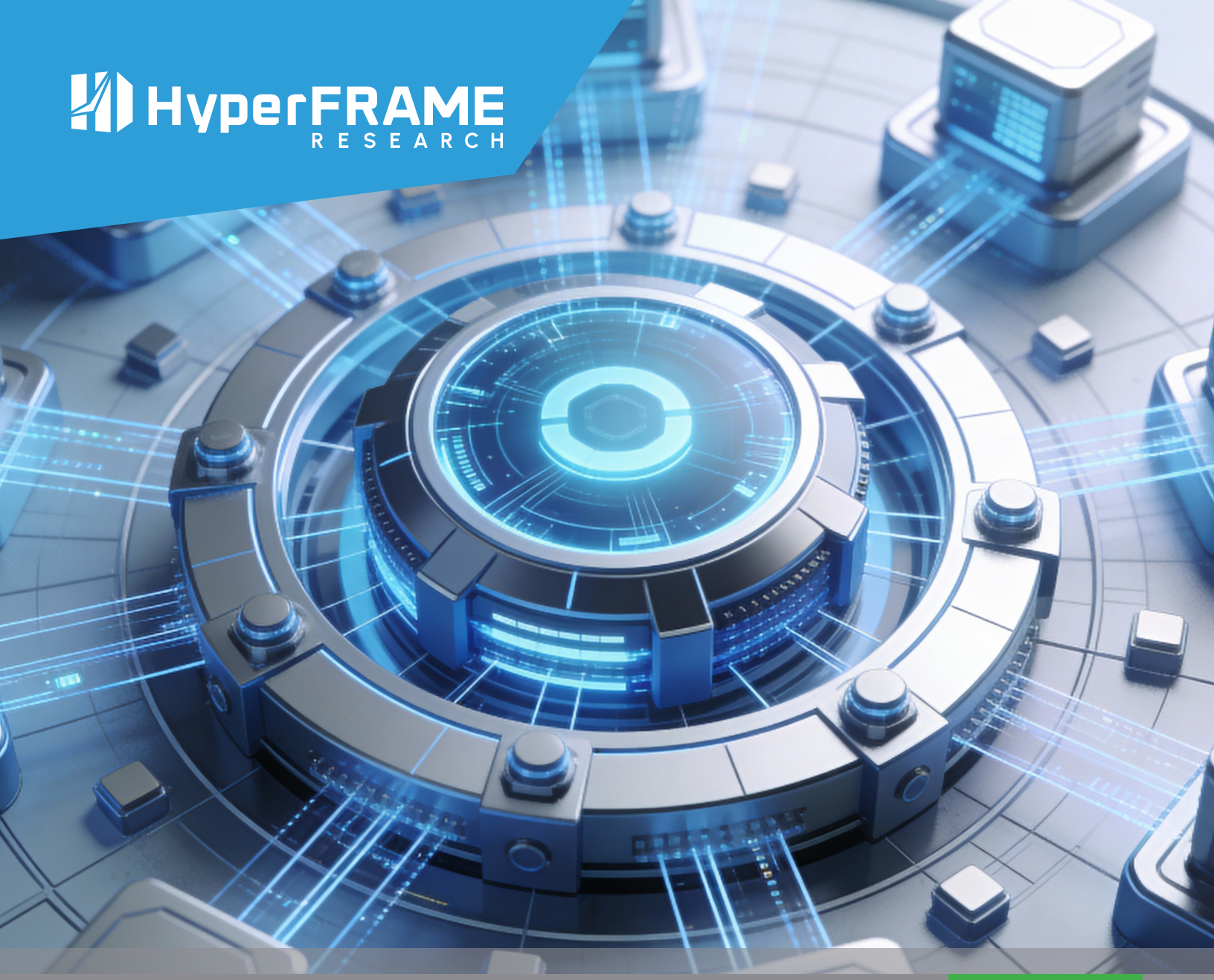
HyperFRAME RESEARCH



RESEARCH BRIEF

2025 Was Big Year For Big Iron With z17 Taking Center Stage

While z17 Garnered the Column Inches, 2025 was the year of AI for Mainframe Workloads.

Authors:

Steven Dickens
CEO & Principal Analyst | HyperFRAME Research

DECEMBER 2025

Executive Summary

2025 has been about AI. Whether it is semiconductors, software, memory, servers, or even geopolitics, the 2025 tech landscape has been overwhelmed with AI, and while LLMs and GPU updates garner column inches in the mainstream media and the nightly news, the mainframe has not been left behind. The z17 debuted in April, almost 61 years to the day after the first s/360 machine broke cover, but despite the longevity of the mainframe, the system is still at the cutting edge of computing.

Public cloud outages brought the Mainframe community out on social media platforms to gloat in 2025. The AWS US East outage or the Google outage earlier this year made workload availability front-page news, with many people realising what a 99.9% availability guarantee actually means in reality. Public cloud SLAs are typically 99.9% (“three nines”) or 8.77 hours of downtime per year, and IBM certifies the mainframe to 99.999999% (“eight nines”), which equates to 315.58 milliseconds of annual downtime. Not all workloads need eight nines of availability, and the cost of providing this level of reliability is prohibitive for many applications, but for those workloads where zero downtime is mandatory, the mainframe still has its place. Put simply, you pay your money, you take your chances.

Artificial Intelligence has fundamentally shifted the mainframe landscape by turning what was once a “manual archaeology” project into a streamlined, high-velocity engineering task. Leveraging agentic AI, AWS Transform now automates the extraction of dense business logic and generates functional test plans, cutting replatforming timelines from years to months. Specialized vendors like Bloop and Heirloom are utilizing Large Language Models (LLMs) to provide real-time code explanation and seamless COBOL-to-Java transpilation, ensuring that the target cloud environment remains maintainable. Meanwhile, Mechanical Orchard and CloudFrame have pioneered “behavioral twin” and “incremental refactoring” approaches, using AI to prove functional equivalence between the legacy system and its modern cloud-native replica.

