

The Modernization Cost Dilemma: Why Cloud Expenses Rise Despite Technical Improvements and How to Achieve Sustainable Economics

Amit Kumar Garg

Independent Researcher, USA

ARTICLE INFO

Received: 13 Jan 2026

Revised: 16 Jan 2026

ABSTRACT

Managed databases, auto-scaling, and cloud-native services guarantee that cloud modernization efforts will enable greater reliability, scalability, and operational efficiency. Nonetheless, the situation is often met with a paradoxical twist to organizations that are seemingly experiencing a successful technical transformation, but the costs of the infrastructure are nonetheless on the rise. Such a phenomenon is due to a root cause of a lack of relationship between technical modernization and operational discipline, in which better infrastructure performance allows inefficiencies to scale up with valid workload. Uncontrolled data collection with no lifecycle management, auto-scaling tools that increase wastage instead of only contributing to business value, global availability designs that increase storage and transfer costs multiple times, obscure cost indicators in controlled service structures, and non-linear disaster recovery configurations are the major cost drivers. These issues have an overrepresented effect on low-margin industries where even a small cost increment has a huge negative effect on profitability and competitive standing. An effective solution means relating data to explicitly managed financial assets with defined ownership and retention policies, bringing down database footprint before introducing further scaling, introducing visibility and accountability to auto-scaling events, introducing lightweight governance frameworks, and designing recovery architectures that trade resilience and efficiency. The features of companies that have sustainable cloud economics are similar: explicit data lifecycle management, proactive database footprint control, engineering of the visibility of the cost impacts and resilience designs that do not make inefficient infrastructure multiply. The key to cloud cost management is ultimately the ability to understand that the costs associated with persistent data are more in the long term than those of temporary compute and that deliberate data practices form the core of cloud operations that are cost-effective over time.

Keywords: Cloud Cost Optimization, Data Lifecycle Management, Auto-Scaling Efficiency, Database Footprint Reduction, Cloud Modernization Economics

1. Introduction

Modernization is usually portrayed as a cost-cutting measure. Organizations that move to managed databases are moving to cloud-native services, introducing auto-scaling, and decommissioning old infrastructure with the hope that efficiency will then follow. Yet many teams experience a frustrating reality: cloud costs continue to rise even after modernization is complete.

According to the Flexera 2023 State of the Cloud Report, organizations continue to struggle with cloud cost management, with enterprises identifying cost optimization as their primary cloud initiative [1]. The challenge is not merely about controlling expenses but understanding where inefficiencies originate after technical transformations are complete. The report reveals that organizations consistently overestimate their actual cloud usage and underestimate the complexity of managing costs across multiple cloud environments. This disconnect between modernization expectations and financial outcomes creates a paradox where technically superior systems generate economically inferior results.

The scale of cloud spending continues to accelerate globally. Worldwide end-user spending on public cloud services demonstrates remarkable growth trajectories, with forecasts indicating substantial year-over-year increases that reflect both adoption acceleration and cost management challenges [2]. This growth encompasses infrastructure as a service, platform as a service, and software as a service segments, all contributing to expanding cloud expenditures that often exceed organizational budgets and expectations.

This outcome is not a failure of cloud platforms or managed services. It is a mismatch between technical modernization and operational discipline. Modern systems make it easier to scale—but they also make it easier to scale inefficiency. This article explains why cloud costs often rise postmodernization, and how organizations can regain control without slowing innovation.

2. The Core Misconception: Modernization Equals Optimization

Cloud modernization fundamentally replaces how systems are run, not how efficiently they behave. Companies spend a lot of money on the migration to managed databases, auto-scaling computer resources, and creating capabilities to replicate globally. These investments provide quantifiable benefits in quality of availability, reliability, and operational ease. Nevertheless, the main fallacy is the belief that modernization of infrastructure should be equated with cost minimization. The reality is far more nuanced, as modernization primarily addresses technical architecture while leaving underlying inefficiencies largely untouched.

