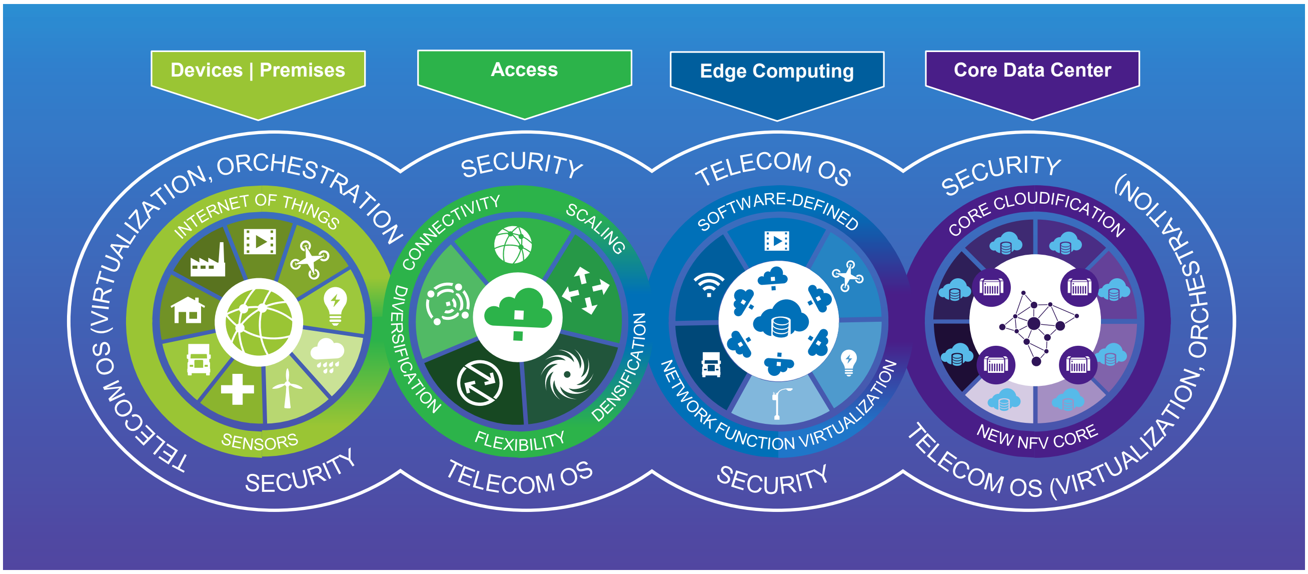
White Paper

Hyperscale and 5G: The Future is Now

By 2025, the digital universe is expected to reach 175 zettabytes, which amounts to 150 times more bytes of data than there are stars in the observable universe. The numbers are staggering, but the hyperscale ecosystem is rising to the challenges in meeting the insatiable demand for memory, bandwidth, computing power, storage, and speed. At the same time, cloudification is blurring the proverbial lines between networks and applications, while 5G technologies are pushing more intelligence to the edge of the network. The hyperscale ecosystem and the breakthrough technologies accompanying it are turning these unprecedented challenges into opportunities for operators worldwide.

Hyperscale Overview

The International Data Corporation (IDC) has defined a hyperscale deployment as five thousand or more servers on a ten thousand ft2 or larger footprint. These proportional requirements belie the true definition of “hyperscale” as the ability to scale rapidly in response to demand. This includes horizontal scaling through the addition of hardware/square footage or vertical scaling through improved bandwidth and efficiency using existing hardware. Data Center Interconnects (DCIs) utilizing 400G or 800G Ethernet connectivity link these massive data centers to one another as well as intelligent edge computing centers. Distributed edge computing reduces system latency by bringing user plane applications and network functionality closer to the use case.



