New Series Annals on Economy, Law and Sociology, Volume 9, Number 1/2026
Romania, June, 2026 ISSN ONLINE 2067-5704

MANAGING DATA TRANSFER COSTS IN HYBRID CLOUD INFRASTRUCTURES: APPLIED RESEARCH

Marius Toderici[1], Diana Cristina Dragomir[2], Mircea Claudiu Fulea[3] and Aurel Mihail Titu[4,5]
[1]National University of Science and Technology POLITEHNICA Bucharest, Bucharest, Romania,
ORCID No. 0009-0007-6065-6502, marius@toderici.ro
[2]Technical University of Cluj-Napoca, 103-105 Muncii Boulevard,
ORCID No. 0000-0001-6833-8815, diana.dragomir@muri.utcluj.ro
[3] Technical University of Cluj-Napoca, 103-105 Muncii Boulevard,
ORCID No. 0000-0001-9640-6379, mircea.fulea@staff.utcluj.ro
[4]Lucian Blaga University of Sibiu, Sibiu, Romania, Corresponding author,
ORCID No. 0000-0002-0054-6535, mihail.titu@ulbsibiu.ro
[5]Academy of Romanian Scientists, 3 Ilfov Street, Bucharest, Romania

ABSTRACT: Hybrid cloud infrastructures offer organizations the flexibility to integrate on-premises resources with public cloud services; however, this integration introduces substantial challenges related to data transfer costs. Charges associated with data egress, inter-region communication, and continuous data replication can escalate rapidly, particularly in multi-cloud scenarios where workloads and datasets are distributed across heterogeneous environments. As a result, unmanaged data movement has the potential to undermine the economic benefits typically associated with cloud adoption. This study investigates strategies for managing and optimizing data transfer costs in hybrid cloud ecosystems, addressing architectural, operational, and financial dimensions. Key approaches include optimizing data locality to ensure that workloads are executed in proximity to their associated datasets, leveraging dedicated interconnect services to achieve predictable bandwidth and reduced egress fees, and applying techniques such as caching, compression, and deduplication to minimize data transfer volumes. Furthermore, workload placement policies and FinOps practices—supported by automation and policy-as-code mechanisms—are identified as critical enablers for enforcing cost-efficient operations and maintaining financial governance. The findings highlight that proactive monitoring, combined with cost-aware architectural design, is essential to balancing performance requirements with economic sustainability. Organizations that embed cost optimization principles into the design and operation of hybrid cloud infrastructures are better positioned to achieve long-term efficiency, scalability, and financial resilience.
KEYWORDS: Cost management, hybrid cloud, data transfer, data egress.

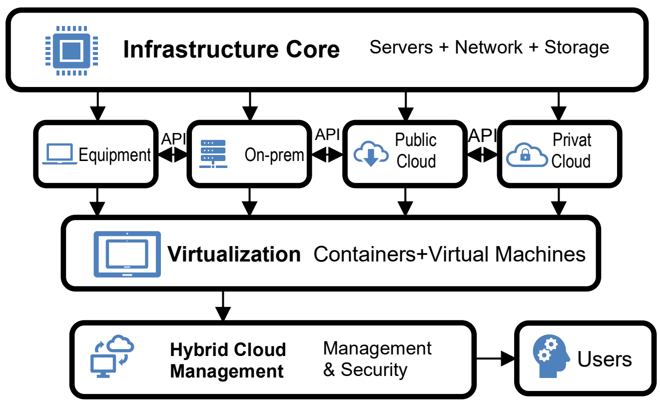
DOI 10.56082/annalsarscieco.2026.1.36

1. INTRODUCTION

Hybrid cloud infrastructures are an architectural model that is increasingly being adopted by modern organizations due to their ability to integrate local (on-premises) IT infrastructure with public cloud services into a unified and interoperable ecosystem. This model allows organizations to leverage the advantages of both paradigms: the control, security, and compliance associated with private environments, and the elasticity, scalability, and distributed access offered by public cloud platforms. From a technological point of view, hybrid cloud architecture is based on interconnecting local infrastructure with Infrastructure-as-a-Service (IaaS) platforms provided by major players such as Amazon Web Services, Microsoft Azure, Google Cloud, or VMware Cloud, through high-capacity communication links capable of supporting intensive data flows.[1]

