Independent market research and competitive analysis of next-generation business and technology solutions for service providers and vendors

HEAVY READING

WHITE PAPER

Edge Computing in Service Provider Networks

A Heavy Reading white paper produced for Advantech Co. Ltd.

ADVANTECH

intel IoT Solutions Alliance Premier

AUTHOR: GABRIEL BROWN, PRINCIPAL ANALYST, HEAVY READING

WHY EDGE COMPUTING?

Change is coming to computing and networking architectures. Centralized cloud – today's dominant paradigm – will remain vital to efficiently store and process information, but as demand for real-time processing and low-latency connectivity increases, edge computing will become progressively more important.

Service providers have a critical role at the edge. With widely distributed networks and associated real estate – central offices, aggregation sites, mobile telephone switching offices (MTSOs), etc. – they are uniquely positioned to deploy edge computing infrastructure that is close to the user and tightly integrated with the transport and access network.

This white paper makes the case for edge computing in operator networks. It discusses the lead use cases for distributed cloud and the hardware requirements in edge locations. By way of example, the paper references a mobile edge computing (MEC) multi-vendor proof of concept (PoC) on "Multi-Service MEC Platform for Advanced Service Delivery," to show how a standards-based edge cloud infrastructure platform can be used to enhance the subscriber experience and optimize network utilization (leveraging analytics, routing, etc.).

Drivers for Edge Computing

The primary driver for edge computing is low latency and associated quality-of-experience (QoE) service requirements. Simply put, content and applications hosted closer to the user can perform better than those hosted far away. For many services, distance is not an issue (e.g., email or messaging), whereas for others (e.g., industrial control systems or connected cars) it can be critical.

In mobile networks there are major efforts to reduce end-to-end latency in both LTE (via Advanced Pro) and, of course, in 5G. This is backed up by new network architectures, including cloud radio access network (C-RAN) and MEC, that will see operators deploy compute and storage capability close to the user. This distributed cloud infrastructure will support networking functions, such as user-plane forwarding, and end-user services, such as applications and content.

Many service providers operate fixed and mobile access networks and increasingly want to integrate their assets to provide common services to customers. A user consuming video over residential WiFi, for example, may want to continue the session on LTE outdoors. The physical and logical network edge may be a good place to enable converged services, and this is driving the concept of multi-access edge computing (MEC).

There are new services and computing models emerging that will make centralized cloud impractical or, in some cases, prohibitively expensive. Often cited examples are connected cars, industry 4.0, virtual reality and "tactile Internet." These types of "next-wave" services represent the major revenue growth opportunity for network operators and depend on low latency to be effective.

Services & Use Cases

There are two major classes of service that can benefit from edge processing: End-user customer services and operator network services (discussed in the following section).

In the end-user category, services can be as simple as faster video delivery to consumers, enterprise use cases such as security for early detection and mitigation of threats, or venue services at campuses, stadiums, offices, etc.; or perhaps more sophisticated services, such as high-performance remote desktop or image processing for services such as video surveillance.

Several future applications that have mainstream potential are expected to need very fast response times between client and server. These may include virtual and augmented reality, local-area factory automation networks (which require on-site controllers) and services, such as electrical grid monitoring, which have stringent performance requirements. Other services, such as Internet of Things (IoT) gateways, may be driven less by latency and more by the need to reduce transmission and acknowledgement times to improve battery life for sensor devices.

Vehicles as Mobile Sensor Platforms

A good example of how edge cloud processing could be transformative is connected vehicles. High-end automobiles are increasingly equipped with sensors for mapping, monitoring, telemetry, traffic, etc. This includes HD cameras, which generate vast amounts of data.

In a recent presentation, Ford Motor Company said that a high-end SUV today generates about 0.5GB of data per hour, but that in the future a small family car would generate 25GB per hour – a 50-fold increase. Not all of this data will need to be transmitted to the network; the bulk of it will be processed locally. However, where real-time or near-real-time responses are required, it may be expensive and impractical to transfer this to a centralized data center.

