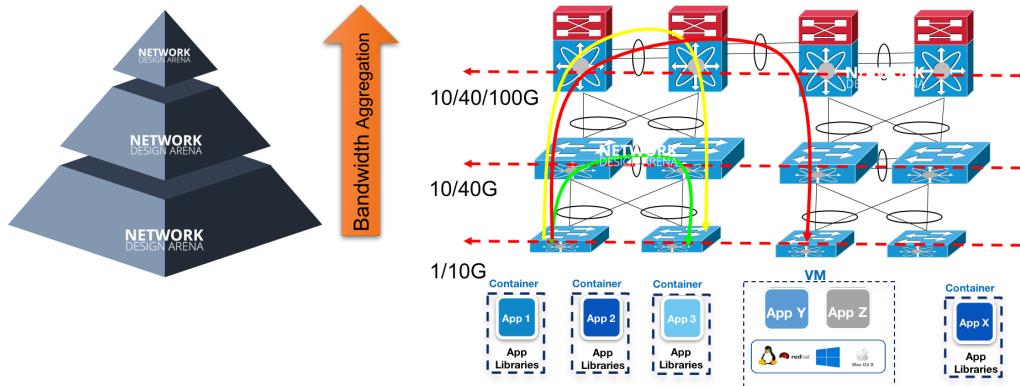


## Why Do We Need Clos Architecture and Considering Only an Overlay SDN Is Not Enough in Today's Data Network?



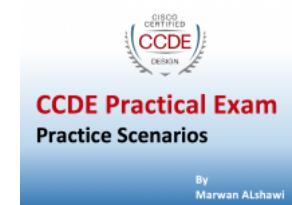
To answer the question in the subject line, let's look at 'how applications in the data center are evolving and the associated impact'. And then analyze 'why the classical network architecture [multi-tiered] is not always the best fit today'.

From a 10000 feet view (business point of view), the data center is the most critical place in the network (PIN) without any doubt, as it's the PIN where organizations (enterprises or service providers) host its digital assets (data and applications). In the digital era, we can say confidently, that data is the most critical asset of today's businesses. It can be even more critical if, the data center represents a core revenue source to the business, such as hosting data center providers (like software as a service [SaaS] and infrastructure as a service [IaaS] providers).

Therefore, you may notice that the design of a data center is always the first PIN to be influenced by the evolution of applications' architecture and requirements. Ideally, it should offer a flexible and reliable transport that is capable enough of accommodating the evolutions of business trends and applications' requirements.

As it shown in the figure below, the way applications are deployed in data centers has been evolving (bare metal > virtualization with VMs > virtualization and micro-services with containers). This change, typically, in response to several factors (technical and non-technical), however, it's outside of this bog scope. You may refer to this blog for more details about containers architecture and its drivers ([Drivers for Containerizing Applications and Container Architecture Overview](#))

### Order Now



### Featured Posts

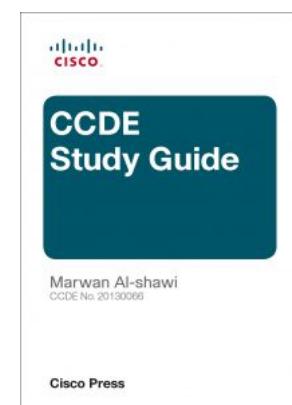
[Certification Selection Dilemma: CCDE, De Cloud Certs](#)

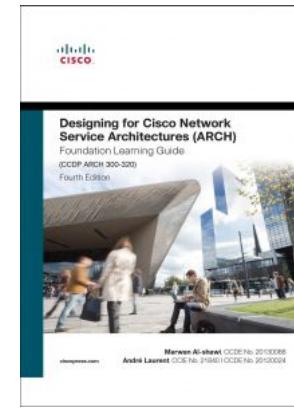
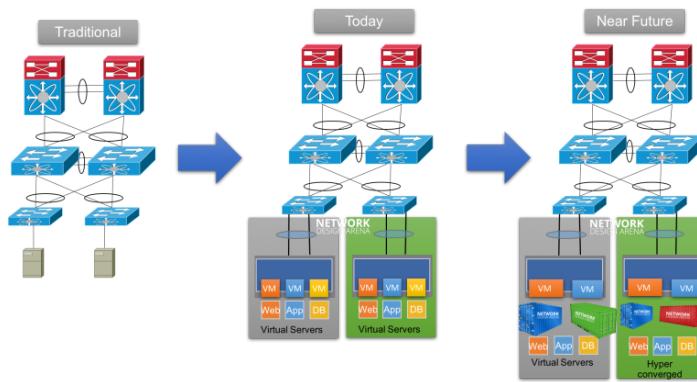
[Cisco SDWAN Design Series-Part-4- The M Cisco SDWAN Policies](#)