The financial implications of this misconception are substantial. Cloud cost optimization strategies that focus exclusively on infrastructure modernization without addressing application-level efficiency typically fall short of projected savings [3]. Organizations discover that migrating workloads to cloud environments without re-architecting applications or optimizing data management practices simply transfers existing inefficiencies to a more expensive operational model. The ease of provisioning resources in cloud environments, combined with the removal of traditional capacity constraints, creates conditions where inefficient practices expand rather than diminish. The teams that are used to resource constrained environment on-premise are now exposed to resource constraints that seem to be limitless, such as storage, computation, and network resources, and they start consuming resources in ways that would not have been possible under the earlier infrastructure paradigm.

This is further aggravated by the fact that modernization undertakings are usually based on functional parity in lieu of functional utility. Migration groups are interested in making sure applications operate properly in the cloud, frequently recreating existing structures and performance trends without underlying reproductivism. Although this strategy of lift and shift simplifies migration complexity and risk, it continues to support inefficient patterns of data access and query designs and unnecessary data retention policies. The result is infrastructure that operates more reliably but at significantly higher cost than anticipated.

Research into managed cloud services adoption patterns reveals that organizations frequently experience unexpected cost escalations during the first twelve to eighteen months following major modernization initiatives [4]. These costs are attributed to several causes, such as the underestimation of data transfer costs, ineffective provisioning of resources, and the absence of visibility about the finer details of consumption. Managed services remove the complexity of infrastructure, which is good in operations, but also hides the cost of architectural choices. Teams are not able to have a direct view of resource utilization, and therefore, it is hard to find and rectify inefficiencies without it growing into huge financial implications.

Therefore, the inefficiencies that used to be confined by on-premises restrictions tend to be given a free hand in the cloud setups. Legacy systems operating with fixed storage capacity naturally enforced data lifecycle management through necessity. When storage approached capacity limits, teams had no choice

but to archive or delete historical data. Cloud environments eliminate these natural constraints, allowing data volumes to grow indefinitely without triggering immediate operational concerns. The infrastructure becomes more capable, scaling seamlessly to accommodate growing demands, but the underlying waste simply scales proportionally with legitimate workloads, creating cost growth that mirrors technical capability rather than business value.

Table 1: The Core Misconception - Modernization vs Optimization [3, 4]

3. The Primary Drivers of Post-Modernization Cost Growth

3.1 Data Gravity Without Data Discipline

The essence of cloud platforms is that the way organizations adopt the concept of data storage is revolutionized since it becomes immeasurably easy to save data with tremendous durability and warranties of data safety. Large cloud providers provide storage solutions with various redundancy levels, automatic cross-availability zone replication, and durability guarantees of enterprise standards. The ease of storage of data, however, is a contrast to the challenge of erasing information. Data deletion remains a human decision requiring explicit action, approval workflows, and often crossteam coordination. This asymmetry between storage ease and deletion friction creates an environment where data accumulates continuously without corresponding cleanup efforts.

Common patterns of unchecked data growth manifest across multiple organizational layers. Transactional databases retain detailed logs and event records spanning years, originally captured for debugging purposes but never removed after their utility expires. Analytical systems create derived datasets and aggregated copies to support various reporting requirements, with each copy persisting independently without lifecycle management. Development and testing environments duplicate production datasets to support realistic testing scenarios, creating multiple full-scale replicas that remain active indefinitely. Each retained byte carries compounding costs, as cloud platforms replicate data across availability zones for durability, maintain multiple backup copies according to retention policies, index data to support query performance, and include all stored data in disaster recovery operations.

The global datasphere continues expanding at exponential rates, with cloud storage representing an increasingly dominant proportion of total data storage [5]. This growth reflects both legitimate business data creation and the accumulation of redundant, obsolete, and trivial data that organizations fail to manage effectively. Data volumes in cloud environments grow without the physical constraints that once limited on-premises expansion, allowing storage consumption to increase continuously until costs become unsustainable. The financial impact of unmanaged data growth compounds over time, as storage costs represent only the most visible component of total data-related expenses.

