Hitachi Vantara

WHITE PAPER

# Harmonizing Compute and Storage in the New World of Infrastructure as a Service — a Symbiotic Relationship

**While compute and storage serve different functions, they are always intertwined. Understanding how they can complement each other in a consumption-based model is crucial for optimizing IT operations.**

## Introduction

In today's digital age, where organizations continually strive for operational excellence, the strategic determination regarding compute and storage infrastructure has become pivotal for business success. Infrastructure modernization and the transition to Infrastructure-as-a-Service (IaaS) models have become the cornerstones for organizations aiming to thrive and succeed. Central to this transformation is the role of compute and storage solutions, ensuring that the deluge of data created, consumed and processed is effectively managed, secured and optimized for the diverse range of applications and services that power modern enterprises.

Every transaction, every customer interaction and every innovation an enterprise undertakes today has data at its core. As we move further into the era of Artificial Intelligence (AI) and cloud-native applications, the amount of data generated and the complexity of managing it only increases. This is both a challenge and an opportunity. The challenge is to efficiently store, retrieve and process this data. The opportunity, on the other hand, is to harness this data, turn it into actionable insights, and ultimately deliver enhanced value to customers and stakeholders.

Compute and storage are not merely interrelated. They also wield considerable influence over each other. We will dive into the dynamic relationship, unearthing the profound impact each has on the other shifting equilibriums based on factors such as scalability, data processing, virtualization, and so on.

### Challenges and Considerations

Organizations are undergoing rapid digital transformation constantly seeking ways to innovate and digitize their operations. This drive often leads to substantial investments in compute infrastructure to power new applications and services in emerging technologies such as artificial intelligence, IoT and big data analytics. These technologies demand robust computational power, driving organizations to procure more powerful servers and processors. However, compute infrastructure investments generate an influx of data as rapid processing of information increases data creation, necessitating larger and more sophisticated storage solutions. Some of the factors that impact and drive an organization's storage infrastructure decisions to match the ever-increasing power of compute are as follows:

| Top Industry Challenges | |
|---|---|
| **Data Proliferation** | High-end compute infrastructure and digitization initiatives are churning data at an unprecedented rate. Organizations are experiencing a constant surge in data generation due to the continual launch of new services and applications powered with swift processing capabilities. This surge in data leads to the need for more extensive, scalable, and advanced storage solutions. According to a recent survey conducted by Statista, the volume of data is projected to exceed 180 zettabytes by 2025, with an average annual growth rate of approximately 25%. However, only a minimal two percent of the data generated in 2020 was retained into 2021. |
| **Data Management Strategy** | As data generation surges, the imperative for effective data management becomes increasingly apparent. Effective data management requires organizations to have well-structured management strategies and solutions. Without robust data management practices in place, organizations may find themselves overwhelmed by the deluge of data and face challenges in data quality, security, accessibility and compliance. |
| **Data Retrieval & Processing** | High-end computing infrastructure such as accelerator cards or GPU-based processors require storage solutions that can keep up with the speed of data retrieval and processing. Slow storage can often become a bottleneck in the quest for faster compute services. |
| **Data Redundancy** | Parallel and distributed computing infrastructure may duplicate data across multiple servers, leading to data redundancy. Organizations must account for deduplication of data when evaluating storage solutions. |
| **Data Protection** | Compute solutions primarily focus on processing power and speed, undermining the critical necessity of data protection. However, data protection should be a parallel priority, if not a higher one. By applying comprehensive measures such as access controls, encryption, backups, archival and compliance to maintain trust and security, organizations can ensure that their data remains secure and uncompromised, even in the face of compute infrastructures that have the potential to strain storage and processing resources. |
| **Data Synchronization** | In solutions where compute and storage are not in sync, there is a risk of data loss or exposure. Storage solutions need to be carefully selected so that they can keep up with the speed of data generated to avoid such data losses. |
| **Data Migration** | Overhauls of compute solutions often change an environment or platform and prompt data migration from one platform to another. An example is from on-premises to public cloud or vice versa. These situations require scalable and elastic storage solutions that can accommodate data growth and integrate out of the box with industry-standard platforms. |
| **Data Reorganization** | Inefficient storage structures may necessitate the need for data reorganization and tiering based on the compute solutions to match the new data volumes, classifications and access patterns. |

## Storage Adaptations in the Age of Compute

# The world's data is growing at an exponential rate, and as a result, compute and storage solutions are evolving rapidly to meet the ever-expanding needs of businesses.

