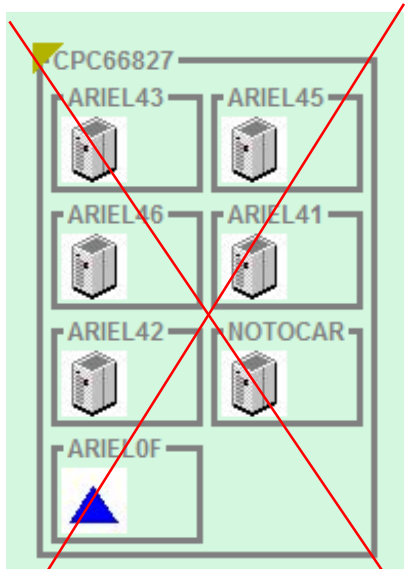
PA – detail on the current, alt CPUs, Healthchecks, sysplex configuration

QM – Quick Migration



QM - change the configuration: new CPU type, move LPARs, new link type, thin interrupt, data center distance.

CP – Capacity Planning



CP – workload growth; not really for sysplex modelling.

Correct the configuration when necessary

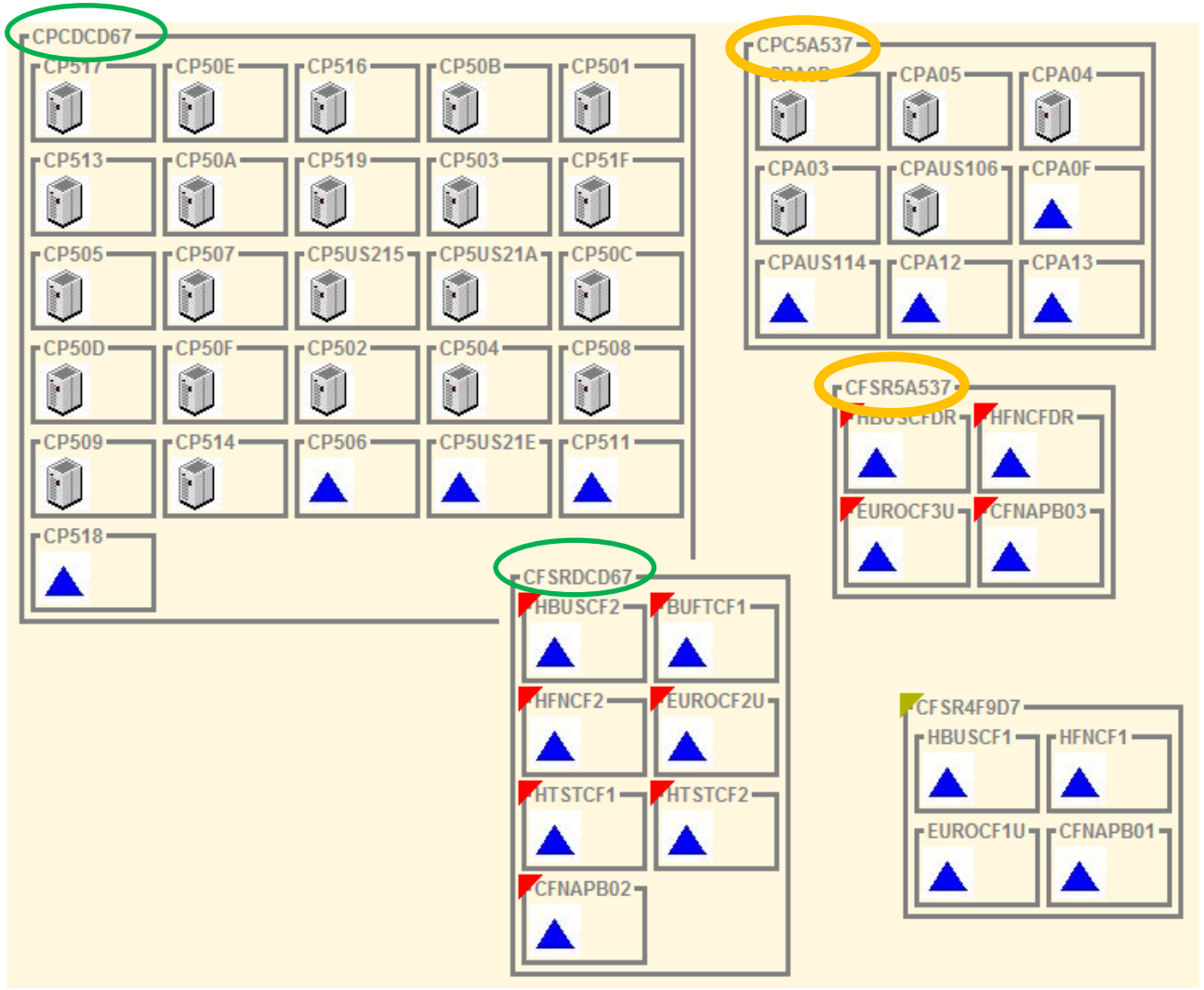
Standalone CF problems

- If a standalone CF has the same serial number as a CPC, these CF partitions ended up on the wrong CPC.
- Internal coupling links on CFSR machines will be red triangled, if a CF is in the wrong place. However, there may be no red triangles, if there are no IC links.
- Weight and engine type will be correct for that partition on the CPC, but not on the CFSR, so *be careful to move the CF to the right LP.*

Link definition problems

- **No longer an issue since z/OS 2.1 SMF 74-4 Channel Path Information**
- Invalid IC link? The CF is in the wrong place.

Correct the Machine layout when necessary



CPCDCD67
CFSRDCD67
CPC5A537
CFSR5A537
CFSR4F9D7

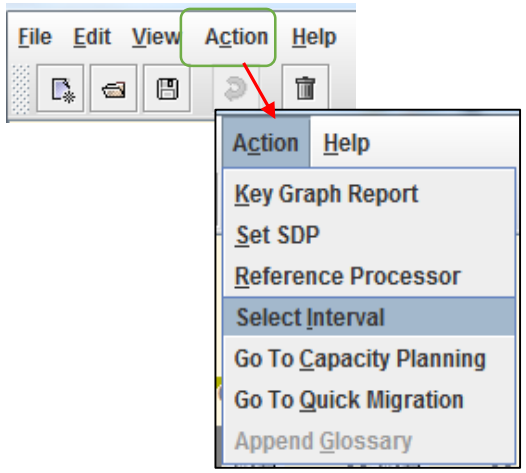
Notice the very similar serial numbers.

There are really only 3 machines in this model.

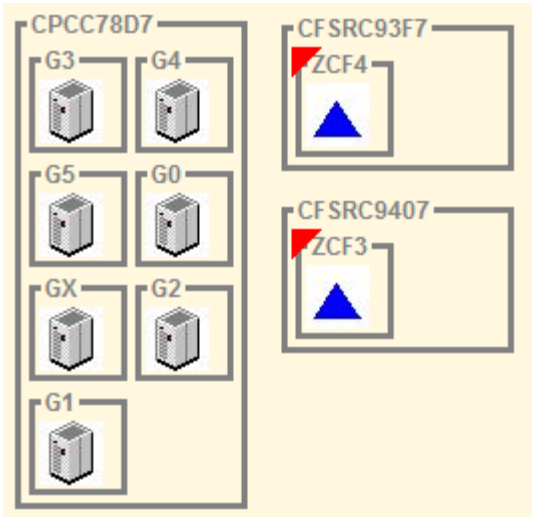
The red triangles are because of IC links on that "extra" machine.

Do Some Preliminary Stuff

1) Select a Study Interval



2) Fix CF red triangles



Open the CF to see what the problem is:

Partition	SYSID:	Reqs/sec	SubChann...	CF Links:	Link Type:	Km	Path Utiliza...
CPCC78D...	CSG3	41.4	28	4	[IFB-1x]	0.0	0.0%
CPCC78D...	CSG4	29.6	28	4	[IFB-1x]	0.0	0.0%
CPCC78D...	CSG5	25.6	28	4	[IFB-1x]	-1.0	0.0%
CPCC78D...	CSGT	8.0	28	4	[IFB-1x]	0.0	0.0%
CPCC78D...	SY2	386.7	28	4	[IFB-1x]	0.0	0.0%
CPCC78D...	VIPA	376.0	28	4	[IFB-1x]	0.0	0.0%

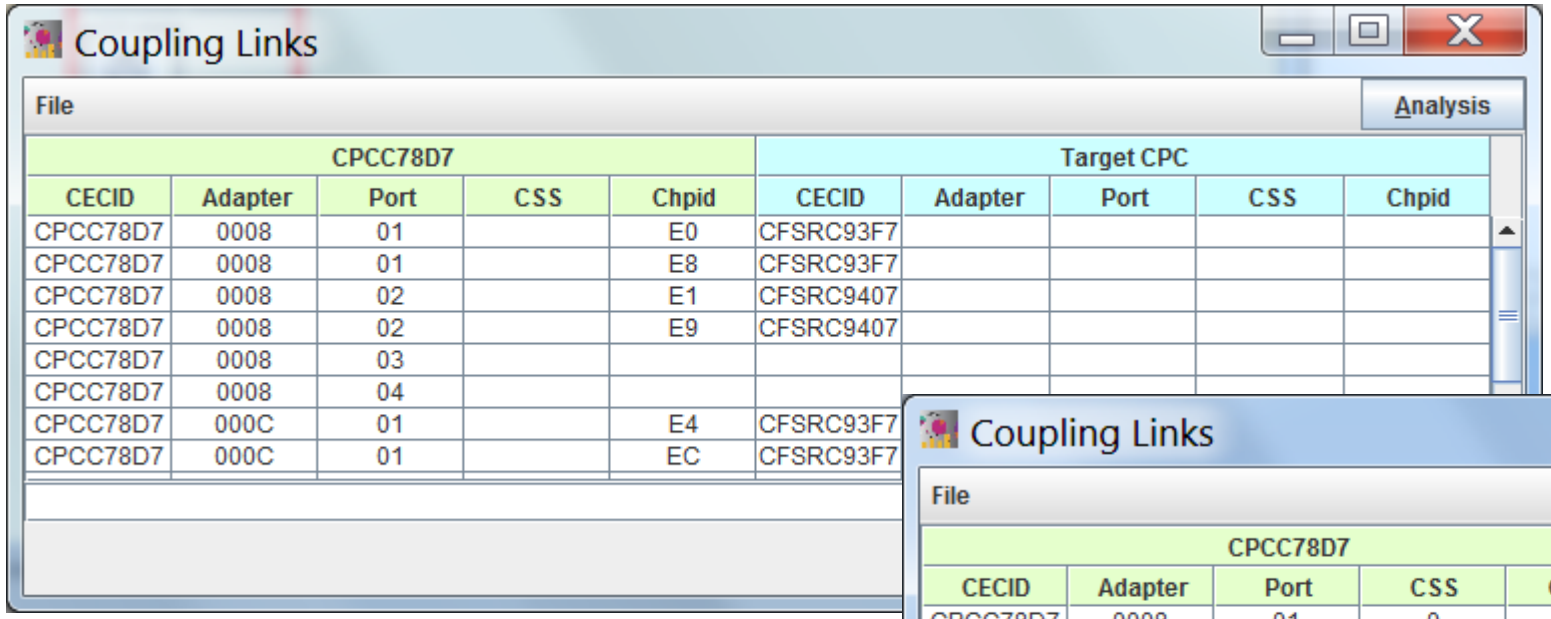
CF Study Interval 2/2/18 09:00 | Link Quick Fix | Set KM (all CFs)

Correct the machine layout

Set Distance

Fix Link Definitions

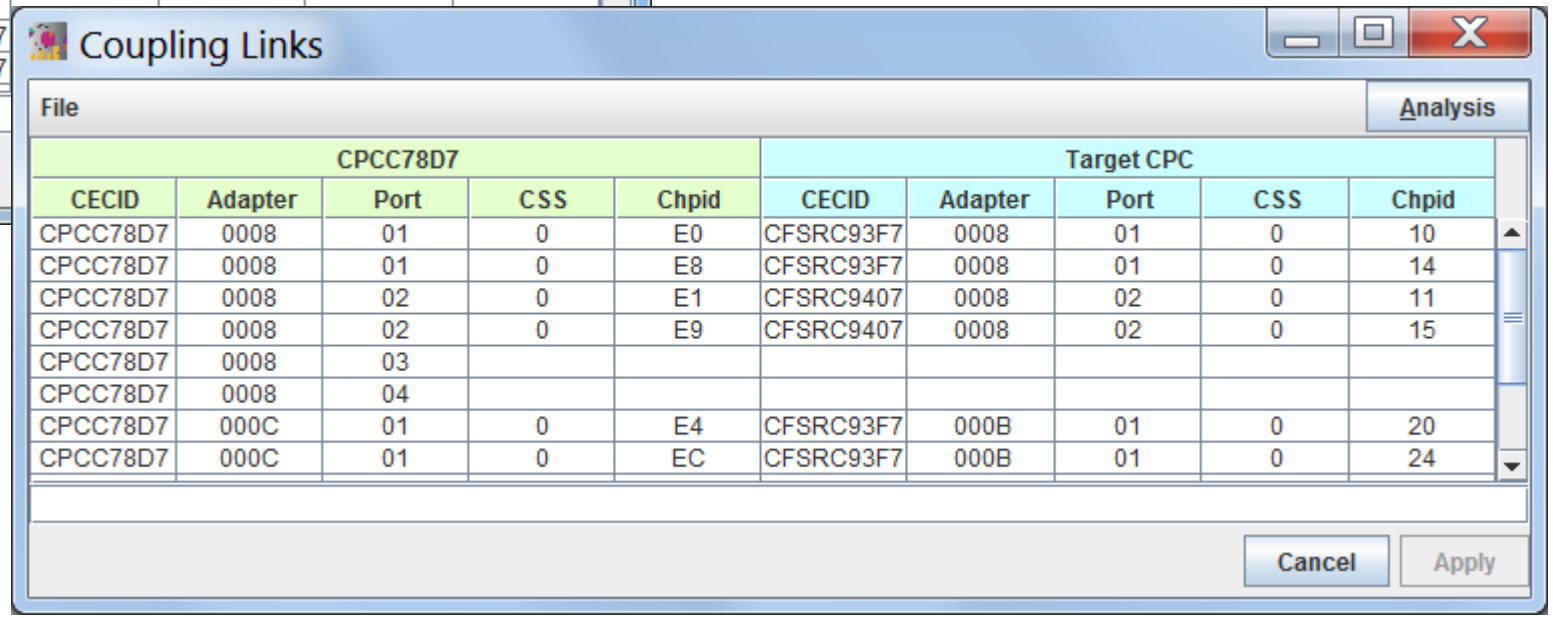
Understand the physical links…



See all of those blank columns?

Load IOCP decks

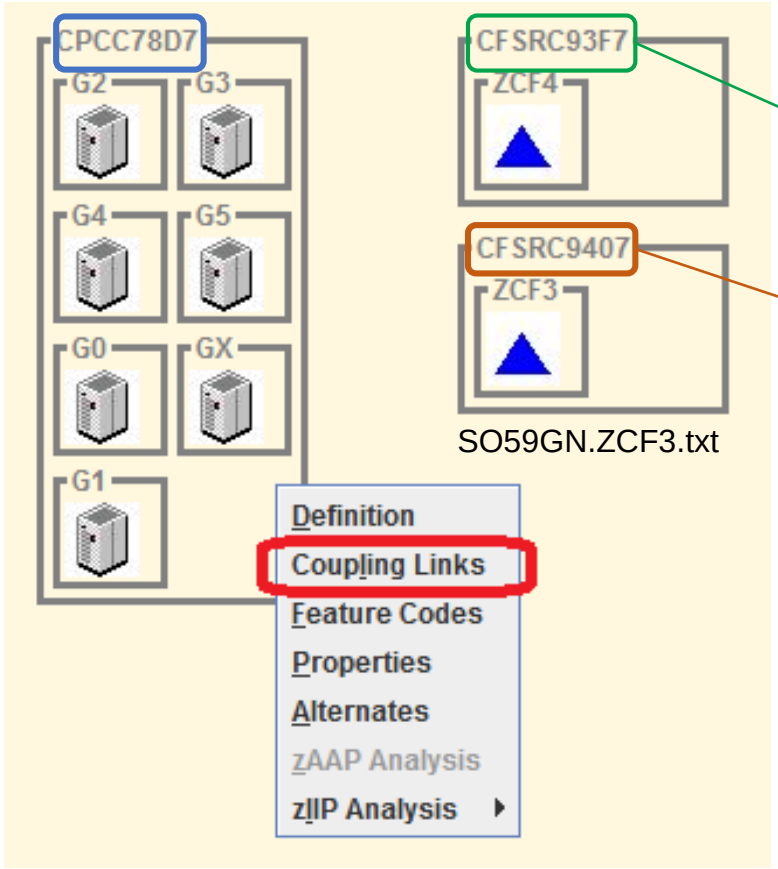
What blank columns?



