# FULFILLING THE PROMISE OF INFRASTRUCTURE CONSOLIDATION

**By**

**Jon Toigo**
**Chairman, Data Management Institute**
**Managing Principal, Toigo Partners International**

## SUMMARY

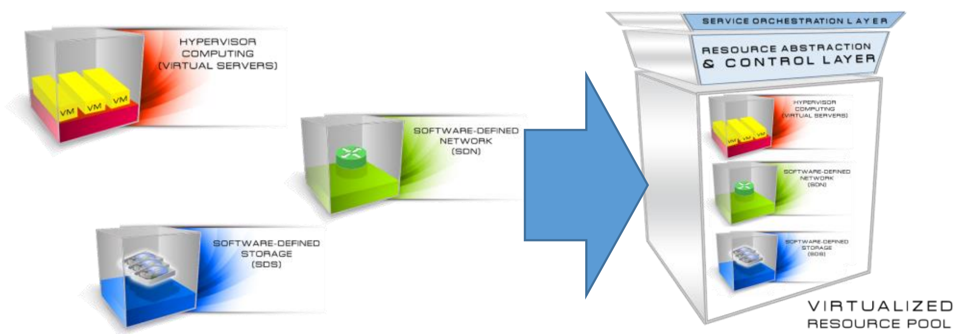
When hypervisor-based computing was pressed forward as application hosting option in the early 2000s, a large part of the value proposition had to do with the benefits of server consolidation that would be realized from server virtualization. Leading analysts claimed that the consolidation value would be so significant that consumers would see a sharp decline in CAPEX spending that would more than compensate for the cost of hypervisor software and services. Plus, with the reduction in physical hardware kit and the resulting reduction in floor space and environmental costs, IT shops would also realize OPEX advantages: fewer servers translating to fewer server administrators, smaller IT staff and reduced labor costs.

To many firms, this was a compelling value case. Only, it didn't materialize as planned. One obstacle to realizing consolidation goals has been the perpetuation of legacy storage arrays and storage area networks (SANs). With the continuing evolution of technologies for deploying converged and hyper-converged infrastructure (HCI), this obstacle might be surmounted in many cases, enabling "hyper-consolidation" along the way. That's where DataCore Software technologies can play a big role.

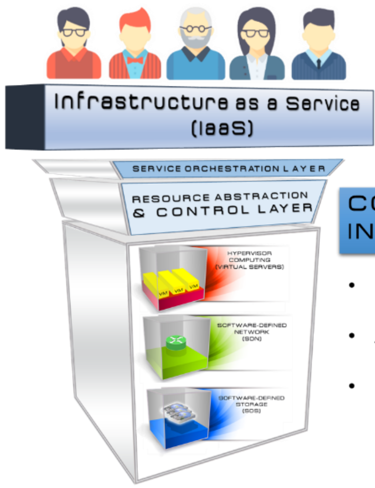
## INTRODUCTION

Hypervisor-based computing was introduced in the early 2000s as a method for virtualizing applications and operating systems and creating them as virtual machines that could be hosted more efficiently on fewer server hosts and move fairly effortlessly from physical server to physical server as a means of load balancing and agile instantiation.

This "software-defined server" technology was later joined by technology initiatives aimed at delivering software-defined networking (SDN) and software-defined storage (SDS). Together, it was hoped that the result would be the achievement of maximum consolidation and maximum deployment agility that would, in turn, usher in an "Infrastructure as a Service" architecture suitable to cloud providers and large data centers.




Essentially, the combination of virtual servers, SDN resources, and SDS assets would create a virtualized resource pool that could be allocated and deallocated rapidly to provide elastic infrastructure support for any workload that required it.




**CONCEPT:**
**INFRASTRUCTURE AS A SERVICE (IaaS)**

- Pooled virtual servers, software-defined networks and software-defined storage
- Automation and orchestration tools for agile deployment of resources to workload
- Lower CAPEX and OPEX costs

Adoption of server virtualization was brisk, with leading industry analysts serving as cheerleaders for the highly profitable server virtualization vendors. One optimistic report, issued in 2007, suggested that CAPEX spending on server hardware in data centers would be so significantly reduced by 2012 that virtual server computing would more than offset its own expense!

