The new math of mainframe modernization

Why modernization is no longer a deferrable investment — and what that means for capital markets firms in 2026

Phillip Matricardi
Deloitte Consulting LLP

For twenty years, technology and operations modernization has been an argument capital markets firms kept winning by deferring. The mainframe was fully depreciated, the cost to replace it was high, and client-facing pain, while real, was manageable. The math favored patience.

The math has changed.

Three shifts have happened at once. AI-assisted engineering has collapsed the cost of the foundational work — data liberation, API enablement, mainframe disentanglement — that has gated every modernization program of the last two decades. The convergence of regulatory mandates and market-structure reforms — T+1 settlement, Treasury clearing, 24x5 processing — has converted what used to feel like theoretical pressure into concrete, dated deadlines. And firms that have begun this work are building new capabilities at an accelerating pace — each improvement enabling the next — Executive decision makers have been told for years that "now is the moment" and are entitled to skepticism. This moment is different, and the reason is specific and testable.

AI is attacking the hard part of modernization

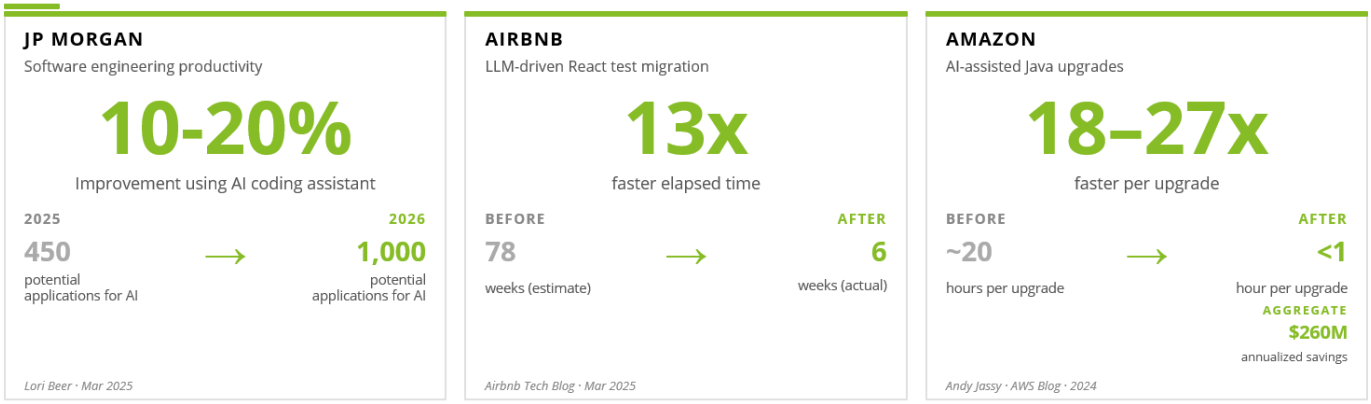
Prior waves of tech transformation — the public cloud wave most recently — may have underwhelmed because the easy parts were easy and the hard parts stayed hard. Many firms lifted and shifted legacy applications to cloud without refactoring, which preserved most of the original constraints and much of the original cost. The real benefit comes from the disentanglement work: breaking the monolith, liberating mainframe data, and building API layers so systems could interoperate on demand rather than in nightly batches. That work was expensive, slow, and required scarce subject-matter experts. So it got deferred.

AI capabilities are evolving fast, and some of the best ways to use them for engineering work are still being discovered — but the direction is clear, and the productivity numbers are no longer speculative. At JPMorgan Chase, tens of thousands of software engineers increased their productivity by 10% to 20% using an internally developed coding assistant, freeing capacity to redirect toward higher-value AI and data work.[1] Airbnb completed a 3,500-file test-framework migration in six weeks against an earlier estimate of seventy-eight weeks.[2] Amazon has used its internal AI assistant to reduce Java application

upgrades from two to three days to less than an hour per application, with aggregate savings of roughly 4,500 developer-years and $260 million in annualized efficiency.[3] These are not marginal improvements; they are order-of-magnitude reductions in the cost of the specific activities that make mainframe modernization expensive.

