SPONSORED BY IC manage

GIGAOM

REPORT

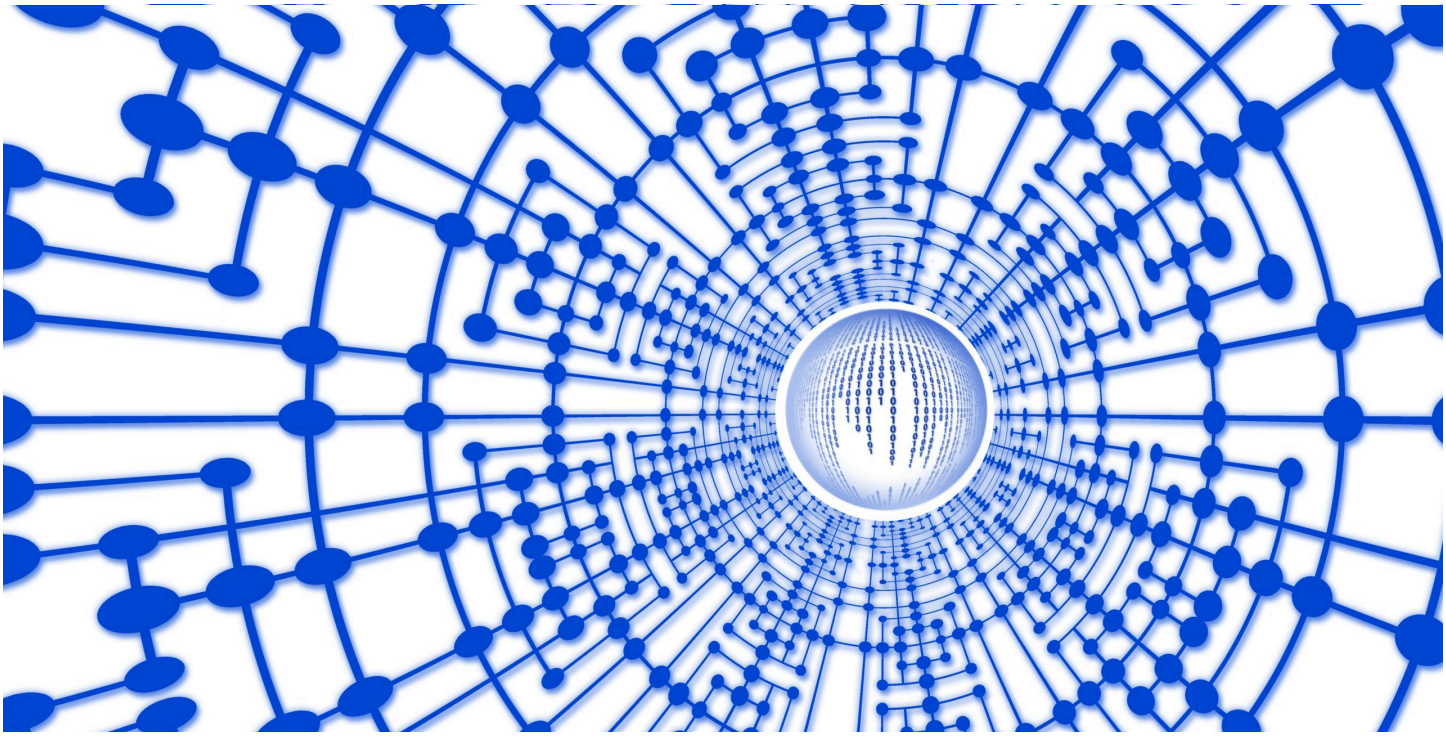
Manage Highly Demanding Storage Workloads in Hybrid Cloud Scenarios with IC Manage Holodeck

v1.0

An Innovative Distributed Caching Technology for High-Performance Workloads

ENRICO SIGNORETTI | FEB 16, 2021 - 2:45 PM CST

TOPICS: DATA STORAGE MULTI CLOUD & HYBRID CLOUD



GIGAOM

Manage Highly Demanding Storage Workloads in Hybrid Cloud Scenarios with IC Manage Holodeck

An Innovative Distributed Caching Technology for High-Performance Workloads

TABLE OF CONTENTS

1. Summary

Alongside the explosive data growth that every organization is experiencing, there is an increasing demand for big data analytics, AI, commercial HPC, and other high-demand workloads. At the same time, enterprises want the flexibility of the cloud while keeping data and costs under control.

