CORNING

Introduction to Data Centers

Overview: What is a Data Center?

A data center (per ANSI/TIA-942) is a building or portion of a building whose primary function is to house a computer room and its support areas. A data center's main functions are to consolidate, manage, and disseminate information technology (IT) or information communication technology (ICT) resources. This includes housing and supporting a network for data processing and operational hosting that provides services such as application and internet access, content distribution, database management, file storage, and data backup. All these services work together to facilitate e-Business among various businesses, consumers, and other administrations (e.g., government agencies).
Some typical software services include:

- Electronic Data Interexchange (EDI)
- Customer Relationship Management (CRM)
- Enterprise Resource Planning (ERP)
- Supply Chain Management (SCM)
- SalesForce Automation (SFA)

In addition to traditional data center services, modern data centers lead the market in search technologies, social networking, and mobile applications, as well as e-commerce. These services extend into the more general world of "cloud computing," as delineated by Infrastructure-, Platform-, and Software-as-a-Service (IaaS, PaaS, and SaaS), where traditional aspects of on-premises services can be outsourced. Figure 1 demonstrates the level of control and management each e-Business service type requires from its target customer. These services run the gamut and include applications, data management, runtime optimization, middleware, O/S, virtualization, server hosting, storage, and complete networking.

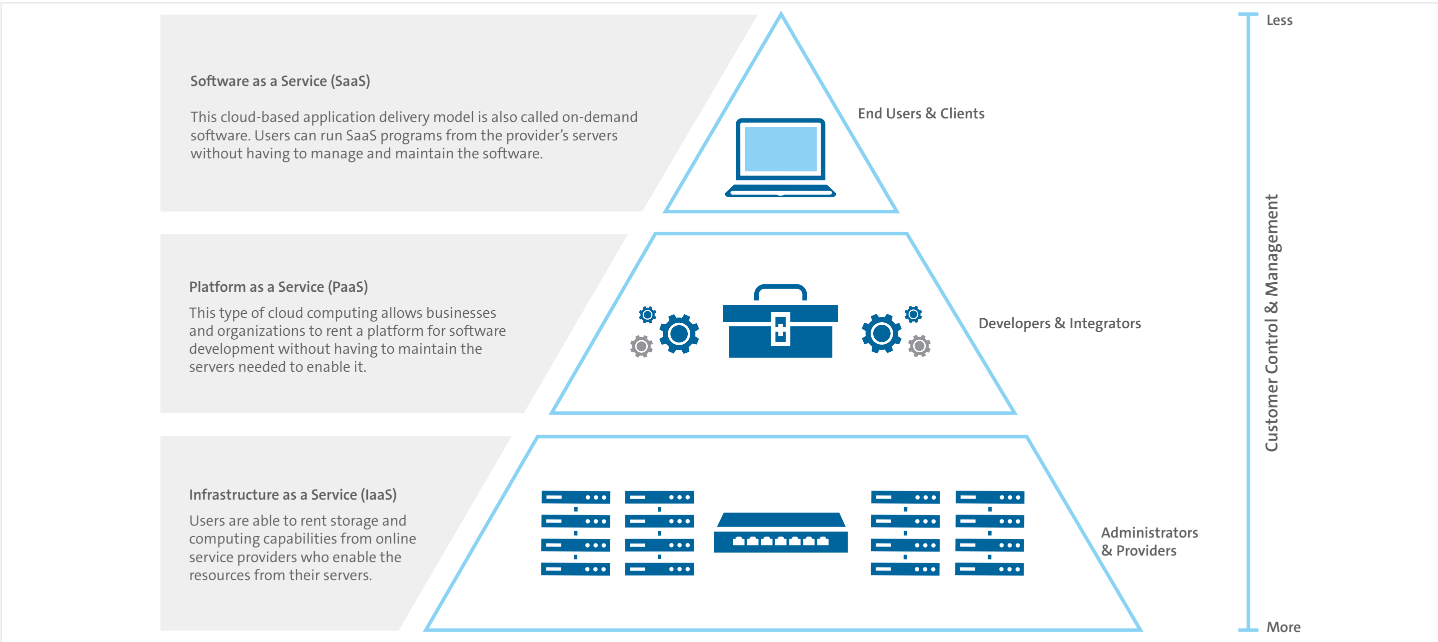

Figure 1: Modern e-Business Services

Beyond information services, data centers also exist as a physically secure facility for uninterrupted service to mission-critical operations. As a result, data centers must ensure the following to provide overall system reliability, resiliency, and redundancy:

- Around-the-clock network availability
- Fail-safe power
- Environmental (including HVAC) controls
- Fire suppression
- Fault detection and monitoring
- Access control, security, and surveillance
- A high-performance fiber optic cabling infrastructure

Data centers were traditionally designated as either enterprise (e.g., corporate) or carrier (e.g., internet) hosting facilities, depending on if they existed to support the private or public domain.