Assisted and autonomous driving is expected to be improved by low-latency, vehicle-to-vehicle communications and vehicle-to-infrastructure roadside services (e.g., for assisted hazard warnings), which by nature require edge processing. There are many initiatives underway around the world in this sector, including the Vehicle-to-X work in 3GPP, the 5G Automotive Association activities and more.

Operator Advantage in Distributed Infrastructure

Network operators have widely deployed access and metro networks. These facilities have been developed over a long period and are expensive to replicate. This gives operators an inherent advantage over large, but relatively centralized, cloud providers because, by definition, the access network is the closest network to the customer. Their objective is to translate this theoretical advantage into a real-world set of services either provided directly to customers or on a wholesale basis to third-party content and application providers.

Operators have a wide range of facilities they are able to use to host distributed compute and storage. These can be converted in micro data centers and interconnected over software-defined networking (SDN)-controlled networks to create highly-distributed, highly-available clouds. Some example locations are as follows:

- **Cell sites and street cabinets:** These assets are generally less than a kilometer from users and often substantially less than that. They are challenging environments for data center hardware because of environmental, security, space and power requirements. Nevertheless, emerging server platforms may make them cost effective.

- **Enterprise server rooms:** Operators typically deploy equipment in the customer premises to provide wide-area network (WAN) services. These can be as simple as Ethernet network interface devices (NIDs), or more elaborate installs of routers, firewalls, optimizers, etc. Already small cell and WiFi make use of these facilities to host local applications needed for venue services.
- **Central office/local exchange:** These generally aggregate traffic from street cabinets and include functions such as DSLAMs for ADSL, cable head-ends, and optical line terminals (OLTs) for fiber to the home (FTTH).
- **BSC and RNC sites:** Mobile operators often have sites that are, or were once, used for 2G base station controllers (BSCs) and 3G radio network controllers (RNCs). These functions have been increasingly centralized over the years, but the facility often remains in operation and underutilized.
- **Transport aggregation sites:** This is where switching and routing equipment is deployed to aggregate access traffic. Operators generally have more than one level of aggregation, depending on their size and geography; there is great variability in these facilities between markets and operators.
- **Cloud RAN hub sites:** Also known as "Baseband Hotels," this is where baseband units for centralized and cloud RAN equipment is deployed. These same facilities can be used to host edge cloud infrastructure and services.
- **Mobile Telephone Switching Office (MTSO):** These are mobile network facilities that house core network equipment; they are generally fairly centralized, although somewhat more distributed than a national data center.

One challenge is that these facilities are available to exploit, but as discussed in **Section 3**, often have specific requirements that make it impractical to deploy generic rack-based cloud infrastructure. In some cases, refurbishment and/or optimized edge data hardware is required.

Resiliency & Reliability at Edge Data Centers

Reliability, resiliency and failover are different at the edge. In a centralized hyper-scale data center, failure of a single system has relatively little impact on service as the cloud management software can rapidly move services to new infrastructure. Indeed, one of the major achievements of cloud computing is the ability to offer high reliability services using low reliability components.

Edge data centers typically don't host a large amount of redundant hardware due to space, power and cost concerns. In this case, a system failure may have a big impact on service availability and user experience. And because these sites are often unmanned, and take time to repair, availability of each NFVI at the edge of the network becomes even more important.

There are two complementary solutions to this:

1. Deploy more reliable carrier-grade equipment (e.g., with internal redundancy and better mean-time-before-failure characteristics), without increasing cost exorbitantly; and
2. Deploy an SDN-controlled network fabric between edge data centers such that workloads can be moved to an adjacent facility in the event of failure.

NETWORK EDGE INITIATIVES

Edge infrastructure can be largely transparent to network protocols and does not necessarily require formal specification by standards development organizations to be effective. Nevertheless, there are benefits from industry collaboration and from the economies of scale associated with an ecosystem of suppliers.

