White Paper

The Value of the Hyperscale Cloud: Understanding the Benefits for NFV

Prepared by

Caroline Chappell
Principal Analyst, Cloud & NFV, Heavy Reading
www.heavyreading.com

on behalf of

www.citrix.com

February 2015
Executive Summary

Hyperscale cloud has been developed by the Internet giants to support the creation and delivery of software-based services at blistering speed, and at the lowest possible cost. Since software is the foundation for service innovation in the 21st century, those players that master this capability will eat markets – the achievement to date of Amazon, Google, Facebook, et al.

The original ETSI network functions virtualization (NFV) vision was to adopt hyperscale cloud architecture and practices to accelerate the creation and delivery of software-based network services, restoring operators’ ability to compete on a level playing field against new Internet rivals.

This vision has become somewhat obscured along the way, due to misunderstandings about the hyperscale cloud itself and how to migrate to it. The telecom industry has spent considerable effort on the development of a virtualized, programmable infrastructure that is fit for NFV. However, it has so far made little effort to address the other two cornerstones of hyperscale cloud: a highly modular (“microservice”-based), scalable and available application architecture and DevOps management approach.

Without all the cornerstones supporting each other, the cloud edifice is weak and doesn’t deliver the transformational benefits enjoyed by rival Internet practitioners. This white paper defines what hyperscale cloud is, and points out the interdependence between each of its three cornerstones. It busts the three big myths about hyperscale cloud that are often used to distract operators and point them toward less transformational virtualization architectures. Finally, it looks at how hyperscale cloud is being applied to the network and what steps operators must take to ensure that they achieve what they hope for and expect from NFV, driving industry change rather than reacting to it.
The Principles of Hyperscale Cloud

Reasons for Building the Hyperscale Cloud

Software is at the heart of service innovation in the 21st century. The hyperscale cloud has been developed by the Internet giants (such as Amazon and Google) and cloud providers (such as Microsoft) for one key reason: They want a platform that supports the continuous, reliable and scalable delivery of software-based services without the cost and speed restrictions imposed by the traditional manipulation of physical IT and networking infrastructure.

The hyperscale cloud is a software-based environment that is abstracted away from physical infrastructure, so that the resources the infrastructure provides can be manipulated rapidly and programmatically, without lengthy procurement cycles and time-consuming truck rolls. Software applications developed to run on the hyperscale cloud environment are built in a certain way that makes them quick and inexpensive to deploy and highly resilient to physical infrastructure failure.

The hyperscale cloud requires new development and operational practices that complement its virtualized environment and new application architecture. Together, these hyperscale cloud properties remove the roadblocks to service innovation, allowing new software to be deployed almost as fast as it can be created.

Network operators have recognized the power of hyperscale cloud and want to tap into it to deliver software-based network services. Today, 80 percent or more of the function in any network appliance is implemented in software. By extracting and re-architecting that software to run on the hyperscale cloud, network operators can leverage the hyperscale cloud’s agile, low-cost and scalable software deployment benefits to build more flexible, innovative and cost-effective software-based networks. This is the vision of NFV.

The presence of the word “virtualization” within the term “NFV” is distracting the industry from the end goal of the initiative: the implementation of hyperscale cloud. Virtualization is one aspect of hyperscale cloud, but on its own, it won’t achieve the transformational benefits envisaged by the ETSI NFV founding members: lower capex, significantly lower opex and several orders of magnitude acceleration in a telco’s speed of doing business.

Telcos need to build hyperscale cloud foundations for NFV if they want the same flexible and cost-efficient business environment that their Internet competitors enjoy.

The Three Cornerstones of Hyperscale Cloud

The hyperscale cloud is very different from a traditional, server-based software execution environment, whether or not those servers have been equipped with hypervisors. Unless companies understand how different the hyperscale cloud is and what changes they will need to make to exploit it, they will not be able to harness it effectively to achieve its benefits. In fact, network operators can make life more complicated for themselves if they try to migrate network functions designed for traditional environments and to apply existing operational practices to the hyperscale cloud.