Network Cloudification: Distributed, Disaggregated, Native Cloud Based, and fully Automated

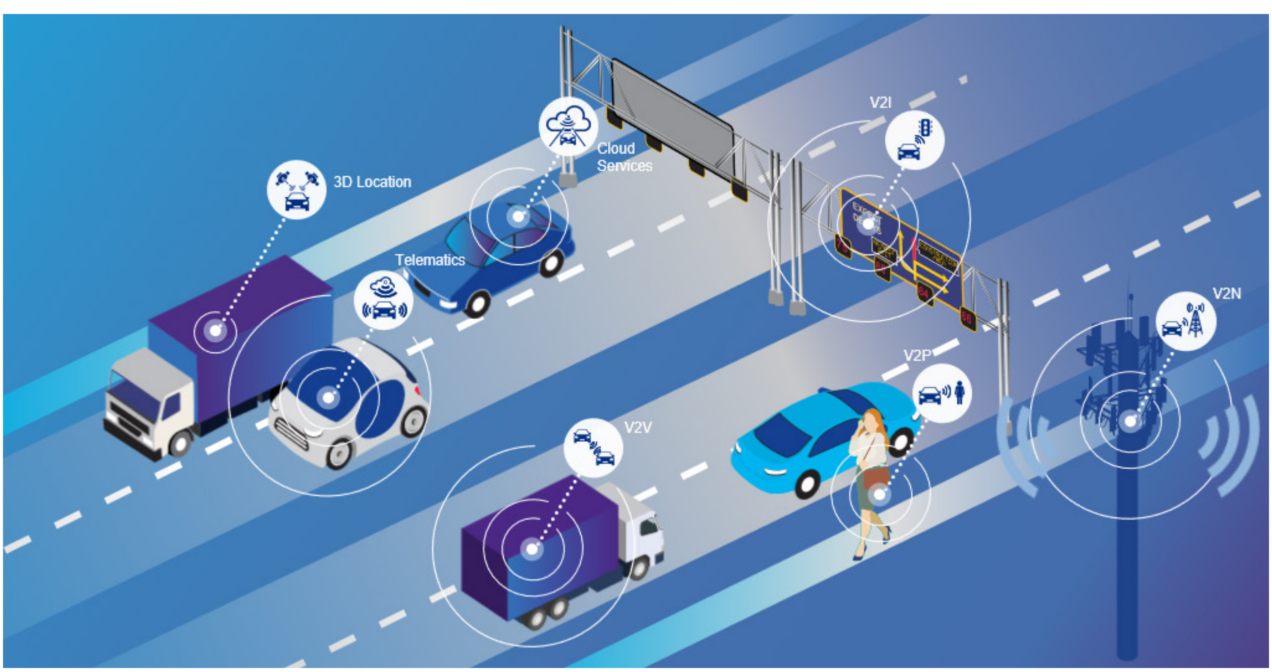
Internet content providers (ICPs), public cloud deployments, and big data storage solutions have become synonymous with hyperscale, with a focus on content driving traditionally rapid deployment timeframes and finite data center lifecycles. The telecom industry has favored a long-game approach to data center deployment, with meticulous planning and comprehensive pre-deployment testing leveraged to achieve five nines (99.999%) availability out of the gate, along with a data center lifespan of ≥ 10 years. The rollout of 5G is placing unprecedented strain on data center infrastructure worldwide, with the introduction of new services, IoT verticals, and the intelligent edge that enables them. Increased end-to-end network complexity is challenging deployment schedules while raising the bar for performance, efficiency, and reliability.

Hyperscale Use Cases

The primary 5G use case categories of Ultra Reliable Low Latency Communication (URLLC), Enhanced Mobile Broadband (eMBB), and Massive Machine Type Communication (mMTC) frame a variety of new verticals with unique requirements and performance expectations for latency, densification, bandwidth, speed, and other essential characteristics. Despite the requisite cloudification and redistribution of intelligence from the core to the edge, hyperscale data centers will continue to play a pivotal role in the scalability, SLA conformance, and network efficiency that supports these verticals.

Advanced Driver-Assistance Systems (ADAS)

