

Composable Infrastructure for Public Cloud Service Providers



Composable Infrastructure Delivers a Cost Effective, High Performance Platform for Big Data in the Cloud

How can a public cloud provider offer a dynamic infrastructure for AI, big data and analytics that delivers bare metal, dedicated instance performance, while also allowing customers to easily scale up or down as needed and capture the economic advantage of a shared resource?

Executive Summary

The public cloud offers significant advantages for deploying traditional enterprise applications. It's fast and easy to start since the equipment is ready to go in the cloud and doesn't need to be purchased and installed by the user. Users can easily scale resources up or down as needed, allowing them to only pay for the equipment they need. And the cloud provider maintains the infrastructure for the enterprise client, while providing the service at an economically attractive cost.

But the economics of the public cloud are built on a shared infrastructure, where multiple customers are running their applications in virtual machines (VMs) on servers, sharing compute, network and storage resources. As the application loads come and go from various customers, the shared resource is highly utilized yielding low cost for all users. However, co-tenancy can introduce variations in performance due to a phenomenon known as the "noisy neighbor", where one or more VMs on a physical host consumes large amounts of compute resources or storage bandwidth resulting in poor performance for other clients' applications.

Big data analytics, in particular, are not well suited to a shared infrastructure. This is one of the principles first developed by hyperscale companies like Facebook and Google. These applications perform best on a scale-out, bare metal architecture, where each client's workload runs on its own, dedicated cluster of machines. While available from cloud providers, a dedicated infrastructure can

take a long time to deploy or modify for each customer, is much more expensive to use since it isn't shared, and doesn't provide the elasticity customers expect from a public cloud service.

So how can a public cloud provider offer a big data infrastructure that delivers bare metal, dedicated instance performance while also allowing customers to easily scale up or down as needed? And at the same time, allow the cloud provider to capture the economics of a shared resource that is better utilized than a dedicated resource?

The answer is Composable Infrastructure. Composable Infrastructure delivers the performance of dedicated, bare metal clusters, but with the elasticity and cost effectiveness public cloud providers have come to rely upon.

The Economics of a Shared Public Cloud Service

Public cloud services running traditional applications in VMs capitalize on the nature of these workflows which come and go on a very predictable basis. An MRP job is run once or twice a day; the rest of the time it's quiet and other VMs on the machine are starting, running or stopping. Many customers can share the same resource. Cloud providers use statistical methods to balance the load across a large number of servers, and since each job is relatively small and short lived, the average experience of users is excellent.

Modern, big data workloads are different. By their nature, they are processing huge amounts of data and are thus CPU or I/O bound for long periods of time. Then, they stop and do nothing for long periods of time. Standard VM-based approaches to sharing resources don't work well for these applications. Resources cannot be shared concurrently because the jobs are very large compared to the available resources. The result is that a big data analytics job is itself a "noisy neighbor" to the other applications sharing the resource, and its performance is severely impacted by other jobs coming in and out while it is running.

One solution the Public Cloud providers can offer for this situation is "dedicated instances" where the user rents 100% of the infrastructure resources 100% of the time. This addresses the "noisy neighbor" problem, but the economics of sharing are lost and the service is quite expensive.

Composable Infrastructure allows a 3rd approach, where resources can be 100% dedicated to a single customer when their workloads are running, but rented out to other users when they are not. It's "time sequencing" sharing rather than "concurrent sharing".

Modern Workloads Require a New Approach

Hyperscale companies like Facebook and Google were the first to develop a new architecture that could store and process a vast quantity of data at low cost. Whereas virtual servers were more than adequate to handle traditional business applications, modern workloads require an architecture that distributes the data, and the processing of that data, across dozens to thousands of bare metal servers. This "scale-out" architecture used by the hyperscale companies is based on some key principles:

- **Deploy Commodity Servers** – Clusters of "industry standard servers" provide vast storage and compute capacity at the lowest possible cost.
- **Shared Nothing** – Each node needs to be independent, equal and parallel to all other nodes in the cluster. This allows for scaling from 10's to 10,000's of nodes.
- **Data Locality** – Keep the data close to the processor so that it doesn't have to travel across the network. This means keeping it on the server's local drives where it's being processed.

