Breaking Out of the Box: How Composable Infrastructure Could Underpin Digital Transformation
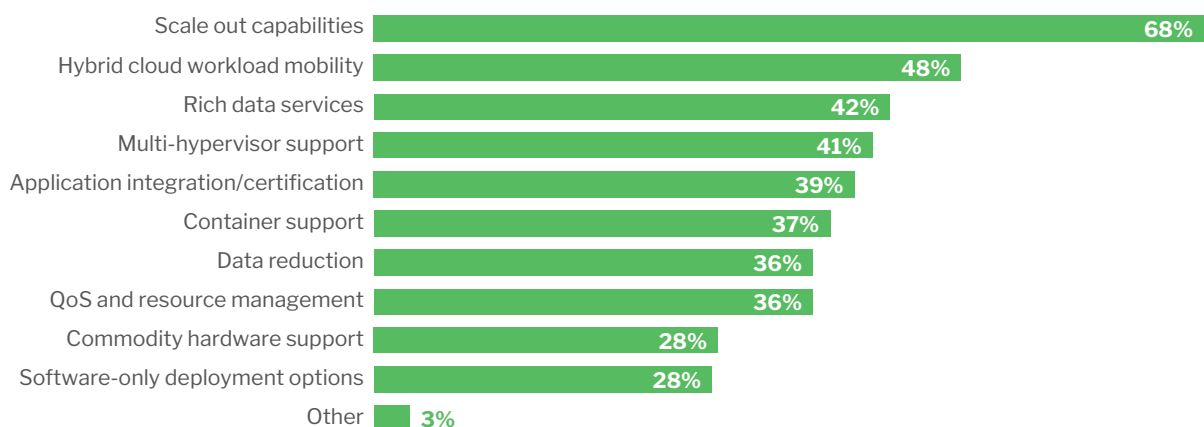
The 451 Take

Whatever you call it – disaggregation, composable infrastructure, software-defined – the concepts are clear enough. New innovations must be applied to system infrastructure to eliminate the rigid, hardware-bound silos that have held companies back from their digital transformation strategies. Organizations need better operational agility through automation and flexible resource scaling if they are to navigate the megatrends that now dominate IT: the changing demands of emerging workloads and the ability to run them wherever it makes the most sense – on-premises, at the edge or in one or more multiple cloud options. New processors, memory, storage and networking technologies are being combined to build the next generation of infrastructure. Access to a fully modernized and optimized foundation of systems technology has never been more important.

Over the past decade, we've seen an evolution of enterprise systems delivery from complex three-tier architectures to pre-integrated packages of server, storage and networking (converged infrastructure) to modular building blocks of server and storage accessed at the hypervisor layer (hyperconverged infrastructure, or HCI). But the rigid hardware architecture underlying these systems has remained largely unchanged over the last 20+ years, making it difficult to adapt to the demands of modern workloads. A more flexible and comprehensive approach to disaggregation is required, something potentially offered by composable infrastructure. Pools of resources (compute, storage, GPU, FPGA) are interconnected over high-speed fabrics, ready to be dynamically 'composed' into bare-metal servers in seconds, without the restraints of physical boxes. Our survey of HCI users identified the most important feature of HCI as scale-out capability – enabling them to add compute and storage capacity without downtime, or to scale resources independently. Other key requirements include extending hybrid multicloud capabilities and meshing platforms with public cloud services, as well as multi-hypervisor support – the latter in a market where several major platforms available today support just one.

Technology Features Important when Selecting HCI Platform

Source: 451 Research's Voice of the Enterprise: Computer Infrastructure, Budgets & Outlook

Q: What technology features are important when selecting a hyperconverged infrastructure platform? (Please select all that apply.)