These advancements highlight a unique industry paradox where two seemingly contradictory truths are simultaneously valid. Put simply: Two truths can be true at the same time. On one hand, the mainframe is growing, with many enterprises increasing MIPS (millions of instructions per second) capacity to handle the surge in high-volume, secure transaction processing driven by digital banking and AI-integrated applications. On the other hand, workloads are moving to the cloud at an unprecedented pace as AI-powered tools lower the risk barrier, allowing companies to migrate non-core or agility-starved applications to services like AWS Transform. This creates a balanced hybrid reality where the mainframe evolves into a high-performance data hub while the cloud becomes the playground for rapid innovation.

One truth does however, remain solely true; the mainframe space is as vibrant and dynamic as any time in the last 61 years, and HyperFRAME doesn’t expect that t change in 2026, in fact the exact opposite. The pace of change and evolution will only increase.

The Spring of Innovation: The z17 Arrives

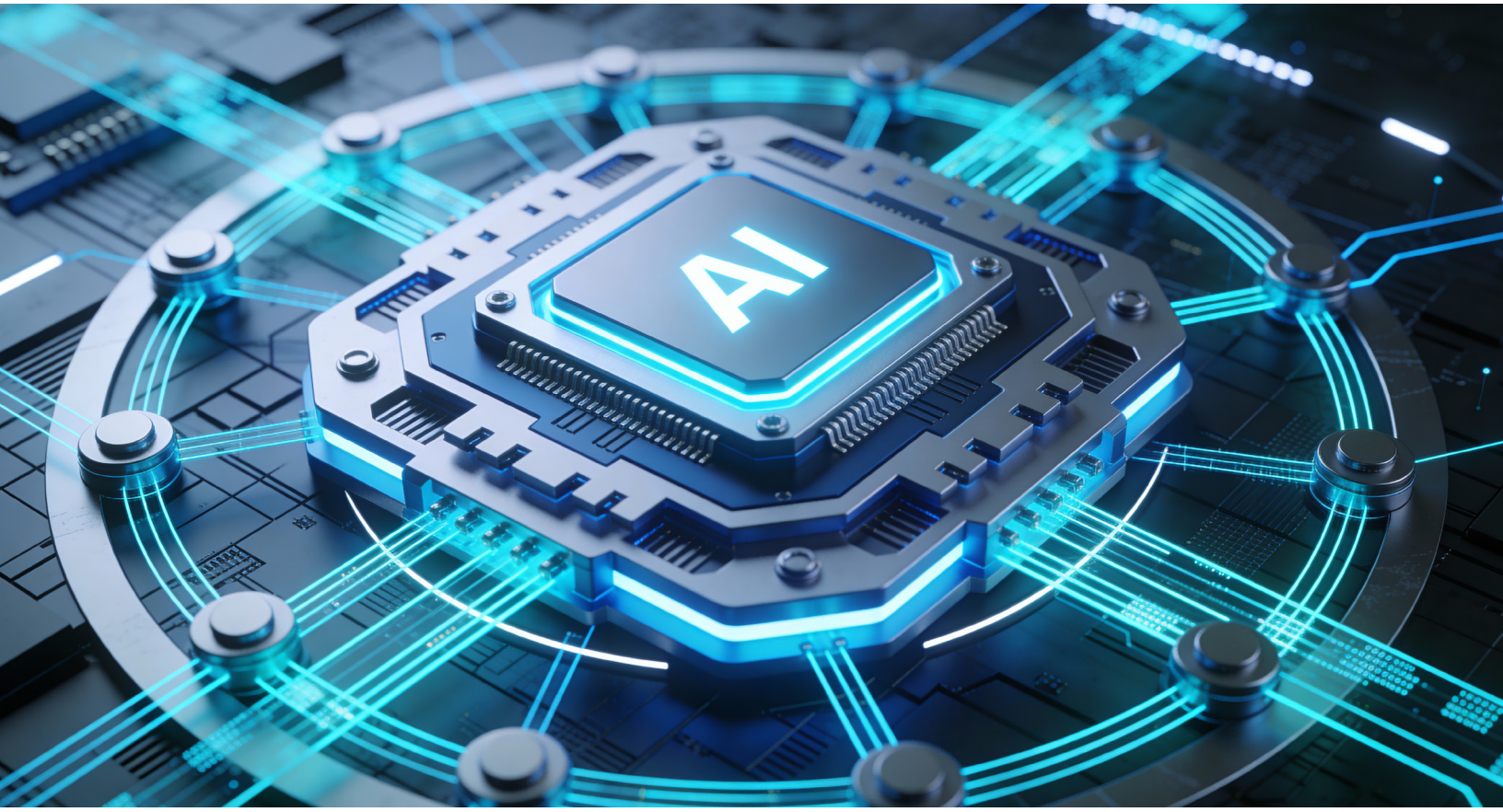
The year truly kicked into high gear on April 8, 2025. This was the day IBM officially unveiled the IBM z17, a system HyperFRAME described as “a quantum leap for transactional AI.” While many expected a simple hardware refresh, the z17 was a statement of intent.

Built on the new Telum II processor, the z17 delivered performance that finally bridged the gap between traditional mission-critical reliability and modern AI demands. With eight 5.5 GHz cores per processor and a 40% increase in L2 cache over the z16, the system isn't just fast, it is optimized for the high-velocity data that lives on the mainframe. HyperFRAME highlighted that the Telum II's on-chip AI accelerator could handle over 450 billion inferencing operations daily. These advancements allow enterprises to perform real-time fraud detection and risk analysis directly in the transaction stream, eliminating the latency and security risks of moving data to the cloud. We have been trying to con the term 'in transaction AI' as a way to articulate how this approach is different from AI deployments in other systems. You can read our deep dive on z17 here or for the real mainframe nerds here is a technical deep dive.

The Q4 Game-Changer: The Spyre Accelerator

If the z17 was the engine, the IBM Spyre Accelerator was the turbocharger. Anticipated throughout the year and finally shipping in Q4 2025, the Spyre Accelerator shifted the conversation from simple inferencing to large-scale Generative AI.

As HyperFRAME noted in its technical deep dives, the Spyre Accelerator is a 75-watt PCIe-connected powerhouse featuring 32 dedicated accelerator cores and 128GB of LPDDR5 memory per card. When scaled to a cluster of eight cards, a single system could command 1 TB of memory and an aggregate bandwidth of 1.6 TB/s. This wasn't just for small models; Spyre was "purpose-built for multiple-model AI," supporting int4, int8, and fp16 datatypes. For the first time, organizations could run Large Language Models (LLMs) and agentic AI directly where their most sensitive data resides.