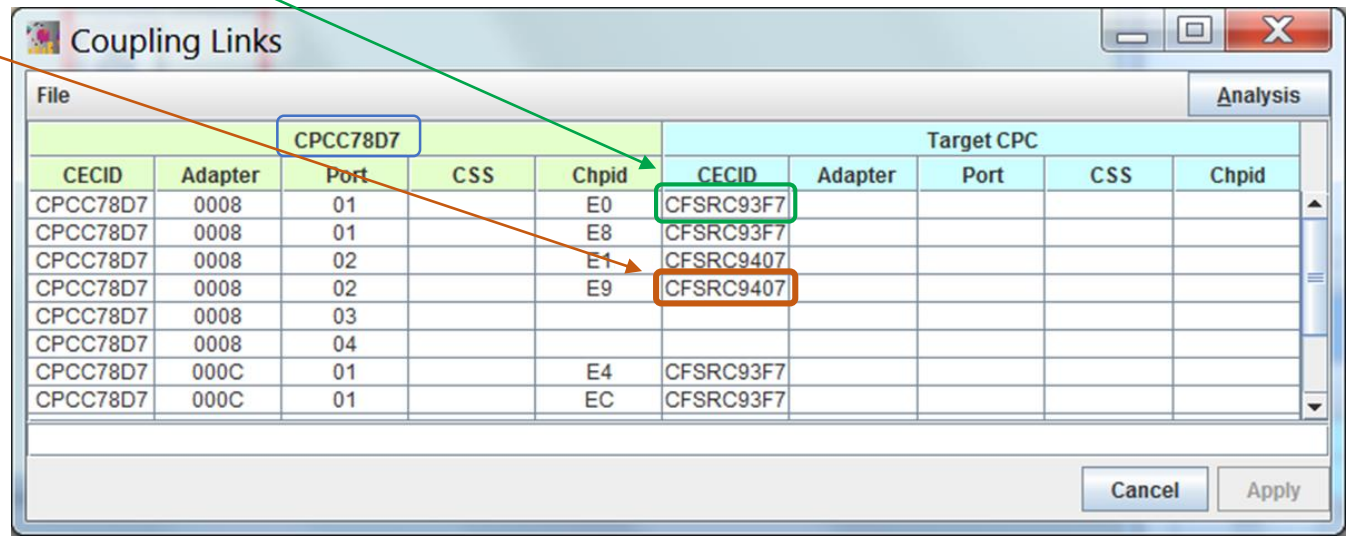
Importing IOCP Information

SO57EN.PROD.CPU.txt

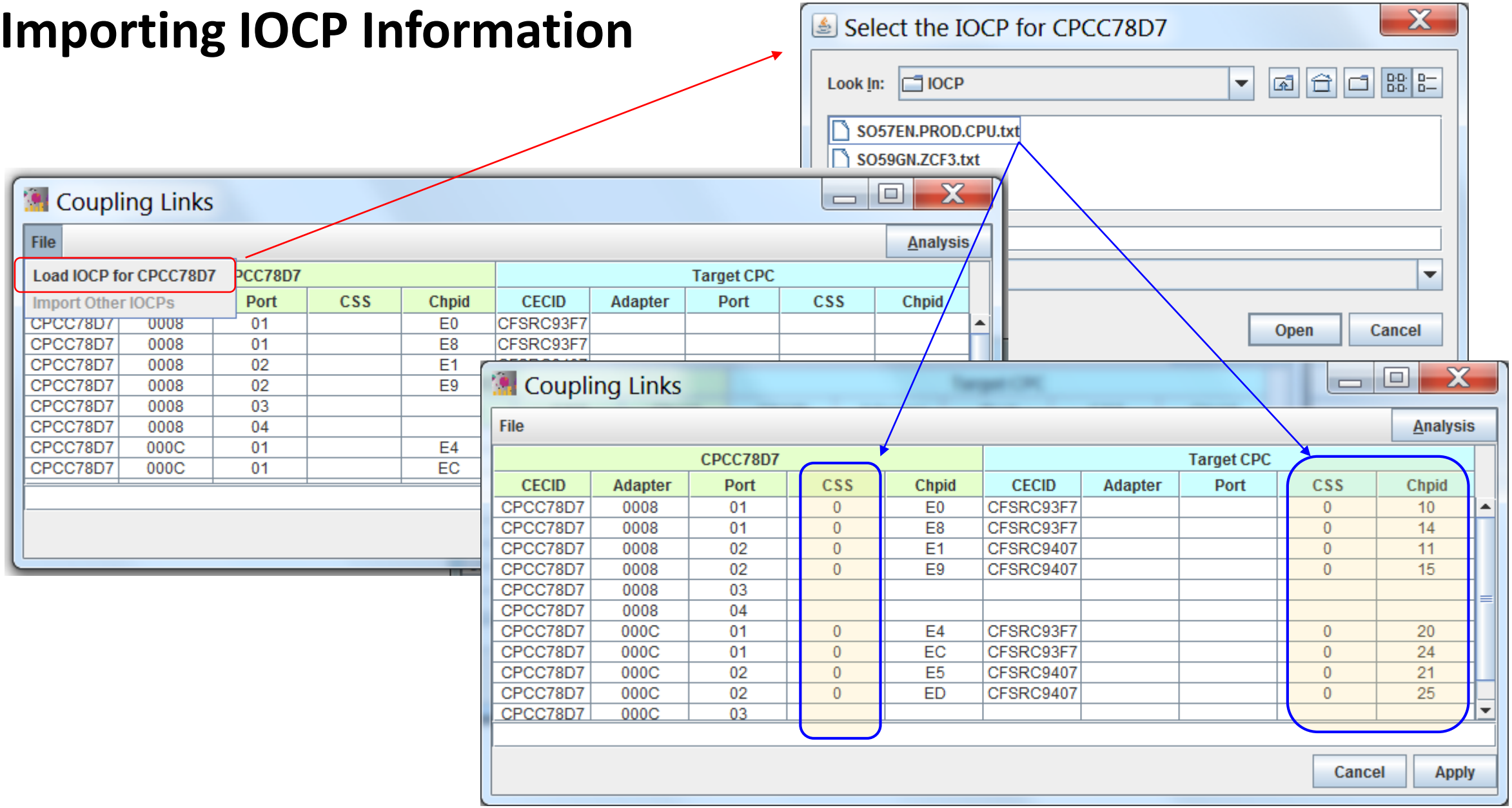
SO60HN.ZCF4.txt



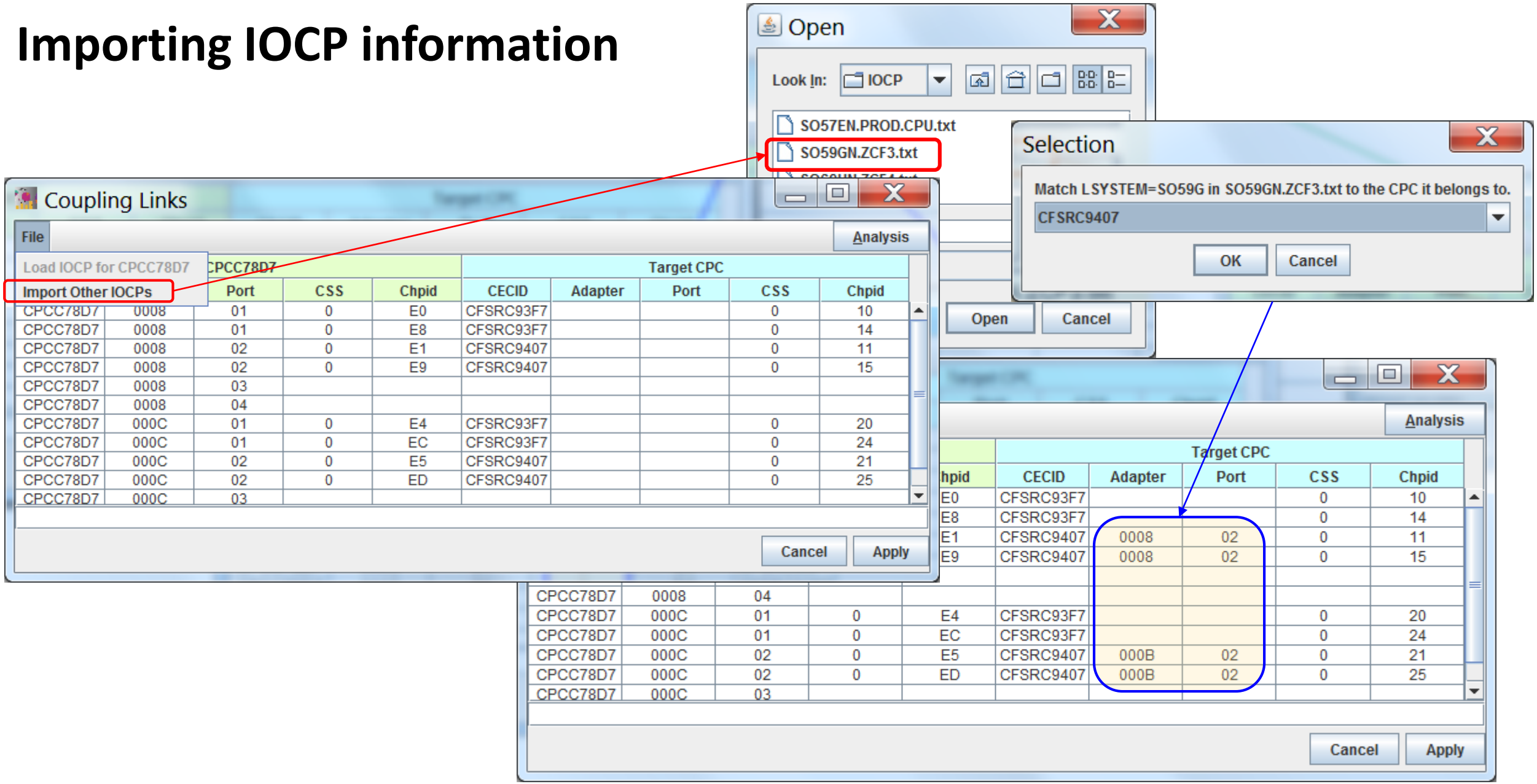
- Know which IOCP goes with which machine.
- Right click within the CPCC78D7 border.
- Select Coupling Links.



Importing IOCP Information



Importing IOCP information



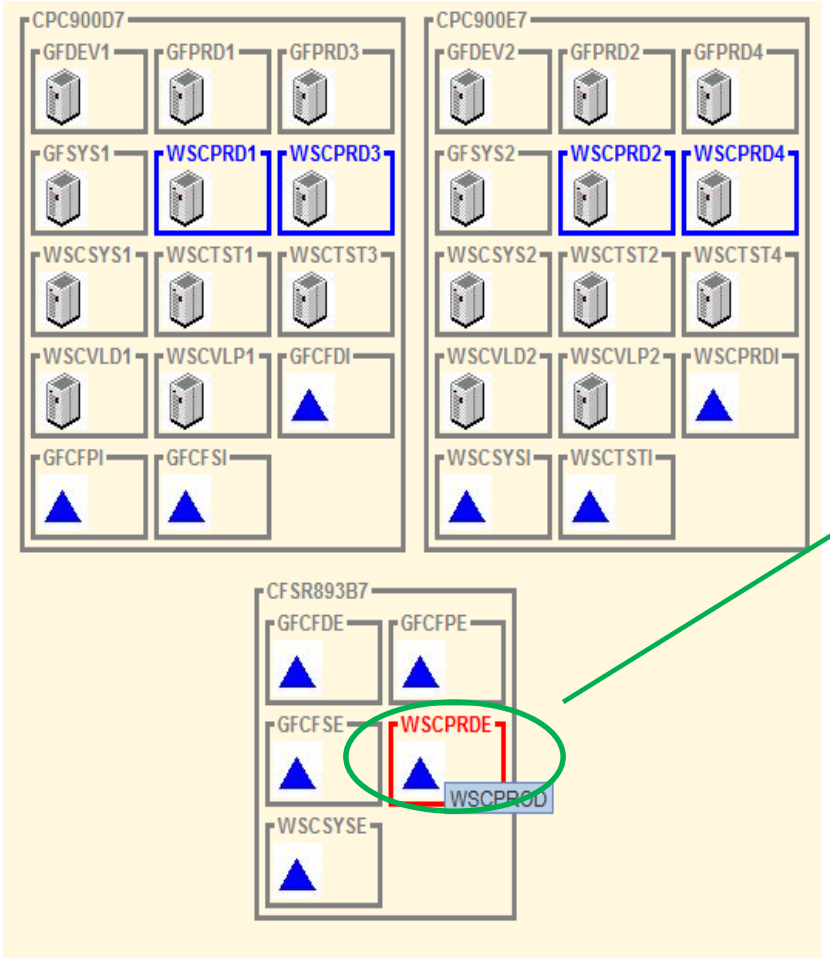
Agenda - Solving Sysplex Questions with zCP3000

- New Technology
- Getting Started
- **Understand problem areas, key resources and intervals**
- Model a new hardware or links or Thin Interrupt
- Coupling Link Considerations

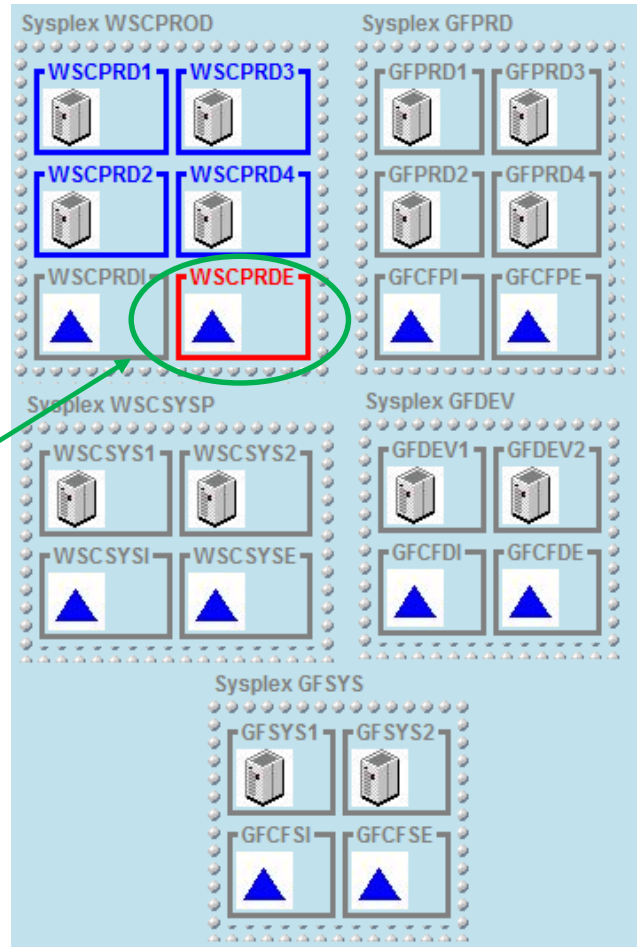
Reference slides include links, graph recommendations, and references.

Understand the Collected Model Data

View->Physical



View->Sysplex Logical



Select WSCPRDE to highlight the other members of the WSCPROD sysplex.

CF Window Summary

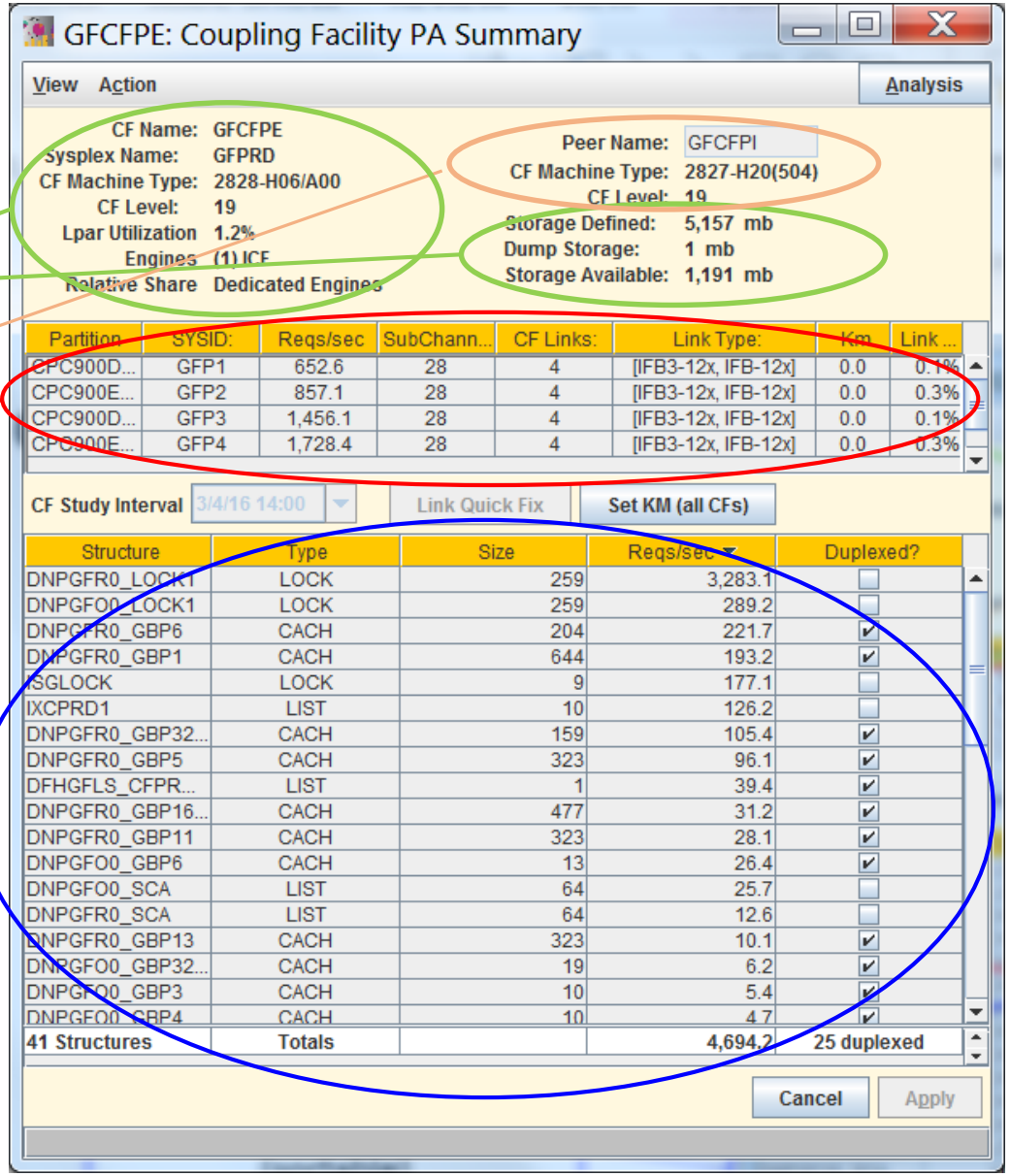
(Main window-> double click a CF)

CF Information

Peer Information

Link Information
Double click row for detail

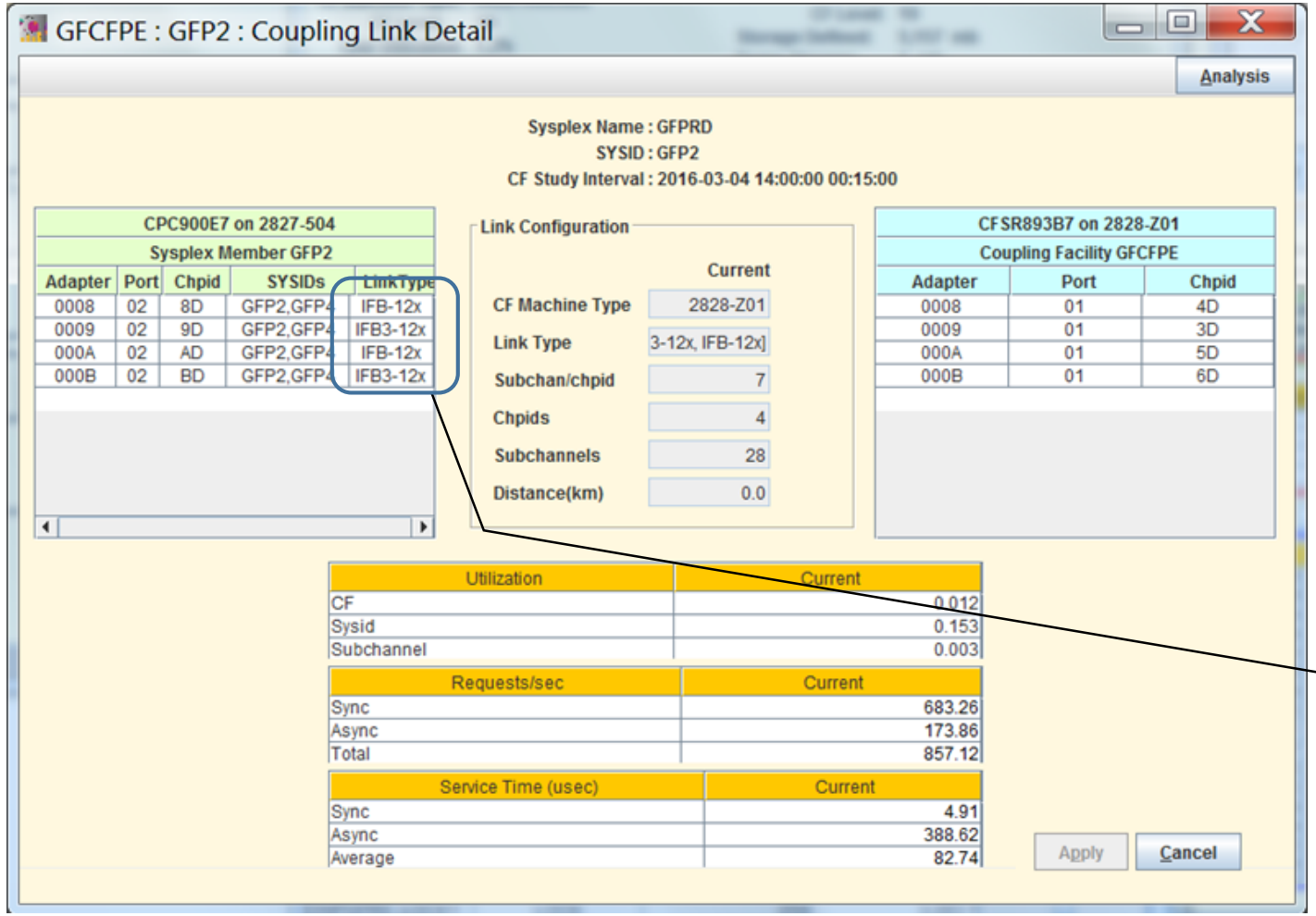
Structure Information (sortable)
Double click row for detail



Graph CF1014 on the CF Window

Coupling Link Detail

(CF window-> double click a link row)



Mixed linktypes

IFB3 – 5000 Mb/sec

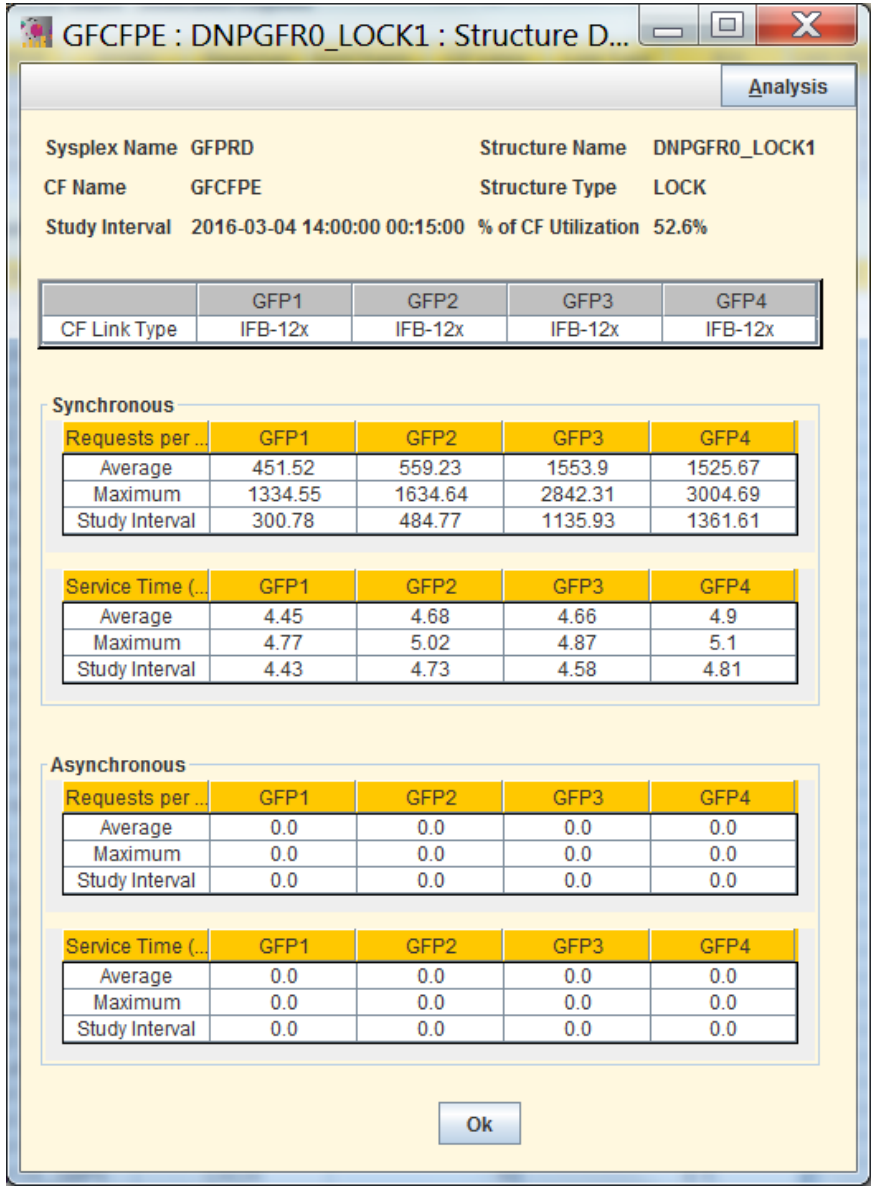
12x – 1000 Mb/sec

Structure Detail

(CF window-> double click a structure)

For critical structures,

- Mostly synchronous?
  (except XCF requests)
- Activity by system
- Include 1-2 reports from 1-2 key structures in the deliverable.



