EDUCATE ‣ NETWORK ‣ INFLUENCE

z/OS Performance Spotlight – Some Top Things You May Not Know

aka Peter and Scott's Tips and Tidbits

z/OS Performance Education, Software, and Managed Service Providers

Creators of Pivotor®

Peter Enrico & Scott Chapman

Enterprise Performance Strategies, Inc.

performance.questions@epstrategies.com

Copyright© by SHARE Association Except where otherwise noted, this work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 license. http://creativecommons.org/licenses/by-nc-nd/3.0/

Abstract

- During this session, Peter Enrico and Scott Chapman will discuss a variety of z/OS performance measurement, analysis, and tuning techniques that may not be commonly known or are not often discussed.

- The key objective of this presentation is to provide the attendee with information they can bring back to their shop and conduct some analysis or tuning exercises. A secondary objective of this session is to help the attendee learn more about the z/OS environment, and how things work. This session is sure to be highly educational!

Contact, Copyright, and Trademarks

Questions?

Send email to performance.questions@EPStrategies.com, or visit our website at https://www.epstrategies.com or http://www.pivotor.com.

Copyright Notice:

© Enterprise Performance Strategies, Inc. All rights reserved. No part of this material may be reproduced, distributed, stored in a retrieval system, transmitted, displayed, published or broadcast in any form or by any means, electronic, mechanical, photocopy, recording, or otherwise, without the prior written permission of Enterprise Performance Strategies. To obtain written permission please contact Enterprise Performance Strategies, Inc. Contact information can be obtained by visiting http://www.epstrategies.com.

Trademarks:

Enterprise Performance Strategies, Inc. presentation materials contain trademarks and registered trademarks of several companies.

The following are trademarks of Enterprise Performance Strategies, Inc.: **Health Check®, Reductions®, Pivotor®**

The following are trademarks of the International Business Machines Corporation in the United States and/or other countries: IBM®, z/OS®, zSeries®, WebSphere®, CICS®, DB2®, S390®, WebSphere Application Server®, and many others.

Other trademarks and registered trademarks may exist in this presentation

Today’s Agenda

- Who we are / what we do (Peter)
- Emerging Areas of Interest
  - Z16 Migrations (Scott)
  - SDC Coefficients and Reevaluating Durations for z/OS 2.5 (Peter)
  - CPENABLE and z/OS 3.1 (Peter)
  - Implicit CPU Protection in z/OS 3.1 (Peter)
  - First Reference Page Faults (Peter)
  - IXGCNFxx Keep local buffers (Peter)
  - Large memory should mean less I/O? (Scott)
  - Scott’s current AI thoughts
- Short Reminders from Continuing Questions and Opportunities
  - XCF transport class simplification (Peter)
  - SRB Update and SMF 30 data (Scott)
  - SuperPAV (Scott)
  - I/O Priority Management (Scott)
  - Record the 98s and 99s (Scott)
  - SMT (Scott)
- Prize drawings! (Jamie)

EPS: We do z/OS performance…

- We are z/OS performance!

- Pivotor
  - Performance reporting and analysis of your z/OS measurements
  - Example: SMF, DCOLLECT, other, etc.
  - Not just reporting, but cost-effective analysis-based reporting based on our expertise

- Performance Educational Workshops (while analyzing your own data)
  - Essential z/OS Performance Tuning
  - Parallel Sysplex and z/OS Performance Tuning
  - WLM Performance and Re-evaluating Goals

- Performance War Rooms
  - Concentrated, highly productive group discussions and analysis

- MSU reductions
  - Application and MSU reduction

z/OS Performance workshops available

During these workshops you will be analyzing your own data!

- WLM Performance and Re-evaluating Goals
  - February 19-23, 2024
- Parallel Sysplex and z/OS Performance Tuning
  - August 20-21, 2024
- Essential z/OS Performance Tuning
  - September 16-20, 2024
- Also… please make sure you are signed up for our free monthly z/OS educational webinars! (email contact@epstrategies.com)

EPS presentations this week

What	Who	When	Where
CPU Critical: A modern revisit of a classic WLM option	Peter Enrico Scott Chapman	Mon 4:00	Salon 12
30th Anniversary of Parallel Sysplex: A Retrospective and Lessons Learned	Peter Enrico	Tue 10:30	Salon 21
z/OS Performance Spotlight: Some Top Things You May Not Know	Peter Enrico Scott Chapman	Tue 1:00	Salon 15
The Highs and Lows: How Does HiperDispatch Really Impact CPU Efficiency?	Scott Chapman	Thu 10:30	Salon 21
Configuring LPARs to Optimize Performance	Scott Chapman	Thu 2:30	Salon 21

Like what you hear today?

- Free z/OS Performance Educational webinars!
  - Have been on hiatus for a couple of months but should be coming back soon
  - Let us know if you want to be on our mailing list for these webinars

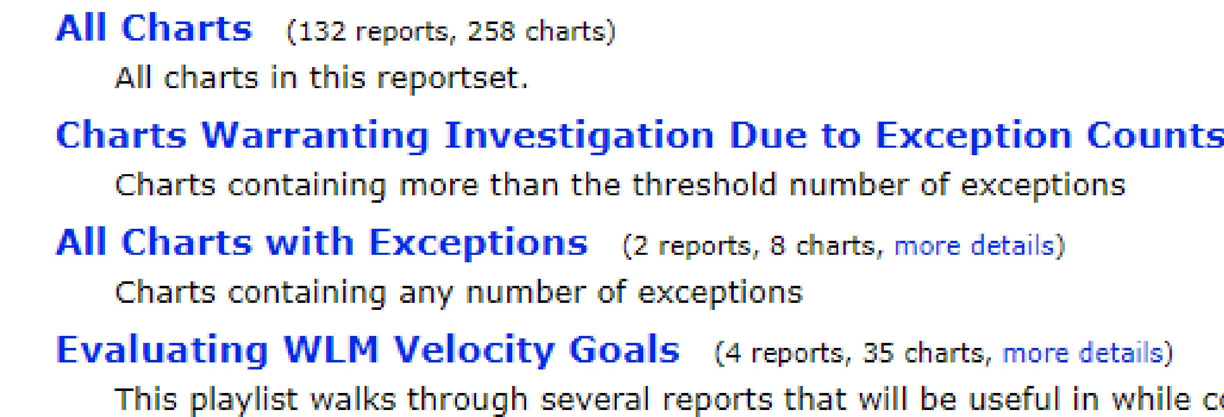
- If you want a free cursory review of your environment, let us know!
  - We're always happy to process a day's worth of data and show you the results
  - See also: http://pivotor.com/cursoryReview.html

Like what you see?

- The z/OS Performance Graphs you see here come from Pivotor™

- If you just a free cursory review of your environment, let us know!
  - We're always happy to process a day's worth of data and show you the results
  - See also: http://pivotor.com/cursoryReview.html

- We also have a free Pivotor offering available as well
  - 1 System, SMF 70-72 only, 7 Day retention
  - That still encompasses over 100 reports!



All Charts (132 reports, 258 charts)

All charts in this reportset.

Charts Warranting Investigation Due to Exception Counts

Charts containing more than the threshold number of exceptions

All Charts with Exceptions (2 reports, 8 charts, more details)

Charts containing any number of exceptions

Evaluating WLM Velocity Goals (4 reports, 35 charts, more details)

This playlist walks through several reports that will be useful in while c

Pivotor – Intelligent Reporting

- Pivotor is our data reporting tool & service designed specifically for z/OS performance reporting
  - Designed and used by z/OS performance experts
  - Processes data from SMF, DCOLLECT, and customer sources
  - Contains hundreds of z/OS performance reports “out of the box”
  - Designed to be easy to use and manage
  - Reports are organized into logical and searchable report sets
  - Features include intelligent exceptions, drill down, search, canned analysis, and so much more
  - Built in expanded helps to help foster education

Comprehensive Report Sets for Immediate Performance Analysis

- Processor Analysis
- MSU, MLC, Usage, Multiplex Analysis
- Storage / Paging Analysis
- Sysplex and Data Sharing Analysis
- Coupling Facility Analysis
- USS Analysis
- IBM MQ Interval
- Environmental Summary Reports

- Workload Manager (WLM) Analysis
- Communication Server TCP/IP, FTP, etc. Analysis
- System Logger Analysis
- DCOLLECT Analysis
- Application Analysis
- Custom Reports (e.g. Mgt Rqmts)
- Customer Application Data
- Batch Analysis

- DASD I/O Subsystem Analysis
- VTS and TMC Analysis*
- Workload I/O Analysis
- DFHSM Analysis
- VSAM and VSAM RLS
- Transaction and Workload Analysis
- GDPS / Global Mirror Analysis
- Other SMF
- Trend / Stats Long term Analysis

- DB2
- IBM MQ
- CICS
- IMS
- WAS WebSphere AS
- IDMS
- File-level I/O
- Root Cause / Performance Debug Analysis
- WLM Algorithm Analysis

>2000 reports “out of the box”

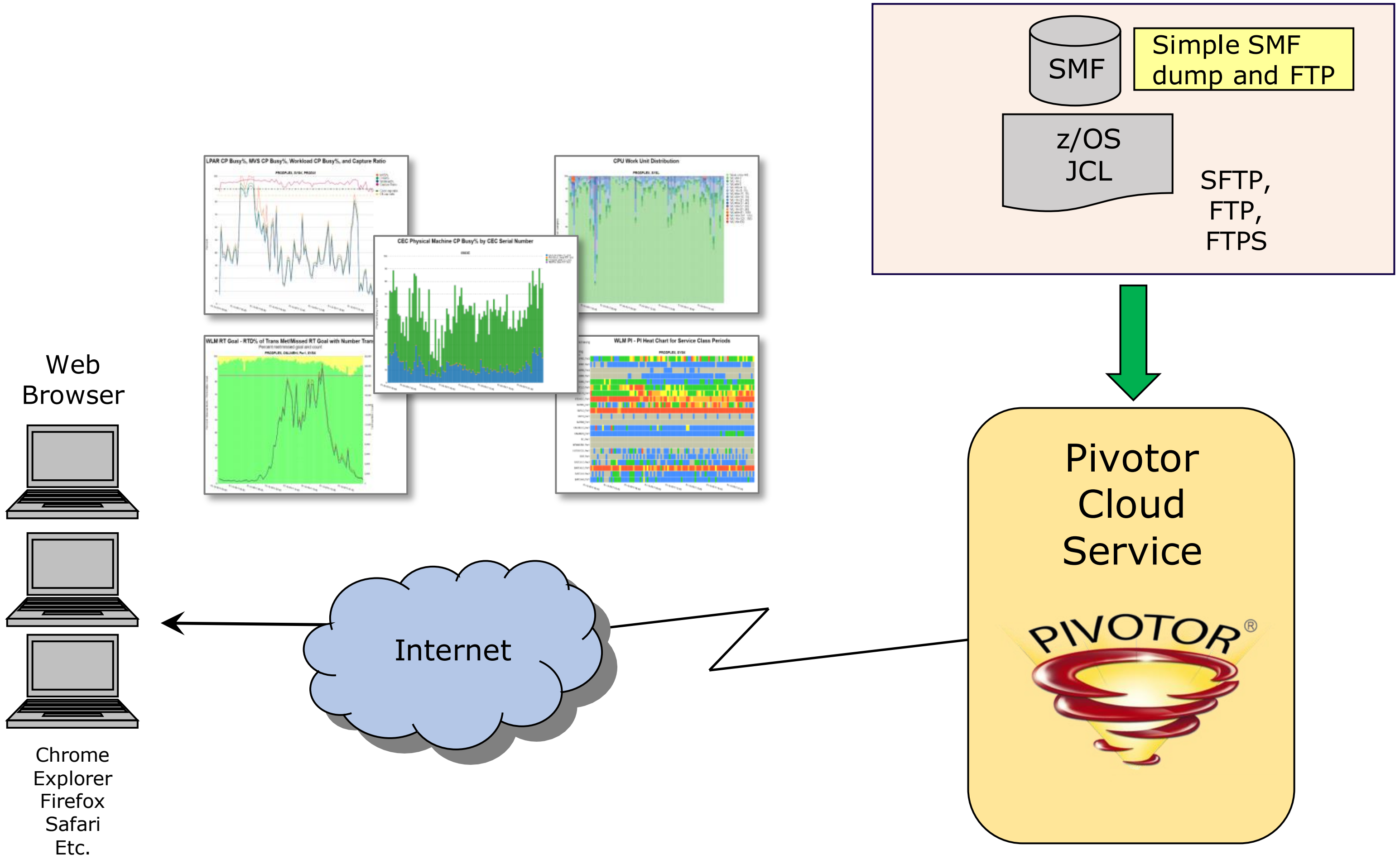
Across multiple timeframes: daily, weekly, monthly, yearly, rolling *n* days, etc.

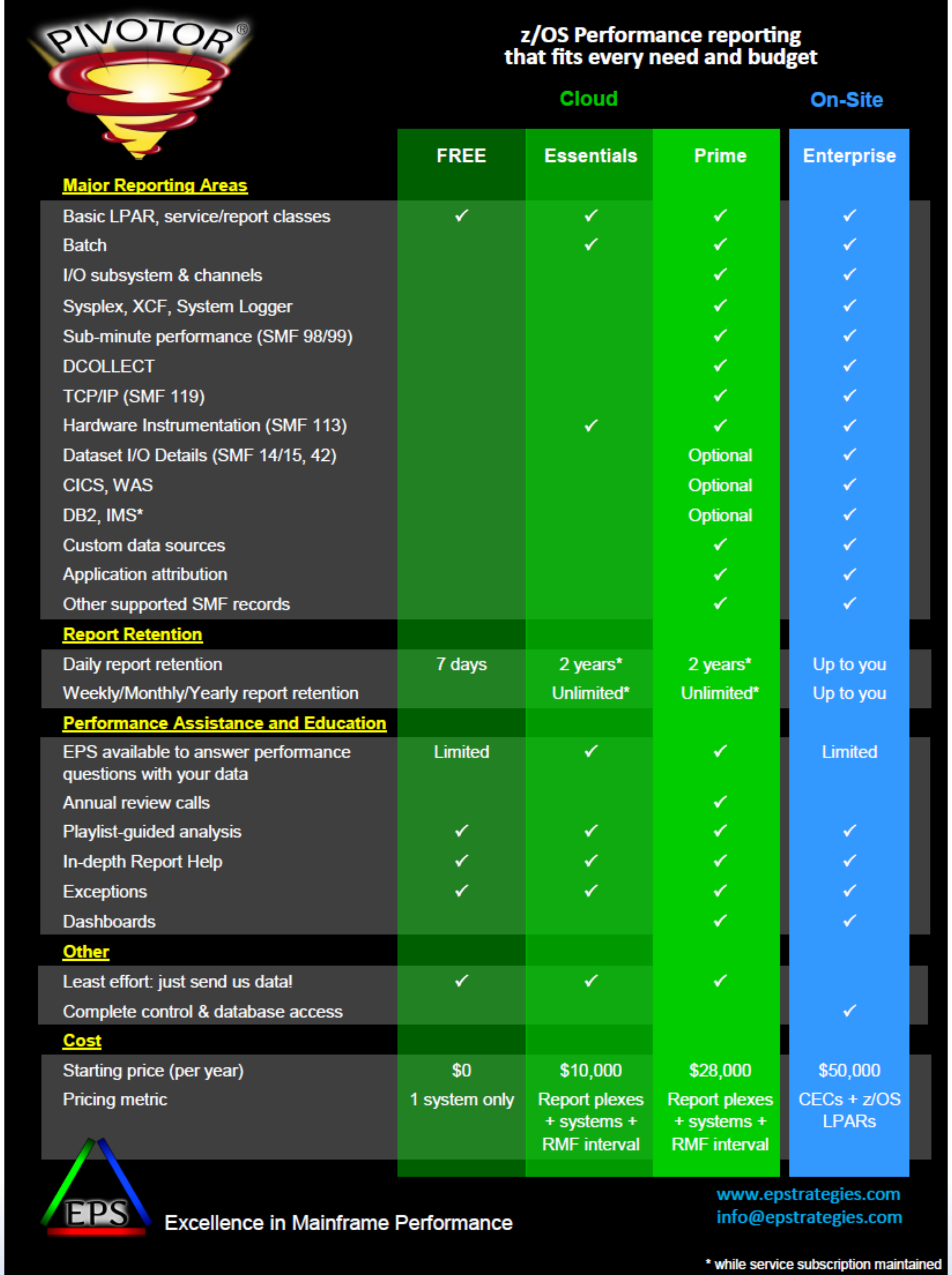
Pivotor Software as a Solution (SaaS)

- Pivotor is offered as both a SaaS or local install
- When SaaS:

SaaS Includes:

- Formal yearly cursory review / discussion
- Ability to ask us performance questions, or for us to look at a particular problem or concern. (support@epstrategies.com)
- We can occasionally look in on your data and performance
- We can participate in performance debug with IBM, or other vendors





z/OS Performance reporting that fits every need and budget

	Cloud			On-Site
	FREE	Essentials	Prime	Enterprise
Major Reporting Areas				
Basic LPAR, service/report classes	✓	✓	✓	✓
Batch		✓	✓	✓
I/O subsystem & channels			✓	✓
Sysplex, XCF, System Logger			✓	✓
Sub-minute performance (SMF 98/99)			✓	✓
DCOLLECT			✓	✓
TCP/IP (SMF 119)			✓	✓
Hardware Instrumentation (SMF 113)		✓	✓	✓
Dataset I/O Details (SMF 14/15, 42)			Optional	✓
CICS, WAS			Optional	✓
DB2, IMS*			Optional	✓
Custom data sources			✓	✓
Application attribution			✓	✓
Other supported SMF records			✓	✓
Report Retention				
Daily report retention	7 days	2 years*	2 years*	Up to you
Weekly/Monthly/Yearly report retention		Unlimited*	Unlimited*	Up to you
Performance Assistance and Education				
EPS available to answer performance questions with your data	Limited	✓	✓	Limited
Annual review calls			✓	
Playlist-guided analysis	✓	✓	✓	✓
In-depth Report Help	✓	✓	✓	✓
Exceptions	✓	✓	✓	✓
Dashboards			✓	✓
Other				
Least effort: just send us data!	✓	✓	✓	
Complete control & database access				✓
Cost				
Starting price (per year)	$0	$10,000	$28,000	$50,000
Pricing metric	1 system only	Report plexes + systems + RMF interval	Report plexes + systems + RMF interval	CECs + z/OS LPARs

