Available online www.ejaet.com

European Journal of Advances in Engineering and Technology, 2021, 8(2):148-154

Research Article ISSN: 2394 - 658X

Optimizing Data Center Performance with Spine-Leaf Architectures: A Guide to Next-Gen Switching

Nikhil Bhagat

Principal Network Engineer
Independent Scholar, Network Engineering
nikhil.bhagat90@gmail.com

ABSTRACT

With today's increasingly complex and large-scale applications, data centers need to adapt to handle ever-expanding performance, low latency communication, and scalability requirements. Data center networks built on a three-tier, top-down architecture no longer meet the bandwidth needs of the future of applications and services. The spine-leaf structure is one solution that counteracts the drawbacks of conventional data center architectures with increased scalability, speed, and versatility. This article describes the history of data center networking, the disadvantages of legacy architectures, and the spine-leaf topology as a way forward. These are covered in more detail explaining the benefits of spine-leaf design to prepare organizations for modern workloads and cloud-native applications. The paper further explains the design considerations a network administrator needs to be aware of when architecting a spine-leaf data center solution.

Keywords: Traditional Three-Tier Architectures, Data Center Architectures, Spine and Leaf Architectures, Best practices.

INTRODUCTION

Data centers are the heart of the digital economy, as they contain the infrastructure to store, process and control huge quantities of data. They provide access to a broad range of services and capabilities, from social networks and e-commerce platforms to Enterprise Resource Planning (ERP) and big data. In the face of faster pace of digital transformation, data centers will become more performant, reliable, and scalable.

Data center networks have traditionally had a three-layer hierarchy including access, aggregation and core. However, the design has its limitations for scalability and performance especially on scales where east-west traffic (server-to-server) continues to evolve. The emergence of cloud, virtualization and big data has rendered the conventional data center network inoperable, and it was time to find a new way to design the data center network.

Spine-leaf design is a next-generation data center architecture layout that delivers more scalable and agile network. In contrast to the traditional three tier architecture, spine-leaf is specifically tuned for heavy east-west traffic and is stable irrespective of the number of servers or the network devices. This paper delves into the traditional data center network design, its drawbacks, the evolution of spine-leaf architecture, how the spine-leaf model can mitigate the limitations of traditional data center architecture, and the design considerations involved in creating a spine-leaf architecture.

TRADITIONAL DATA CENTER NETWORKS

The data center network design has been around for decades and was originally based on enterprise campus networks, but was later adapted to more specialty datacenter networks. It is typically built with a three-tier structure that has the following elements:

A. Access Layer

It's the bottom layer where the servers and storage devices interface with the network. The access layer connects the endpoints, and largely uses Layer 2 (Ethernet) switching.

B. Distribution Layer

Aggregation layer between access and core layers – The aggregation layer pools traffic from multiple access layer switches. It handles policy administration, VLAN aggregation, Layer 3 routing and links multiple access layer switches together. It performs layer 3 routing and uplinks to the Core Layer routers.

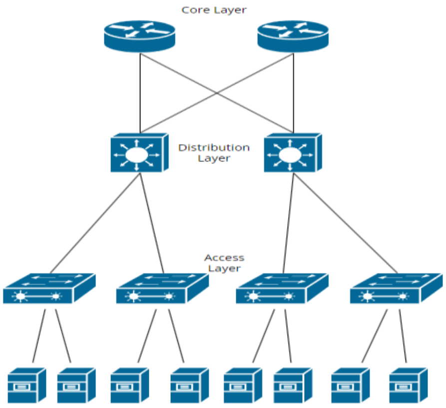

Fig. 1 *Traditional Data Center Three-Tier Network Architecture*

C. Core Layer

The core layer is the high level layer in the standard model and supports fast switching and routing between aggregation layers. The core layer also reaches the data center to outside networks, like WAN or the internet.

CHARACTERISTICS OF TRADITIONAL DATA CENTER NETWORKS:

A. Primary Focus on North-South Traffic

The traditional data center three tier architecture was designed for optimal north-south traffic flow (i.e., data travelling from the servers to other clients or the internet). This traffic converges from the access layer all the way to the aggregation and core layers.

B. Oversubscription

Oversubscription is a common theme of an older architecture especially at aggregation and core level. That's when the combined bandwidth of access layer switches exceeds the aggregation and core layers. This in turn causes traffic congestion and bottlenecks.