[Cisco SDWAN Design Series-Part-3- Contrc Planes Logic](#)

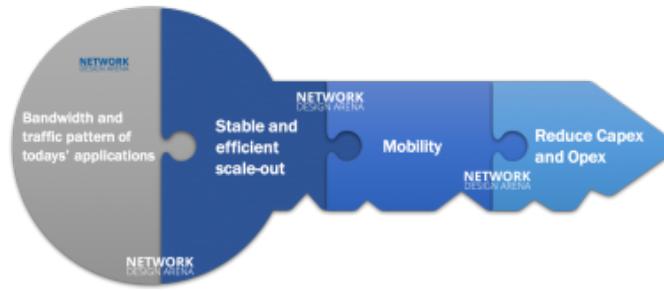
[Cisco SDWAN Design Series-Part-2- Archite Components](#)

[Cisco SD-WAN – Network Evolution for the and Digital Era –](#)

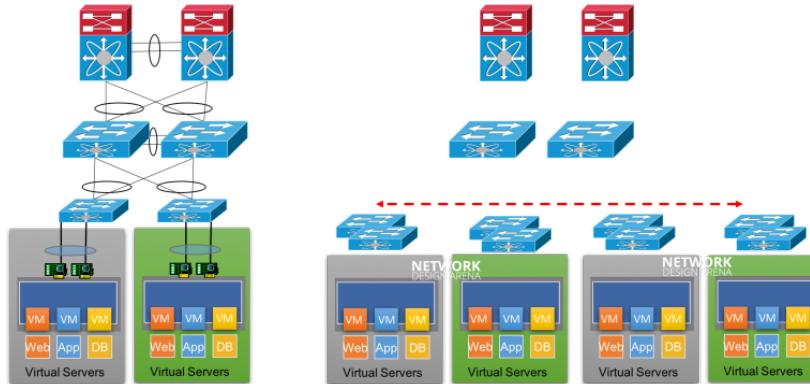




When designing any data center network, we will always need to make sure it is capable to at least, satisfy the four key architecture elements shown below. (the importance of each one can vary based on the business priorities and objectives)



Since more than a decade, systems and applications virtualization become the standard way of deploying applications. whether it's a cloud deployment or only VMs based architecture. As we know with this model, a single physical machine will run multiple virtualized servers/applications. Which means, more bandwidth per host is required. Besides, scaling out the number of machines indicates, we will need a HW and throughput scale up at the aggregation and core layers.



In fact, the new business demands and technology trends in the digital transformation era, are changing the role of IT and introducing new challenges to support new applications architecture such as micro-services, applications mobility, big data applications, as well as the ability to provide dynamic policy provisioning. In such new DC environment(as highlight above) bandwidth demand is much higher, also, it is obvious that that the traffic pattern is unpredictable compared to traditional data centers, especially when containerized applications with micro-services architecture are used or big data applications with distributed HDFS that need ultra-low latency of east-west packet forwarding to provide the expected performance.

## Recent Posts

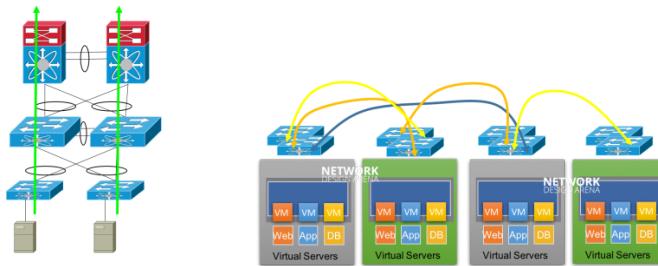
Certification Selection Dilemma: CCD DevNet or Cloud Certs

SDDC (ACI) and SDWAN -Better Toge

Cisco SDWAN Design Series-Part-4- T Magic of Cisco SDWAN Policies

The Road to HybridCloud

Cisco SDWAN Design Series-Part-3- C Data Planes Logic



Traditionally, the DC network is based on the classic three-tier hierarchy and almost always applications' traffic pattern is north-south.

