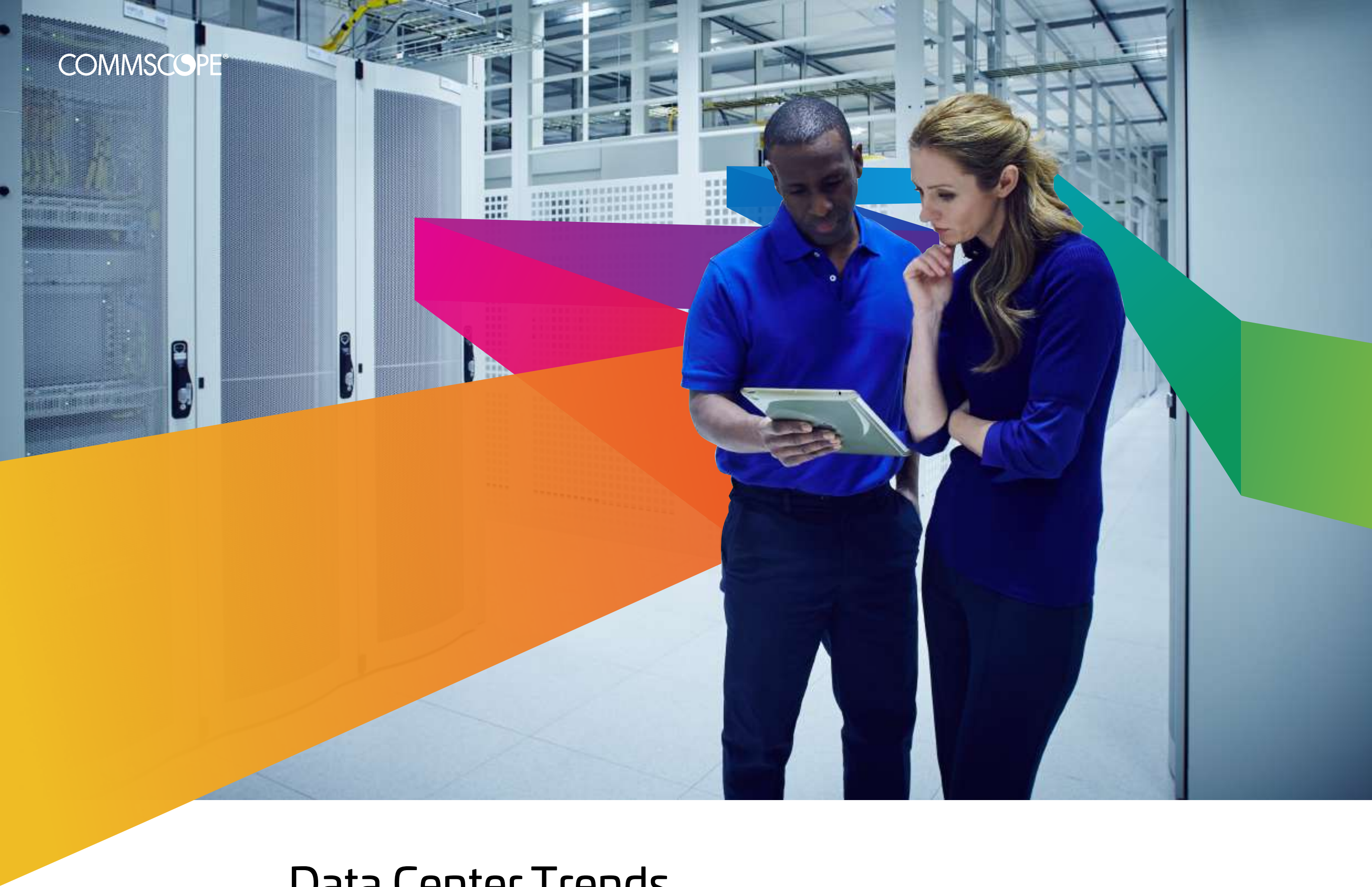


Data Center Trends

Contents

Authors

James Young

Director of the Data Center division for Enterprise Solutions at CommScope, overseeing the strategy and providing leadership to technical and sales teams worldwide. James has acquired extensive experience in the sale of OEM products, network solutions and value-added services through sales environments of direct and indirect channels. He is Communication Distribution Designer (RCDD) and a certified data center design professional (CDCP).

Kam Patel

Director of Hyperscale and Service Provider Data Center Solutions. Kam has been with CommScope for more than 25 years in a variety of Business Development, Engineering, Marketing, Operations, Product Management and Strategy roles and holds more than 20 patents for network equipment.

Jason Bautista

As Solution Architect for Hyperscale and Multi-Tenant Data Centers, Jason is responsible for data center market development for CommScope. He is monitors trends in the data center market to help drive product roadmap strategy, solutions and programs for Hyperscale and Multi-Tenant data center customers.

Ken Hall

Data Center Solutions Architect for Cloud-Scale/Hyperscale Solutions. Ken has been with CommScope for more than 30 years in a variety of Global Program, Project Management, Strategy, Business Development, Marketing and Technical Sales roles.

Ken has extensive experience with Network OEM equipment efficiency (board level through structured cabling) and has 9 patents to date for fiber optics and cable management solutions. He is a Registered Communication Distribution Designer (RCDD) and Network Technology Specialist (NTS).