Hyperscale cloud has three interdependent cornerstones, all of which need to be present to realize the benefits:
1. A virtualized infrastructure that can be programmatically controlled

In the hyperscale cloud, virtualized infrastructure, servers, storage and network connectivity exist as a set of digital representations, insulated from physical hardware through a virtualization layer. Hypervisors or any other type of virtualization technology, such as Linux containers or software-defined overlay networks, are merely the means of creating virtual abstractions of infrastructure resources.

Virtualization is the lubricant in the hyperscale cloud that makes possible programmability – the manipulation of the virtualized infrastructure resources at different levels of granularity through application programming interfaces (APIs). Programmability implies the automated execution of instructions, so that hyperscale cloud resources are manipulated and configured automatically rather than manually. The high levels of programmed automation possible in the hyperscale cloud mean that infrastructure changes can be implemented fast, often and with a high level of reliability.

2. A highly modular, scalable and available application architecture

The hyperscale cloud has driven the development of a new software architecture design pattern in which applications are composed from small, decoupled “microservice” components, each of which focus on a specific task. Microservices communicate with each other through APIs, and each instance of a microservice is interchangeable with any other instance of itself – a concept called “fungibility.”

For example, virtual resources, Web server, application server and database microservices may be composed into applications at runtime. Each microservice is replaceable, so an application can be deployed with n number of each type of microservice, depending on the capacity required in that tier of the application.

Orchestrating lots of small, replaceable components is inherently more flexible than trying to manage large, monolithic blocks of software, so instance “fungibility” supports scalability. For decades in the telecom industry, scale has been linked to size: If operators want to scale a network element, they buy a bigger box. The hyperscale cloud application architecture scales by adding more microservice instances to an
application. This is more efficient, as new instances are created only when they are needed and only to meet the demands on the application at a particular time. When demand drops, instances can be scaled back and resources reclaimed.

A microservice-based application architecture also creates **system-level high availability**. In the fungible world of the hyperscale cloud application architecture, individual microservices are never single points of failure. Since an application is composed from multiple, replicated software units, state is held at the application level, not at the level of a specific instance. This means that any instance can fail without affecting the overall availability of the application.

The entire hyperscale cloud naturally operates in active-active mode, with each microservice instance providing backup to the other(s). When an instance fails, an equivalent instance can immediately be spun up in its place, restoring the overall fault tolerance of the application to its previous level. Some of the most impressive demonstrations of the power of applications to withstand failure in the hyperscale cloud come from Netflix's Simian Army.

The hyperscale cloud application architecture achieves high availability at much lower cost than a traditional environment. In the latter, availability depends on the performance and health of an application and its supporting physical server; or, in a networking context, a network appliance. If the application/service/appliance fails, the service it performs is unavailable. Each network appliance is expected to have five-nines reliability to prevent this, and/or be paired with a mirror-image application server/appliance, just in case of failure. High-availability pairing is not only very expensive, due to the need to overprovision capacity by a massive 100 percent, but there are also risks involved in failing over to the redundant server/appliance.

High availability is not just about unplanned failure; it is also needed to ensure the application keeps running during planned outages, such as for software upgrade and maintenance purposes. In the hyperscale cloud, an out-of-date microservice can be "killed" and immediately replaced with a new, upgraded instance without affecting the overall execution of the application. This capability powerfully supports high availability during planned lifecycle management activities.

3. A "DevOps" approach to managing and orchestrating hyperscale cloud infrastructure and applications

Because hyperscale cloud is a new platform for running applications, it requires a new operational mindset and practices. A DevOps approach is a key requirement for the successful operation of a hyperscale cloud.

Fundamentally, DevOps is about developers taking more responsibility for the operation and management of their applications at runtime. In a DevOps model, developers cooperate with operational teams to deploy their applications. This contrasts with a traditional model where, once developed, applications are "thrown over the wall" to operations staff, who are then responsible for scoping the execution environment the application would need.