Emerging technologies, hybrid cloud and the ascent of object storage are transforming the storage landscape, paving the way for robust, scalable, sustainable and eco-friendly data storage solutions. According to statistics published by Statista, the installed base of global data storage capacity, keeping pace with the data explosion, is expected to increase from 6.7 zettabytes (ZB) in 2020 to around 16 zettabytes in 2025 with a forecasted compound annual growth rate of 19.2 percent.

To comprehend the dynamics of storage allocation, consider a pool of 1000 Virtual Machines (VMs) running both stateful and stateless applications. For simplicity, let's assume that half of these VMs are stateful with each application on these VMs consuming 100 GB of storage. Applying a High Availability (HA) factor of 3, we have 167 stateful VMs (500 divided by 3), dragging 16.7 terabytes of storage. The narrative

extends beyond the immediate storage needs, prompting us to contemplate the trajectory of storage volume growth. The opportunity lies in creating storage solutions that not only meet the present demands but also scale seamlessly to accommodate the anticipated surge in storage volumes as these applications evolve.

Further, let's explore, below, some of these industry trends contributing to this growth:

## Virtualization & Containerization

As organizations seek agility, flexibility and efficiency, the virtualization of compute resources has become a catalyst for the evolution of software-defined storage (SDS) solutions. The rise of virtualization has decoupled applications from physical infrastructure, enabling dynamic allocation and scaling of compute resources. This newfound agility sets the stage for broader transformation across IT stacks, including storage.

In the traditional model, storage was tethered to specific servers, leading to inefficiencies, but the decoupling of compute from physical hardware has liberated data from these constraints. SDS represents a fundamental shift in how we utilize and manage storage infrastructure.

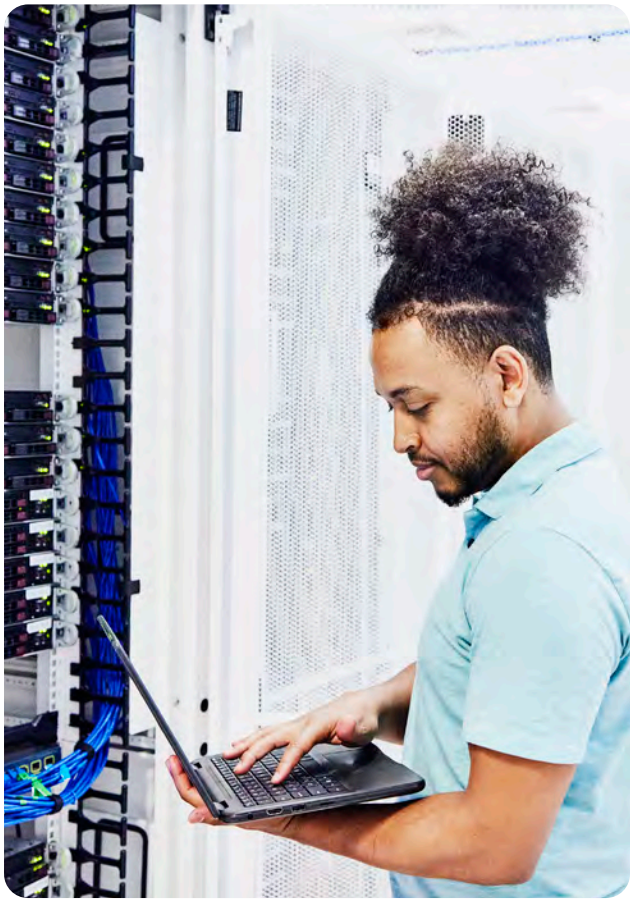
Organizations often struggle to manage their heterogeneous infrastructure and end up using disparate tools provided by respective vendors to achieve desired results. Virtualization allows organizations to build a common unified platform for both compute and storage, using the respective abstractions to manage all their infrastructure needs in one place. Unlike the traditional storage solutions that are tied to specific hardware, SDS abstracts capabilities from the underlying physical devices, offering unprecedented flexibility, scalability and efficiency. It empowers organizations to adapt to changing demands in data and ensure that their storage infrastructure remains responsive to meet the dynamic data growth.

SDS offers a centralized control plane that can be used by administrators to define and modify storage policies and configurations across the entire infrastructure. It decouples the logical storage from the physical hardware on which they are deployed, creating a virtual layer. This allows data to be stored and retrieved without being dependent on any specific storage device. SDS is often complemented with advanced automation and orchestration capabilities that allow for dynamic provisioning and deprovisioning, replication and migration of data without the need for manual intervention, reducing human errors. It is highly scalable and as the data needs grow, additional storage can be provisioned dynamically into the existing infrastructure without disrupting operations. It runs on commodity hardware that can significantly lower the total cost of ownership while still providing enterprise grade capabilities.

