Is a Sub-Capacity CPC the Right Move for You?

Frank Kyne, Watson & Walker

Welcome

- Who are we?
  - Watson & Walker founded in 1988 by **Cheryl Watson** & **Tom Walker**
  - Publisher of *Cheryl Watson's Tuning Letter* and *CPU Charts* since 1991.
  - After the Tuning Letter, our primary focus is on helping our customers understand their software bills and select the pricing and technical options that deliver the best value for them.
  - We are completely independent, not beholden to any vendor, so we can offer objective information based on our collective experience and what we see in other customers, thereby allowing our clients to make a fully informed decision.
  - For more information, see www.watsonwalker.com.

Agenda

- Introduction
- Why a Sub-Cap model might be a good fit for you
- What are the challenges?
- Evaluation Criteria
- Summary and Questions

Introduction

Introduction

- I want to thank **Todd Havekost** for all his help and support and enthusiasm. Apart from all his help with this webinar AND his very informative articles in every Tuning Letter, Todd is also my one-man encyclopedia for all things IntelliMagic Vision-related.
  - Most of the SMF charts in this presentation were created by Todd, or with Todd's assistance.
- Also want to thank members of IBM's Z performance team for their ongoing help and support and patience.
- This presentation is based on our work with helping a number of customers evaluate upgrades to sub-cap CPC models.

Introduction

- IBM has had ‘enterprise class’ mainframes with multiple speed ranges (4xx, 5xx, 6xx, 7xx) going back as far as the z9 in 2005.
  - The 4xx, 5xx, and 6xx ranges are known as sub-capacity models.
- These were originally created in order to give smaller customers more granular upgrade options.
- However, a side effect of having more, slower, PUs is that the amount of cache per MIPS is higher, and the number of PUs to deliver a given number of MIPS is higher.
- In an environment where effective use of cache is *one* of the determinants of how much work a CPC can do, these sub-cap CPCs have been known to deliver more capacity than expected.

Introduction

- One much-publicized example was NASCO, who upgraded from a z13 709 to a z14 523.
  - NASCO's **Paul Snyder** and **Dave Laaker** kindly shared their experiences with our Tuning Letter subscribers in Tuning Letter 2018 No. 3.
- NASCO had an Average-to-High RNI workload. Their LPAR topology was tuned as well as possible, but with a total of only 9 GCPs, the number of possible Vertical High CPs was limited.
- Based on their workload category, IBM's zPCR tool showed 11,424 MIPS for their 709, and 11,848 MIPS for the target z14 523 – a 3.7% increase in capacity.
- But because of the increased number of PUs, the increased amount of cache, ***and*** how their workloads and LPAR configuration interacted with the 523, their peak R4HA dropped by 22% after the move.

Introduction

- **NOT EVERYONE WILL ACHIEVE THESE RESULTS**.
- AND, remember that an x% reduction in peak R4HA does NOT mean the same percent reduction in software costs.
- However, the *potential* savings are such that we believe that sub-cap CPCs should at least be *evaluated* by anyone upgrading to a CPC with less than 20-25K General Purpose CP MIPS.
- We see a growth in the number of large sites that have a mix of larger CPCs for production, and smaller CPCs for development – those smaller CPCs *might* be excellent candidates for a sub-cap model.
- In particular, PLEASE don't just automatically purchase a 7xx model 'because that is what we've always done'. The potential savings are too large to ignore.

Why a Sub-Cap Model Might Be a Good Fit for you

Might a Sub-Cap CPC Be a Good Fit For You?

- *Our* clients' experiences with moving from one speed range (a 7xx, for example) to a smaller one (4xx, 5xx, or 6xx) have been very positive.
  - Throughput and response times have exceeded expectations.
    - We are only aware of *one* (special) case where the CPC fell short of expectations.
  - Reported MSU consumption for the same work is typically lower than projected.
    - This means a smaller software bill to do the same amount of work.
  - CPC Upgrades are based on a price per Average MI Workload MIPS.
    - Combining the greater granularity with the potential for a sub-cap model to deliver more capacity than expected, you *might* be able to reduce upgrade costs by purchasing a smaller upgrade than you had planned on.
    - At a minimum, the more granular upgrade options on sub-cap models *might* help you find a better fit than if you are limited to selecting a 7xx model.

Might a Sub-Cap CPC Be a Good Fit For You?

- But, there *are* limitations of sub-cap CPCs:
  - You are limited to a max of 34 general purpose CPs on a z15, fewer on previous generations.
  - Because the general purpose CPs are slower, the maximum general purpose CP MIPS is limited:
    - z15 4xx 6382 Average MI Workload MIPS
    - z15 5xx 18057 Average MI Workload MIPS
    - z15 6xx 25887 Average MI Workload MIPS
  - Most important limitation is the per-CP Speed:
    - z15 401 267 Average MI Workload MIPS
    - z15 501 781 Average MI Workload MIPS
    - z15 601 1151 Average MI Workload MIPS
    - (z15 701 is 2055 Average MI Workload MIPS)

Might a Sub-Cap CPC Be a Good Fit For You?

- There are also positive aspects:
  - Special purpose engines (zIIP, IFL, ICF) *always* run at full 7xx speed.
  - The limit of 34 engines applies *only* to GCPs. For example, you could have a z15 434 with 34 GCPs *and* 68 (full speed) zIIPs.
  - For CBU or OOCoD purposes, you are not limited to upgrades within the same speed range – a 510 could be CBUed to a 710 if you wish.
  - On z15, System Recovery Boost increases the speed of sub-cap GCPs during shutdown and recovery to the speed of a 7xx.
  - From a performance perspective, the cache/MIPS is higher on a subcap. Even though the engine speed is lower, the size of each cache is the same as the full speed 7xx model, regardless of speed range.
  - Having the same number of MIPS spread over more CPs opens the possibility of having more Vertical High CPs.

What are the challenges?

What are the challenges?

- There is no free lunch. Evaluating a move to a sub-cap CPC is more work than moving to the equivalent newer model.
- As mentioned previously, the engine speed can be a challenge if you have large monolithic workloads that can't use multiple CPs:
  - For example, long running batch programs that consume large amounts of CPU time *and* that are on the critical path.
  - Some ISV products that consume a large amount of CPU on a single TCB.
  - Large non-threadsafe CICS regions that do, or could, encounter a large amount of contention on the QR TCB.
  - Online transactions that consume large amounts of CPU time, but are still considered to be 'interactive'.

What are the challenges?

- Another challenge is *identifying* those situations.
  - Analysis of SMF Type 30.4 (Step end) records can help you find large batch job steps. But the SMF 30 records won't tell you if the job is on the critical path or not. If it is not, *maybe* no one cares how long it runs?
  - For *server-type* address spaces, we recommend using the SMF30_Highest_Task_CPU_Percent field in the SMF type 30 subtype 2/3 (interval) records. This can help identify jobs/started tasks that are using a large percent of an engine in your current CPU.
  - If you have a lot of CICS regions, you might have to process a LOT of CICS SMF data to identify the large regions that also have a lot of QR contention or very CPU-intensive transactions.

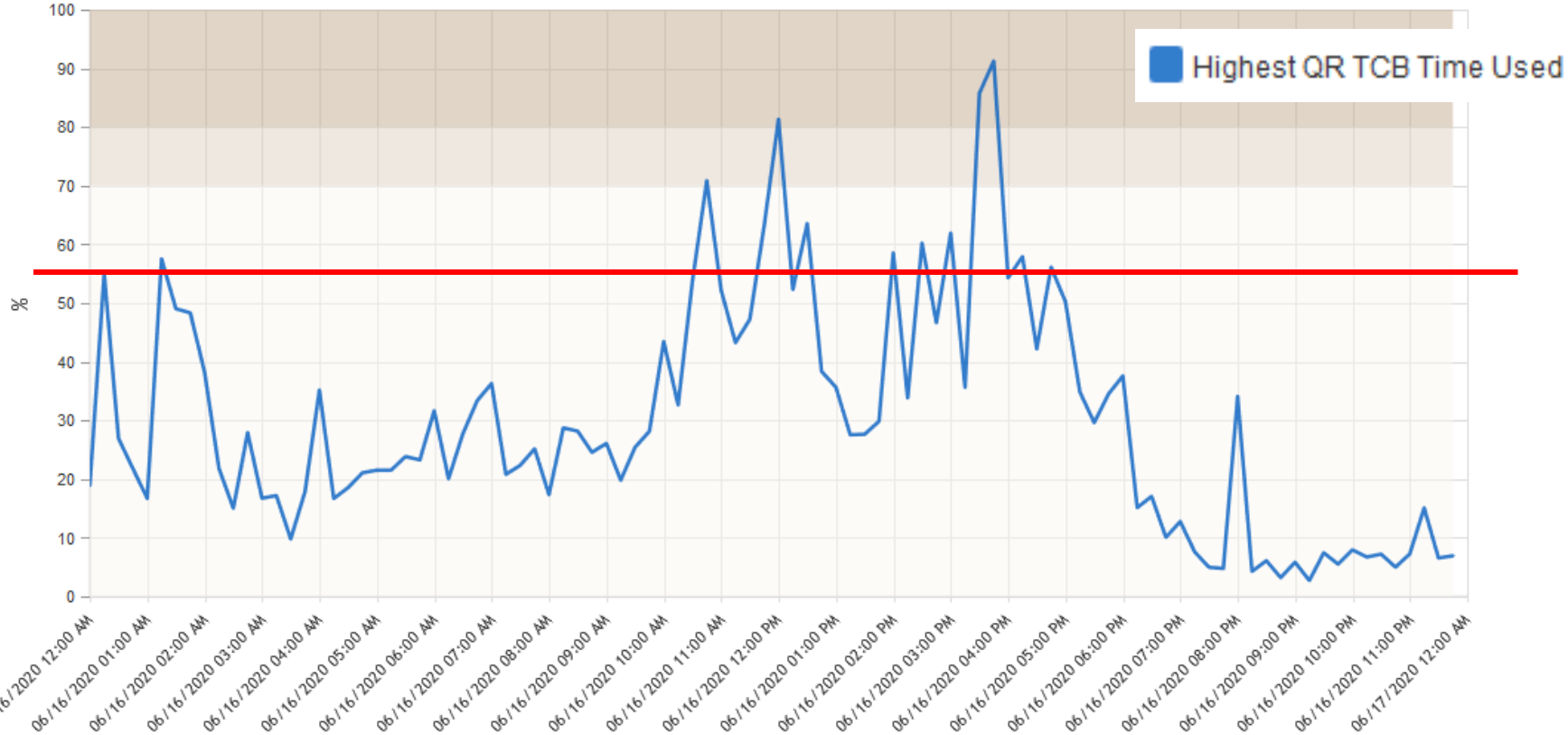
Region Health

For CICS Group ‘Group 1’ by Generic Applid

Generic Applid	Highest QR TCB Time Used	Peak active user transactions	Times at MAXTASK	SOS on any of the DSAs	SOS on any of the EDSAs	Waits for VSAM File Strings	Waits for LSR Pool Strings
Regn1	⚠	■	■	■	■	■	■
Regn2	■	■	■	■	■	■	■
Regn3	⚠	■	■	■	■	■	■
Regn4	■	■	■	■	■	■	■
Regn5	■	■	■	■	■	■	■
Regn6	■	■	■	■	■	■	■