In the hyperscale cloud, the virtual resources for an application's execution environment already exist as microservices that have been predefined by operations staff and can be programmatically instantiated in minutes by the developer. In turn, developers define for operations teams what aspects of an application they need to monitor at runtime and what the rules are for triggering the scaling and healing of a particular application component (microservice).
Automation is key to DevOps. Hyperscale cloud operations are too numerous, change too fast and the complexity associated with managing tens of thousands of microservices too great for manual operations. A hyperscale cloud is an “auto-mate first” environment. Companies that practice DevOps are as concerned about automation as network operators are about five-nines reliability. Indeed, hyperscale cloud operators achieve highly cost-efficient application lifecycle management, five-nines application availability and rapid onboarding and/or modification of applications in minutes as a result of operational automation.

Debunking the Myths Around Hyperscale Cloud

There are several common misunderstandings associated with hyperscale cloud.

**Myth No. 1: Hyperscale cloud cornerstones can be added incrementally**

To date, the NFV initiative has focused strongly on one cornerstone of hyperscale cloud: the virtualized, programmable infrastructure. It has largely ignored the others because they are even harder to achieve: software-based network function needs to be re-architected as collections of microservices and operational practices need to be changed and automated, with disruption to existing organizational culture and skill sets.

There is a misplaced belief, fostered by industry vested interests, that a virtualized environment alone will deliver early NFV benefits, and that the other aspects of the hyperscale cloud – its application architecture and DevOps operational approach – can be added a long way down the road. At present, most virtual network function (VNF) owners typically take an existing network function and run it unchanged on top of virtualized infrastructure. This approach has two advantages for VNF vendors: it requires minimal effort (and thus cost), and the vendor can control the VNF’s operational environment, which is often a proprietary implementation of a virtualized infrastructure, rather than a hyperscale cloud.

Unfortunately, there is no shortcut to building the hyperscale cloud. All three cornerstones have to be present from the start if the project is to succeed. Operators may find that the costs of running complex, performance-sensitive network functions on a virtualized infrastructure, without DevOps-based automation or a microservice architecture approach to high availability and scalability, far outweigh the expected benefits. This is because NFV’s benefits are predicated on hyperscale architecture and practices, not on virtualization alone.

**Myth No. 2: Hyperscale clouds are large and centralized**

Because Amazon, Google, et al., run hyperscale cloud in vast data centers, and because of the connotations of the prefix “hyper,” many operators assume that a hyperscale cloud will not be relevant for the kinds of networking use cases they want to deploy. These include running very small-scale clouds in highly distributed locations, such as central offices or cell sites.

In fact, the term "hyperscale" refers to the fact that the cloud architecture is highly scalable, from the very small scale – one or two server blades – to massive scale data centers. The power of the hyperscale cloud is based on the fact that the same application – or VNF – can run, unchanged, at any scale, anywhere in the hyperscale cloud infrastructure.
Myth No. 3: Hyperscale cloud architectures always oversubscribe resources

Since virtualized resources in a hyperscale cloud infrastructure are programmatical-ly allocated to applications, cloud operators can control that allocation accord-ing to business policies and parameters such as quality of service and cost.

Public cloud providers are the best-known users of hyperscale cloud architectures. To achieve their very low usage charges, public cloud providers often choose to oversubscribe the allocation of cloud resources, such as CPU cycles and memory. Oversubscription allows them to pack applications onto their infrastructure in a highly cost-effective way. Knowing this, many public cloud customers routinely “kill” the lowest-performing 5 percent of application instances and use the principle of fungibility, together with DevOps automation, to spin up new ones in an area of the infrastructure that is less subscribed.

