



# International Journal of Multidisciplinary Research and Growth Evaluation



International Journal of Multidisciplinary Research and Growth Evaluation

ISSN: 2582-7138

Received: 17-05-2020; Accepted: 14-06-2020

www.allmultidisciplinaryjournal.com

Volume 1; Issue 2; May-June 2020; Page No. 47-54

## Mitigating Data Transfer Bottlenecks on Google Cloud: Best Practices for Cloud-to-Cloud and Cloud-to-On-Premises Connections

Tulasiram Yadavalli

Computer Science and Engineering, USA

Corresponding Author: Tulasiram Yadavalli

DOI: <https://doi.org/10.54660/IJMOR.2020.1.2.47-54>

### Abstract

Efficient data transfers between cloud platforms and on-premises systems are critical for optimizing real-time analytics, particularly when leveraging tools like BigQuery. Common challenges such as network latency, scalability, and managing large datasets can hinder performance. By implementing best practices like optimizing BigQuery configurations, automating ETL pipelines with tools such as

Google Dataflow, and utilizing real-time data streaming, organizations can improve transfer efficiency and reduce latency. These strategies ensure seamless data ingestion, enabling faster and more accurate data processing, better scalability, and enhanced decision-making in cloud environments.

**Keywords:** Data Transfer Bottlenecks, Data Transfer Optimization, Google Cloud, Latency Reduction, ETL Pipelines, On-Premise, VPN, Load Balancing, Encryption

### 1. Introduction

The rapid expansion of cloud infrastructure in recent years has introduced significant challenges related to efficient data transfer, especially for organizations utilizing Google Cloud for hybrid and multi-cloud strategies.

Mitigating data transfer bottlenecks and optimizing latency between Google Cloud, other cloud platforms, and on-premises systems is crucial for real-time operations. Google Cloud offers various transfer methods—Google Cloud Interconnect, VPN, and Direct Peering—that facilitate diverse data movement needs.

These methods have specific benefits and trade-offs, with performance, cost, and security considerations varying based on the organization's connectivity requirements<sup>[1]</sup>.

To address bandwidth limitations, data compression and deduplication have been widely recommended by researchers as effective ways to maximize transfer speeds and minimize data volume.

These techniques are essential for organizations handling large data sets, as they reduce transfer times by removing redundant data and compressing files to better utilize available bandwidth<sup>[2]</sup>.

However, encryption—vital for maintaining data security in transit—often adds latency, creating additional bottlenecks that require strategic balancing to achieve optimal performance.

Studies suggest that encryption methods can slow down data transfer due to the computational resources required, especially in high-volume transfers<sup>[3]</sup>.

Moreover, utilizing Google's Load Balancing services can streamline data transfer by dynamically distributing workloads across multiple paths. This approach helps maintain high throughput during peak demand periods, avoiding bottlenecks that can occur when bandwidth is limited or when traffic is concentrated on a single route<sup>[4]</sup>.

By distributing data across different nodes, Google Cloud's Load Balancing services also enhance resiliency and enable better handling of large-scale transfers without sacrificing speed<sup>[5]</sup>.

Previous studies highlight the importance of selecting appropriate transfer options based on workload and performance requirements, with Google Cloud Interconnect providing dedicated, high-bandwidth connections suited for data-intensive Applications, while VPN and Direct Peering offer flexibility and cost-effectiveness for smaller or less latency-sensitive tasks<sup>[6]</sup>. Partitioning and clustering techniques, particularly in Google BigQuery, further aid in reducing latency, improving both data access and query performance to support timely analytics<sup>[7]</sup>. Understanding how to manage workload elasticity and

Scaling in Google Cloud environments can also mitigate latency issues by dynamically adjusting resource allocation based on data volume and system demands [8].

Finally, case studies reveal that effective ETL (Extract, Transform, Load) optimization can significantly impact data flow efficiency, with Google Dataflow enabling both stream and batch processing to transform data before ingestion, reducing the risk of delays in analytics pipelines [9].

These practices are critical in ensuring that cloud and hybrid infrastructures can deliver high-performance data transfers to support modern analytics needs [10].

## 2. Literature Review

As cloud ecosystems have expanded, so has the demand for flexible, interoperable connections across multiple cloud platforms.

