

Using AI for capacity planning and performance management

Domenico D'Alterio
Principal Product Manager,
AIOps for IBM Z
ddalteri@it.ibm.com

John Baker
Principal Consultant, Z Performance,
AIOps for IBM Z
johnb@ibm.com

Mainframe operations teams are under pressure to keep critical services running – but **growing complexity** and **fragmented tools** make it harder than ever

the average hourly cost enterprises incur due to downtime

\$300K+

AI-enabled insights can augment established practices to help teams manage challenges more proactively.



Customer expectations are higher than ever

Customers expect always-on digital services and Z is the backbone.



Operational risk is rising

Increasing workload volume, more complex hybrid architectures, growing SME shortages



Reactive operations no longer scale

Teams can't manually correlate data from dozens of tools fast enough to keep pace.

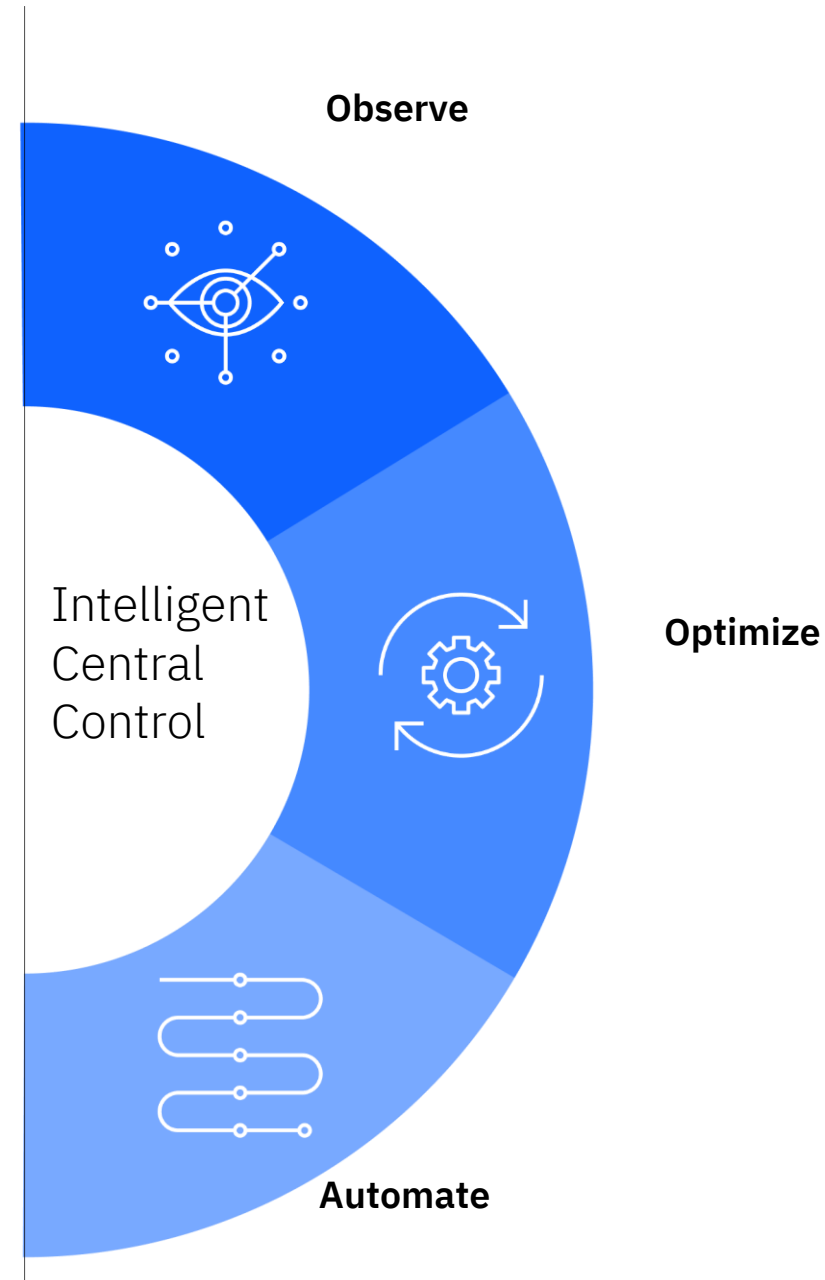


AI-assisted operations are now essential

Organizations must shift from manual incident chasing to proactive, AI-driven resiliency.

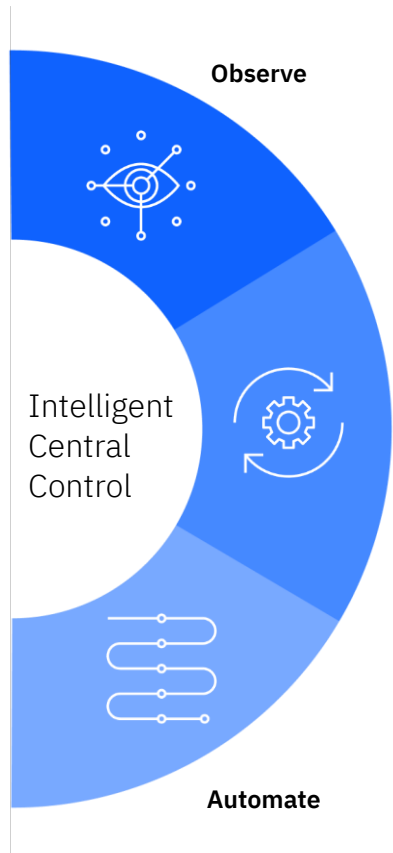
IBM can help you simplify operations and enhance resiliency

- Simplified, AI-powered operations that provide observability, optimized performance and increased automation.
- AI-driven operations that anticipate issues through OpenTelemetry-enabled insights.
- Delivers operational resilience, reduced MTTR, and maintains continuous availability.



