

IBM Worldwide Systems Center Performance Checkup Recommendations

February 23rd, 2026
1:15pm-2:15pm

Stephanie DeLuca

Stephanie.Deluca@ibm.com

IBM Z Performance Specialist, IBM Worldwide Systems Center (WSC)

Kwasi Abrefa-Kodom

Kwasi.Abrefa-Kodom@ibm.com

IBM Z Performance Specialist, IBM Worldwide Systems Center (WSC)

About the IBM Worldwide Systems Center

IBM Z Worldwide Systems Center (WSC)

Drive IBM Z platform adoption, growth and retention
with deep technical expertise in new and existing IBM Z capabilities

The WSC experts provide SME support and act as consultative, trusted advisors to IBM Z clients. Our experts establish best practices and develop scalable, reusable demos, tools, methodologies and enablement content for IBM Z. We showcase and use these assets to upskill the IBM Z community by mentoring, instructing, and advising clients, business partners and IBMers. Our SMEs are key systems partners for product development including Early Support Programs and critical situation support. We also perform economics analysis, execute competitive performance benchmarks and comprehensive health checks.

To engage the IBM Z WSC
contact your local IBM account team
or IBM Business Partner



Agenda

Session Objectives

Value of Proactive Performance Management

z/OS Performance Checkup

- CPC Utilization

- Workload Utilization and Growth

- LPAR Configuration

- z/OS Capture Ratio

- zIIP Usage and IIPCP Time

- Memory Allocation and Available Storage

- WLM Performance Index

Summary

Session Objectives

Discuss how performance directly ties to business outcomes

Recognize the value of proactive performance management

Identify and understand key performance metrics that impact z/OS efficiency

Leave with an actionable checklist and techniques to apply in your environment

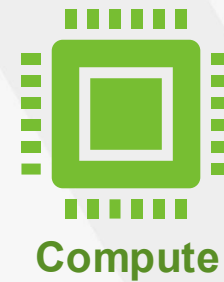
What is Performance Management for IBM Z

The goal of performance management is to **manage and monitor existing resources** to best meet performance objectives and Service Level Agreements (SLAs)

- Online peak transactions per second or load
- Transaction response time
- Batch window requirements for throughput and elapsed time

Involves **monitoring, analyzing, and reporting** on system KPIs including CPU usage, batch volumes, transaction load, and response time components

Resources include:



Monitoring includes:



Performance Tuning

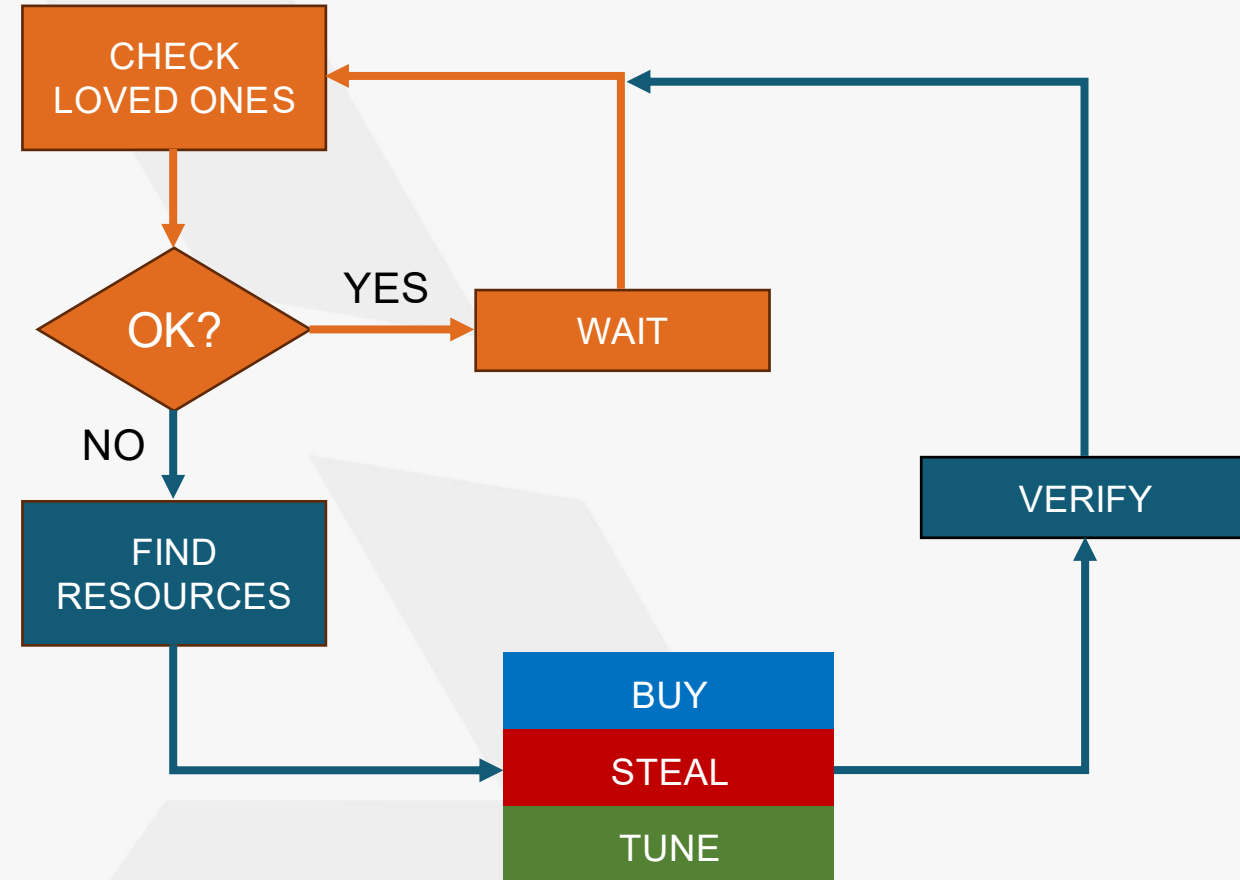
Performance tuning is the process of **optimizing the configuration and system parameters to improve overall performance and efficiency** of workloads and applications

- Especially to ensure that business goals, availability objectives, and SLAs are consistently met

Tuning involves:

- Monitoring, analyzing, and reporting on data from Systems Management Facilities (SMF)
- When there's a problem, identifying bottleneck(s)
- Adjusting parameters one at a time
- Verifying results of the change

Buy, Steal, or Tune?



Value of Proactive Performance Management



Maximize the efficiency of limited system resources
(Do more with less, less saves \$\$)



Identify and resolve potential issues before they impact availability



Support more accurate capacity planning and long-term growth



Enable faster troubleshooting and problem diagnosis, reducing MTTR



Shift away from reactive, crisis-driven work toward strategic performance management