A Big Year For IBM Mainframe Software

AS IBM made the headlines with the new z17 and Spyre accelerator, the mainframe software business under a newly minted General Manager, Skyla Loomis went from strength tho strength. We covered IBM Concert, watsonx and z/OS 3.2 earlier in the year,and how AI is impacting the mainframe.

IBM Concert for Z - IBM made moves earlier in the year when it launched Concert for Z. The solution aims to deliver an AI-powered IT operations management hub designed to simplify the complexity of managing modern mainframe environments. The platform addresses common challenges like data overload and fragmented visibility by aggregating and correlating operational data into a single, unified interface. By leveraging generative AI and built-in IBM Z expertise, it provides real-time insights and "expert advice" to help teams detect anomalies and identify root causes faster. One of its key features includes automated event correlation, which reduces system noise by grouping and prioritizing critical alerts to prevent service disruptions. The solution also streamlines maintenance by tracking software updates and providing AI-driven remediation recommendations to close mainframe skills gaps. Ultimately, IBM Concert for Z aims to maximize business resiliency and reduce application downtime through proactive risk management and a centralized command console.

Z/OS 3.2 - z/OS continued to evolve to align with the technical underpinnings of the mainframe hardware. IBM launched z/OS 3.2 in July this year. While z/OS releases are hardly moments for revolution these days IBM's z/OS 3.2 represents an evolution of the mainframe, specifically designed to harness the power of the IBM z17 and its on-chip AI acceleration to process over 450 billion inferences daily. By embedding AI directly into the operating system, z/OS 3.2 looks to enable real-time fraud detection and predictive workload management while keeping sensitive data secure on-platform. The release significantly also looks to lower the barrier for a new generation of talent through the introduction of REST APIs, Python support, and cloud-native integrations that align mainframe operations with modern hybrid cloud workflows. Furthermore, it prioritizes long-term resilience by implementing NIST-standard quantum-safe cryptography and AI-driven threat detection to protect against both current and emerging cyber risks. You can read more about HyperFRAME's analysis of the 3.2 release here.

watsonx for IBM Z - In 2025, IBM has significantly matured its watsonx for IBM Z strategy, shifting the conversation from experimental pilots to a deeply integrated "agentic" reality where AI operates natively within the mainframe ecosystem. This evolution is anchored by the general availability of the Spyre Accelerator, a dedicated hardware solution that allows the watsonx Assistant to run Large Language Models directly on the platform to ensure maximum data residency and minimal latency.

The introduction of watsonx Code Assistant for Z 2.8 has revolutionized the modernization lifecycle by utilizing autonomous agents to handle everything from initial dependency mapping to functional verification. Furthermore, by integrating the Granite 4.0 model, fine-tuned specifically for COBOL and Z-specific middleware, IBM is seeking to enable the generation of highly optimized, enterprise-grade code that adheres to strict mainframe performance standards.

IBM also expanded its reach into the most challenging legacy environments by launching support for Assembler code explanation and introducing watsonx Code Assistant for i to address the RPG skills gap. Beyond code generation, the platform now features sophisticated AI-powered Business Rule Discovery, which extracts embedded logic from "black box" applications and transforms it into human-readable documentation. While not IBM Z related this evolution by IBM will only bolster the underlying Granite LLM and give the watsonx a larger footprint for innovation, which will subsequently up level the mainframe space.

To mitigate the risks of refactoring, IBM deployed Automated Validation Assistants that generate and execute test cases for both COBOL and Java to prove functional equivalence between old and new systems. All of these capabilities are overseen by watsonx.governance, which allows highly regulated industries to audit AI decision chains directly within their existing Z management frameworks.

Software Ecosystem: Broadcom MSD, BMC Software Rocket, and the Open Revolution

2025 wasn't just a hardware story. The software ecosystem surrounding the mainframe matured at a breakneck pace. HyperFRAME tracked how the Mainframe ecosystem united Around z17, with key vendors providing day-one support for the new hardware. While z17 support was obviously crucial for the big three of Broadcom, BMC and Rocket, they all made significant investments in new functionality and products in 2025, the highlights of which are outlined below.

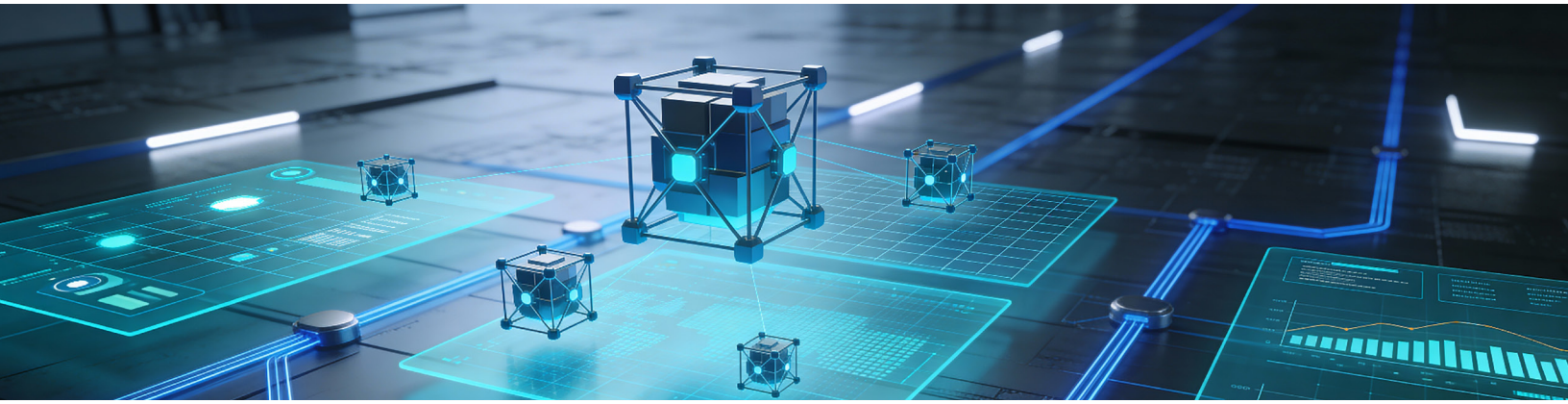
Broadcom

Throughout 2025, Broadcom's Mainframe Software division has executed a sophisticated strategy that acknowledges a fundamental reality of the modern enterprise: the mainframe is no longer an isolated silo, but a core component of a hybrid cloud ecosystem that requires modern talent and seamless integration. By focusing on the intersection of user experience, automated intelligence, and workforce development, Broadcom is attempting to future-proof the platform against the looming challenges of talent attrition and increasing operational complexity.