Contemporary cloud services often provide Infrastructure-as-a-Service (IaaS) to meet dynamic user needs; however, no single provider offers infinite resources, spurring the development of inter-cloud environments. These environments allow cloud platforms to share infrastructure, enhancing scalability and resource flexibility but creating new complexities in data transfer management and security across providers.

Research highlights the importance of secure communication frameworks in multi-provider settings, with peer-to-peer overlays and secure admission protocols providing low-latency, efficient solutions for secure, cross-cloud data transfers [1].

The surge in digital data storage has amplified the need for deduplication to manage redundant information, which can strain cloud storage systems.

Deduplication is particularly advantageous for optimizing bandwidth in data transfers, as it eliminates duplicate data before transfer, decreasing load and reducing costs.

Recent research categorizes deduplication methods for various data types—text, image, and video—highlighting unique challenges for each, thereby supporting efficient data handling and facilitating seamless, large-scale data movement across cloud environments [2].

While encryption is critical for securing data in transit, it often introduces latency due to the computational demands required, especially for high-volume data transfers.

Research highlights that managing data confidentiality is particularly challenging when dealing with sensitive data streams from resource-limited devices in critical applications. Selective encryption methods have been explored to balance security and performance, using efficient key management and refreshment to maintain confidentiality without disrupting data flow, thus mitigating latency concerns while protecting data integrity [3].

Utilizing load balancing is essential in cloud environments with high user demands, as it efficiently manages data transfers by dynamically redistributing workloads. Techniques like Distributed Dynamic FTP (DDFTP) exploit replicated file storage across multiple servers, optimizing download speeds by retrieving file segments simultaneously from various sources. This approach not only maximizes network bandwidth utilization but also maintains smooth load distribution across diverse, dynamically changing servers, which is crucial for enhancing transfer rates and managing resource fluctuations efficiently [4].

Distributed systems are fundamental to cloud computing, enhancing scalability, resilience, and efficiency across cloud

environments. By distributing resources and processes across multiple nodes, they enable load balancing and fault tolerance through redundancy and data replication. These systems are pivotal in modern cloud infrastructures, supporting microservices, containers, and serverless computing. As cloud platforms like Google Cloud continue to integrate distributed systems, they enable large-scale data transfers, ensuring high availability, seamless scalability, and enhanced performance despite challenges like network latency and security [5].

Previous research emphasizes the importance of aligning cloud transfer options with specific application needs. For high-bandwidth, data-intensive tasks, Google Cloud Interconnect offers dedicated connections that support optimal performance, while VPN and Direct Peering provide cost-effective, flexible solutions for less demanding workloads. The growing integration of IaaS with high-performance computing (HPC) resources also highlights the potential for cloud platforms to bridge the gap between big data and scientific computing, enhancing scalability and performance across diverse applications [6].

Partitioning and clustering techniques, particularly in Google BigQuery, significantly enhance data processing efficiency, reducing latency and improving query performance for large data sets. These methods are particularly valuable for sessionized data, which involves analyzing user activity in discrete sessions. By optimizing data storage and enabling faster access to large, historical data, Google BigQuery supports timely analytics, facilitating informed business decisions. The cloud-based infrastructure provides a scalable, cost-effective solution compared to traditional on-premise systems for handling extensive sessionized data [7].

Elasticity in cloud computing is a critical feature that allows resources to be dynamically allocated or removed based on fluctuating demand. This capability is essential for optimizing performance and reducing latency in Google Cloud environments. By scaling resources "on the fly," cloud platforms can efficiently handle varying workloads, ensuring that data volume and system demands are met without over-provisioning. Studies on cloud elasticity have highlighted the importance of tools, metrics, and evaluation methods to maximize resource utilization and address performance challenges [8].

Effective ETL (Extract, Transform, Load) processes are crucial for managing and integrating data from heterogeneous sources, ensuring that data is cleaned, transformed, and loaded efficiently into data warehouses.

Case studies highlight the importance of optimizing these workflows to reduce processing time and improve overall performance. Google Dataflow, for instance, supports both stream and batch processing, allowing for efficient transformation of data before ingestion. This optimization reduces delays in analytics pipelines and enhances the speed and reliability of data flows [9].