Z/OS PERFORMANCE CHECKUP

Performance Check #1: CPC Utilization

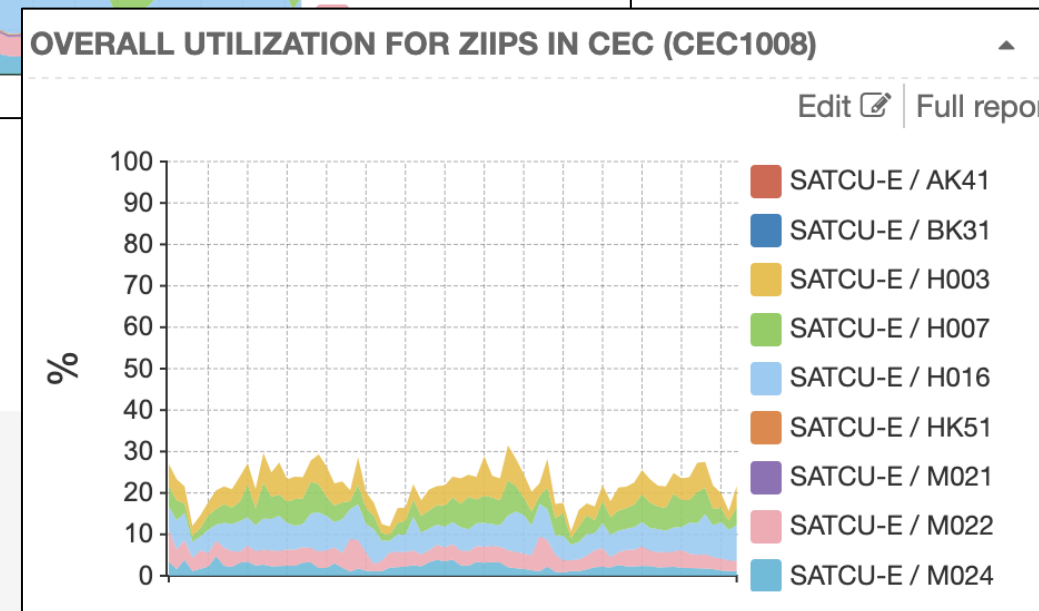
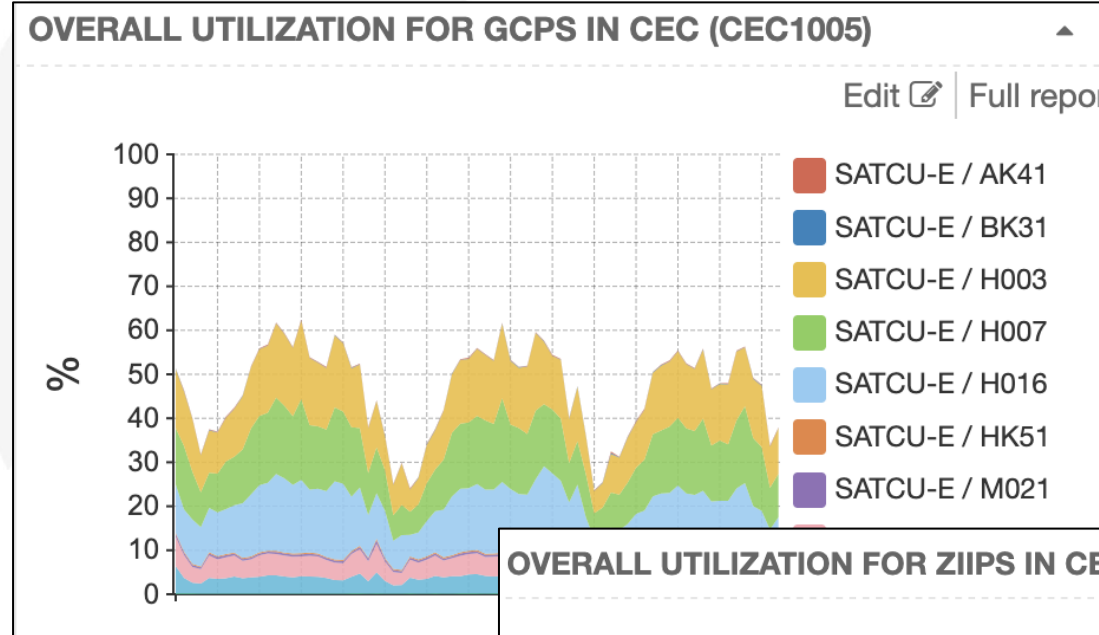
How busy is each engine pool

Establishes a baseline

- Is the CPC lightly utilized with plenty of capacity, or running at higher utilizations where CPU contention and certain performance pressures are more likely
- Tip: *High* utilization will be different for everyone!

Identifies which LPARs are driving the busy, helping identify those with *potentially* greater influence on overall performance

Reveals long term utilization and growth trends that are critical for informed capacity planning decisions

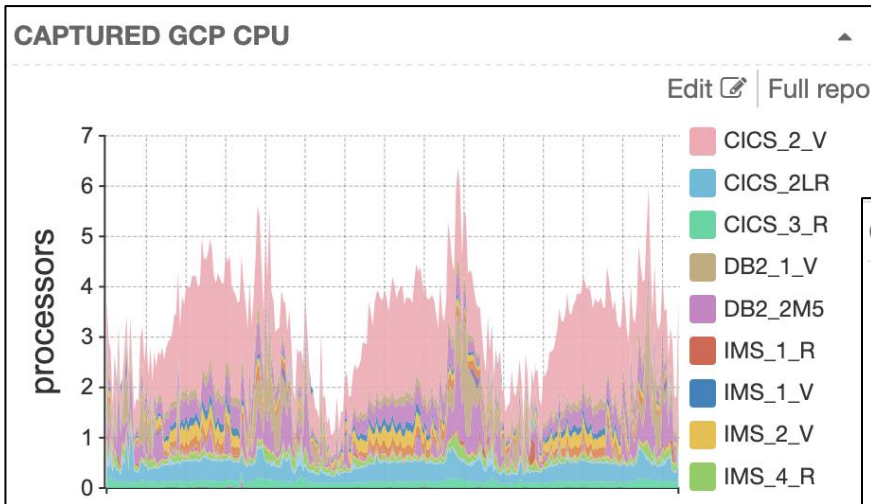


Performance Check #2: Workload Utilization

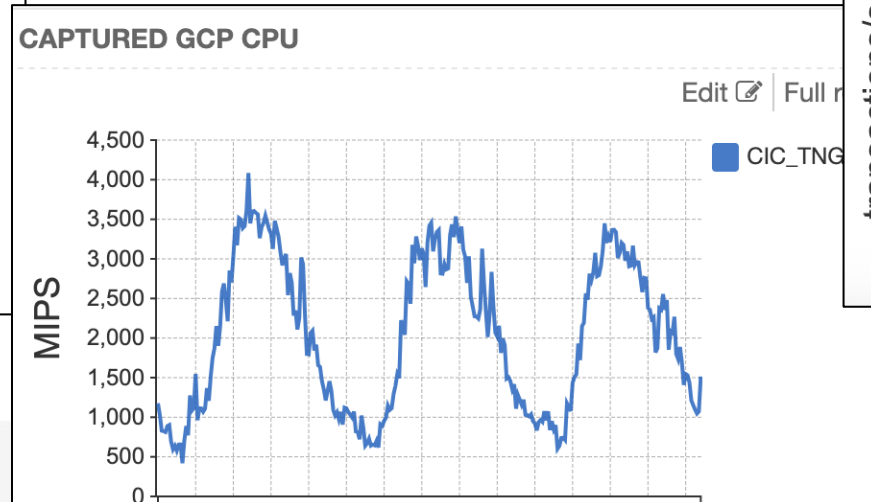
Service or report class utilization and transaction rates

Can then drill down into which workloads are driving the CPU busy

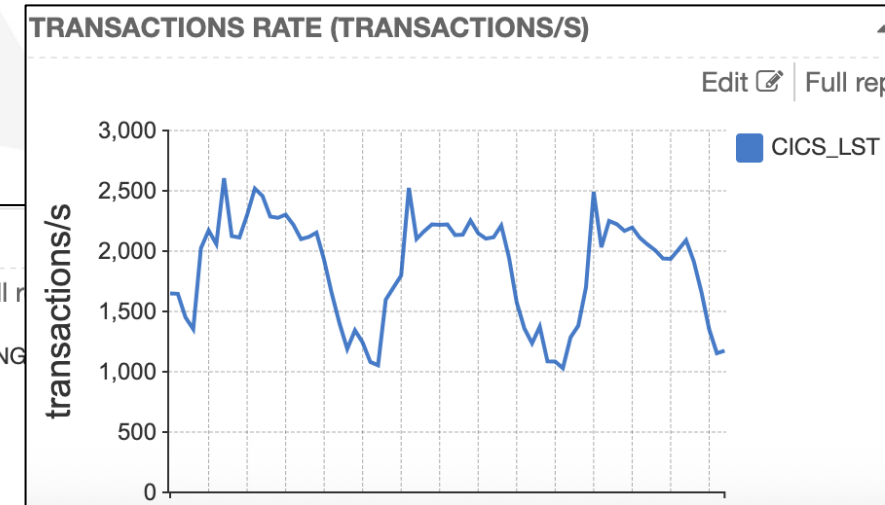
- Are there primary batch vs online LPARs? Which workloads run during peak busy?
- Identify unexpected spikes, peak periods, and long-term growth trends
- Use report classes, transaction rates, and/or jobs for further granularity