C. Hierarchical Structure

The traditional data center network design is hierarchical with layers that feed each other, leading to centralized traffic control.

This was acceptable for earlier data centers where the majority of traffic had north-south dominance i.e. client to server communication and vice versa. As current applications move towards a distributed environment with more server to server (east-west) communication, its flaws have become more apparent.

DRAWBACKS OF USING A TRADITIONAL DATA CENTER NETWORK DESIGN

Traditional data center networks have served as the foundation for enterprise IT for decades, but they have some downsides, especially in the case of modern workloads that need better performance, scalability, and agility. These constraints are more severe in the cloud, on large virtualized data centers and at big data analytics platforms. Among the most challenging problems of traditional data center networks are:

A. Scalability Challenges

The three-tier architecture of traditional networks is not scalable in large data centers. Increasing servers and devices adds further complexity within the network. To scale up, more aggregation and core switches are required, which will add costs and complexity. Furthermore, the hierarchy creates bottlenecks in the aggregation and core layers that are unable to accommodate the increasing east-west traffic.

B. Performance Challenges with East-West Traffic

Modern data centers primarily process high amounts of east-west traffic; that is between servers inside the data center rather than between servers and external customers. The standard data center configuration is designed to be able to process a north-south oriented traffic and is only partially capable of processing large east-west traffic. Data that needs to migrate between servers typically goes through several levels of switching (access, aggregate, core) which adds latency and can cause bottlenecks.

C. High Resource Utilization

Traditional data center networks are notoriously resource intensive due to oversubscription whereby upstream aggregation and core layers lack bandwidth to accommodate all the traffic of the access layer. Therefore, the network becomes congested and slowdowns resulting in unsatisfactory performance for bandwidth-intensive applications.

D. Complex Network Management

The management of a traditional data center networks become more challenging as the infrastructure continues to grow. The different layers of switches and routers with their own configuration and administration makes it harder to have standardized policies. This complexity also creates more potential for misconfiguration, which may cause downtime affecting the organization's availability.

E. Lack of Agility

The hierarchical nature of traditional networks can't scale rapidly to evolving business needs. Migrating servers, patching the network or rewiring infrastructure to support new application requires effort and significant planning. The lack of agility is a major disadvantage in today's dynamic IT environments, where businesses are expected to rapidly scale, configure, migrate and maintain the network.

F. Challenges with Latency sensitive workloads

As data typically has to go through multiple layers in the network, legacy architectures cause more latency that can negatively impact the application performance. When new applications such as distributed databases and big data applications get more latently aware, older networks find themselves unable to cope with this need.

These limitations have prompted data centers to seek out new architectures capable of handling the demands of high performance latency-sensitive east-west workloads.

INTRODUCTION TO SPINE-LEAF ARCHITECTURE

Spine-leaf design is a new data center networking architecture that fixes many of the problems with the traditional data center three tier layout. It's designed around a two-layer structure with spine switches serving as the foundation of the network and leaf switches routing directly to servers and storage units.

A. Architecture Overview

• **Leaf Layer**: In spine-leaf, the leaf layer comprises top-of-rack (ToR) switches that bridge directly to servers and storage units. Every leaf switch also ties in with every spine switch in the system. The leaf layer serves as the interface layer in this design, allowing you to communicate with the servers on the network.

• **Spine Layer**: The spine layer has high-performance switches which make up the center of the network. Every leaf switch goes to each spine switch in a fully mesh topology where traffic can traverse the fastest route between any two switches on the network. The spine layer allows high bandwidth, low latency linkage between all leaf switches.

B. Characteristics of Spine-Leaf Architecture:

• **Clos Topology:** The spine-leaf scheme borrows the Clos topology named after Charles Clos, a telephone engineer who came up with the same idea for phone switches. In a Clos network, all leaf switches are the same distance apart, so an organization can mitigate any outages regardless of where the traffic originates or is destined towards.