However, as the degree of integration between on-premises and cloud environments increases, the complexity of managing the costs associated with these infrastructures increases significantly. A key element, often insufficiently analyzed, is the costs generated by data transfer between hybrid infrastructure components, particularly those related to public cloud egress traffic (data egress fees). These costs are influenced by multiple factors, such as the volume and frequency of transfers, architectural topology, the geographical location of resources, and the commercial policies of cloud service providers,

Figure 1 Hybrid cloud infrastructure



making it difficult to estimate them accurately in the initial design and planning phases.

In the absence of adequate monitoring and control mechanisms, data transfer costs can become a dominant component of operating expenses, affecting the economic sustainability of hybrid cloud solutions. This problem is exacerbated by the transition from traditional CapEx investment models, characteristic of on-premises infrastructures, to variable OpEx cost models, specific to cloud services. In this context, effective management of data transfer costs requires an integrated approach that combines technical, operational, and financial perspectives through the use of governance and optimization practices such as FinOps.[2]
Given these considerations, analyzing and optimizing the costs associated with data transfer in hybrid cloud infrastructures is becoming a topic of major interest for both academia and IT practitioners. This article aims to address this issue by highlighting the factors that determine data transfer costs and exploring management and optimization approaches designed to support the sustainable adoption of hybrid cloud infrastructures.

2. DATA TRANSFER AND NETWORK COSTS

The costs associated with data transfer and network traffic in cloud environments are inherently more difficult to estimate than those related to compute or storage resources, which are typically priced using fixed or clearly defined unit rates. Network costs are influenced by multiple dynamic factors, including traffic direction, volume, destination, latency requirements, and security constraints. Moreover, these costs also serve as a mechanism for cloud service providers to regulate and limit excessive data movement outside their platforms, as outbound traffic places significant demand on communication infrastructure, bandwidth capacity, security mechanisms, and network performance.[3]
In traditional on-premises environments, infrastructure-related costs—such as servers, storage systems, networking equipment, cooling, electrical power, and physical space—are generally incurred as separate capital investments. These expenditures are often made at different times and sourced from multiple vendors, but they are largely fixed in nature. As a result, despite their long-term financial impact, such costs are relatively predictable and easier to forecast within budget planning processes.
By contrast, cloud computing operates predominantly under a variable, consumption-based pricing model, where resources are billed according to actual usage. While this approach offers flexibility and cost efficiency for workloads with variable or unpredictable demand, it can also lead to significant cost overruns in the absence of effective optimization strategies. Misconfigurations—such as the lack of a Content Delivery Network (CDN), unnecessary data replication, improper regional placement of storage resources, or excessive outbound traffic—can rapidly amplify network-related expenses.
Among network-related charges, data egress costs are often perceived as particularly high. First, they are billed explicitly on a per-gigabyte basis, whereas equivalent data transfers in on-premises environments are typically absorbed within broader infrastructure costs and are not itemized separately. Second, data egress charges are highly visible in cloud billing statements under labels such as Data Egress or Outbound Traffic, which accentuates their perceived financial impact. Finally, these costs can escalate quickly when application architectures are not designed for efficient data flow or when caching and content distribution mechanisms, such as CDNs, are not employed.
Consequently, understanding the drivers of data transfer costs and adopting traffic-aware architectural designs are essential for achieving financial sustainability in cloud and hybrid cloud environments. Effective cost management requires not only technical optimization but also continuous monitoring and alignment between architectural decisions and financial governance practices. Data egress costs are perceived as particularly high because:

- They are charged per GB transferred, unlike on-premises costs, which do not explicitly reflect data transfers.
- They are directly visible on the invoice (e.g., "Data Egress" or "Outbound traffic"), which emphasizes their financial impact;
- They can increase rapidly if the application architecture is not optimized for efficient traffic or if a content delivery network (CDN) is not used.