Excellence in Mainframe Performance

www.epstrategies.com
info@epstrategies.com

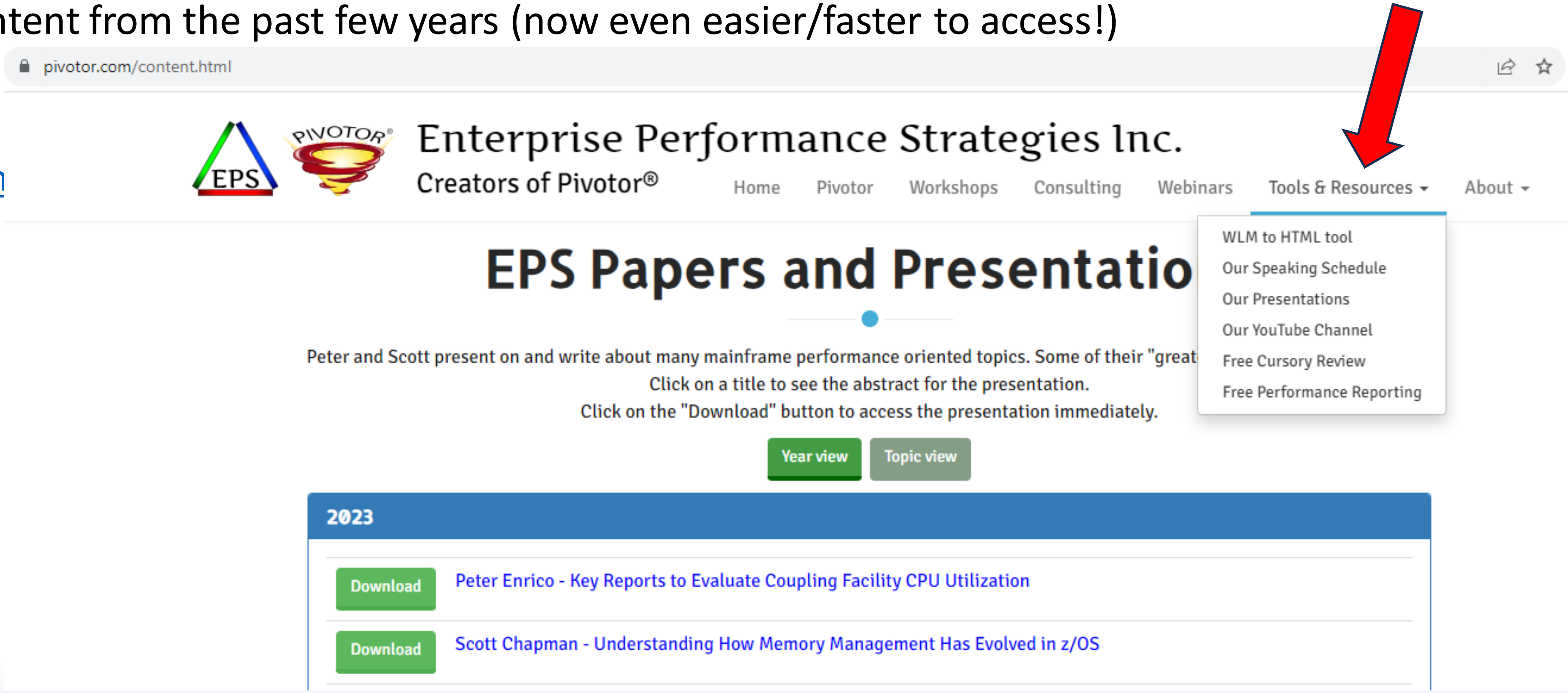
* while service subscription maintained

- Pivotor pricing is clear and affordable

More Free Things!

- On our web site click on Tools & Resources to access:
  - WLM to HTML Tool
    - Get your WLM policy in a useful and usable HTML format
  - Our Presentations
    - Lots of great content from the past few years (now even easier/faster to access!)

https://www.epstrategies.con
https://www.pivotor.com/
(Same site behind both URLs)



Enterprise Performance Strategies Inc.
Creators of Pivotor®

EPS Papers and Presentatio

Peter and Scott present on and write about many mainframe performance oriented topics. Some of their "great
Click on a title to see the abstract for the presentation.
Click on the "Download" button to access the presentation immediately.

2023

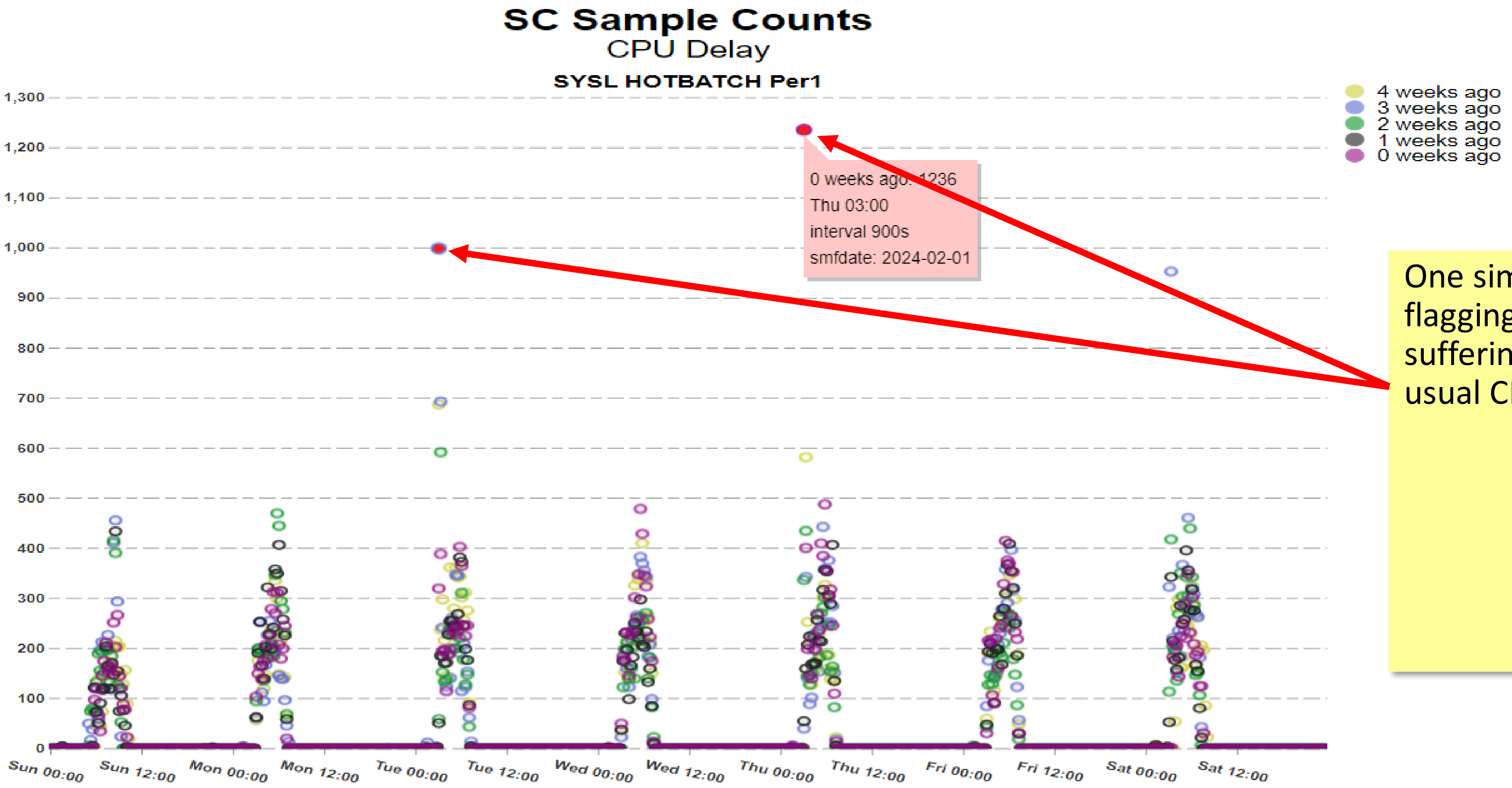
Peter Enrico - Key Reports to Evaluate Coupling Facility CPU Utilization

Scott Chapman - Understanding How Memory Management Has Evolved in z/OS

Announcement!

Pivotor Outlier Detection & Analysis

- Newly rolling out to our customers right now!
- Uses combination of Machine Learning techniques to find outliers (aka anomalies) at scale while limiting or avoiding problems inherent in previous techniques
  - Running against dozens of metrics on daily basis
- Expect it to be useful for:
  - Problem determination (including around a timeframe)
  - Early warning signals
- Webinar coming up next week (March 12th) to discuss in detail
  - See https://www.pivotor.com/webinar.html

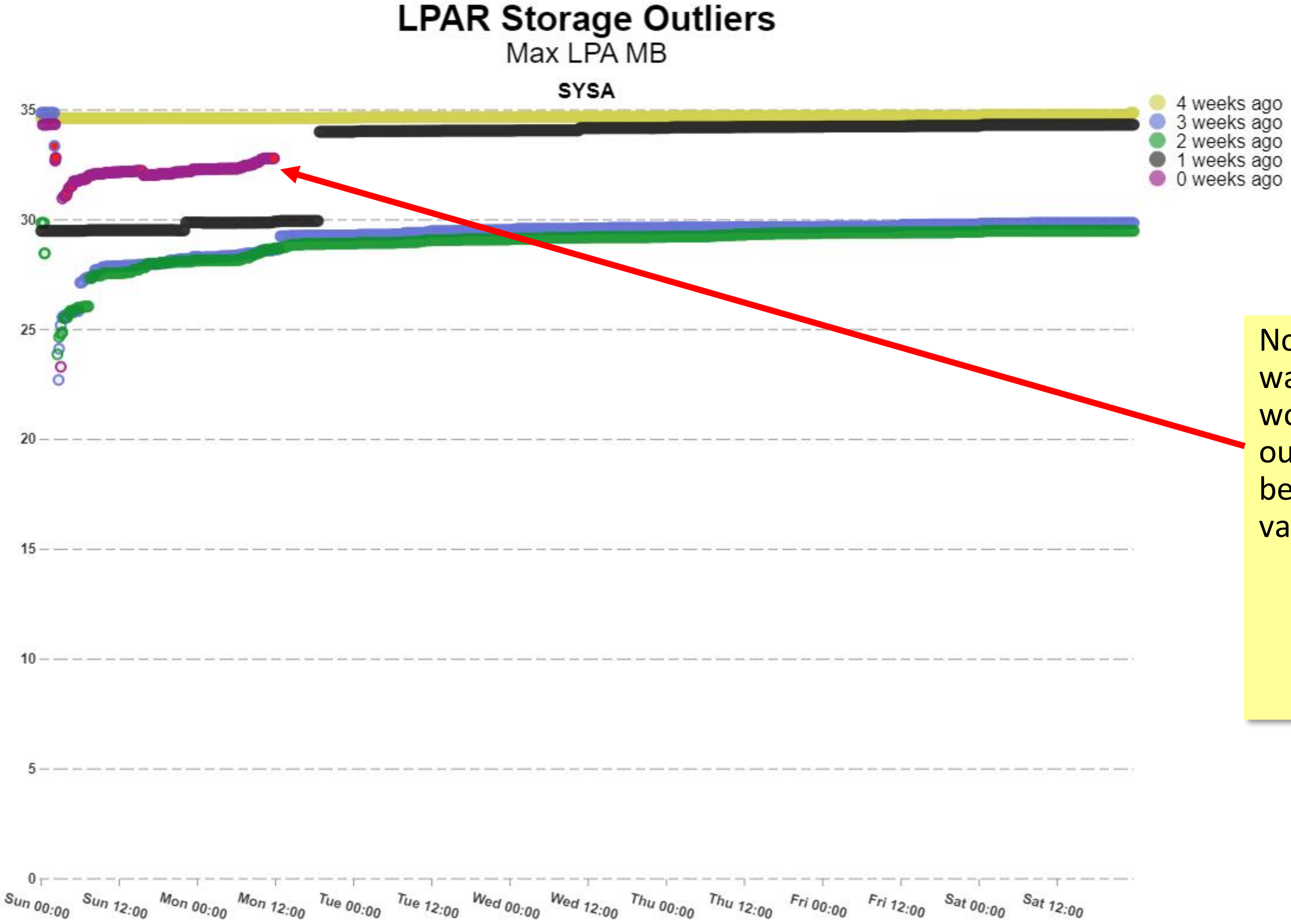


SC Sample Counts

CPU Delay

SYSL HOTBATCH Per1

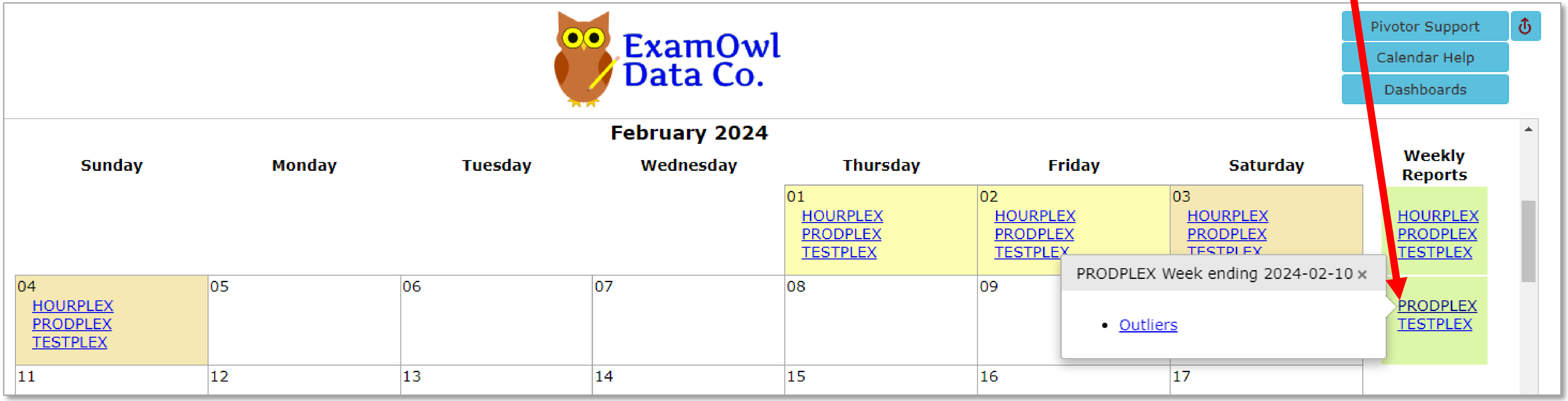
One simple example flagging a service class suffering more than usual CPU delays.



Note that because of the way the ML algorithms work, we can find outliers that might be between common values.

For Pivotor Customers…

- Attend the upcoming webinar!
- Reports run daily as a week-to-date report, so are under weekly reports
  - Should be there now for most of you
- Let us know what you think!



ExamOwl Data Co.