Enterprise data centers are privately owned and operated by companies, institutions (e.g., universities), governments, or other business entities and organizations. Enterprise facilities provide internal data transactions and processing, web-based services through either intranets or extranets, and are supported and managed by internal IT support.

Carrier data centers are traditionally owned and operated by "TELCOs," traditional telephone companies or telecommunications service providers, or unregulated competitive service providers who serve the greater public and offer outsourced IT services through internet access. These facilities are usually either **colocation (co-lo)** or **multitenant data centers (MTDC)**, where space is rented out to various clients to host services and/or provide networking equipment to connect to a service provider. Space can be allocated according to need, whether it's just a server rack or cabinet, a complete purpose-built cage, or an entire data hall.

In addition to the two traditional designations, a distinct facility class has emerged known as **Hyperscale Data Centers (HDCs)**, which are much larger than any of their predecessors. These operators include traditional carriers but are mainly comprised of today's dominant e-commerce and cloud service providers (e.g., Amazon, Apple, Facebook, Google, IBM, and Microsoft). These facilities may have several hundred thousand servers to support a wide variety of external customers and clients, or they may be dedicated to a single enterprise, where scalability is just as critical as capacity. Moreover, cloud service providers have become the primary sources of international bandwidth demand. They now account for more than half of all used international capacity (according to TeleGeography's "The State of the Network," 2020 edition).

Figure 2 depicts the different types of facilities that now fall under the general term of "cloud services," compared to the traditional in-house enterprise services.

Enterprise	In-House		▪ Private ownership ▪ Enterprise large organizations	Organizations can design, build, and operate their own facilities
Cloud	Colocation "co-lo"		▪ Customers own hardware ▪ Outsource facility and internal systems maintenance	In a multitenant data center, colocation space can be sold to enterprises by the rack, cabinet, or cage
	Wholesale Data Center		▪ Sell large space ▪ Supplies facilities maintenance	The provider can sell data center space in larger capacities vs. co-lo
	Dedicated Hosting		▪ Servers are **not** shared ▪ Customer controls server	The provider operates and/or rents server capacity to single customers
	Managed Hosting		▪ Hardware owned by customer or provider ▪ Many services provided	The provider operates servers and storage for its customers
	Shared Hosting		▪ Multitenant Applications	Customers share server capacity

Figure 2: Data Center Types and Services

The power ratings of hyperscale data centers are tens to hundreds of megawatts, while the power ratings for edge data centers are tens to hundreds of kilowatts.

A contrary and seemingly contradictory trend to HDC growth is the burgeoning **Edge Data Center (EDC)** market. The "edge" of traditional operator networks, such as TELCO central offices or MSO headends, is where operators connect to their customers. The "edge" of modern telecommunication networks entails those locations that position more data closer to end-users and clients to improve latency, bandwidth, and **quality-of-service (QoS)** issues. This could involve deployments spanning the following network- and establishment-based installation sites: cell tower and radio shelters, access and aggregation points (including single OSP cabinets and plug-in containers), enterprise, industrial, retail, financial, transportation (e.g., railroad) locales, related municipal vicinities (e.g., light poles), and of course traditional multirack lineups in core or cloud facilities.

Bell Labs predicts that 60% of servers will eventually be placed in EDCs by 2025. Additionally, Gartner predicts that by 2025, 75% of enterprise-related data will be created and processed at the edge.

Machine-generated traffic (including machine-to-machine communications) and increasing **Internet-of-Things (IoT)** device installations at the edge will comprise many of the drivers toward these trends. Service providers will have to improve network functionality to support faster and greater data processing for technologies, such as **virtual reality (VR)**, connected vehicles, telemedicine, immersive media, and full-scale smart city developments. In fact, cloud providers can be thought of as extensions of edge computing or the gateways to edge resources.