Pedro Torres

Pedro Torres is director of technical sales, Europe and CALA, for CommScope Mobility Solutions. He previously worked for Ericsson as RAN CTO for the Telefonica Global account. He had prior roles in R&D, system engineering, and IP and RAN solution sales. He holds a Master of Science degree in telecommunications engineering from Madrid Polytechnic University and an executive leadership program at UNC.

The Year Ahead: What's Impacting the Data Center?

There's no such thing as business as usual in the data center. As data center professionals you must always be one step ahead of the curve to meet demands for ubiquitous connectivity, lower latency, and reliability. With global IP traffic reaching 5 Zettabytes and customer's demand for connectivity placing huge pressure on the network - what challenges will data centers face in 2020 and beyond?

We've pulled together insight from data center experts to paint a picture of what topics we can expect to hear more about in the coming year. Our experts examine how technological trends such as the 5G, Internet of Things, 400G and increasing cloud usage are impacting data center design and strategy and what that could mean for you.

At CommScope we understand your needs for faster deployments, global quality and efficient scalability – let us help you build the networks of tomorrow.

1 Leveraging AI in the Data Center: Challenges and Opportunities

James Young

The potential impact of artificial intelligence on the data center has been called '*game-changing*' and '*limitless*'.

While it may be too early to bestow such accolades, there is demonstrative proof that AI can provide significant benefit in terms of enhancing data center operations.

What we know thus far is that hyperscale data centers are ramping up their processes to support the development and use of AI, as well as machine learning, a subset of AI. Hyperboles aside, the technology is expected to optimize a wide variety of future data center operations including planning and design, workload management, ensuring uptime, and controlling costs.

By 2022, IDC predicts that 50 percent of IT assets in data centers will be able to run autonomously because of embedded AI functionality.

In addition to executing complex processes with ultra-low latency, AI and machine learning are critical to automating routine maintenance and monitoring in order to free skilled data center professionals to focus on optimization and planning. According to Gartner, more than 30 percent of data centers that don't deploy AI and machine learning won't be operationally and economically feasible by 2020.

AI will also help organizations stay ahead of increasing data storage and processing requirements. Other important applications in the data center will include:

- Improving security
- Conserving energy
- Reducing downtime
- Optimizing server workloads
- Monitoring equipment

These applications differ significantly from traditional hyperscale workloads [AS1] like search for example. Optimizing AI processes to support highly-parallel processes and feeding them vast amounts of data will require re-designing architectures used to connect storage, compute, switching and physical network elements. What, exactly, should data center managers be planning for?

> More than 30 percent of data centers that don't deploy AI and machine learning won't be operationally and economically feasible by 2020
>
> — Data Center Knowledge, May 2019

Designing the new AI processes will involve servers using specifically adapted accelerators that can process complex analytics on a vast amount of raw data. At the same time, ultra-low latency demands will require a flattening in the network layers. By eliminating top-of-rack switches in the server network in favor of next generation switches with the capacity to connect many racks of servers, data centers will be able to reduce network latency while network capacity continues ramping up. These new switches will provide much higher server connection speeds while lowering the number of switches required saving capital and most importantly lowering power and operating expenses.

Supporting the new AI-enabled designs poses a variety of challenges. Building the AI models that will be needed requires servers capable of processing vast amounts of data. This in turn, requires links capable of feeding the servers at higher and higher lane rates. Network connection speeds are climbing; 50G is common today and 100G is set to follow very quickly. Fueled by chipsets like the recently introduced Tomahawk 4, new generations of switches will directly support a much larger number of servers. The result will be networks and compute/storage that is optimized to deliver AI insights with the lowest time and power consumption.

In response to requests from hyperscale operators the IEEE has recently completed work on new 400G transceiver standards that will provide off server connections based on 50G Ethernet (set to publish early 2020). Each new 400G SR8 transceiver will support eight server connections. These low cost and low power VCSEL-based transceivers use multi-mode structured cabling and can combine with 12.8T switches, scaling to support hundreds of servers per switch. These new switches and transceivers are already being deployed in hyperscale data centers today.

Work currently underway is projected to double the server connection speed to 100G within a couple of years. The IEEE recently formed a study group to explore low cost in-row server networks. Transceiver manufacturers II-VI and Broadcom have recently announced 100G plans to introduce VCSELs to support these networks targeted for market release in 2021, just in time to support the volume deployment of 100G severs.

Structured optical cabling networks are being deployed to manage the 16 fibers that the transceivers will use to connect eight servers. These networks will provide better support for the density that 12.8T switches provide, reducing deployment times and paving the way for future 100G networks that will follow in a few short years.

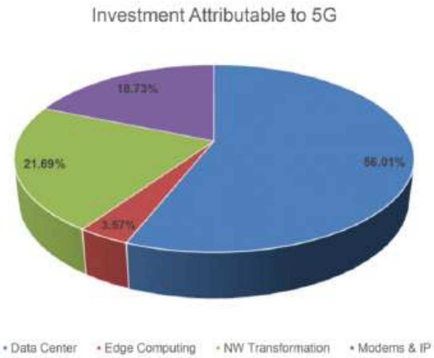
Ultimately, the efforts to enable greater use of AI in the data center will likely lead to much improved latency performance, lower capital costs and, very importantly, a reduction in the amount of power required to operate high capacity networks. To realize these benefits, however, data centers must adapt their cabling plant. This involves delivering more fiber from the core to the switch while keeping the inside cabling plant manageable.

