Cloud Networking:
Design Patterns for "Cloud-Centric" Application Environments

January 2009

The Emergence of Cloud Computing

Cloud Computing took the infrastructure landscape by storm in 2008. The economic argument for the cloud is unquestionable: increased server utilization, optimized power and cooling, outsourced management of complex systems, elastic scaling capabilities to enable instantaneous scale up and down, and rapid app development.

Public Clouds, Platform-as-a-Service (PAAS), and Enterprise Private Clouds

While the "why" behind cloud computing is clear, the "where" is still a work-in-progress. The early "Public Cloud" offerings from Amazon.com have taken a building block approach to the problem. In the Amazon web services paradigm, customer access to their compute and storage nodes comes closest to having an on-site box. Indeed, the Amazon offering is much more a *deployment* platform than a *development* platform. However, it does not mandate any application changes and enables the user to mix and match infrastructure services while meeting the economic goals set forth above. One can expect to see every hosting company in the world offer an analogous service. The Platform-as-Service (PAAS) offerings from Google (AppEngine), Salesforce.com (Force.com) and Microsoft (Azure) provide a Cloud Computing environment at a higher abstraction layer and include capabilities like a database API and integration API. These are true development platforms. The advantage to the end user is that the platforms are built to natively scale. However, in order to take advantage of the native scalability, the user must live within the constraints of the app dev environment. While the public cloud and PAAS paradigms provide compelling economics, they lack the singular quality that is central to many enterprises – control. Enterprise clouds aim to achieve the same level of elasticity as cloud providers provide without raising the security, manageability and support issues that typically emerge with any 3rd party provider. Central to the enterprise cloud paradigm is an orchestrator that controls resources across virtualized compute, storage and network nodes in a policy-based manner. Offerings such as VMWare's vCloud initiative accomplish this by orchestrating on-site virtualized environments with off-site virtualized environments.

Cloud Paradigms and Cloud Networking

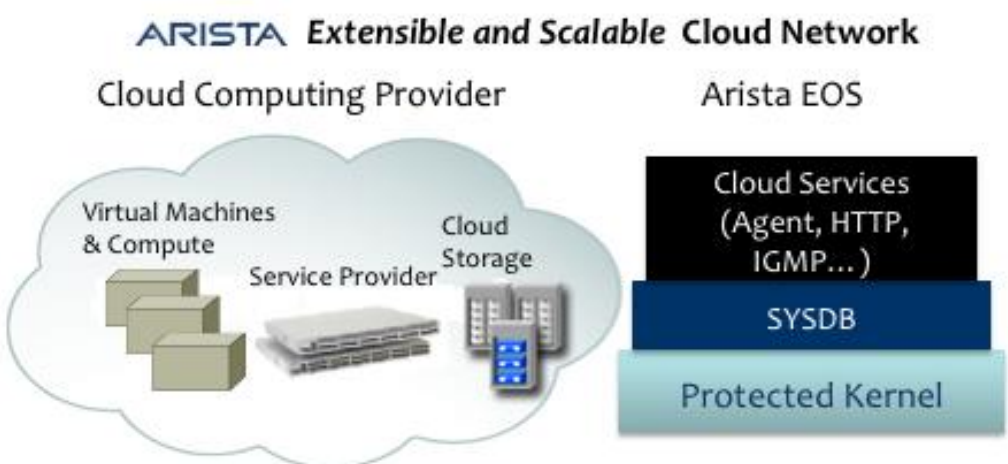
While "where" a cloud environment runs remains varied, "how" an application can be built, deployed, run and scaled in the cloud in a manner that integrates seamlessly with existing infrastructure and application paradigms is a significant challenge and a source for innovation. Central to the re-thinking of cloud infrastructure is the network. In the context of a public cloud, PAAS offering, or enterprise cloud - the network provides the connective tissue between the various infrastructure tiers and is the critical scaling factor in an environment that must be built to support an order of magnitude more scale than today's data centers. The attributes of this cloud network – scalability, low latency, reliability, manageability and low cost require a re-design of the basic switching fabrics prevalent in the data center. This paper aims to outline 7 application and infrastructure use cases that are representative of the new clouds, and details the role of the "Cloud Network" in those environments.

Design Pattern #1: Cloud Hosting and "PAAS" Providers

Key Challenges

The extraordinary scale that Cloud hosting and PAAS providers face cannot be understated. As Google crosses 1M+ servers and the bandwidth running on Amazon's EC2/S3 service exceeds that of Amazon.com, any network to serve public clouds and PAAS providers must be able to scale to the demands of the compute environment. Central to the cloud computing architectural paradigm is the ability to tolerate single node failures without impacting the overall system and for failure scenarios in any infrastructure element (compute, network, storage) to not meaningfully impact availability. However, today's network architecture binds its state environment to the processing nodes – as a result increasing the time for recovery in the event of failure scenarios.

Network architects must re-think the traditional close coupling between state information contained in network nodes and the processing engine. Indeed, network state information needs to be separated from the processing engine to enable automatic scale-out and resiliency.