EDCs will also complement the distributed and locally integrated mobile network currently represented by 5G, which is essentially a first-generation edge application. For more information on this topic, consult TIA's position paper "Edge Data Centers" and related TIA and AFCOM white papers on the drivers and advantages of EDCs. For information on projects to transform the network edge into an agile service delivery platform, refer to the CORD (Central Office Re-architected as a Datacenter) platform, which seeks to leverage SDN, NFV, and cloud technologies toward the conversion of existing central-office structures into edge data centers.

Globally, markets for these services will continue to grow and compel more investment in the public, edge, and hyperscale data center spaces. Although these trends portend less spending for on-premise IT, millions[†] of data centers still exist within that range, from small enterprise server closets and rooms to commercial MTDCs, as well as today's larger HDC sites. The long-term growth trend is unambiguous with regard to data center providers and increasing markets.

[†]"Datacenters of the future," 451 Research Advisory Pathfinder Report, May 2017

Data Center Functional Areas

Returning to the architectural aspects of a data center facility, the functional areas and typical components of a data center can be broken down into the following zones with their respective physical and logical network communication roles:

1. Wide Area Network (WAN) Connectivity

- Physical: Service/access provider network demarcation
- Logical: Network ("internet") packet routing

2. Local Area Network (LAN) Connectivity

- Physical: Customer/client server front-end access
- Logical: Data link frame switching and routing

3. Storage Area Network (SAN) Connectivity

- Physical: Data volume device back-end access
- Logical: Data link frame switching and block-level transfers

The major elements of data center structure and topology will be examined throughout this guide in the context of ANSI/TIA-942-B, "Telecommunications Infrastructure Standard for Data Centers," but **Figure 3** below provides a high-level introductory view.