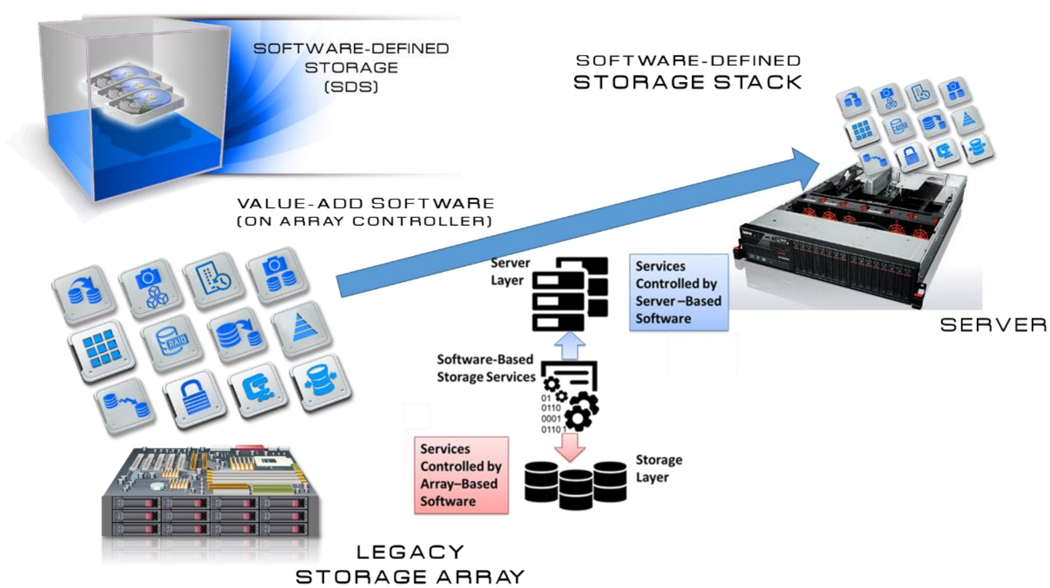
Only, it didn't quite work out that way. By 2013, spending on data center hardware had continued to accelerate. Part of the problem was that, after the "low hanging fruit" of web servers and file servers had been virtualized and consolidated onto fewer machines, more demanding applications began to challenge the model. Servers had only a fixed complement of resources to share in an multi-tenancy arrangement – even with technologies such as hyper-threading that were intended to support multiple virtual machines (VMs) on a multicore processor of a single host system. Simply put, the more resources required by a VM, the fewer the number of VMs that a single server could support.

Resource constraints went beyond the speed of the CPU and speed and quantity of memory that could be shoe-horned into a server chassis. The inability of the multicore chip to support the efficient unloading of I/O from multiple, concurrently running, VMs became a log jam. Each VM, nested in its own logical core, had to wait its turn to have its I/O request handled – analogous to many customers waiting in a queue to get the next taxi. This was owed to the perpetuation of sequential I/O processing methods, inherited from single-core ("unicore") CPU-based computers of the prior 30 years. Slow, sequential, I/O handling created latency that showed up in the form of poor application performance to end users, casting aspersions on the hypervisor computing model generally.

Hypervisor vendors attributed application performance problems to resource oversubscription (using servers that had too few resources to share) and to storage I/O chokepoints supposedly created by "legacy storage" – that is, shared storage topologies like SANs. The latter explanation was debatable, but the former led to many recommendations and best practices that restricted the number of VMs that should be hosted concurrently on servers. This had the impact of sharply reducing the number of virtual machines that could be hosted on a server, and by extension, the consolidation value of server virtualization.

