Composable Data Infrastructure In The Digitally Transformed Enterprise

Situation Analysis

Digital transformation is about leveraging technology to respond to the needs of the market faster. The result could be realized through operational efficiencies, faster time to market with new products, or simply responding to customers' needs more quickly.

Leveraging technology, however, is about more than just deploying the latest compute, network, and storage infrastructure. It's about provisioning these resources in real time to deliver optimal performance in a resilient and secure fashion.

Many organizations have found composable infrastructure – the abstraction of compute, storage, and networking managed through a centralized control plane – to be the best method of deploying and managing infrastructure for real-time consumption by DevOps engineers embedded in business units.

Composability is the core technology paradigm that allows the largest cloud providers to service customers at scale. It is what enables the "select-and-go" model for users, delivering the ideal platform with the requisite performance, resilience, and security requirements.

While composable infrastructure has greatly enabled the modern organization to drive toward improved speed and agility through the cloud (and, to a lesser extent, on-premises), a new class of composability – data infrastructure composability – aims to drive even greater performance and resilience.

This research brief will explore this new class of composability and how companies like Volumez use it to deliver unheard-of application performance levels.

A Short History of Composable Infrastructure

Moor Insights & Strategy (MI&S) defines composable infrastructure as the disaggregation of computing, networking, and storage, where each resource is dynamically allocated to support the needs of applications and workloads.

Composable infrastructure can mean different things to different organizations. For some, it's the fully integrated hardware and software stack that IT solutions vendors such as Hewlett Packard Enterprise (Synergy), Dell (PowerOne), Cisco (UCS), and others deliver.

For others, composable infrastructure refers to software-defined infrastructure (SDI) and software-defined datacenter (SDDC) solutions. VMware's mainstreaming of virtualization drove the concept of allocating resources automatically. This led to hyperscalers and cloud providers leveraging tools in the open-source community to perform orchestration at scale.

Open-source tools such as Ansible, Puppet, Chef, Terraform, and Ceph became popular for enabling enterprise organizations to bring this cloud-like capability into the datacenter. And a strong argument can be made that open-source orchestration tool Kubernetes truly brought composable infrastructure into the mainstream. These tools and open-source projects established what is today a vibrant market.

While the notion of disaggregating computing, networking, and storage resources can be traced back to the IBM mainframe, the more modern adoption of composable infrastructure is rooted in cloud computing, where real-time orchestration of resources is critical to the success of every major cloud provider.

Why Is Composable Infrastructure Critical?

MI&S sees composable infrastructure filling three critical needs. The first is driving operational velocity in the business. The ability of embedded DevOps teams to quickly and easily allocate infrastructure resources accelerates application development considerably. And this, in turn, allows levels of speed *and* agility not realized through traditional methods.

The second reason composable infrastructure is critical relates to the innovation cycles of hardware and software. Software evolves faster and more frequently than hardware. In this regard, the decoupling of resources introduced by composability removes dependencies on underlying platforms.

Additionally, composable infrastructure enables organizations to take advantage of hardware innovations that take place faster than traditional IT lifecycle refresh rates. Perhaps no better example of this is the performance of PCIe/NVMe flash doubling every 18 months, or the release of graphics processing units (GPUs) for workload

acceleration that is out of phase with an organization's refresh cycle. Through composability, these resources can be made immediately available for use by new and existing applications across the enterprise.

Finally, composable infrastructure enables organizations to fully leverage the power of high-powered (and high-priced) top-bin central processing unit (CPU) and GPU resources. By increasing the utilization of these resources, businesses can realize a higher return on Capital Expenditure (CapEx) and Operational Expenditure (OpEx) investments.