February 2024

Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Weekly Reports
				01 HOURPLEX PRODPLEX TESTPLEX	02 HOURPLEX PRODPLEX TESTPLEX	03 HOURPLEX PRODPLEX TESTPLEX	HOURPLEX PRODPLEX TESTPLEX
04 HOURPLEX PRODPLEX TESTPLEX	05	06	07	08	09		PRODPLEX TESTPLEX
11	12	13	14	15	16	17	

PRODPLEX Week ending 2024-02-10 ×

- Outliers

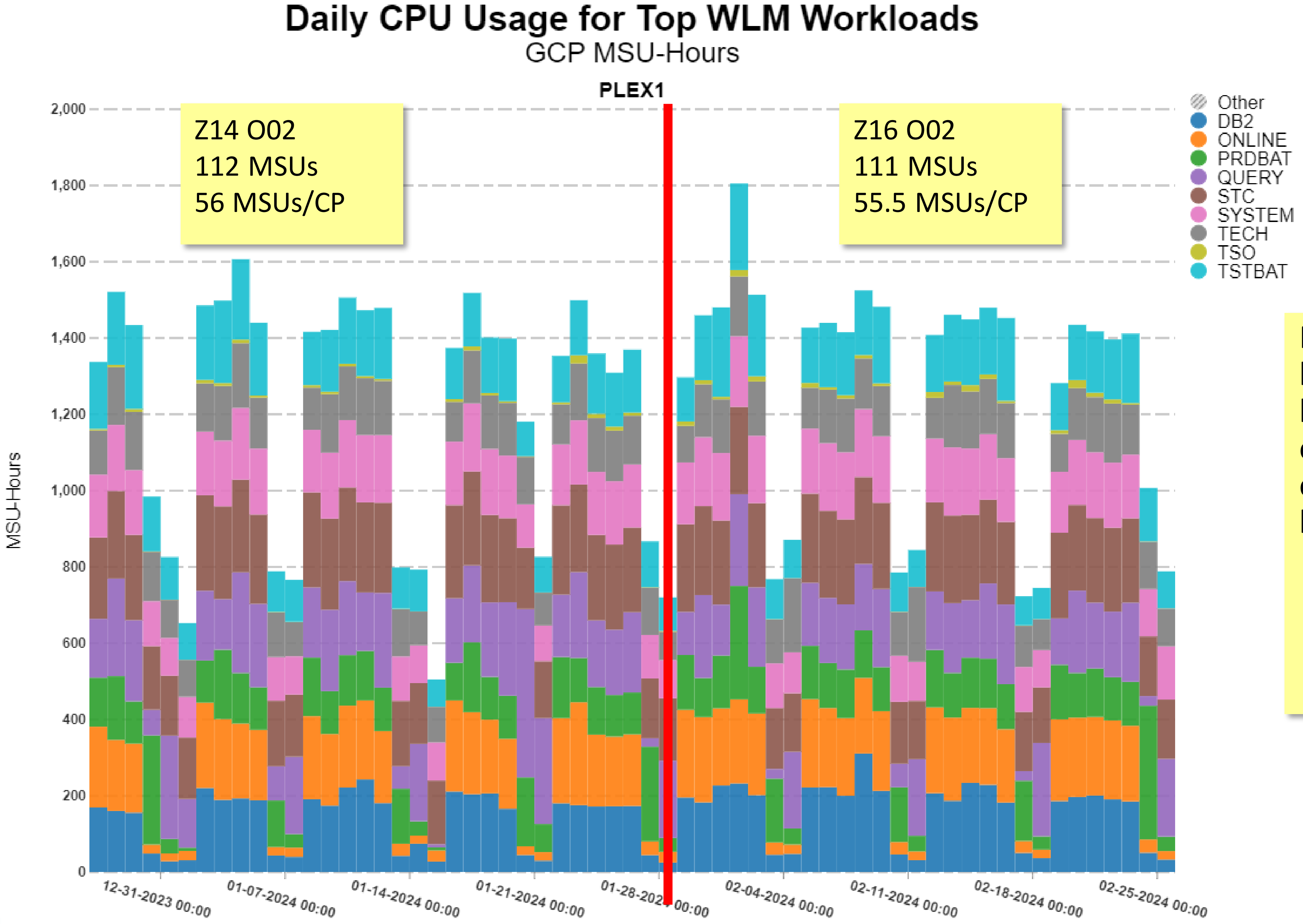
Emerging Areas of Interest

New things coming and things we're actively keeping an eye on

z16 Migrations

TLDR: Mostly going as expected

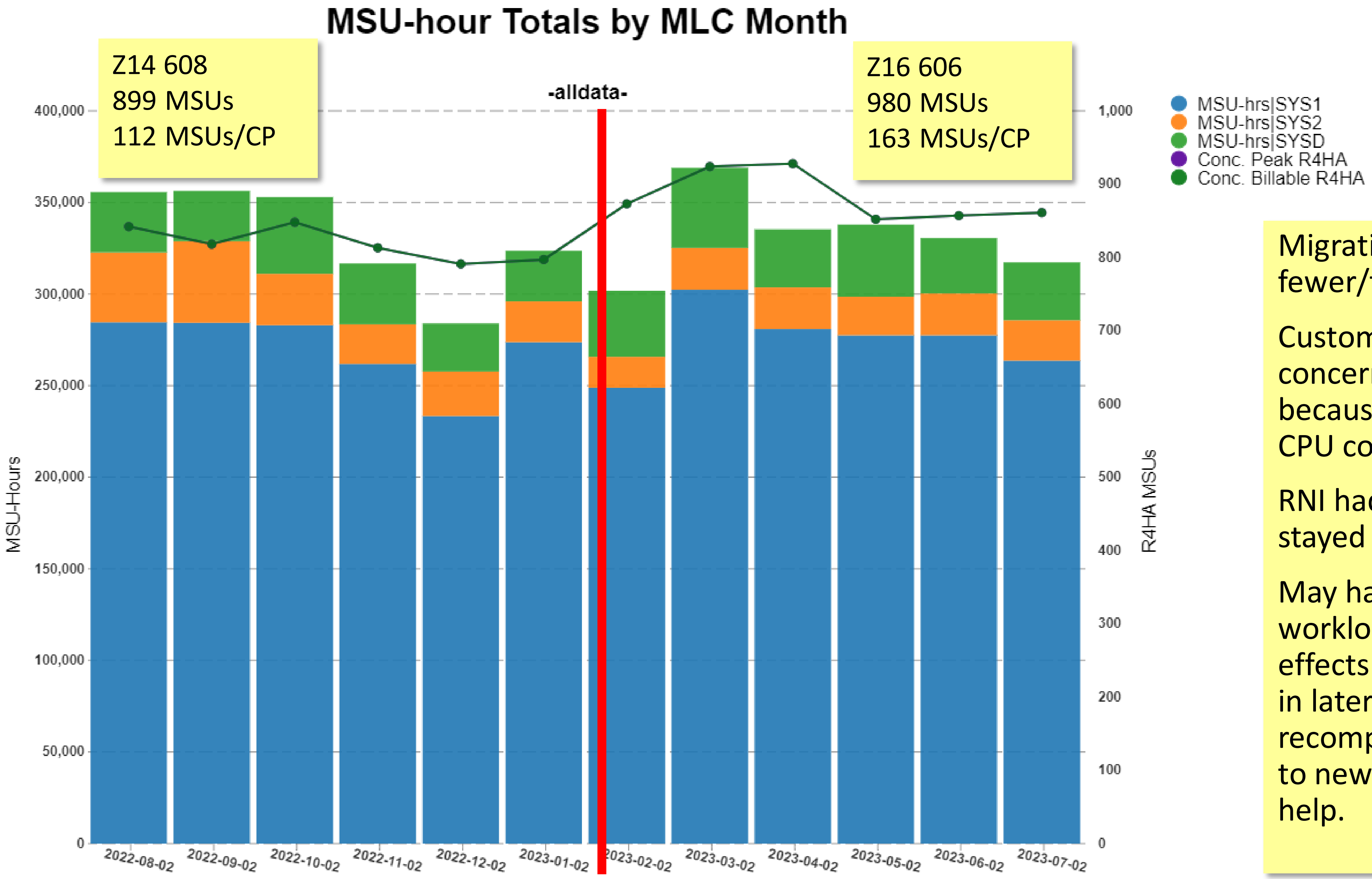
- There was some questions about how the significant cache design change would behave in real life
- For the migrations we've seen, it seems that migrations to the z16 have been pretty much along (our) expectations
  - Except for the one customer that did contact us that saw higher MSU consumption, but the had moved to fewer/faster CPs
- In general, fewer/faster CPs are likely to be worse for overall system efficiency
  - Thought the larger L2 cache size might mitigate this, but… maybe not
- More/slower (or more/faster!) better for efficiency
- Staying with same number of CPs is the conservative approach



Z14 O02
112 MSUs
56 MSUs/CP

Z16 O02
111 MSUs
55.5 MSUs/CP

Recent migration, very low risk because they kept the same engine count and overall capacity rating. Looks like it went fine.



Z14 608
899 MSUs
112 MSUs/CP

Z16 606
980 MSUs
163 MSUs/CP

Migration to fewer/faster.

Customer had some concerns after migrating because of increased CPU consumption.

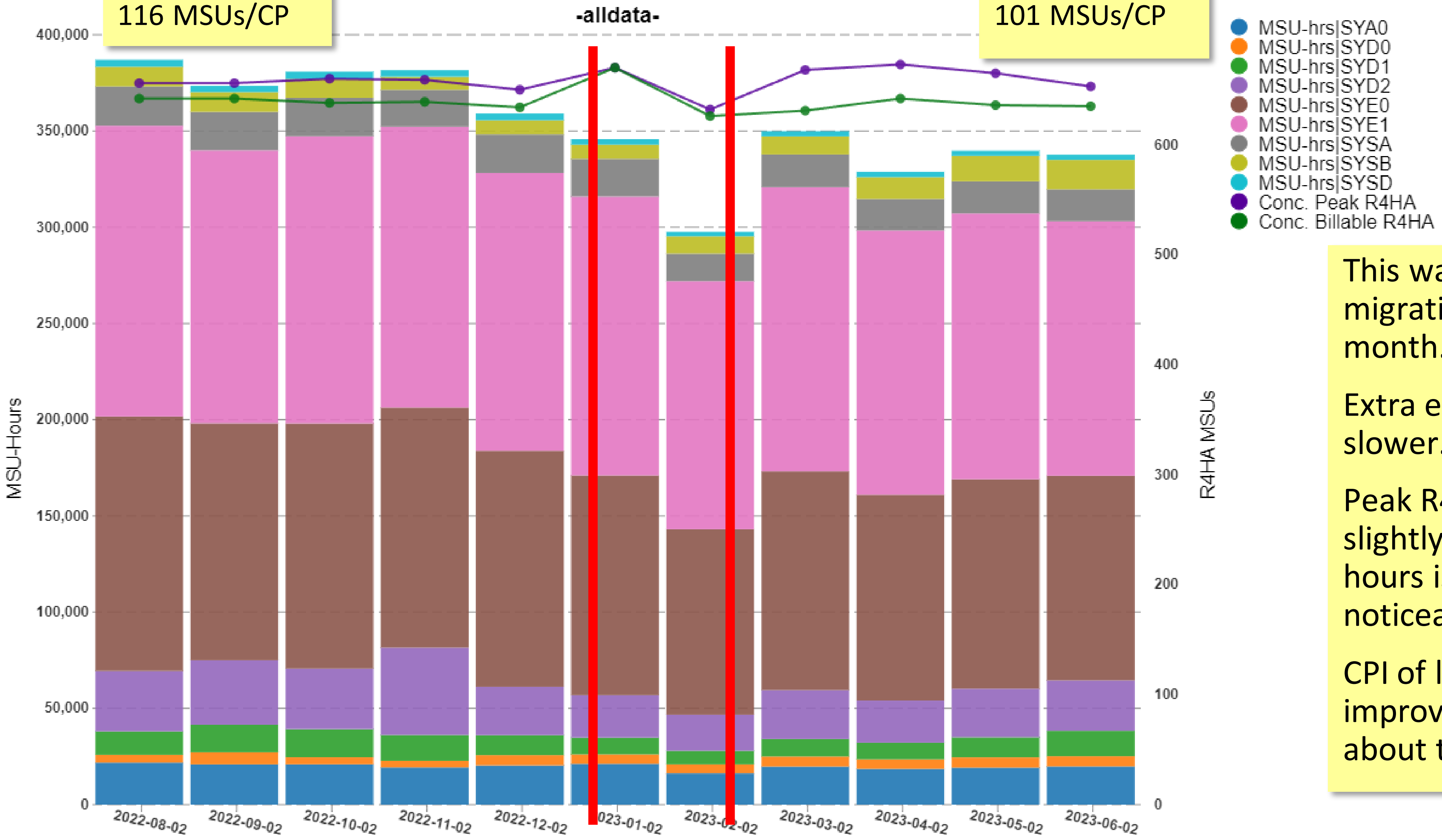
RNI had gone up, CPI stayed similar.

May have been workload related as effects seemed reduced in later months. Also, recompiling after moving to new architecture can help.

MSU-hour Totals by MLC Month

Z14 606
696 MSUs
116 MSUs/CP

Z16 507
706 MSUs
101 MSUs/CP



This was a phased migration over about a month.

Extra engine only slightly slower.

Peak R4HA did go up slightly, but total MSU-hours is down noticeably.

CPI of larger systems did improve, RNI stayed about the same.

z/OS 2.5 Service Definition Coefficients

Like goals, durations need to be periodically re-evaluated
(but many haven't!)

Service Definition Coefficients Updates

- Recommended values by EPS since about 2018 (maybe earlier)
  - CPU=1, SRB=1, IOC=0 MSO=0
  - Summary of reasoning: Aging a transaction based on I/O no longer made much sense since I/O priority management mattered much less due to advent of PAVs, and most I/O processing is also outside the z/OS operating system. So why age a workload based on its I/O characteristics. It is CPU that matters.
- z/OS 2.5 the SDCs go away, and the values will default as follows
  - CPU=1, SRB=1, IOC=0. MSO=0
  - Basically, it is durations are now based on CPU and SRB service units, and not longer based on the concept of 'service'.
- Most customers are using 1,1,0,0
  - If you haven't made the transition yet, read next slides…

IBM’s z/OS 2.5 Migration Step

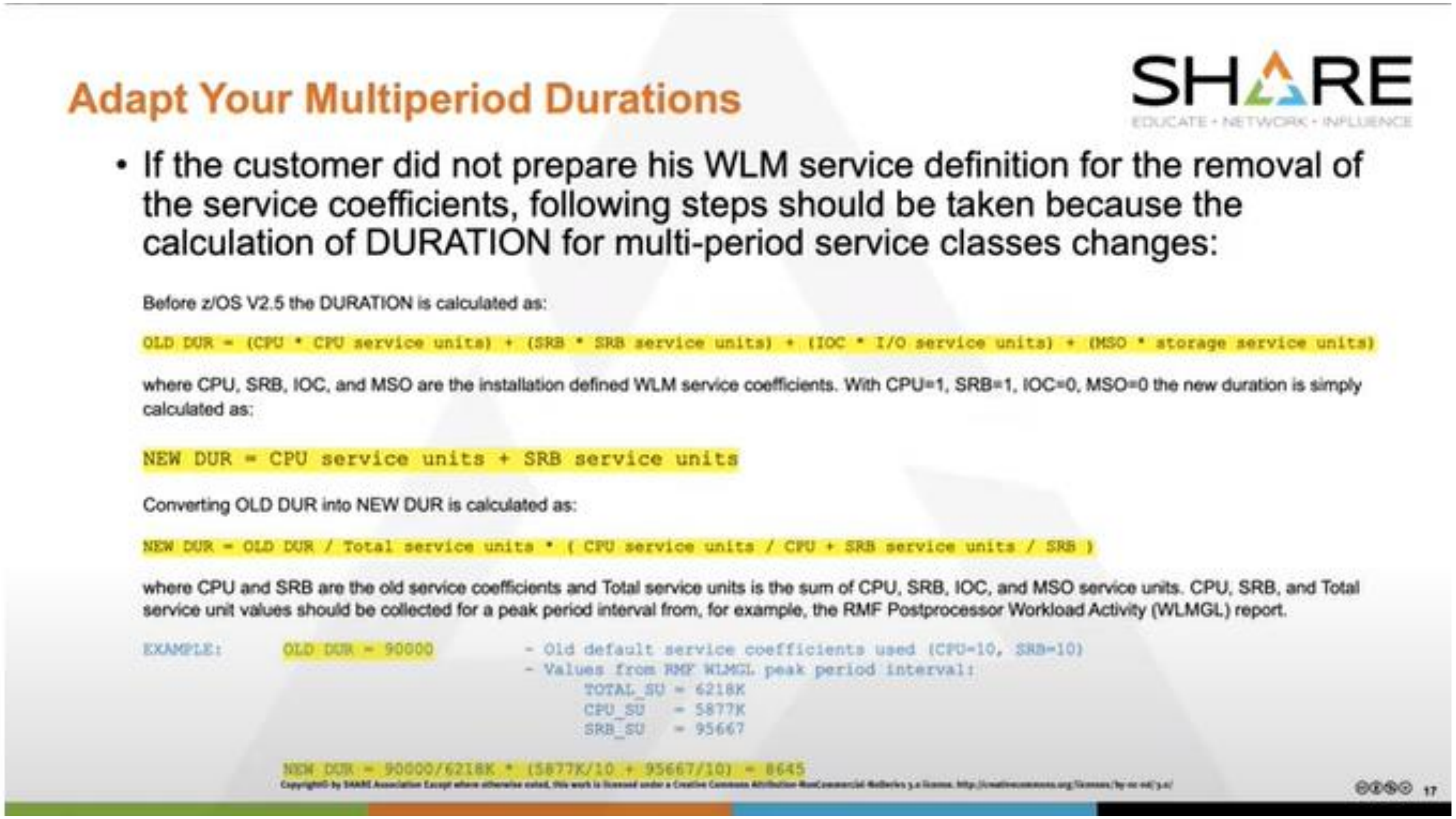
The following is an excerpt from SHARE presentation:
PERFORMANCE INFRASTRUCTURE IMPROVEMENTS IN Z/OS V2.5 WLM
Presenter:
ANDREAS HENICKE (IBM WLM)

Presentation discusses the z/OS 2.5 migration steps suggested to migrate your period durations prior to migrating to z/OS 2.5.

Basically, IBM is suggesting to take CPU and SRB ‘service’, divide by your current SDCs to convert to ‘service units’. Then take the sum of those values and multiple them by the ratio of current duration to service consumed.

Or put a little simpler…
Blah, blah, blah…

Feel free to take this approach, but a bit to complicated for me.