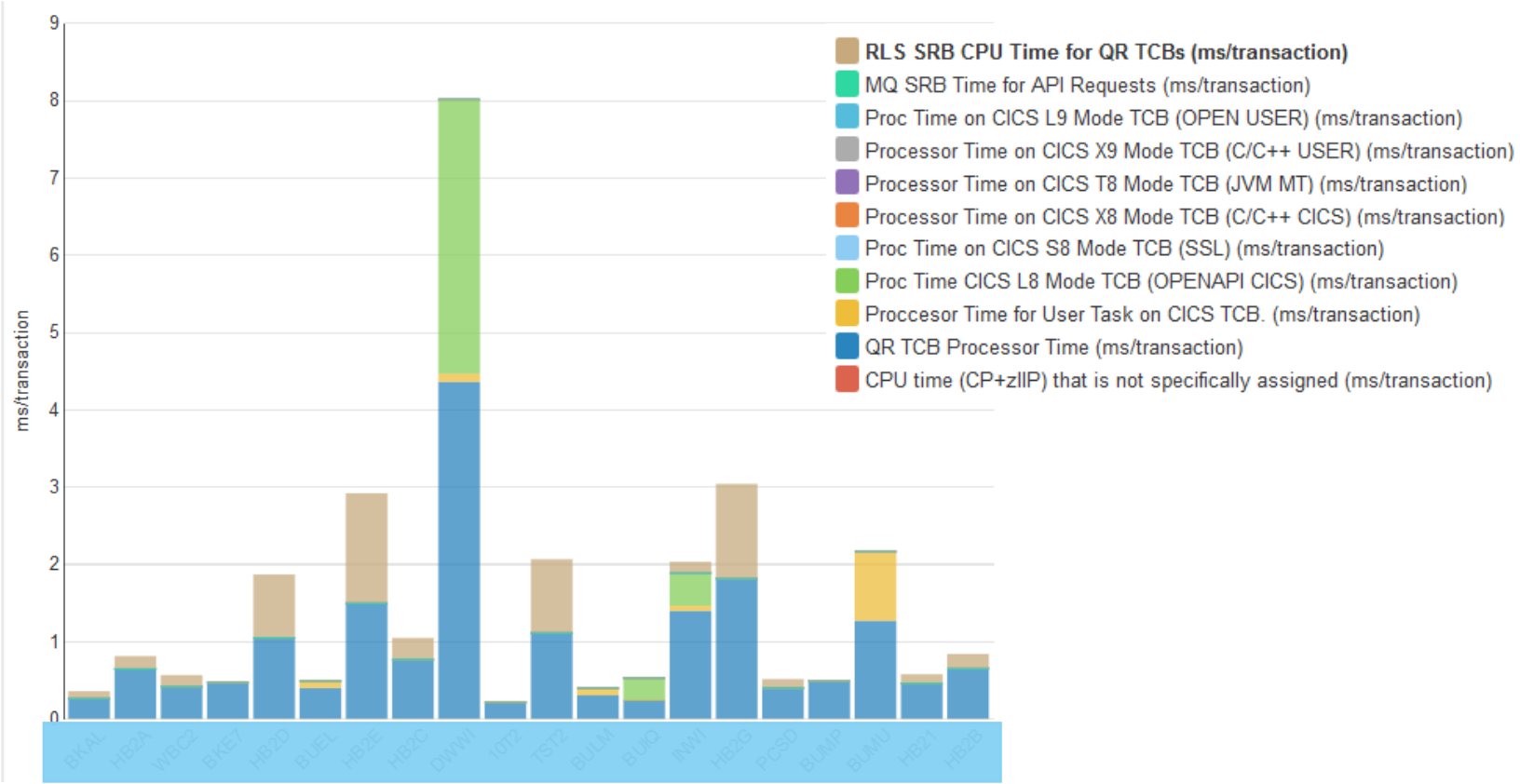
Highest QR TCB Time Used

For CICS Group ‘Group 1’, for Generic Applid ‘Regn1’



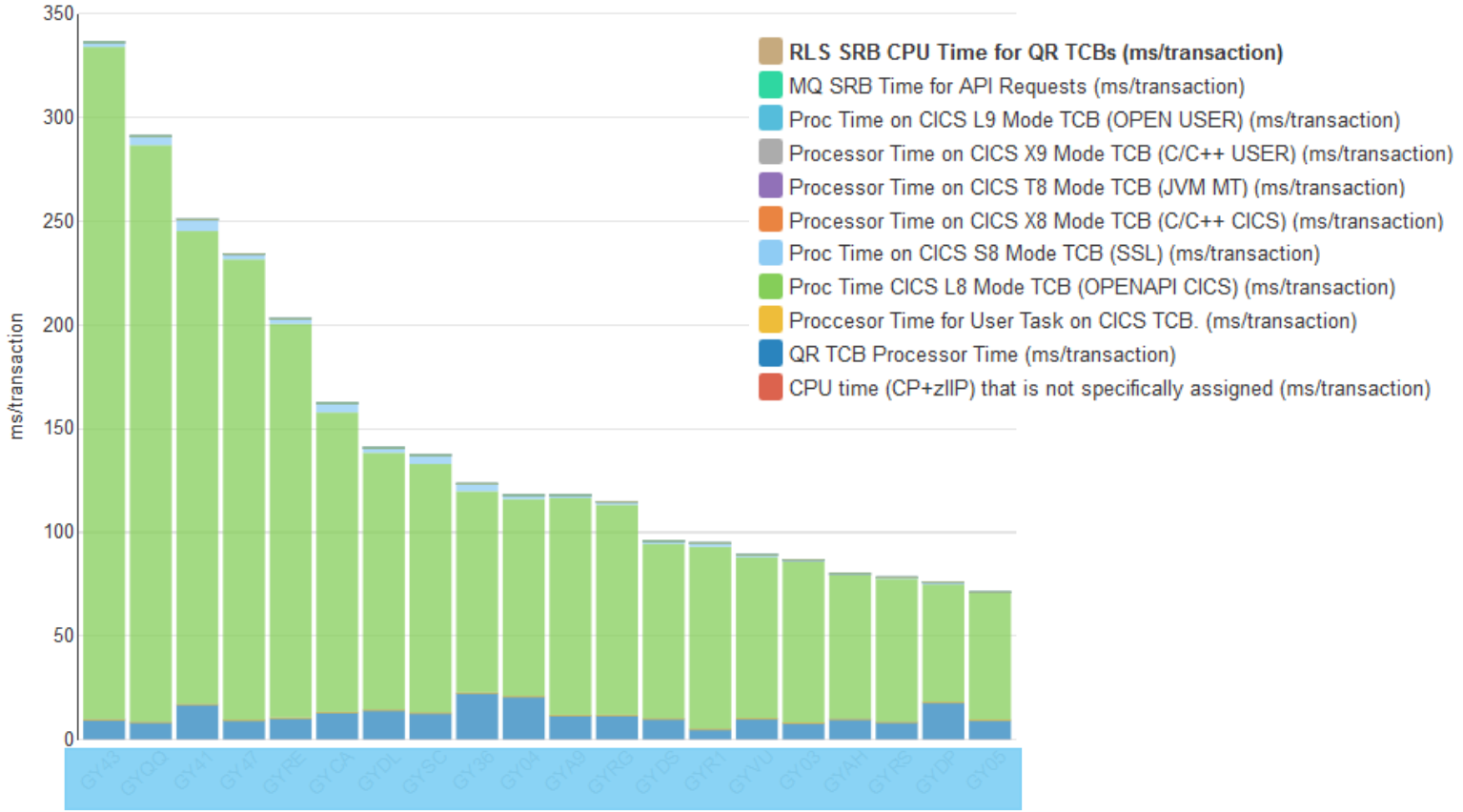
CICS CPU per Transaction

Top 20 Transactions by Transaction Rate



CICS CPU per Transaction

Top 20 Transactions by CPU per Transaction



What are the challenges?

1. Traditional upgrades, to the same speed range in a newer generation, meant that the new GCP speed was at least as fast as your current CPC, so engine speed was never a concern. This is not the case when moving to a model with slower CPs.

2. The IBM capacity planning tools, zPCR and zCP3000, are designed to address the most common upgrade scenarios - most customers upgrade to a similar model in the next generation (z13 710 to z14 710). Upgrades *between* speed ranges (e.g. 710 to 620) are less common.
   - Because a move to a sub-cap model is likely to result in more cache and potentially more VH CPs, it is possible that the workload categorization of your systems will change. This relationship is very complex, and the IBM tools are not designed to model how a dramatic change in the number and speed of cores could change the workload categorization.
   - If the IBM tools can't give you accurate predictions of the capacity of your target CPC, that makes the upgrade feel riskier.

Evaluation Criteria

- We have NOT found any *one* attribute that answers this question. Based on work with a number of customers that have moved to sub-cap models, we came up with **8** criteria that, *when viewed together*, give a good indication of how successful a move to a sub-cap CPC is likely to be:
  - Current LSPR Workload characterization.
  - Complexity of the current LPAR configuration.
  - Benefit of 'Vertical High' CPs as seen in current workloads.
  - Percent of work currently running on Vertical High CPs.
  - Will the processor cache size and design of the target CPC be a good fit for workload profile?
  - Would the lower CP speed result in large LPARs spanning drawers?
  - Do the large LPARs have a mix of long- and short-running work units?
  - Are there work units queued at times of high CPU utilization?