2 | 5G and the role of the data center


James Young

Pedro Torres

To ask how 5G will affect life inside the data center is analogous to asking how a city would stand up to a natural disaster. The answer is, it depends—on the city and the storm. Make no mistake, 5G will most definitely alter how data centers are designed and, in some cases, will change the role they play in the larger network. By some estimates, data centers will be spending over half their operating budget to support 5G by 2025. The $64,000 question is, “where will the money be spent and for what?” To be more specific means digging a bit deeper. Grab your shovel.



5G: Promises and problems

Let's start with what we know. From a technical perspective, 5G will have several defining characteristics. Most obvious is the use of the new 5G New Radio (NR) air interface that will exploit new spectrum and provide latency capabilities in the single-digit millisecond range.

The enhanced performance will drive deployment of billions of edge-based connected devices and create the need for flexible user-centric networks. To deal with the crush of new data, networks will rely heavily on virtualized architectures like network slicing, and other cloud-based technologies such as AI and machine-type learning.

The wild card in all of this are the applications that 5G will enable. Self-driving vehicles, industrial automation, facial recognition, machine-to-machine communications and a mind-numbing range of smart city apps—you name it and it's in somebody's pipeline. Problem is, they have a wide variety of requirements regarding latency, reliability and the volume/type of data traffic generated. Unless you know the parameters of the problem, it's tough to speculate on how it will affect the data center. We can, however, provide some insight based on CommScope's experience and perspective. Here are a couple of key trends we see affecting the data center ecosystem, and a question that may be at the heart of the matter.

Moving compute and storage to the edge

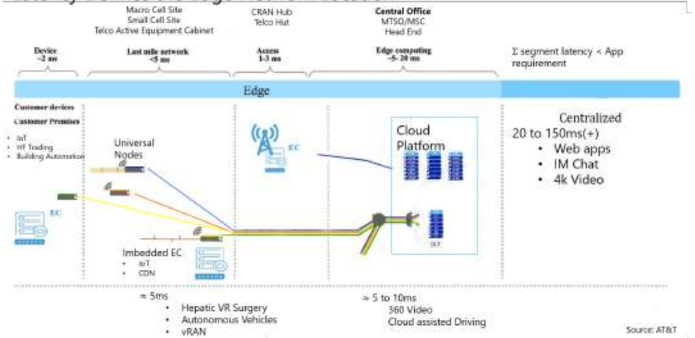
User-centric networks involve pushing the compute and storage resources closer to the users and connected devices. The only way to meet the ultra-reliable, low latency requirements will be to deploy these edge nodes as mesh networks with an east-to-west flow and parallel data paths. In some cases, these nodes may be big enough to classify as pod-type data centers or micro data centers, similar to those being used by the telecom and cable providers. The bigger question is what effect this will have on the core data center. Again, it depends.

Cloud-scale data centers as well as the larger enterprise facilities may be only slightly affected, as they are already using distributed processing and are designed to handle the increased data flow from the edge.

A bit more disruption may occur among the retail multi-tenant data centers (MTDCs), which have traditionally grown in response to rising demand for cloud-scale services. In order to maintain that relationship, look for retail MTDCs to re-locate closer to the edge in order to provide regional points of presence for the cloud-scale facilities.

Perhaps the biggest changes will occur among the service providers as they refine the relationship between their core data centers and the evolving central offices and centralized RAN (CRAN) hubs. Increasing virtualization of the core networks and the radio-access networks is necessary to handle the anticipated 5G data flow and enable service providers to easily move compute and storage capacity where it is needed.

Latency Defines the Edge Network Location



The effects of virtualization AI and machine learning

Another potential disrupter will be the increasing use of virtualization, artificial intelligence (AI) and massive machine learning. These technologies will require accelerated server speed and higher network capacity to enable greater volume of increasingly sophisticated edge services. Building the data models will require processing massive data pools which, in most cases, are best matched to core data center capabilities.

Most of the data that will be used to develop AI models will come from the edge. This hints at a potential change in how the larger cloud-scale data centers will support the network. One scenario involves using the horsepower of the core data center to assemble data from the edge to develop the models. The finished models would then be pushed out to deliver localized low-latency services. The process would then be repeated creating a feedback loop that refines the operating model.

In the service provider networks, increasing virtualization may have a more direct effect in the core data center. As wireless and wireline networks become more virtualized and converged, the business case for a single physical layer infrastructure grows stronger.

The big question is the degree of convergence that will occur between the core network and the RAN and where this will happen. Central office or data center?

Prioritizing and balancing the data load

Of course, the degree of change within the various data center environments will depend on the application requirements. The data traffic generated by the billions of sensors and devices may produce a steady stream of data while others will be delivered intermittently or produced in irregular bursts of information. How can the data collection and processing be optimized? How much data should remain local to the edge device or edge node versus needing to be processed in the core data center?

Once these questions are answered, network engineers need to determine the best way to move the data through the network. Different latency and reliability requirements require the ability to prioritize data traffic at a very granular level. What can be off-loaded onto the internet via local Wi-Fi versus having to be backhauled to the cloud service provider (CSP) data center? And remember, edge networks must fit into a financial model that makes it profitable.