COMPOSABLE DATA INFRASTRUCTURE – COMPLETING THE SOLUTION

Composable infrastructure is a key element of the cloud operating model and a must-have for any IT organization supporting the digitally transformed business. However, while it has delivered several tangible benefits to organizations, MI&S sees areas where organizations continue to struggle:

- ***Enterprise-grade data services with predictable performance****:* Performance and resiliency are critical for the modern business. This is especially true for workloads running in the cloud, where data service features typically lag the capabilities of high-end on-premises enterprise solutions. Additionally, organizations have had to account for variations in latency, performance, and resilience, and a lack of predictability across cloud providers. Having to integrate with multiple cloud provider application programming interfaces (APIs) also adds a layer of platform engineering overhead that inhibits multi-cloud strategies and lays the ground for cloud vendor lock-in.

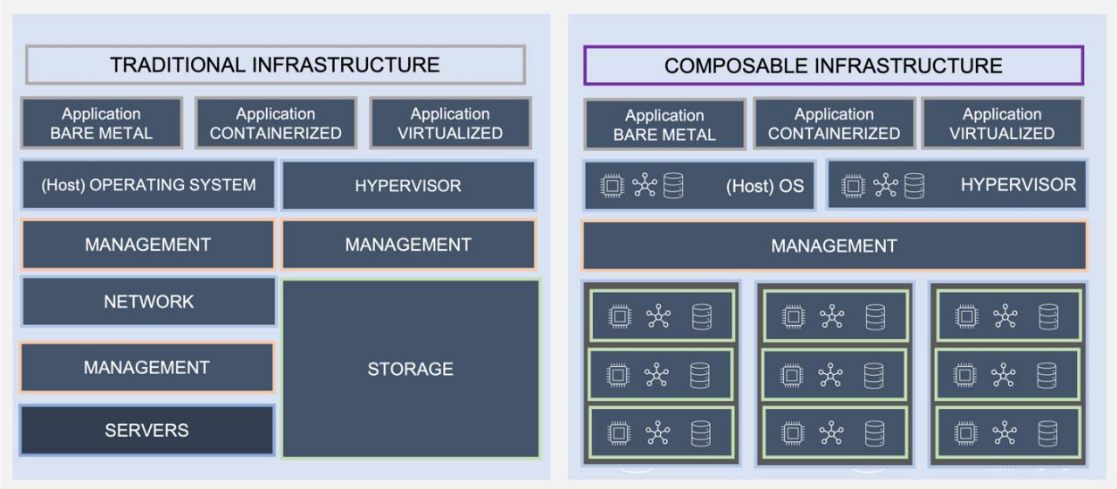
- ***The separation of the control plane and the data plane****:* Storage control and data planes are historically tightly coupled, impacting platform scalability and extensibility. Tight coupling hinders rapid innovation across hardware and software, impacting the ability to continuously deliver new features and capabilities.

  For example, Kubernetes provides that strict separation of data and control planes, resulting in scale, resilience, and the ability to deliver continuous updates. This same abstraction that exists for compute and networking is required in the storage infrastructure layer.

- ***Specialized resources****:* As budgets continue to flatten and even shrink, investing in the resources – technology and people – to deliver enterprise data services is increasingly difficult. Specialized storage engineers managing bespoke solutions are a significant expense and can create vendor lock-in.

These challenges have given life to composable data infrastructure, a software layer that allows cloud teams to deliver predictable high-performance and enterprise-grade data services to users across the organization. Composable data infrastructure enables this capability without the need for specialized storage engineering skillsets. Instead, a declarative interface enables platform or DevOps engineers to define application input/output (I/O) performance and resilience requirements dynamically at runtime.

FIGURE 1: RACK-LEVEL COMPOSABLE INFRASTRUCTURE



Composability drives greater performance and greater resource utilization

Source: Moor Insights & Strategy

Composable data infrastructure should respond to a user making a declarative input of the desired state by dynamically and quickly constructing an application data path that provides the performance and resilience specified in the request. This contrasts with traditional infrastructure, which requires storage specialist knowledge and imperative programming of each step in the storage provisioning process.