With today's modern applications architectures and requirements, this classic three-tier DC network architecture may lead to significant oversubscription, possible latency and blocking architecture, which is unsuitable for the requirements of today's workloads and applications requirements.

Oversubscribe hosts connected to access switch ports is common approach or practice to, as long as these hosts can utilize the ports at a full line-rate performance and the switching throughput/fabric is non-blocking. but this is not always the case.

There are three levels of potential oversubscription:

**Port oversubscription:** this is simply when the ingress capacity exceeds the egress capacity. For instance, for 48x 10G attached servers, the access switch needs at least 480 Gbps of port capacity toward the upstream or distribution layer to provide 1:1 oversubscription (1:1 here meaning zero or no subscription). However, if the access switch has only 2x 10G uplinks, in this case the oversubscription is 24:1.

Note: Although you won't need to provide 1:1 performance, oversubscription in the data center not like campus network it needs to be minimized, ideally 4:1 or 3:1, depends on the scale and the hosted applications.

**Switch oversubscription:** this type of oversubscription happens at the device level, when total supported device switching bandwidth of a switch is less than the aggregate bandwidth simultaneously coming from the ingress switch ports. This will typically lead to partial line-rate performance of some of the port access ports.

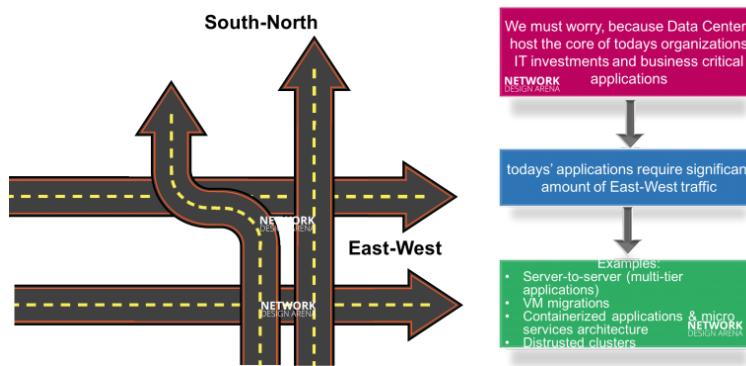
**Network oversubscription:** this type of oversubscription takes place when traffic passes through different devices, normally when traffic traverses through the network tiers where bandwidth consolidation/aggregation happens at the higher layers, as a result the ingress bandwidth will always be greater than the egress bandwidth.

Blocking on the other hand, referred to the situation when the 'ingress and egress' bandwidth capacity available at the ports level, but the switch itself is capable to perform the forwarding at the desired rate due to hardware or queuing inefficiencies.

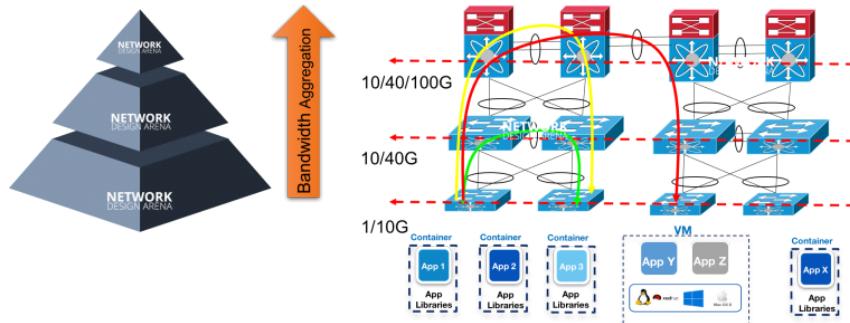
That's why you see the Cisco Nexus 9500 platform, is designed with deep buffer to provide nonblocking, low-latency, line-rate performance with high 10 and 40 Gigabit Ethernet port density similarly the Cisco Nexus 9300 platform switch consists of one network forwarding engine (NFE) and one application leaf engine (ALE) or ALE-2 to provide additional buffer space and facilitates advanced network functions.

There is nothing wrong with the three-tier architecture, but you can think of it as an architecture model was designed for different needs, scale and requirements.