Stay tuned for the answers

Even as the first 5G networks services are beginning to go live, these questions are still very much up in the air. From the data center manager to the CIO, learning how to mine the capabilities of 5G will require a lot of on-the-job training. The avalanche of new data from the network edge necessitates high compute and storage horsepower. But exactly how much horsepower will data centers will need and where will they need to be located? You guessed it: that depends.

One thing I can promise, though, is that CommScope will be among the first to understand what changes are coming and how it affects data centers of all types. After all, it's our job to know what's next. Stay tuned…

3 | Data centers must adapt to higher and higher fiber counts


Jason Bautista

Ken Hall

The volume of digital traffic pouring into the data center continues to climb; meanwhile, a new generation of applications driven by advancements like 5G, AI and machine-to-machine communications is driving latency requirements into the single millisecond range. These and other trends are converging in the data center's infrastructure, forcing network managers to rethink how they can stay a step ahead of the changes.

Traditionally, networks have had four main levers with which to meet increasing demands for lower latency and increased traffic.

- Reduce signal loss in the link
- Shorten the link distance
- Accelerate the signal speed
- Increase the size of the pipe

While data centers are using all four approaches at some level, the focus - especially at the hyperscale level - is now on increasing the amount of fiber.

Historically, the core network cabling contained 24, 72, 144 or 288 fibers. At these levels, data centers could manageably run discrete fibers between the backbone and switches or servers, then use cable assemblies to break them out for efficient installation. Today, fiber cables are deployed with as many as 20 times more fiber strands — in the range of 1,728-, 3,456- or 6,912 fibers per cable.

Higher fiber count combined with compact cable construction is especially useful when interconnecting data centers (DCI). DCI trunk cabling with 3,000+ fibers is common for connecting two hyperscale facilities, and operators are planning to doubling that design capacity in the near future. Inside the data center, problem areas include backbone trunk cables that run between high-end core switches or from meet-me rooms to cabinet-row spine switches.

Whether the data center configuration calls for point-to-point or switch-to-switch connections, the increasing fiber counts create major challenges for data centers in terms of delivering the higher bandwidth and capacity where it is needed.

The massive amount of fiber creates two big challenges for the data center.

The first is, how to deploy it in the fastest, most efficient way: how do you put it on the spool, how do you take it off of the spool, how do you run it between points and through pathways? Once it's installed, the second challenge becomes, how do you break it out and manage it at the switches and server racks?

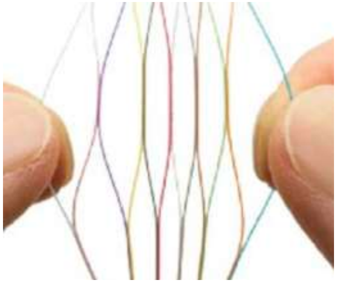
Rollable ribbon fiber cabling

The progression of fiber and optical design has been a continual response to the need for faster, bigger data pipes. As those needs intensify, the ways in which fiber is designed and packaged within the cable have evolved, allowing data centers to increase the number of fibers without necessarily increasing the cabling footprint. Rollable ribbon fiber cabling is one of the more recent links in this chain of innovation.

Rollable ribbon fiber cable is based, in part, on the earlier development of the central tube ribbon cable. Introduced in the mid-1990s, primarily for OSP networks, the central tube ribbon cable featured ribbon stacks of up to 864 fibers within a single, central buffer tube. The fibers are grouped and continuously bonded down the length of the cable which increases its rigidity. While this has little affect when deploying the cable in an OSP application in a data center a rigid cable is undesirable because of the limited routing restrictions these cables require.

In the rollable ribbon fiber cable, the fibers are attached intermittently to form a loose web. This configuration makes the ribbon more flexible, allowing manufacturers to load as many as 3,456 fibers into one two-inch duct, twice the density of conventionally packed fibers. This construction reduces the bend radius making these cables easier to work with inside the tighter confines of the data center.

Inside the cable, the intermittently bonded fibers take on the physical characteristics of loose fibers which easily flex and bend making it easier to manage in tight spaces. In addition, rollable ribbon fiber cabling uses a completely gel-free design which helps reduce the time required to prepare for splicing, therefore reducing labor costs. The intermittent bonding still maintains the fiber alignment required for typical mass fusion ribbon splicing.

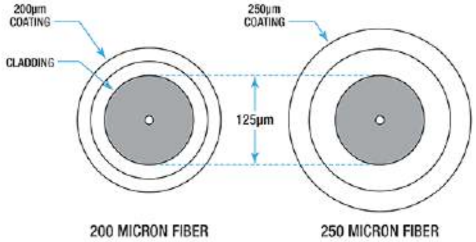


Rollable ribbon fiber is bonded at intermittent points

Source: ISE Magazine

Reducing cable diameters

For decades, nearly all telecom optical fiber has had a nominal coating diameter of 250-microns. With growing demand for smaller cables, that has started to change. Many cable designs have reached practical limits for diameter reduction with standard fiber. But a smaller fiber allows additional reductions. Fibers with 200-micron coatings are now being used in rollable ribbon fiber and micro-duct cable.



For optical performance and splice compatibility, 200-micron fiber features the same 125-micron core/cladding as the 250-micron alternative.

Source: ISE Magazine

It is important to emphasize that the buffer coating is the only part of the fiber that has been altered. 200-micron fibers retain the 125-micron core/cladding diameter of conventional fibers for compatibility in splicing operations. Once the buffer coating has been stripped, the splice procedure for 200-micron fiber is the same as for its 250-micron counterpart.