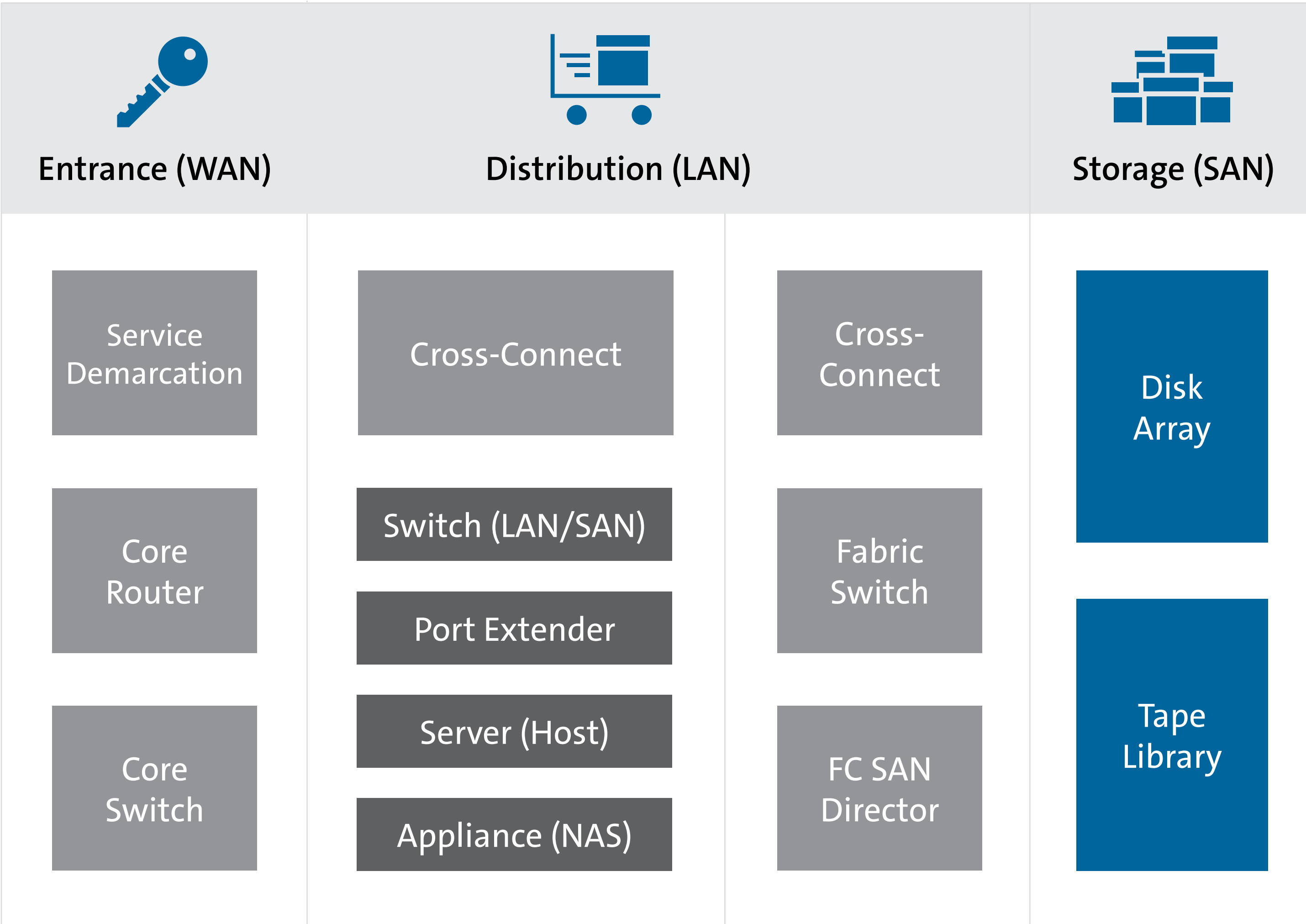

Figure 3: Data Center Functional Areas or Zones

Note: Components shown are not an exhaustive representation, but provided for illustrative and conceptual purposes only.

Entrance Access Zone

This area of the data center is sometimes referred to as the "point-of-presence" (PoP) or "meet-me" room. However, it is more specifically a physical **entrance room (ER)**, as defined by ANSI/TIA-942, where the service provider's interbuilding cabling connects to the customer-owned intrabuilding cabling. This entrance access zone enables easy passage to the service provider's telecommunications network. This area may contain the provider's demarcation hardware and equipment, including internet gateways and routers, as well as traditional **PBX** or **voice gateways** and **T-3 multiplexers**. For this reason, this zone is historically associated with the "core" layer of the network architecture. Data centers may have multiple entrance rooms for redundancy purposes or accommodating access provider-provisioned circuit cable length restrictions. This zone is often outside the main computer room for administration and security purposes, but it may coincide with the main computer room or distribution (areas) in smaller data centers.

Distribution Network Zone

The ER interfaces with the rest of the data center (i.e., the "computer room") through the **Main Distribution Area (MDA)**. The MDA may be adjacent to or part of the ER. This area provides the front-end connection to the primary functional components of the data center: servers.

Server types may include application, catalog, database, gaming, file, mail, media, virtual, and web servers. Network equipment located in the MDA consists of core routers, LAN and SAN switches (including “spine” and “interconnection,” or “leaf”), high-performance computing (HPC) switches, and software-based virtual machines (VMs) or hypervisors for server virtualization. This zone also serves as the main cross-connect (MC), which is the central point of distribution for the facility’s cabling infrastructure. For this reason, the MC may also be known as the *central* or *main patching location*.

The MDA can connect to lower layer distribution areas, which establishes a hierarchy designed to efficiently and effectively establish connectivity between all the data center components. These other distribution areas and major elements are further defined in TIA-942-B.

Storage Zone

This area of the data center (if it exists) provides the back-end connection to data. **Storage Area Networks[1] (SANs)** are part of the standard data center infrastructure because it offers:

- Administration flexibility
- Consolidation, management, and scalability advantages
- Access, back-up, movement, automation, **virtualization[2]**, and **software-defined storage (SDS)** capabilities

For these reasons, SANs may be preferred over conventional **direct-attached storage (DAS)** and **network-attached storage (NAS)** file-based devices.

SANs establish switches and directors between servers and storage devices, enabling comprehensive network connectivity so that servers can share a common storage utility (which may be remotely located).

Storage areas typically contain devices based on block-level storage, such as disk (JBOD/RAID), tape, and optical storage drives. Modern SANs also include the concept of **object storage[3]**, in which the data interaction occurs at the application level through an **API** instead of direct access via the operating system. Object storage devices act as modular units that can be aggregated into larger storage pools across locations. Physically, object-based storage arrays are the same types of disk units found in conventional SANs but are accessed via HTTP.