In theory a hybrid cloud strategy, for example, to take advantage of cloud bursting for additional processing power looks like the right choice; however, keeping data far from compute resources is very inefficient due to the increased latency. What's more, high-performance cloud storage is very expensive, while egress fees and other hidden costs can quickly discourage its use for interactive workloads.

Moving data is also difficult, expensive, and inefficient. Even thinking about selecting a data set and moving it to a remote place can be challenging, especially if we think that we need the resulting data back after the compute job is ended. The sticky nature of data is usually described as "data gravity." This is not news, but it is very important to understand before looking for a solution.

Every organization hopes to consolidate all its data in a single place, to make it more accessible and controllable, and to avoid the creation of data silos. But as already mentioned, moving large data sets back and forth is expensive and impractical, and accessing them remotely is challenging and inefficient as well. This is an issue that is becoming even more severe with edge-cloud and multi-cloud infrastructures and data repositories that can be located far from the applications. Expensive connectivity options can alleviate some of these issues, like bandwidth requirements, but they are not dynamic and cannot bring data closer to the CPU.

Financially speaking, this is a nightmare:

- Inefficiency introduced by latency results in poor CPU utilization, longer processing time, and higher compute costs.
- Moving entire data sets to the compute nodes for the job and then back when they are processed can dramatically increase cloud storage bills because of egress fees and the number of transactions necessary to complete the job.

Costs increase and become unpredictable, posing several risks to the business and raising doubts about the sustainability of this model for the future.

Cloud providers built formidable virtual compute infrastructures that include acceleration options like FPGAs and GPUs for reasonable prices. Unfortunately, cloud storage often offers less flexibility, lacking the mechanisms to create cost-effective solutions for highly demanding workloads. The most interesting and cost-effective approach to bringing data close to the CPU, and to reducing latency with high-capacity volumes, is to separate the performance and access layer from capacity and protocol dependency. By doing so, it is possible to build a caching mechanism that works across long distances

and the Internet, pushing hot data to the CPU and the applications that need it.

IC Manage Holodeck takes a unique approach to caching that enables users to take advantage of flash memory installed in the compute nodes and shared across high-speed local networks to intelligently cache data and metadata efficiently. The result: IC Manage enables users to accelerate many highly demanding workloads, both on-premises and in the cloud, to provide an impressive ROI quickly.

2. The Problem

Human and machine-generated data are growing, but to address this trend we must access data properly, no matter where it is stored. In fact, no matter how we store it, in a single big data lake or across smaller repositories, it is highly likely that compute power is not always close to the data. There are several reasons for this:

- Specific applications are available only in the cloud, but the data is on-premises.
- Compute power in the cloud is cheaper and paid for on a consumption basis.
- Data is generated by several sources, consolidated in one location, and accessed by different, globally dispersed applications.
- The organization is moving from a single or hybrid cloud approach to multi-cloud.

Data has gravity, and moving it, especially large quantities of it, is often not feasible. Creating a copy of data for a single job is expensive, takes time, and can create security issues. In some cases, it is not even possible to isolate a specific data set before starting a complex job. At the same time, accessing data remotely is very challenging as well.

There is a combination of factors that make remote data access for data-intensive workloads difficult and expensive:

- **Latency.** When data has to traverse long distances and different networks to reach the compute node, every IO operation puts the CPU in a waiting mode. This severely limits overall system efficiency and ability to execute complex jobs in a reasonable time.
- **Poor CPU utilization.** This can have a major impact on cloud costs. Compute power is paid for by the minute and every minute of wasted time lands directly on the bottom line of the cloud bill at the end of the month. Depending on effective CPU utilization, costs can be up to three or four times higher than normal with longer times to get a result.
- **Protocols.** Usually unstructured data at the base of modern big data workloads is accessed via file sharing protocols like NFS or SMB, which are not designed to work over long distances and internet connections. The communication is very chatty and inefficient, creating additional issues. S3 is an alternative from an ease of access point of view, but HTTPS, used as a transport layer, is highly inefficient for interactive workloads. Again, poor network utilization heightens cost. On one hand, we need more bandwidth than is effectively necessary to move data, and on the other, the inefficiency of the protocol introduces costs related to the number of operations involved in browsing and accessing data. Many providers charge additional fees for IO operations, a situation that can easily become unsustainable with small files.
- **Size of filesystems.** Most big data applications are based on files and, again, need to access large amounts of data. Even when the data set is a fraction of the total, just navigating the file system can