New chipsets are further complicating the challenge

All servers within a row are provisioned to support a given connection speed. But in today's hyper-converged fabric networks it is extremely rare that all servers in a row will need to run at their max line rate at the same time. The difference between the server's upstream bandwidth required and the downstream capacity that's been provisioned is known as the oversubscription, or contention ratio.

In some areas of the network, such as the inter-switch link (ISL), the oversubscription ratio can be as high as 7:1 or 10:1. A higher ratio is chosen to reduce switch costs but the chance of network congestion increases with these designs.

Oversubscription becomes more important when building large server networks. As switch to switch bandwidth capacity increases, switch connections decrease. This requires multiple layers of leaf-spine networks to be combined to reach the number of server connections required.

Each switch layer adds cost, power and latency however. Switching technology has been focused on this issue driving a rapid evolution in merchant silicon switchin ASICs.

On December 9, 2019, Broadcom Inc. began shipping the latest StrataXGS Tomahawk 4 switch, enabling 25.6 Terabits/sec of Ethernet switching capacity in a single ASIC. This comes less than two years after the introduction of the Tomahawk 3 which clocked in at 12.8Tbps per device.

These ASICs have not only increased lane speed, they have also increased the number of ports they contain. Data centers can keep the oversubscription ratio in check. A switch built with a single TH3 ASICs supports 32 400G ports. Each port can be broken down to eight 50GE ports for server attachment. Ports can be grouped to form 100G, 200G or 400G connections. Each switch port may migrate between 1-pair, 2-pair, 4-pairs or 8-pairs of fibers within the same amount of QSFP footprint.

While this seems complicated it is very useful to help eliminate oversubscription. These new switches can now connect up to 192 servers while still maintaining 3:1 contention ratios still maintaining 8 400G ports for leaf-spine connectivity! This switch can now replace 6 previous-generation switches.

The new TH4 switches will have 32 800Gb ports. ASIC lane speeds have increased to 100G. New electrical and optical specification are being developed to support 100G lanes. The new 100G ecosystem will provide an optimized infrastructure which is more suited to the demands of new workloads like maching learning or AI.

The evolving role of the cable provider

In this dynamic and more complex environment, the role of the cabling supplier is taking on new importance. While fiber cabling may once have been seen as more of a commodity product instead of an engineered solution, that is no longer the case. With so much to know and so much at stake, suppliers have transitioned to technology partners, as important to the data center's success as the system integrators or designers.

Data center owners and operators are increasingly relying on their cabling partners for their expertise in fiber termination, transceiver performance, splicing and testing equipment, and more. As a result, this increased role requires the cabling partner to develop closer working relationships with those involved in the infrastructure ecosystem as well as the standards bodies.

As the standards surrounding variables such as increased lane speeds multiply and accelerate, the cabling partner is playing a bigger role in enabling the data center's technology roadmap. Currently, the standards regarding 100GE/400GE and evolving 800Gbs involve a dizzying array of alternatives. Within each option, there are multiple approaches, including duplex, parallel and wavelength division multiplexing – each with a particular optimized application in mind. Cabling infrastructure design should enable all of these alternatives.

All comes down to balance

As fiber counts grow, the amount of available space in the data center will continue to shrink. Look for other components, namely servers and cabinets, to deliver more in a smaller footprint as well. Space won't be the only variable to be maximized. Combining new fiber configurations like rollable ribbon fiber cables with reduced cable sizes and advanced modulation techniques, network managers and their cabling partners have lots of tools at their disposal. They will need them all.

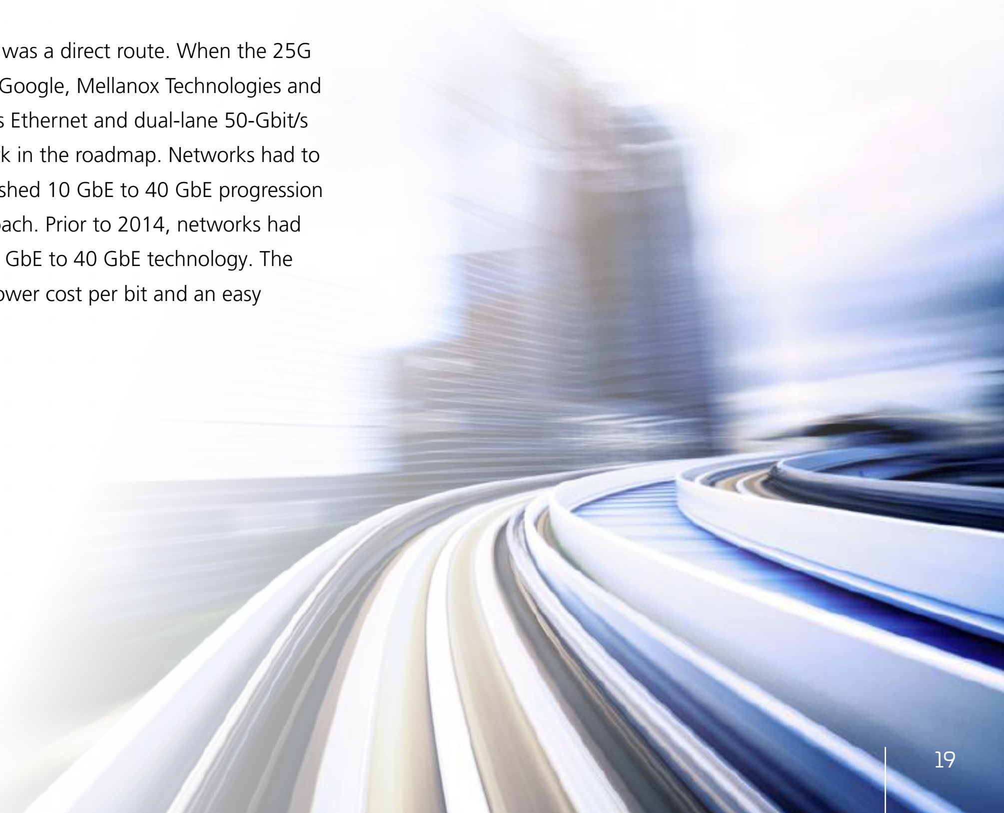
If the rate of technology acceleration is any indication of what lies ahead, data centers, especially at the hyperscale cloud level, better strap in. As bandwidth demands and service offerings increase, and latency becomes more critical to the end-user, more fiber will be pushed deeper into the network.