Modernizing the Interface: Data Management & AIOps

The October 2025 launch of the Centerstage Web UI marks a pivotal moment in Broadcom's data management roadmap. By introducing a modern, consistent interface for database tools—starting with Db2 and slated for broader rollout—Broadcom is addressing the cognitive load that often hinders productivity. From an analyst's perspective, this is less about aesthetics and more about standardizing the operational language of the mainframe. When new and expert DBAs can collaborate within the same visual framework, the "tribal knowledge" of senior staff is more easily transferred to junior members, effectively accelerating the onboarding process and reducing the risk of human error during critical database tasks.

This push toward modernization extends deeply into the AIOps domain with the release of WatchTower 1.3. The update emphasizes "observability with context," moving beyond simple monitoring to provide actionable insights tailored to specific business needs. The New Rule Builder component is particularly significant; by allowing users to build Monitoring Rules directly within dashboards, it empowers IT teams to include or exclude resources based on highly granular criteria. This level of filtering, supported by enhanced time-series and scalar APIs, ensures that system programmers are not overwhelmed by "alert fatigue." Furthermore, the enhanced Alert Rules provide administrators with the flexibility to match alerts to a much broader set of environmental conditions, allowing for a proactive rather than reactive posture in maintaining system uptime.



Bridging the Platform Divide: DevX & Security

In the realm of Developer Experience (DevX), the August 2025 Code4z Expansion is perhaps the most significant bridge built between the mainframe and distributed worlds. By delivering comprehensive Z-language support that integrates directly with Endevor® and Git, Broadcom is meeting developers where they already live: in Visual Studio Code. This move effectively commoditizes mainframe development skills, allowing a developer who is comfortable in a distributed environment to contribute to mainframe projects using familiar extensions and debugging tools. The inclusion of seamless abend analysis within this familiar IDE doesn't just boost productivity; it fundamentally improves code quality by making complex mainframe failures easier to diagnose and resolve for a wider range of personnel.

Parallel to this development agility is a renewed focus on rapid-response security. The April 2025 release of Compliance Event Manager 7.0 addresses the critical "time-to-protection" metric. In an era where ransomware and insider threats move at machine speed, the ability for a customer to activate mainframe monitoring across all LPARs within minutes of installation is a major operational advantage. By offering a blend of out-of-the-box monitors for immediate threat detection and the ability to define custom monitors for specific compliance requirements, Broadcom is providing a security framework that is both accessible for generalists and deep enough for specialized cybersecurity analysts.

The "Beyond Code" Strategy: Solving the Human Capital Crisis

While technical enhancements are vital, Broadcom's most insightful move remains its "Beyond Code" initiative, which treats the mainframe talent shortage as a technical debt that must be paid. The expansion of the Vitality Residency Program in 2025 into Australia, New Zealand, and Europe signals a global commitment to workforce resiliency. This program is a clever piece of ecosystem engineering: Broadcom invests up to a year in recruiting and training new talent at low to no cost for the customer, providing both classroom instruction and hands-on residency mentoring.

By the time a candidate is ready for full-time hire, they have been "vetted" in the customer's own environment, effectively eliminating the high cost and high risk usually associated with mainframe recruitment. Achieving record year-over-year growth in this program suggests that the market is hungry for this type of partnership-first approach. From an industry standpoint, Broadcom is not just selling software; they are manufacturing the skilled labor required to run that software for the next twenty years. This holistic approach to change management and operational optimization suggests a long-term vision where the vendor's success is inextricably linked to the operational health and staffing stability of the client.

Rocket Software

Back in May, we covered Rocket Software's announcement focusing on bridging the IT skills gap and enhancing mainframe performance by infusing AI and automation across its core

portfolio. To boost developer productivity, the company introduced AI-powered assistants for MultiValue and Uniface that use generative AI to explain, autocomplete, and generate code, potentially reducing onboarding time from months to weeks. On the operations side, Rocket TMON now utilizes machine learning to proactively identify performance anomalies and KPIs before they impact system stability. Additionally, Rocket Zena empowers non-technical users to automate their own processes through self-service tools, while Rocket EDX streamlines document management using natural language search capabilities. These innovations collectively aim to modernize the mainframe experience and strengthen system resilience without requiring organizations to undergo risky, wholesale infrastructure replacements.

Next Evolution DevOps Tools

Rocket Software has announced a new generation of DevOps innovations specifically designed to enhance the developer experience (DevEx) across mission-critical systems like IBM Z and IBM i. By integrating AI-driven insights and deeper automation, the suite aims to bridge the growing IT skills gap by making heritage systems more accessible to modern developers. These tools allow teams to utilize familiar environments such as Git, VS Code, and Azure Pipelines while maintaining the stability of core mainframe infrastructure. The release emphasizes a unified approach to DevOps that spans across hybrid cloud, distributed, and on-premises platforms. Key features include AI-powered code explanation for COBOL and proactive anomaly detection to reduce late-stage debugging costs. Ultimately, the initiative seeks to help organizations accelerate value delivery without being hindered by outdated tools or talent shortages.

Rocket DataEdge Hybrid Cloud Integration

Earlier in 2025, the introduction of Rocket DataEdge marked a significant step in unifying data across mainframe, distributed, and cloud environments to support advanced AI and analytics. This structured data suite addresses a critical bottleneck where only a fraction of IT leaders currently capitalize on core transactional data for their AI models. By providing seamless discovery and real-time synchronization, DataEdge is built to enable organizations to build high-quality, scalable datasets from disparate sources. The suite incorporates tools for data replication, intelligence, and virtualization to minimize the complexity and errors typically associated with hybrid infrastructure. This unified metadata management approach ensures that decision-makers have a complete, timely view of their business operations. Consequently, the solution empowers enterprises to execute forward-looking data strategies that maximize competitive advantage in a rapidly growing hybrid cloud market.