generate a large number of operations on metadata. A combination of protocol inefficiencies, latency, and bandwidth can easily undermine the ability to work with large file systems, and just browsing the file system can be challenging and time consuming.

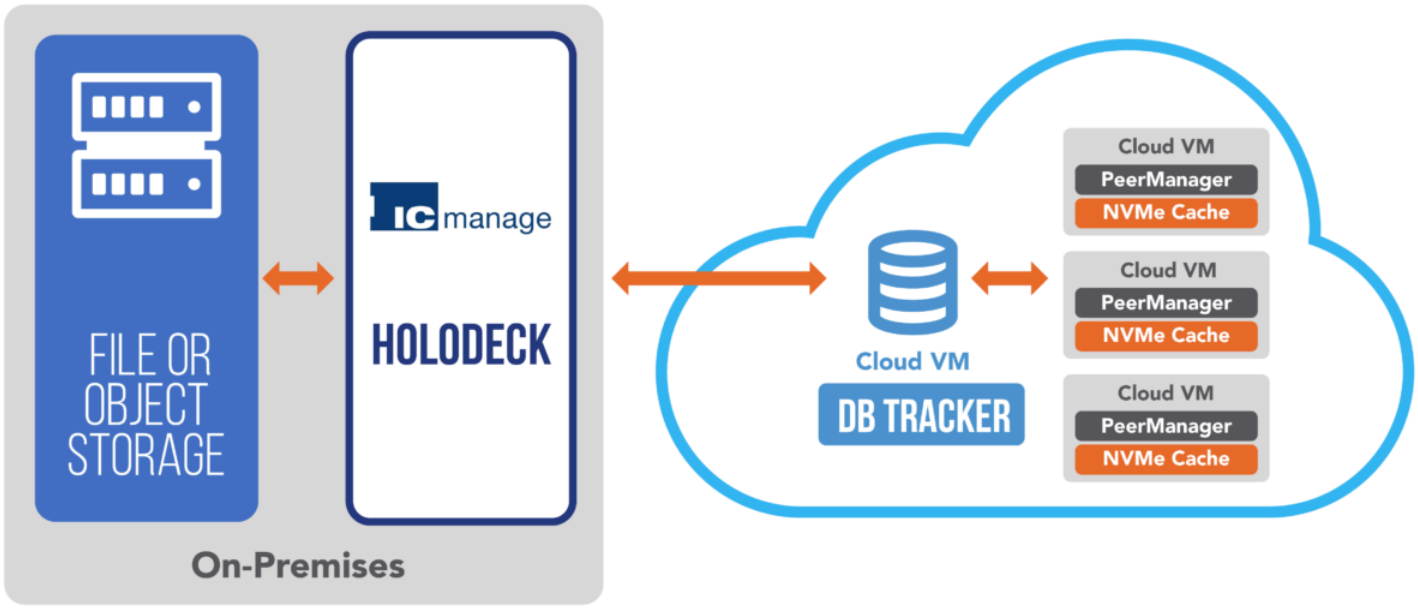
This is not a problem of pure cost but of TCO as well. In fact, we may need to change directory structures, processes, and application logic as a workaround. All this is incredibly expensive and not feasible in most cases.

Throughput is another issue that can be solved partially with better and more expensive connectivity, however inefficiencies introduced by latency and protocols not designed for long-distance and Internet communications can undermine this type of investment. And that is without accounting for the time and investment needed to establish sufficient bandwidth between different locations.

Combined, these factors contribute to increasing the time that compute nodes wait for data, making it difficult and expensive to operate on data in hybrid and multi-cloud scenarios. The result: an unfavorable overall cost structure and unpredictable performance across different workloads.