3. INBOUND AND OUTBOUND TRAFFIC BETWEEN CLOUD ENVIRONMENTS

Inbound traffic—that is, data transferred to the cloud infrastructure (also known as data ingress)—is, in most cases, free of charge, regardless of the volume transferred, for all major cloud providers. Outbound traffic (data egress), on the other hand, is treated differently. Regardless of the destination – the public internet, local data centers (on-premises), physical offices, or other cloud regions – this type of traffic is charged based on several factors: the cloud provider, the service used, and the network route involved.
One aspect that is often overlooked when evaluating the costs of cloud services is data egress, which can

generate significant expenses depending on the network architecture. Cloud service providers monitor and charge for network traffic leaving a defined area of their infrastructure—whether it's data transfer between different regions, between availability zones within the same region, or even between separate virtual networks within the same region. These costs can vary depending on the volume of data transferred, the destination location, and the network route used, making them a significant factor in total cloud operating expenses.

Generally, data egress is charged at a fixed price per gigabyte up to a certain volume (monthly threshold), after which progressive discounts may apply for larger volumes of data transferred. Cloud data egress (transfer or data egress) refers to any data transfer that leaves a cloud provider's infrastructure and reaches an external location. These locations may include:

- a local (on-premises) data center;
- another cloud data center in a different region;
- another Availability Zone (AZ) within the same region;
- another virtual network (VPC, VNet) in the same cloud;
- or the public internet (e.g., to end users, external APIs, etc.).

Data egress fees can have a significant impact on total cloud infrastructure costs, especially for organizations that:

- process large volumes of data;
- perform cross-region replication;
- offer streaming, offsite backup, or content delivery (CDN) services;
- or frequently transfer data to external users or partners.

Outbound cloud data traffic can manifest in multiple forms. One common example involves data transfers between virtual machines within a cloud environment, such as moving information from a database to a spreadsheet accessed by a business user. While such internal transfers contribute to network usage, a more substantial and often overlooked source of cost arises from the delivery of static content—including images, videos, and CSS/JavaScript files—stored in the cloud and accessed via web or mobile applications.

A widely adopted architectural practice is to store static assets in object storage services, separate from the core web application, due to the performance, scalability, and development efficiency benefits this approach provides. In this model, web applications retrieve static resources directly from object storage services such as Amazon S3, Microsoft Azure Blob Storage, or Google Cloud Storage. These resources are accessed through HTTP endpoints and are not cached or stored locally on the application servers.

However, each access to such resources incurs multiple cost components. First, read operations against object storage are billed based on the number of access requests. Second, the delivery of content to end users generates outbound network traffic from the cloud provider to the public internet, which is charged as data egress. When multiplied across large user bases or high-frequency access patterns, these combined costs can represent a significant portion of total operational expenditures.

This cost structure underscores the importance of optimizing content delivery mechanisms—such as employing caching strategies or integrating content delivery networks (CDNs)—to reduce redundant storage reads and minimize outbound traffic in cloud and hybrid cloud environments.

However, each access of this type generates double costs:

- for read operations from storage space;
- for traffic leaving the cloud to end users, i.e., to the internet.

Tabel 1. Output types and pricing (Source: Google Cloud)

Traffic type	Description	Estimated cost ($/GB)
In the same area/network (VPC)	Internally, without leaving the local infrastructure	Free
Between zones (in the same region)	AZ → AZ	0.01
Between regions	E.g.: europe-west1 → us-central1	0.08 – 0.15
To the internet	E.g.: data delivered to customers	0.12 – 0.23 (depending on volume and region)
To another cloud / peering	Trough interconnections (VPN, Interconnect)	0.01 – 0.05

Therefore, it is essential for organizations to closely monitor these costs and optimize the use of content storage and delivery services by:

- Reducing inter-regional transfers – keeping application components in the same region;
- Using CDN content distribution networks – for static web content (Cloudflare, AWS CloudFront, Azure CDN);
- Enabling local caching – through edge nodes or application-level caching;
- Planning internal traffic – use the same zone and virtual network for VMs;
- Estimating and continuous monitoring – through cloud analytics tools (e.g., GCP Cost Explorer, AWS Cost Anomaly Detection).[4]