AI is attacking the hard part of modernization

Three recent proof points: AI-assisted engineering is producing order-of-magnitude shifts, not marginal gains



AI capabilities are evolving fast and the optimal way to use them is still being discovered — but the direction of travel is clear.

Deloitte Consulting LLP

Translated into capital-markets economics: a mainframe disentanglement that would have been priced at $150–$200 million over four or more years just a few years ago could now price materially lower, with delivery timelines compressed by a third or more, according to Deloitte's own delivery experience. The business case rejected at the old number is a different conversation at the new one.

The implication is straightforward. Programs that were economically prohibitive are now merely expensive. Programs that were expensive are now tractable. This argument deserves a fresh run through your investment committee.

The calendar has become the operations leader's problem

The second shift is about the operating environment. Over the next twenty-four months, the operations function in capital markets could face a set of deadlines that, together, cannot be met with quarterly release cycles and overnight batch processing.

Treasury cash clearing compliance arrives December 31, 2026, with repo clearing following on June 30, 2027.[4] NSCC transitions to a 24x5 processing schedule on June 28, 2026, with national exchanges expected to follow into 2027.[5] The UK and EU are targeting October 2027 for T+1 settlement — coinciding with the U.S. Treasury clearing implementation and exposing firms with global books to simultaneous change.[6] The SEC's 2026 examination priorities flag operational resiliency, Regulation SCI incident response, and the new Regulation S-P amendments as areas of heightened focus, with the Regulation S-P compliance date of June 3, 2026 for smaller firms not subject to the December 2025

deadline.[7] All of these imply capabilities that mainframe-bound environments cannot readily produce: real-time observability, rapid recovery, and evidence-grade audit trails across distributed systems.

None of these demands are new. What is new is their concurrence. A firm running its books-and-records on a monolithic core with quarterly releases, reconciling through overnight batches, and depending on a horizontal technology group to prioritize across dozens of downstream applications cannot meet the cumulative demand. The question stops being “should we modernize” and becomes “which deadline will we miss first.”

On vendors: observe the direction, preserve the optionality

Any honest discussion of capital markets operations modernization must acknowledge that the most mainframe-bound parts of the function — books-and-records, custody, clearing, corporate actions, cost basis, asset servicing — are overwhelmingly delivered by a small set of outsourced providers who run their own mainframes. Clients’ improvement aspirations have historically run up against their providers’ release cycles, data access patterns, and willingness to customize.

The major providers are now, visibly, making capital commitments to shift to cloud, expand their API offerings, and unlock AI-enabled engineering. That is genuinely good news and deserves recognition. It is also worth noting that the same structural constraints that have made inhouse modernization a decade-long endeavor apply to the providers running those workloads on behalf of the industry. AI can help accelerate the path; it does not shorten it to zero. Roadmaps that assume otherwise — the client’s own or a vendor’s — will require patience.

The corollary deserves equal candor. For a generation, the build-versus-buy decision for core operations infrastructure was decided before the meeting started: building anything resembling a books-and-records system, a cost-basis engine, or a corporate-actions processor at scale was prohibitively expensive. That calculus has not flipped, but it may be moving. The rational response is not

wholesale replacement — the providers' scale, regulatory track record, and operational maturity remain substantial, and replacement risk is still asymmetric. The rational response is to preserve optionality. Ensure your data, integrations, and client-facing experiences are sufficiently decoupled from any single vendor's roadmap — so that you retain strategic choices your predecessors did not have.

What "acting now" actually means

The practical agenda is unglamorous and concrete — and it is the same agenda that makes AI deployment possible, which is why it aligns well with current investment appetite.