But oversubscription of resources is a choice that hyperscale cloud makes possible: In other words, oversubscription is a programmable property, not an inherent feature of hyperscale cloud. This means operators can choose how to allocate resources to VNFs depending on a VNF’s requirements for processing power, memory and network I/O. The hyperscale cloud can just as easily support VNFs with stringent performance and quality-of-service requirements as it can best-effort IT workloads.
Preparing for Hyperscale Cloud-Based NFV

Applying Hyperscale Cloud to the Network

Many vendors and operators want to start their NFV journey by virtualizing user plane network functions that forward/manipulate IP packets, such as load balancing, firewall, DPI, encryption and content filtering functions. Layer 4-7 functions represent a good starting point to apply hyperscale cloud concepts, since they are:

- Fundamental to the secure and resilient operation of the cloud, so they are needed to support a hyperscale cloud implementation.
- Network services that enterprise users want to purchase in a self-service, pay-per-use model. Telcos therefore want a faster and more cost-efficient way of delivering these services, based on hyperscale cloud properties of virtualization, automation, reliability and scale.
- Network functions needed at the customer edge of the network that can be expensive to deliver physically, requiring manual installation. Layer 4-7 services are closely associated with the NFV virtual CPE (vCPE) use case, which many operators want to adopt both to cut the cost of delivering these network functions to enterprise customers, and to raise more revenue by providing such customers with a greater choice of network-based services. This use case is well supported by hyperscale cloud: VNFs can be rapidly provisioned and managed automatically on virtualized infrastructure at the edge of the network using a DevOps approach; and, if properly implemented using the hyperscale cloud application architecture, the VNFs will run reliably and scalably on that infrastructure, minimizing truck rolls and operational cost.

The application delivery controller (ADC) can encompass multiple Layer 4-7 network functions, and is an example of a VNF that vendors are re-architecting for the hyperscale cloud. Today, virtualized ADCs are large, monolithic VMs that integrate all the functions and serve multiple tenants. This creates complexity, especially configuring and maintaining multiple different policies, features and licenses for the different needs of each tenant.

If the ADC is re-architected for hyperscale cloud, a virtual instance(s) can be attached to each tenant’s application, just containing the specific Layer 4-7 functionality needed by that application. These virtual ADCs are instantiated through DevOps processes. There could be thousands of such virtual instances running effectively as “microservices” that protect individual applications in a customized way.

DevOps automation means that the overhead in spinning up and maintaining multiple variants of a VNF instance may be considerably lower in the hyperscale cloud than deploying and managing a single, multi-tenant virtualized or appliance-based ADC. This illustrates the power of the hyperscale cloud to improve service delivery while reducing operational cost.

Core Network Functions

Network subsystems and control plane functions are now beginning to be re-engineered for the hyperscale cloud. An example is the open source Clearwater IMS, which has a CSCF designed for the hyperscale environment. Some policy controllers and online charging systems are following suit. At least one virtual Evolved Packet
Core (EPC) product has been developed to run in the hyperscale cloud. Any VNF can benefit from hyperscale cloud properties, providing it has adopted a microservice application architecture and is managed using the DevOps “automate first” paradigm.

**Steps to Adoption of Hyperscale Cloud**

Recognizing that hyperscale cloud will underpin any business delivering software-based services in the future, far-sighted operators participating in ETSI NFV are championing this approach. However, the industry in general is currently focusing on virtualizing infrastructure, which is just one cornerstone of hyperscale cloud.

Operators that want to achieve the anticipated benefits of NFV need to do more to support the holistic implementation of hyperscale cloud in an NFV context. For example, they will need to:

1. **Acquire the skill sets needed for “automate first”**

Recruiting people with the right skill sets for hyperscale cloud and integrating them into IT and network operations teams is likely to be the largest challenge facing operators adopting the hyperscale cloud.

Today, network operations staff often don’t have an “automate first” mindset. They are used to managing dashboards and asking network equipment vendors to make software changes in business/operations support systems (B/OSS) or element management systems. However, they do have a very deep understanding of the network as a system – a set of interdependent network functions – which is very different from managing discrete “cookie-cutter” IT applications.

Operators need to marry network operations expertise with new hyperscale cloud operations skills brought in by new recruits and/or their IT department. They need to combine the operational knowledge of their network engineers with the “automate first” mindset of DevOps-trained staff to ensure that VNFs and the network services that contain them take advantage of the scale, resilience and cost benefits of the hyperscale cloud.

2. **Influence the market to architect VNFs for hyperscale cloud**

Incumbent VNF vendors have no incentive to change their ways and re-architect their functions unless operators flex their market muscle. Operators need to let vendors know they are serious about hyperscale cloud and press them for a roadmap and timeline in which they can expect VNFs that have been engineered for it.