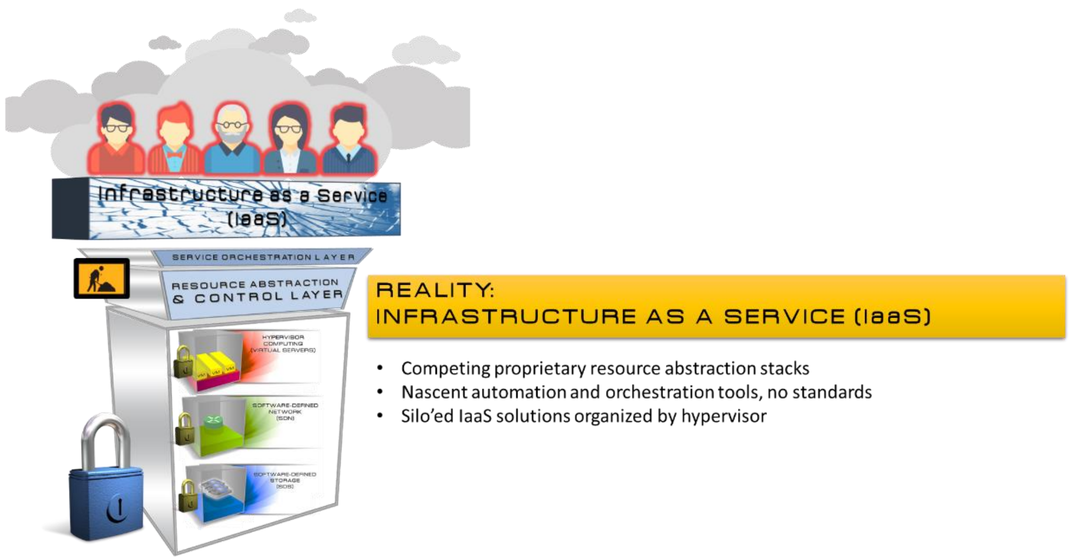
Whether or not legacy storage itself actually played a role in VM performance shortcomings, it was quickly targeted for replacement using "simpler" deployment topologies. Software-defined storage gained credibility as a solution to the complexity, cost and efficiency of storage infrastructure.

Software-defined storage (SDS) was originally described as the abstraction of storage services away from array controllers and into a software stack residing on the server host. SDS was not a new concept. It had been offered as early as 1993 in the IBM mainframe community in the form of System Managed Storage, and by Digital Equipment Corporation a few years later on several of its operating systems. In 1999, DataCore Software offered its SANsymphony technology as a means to virtualize heterogeneous storage hardware, regardless of kit brand or interconnecting plumbing, to create a virtual storage pool that could be allocated and deallocated at will – with associated capacity management, data security and data protection services.




On its face, the software-defined storage advocates were attempting to centralize storage services and resources for more efficient management and distribution to workload. The ability to accomplish this task was hampered in heterogeneous (multi-vendor) storage infrastructure by proprietary obstacles in many storage arrays that limited their management in common. So, SDS had long been a source of debate in the industry, mainly between independent software developers like DataCore Software, and hardware vendors who had little interest in working and playing well with their competitors.

When hypervisor vendors began calling storage service abstraction concept “software-defined storage” they were simply putting a new tag on an old concept.  The idea of hypervisor-controlled SDS was to provide a uniform “soup to nuts” infrastructure model that would ensure the full compatibility of, and greatest efficiency from, the entire hardware/software stack. Coincidentally, the idea created an opportunity for hypervisor vendors to lock in their customers and lock out their competition, so it was not that much unlike the rationale of the monolithic array and SAN hardware vendors.




The net result of the proliferation of proprietary, hypervisor-centric, virtualization and software-defined technology stacks was that the Infrastructure as a Service model, with all of its promised consolidation value, became fractured.

