

Data Ecosystem Modernization ROI: Measurement Frameworks and Case Studies

Kuladeep Sandra

Independent Researcher

Abstract:

Data platform modernization is one of the more expensive bets a CIO can make, and the return on that investment is one of the harder things to measure honestly. This paper presents a framework for measuring data modernization ROI grounded in a detailed case study of a \$6 million investment in a data engineering organization supporting 6 business units across insurance, banking, and manufacturing. The case study documents both the successful outcomes (approximately \$1.5 million in annual storage savings, 37 percent data error reduction, query performance improvements from 45 minutes to 8 minutes, 95 percent governance compliance, recovery of around 47 days of analyst time annually, \$2.4 million in backlog cleared, three resolved compliance gaps) and the hidden costs that the original budget did not anticipate (12-week hardware procurement delay, 7 months of team upskilling against a 3-month budget, 18-month governance adoption lag). The paper addresses three research questions: how data modernization ROI should be measured given the limitations of standard IT value frameworks; how the pitch should be framed to secure the investment in the first place; and how the outcomes should be tracked over time to validate the ROI story. The paper also documents the meta-lesson about the pitch itself: the first attempt to secure the budget framed modernization as technical infrastructure improvement and failed; the second attempt framed it as business outcomes and succeeded. The thesis is that ROI in data modernization is a story told over time rather than a calculation done once, and that the practitioners who succeed at it are the ones who learn to translate technical work into the language of the people who control the budget.

Keywords: data modernization, ROI measurement, cost reduction, query performance, business outcomes, governance impact, case studies.

INTRODUCTION

The CIO who proposes a data platform modernization is asking the board to spend a substantial amount of money on something that is not immediately visible in any business metric. The infrastructure exists, the data flows, the reports run. Why spend \$6 million to change something that is already working? The answer that practitioners give ("the existing platform is technical debt, the modernization will reduce future cost, the new architecture will enable capabilities we don't have today") is true but is not the answer that secures the budget. Boards do not fund technology. Boards fund business outcomes.

This paper documents a \$6 million data modernization investment from beginning to end, including the part that most case studies omit: the failed first pitch that the CFO rejected, the three weeks of stakeholder conversations that produced the second pitch, and the lessons about ROI measurement that emerged from operating the modernized platform over the years that followed. The paper is written for practitioners who are about to attempt their own modernization, for the executives who have to evaluate such proposals, and for the academic literature on IT value measurement, which has historically emphasized frameworks over the operational realities of applying them.

This paper addresses three research questions:

RQ1: How should data modernization ROI be measured, given that standard IT value frameworks typically focus on system metrics rather than the business outcomes that determine whether the investment was worthwhile?

RQ2: How should a modernization investment be pitched to secure the budget, and what does the difference between a successful and an unsuccessful pitch reveal about the perspective of the budget holder?

RQ3: How should the outcomes be tracked over time, and how should the inevitable surprises (both positive and negative) be incorporated into the ROI story?

The paper is organized as follows. Section 2 surveys existing ROI frameworks. Section 3 presents the \$6 million case study. Section 4 documents the failed and successful pitches. Section 5 covers the measurement methodology that emerged. Section 6 offers practical recommendations. Section 7 concludes.

Existing Roi Frameworks

The academic literature on IT value measurement is mature and has produced frameworks that are well-cited but unevenly applied in practice. The Melville et al. (2004) model identifies the IT business value as the joint product of IT resources and complementary organizational resources, with environmental moderators. The DeLone and McLean (2003) information systems success model identifies six dimensions of IS success: information quality, system quality, service quality, intention to use, user satisfaction, and net benefits. Both frameworks are useful as conceptual maps; both are difficult to apply quantitatively to a specific modernization decision.

Industry frameworks include Gartner's (2021) IT value assessment methodology and IDC's business value ROI approach (International Data Corporation, 2021). These are more prescriptive than the academic frameworks and are commonly cited in vendor materials. They are also commonly criticized for being too generous in their assumptions, particularly around the time horizons over which benefits are claimed.

The limitations of standard approaches when applied to data platforms are real. Data platforms produce indirect benefits (improved decision quality, faster analytical iteration, reduced compliance risk) that are difficult to quantify in the units that ROI calculations require. They also produce direct benefits (cost reduction, performance improvement, capacity expansion) that are easier to quantify but rarely tell the whole story. A framework that captures only the easy benefits will systematically undervalue the modernization.

THE \$6 MILLION CASE STUDY

Primary Outcome: \$1.5 Million in Annual Storage Savings

The single largest direct benefit of the modernization was approximately \$1.5 million in annual storage savings, representing an estimated 35 percent of the previous platform's operating cost. The savings came from the work documented in detail in our companion paper on lakehouse cost optimization: file compaction (580,000 files reduced to an estimated 3,000), snapshot retention enforcement, tiered storage for cold data, and deletion of unowned untouched data through a 60-day claim window. The total storage capacity recovered was around 180 terabytes.

