

Banking AI Needs Core Data. Most Banks Still Can't Reach It.

How Zafin and VirtualZ help banks move usable mainframe data into modern analytics and AI environments — without disrupting the core mainframe system of record.



Executive summary

Every bank is racing to put AI to work. The obstacle is rarely the models — it's the data. Much of what would make banking AI valuable (balances, transactions, customer history and operational context) still lives on the mainframe, encoded in a decades-old format that modern analytics and AI tools cannot easily read directly. For the platforms that serve these banks, reaching that data has often meant slow, brittle, multi-step extraction pipelines. That data-access gap, not the algorithms, is one of the most practical constraints on scaling banking AI today.

This paper describes how Zafin, working with VirtualZ Computing, is helping close that gap in production. Through Zafin IO, Zafin connects legacy and modern systems so regulated institutions can modernize progressively without disrupting critical operations. With VirtualZ's PropelZ capabilities, Zafin can decode mainframe data that modern tools cannot easily read, move usable copies into modern data, analytics and AI environments — and land it in a form its teams can query directly. From there, Zafin layers AI-assisted investigation on top — collapsing a slow, multi-step process into a more streamlined workflow and helping turn incident investigation from an hours-long effort into a minutes-long one for a top-tier bank.

The result is a repeatable model for how banks and their platform partners can make core data AI-ready and usable in modern analytics environments — without a disruptive, multi-year core migration or application rewrite.

The data your AI needs most is the hardest to reach

This is not a niche problem — it sits on top of the infrastructure that runs global finance. According to the IBM Institute for Business Value (in collaboration with Oxford Economics, survey of 2,551 global IT executives), the

mainframe powers 70% of global transactions on a value basis.¹ IBM further reports that 77 of the world's top 100 banks depend on the mainframe for their operations.² This is where the industry's most valuable data already is.

And the mainframe platform is not declining — it is being actively invested in. The global mainframe modernization market is projected to grow from USD 8.39 billion in 2025 to USD 13.34 billion by 2030, a compound annual growth rate of 9.7%, driven by demand for agility, resilience, and digital innovation as enterprises integrate legacy core systems with cloud, AI, and APIs.³

Banks are not retiring the mainframe; they are connecting it to the cloud. In BMC's 2025 Mainframe Survey, among financial-services firms prioritizing cloud technologies, 47 percent reported connecting the mainframe to cloud-based workloads.⁴ The mainframe is becoming a first-class source for modern analytics and AI, and the addressable opportunity is to make its data usable at that scale.

But the same longevity that makes the mainframe reliable also makes its data hard to use: much of it is stored in files whose layouts were historically described by hand-maintained COBOL copybooks — a structure that modern SQL engines, cloud data platforms, and AI models were never designed to read. The data is both indispensable and, for AI purposes, largely out of reach.

That creates a daily operational tax. When something goes wrong — a reconciliation break, a disputed fee, a data-quality question — the teams responsible for resolving it often cannot simply query the underlying mainframe files. They must first extract the data, decode it, transform it, and stage it somewhere usable, frequently across several tools and several hours, before investigation can even begin. Every step adds fragility, latency, and cost, and every hour spent decoding is an hour not spent resolving.

¹ IBM Institute for Business Value (in collaboration with Oxford Economics), "AI drives mainframe innovation," IBM, November 18, 2025. <https://www.ibm.com/think/insights/ai-drives-mainframe-innovation>

² IBM, "Transforming mobile banking with real-time data," IBM, November 19, 2025. <https://www.ibm.com/new/product-blog/mobile-banking-with-real-time-data>

³ MarketsandMarkets, "Mainframe Modernization Market worth \$13.34 billion by 2030," August 21, 2025. <https://www.marketsandmarkets.com/PressReleases/mainframe-modernization.asp>

⁴ BMC Software, "2025 Mainframe Survey Key Financial Services Takeaways," September 16, 2025. <https://www.bmc.com/blogs/mainframe-survey-financial-services-key-takeaways/>

For the fintech and modernization platforms that serve these banks, the problem compounds. They inherit the bank's mainframe data as a critical input, and their ability to deliver value — analytics, reconciliation, AI-driven insight and banking execution — is gated by how quickly and reliably they can make that data readable. The pipeline becomes the bottleneck, and the AI on the other end of it only ever sees a fraction of what it could.