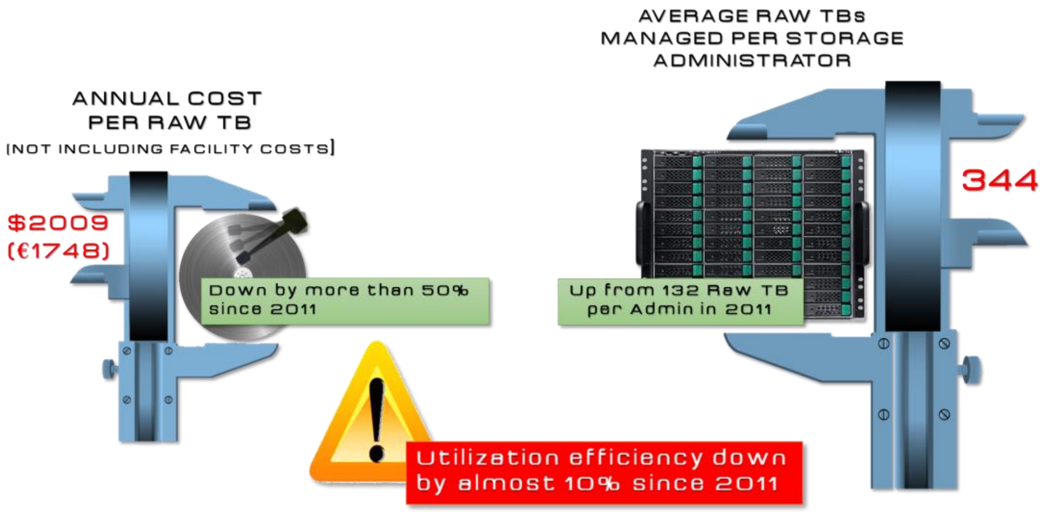
New and proprietary hardware kit proliferated, representing itself as “converged” infrastructure because it could be deployed behind a “brand X” virtual server host to serve up storage capacity to virtual machines from that hypervisor on a dedicated basis (supposedly in a manner superior to shared storage).  In most respects, these products represented an arrangement between the hypervisor vendor and the hardware vendor to split functionality between the hardware controller and the SDS software stack in the hypervisor, resulting in a lock-in system.

Later, the concept of a hyper-converged infrastructure was introduced – essentially establishing a model in which all compute, network and storage would be delivered by a single rack-mountable appliance acting as an HCI cluster node. Most hypervisor-centric HCI solutions required a minimum number of storage nodes -- usually three -- each with identical hardware (that needed to be obtained from a certified hardware list) and its own software license. This had the effect of increasing the cost of the nodal hardware significantly and of pushing storage capacity demand upwards by as much as 300% in heavily virtualized server environments, according to some analysts.

Moreover, in shops where more than one brand of hypervisor was being used to support workload, customers discovered that technology silos had been created behind each type of hypervisor that prevented resource sharing. Silo-ing system resources behind hypervisors compromised infrastructure consolidation goals since high availability platforms (essentially, server clusters with VM template cut and paste) needed to be designed for exclusive use of a single hypervisor vendor's workload.

Also, as in earlier days of monolithic and proprietary storage hardware, "identicality" became an issue again. Data from workload created under hypervisor A could only be stored on hypervisor A's technology stack, and mirrored or replicated to a second instantiation of identical software and hardware. Because of the proprietary nature of early hypervisor-centric SDS, the data of different workloads virtualized with different hypervisors could not share available storage capacity. Redundant capacity needed to be deployed behind different silos to accommodate all the data.

An early indication of the inefficiencies introduced by hypervisor-centric converged and hyper-converged infrastructure came in a comparative study by one industry analyst of the impact of segregating storage behind virtual servers and their hypervisors. At first glance, the numbers appeared to support to arguments of the hypervisor vendors. The cost to manage a terabyte of storage, not including facility costs, had dropped to about $2009 – down more than 50% from the cost in 2011, without the advent of converged and hyper-converged. Moreover, using tools provided by the hypervisor vendor as part of SDS stack administration, the number of TBs of storage that a single administrator could manage in a year grew from 212 in 2011 to 344 by 2016.[i]




The “gotcha” in the data, however, proved to be a bit more elusive to the analysts. Looking across the entire enterprise, silo’ing storage resources behind hypervisors and virtual servers had led to a net reduction of almost 10 percent in storage utilization efficiency. This reflects the inability to share capacity and services across isolated storage silos in order to achieve greater efficiencies and economies of scale. This metric is worth watching in the future as it is the same problem that the industry sought to address with greater resource sharing topologies (SAN and NAS) in the previous decade.

Bottom line: the industry is currently debating the merits of discrete (aka legacy) storage, converged storage, and hyper-converged infrastructure models as an approach that will realize the goal of infrastructure consolidation that led to the widespread adoption of virtual server technology. Thus far, innovations have been seized upon by hypervisor software vendors and, in some cases, by storage hardware vendors to bolster other proprietary objectives. Needed is a sensible way to address the challenges introduced by workload virtualization that will optimize application performance while containing expanding IT costs.