The hyperscale and cloud-based facilities are under increasing pressure to deliver ultra-reliable connectivity for a growing number of users, devices and applications. The ability to deploy and manage ever higher fiber counts is intrinsic to meeting those needs. The goal is to achieve balance by delivering the right number of fibers to the right equipment, while enabling good maintenance and manageability and supporting future growth. So, set your course and have a solid navigator like CommScope on your team.

4 | 16-fiber MPO provides attractive path to 400GbE

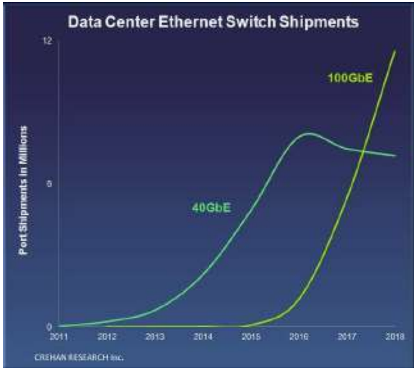
Kam Patel

Nobody ever said the Ethernet Roadmap was a direct route. When the 25G Ethernet Consortium (Arista, Broadcom, Google, Mellanox Technologies and Microsoft) proposed single-lane 25-Gbit/s Ethernet and dual-lane 50-Gbit/s Ethernet in July 2014, it created a big fork in the roadmap. Networks had to choose between sticking with the established 10 GbE to 40 GbE progression or adopting the 25 GbE to 50 GbE approach. Prior to 2014, networks had built their migration plans around the 10 GbE to 40 GbE technology. The introduction of the 25 G lane offered a lower cost per bit and an easy transition to 50 G, 100 G and beyond.



25G/50G lanes create new challenge

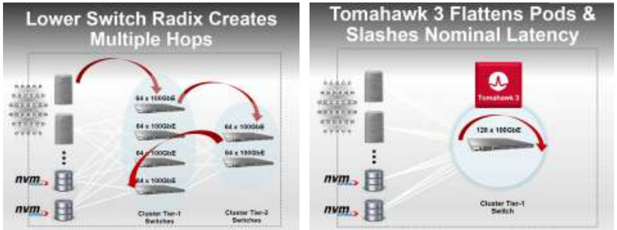
If port sales of the competing technologies are any indication, the industry has spoken. Boosted by the 25GbE technology, sales of 100GbE ports are on the rise while 40GbE switches have been declining ever since 2016. Much of the growth in 25GbE lane adoption is being fueled by larger hyperscale and cloud-based data centers where 100GbE singlemode core-to-core links are standard and provide an easy jump to 400GbE.



50G (8 port)	Module	Switch - Switch	Breakout
2 X 200		MPO 16, MPO 8, LC	DC, SN, CS
4 X 100	QSFP-DD, OSFP	DC, SN	
8 X 50		MPO 16	
2 X 50	SFP-DD		SN, DC

The increase in speeds is also being driven by new merchant silicon (ASIC) chips that enable the next generation of switches to support 32 ports of 400 Gbps, based on 50GbE lane speeds.

As switch ASIC chips get faster and add more switching capacity, data centers must find the most cost-effective way to distribute the increased capability to the front panel of the switch. The goal is to be able to support 100GbE, 200GbE and 400GbE switch-to-switch links while also allowing these switches to provide the option of connecting more servers.

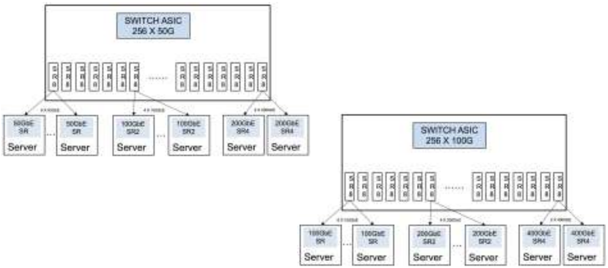


Servers connected through a single switch provide much lower latency, a key driver for application performance.

Source: Broadcom

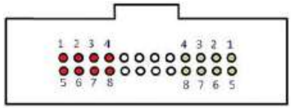
The value of 16-fiber technology

Until recently, the primary method of connecting access switches and servers within the data center has involved 12- or 24-fiber connectivity, typically delivered using multi-fiber push-on (MPO) connectors. The introduction of octal technology (8 switch lanes per switch port) enables data centers to match the increased number of 50G lanes (currently 256 per switch ASIC).