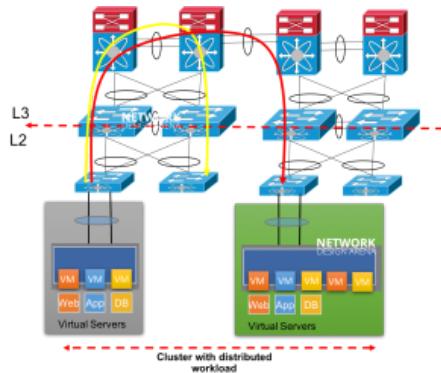
Is it something that we should worry about?



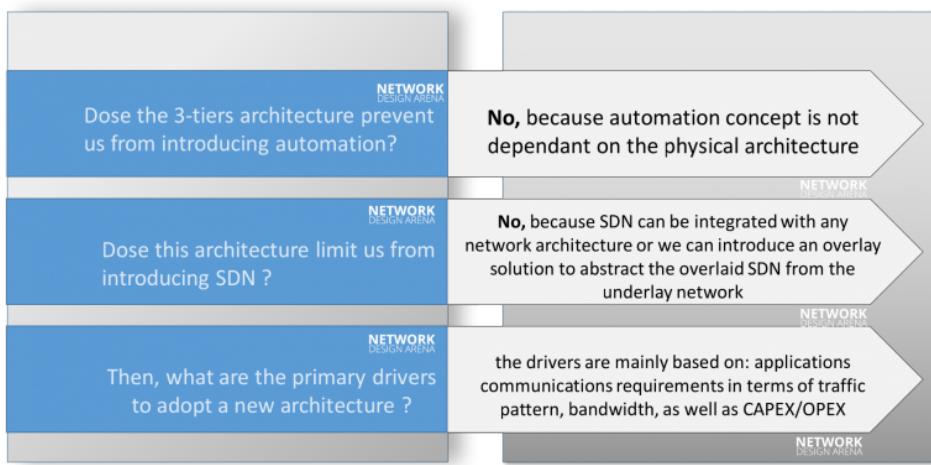
With this architecture the higher the traffic moves in the network hierarchy (bottom-top) of the three-tier network, the more bandwidth oversubscription/sharing will be among the switches in the access layer (aka. bandwidth aggregation), as illustrated in the figure below. This can get worse with the use of containerized applications with micro-services architecture, as east-west traffic will be extremely high and nondeterministic.



Same implications are applicable, when there are stretched applications' clusters (within the same DC across different access nodes) this architecture



However, this does not mean every data center need to change its network architecture, it depends on the applications and scale. Also, the following figure highlight some facts.



That being said, if you are building a modern DC network, with modern applications architecture and you are planning to rely on software overlays (SDN approach) **without optimizing the actual underlay network architecture, then you should expect a possible scale and performance degradation** as the DC grow with more applications and traffic load, because the overlay cannot optimize the actual underlay network once the pack has been put in the wire (is the switching HW capable to provide non-blocking forwarding? What about ports oversubscription?, switch and network oversubscription?).

Therefore, you need to start with architecture of the underlay network, like when civil engineers build a high-rise building, they start with a strong foundation.

That's why today Clos architecture also commonly known as "Spine and Leaf" architecture is becoming the most wildly used data center architecture.

Now, we know why do we really need a new architecture, the subsequent blog, will discuss Clos architecture in detail.

Categories :

Architect

Data Center and Cloud



Marwan Al-shawi – CCDE No. 20130066, Google Cloud Certified Architect, AWS Certified

Solutions Architect, Cisco Press author (author of the Top Cisco Certifications' Design Books "CCDE Study Guide and the upcoming CCDP Arch 4th Edition"). He is Experienced Technical Architect. Marwan has been in the networking industry for more than 12 years and has been involved in architecting, designing, and implementing various large-scale networks, some of which are global service provider-grade networks. Marwan holds a Master of Science degree in internetworking from the University of Technology, Sydney. Marwan enjoys helping and assessing others, Therefore, he was selected as a Cisco Designated VIP by the Cisco Support Community (CSC) (official Cisco Systems forums) in 2012, and by the Solutions and Architectures subcommunity in 2014. In addition, Marwan was selected as a member of the Cisco Champions program in 2015 and 2016.

### 3 Comments

**Cem Akilli** says:

October 2, 2018 at 3:58 am