Adapt Your Multiperiod Durations

- If the customer did not prepare his WLM service definition for the removal of the service coefficients, following steps should be taken because the calculation of DURATION for multi-period service classes changes:

Before z/OS V2.5 the DURATION is calculated as:

```
OLD DUR = (CPU * CPU service units) + (SRB * SRB service units) + (IOC * I/O service units) + (MSO * storage service units)
```

where CPU, SRB, IOC, and MSO are the installation defined WLM service coefficients. With CPU=1, SRB=1, IOC=0, MSO=0 the new duration is simply calculated as:

```
NEW DUR = CPU service units + SRB service units
```

Converting OLD DUR into NEW DUR is calculated as:

```
NEW DUR = OLD DUR / Total service units * ( CPU service units / CPU + SRB service units / SRB )
```

where CPU and SRB are the old service coefficients and Total service units is the sum of CPU, SRB, IOC, and MSO service units. CPU, SRB, and Total service unit values should be collected for a peak period interval from, for example, the RMF Postprocessor Workload Activity (WLMGL) report.

```
EXAMPLE:     OLD DUR = 90000          - Old default service coefficients used (CPU=10, SRB=10)
                                      - Values from RMF WLMGL peak period interval:
                                           TOTAL_SU = 6218K
                                           CPU_SU   = 5877K
                                           SRB_SU   = 95667

             NEW DUR = 90000/6218K * (5877K/10 + 95667/10) = 8645
```

Copyright© by SHARE Association Except where otherwise noted, this work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 license. http://creativecommons.org/licenses/by-nc-nd/3.0/

Peter’s Approach to Migrating SDCs to New z/OS 2.5

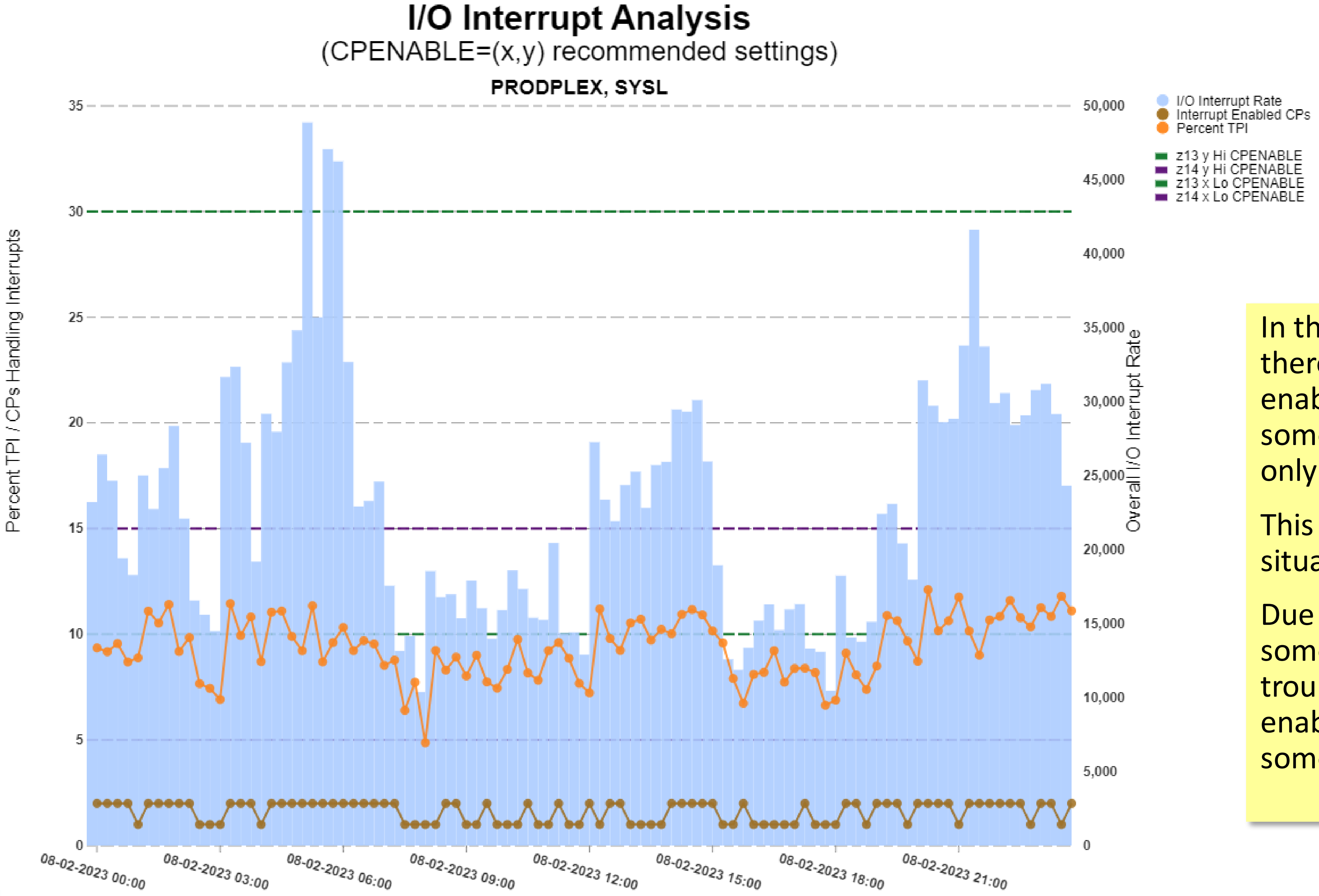
- Understand that most durations for multiple periods are usually wrong to begin with.
  - If you feel yours are correct, then do this exercise

- My general approach is a follows:
  1. Determine your current SDCs
  2. Remember the reason you are defining a multiple period service class
  3. Determine your current multiple period service classes
     - Most likely multiple periods are only being used for the following interactive workloads or certain batch
       - TSO, Interactive OMVS, DDF, WAS CB, Batch (sometimes)
  4. Determine which multiple period service classes are consuming I/O service and how much
  5. Then ignore any sort of duration migration exercise for the following enclave workload types since these enclave workloads do not consider I/O service
     - DDF
     - WAS CB
     - So will be left with workloads such as eft with only TSO, interactive OMVS, and Batch,
  6. Revisit duration
     - Either start fresh (which should be done for many periods regardless of this change)
     - Ignore and accept
     - Tweak

CPENABLE in z/OS 3.1

CPENABLE

- CPENABLE in IEAOPTxx sets the low and high threshold for disabling / enabling processors for handling I/O interrupts
- z13 and below recommendation is (10,30)
- On z/14 and above the recommendation is (5,15)
  - Prior to z/14 all no-work wait CPs were enabled for interrupts
  - z/14+ rely solely on WLM/SRM to set the number of CPs enabled for interrupts
- The goal of this change was to better ensure 2 CPs are enabled for handling I/O interrupts
  - Single CP enabled for I/O interrupts puts LPAR at greater risk of delaying I/O
  - Sometimes with quite problematic results – having 2 is partly risk mitigation
- We've sometimes recommended even more aggressive settings (e.g. 3,10)



In this case, sometimes there were 2 CPs enabled for interrupts, sometimes there was only a single CP.

This is a fairly common situation.

Due to arrival patterns, some systems have trouble getting a second enabled even with something like (3,10).

CPENABLE in z/OS 3.1

- In z/OS 3.1 minimum CPs enabled will raise from 1 to 2
  - The only z/OS 3.1 LPAR we've seen data for only had 1 online processor ☹
- New CPENABLE option of SYSTEM will take IBM's recommendation for the generation of hardware the system is running on
- Evaluation of enabled CPs will change from 20 seconds to 2 seconds

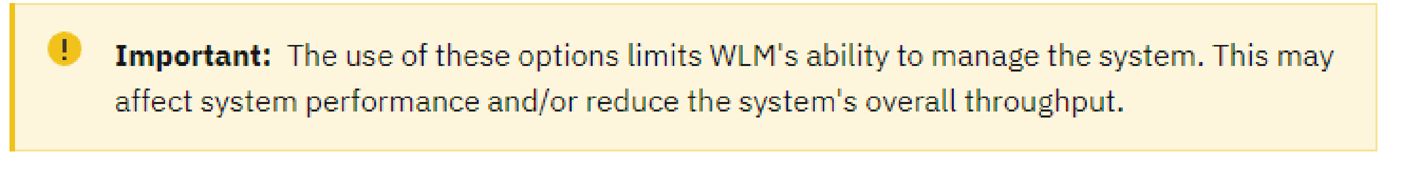
- We think this is a great change!
  - Will be able to specify CPENABLE=SYSTEM and probably not worry about it
  - A lot of I/O can happen in 20 seconds so changing to every 2 seconds (same as HiperDispatch cycle) makes sense
    - Extra path length seems like it would be pretty minimal

Implicit CPU Protection in z/OS 3.1

Also, see our presentation from Monday!

CPU Critical aka Long-term CPU Protection

- Long-time option in your WLM service definition
- Enabled by setting YES for CPU Critical on a Service Class
  - Must be a single-period SC and cannot be discretionary
- Ensures that the CPU Critical SC always has a dispatching priority that's greater than the DP of lower importance service class periods
- Note some small amount of lower-importance work may still get higher DP:
  - Due to promotion for locks, resource contention, etc.
  - Small consumers
- General recommendation has been to avoid this option
  - Allows WLM to make better decisions about balancing overall work throughput to best meet the goals of all work



Important: The use of these options limits WLM's ability to manage the system. This may affect system performance and/or reduce the system's overall throughput.

New IBM Defaults in z/OS 3.1

- New option for “Implicit” Long-Term CPU Protection
  - In other words, CPU Critical without having to specify it on every SC definition
- Default is “On” for importance 1 service classes
  - Optional, but “Off” for importance 2 service classes
- We think “On” for importance 1 workloads is a bad default
  - Could significantly change the dispatching priority of work in the system
  - Goes against historical practices of not changing defaults that change behavior
- DP/Importance inversions are common
  - I.E. Lower Importance work running with a DP above higher importance work
  - Not all such inversions are problematic
  - Not all importance 1 work really should be importance 1

Our thoughts

- We don't see the need for this change
  - A significant part of the premise of WLM was that it would manage dispatching priorities and could intelligently move them in possibly counter-intuitive ways to better balance throughput for diverse workloads
  - If you want, you can make all importance 1 work CPU Critical today
- We'd recommend turning this off for z/OS 3.1 and wish that was the default
- If you want to go to z/OS 3.1 with it on, we might suggest
  1. Evaluate which workloads are at risk
  2. Before 3.1, incrementally add CPU Critical to importance 1 workloads
     - If something goes wrong you can back out your change and z/OS 3.1 doesn't get the blame
- We do sometimes recommend CPU Critical, but it's an exception, not the rule
- Emerging area of study, we might refine our recommendations over time

First Reference Page Faults Decrease Capture Ratios

What is a first reference page fault?

- Demand Page Faults
  - Typically, virtual frames are backed by real storage
  - If there is stress on storage, a real frame could be paged out to auxiliary storage
  - When that frame is re-referenced, this is known as a demand page fault
  - Demand Page Fault:
    - When a referenced page of virtual storage is not backed by a frame in central storage, a page fault occurs. This requires z/OS to retrieve the page from auxiliary storage and bring it into central storage.

- First Reference Page Fault
  - When a referenced page of virtual storage is not ***YET*** backed by a real frame in central storage, a first reference page fault occurs
  - It is the 1st reference page fault that drives Dynamic Address Translation (DAT), and the real frame is associated with the virtual address

Capture Ratios and 1st Referenced Page Faults

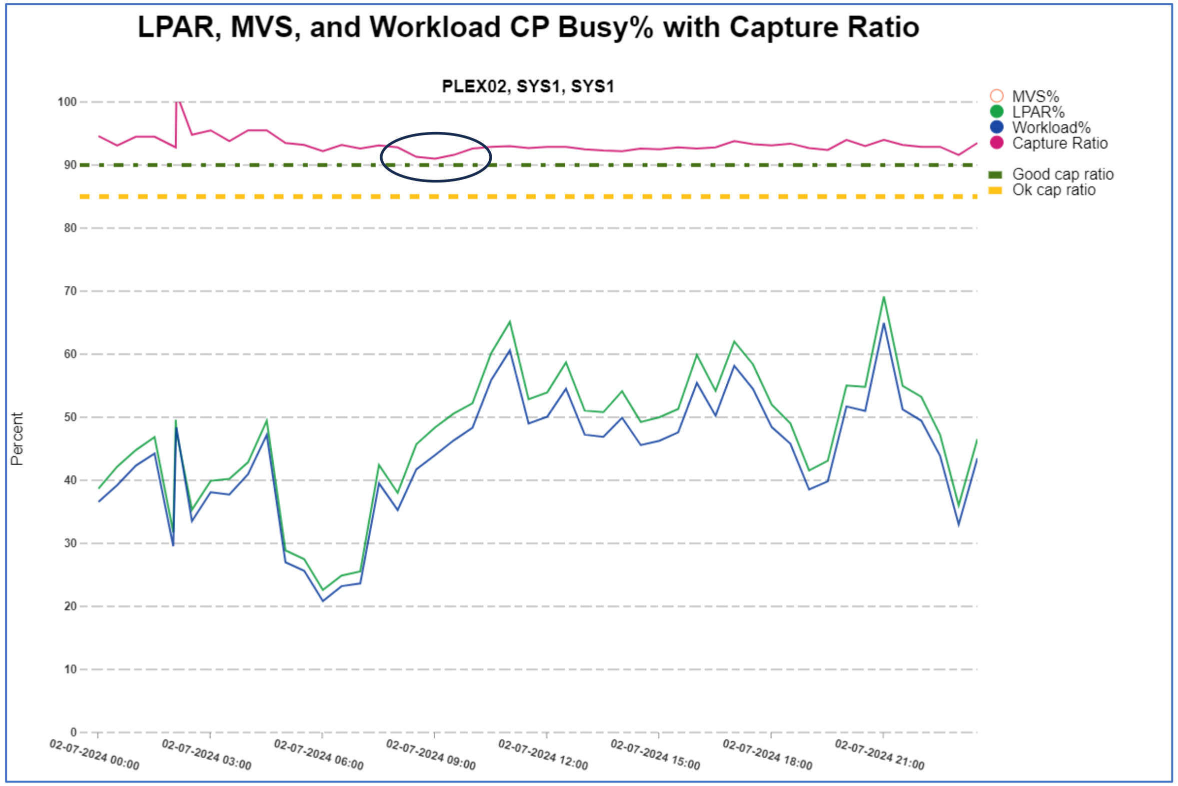
- IBM WCS says that 1st Reference Page Faults contribute to uncaptured times
  - And that 1st Reference Page Fault rates above 100,000 per second should be considered problematic

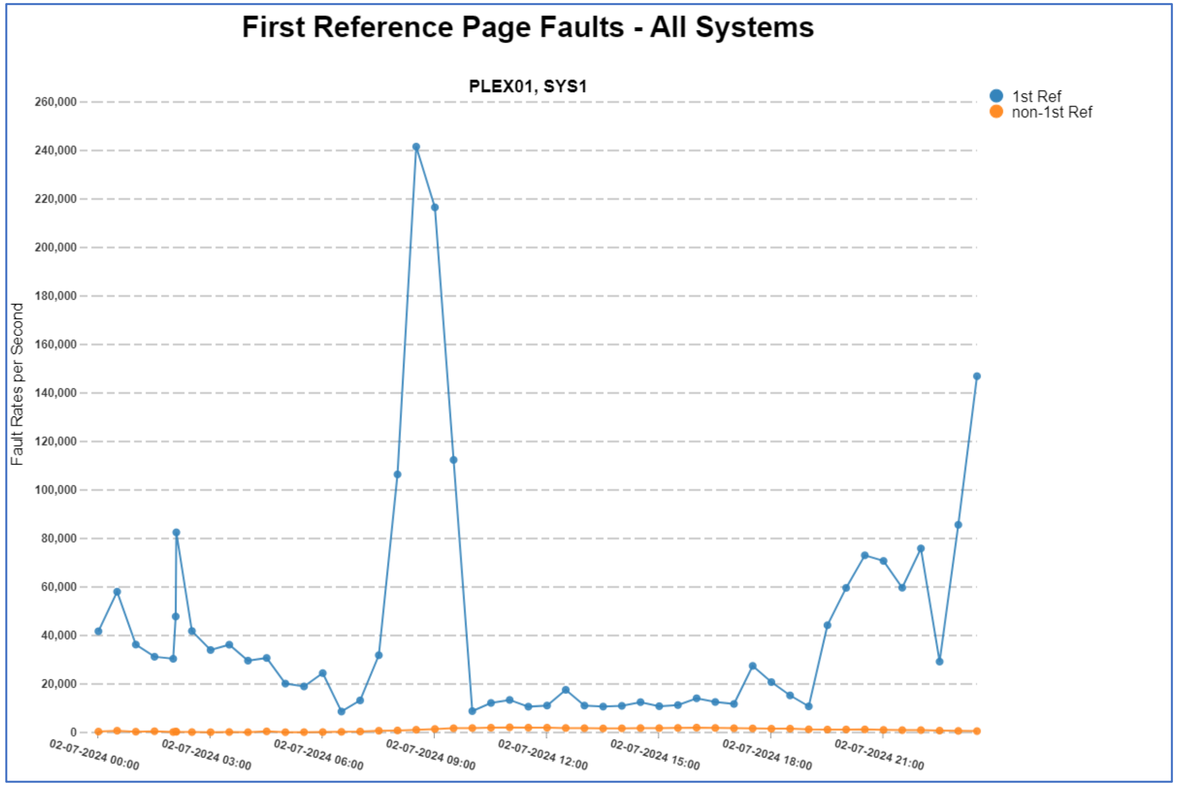
- Comments:
  - There is not much that can be done by customers to alleviate 1st Reference Page Faults
    - Perhaps recode applications to get less storage?
    - However, correlating them to capture ratios can be helpful to explain some of the uncaptured times
  - So many things contribute to uncaptured times, that is tough to see the direct correlation
  - Just understand this, and if investigating low capture ratios, then consider analyzing your 1st reference page faults to *maybe* help explain.