LINUX IS THE DATA PLANE

A direct path from storage (data) to the application is novel. The storage market comprises highly performant solutions built on technology designed to reduce latency as much as possible through proprietary software running on dedicated storage controllers.

Solutions based on storage controllers have reached a dead end with no resolution for issues like overcoming metadata server scalability and improving efficiency of cluster locks. A transition to controller-less Linux-based storage architectures will drive business value in four dimensions:

- ***Scalability of Cost***: PCI/NVMe speeds are increasing exponentially – doubling approximately every 18 months – while CPU core speeds are only increasing linearly. Continuing to scale controller-based architectures to keep up with media, therefore, requires an exponentially increasing investment in controller hardware.

- ***Scalability of Performance***: Using a 100% Linux data path without storage controllers allows for a direct network path between the applications stacks and the media. This takes full advantage of the enormous input/output operations per second (IOPS) available in modern solid state drives (SSDs) and utilizes the full bandwidth and low latency over non-volatile memory (NVMe) between the media and the Linux kernel on application servers, thus providing unparalleled low latency, high bandwidth, and unlimited scalability.

- ***Features and Capabilities***: The modern Linux kernel natively includes primitives required to render enterprise-grade data service features – such as snapshots, thin provisioning, and quality-of-service (QoS) – with significantly more advanced capabilities than proprietary storage software running on dedicated storage controllers.

- ***Standardization and Quality***: Standardizing on the Linux kernel for the data path aligns enterprise storage with the same code base that is powering trillions of devices around the world. While enterprise storage vendors have extremely robust quality control systems, the Linux kernel is by far the most heavily scrutinized source code in the world. Replacing proprietary closed software with widely deployed open-source software reduces exposure to software bugs and unilaterally improves software quality.

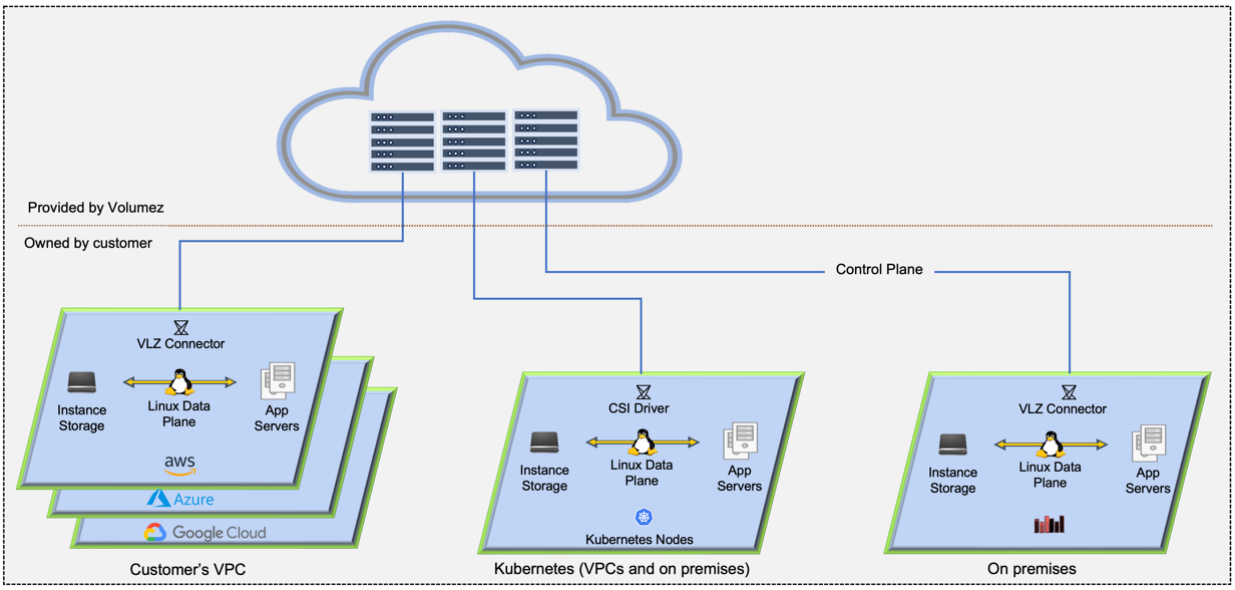
MI&S believes this paradigm shift in the storage data plane will usher in a new class of composable data infrastructure providers that enable data-driven organizations to improve agility, optimize the efficiency of their cloud operations and spend, and deliver a faster, more responsive application experience to their customers. Further, MI&S sees Volumez as the pioneer in this new data services space.