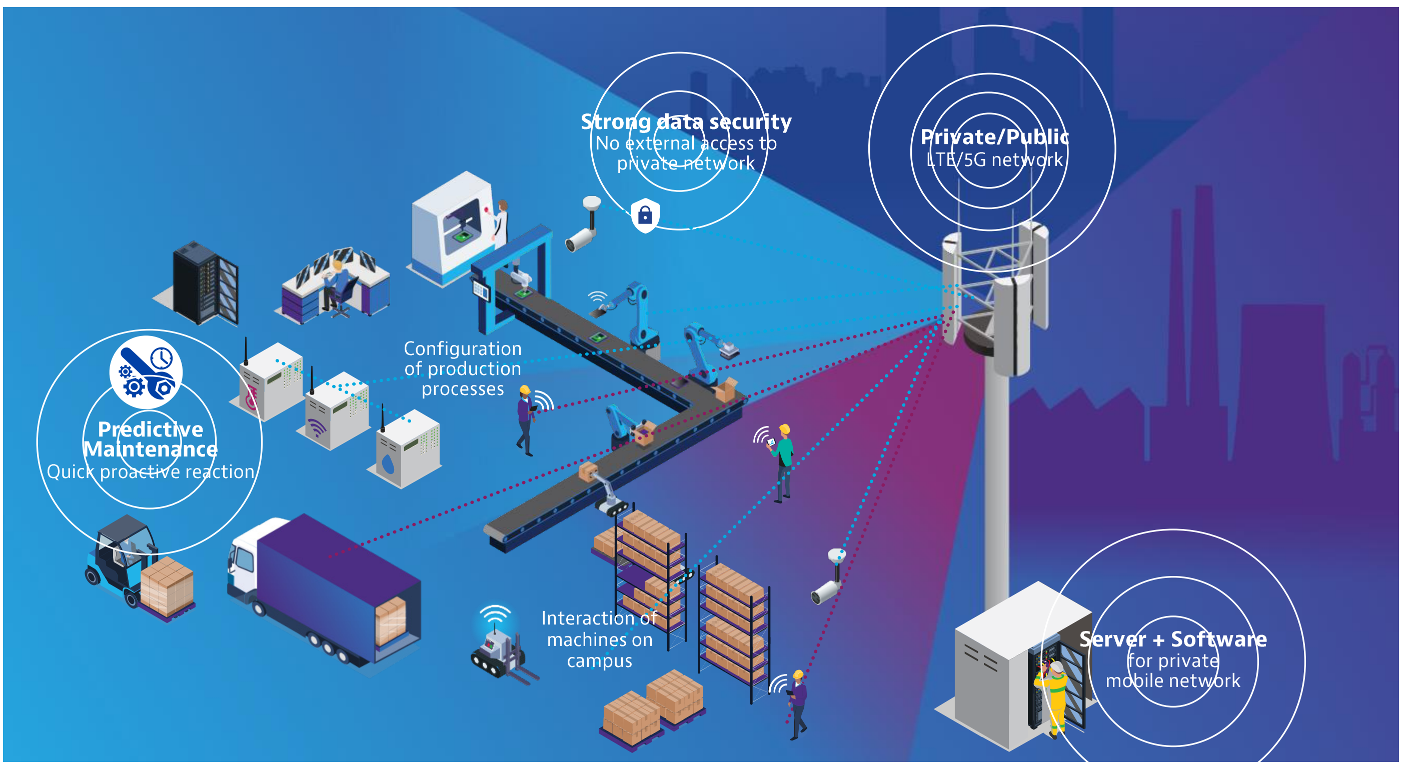
ADAS applications have progressed from vehicle-specific automatic braking, blind spot detection, and collision avoidance systems to encompass vehicle-to-vehicle, vehicle-to-network, vehicle-to-pedestrian, and vehicle-to-infrastructure communication, along with supporting elements such as telematics for infotainment, traffic control infrastructure, and 3D geolocation. Each element of the highly anticipated conglomeration of ULLC applications known as "Vehicle to Everything" (V2E) requires a well-orchestrated network slice flowing from Core to RAN. In general, the autonomous edge supports low latency ADAS requirements while the Core provides support for artificial intelligence (AI) rules, software updates, and traffic conditions as patterns develop and responses are optimized. The emergence of ADAS and connected cars highlights the importance of system latency in the 1-2ms range as a gating requirement for performance and driver safety.