4. DATA TRANSFER BETWEEN REGIONS

Each cloud service provider applies distinct pricing models and policies for data egress, making careful architectural planning a critical requirement in hybrid and multi-cloud environments. A cloud region represents a geographically defined area that comprises multiple interconnected data centers. Within each region, cloud providers deploy one or more Availability Zones (AZs), which are physically isolated locations designed to ensure redundancy, fault tolerance, and high service availability.[5]

Although major cloud providers offer high-bandwidth, low-latency connectivity between regions to support globally distributed and highly available applications, inter-region data traffic is metered and billed, typically based on the volume of data transferred, measured in gigabytes (GB). Cross-region data transfer refers to the movement of data from one geographic region to another within the same cloud provider—for example, transferring data from Europe-West1 (Belgium) to US-East1 (South Carolina) in Google Cloud Platform. Such transfers can significantly increase operational costs, particularly in scenarios involving frequent replication, global load balancing, or disaster recovery configurations.

Availability Zones within the same region are distinct physical locations, each equipped with independent power supply, networking, and cooling infrastructure. Their primary purpose is to enable fault isolation and service continuity in the event of localized failures. While deploying applications across multiple availability zones is considered a best practice for improving resilience and reliability, data transfer between zones is often subject to additional charges. For example, both Google Cloud and Amazon Web Services typically charge approximately USD 0.01 per GB for traffic exchanged between availability zones within the same region.

Across major cloud providers—including AWS, Microsoft Azure, and Google Cloud—outbound data transfer is consistently treated as a billable resource. In general, data transfers between different regions (inter-region traffic) incur higher costs, while transfers within the same region but across different availability zones are either free or billed at a significantly lower rate. Conversely, inbound data traffic (data ingress) is typically free of charge, whereas outbound traffic (data egress) is chargeable, reflecting the cost of network utilization and infrastructure maintenance borne by the cloud provider.

These pricing characteristics highlight the importance of workload placement, regional architecture design, and data locality considerations when designing hybrid cloud infrastructures. Poorly planned inter-region or inter-zone data flows can result in substantial and often unexpected network costs, emphasizing the need for continuous monitoring and cost-aware architectural decisions.

Cloud providers (AWS, Azure, Google Cloud) charge per GB for data leaving one region and entering another (Table 2).

In general:

- Transfer between different regions (inter-region) is charged;
- Transfers within the same region but between different zones are either free or much cheaper;

Inbound traffic is free, but outbound traffic is chargeable.

Tabel 2. Costs of data transfer between regions (Source: Google Cloud)

Transfer type	Example	Estimated Cost $/GB
Inter-region, standard	Europe → USA	0.12
Intra-region, between zones	Belgium zone A → zone B	0.01-0.02
Within the same VPC	The same region	free

5. USE OF CONTENT DELIVERY NETWORKS (CDN)

A CDN (Content Delivery Network) fetches website resources once and caches them in geographically Content Delivery Networks (CDNs) constitute a fundamental component of modern cloud and hybrid cloud architectures, with a significant impact on both application performance and data transfer cost optimization. A CDN operates by retrieving content from origin servers and caching it at geographically distributed edge locations positioned closer to end users. This architectural approach reduces the physical distance that data must travel, thereby decreasing network latency and improving response times.

In hybrid cloud infrastructures, where applications and data are distributed across on-premises environments and multiple cloud regions, CDNs play a critical role in minimizing data egress from public cloud platforms. Without distributed caching mechanisms, each request for static content would require a data transfer from the primary cloud region, generating recurring network charges and increasing load on origin infrastructure. By serving content directly from edge caches, CDNs significantly reduce redundant outbound traffic, leading to measurable reductions in operational network costs.[6]

Beyond economic benefits, CDNs contribute substantially to performance optimization. Reduced latency and faster content delivery enhance user experience, which is particularly important for globally distributed web and mobile applications. Research has demonstrated a strong correlation between page load times and key performance indicators such as user retention, conversion rates, and overall satisfaction. Consequently, the adoption of CDNs should be viewed not only as a cost-control mechanism but also as a strategic enabler of service quality and competitiveness.