Evaluation Criteria

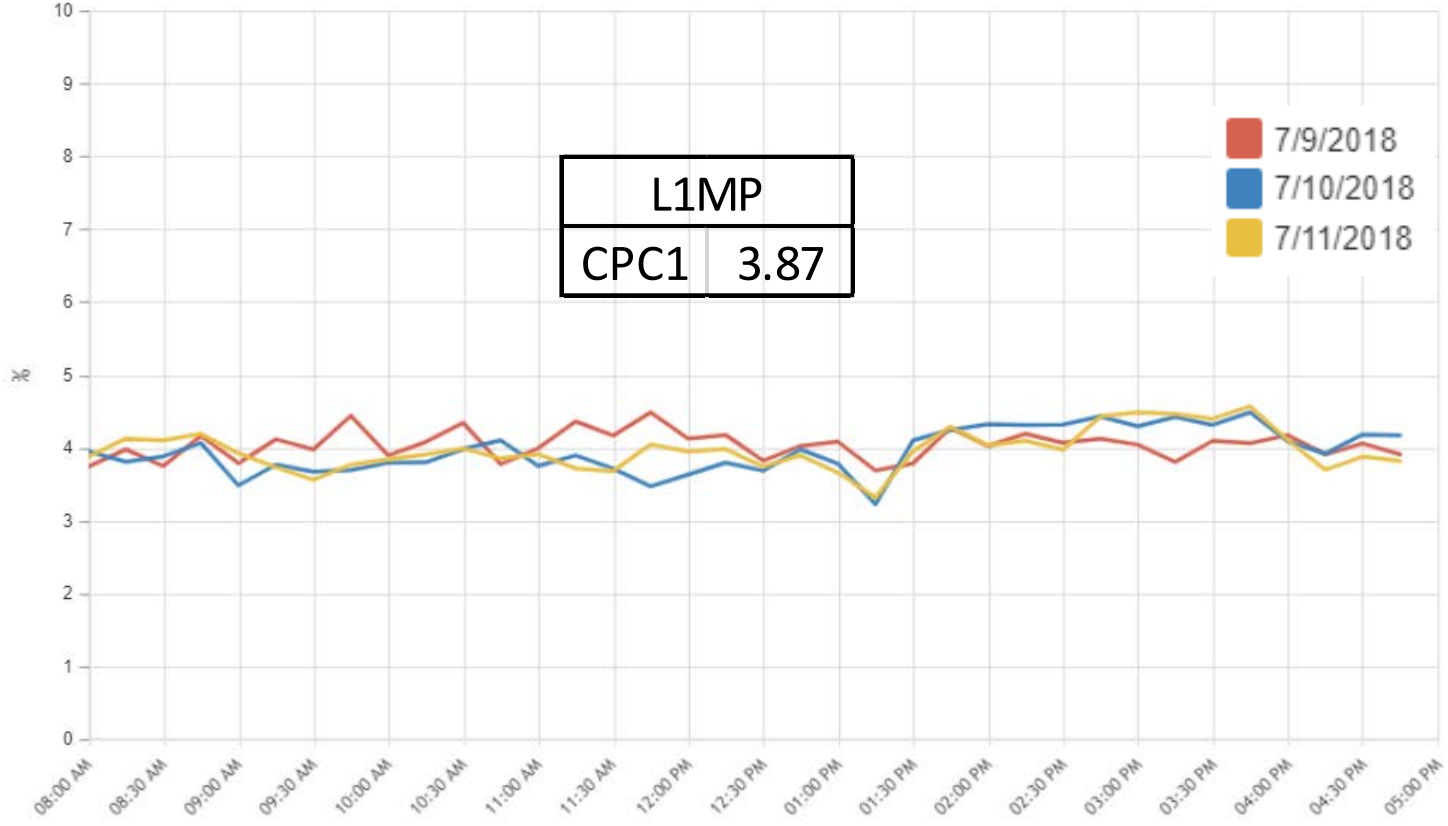
LSPR Workload Characterization

- Categorizes workloads as High, Medium or Low in terms of the demand they place on the processor cache hierarchy.

- Determined by two metrics:
  - L1MP (Level 1 Miss Percentage)
  - Numeric RNI (Relative Nest Intensity)

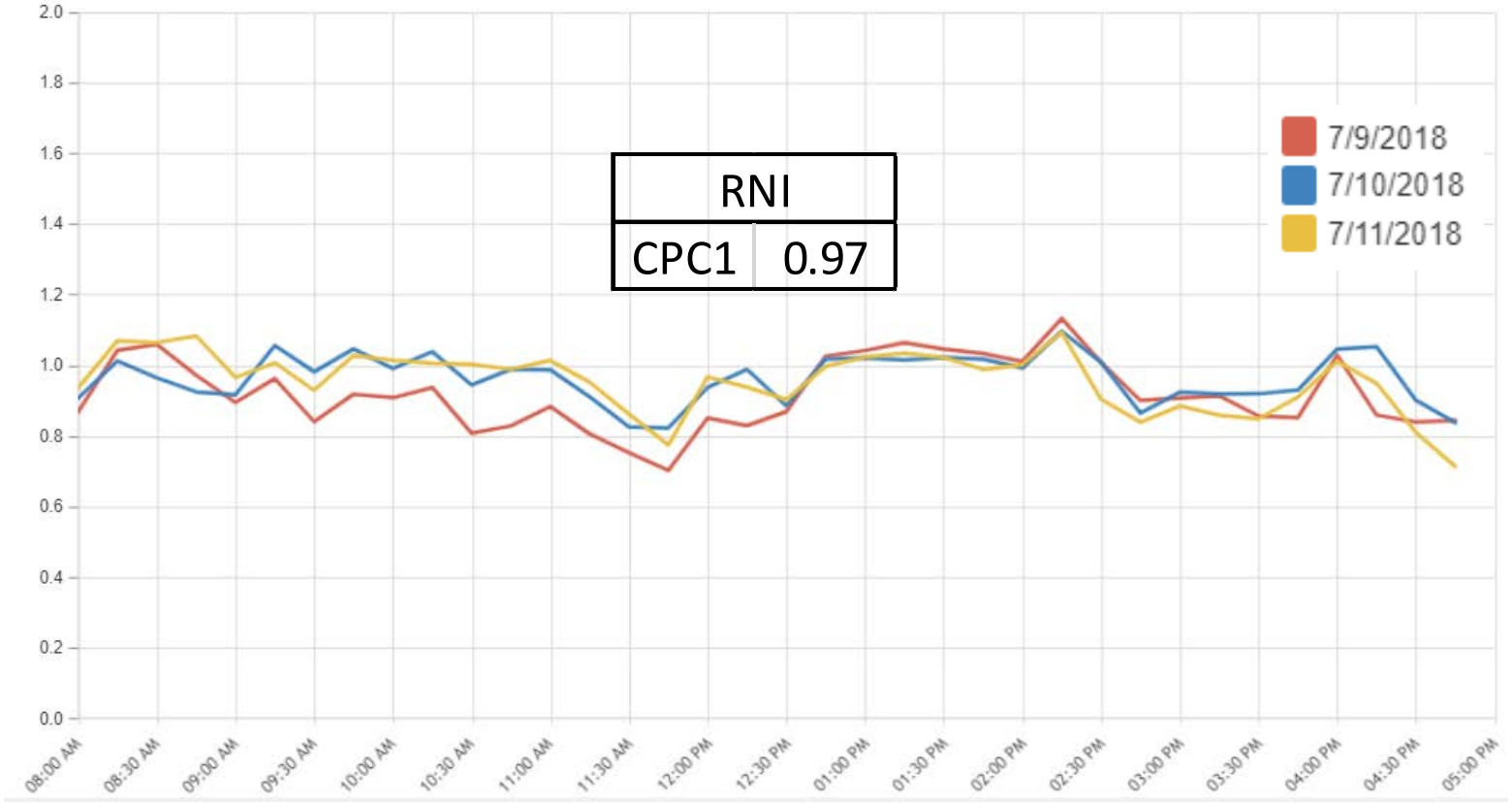
Level 1 Cache Miss Percentage

For Processor Complex Name ‘CPC1’



Relative Nest Intensity

For Processor Complex Name ‘CPC1’



IBM LSPR Workload Characterization Table

L1MP	Numeric RNI (or "Nest Depth Metric")	RNI-based LSPR Workload Characterization
< 3	>= 0.75 < 0.75	AVERAGE LOW
3 to 6	> 1.0 0.6 to 1.0 < 0.6	HIGH AVERAGE LOW
> 6	>= 0.75 < 0.75	HIGH AVERAGE

L1MP=3.87% ; Numeric RNI=0.97 (Day shift)
Workload: AVERAGE (or "AVERAGE-HIGH")