Building blocks for modern Z operations

Integrated capabilities help you improve performance, resilience and cost efficiency.



Observe

Deliver enterprise-wide visibility across IBM Z and hybrid applications to support AI-assisted issue detection, diagnosis, and trend-based prevention.

- Efficiency-Driven Z Observability
- Rapid Root Cause Analysis
- Cost & Utilization Insights
- End-to-End DevOps Lifecycle Tracking
- Proactive Trend Analysis & Visualization



Optimize

Maximize mainframe availability and performance through correlated insights, continuous tuning, and AI driven optimization across Z + hybrid environments.

- App & Subsystem Performance Optimization
- Continuous Performance Assurance
- Cross-Environment Resource Optimization
- Multi-Source Data Correlation



Automate

Apply agentic AI and guided automation to reduce MTTR, boost staff productivity, and sustain resiliency across mainframe operations.

- Automated Remediation
- Automated Resilience & Recovery
- AI-Powered Low-Code Workflows
- Repeatable Runbooks

The future of performance and capacity planning on IBM Z

Today's global, 24/7 economy demands continuous availability (zero service disruptions) from business applications. A lack of visibility into critical applications and understanding of the IT infrastructure makes this difficult to achieve.

IBM provides AI, agents, forecasting, and what-if technology to empower mainframe sites to prevent service disruptions before they occur and utilize measurement data insights to lower costs and optimize business applications.



Reduce downtime; identify risks and inefficiencies

Automated site health and risk detection flags hidden risks, both current and upcoming.

Automated site health uses AI based forecasting to alert on future usage that exceed requirements.

What-If analysis assists in picking optimal configurations given future workloads.



Safely Optimize Costs without impacting service

Easily optimize your workloads and site configuration to optimize CPU consumption.

Eliminate redundant tooling and silos while enhancing cross-departmental collaboration with end-to-end visibility across your infrastructure.

Use forecasting and what if to choose a configuration that minimizes service disruptions.



Augment Staff and Accelerate Training

Augment the effectiveness of staff with interactive, customizable, and shareable reports and dashboards, built-in and explanations, and extensive drill downs..

Expedite learning, promote collaboration, and enhance analytical effectiveness with force-multiplying technology

This strategic value is only unlocked with smart data integration between the silos and by enabling cross-infrastructure views of specific application / business workloads

IBM Z IntelliMagic Vision for z/OS



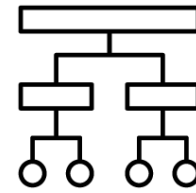
End-to-end
z/OS infrastructure
AI-driven analytics for
performance management