Phishing-Resistant Secure Host Connectivity

Rocket Software, at the beginning of the year, launched Rocket Secure Host Access, a security-first terminal emulator designed to protect “green screen” applications from modern cyber threats like phishing and fake worker attacks. The solution is unique in its ability to integrate legacy host connectivity with contemporary Identity and Access Management (IAM) systems, supporting Multi-Factor Authentication (MFA) and Single Sign-On (SSO). This development is particularly timely given the rising costs of data breaches and the introduction of strict regulatory mandates like DORA and 23 NYCRR Part

500. By centralizing management and providing password-less access, the tool helps organizations secure their most sensitive mainframe data with the same rigor as the rest of their IT estate. It offers both desktop and zero-footprint web options to ensure high availability and flexibility for diverse workforces. Overall, the solution serves as a critical bridge for enterprises looking to align their heritage systems with modern cybersecurity standards.

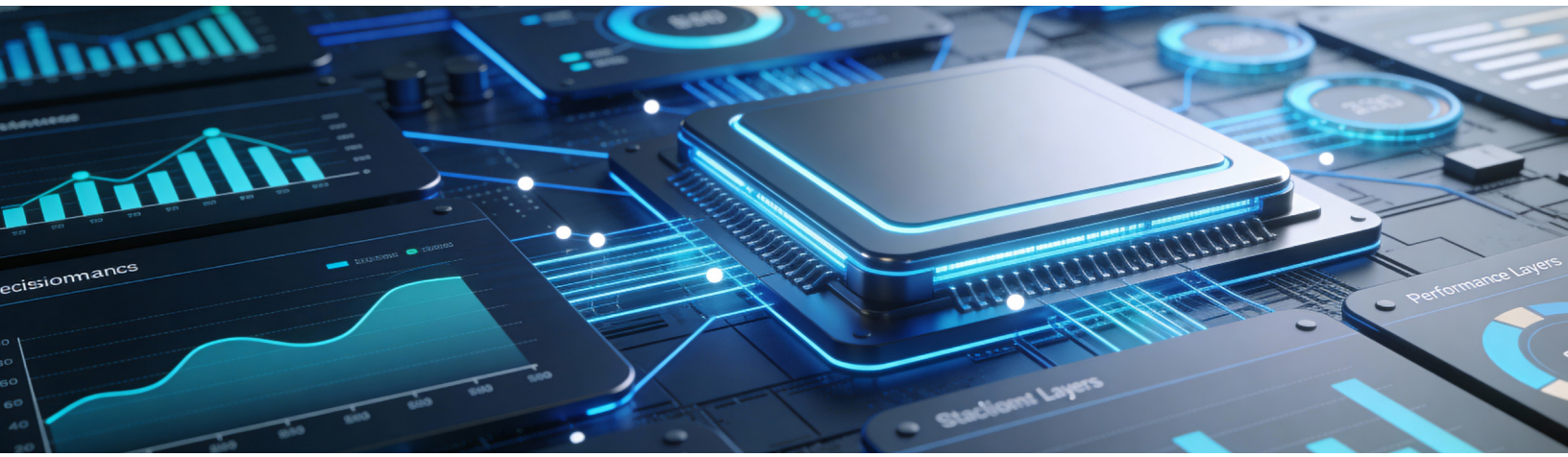
BMC

BMC entered 2025 leaner and more focused on the mainframe after it announced late in 2024 that it was splitting into two companies. This renewed focus has been evident in how the company continues to innovate on a quarterly cadence. We covered the 44th quarter of innovation in a report in April.

The 2025 BMC Mainframe Survey (the 20th the company has run over the years) highlighted five critical metrics that stood out, and for me at least, define the current state of the industry, starting with a record 97% of respondents who now view the mainframe as a viable, long-term platform for new growth. This confidence is bolstered by a massive generational turnover, with Millennials and Gen Z now comprising 66% of the mainframe workforce, effectively debunking the myth of a "skills gap" crisis. This new workforce is managing an expanding environment, as 72% of organizations report an increase in general-purpose capacity driven by new application development. Modernization is also happening at a rapid pace, with 65% of organizations already implementing generative AI and 72% of shops now utilizing Java as their primary language for writing new mainframe applications.

From our conversations in 2025, BMC is strategically looking to reshape the mainframe experience by prioritizing the seamless integration of artificial intelligence directly into the user's existing environment. A cornerstone of this evolution is the BMC AMI Assistant, which leverages AI-driven knowledge expert chat capabilities across both operations and development solutions to provide real-time, contextual support. Perhaps most impactful is the AI-driven code conversion capability, which stands as a premier tool for modernization. The true insight lies in the architecture of this rollout; rather than treating AI as a standalone product, BMC "infuses" it into the daily workflows of developers and operators. This approach ensures that sophisticated insights are as accessible as a simple right-click, drastically reducing friction and saving time. Furthermore, the strategy empowers organizations through flexibility, offering a curated library of Large Language Models (LLMs) optimized specifically for mainframe logic while simultaneously supporting a Bring Your Own LLM (BYOLLM) model for bespoke enterprise needs.

Another thing that stood out to HyperFRAME in 2025 was BMC AMI Cloud Datasets, the technology acquired by BMC when it bought Model9 a few years ago, and how it is enhancing data management. BMC AMI Cloud Data Sets (CDS) represent a major shift in mainframe storage by allowing applications to read and write data directly to cloud object storage without requiring any changes to existing Job Control Language (JCL). By emulating a virtual Tape Library, this solution enables organizations to bypass expensive physical and virtual tape hardware, leading to significant reductions in infrastructure and licensing costs. The technology is designed for high efficiency, utilizing zIIP engines to offload CPU-intensive tasks and minimize general-purpose MIPS consumption. Beyond simple storage, it looks to accelerate operational windows with backup speeds up to 16 times faster (according to BMC internal benchmarks) than traditional methods, while providing modern security through immutable backups and encryption. Ultimately, CDS aims to bridge the gap between legacy environments and modern architecture, turning siloed mainframe data into a flexible resource ready for cloud-based AI and analytics.