All service classes



Report class CIC_TNG2 contributing to CICS_2_V



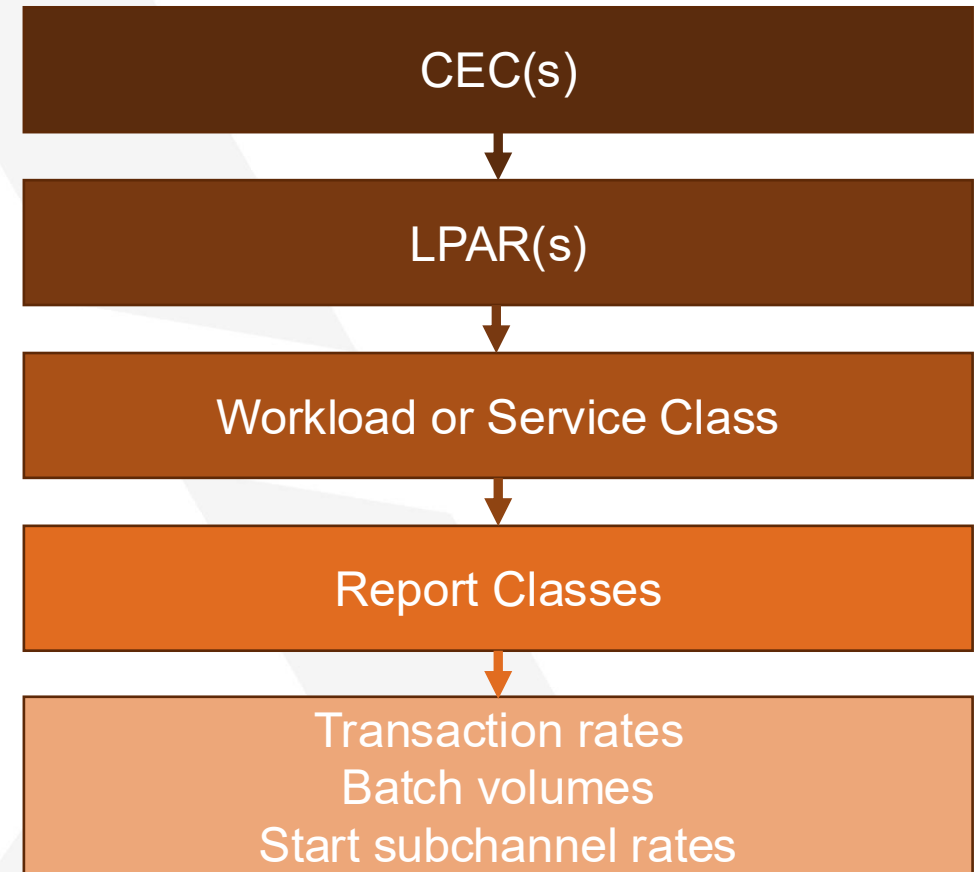
CICS transaction rate

Performance Check #1-2: CPC and Workload Utilization

Actions

Use a top-down approach to get a complete picture of system utilization and identify what 'normal' is

- Monitor and report long term utilization and growth at the CPC, LPAR, and service or report class levels
 - *Critical* for successful capacity planning
 - Tip: Use report classes as much as possible for more granular data and reporting
- Monitor and report on other load indicators
 - Transaction rates, batch volumes, start subchannel rates
 - Demonstrate to the enterprise how much work is being done and identify growing business demand
- Enable alerting for workload spikes that require attention



Performance Check #3: LPAR Configuration

Setting LPAR weight and Logical CPs

In a shared z/OS environment, the distribution of processor capacity to your LPARs is determined by the **LPAR weights** and **logical CPs**

- Defined in the Hardware Management Console (HMC) for each engine pool (GCPs, zIIPs, IFLs)
- Used to ensure each LPAR has sufficient capacity to meet performance objectives of their workloads
- Will be based on the importance of work running on that LPAR

LPAR weight:

used to set guaranteed processor capacity

LPAR fair share:

weight converted to a percent of total capacity

Engines by weight:

Fair share converted to physical engines

Logical CP count:

Maximum # of engines that could be used if other LPARs are not using their full share

IBM z16 709

LPAR name	LPAR Weight*	Fair share %	GCPs by weight	Logical CPs*
SYS1	700	70%	6.3	9
SYS2	300	30%	2.7	5
Total	1000	100%	9.0	14

*Defined in HMC, helpful if weight adds up to 1000

*Only LPARs with shared CPs are used in calculation

Performance Check #3: LPAR Configuration

Setting LPAR weight and Logical CPs

Best Practice

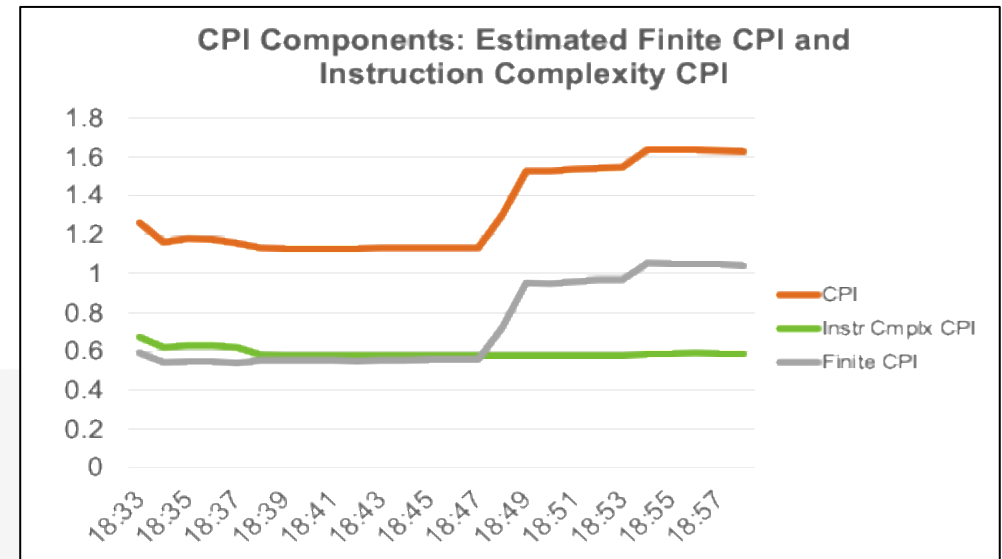
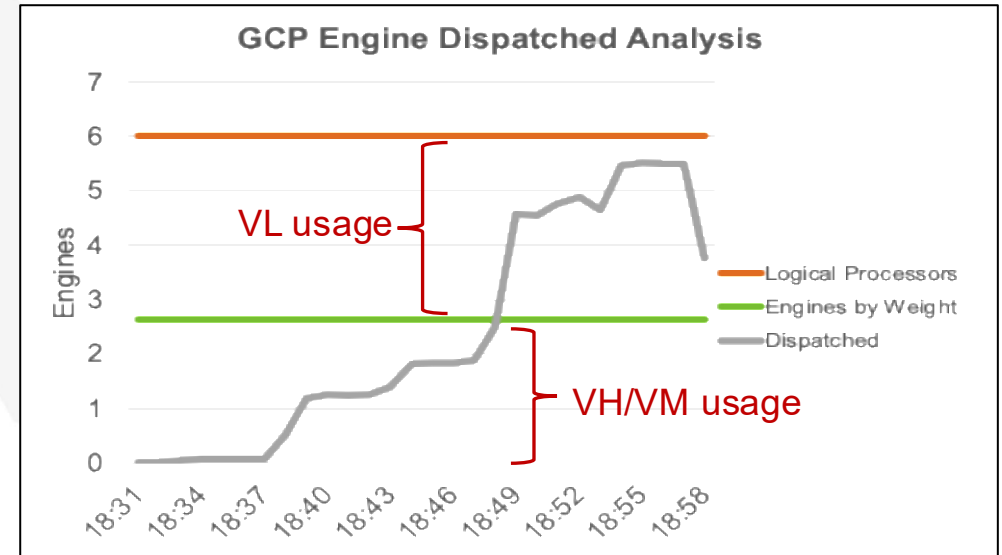
Weights should be set to meet capacity requirements, based on business goals

Logicals should be set to meet the weight, plus between 0-2 additional logicals based on LPAR size and appropriate whitespace

- Do not define all logicals on the CEC to each LPAR
- Optimize for Vertical Highs (VH)
- Define at least 25% whitespace (small LPARs) up to two additional logicals to handle spikes

Why?