[1]The Storage Networking Industry Association (SNIA) defines the storage-area network (SAN) as a network whose primary purpose is the transfer of data between computer systems and storage devices and among storage devices.

[2]Storage Virtualization (per SNIA) is the act of abstracting, hiding, or isolating the internal function of a storage (sub) system or service from applications, compute servers or general network resources for the purpose of enabling application and network independent management of storage or data. Storage virtualization is a conceptual extension of server virtualization.

[3]Object storage is ideal for unstructured data in which the data is read but not written to, such as static web pages, backups, images and photos, music, video and other multimedia files. Several online collaboration services have emerged to take advantage of these sorts of databases.

Data Center Design Considerations

Common requirements for data centers are influenced by all of the following considerations:

- Architectural
- Electrical
- Environmental
- Fire protection
- Water infiltration
- Access and physical security
- Overall energy efficiency, which might be the single-most critical factor in successful data center design

To better understand these considerations, consult the ANSI/TIA-942-B Telecommunications Infrastructure Standard for Data Centers and other industry available resources, such as BICSI's data center design and implementation best practices and standards.*

Organizations such as ASHRAE (American Society of Heating, Refrigerating and Air-Conditioning Engineers) can also offer overall insights on energy standards and commissioning for data centers. ASHRAE provides verification and documentation methods for ensuring that the facility and all of its systems and assemblies are planned, designed, installed, tested, operated, and maintained to meet owner requirements.

Other sources of information for the design, installation, and operation of data centers include the Uptime Institute, LEED (Leadership in Energy and Environmental Design), and AFCOM (Advancing Data Center and IT Infrastructure Professionals).

This guide is not intended to be a comprehensive handbook on facility construction and planning or the design and maintenance of the relevant systems. These are comprehensive and coordinated project management and professional engineering undertakings that bear on multiple industry standards, codes, and regulations.

Rather, this guide focuses on the relevant telecommunications aspects, including data center pathways, spaces, and related topologies, as well as prevailing fiber optic cabling systems. Adequate planning and consideration during facility construction or renovation are crucial, however, to cost-effectively support growth and changes in the data center telecommunications infrastructure.

Lastly, this guide is not intended to guide the complete cost analysis for a data center. There are data center capital cost calculators that allow users to input physical infrastructure parameters toward exploring different cost scenarios. These tools utilize load, redundancy, density, key power, and cooling requirements to determine rack quantities, floor space, and total capital costs. The calculations are usually used to compare how various architectures will operate for specific locations and configurations. Such analyses are typically carried out through the normal bidding process by general construction contractors, subcontractors, and equipment vendors at the design documentation completion levels. For data center project stakeholders or design and construction teams, the cost is naturally the paramount consideration.

With the preceding disclaimer stated, several design parameters are crucial to an overall successful design, implementation, and operation of a data center over its life. These include metrics such as availability, reliability, maintainability, and scalability, as well as other criteria such as capacity, efficiency, density, and of course, capital and operational costs.

In today's ever-connected, always-on world of e-commerce, quantifiable attributes can be established to allow some predictive control over acceptable network downtime regarding facility resources. These attributes will be further explored throughout the remainder of this guide.

All data centers have four key items to consider when evaluating infrastructure with network efficiency as the ultimate goal:

1. **Manageability**
2. **Flexibility**
3. **Scalability**
4. **Reliability and Availability**

In several ways, these considerations are related. But it's useful to draw boundaries when organizing the design approach toward specific challenges and changes that inevitably occur throughout any data center's life cycle.

The common acronym of **"MAC"** (Moves, Adds & Changes) pervades even the modern data center and describes the ever-unfolding reconfiguration of data center architecture and topology.

*BICSI's prevailing Data Center resource publications include the following: 1) ANSI/BICSI 002-2019 Data Center Design and Implementation Best Practices; 2) BICSI 009-2019; Data Center Operations and Maintenance Best Practices; and 3) Essentials of Data Center Projects

Figure 4: Data Center Design Considerations

Manageability

End users want a high-performance, low-profile solution for the effective operation of the network. Manageability is essential toward this goal; without it, the cabling infrastructure takes over the data center in a short amount of time. To increase control over the data center infrastructure, **structured cabling** should be implemented. "Structured cabling" is defined as building or campus telecommunications cabling infrastructure that consists of several standardized smaller elements called subsystems. For it to be effective, structured cabling must be organized in such a way that individual fibers are easy to locate, MACs are easily managed, and there is ample airflow around cabling.