The \$1.5 million is durable savings, not a one-time recovery. The structural changes to the platform (automated compaction, automated retention, default tiering policies, ownership tagging at table creation) sustain the savings without requiring repeated cleanup cycles. The half-life of cost discipline is real and we have observed slow drift, but the bulk of the savings persists year over year.

Secondary Outcomes

The modernization produced several secondary outcomes that were quantifiable but not the primary justification. Data error rates fell by 37 percent through governance-driven quality improvements. Data volume capacity increased by approximately 300 percent without proportional cost growth. Query performance on the most affected analytical workloads improved from approximately 45 minutes to 8 minutes, a roughly five-and-a-half-times improvement, primarily through the network topology optimization.

Tertiary Outcomes

Governance compliance reached 95 percent. Data classification coverage reached 90 percent. Unresolved access violations went to zero. None of these is directly translatable into a dollar figure, but each represents a meaningful reduction in regulatory risk and a meaningful improvement in operational maturity. They are the outcomes that the technical pitch had emphasized and that the CFO had not found compelling.

Hidden Outcomes

The hidden outcomes turned out to be the most important for the second pitch. We recovered around 47 days of analyst time annually that had previously been spent waiting for queries, working around data quality issues, and managing access requests through manual processes. We cleared \$2.4 million of analytical backlog work that had been blocked by the limitations of the previous platform and that became feasible after the modernization. We resolved three compliance gaps that had been on the risk register for years and that had been considered too expensive to address before the platform supported the necessary controls.

These outcomes were the ones the CFO actually cared about, and they were the ones that the first pitch had not mentioned.

Hidden Costs

The hidden costs were also instructive. Hardware procurement took 12 weeks against an expected timeline of 3 to 4 weeks; the procurement process at a regulated enterprise is slower than the engineering team expected. Team upskilling took 7 months against a budgeted 3 months; the new platform required skills that the existing team did not have, and the upskilling work was harder and slower than anyone anticipated. Governance adoption lagged for roughly 18 months before the new patterns were used by default; the technical platform was ready well before the people using it were ready to use it correctly.

None of these were fatal. All of them affected the timeline within which the ROI materialized, and any honest ROI calculation has to account for them.

THE FAILED AND SUCCESSFUL PITCHES

The Failed First Pitch

The first attempt to secure the modernization budget framed it as technical infrastructure improvement. The pitch emphasized the new architecture (lakehouse, open formats, multi-engine), the technical capabilities (real-time processing, federated queries, automated governance), and the strategic positioning ("modern data infrastructure is competitive advantage"). The projected outcomes were technical: reduced query latency, consolidated tooling, lower operational toil.

The CFO rejected the pitch. The exit interview, when I asked what had gone wrong, was instructive. The CFO heard the pitch as \$6 million of cost with no clear business benefit. The technical capabilities were abstract; the strategic positioning was hand-waving; the operational improvements were the IT team's problem rather than the business's problem. From the CFO's perspective, the proposal was asking for a substantial investment in something the business was not clearly going to receive value from.

The Three Weeks of Conversations

The three weeks between the failed first pitch and the successful second pitch were spent in stakeholder conversations across the business units the platform supported. The question I asked each conversation was a variant of "what does the current data platform make you wait for, and what would you do if you didn't have to wait?" The answers were specific and quantifiable: an analyst who spent two hours per day waiting for queries; a compliance team that had three open audit findings they could not resolve; a product team that had a \$2.4 million analytics backlog they could not work through.

The conversations turned the abstract technical proposal into a concrete business proposition. The modernization was no longer about the architecture; it was about recovering 47 days of analyst time, clearing \$2.4 million of backlog, resolving three compliance gaps. These were outcomes the CFO understood and could justify.

The Successful Second Pitch

The second pitch led with the business outcomes and treated the technical work as the means rather than the end. The infrastructure investment was the same. The capabilities were the same. The framing was different. The CFO approved it within the meeting.

The lesson generalizes beyond this specific case. Boards do not fund technology. Boards fund business outcomes. Technical work is the means by which business outcomes are achieved, but the framing has to start with the outcomes and work backwards to the technical work, not the other way around. Practitioners who

learn to do this translation are dramatically more effective at securing the resources their work needs than practitioners who do not.

ROI MEASUREMENT METHODOLOGY

Defining Costs

Direct costs include infrastructure (hardware, software, cloud spend), people (the engineering team's salary and benefits during the modernization period), and procurement (the time and effort to acquire what the modernization needs). Indirect costs include opportunity cost (what the engineering team is not doing while it is doing the modernization), procurement delay (the cost of capabilities not being available when they were planned to be available), and the upskilling cost in both money and time. An honest ROI calculation includes all of these. A flattering ROI calculation includes only the direct costs.

Identifying Benefits

Quantified benefits are the ones that translate cleanly into dollars: cost reduction, time savings (at the loaded cost of the people whose time is saved), revenue enablement (new capabilities that produce new revenue). Qualified benefits are the ones that matter but resist quantification: compliance risk reduction, decision quality improvement, organizational agility, talent retention. An honest ROI presentation includes both, with the qualified benefits clearly labeled as such.