The integration of CDNs into hybrid cloud environments typically requires minimal architectural adjustments, such as configuring access endpoints, defining caching policies, and managing security certificates. Although these changes may introduce an initial implementation effort, the long-term gains in cost efficiency and performance generally outweigh the associated complexity. Moreover, modern CDN platforms provide advanced monitoring and analytics capabilities, allowing organizations to quantify cost savings, analyze traffic patterns, and continuously refine optimization strategies.

CDNs also play a significant role in strengthening the security posture of hybrid cloud applications. Many CDN providers offer built-in protections such as Distributed Denial of Service (DDoS) mitigation, Web Application Firewalls (WAF), and traffic filtering mechanisms. These features help reduce attack surfaces, improve service availability, and protect critical resources, which is particularly important in hybrid architectures characterized by increased complexity and exposure.[7]

From a cost management perspective, it is important to note that several CDN providers maintain strategic partnerships with major cloud service providers. These agreements can result in reduced or even eliminated data transfer fees between the cloud platform and the CDN network. For organizations delivering large volumes of static or multimedia content, such partnerships can significantly lower total operating costs. As a result, the selection and integration of an appropriate CDN solution become strategic decisions with direct implications for financial efficiency.

In conclusion, the use of Content Delivery Networks within hybrid cloud infrastructures represents a best practice for reducing data egress costs, improving application performance, and enhancing security. When combined with structured governance and financial management frameworks such as FinOps, CDNs enable organizations to achieve greater cost transparency and to maximize the economic and operational benefits of hybrid cloud architectures.

6. CONCLUSIONS

Hybrid cloud infrastructures provide organizations with a flexible and scalable framework for modernizing IT systems by enabling the integration of on-premises resources with public cloud services. Nevertheless, the operational and technological advantages of this model are accompanied by significant challenges related to cost management, particularly in the area of data transfer. Unlike compute and storage resources, network-related costs are inherently more difficult to estimate and can increase rapidly in the absence of optimized architectural designs and effective control mechanisms.

The analysis presented in this study demonstrates that outbound cloud traffic (data egress) represents one of the primary contributors to operational expenditures in hybrid cloud environments. These costs are influenced by multiple factors, including architectural topology, geographic distribution of resources, replication frequency, and application workload characteristics. While inter-regional and inter–availability zone data transfers are essential for ensuring high availability, fault tolerance, and resilience, they can impose a substantial financial burden if not carefully planned and managed.

The findings emphasize the importance of adopting comprehensive data traffic optimization strategies, such as co-locating application components, minimizing inter-regional data movement, leveraging content delivery networks (CDNs), and implementing caching mechanisms at both application and network levels. In addition, continuous monitoring of network utilization, coupled with the systematic use of cost analysis and anomaly detection tools provided by cloud service providers, is critical for identifying inefficiencies and preventing uncontrolled cost escalation.

In conclusion, effective management of data transfer costs in hybrid cloud infrastructures extends beyond purely technical considerations and requires robust governance and financial oversight. An integrated approach involving close collaboration between IT and finance teams is essential for achieving transparency and accountability. Frameworks such as FinOps offer a structured methodology for aligning actual resource consumption with business value and economic impact, thereby supporting the sustainable and cost-efficient adoption of hybrid cloud architectures. Future research may explore the development of predictive cost modeling techniques and evaluate the influence of emerging interconnection technologies and intelligent orchestration mechanisms on reducing data transfer costs.