Perhaps no environment requires effective structured cabling more than the data center. With limited tolerance for downtime or network failure, the data center's owners and operators are among the main consumers of training resources devoted to structured cabling. The reason is clear – even as fewer traditional data centers are being built in favor of outsourcing to the cloud, there are still physical structures enabling the cloud, and these structures need to be cabled. The key benefit of structured cabling is that the user regains control of the infrastructure rather than living with an unmanageable buildup of patch cords and unidentifiable cables.

Fortunately, structured cabling can be objectively defined and is a central tenet of ANSI/TIA-942-B.

Flexibility

Flexibility bears many data center cabling system aspects, but it ultimately addresses services that will be supported throughout the data center. Flexible cabling infrastructure allows quick-and-easy changes with little to no impact on the data center's day-to-day operation. It also reduces the risk that tomorrow's technology will render the network infrastructure obsolete.

Network churn is a key concern for network designers and administrators. Workstation relocations and network reconfigurations all increase the strain on the network. A flexible facility is one that accommodates the useful life of the cabling in terms of data throughput, fiber performance, facility size, occupant population, equipment and electronics upgrades. Its flexibility should also extend to anything else that may necessitate reconfiguring the connectivity between client devices and the owner's networking gear.

The star topology promoted by ANSI/TIA-942 meets a variety of application requirements and is the origin of flexible design. Flexibility is also about the compromise between signal degradation for passive systems and the simplification of MACs. In terms of cable routing infrastructure, there is a general trend toward overhead cable trays suspended directly from the ceiling to facilitate the addition and removal of cabinets and racks of various heights while freeing up space below raised tiles to meet airflow and cooling requirements.

Redundancy in cable routes is another means of implementing flexibility in the data center. There are different levels or tiers of redundancy to which a facility can be designed to allow it to meet specific needs. Of course, in today's data center, using a single type of cabling (preferably fiber optic) can simplify both management and flexibility to support new applications.

Scalability

Scalability of the data center is essential for migration to higher data rates and for adding capacity without major disruption of operations. The initial data center must be designed so it can be scaled quickly and efficiently as requirements change. The data center has always been the major hub for data storage and processing in the network. As cloud computing and virtualization have become more mainstream, there will be increased data transportation, storage, and processing requirements demanded of the data center. Understanding these new applications and the system requirements they drive, particularly for transmission speeds, is important to ensure that the data center can evolve and scale to support new platforms and data rates.

Moreover, the term “hyperscale” underscores the importance of scalability in the modern data center. Hyperscale ultimately refers to the computer architecture’s ability to scale in order to respond to increasing workload demands. This can manifest itself in two types of scaling: vertical and horizontal. Vertical scaling entails adding more processing power to existing machines, whereas horizontal generally means adding to the count of overall machines in the computing environment or various application spaces. Both types of scaling directly impact cable infrastructure in terms of the number of data rates available on transmission links or channels.

The general trend in data centers concerning transmission speeds is ever-increasing multiples of 10 and 100-gigabit deployments. These deployments can scale up to hundreds, if not thousands, of ports using two-fiber connections. This results in hundreds and thousands of fibers per chassis that must be managed through a rack-mountable network cabling infrastructure. Deployments of 400 Gb and 800 Gb platforms will have to support even more fibers in one chassis. Furthermore, higher data rates also tend to drive away from traditional duplex transmission to parallel optics. Wavelength division multiplexing (WDM) implementations may be the possible exception with some transmission schemes. Still, high data rates generally mean more optical fibers – especially in modern dual-layer-mesh or “spine-and-leaf” architectures, where network switches are fully cross-linked to one another for higher data throughput and resiliency.

Many transmission rates already require a multifiber MPO-type connector for interface with electronics – 8, 12, 16, 24, and 32 fibers are all conceivable options, depending on the specific standard. The transceiver landscape is already mature for 100G speeds and above with a full suite of solutions. The move to higher-fiber densities per chassis and MPO-connector-based patching will require unprecedented patch panel density and patch cable management at the interface between network cabling infrastructure and electronics.