INDUSTRY EVIDENCE

Industry research bears this out: in Celent's 2025 Decision Intelligence and Data Survey, **60% of Tier-1 bank responses** identified existing/legacy data warehouse architectures as a barrier to improving decision intelligence, even as those banks migrate to more modern cloud-based data platforms.⁵

The data exists. Reaching it is the problem.

The approach: decode at the source, query everywhere

Zafin's insight was simple: mainframe data does not need to be trapped behind a long migration effort before it can create value. Working with VirtualZ, Zafin can decode it, move usable copies into modern analytics environments, and make it queryable for the teams handling reconciliation, investigation and AI-enabled analysis — while the IBM Z core remains the trusted system of record.

Working with VirtualZ, Zafin built an internal application — Zafin File Explorer — that uses PropelZ underneath to decode mainframe EBCDIC files directly and give teams a faster way to inspect and query the data. When a bank's file arrives, the tool matches it to the right schema, spins up PropelZ to decode it, and lands the result in a form the team can immediately work with. For large

⁵ Celent, "Decision Intelligence and Data Survey 2025" (n=106), cited in Celent/InterSystems, "It's All About the Data!" 2025.

files, it runs multiple PropelZ instances in parallel, splitting the file by record length so that datasets of millions of rows open in minutes rather than hours.

Getting the schema right used to be the hard part. Traditionally, reading a mainframe file meant hand-building and maintaining a COBOL copybook to describe every field — finicky, expertise-dependent work that breaks the moment a source layout changes. PropelZ increasingly removes that burden through automatic metadata discovery: for database sources it reads the structure directly, needing no copybook at all, and where a copybook is useful it can generate one automatically rather than requiring it to be engineered by hand. The pipeline adapts to the data instead of the team adapting the data to the pipeline.

Once the data is decoded, it is simply data. Zafin's teams can run SQL queries against it, filter and join across multiple files, and — through an AI layer Zafin built on top — ask questions in natural language and get answers back as executable queries. A support engineer investigating an incident no longer needs to know the physical layout of a mainframe file; they can ask what they need to know and interrogate the answer.

Crucially, this requires no change to the bank's mainframe applications and no rip-and-replace. PropelZ is configuration-driven: when a new data attribute is needed, the team changes the configuration rather than writing new extraction code. The mainframe keeps doing what it does best; the data it produces simply becomes usable — and AI-ready — in the modern environments where Zafin and the bank need to work with it.

"With PropelZ, we can move usable, decoded copies directly so our teams can perform reconciliation and analytics. That removes several stages from the pipeline and gives us a more direct path from core data to operational insight. We're seeing stronger performance and handling larger transaction volumes with room to spare. For banks, the value is clear: they can preserve the IBM Z system of record while making critical data more usable for modern analytics, AI-enabled investigation and banking execution. Many of the banks we serve run on mainframe systems, and we're their modernization partner — this is a core part of how we deliver that."

Shahir Daya • Chief Product & Technology Officer, Zafin

The results: from hours to minutes, and from extraction to insight

For the top-tier banks Zafin serves today, the change is tangible. Investigation work that once required a slow, multi-stage extraction now begins almost immediately: a multi-million-row file opens in minutes, and the team moves straight to querying it. The pipeline that once stood between a question and its answer is now significantly more direct.

The business value shows up in four places:

- **SPEED TO RESOLUTION.**

Incident investigation compresses from an hours-long, multi-tool effort into a fast, single-environment workflow. When a bank operates against tight service-level commitments, hours recovered before a deadline are the difference between a non-event and a breach.

- **FEWER MOVING PARTS.**

Removing three to four steps from the pipeline removes three to four things that can fail, slow down, or require specialized knowledge to maintain. A simpler pipeline is a more reliable one.