3. **Educate the organization in the properties provided by the hyperscale cloud**

The separation of software-based network function from proprietary hardware is itself a big step. Operators are concerned because they no longer have a single vendor responsible for the end-to-end performance and availability of the function and the virtual and physical infrastructure it runs on.

Properties of the hyperscale cloud, such as the scalability and system-level redundancy delivered through its application architecture, can alleviate many of these fears. However, the hyperscale cloud requires a completely different – and currently alien – DevOps way of thinking and operating. The DevOps approach requires culture change while new operational practices will lead to organizational change.
The organizational boundaries between IT and network need to be dismantled, as do the barriers between network function developers and network operations if hyperscale cloud is to be applied to the network.

Network operators have a large educational task ahead of them. They need to explain the value of doing things differently with hyperscale cloud and actively champion it as an enabler of their future, software-based network.

4. Consult market players with hyperscale cloud knowledge

To help them acquire and integrate the right skill sets for – and educate their organizations in the benefits of – hyperscale cloud, operators need to turn to industry players with deep knowledge in this area. Hyperscale cloud originated in the IT domain, and IT companies that have contributed to its evolution over a number of years have the best insights into what operators need to deploy it successfully. Players that have not only been involved on the ground floor of hyperscale cloud development but also have knowledge of running network functions in this environment are rare at this stage of the market, and are worth seeking out.
Conclusion

Hyperscale clouds are the future execution environments for software-based services because of the extreme speed and low cost with which they enable such services to be created and launched. Any business will need to leverage hyperscale cloud properties to remain competitive in the 21st century.

Hyperscale cloud does require a mindset shift both within the IT industry, and certainly as a platform for running software-based (virtual) network functionality. It is also not possible to adopt hyperscale cloud in an evolutionary manner: All three of its cornerstones – a virtualized, programmable infrastructure, highly modular and scalable application architecture and a DevOps management approach – need to be present together, supporting each other, for the transformational benefits of hyperscale cloud to be realized.

There are a number of misunderstandings about hyperscale cloud architectures that need to be debunked. For example, hyperscale cloud is not equivalent to public cloud, but public cloud is one cloud delivery model that a hyperscale cloud architecture will support. Hyperscale clouds are not necessarily large – they can be highly distributed across tens or thousands of smaller “data centers,” including telco locations for micro data centers, such as central offices or base stations.

The industry is at the beginning of transforming network functionality for the hyperscale application architecture, starting with Layer 4-7 network functions, such as ADCs and load balancers, which are fundamental to the secure and resilient operation of the hyperscale cloud itself. New entrants have also proven the value of hyperscale cloud as a low-cost execution environment for complex, core VNFs such as IMS or the EPC.

However, more needs to be done to facilitate hyperscale cloud adoption if telecom operators are to compete successfully, and on a level playing field, with their Internet competitors. Operators need to acquire the skill sets and “automate first” mindset needed for hyperscale cloud. They can put pressure on network function vendors to re-architect their software in a hyperscale application architecture model. They will have to educate and evangelize to their organizations to appreciate the value of migrating to the hyperscale cloud, and they need to seek the right advisers to help them build hyperscale cloud.

The evidence that hyperscale cloud delivers transformational business benefits is already out there in the market, demonstrated by the most valuable and successful companies in the world. The transition to hyperscale cloud isn’t easy, but the prizes are great. Those telcos that step up to the challenge will be the ones that not only survive the disruptive forces of software-based innovation, but also master them.
About Citrix

Citrix (NASDAQ: CTXS) is leading the transition to software-defining the workplace, uniting virtualization, mobility management, networking and SaaS solutions to enable new ways for businesses and people to work better. Citrix networking solutions for mobile operators elevate subscriber quality of experience, economically scale data and control planes, and enable the transition to next-generation networks. Citrix solutions are in use at more than 330,000 organizations and by more than 100 million users globally. Learn more at www.citrix.com.