Research into cloud computing resource optimization demonstrates that data management practices significantly impact total cost of ownership, with storage-related expenses often representing higher long-term costs than compute resources [6]. The study reveals that organizations frequently underestimate the full cost implications of data retention, focusing on nominal storage prices while overlooking ancillary costs associated with data replication, backup, indexing, and recovery operations. Over time, data gravity becomes the dominant cost driver in cloud environments, independent of compute efficiency or application optimization efforts. The sheer volume of retained data creates its own momentum, making cleanup efforts increasingly difficult as datasets grow larger and interdependencies become more complex.

3.2 Auto-Scaling Scales Waste as Well as Value

Auto-scaling capabilities represent one of cloud computing's most celebrated features, enabling applications to respond dynamically to changing demand without manual intervention. Organizations also apply auto-scaling policies to guarantee the availability of applications at times of traffic peaks, and consider a smooth scaling as a sign of maturity in the architecture. Systems scale compute resources, database capacity, and network bandwidth based on utilization metrics, and costs increase in proportion to resource consumption. Nevertheless, this success at a glance can frequently conceal structural inefficiencies that auto-scaling only increases instead of eliminating.

The fundamental challenge with auto-scaling in practice lies in what remains unnoticed during scaling events. Applications issue queries that scan entire tables when indexed lookups would suffice, transferring and processing orders of magnitude more data than necessary. Database schemas accumulate indexes created to optimize specific queries, with each additional index imposing write amplification costs that multiply as systems scale. Cold data stored alongside hot data participates in routine operations despite rarely being accessed, consuming resources during every scaling event. Auto-scaling responds to these inefficiencies by provisioning additional capacity, effectively hiding the waste behind expanded resource allocation. Teams observe that applications handle peak loads successfully and interpret this as validation of their architecture, while the underlying inefficiency remains unexamined.

Research into cloud-native application performance demonstrates that query optimization and data access patterns significantly impact resource consumption in auto-scaled environments [7]. The study examines how applications behave under dynamic scaling conditions, revealing that inefficient data access patterns consume disproportionate resources compared to optimized implementations. Applications that scan large datasets unnecessarily, retrieve excessive columns in query results, or perform redundant computations trigger scaling events that could be avoided through architectural improvements. The research quantifies how query optimization reduces resource requirements substantially, demonstrating that efficiency gains at the application layer deliver more sustainable cost reductions than infrastructure scaling alone.

The financial implications become particularly pronounced in environments where auto-scaling operates continuously. Cloud cost management analysis reveals patterns where organizations maintain consistently elevated resource utilization driven by inefficient application behavior rather than legitimate demand growth [8]. Database instances scale to accommodate query patterns that could be optimized, compute resources expand to handle workloads containing redundant processing, and network capacity increases to transfer unnecessarily large result sets. Auto-scaling obscures these inefficiencies by treating all resource demands as legitimate, provisioning capacity without distinguishing between value-creating work and waste. The result is infrastructure costs that scale predictably with application activity, but without corresponding business value creation, as much of the scaled capacity addresses inefficiency rather than actual customer demand.

3.3 Global Availability Comes With Permanent Multipliers

The availability features of modern cloud architectures are becoming more of a global offering to provide low-latency access to users in every part of the world and resilience against a regional outage. Organizations add multi-region database setups, cross-region replication of the most important datasets, and geo-redundant backup systems to achieve the goals of availability and disaster recovery. These architectural designs radically enhance the reliability of the applications, as users will have the same experience of the application in any location, and because the data is available even during a significant failure of the infrastructure. However, these resilience features introduce cost multipliers that many organizations underestimate during architecture design phases.

The fundamental economics of global availability stem from data replication requirements. Data that once existed in a single location now exists across multiple geographic regions, each hosting complete

or partial replicas depending on architectural decisions. Cloud providers charge for storage in each region independently, meaning that a three-region deployment triples baseline storage costs before considering additional factors. Beyond storage multiplication, cross-region data transfer incurs substantial charges that accumulate with every synchronization operation [9]. Database writes replicate across regions, triggering data transfer charges in both directions. Backup operations in multi-region configurations create redundant copies in each region, further multiplying storage costs. When organizations implement these patterns without carefully managing the underlying dataset size, the cost multipliers apply to bloated data volumes, amplifying existing inefficiencies across every region and availability zone.