LSPR Workloads and Sub-Cap CPCs

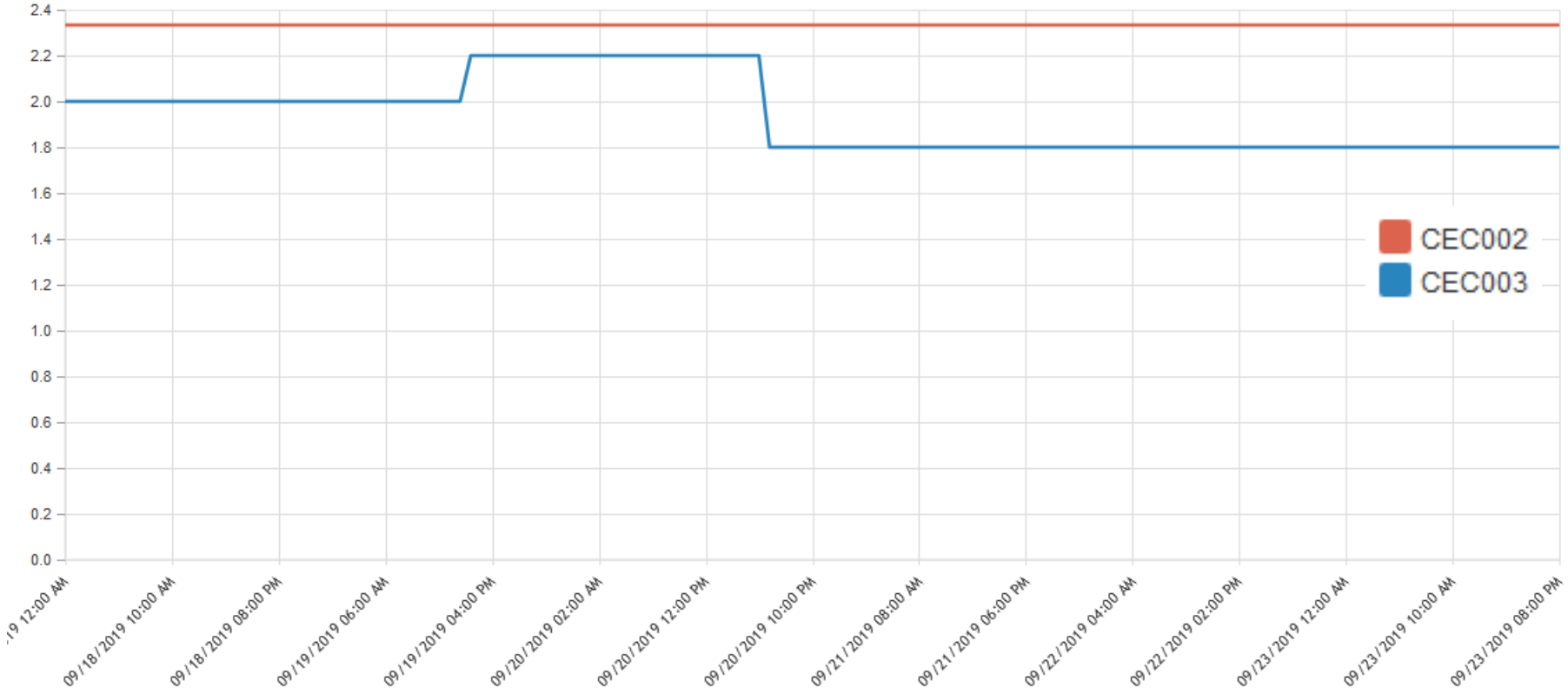
- Processor cache configuration advantages with sub-cap CPCs.
  - More CPs provide more processor cache (chip-level L1 and L2, possibly shared L3).
  - More CPs translates into additional Vertical High CPs.
- LSPR workload category indicates potential for improvement.
  - “Low” workloads are already operating efficiently with existing cache topology.
  - “Average” and “High” workloads have more potential to benefit from enhanced cache configuration of sub-cap CPCs.

Complexity of the LPAR configuration

- CPC logical to physical ratio
- Vertical CP configuration (mix of VHs, VMs, and VLs by LPAR)
- How the CPU usage by each LPAR over time compares to its guaranteed capacity (“Engine Dispatch Analysis”)
- Impact of capping (if used) on vertical CP configuration

Logical to Physical CP Ratio

By Processor Complex Name

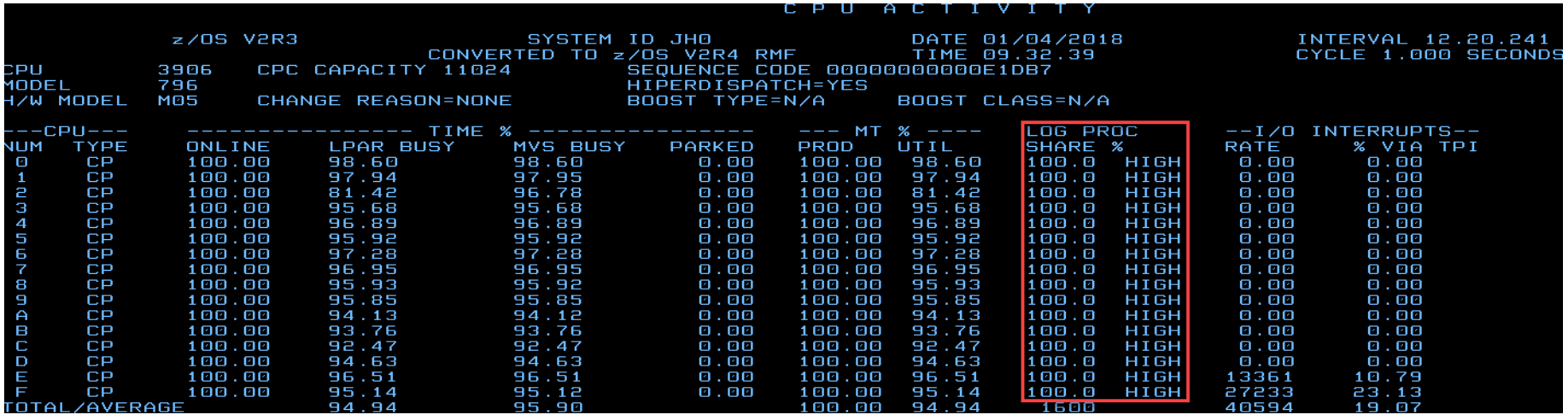


Vertical CP Configuration

- Based on LPAR weights and the number of physical CPs, PR/SM assigns logical CPs as
  - Vertical High (VH) – 1-1 relationship with physical CP
  - Vertical Medium (VM) – has at least 50% share of a CP
  - Vertical Low (VL) – has no guaranteed share, exists to use capacity from “donor” LPARs not using their share
- Work running on VHs has higher probability of cache hits
- Work running on VMs & VLs is subject to being dispatched on various CPs and contending with other LPARs

Vertical CP Configuration

- The RMF PP CPU report shows the ‘polarity’ (VL, VM, VH) of each CP:



C P U A C T I V I T Y

z/OS V2R3 SYSTEM ID JH0 DATE 01/04/2018 INTERVAL 12.20.241
CONVERTED TO z/OS V2R4 RMF TIME 09.32.39 CYCLE 1.000 SECONDS

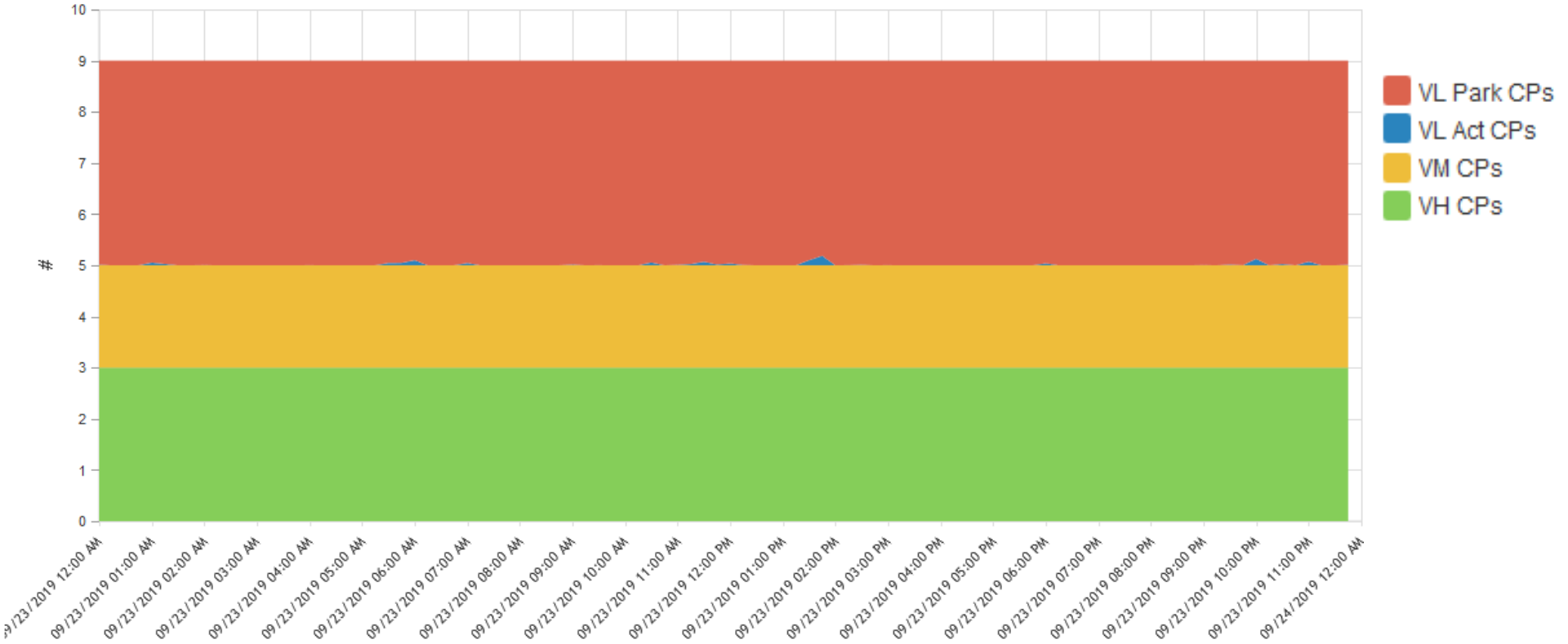
CPU 3906 CPC CAPACITY 11024 SEQUENCE CODE 00000000000E1DB7
MODEL 796 HIPERDISPATCH=YES
H/W MODEL M05 CHANGE REASON=NONE BOOST TYPE=N/A BOOST CLASS=N/A

---CPU---		TIME %				--- MT % ---		LOG PROC		--I/O INTERRUPTS--	
NUM	TYPE	ONLINE	LPAR BUSY	MVS BUSY	PARKED	PROD	UTIL	SHARE %		RATE	% VIA TPI
0	CP	100.00	98.60	98.60	0.00	100.00	98.60	100.0	HIGH	0.00	0.00
1	CP	100.00	97.94	97.95	0.00	100.00	97.94	100.0	HIGH	0.00	0.00
2	CP	100.00	81.42	96.78	0.00	100.00	81.42	100.0	HIGH	0.00	0.00
3	CP	100.00	95.68	95.68	0.00	100.00	95.68	100.0	HIGH	0.00	0.00
4	CP	100.00	96.89	96.89	0.00	100.00	96.89	100.0	HIGH	0.00	0.00
5	CP	100.00	95.92	95.92	0.00	100.00	95.92	100.0	HIGH	0.00	0.00
6	CP	100.00	97.28	97.28	0.00	100.00	97.28	100.0	HIGH	0.00	0.00
7	CP	100.00	96.95	96.95	0.00	100.00	96.95	100.0	HIGH	0.00	0.00
8	CP	100.00	95.93	95.92	0.00	100.00	95.93	100.0	HIGH	0.00	0.00
9	CP	100.00	95.85	95.85	0.00	100.00	95.85	100.0	HIGH	0.00	0.00
A	CP	100.00	94.13	94.12	0.00	100.00	94.13	100.0	HIGH	0.00	0.00
B	CP	100.00	93.76	93.76	0.00	100.00	93.76	100.0	HIGH	0.00	0.00
C	CP	100.00	92.47	92.47	0.00	100.00	92.47	100.0	HIGH	0.00	0.00
D	CP	100.00	94.63	94.63	0.00	100.00	94.63	100.0	HIGH	0.00	0.00
E	CP	100.00	96.51	96.51	0.00	100.00	96.51	100.0	HIGH	13361	10.79
F	CP	100.00	95.14	95.12	0.00	100.00	95.14	100.0	HIGH	27233	23.13
TOTAL/AVERAGE			94.94	95.90		100.00	94.94	1600		40594	19.07

- The information is contained in the type 70.1 and 99.14 SMF records.