V2E leaves little margin for error and requires a very well-orchestrated network slice

Factory Automation

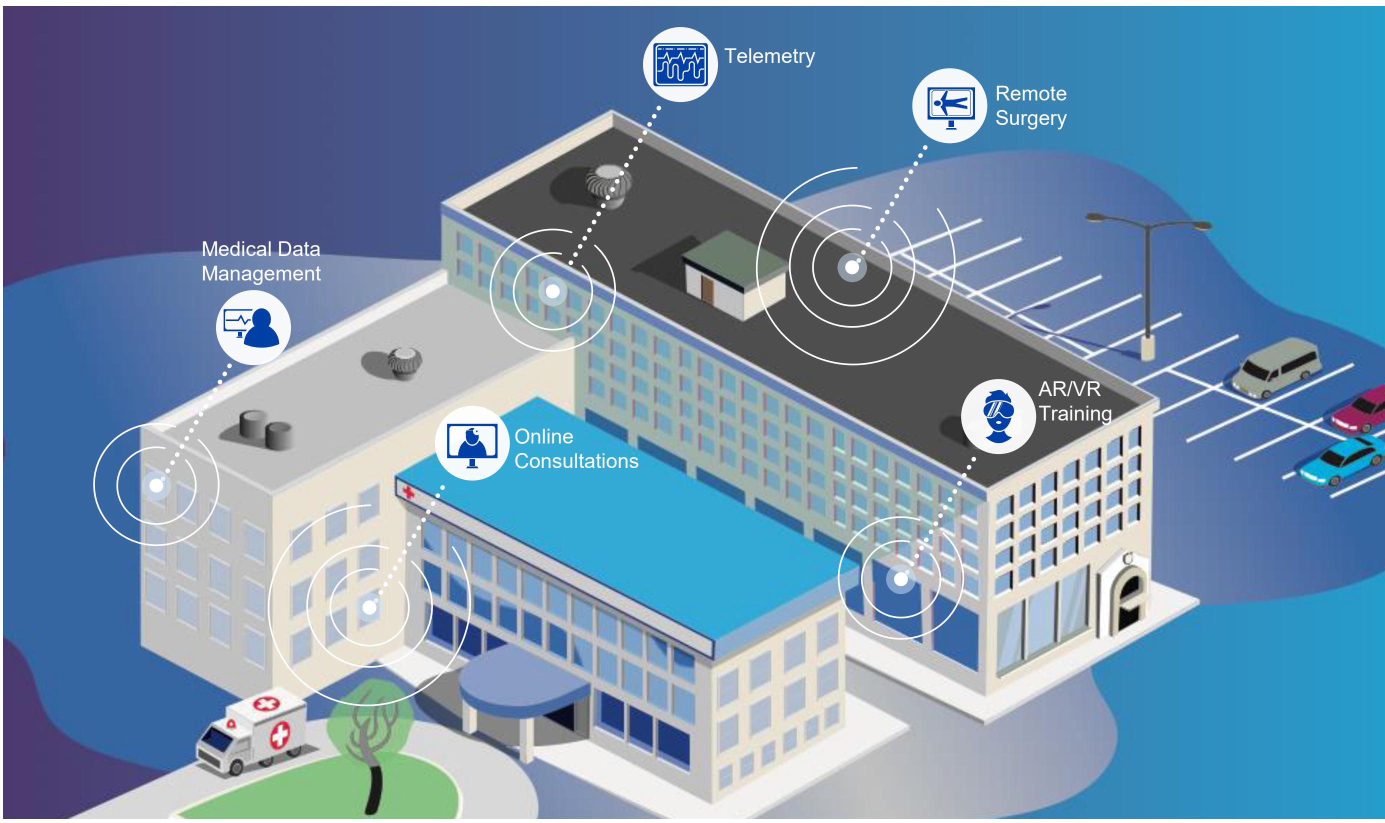
The manufacturing sector has recognized the potential of private edge cloud computing to provide a responsive central nervous system connecting breakthrough robotics, material handling, and predictive maintenance applications. This market has expanded as 4G transitions to 5G and private companies leverage URLLC to enable next-generation factory automation or "Industry 4.0". High bandwidth is required to support smart factory data rates from 1-20 Gb/sec. Densification of real-time IoT sensors can number in the millions for a single deployment. Hand tools as elemental as screwdrivers are now being converted into intelligent devices, continuously streaming torque, position, and calibration data back through the network. Data center software plays a pivotal role in coordinating and analyzing IoT sensor data and optimizing predictive maintenance trigger points.