The compounding effect of global availability on unoptimized datasets creates exponential cost growth. An organization retaining unnecessary historical data in a single-region deployment faces straightforward storage costs. The same organization implementing three-region active-active replication immediately triples those storage costs, plus incurs continuous data transfer charges for synchronization. Adding geo-redundant backups with multi-region retention multiplies costs further, as each backup copy exists in multiple locations.

The mathematical relationship between data volume and total cost becomes multiplicative rather than additive, meaning that every megabyte of unnecessary data retained carries costs across multiple dimensions: primary storage in each region, backup copies in each region, cross-region transfer for synchronization, and disaster recovery storage in geographically distant locations.

Organizations implementing global availability strategies frequently discover that resilience improvements come with financial implications that substantially exceed initial projections [10]. The analysis of multi-region deployment costs reveals that data transfer charges, often overlooked during planning phases, can represent significant portions of total cloud expenditure. These transfer costs occur not only during initial data replication but continuously as applications generate writes that propagate across regions. Also, globally distributed architecture backup and disaster recovery operations use significantly more resources than those of a region since data has to be moved over geographic boundaries and stored in multi-redundant locations. In cases where the underlying dataset has large volumes of outdated or useless information, they multiply the cost of waste as easily as useful data, and may lead to a scenario where resilience architecture ASAP is used to multiply inefficiency instead of merely securing vital information.

Cost Driver	Mechanism	Impact on Total Costs
Data Gravity Without Discipline	Logs, events, and duplicates persist indefinitely; every byte is replicated, indexed, and backed up	Becomes a dominant cost factor independent of compute efficiency
Auto-Scaling Waste	Inefficient queries, index sprawl, and cold data in hot paths scale with capacity	Masks inefficiency through capacity expansion rather than optimization
Global Availability Multipliers	Multi-region replication, cross-region transfers, geo-redundant backups	Multiplies storage and transfer costs across all regions and zones
Obscured Cost Signals	Managed services abstract infrastructure; teams lose visibility until invoices arrive	Storage grows unchecked, and replica counts become permanent
		defaults

Non-Linear Recovery Costs	Large datasets require extended restoration time and elevated compute during recovery	Data volume dominates recovery time and operational risk during incidents
---------------------------	---	---

Table 2: Primary Drivers of Post-Modernization Cost Growth [5, 6]

4. Why This Hits Low-Margin Businesses Hardest

In industries operating with thin profit margins, such as retail, logistics, food services, and online marketplaces, cloud cost increases exert disproportionate pressure on business viability. These sectors typically operate with profit margins measured in single-digit percentages, where small absolute cost increases translate to significant margin erosion. A retail operation with a three percent net margin finds that a modest increase in infrastructure costs can eliminate substantial portions of profitability, requiring corresponding revenue increases just to maintain previous earnings levels. The mathematics of low-margin businesses means that cost reductions deliver outsized business impact compared to equivalent revenue increases, as saved costs flow directly to the bottom line without the variable costs associated with generating additional sales.

The competitive dynamics in low-margin industries intensify the financial pressure from cloud cost overruns. Price and operational efficiency are the key elements behind competition between companies in these industries and the capacity to raise costs without loss of market share is minimal. A marketplace that cannot regulate infrastructure costs is at a strategic dilemma, whether to take the costs and lower the profitability, or raise costs and prices at the risk of being outcompeted. Both options dilute the business's stand, because lower margins also restrict the ability to reinvest funds in the business when the margins are down, but hikes in prices make the platform less appealing compared to other businesses. This dynamic creates situations where technically superior cloud infrastructure paradoxically weakens competitive position by consuming margin that could otherwise fund growth initiatives, market expansion, or competitive pricing strategies.