PLEX1007 Sysplex Overview

(Sysplex window->Analysis)

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

Graph PLEX1007 on the Logical Sysplex Window

PLEX1012 CPC Coupling Links Summary

(Right Click CPC->Coupling Links->Analysis)

CPCC78D7					Target CPC				
CECID	Adapter	Port	CSS	Chpid	CECID	Adapter	Port	CSS	Chpid
CPCC78D7	0008	01	0	E0	CFSRC93F7	0008	01	0	10
CPCC78D7	0008	01	0	E8	CFSRC93F7	0008	01	0	14
CPCC78D7	0008	02	0	E1	CFSRC9407	0008	02	0	11
CPCC78D7	0008	02	0	E9	CFSRC9407	0008	02	0	15
CPCC78D7	0008	03							
CPCC78D7	0008	04							
CPCC78D7	000C	01	0	E4	CFSRC93F7	000B	01	0	20
CPCC78D7	000C	01	0	EC	CFSRC93F7	000B	01	0	24
CPCC78D7	000C	02	0	E5	CFSRC9407	000B	02	0	21
CPCC78D7	000C	02	0	ED					
CPCC78D7	000C	03							
CPCC78D7	000C	04							

Table 1. Coupling Link Summary for CPCC78D7

CPCC78D7				Target				
Adapter	Port	Chpid	Partitions	CPC	AID	Port	Chpid	Partition Names
0008	01	E8	ZZPLEX2: VIPA GLF5 GLFT GLF4 GLF3 GLF2	CFSRC93F7	0008	01	14	ZZPLEX2: ZCF4
0008	01	E0	ZZPLEX2: VIPA GLF5 GLFT GLF4 GLF3 GLF2	CFSRC93F7	0008	01	10	ZZPLEX2: ZCF4
0008	02	E9	ZZPLEX2: VIPA GLF5 GLFT GLF4 GLF3 GLF2	CFSRC9407	0008	02	15	ZZPLEX2: ZCF3
0008	02	E1	ZZPLEX2: VIPA GLF5 GLFT GLF4 GLF3 GLF2	CFSRC9407	0008	02	11	ZZPLEX2: ZCF3
000C	01	EC	ZZPLEX2: VIPA GLF5 GLFT GLF4 GLF3 GLF2	CFSRC93F7	000B	01	24	ZZPLEX2: ZCF4
000C	01	E4	ZZPLEX2: VIPA GLF5 GLFT GLF4 GLF3 GLF2	CFSRC93F7	000B	01	20	ZZPLEX2: ZCF4
000C	02	E5	ZZPLEX2: VIPA GLF5 GLFT GLF4 GLF3 GLF2	CFSRC9407	000B	02	21	ZZPLEX2: ZCF3
000C	02	ED	ZZPLEX2: VIPA GLF5 GLFT GLF4 GLF3 GLF2	CFSRC9407	000B	02	25	ZZPLEX2: ZCF3

Table 3. Summary of Coupling Link Adapters on CPC CPCC78D7

CF Health Check

(CF window->Analysis->CF1000)

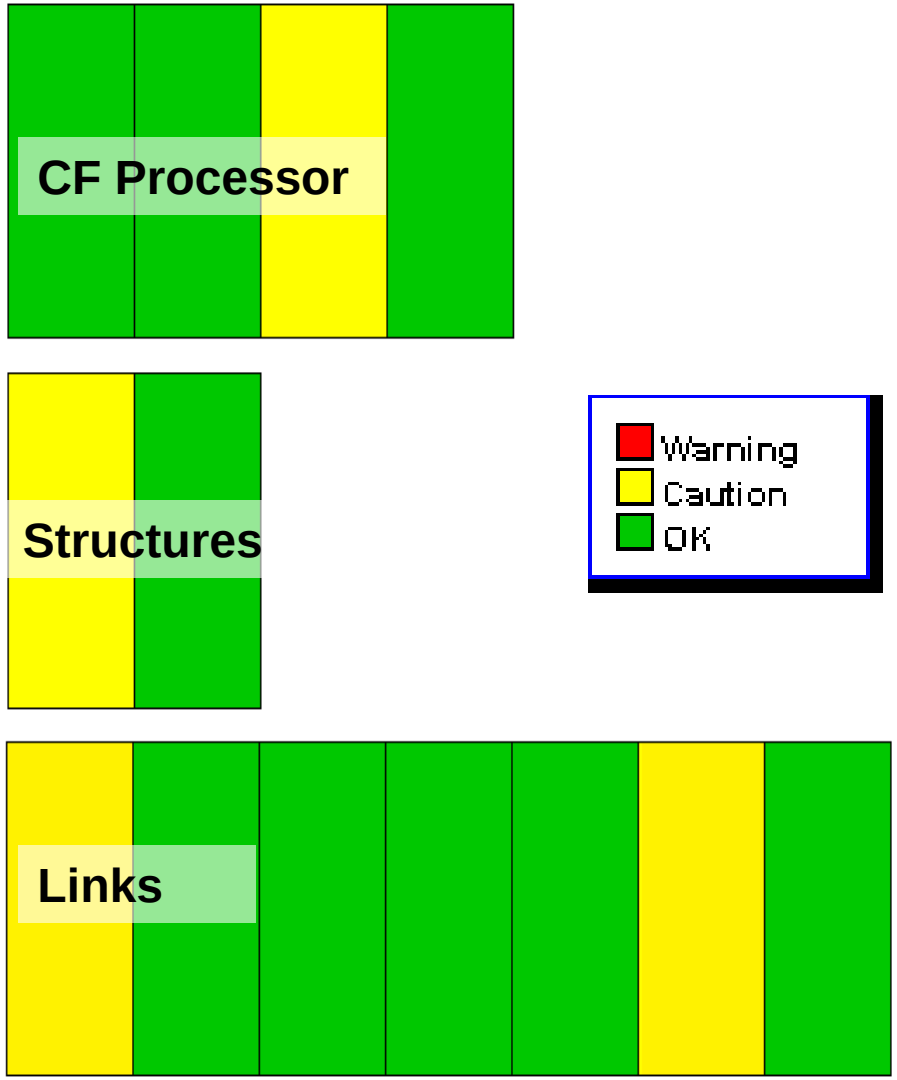
A single report giving insights on several key areas.

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

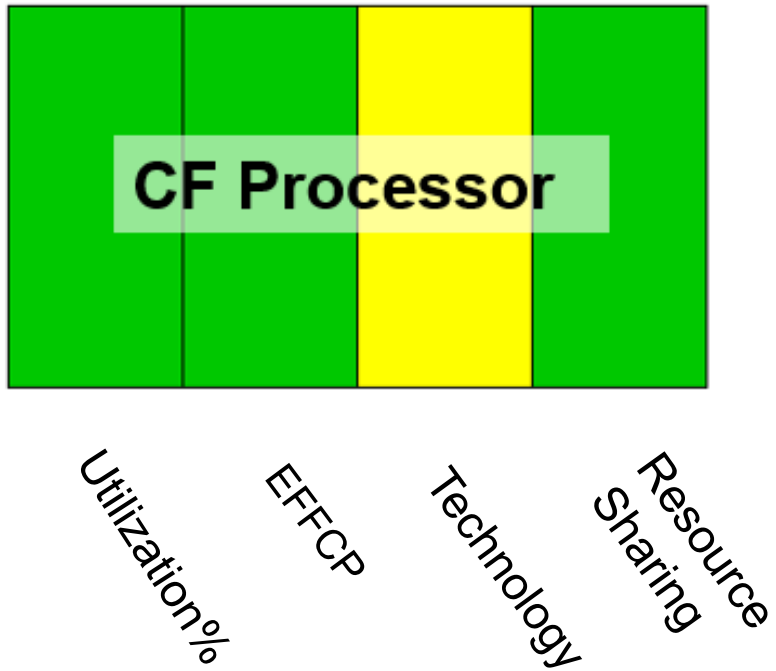
CF Health Check – CF Processor

Rules: if Thin Interrupt

1. Utilization : warn if cfutil>90%, and effcp>=entitled CPs
2. (EFFCP rule is n/a and will not appear)
3. Hardware should be current generation.
4. Are all resource sharers using Thin?

Rules: if not Thin Interrupt

1. LPAR Utilization should be <50%
2. Effective Engines should be >= 1
3. Hardware should be current generation.
4. Sharing, with more than one PROD



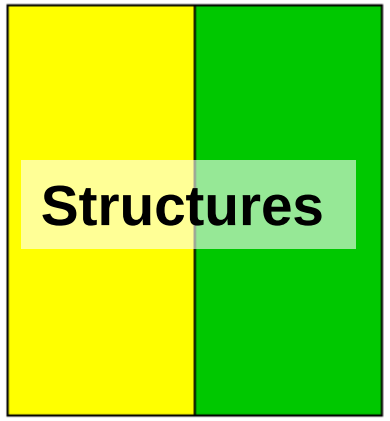
Graph CF1000 on the CF Window

CF Health Check - Structures

Lock Contention

Lock Requests had to wait because the lock was already held.

Fix in the Application, if possible.



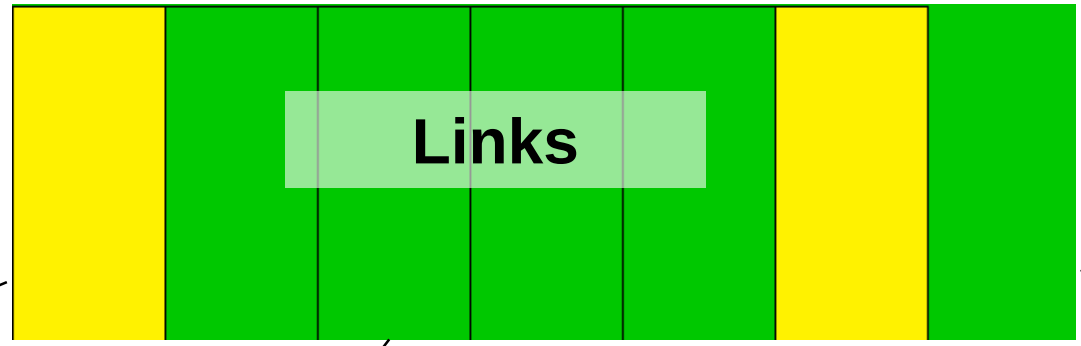
False Lock Contention

2 locks hash to the same index.

Increase the Structure size.

Structures	Type	Average Req/sec	Lock Contention	False Lock Contention
DNPDBR0_LOCK1	LOCK	11,742.6	2.02%	0.70%
DNPDBO0_LOCK1	LOCK	59.7	0.04%	0.02%

CF Health Check - Links



Link Technology

Should be IC, ICA, or CE-LR

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

z/OS sees 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.

Mixed Linktypes

More than one speed between this CF and a sysplex member.

Graph CF1000 on the CF Window

Service Time

The metric that you should be looking at.

From: when z/OS issues a CF request,
To: when the return is recognized by z/OS.

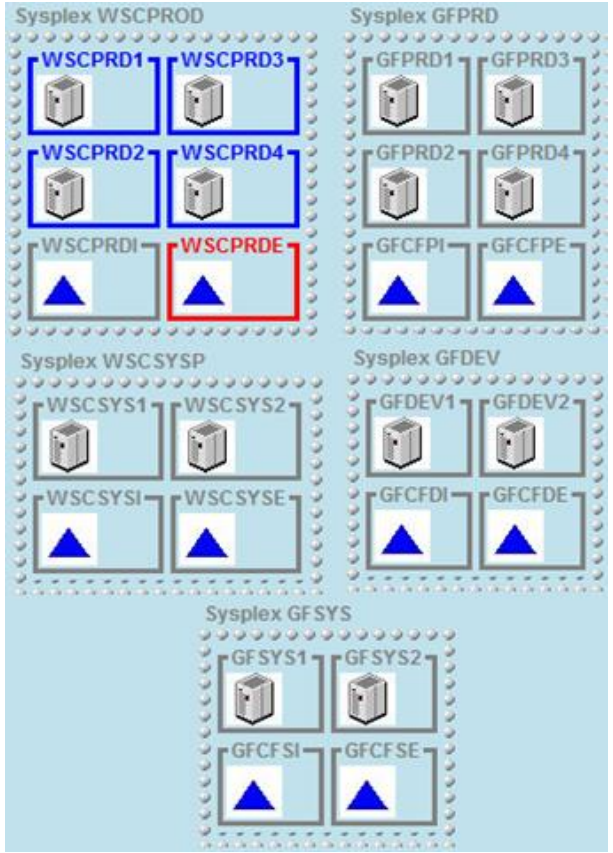
It is comprised of:

- Host Hardware
- Link Latency
- Data Transfer*
- CF Busy time
- Does not include queue time.

* Actual Data Transfer size is unknown.

Service Time

(Main window->View->Logical Sysplex)



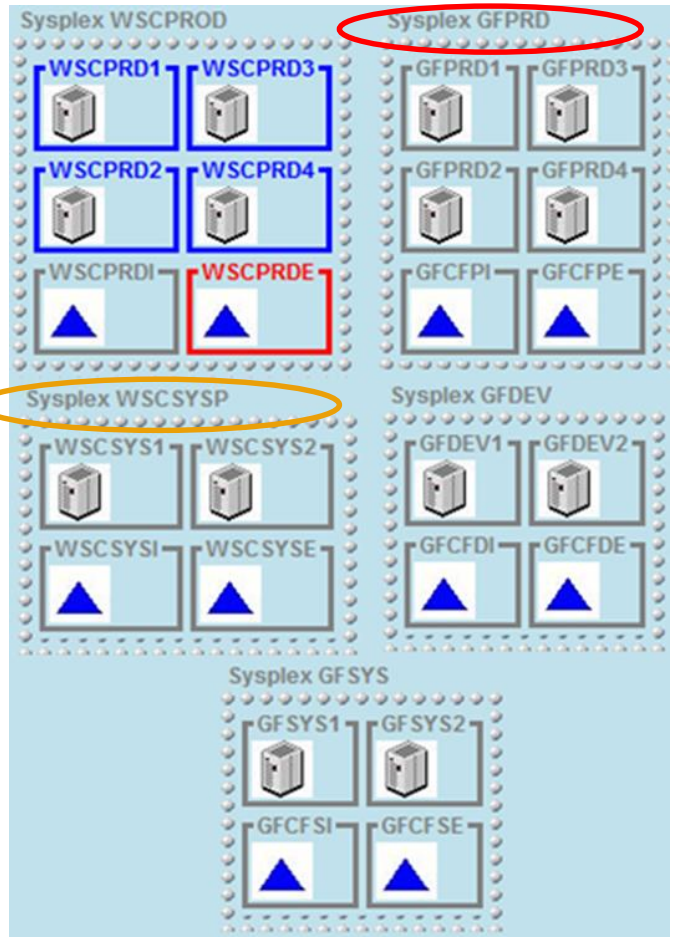
The metric to look at,
but where to start looking?

10 CFs
14 z/OS members
~= 350 structures
180 intervals
Locks, Cache, List ops
Sync and Async

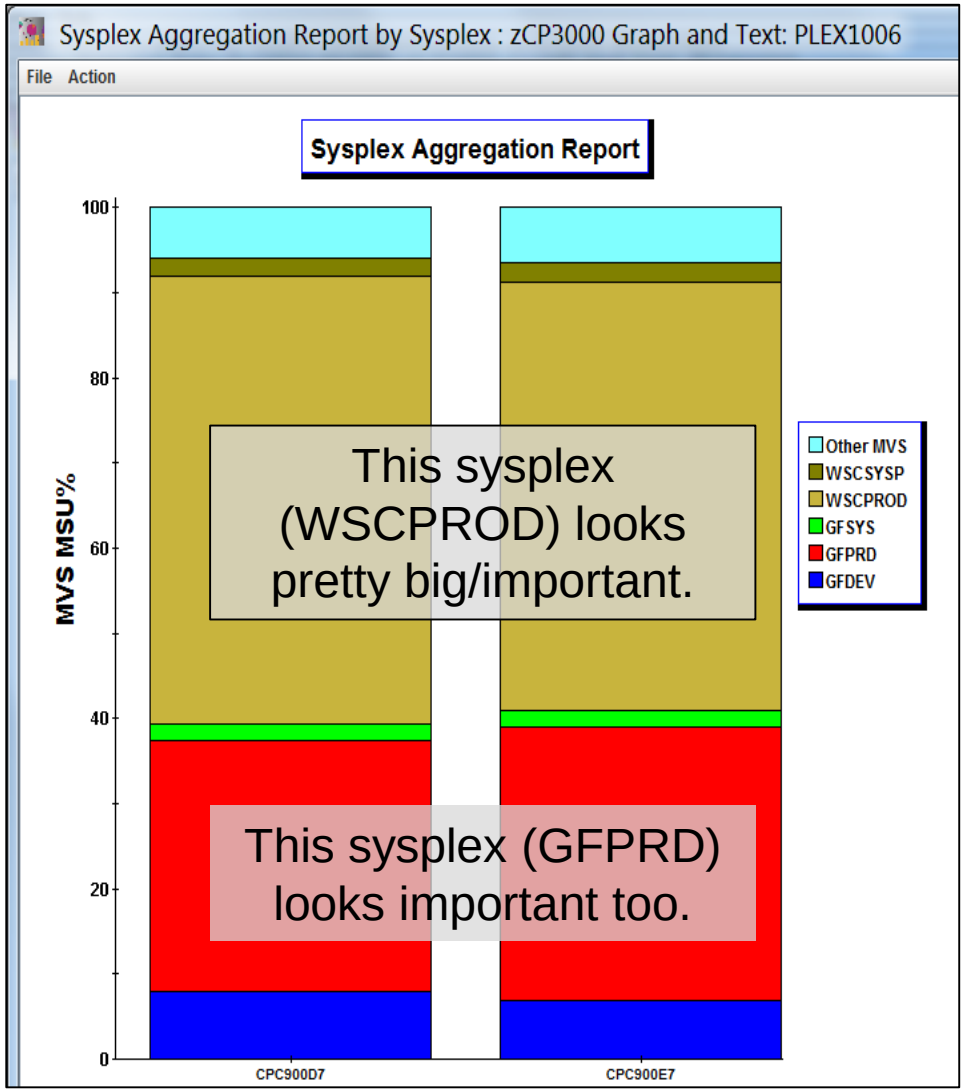
- **Difficult to scope.**
- **Difficult to characterize.**
- **Difficult to summarize.**

Where to start?

