

Mass-Scale Infrastructure for Cloud

Benefits

With applications and workloads supporting the proliferation of AI/ML and 5G as well as IoT generating increasing levels of traffic, data centers are being challenged with an increasing demand for edge computing and content disaggregation to keep up with the latency requirements. The demand shift means web scalers cannot continue to rely on centralized data center designs that require long transport paths and take a long time to build or modify. With a near-instantaneous response demanded by consumer applications, even moderate amounts of transport distance are too much. Web scalers need the next iteration of data center to be economically repeatable at the edge, massively scalable, environmentally sustainable, programable, and simple to operate.

Cisco overcomes these challenges and has developed four tenets that should be considered as web scalers expand their data centers. The four tenets are:

1. Performance: Systems and their associated ASICs need to be leading edge. They need to be space, power, and cooling efficient, as well as able to handle the dynamic nature of traffic regardless of role (spine, leaf, or top of rack).

2. Uniformity: To be economically feasible, new data centers need to be uniform in their deployments. This allows them to be simple to manage, utilizing common interfaces across hardware and software.

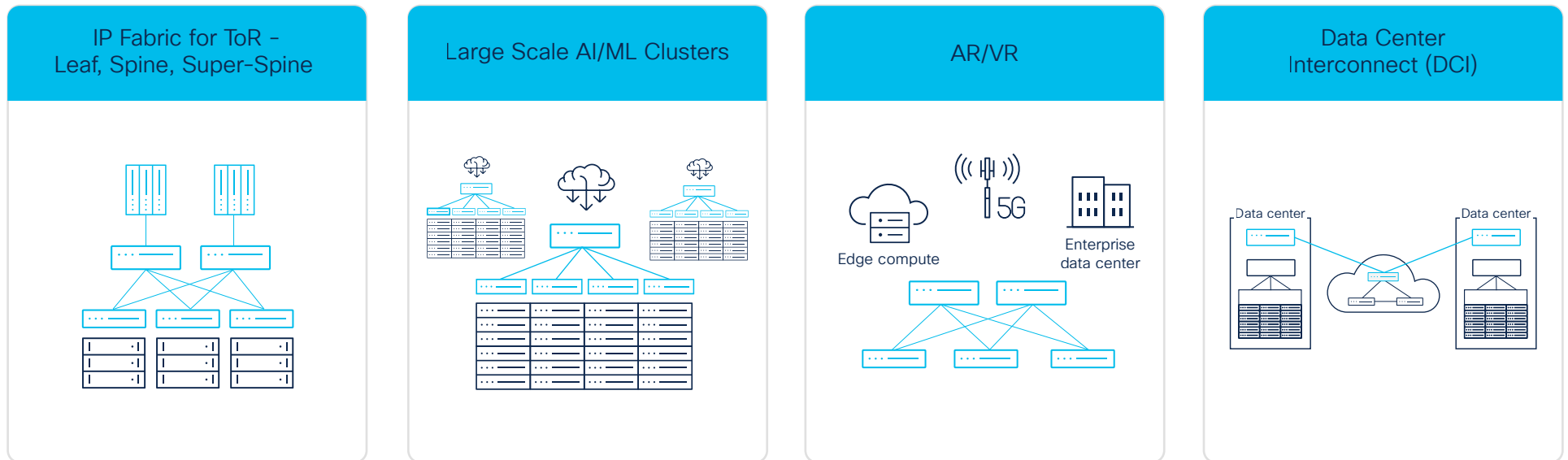
3. Scalability: Systems and software need to be able to meet current and future demand without service impacting upgrades.

4. Choice: Web scalers need to be able to operate their data centers as they see fit with choice of NOS, hardware, and a silicon that can perform to their standards.

The Cisco Mass-scale Infrastructure for Cloud portfolio focuses on helping web scalers grow revenue, optimize costs, and mitigate risk. It accomplishes this by improving operational efficiency, increasing the utilization of assets, and reducing the network's environmental impact. The key to realizing these business outcomes, especially at massive scale, is simplicity. The innovations introduced with the Converged SDN Transport architecture extend to data center deployments and help satisfy the four tenets above with a reduction in the number of moving network parts to provide a significantly simpler, automated, and programmable architecture that delivers immediate savings and sustainable operations into the future.

Figure 1. Mass Scale Infrastructure for Cloud for a variety of network use cases

Web Scale Data Center Use Cases



The Cisco Mass-scale Infrastructure for Cloud portfolio offers

- 1. Performance:** The Cisco 8000 Series is built on the groundbreaking Cisco Silicon One and offers massive performance in a small 7-nm ASIC platform. Web scalers can reduce their footprints, simplify operations, and scale to meet increasing traffic demands. Innovations in pluggable optics will help drive the simplification of the data center.
- 2. Uniformity:** Web scalers can have uniformity in ASICs with Silicon One. They can have Cisco 8000 Series in the transport core and data center networks, and they can have a singular NOS with IOS XR throughout.
- 3. Scalability:** The expansion of the Cisco 8000 Series to include 800G keeps us at the leading edge of network expansions. These platforms offer the ability to scale from 100G to 400G to 800G without the need to rip and replace and align with the network's upgrade cycle. With available programs to improve cost management and flexible port licensing, upgrades become a programming change to a port versus a truck roll for physical changes or onsite configurations. This allows for more predictable cost structures.

Learn more

Meet your network—simplified.

[Mass Scale Infrastructure for Cloud](#)

4. Choice: We provide flexibility and can help with investment protection. The portfolio offers choice in NOS with IOS XR or SONiC, allowing web scalers to use what is best for them. Silicon One also offers the ultimate choice for web scale providers. Providers can buy the Cisco 8000 Series with Silicon One, or they can use Silicon One within a custom system. This disaggregation of NOS and ASIC means providers can choose what best suits their environment.

This new hardware and technical innovations will have the greatest impact on the following web scale use case deployments:

- IP fabric for ToR/leaf, spine, and super spine roles
- Large-scale AI/ML clusters
- Augmented Reality (AR)/Virtual Reality (VR)
- Data Center Interconnects (DCIs)