Example : Tough to see any correlation

Capture Ratios for System

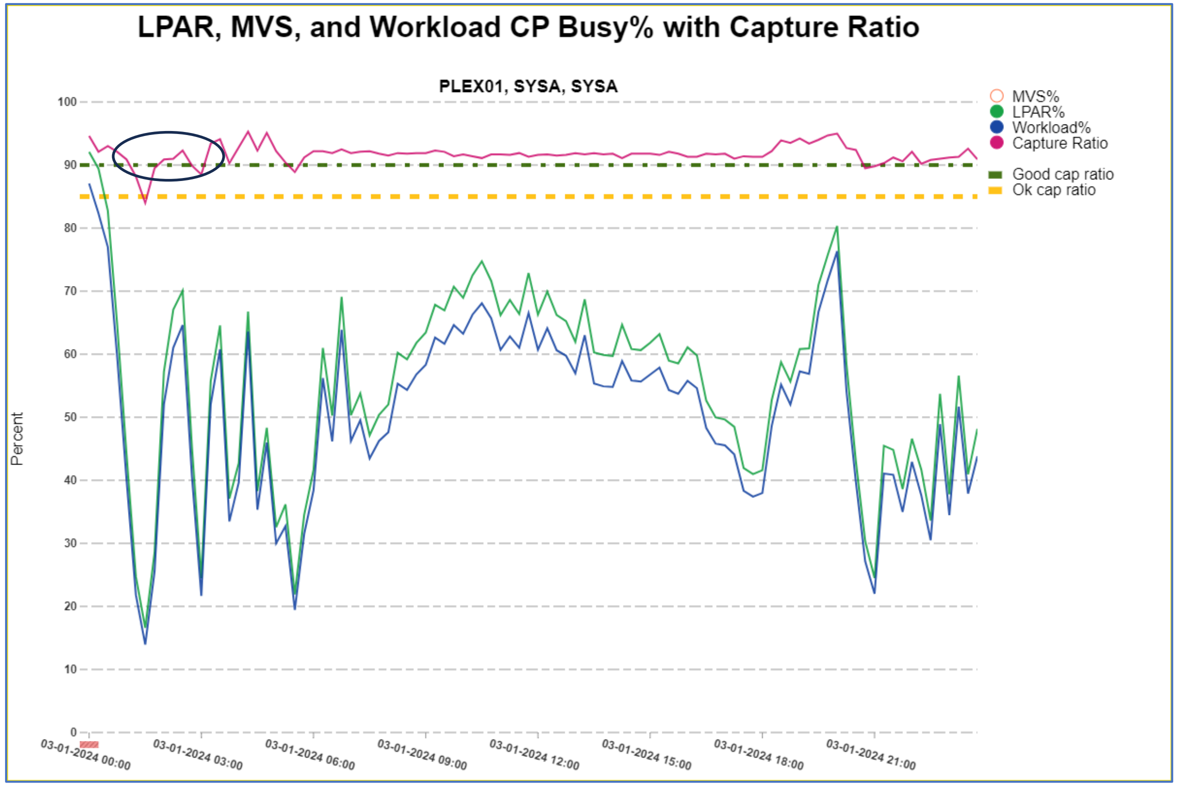
1st Reference Page Fault Rates



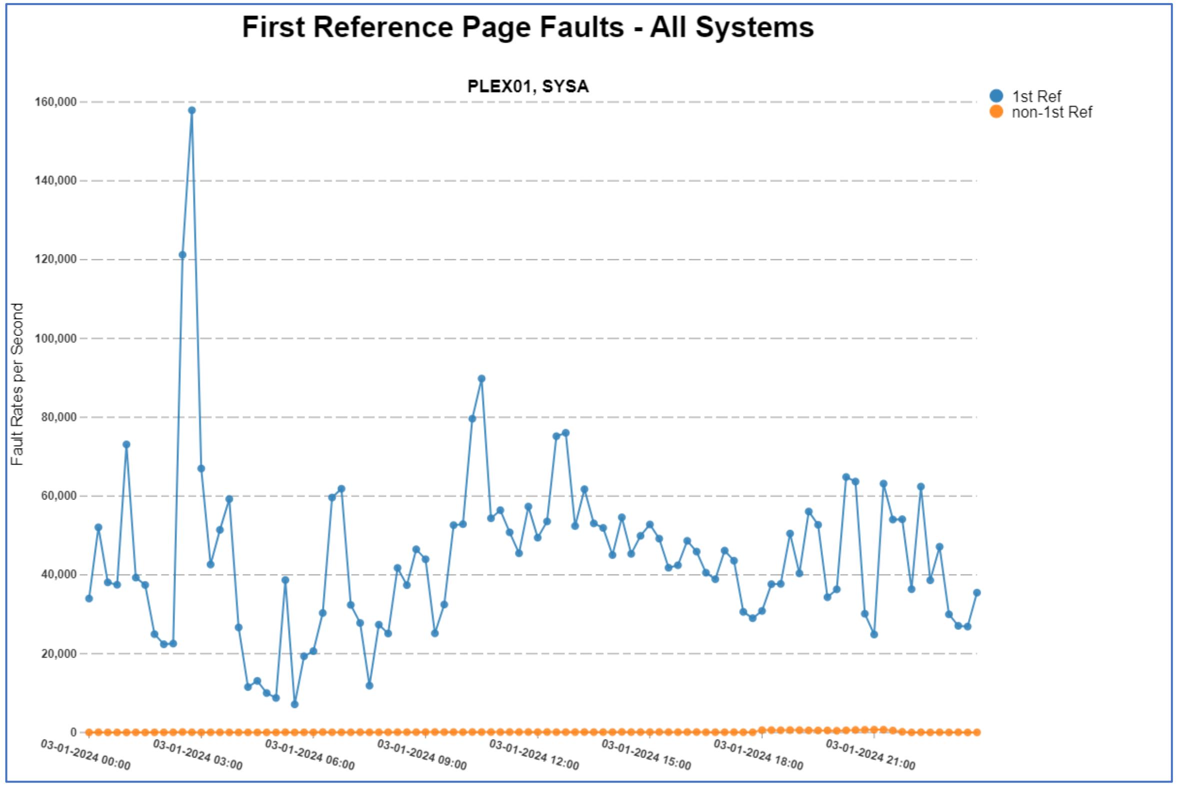


Example : Tough to see any correlation

Capture Ratios for System



1st Reference Page Fault Rates



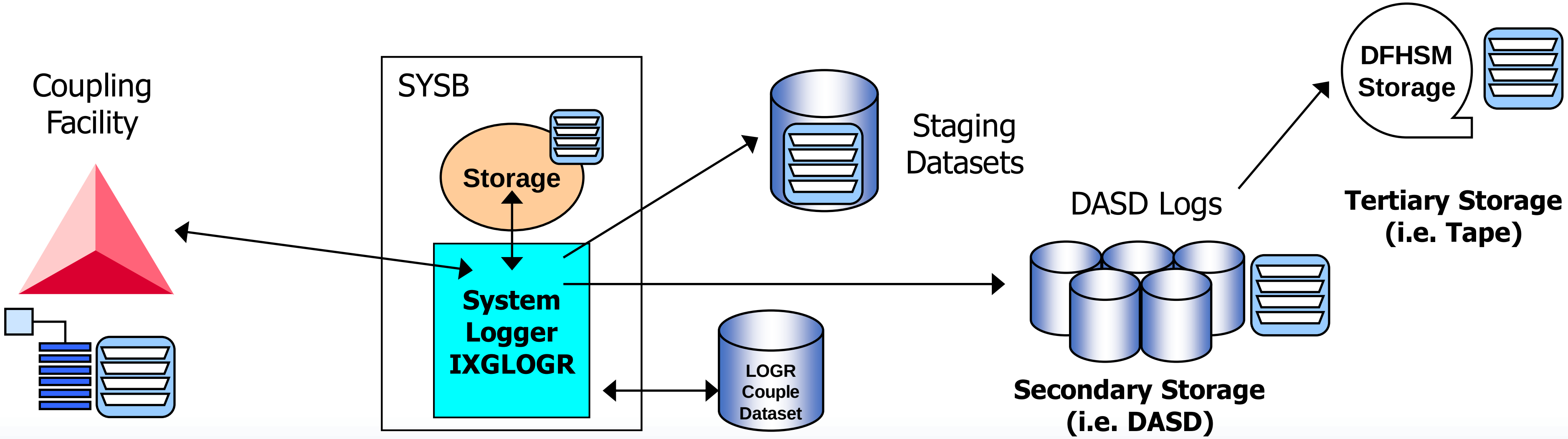
New z/OS System Logger IXGCNFxx Parameter

KEEPLOCALBUFFERS(NO | YES

Targeted to alleviate the uncaptured time due to 1st reference page faults

Introduction to z/OS System Logger

- □ z/OS System Logger - Component of z/OS that provides logging services
  - IXGLOGR – key system address space for logger functions
  - Interim Storage - Primary storage used to hold the log data that has not yet been offloaded
    - □ What 'interim storage' is depends on how the log stream has been setup
    - □ Examples of include central storage (via a data space), Coupling Facility, Staging data sets
  - Secondary Storage - generally DASD
  - Tertiary Storage – generally Tape medium



New system Logger IXGCNFxx Parm

(APAR OA63551)

- PROBLEM DESCRIPTION:
  - New function to reduce page faults caused by IXGWRITE requests that were submitted after a log stream offload occurred.

- RECOMMENDATION:
  - Delays in completing IXGWRITE requests can occur as a result of page faults associated with system logger local buffers used by IXGWRITE processing.

- Comments
  - A new IXGCNFxx parmlib option will be introduced to keep the real frames that back the local buffers when the storage for the local buffers are freed after a log stream offload.
  - Keeping the real frames reduces page faults that will occur when the local buffers are reused during subsequent IXGWRITE requests. This will result in an increase of real storage associated with the System Logger address space.

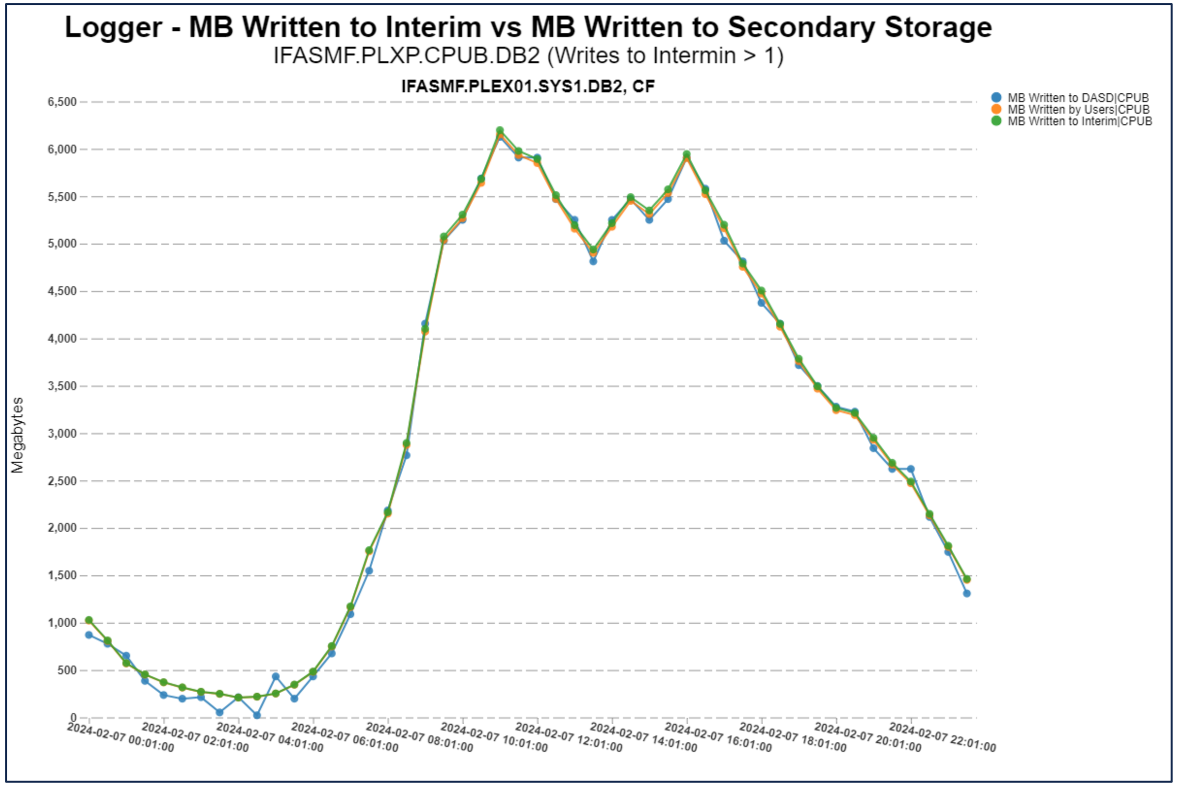
New IXGCNFxx KEEPLOCALBUFFERS Parm

KEEPLOCALBUFFERS(NO | YES)

- Specifies whether the system will request to keep the real frames backing the local buffers used as interim storage when it is freed. Keeping the real frames reduces page faults that will occur when the local buffers are reused during subsequent IXGWRITE requests.
- Note: Local buffers are data space areas associated with the system logger address space, IXGLOGR. Specifying KEEPLOCALBUFFERS(YES) may result in systems experiencing increased paging.
- Evaluate your real memory requirements to ensure unacceptable paging does not occur by reviewing the amount of real memory consumed by the system logger address space, IXGLOGR.
- The following options are possible:
  - NO - Indicates that the system will not keep the real frame used to back local buffers when the buffer storage is freed.
  - YES - Indicates that the system will request to keep the real frame used to back local buffers when the buffer storage is freed .
- You can use the DISPLAY LOGGER,IXGCNF,MANAGE command to view the parameter settings for configuring the system logger.
- Default: NO

Example: Logger Offloads of SMF

MBs of SMF offloaded

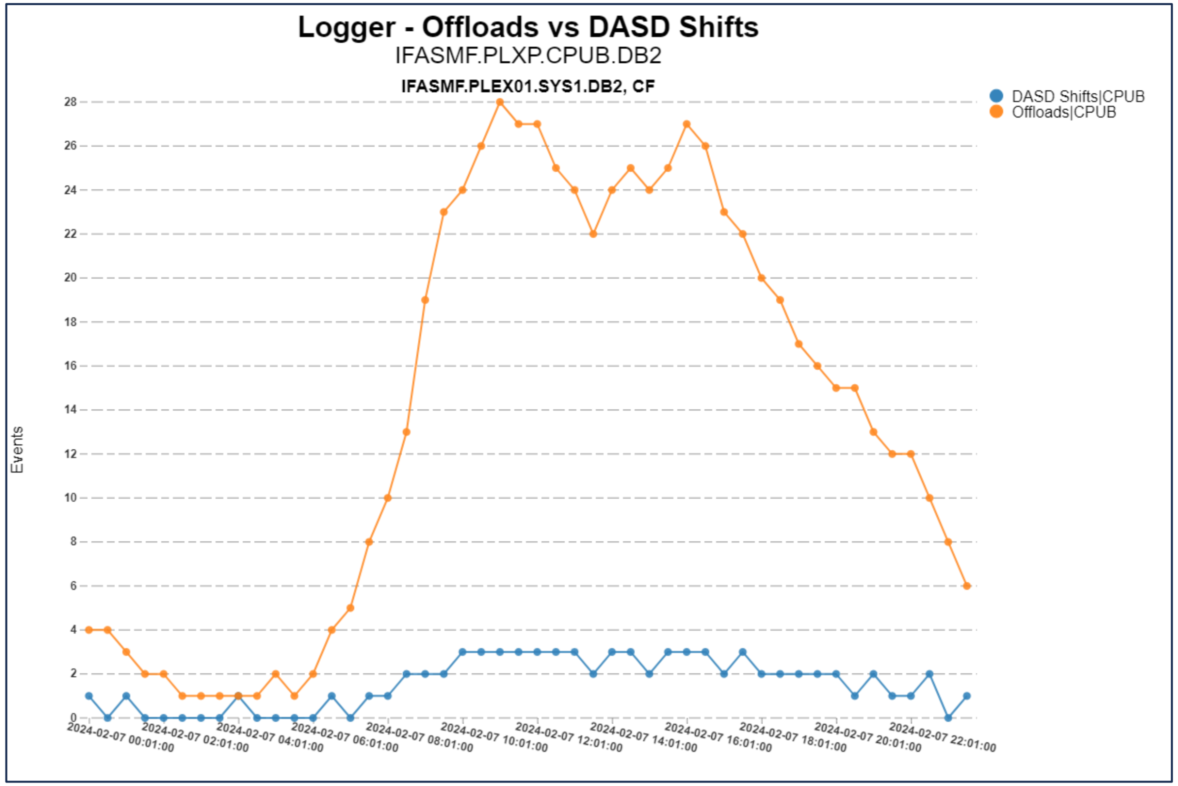


Logger - MB Written to Interim vs MB Written to Secondary Storage

IFASMF.PLXP.CPUB.DB2 (Writes to Intermin > 1)