(Main window->View->Logical Sysplex)

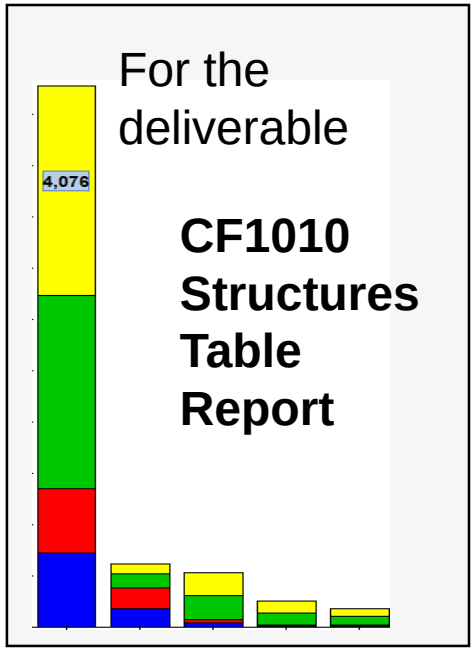


(Sysplex window->Analysis->PLEX1006)

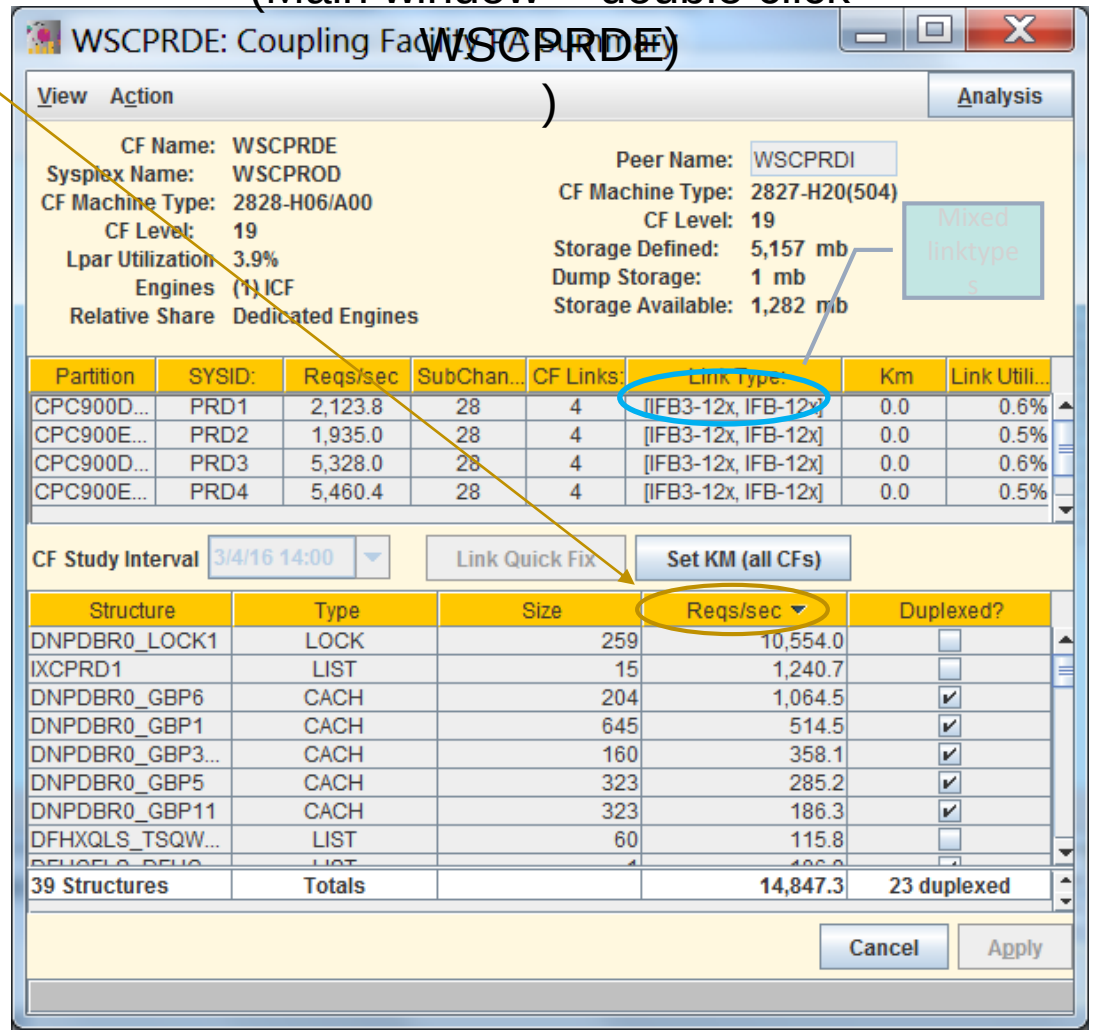


Focus on 1-2 key structures on key CFs

- Click on the column head to sort.
- Double click the row to see more about LOCK1
- Note: **mixed linktypes**



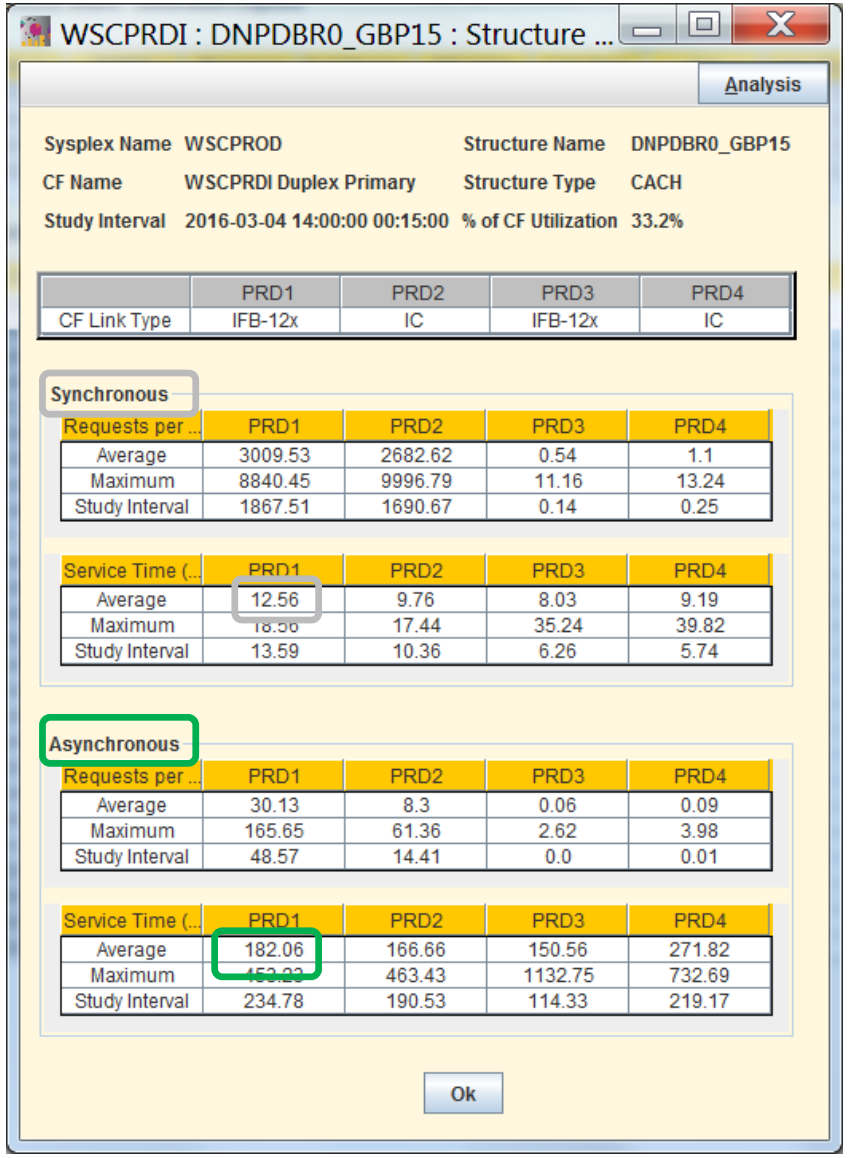
(Main window-> double click WSCPRDE)



Service Time by Structure

- 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.
- HUGE difference between sync and **async** service time (13 vs **182** mics).
- Service Time graphs available from the CF window, the Link window, and the Structure detail window.

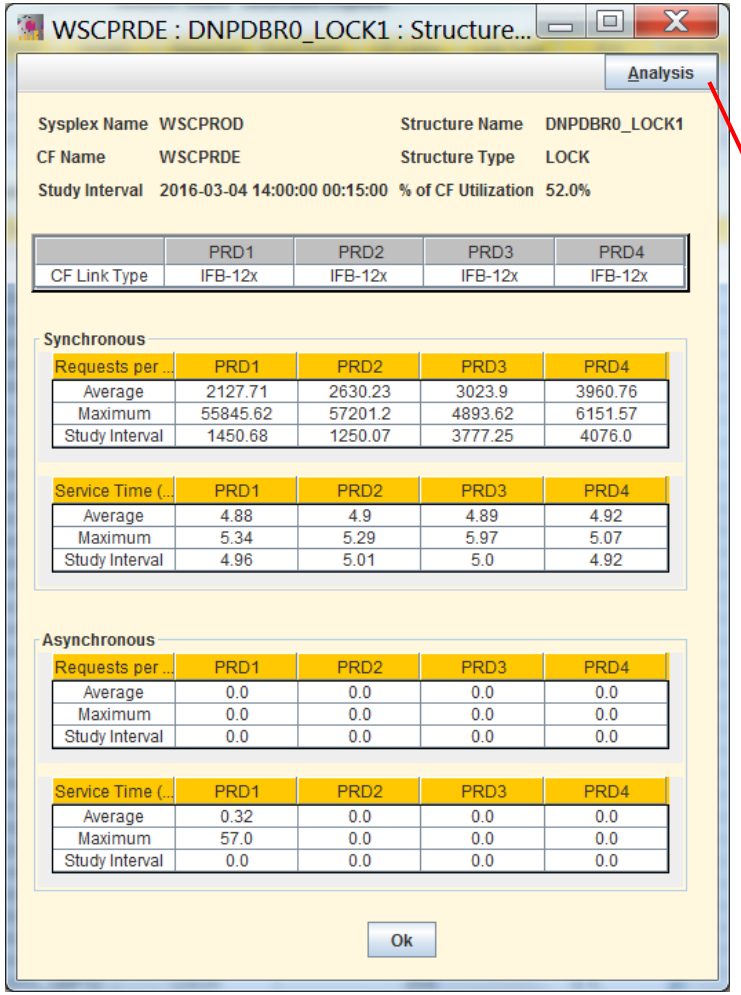
(CF window-> double click DNPDBR0_GBP15)

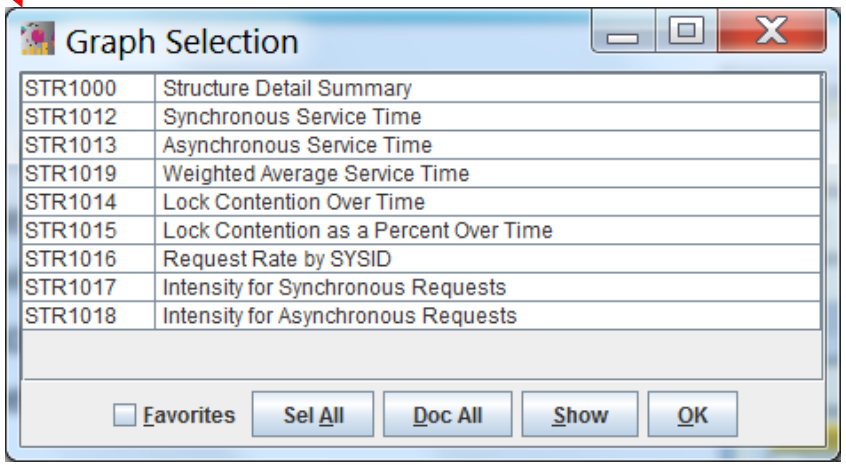


Structure Detail Window – also report STR1000

Service Time for Important Structures

(WSCPRDE window-> double click DNPDBR0_LOCK1)





- Pick a busy structure.
- Double click to look at structure detail.
- Note which links are busiest.
- Are most requests running synchronously?
- Are they getting good service time?
- Maybe include STR1000 or STR1012 in the customer deliverable, to show that it's happy.

Identify key coupling links

Which Sysplexes->CFs are important?

WSCPRDI, WSCPRDE on sysplex WSCPROD

GFCFPI, GFCFPE on sysplex GFPRD

- Sort the links by request rate (CF window, click on the column header)
- Look at the CF Structures Report to see which links are busiest for the busiest (aka most important) structures.
- Look at anything shows up yellow in the Subchannels part of the CF Health Check.

10 CFs
14 z/OS members
~= 350 structures
180 intervals

Can’t one number tell me if the links are ok for this CF?

No, because …

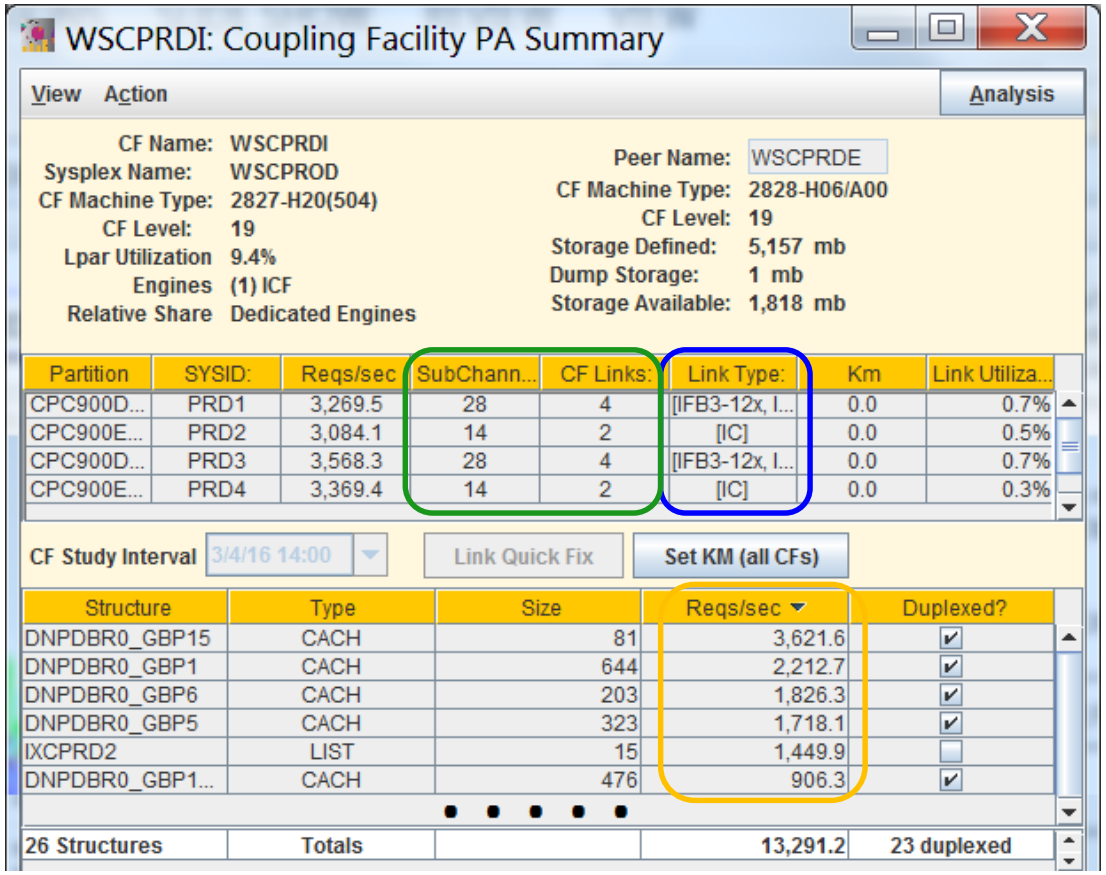
Different # of Subchannels and requests per z/OS system

Different link speeds

Variation in data transfer sizes.

Uneven distribution of requests from z/OS members.

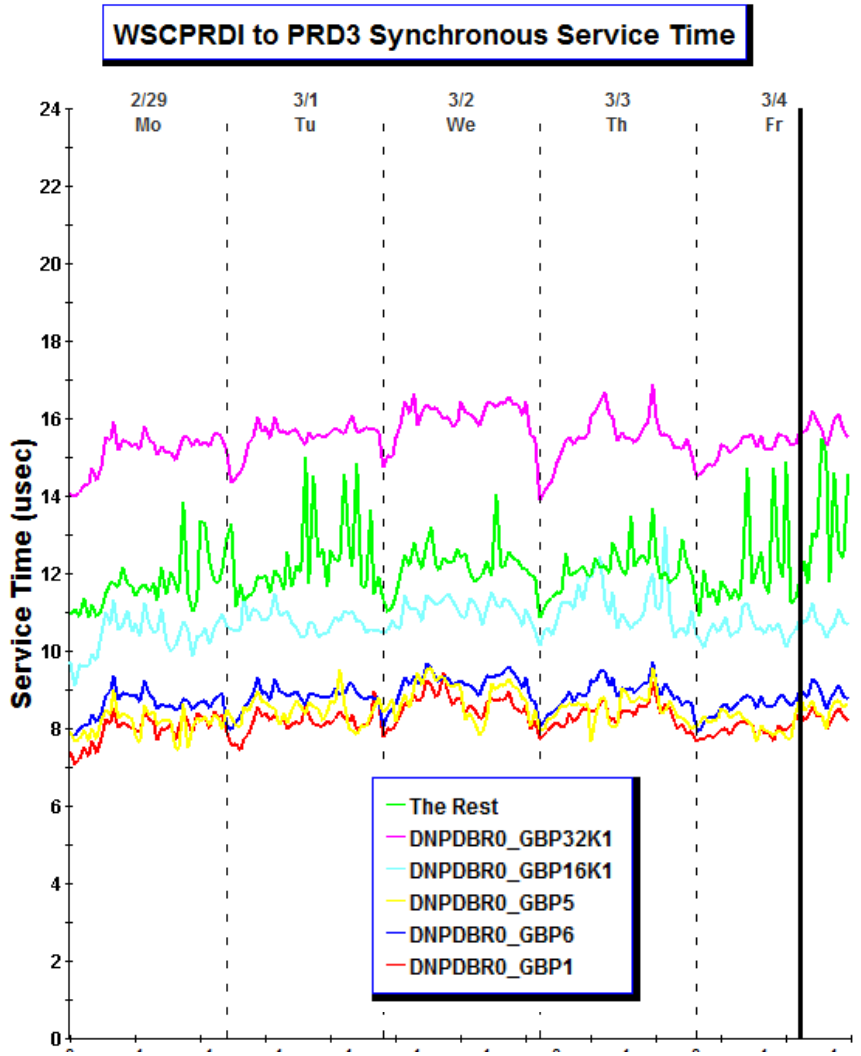
(main or sysplex window-> double click WSCPRDE)



Service Time by Link

- If a key structure has bad service time on a specific link, look at the service time for all traffic on that set of links.
- (there are 4 “links” running between WSCPRDI to PRD3).
- Which means there are 4 chpids: 88, 98, A8, B8
- Those 4 chpids run through 4 different adapters on CPC900D7.
- ***It would be nice to look at the physical link as well.***

(CF window->PRD3 link->Analysis->CFL012)



Sysplex Graph Recommendations

See graph index in reference slides.
Use zCP3000 “favorites” function.

PA Mode

PA and QM Mode

QM only

Probably in every study

- zCP3000 main view : Topology
- per CPC (Coupling Links window): Coupling Links Summary
- Sysplex Logical window : Sysplex Topology, Coupling Summary
- For each CF : CF Summary, CF Health Check, CF Migration Summary report

Look for areas of interest

- Thin Interrupt Effect
- CF Structures Table
- CF Link Topology
- Coupling Adapter Summary
- 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 - Solving Sysplex Questions with zCP3000

- New Technology
- Getting Started
- Understand problem areas, key resources and intervals
- **Modeling new hardware or links or Thin Interrupt**
- Coupling Link Considerations

Reference slides include links, graph recommendations, and references.

Modeling configuration changes

You must be in QM Mode

- Resolve red triangles and fill in IOCP info before going to QM mode.
- Multiple changes (machine, linktype, dispatch), in any order.
- Configuration changes may then affect:
  - ICF utilization
    - w/ the speed of the CPC or alt
    - nway change
  - Service time when:
    - Linktype or distance changes
    - CF machine type changes via drag and drop
    - CF Dispatch mode changes (ie Thin Interrupt vs dedicated)
  - % ops executing synchronously
  - Channel path utilization
  - z/OS utilization changes will be very small

Service Time vs Sysplex Overhead

Sysplex Op service time 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

Sysplex Overhead (on z/OS) is dependent on

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

Service time does not include queue time.
-> see CF Health Check, Subchannels, Delayed Requests

* (not measured)

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
z14	3906-701	26.00

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

Cookbook steps to model new hardware

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

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

- Fix red triangles, "extra" standalone CF machines, km distance
- For standalone CFs : model, engine type, and LPAR weight
- Import IOCP information, when available.
- Generate some reports, like the Health Checks for key SYS and CF

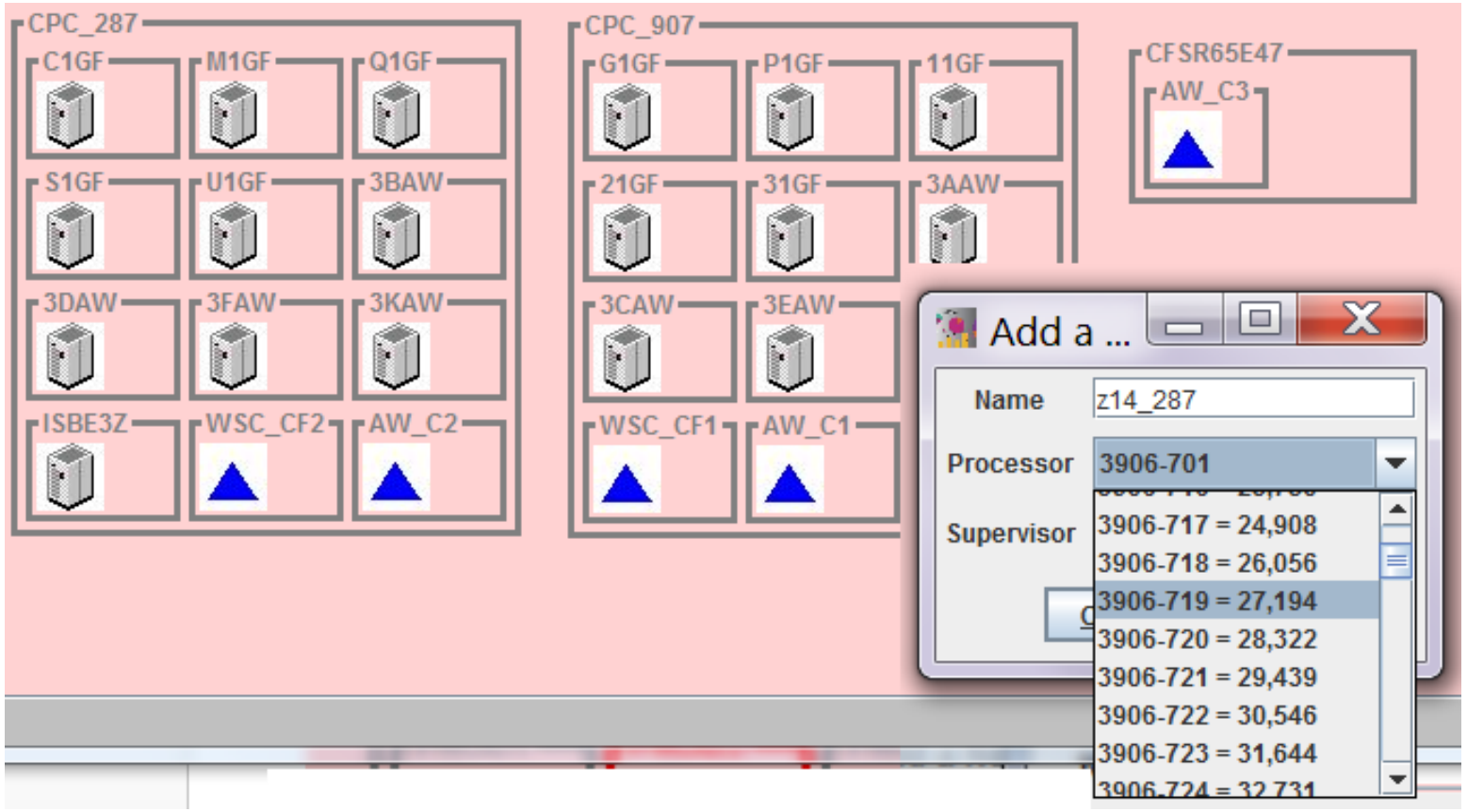
Save then go to QM mode.