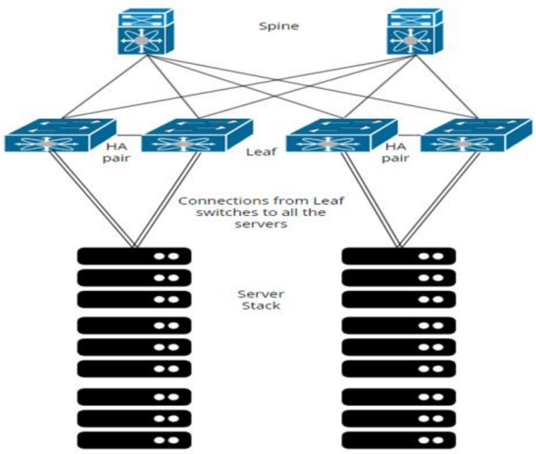

Fig. 2 *New Spine and Leaf Data Center Architecture*

• **Non-Blocking Design:** Unlike traditional networks, where traffic may be congested at aggregation or central nodes, spine-leaf design is not blocking. All the leaf switches lead directly to all the spine switches, and the network can be scaled horizontally by adding spine switches without bottlenecks.

• **East-West Traffic Optimization:** Fully meshed spine-leaf networks are optimized for east-west traffic, so they are suitable for today's data centers where most traffic is centered towards east-west communication between servers instead of north-south with external clients.

ADVANTAGES OF USING SPINE-LEAF ARCHITECTURE OVER OLD DATA CENTER NETWORK DESIGN

The spine-leaf design offers several key advantages over conventional data center network configurations, making it the ideal solution for contemporary data centers aiming to serve massive low-latency workloads. Some of the significant advantages include:

A. Scalable

Spine-leaf architectures naturally scale more than three-tier architectures. As the data centers expand, more leaf and spine switches can be added without impacting the existing network. Each leaf switch connects to all other spine switches and there are options to add more spine switches in order to broaden the network's total bandwidth. This horizontal scaling framework enables virtually limitless growth at predictable performance.

B. Optimized for East-West Traffic

Modern applications like distributed databases, big data systems, and microservices generate a good amount of east-west traffic. Spine-leaf design is crafted to manage this traffic effectively. By having each leaf switch connected to all spine switches, data travels the shortest route between any two servers which will reduce latency and increase performance.

C. Low Latency

The spine-leaf architecture drastically lowers latency compared to traditional hierarchical architecture. As each leaf switch is directly wired to each spine switch, data only needs to go through one leaf switch and one spine switch to arrive. This two-layered design reduces the number of switches the data has to go through, which means applications can operate at lower latency and response time.

D. Optimized Bandwidth Utilization

Oversubscription is one of the most common issues with the traditional three tiered design data center networks: the traffic at the access layer is more than the aggregation and core layers. In a spine-leaf network, a non-blocking scheme ensures that each leaf switch has a direct route to every spine switch, eliminating bottlenecks and optimizing bandwidth utilization.

E. Simplified Network Management

Spine-leaf architectures are easier to operate, maintain and troubleshoot than three-tier models. As it is a flatter and lower-layer network, there are fewer layers of switches to manage. This simplifies the process of managing mistakes and deploying consistent policies across the network. It is also easier to diagnose a spine-leaf network as there are fewer potential failure points.

F. High Availability

Spine-leaf architecture is built on a full mesh model that is fully redundant. When a spine switch fails, traffic will automatically be rerouted through a second spine switch, without any impact on network traffic. Likewise, when a link between a leaf switch and a spine switch fails, traffic is automatically rerouted via a redundant available to provide high availability and fault tolerance.

G. Consistent Performance

The performance within the traditional three tier data center networks depends on the device location and the amount of traffic going through the network. Spine-leaf on the other hand ensures consistency across the whole data center. Since every leaf switch is exactly halfway between each spine switch, every device receives the same latency and bandwidth irrespective of location.

H. Cost Efficient

The spine-leaf system scale horizontally by adding more leaf and spine switches, making it a cheap option for growing data centers. As the organization grows, they do not need to rebuild the entire network; they can simply add new switches to the existing infrastructure as needed to scale incrementally.

SCENARIOS WHERE SPINE-AND-LEAF ARCHITECTURE OUTPERFORMS TRADITIONAL THREE-TIER DESIGN

A. Cloud-Based Workloads

Spine-and-leaf architectural design offers high scalability, and can be used on cloud-native workloads with high scale. As it has a full mesh core design, spine or leaf switches can be easily added without slowing performance down. Alternatively, the three-tier model (core, aggregate, access) slows down as more devices are deployed, resulting in costly and expensive upgrade cycles. This hierarchical design restricts its dynamic scaling potential, making the three-tier architecture inapplicable for growing, scalable, cloud-based workloads.