IFASMF.PLEX01.SYS1.DB2, CF

Number of offloads



Logger - Offloads vs DASD Shifts

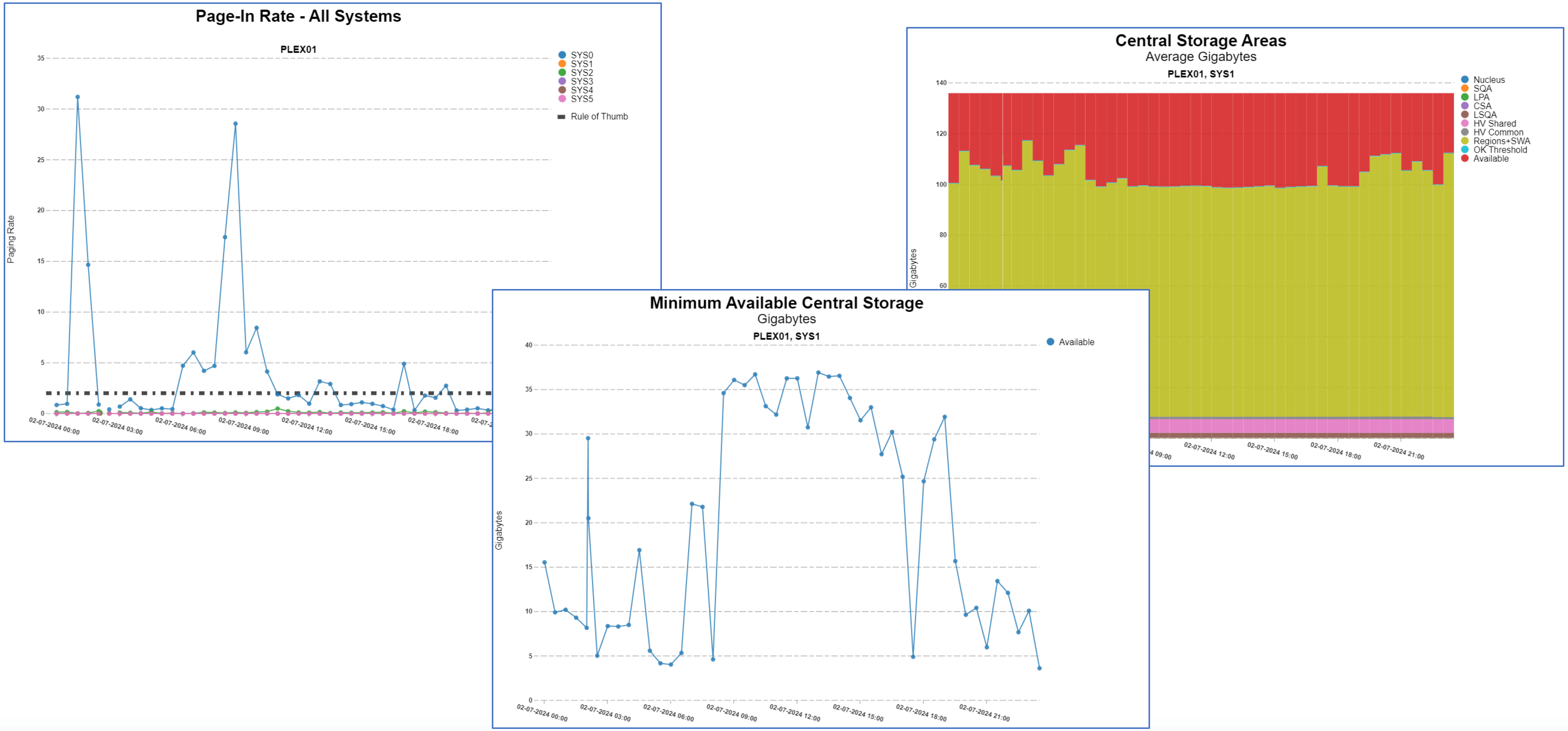
IFASMF.PLXP.CPUB.DB2

IFASMF.PLEX01.SYS1.DB2, CF

Example: Logger Offloads of SMF

Paging

Average and Min Storage Available



Large memory should mean less I/O

See also: Scott’s presentation from last SHARE

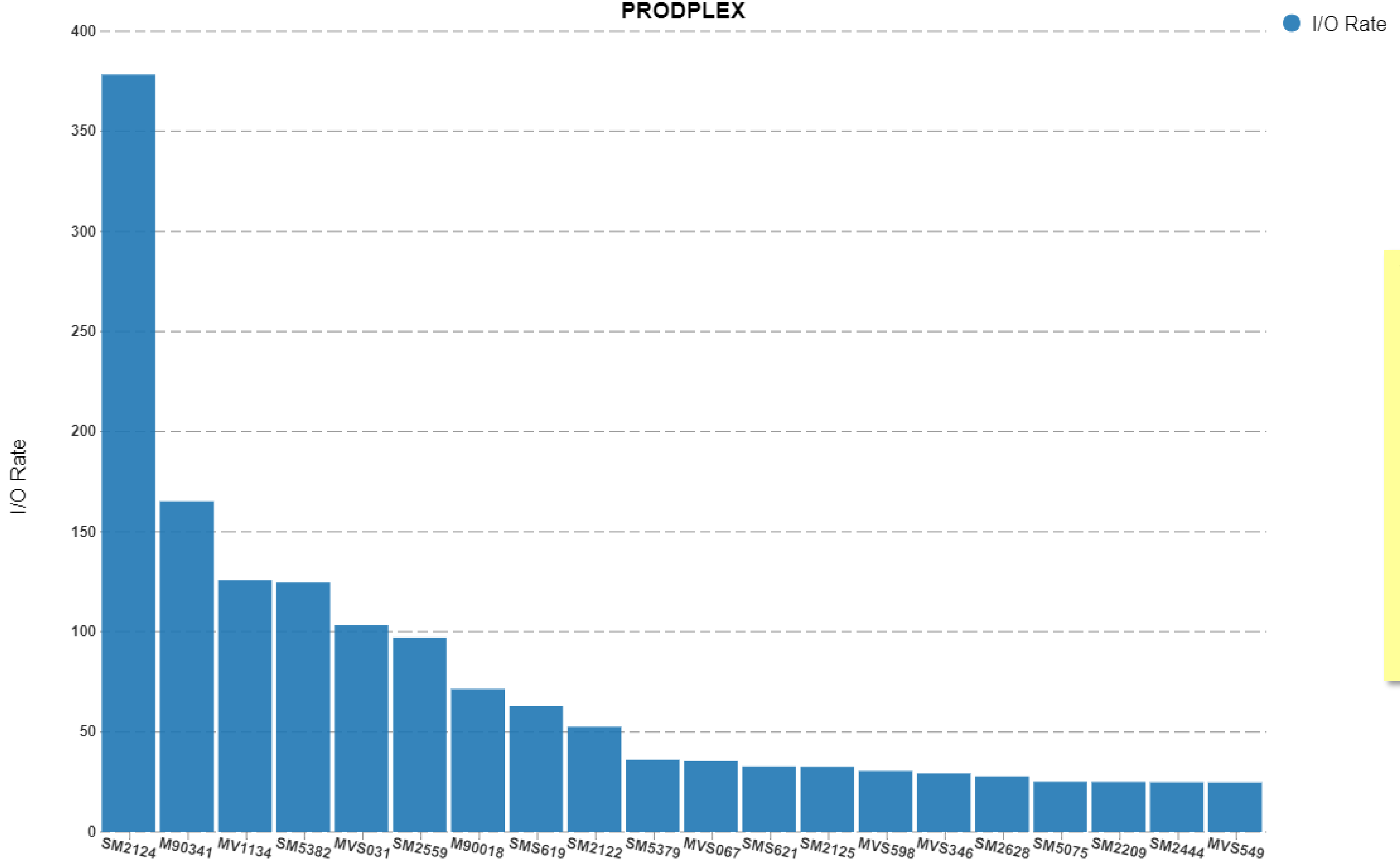
Memory and I/O

- We see systems with lots of memory free and yet they're doing significant amounts of I/O
- We've been saying for a long while things like "make your BPs bigger"
- But lately we've been trying to look deeper to point out opportunities
  - How much data is really on those busy volumes?
  - Which specific datasets are getting lots of read I/O

LVs with Highest I/O Rates

(Averaged Over Period of Study)

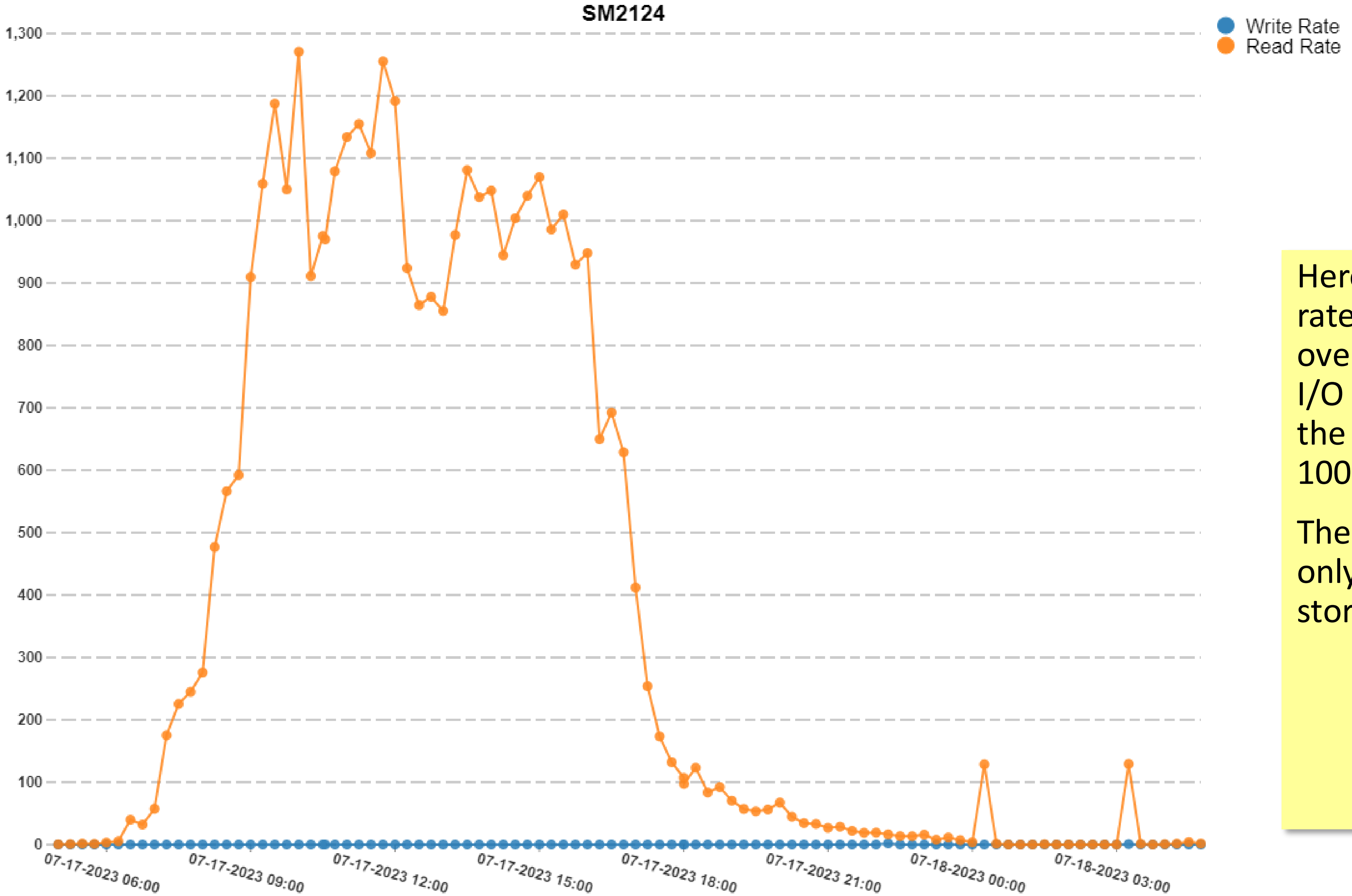
PRODPLEX



This Pivotor report shows the top volumes by I/O rate over the day.

375 IOPS doesn't sound too interesting but note that is an average I/O rate over 24 hours.

Logical DASD Volume Explorer

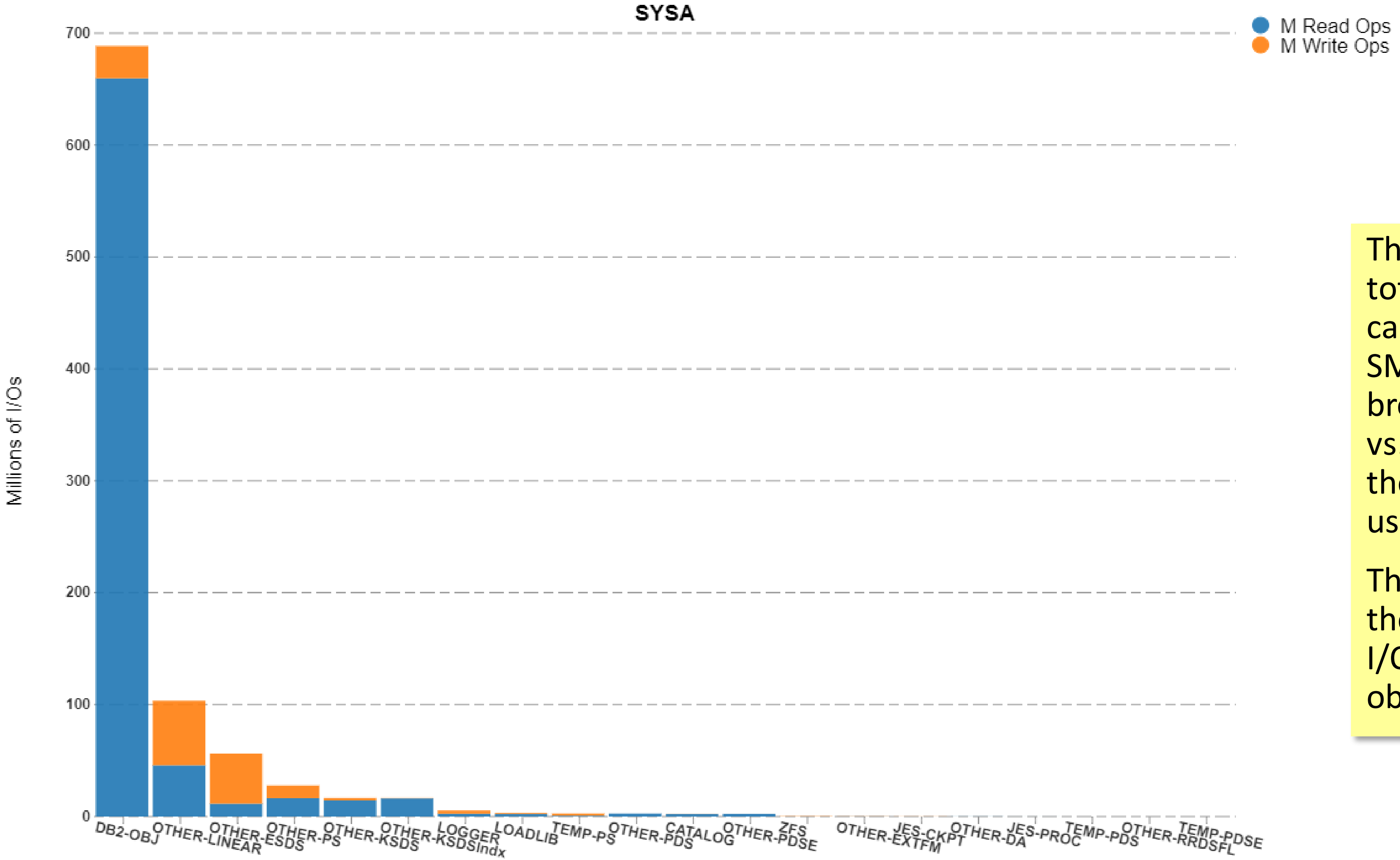


Here's the read and write rate for a particular volume over time. Virtually all the I/O is read I/O, and during the day it is doing over 1000 IOPS.

The kicker: this volume only has 1.5 GB of data stored on it!

Top Dataset I/O Counts by Dataset Usage

Total I/Os for Study Period



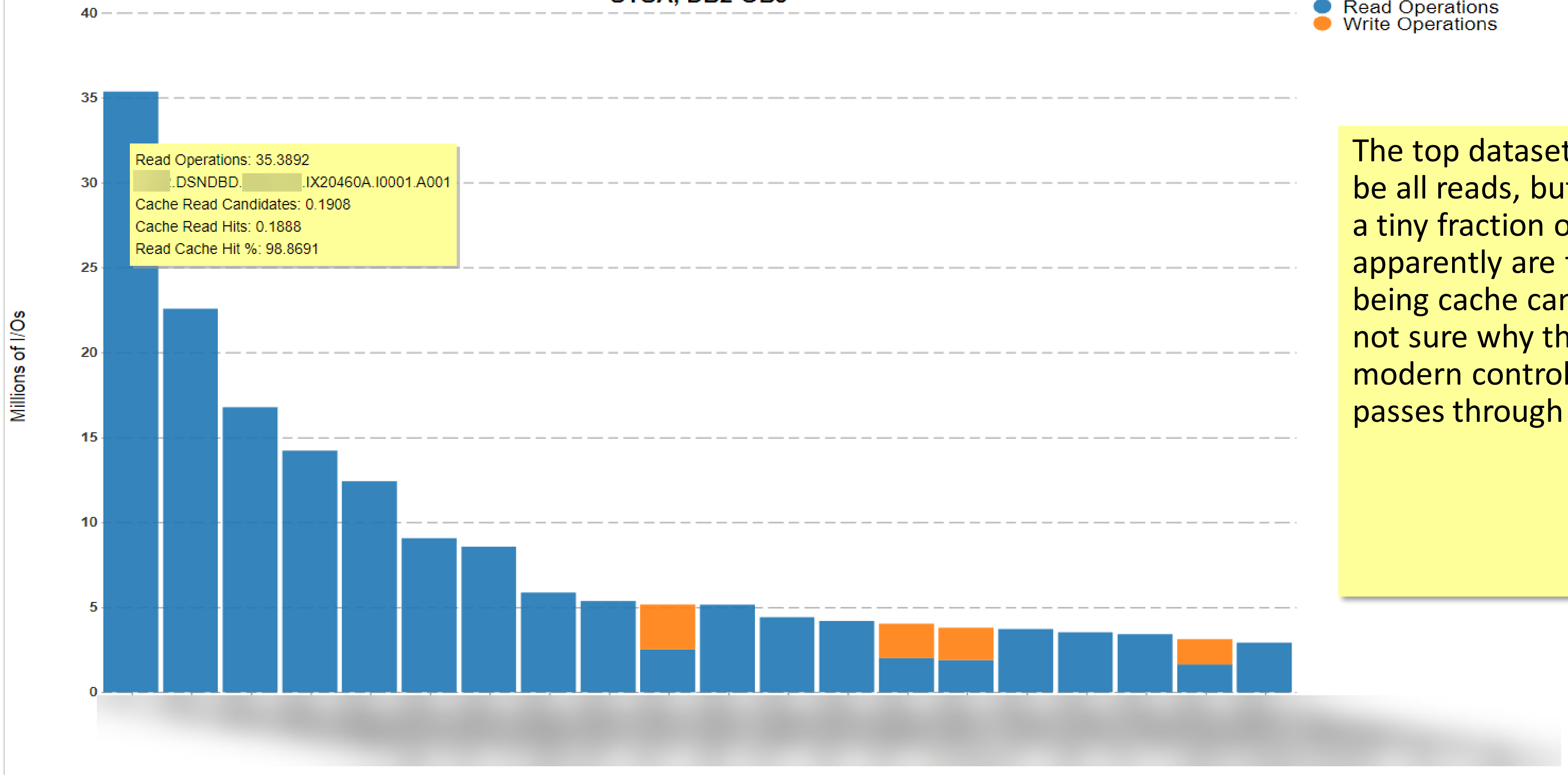
This reports looks at the total I/O over (in this case) a day from the SMF 42 records and breaks it down by reads vs. writes and by what the dataset is (probably) used for.