2.1 The Solution

IC Manage Holodeck is a software-defined storage solution that addresses the challenges discussed above, enabling users to access data quickly, no matter the size or distance of the file system. It creates a virtual projection of remote storage resources on any cloud or on-premises compute nodes. It is based on a scale-out distributed caching architecture that takes advantage of flash memory installed in the nodes of the compute cluster. All nodes contribute with the amount of free flash memory they have locally to create a shared caching pool. Each single node has access to the entire cache, which is shared across the cluster through high-speed Ethernet connections.



Source: GigaOm 2021

Figure 1. The Holodeck Architecture

Thanks to this mechanism, every node added to the compute cluster contributes additional resources and increases overall cluster performance. And because the flash memory installed in the compute nodes is relatively cheap, the user can add an abundant amount of memory when needed.

The shared cache enables applications to read and write hot data quickly and keeps the CPUs engaged with active tasks. The result is better CPU utilization, which can be leveraged in two ways:

1. Use fewer compute nodes to do the job in the same amount of time
2. Use the same number of nodes to get results faster

No matter which alternative the user chooses, from a financial perspective, the cloud bill goes down.

The caching mechanism separates performance from capacity, but IC Manage Holodeck does more than that and optimizes the communication aspect as well. In fact, an additional software component, The Holodeck gateway, is installed next to the storage system, acting as a protocol translator to abstract physical storage access and make it Internet- and latency-friendly.

The gateway eliminates all the unnecessary chatting and optimizes operations whether the data is stored in an object store or in a file system. Holodeck gateway also optimizes metadata handling, giving the compute nodes complete and immediate visibility into the remote file system without needing to continuously communicate while navigating through it. This eliminates another big problem and makes any type of file or object store accessible efficiently from practically anywhere.

By optimizing end-to-end communication and keeping hot data next to compute nodes, latency issues are drastically reduced, making it possible to run several workloads efficiently over long distances, with limited bandwidth, or just by accelerating the storage system. Again, all these improvements contribute to decreased costs as well.

3. The Benefits

IC Manage Holodeck is a software-defined storage solution that can accelerate several types of latency-sensitive workloads easily and can be deployed for many different use cases in scenarios such as:

- **Hybrid cloud:** Many users have huge amounts of data stored on-premises and they prefer to use compute resources in the cloud to avoid CapEx spending while getting the right infrastructure for every workload. Cloud resources can be consumed and paid for by the minute, minimizing investment and optimizing budgeting. Thanks to IC Manage technology, latency is reduced, and compute efficiency is highly optimized, driving down cost and helping users get results faster and spend less money to do it.
- **Multi-cloud:** IC Manage Holodeck is a software solution and can be installed on compute nodes no matter where they are, enabling optimized data access across clouds while limiting egress fees and lock-in risks.
- **NAS acceleration:** Thanks to its caching technology, IC Manage Holodeck can be used to accelerate NAS or object storage systems. Hot data is moved to the shared cache in the compute cluster, minimizing IO activity on the NAS system. The user can take advantage of existing infrastructure, separate performance from capacity for better system tuning, and avoid expensive hardware refreshes and data migrations.
- **Edge:** Many organizations are investing heavily in the cloud and consolidating data there, but they often need to access it from edge locations. Limiting egress fees while keeping up speed of access is paramount in this scenario, especially when bandwidth is limited. IC Manage Holodeck optimizes communication between the cloud and the edge, limiting unnecessary traffic and moving only hot data to the remote location. This has a huge impact on bandwidth optimization, performance, and cost.

From a business perspective, IC Manage Holodeck provides a quick return on investment (ROI) and has a positive impact on infrastructure TCO:

- **Ease of deployment and use:** Holodeck is a scale-out software solution that is easy to install, configure, and automate. It can be installed in minutes on hundreds of nodes and is transparent to applications and end users. There is no learning curve and applications work seamlessly, just faster.
- **Cost-effective:** Holodeck leverages flash memory resources in the compute nodes. Each node of the cluster contributes to add capacity and performance while optimizing access to the storage system. Investment is limited and the availability of software subscriptions helps customers to choose between CapEx and OpEx models.
- **Investment protection:** IC Manage Holodeck enables users to improve performance of existing storage systems and extend their life, without paying for expensive upgrades over time. By separating capacity from performance, the user can get better infrastructure flexibility and focus on

the best $/GB for the storage system, while getting $/IOPS from Holodeck when necessary.