Cloud and hybrid infrastructures are essential for supporting modern data analytics, offering high-performance data transfer capabilities that are crucial for Big Data applications. The architecture of these infrastructures typically includes cloud-based Big Data systems, which integrate data management, analytics technologies, and security features. Leveraging platforms such as SlipStream for multicloud deployment, organizations can automate the management and scaling of complex applications, enhancing their ability to meet the demands of data-intensive applications while



**Inefficient data compression and deduplication techniques**

Data compression and deduplication are essential for reducing data transfer times and maximizing bandwidth usage. However, many organizations struggle with inefficient techniques, leading to longer transfer times and higher costs.

**Challenges with data compression and deduplication:**

- **Inadequate compression algorithms:** Weak or improperly configured compression algorithms can result in poor data reduction, leaving much of the available bandwidth underutilized.
- **Ineffective deduplication strategies:** Without proper deduplication, redundant data continues to be transferred, unnecessarily consuming bandwidth and increasing overall transfer times.
- **Increased storage requirements:** Failure to adequately deduplicate data can also result in bloated storage requirements, which further strains network resources.
- **Incompatibility with diverse data types:** Different data types (e.g., images, videos, and text) require tailored compression and deduplication strategies, making it harder to standardize the process for all use cases.

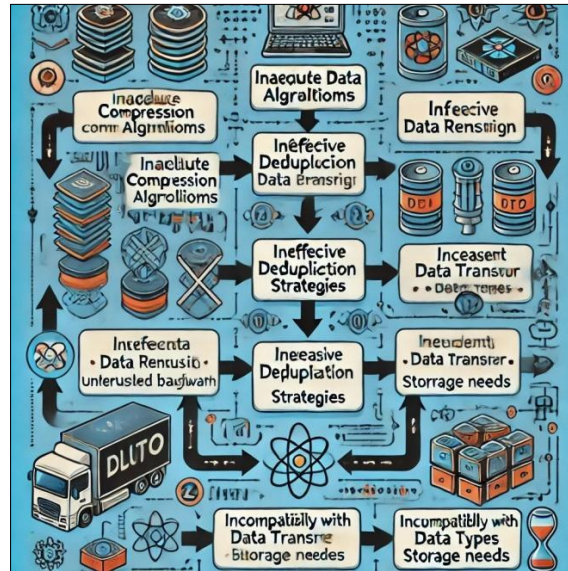


Fig 2: shows the challenges in data compression and deduplication.

**Load balancing limitations during peak demand**

Google Cloud offers Load Balancing services that can optimize data transfers by distributing workloads across multiple paths. However, improper configuration, resource limitations, or high traffic during peak times can result in inefficient load balancing and performance issues.

**Problems with load balancing:**

- **Uneven traffic distribution:** Poorly configured load balancing can result in some paths receiving more traffic than others, leading to congestion and bottlenecks.
- **Limited scalability during high demand:** Load balancers may struggle to handle surges in traffic, especially during peak demand periods, causing slowdowns and delays.
- **Resource contention:** During periods of heavy data transfer, load balancing may not be able to allocate resources efficiently, further exacerbating transfer delays.
- **Impact on performance:** The inability to effectively distribute traffic during peak times can lead to data transfer bottlenecks, ultimately affecting user experience and system performance.

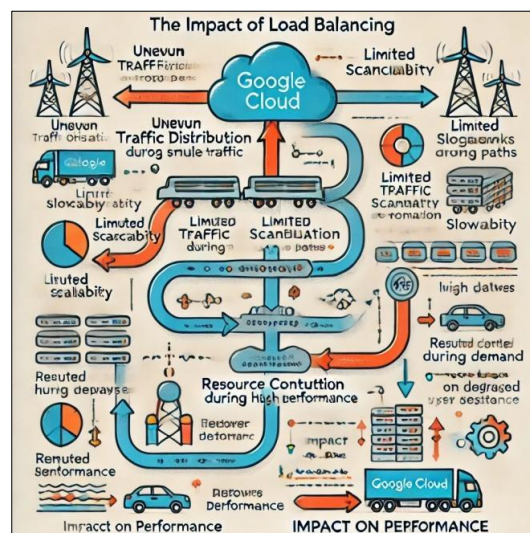


Fig 3: shows the impact of improper load balancing in Google Cloud.