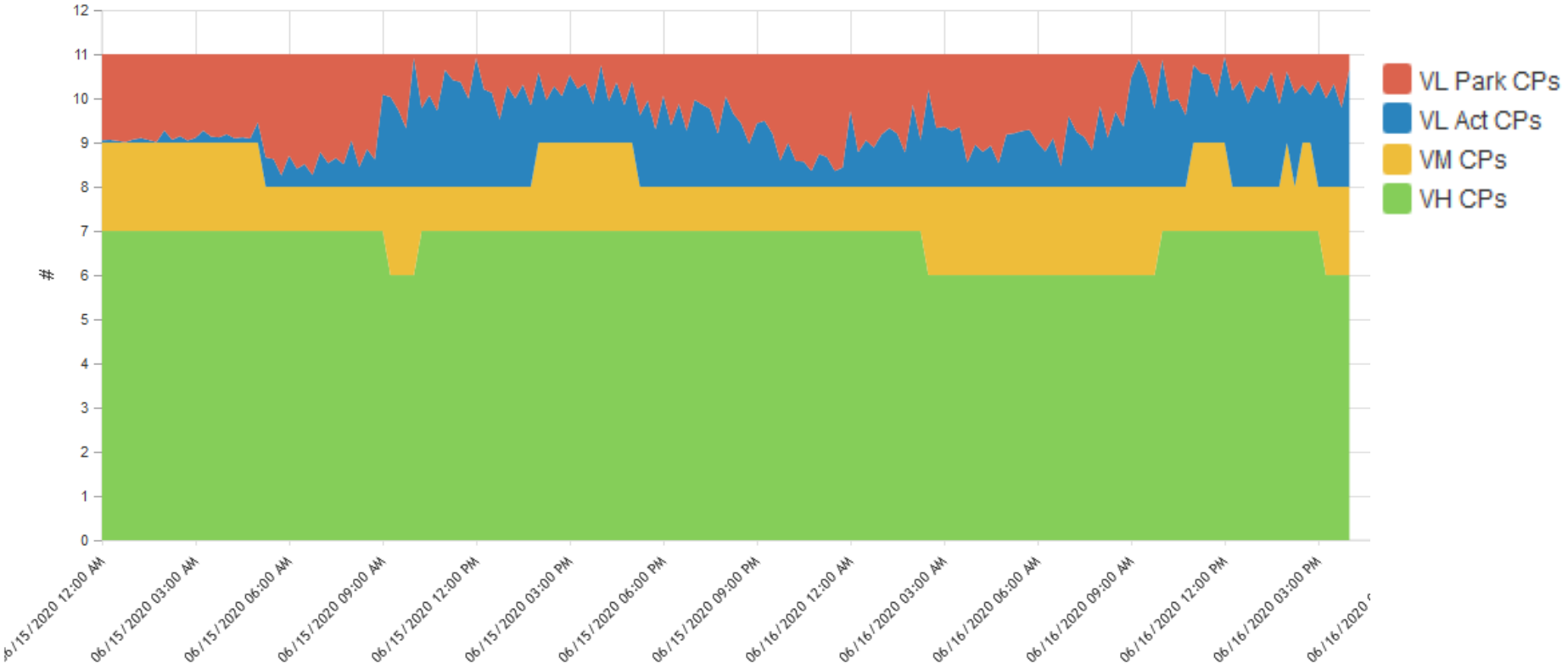
Vertical CP Configuration

For System ID ‘SYS3’



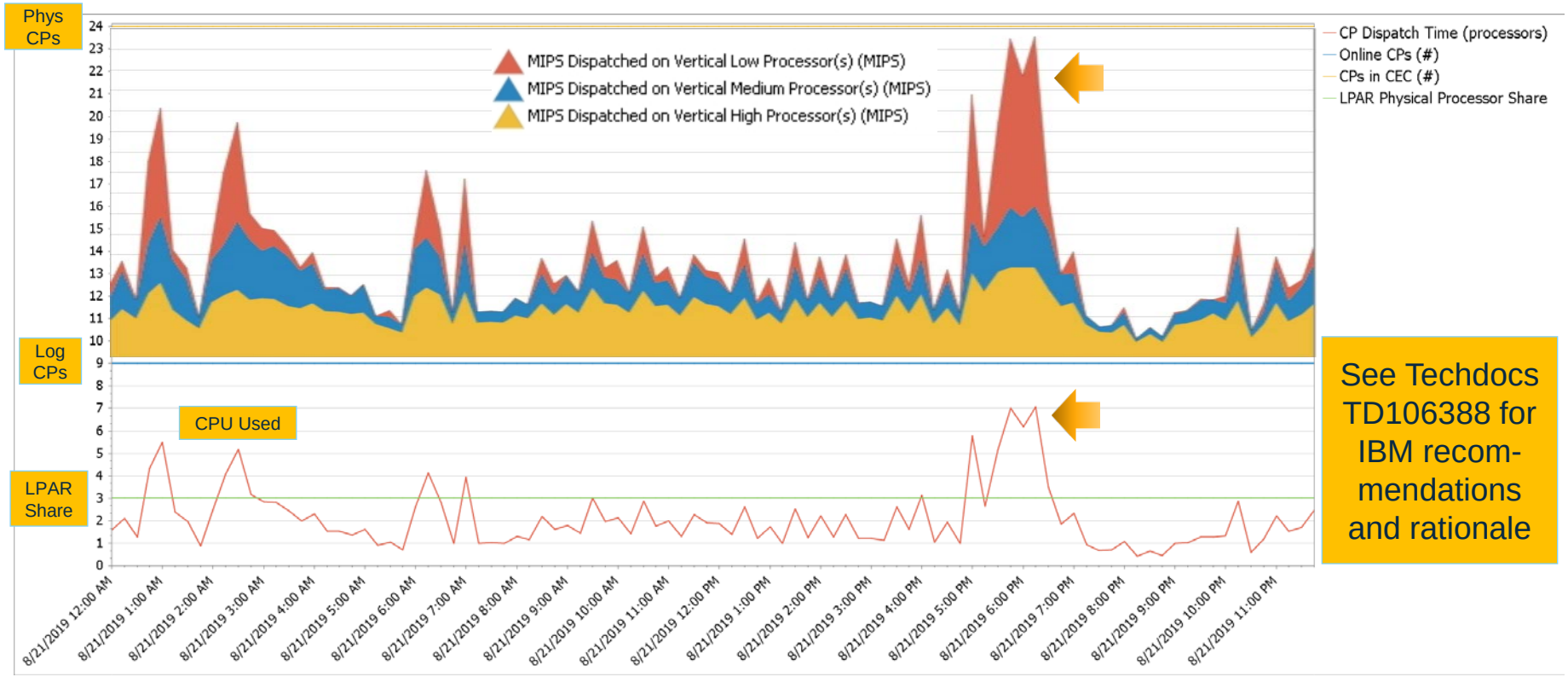
Vertical CP Configuration (IRD)

For System ID ‘SYS4’



Engine Dispatch Analysis

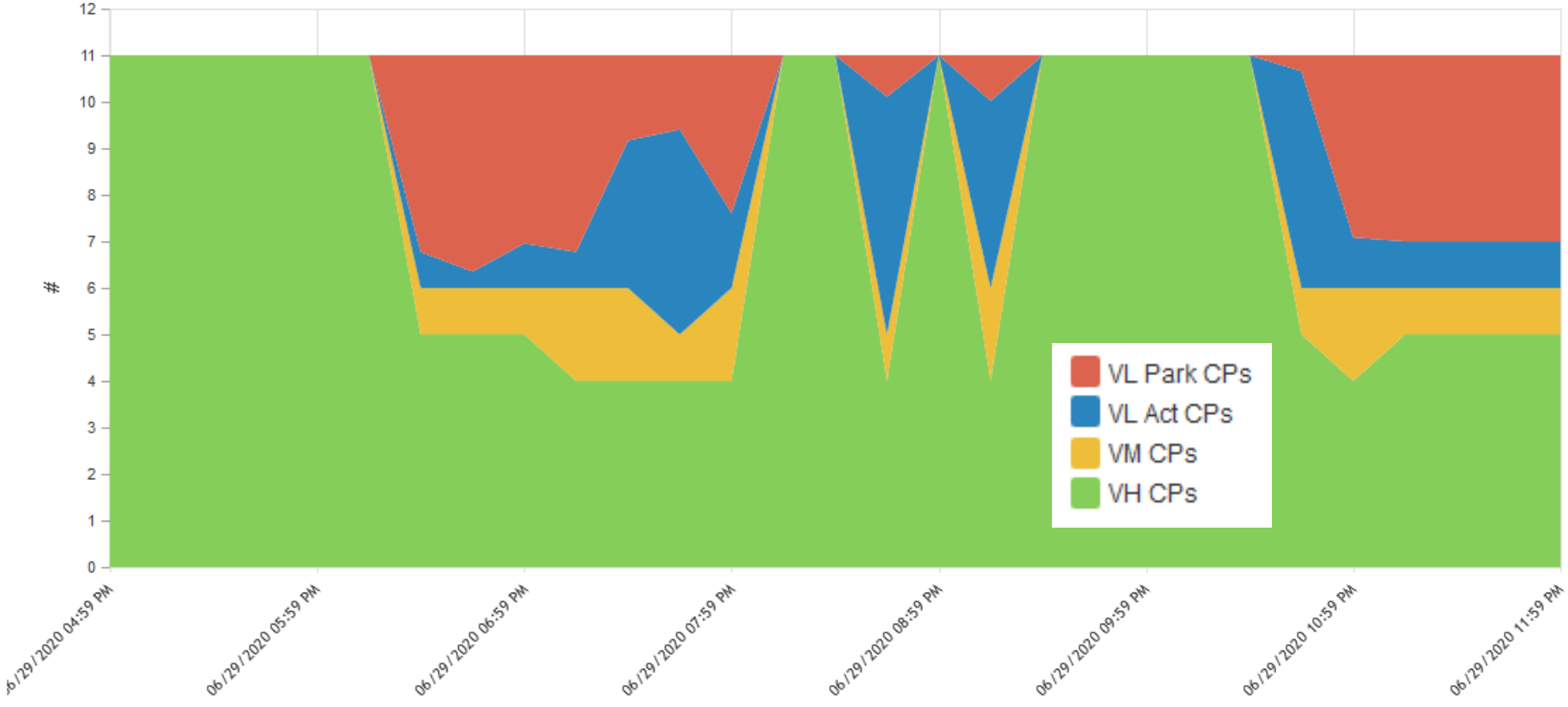
For System ID 'SYS5'



