Cloud networking is a compelling way to deliver web-based and non-web-based applications that better utilize the physical infrastructure, while lowering costs by moving from silos of expensive customized servers to racks of cost-effective ones.
Overview

Cloud computing is a compelling way to deliver web-based and non-web-based applications that better utilize the physical infrastructure, while lowering costs by moving from silos of expensive customized servers to racks of cost-effective ones. Businesses of all sizes are leveraging cloud infrastructure for collaborating with their customers, partners and suppliers. Due to the elastic nature of clouds, a particular application may be delivered from one physical location at a given moment but from a different, potentially far-away location in the same cloud at a later moment. Web-based applications in clouds include internet-facing enterprise applications serving thousands of users, as well as large-scale social networking sites with millions of subscribers.

Performance and scaling are very important for all modern web applications due to the high bandwidth needed on a per transaction basis (thanks to rich Web 2.0 page content), number of transactions and transaction rate. Oversubscriptions and bottlenecks of any sort can lead to long response times or even dropped connections, resulting in a frustrated subscriber base. Any outage related to network or network services can cause application outages, which can mean loss of revenue and hence the need for more robust networking.

What implications do these new interaction models have on L4-L7 application delivery services infrastructure? For one, a high degree of scalability and handling of peak loads will be required as larger sets of users access cloud applications. Second, the always-on nature of cloud applications requires resiliency in L4-L7 cloud networking services. Third, the agile nature of cloud applications themselves require that cloud-networking services be dynamically provisioned and “follow” application workloads wherever they meander. And finally, the fundamental cloud tenant of efficiency requires that cloud network services be shared across different applications and user communities without unduly increasing the physical infrastructure and/or its complexity.

New Deployment Model for Cloud Networking Services

Cloud networking services are built with two canonical components:

- Cloud Network
- L4-L7 Service Network

A robust cloud networking foundation based on Arista’s 71xx series of wire-rate, low-latency and high density 1Gb/10Gb Ethernet switches is vital. The Arista cloud network design is built using a two-tier spine-leaf architecture that maintains uniform cross-sectional bandwidth. Unlike the highly oversubscribed design in traditional data center networks, latency and application response time in the Arista cloud network is significantly improved by 10-20X, minimizing delays due to queuing and/or dropping of TCP packets.

A typical L4-L7 network is built via separate physical constructs of appliances in the data center. To extend this approach to a cloud environment, a more in-band network deployment model, i.e., one that is highly congruent with the agile and efficient nature of clouds, is required.
In Arista’s cloud networking, L4-L7 services can be embedded directly in the network fabric itself, as shown in Figure 1. In effect, L4-L7 application services are embedded in the switch itself, through Arista’s EOS (Extensible-OS) Service Agent called ESA.

With the EOS Service Agent (ESA), cloud-networking services can be implemented by simply loading software on Arista’s 7xxx switches and enabling the functionality. An alternative model is adding service modules in a modular Ethernet switch. Modules built under this legacy enterprise approach tend to use custom hardware and proprietary software, and consume expensive switch slots. And even with this high level of investment, modular switch-based solutions do not provide the application-layer awareness and intelligence necessary to improve the performance or availability of web applications, nor do they meet the requirements of modern virtualized data centers and cloud environments.

It is the innovative, open and extensible system architecture of Arista 7xxx series of switches that make this deployment model possible. The logical system architecture is shown in Figure 2. The Arista Extensible Operating System (EOS) allows network switching functionality and L4-L7 network services functionality to work together in tandem without impacting switching functions on data-plane, control-plane or management-plane level.
Workflow for Cloud Networking Services

Consider an example where the embedded cloud network services are available on an Arista 7xxx leaf switch. When user initiates a web request, as depicted in Figure 3, the corresponding Arista switch that front-ends the web application forwards that request flow to ESA. ESA provides the necessary services, e.g., server load balancing to identify the right application resource and translate the destination virtual IP address (VIP) to a real IP address (RIP). It then sends the flow back to the switch for forwarding to the appropriate virtual machine on which the web application instance resides.