Factory automation increasingly includes public/private 5G components

Connected Health

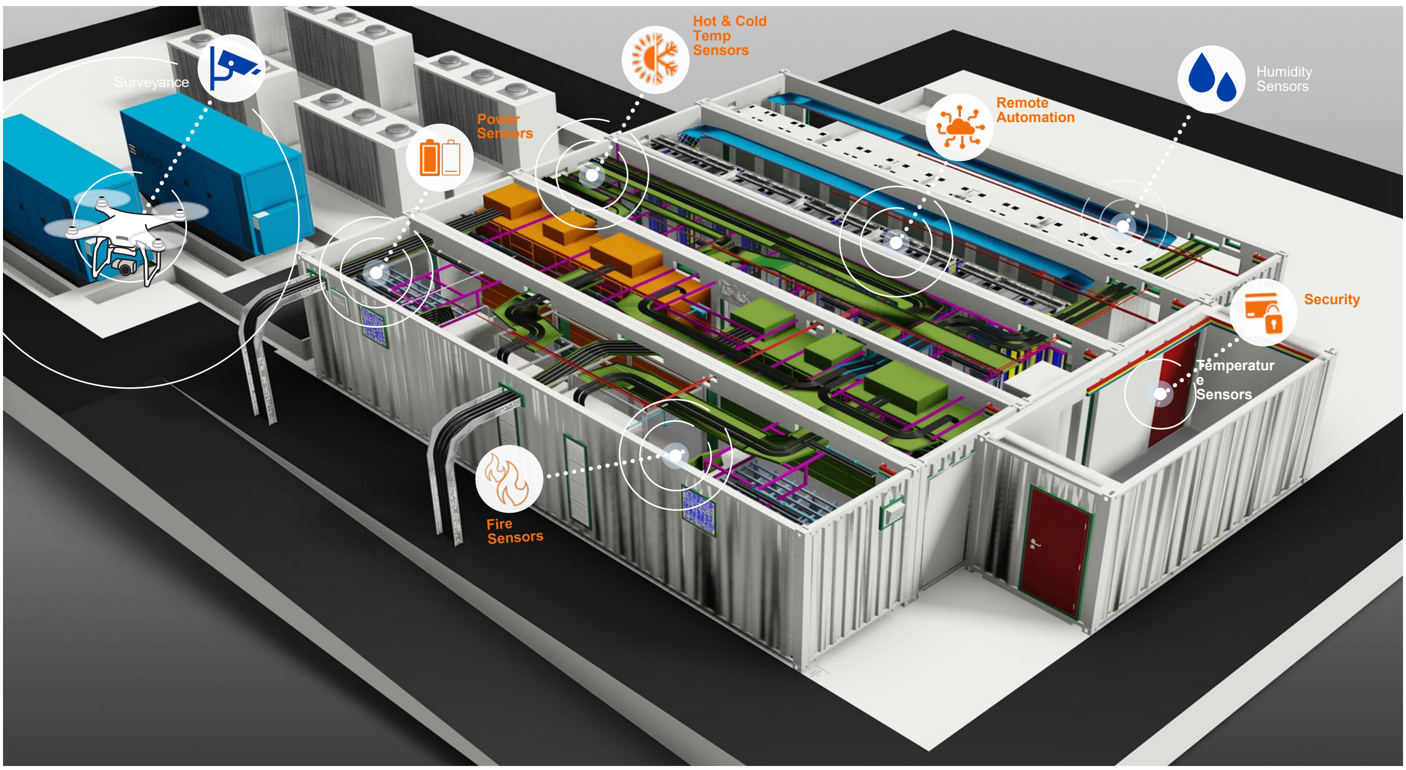
The healthcare industry presents an equally multi-faceted set of verticals encompassing online consultations, remote surgeries, AV/AR training for nurses, telemetry, and data management applications, each with disparate latency, bandwidth, privacy, and mobility requirements. While routine health checkups and record transfers may be more forgiving with respect to latency and reliability, cutting edge remote surgical procedures underscore the need for end-to-end network monitoring and assurance. The coordinated data analytics, AI, and machine learning (ML) capabilities of the hyperscale data center and edge computing locations must flawlessly address these flexible requirements to ensure patient safety and mitigate healthcare provider liability.