1. Work runs most efficiently when running within your defined weight using VHS and VMs
 - CPI == Efficiency == MIPS == \$\$
2. Vertical Lows are *not guaranteed*
 - If your production LPAR(s) are relying on them, and the system gets busy, there will be performance impact
3. Several more reasons to limit logicals in Appendix A



Performance Check #3: LPAR Configuration

Setting LPAR weight and Logical CPs

Do the LPAR definitions match business priorities and follow best practices for efficiency?

Actions

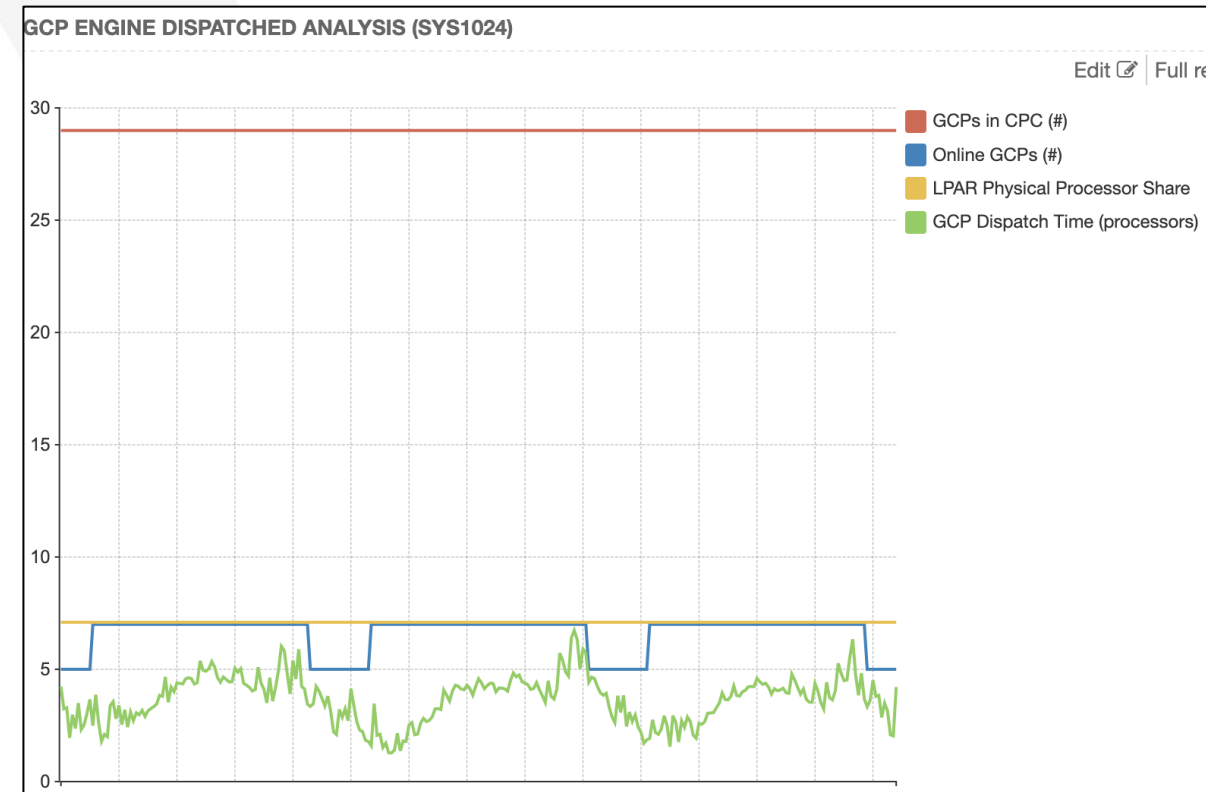
Review LPAR weights compared to actual usage

- Evaluate LPARs consistently using more than their weight
 - If other LPARs get busier and start using their share, am I okay with the impact to these LPARs?
- Evaluate LPARs consistently using far less than their weight
 - Does someone else need it more?
- Hint: Chart shows dispatch time, fair share, and logicals all in terms of engines

Follow best practice in optimizing for Vertical Highs and limiting total logicals

Consider dynamic weight adjustments for workloads where priorities shift

- Ex. Online LPAR may be more important during the day where batch LPAR is more important at night



Performance Check #4: z/OS Capture Ratio

z/OS capture ratio is a common indicator of overall system health

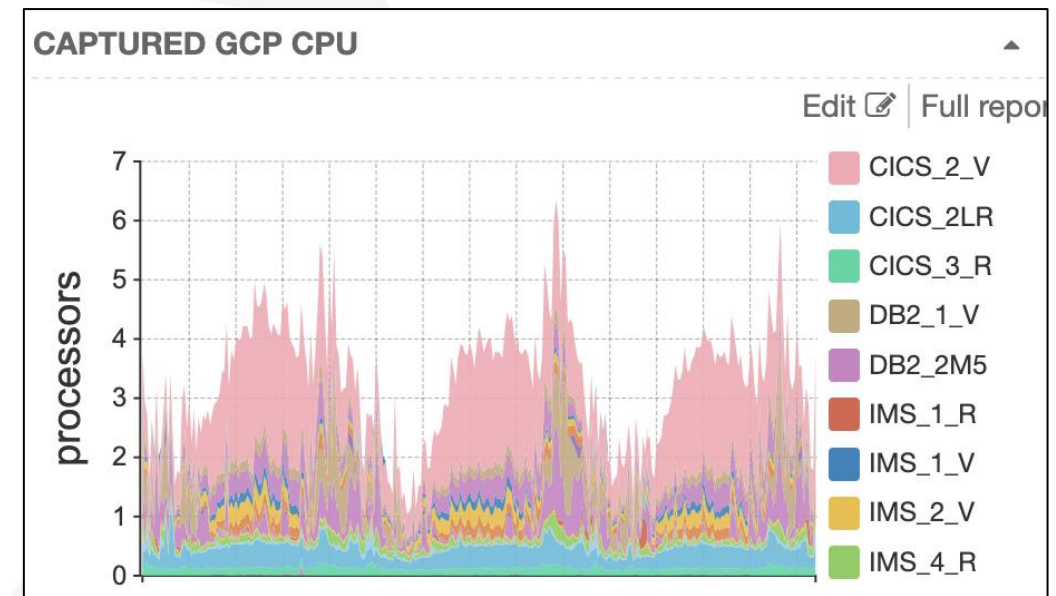
$$= \frac{\text{Captured CPU time}}{\text{Captured + uncaptured CPU time}} = \frac{\text{CPU time that can be attributed to a workload}}{\text{Total dispatched CPU time}} = \frac{\text{SMF 72 (Workload activity data)}}{\text{SMF 70 (Total dispatched CPU time)}}$$

Uncaptured time on z/OS includes necessary system work that can't be directly attributed to one specific workload, transaction, or job

- There will always be some uncaptured time

Typical capture ratio on a well-tuned system should be above 88%

- A lower capture ratio indicates higher than normal system overhead
- *Note:* Low utilization periods may show lower capture ratios on otherwise healthy systems