CICS



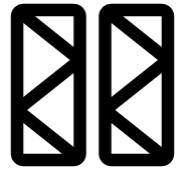
Db2



IMS



MQ



Systems



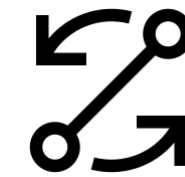
TCP/IP



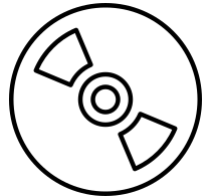
WebSphere



z/OS
Connect



Disk &
Replication



Virtual Tape

An IntelliMagic Customer - z/OS Infrastructure

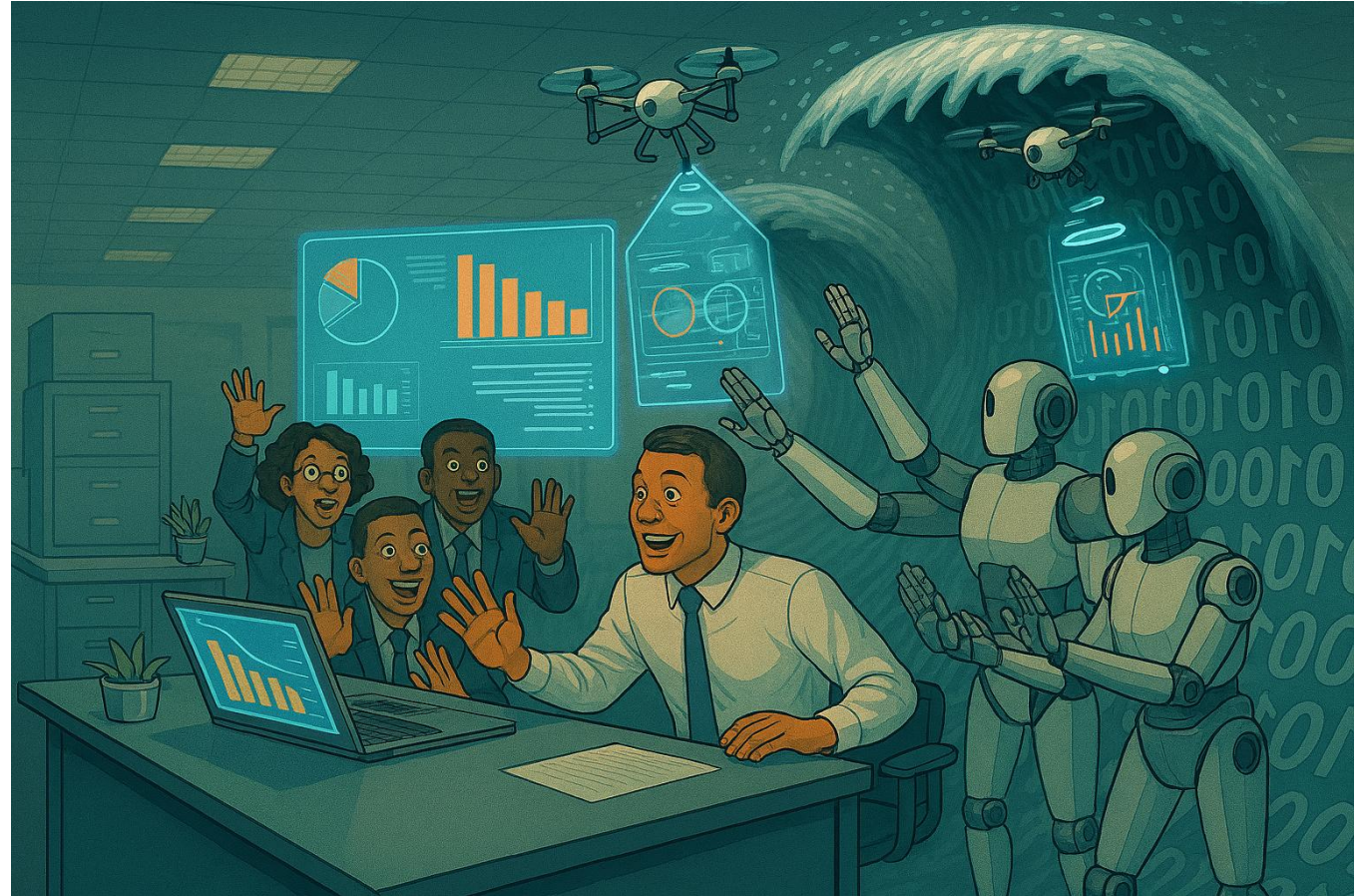
- SMF/RMF/CMF/BVIR/IMS Logs Daily Pro
- Daily: 929 million+ records
- 24 hrs at 15 Minute Intervals = **672 intervals** per week
- Each record type contains 5-50 metrics (e.g. 15 metrics per record)
- $929m \times 15 = 13.9+$ billion interrelated metrics per day



Tsunami:

..an arrival or occurrence of something in overwhelming quantities.

An IntelliMagic Customer - z/OS Infrastructure



Tsunami:

..an arrival or occurrence of something in overwhelming quantities.

AI can help analyze in
better ways

Using AI and Machine Learning to Detect Unexpected Changes in Performance and Usage