### Complexities in selecting the right transfer method

Google Cloud offers various data transfer methods, including Google Cloud Interconnect, VPN, and Direct Peering. Each method has distinct benefits and trade-offs, making it challenging to select the optimal one for specific use cases.

#### Challenges with selecting the right method:

- **High-bandwidth needs:** For applications that require high throughput, Google Cloud Interconnect may be the best option, but it may not be cost-effective for smaller or less demanding tasks.
- **Cost vs. performance trade-off:** VPN and Direct Peering may offer cost-effective solutions for smaller workloads but may not deliver the performance needed for large-scale or time-sensitive data transfers.
- **Compatibility issues:** Each transfer method has specific requirements and limitations, such as geographical restrictions or security constraints that can complicate the decision-making process.
- **Lack of flexibility:** Some methods may be too rigid or lack flexibility in terms of scalability, preventing organizations from adapting their transfer methods to changing needs or traffic loads.

### Challenges in multi-cloud and multi-provider data management

The increasing adoption of multi-cloud and hybrid cloud environments brings about significant challenges in managing cross-platform data transfers, data security, and system compatibility.

#### Problems in multi-cloud and multi-provider environments:

- **Data compatibility:** Ensuring compatibility between different cloud providers, such as Google Cloud, AWS, and Azure, can be complex, particularly when using proprietary technologies.
- **Security concerns:** Different security protocols across platforms can create gaps in protection, potentially exposing sensitive data to risks during transfer.
- **Performance inconsistencies:** Varying network architectures and transfer protocols across cloud platforms can lead to inconsistent transfer speeds and increased latency.
- **Management complexity:** Managing resources and workflows across multiple cloud environments can be cumbersome and difficult to optimize, requiring specialized knowledge and tools.

These are the core problems organizations face when trying to optimize data transfers across Google Cloud, other cloud platforms, and on-premises systems. Addressing these bottlenecks and latency issues is crucial to improving data flow, ensuring real-time operations, and maintaining high performance across cloud environments.

### 4. Solution: Optimizing data transfer for efficiency and performance

In light of the various bottlenecks and latency challenges identified in the previous section, it is crucial for organizations to implement effective solutions to overcome these obstacles.

The following proposed solutions aim to enhance data transfer speeds, reduce latency, and optimize overall performance across cloud and hybrid environments.

By utilizing Google Cloud's tools, optimizing transfer

methods, and implementing best practices, organizations can maximize their bandwidth and efficiency while ensuring data security and compliance

### Maximizing bandwidth and reducing transfer times

To address bandwidth constraints and improve transfer speeds, organizations can utilize a combination of data optimization techniques and cloud infrastructure enhancements. These methods focus on maximizing available bandwidth, ensuring efficient use of network resources, and minimizing transfer delays.

#### Strategies for maximizing bandwidth and reducing transfer times:

- **Data Compression:** Utilize advanced compression algorithms to reduce the size of data being transferred, thereby making better use of available bandwidth and shortening transfer times.
- **Data Deduplication:** Implement deduplication techniques to eliminate redundant data before transfer, reducing unnecessary data traffic and optimizing bandwidth usage.
- **Parallel Transfers:** Leverage parallel data transfer techniques to split large files into smaller chunks and transmit them simultaneously, reducing overall transfer time.
- **Optimized Transfer Protocols:** Use protocols like Google Cloud's Transfer Service, which optimizes large-scale data migration and minimizes latency.

### Utilizing advanced networking options for low latency

Reducing latency in multi-cloud and hybrid environments requires careful consideration of network architecture and the selection of appropriate data transfer methods. By optimizing the underlying network configuration and using the right cloud services, organizations can minimize delays and ensure faster data transfers.

#### Approaches to minimizing latency:

- **Google Cloud Interconnect:** Utilize Dedicated Interconnect or Partner Interconnect to establish high-speed, low-latency connections between on-premises systems and Google Cloud, bypassing public internet congestion.
- **Direct Peering:** Set up Direct Peering with Google Cloud to enable a private connection, ensuring faster and more reliable data transfers between Google Cloud and other cloud providers or on-premises systems.
- **VPN Optimization:** For organizations using VPNs, optimize configurations to reduce overhead and enhance throughput, ensuring that the VPN connection is not a limiting factor in data transfer speed.
- **Edge Caching:** Implement edge caching to store frequently accessed data closer to users or systems, reducing the need for repeated long-distance data transfers and improving response times.