First, data accessibility. Core operational data — positions, transactions, corporate actions, reference data, client master, reconciliation exceptions — must be accessible outside the mainframe in forms that modern applications, analytics and large language models can consume. Lakehouse architectures, streaming patterns, and well-governed data products are the correct technical answer. Clear ownership of each data domain is the organizational answer, which presents its own set of challenges to established organizations that we will not elaborate on here.

Second, APIs that enable action, not just access. A reporting API exposes data for consumption — it lets you read data, but not change it. An action API — versioned, documented, permissioned, SLA-backed — lets agents and applications actually do work: place a trade, update an account, raise an exception, post a journal entry. The value frontier in AI has moved from retrieval to agentic action, and firms whose internal systems are reachable only through mainframe screens or nightly file inputs will struggle to participate, regardless of how much they spend on models.

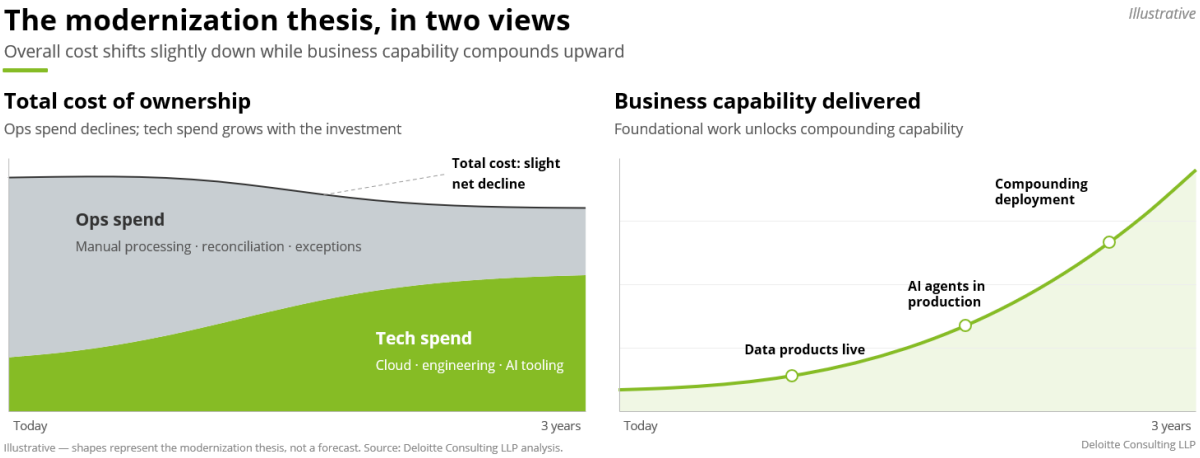
Third, monolith decomposition, done in waves. A "big bang" replatforming that replicates the entire legacy system in the cloud before delivering business value can take several years. Massive programs like this can also lose organizational momentum and funding before any business value is delivered. An incremental approach — separating discrete business-function oriented products from the monolith, delivering them as cloud-native capabilities, and back-syncing to legacy until surrounding products catch up — is slower on paper but could be faster in practice, because it produces the early wins that sustain organizational belief in the program.

Two short examples from our work. At one large retail brokerage, a twenty-million-LOC monolith was decomposed into eighteen product families. Now a portfolio analysis tool previously blocked by overnight batch integration ships enhancements every three weeks. At a wirehouse, a cost-basis product was blocked for years because one vendor could only consume data after another vendor's overnight processing; reorganizing around data product ownership and refactoring the enrichment to cloud made real-time cost basis possible. The lesson is not that vendors are obstacles; it is that without decoupled data and clear internal ownership, firms can only solve these problems on their vendor's timeline.

Why the gap is now widening non-linearly

The uncomfortable implication is that firms which have already liberated their data and built their API layers are not merely ahead by the size of their investment. They are ahead by a factor that compounds.