- **Added flexibility and agility:** Thanks to Holodeck, every organization can react quickly to emerging business needs. The separation of performance from capacity, and the optimization of all communication between the two layers, enable IT managers to transform their infrastructure quickly and make it ready to support any business initiative, no matter whether it involves big data analytics, a new AI project, or HPC workloads for accelerating R&D.

4. The Market Landscape

IC Manage Holodeck is not alone in the market. Other vendors, including cloud providers and storage manufactures, have similar solutions, but Holodeck offers unique characteristics that minimize the risk of lock-in while providing better scalability and performance. Above all, Holodeck is vendor-agnostic and works with both file and object storage, providing more freedom of choice around configuration and deployment layouts.

Table 1. Solutions Compared

	IC Manage	DDN IME	NetApp FlexCache	MS Azure Avere	MS Azure HPC Cache
Scale-Out	YES	YES	N/A	YES	N/A - not really scalable
R/W	YES	YES	NO	YES	YES
No Lock-in	YES	NO	NO	NO	NO
File+Object	YES	NO	NO	YES	YES
Flexibility (on-prem + hybrid + cloud)	+++	+	++	++	+
Cost	$$	$$$	$	$$$	$$

Holodeck answers the challenges posed by complex, hybrid, and evolving infrastructures:

- It provides flexibility and simplicity while keeping costs down.
- It is a software solution that doesn't need specific hardware to work.
- It can be deployed on-premises and in the cloud.
- It can be configured to work with both pass-through and write-back policies to further optimize IO activity.
- It supports files and object backends.
- It is cost-effective and less expensive than dedicated hardware solutions, allowing the user to optimize the investment as well.

When compared to other solutions on the market, IC Manage is the only one that combines all these characteristics, providing the best and quickest ROI to the user.

5. Conclusion

Holodeck is a unique product that can accelerate several types of highly demanding workloads across a range of deployment scenarios. Its unique characteristics enable users to protect their investments in existing infrastructures while pursuing a hybrid or multi-cloud strategy.

The IC Manage solution shows impressive improvements in infrastructure efficiency, enabling users to do more with less and providing immediate ROI. Primary use cases include workload acceleration for big data and HPC applications including AI, EDA, analytics, genomics, video, and image rendering, and more, across cloud, hybrid cloud, and on-premises deployments. But it is also worth noting that there are several secondary use cases for this caching technology, including edge, work from home, and remote office/branch office scenarios in which users need to access data quickly and efficiently.

6. About Enrico Signoretti

Enrico has 25+ years of industry experience in technical product strategy and management roles. He has advised mid-market and large enterprises across numerous industries and software companies ranging from small ISVs to large providers.

Enrico is an internationally renowned visionary author, blogger, and speaker on the topic of data storage. He has tracked the changes in the storage industry as a Gigaom Research Analyst, Independent Analyst and contributor to the Register.

7. About GigaOm

GigaOm provides technical, operational, and business advice for IT's strategic digital enterprise and business initiatives. Enterprise business leaders, CIOs, and technology organizations partner with GigaOm for practical, actionable, strategic, and visionary advice for modernizing and transforming their business. GigaOm's advice empowers enterprises to successfully compete in an increasingly complicated business atmosphere that requires a solid understanding of constantly changing customer demands.

GigaOm works directly with enterprises both inside and outside of the IT organization to apply proven research and methodologies designed to avoid pitfalls and roadblocks while balancing risk and innovation. Research methodologies include but are not limited to adoption and benchmarking surveys, use cases, interviews, ROI/TCO, market landscapes, strategic trends, and technical benchmarks. Our analysts possess 20+ years of experience advising a spectrum of clients from early adopters to mainstream enterprises.

GigaOm's perspective is that of the unbiased enterprise practitioner. Through this perspective, GigaOm connects with engaged and loyal subscribers on a deep and meaningful level.

8. Copyright