### Ensuring efficient data encryption without compromising performance

While encryption is critical for data security, it should not significantly hinder data transfer performance.

To strike the right balance between security and efficiency, organizations can adopt encryption strategies that minimize overhead while ensuring compliance with data protection regulations.

### Solutions for efficient encryption and minimal impact on performance:

- **Hardware-Accelerated encryption:** Leverage specialized hardware, such as Google Cloud's Confidential VMs or customer-supplied encryption keys, to accelerate encryption and decryption processes without sacrificing performance.
- **End-to-End encryption at rest and in transit:** Use efficient encryption methods that provide robust protection without excessively slowing down transfers, ensuring that data remains secure throughout its journey.
- **Optimized key management:** Implement cloud-native key management solutions that streamline key rotation and management, reducing the impact on performance while maintaining high security standards.

### Optimizing load balancing for scalability and peak demand

Load balancing plays a critical role in managing data transfer efficiency, particularly during periods of high traffic or peak demand. Proper configuration and scaling of load balancing resources can ensure that data is evenly distributed, reducing the risk of congestion and ensuring optimal performance.

#### Best practices for optimizing load balancing:

- **Google cloud load balancing:** Use Google Cloud's global load balancing solutions to distribute traffic across multiple resources, ensuring that no single instance becomes a bottleneck during high-demand periods.
- **Auto-Scaling:** Implement auto-scaling policies that automatically adjust resources based on traffic loads, ensuring that your infrastructure can handle fluctuations in demand without sacrificing performance.
- **Global Distribution:** Utilize global load balancing to direct traffic to the nearest available resources, reducing latency and improving the user experience for geographically dispersed users.

### Selecting the right data transfer method for each scenario

Choosing the correct data transfer method is crucial for balancing cost, performance, and scalability. By understanding the specific needs of each data transfer scenario, organizations can select the most appropriate solution to meet their goals.

#### Key strategies for selecting the right transfer method:

- **Google cloud interconnect:** For high-volume or latency-sensitive transfers, Google Cloud Interconnect provides a dedicated, private connection with high throughput and low latency, ideal for critical business operations.
- **VPN:** Virtual Private Network (VPN) is an effective solution for organizations that require secure data transfers over public internet connections, particularly when low-cost or moderate-performance solutions are acceptable.
- **Direct Peering:** Direct Peering enables private connections between cloud providers or between Google Cloud and on-premises systems, optimizing performance while maintaining data security.
- **Hybrid data transfer methods:** In hybrid cloud environments, leverage a combination of these methods to optimize data transfer based on the type of data and transfer requirements, ensuring cost-effective and high-performance operations.

### Managing Multi-Cloud transfers for seamless integration

For organizations that operate in multi-cloud environments, it's essential to have strategies in place to manage cross-platform data transfers efficiently. With proper integration and data management, businesses can overcome the challenges of compatibility, security, and performance inconsistency.

#### Solutions for effective multi-cloud and multi-provider integration:

- **Cloud-to-Cloud Interconnects:** Use inter-cloud connections and APIs to facilitate seamless data transfers between Google Cloud and other cloud providers, minimizing latency and maximizing transfer efficiency.
- **Unified management platforms:** Implement multi-cloud management platforms to streamline monitoring and control of data transfers across various environments, ensuring that resources are efficiently allocated.
- **Cross-Platform security protocols:** Adopt unified security standards and protocols (such as end-to-end encryption) across all platforms to maintain a consistent security posture without sacrificing performance.
- **Data Standardization:** Ensure data formats and structures are compatible across platforms to avoid delays caused by data conversion or transformation during transfers.

These proposed solutions aim to address the key bottlenecks and latency issues identified earlier, ultimately enhancing the efficiency, speed, and reliability of data transfers across Google Cloud, other cloud platforms, and on-premises systems.

By implementing these strategies, organizations can optimize their data transfer processes, ensuring smoother operations, reduced costs, and improved business outcomes.