A firm with accessible data and versioned APIs can deploy an AI agent against a reconciliation exception pattern in weeks. A firm whose data is trapped behind nightly extracts and whose systems are reachable only through terminal-emulated "green screens" cannot deploy that agent at all — not until it first completes the foundational modernization work. That work is itself accelerated by the same AI tooling the modernized firm is already using. The result is a compounding gap: the leader moves faster because it has the tools; the laggard moves slower because it does not.



For most of the last two decades, operations leaders could reasonably assume that a competitor's lead was capped by the same technological constraints that capped their own catch-up. That assumption no longer holds. The constraints have unevenly loosened, and they tend to loosen faster for firms that have already done the foundational work. The window in which firms can recover from deferred decision making is narrowing. At some point, it will close.

There is also a cost looming on the calendar that deserves to be included in the business case. Firms that arrived at T+1 unready in 2024 incurred tens of millions in emergency remediation —Treasury cash clearing and UK/EU T+1 are the next two instances of the same risk, with revenue exposure on top.

What to do before this time next year

The concrete ask of leadership teams in the first half of 2026 is not a program. It is three decisions. Identify the two or three core domains whose data your firm most needs to be accessible outside the mainframe, and commit to liberating them on a dated plan. Identify the internal systems whose functionality your firm most needs to be reachable by agentic action, and commit to publishing versioned APIs that can act on them. Identify the vendor dependencies whose roadmaps most constrain

your own, and define the decoupling posture — data, integration, and experience — that preserves your optionality regardless of how those roadmaps evolve.

None of these is a commitment to a finished state. All of them are commitments to begin, on a timeline that reflects forcing functions the industry is already committed to. Firms that make these decisions now will find, three years from today, that the AI-enabled capability map available to them is qualitatively different from the one available to firms that waited another budget cycle. The math has changed. The calendar has become specific. The question is whether your firm will still own the decision when it becomes unavoidable, or whether the decision will be made for you.

Endnotes

1. Haripriya Suresh, "JPMorgan engineers' efficiency jumps as much as 20% from using coding assistant," Reuters, March 13, 2025, citing remarks by Lori Beer, global Chief Information Officer of JPMorgan Chase.
   https://ca.finance.yahoo.com/news/jpmorgan-engineers-efficiency-jumps-much-190410445.html
2. Charles Covey-Brandt, "Accelerating Large-Scale Test Migration with LLMs," The Airbnb Tech Blog (Medium), March 13, 2025.
   https://medium.com/airbnb-engineering/accelerating-large-scale-test-migration-with-llms-9565c208023b
3. Andy Jassy, "One of the most tedious (but critical tasks) for software development teams is updating foundational software…," LinkedIn, August 22, 2024. Aggregate developer-year and dollar savings figures also reported in "Amazon Q Developer just reached a $260 million dollar milestone," AWS DevOps Blog, August 2, 2024.
   https://aws.amazon.com/blogs/devops/amazon-q-developer-just-reached-a-260-million-dollar-milestone/
4. U.S. Securities and Exchange Commission, "SEC Extends Compliance Dates and Provides Temporary Exemption for Rule Related to Clearing of U.S. Treasury Securities," Press Release 2025-43, February 25, 2025.
   https://www.sec.gov/newsroom/press-releases/2025-43
5. DTCC, "The Shift to 24x5 Trading: What It Means for U.S. Equity Markets," 2025.
   https://www.dtcc.com/-/media/Files/Downloads/Transformation/theshiftto24x5trading.pdf
6. SIFMA, "Operations Conference & Exhibition Debrief," May 14, 2025.
   https://www.sifma.org/research/insights/operations-conference-exhibition-debrief
7. U.S. Securities and Exchange Commission, Division of Examinations, "Fiscal Year 2026 Examination Priorities," November 17, 2025; Regulation S-P amendments effective dates per SEC Advisers Act Release No. 6604 (May 16, 2024).
   https://www.sec.gov/files/2026-exam-priorities.pdf