Cloud optimization strategies become existential imperatives rather than efficiency projects for lowmargin businesses. A logistics company discovering that database bloat has doubled infrastructure costs cannot simply accept this as the price of modernization. The cost inflation could be the distinction between lucrative activities and losses to the extent urgent redemption becomes necessary, irrespective of the priorities in engineering or promises in the roadmap. Cloud cost management should be approached with the seriousness given to the core product development by these organizations, since the uncontrolled infrastructure costs are a direct threat to the sustainability of the business. This pressure to maximize is further increased by the fact that there are still few alternatives since moving back to on-prem infrastructure is both costly in terms of transition costs and operation complexity, and there is an ever-increasing pressure of cloud costs that are decreasing the margin needed to ensure business survival and growth in competitive markets.

Industry Characteristic	Cloud Cost Challenge	Business Consequence
Single-digit profit margins	Small absolute cost increases eliminate substantial profitability portions	Infrastructure costs directly threaten business viability
Price-based competition	Limited ability to pass costs to customers without market share loss	Must absorb costs or accept competitive disadvantage
Operational efficiency	Cost increases consume the margin needed for growth and competitive	Technically superior infrastructure

focus	pricing	weakens competitive position
Strategic imperatives	Cost management becomes existential rather than an efficiency initiative	Optimization is required for survival rather than improvement

Table 3: Impact on Low-Margin Businesses [7, 8]

5. Fixing the Problem: From Infrastructure Optimization to System Efficiency

Cloud cost control fundamentally requires shifting perspective from infrastructure management to system efficiency. Organizations must recognize that sustainable cost optimization does not emerge from simply turning features off or reducing service levels. Instead, effective cost management aligns system behavior with economic intent, ensuring that every resource consumed contributes to business value creation. To achieve this alignment, one needs to put in place clear ownership of data assets, lifecycle management policies, and business justification visibility of resource usage.

5.1 Treat Data as a Financial Asset, Not a Byproduct

All datasets in an organization are to be clearly owned, and the responsibility of management of a particular dataset is assigned to an individual or a group of people. Orphaned data will always continue to live since no one is in charge or answerable to decide on its further usefulness. To drive proper data management, it is necessary to define the purpose of all the various data sets, set up the retention horizons depending on the business and regulatory needs, and deploy policies of data lifecycle change that will involve the automatic movement of data to the right storage level, and then ultimately its disposal. If a dataset cannot demonstrate clear business justification for its ongoing storage costs, it does not belong in production systems consuming premium storage and participating in replication and backup operations.

Deleting unused data is often one of the most immediate and high-impact cost reduction opportunities for many organizations. Historical logs generated during debugging sessions, event streams captured for analytics that were never implemented, and duplicate datasets created for projects that concluded months or years ago all represent pure waste, consuming resources without providing value. Organizations implementing systematic data cleanup initiatives frequently discover that thirty to forty percent of stored data serves no current business purpose, offering immediate cost reduction opportunities without impacting operational capabilities or customer experiences.

5.2 Reduce Database Footprint Before Scaling Further

Smaller databases deliver compounding benefits across multiple dimensions simultaneously. Reduced database footprint improves query performance by decreasing the data volume scanned during operations, accelerates replication by minimizing the data transferred between nodes, shortens recovery time by reducing the dataset that must be restored during failures, and decreases costs across storage, compute, network, and disaster recovery operations. This combination of performance, reliability, and cost benefits makes database footprint reduction one of the highest-value optimization activities available to engineering teams.

The first step that should be taken in footprint reduction strategies involves the identification and elimination of outdated information, such as completed transactions not required in the retention rules, records overridden by newer records, and archived data stored in production databases inappropriately. After data cleanup, teams are requested to analyze schema design to determine recommendations to minimize data that is stored in the database with normalization, compression, or transfer of data that is rarely touched to the relevant levels of storage. The savings of the compound of footprint reduction are not limited to the direct cost of storage, including decreased backup storage, decreased disaster

recovery time, decreased replication bandwidth usage and enhanced application performance that can lower compute requirements.