Connected health environments interested in patient safety and privacy will not tolerate latency

Unmanned Data Centers

The next generation of data centers, particularly at the edge, will be overwhelmingly lights out (unmanned). This new reality has accelerated the development of the data center IoT vertical, with its own unique and dedicated network slice. The data-center-as-use-case leverages the same real time 5G IoT sensing and remote automation that is redefining manufacturing plants, smart ports, retail centers, and stadium venues. Robots or drones can effectively perform important surveyance tasks within the center, while ULLC can be used to queue automatic links to service dispatch. Much like smart port sensors within shipping containers, strategically deployed temperature and humidity sensors can feed back important environmental data to automate and expedite hardware and HVAC adjustments.



Data centers are moving to a "lights out" and proprietary IoT model

Hyperscale Industry Challenges

Among the more valued and transferable resources on Earth, data bears the dual distinction of being both infinite and man-made. The infinite nature of data is contrasted by the finite resources required to create, process, transport, and store it. With data centers already consuming approximately 3% of the world's electricity and contributing 2% of the total greenhouse gas emissions, conservation and efficiency have moved to the forefront. Human labor and expertise are additional finite resources that must be continually available to operate, maintain, and troubleshoot data center hardware, software, fiber, and facilities.

The resource constraints associated with hyperscale computing underscore the challenges of scalability, as 5G emerges and the demand for telecom, ICP, and big data storage applications rise exponentially. Continuing to scale horizontally has proven to be physically and environmentally unsustainable. Each data center deployment or expansion also requires a proportional increase in fiber count, introducing additional exposure to fiber damage, contamination, and vandalism which can lead to service disruption, steep SLA violation penalties, and costly repairs. Fiber installers spend up to 20% of their time troubleshooting, with DCI issues being notorious for high mean-time-to-repair (MTTR) and recovery costs.

Increased demand has also intensified deployment schedule pressures, with fiber, RAN, and RF verification testing sometimes truncated or omitted to meet aggressive delivery timelines. Disaggregated networks are assumed to be more resilient by nature, with work arounds available for system failures. However, the cumulative cost of compromised network operation, troubleshooting, and resolving issues on a live network can be significantly higher than more proactive pre-deployment testing strategies. DCI bandwidth is receiving a much-needed boost from 400G and 800G Ethernet connectivity enabled through higher order modulation. This transition also heightens the importance of standards-based Y.1564 Ethernet service activation, bit-error rate testing, and network slice transport assurance and optimization. Test access and visibility can be impeded by the growing trend towards lights out data centers.

Some of the more arduous hyperscale challenges for the years ahead will be driven by the intricacy of distributed, disaggregated, and cloudified 5G networks. As the triple constraint theory assures, data center costs will continue to escalate so long as scale and complexity continue along their current trajectories. Virtualized RAN, Massive MIMO, and antenna beamforming further complicate RF and network performance testing, introducing new spectrum analysis, demodulation, and SLA conformance challenges. To overcome these pre-determined pitfalls and ensure the full promise of 5G, end-to-end network slicing must be orchestrated seamlessly for each unique vertical. This immense challenge resists any "swivel-chair" or siloed approaches to network management. Outdated modes of data center and network testing and assurance run contrary to the objectives of fully automated and programmable network slicing and edge computing. Critical 5G IoT use cases leave no margin for error with respect to SLA conformance and reliability.

Addressing the Hyperscale Pain Points

Historically, many of the greatest technical challenges have been resolved through a combination of innovation, out-of-the-box thinking, and cooperation. Although the challenges presented by hyperscale computing do not represent existential threats, they do require a similar melding of tactics to responsibly ensure the sustainability of hyperscale computing and the promise of 5G for generations to come.

Efficiency

An experimental Microsoft data center was retrieved off the coast of Scotland in the Summer of 2020 after being submerged for over two years. This experiment provided evidence of vastly improved cooling efficiency and reduced component corrosion for underwater data center deployments. The Climate Neutral Data Centre Pact in Europe takes an equally ambitious stance, with a stated goal of data center carbon neutrality by 2030. While ongoing investigations into green energy conversion and aspirational seed change experiments are needed, the current hyperscale ecosystem also presents immediate opportunities to incrementally improve efficiency and reduce consumption.