Incremental Modernization and COBOL Transformation

Enable low-risk, incremental modernization while expanding your available talent pool and optimizing critical workloads. By selectively transforming legacy COBOL, while meticulously preserving decades of proven business logic, you can bridge the gap between traditional mainframe reliability and modern agility.

Utilizing AI-powered code explanation and automated refactoring, these solutions translate complex syntax into human-readable insights. This process yields sustainable, high-quality Java code through the combined power of BMC AMI Assistant and BMC AMI DevX, allowing your organization to transition at its own pace without the "rip-and-replace" risks typically associated with core system updates.

Knowledge Empowerment and Talent Onboarding

Accelerate problem-solving, bridge persistent skills gaps, and onboard new talent significantly faster with the innovative knowledge expert capability of BMC AMI Assistant. This tool serves as an on-demand mentor, delivering trusted, highly contextual answers to complex mainframe questions directly within the developer's existing workflow.

By eliminating the need to toggle between disconnected documentation and terminal screens, it empowers cross-functional teams to maximize mainframe performance and drive superior business outcomes. This shift ensures that even early-career engineers can navigate the mainframe with the confidence of a seasoned expert, securing the future of your platform.

Hybrid Visibility and Enterprise Observability

BMC is looking to drive comprehensive visibility across your increasingly complex, hybrid environments to resolve performance bottlenecks faster and facilitate proactive, data-driven decision-making. By integrating mainframe operations metrics into leading enterprise observability platforms, such as Elastic and Splunk enterprises can break down traditional data silos.

This is achieved through the implementation of OpenTelemetry (OTel) within BMC AMI Ops, providing a unified "single pane of glass" for your entire infrastructure. This standardized approach allows IT operations teams to monitor mainframe health alongside cloud-native applications, ensuring end-to-back resiliency and faster root-cause analysis across the modern enterprise stack.

If you want to delve deeper into BMC solutions, check out this video with BMC execs, Priya Doty and Matt Whittbourne.

The Summer of the Emperor: LinuxONE 5

In May, at the IBM Think conference, the focus shifted to the LinuxONE 5 Emperor, read HyperFRAME's coverage of the announcement here. Celebrating the 10th anniversary of the LinuxONE brand, this release proved that the s390x architecture remains the premier "Linux Fortress."

The Emperor 5, also powered by Telum II, focused heavily on the "custody of digital assets." HyperFRAME noted that while these deployments aren't yet numerous, the unique security requirements of digital asset infrastructure make LinuxONE "almost uniquely qualified" to support them. With 99.999999% availability and the ability to leverage the Spyre Accelerator for secure AI-driven Linux workloads, the Emperor 5 solidified its place in the hybrid cloud.

Digital Asset Haven - Mainframe Tech For Crypto Custody

The LinuxONE hype train continued later in the year with the launch of Digital Asset Haven, a comprehensive, enterprise-grade platform specifically designed to help financial institutions and governments scale their digital asset operations with confidence. Developed in a strategic partnership with the wallet infrastructure provider Dfns, the solution unifies the often fragmented stages of custody, transaction execution, and settlement into a single, cohesive workflow. A primary technical pillar of this launch is the implementation of Multi-Party Computation (MPC), which cryptographically splits keys to ensure there is no single point of failure in the management of digital funds. This security is further hardened by leveraging IBM Z and LinuxONE hardware, utilizing the Crypto Express 8S HSMs to perform transaction signing within a physically tamper-resistant environment. According to the Research Note from HyperFRAME Research, the platform's Transaction Lifecycle Management feature is a standout, providing automated routing and monitoring across more than forty different public and private blockchains. To assist with the complex regulatory landscape, the platform includes pre-integrated services for essential compliance tasks such as identity verification (KYC) and anti-money laundering (AML) monitoring. HyperFRAME highlights that by making the solution available through hybrid SaaS and on-premises models, IBM is directly addressing the stringent data residency and sovereignty requirements of the world's most regulated entities. Ultimately, this launch is framed as a strategic move to unlock the trillion-dollar digital asset custody market by offering the same level of resilience and governance that institutions have traditionally expected from "Big Iron" banking rails.

Replatforming to the Cloud

AWS is coming for the Mainframe platform, and during the re:Invent show went as far as to literally blow up an old server rack to highlight the challenge with technical debt.

AWS Transform for Mainframe workloads represents a shift toward "agentic AI" in the cloud migration space. This service utilizes specialized AI agents to automate the complex, manual stages of modernization, including the reverse engineering of monolithic COBOL codebases into cloud-optimized Java microservices. According to our analysis, the introduction of the "Reimagine" capability allows organizations to not just lift-and-

shift, but fundamentally re-architect applications by extracting business logic and generating microservice specifications. The platform further streamlines the migration journey with automated testing features that generate test plans and validation scripts, significantly reducing the risk of regressions. HyperFRAME's analysis highlights that these AI-driven workflows can accelerate modernization timelines from years to just months, as demonstrated by early adopters like Toyota and BMW. A key technical differentiator is the "self-debugging" cycle, where the AI agents iteratively analyze modernization failures and refactor their own approach until they achieve functional equivalence. Furthermore, the service is architected to be composable, allowing enterprises to integrate their own proprietary documentation and institutional best practices into the AI's transformation patterns. By reducing technical debt through such high-fidelity automation, AWS Transform aims to free up developer bandwidth for innovation rather than maintenance. Ultimately, HyperFRAME views this launch as a critical milestone that helps remove the historical "knowledge silos" and psychological barriers associated with exiting the mainframe platform.

AI-Powered Replatforming Efforts Gain Traction

The modernization of what a myriad AI-powered replatforming and refactoring vendors call 'legacy workloads' or the "black box" COBOL, JCL, REXX, and PL/1 systems that still power 70% of global financial transactions, is undergoing a paradigm shift. Moving away from risky manual rewrites or stagnant "lift-and-shift" migrations, a new cohort of vendors, including Bloop, Mechanical Orchard, Astadia, CloudFrame, and Heirloom Computing is leveraging AI to turn refactoring and subsequent replatforming to the cloud into a data-science discipline.