VOLUMEZ – A NEW DATA MANAGEMENT PARADIGM

Volumez, a SaaS-based cloud volume orchestrator for Kubernetes and virtual machine-based architectures, is the first mover in this composable data infrastructure space. It composes raw cloud storage media, such as Amazon EC2 Instance Store, Google Compute Local SSDs, and Azure Ephemeral Disks, into high-performance Linux-based data paths for stateful cloud-based applications.

Volumez is designed for DevOps and platform engineers who need to automate creation of application data paths with low latency, geographic resilience, and economic cost-performance, but may not have deep domain expertise in Linux-based distributed storage systems. Users simply define *declarative policies* that specify an application's performance and resilience requirements, and the Volumez Engine composes the data path.

FIGURE 2: THE VOLUMEZ ARCHITECTURE



Source: Moor Insights & Strategy

CONTROL PLANE

The Volumez control plane provides the user interface, API server, monitoring, observability, provisioning, and data management functions. Users register application servers with the control plane by deploying the Volumez CSI driver on Kubernetes clusters or Volumez Connector Service on virtual machines.

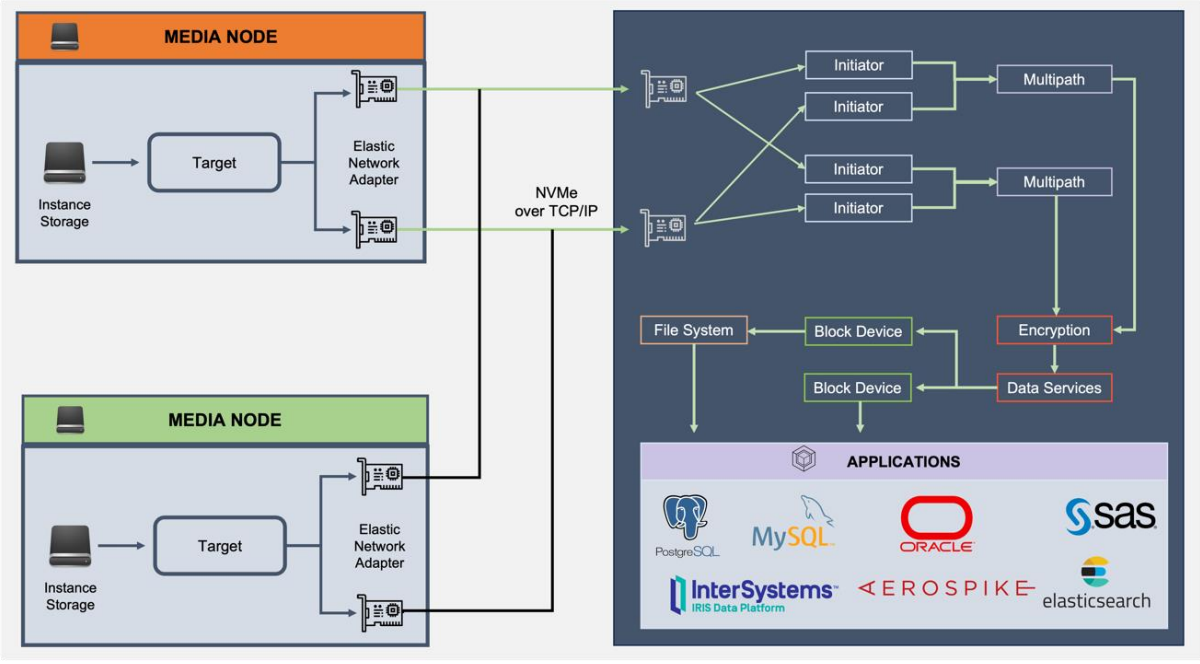
The Volumez Connector is a user space service that issues data path configuration commands and monitors the server data path for drift. The Connector maintains an encrypted connection to the Volumez control plane.