- Remote data center monitoring reduces the carbon footprint associated with human transportation and troubleshooting and expedites problem resolution by automatically detecting inefficiencies and malfunctions. Incorporation of artificial intelligence for server and optical power monitoring adroitly scales back power consumption when equipment is idle or running at lower capacity levels.
- The ongoing transition to network function virtualization (NFV) and software defined networks (SDN) continues to unseat active system electronics and their inherent energy consumption. This paradigm shift also enables greater decentralization of cloud networks. The elasticity of distributed clouds also brings improved efficiency and profitability.
- Real-time IoT sensing enabled by 5G hyperscale deployments makes lights out operation more efficient and reliable. Reduced reliance on humans for data center monitoring, troubleshooting, and repairs also allows for greater flexibility in hyperscale data center geographies. Colder, more remote climates that provide substantial cooling efficiency benefits can now be considered viable locations.

Proactive Testing

Despite the frequent mischaracterization of testing as a driver of additional time and resources, data center testing can improve efficiency, scalability, and schedule performance. New test solution categories and capabilities provide the requisite speed, intelligence, and accuracy needed to address the scale and complexity of hyperscale data centers and 5G uses cases with dramatically improved testing ROI. For example, comprehensive pre-deployment fiber testing and automated fiber monitoring and diagnostics introduce minimal installation schedule impact while significantly reducing unplanned service degradation instances post-deployment. MPO-based multi-fiber testing utilizes intuitive cloud-based workflows to address the speed and accuracy requirements of hyperscale installations. Network traffic emulation and active testing are additional proactive test capabilities used to optimize quality and reliability during design and deployment phases. Variable traffic loads can be used to “stress test” the network and safeguard against potential real-world failures. Similar pre-deployment testing for new verticals involving 5G would include RF testing and certification, RAN transport or Xhaul connection calibration and verification, along with application and slicing analytics and emulation. This emphasis on upstream testing and validation improves time to market and out-of-box performance, while reducing unplanned outages, troubleshooting, and updates after deployment.

Automation

Automation is a fundamental output of hyperscale 5G that must also be utilized across the hyperscale ecosystem to mitigate the complexity of virtualized, decentralized, and cloudified networks with high expectations for precise, real-time responsiveness. Automation must seamlessly extend from system performance to system test and troubleshooting. Data analytics, AI, and ML drive the orchestration needed to build optimized network slices deploying App functions. Cloud-enabled test automation can be used to perform high-performance throughput and BER testing between data centers and verify network slice integrity from end to end. Self-healing is a key attribute of autonomous networks, as advanced analytics and AI provide the requisite neural pathways and cognitive insights needed to self-correct without human intervention. Intelligence residing in the data center also facilitates self-adjusting and self-optimizing RAN to instantaneously adjust cell bandwidth, location, tilt, and azimuth.