See Techdocs TD106388 for IBM recom-mendations and rationale

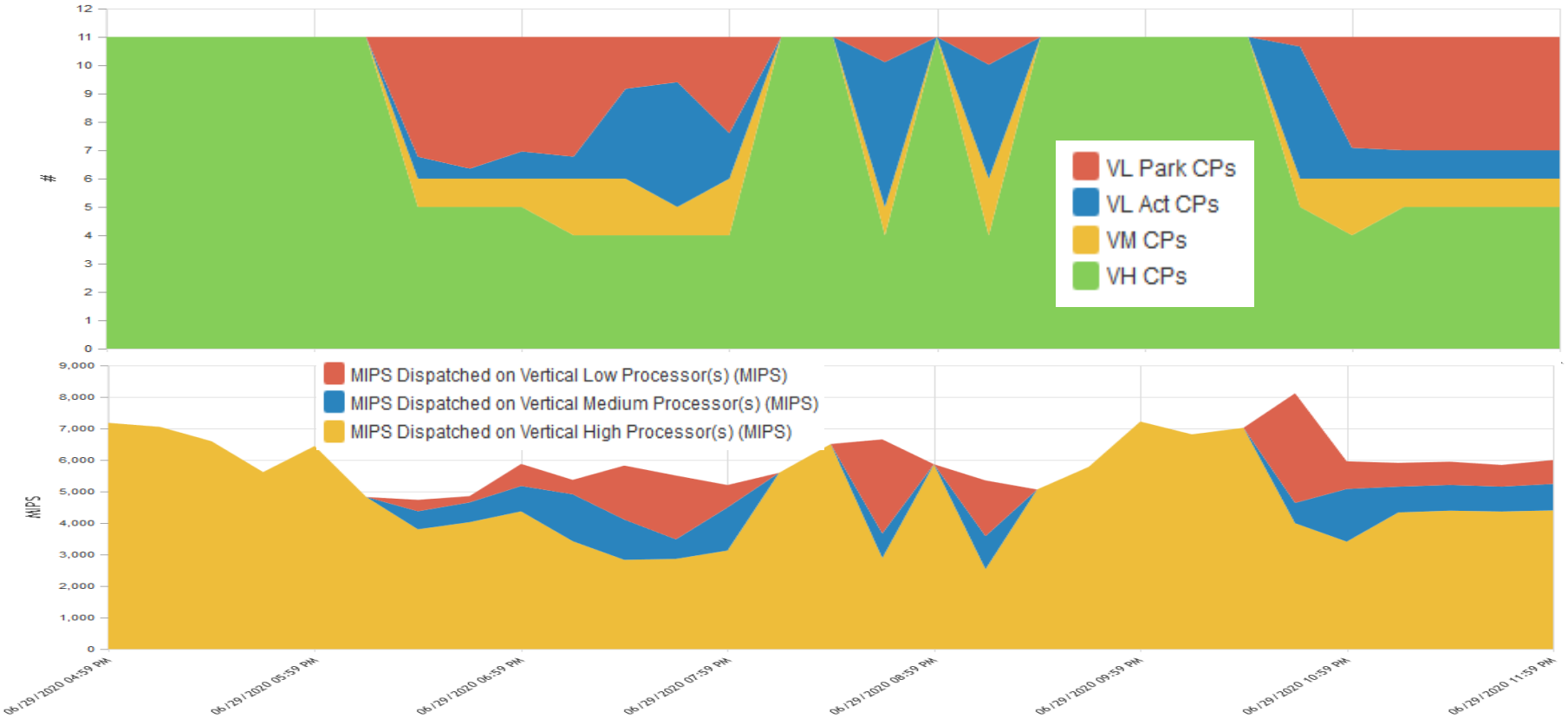
Vertical CP Configuration – Capping

For System ID ‘SYS6’



Dispatched MIPS by Vertical CP – Capping

For System ID ‘SYS6’

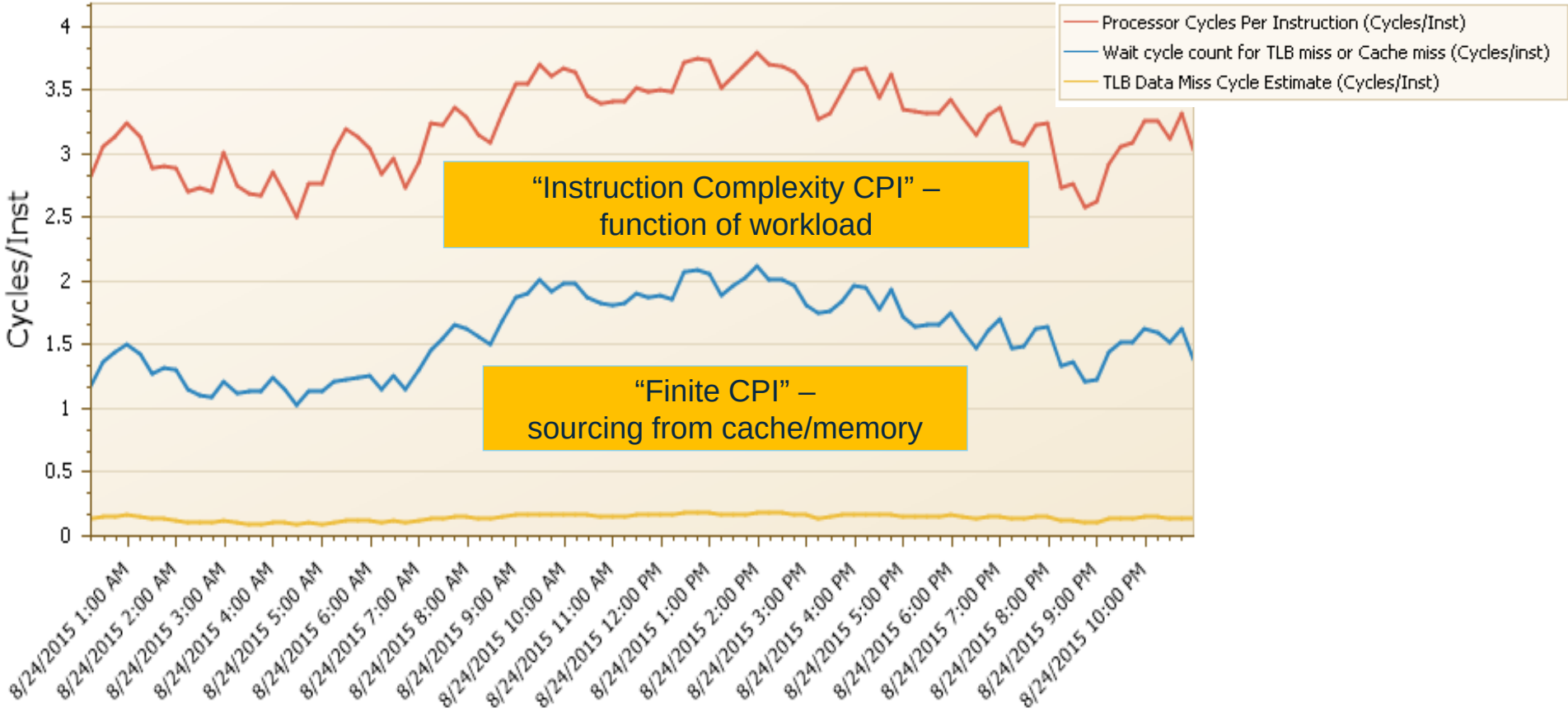


Vertical High CPs

- The larger number of CPs in a sub-cap CPC provide the *potential* to have more Vertical High CPs.