Servers in this model have direct attached storage drives, whereas the typical public cloud architecture uses NAS or object storage systems with servers accessing storage over the network. The latter can suffer from a severe storage and network bottleneck in a scale-out environment as it does not satisfy the Data Locality and Shared Nothing principles.

Fortunately, there’s a new solution that combines the cost effectiveness and elasticity of a virtualized service with the performance and predictability of a dedicated instance. Public cloud providers can deliver a big data service that won’t disappoint on performance and can easily be time sequence shared between multiple customers.

Composable Infrastructure Checks All the Boxes

Hyperscale companies have shown how a scale-out architecture based on some key principles provides the optimal infrastructure for modern workloads. But public cloud providers are hard pressed to deliver this architecture in a cost-effective and agile manner.

An innovative data center architecture known as Composable Infrastructure was developed to directly address the issue of providing dynamic resource pools for modern workloads requiring a scale-out environment. Composable Infrastructure is designed to support all of the hyperscale infrastructure principles, while delivering the flexibility and agility users associate with the public cloud. It addresses the shortcomings of a dedicated instance cloud service, including rapid deployment time, a fully dynamic resource pool and a cost-effective service. Composable Infrastructure delivers the best of both virtualized and dedicated instance services, giving public cloud providers a superior offering for modern workloads.

	Virtualized Public Cloud Service	Composable Infrastructure	Dedicated Instance Cloud Service
Cost Effective	X	X	
Elasticity of Resources	X	X	
Fast Time-to-Deployment	X	X	
Scale Compute and Storage Independently	X	X	
Dedicated, Bare Metal		X	X
Predictable Performance		X	X
Shared Nothing		X	X
Data Locality		X	X

Composable Infrastructure brings together the best of "virtualized" and "dedicated instance" cloud services.

Composable Infrastructure Delivers Agility by Enabling “Infrastructure as Code”

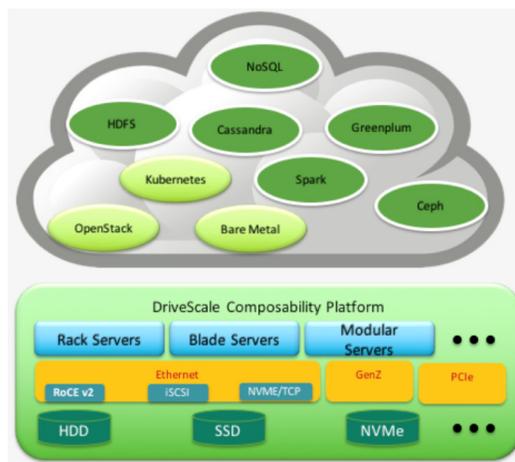
Composable Infrastructure works by combining separate compute and storage resources into “physical” servers and clusters under software control. This allows cloud providers to quickly deploy infrastructure that is optimized for each customer’s workload, and adjust that infrastructure quickly and easily as needed. The result is a much more efficient use of the data center resource pool, and an infrastructure that’s as agile as a virtualized service.

Composable Infrastructure disaggregates compute and storage by replacing standard servers and their internal disks with pools of commodity compute and storage resources consisting of diskless servers and commodity flash or hard disk drives. Operators can easily combine and recombine those resources into software-defined physical nodes and clusters that are indistinguishable from standard servers to the software running on them. There are no abstraction or virtualization layers, and performance is equivalent to bare metal servers with direct-attached storage.

Now public cloud providers can offer a superior service for customers looking to deploy modern workloads on a scale-out infrastructure with all the benefits of a dedicated instance service including predictable, bare metal performance, while adhering to the hyperscale principles of data locality, shared nothing and the use of commodity compute and storage.

Unlike dedicated instances, Composable Infrastructure also provides the benefit of agility. A customer’s data set will reside in a group of disks, which they will rent on a permanent basis. Customers can attach those disks to a set of servers at any time and “spin up” clusters to run their job for as long as needed. The number of servers attached to the data set can be changed depending on the type of workflow and speed of completion needed. Once the work is done, the cluster is “freeze dried”, and drives are disconnected from all the servers which are put back into the compute pool to be used by the next customer and attached to other datasets.