Based on mix of models on top of state space model








Learns from historic performance and usage

Report using exceptions that are flagged with yellow and red indicators

WLM Importance Rating Change Detection

For System ID 'A004' by Importance

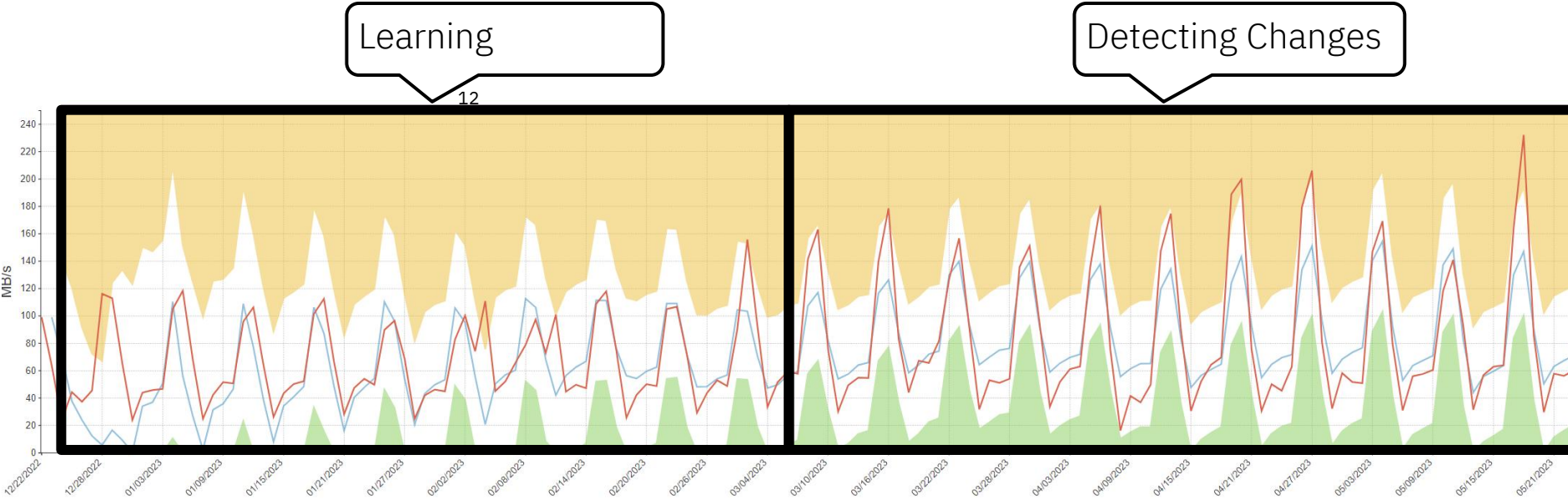
Drilldown 

Importance 	Performance Index 	Performance Index Change (std devs) 	All CP processor time used (Classic) (s) 	All CP processor time used (Classic) Change (std devs) 	zIIP Processor Usage (s) 	zIIP Processor Usage Change (std devs) 
3	1.82	2.23	14.01	-0.11	10.056	-0.04
5	1.76	1.40	8.45	2.26	10.996	1.57
1	1.39	1.37	440.22	1.65	744.445	2.30
4	0.95	2.23	147.73	0.56	39.371	0.26
Disc	0.81	0.26	98.96	0.28	7.882	-0.60
2	0.73	0.62	249.69	0.81	43.135	0.81
0-Sys	0.00	0.00	487.05	2.93	3.343	0.87

Using AI and Machine Learning to Detect Unexpected Changes in Performance and Usage

Based on mix of models on top of state space model

Learns from historic performance and usage



Using AI to flag risks to infrastructure and application availability

- > IDUG
- > Workshop
- > WSC
- > Ungrouped
- Reports**
- Health
- Health Insights
- Health Exceptions
- DSS
- Groups
- Tiers
- Systems and WLM**
- Log Streams
- Coupling Facility
- XCF
- Db2
- CICS
- MQ
- TCP/IP
- FICON and Channels
- TS7700
- > Change Detection
- > Configuration
- > Topology
- > Rating by Time
- > Applications

WLM Importance ratings per z/OS Sysplex [rating: 2.56]

for all Service Classes by z/OS Sysplex ID
Rating based on Service Class Period Statistics data using Service Class Period Thresholds

Chart
Grid
Thresholds
Drilldown

z/OS Sysplex ID	Performance Index	Captured GCP CPU	z/OS Eligible CPU Exec on		I/O Response		Page-in Rate
			z/OS CPU	GCP	I/O Rate	Time	
PLEXA000	!	●	●	■	●	■	■
PLEXH000	!	●	●	!	●	■	■
PLEXM000	⚠	●	●	■	●	■	■

Description

The WLM Health chart by Importance shows whether the service classes meet the specified WLM goals for each WLM Importance level. If the service classes meet the goals, the Performance Index will be 1, and the corresponding bubble green. When the goals are not met, the bubble will be yellow or red. Drill-downs are available for a health chart per service class, or for by-time minicharts, allowing you to investigate whether the performance problems that cause the goals to be missed are processor or storage related. WLM goals are not maintained for system goals (0) nor for discretionary work (99). The Health chart also shows whether there is a significant amount of z/OS eligible work running on CPs, and finally whether the I/O response time is good.

Field Description

Performance Index

The performance index for the service class. When multiple service class (periods) are summarized, the weighted average will be shown based on the using samples for workloads with a velocity goal. When other or mixed goal type workloads are averaged, a plain unweighted average is used of all service or report class periods. The performance index will be maximized at 5.

The calculated Workload PI. Basically, a value greater than 1 means that targets are not met. The workload PI is calculated differently for each goal type. When the goal types are different between the different types of data, use a weighted average based on wkldusing (R723CTOU) and wklddelay(R723CTOT)is using IntelliMagic Vision. When the goal type is Velocity, calculate the PI by calculating velocity.