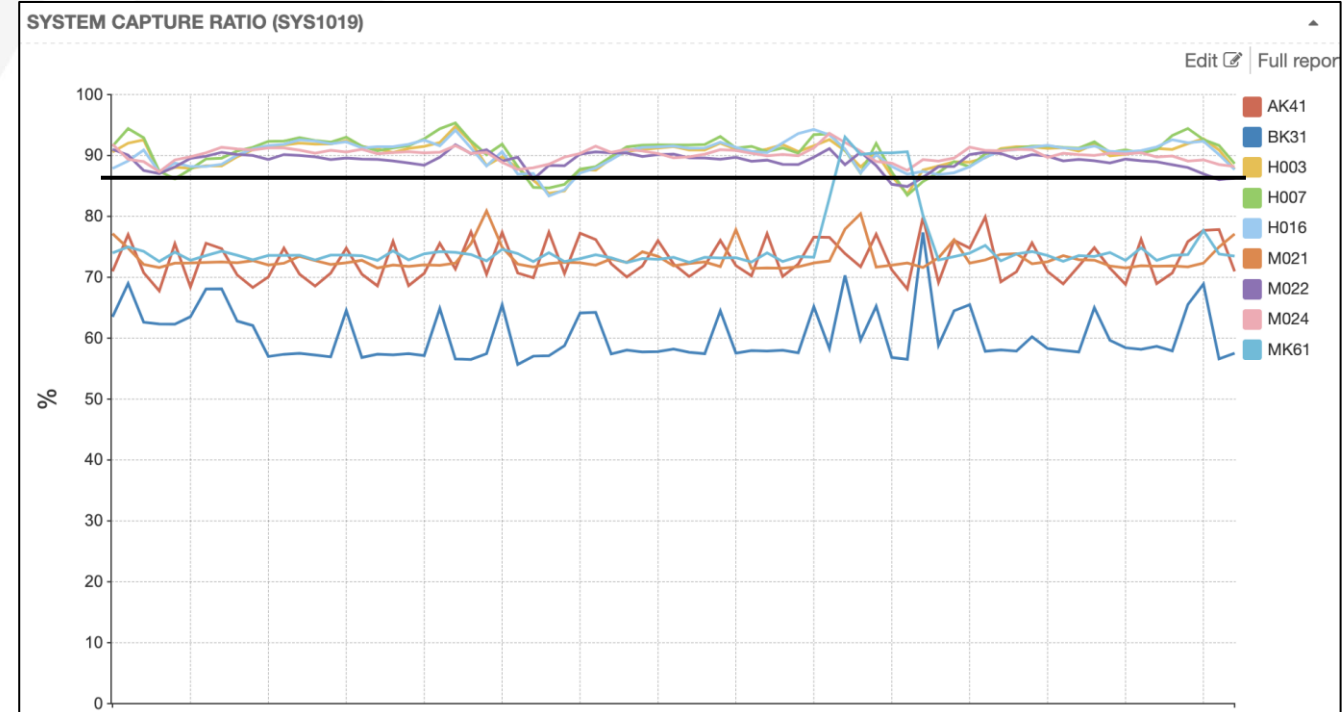
Performance Check #4: z/OS Capture Ratio

Is the capture ratio lower than it should be, indicating high system overhead?

Actions

If consistent capture ratios below 88% are observed, especially on larger LPARs, investigate common causes of uncaptured CPU time to gain efficiency

- WSC recommends using system wide capture ratio that includes both GCP and zIIP dispatch time
- z/OS Capture Ratio is reported in many performance monitoring products including RMF
 - Service times in Workload Activity Report (SMF 72) divided by Total Dispatch Time from CPU Activity Report (SMF 70)
- See Appendix B for common causes of low capture ratio



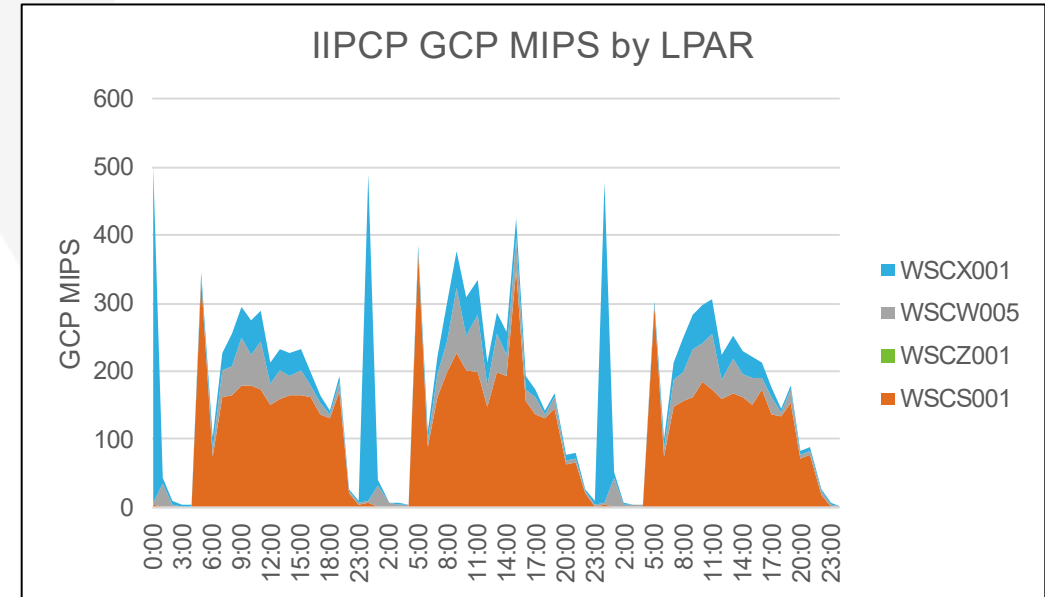
Performance Check #5: zIIP Usage and IIPCP Time

zIIPs allow select workloads to take advantage of additional processing capacity **without impacting their MSU rating or software usage charges**

- zIIP-eligible workloads include Java, Python, zCX, portions of Db2, Machine Learning for z/OS, and more

zIIP-eligible workloads can also run on GCPs when there's not sufficient zIIP capacity

- “Needs help” algorithm:** Wait to see if zIIP becomes available, and if not, ask for help from other engines including GCPs
- IIPCP time (crossover):** Time that zIIP eligible work spends running on GCPs, now contributes to software costs



```

SERVICE CLASS=CICSPROD

SERVICE TIME      ---APPL %---
CPU      377.378      CP      411.47
SRB       1.257      IIPCP   15.62
RCT       0.000      IIP     219.59
IIT       0.000      AAPCP    0.00
HST       0.000      AAP      N/A
IIP     131.753
AAP      N/A
    
```

