INSIGHT

HPE Synergy: Composable/Disaggregated Infrastructure for the Enterprise

Ashish Nadkarni

IDC OPINION

HPE Synergy is a new class of system that falls under a category known as Composable/Disaggregated Infrastructure – an emerging category of datacenter infrastructure that seeks to (dis)aggregate compute, storage, and networking fabric resources into shared resource pools that can be available for on-demand allocation (i.e., "composable"). The HPE Synergy system is a full-stack system comprising the disaggregated hardware platform and a suite of management software that allows rapid provisioning and seamless ongoing management of resources. HPE is targeting Synergy for IT organizations that require a common (read: flexible and agile) infrastructure to host next-generation applications (NGAs) and current-generation applications (CGAs) – whose design principles and therefore operational characteristics are fairly divergent:

- CGAs assume infrastructure resiliency and require computing technologies like virtualization and clustering that provide portability and transparently preserve application state. CGAs include ERP, CRM, and other enterprise applications and utilize shared storage and relational databases.
- NGAs leverage application components that are stateless, share nothing, horizontally scalable, and delivered via lightweight containers. They also leverage application and/or data layer processes that do not assume infrastructure resilience. Some prominent examples of these NGAs are mobile, online gaming, IaaS/SaaS, IoT applications, and Big Data, and they utilize newer methodologies like DevOps and Continuous Integration and Development (CI/CD).

August 2016, IDC #US41652216

IN THIS INSIGHT

This IDC Insight provides an assessment of the set of technologies and solutions that are collectively classified as "Composable/Disaggregated Infrastructure." It also discusses workloads, applications, and customers that are driving demand for these technologies and solutions.

The term *Composable/Disaggregated* is used to imply that "Composable" and "Disaggregated" are complementary to each other and denote two different things. The two terms are not interchangeable, even though a "/" may imply otherwise. According to IDC's taxonomy on Composable/Disaggregated Infrastructure:

- "Composable" implies the ability to create a set of virtually consumable infrastructure entities from physically disparate resources "disaggregated" at a component level via unified low-level application programming interfaces (APIs). "Composability" is driven at the software (API) level, and "Disaggregation" is primarily driven at the hardware level. To fully implement the design principles of Composable Infrastructure, the hardware it operates on has to support partial or full disaggregation (in which resources are pooled down to a component level).
- "Composable" and "Disaggregated" systems have different evolutionary trajectories. The enabling technologies for these systems are in different stages of maturity: some such as common APIs offering infrastructure-as-a-code capabilities are becoming available, while others such as silicon photonics are still in development. Therefore, "Composable Infrastructure" software can hypothetically operate on any type of hardware as long as that hardware supports the Composable API.
- IDC views Composable/Disaggregated systems as an evolution of converged and hyperconverged infrastructure. While the new technology is a significant leap forward, the gist of this evolution is: The hardware side is moving toward disaggregation, while the software side is moving toward composability — via a unified API-based provisioning, orchestration, and automation layer.
- The Composable stack, which is basically a collection of software tools and stacks, will be tracked under an existing or a newly defined "infrastructure software" functional market (to be determined). The "disaggregated" hardware stack (i.e., partially or fully disaggregated hardware) for now will be tracked in the existing computing platforms (aka server) markets. If there are external systems like storage arrays and networking equipment that are part of this infrastructure, they will be tracked in their respective markets. The hardware itself could be considered to be in a category of its own, but it is too early to say so.

For more details, refer to *Composable/Disaggregated Infrastructure and Rackscale Architectures — Market Background, Trends, and Taxonomy, 2016* (IDC #US41633516, August 2016).

SITUATION OVERVIEW

With the transition to the 3rd Platform, IDC is seeing a rise of a new class of workloads — collectively referred to as "next-generation applications" (some industry taxonomies refer to such applications as "cloud-native applications," or CNAs). These emerging applications have distinct characteristics different from that of "traditional" enterprise applications (referred to as "current-generation applications") and, as such, have unique infrastructure requirements. Some prominent examples of these NGAs are mobile, online gaming, IaaS/SaaS, and Big Data.