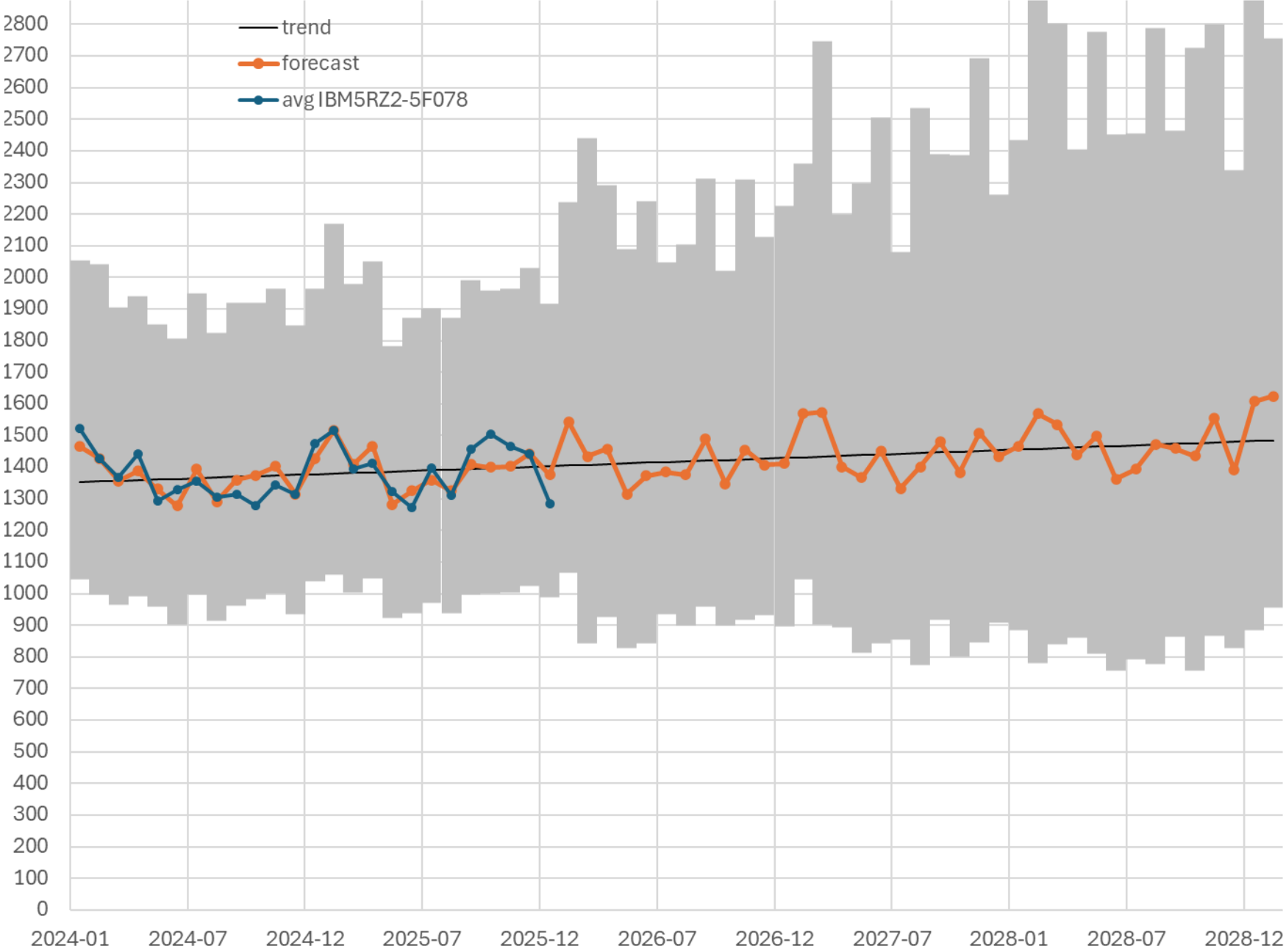
Add to: Collected Dashboard Favorites Edit report

Single day 4/1/2024
Interest group IGT, All sysplexes, All shifts
Reporting interval Measurement

Forecasting

Predicting years ahead
Learning from history
Seasonality

- Answer questions like:
- Will I run out of capacity in my CPCs before I will replace them?
 - Will I stay within my TFP contract?



IBM IntelliMagic agent for Z

The **IBM IntelliMagic agent for Z** empowers performance analysts to find relevant information in the vast Z performance space. The **IBM Z IntelliMagic agent for Z** is an experienced co-worker that assists users in quickly finding relevant reports for conducting performance analysis and resolving issues faster. The agent understands **natural language queries**, providing a conversational interaction experience.

"With IBM IntelliMagic agent for Z, users can find the nuggets that help them solve day to day problems"

Vision for IBM IntelliMagic agent for Z

Strategic Vision:

AI that guides and supports you through the tasks related to using IBM Z IntelliMagic Vision for z/OS.

Giving helpful suggestions, can point out things of interest and can summarize and analyze your z/OS environment.

As if an experienced colleague is sitting right next to you.