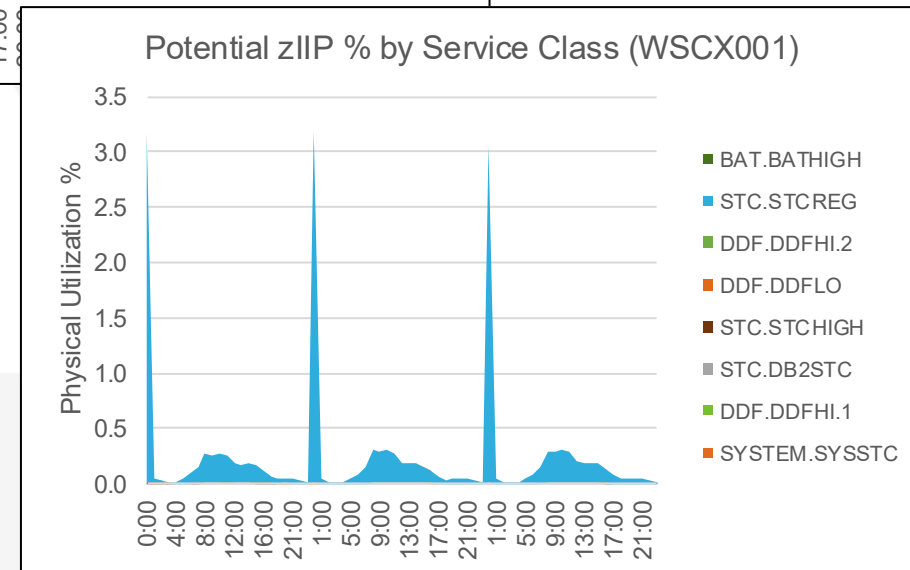
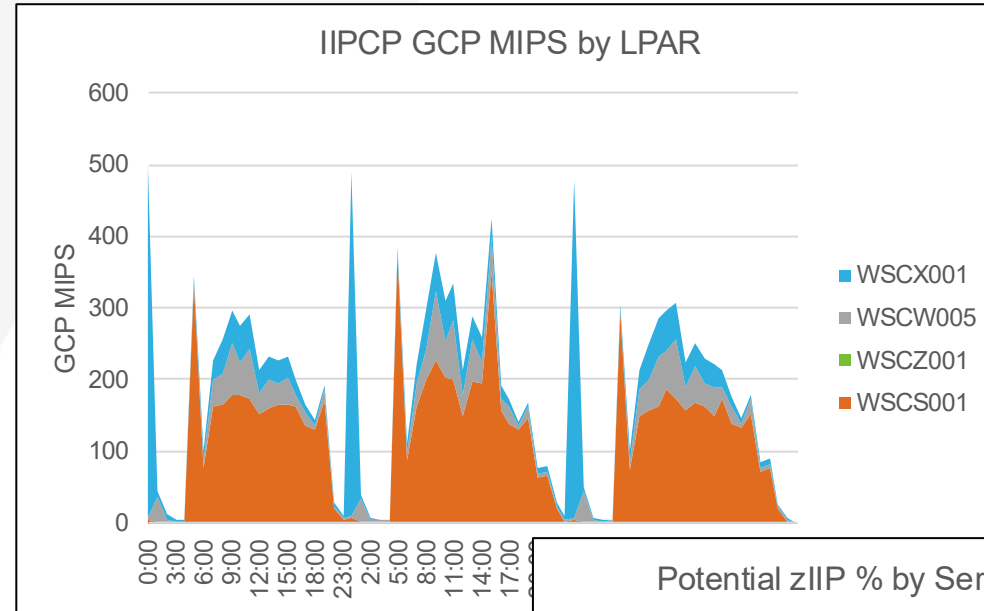
Performance Check #5: zIIP Usage and IIPCP Time

Is IIPCP time costing a significant amount of MIPS (and MSUs)?

Actions

If it's surpassing what's acceptable for your organization, consider the following options to reduce crossover:

1. **Buy more zIIPs:** Best performance
2. **Increase zIIP weight of the partition:** Stealing from someone else
3. **Enable SMT:** Get more zIIP capacity at the expense of response time, until you're able to buy more zIIPs
4. **Tuning to limit crossover:**
 - Service Class Honor Priority=NO: Limit specific service classes from running on GCPs.
 - Increase IEAOPTxx parameter, ZIIPAWMT: Specifies how long work will wait before asking for help – increase *slowly*



Performance Check #6: Memory Usage and Available Storage

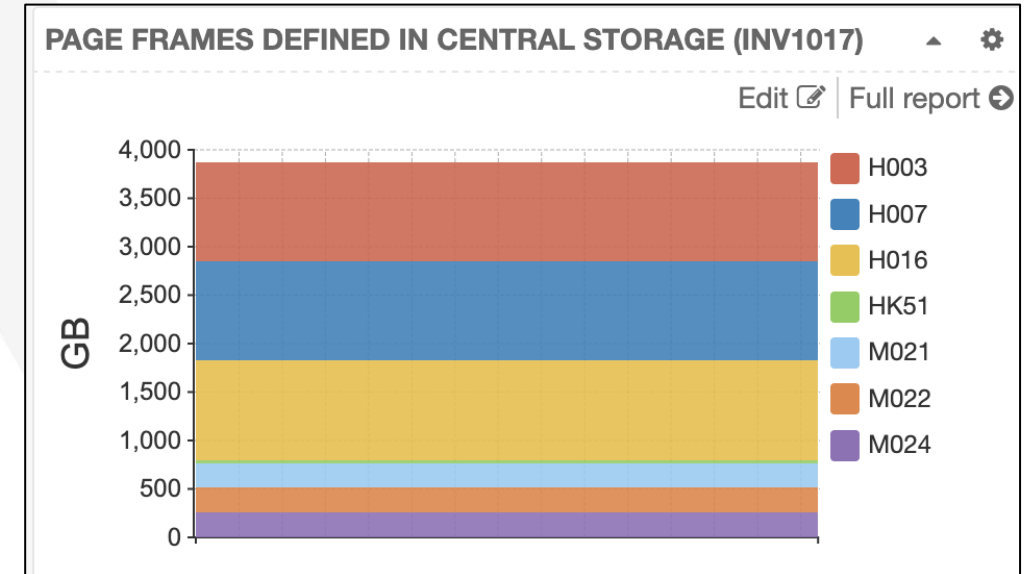
Memory (central storage) is allocated to each LPAR via the HMC

- Used for all applications, programs, and code running on the LPAR

Available Storage refers to the amount of central storage frames that are free for programs and applications to use when their data is brought in from Aux

If the available frame count drops extremely low, it can indicate storage contention and lead to page stealing

- Takes a frame assigned to an *active* user and makes it available for other work
- z/OS uses a least recently used algorithm for page stealing
- Some situations, like a system dump, can lead to significant number of active frames being paged out to aux, impacting performance



Processor Complex Serial Number	System ID	Processor Storage (GiB)
IBM-7BED8	H003	1,024.0
IBM-7BED8	H007	1,024.0
IBM-7BED8	H016	1,024.0
IBM-7BED8	HK51	32.0
IBM-7BED8	M021	256.0
IBM-7BED8	M022	256.0
IBM-7BED8	M024	256.0

Performance Check #6: Memory Usage and Available Storage

Is there sufficient available storage on each LPAR?

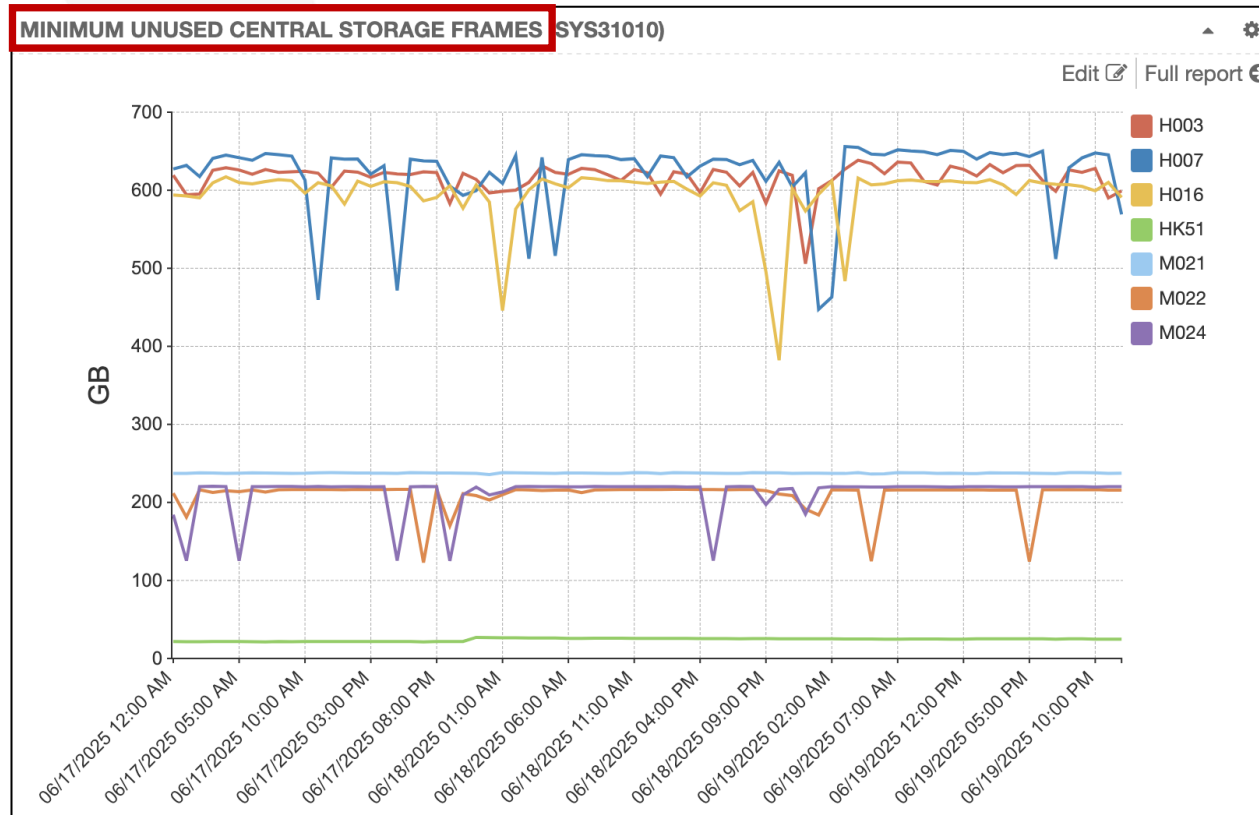
- Check using *minimum* available frame count, not average