In Kubernetes environments, the Volumez CSI driver enables management of persistent volumes through the Kubernetes API and tools such as kubectl. Users need only specify a policy name in the Persistent Volume Claim, and Volumez does the rest automatically.

Data Plane

The Volumez data plane is composed of raw NVMe instance storage media and off-the-shelf Linux instances, both provided by every major cloud provider. Unlike scale-out software-defined storage (SDS), Volumez does not inject proprietary drivers into the data path. A typical Volumez data path on Linux, shown below, is composed of industry-standard components such as LVM2, Linux-RAID, dm-thin, and dm-crypt. Volumez uses encrypted NVMe-over-TCP or iSCSI for communication between nodes and to maintain resilience across availability zones.

Figure 3: Volumez Removes The Storage Controller



The Volumez architecture utilizes the Linux data path, removing latency

Source: MI&S from Volumez source material

The Volumez data plane is unique in the industry for its lack of storage controllers – clusters of dedicated servers that provide functions such as RAID, thin provisioning, and snapshots but add cost, latency, and I/O bottlenecks that limit performance and scalability. Volumez moves these functions to the Linux kernel on the application server, dramatically improving I/O performance as data services are rendered by the same CPU cores as the application while the data sections are in CPU cache.

The design focus of Volumez seems to be centered on five principles:

1. ***Predictability*** – As organizations move workloads from traditional IT environments to the cloud, they typically trade increased agility for a loss of predictability, both in application performance and operating cost. Arguably, the most strategic capability of Volumez is bringing uniform guaranteed I/O performance and total elimination of the noisy neighbor problem with predictable data storage costs across every cloud provider.

2. ***High Performance*** – Volumez claims to deliver 1.5M IOPS and 12GB/s of throughput per volume, guaranteed at latency no greater than 500 uSec on any supported public cloud. This level of application I/O performance is simply not achievable by incumbent controller-based cloud storage technologies. This step-function reduction in I/O latency has two strategic effects: First, it improves CPU efficiency by reducing I/O wait time, which translates immediately into infrastructure cost savings. Second, it makes existing applications faster and more responsive without refactoring, which translates directly into improved customer experience.

3. ***Manageability*** – The Volumez UX is clearly designed for DevOps and platform engineering teams that do not want to think about underlying storage technologies and may not have deep storage I/O expertise. The two critical elements of composability – orchestration and observability – are prominent in Volumez user workflow. Users simply define the required capacity, IOPS, bandwidth, latency, and zone resilience for an application and click "create." Within seconds, a file system attaches to the application pod or server. From an observability perspective, the Volumez interface is a very clean console that allows engineers to easily monitor real-time and historical I/O performance across the datacenter, from very granular application performance to performance in aggregate. Similarly, data services such as fast snapshot and restore are continuously available and activated with an API call or click in the UI.

4. ***Scalability*** – As a cloud-native architecture, Volumez counts scalability as a central feature of its design. Each application server manages its own data services and maintains a direct connection to its raw NVMe instance storage. By eliminating storage controllers from the data path, there is no shared state between applications and infinite, razor-flat scalability from single servers to planetary-scale distributed applications.