Furthermore:

- CGAs typically support internal operations, including business continuity and ensuring service uptime. CGAs assume infrastructure resiliency and require computing technologies like virtualization and clustering that provide portability and transparently preserve application state. CGAs include ERP, CRM, and other enterprise applications and utilize shared storage and relational databases. Currently, IDC views CGAs to be in a mode of consolidation, where IT organizations are seeking ways to drive opex efficiencies. Technologies such as virtualization and converged systems enable IT to deliver on such efficiencies.
- NGAs on the other hand are often built to run in cloud environments and on devices and require IT to adapt to high rates of change and uncertainty in performance and capacity not seen with CGAs. NGAs leverage application components that are stateless, share nothing, horizontally scalable, and delivered via lightweight containers. They also leverage application and/or data layer processes that do not assume infrastructure resilience. Some prominent examples of these NGAs are mobile, online gaming, IaaS/SaaS, IoT applications, and Big Data, and they utilize newer methodologies like DevOps and Continuous Integration and Development.

Whereas CGAs typically scale by increasing the number of virtual machines running on systems, NGAs scale horizontally – increasing performance and/or capacity by adding servers. The triggers for demand in growth are also distinct. Generally speaking, enterprise applications, such as financial/accounting and email, are tied to organizational metrics, such as growth of employees, customers, and revenue. The capacity planning, while still not an easy task, is typically months or years. The metrics are known quantities and more often than not involve processing structured data sets. In contrast, the capacity planning for next-generation applications is less predictable and involves nonlinear scaling of the service levels. The capacity planning is typically days or weeks, and it is also extremely common to react to unanticipated changes. The instantaneous fluctuations in demand, either up or down, require the rapid provisioning and reprovisioning of resources to the applications.

Given the vast differences in requirements between CGAs and NGAs, the challenge for IT is to deploy an infrastructure that works for both. Solving this challenge is a function of the mandate for IT to remain relevant in the era of cloud – to be cost focused and operations driven for CGAs and to be agility focused and line-of-business driven for NGAs.

HPE Synergy: The First Instantiation in a New Class of Infrastructure

In late 2015, HPE announced a new class of systems known as Synergy that is built to further its vision of a fully composable datacenter infrastructure. Synergy is designed to serve as a bridge for businesses that are in the thick of CGAs but are rapidly transitioning to supporting NGA environments. From that side, it serves as the best of both worlds – for applications with infrastructure resiliency requirements and for applications that don't assume infrastructure resiliency. The three key foundational design elements of HPE Synergy that enable IT organizations to deploy NGAs at "cloud speed" or take a more traditional approach to deploy highly resilient infrastructure for CGAs are:

- **Unified API** allows organizations to implement infrastructure as code – a single line of code that abstracts all infrastructure. This enables developers to integrate infrastructure provisioning commands directly into the application development process, which in turn allows a quicker deployment of applications.

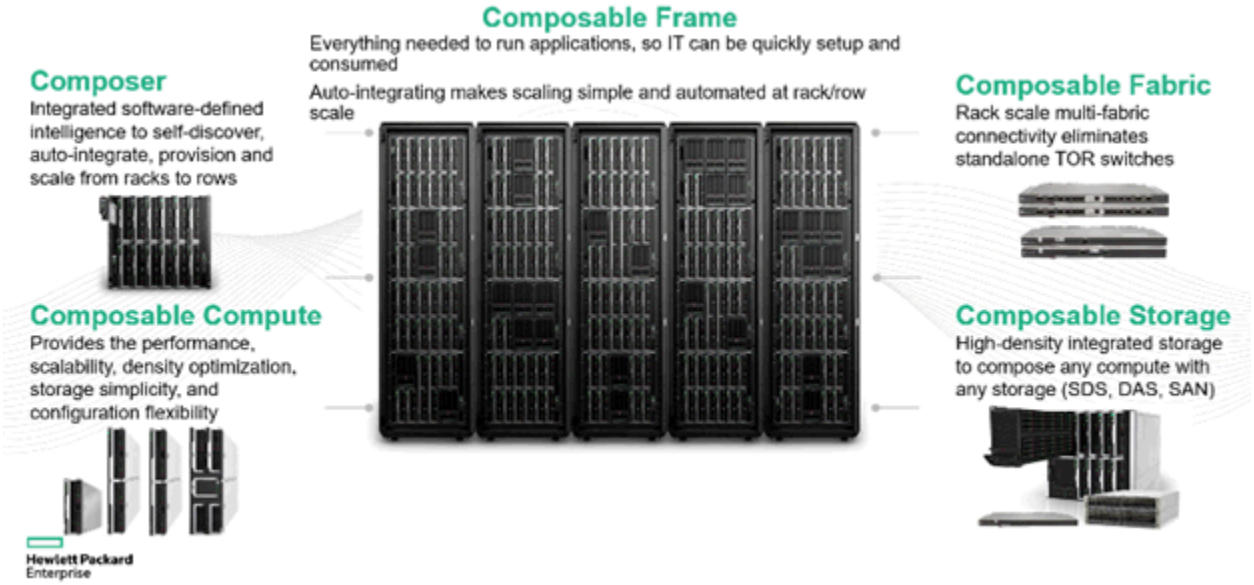
- **Software-defined intelligence** enables organizations to take a template-driven approach to workload composition and management. This minimizes hiccups caused by operational activities such as OS patching and firmware upgrades, thereby improving IT efficiency and reducing opex costs.
- **Fluid resource pools** provide the ability to compose pseudo-physical, virtual, and containerized computing units from compute (CPU and memory) and data persistence (disk and flash) resource pools. This reduces the wastage and overhead caused by overprovisioning of resources, and thus capex costs.