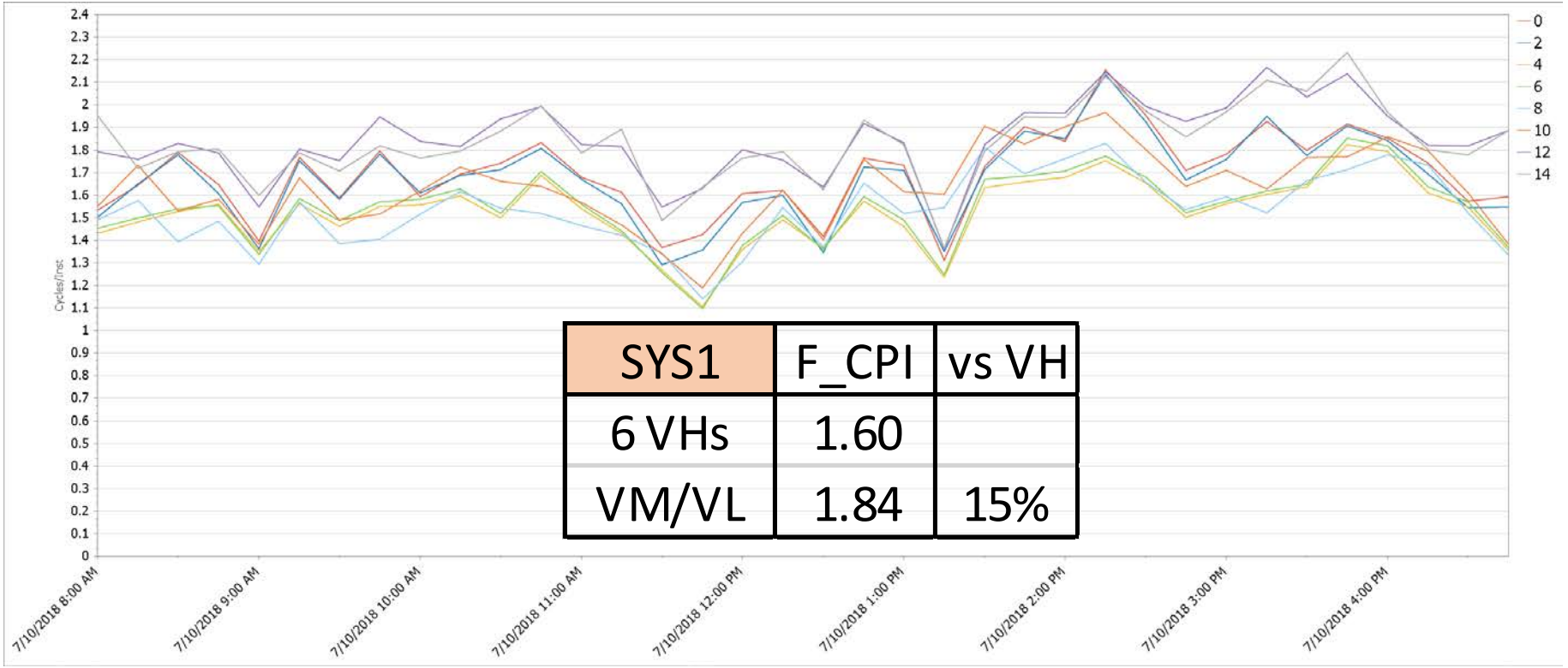
- Two considerations in evaluating potential benefit:
  - How much does your workload benefit from running on VHs?
  - How much of your work that *isn't* currently running on VHs could benefit if moved to a VH CP?

Cycles Per Instruction



Finite CPI by Logical CP

For System ID 'SYS1'

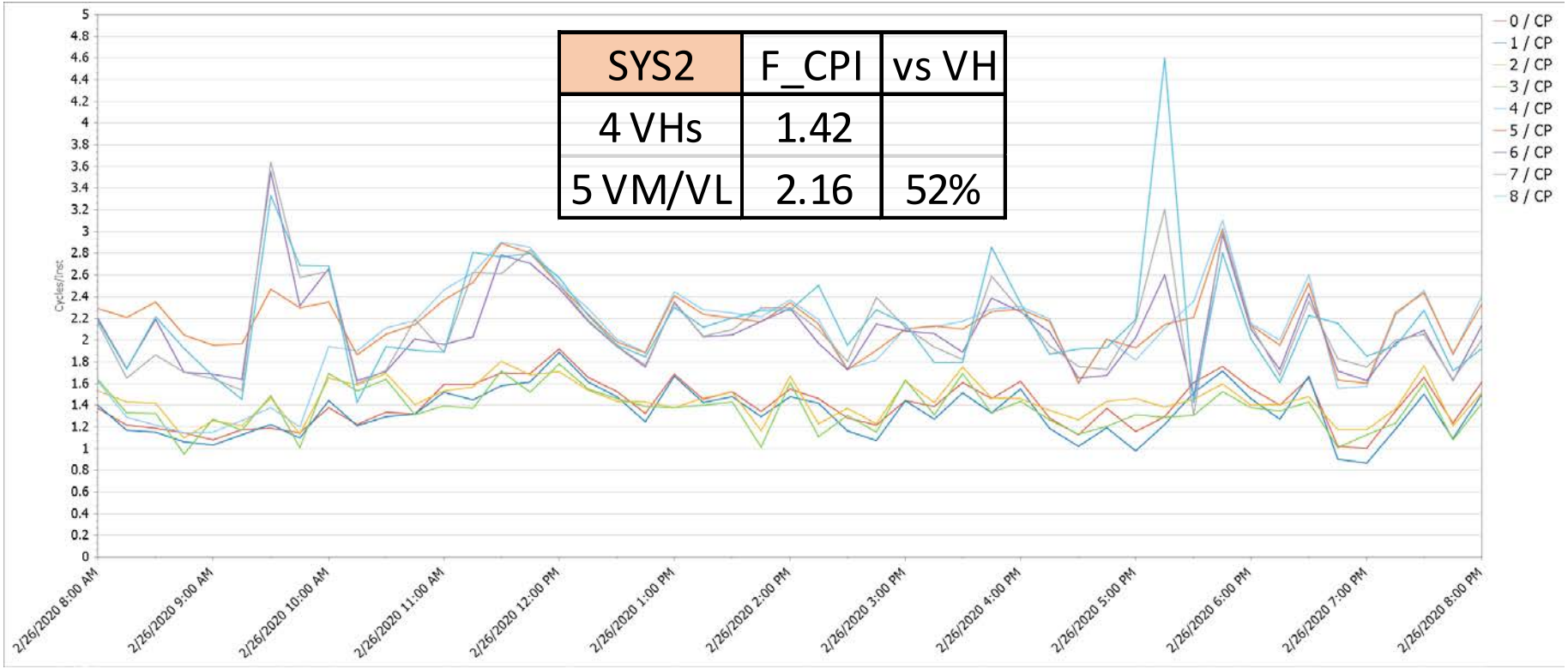


SYS1	F_CPI	vs VH
6 VHs	1.60	
VM/VL	1.84	15%

Finite CPI by Logical CP

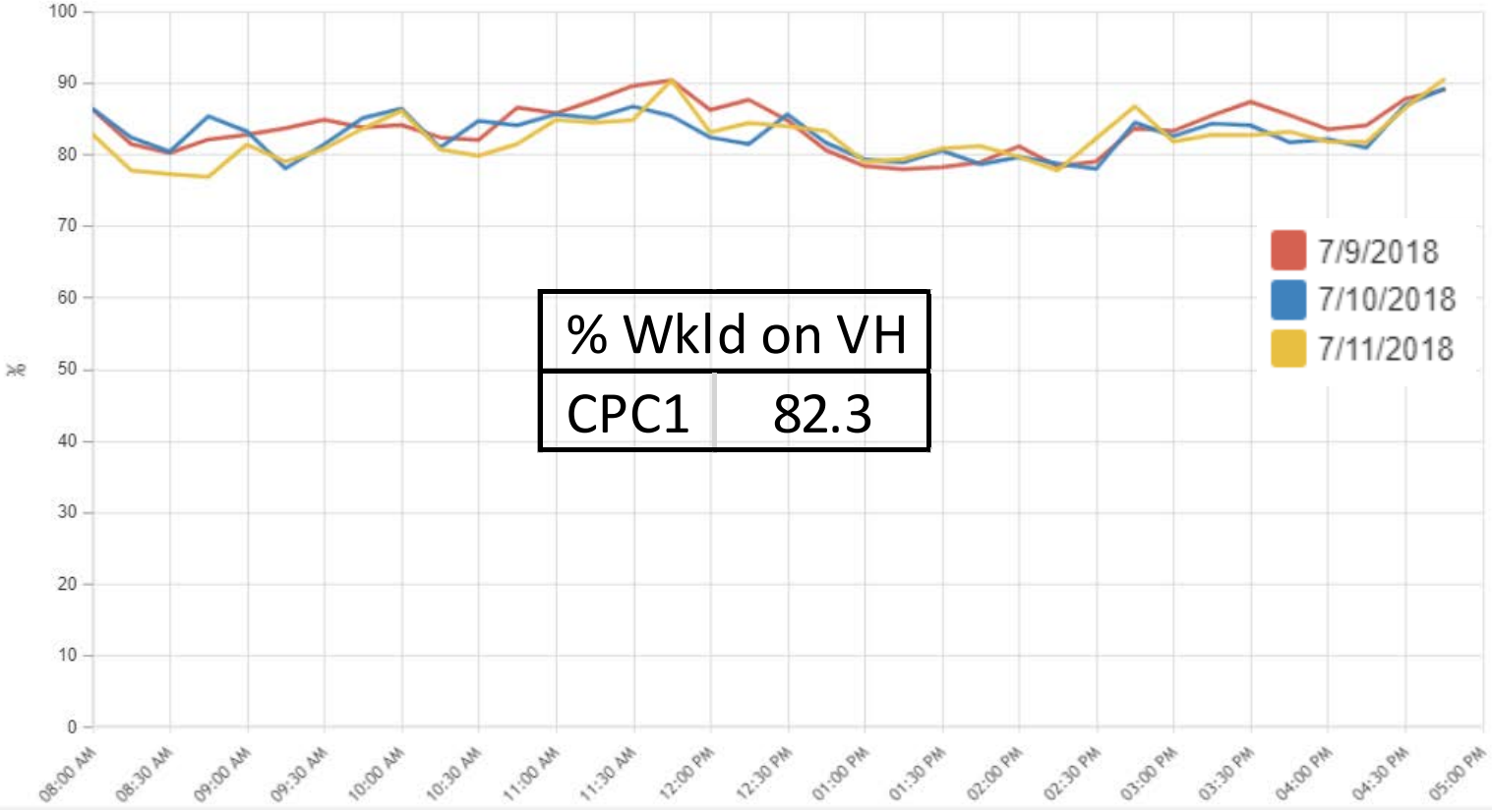
For System ID ‘SYS2’