B. East-West Traffic in Distributed Applications

Modern data centers experience high east-west traffic with the data traversing from server to server within the data center, especially in Hyperconverged and Distributed Applications (HCI), and distributed applications. Spine-and-leaf structures accommodate such traffic, providing multiple routes between the servers on each leaf-to-spine connection, delivering low latency, high bandwidth performance. The three-tier architecture is, however, built for north-south traffic (server-to-outside-client), and thus does not accommodate growing east-west traffic of microservices and containerized applications.

C. Low-Latency Requirements within Big Data and AI Workloads

Big data applications, AI and machine learning also demand very low latency and high-bandwidth networks to sift through large data sets at a great speed. Spine-and-leaf solutions satisfy these requirements by skipping hops between servers (normally two) instead of several. The full-mesh architecture supports higher data rates and decreases latency for real-time application. The traditional three-tiered approach adds latency when traffic goes through additional hops, and is less suitable for high performance computing.

D. High Availability in Business-Critical Environments

Spine-and-leaf design offers better fault tolerance through multiple redundant leaf/spine switch paths. When a spine switch goes down, traffic can be routed along new routes and network will remain available without compromising performance. Equal-cost multi-path (ECMP) routing improves reliability even further by spreading out traffic along all paths. The three-tier model, on the other hand, creates single points of failure (e.g., in the aggregated layer and the central level) leading to a much more vulnerable network.

E. Software-Defined Networks (SDN)

Spine-and-leaf architecture can easily be blended with SDN because of its flat and predictable topology, which simplifies the automation and control of the network. In spine-and-leaf networks, centralized SDN controllers can help control traffic and optimize resources. On the other hand, multi-layer three-tier architecture makes SDN implementation much more complicated as traffic must be handled at multiple network layers.

F. Support for Virtualized Environments

Virtualization tools like virtual machines (VMs) and containers create mobile workloads that move frequently from one host to the next. Spine-and-leaf is flexible, and the high-bandwidth, low-latency architecture lends itself well to such environments where performance doesn't vary between different workloads. The traditional three-tier hierarchical architecture introduces latency and bottlenecks when workloads shift, making it impractical for dynamic, virtualized environments.

SPINE AND LEAF ARCHITECTURE DESIGN CONSIDERATIONS

A. Capacity Planning

This spine-and-leaf structure delivers excellent scalability, both in terms of capacity and infrastructure. To create a scalable system, an organization needs to account for the number of spine and leaf switches. The spine switches on the full-mesh topology communicate with the leaf and the spine switches, and the organization could scale by adding additional spine switches (to increase the bandwidth) or leaf switches (to add more devices). Oversubscription, or the excess of egress/ingress bandwidth, must be minimized so as to prevent congestion. However, some oversubscription may be tolerated, depending on the cost and requirements of the application. For the future growth, organizations should also look into port density at leaf and spine switches.

B. Traffic Flows

Understanding traffic flows is essential before planning to build a spine-and-leaf design. Today's data centers carry a high amount of east-west traffic (between servers) on the hyper-converged and distributed architecture. Spine-and-leaf is optimized to carry east-west traffic with fewer hops and less latency. Although the importance of north-south traffic (incoming and outgoing data from external clients to data centers) has faded away, it still has to be carefully managed for seamless ingress and egress, with proper placement of firewalls and load balancers.

C. Fault Tolerance

A feature of spine-and-leaf construction that improves fault tolerance is redundancy. Spine and leaf layer redundant switches ensure high availability during a switch or a link failure. ECMP Equal-Cost Multi-Path routing spreads traffic among multiple paths to optimize, load balance and guarantee flow during a link failure. Redundant leaf-spine switches eliminate single point failure, maintaining reliability.

D. Bandwidth Requirements

The requirement of high bandwidth and low latency is at the heart of spine-and-leaf design. Spine switches needs to handle aggregated traffic across all the leaf switches, with links such as 40G, 100G or 400G depending on the bandwidth requirements. The two-hop architecture (leaf spine leaf) minimizes latency, making this the ideal architecture for real-time applications such as analytics and high-performance computing.