The customer only pays for the compute time they use. Clusters can easily and quickly be spun up or down, with nodes being added or removed from each cluster by software command. Compute is “on demand” and storage resources can be independently scaled to the data set’s requirements, enabling each customer’s cluster to be right-sized for its workload, thus eliminating underutilized resources or the need for overprovisioning. With Composable Infrastructure, service providers can better plan their capacity for all customers allowing them to have resources ready for on-demand deployment.



Composable Infrastructure

Composable Infrastructure offers additional benefits for the cloud provider:

- Lower cost upgrades - Since compute and storage resources are disaggregated, each can be upgraded independently. For example, servers can be upgraded to the latest Intel processors without having to replace the internal drives used with standard servers. Substantial savings can be realized by extending component life cycles.
- Avoid or fix failures faster - When a server fails, the drives attached to it can quickly and automatically be reassigned to another server, keeping the customer's application from experiencing any degradation. Similarly, drive failures can be avoided by detecting errors before the drive completely fails, copying the data to a new drive and recomposing it to the server with the failing drive.
- Reduced data center footprint - Rack space, power and cooling costs can all be reduced with the ability to use diskless servers and commodity storage. Diskless servers and commodity storage allow denser packaging of compute and storage resources than standard servers with internal drives.

Composable Infrastructure Offers Substantial Cost Savings vs. Dedicated Instances

Modern, big data workloads run on clusters of computers, whereas traditional applications share the resources of a single, virtualized server. Since they inherently require more equipment to run, modern workload infrastructure will cost more than infrastructure for traditional applications.

But Composable Infrastructure can dramatically lower the cost of a public cloud service supporting modern workloads as compared to a dedicated instance offering. This is achieved in the following ways:

- Using commodity servers and storage vs more expensive servers
- Right-sizing the infrastructure for each workload and maximizing utilization of all resources
- Breaking down the silos between customers and workloads so that resources can be easily moved and shared between them
- Lowering operating costs since changes to the infrastructure are done via software control rather than manual reconfiguration
- Lowering upgrade costs by separating compute and storage upgrades
- Preventing and responding to failures faster, while improving SLA delivery

Just as virtualization allows clients to share a server, Composable Infrastructure lets them share the compute and storage resource pools from which they build clusters. With these cost advantages, public cloud providers can deliver a much more cost-effective service for modern workloads.

Ease of Implementation and Use

Composable Infrastructure can quickly and easily be integrated into any cloud provider's environment. It works with any industry standard equipment, so service providers can continue to procure servers and storage from the vendor of their choice. There is no forklift upgrade. Composable Infrastructure can be implemented incrementally as new racks are deployed, or existing racks are upgraded. Clusters can consist of a combination of traditional servers with internal storage and new software composed nodes.

Composable Infrastructure is easy to manage with a management console providing a view of all resources in each data center. Composable Infrastructure can be scripted to manage events and automate redeployment of resources. It can also be integrated with existing data center management software tools via a RESTful API.

Summary

Composable Infrastructure is a new solution that enables public cloud providers to deliver a superior service for today's modern workloads running on clusters of computers. It combines the cost-effectiveness, instant availability and agility users have come to expect from virtualized cloud offerings, with the predictable, bare metal performance of a dedicated instance service.

Learn more about Composable Infrastructure at <http://www.drivescale.com/>.



DriveScale is the leader in Composable Infrastructure for enterprise and cloud data centers. The DriveScale Composable Platform disaggregates compute and storage resources, and quickly and easily composes them into infrastructure optimized for dynamic, data-intensive workloads such as big data, analytics and AI. IT leaders can perform capacity planning in real time, accelerate infrastructure deployment and optimize resource usage, while reducing costs. DriveScale's platform is open and works with a large ecosystem of alliance partners. Founded and led by technologists with deep roots in IT architecture who built enterprise-class systems for Cisco, Sun Microsystems and NetApp



DriveScale, Inc
1230 Midas Way, Suite 210
Sunnyvale, CA 94085

Main: +1 (408) 849-4651
www.drivescale.com