### 5. Recommendations

To effectively address the challenges and bottlenecks discussed in previous sections, organizations must adopt a combination of strategies and best practices to optimize data transfer efficiency, minimize latency, and ensure scalability. By leveraging Google Cloud's tools and services in conjunction with industry best practices, organizations can ensure that their data flows smoothly between cloud environments, improving operational performance and reducing costly delays.

The following recommendations outline key strategies to maximize transfer efficiency and mitigate the challenges associated with cloud-to-cloud and cloud-to-on-premises data transfers.

#### Optimize network architecture for seamless data transfers

A well-designed network architecture is critical to ensuring efficient data transfers. By choosing the right connection types and optimizing network performance, organizations can reduce latency and avoid bottlenecks that impede data flow.

#### Recommended network architecture optimizations:

- **Google cloud interconnect:** Utilize Dedicated Interconnect or Partner Interconnect for high-performance, low-latency connections between Google Cloud and on-premises or third-party cloud environments. This ensures a private, secure, and faster

data transfer channel, bypassing the public internet.

- **Direct Peering:** Set up Direct Peering with Google Cloud to establish a private connection with other cloud providers or on-premises systems. This reduces congestion on the public internet and improves data transfer speed.
- **Cloud VPN:** If using VPNs, configure them for optimal performance by selecting the right instance types and adjusting MTU settings to prevent fragmentation, reducing latency and improving throughput.

#### Enhance data transfer efficiency with compression and deduplication

Optimizing the data itself before transfer can significantly improve speed and reduce the strain on available bandwidth. Compression and deduplication are effective techniques to reduce the amount of data being transferred, ensuring that only necessary information is sent.

##### Key methods for enhancing data transfer efficiency:

- **Data Compression:** Implement compression algorithms (such as gzip or Snappy) to reduce the size of data files before they are transferred. This minimizes the amount of bandwidth required and speeds up the overall transfer process.
- **Data Deduplication:** Use data deduplication to eliminate duplicate data before transfer, ensuring that only unique information is sent across the network. This reduces unnecessary load on the network and improves transfer efficiency.
- **Optimized data formats:** Consider using efficient data formats (e.g., Parquet or Avro) that are both compressed and optimized for analytical workloads. These formats minimize data size while maintaining the integrity of the transferred information.

#### Implement Low-Latency data transfers with real-time ingestion

To ensure that data is transferred and available for analysis in near real-time, organizations should utilize streaming technologies and optimization techniques that reduce latency and ensure timely data access.

##### Recommendations for minimizing latency in data transfers:

- **Streaming data ingestion:** Leverage real-time streaming technologies such as Google Cloud Pub/Sub and BigQuery's streaming inserts to continuously ingest data with minimal delay. This ensures that newly generated data is available for analytics as soon as it arrives.
- **Partitioning and clustering in BigQuery:** Use BigQuery's partitioning and clustering features to optimize how data is queried, reducing latency when accessing large datasets and improving query performance.
- **Regionalization:** Store data in regions close to where the processing occurs to minimize network latency. This helps to optimize performance by reducing the distance data must travel between sources and destinations.

#### Utilize auto-scaling and elastic resources for consistent performance

As data volumes fluctuate, ensuring that infrastructure can scale dynamically is critical to maintaining consistent performance. Google Cloud's elastic resources and auto-

scaling capabilities provide the flexibility to handle spikes in data transfer demands.

##### Strategies for scaling resources efficiently:

- **Auto-Scaling Resources:** Configure auto-scaling for both compute and storage resources in Google Cloud to automatically adjust to fluctuations in workload, ensuring consistent performance during peak times.
- **Elastic bigQuery resources:** Take advantage of BigQuery's ability to scale horizontally to handle larger datasets and higher query demands. This ensures that organizations can process large amounts of data without compromising speed or accuracy.
- **Resource Prioritization:** Implement workload prioritization to ensure that high-priority processes or queries are allocated sufficient resources to avoid delays during heavy workloads.

#### Strengthen security without impacting performance

While security is crucial, it's essential that encryption and other security measures do not excessively slow down data transfers. By using efficient security protocols, organizations can ensure secure transfers without compromising performance.