E. Physical Layer planning (Cabling)
Spine-and-leaf deployments require efficient physical design and wiring. Leaf switches must be placed near servers in top-of-rack (TOR) or end-of-row (EOR) modes. Spine-and-leaf layouts involve high-density cabling, with each leaf switch connecting to each spine switch. Proper cable management makes managing the infrastructure easier.

F. Selecting right Switches
It is important to select appropriate Spine and leaf switches. Spine switches handle traffic from all the leaf switches while Leaf switches provide port density for server connections. This involves choosing switches based on performance and cost, as well as software options for automation and telemetry.

G. Automation and Monitoring Tools
It takes automation to operate large spine-and-leaf systems at scale. Software-Defined Networking (SDN) allows centralization and automation automates functions such as provisioning and traffic flow. Monitoring solutions with real-time telemetry helps to improve the network efficiency by detecting issues at an early stage.

H. Connectivity with the External Environment
When communicating with the external network, the edge routers, firewalls, load balancers need sufficient bandwidth to handle traffic. For hybrid cloud solutions, bandwidth and security is needed to manage traffic between the data center and the cloud providers.

I. Security Considerations
Secure spine-and-leaf architectures with segmentation by strategically placing firewalls and load balancers in the network. Distributed firewalls help secure the network while load balancers provide a scalable way to handle traffic spikes.

CONCLUSION

The exponential growth in data center requirements, resulting from cloud computing, big data, virtualization, and the newer application platforms has made three-tier data center networks obsolete. Spine-leaf architecture provide scalable, adaptable, high-performance platform designed to scale to the needs of next-generation data centers.

With its fully meshed, non-blocking architecture, the spine-leaf design delivers predictable, low-latency results, even during the growing east-west demand. It solves many of the issues with traditional hierarchical networks (oversubscription, latency, and limited scalability). Furthermore, the spine-leaf design improves network management efficiency while offering greater redundancy and availability, making it the best option for companies seeking to maximize data center performance.

As data centers grow and scale, spine-leaf architecture will become the foundation of next-generation networking, enabling organizations to keep pace with modern workloads and deliver the performance, scalability, and reliability required by the digital economy.

REFERENCES

[1]. P. Jones, D. Comfort, and D. Hillier, "The changing geography of data centres in the UK," Geography Journal, vol. 1, no. 2, 2019.

[2]. N. Ilyadis, "The evolution of next-generation data center networks for high capacity computing," IEEE Communications Magazine, vol. 48, no. 7, pp. 138-144, July 2010.

[3]. "The case for a leaf-spine data center topology," White Paper, Dell Technologies, 2015.

[4]. "Cisco's massively scalable data center network fabric white paper," Cisco White Paper, Cisco Systems, 2016.

[5]. J. Lan, Z. Liu, L. Li, and X. Xu, "A summary of the large-scale access convergence network structure," in Proc. IEEE Int. Conf. Commun., Shanghai, China, 2019, pp. 645–652.

[6]. "Data Center Multi-Tier Model Design," Cisco Design Guide, Cisco Systems, 2014.

[7]. B. Dab, I. Fajjari, D. Belabed, and N. Aitsaadi, "Architectures of data center networks: Overview," IEEE Communications Surveys & Tutorials, vol. 20, no. 1, pp. 45-62, 2018.

[8]. S. Das, G. Parulkar, and N. McKeown, "Rethinking IP core networks," in Proc. ACM HotNets-VII Workshop, Calgary, Canada, Oct. 2008.

[9]. N. Bitar, S. Gringeri, and T. J. Xia, "Technologies and protocols for data center and cloud networking," IEEE Communications Magazine, vol. 51, no. 9, pp. 60-68, Sept. 2013.

[10]. D. Bernstein and E. Ludvigson, "Networking challenges and resultant approaches for large scale cloud construction," in Proc. IEEE Int. Symp. Cloud Comput., San Francisco, CA, USA, 2010, pp. 99-106.

[11]. U. H. Rao, "Challenges of implementing network management solutions," IEEE Network, vol. 28, no. 4, pp. 15-19, July 2014.