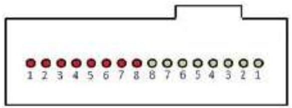


With 32 ports per 1U switch network managers can provide bandwidth of 400G per port. This is where 16-fiber interfaces offer a distinct advantage over 12-fiber or 24-fiber solutions. A 12-fber MPO supports a maximum of 6 lanes, while a 24-fiber MPO connector supports the required sixteen fibers—eight transmit and eight receive—but leaves a third of the 24 MPO connections unused. The sweet spot, then, seems to be a 16-fiber MPO connector. Eight transmit and eight receive fibers, each running at 50 Gbps would support 400GbE while utilizing 100% of the connector's capacity.

The value of the 16-fiber approach is the added flexibility it affords. It enables data centers to take a 400G circuit and divide it into manageable chunks using the current predominant technology of 50 GbE. For example, a 16-fiber connection at the switch can be broken out to support up to eight servers connecting at 50 Gbps, or up to four 100 Gbps servers with 100G connections.



24 fiber MPO



16 fiber MPO

16-fiber MPO connectors are just beginning to be deployed with the broad market expected to emerge over the next 12 to 18 months. The 16-fiber MPO is nearly identical to existing 12- and 24-fiber MPOs. The main difference being that the pins of the 16-fiber connector are smaller, and the connector's body is keyed differently to eliminate confusion with the 12-fiber or 24-fiber units. 16-fiber connectivity is supported with a 2 x 12-fiber (MPO24) configuration, or a single row of 16 fibers (MPO16). While the single row MPO design aids polarity management, 12- and 24-fiber MPOs are more popular solutions.

As servers become more powerful (and power hungry), the number of servers per rack is expected to decrease.

Positioning these switches at the top of server racks results in unused (and very expensive) switch ports sitting idle at the top of the server rows. By using 16-fiber connectivity to breakout these new 400G switch ports, data centers can consolidate switching equipment into the middle or end of the server row and make better use of their available capacity.

At the same time, server connection speeds are increasing. In response, data centers are replacing direct-attach-copper (DAC) server connections with active-optical-cables (AOCs) whose fiber-based optics increase the reach of these connections. AOCs, however, use non-standard electronics integrated into the assembly. So, while the AOC extends the link span, the entire link has to be replaced anytime speeds increase. A structured cabling approach using 16-fibers enables easy upgrades so that only the transceivers need replacing. This flexible design accommodates server speed upgrades with lower capital equipment costs while maintaining the efficient, low latency performance today's server networks require.

Adding more tools to the toolkit

The IEEE recently completed the 802.3cm 16-fiber multimode standard (set for publication early in 2020) with the 400G SR8 designed for in-row 50G server connections. Singlemode schemes are also being developed with two 400G-DR4s in one package or eight 100G lanes on 8 fiber pairs which takes advantage of the 16-fiber technology.

With either multimode or singlemode, multiplexing also plays a role in meeting capacity and latency requirements. The IEEE also added a new SR4.2 400G multimode standard, which is essentially four 100G bi-directional (BiDi) links combined in a single module. When deployed on OM5 multimode, which is optimized for short wavelength division multiplexed (SWDM) links, this standard provides for link spans up to 150 meters, 50-percent longer than currently available.

Ultimately, the right solution depends on a variety of variables —budget, location, application requirements and service level agreements to name a few. And not everybody is trying to solve the same problems. Technologies like 16-fiber connectivity become another tool in the toolkit available to help data center operators get where they need to go as efficiently as possible.

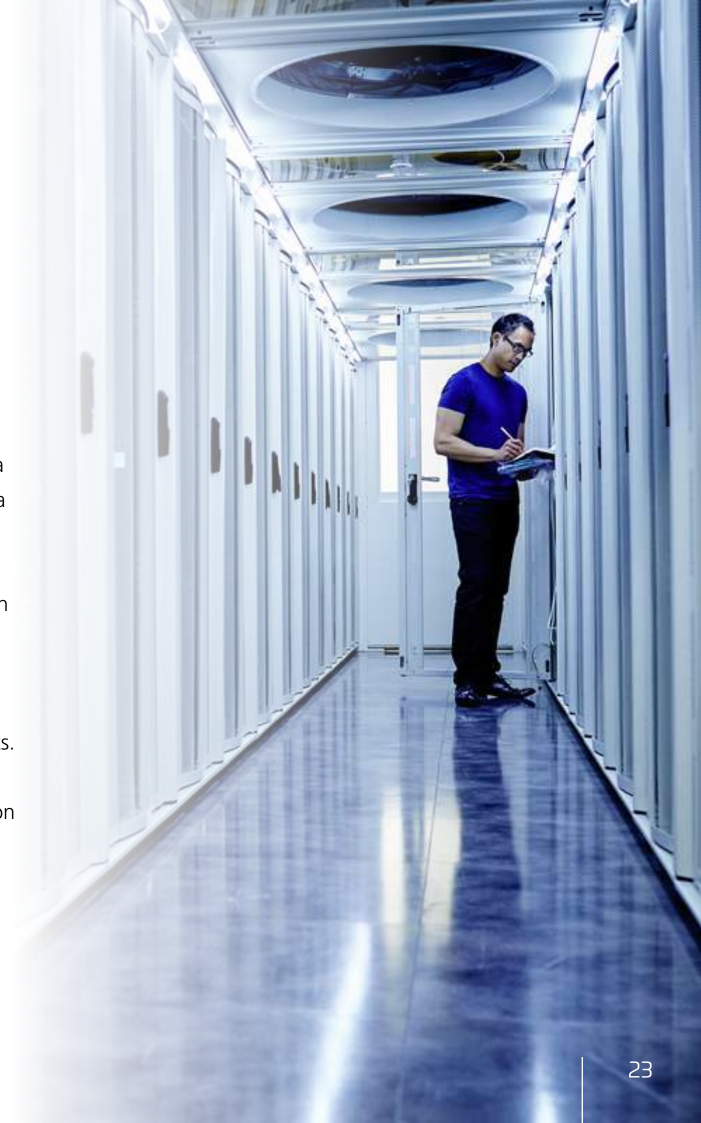
5 | When Does One Person's Edge Become Another's MTDC?