Actions if too low

1. Allocate additional memory to the LPAR
 - If not all memory on CPC was defined
2. Take memory from another LPAR that doesn't need it
 - Identify LPARs that have more than enough memory defined
3. Make application changes to reduce contention
 - Are there programs or applications that can be limited in how much memory they're allowed to use
4. Tip: During processor upgrade, go big on memory

Best Practice

- Available frames should be at least as large as MAXSPACE parameter (how much memory a system dump can use)
- Demand paging rates should be at or near 0



Performance Check #7: z/OS WLM Performance Index

Workload Manager is a base component of z/OS that **manages the distribution of resources** to competing workloads on the system

- Manages use of CPU, I/O, storage, and more

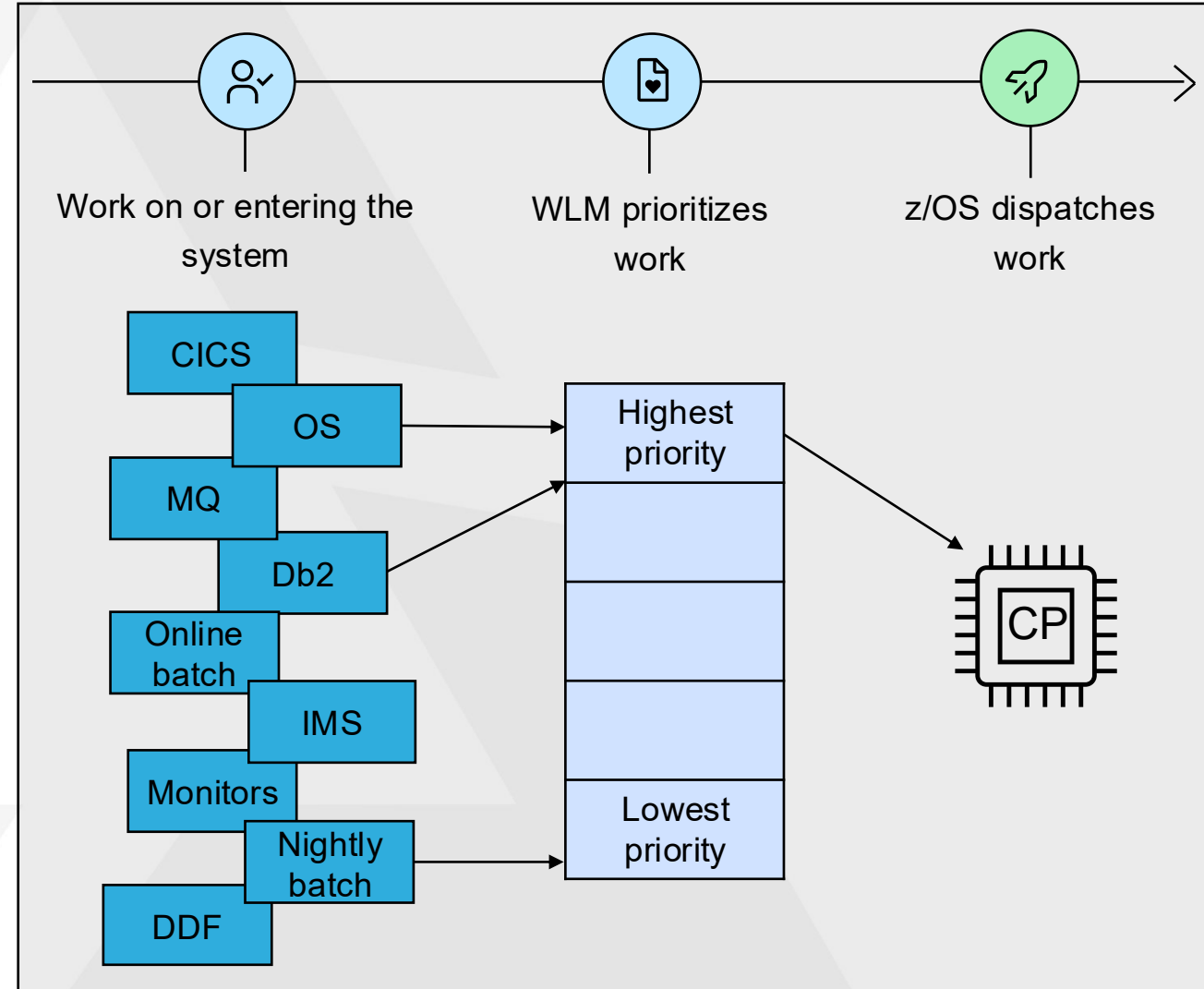
How it works:

- Define expectations (goals) for how work should perform in a WLM policy
- WLM continuously monitors and reports on goal achievement for each defined workload (**Performance Index**)
- WLM will dynamically adjust resources to ensure workloads are meeting these performance goals

Example WLM policy and more information:

<https://www.ibm.com/support/pages/wsc-zos-workload-manager-starter-policy-and-recommendations>

Ex) WLM CPU management



Performance Check #7: z/OS WLM Performance Index

Performance Index (PI)

Metric for goal achievement for each service class period

A local PI and a sysplex PI are calculated for each service class

Used in combination with importance to set a dispatch priority for the service class



PI less than one

Service class is **overachieving its goal**

Can be designated as a donor where WLM may steal resources to help another service class

Goal	90% in 0.500s
Actual	90% in 0.250s
PI	0.5



PI equal to one

Service class is **exactly meeting its goal**

The service class should have a PI close to one when performance is acceptable

Goal	90% in 0.500s
Actual	90% in 0.500s
PI	1.0



PI greater than one

Service class is **missing its goal**

Can receive resource from a donor where WLM will attempt to help the service class

Goal	90% in 0.500s
Actual	90% in 1.000s
PI	2.0

Performance Check #7: z/OS WLM Performance Index

It's crucial to review the goals in your WLM policy periodically to ensure they're **accurate, realistic and match business expectations** by evaluating the PI of each service class

What's the impact?

When **overachieving**, WLM may take resources away to help other workloads, even if they are at a lower importance

- If the processor isn't busy today, this may not be a big deal because everyone is getting the capacity they need
- As soon as there's a demand spike? This can have a significant impact.

When **constantly missing** goals, WLM may implement 'skip clock' and ignore a service class for some time to help other workloads

- In an effort to find a workload where it can make a more beneficial change

Performance Check #7: z/OS WLM Performance Index

Does the WLM goal match business expectations?