Figure 1 illustrates the main components of the HPE Synergy system. They are:

- **Composable frame (includes compute, fabric, storage, and management modules):** The frame houses compute and storage modules. Compute modules provide CPU and memory capacity, while storage modules contain shareable DAS drives and flash. Each node contains up to 200 drives. Apps like Hadoop and email that require shared nothing capacity or an SDS block-storage software, like StoreVirtual VSA, can utilize these drives for data persistence. Fabric modules allow frame expansion and external connectivity. The fabric modules can uplink to the datacenter directly – 10G or 40G or FC. Internally, it runs on a shared PCIe fabric and can run across multiple frames.
- **Composable software suite:** Two key elements of this software stack are the HPE OneView Composer and Image Streamer. Composer is the provisioning engine that allows resources from the frame to be carved into pseudo-physical compute and persistence units. Image Streamer is the software that allows the operating platform to be provisioned onto these pseudo-physical units.

FIGURE 1

HPE Synergy Core



Source: HPE, 2016

HPE Synergy also supports remote object, file, or SAN storage to be connected to the frame. In the case where the external storage supports the composable API (such as HPE 3PAR StorServ arrays), the provisioning domain can be extended to include such storage.

While in theory any hardware could be disaggregated, in reality the hardware has to be truly disaggregated. This is where the HPE Synergy design is truly different. It is stateless from a data perspective — so the OS images can float anywhere in the frame and from a configuration state perspective — to ensure that images can be altered on the fly.

FUTURE OUTLOOK

From a systems perspective, any Composable/Disaggregated architecture is made up of two parts: the first is the ability to (dis)aggregate IT resources into compute, storage, and fabric pools, and the second part is to compose consumable resources from such disaggregated pools via a unified API. This means IT organizations now have to manage a greater number of smaller assets. This could actually result in more complexity. Intelligent software is therefore needed to manage all the distinct assets and to compose the optimal configuration for a specific application. All the elements inside the installed infrastructure are pooled and require management. This requires:

- **Discovery and location of resources.** Monitoring and life-cycle management software are necessary to provide full awareness of the hardware assets and application workloads.
- **Self-discovery, provisioning workloads, orchestration, and healing.** It will be essential to have visibility into the utilization of the discrete resources and understand the load on the resource elements.
- **Orchestration layer.** It will be essential to enable a catalogue compute, storage, networking, and memory in an orchestration layer and also define resource requirements for specific applications and compose them in a set that is optimized for the workload.

IDC sees these as necessary factors for Composable/Disaggregated systems to be successful in the enterprise — in speeding up provisioning times, improving IT utilization, and simplifying overall IT operations. For the foreseeable future, IDC expects systems vendors like HPE to dominate this space — primarily because of their decades of collective experience in building enterprise infrastructure platforms.

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

Global Headquarters

5 Speen Street
Framingham, MA 01701
USA
508.872.8200
Twitter: @IDC
idc-community.com
www.idc.com