James Young

Challenges and opportunities in multi-vendor, multi-carrier networks

Till now, the lines of demarcation influencing the design and role of a data center were, at least somewhat, clear and consistent. The local service area tended to reach out about 150 miles, give or take. Inside the data center, resources such as compute and storage capacity, as well as the links connecting the data center to the access network, were designed based on a predictable traffic load.

This is especially true within the multi-tenant data center (MTDC) where physical location is often dictated by the latency requirements of its tenants. An MTDC that is proximal to a stock exchange, for example, has a high value to firms that require low latency access to that exchange. The location of the users typically defines the network's edge while the data center's location is dictated more by latency. Much of this is now changing.



Contraction at the edge

Deployments of 5G and IoT are beginning to ramp up, enabling a new order of applications requiring ultra-reliable, low-latency performance. One effect is the shrinking of the traditional local service area, which brings the data center ever closer to the network edge. Today, the lines—and even the roles—separating the two are starting to blur.

Of course, this trend is not entirely new. For years, content providers have deployed more and more resources closer to their subscribers to support content caching for improved savings and reduced latency. Now, however, other types of networks are finding strong use cases and are doing the same. While the MTDCs must find a way to reposition themselves, the carrier networks—many of whom have struggled with falling ARPU—are finding new opportunities in the increased edge-based activity.

Ultra-reliable, low latency becomes a capacity issue

Driving the concept further are IoT machine-to-machine latency requirements and the deluge of data that billions of connected devices will produce. In other words, increasing capacity is critical. Yet, there is a limit to how much fiber can be deployed, so operators must look at other methods for adding bandwidth. Wavelength division multiplexing may be one piece of the puzzle, another is to continue to reduce the distance the data must travel.

There are a number of strategies for shortening the data path. From a network design perspective, operators will need to continue to increase the amount of east-west traffic as opposed relying on longer north-south hops back and forth between the data centers and the edge. This will also require more parallel links to meet the higher reliability requirements.

The bottom line is that networks will need to keeping building out their edge-based resources in order to consume more data locally. This not only enables them to meet the URLL demands, it can become an effective strategy for conserving bandwidth.

Cloud integration and effects in the data center

All this will affect the design and, to a certain extent, the role of MTDCs. As the network service area contracts, the resources deployed at the edge will be much better equipped to handle the performance demands than the traditional (regional) MTDC solutions. In addition, requirements for lower cost, smaller footprint and smaller service areas will further challenge the existing MTDC business models.

The cloud will play a key role as carriers and content services providers look to adapt to this new environment. CSPs will gravitate to large cloud deployments while smaller cloud instances dominate at the edge. The major challenge will be extending a distributed cloud structure across many geographic locations while providing service automation and maintaining security control.

While the types of data traffic being shuttled between the edge and the data centers at the core will change, it should have little effect on the existing software and control systems. The infrastructure, especially the amount of fiber, will absolutely need to change. Edge-to-core traffic will drive the need for more Ethernet. Fiber will be the key to success with evolving high density fiber cable and apparatus solutions emerging to meet these requirements.

There is also a role for long haul high capacity options, as well as the bandwidth-boosting effects of wavelength division multiplexing application like coarse wave division multiplexing (CWDM) and dense wave division multiplexing (DWDM).

It is important to note that these next gen networks will not be created from the ground up; operators will be adapting what they have, leading to hybrid multivendor systems featuring a mix of new and existing components. The infrastructure will be complex and getting it right is a tall order. But if done well, it will lead to a more efficient and simplified network, able to grow and adapt to meet new and unimagined demands.

What's Next?

Although there are still some questions around exactly what networks of the future will look like, the evolution of the data center is starting to bring this future into sharper focus.

Data centers will continue to grow with global internet traffic increasing up to three times by 2021. Network architectures will evolve to meet these demands - with 400Gbps speeds forcing interconnected architectures that support virtualization and unimagined transmission speeds. User-centric networks will require resources to be pushed closer to the users and the connected devices they are using while 5G will require higher compute and storage horsepower.

Future proofing the data center will continue to be a priority for data center operators in the coming years. At CommScope we stay on top of the ever-changing trends and are here to support you and your networks' growth today and into the future.

commscope.com

Visit our website or contact your local CommScope representative for more information.

EB-114170-EN