## Unstructured Data – Ascent of Object Storage

Traditional data storage is primarily based on block and file storage as most of the data generated has been structured. However, with the advent of AI/machine learning (ML) and NoSQL databases that store and process unstructured data, these storage systems struggled to efficiently manage the sheer volume and complexity of such data in contemporary computing environments.

Object storage is a shift in how data is stored and retrieved. Instead of storing data as blocks or files, object storage stores data as objects that encapsulate, data, metadata and a unique identifier or key. This key or identifier allows data to be easily indexed, retrieved and organized, even in massive data sets. It is agnostic to the type or format of data, making it ideal for storing unstructured data produced by modern-day



applications. The rich metadata included with each object allows for efficient indexing, cataloging and search, making it ideal for use cases such as AI\ML where search based on attributes is of paramount importance.

## Hyperconverged Compute Solutions

GPUs are known for their exceptional parallel processing. Their speed and responsiveness create a need for equally high-performing storage. Traditional hard disk drives can cause a bottleneck in data transfer rates, leading to lower GPU utilization. Hyperconverged infrastructure solutions have emerged as a response to the increasing demand for efficient and scalable data storage to match the rapid progress in compute technologies. Hyperconverged infrastructure is fairly innovative and integrates compute, storage and networking components into a single unified box. This consolidation streamlines operations and simplifies architecture, as it eliminates the need to manage disparate systems and hardware.

Hyperconverged systems provide elasticity because they are inherently scalable, allowing seamless additions of new nodes to the cluster when storage/compute/network needs grow. It doesn't necessitate a significant overhaul of infrastructure, as is often the case with traditional solutions. With NVMe, drives that communicate directly with CPU/GPU, storage can be extremely fast with storage and retrieval occurring at lightning speeds, taking resource efficiency to new heights.

## Infrastructure as a Service (IaaS)

The concept of Infrastructure as a Service (IaaS) represents a metamorphic evolution in IT consumption models, offering a spectrum of services under its umbrella. Storage as a Service (STaaS) follows the IaaS model where everything including compute is exposed as a service, allowing organizations to access and utilize storage without having to own the underlying infrastructure. STaaS uses capabilities of cloud infrastructure, public or private, to provide storage resources rather than relying upon on-premises storage solutions. This provides flexible and scalable storage services, eliminating the need for organizations to invest in and maintain their own storage hardware. It reduces upfront capital expenditures and streamlines data management with on demand storage.

Many STaaS providers offer pay-as-you-go models bridging the consumption gaps between usage and overprovisioning or underutilization, often saving significant costs. STaaS platforms ensure that data is accessible from virtually anywhere with a connection. These platforms include robust data protection features such as redundancy, backup and recovery along with built-in security measures to prevent any loss or corruption of data, ensuring resiliency and business continuity.

## Hybrid and Multi Cloud Compute Solutions

Hybrid and public cloud compute solutions are gaining prominence as a strategic approach for organizations. Many organizations recognize the value of combining on-premises and cloud resources to meet their dynamic requirements. Hybrid and multi-cloud storage approaches have emerged as transformative solutions to match the compute solutions, offering businesses a dynamic, efficient and scalable way to manage their data. These approaches combine the strengths of multiple cloud and on-premises environments and enjoy the best of both worlds, balancing performance, costs and redundancy, enabling them to achieve unprecedented levels of resiliency and flexibility.

Hybrid cloud storage blends on-premises infrastructure with cloud-based resources, giving organizations the ability to integrate their existing data centers and storage solutions with public or private cloud-based solutions. It allows them to use cloud for non-sensitive or archival data, freeing up on-premises storage for critical applications and workloads. It provides flexibility to move data between on-premises and cloud storage as requirements change, maintaining the flexibility to adapt quickly.

Multi cloud storage also allows organizations to spread their data and applications across different clouds, avoiding vendor lock-ins and improving availability. This diversification minimizes the risk of service outages and reliance on a specific vendor providing price flexibility.

## Internet of Things and Edge Computing

Internet of Things (IoT) is a network of interconnected devices that continuously collect and exchange information, generating vast amounts of data especially in applications like industrial IoT, healthcare, etc. This data includes sensor readings, telemetry, environmental data and more. Storing and managing this large amount of data requires robust and scalable storage solutions. Many of these applications require real-time or near real-time processing of data for timely decision-making, adding further pressure on the storage infrastructure to handle high-speed ingestion and retrieval.