## THE SOLUTION

DataCore Software has long offered unique solutions to many of the challenges cited above. Not only did its “SDS software” (called “storage virtualization”) provide a shared platform for the data from workloads virtualized by any hypervisor (as well as non-virtualized workload data), it also facilitated

- The inclusion of existing storage technology investments (so-called legacy storage arrays and SANs) into emerging SDS infrastructure
- Support for any and all storage-to-server interconnect topologies and automated traffic balancing between interconnects
- An adaptive DRAM caching technology to ameliorate any storage I/O latency
- Adaptive parallel I/O technology to alleviate latencies from raw I/O request handling

The next evolution of DataCore’s SDS is seeing improvements in the simplification of its deployment as part of a converged infrastructure or hyper-converged infrastructure solution.

Converged infrastructure, which involves the mapping storage resources to specific Virtual Machines or to database workloads, is very easy to facilitate with DataCore as it has always been part of the company’s core SANsymphony technology.

The DataCore SANsymphony Software-Defined Storage Platform enables the creation of a converged server SAN that can be easily connected to any virtual server host. DataCore’s SANsymphony Software-Defined Storage Platform provides

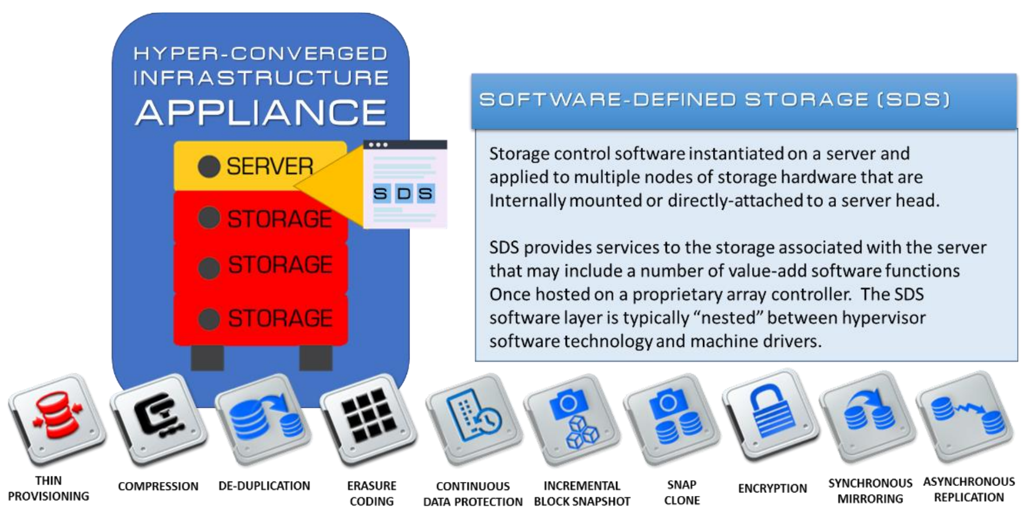
- A minimum 2 node (rather than 3 node) storage cluster, without component identicality requirements
- Simple assignment of service-rich volumes to workload
- Active-active clustering between Converged Server SAN instances
- Integrated management with DataCore converged and hyper-converged infrastructure platforms as well as DataCore Virtual External SANs, deployed across legacy infrastructure

As with any converged infrastructure, DataCore’s solution dedicates resources to specific workload. Moreover, dedicated volumes are equipped with a user-defined set of data protection and capacity management services taken from DataCore’s robust SDS stack, so the appropriate “tier” of storage can be delivered to exactly the right data and workload all the time. And scaling storage volumes is a snap with DataCore: growing a volume simply requires adding more capacity to storage nodes or another chassis of storage to the DataCore cluster platform. The difference between doing converged storage with DataCore, of course, is that you need not be locked into a proprietary hardware vendor when you want to add nodal hardware.