5. ***Resilience*** – The cloud is a dynamic environment where infrastructure issues from partial degraded performance of a server or NVMe media to full zone failures must be anticipated and accounted for in well-architected applications. Volumez abstracts the complexity of cloud data resilience by simply asking DevOps to specify the number of simultaneous media and zone failures against which an application must be resilient. Volumez's control plane algorithm automatically creates a data placement and network connection strategy that guarantees resilience of application data. After provisioning, the control plane continuously monitors the infrastructure for changes and failures and automatically relocates data copies as needed to maintain resilience.

While MI&S did not independently run tests to verify the performance numbers, we did witness the declarative interface for performance and resilience requirements, mounting of cloud volumes, and the performance of Volumez with small block random read and write operations. The ease of use was simply astonishing, and the performance was confirmed.

VOLUMEZ IN THE DIGITALLY TRANSFORMED BUSINESS

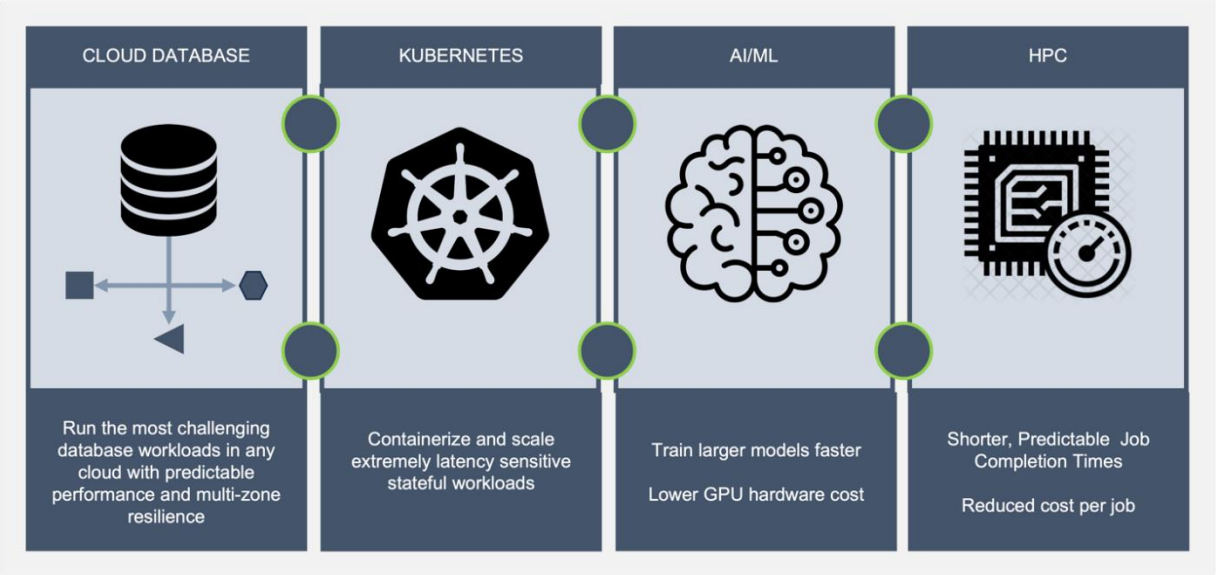
New technology without real-world benefits means nothing, as any experienced IT professional understands. As a result, MI&S always looks for the return on investment (ROI) for any technology solution it evaluates. And in the case of a new market maker like Volumez, we take a more skeptical view.

While companies have employed composable infrastructure to aid in digital transformation, the real benefits realized have been heavily indexed toward computing and networking. Volumez brings these composable benefits to data storage – predictability, performance, manageability, and scale across multiple clouds with point-and-click simplicity.

The ROI factor comes by way of infrastructure efficiency and application development velocity. MI&S believes organizations that employ composable data infrastructure such

as Volumez will realize a significant improvement in cloud resource utilization, especially expensive CPU and GPU resources, along with reduction in the operational resources required to develop, test, and deliver new features and applications to their customers.