IoT devices often utilize edge computing to process data locally before sending it to centralized servers. The local processing helps filter and analyze data closer to the source, reducing the amount of data that needs to be transmitted and stored centrally. This impacts the storage architecture, requiring distributed storage solutions where data is stored both locally and in centralized repositories based on access patterns, characteristics of the datasets and scalability requirements. Ensuring data consistency (strong or eventual) across different locations can be complex.

## Quantum Computing

Quantum computing represents a paradigm shift and is a disruptive computing technology that demands new approaches to data storage and processing. In quantum computing, the qubits (short for quantum bit) replace the traditional bits, which can represent multiple states simultaneously including ones and zeros. This allows quantum computers to process and generate vast amounts of information in parallel.

Unlike the traditional bits that are stable and easily stored in traditional storage, qubits are highly sensitive to environmental factors such as temperature, which can cause loss of qubits or coherence and, consequently, the information stored in it. As such, quantum storage media, such as superconducting circuits or trapped ions must be stored at extremely low temperatures to preserve the fragile quantum states. Scaling quantum storage to match the quantum computers that are capable of processing larger qubits remains a challenge. As this technology advances, addressing challenges in storing and maintaining the integrity of qubits will be a crucial step forward for realizing its full potential.

## Strategy for Success

As compute and storage infrastructure needs grow, organizations seek seamless, integrated solutions that combine powerful compute resources with efficient storage capabilities, all managed though a user-friendly interface. Hitachi Vantara fulfills this growing demand for integrated solutions by providing a unified platform that not only delivers exceptional performance but also streamlines management.

EverFlex from Hitachi provides IaaS solutions and combines Hitachi's expertise in both compute and storage to offer tangible benefits in performance, cost efficiency and scalability, meeting the dynamic demands of your applications. Our close partnership and certifications with Cisco further enhance this unified platform approach and capabilities. Our IaaS solutions are bundled with hardware, software, storage, a control plane and related networking and managed services, reducing the hassles of traditional IT setups to simplify procurement and maintenance. This reduces deployment times and compatibility issues to drive significant TCO savings. The solutions integrate out of the box with industry standard storage products across different storage classes such as file, block and object. It involves creating a single control plane and data plane across block, file, object, cloud, mainframe and SDS workloads, managed by a single AI-enabled software stack. The solutions operate seamlessly in both converged and hyperconverged storage space and can be managed through a control plane that uses orchestration and automation to reduce human errors.

Hitachi recognizes the need for tailored and adaptable solutions and we endeavor to create tools that extend the functionalities of existing systems. Hitachi's Storage Plugin for Containers (HSPC) is a plugin that is compatible with the industry standard Container Storage Interface (CSI) and allows it to integrate with any CSI-compatible storage device. Hitachi Replication Plug-In for Containers (HRPC) is another plugin that enables enterprise data services in container environments for stateful storage with the Hitachi VSP and E-Series family.

All our offerings are backed with Hitachi's managed services, delivered as IaaS and pay-per-consumption pricing models that provide predictable expenses that have delivered years of successful service to its clients with guaranteed SLAs. The pay-as-you-go model offers IaaS flexibility with predictable expenses.

## Conclusions and Final Thoughts

The relationship between compute and storage is both intricate and impactful, shaping the outline of digital transformation initiatives. At the heart of this synergy lies the dynamic nature of data and processing power, creating a symbiotic bond. The ascent of cloud computing acts as a catalyst, and as organizations embrace cloud-native solutions, the integration and harmonization of compute and storage become more seamless. Vendors and solution providers recognize the evolving needs of organizations offering integrated solutions that cater to both computational and storage needs.

In conclusion, as computation demands evolve with innovation, and data proliferates, understanding how compute drags storage is pivotal for organizations seeking efficiency, resilience and optimal resource utilization as part of their strategic IT planning. Organizations that grasp the intricacies of this relationship are poised to unlock the full potential of their digital infrastructure solutions. While embracing scalable architectures, organizations should leverage advancements in storage technologies and adopt intelligent data management practices.

**Learn More →**

Click here to learn more about EverFlex solutions from Hitachi Vantara, delivered as a cost-saving, fully managed infrastructure-as-a-service model.

Hitachi Vantara
A Hitachi Group Company

**Corporate Headquarters**
2535 Augustine Drive
Santa Clara, CA 95054 USA
hitachivantara.com | community.hitachivantara.com

**Contact Information**
USA: 1-800-446-0744
Global: 1-858-547-4526
hitachivantara.com/contact



HV-BTD-WP-Harmonizing-Compute-and-Storage-18Jan24-A