Multi-access Edge Computing (MEC)

MEC is a formally approved Industry Specification Group (ISG) within the European Telecommunications Standards Institute (ETSI). It was initially called "Mobile Edge Computing" and targeted mobile networks, but is now in the process of being re-chartered and renamed to focus on *multi-access* services for both fixed and mobile access. This re-focus is partly due to the increasing opportunity of edge computing in general, and partly because converged operators would prefer to pursue a single, generic edge computing strategy across networks. If services themselves are access-independent, it makes sense to have a cloud infrastructure that reflects this.

The ISG has been influential in building the case for edge computing on the business and technical side. It has taken a lead on identifying potential services and created a detailed architecture with interface specifications and application programming interfaces (APIs). A major development was aligning the MEC with the ETSI network functions virtualization (NFV) initiative, so that operators could deploy a common cloud networking environment in the core and edge for NFV. This will have major operational benefits related to virtualized network function (VNF) onboarding and portability to service orchestration and to resilience and load balancing. The MEC concept is being developed globally, but probably has strongest support in Europe.

Central Office Re-architected as Datacenter (CORD)

The CORD initiative is hosted by the Linux Foundation inside the On.LAB group that is also developing the open source ONOS SDN Controller. It aims to specify a full-stack edge infrastructure, including hardware, virtualization layer and management tools. As the name makes clear, the objective is to repurpose telco central offices into distributed data centers.

CORD was initially focused on virtualizing OLTs for residential fiber access, but has broadened its remit considerably to include mobile networks and enterprise services. The reference architecture combines off-the-shelf servers, white-box switches and disaggregated access technologies with open source software to provide an extensible service delivery platform.

So far, CORD has been primarily a North American initiative, with notable support from the two main U.S. wireline incumbents, AT&T and Verizon; however, there appears to be growing support from some major Asian carriers, including China Unicom and NTT Docomo, both of which are active participants.

Distributed NFV

NFV is a way to separate telecom application software from the underlying hardware. The architecture includes an infrastructure platform (NFVI) with a virtualization layer (virtualized infrastructure manager, or VIM), a management an orchestration suite (MANO) and VNFs. It is in effect the industry-standard reference architecture for software-centric service provider networks and informs the development of "cloud native" networking.

The type of platform on which VNFs are deployed, and where they are located, is dependent on several factors. In line with cloud model, many VNFs – particularly control-plane functions or those related to subscriber data management – are hosted in centralized locations where the requisite security and support is available.

However, an increasing number of functions are also being hosted closer to the edge. This is particularly the case for user-plane functions and those VNFs required for low-latency end-user services. In mobile core networks, for example, control user-plane separation (CUPS) lends itself to distributed deployment of forwarding nodes, which can optimize routing and apply user-plane traffic processing, e.g., for content optimization, policy enforcement or security.

This model is being deployed for 4G LTE and forms the basis of the next-generation 5G architecture currently under development. Another example is fixed and mobile voice over IP (VoIP) networks where session border controllers (SBC) and media gateways (MGs) are similarly suited to some level of distribution.

In the access network, there are also changes underway. Cloud RAN is driving a new edge data center model that is likely to see VNFs and end-user services hosted alongside baseband units and virtualized cloud RAN controllers. And in fixed access the CORD model discussed above provides an analog, with virtualized OLTs in GPON access networks, for example – and as with the cloud RAN case, once OLT functions are virtualized it will make sense to apply services at the same location.

MEC PoC on Advanced Service Delivery

A good example of how MEC and NFV can combine to offer network-based edge computing is the ETSI MEC proof of concept (PoC) on "Multi-Service MEC Platform for Advanced Service Delivery."

Beyond specification work, one of the major activities of the MEC ISG is to run a PoC program that aims to establish the viability of the MEC architecture for a wide range of use cases. It is specifically a multi-vendor program; each PoC must have an operator sponsor and multiple vendor participants to be approved. Results are shared with the wider industry and PoCs are often demonstrated at industry events.