- Add a new z14
- Move partitions to the new CPC → service time recalc
- Define engines and lpars of the new z14 → service time recalc
- Upgrade link types → service time recalc
- Change dispatch mode → service time recalc

Look at the CF Migration Report →

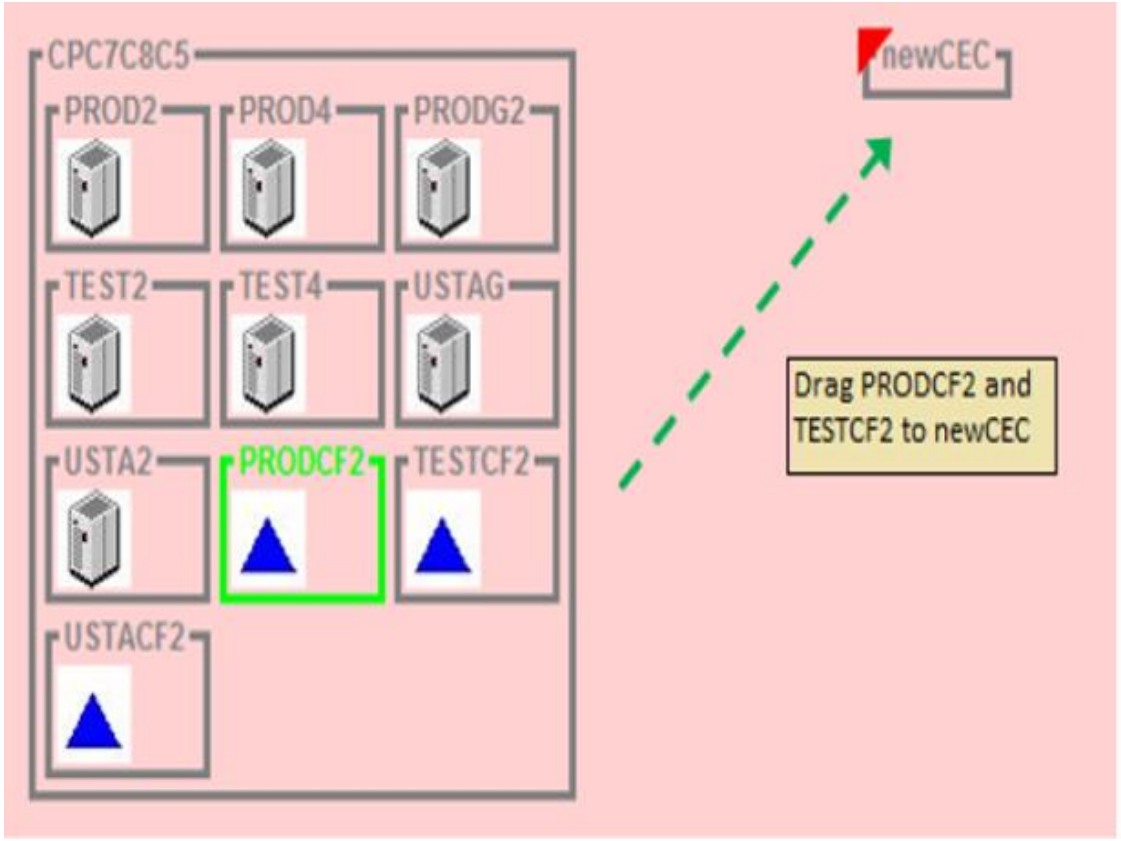
Change the CF machine type

Step 1 – add the new machine



Change the CF machine type

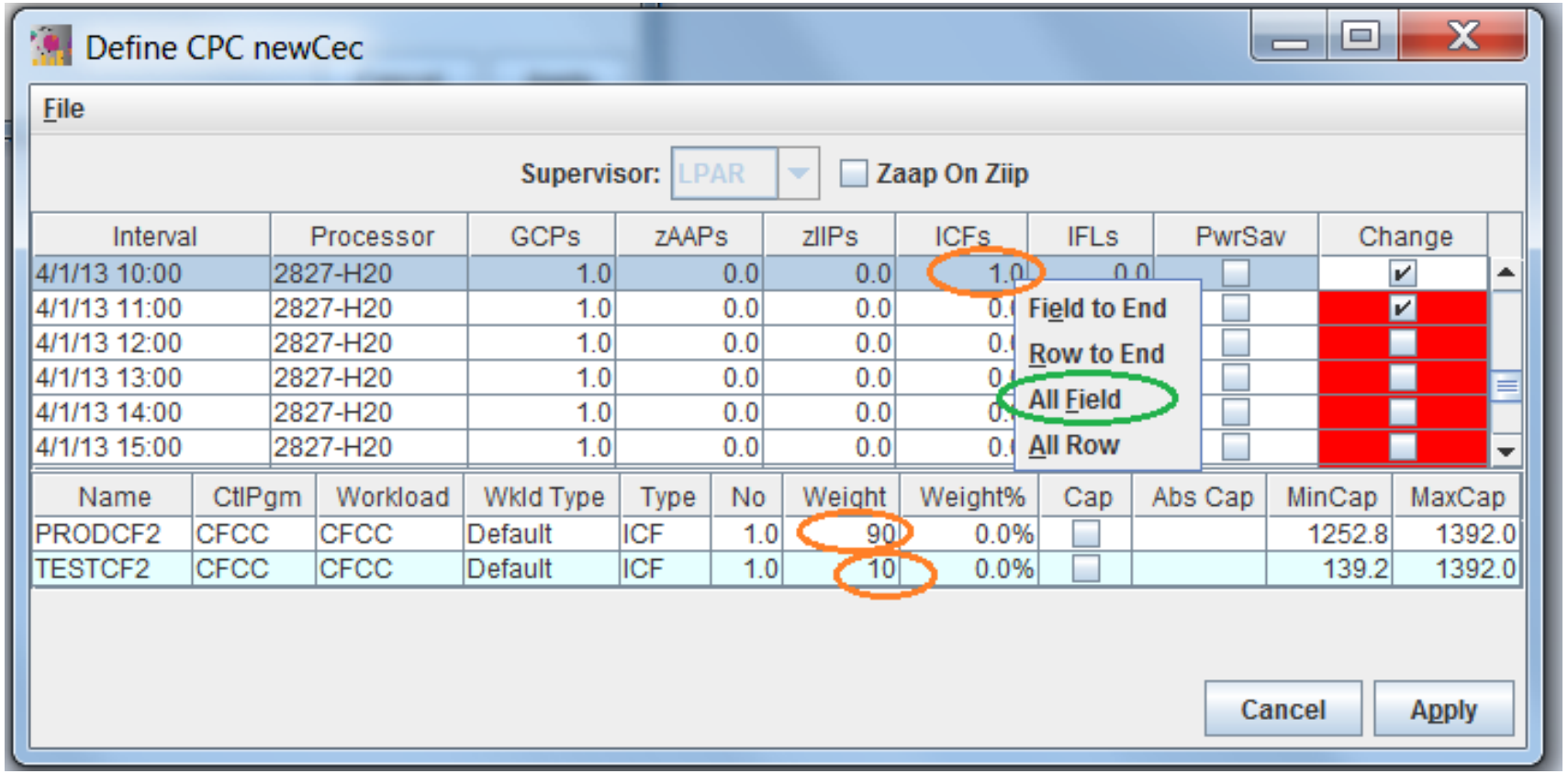
Step 2 – move some partitions there



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

Change the CF machine type

Step 3 – define real engines and partitions



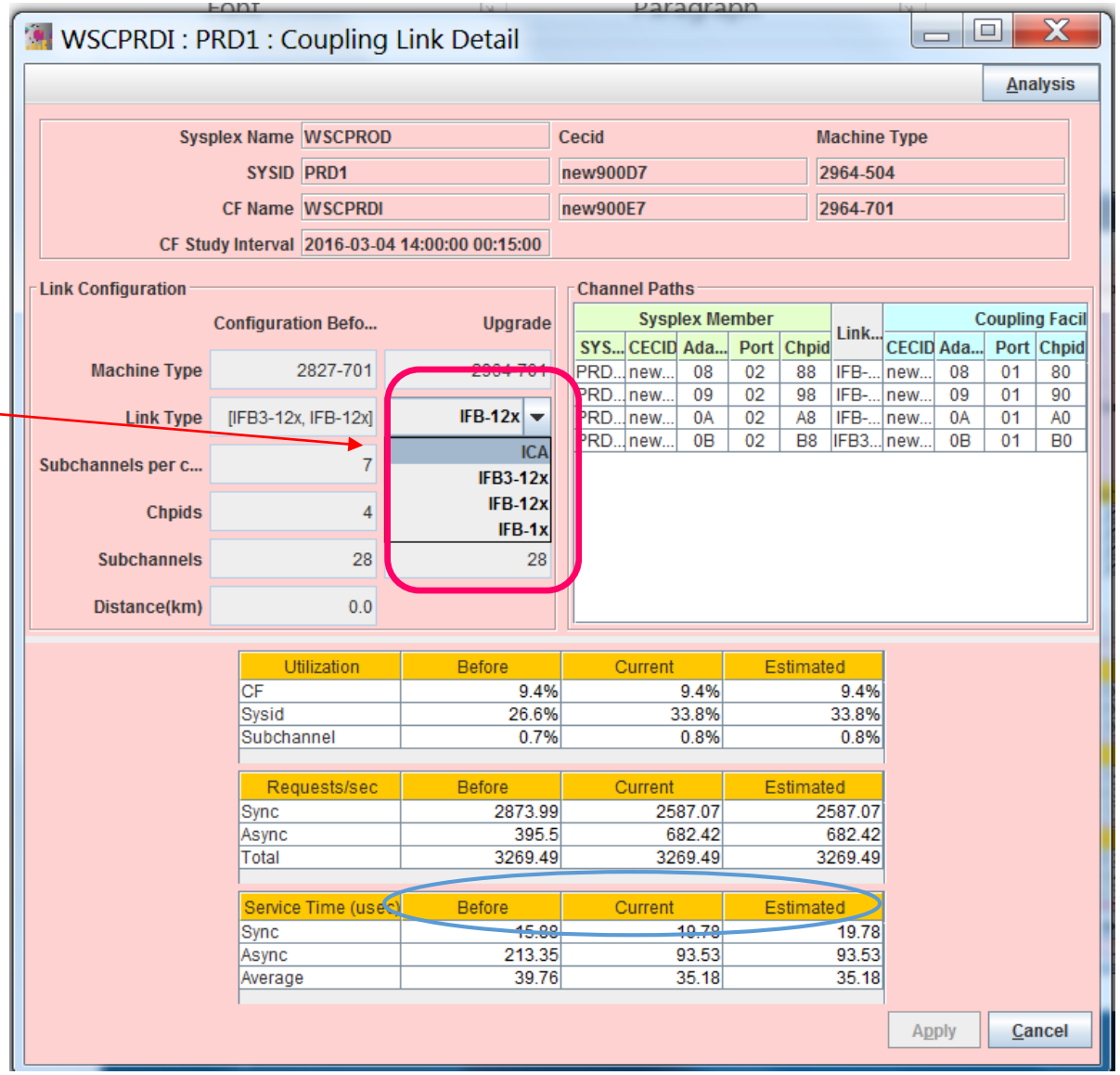
Change the Coupling Links

- Double click a link on the CF window.
- Click “Upgrade Link”
- Change the linktype
- Change #chpids
- Change Subchannels*

“Before” is PA mode.

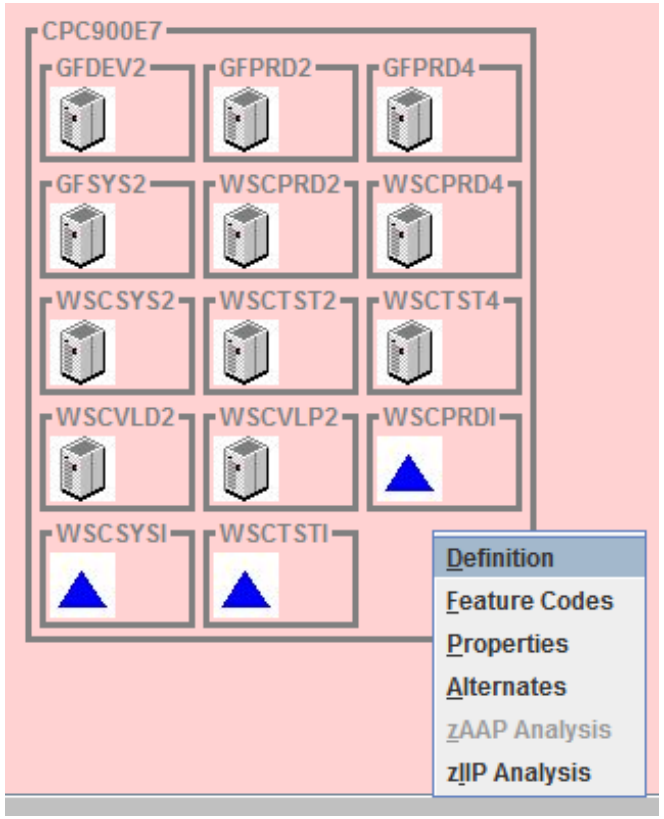
“Current” shows the changes so far in QM.

“Estimated” shows the effect of the new change.



Change from Dedicated to Thin Interrupt

Step 1 – change the LPAR definition



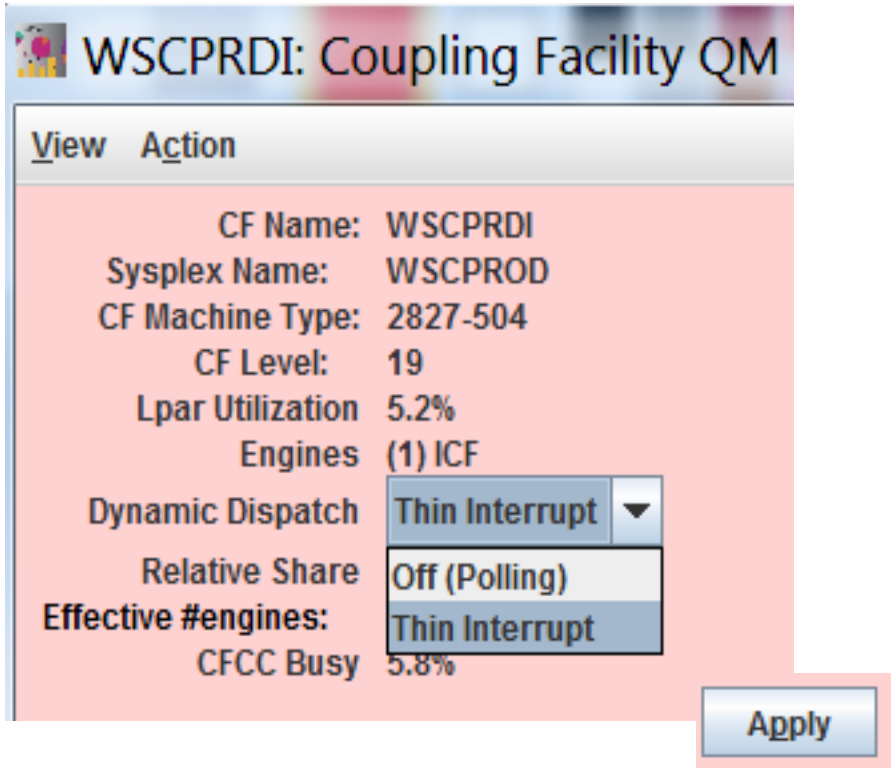
Name	CtlPgm	Workload	Wkld Type	Type	No	Weight	Weight%	Cap
GFDEV2	z/OS-2.1	High	CPU-MF	CP	2.0	94	10.6%	
				zIIP	1.0	60	7.2%	
GFPRD2	z/OS-2.1	Average	CPU-MF	CP	2.0	49	5.5%	
				zIIP	2.0	170	20.5%	
GFPRD4	z/OS-2.1	Average	CPU-MF	CP	2.0	163	18.4%	
				zIIP	2.0	170	20.5%	
GFSYS2	z/OS-2.1	Average	Default	CP	1.0	24	2.7%	
				zIIP	1.0	10	1.2%	
WSCPR...	z/OS-2.1	High	CPU-MF	CP	2.1	147	16.6%	
				zIIP	2.0	170	20.5%	
WSCPR...	z/OS-2.1	Average	CPU-MF	CP	2.1	296	33.4%	
				zIIP	1.0	170	20.5%	
WSCSYS2	z/OS-2.1	Average	CPU-MF	CP	1.0	42	4.7%	
				zIIP	1.0	20	2.4%	
WSCTST2	z/OS-2.1	Average	Default	CP	1.0	70	7.9%	
				zIIP	1.0	60	7.2%	
WSCTST4	z/OS-2.1	Average	Default	CP	1.0	47	0.0%	
WSCVLD2	z/VM-6.3	Average/...	Default	IFL	2.0	400	44.4%	
WSCVLP2	z/VM-6.3	Average/...	Default	IFL	2.0	500	55.6%	
WSCPRDI	CFCC	CFCC	Default	ICF	2.0	70	Ded	
WSCSYSI	CFCC	CFCC	Default	ICF	1.0	10		
WSCTSTI	CFCC	CFCC	Default	ICF	1.0	20		

Field to End
All Field

Change weight from Ded to 70.
Change LCPs from 1 to 2. → Why?

Change from Dedicated to Thin Interrupt

Step 2 – change the dispatch mode for WSCPRDI

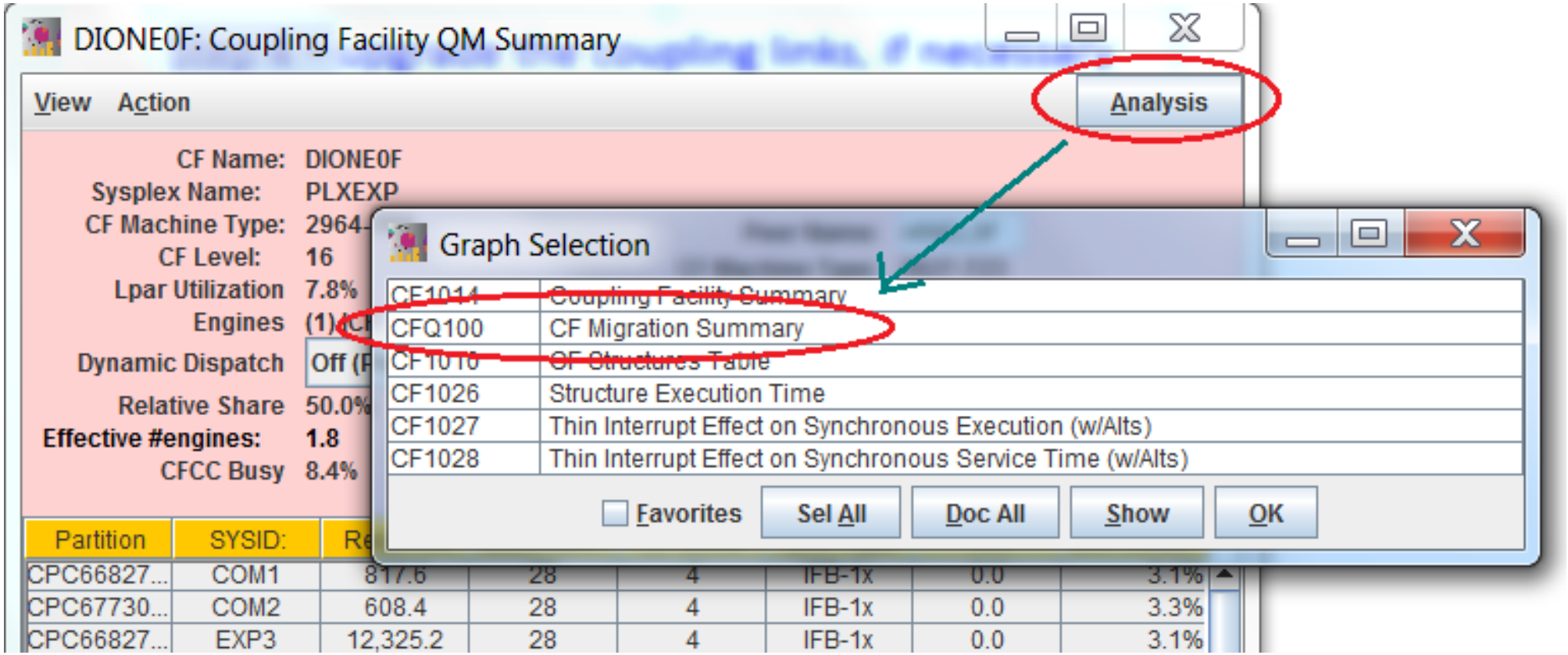


LPAR definition must already be set to sharing.

Thin Interrupt service time is similar, slightly worse than for dedicated.

Generate a before and after report

Last Step : generate a CF Migration Summary for each major CF.



CFQ100 CF Migration Summary Report

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)

CFQL001 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

CF Link Window

CFQL001 Report

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.