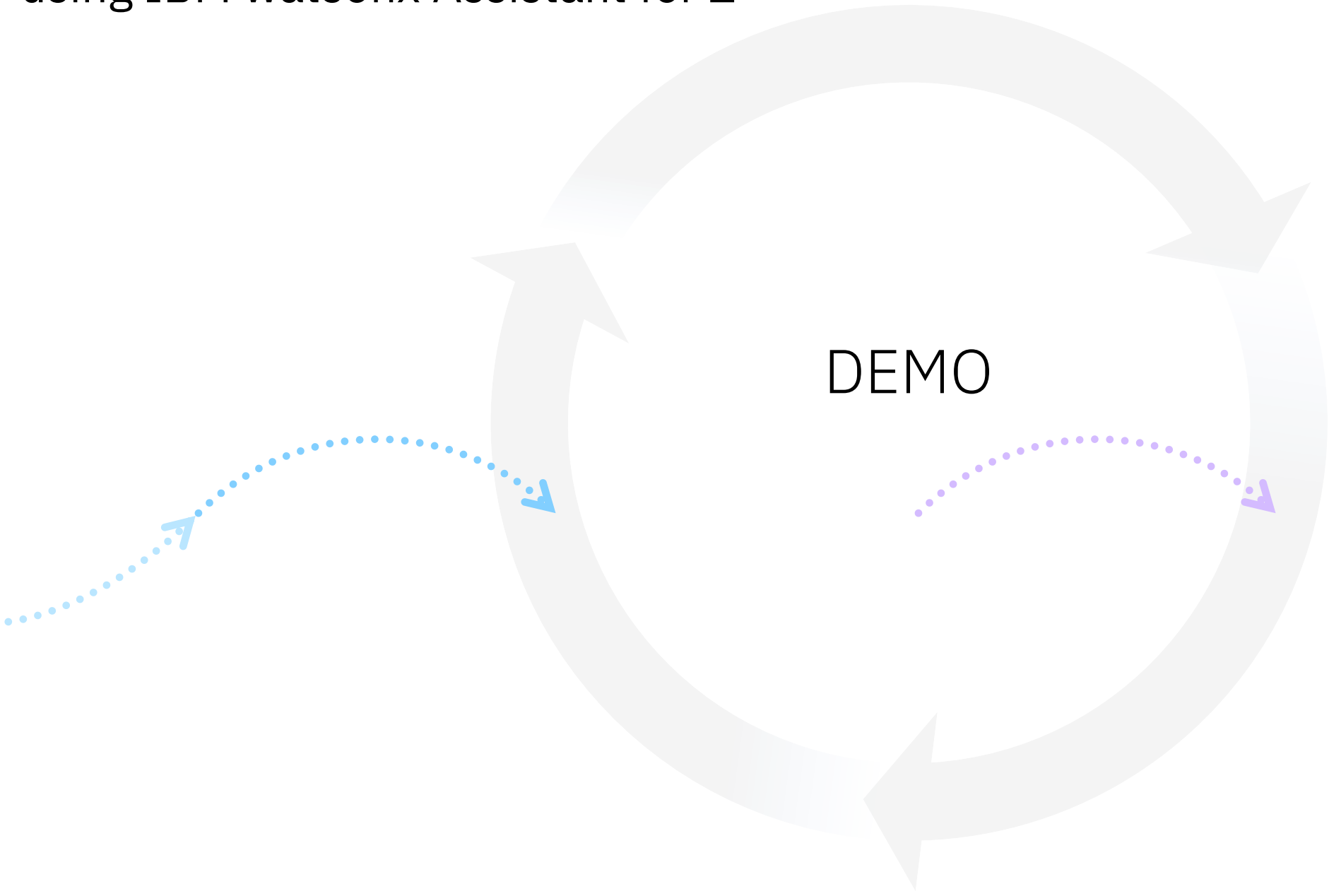
Enabled by:

Using Agents that have intimate knowledge of the structure and capabilities of IntelliMagic, and that can access reports and data from IntelliMagic.

This combined with supporting documents about how to use IntelliMagic as well as general z/OS knowledge.

Allow the AI to support the user with best practices and informed guidance.

Changing the way you engage and interact with IBM Z IntelliMagic Vision using IBM watsonx Assistant for Z



IBM Z IntelliMagic Agent

Roadmap

Today - GA

Help users finding information in IBM Z IntelliMagic Vision for z/OS based on current report and user query

Support Spyre

Near Term

Explain - Summarize anomaly based on data and drill downs

Future

Use history as additional context answering questions

Adding knowledge of environment (as know by IntelliMagic)