The "Multi-Service MEC Platform for Advanced Service Delivery" PoC included six companies: Advantech, Brocade, Gigaspaces, Saguna, Vasona and Vodafone. The aim was to demonstrate the ability to support multiple MEC platforms and applications residing on shared and common computing infrastructure. Each application should provide a different value-add to the traffic traversing platform.

The hardware platform in this PoC was an Advantech Packetarium XLc blade server, which is a hardened, high-density platform suited to deployment and operation in distributed data center locations. The other vendors provided infrastructure software, VNFs and services. Traffic was dynamically routed, by access point name (APN) or IP address range, to the MEC platforms, where the service was applied. Services tested included data analytics, traffic optimization and service chaining.

EDGE DATA CENTER REQUIREMENTS

Scale and latency are the major differences between centralized and edge data centers and this, more than any other factor, drives operator deployment decisions. Centralized, hyper-scale data centers offer greater economies of scale, whereas a distributed data center in a central office is optimized for performance and transport efficiency. In practice, it is likely that many future services will benefit from the right balance of distributed and centralized cloud infrastructure. Through more efficient design, with the right hardware platforms, distributed data centers become more attractive to operators seeking to run internal services, and to content and application providers seeking access to end users.

Deploying Distributed Data Centers

Alongside the service strategy, there are several high-level factors that impact how operators think about a distributed data center strategy. Four of the most important are outlined in **Figure 1**.

Figure 1: Distributed Data Center Deployment Issues

Issue	Key Factors and Decision Points
Demand Forecasting	Clients pay for more resources (e.g., CPU, storage) as they need it, making it difficult to predict future demand based on past behavior. Means the operator has to deploy infrastructure in advance of customer demand.
Operations & People	Edge locations are lightly supported (not many people on site). May require additional staffing costs. Underlines need for very low maintenance equipment. Access to facilities can be an issue.
Cloud Platform	Applications and services should be portable between edge data centers and centralized data centers. Needs common deployment and management environment across centralized and edge cloud infrastructure.
Network Fabric	Networking underlay and overlays using a common SDN-controlled fabric are need to connect edge cloud locations. Features such as near-real-time migration to enable software-defined resiliency make this critical.