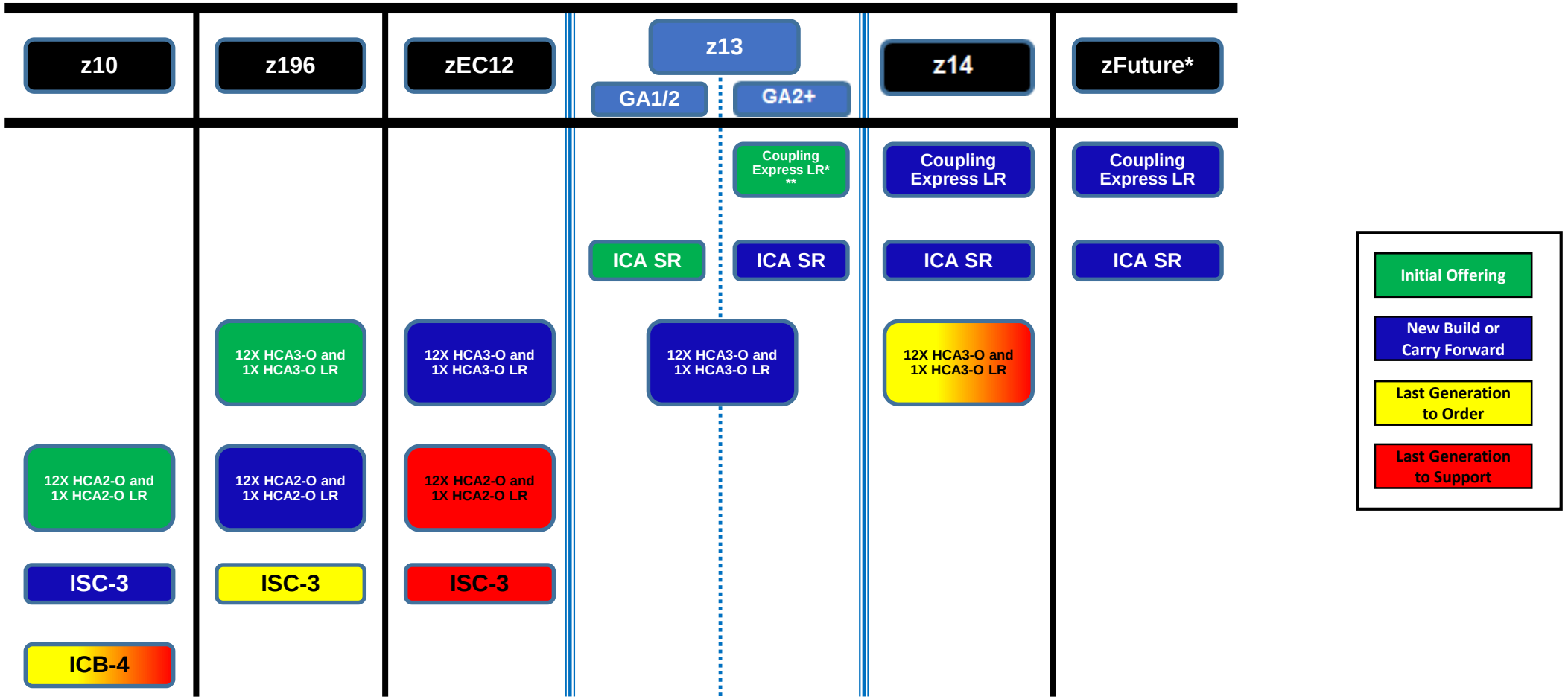
Graph CFQL001 on the CF Link Window (RMF only)

Agenda - Solving Sysplex Questions with zCP3000

- New Technology
- Getting Started
- Understand problem areas and key resources
- Model a new hardware or links or Thin Interrupt
- **Coupling Link Considerations**

Reference slides include links, graph recommendations, and references.

Coupling Link Roadmap – High End (HE) Systems

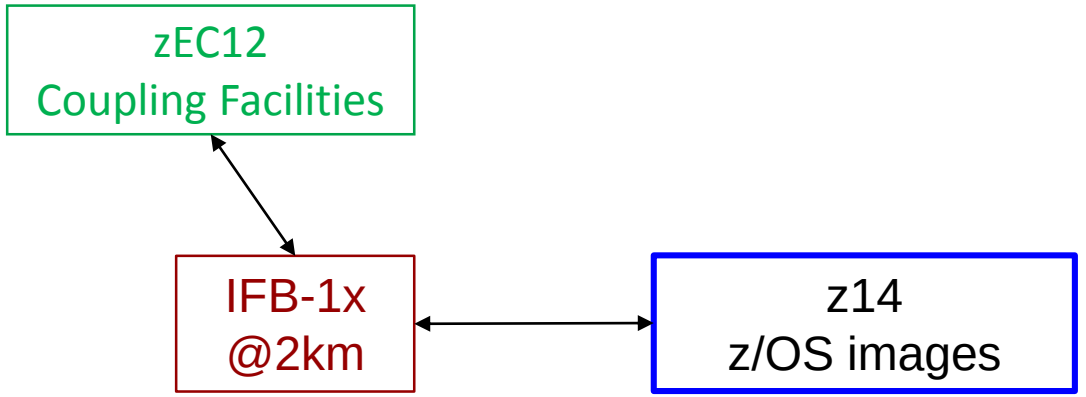


InfiniBand Link Migration problem#1

Fast migration to ICA SR short-distance coupling link technology.
Really fast migration to CE LR long-distance coupling.

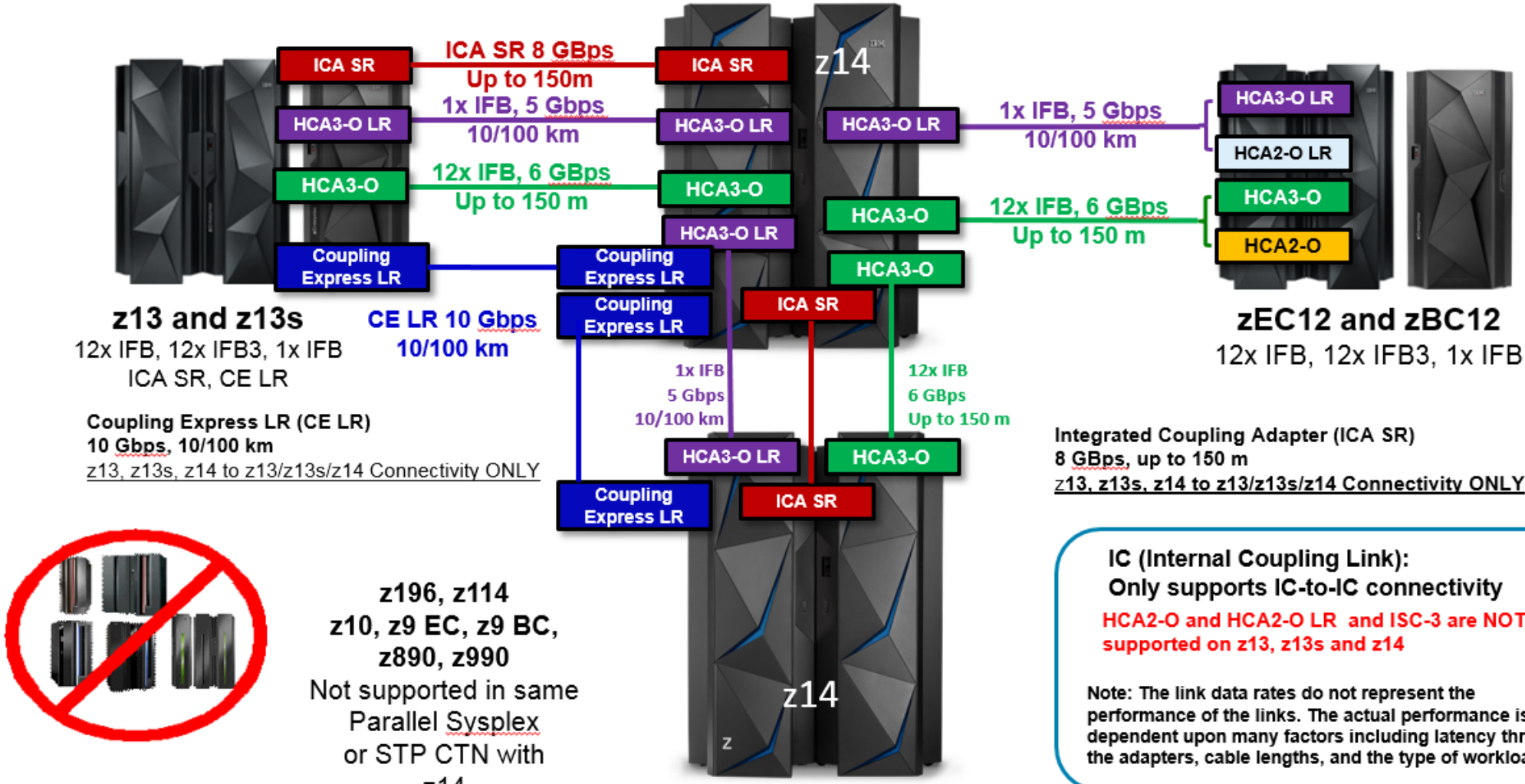
- ICA links : z13 and up
- CE-LR links : z13 and up
- IFB3 : z196,z114 through z14
- IFB-12x : z9-EC through z14
- IFB-1x : z10-EC through z14

- ISC-3 : Jurassic period through zEC12
- zEC12/zBC12 WFM Dec 31, 2017

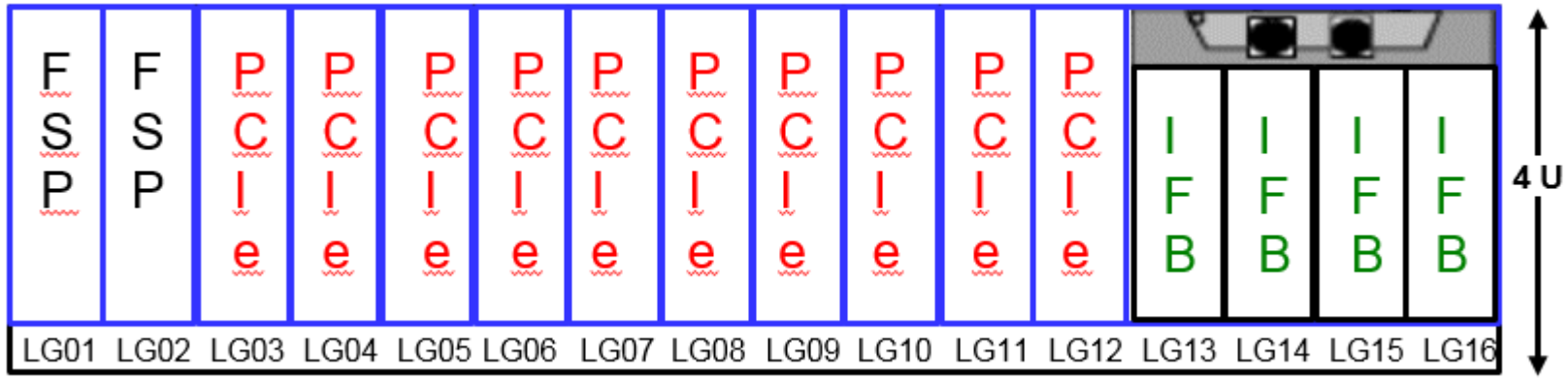


Standalone CF will have to upgrade before z14 PROD can.

z14 Coupling Connectivity



z14 CPC Drawer I/O Fanout



PCIe Fanout Slots (Ten), slots LG03 – LG12 support:

- Up to 10 one-port PCIe 16 GBps I/O fanouts to support up to 10 domains in 32-slot PCIe I/O drawers
- Up to 10 ICA (short range) or CE-LR (long range) two-port coupling fanouts

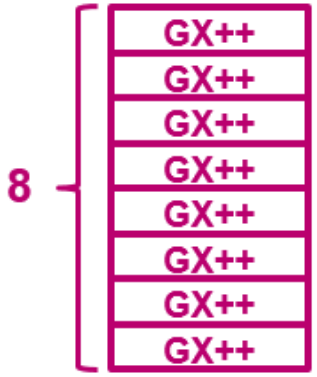
IFB Fanout Slots (Four), LG13 – LG16, can support:

- Up to four HCA3-O 12x InfiniBand coupling fanouts, 8 12x 6 GBps links – Two per fanout
- Up to four HCA3-O LR 1x InfiniBand coupling fanouts 16 1x 5 Gbps links – Four per fanout

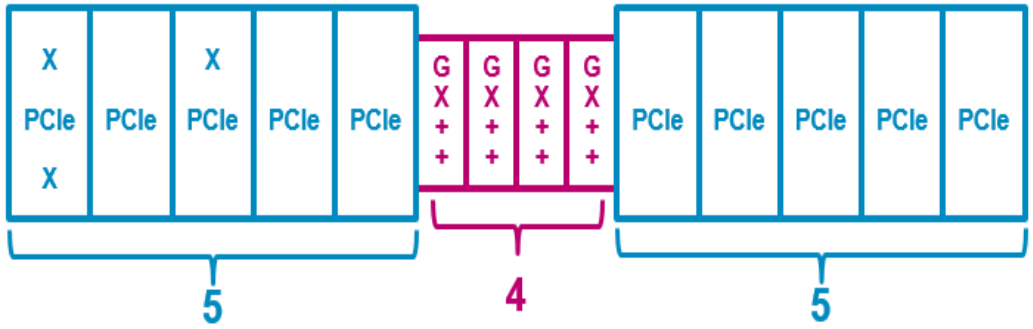
* z14 is the LAST IBM Z server to support InfiniBand features (SoD)

InfiniBand Link Migration problem#2

zEC12 / z196 Book



z13/z14 Drawer



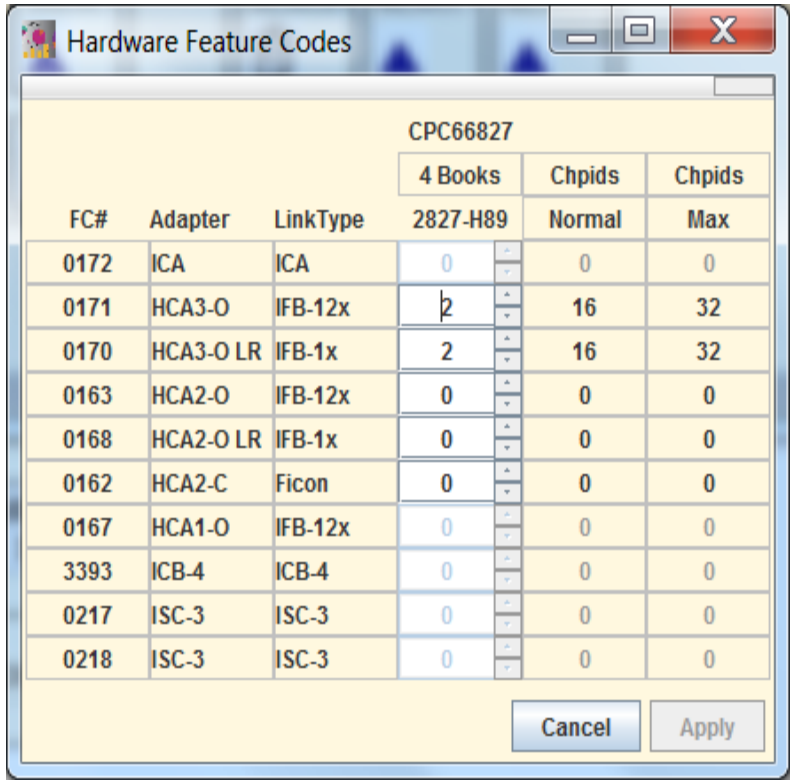
InfiniBand uses GX++ fanouts

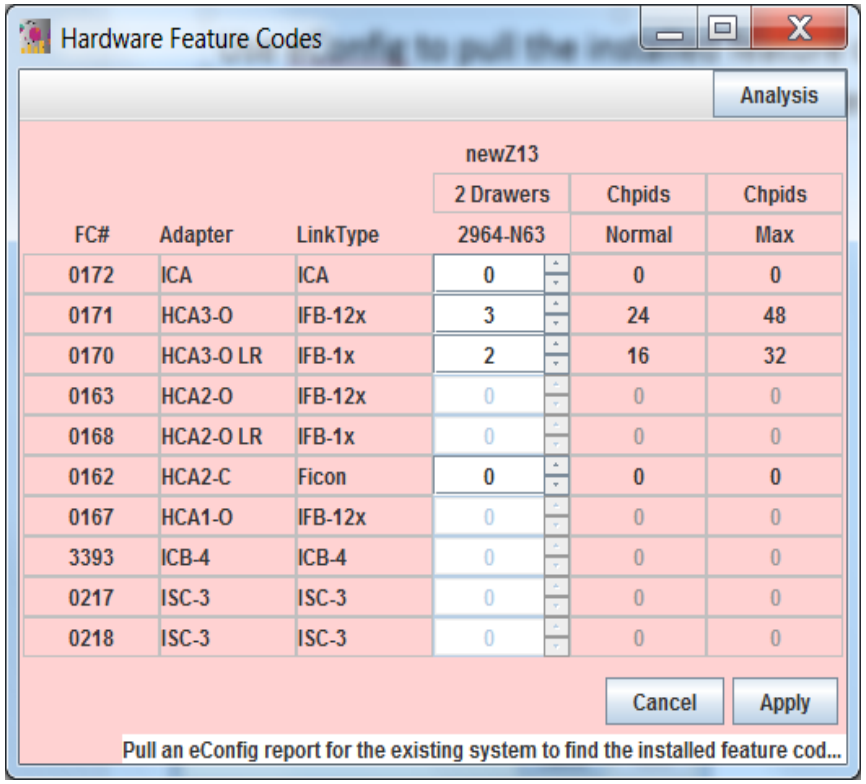
ICA-SR and CE-LR use PCIe fanouts

Migrating the same number of IFB coupling links to z13/z14 from z196/zEC12 may require additional CPC drawers

InfiniBand Link Migration

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.





Adding drawers just for the Gx++ fanouts?

- Migrate 12x and IFB3 to (faster) ICA type links.
- Migrate 1x to (faster) CE Long Range links.
- ICA and CE-LR are z13 and above only.
- Eliminate FICON Express8 (also uses GX++)
- Consolidate chpids on to fewer IFB/ICA links

→ Max 4 chpids per port for IFB3, ICA, CE-LR

InfiniBand Link Migration problem #3

IFB-12x and 1x allowed up to 8 chpids per link.
IFB3 protocol allowed max 4 chpids per link.

ICA and CE-LR allow max 4 chpids per link.

IFB links* with >4 chpids need to be reconfigured.

How do you know if you have this problem?

- CF1000 CF Healthcheck Links rule about mixed speed (some IFB3, some IFB-12x)
- PLEX1012 Sysplex Coupling Links Summary lists the chpids per adapter/port

PLEX1012 Sysplex Coupling Links Summary

These 5 chpids will run at 12x not IFB3 speed

Sysplex Member					LinkType	Coupling Facility				
CECID	Adapter	Port	Chpids	Sys/CF		CECID	Adapter	Port	Chpids	Sys/CF
CPC900D7	0008	01	80	[GFP1, SYS1, GFP3, PRD3, PRD1, GFD1]	HCA3-O				40	[WSCSYSE]
			82	[GFP1, GFP3, PRD3, PRD1, GFD1]	HCA3-O	CFSR893B7	0008	02	42	[WSCPRDE]
			84	[GFD1]	HCA3-O				44	[GFCFDE]
			85	[GFP1, GFP3, GFD1]	HCA3-O				45	[GFCFPE]
CPC900D7	0008	02	88	[PRD3, GFCFDI, GFCFPI, PRD1, GFCFSI]	HCA3-O	CPC900E7	0008	01	80	[WSCPRDI]
			89	[SYS1, PRD3, GFCFDI, GFCFPI, PRD1, GFCFSI]	HCA3-O				81	[WSCSYSI]
CPC900D7	0009	01	90	[GFP1, SYS1, GFP3, PRD3, PRD1, GFS1, GFD1]	HCA3-O	CFSR893B7	0009	02	50	[GFCFSE]
			92	[GFP1, GFP3, PRD3, PRD1, GFS1, GFD1]	HCA3-O				52	[WSCSYSE]
			94	[GFD1]	HCA3-O				54	[WSCPRDE]
			95	[GFP1, GFP3, GFD1]	HCA3-O				55	[GFCFDE]
			96	[GFP1, GFP3, GFS1, GFD1]	HCA3-O				56	[GFCFPE]
CPC900D7	0009	02	98	[PRD3, GFCFDI, GFCFPI, PRD1, GFCFSI]	HCA3-O	CPC900E7	0009	01	90	[WSCPRDI]
CPC900D7	000A	01	A2	[GFP1, GFP3, PRD3, PRD1, GFS1, GFD1]	HCA3-O	CFSR893B7	000A	02	32	[WSCPRDE]
			A4	[GFD1]	HCA3-O				34	[GFCFDE]
			A5	[GFP1, GFP3, GFD1]	HCA3-O				35	[GFCFPE]
			A6	[GFP1, GFP3, GFS1, GFD1]	HCA3-O				36	[GFCFSE]
CPC900D7	000A	02	A8	[PRD3, GFCFDI, GFCFPI, PRD1]	HCA3-O	CPC900E7	000A	01	A0	[WSCPRDI]
			AA	[SYS1, PRD3, GFCFDI, GFCFPI, PRD1]	HCA3-O				A2	[WSCSYSI]
CPC900D7	000B	01	B2	[GFP1, GFP3, PRD3, PRD1, GFD1]	HCA3-O	CFSR893B7	000B	02	B2	[WSCPRDE]
			B4	[GFD1]	HCA3-O				B4	[GFCFDE]
			B5	[GFP1, GFP3, GFD1]	HCA3-O				B5	[GFCFPE]
CPC900D7	000B	02	B8	[PRD3, GFCFDI, GFCFPI, PRD1]	HCA3-O	CPC900E7	000B	01	B0	[WSCPRDI]

Table 1. Coupling Link Summary for CPC900D7

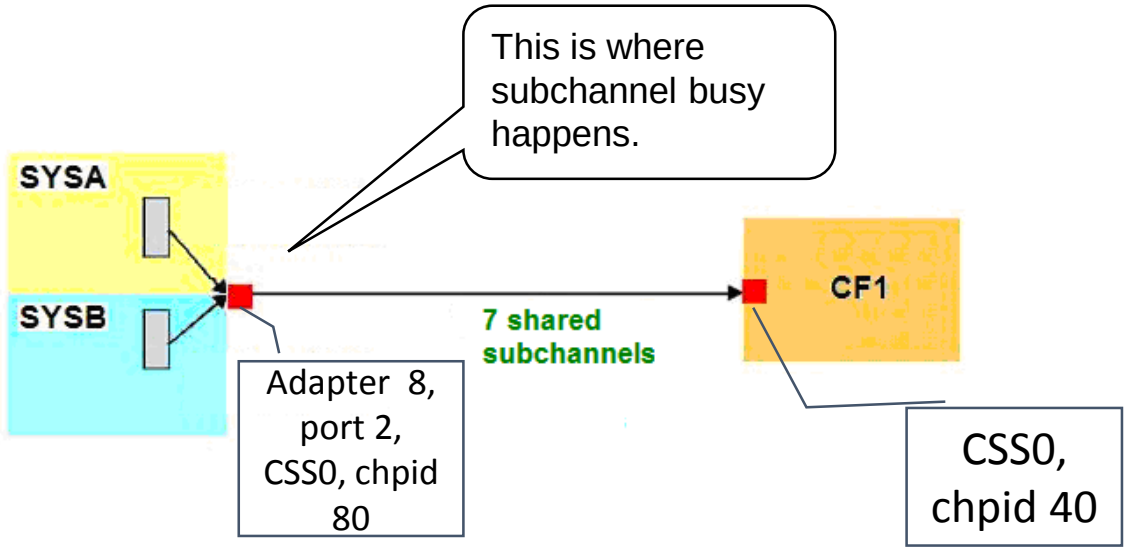
Import from IOCP

You can’t migrate these 5 chpids directly from InfiniBand to ICA-SR type links.