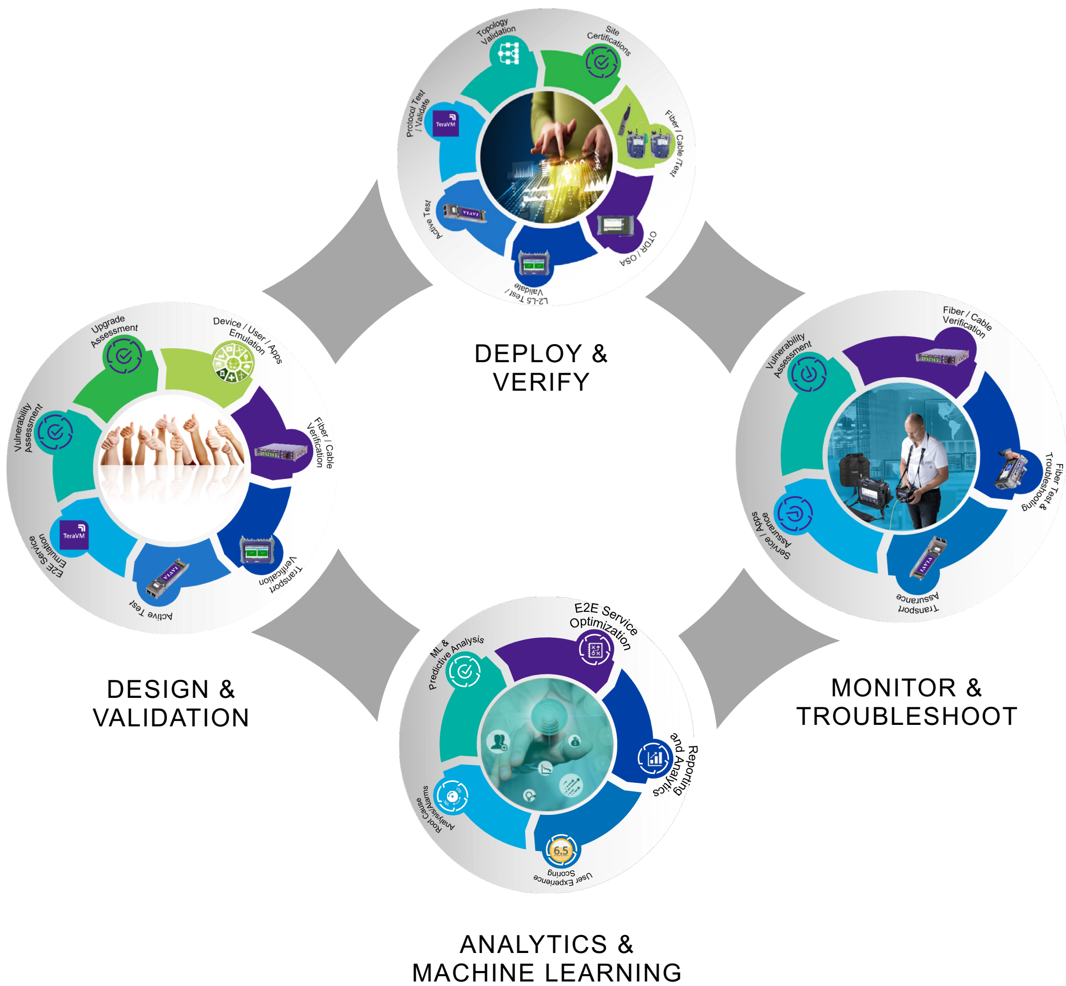
Open RAN

Open Radio Access Network (O-RAN) architecture, empowered by the principles of intelligence and openness, is quickly becoming the new frontier for data centers. Traditionally, RAN deployments were closed and proprietary, relying on a single vendor for a complete RAN solution. RAN architecture has now shifted towards 1st level disaggregation, with more virtualization of RAN elements and some disaggregation, but still proprietary with no interoperability. Open RAN calls for complete disaggregation, with multi-vendor interoperability, open APIs, and intelligence embedded in every layer of the RAN.

This distribution of intelligence from Core to Edge to RAN also provides unprecedented opportunities for ultra-low latency verticals. The data center becomes a more active part of the network and services themselves as the computing environment moves closer to the user and a self-managed, zero-touch network from end-to-end becomes feasible. Open RAN will become cloudified as more automation, machine learning, and interoperable NFV modules are introduced, furthering the transformation of the 5G network into a “cloud of clouds”. Perhaps most importantly, O-RAN is setting new precedents for cooperation, interoperability, and reduced development lifecycles that resonate throughout the 5G network and hyperscale ecosystems.

Optimization Solutions for the Next Generation of Hyperscale

VIAVI offers fully integrated and interoperable solutions for network and data center testing, performance optimization, and service assurance. Designed to support the most complex networks and services of today and tomorrow, our solutions help you get the best possible performance from your investments. We offer network and service test automation solutions spanning mobile, wireline, and cloud-based environments through standardized operation, optimization and troubleshooting as well as emerging solutions that leverage dynamic, machine-learning orchestration and support.



VIAVI provides the multi-dimensional visibility, intelligence, and insight you need to efficiently manage physical and virtual environments and profitably deliver optimum service levels, transition to new technologies, and launch innovative services

There's a reason that more than 4000 global customers turn to VIAVI to help them address their toughest network, application, and service challenges. With technical mastery in networking and deep knowledge of what it takes to deliver leading-edge communications and IT services, VIAVI makes it easier to stay ahead of the accelerating pace of technology change, compete more effectively in the market, and deliver greater value to your customers.

VIAVI Solutions

Contact Us **+1 844 GO VIAVI**
(+1 844 468 4284)

To reach the VIAVI office nearest you,
visit viavisolutions.com/contact

hyperscale-and-5G-wp-xpf-nse-ae
30193218.900.1021