The primary challenge of so-called legacy systems is not just the ancient syntax, but the "lost intent" buried within millions of lines of undocumented code. Bloop is aiming to address this at the discovery layer. By applying semantic search and Large Language Models (LLMs) to legacy codebases, Bloop allows architects to treat a mainframe like a searchable database. This analytical layer is critical because it identifies the "dead code" and hidden dependencies that often cause migration projects to spiral in cost. Instead of guessing which modules are vital, developers can use Bloop to map the functional topography of the system before a single line of code is moved.

Where Bloop provides the map, Mechanical Orchard provides a new philosophy of replication. Their "Behavioral Twin" approach is a sophisticated departure from traditional transpilation. By using AI agents to observe a system's live execution, they capture the output behavior rather than just the source code. This is an analytical breakthrough: it acknowledges that the original source code may no longer reflect the reality of how the system functions in production. By recreating these behaviors in modern, cloud-native environments, Mechanical Orchard bypasses the "garbage in, garbage out" risk inherent in translating old, inefficient COBOL directly into Java.

CloudFrame and Astadia focus on the bridge between deterministic precision and generative flexibility. Legacy systems require zero-tolerance for error, a requirement that traditional GenAI often fails due to hallucinations. CloudFrame solves this through "Agentic AI" that operates within a controlled framework, using a cross-model approach to refactor code into Java that is both functionally identical and architecturally modern. Astadia complements this with its FastTrack automation, which uses AI to handle the "beautification" of code. In traditional migrations, automated tools often produced "Jobol", Java that still looked and acted like COBOL. Astadia's AI-driven approach refactors the logic into truly idiomatic, object-oriented code that modern developers can actually maintain.

Finally, Heirloom Computing bridges the gap between the old world and the new through high-velocity transpilation to the Java Virtual Machine (JVM). Their AI-enhanced platform allows for an "incremental" modernization strategy. This is analytically superior to a "big bang" migration because it allows legacy logic to run on modern infrastructure while being refactored piece-by-piece. By placing the legacy code on a modern stack immediately, Heirloom enables companies to stop paying "mainframe taxes" (MIPS costs) while they take their time with deeper AI-driven refactoring.

Together, these companies represent a shift toward Deterministic AI. By combining the pattern-matching capabilities of LLMs with the strict rules of compilers and behavioral observation, they are effectively "de-risking" the most complex IT challenge of the 21st century. They are not just moving code to the cloud; they are reclaiming the business logic trapped in the mainframe, making it readable, scalable, and ready for the next generation of digital innovation.

Workloads Are Like Journeys

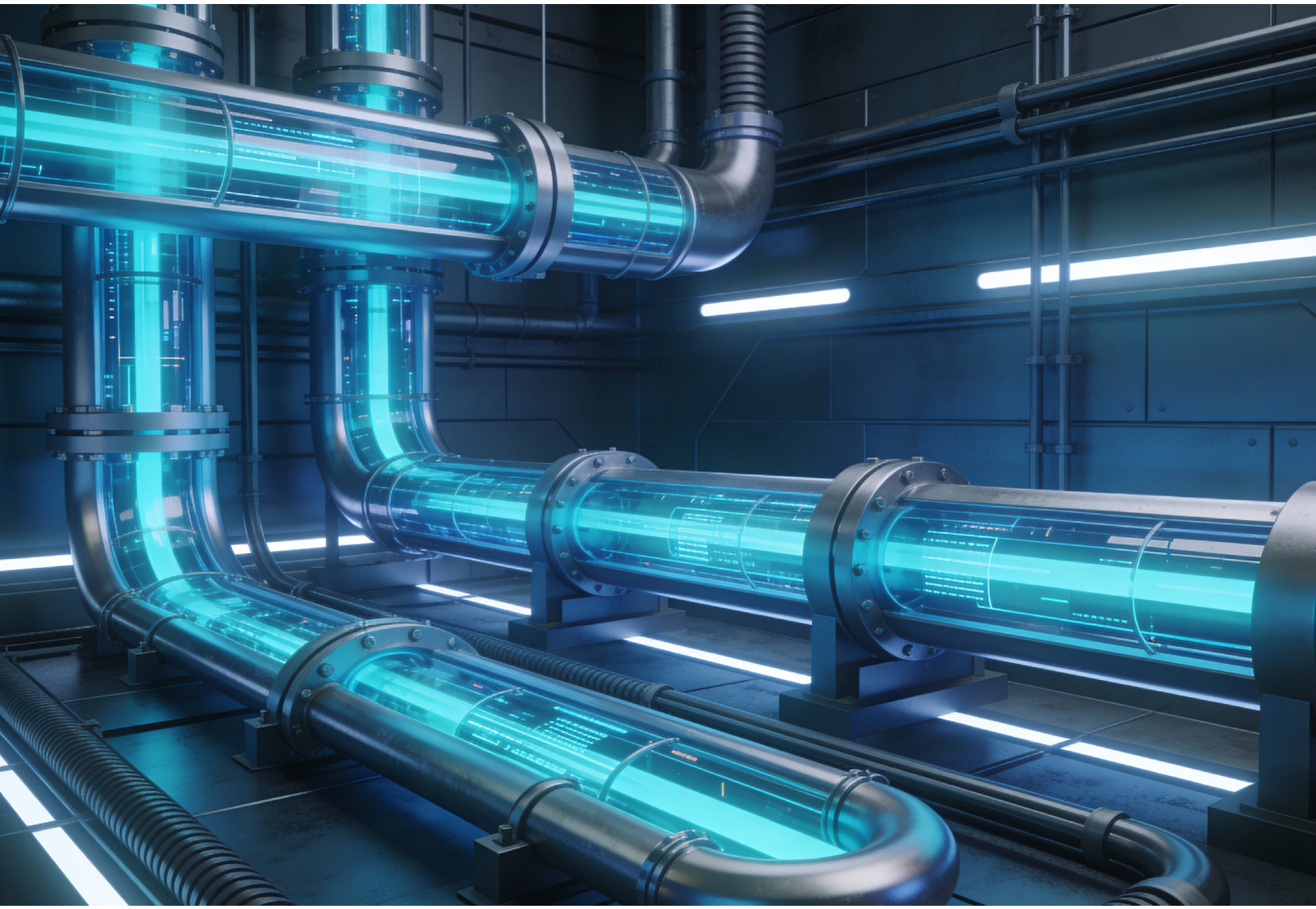
A long-held belief by HyperFRAME is that in the world of enterprise computing, managing application workloads is remarkably similar to orchestrating a global transportation network. Running a standard x86 server is akin to owning

a personal car; it offers great individual flexibility and a low entry cost, but it lacks the structural efficiency for mass-scale movement. To further stretch the analogy, renting a car is akin to consumption-based models like HPE GreenLake, Dell APEX, or Lenovo's TruScale. Going even further, the public cloud compute instances would be analogous to an Uber or Lyft in this analogy framework.