SYS2	F_CPI	vs VH
4 VHs	1.42	
5 VM/VL	2.16	52%



% Workload Executing on Vertical High CPs

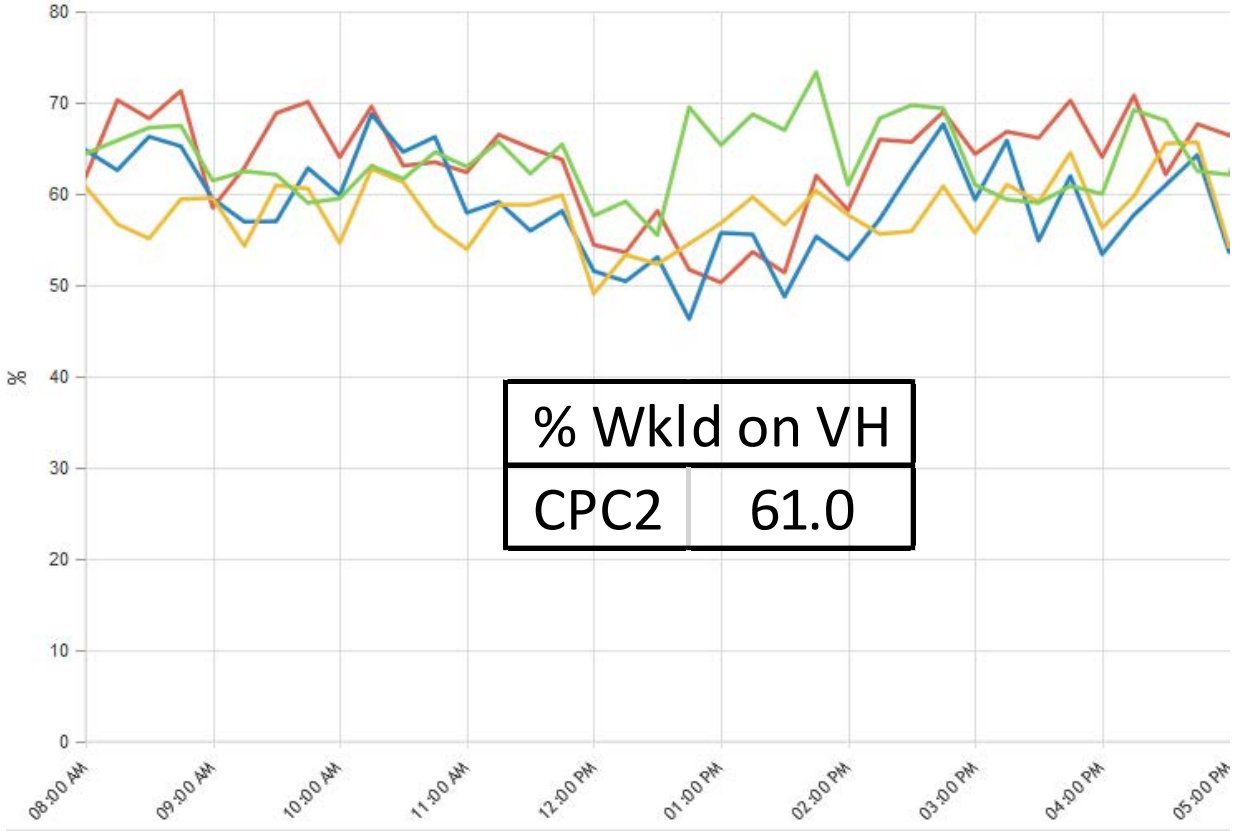
For Processor Complex Name 'CPC1'



% Wkld on VH	
CPC1	82.3

% Workload Executing on Vertical High CPs

For Processor Complex Name 'CPC2'



% Wkld on VH	
CPC2	61.0

How does new CPC Cache Design fit workload?

- IBM changes the CPC cache and memory design with every new CPC generation.
  - Sometimes the change is significant, like zEC12 to z13, and sometimes it is more evolutionary, like z13 to z14.
- Cache *sizes* change, cache design changes (moving TLB into L1 cache, for example), if you move to a sub-cap model the *number* of caches change, and the number of cores per chip (and therefore, the number of cores sharing the same L3 cache) changes.

How does new CPC Cache Design fit workload?

- You can use the information in the SMF 99.14 records (or on the HMC) to understand your current logical CP – to – physical chip mapping.
- Then adjust the number of required logical CPs based on the relative CP speeds to identify the number of logical CPs in each LPAR.
- With that information, identify the impact on your large important LPARs:
  - Will all the CPs fit in a single chip now?
  - Will the lower CP speed force an LPAR beyond the capacity of one chip/cluster/drawer?
- Use **Alain Maneville's** LPAR Design tool to help create the LPAR configuration for your target CPC.

LPAR Topology

Systems, Paging, WLM » Processor Reporting » 4HRA and Polarity » LPAR Config » Logical Processors

Processor Complex and LPAR information

For System ID A020

Add to: Collected | Dashboard | Favorites

Custom 1/16/2020 10:00 AM - 1/16/2020 11:00 AM
Interestgroup **IGT, All Sysplexes, All shifts**
Reporting interval **Measurement**

Processors, LPARs and CECs with Hardware data

CEC Serial	CEC Name	Processor Type	Processor Architecture	Processor Speed (Cycles/microsec)	Relative Nest Intensity	Level 1 Cache Miss (%)	IBM RNI Workload Category	Estimated TLB1 CPU Miss % of Total CPU (%)
IBM-AF27	CEC-A	CP	z14	5,208.00	0.926	3.253	Average	2.045
IBM-AF27	CEC-A	zIIP	z14	5,208.00	2.207	2.542	Average	4.018

Logical Processors assigned to LPAR

System ID	Processor ID	Logical Processor/Core ID in LPAR	Processor Type	Polarization	Core Capacity	Chip Id	Node/Book Id	Drawer Id	Logical Processor Flags
A020	0000	0000	CP	Vertical High	2,000,000	3	1	2	0
A020	0002	0001	CP	Vertical High	2,000,000	3	1	2	0
A020	0004	0002	CP	Vertical Medium	1,703,125	3	1	2	0
A020	0006	0003	zIIP	Vertical Medium		1	2	3	0
A020	0007	0003	zIIP	Horizontal or Dedicated	0				0

Does the LPAR have a mix of long- and short-running work units?

- CPU hogs *will* run for longer on a slower CPC.
- However, if you have a mix of short- and long-running work units, the smaller work units can use the additional CPs to run in parallel with the CPU hogs.
- We list all job steps by CPU time and agree a ‘concern threshold’.
  - Owners of job steps that take *less* than the threshold number of seconds probably will not notice the increased CPU time.
    - It is not unusual to find that less than 1% of job steps use more than 1 *second* of CPU time.
  - Job steps *over* the concern threshold are investigated to determine if they would be an issue.

Mix of long- and short-running work units on LPAR?

- For server address spaces, can they service their load with a slower CP?
  - For these, we use the SMF30_Highest_Task_CPU_Percent field to identify programs that use a large portion of a CP during any SMF interval.

Steps with Highest Pct CPU Used

by Address Space Name with filter on Address Space Name

Add to: Collected | Dashboard | Favorites | Edit report

Single day 6/29/2020
Interestgroup **IGT**, **All Sysplexes**, **All shifts**
Reporting interval **Measurement**

Address Space Name	z/OS Program Name	Stepname from EXEC card	Total CP usage (s)	zIIP Time (s)	Execution Time (s)	Step Start Date and Time	Step End Time	Service Class	Highest Percentage of CPU Use for any Task in Address Space (%)
JDC1294	IKJEFT01	DB2IKJE	2,791.919	0.000	2,850.00	6/29/2020 2:10:17 AM	6/29/2020 2:57:47 AM	SLMPROD	98
JTH1ADTH	DFSERA10	STEP1	5,317.158	0.000	5,514.00	6/29/2020 1:00:29 AM	6/29/2020 2:32:23 AM	SLMPROD	97
JMI1615B	BPXPRECP	*OMVSEX	973.319	0.000	1,176.00	6/29/2020 10:41:22 PM	6/29/2020 11:00:58 PM	OMVS	85
JYG1AREG	BPXPRECP	*OMVSEX	1,627.478	0.000	4,138.00	6/29/2020 7:20:26 PM	6/29/2020 7:54:55 PM	OMVS	77
JGY822PE	IKJEFT01	GYRKTAA	926.379	0.000	1,016.00	6/29/2020 11:01:57 PM	6/29/2020 11:18:53 PM	SLMPROD	72

Are there work units queued at times of high CPU utilization?

- The last criteria that we look at is the level of queueing when the CPU utilization is at its peak – that is, if more CPs were available, could they be utilized?
- This is related to the job mix, but provides a more comprehensive view over a longer timeframe.
- CPU Queue data comes from SMF Type 70 records.

Summary

Summary

- **There is no magic bullet** indicator of whether a sub-cap CPC would work well for you, or how much benefit it would provide compared to your ‘normal’ upgrade path.
  - Evaluating an upgrade to a CPC with slower speed general purpose CPs is definitely more work than just moving to the newer version of whatever you have today.

- However, based on the improved performance and reduced overall cost to deliver the service that some customers have observed, and the amount of money that you will be spending on the upgrade and associated software bills, investing the time to evaluate all your options *would* seem to be a wise move.

Summary

- Knowing what to look for, and having powerful tools to help you extract the required information, make such an evaluation possible.
- IntelliMagic Vision makes nearly all the information you need easily accessible.
- If you would like assistance with this, or would simply like to have an independent group analyze all your upgrade options, please contact us at technical@watsonwalker.com.

Upcoming zAcademy Session

IntelliMagic zAcademy Upcoming Session(s)

All sessions start at 9am CT | 10am ET | 4pm CEST

Previous sessions and recordings can be found below >

September

15

Your datacenter under TFP: the new rules of measurement

Cheryl Watson of Watson & Walker and John Baker of IntelliMagic

Tailored Fit Pricing is no longer an abstract concept as many datacenters have adopted this new pricing model. For those that held back, recent enhancements from IBM have made TFP potentially more attractive. While TFP has the potential to greatly simplify software costs, the Capacity Planner/Performance Analysts (CP/PA) needs to adjust many of their tried and true methods for tracking – and controlling – MSU consumption. Are you paying too much? Have your measuring tools and techniques adapted to these new rules? Join Cheryl Watson and John Baker as they dive into this critical subject. They've been known to have an opinion or two.

Register

Resources

- Todd Havekost, Lessons on Optimizing Processor Cache From z15 Upgrade Case Studies, IntelliMagic zAcademy Session #5 video, 5/12/2020
- Todd Havekost, Impact of z14 on Processor Cache and MLC Expenses
- Frank Kyne, "Customer Sub-capacity CPC Experience", Cheryl Watson's Tuning Letter 2018 #3, pp. 57-75
- Frank Kyne, "A Holistic Approach to Capacity Planning", Cheryl Watson's Tuning Letter 2015 #4, pp. 55-75
- Alain Maneville, LPAR Design Tool
- Cheryl Watson, CPU Charts (for every CPC from z900 to z15 T02)
- IBM, zBNA Tool for modeling impact of changing CP speeds on batch jobs

Thank you!