FIGURE 4: VOLUMEZ WORKLOAD VALUE-ADD



Volumez delivers distinct value for modern workloads

Source: Moor Insights & Strategy

Finally, Volumez is easily consumable and scalable. MI&S witnessed Volumez being deployed on AWS and Azure in seconds and then scaling to workload I/O performance that exceeds incumbent cloud storage solutions. As the needs of an organization grow, scaling the data environment is simply adding more NVMe instance storage. With Linux itself as the data plane, there can be no vendor lock-in on any cloud at any scale.

Regarding workload affinity, MI&S believes Volumez delivers an authentic and impressive value for the workloads that drive the modern business, including AI/ML, cloud databases, high performance computing, and stateful Kubernetes. Each of these workloads is unique in its performance requirements, yet they all share the need for scalable, resilient, and predictably performant data services. These are all capabilities that MI&S has seen demonstrated in the Volumez solution.

SUMMARY

In the digitally transformed world, data drives business, and a new class of applications with unique and competing compute, network, and storage requirements have populated the datacenter.

While composable infrastructure has driven significant efficiencies in the application and utilization of compute and networking infrastructure, data infrastructure has historically been bottlenecked by controller-based scale-up and scale-out storage architectures that cannot keep up with the exponentially increasing performance of NVMe media.

By eliminating storage controllers and leveraging the Linux kernel itself as the storage data path, technology companies such as Volumez are set to revolutionize the cloud industry with solutions that vastly outperform incumbent technologies and deliver significant improvements in cloud operating efficiency.

There are a variety of compelling storage and data solutions in the market, each delivering incremental improvements in performance and efficiency. But Volumez, as the first mover in a new class of technology called composable data infrastructure, delivers a step-function improvement in infrastructure performance and efficiency by unlocking the inherent capabilities of storage media, networking, and compute in the cloud.

For enterprise organizations operating in the cloud, the advanced capabilities of composable data infrastructure translates into enhanced customer experience with more responsive applications, faster time to market for delivering new products and features to customers, and ultimately, competitive advantage.

For more information on Volumez, please visit volumez.com.

IMPORTANT INFORMATION ABOUT THIS PAPER

CONTRIBUTORS
Matt Kimball, Vice President and Principal Analyst, Datacenter Compute and Storage

PUBLISHER
Patrick Moorhead, CEO, Founder, and Chief Analyst at Moor Insights & Strategy

INQUIRIES
Contact us if you would like to discuss this report, and Moor Insights & Strategy will respond promptly.

CITATIONS
This paper can be cited by accredited press and analysts but must be mentioned in the context, displaying the author's name, title, and "Moor Insights & Strategy." Non-press and non-analysts must receive prior written permission from Moor Insights & Strategy for any citations.

LICENSING

DISCLOSURES
Volumez commissioned this paper. Moor Insights & Strategy provides research, analysis, advising, and consulting to many high-tech companies mentioned in this paper. No employees at the firm hold any equity positions with any companies cited in this document.

DISCLAIMER
The information presented in this document is for informational purposes only and may contain technical inaccuracies, omissions, and typographical errors. Moor Insights & Strategy disclaims all warranties regarding such information's accuracy, completeness, or adequacy and shall have no liability for errors, omissions, or inadequacies. This document consists of the opinions of Moor Insights & Strategy and should not be construed as statements of fact. The views expressed herein are subject to change without notice.

Moor Insights & Strategy provides forecasts and forward-looking statements as directional indicators, not as precise predictions of future events. While our forecasts and forward-looking statements represent our current judgment on the future, they are subject to risks and uncertainties that could materially cause actual results to differ. You are cautioned not to place undue reliance on these forecasts and forward-looking statements, which reflect our opinions only as of this document's publication date. Please remember that we are not obligating ourselves to revise or publicly release the results of any revision to these forecasts and forward-looking statements in light of new information or future events.