For those who need temporary or scalable power, the public cloud acts as a rental car service, providing the same underlying x86 architecture with the convenience of on-demand access without the long-term commitment of ownership. However, just as you wouldn't try to move an entire city's population using thousands of individual sedans, high-volume transactional workloads require a different class of infrastructure altogether. The IBM mainframe is the high-speed rail of the data center, designed to move massive volumes of "passengers", or data packets, with incredible throughput and a predictable schedule that never deviates. For the most critical global operations, the mainframe evolves into the equivalent of a commercial airliner, where every single critical system, from the engines to the flight computers, is built with dual or triple redundancy to ensure the journey never stops.

Just as an airplane is designed for "continuous availability" the z17 architecture allows for hardware repairs and upgrades without ever landing the application. Ultimately, while cars and rental services are perfect for personal errands and point-to-point agility, the "Big Iron" remains the heavy-lift transport system that keeps the global economy in motion.

Put simply, workloads are like journeys, and just because a car exists doesn't mean planes and trains are invalid forms of transport.



Looking Ahead

Retrospective analysis of the mainframe in 2025 reveals a fundamental paradigm shift: the mainframe has transitioned from a target of modernization to the primary engine of enterprise-wide transformation. The industry has moved past the "legacy" trope, recognizing that the most efficient path to digital maturity is not through platform replacement, but through the strategic orchestration of the mainframe's unparalleled throughput and security.

The dual introduction of the IBM z17 and the Spyre Accelerator effectively rebranded the platform as a high-density "AI Factory," capable of localizing generative AI within the same cryptographic boundary as the world's most sensitive data. By synthesizing Broadcom's push for open observability standards, Rocket Software's dev-centric modernization frameworks, and IBM's specialized silicon, the ecosystem has validated a "gravity-based" data strategy, performing complex inferencing exactly where the records of truth reside.

At the heart of this hardware revolution was the Telum II processor, which became the year's most significant chip by integrating AI acceleration directly into the compute core to eliminate the "latency tax" associated with external GPUs. This was bolstered by the Spyre Accelerator, the year's hottest hardware expansion, which utilized 32-core clusters to enable Retrieval-Augmented Generation (RAG) and LLM execution directly on-platform.

From a software perspective, the "island" status of the mainframe was finally dismantled by the widespread adoption of OpenTelemetry (OTel) and open-source standards, a trend that unified Z observability with modern cloud-native stacks.

In the mainframe ecosystem, two seemingly contradictory trends are coexisting to define a new era of hybrid computing. On one hand, overall MIPS consumption is growing as enterprises integrate artificial intelligence directly into their core transactional workflows. This growth is driven by the need for real-time inference on the mainframe, allowing businesses to detect fraud or authorize payments without the latency of moving data off-platform. Simultaneously, applications that are no longer a strategic fit for the mainframe are increasingly being migrated to cloud-native environments.

This off-ramping is being accelerated by AI-powered refactoring tools that can automatically translate legacy code into modern, scalable microservices. Consequently, the mainframe is becoming more specialized and powerful for high-value tasks while shedding technical debt that belongs elsewhere. This dual movement creates a "best-of-both-worlds" architecture where the mainframe acts as a high-performance hub within a broader cloud ecosystem. Rather than a total migration or a static status quo, the industry is witnessing a sophisticated rebalancing of workloads based on performance and cost-efficiency. Ultimately, the synergy between mainframe growth and cloud expansion proves that modernization is not a zero-sum game but a strategic alignment of technology to purpose.

Put simply, two truths are true at the same time, and those who hold only one truth to be true are missing the whole picture. The mainframe is growing, and workloads are moving off that aren't suited to the platform. AI is powering both truths!

ABOUT HYPERFRAME RESEARCH:

HyperFRAME Research delivers in-depth research and insights across the global technology landscape, spanning everything from hyperscale public cloud to the mainframe and everything in between. We offer strategic advisory services, custom research reports, tailored consulting engagements, digital events, go to market planning, message testing, and lead generation programs.

Our industry analysts specialize in rigorous qualitative and quantitative assessments of technology solutions, business challenges, market forces, and end user demands across industry sectors. HyperFRAME Research collaborates closely with your Analyst Relations, Product, and Marketing teams to build and amplify your thought leadership, positioning your expertise to enhance brand and product recognition. Through content that engages readers, viewers, and listeners alike, we ensure your voice resonates across channels.

CONTACT HYPERFRAME RESEARCH:

Steven Dickens
CEO & Principal Analyst | HyperFRAME Research

Email Address:

steven.dickens@hyperframeresearch.com

Telephone Number:

+1 845 505 1678

X: @StevenDickens3

LinkedIn: Steven Dickens

BlueSky: Steven Dickens

CONTRIBUTORS

Steven Dickens
CEO & Principal Analyst | HyperFRAME Research

INQUIRIES

Contact us if you would like to discuss this report and HyperFRAME Research will respond promptly.

CITATIONS

This paper can be cited by accredited press and analysts, but must be cited in-context, displaying author's name, author's title, and "HyperFRAME Research." Non-press and non-analysts must receive prior written permission by HyperFRAME Research for any citations.

LICENSING

DISCLOSURES

HyperFRAME Research provides research, analysis, advising, and consulting to many high-tech companies, including those mentioned in this paper. No employees at the firm hold any equity positions with any companies cited in this document.


HyperFRAME
RESEARCH