Can clarify terminology used

Can suggest other areas to look at, based on WXA4Z

Explain

Summarize anomaly based on data and drill downs

Why

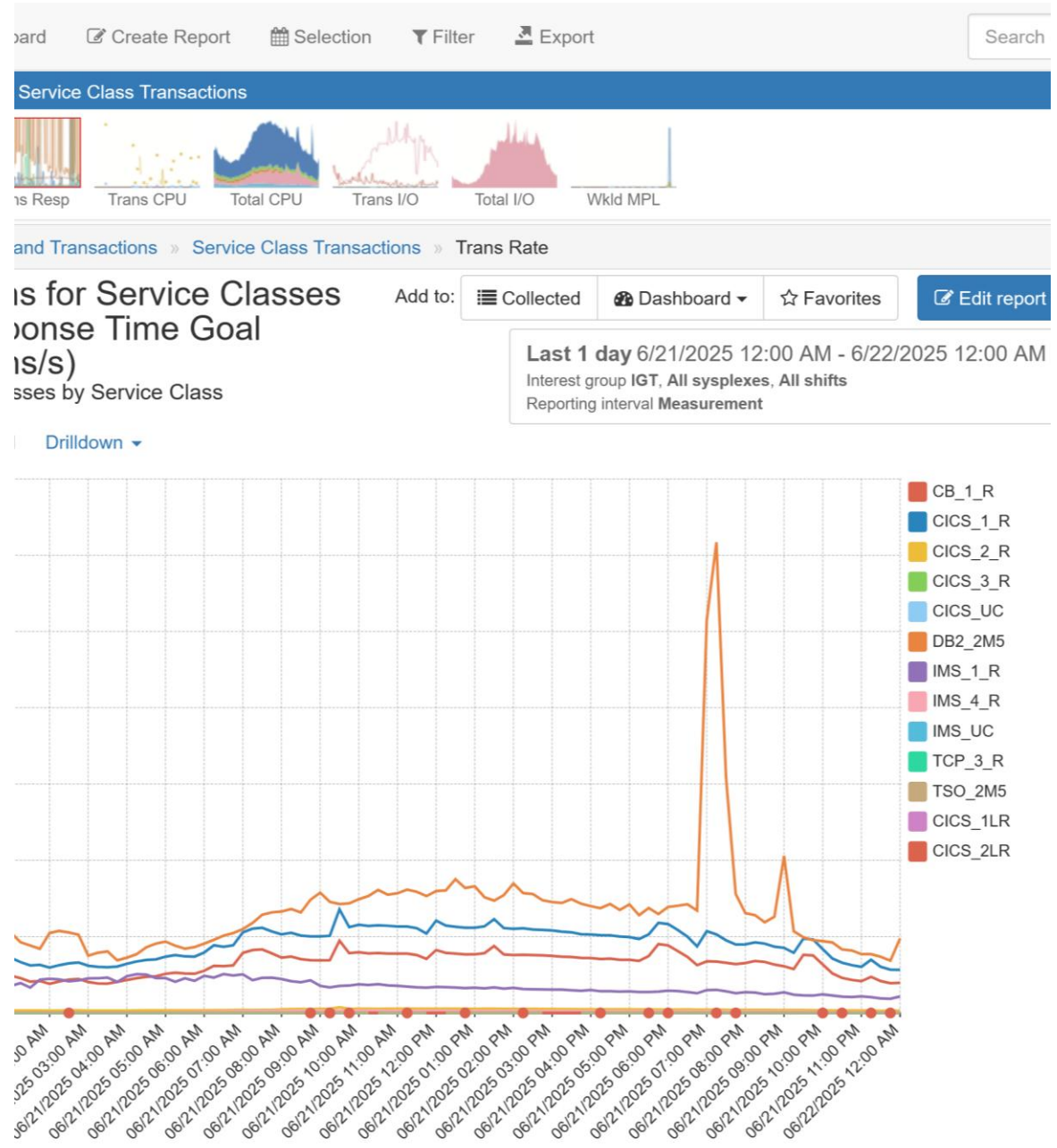
Spikes found at an aggregated level are typically only caused by a few underlying factors.

Currently finding these factors requires multiple drill-downs, through multiple levels.

This is time consuming.

What

Let AI do the drill-down and analysis work and report back on what the contributing factors are.



Show

What

User can select an anomalous point and ask for an explanation.

Agent gives a summary of what the underlying issue might be.

This summary appears in the chat and contains clickable previews of relevant reports.

How

Agent navigates through IntelliMagic Vision. It inspects the content of the reports and uses its knowledge to select the most relevant drill-down paths to inspect further.

It collects relevant findings and summarizes them back to the user.

The screenshot displays a monitoring interface for the period 'Last 1 day 6/21/2025 12:00 AM - 6/22/2025 12:00 AM'. The chart shows a significant spike in activity at 7:15 AM. A blue arrow points from this spike to a chat window on the right. The chat window, titled 'watsonx Assistant for Z 2:58 PM', contains three report cards:

- Transactions for Service Classes with a Response Time Goal (transactions/s) for all Service Classes by Service Class**: At the service-class level, the DB2_2M5 class carries more than half of all transactions, contributing 54.8% of the 13,504 total, far above the average share of 9.1% and the mean total of 7,760, indicating a pronounced load.
- Transactions for Service Classes with a Response Time Goal (transactions/s) for Service Class 'DB2_2M5' by Report Class**: Within DB2_2M5, the report class DBDB2000 alone accounts for all 7,404 transactions, exceeding the mean of 3,099 and the standard deviation of 1,243, showing it dominates the workload.
- Transactions for Periods with a Response Time Goal (transactions/s) For Service Class 'DB2_2M5', for Report Class 'DBDB2000' by System ID**: At the system level for DBDB2000, system H002 supplies 21.7% of the 7,404 transactions, higher than the mean share of 16.7% and the standard deviation of 1,243, marking it as the primary system source.

By period, period 1 carries almost all of the 7,404 transactions (99.997%), well above the...

Type something...

Details

Background

Will be part of the IBM IntelliMagic agent for Z.

Core algorithms developed by IBM Research.

Uses domain knowledge embedded in IBM Z IntelliMagic Vision.

Elements

Steps that support the conclusion.

Clickable evidence, linking back to the real data.

Summary, based upon the steps shown above.