Shared links vs shared channel paths

(a primer)

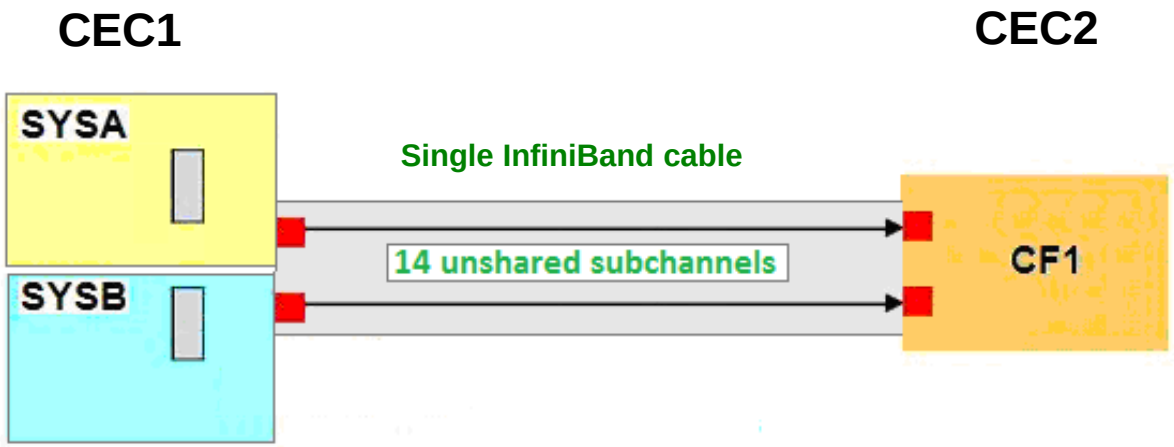
Shared Channel Paths



IOCP Definition of a shared cf link:

```
CHPID PATH=(CSS(0),80),SHARED,
          PARTITION=((WSCSYSA,WSCSYSB),(=)),
          CPATH=(CSS(0),40),CSYSTEM=EXTRNCF1,
          AID=08,PORT=1,TYPE=CIB
```

Multiple CHPIDs on a single InfiniBand Link



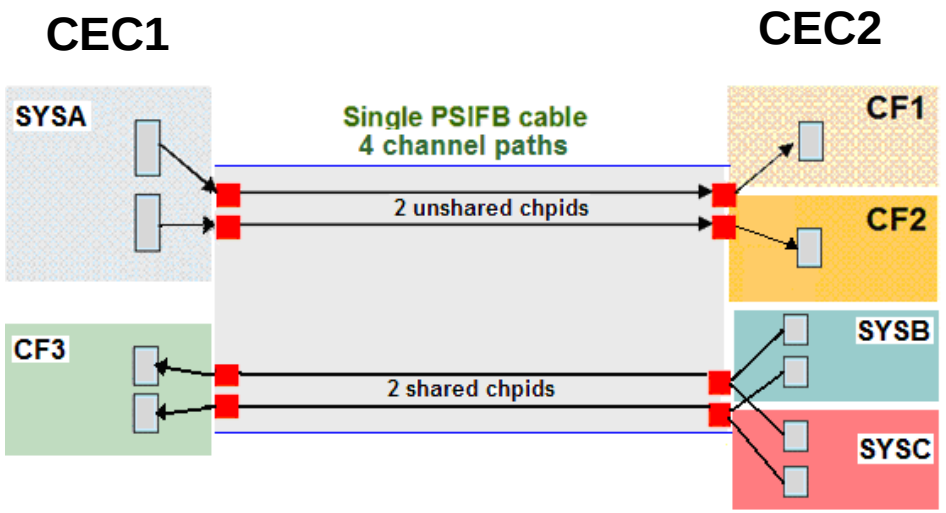
This InfiniBand link carries 2 unshared chpids.

- **No subchannel busy** because the chpids are unshared.
- This physical link can carry 2 more chpids.*
- Generally, the same adapter type on both ends.**

Shared Chpids vs Shared Links

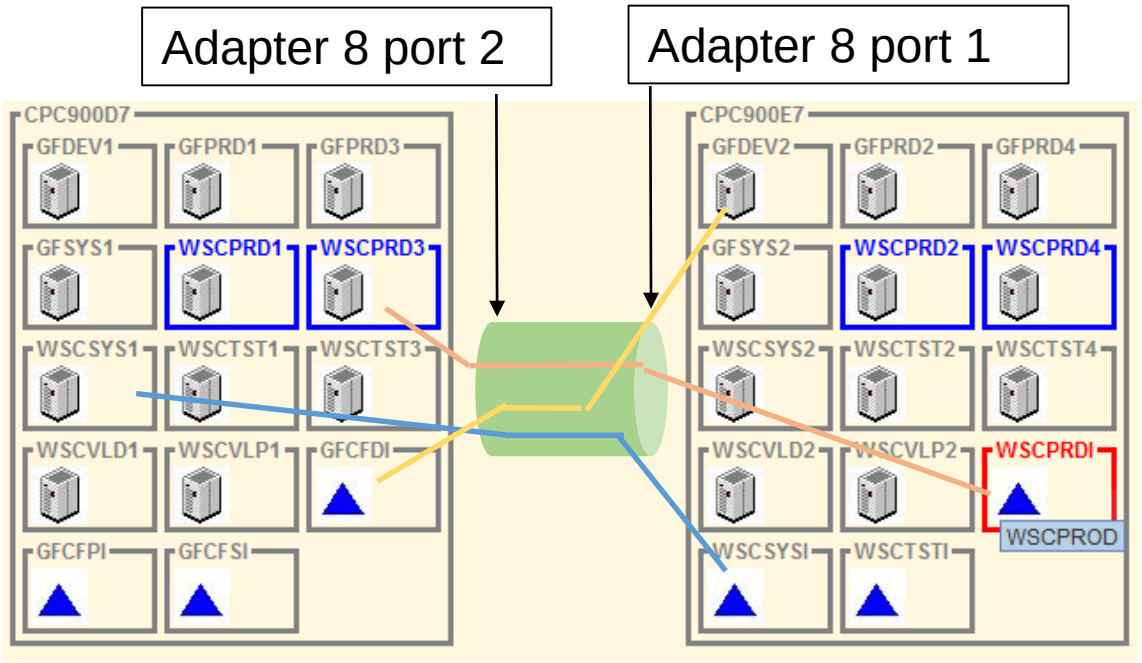
This InfiniBand link carries 4 chpids.

There is also a 2nd port (physical link) not shown.



- InfiniBand will let you configure more than 4 chpids per port, but you should not. ICA allows max 4.
- SYSA has one chpid each to CF1 and CF2; it 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 2 chpids to CF3, but they are shared. CF3 sees 2 chpids.

Shared Physical Links

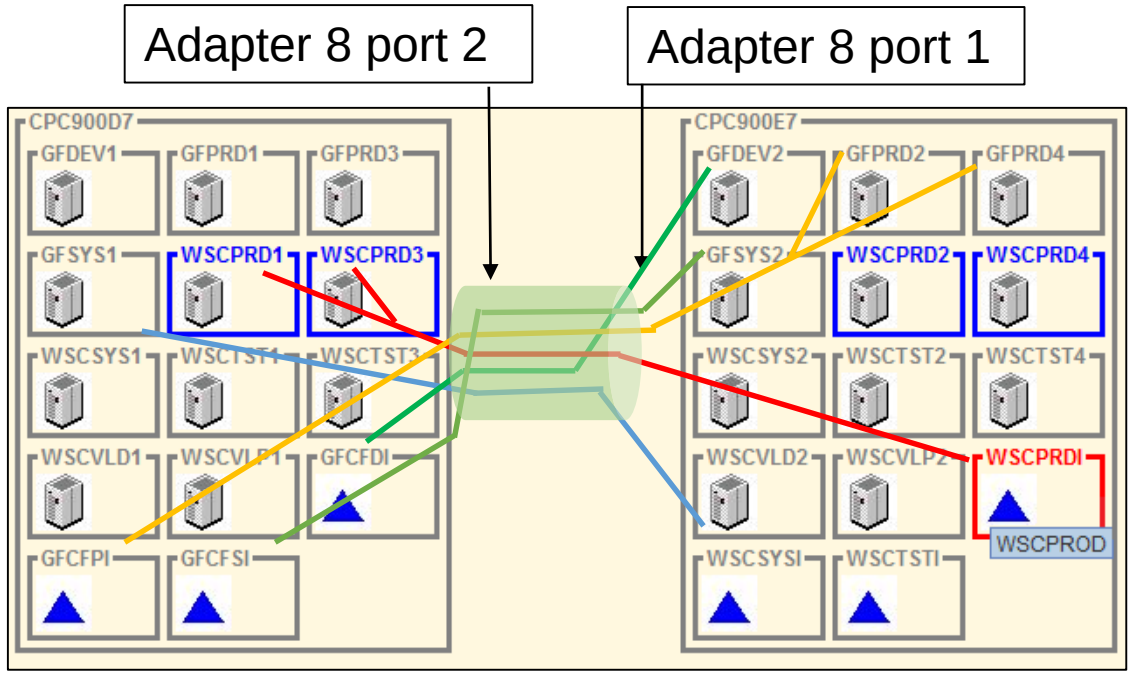


A simplified view.

- all of the chpids running through an adapter/port.
- multiple sysplexes
- z/OS partitions can share chpids
- SYSes and CFs on either end, or both.

Shared Physical Links

This link is carrying 5 chpids, so all 5 will run in 12x mode not IFB3.



Chpid 88 on 900D7 adapter 8 port 2, connecting to chpid 80 on 900E7 adapter 8 port 1.

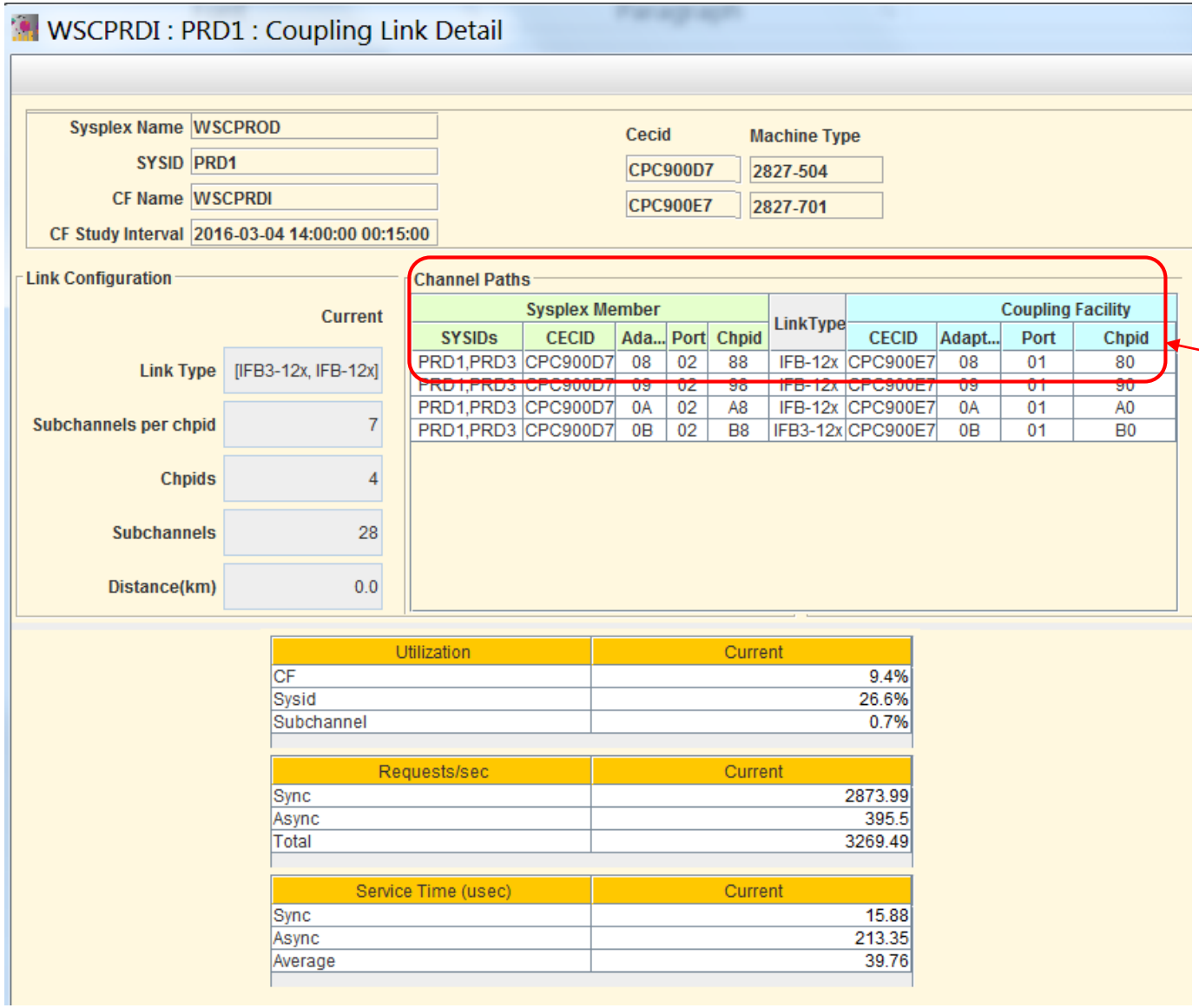
Chpid 89 on 900D7 adapter 8 port 2, connecting to chpid 81 on 900E7 adapter 8 port 1.

Chpid 8D on 900D7 adapter 8 port 2, connecting to chpid 85 on 900E7 adapter 8 port 1.

Chpid 8E on 900D7 adapter 8 port 2, connecting to chpid 86 on 900E7 adapter 8 port 1.

Chpid 8F on 900D7 adapter 8 port 2, connecting to chpid 87 on 900E7 adapter 8 port 1.

Shared Links, Shared Chpids



WSCPRDI : PRD1 : Coupling Link Detail

Sysplex Name	WSCPROD
SYSID	PRD1
CF Name	WSCPRDI
CF Study Interval	2016-03-04 14:00:00 00:15:00

Cecid	Machine Type
CPC900D7	2827-504
CPC900E7	2827-701

Link Configuration

	Current
Link Type	[IFB3-12x, IFB-12x]
Subchannels per chpid	7
Chpids	4
Subchannels	28
Distance(km)	0.0

Channel Paths

Sysplex Member					LinkType	Coupling Facility			
SYSIDs	CECID	Ada...	Port	Chpid		CECID	Adapt...	Port	Chpid
PRD1,PRD3	CPC900D7	08	02	88	IFB-12x	CPC900E7	08	01	80
PRD1,PRD3	CPC900D7	09	02	98	IFB-12x	CPC900E7	09	01	90
PRD1,PRD3	CPC900D7	0A	02	A8	IFB-12x	CPC900E7	0A	01	A0
PRD1,PRD3	CPC900D7	0B	02	B8	IFB3-12x	CPC900E7	0B	01	B0

Utilization	Current
CF	9.4%
Sysid	26.6%
Subchannel	0.7%

Requests/sec	Current
Sync	2873.99
Async	395.5
Total	3269.49

Service Time (usec)	Current
Sync	15.88
Async	213.35
Average	39.76

Why are 3 of these chpids in 12x mode?

Chpid 88 runs on that link we just saw, which carries 5 chpids.

Noted in
CF1000 CF Health Check
PLEX1012 Links Summary

CF1000 CF Healthcheck for WSCPRDI

Links Rule 6 notices that there are different speed links.

Links Rule 6 - Mixed Link Speeds - Caution

For some WSCPROD sysplex members, link speeds between WSCPRDI and that sysplex member operated at different speeds. This may be as intended, however it can make transaction speeds less consistent.

SYSID	links	chpids	Mbs/sec	LinkType
PRD1	1	[B8]	5,000	IFB3 protocol, adapter type HCA3-O
	3	[88, A8, 98]	1,000	IFB-12x linktype, adapter type HCA3-O
PRD2	2	n/a	9,400	IC IC
PRD3	1	[B8]	5,000	IFB3 protocol, adapter type HCA3-O
	3	[88, A8, 98]	1,000	IFB-12x linktype, adapter type HCA3-O
PRD4	2	n/a	9,400	IC IC

PLEX1012 Coupling Links Summary

Shows you why chpid 88 is running in IFB-12x, not IFB3, link speed.