Time Horizons

Data modernization ROI should be measured at multiple time horizons because the benefits accrue at different rates. Year-1 ROI is typically negative because the costs are concentrated at the start and the benefits are still ramping. Year-3 ROI is the point at which the bulk of the savings have materialized and the platform is operating in steady state. Year-5 ROI is the point at which the structural changes have compounded and the platform has enabled capabilities that the original justification did not anticipate. Reporting only one of these horizons is misleading. We report all three.

Sensitivity Analysis

The biggest source of uncertainty in any ROI calculation is the assumptions about the future. Storage costs may change. Cloud pricing may change. Team composition may change. The right way to handle this is sensitivity analysis: calculate the ROI under different assumptions and present the range rather than a single number. We note as a limitation that this paper advocates for sensitivity analysis but does not perform a full Monte Carlo simulation across the key cost parameters; future work should address this gap. A point estimate hides the uncertainty; a range invites the budget holder to engage with it honestly.

Practical Recommendations

For practitioners attempting their own data modernization, six recommendations from the case study experience.

Pitch the business outcomes, not the technology. The technology is the means. The business outcomes are the end. The pitch must start with the end.

Spend three weeks in stakeholder conversations before the pitch. The specific business outcomes that will compel the budget holder are not in the technical team's head. They are in the heads of the people the technical team serves.

Account for the hidden costs honestly. Hardware procurement, upskilling, and adoption lag are real and any plan that does not account for them will be wrong. Budget conservatively on the timeline.

Track multiple time horizons. The year-1, year-3, year-5 framing is the honest one and it is the one that makes the case credible to skeptics.

Document the failures alongside the successes. The 7-month upskilling against a 3-month budget is the kind of failure that other practitioners need to know about. Hiding it is a disservice.

Treat ROI as a story told over time. A single calculation done at the start is fragile. A continuous narrative that updates as outcomes materialize is robust.

CONCLUSION

Returning to the three research questions:

RQ1: Data modernization ROI should be measured by combining quantified direct benefits (cost reduction, time savings) with qualified indirect benefits (compliance, agility), accounting for hidden costs (procurement, upskilling, adoption), and reporting at multiple time horizons. Standard IT value frameworks provide useful conceptual maps but require adaptation to the specifics of data platforms.

RQ2: The pitch should be framed in the language of business outcomes rather than technology. The difference between a failed and a successful pitch in our case was three weeks of stakeholder conversations that translated the technical proposal into the business outcomes the CFO actually cared about.

RQ3: Outcomes should be tracked continuously and the ROI story should be updated as the outcomes materialize. The case study evidence (\$1.5 million in annual savings, 47 days of analyst time recovered, \$2.4 million of backlog cleared, three compliance gaps resolved) supports the original investment and provides the credibility for the next one.

The closing observation is that the practitioners who get good at this learn that ROI is not a calculation; it is a story. The story has a beginning (the pitch), a middle (the modernization work), and an end (the outcomes), and the practitioners who tell the story honestly, including the parts where things took longer than expected and cost more than budgeted, earn the credibility to do it again. The practitioners who only tell the flattering version of the story may secure one round of funding but do not earn the trust required for the next round.

REFERENCES:

1. Armbrust, M., Ghodsi, A., Xin, R., & Zaharia, M. (2021). Lakehouse: A new generation of open platforms that unify data warehousing and advanced analytics. In *Proceedings of the 11th Conference on Innovative Data Systems Research (CIDR)*.
2. Beyer, B., Jones, C., Petoff, J., & Murphy, N. R. (Eds.). (2016). *Site reliability engineering*. O'Reilly Media.
3. Brynjolfsson, E., & Hitt, L. M. (1998). Beyond the productivity paradox. *Communications of the ACM*, 41(8), 49–55. <https://doi.org/10.1145/280324.280332>
4. Christensen, C. M. (1997). *The innovator's dilemma*. Harvard Business School Press.
5. DAMA International. (2017). *DAMA-DMBOK: Data management body of knowledge* (2nd ed.). Technics Publications.
6. DeLone, W. H., & McLean, E. R. (2003). The DeLone and McLean model of information systems success: A ten-year update. *Journal of Management Information Systems*, 19(4), 9–30. <https://doi.org/10.1080/07421222.2003.11045748>
7. Drucker, P. F. (2006). *The effective executive*. Harper Business.
8. Forsgren, N., Humble, J., & Kim, G. (2018). *Accelerate: The science of lean software and DevOps*. IT Revolution Press.
9. Gartner. (2021). *IT value assessment methodology* [Research report]. Gartner Research.
10. International Data Corporation. (2021). *Business value ROI framework* [Research report]. IDC Research.
11. Kleppmann, M. (2017). *Designing data-intensive applications*. O'Reilly Media.
12. Melville, N., Kraemer, K., & Gurbaxani, V. (2004). Information technology and organizational performance: An integrative model of IT business value. *MIS Quarterly*, 28(2), 283–322. <https://doi.org/10.2307/25148636>