- **A STRAIGHT LINE INTO THE ANALYTICS ENVIRONMENT.**

Zafin validates fees and rates, reconciles the bank's data, and builds its dashboards — removing the intermediate staging-and-export hop entirely. The teams serving the bank spend less time moving data and more time acting on it, and the data arrives sooner in the place where the work actually happens.

- **AI LEVERAGE OF DATA THAT WAS PREVIOUSLY OFF-LIMITS.**

Once mainframe data is decoded and queryable, it becomes usable by the same AI tooling banks are investing in everywhere else — turning a historically opaque data source into one that can be interrogated in natural language.

For Zafin, the approach scales to support its customer base. Nearly every bank it serves runs a mainframe, and the same pattern — move usable data into modern environments, make it queryable and layer AI-assisted investigation on top — is repeatable across them. What began as a way to serve one bank can now be offered to other customers with comparable needs.

Why it works: PropelZ as the decoding engine

PropelZ is purpose-built for exactly this problem: reading legacy mainframe data structures and delivering them, faithfully and at speed, into modern targets. It handles the full range of mainframe file types, and — through automatic metadata discovery — it minimizes the traditional dependence on hand-built COBOL copybooks, reading a source's structure directly or generating the layout automatically. It moves the decoded data to modern destinations — relational databases, cloud object storage, and, natively, Databricks — with a modern, efficient architecture rather than the staging-heavy design of older ETL tools. Delivering data straight to the analytics platform, rather than through a chain of intermediate hops, is exactly the kind of shortcut that older ETL designs cannot make.

Two properties make it a fit for banking specifically. First, it is designed for high throughput on large volumes, with parallelism that lets a single large file be decoded across multiple workers. Second, on the mainframe itself, its processing is largely eligible for specialty-engine (zIIP) offload, meaning heavy decoding workloads can run without inflating the bank's mainframe software costs — a material consideration when the data volumes run to a terabyte a day.

The result is an engine that a platform like Zafin IO can build on with confidence: fast enough for production volumes, faithful enough for financial data, and economical enough to run at scale.

What this means for banks and their platform partners

The Zafin collaboration points to a broader pattern for making core data AI-ready — one that sidesteps the risk and multi-year timeline of core replacement.

Banks do not have to choose between keeping the mainframe and feeding their AI. The mainframe can remain the reliable system of record while its data becomes fully available to modern analytics and AI, decoded on demand rather than trapped behind a brittle pipeline. And the platforms and systems integrators that serve banks can offer this as a capability — faster investigation, AI-ready data and a simpler integration path — without asking the bank to rewrite core applications or destabilize critical operations.

For the senior technology leaders weighing where to spend AI and modernization budget, the appeal is that this is incremental, low-disruption, and already proven in production at a top-tier institution. It is not a bet on a future state; it is a capability available now.

About the partners



Zafin is an AI platform company for regulated institutions. Built from more than 20 years of operating in demanding regulated technology environments, Zafin brings structure, intelligence and accountability to how critical work, systems and decisions are created, governed and improved. Zafin's portfolio includes Zafin AIOS, the control plane for governed agentic work; the Zafin Banking Platform, the AI-powered banking execution platform for product, pricing, offers, billing, loyalty, relationship and deal logic; and Zafin IO, the integration platform for regulated institutions.

In this work with VirtualZ, Zafin IO provides the strategic integration frame: connecting legacy and modern systems so institutions can modernize progressively without disrupting critical operations. The resulting data can support analytics, reconciliation, AI-enabled investigation and downstream banking execution workflows across the Zafin Banking Platform.

zafin.com



VirtualZ Computing is the maker of PropelZ, a mainframe data access and modernization product that decodes legacy mainframe data — including EBCDIC files — and delivers it, at speed, to the modern databases, cloud platforms, and analytics tools where banks and their partners put it to work. PropelZ's automatic metadata discovery minimizes the traditional dependence on hand-built COBOL copybooks, so teams can move data with a SQL query rather than a copybook engineering exercise.

virtualzcomputing.com