5.3 Make Auto-Scaling Visible and Accountable

Implementing auto-scaling must include a high level of observability so that the events of scaling can become visible to the engineering teams, as well as the business stakeholders. Each scaling event must produce metrics and logs that can be used to analyze events that occurred after the event, and answer questions such as: what did the scaling happen because of, was the trigger an indication of legitimate demand growth, or was the trigger an indication of application inefficiency, and does the scaling pattern indicate a recurrent behavior that needs architectural consideration, or is the scaling pattern an anomalous spike that justifies a temporary increase in capacity.

Effective auto-scaling governance establishes regular review processes where teams examine scaling patterns and investigate underlying causes. Cost-aware scaling policies should incorporate efficiency metrics alongside availability targets, recognizing that unlimited scaling capability does not represent optimal architecture if the scaling addresses inefficiency rather than genuine demand. Organizations should implement alerting on scaling events that exceed expected patterns, triggering investigations into whether application optimization could reduce or eliminate the scaling requirement. This approach treats scaling as a signal requiring investigation rather than purely an operational success metric.

5.4 Introduce Lightweight Data Governance

Data governance need not require heavyweight processes or extensive bureaucracy to deliver value. Minimum viable governance: defines explicit data domains, in line with business capabilities, assigns explicit ownership of each domain with specific roles of data quality management and data lifecycle management, and records access patterns and expectations of retention of various data categories. These governance fundamental components ensure that the environment cleaned up does not degrade with the lapse of time because newer data is created under the same distribution and ownership patterns as opposed to being created naturally and unsupervised with time.

The governance structures must be designed to prevent rather than cure so that the practices should be designed to prevent the emergence of future problems in data management instead of concentrating on the clean up of the current situation. This involves making data lifecycle plans mandatory when developing new features, ownership assignment mandatory before datasets are put into production, and making retention policies automatically enforced instead of having them manually cleaned up. Lightweight governance is a balance between structure and agility where there is enough guidance that does not result in issues of cost occurring again but does not result in process overhead that slows the velocity of development.

5.5 Design for Recovery, Not Just Availability

High availability and recoverability are two similar architectural issues. High availability reduces downtime by using redundancy and failover features so that it can stay online in case one of the components fails. Recoverability is concerned with having the capability to recover operations after disastrous failures, with specific interest in recovery time and data integrity, rather than ongoing availability. Although high availability minimises the occurrence of recovery operations, it does not remove the need to achieve good recovery capabilities in case failures surpass redundancy provisions.

Recovery-driven design principles are to have a limited size of data volumes that make recovery time predictable, to continuously test recovery processes to ensure they are working, and to make sure that recovery replicas and backups are not present by default. Smaller data sets decrease recovery time and recovery cost in case of incidents, as restoration operations are done faster in case of incidents, utilizing fewer resources. The design of recovery architectures in an organization should be such that when there is realistic data in the system, it can recover essential services within predefined time frames instead of depending on an infinite availability of resources during the event of a crisis. This strategy acknowledges

that the cost of disaster recovery increases with the amount of data being recovered and that disaster recovery can only be effective in cases where recovery procedures are regularly tested with representative data sets and not with ideal architectural models that often do not represent the actual operational reality.

Solution Area	Implementation Strategy	Compounding Benefits
Data as a Financial Asset	Assign ownership, define purpose, and establish retention horizons for all datasets	Immediate cost reduction through deletion of unused logs and historical events
Database Footprint Reduction	Remove obsolete data, optimize schema, and migrate cold data to appropriate tiers	Improves performance, replication speed, and recovery time, and reduces costs simultaneously
Auto-Scaling Accountability	Make scaling observable, attributed, and reviewed; investigate triggers and patterns	Distinguishes legitimate demand from inefficiency; enables cost-aware scaling
Lightweight Data Governance	Define domains, assign ownership, establish access and retention expectations	Prevents cost problems from recurring after cleanup initiatives
Recovery-Focused Design	Test recovery paths, bound data volumes, and make copies intentionally	Reduces recovery time and cost during incidents through smaller datasets

Table 4: Solutions for System Efficiency [9, 10]