##### Best practices for secure and efficient data transfers:

- **Hardware-Accelerated Encryption:** Utilize Google Cloud's hardware-accelerated encryption technologies to speed up the encryption and decryption process without sacrificing security.
- **End-to-End Encryption:** Employ end-to-end encryption during data transfers to protect sensitive information while minimizing the impact on performance by using optimized encryption algorithms.
- **Key management optimization:** Implement efficient key management practices, such as using Google Cloud's Key Management Service (KMS) to automate key rotation and manage encryption keys at scale, without introducing delays in the transfer process.

#### Optimize data transfer tools for multi-cloud and hybrid environments

In multi-cloud and hybrid environments, it's important to have seamless integration between different cloud platforms and on-premises systems. Choosing the right transfer tools and configuring them correctly can ensure smooth, efficient data flows.

##### Recommendations for optimizing multi-cloud and hybrid data transfers:

- **Cross-Platform transfer solutions:** Use tools like Google Cloud Storage Transfer Service or Transfer Appliance to facilitate fast, secure data movement between cloud platforms, ensuring seamless transfers across different environments.
- **Unified data management platforms:** Implement unified data management tools to monitor and control data flows across all platforms, ensuring visibility, compliance, and efficiency during the transfer process.
- **Data standardization across platforms:** Standardize data formats and structures across cloud and on-premises environments to ensure compatibility, minimize the need for data transformations, and reduce delays during data transfers.

By implementing these recommendations, organizations can significantly enhance their data transfer performance,

minimize bottlenecks, and ensure that their cloud and hybrid environments are optimized for speed, scalability, and security.

These best practices will not only reduce latency but also enable faster decision-making and more efficient operations, providing a competitive edge in today's data-driven landscape.

## 6. Conclusion

The discussion in this paper highlights the importance of optimizing data transfers between Google Cloud, other cloud platforms, and on-premises systems for improved performance and scalability.

Addressing challenges like network congestion, data quality, and resource inefficiencies is essential for ensuring seamless data movement and real-time analytics.

By implementing strategies such as robust data validation, optimized BigQuery configurations, auto-scaling resources, and secure encryption practices, organizations can overcome these obstacles and leverage Google Cloud's full potential.

These best practices not only enhance data ingestion efficiency but also empower businesses to make timely, data-driven decisions, driving operational success and supporting sustainable growth in an increasingly digital world.

## 7. References

1. Sajjad A. A secure and scalable communication framework for inter-cloud services [dissertation]. London: City University London; 2015.
2. Kaur R, Chana I, Bhattacharya J. Data deduplication techniques for efficient cloud storage management: A systematic review. *The Journal of Supercomputing*. 2018;74:2035-2085.
3. Puthal D, Wu X, Surya N, Ranjan R, Chen J. SEEN: A selective encryption method to ensure confidentiality for big sensing data streams. *IEEE Transactions on Big Data*. 2017;5(3):379-392.
4. Mohamed N, Al-Jaroodi J, Eid A. A dual-direction technique for fast file downloads with dynamic load balancing in the cloud. *Journal of Network and Computer Applications*. 2013;36(4):1116-1130.
5. Kommera AR. The role of distributed systems in cloud computing: Scalability, efficiency, and resilience. *NeuroQuantology*. 2013;11(3):507-516.
6. Younge AJ. Architectural principles and experimentation of distributed high-performance virtual clusters [dissertation]. Bloomington: Indiana University; 2016.
7. McGinley R, Etter J. Storage and analysis of big data tools for sessionized data. [Journal Name]. 2015. [Pending journal name and volume details].
8. Coutinho EF, de Carvalho Sousa FR, Rego PAL, Gomes DG, de Souza JN. Elasticity in cloud computing: A survey. *Annals of Telecommunications*. 2015;70:289-309.
9. Ali SMF, Wrembel R. From conceptual design to performance optimization of ETL workflows: Current state of research and open problems. *The VLDB Journal*. 2017;26(6):777-801.
10. Demchenko Y, Turkmen F, de Laat C, Hsu CH, Blanchet C, Loomis C. Cloud computing infrastructure for data-intensive applications. In: *Big Data Analytics for Sensor-Network Collected Intelligence*. Academic Press; 2017. p. 21-62.