Higher fiber counts needed to support higher data rates will drive much greater cable density, not only in terms of number of fibers per cross-sectional area but also in the number of cables that can be secured and managed at the patch panel interface. The most successful data centers will be those that incorporate a high degree of density, flexibility, and scalability into their network cabling infrastructure.

Reliability and Availability

Data center structured cabling solutions must provide stability and enable 24/7 uptime. Uptime requirements are segmented by the “tier” level implemented in the data center, ranging from Tier 1 to Tier 4, as defined by the Uptime Institute. Tier 4 data centers have uptime requirements of 99.995%, less than one half-hour per year. Per BICSI, reliability is the probability that a piece of equipment or a system will perform its intended function under specified conditions for some time without failure. Reliability can be expressed as a percentage, using mean-time-between-failures (MTBF) data associated with electronics or other ancillary equipment or active systems.

Likewise, the data center components’ maintainability can be described with mean-time-to-repair (MTTR) data, which quantifies the probability that a component can be repaired in a specified amount of time. Overall data center availability is often associated with the likelihood that facility systems or components are operational and available at specified times and can be expressed as hours per year or related percentages. So, there are means to quantify acceptable risks with data center downtime and plan accordingly with the infrastructure.

Lastly, beyond the current system resiliency and redundancy, overall data center resources should be responsive to ever-changing business requirements that may entail rapid, unanticipated growth, improvements, and other technological innovations.

Network Efficiency

The term "efficiency" is usually associated with energy usage. The design of telecommunications cabling, pathways, and spaces can significantly influence overall efficiency in this context, especially as it applies to airflow convection. However, in this instance, network efficiency pertains to the speed and ease with which MACs can be accommodated incrementally at the lowest capital and operational cost. Data centers have seen significant growth in size and numbers in the past few years. As networks continue to evolve and move toward 100, 200, and 400-gigabit data rates, data centers should continue to see substantial growth in the future. Due to the considerable growth in data centers, there is a need to have simple, efficient cabling solutions that maximize space and facilitate reduced installation time and costs.

Corning's Data Center Solutions Provide Three Tiers of Value

Figure 5: Data Center Product Solutions Tiers of Value

Product Solutions

As illustrated in Figure 5, there are three tiers of value that Corning's data center solutions provide. Preterminated solutions are the preferred solution, as they allow for higher fiber density, reduced installation time, and the ability to facilitate MACs easily. Preterminated optical fiber cabling solutions streamline the process of deploying an optical network infrastructure in the data center. A modular design guarantees compatibility and flexibility for all optical connectivity and easily scales as demands dictate and requirements change. The preterminated solutions also manage fiber polarity, virtually eliminating it as a concern in network design, installation, or reconfiguration.

As alluded to before, technology road maps indicate that transmission speeds ranging from 10G to 400G will be based on either 2- or 8-fiber connectivity solutions. Base-8 preterminated solutions, such as Corning EDGE8® solutions, provide increased system density while optimizing network scalability, improving link performance, and utilizing 100% of the fiber (vs. conventional 12 F MPO-based solutions that may strand the center four fibers). Custom-engineered components also enable simple integration into common SAN directors and switches, while the preterminated components allow for reduced installation time and faster MACs. Furthermore, a Base-8 solution can improve network efficiency in the data center in the following areas:

- Increased asset utilization, with reduced jumper complexity and elimination of stranded cabling assets
- Technology adoption, due to 100% fiber utilization without the need for conversion modules, leads to cost and loss savings
- Risk avoidance, providing a simple path to 40G, 100G, and even 400G

Final Thoughts

To successfully meet the data center's requirements and demands, it is important to examine its topology and components. If chosen correctly, topology and components can come together to optimize manageability, flexibility, and scalability – creating a cost-effective and efficient network that saves time and money in both design and implementation.

A well-planned infrastructure can last 15 to 20 years and will have to be operational through multiple generations of system equipment and data-rate increases.

CORNING

Corning Optical Communications LLC • 4200 Corning Place • Charlotte, NC 28216 USA
800-743-2675 • FAX: 828-325-5060 • International: +1-828-901-5000 • www.corning.com/opcomm