Hosting databases, whether virtualized or not, can also benefit from convergence. With databases increasingly going "in-memory" – being run entirely from memory components without the latency from making data inquiries via disk accesses – technology is needed to optimize memory reads and writes. DataCore has already demonstrated a new software technology, Adaptive Parallel I/O, which can work as part of its SDS stack to achieve mind-bending raw I/O speeds from today's multi-core server CPUs. Adding this capability to DataCore's already impressive converged infrastructure solution and DataCore becomes the obvious candidate for high performance and in-memory database workloads: extraordinary performance using the gear that a company already owns.

When it comes to hyper-converged infrastructure, DataCore also provides the technology. HCI is finding increased popularity in small office environments, or in remote offices, or behind specific applications such as Virtual Desktop. However, as the technology matures, one can envision scenarios in which HCI appliances become high performance building blocks for highly scalable and highly agile infrastructure.

Essentially, HCI is a productization of SDS in which server and storage hardware are pre-integrated with an SDS stack to deliver a unified or "hyper-converged" appliance. DataCore has ported its software offerings to HCI and is currently working with many brand name server manufacturers to deliver readily-deployed appliances to market.




A basic HCI appliance has the footprint of a rack-mountable server. The basic kit uses internal storage or a server head with a JBOD array for disk or SSD or both. With current generation multi-terabyte HDDs and SSDs, considerable capacity can be supported in a “pizza box” sized unit.

While current deployments of HCI technology may be limited to small and remote business offices, where their limited footprint and remote management capabilities make them ideal for basic workloads, the future is looking bright for HCI in the enterprise data center. One reason is DataCore’s Adaptive Parallel I/O software that can make a DataCore Hyper-Converged Virtual SAN deliver application performance metrics rivaling those of very expensive supercomputing infrastructures – using the server CPU you already have!

DataCore’s stellar Storage Performance Council™ benchmarks in 2015 and 2016 were operated from an HCI appliance comprising a 2u rack-mount Lenovo server with internal SAS disk and Samsung SSDs. The company has just achieved extraordinary SPC-1™ IOPS™ measurements that place it in the top three performers in the market, but at a fraction of the cost of the competitors on a per IOPS basis. The result is an industry-leading high performance HCI appliance that features a low cost, low power consumption, and low floor space footprint enabling significant physical equipment consolidation.

While it is tempting to suggest that features such as parallel I/O enable greater VM density, the number of VMs that can be hosted are still gated by the CPU and memory. With DataCore converged and hyper-converged SAN technology, consumers can optimize the number of VMs that an individual appliance can host so that promised VM densities are more achievable. This is a more sensible claim than competitors who promise a particular VM count absent any reference to the type of workload that is being virtualized.

DataCore-based HCI appliances can plug and play with DataCore-virtualized legacy infrastructure, if desired. Plus, DataCore HCI appliances will support not only server hypervisors, but also containers and in-memory databases, and they can be managed in common regardless of their hypervisor or workload.

## CONCLUSION

The promised infrastructure consolidation narrative of early hypervisor software evangelists failed to take into account the proprietary nature of hypervisor-controlled hardware/software stacks or the limiting effects of sequential I/O execution and silo'ed hardware resources. To address these limiters and to derive "hyper-consolidation" from hyper-convergence, IT planners need to consider SDS solutions from DataCore Software.

With DataCore Software, not only servers, but also storage infrastructure, can be consolidated in a way that realizes peak optimization levels. For smaller firms, the flexible hardware and workload support of DataCore's SDS platform or hyper-converged infrastructure appliances can provide a top notch data center in a very small footprint. For the largest companies and cloud service providers, DataCore SDS and DataCore-based HCI platforms can become "atomic units" of compute that can be deployed quickly to meet changing workload demands and managed centrally to reduce administrative costs.

DataCore Software technology was among the first in what is now called software-defined storage. It is ten generations ahead of the preponderance of SDS software that has been delivered to market to date and features an SDS stack component that none of its competitors offer: adaptive parallel I/O that can make all the difference in reaching latency reduction and application performance goals even with a fully populated virtual server platform.

Worth a close look.

---

[i] Gartner, IT Key Metrics Data 2016: Key Infrastructure
Measures: Storage Analysis: Multiyear Published: 14 December 2015