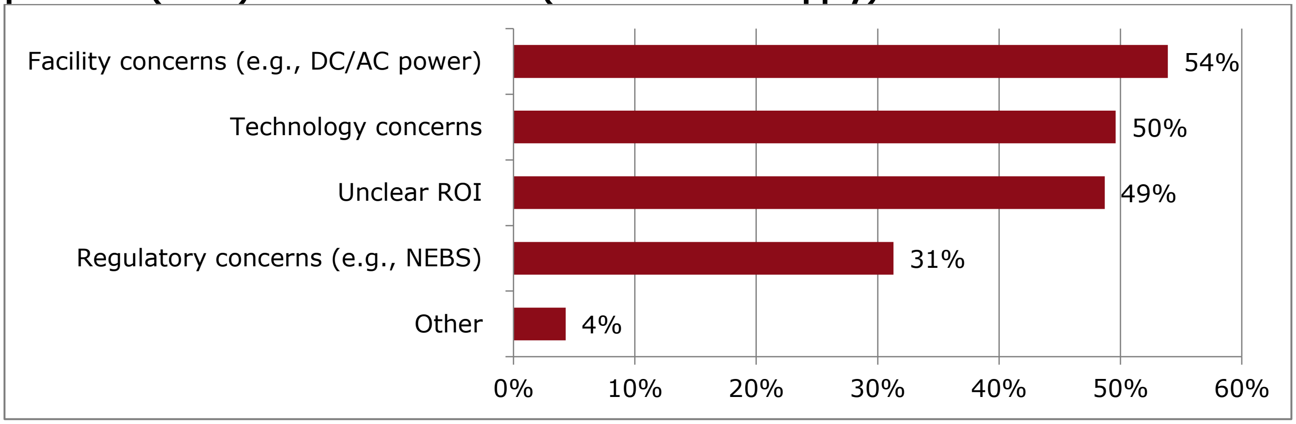
Source: Heavy Reading

These strategic technology and business issues should be assessed before embarking on a widespread deployment of edge data center, albeit informed decision making is predicated on experience of running live systems in pilot deployments. Operators with a view on how they would like to work with cloud service providers (of all kinds) and how they would like to go forward with NFV are better placed to make these decisions. The deployment of an SDN-controlled IP fabric to interconnect locations, for example, is critical for successful deployment and operation.

Edge Data Center Conversion

Distributed facilities themselves must also be converted to house data center equipment. One important question is how operators can reduce the cost of conversion, and ongoing opex, by deploying hardware optimized for these locations, yet maintaining compatibility with their centralized cloud infrastructure for application portability. **Figure 2**, from a 2016 Heavy Reading operator survey, shows how facilities concerns are the biggest challenge with converting central office locations into distributed cloud data centers.

Figure 2: What are the biggest barriers to converting central offices or points of presence (PoPs) into data centers? (choose all that apply)



Source: Heavy Reading Telco Data Center Operator Survey, June 2016, n=112

These brownfield environments, which have accumulated over years for different purposes, are generally not suitable for standard data center racks, especially in terms of power, heat and space constraints without conversion. This can be costly, depending on the type of equipment the operator would like to install. **Figure 3** outlines the major challenges.

Figure 3: Major Challenges to Data Center Conversion

Challenge	Key Factors to Conversion
Regulation	Some markets require NEBS certification in telco facilities and have specific security requirements. There may be an impact on grandfathered equipment at some locations (putting previously certified compliant locations in jeopardy).
Power Supply	Telecom gear is generally DC-powered, whereas many IT platforms require AC. Changes to power supply infrastructure is a very common issue.
Space & Size	Space can be challenging at strategically located facilities, especially to support both edge data center and traditional central office services. Need to check access – e.g., to ensure that standard rack sizes can fit.
Weight & Structural Issues	Footprint of IT infrastructure varies from regular telco gear, causing alterations to floor layouts. Facilities may have limited weight-bearing capability and require upgrading.
Fire Suppression	Fire suppression is required for AC power-based equipment since they are not NEBS. May need to deploy self-contained racks with integrated fire suppression.

Source: Heavy Reading

One of the main barriers to converting central offices into data centers is that there is no rack concept in central offices. Power and cooling per square meter, load and weight-bearing on the floor in existing buildings are, as a result, similarly challenging. This is compounded because power and cooling limitations can result in underutilization of floor space, as only a part of it can be filled with data center racks.