This site is not unusual: the vast majority of the I/O is reading from DB2 objects.

Top Dataset I/O counts by Dataset Usage

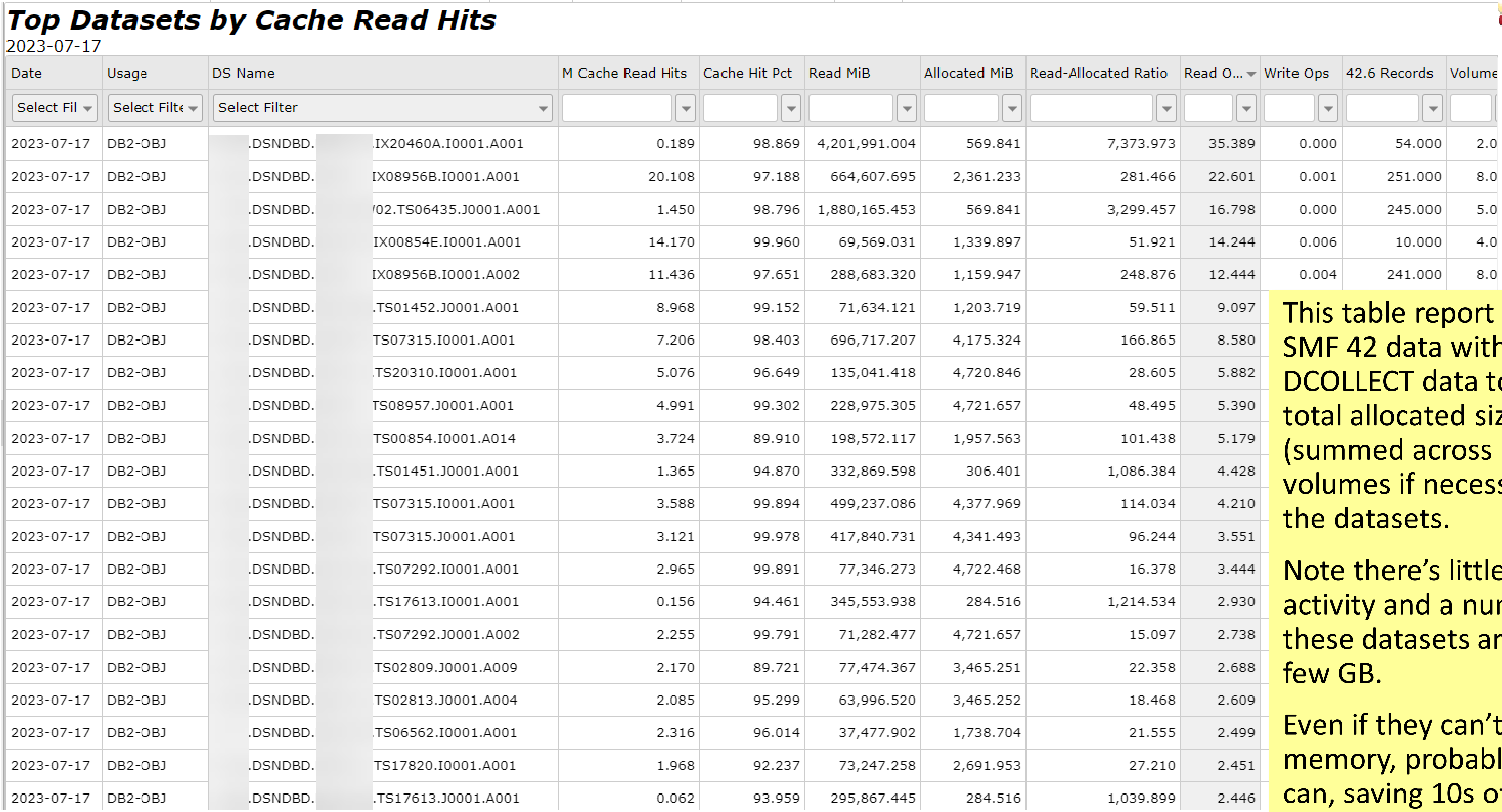
Total I/Os for Study Period

SYSA, DB2-OBJ



Read Operations: 35.3892
[redacted].DSNDBD.[redacted].IX20460A.I0001.A001
Cache Read Candidates: 0.1908
Cache Read Hits: 0.1888
Read Cache Hit %: 98.8691

The top dataset appears to be all reads, but oddly, only a tiny fraction of those apparently are flagged as being cache candidates. I’m not sure why that is. But in modern control units all I/O passes through cache.



Top Datasets by Cache Read Hits

2023-07-17

Date	Usage	DS Name	M Cache Read Hits	Cache Hit Pct	Read MiB	Allocated MiB	Read-Allocated Ratio	Read O...	Write Ops	42.6 Records	Volume
Select Fil	Select Filt	Select Filter									
2023-07-17	DB2-OBJ	[redacted].DSNDBD.[redacted].IX20460A.I0001.A001	0.189	98.869	4,201,991.004	569.841	7,373.973	35.389	0.000	54.000	2.0
2023-07-17	DB2-OBJ	[redacted].DSNDBD.[redacted]IX08956B.I0001.A001	20.108	97.188	664,607.695	2,361.233	281.466	22.601	0.001	251.000	8.0
2023-07-17	DB2-OBJ	[redacted].DSNDBD.[redacted]/02.TS06435.J0001.A001	1.450	98.796	1,880,165.453	569.841	3,299.457	16.798	0.000	245.000	5.0
2023-07-17	DB2-OBJ	[redacted].DSNDBD.[redacted]IX00854E.I0001.A001	14.170	99.960	69,569.031	1,339.897	51.921	14.244	0.006	10.000	4.0
2023-07-17	DB2-OBJ	[redacted].DSNDBD.[redacted]IX08956B.I0001.A002	11.436	97.651	288,683.320	1,159.947	248.876	12.444	0.004	241.000	8.0
2023-07-17	DB2-OBJ	[redacted].DSNDBD.[redacted].TS01452.J0001.A001	8.968	99.152	71,634.121	1,203.719	59.511	9.097			
2023-07-17	DB2-OBJ	[redacted].DSNDBD.[redacted]TS07315.I0001.A001	7.206	98.403	696,717.207	4,175.324	166.865	8.580			
2023-07-17	DB2-OBJ	[redacted].DSNDBD.[redacted].TS20310.I0001.A001	5.076	96.649	135,041.418	4,720.846	28.605	5.882			
2023-07-17	DB2-OBJ	[redacted].DSNDBD.[redacted]TS08957.J0001.A001	4.991	99.302	228,975.305	4,721.657	48.495	5.390			
2023-07-17	DB2-OBJ	[redacted].DSNDBD.[redacted]TS00854.I0001.A014	3.724	89.910	198,572.117	1,957.563	101.438	5.179			
2023-07-17	DB2-OBJ	[redacted].DSNDBD.[redacted].TS01451.J0001.A001	1.365	94.870	332,869.598	306.401	1,086.384	4.428			
2023-07-17	DB2-OBJ	[redacted].DSNDBD.[redacted]TS07315.I0001.A001	3.588	99.894	499,237.086	4,377.969	114.034	4.210			
2023-07-17	DB2-OBJ	[redacted].DSNDBD.[redacted]TS07315.J0001.A001	3.121	99.978	417,840.731	4,341.493	96.244	3.551			
2023-07-17	DB2-OBJ	[redacted].DSNDBD.[redacted].TS07292.I0001.A001	2.965	99.891	77,346.273	4,722.468	16.378	3.444			
2023-07-17	DB2-OBJ	[redacted].DSNDBD.[redacted].TS17613.I0001.A001	0.156	94.461	345,553.938	284.516	1,214.534	2.930			
2023-07-17	DB2-OBJ	[redacted].DSNDBD.[redacted].TS07292.J0001.A002	2.255	99.791	71,282.477	4,721.657	15.097	2.738			
2023-07-17	DB2-OBJ	[redacted].DSNDBD.[redacted]TS02809.J0001.A009	2.170	89.721	77,474.367	3,465.251	22.358	2.688			
2023-07-17	DB2-OBJ	[redacted].DSNDBD.[redacted]TS02813.J0001.A004	2.085	95.299	63,996.520	3,465.252	18.468	2.609			
2023-07-17	DB2-OBJ	[redacted].DSNDBD.[redacted].TS06562.I0001.A001	2.316	96.014	37,477.902	1,738.704	21.555	2.499			
2023-07-17	DB2-OBJ	[redacted].DSNDBD.[redacted]TS17820.I0001.A001	1.968	92.237	73,247.258	2,691.953	27.210	2.451			
2023-07-17	DB2-OBJ	[redacted].DSNDBD.[redacted].TS17613.J0001.A001	0.062	93.959	295,867.445	284.516	1,039.899	2.446			

This table report joins the SMF 42 data with the DCOLLECT data to get the total allocated size (summed across multiple volumes if necessary) of the datasets.

Note there's little write activity and a number of these datasets are only a few GB.

Even if they can't all go into memory, probably some can, saving 10s of millions of I/Os.

How will AI change what we do?

Scott’s AI Thoughts

- There’s going to be a lot of interesting applications for AI over the next several years
  - Most of which have nothing to do with managing z/OS performance
  - Interesting questions and uncertainty in the realms of ethics, legal liabilities, and potential regulation for at least some use cases
- z/OS performance analysts are not going to be put out of a job tomorrow
  - There’s a lot of exterior factors that come into play in managing a system that is not captured in the performance data about the system
    - Not all dispatching priority inversions are bad, not all “bad” goals are wrong
    - Sometimes we intentionally restrict performance for various reasons
  - Anybody(thing) evaluating your system should be asking “what” and “why” and explaining “what” and “why” as well!
- AI Code generation works surprisingly well and can make us more efficient

Batch Management

- "AI-powered Workload Manager (WLM), designed to intelligently predict upcoming batch workload and react accordingly to optimize system resources in a proactive way. This AI capability represents the first use case that leverages the AI Framework for IBM z/OS." (IBM announcement)
- But … predicting upcoming batch workloads and proactively managing initiators has been a thing in the past without AI
  - E.G. ThruPut Manager Automation Edition
  - And z/OS Performance Analysts have been doing this with Actual Intelligence
- Nonetheless, this is an interesting area to explore and may be useful
  - Given how reluctant people were to move to WLM-managed inits… it will be interesting to see the uptake on AI-managed initiators!
    - Recent "survey": about half of all plexes had some WLM-managed inits, and in those plexes, about half of the job classes were WLM-managed

Continuing Questions & Ongoing Opportunities

Things we’re talking about with people

XCF Transport Class Simplification

z/OS v2.4 XCF Transport Class Simplification

- z/OS 2.4. Eliminates the need to define size only transport classes
  - Segregation of messages purely by size
  - XCF transport classes more self-managing and self-tuning
    - No longer need to tune and optimize XCF transport classes message sizes to match the signaling workload characteristics
    - Also results in decreased number of path definitions, etc.
  - No longer static definition for assignment of resources
    - System automatically applies resources where needed
  - Avoid performance and resiliency impacts from poorly-tuned transport class sizes
    - Also, improve resiliency by avoiding monopolization of message buffer space
  - New/improved statistics for reporting message path utilization, signal counts, and no-buffer conditions

- Later planned support will address group segregation
  - Isolation of ill-behaved members to avoid sympathy sickness
  - One member will not negatively impact signal delivery of other members

New _XCFMGD (pseudo) transport class

New control XTCSIZE to enable/disable new support

- Basically, a chicken switch
- When set to DISABLED, XCF signaling resources are managed as they were prior to z/OS 2.4
- When set to ENABLED, _XCFMGD transport class used
- Can disable or enable the XTCSIZE switch dynamically with the SETXCF FUNCTIONS operator command

```
SETXCF FUNCTIONS,DISABLE=XTCSIZE
SETXCF FUNCTIONS,ENABLE=XTCSIZE
```

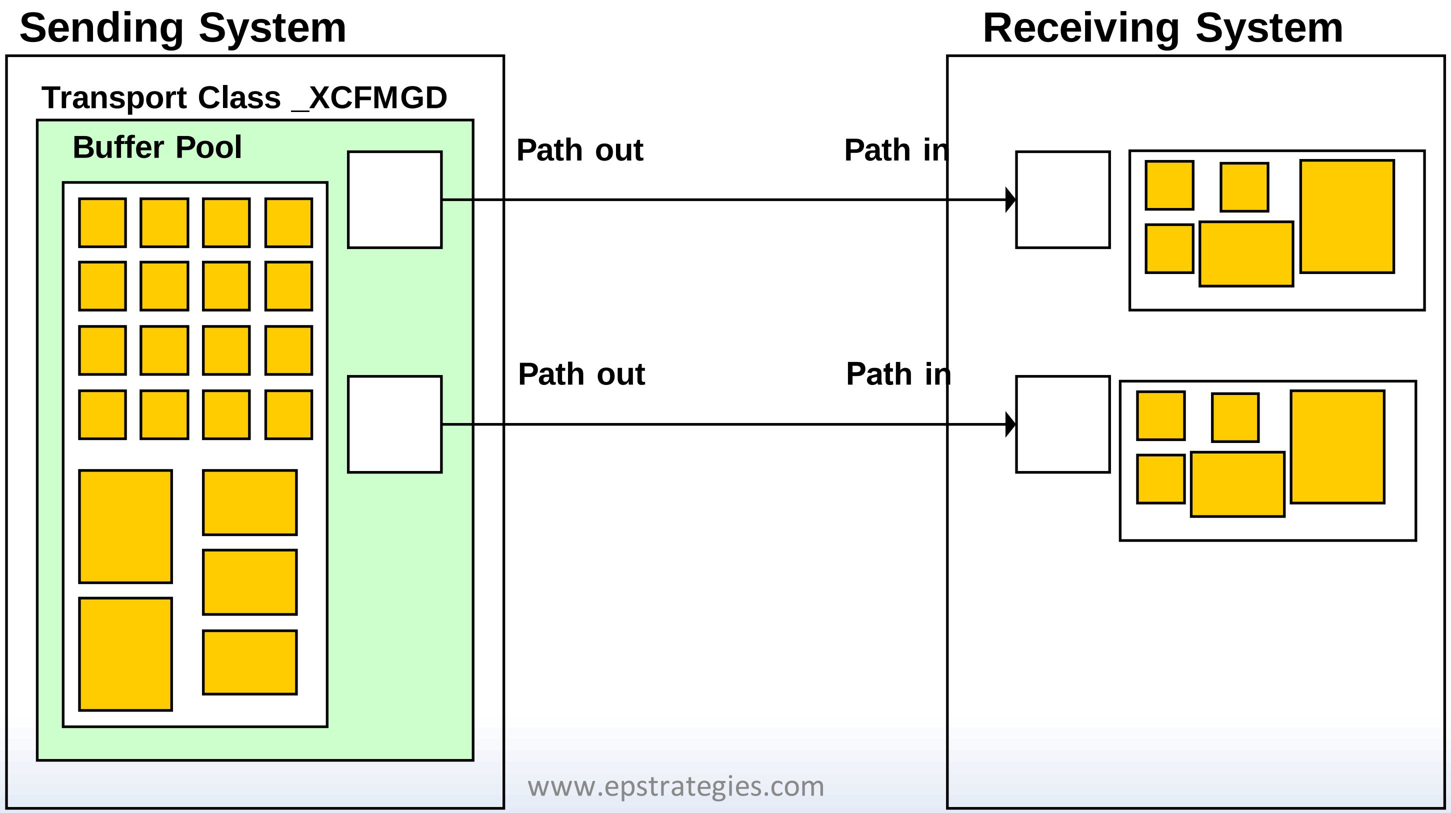
New _XCFMGD (pseudo) transport class in COUPLExx member of parmlib

- Implicitly defined by XCF (thus, it always exists)
  - Will not be used if XTCSIZE is DISABLED or if target system is pre-z/OS V2R4
  - Installation cannot directly control its attributes (classlen=0, XCF determines MAXMSG)
  - When XTCSIZE is ENABLED, all paths in the “XCF Managed” classes are logically reassigned to the _XCFMGD transport class
- Algorithm uses the “best fit” buffers on the send side
  - Maximizes number of signals that can be accepted for a given MAXMSG limit to better handle bursts of activity and delays
  - As a reminder, traditional classes generally use the “defined size” which might not be best fit
- Paths run at the maximum signal size
  - Thus, any message can be transmitted without any additional overhead
    - Never need to re-negotiate signal size (or tune) the signal paths

Example of using _XCFMGD

Sending System
- Obtain an outbound message buffer
- Select path on which to send message
- Initiate transfer of message over path to target system

Receiving System
- Receives message on inbound path
- Places message into inbound buffer
- Coordinates message to be delivered to target member (i.e. the application)



Migration notes:

- May need to maintain your old COUPLExx XCF definitions
  - It is very likely that when migrating to z/OS 2.4, not all systems in the Sysplex will be migrated at the same time
  - Thus, it is very possible that during migration to z/OS 2.4 that some systems in the Sysplex will be back=level
- The traditional transport class definitions intended to manage size segregation should be maintained until all systems in the Sysplex are running z/OS V2R4
- Lesson, do not remove all setup from COUPLExx member until all systems are migrated
  - In addition, keeping the old definitions will allow for selective XTCSIZE disablement

COUPLExx member changes

Pre-z/OS 24

```
CLASSDEF CLASS(DEFAULT) CLASSLEN(956)
CLASSDEF CLASS(MSG08K)  CLASSLEN(8124)
CLASSDEF CLASS(MSG16K)  CLASSLEN(16316)
CLASSDEF CLASS(MSG24K)  CLASSLEN(24508)
CLASSDEF CLASS(MSG32K)  CLASSLEN(32700)

PATHIN  STRNAME(IXCSIG1,IXCSIG2) MAXMSG(2000)
PATHIN  STRNAME(IXCSIG3,IXCSIG3B,IXCSIG4,
                IXCSIG5,IXCSIG6)

PATHOUT STRNAME(IXCSIG1,IXCSIG2) CLASS(DEFAULT)
PATHOUT STRNAME(IXCSIG3)         CLASS(MSG08K)
PATHOUT STRNAME(IXCSIG3B)        CLASS(MSG08K)
PATHOUT STRNAME(IXCSIG4)         CLASS(MSG16K)
PATHOUT STRNAME(IXCSIG5)         CLASS(MSG24K)
PATHOUT STRNAME(IXCSIG6)         CLASS(MSG32K)
```

z/OS 2.4 +

```
PATHIN  STRNAME(IXCSIG1,IXCSIG2) MAXMSG(2000)
PATHIN  STRNAME(IXCSIG3,IXCSIG3B,IXCSIG4,
                IXCSIG5,IXCSIG6)

PATHOUT STRNAME(IXCSIG1,IXCSIG2)
PATHOUT STRNAME(IXCSIG3)
PATHOUT STRNAME(IXCSIG3B)
PATHOUT STRNAME(IXCSIG4)
PATHOUT STRNAME(IXCSIG5)
PATHOUT STRNAME(IXCSIG6)
```

(Or could just leave the definitions alone, and XCF will ignore if XTCSIZE is enabled)

Other comments / notes

SRB Updates (and SMF 30s)

- Continue to see customers not leveraging System Recovery Shutdown Boost
  - Does not get invoked automatically, you have to update your procedures
  - Maybe shutdown time is not a pain point for most sites?
- Initial problem with SMF 30 and SRB was that if you weren’t syncing your intervals, you wouldn’t get new SMF 30 (and presumably others) interval records for boost periods
  - Looks like that’s now fixed
- New problem observed:
  - During IPL boost, rarely, some SMF 30 records may have incorrect interval end times (such as before the interval begin!) and may have multiple records written
  - Most such records seem to contain little to no utilization though
  - Is relatively infrequent, but have seen it across multiple customers
- As always, exclude boost periods from performance analysis!