7. REFERENCES

1. Polinati, A., K., *Hybrid Cloud Security: Balancing Performance, Cost, and Compliance in Multi-Cloud Deployments*, Cryptography and Security, arXiv:2506.00426, (2025).
2. Deochake, S., *ABACUS: A FinOps Service for Cloud Cost Optimization*, Technical Report / Service Framework Description, (2024).
3. Mazhelis, O., Tyrväinen, P., *Economic Aspects of Hybrid Cloud Infrastructure: User Organization Perspective*, Information Systems Frontiers, vol. 14, pp. 845–869, (2012).
4. Eiers, W., Sankaran, G., *Quantifying Permissiveness of Access Control Policie*s, Proceedings of the 44th IEEE/ACM International Conference on Software Engineering (ICSE 2022), IEEE eds., Pittsburgh, PA, (2022).
5. Benmammar, B., Asma, A., *Intelligent Network Management and Control: Intelligent Security, Multi-Criteria Optimization*, *Cloud Computing, Internet of Vehicles, Intelligent Radio*, Wiley eds., Hoboken, NJ, (2021).
6. Tyagi, A., *Optimizing digital experiences with content delivery networks: Architectures, performance strategies, and future trends*. World Journal of Advanced Research and Reviews, https://doi.org/10.48550/arXiv.2501.06428, (2020).
7. Karthikeyan, R., Jha, K., Thakur, R., Tripathy, A., Sunil, Dr., Rao, P. S., *Performance Profiling of Encrypted Traffic in Content Delivery Networks, Journal of Internet Services and Information Security*, vol. 15, pp. 431–444, DOI: 10.58346/JISIS.2025.I4.031., (2025)
8. Lee, S., Manjunath, D., Joo, C., *On the Economic Effects of CDN-Mediated Delivery on Content Providers*, IEEE Transactions on Network and Service Management, vol. 19, no. 4, pp. 4449–4460, DOI: 10.1109/TNSM.2022.3162883, (2022)
9. Thakur, M. A., R. G., *User Identity & Lifecycle Management Using LDAP Directory Server on Distributed Network*, Proceedings of the International Conference on Pervasive Computing (ICPC), IEEE eds., Pune, India, (2015).
10. Thakur, M. A., R. G., *User Identity and Access Management Trends in IT Infrastructure – An Overview*, Proceedings of the International Conference on Pervasive Computing (ICPC), IEEE eds., Pune, India, (2015).
11. Kihara, M., S. I., *Security and Performance of Single Sign-On Based on One-Time Pad Algorithm*, Cryptography, (2020).
12. Banciu, D., *Infrastructură de Tip Cloud pentru Instituțiile Publice din România – ICIPRO*, Big Data, The Cloud, Analytics – Triunghiul Profitabilității, București, Romania, (2016).
13. Mohammadinodoushan, M., B. C., *Resilient Password Manager Using Physical Unclonable Functions*, IEEE Access, vol. 9, pp. 17060–17070, (2021).
14. Uddin, M., S. I.-N., *A Dynamic Access Control Model Using Authorising Workflow and Task-Role-Based Access Control*, IEEE Access, vol. 7, pp. 166676–166689, (2019).
15. Banciu, D., *Cercetare – Dezvoltare – Inovare prin Soluții Cloud*, Mobile Innovation 2015, Bucharest, Romania, (2015).
16. Titu, M. A., Deac-Suteu, D., Toderici, M. I., *Catalogul de Servicii Informatice – Instrument pentru Îmbunătățirea Calității și Securității Suportului Informatic într-o Companie*, Acta Technica Napocensis, Cluj-Napoca, Romania, (2021).
17. Majumdar, S., Madi, T., Wang, Y., Jarraya, Y., Pourzandi, M., Wang, L., Debbabi, M., *Security Compliance Auditing of Identity and Access Management in the Cloud: Application to OpenStack*, International Conference on Cloud Computing Technology and Science (CloudCom), IEEE eds., Vancouver, Canada, (2015).
18. Sofradžija, H., Alibegović, A., Adžajlić-Dedović, A., *The Era of Digital Transformation: A World Taken Over by AI, and the Adventure of Technology in the Culture of Real Virtuality*, Lecture Notes in Networks and Systems, vol. 1484, Springer eds., Sarajevo, Bosnia and Herzegovina, (2025).
19. Guida, G. C., Limongiello, M., An Evolved *Bridge Digital Twin Framework: A New Paradigm for Infrastructure Monitoring and Management*, Lecture Notes in Networks and Systems, vol. 1484, Springer eds., Sarajevo, Bosnia and Herzegovina, (2025).
20. Sarwar, M. I., Abbas, Q., *Digital Transformation of Public Sector Governance with IT Service Management – A Pilot Study*, IEEE Access, vol. 11, pp. 6490–6512, (2023).