Actions

Identify service class periods that are consistently missing or overachieving and adjust goals appropriately

- Use z/OSMF WLM Policy Advisor or any tool that processes SMF 72.3 records and reports on PI

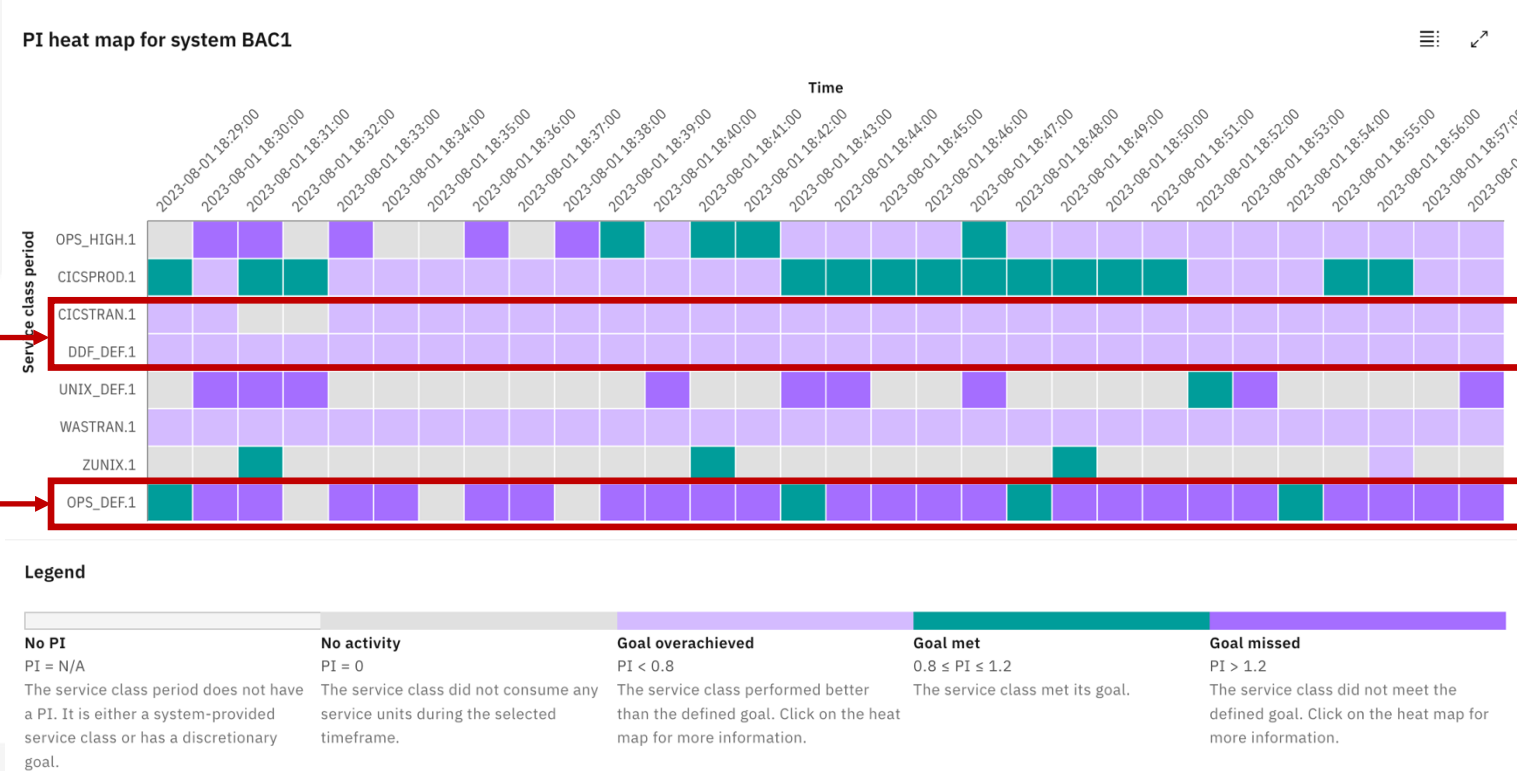
PI consistently less than 0.8 (overachieving goal)

- Is the goal accurate/realistic?
 - No: Make the goal more aggressive, to protect the work and match business expectations
 - Yes: No change, okay with work being marked to donate resource to others
- Keep in mind the minimum PI for a percentile response time goal is .5

PI consistently over 1.2 (missing goal)

- Is the goal accurate/realistic?
 - No: Make the goal less aggressive
 - Yes: May need a higher importance or more resources for it to achieve the required performance

z/OS 3.1+ WLM Policy Advisor



Summary

- Regular performance checkups are critical to ensuring your z/OS environment runs efficiently and supports business objectives
 - Maximize throughput and efficiency
 - Support informed capacity planning decisions
 - Resolve issues before they impact availability
 - Enable faster troubleshooting
- z/OS Performance Checkup
 - 1-2. Monitor and report CPU Utilization and Workload Growth
 3. Review LPAR configuration for potential performance gains
 4. Check for low capture ratios and investigate for efficiency improvements
 5. Monitor IIPCP time and evaluate options for MSU savings
 6. Check available storage levels to reduce paging impacts
 7. Regularly review WLM Performance Index and Service Class goals for peak performance and to prevent availability problems

Experience more with IBM



Visit us at the IBM Booth #113

After a full day of technical sessions, take a break with us!

Connect with our experts, snap a photo with the z17 Plexi or the latest Telum II, and get an up-close look at our Spyre Accelerator.

Come back each day for fresh topics and demos at our expert stations.

Think 2026

Join 5000+ senior business and technology leaders who are seizing the AI revolution to unlock unprecedented growth and productivity at **Think 2026**.

Find out more information using the QR code below.



IBM Digital Asset Haven

IBM Digital Asset Haven is the operational backbone for financial institutions and regulated enterprises entering the digital asset economy.

Find out more information using the QR code below.



Thank You

- Questions?



© 2026 International Business Machines Corporation IBM and the IBM logo are trademarks of IBM Corporation, registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on ibm.com/trademark.

This document is current as of the initial date of publication and may be changed by IBM at any time. Statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

THIS DOCUMENT IS DISTRIBUTED "AS IS" WITHOUT ANY WARRANTY, EITHER EXPRESS OR IMPLIED. IN NO EVENT, SHALL IBM BE LIABLE FOR ANY DAMAGE ARISING FROM THE USE OF THIS INFORMATION, INCLUDING BUT NOT LIMITED TO, LOSS OF DATA, BUSINESS INTERRUPTION, LOSS OF PROFIT OR LOSS OF OPPORTUNITY.

Client examples are presented as illustrations of how those clients have used IBM products and the results they may have achieved. Actual performance, cost, savings or other results in other operating environments may vary.

Not all offerings are available in every country in which IBM operates.

It is the user's responsibility to evaluate and verify the operation of any other products or programs with IBM products and programs.

The client is responsible for ensuring compliance with laws and regulations applicable to it. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the client is in compliance with any law or regulation.

Your feedback is important!

Submit a session evaluation for each session you attend:

www.share.org/evaluation





APPENDIX

Appendix A

WSC Best Practice – Number of Logical CPs defined to an LPAR

IBM WSC best practice whitepaper:

<https://www.ibm.com/support/pages/node/6354843>

This document reviews the IBM Best Practice Recommendation for setting the number of Logical CPs in a z/OS LPAR.

The IBM Best Practice recommendation is to not configure all of the available logical CPs to each partition. Rather, we recommend the number of logical CPs defined be equal to the number of logical required to satisfy the relative weight of the LPAR, plus between 0 and 2 additional logical CPs based on LPAR size and appropriate whitespace.

Why?

- **Work runs most efficiently if you run within defined weight using VHS and VMs**
- LPAR Busy value displayed on online monitors is relative to number of LCPs
- LPAR time slice is sensitive to number of logicals, fewer logicals will lead to longer time slices
- Reduce the impact of a CPU Loop, fewer logicals limits potential impact
- Certain z/OS operations need to be done even for parked logicals
- Additional system resources utilized for each logical processor

Appendix B

Causes of low capture ratio

Common Causes of Uncaptured CPU Time

- High page fault rates
 - Full preemption
 - Suspend lock contention
 - Spin lock contention
 - GETMAIN/FREMAIN being done in interrupt handlers or the dispatcher
 - Branch Tracing
 - IRB queuing with a large subtask tree
 - Inability to queue IRBs to a task
 - SLIP processing
 - Long internal queues
 - Affinity processing
 - Account code verification
 - Fragmented storage pools
 - Inefficient ACS routines
 - Symbolic Substitution
- ← Most Common

- ★ Generally look for a capture ratio in the 88-95% range
- ★ Use SMF 30, subtype 4,5 to get information on CPU time spent in the initiator to determine if there are areas for improvement
- ★ New z/OS 1.12 SMF 30 fields to characterize batch times