Feature	Percentage
Scale out capabilities	68%
Hybrid cloud workload mobility	48%
Rich data services	42%
Multi-hypervisor support	41%
Application integration/certification	39%
Container support	37%
Data reduction	36%
QoS and resource management	36%
Commodity hardware support	28%
Software-only deployment options	28%
Other	3%

With all the new layers of abstraction – virtualization, containers and cloud – it's easy to forget that underneath it all, the physical infrastructure still matters. 451 Research has defined invisible infrastructure as 'the evolution of IT and communications infrastructure to meet the demands of modern, digital organizations, enabling technology consumers to assemble, access and pay for digital services in a simple, seamless and automated manner, without specific knowledge of the underlying physical infrastructure.' But to get there, the physical infrastructure underneath must be more specific, more adaptable, more accessible, more scalable and more efficient to both run and manage. The era of 'good enough' general-purpose system units is passing.

The 451 Take (continued)

Customers won't make significant changes to their infrastructure without solid business reasons to justify the time, expense and risk. The foremost of these is the woeful average utilization rate of systems infrastructure. Typically, current resources are being run at about 30% of capacity – this is an anecdotal estimate, but in our view, it's a generous one. That represents a huge cost premium and a poor efficiency rate. To bring the rates up, systems must be sized and balanced for optimum use, and overprovisioning must be carefully managed.

New workloads, very different from the monolithic 'heritage' apps with their traditional text, structured data and emphasis on transaction processing that still dominate enterprise computing, are now becoming mainstream. They are heavy with multimedia elements, rely on rapid access to large and disparate data sources, and (most importantly) are written in the form of micro-services, digging deeper into the software stack to access a common set of data and cloud services optimized for the infrastructure underneath.

That infrastructure might be located within an enterprise datacenter and/or in a private or public cloud. That 'and/or' is particularly important because it's now clear that some of those heritage applications, and even some cloud-native applications, will stay on-premises where the customer maintains control and continuity and can choose precisely what infrastructure resources are required. Specialist service providers and public clouds, often more than one, will increasingly be used for greenfield projects or to take advantage of a particularly advantageous mix of domain expertise, services, performance, cost and location.

We are looking toward a more fundamental and groundbreaking disrupter that will become the transformational face for most infrastructure environments. That role will be filled by platforms (likely composable) that disaggregate all resources across common fabrics and provide both generalized and specialized flexibility for any environment, regardless of size, scope or workload mix.

Business Impact

FLEXIBILITY OF PRIVATE CLOUDS CAN MATCH THE PUBLIC CLOUD. One of the hallmarks of public cloud is flexible resource consumption; private clouds built on disaggregated technology will be better positioned to mimic that flexibility on-premises, without the restriction of scaling storage and compute in lockstep.

ADAPTING TO NEXT-GENERATION WORKLOADS. An emerging set of data-intensive applications, including AI and edge computing, will increasingly rely on automation if they are to run efficiently while hiding some of the complexity. The current generation of static hardware can be limiting for these workloads without massive overprovisioning.

OPEN ARCHITECTURES MEAN MORE RAPID ACCESS TO THE LATEST TECHNOLOGY. Composable infrastructure should support a mixture of bare-metal, virtualized and cloud-native containerized resources. And fabric independence enables a path to the emerging generation of interconnects capable of connecting and pooling high-bandwidth memory and accelerators from multiple sources.

Looking Ahead

We are heading toward full disaggregated infrastructure, where pools of resources become available as dynamically 'composable' resources, assembled to match the demands of application workloads in real time, and then reassembled for the next set of demands. The system platforms underneath will be the scale-out clusters of commodity servers, storage, accelerators and networking already in use at the largest hyperscalers, with most of the specialist features not included, therefore transferring much of the intelligence and optimization to the software layer. From there, specialist resources can rejoin the mix – smart network interface cards and remote direct memory access to handle the huge increase in server-to-server east-west traffic, critical resources communicating over high-speed fabrics (NVME, GPU, FPGA and storage-class memory) to run demanding transaction processing and I/O-hungry applications at maximum performance, as well as custom accelerators for more efficient processing of security, analytics, imaging, deep learning and AI.