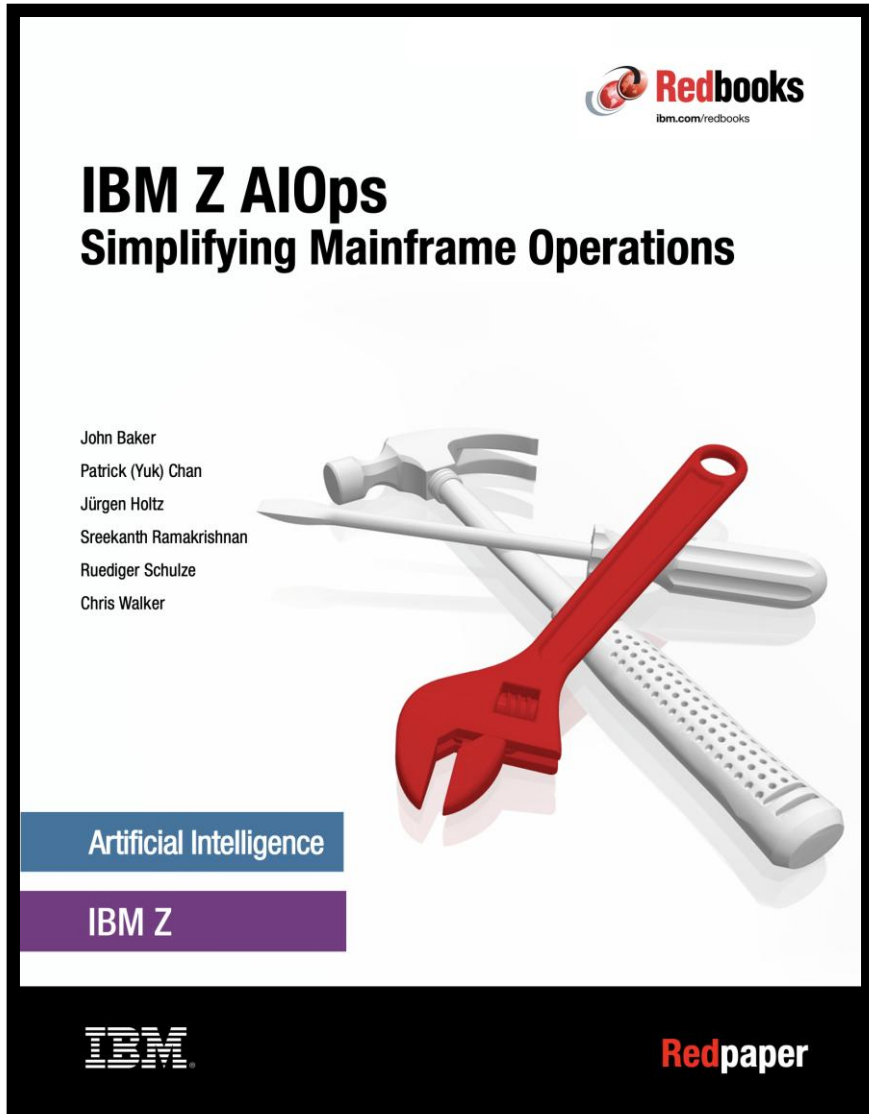
Observability and AIOps for IBM Z sessions at SHARE Orlando

Day	Time	Title	Featured Products
Monday	9:45 am	Data Center Automation - Z System Automation and Agentic AI	Z System Automation
	10:30 am	Using OpenTelemetry to Integrate the Mainframe Into Your Enterprise-Wide Observability Platform	Z Observability Connect, Instana
	1:15 pm	BYOD Lab: WXA4Z Agentic Hands-on Workshop	watsonx Assistant for Z
	1:15 pm	Intelligent Automation of a Hybrid Data Center With Next Generation of Z System Automation	Z System Automation
	2:30 pm	IBM Z NetView Technical Updates	Z NetView
	3:45 pm	Simplify Mainframe Operations with the Latest IBM Z OMEGAMON Enhancements	OMEGAMON
Tuesday	10:30 am	Stronger, Smarter Operations: How BPER Reinvented Data Resiliency Management with IZBR	Z Backup Resiliency
	2:30 pm	What's New in IBM Z Workload Scheduler V.10.2.x and Roadmap	Z Workload Scheduler
Wednesday	9:15 am	IBM Concert for Z, an AI-Powered Mainframe Resilience Platform: Solutions vs. Alerts	Concert for Z
	1:45 pm	BYOD Lab: AI Enabled Proactive Monitoring to Get the Most From Your System With IBM Concert for Z	Concert for Z
	2:30 pm	IBM Z Cyber Vault Explained - Soup to Nuts and Nose to Tail	Z Backup Resileincy
	3:45 pm	Using AI for Capacity Planning and Performance Management in Z	IntelliMagic
Thursday	10:30 am	IBM Z NetView Hints and Tips	Z NetView

Connect with IBM technical leaders and product management team:

Concert for Z	OMEGAMON	Z Observability Connect	NetView	Z System Automation	Z Workload Scheduler	IMS Tools
Domenico D'Alterio	Ash Mahay	Instana	Derrick Washington	Johannes Hausch	Domenico D'Alterio	Tracy Dean
Fabrizio Miatto		Chris Walker			Wolfgang Schaeberle	

New Redpaper



How can you simplify mainframe operations? To answer this question, this IBM Redbooks publication draws on Lean Thinking, which focuses on identifying waste and strain in any process.

By applying AI and machine learning technologies to mainframe operations, you can deliver improved efficiency and effectiveness.

24

Download your copy:

<https://ibm.biz/Redpaper-SMO-2025>



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**.

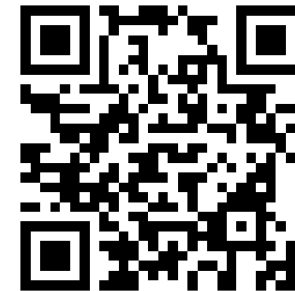
Find out more information using the QR code below.



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.



Your feedback is important!

Submit a session evaluation for each session you attend:

www.share.org/evaluation