![Figure 3: Workflow through ESA](image)

Attributes of Cloud Networking Services:

Embedding L4-L7 application-level intelligence in a resilient cloud network provides the following key benefits:

1. **Latency and Scale**: The ESA approach reduces overall network latency compared to a physically separate appliance, as packets do not need to traverse physical ports where buffering and other interface related latencies are added twice (switch to appliance and appliance to switch). Since each rack has its own top of rack 7xxx series switch with ESA, scaling is built in. Each time a new rack is added, it is ready for both cloud networking and cloud network services.

2. **High Availability (HA)**: This resilience of Arista’s EOS architecture is inherently extended to ESA where ESA is considered a separate application with its own fault-isolated container. ESA can be loaded, enabled, disabled, removed or upgraded without any impact to switching functionality. No new architectural considerations need to be made when enabling ESA for such a rack, as the “rack-level” failure domain remains unchanged. Alternatively, for higher availability deployments, two 7xxx switches can be configured as an active pair and the corresponding ESAs can also establish HA relationship for 1:1 application service availability.

3. **Agile Operation**: As workloads migrate within the embedded cloud network service infrastructure, L4-L7 services continue to be applied to migrated workloads through ESA. Thus, servers or even an entire rack can
be retired while maintaining consistent application interaction experience. As web applications scale up or down on-demand (i.e., as virtual machines for a given application are added or deleted), ESA continues to provide L4-L7 services to these applications in the network and virtual machines (VMs) get the same service wherever they are.

4. **Lower cost**: When there are no physical appliance devices to deal with in an embedded cloud networking services, the overall infrastructure cost is lowered due to better utilization of rack space, reduced power consumption and cooling, no cabling costs incurred for wiring switch and appliance, freed up switch and optical ports cost, and savings on overall hardware cost.

5. **Operational Simplicity**: Arista EOS simplifies the overall management of cloud networking services. For example, an IT administrator can control and/or monitor ESA (e.g., enable/disable, health check) directly from the switch CLI and also manage the ESA software instance (e.g., load, remove, upgrade) without impacting switching operations.

**Citrix NetScaler VPX on Arista Switches**

Arista Networks and Citrix have formed a technology partnership to integrate Citrix® NetScaler® VPX virtual appliance technology into Arista’s 7xxx series of switches for providing L4-L7 application delivery services in the network fabric. NetScaler VPX provides the same features as Citrix® NetScaler® appliances, but in a virtual appliance. NetScaler VPX virtual appliance is integrated as an ESA agent on Arista’s 7xxx switches. The switch and virtual appliance combination, as illustrated in Figure 4, now provide:

- Ethernet connectivity to all physical servers in a rack
- Line-rate, low latency, resilience switching (with IEEE 802.1Q trunking and IEEE 802.3ad link aggregation)
- High-speed L4-L7 load balancing and intelligent application and server health monitoring
- Much faster application response time through web content caching, compression and TCP optimization
- Better security with an ICSA-certified application firewall and acceleration of SSL encryption and decryption

![Figure 4: Cloud Network services L2-L7 architecture based on Arista switches with Embedded Citrix Netcaler VPX Agent](image-url)
The NetScaler VPX virtual appliance is enabled through EOS’ command line interface (CLI). All NetScaler VPX instances, Arista-embedded as well as non-embedded, can be then managed and monitored centrally using Citrix’s Command Center.

IT organizations now have the flexibility to deploy physical appliances, virtual appliances and/or switch-embedded appliances to optimally architect their cloud and data center networks as well as minimize the total cost of operation (TCO). For example, an IT organization may deploy a combination of physical and ESA-based NetScaler VPX virtual appliances, depending on the granularity of coverage across its entire web application infrastructure.

Summary

The Arista 7xxx Series of switches feature the industry’s highest density 10Gb Ethernet switching solution with industry’s best footprint, latency and scale. When internally coupled with Citrix NetScaler VPX virtual appliance solutions, the combination accelerates web application performance and secures applications from attack, all while enabling cloud computing.

NetScaler is an industry leading web application delivery solution that accelerates performance, provides L4-7 traffic management, protects applications and off-loads servers from compute-expensive tasks to ensure application availability, increased security and substantially lower costs. It reduces TCO of web application delivery, optimizes the user experience and makes applications run faster.

A thoughtful and seamless L2-L7 cloud networking services foundation with breakthrough price-performance enables 1Gb/10Gb Ethernet to be deployed everywhere in the data center, which can significantly improve server utilization and consequently data center power efficiency without compromising application performance.