Conclusion

Fundamental changes in reliability, scalability, and the speed of operation. Cloud modernization provides organizations with the ability to develop and run systems that would otherwise have been unattainable or prohibitively costly when operating within a traditional infrastructure setting. But the modernization is not necessarily a better way to be more efficient economically. Modernization can be expected to drive the increase in cost at a very rapid rate without the discipline in its operations and specific cost management practice, since there were natural restraints that once held back inefficiency. The same characteristics that make cloud platforms useful, infiniteness, global availability, automated provisioning, also make waste to scale with value-creating workloads. The road to sustainable cloud economics means making the modernization process of infrastructure transformation. Organizations do not need to merely switch the mode of operation of their infrastructure, but rather they need to undergo a fundamental transformation in the way that they handle their data, manage their resources, and bring visibility to the business value that their cloud expenditure brings. This change includes the implementation of data lifecycle, creating clear retention and deletion policies, proactively managing database footprint so that systems are sized appropriately to their workload, making cost indicators visible to engineering teams so that decisions about resource consumption incur an economic cost consideration, and being resiliently designed in an efficient manner as opposed to with redundancy as a matter of default. When modernizing, organizations that effectively manage the cost of the cloud do not incur information without using a policy; they manage data as an asset, with clear ownership and lifecycle management policies. They are actively monitoring database footprint as they are aware that small systems are more efficient, cheaper, and recover more quickly than fat systems. They provide the engineering teams with a good visibility of the cost of an architectural decision, and they maintain a feedback loop that promotes efficient design patterns. They use resilience architecture in a considered manner such that they offer the required redundancy, but not blindly increasing inefficiency between

geographical regions and availability zones. Above all, they see cloud platforms as an opportunity to enliven business opportunities instead of infinite sources that are made without thinking about their cost-effectiveness. In cloud environments, the storage costs of organizations eventually cost more than what is executed, so the management of cloud costs becomes sustainable only when conscious data practices are applied, so that only the useful information remains in the production systems.

References

- [1] Tanner Luxner, "Cloud computing trends and statistics: Flexera 2023 State of the Cloud Report," Flexera, 2023. [Online]. Available: <https://www.flexera.com/blog/finops/cloud-computing-trendsflexera-2023-state-of-the-cloud-report/>
- [2] Express Computer, "Gartner Forecasts Worldwide Public Cloud End-User Spending to Reach \$679 Billion in 2024," 2023. [Online]. Available: <https://www.expresscomputer.in/news/gartner-forecastsworldwide-public-cloud-end-user-spending-to-reach-679-billion-in-2024/105778/>
- [3] Anish Anurag, "What Is Cloud Cost Optimization and Why It Matters," Qwegle, 2025. [Online]. Available: <https://www.qwegle.com/cloud-cost-optimization-2025/>
- [4] Greg Zwakman and Liam Eagle, "Managed services market forecast: Managed cloud services coming to the fore," S&P Global, 2023. [Online]. Available: <http://spglobal.com/marketintelligence/en/news-insights/research/managed-services-market-forecast-managed-cloud-servicescoming-to-the-fore>
- [5] David Reinsel et al., "The Digitization of the WorldFrom Edge to Core," IDC White Paper, 2018. [Online]. Available: <https://www.seagate.com/files/www-content/our-story/trends/files/dataageidc-report-final.pdf>
- [6] Lizhe Wang et al., "Scientific Cloud Computing: Early Definition and Experience," 2008 10th IEEE International Conference on High Performance Computing and Communications, 2008. [Online]. Available: <https://ieeexplore.ieee.org/document/4637787>
- [7] Jie Zhu et al., "Sky Computing: Accelerating Geo-distributed Computing in Federated Learning," arXiv:2202.11836, 2022. [Online]. Available: <https://arxiv.org/abs/2202.11836>
- [8] Kent Bennett et al., "State of the Cloud 2023," BVP Atlas. [Online]. Available: <https://www.bvp.com/atlas/state-of-the-cloud-2023>
- [9] Amazon Web Services, "Amazon S3 Pricing." [Online]. Available: <https://aws.amazon.com/s3/pricing/>
- [10] IBM Education, "What is cloud cost optimization?" [Online]. Available: <https://www.ibm.com/think/topics/cloud-cost-optimization>