[12]. H. Kim and N. Feamster, "Improving network management with software-defined networking," IEEE Communications Magazine, vol. 51, no. 2, pp. 114-119, Feb. 2013.

[13]. S. M. Rumble, D. Ongaro, R. Stutsman, M. Rosenblum, and J. K. Ousterhout, "It's time for low latency," Commun. ACM, vol. 55, no. 9, pp. 50–56, Sept. 2012.

[14]. S. Walklin, "Leaf-spine architecture for OTN switching," Journal of Optical Networking, vol. 19, no. 4, pp. 76-85, April 2017.
[15]. A. Saleh, "Evolution of the architecture and technology of data centers towards exascale and beyond," IEEE Commun. Mag., vol. 52, no. 9, pp. 136–142, Sept. 2014.
[16]. M. Al-Fares, A. Loukissas, and A. Vahdat, "A scalable, commodity data center network architecture," in Proc. ACM SIGCOMM, Seattle, WA, USA, 2008, pp. 63-74.
[17]. V. Harsh, S. A. Jyothi, and P. B. Godfrey, "Spineless data centers," in Proc. USENIX Conf. Networked Systems Design and Implementation (NSDI), Boston, MA, USA, 2016, pp. 323-336.
[18]. H. R. H. Andersson, A. D. Blanche, and T. Lundqvist, "Flipping the data center network: Increasing east-west capacity using existing hardware," ACM SIGCOMM Comput. Commun. Rev., vol. 49, no. 3, pp. 32–37, July 2019.
[19]. M. Alizadeh and T. Edsall, "On the data path performance of leaf-spine datacenter fabrics," IEEE Communications Letters, vol. 18, no. 7, pp. 114-118, July 2014.
[20]. Y. Chen and J. Wu, "Joint coflow routing and scheduling in leaf-spine data centers," IEEE Trans. Parallel Distrib. Syst., vol. 28, no. 11, pp. 3181-3194, Nov. 2017.
[21]. T. Gomes, D. Tipper, and A. Alashaikh, "A novel approach for ensuring high end-to-end availability: The spine concept," IEEE Trans. Netw. Serv. Manag., vol. 14, no. 3, pp. 792-805, Sept. 2017.
[22]. K. C. Okafor, I. E. Achumba, G. A. Chukwudebe, and G. C. Ononiwu, "Leveraging fog computing for scalable IoT datacenter using spine-leaf network topology," IEEE Internet Things J., vol. 5, no. 6, pp. 4926-4935, Dec. 2018.
[23]. A. Hakiri, A. Gokhale, P. Berthou, D. C. Schmidt, and T. Gayraud, "Software-defined networking: Challenges and research opportunities for future internet," IEEE Commun. Mag., vol. 53, no. 1, pp. 36-43, Jan. 2015.
[24]. R. R. Reyes and T. Bauschert, "Infrastructure cost evaluation of intra-data centre networks," Journal of Optical Networking, vol. 20, no. 2, pp. 112-123, Feb. 2020.
[25]. L. Paraschis, "Advancements in data-center networking, and the importance of optical interconnections," Optical Fiber Communications Conference, San Francisco, CA, USA, Mar. 2018.
[26]. M. Nurujjaman, S. Sebbah, and C. Assi, "Design of resilient ethernet ring protection (ERP) mesh networks with improved service availability," IEEE/ACM Trans. Netw., vol. 26, no. 4, pp. 1772-1784, Aug. 2018.
[27]. T. Zhang, L. Linguaglossa, P. Giaccone, L. Iannone, and J. W. Roberts, "Performance benchmarking of state-of-the-art software switches for NFV," in Proc. IEEE Conf. Netw. Software (NetSoft), Paris, France, 2017, pp. 1-6.
[28]. A. Nunez, J. Ayoka, M. Z. Islam, and P. D. Ruiz, "A brief overview of software-defined networking," IEEE Commun. Surveys Tuts., vol. 21, no. 2, pp. 82-101, 2019.
[29]. W. Chao, S. Horiuchi, and K. Tayama, "A resource design framework to realize intent-based cloud management," IEEE Cloud Computing, vol. 4, no. 1, pp. 12-23, Jan. 2017.
[30]. S. K. Majhi and P. Bera, "Designing an adaptive firewall for enterprise cloud," IEEE Access, vol. 7, pp. 24734-24746, 2019.