Sysplex Member					LinkType	Coupling Facility				
CECID	Adapter	Port	Chpids	Sys/CF		CECID	Adapter	Port	Chpids	Sys/CF
CPC900D7	0008	01	80	[GFP1, SYS1, GFP3, PRD3, PRD1, GFD1]	HCA3-O	CFSR893B7	0008	02	40	[WSCSYSE]
			82	[GFP1, GFP3, PRD3, PRD1, GFD1]	HCA3-O				42	[WSCPRDE]
			84	[GFD1]	HCA3-O				44	[GFCFDE]
			85	[GFP1, GFP3, GFD1]	HCA3-O				45	[GFCFPE]
CPC900D7	0008	02	88	[PRD3, GFCFDI, GFCFPI, PRD1, GFCFSI]	HCA3-O	CPC900E7	0008	01	80	[WSCPRDI]
			89	[SYS1, PRD3, GFCFDI, GFCFPI, PRD1, GFCFSI]	HCA3-O				81	[WSCSYSI]

Sysplex Member					LinkType	Coupling Facility				
CECID	Adapter	Port	Chpids	Sys/CF		CECID	Adapter	Port	Chpids	Sys/CF
CPC900E7	0008	01	85	[GFP2, GFP4, GFD2, GFS2]	null	CPC900D7	0008	02	8D	[GFCFSI]
			86	[GFD2]	null				8E	[GFCFDI]
			87	[GFP2, GFP4, GFD2]	null				8F	[GFCFPI]

(because there are 5 chpids configured on CPC900D7/8/2 link to CPC900E7/8/1)

CFL005 CF Link Topology Report

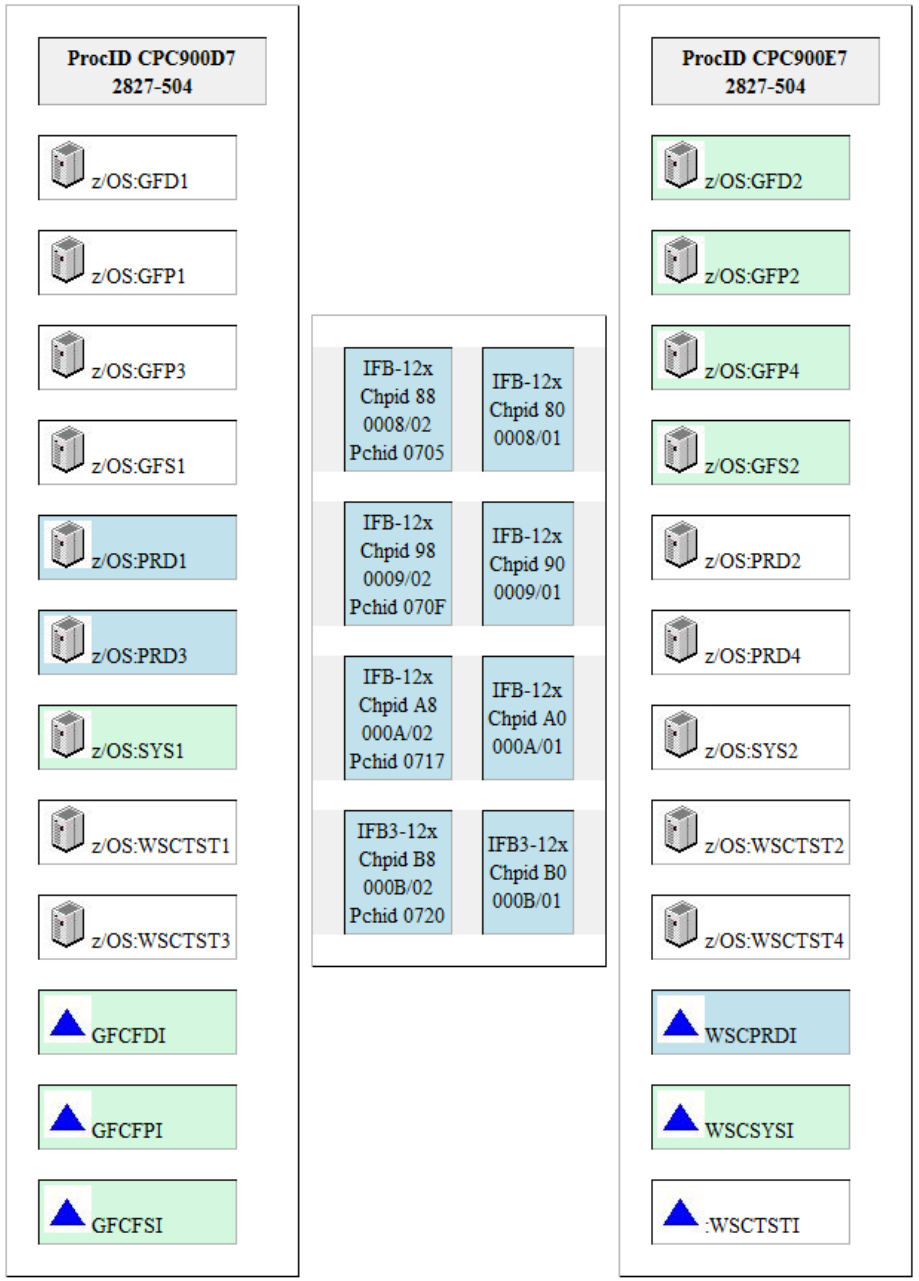
Looking at the link from CF WSCPRDI To z/OS PRD1

CHPID Sharing

PRD1 and PRD3 CPC900D7 share 4 chpids to the WSCPRDI coupling facility on CPC900E7, running over 4 different physical links.

Link Sharing

Other z/OS systems and coupling facilities are using other chpids which also run over these same 4 physical links.



CF Link Migration

- **Start planning the move to ICA and CE-LR now.**
- **Moving from IFB to ICA links?** both ends must be z13 and above.
- **Migrating IFB links to z13?** z13 cannot carry as many IFB per book.
- Ask for current IOCP decks for all machines, including standalone CF machines.
- Replace chpids one for one.

- **CF Health Check – Subchannel Busy**? Configure more chpids and share less. Or reduce chpids per CF-SYS connection, but share with fewer partitions.
- One power source per adapter (not port), so minimum two adapters per CPC.
- Maximum 4 chpids per port → **CF Health Check – Subchannel Mixed Linktypes**

Support in zCP3000

- Channel Path data is already in CP3KEXTR and zCP3000.
- Look for problems flagged in the CF Healthchecks.
- Use the physical link based reports, like topology.
- Modeler will need to supply some configuration information.

Performance data is currently reported for CF-SYS. It is *NOT* reported per channel path. There is no way to tell you, for example, how much traffic is going over chpid 88.

Solving Sysplex Questions with zCP3000

Reference Slides

- Recommended Graphs and Graph Index
- Connectivity Reference : Max Supported Coupling Links
- What is Coupling Thin Interrupt, anyway?
- Dispatch Modes, Effective CPs
- Urls, Contacts, and how to stay informed.

Sysplex Graph Recommendations

Probably in every study

- zCP3000 main view : Topology
- per CPC (Coupling Links window): Coupling Links Summary
- Sysplex Logical window : Sysplex Topology, Coupling 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
- Coupling Adapter Summary
- 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

Index to recommended Graphs

zCP3000 main view

- INV1012 Topology

zCP3000 sysplex view

- PLEX1007 Sysplex Overview
- PLEX1008 Sysplex Coupling Summary
- PLEX1005 Sysplex Aggregation

CPC Coupling Links Window

- PLEX1012 Coupling Links Summary

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

PA Mode

PA and QM Mode

QM only

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

CF Connectivity Reference

Max Supported Coupling Links									
	z14	**z13s**	**z13**	**zEC12**	**zBC12**	**z196**	**z114**	**z10**	**z9**
IC		32	32	32	32	32	32	32	32
CE LR	64	32	32	n/a	n/a	n/a	n/a	n/a	n/a
ICA SR	64	32	40	n/a	n/a	n/a	n/a	n/a	n/a
HCA3-O LR (1x)	64	32	64	64	32	48	32	n/a	n/a
HCA3-O (12x)	32	16	32	32	16	32	16	n/a	n/a
HCA2-O LR (1x)	n/a	n/a	n/a	32	12	32	12	32	n/a
HCA2-O (12x)	n/a	n/a	n/a	32	16	32	16	32	n/a
HCA1-O (12x)	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	16
ISC-3	n/a	n/a	n/a	48	32	48	48	48	48
Max chpids	512	256	256	128	128	128	128	64	64

Link = Port (2 ports per adapter, except 4 port IFB-1X on HCA3-O LR)

Coupling Thin Interrupt

Starting with zBC12 and zEC12-GA2 the arrival of an unsolicited signal on a Coupling Facility link will generate an interrupt.

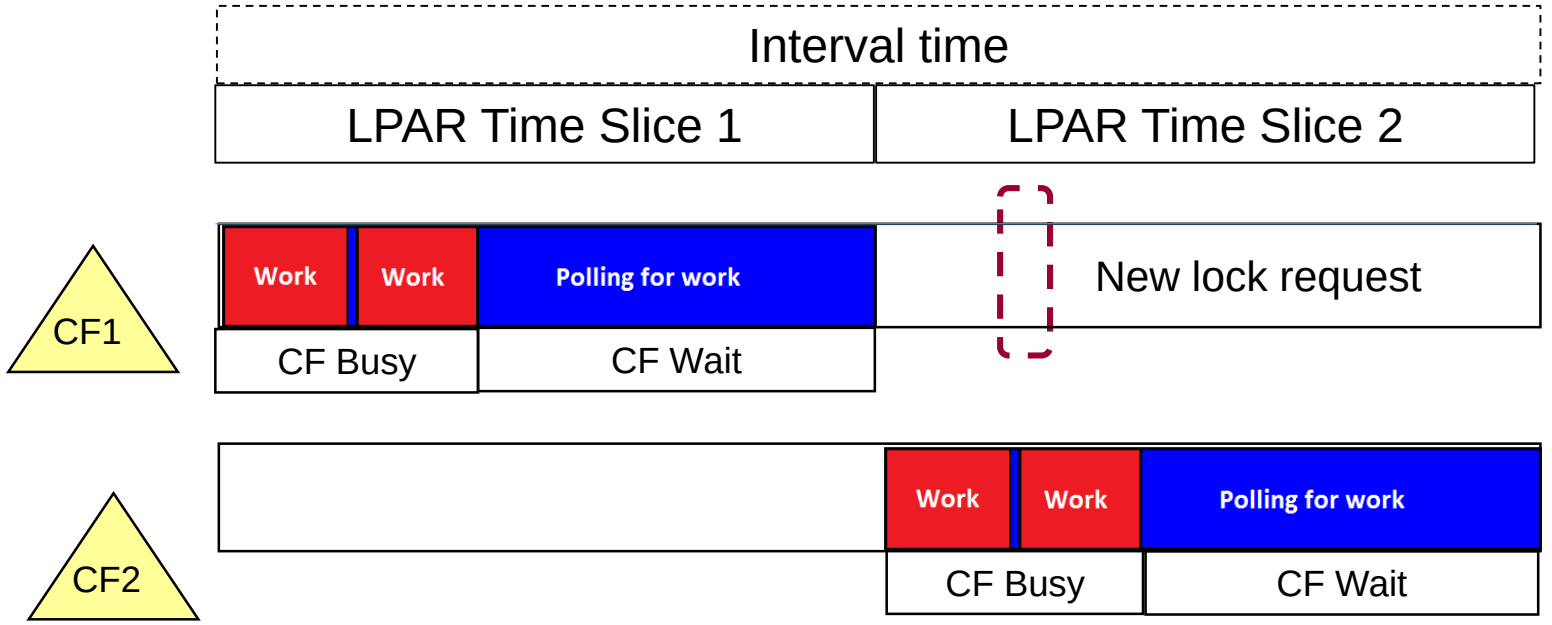
- Both z/OS and the Coupling Facility can exploit this.
- CF and z/OS exploitation are independent of each other.
- z/OS support mainly benefits low weight partitions
- For Coupling Facilities, it benefits CFs sharing CPU resource
- It does not benefit CFs with dedicated engines.

IBM recommends:

- Use dedicated engines for production CFs
- Otherwise enable Coupling Thin Interrupt for all CFs

CF Dispatch mode - Polling

"Regular" Sharing - **Both CF1 and CF2 have DYNDISP=OFF**



Default LPAR Time Slice

12.5 – 25 ms

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

EFFCP =~ 0.5 engines

CF (LPAR) Utilization = 17.5%

CFCC Busy = 35%

Big difference in scale : microsec (sysplexOp) vs millisec(LPAR slice)

Effective CPs vs Entitled CPs

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

EFFCP =~ 0.5 engines

CF (LPAR) Utilization = 17.5%

CFCC Busy = 35%

Entitled engines = relative weight / # of shareable engines of that type.

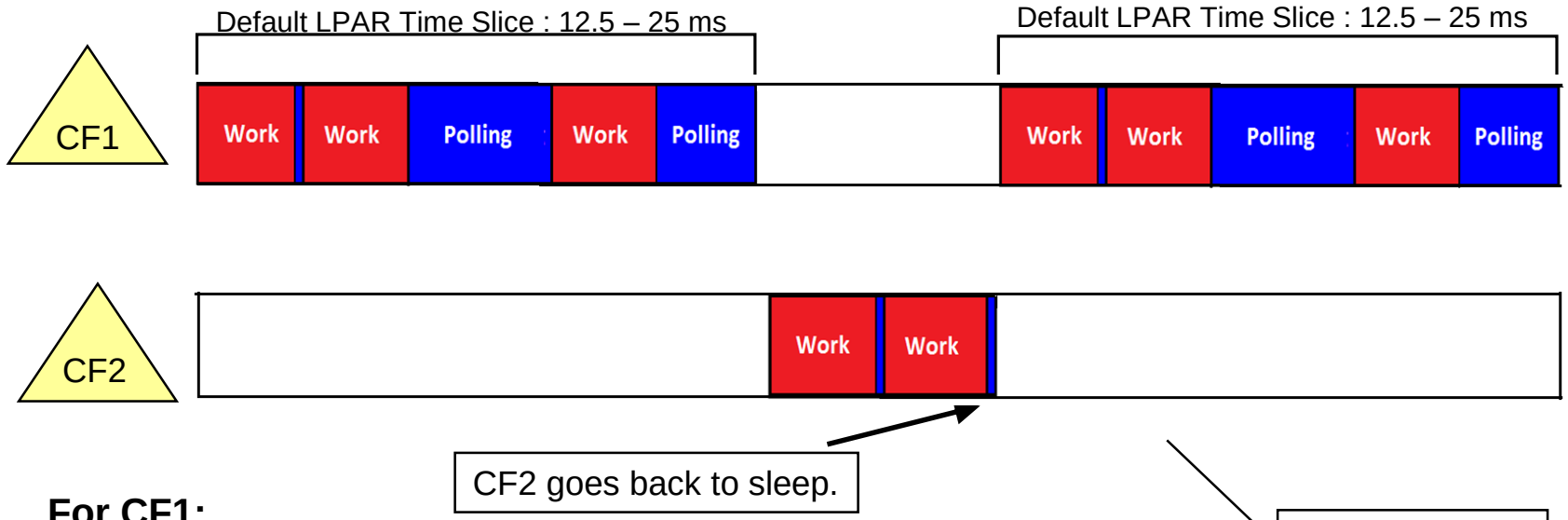
- The engine equivalent of what the partition is entitled to.
- Not the same thing as capping.
- Partition can receive more or less than entitlement.

It's what LP mgmt is trying to deliver, if the partition wants it.
And the partition is configured to be able to use it.

Don't ignore this zCP3000 msg:
"Weight indicates more capacity than can be provided with the LCPs defined for ICF pool".

CF Dispatch mode - Dynamic CF Dispatch

CF1 has DYNDISP=OFF & CF2 has DYNDISP=ON



For CF1:

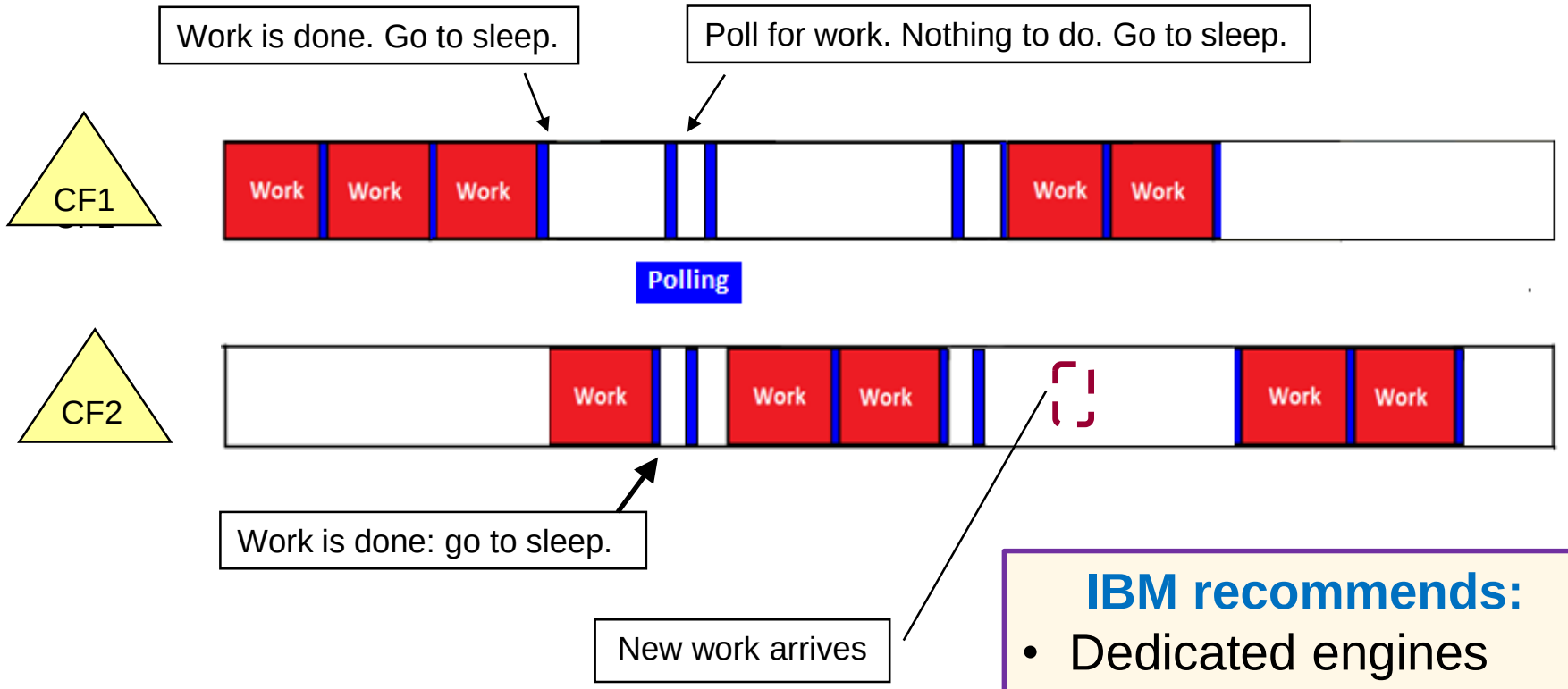
EFFCP =~ 0.85 engines

CF (LPAR) Utilization = 45%

CFCC Busy = 56%

CF Dispatch mode - Thin Interrupt

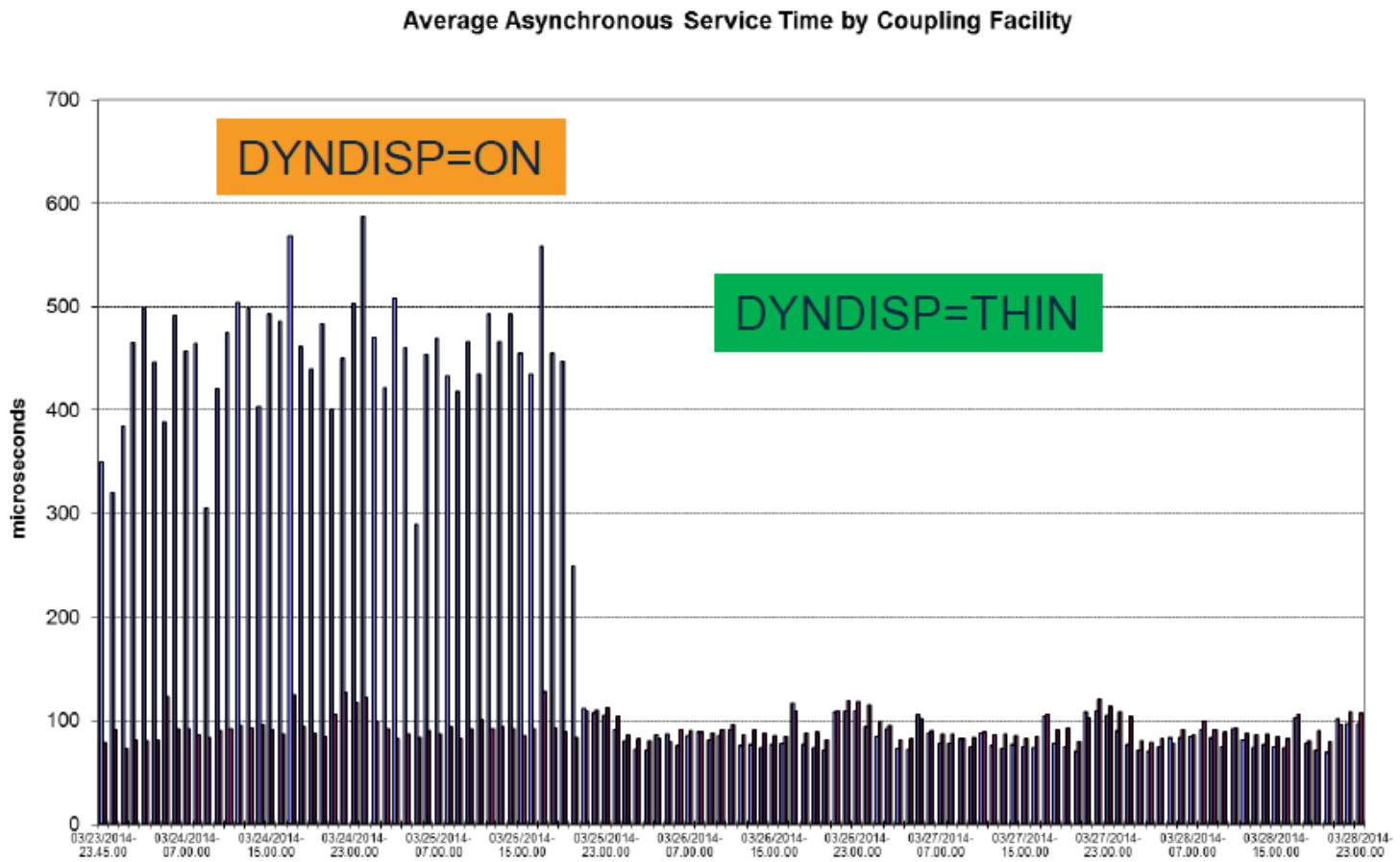
Both CF1 and CF2 have DYNDISP=THININTERRUPT



IBM recommends:

- Dedicated engines
- Otherwise Thin Interrupt

Impact of Thin Interrupt



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

How to stay informed

zCP3000 Google Group (private)

https://groups.google.com/forum/#!forum/zcp3000

Release announcements, comments, urgent problems, ideas, whatever.

IBMers

Program WebPage	Download	User's Guide	Online Education	Change Notification	
zCP3000 v06/16	Download 06/24/2016	01/31/2013	zCP3000 Library Demos and Labs	Subscribe	Unsubscribe

ALWAYS pull down the latest version of CP3KEXTR and zCP3000 when starting a new study.

Useful links

	IBM W3	PartnerWorld
zCP3000 Download & Doc	PRS1772	PRS1865
zCP3000 Demos and Labs	PRS5318	PRS5319

Barbara Weiler's white paper on Coupling Thin Interrupts
https://www-03.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP102400

Frank Kyne: The Skinny on Coupling Thin Interrupts
https://share.confex.com/share/123/webprogram/Session15602.html

Other 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