SuperPAV

- IOSQ time is rarely a significant component of I/O response time, but we still sometimes see some
- SuperPAV generally eliminates the little remaining IOSQ time
  - SuperPAV enables sharing of PAVs between LCUs, effectively allowing access to more PAVs for each volume
- If your DASD is less than even 5 years old, it almost certainly supports SuperPAV
  - Check with your DASD vendor and enable in IECIOS: HYPERPAV=XPAV
  - Can be done dynamically, so easy change

I/O Priority Management

- A few (several?) years ago we made the recommendation that most customers should disable I/O Priority Management in WLM
  - Recommendation had been for ~20 years to enable it
  - Changing reality of I/O meant that having it enabled inflated velocities
- At the time we said probably 90% of sites shouldn’t have it enabled
- Having seen even more data over the years, that’s probably now >99%
  - It makes WLM focus on just CPU using and delays
  - May have to revisit/reset your velocity goals when you do this though
  - “Worst” case is that turning it off makes no difference
- IBM is also now recommending to turn off I/O Priority Management
  - Except the manual hasn’t been updated 🤦

Record the 98s and 99s

- They provide insights into performance at a sub-minute level
    - 10 second WLM Policy Adjustment interval
    - 2 second HiperDispatch interval
    - 5-60 second High Frequency Throughput Statistics
- Yes, you're not going to look at them every day, but they can be quite useful for problem determination: especially for transient problems!

SMF 98/99 records to Include

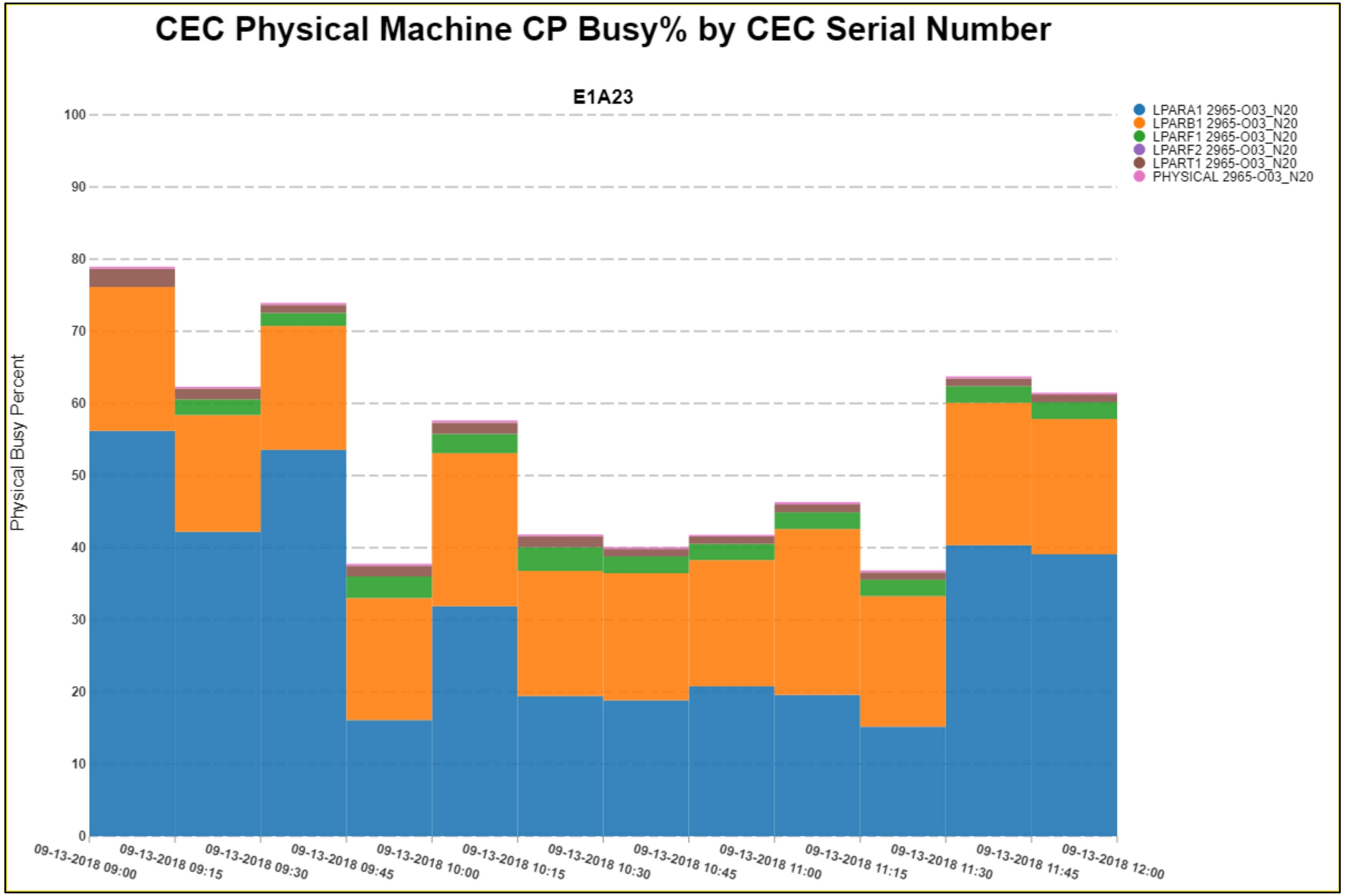
```
In SMFPRMxx:

HFTSINTVL(15)
```

- SMF 98 High-frequency Throughput Statistics (HFTS)
  - IBM recommendation is to record on 5 second interval
    - Can use 5, 10, 15, 20, 30 or 60 seconds
    - 5 second interval is about 400MB-500MB/system/day
- SMF 99 SRM/WLM details
  - Our minimum recommended subtypes: 6, 10, 11, 12, 14
    - These will be around 50-150MB/system/day
  - Subtype 1, 2, and 3 can be quite useful, but can be more voluminous
    - These can be 1-1.5GB/system/day
  - Pivotor customers: send them if you're collecting them!
  - Subtype 13 is somewhat voluminous but is undocumented "IBM use only"
    - 150-200MB/system/day
    - We recommend you turn off subtype 13s until/unless IBM asks for them

None of these records represent data you will look at every day, but it's nice to have them available when you need them!

Classic CEC Utilization Transient Performance Problem



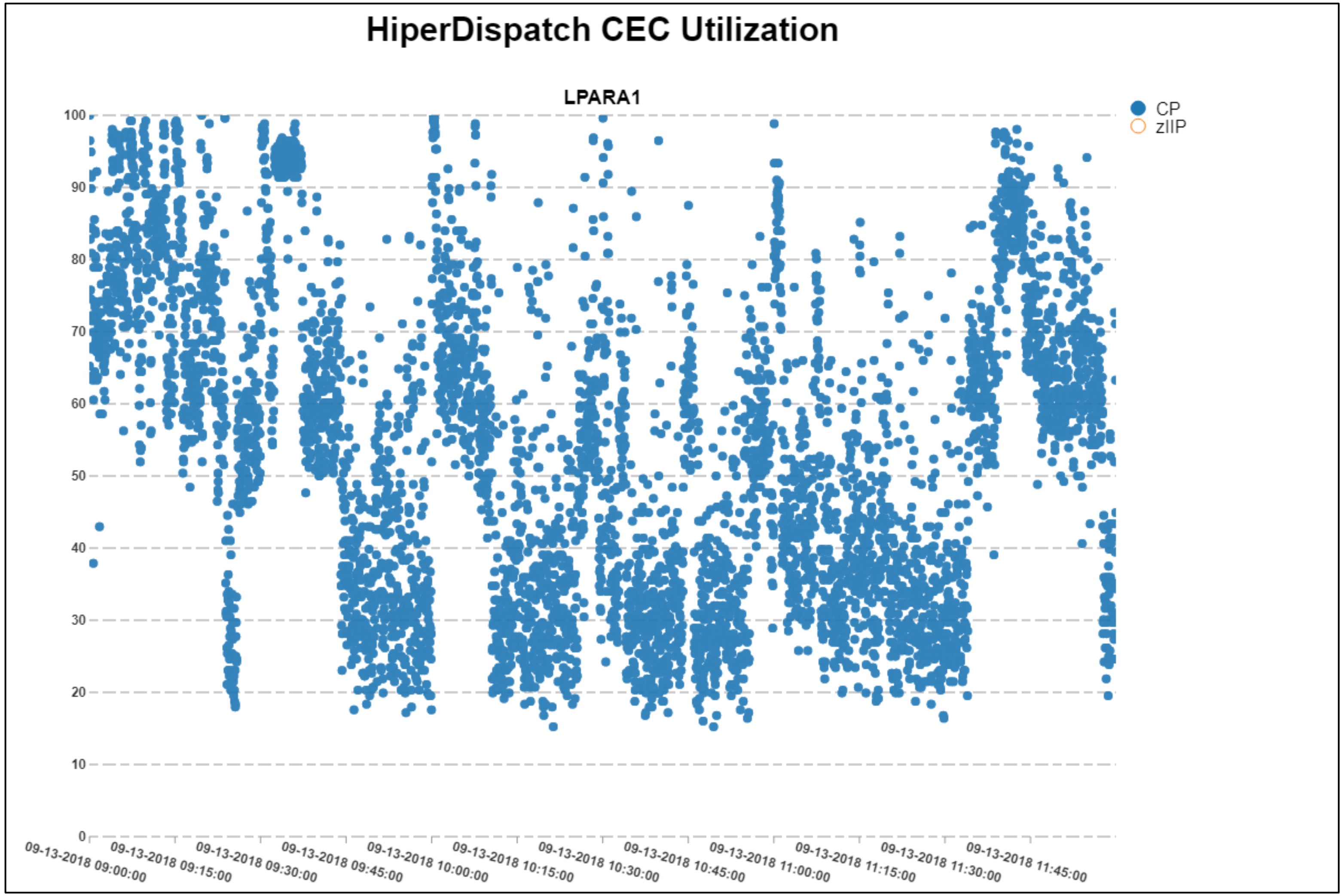
Problem Statement:

System Seemed to Freeze / Stall / things too a long time, but we have lots of available capacity

This is just a standard view of CEC Utilization, here we've narrowed in to just 3 hours in the morning, where it doesn't appear there's really any capacity concerns.

This chart is generated from data that comes from the SMF 70 records. In this example, the measurement intervals are 15 minutes.

Classic CEC Utilization Transient Performance Problem



High Frequency CEC Utilization:

This also is a CEC utilization chart for the same 3 hours as the previous chart.

This data comes from the from the SMF 99.12 HyperDispatch records.

The CEC utilization is at 2-second measurement interval.

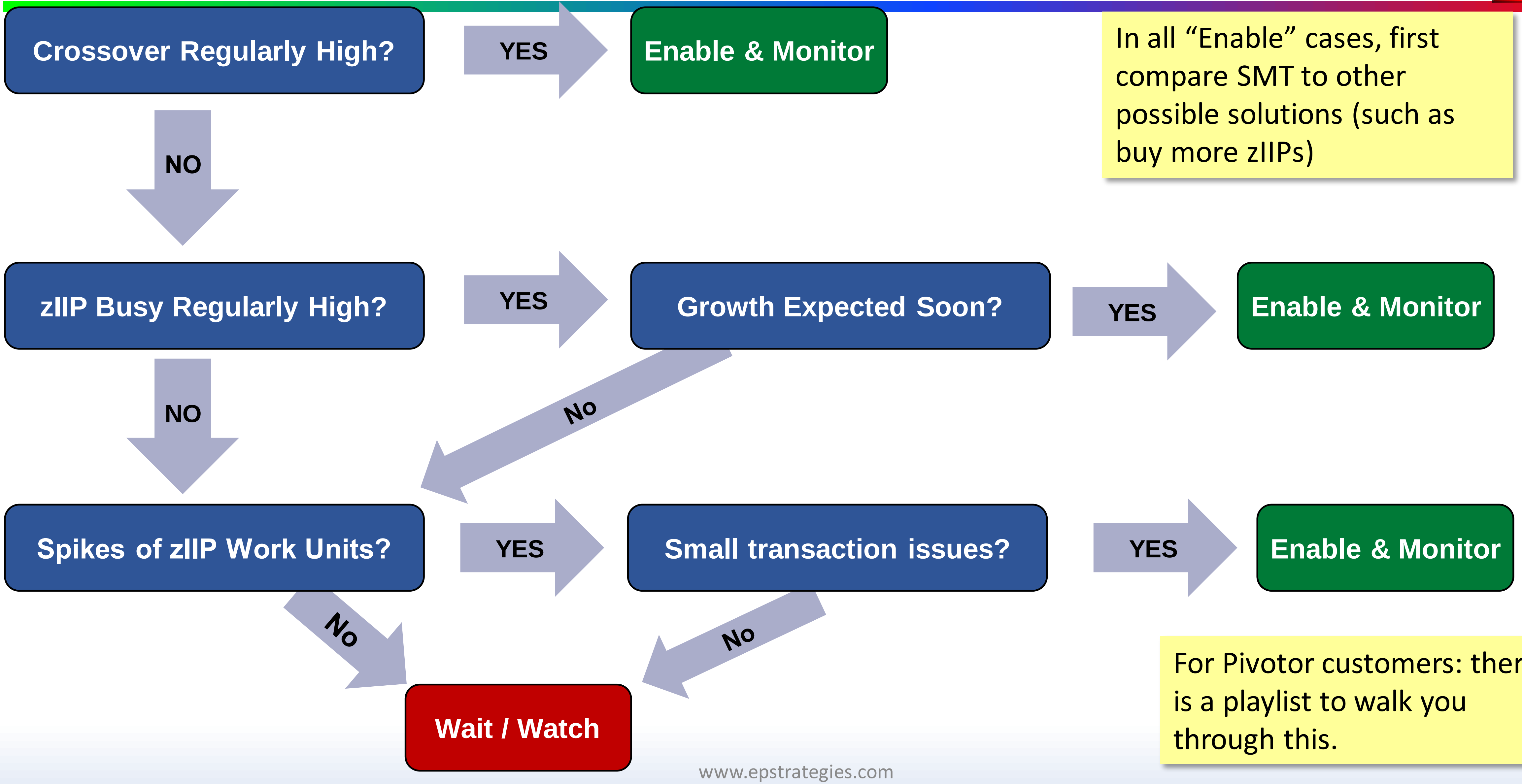
Note that this tells a different story than the 15-minute RMF intervals.

SMT

- Should I enable SMT?
  - Probably not (but sometimes, yes)
- We sometimes see customer with SMT enabled “just because”
  - That’s probably “ok” but it’s probably also unnecessary
- In some cases, unnecessary use of SMT might be sub-optimal
  - Remember z/OS densely packs the cores so even if you have a relatively high number of unused zIIP cores, with SMT enabled the work will be assigned to an in-use core first
- Our general recommendation: only enable SMT when actually needed
  - Leave SMT in your bag of tricks ready to be used when the need develops
  - SMT also makes detailed capacity planning for zIIPs effectively impossible.
- See also Scott’s SMT presentation on our website

https://www.pivotor.com/content.html

SMT Enablement Flowchart



In all “Enable” cases, first compare SMT to other possible solutions (such as buy more zIIPs)

For Pivotor customers: there is a playlist to walk you through this.

Wrap-up

- We hope you enjoyed this and that you’ve learned something
- Let us know if you like this potpourri of topics format
- We’ll be around now and all week for questions

- Questions?

- Please visit our website: www.epstrategies.com
  - Past presentations
  - WLM to HTML tool
  - More information about Pivotor
  - Future educational webinars