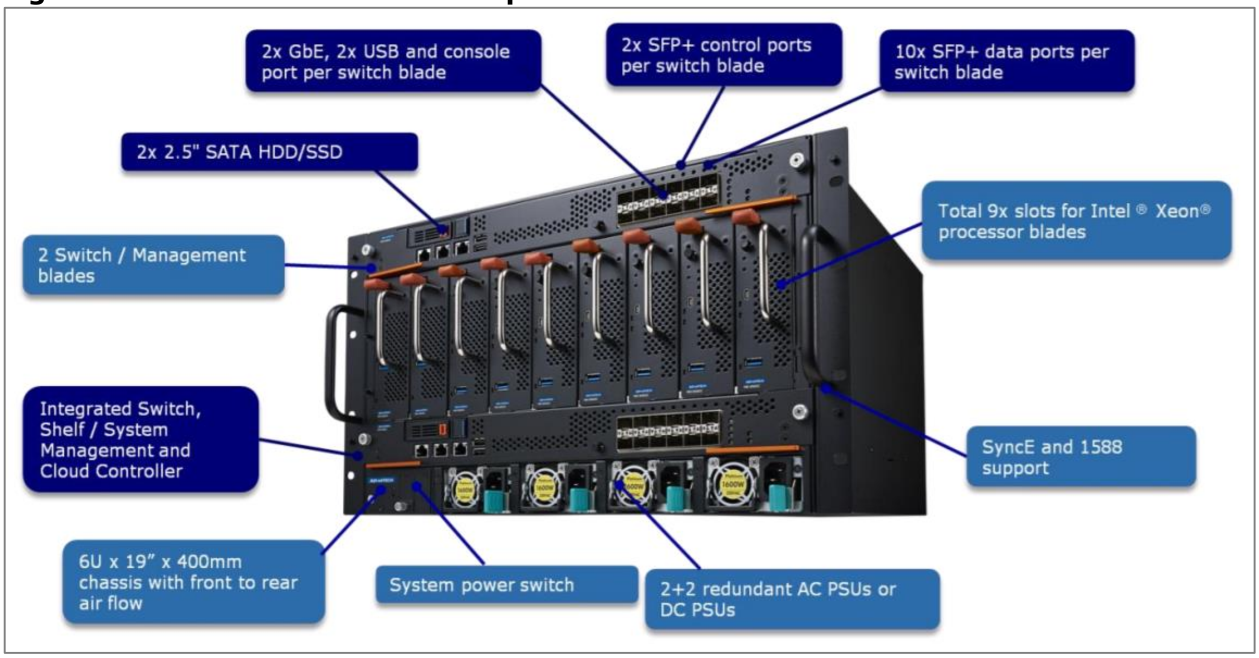
Working in the operator's favor is the fact that the most useful facilities, in high-value locations, have probably been well maintained and may not require major upgrades. Or only a portion of facility may need converting to AC power, for example. It is also worth noting that Europe and Asia do not have the same NEBS-related concerns as U.S. operators, perhaps making their transition a bit easier.

Hardware Platforms for Distributed Cloud

Given the deployment constraints and performance requirements of edge cloud services, network operators need optimized equipment to make distributed data centers sufficiently reliable and economic. As discussed, these locations are often only lightly staffed, or not staffed at all, and unlike centralized data centers, hardware must meet strict uptime requirements, be power-efficient, and on occasion support NEBS criteria. However, this infrastructure should be generic from the application or VNF perspective, so that it can form part of a network-integrated cloud.

An example platform designed to operate in data centers, central offices and telecom rooms at the edge of the network is the 6RU x86 Carrier Grade Blade Server shown in **Figure 4**. As an x86-based platform, it can be deployed as an NFVI and is capable of running the full range of NFV and MEC services.

Figure 4: Blade Server Platform Optimized for Distributed Data Centers



Source: Advantech

Key features include:

- 9 front slots to host hot-swappable Intel CPU blades
- 9 slot mid-plane providing dual star connectivity between switch boards and blades
- 2 x ESP-9002C Switch/Management Modules
- System input/output (I/O) per switch blade: 10 x 10G data fabric interfaces
- Up to 2 x 2.5" hot-swappable HDD trays for SATA disks/SSD per switch blade as integrated system NAS (optional)
- Power density, efficiency and cooling: The 6RU product above offers 400W per RU power footprint and can be deployed in industry standard 19" racks
- Designed to comply with NEBS Level 3

ABOUT ADVANTECH

Advantech Networks & Communications Group provides the industry's broadest range of communications infrastructure platforms, scaling from one to hundreds of Intel cores, consolidating workloads onto a single platform architecture and code base. The group's technology leadership stems from its x86 design expertise combined with high-performance switching, hardware acceleration and innovative offload techniques.

For the new IP infrastructure, Advantech's NFV Elasticity framework extends NFV to the network edge by supporting scalable carrier-grade platforms that run VNFs anywhere in the network. Operators, integrators and software-vendors can then rapidly validate the latest NFVI for applications, such as vE-CPE, SD-WAN, MEC, Fog and IoT, and benchmark both applications and VNFs using Advantech's Remote Evaluation Service. For more information, visit www.advantech.com/nc.