Solution

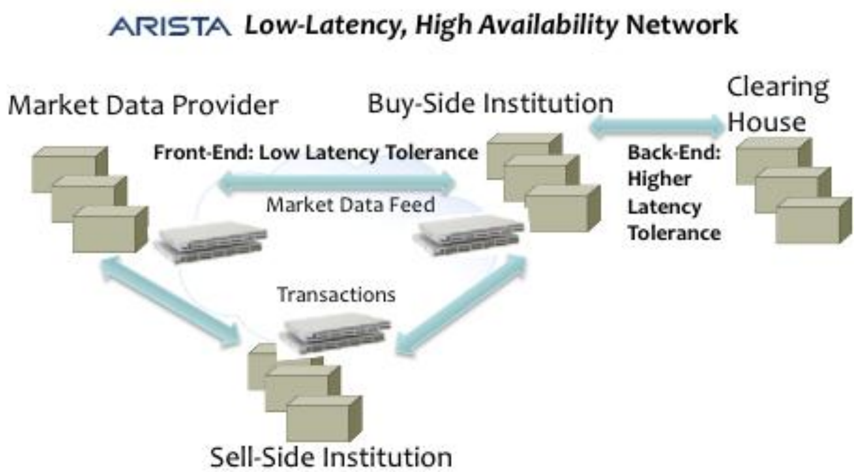
Arista's revolutionary EOS software system is at the heart of a scalable, resilient and fault-tolerant network. The software architecture provided by Arista's EOS separates state from processing resources and enables in-service upgrades. It also enables fault containment and stateful restart. The system leverages *NIX based tools for the compute cloud to provision the network.

The EOS architecture marks the first time that networks have embraced the paradigm that applications have embraced for a long time – separation of state from processing.

Design Pattern #2: Market Data and Electronic Trading

Key Challenges

The Financial Services industry has had a tumultuous 2008. Nonetheless, the rapid growth of electronic trading is a trend that is here to stay. Financial networks have become high performance clouds – across market data providers, buy side institutions, sell-side institutions, exchanges and regulators. Although not typically described as "clouds," they share the common attributes of large scale, rapid infrastructure growth and applications that are delivered on-demand. The industry is challenged by the extraordinary growth in the volume of electronic transactions and the enlargement of content per "tick" and per "trade." While transaction volumes and sizes grow, no institution can afford to sacrifice network latency or reliability. Indeed, these attributes are directly correlated to trade execution, "best price" and revenues. Latency-tolerance is particularly low in the "front-end" of trade processing where firms have gone as far as co-locating with exchanges to receive direct feeds. Central to the scalability of these networks is rapid broadcast of data to a large number of parties. This requires robust multicast transmission to an increasing number of groups. Faced with extraordinary scale and performance pressures in their environments, financial institutions must respond in a manner that continues to shave off latency, increase reliability, support growth of multicast groups and do all of this while keeping costs down.



Solution

Arista's low-latency switches permit sub-microsecond latencies, providing advantages to every member of the financial services ecosystem. At every node, shaving microseconds off provides a competitive advantage and reduces risk. Financial institutions have typically resorted to Infiniband-based solutions to achieve latency gains. The Arista alternative provides an alternative to traditional store and forward Ethernet latency, while avoiding costly islands of Infiniband. It also provides robust, reliable multicast support for upto 2,048 groups to enable scalable multicast deployments. The Arista solution is particularly applicable to the front-end of trade processing where latency is crucial. In co-located environments, a high performance Arista Layer 2 network can further optimize latency.

Design Pattern #3: High Scale Web Environments

Highly scalable web environments such as content sites, e-commerce sites and gaming properties have had to effectively build cloud computing environments of their own in order to scale while maintaining page load times. In fact, page load times are directly correlated to user engagement and ultimately to revenues. Tolerance among users for response time on the web declines year over year with increased penetration of broadband and increased reliance on the web for business or life-critical activities. Increasingly demanding users are coupled with the increase in personalization and dynamic content that is essential to delivering a highly relevant end-user experience. Personalization and dynamic content delivery causes increased pressures on the database layer where most transactions get hung up. Database bottlenecks have fostered the use of object caches such as Memcached where 90+% of latency comes from network I/O. Indeed, the scale out demands are no longer just limited to the front-end of a web environment but also to the back-end which requires a dense interconnect.

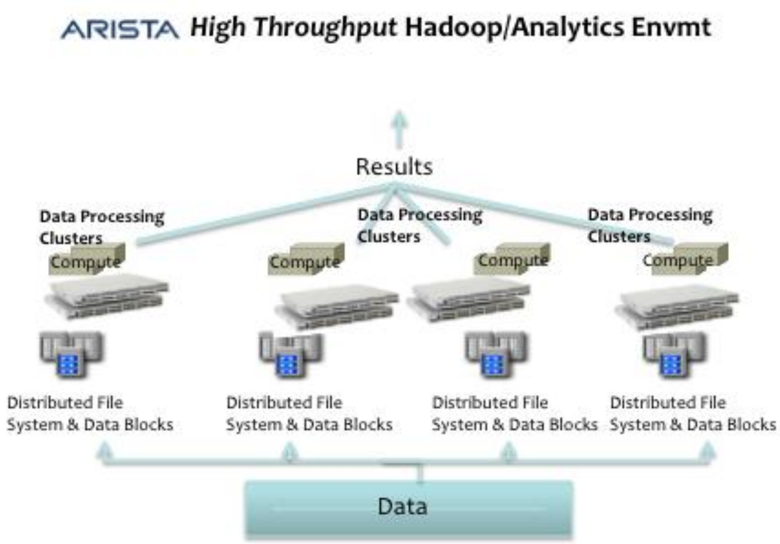
Solution

Given the significant dependence of web site performance on object caches like Memcached and the importance of network I/O in Memcached seek times (often accounting for 90+% of round trip times), Arista's switches with 10Gig capabilities will have a significant positive impact on page load times and user experience for the 90+% of requests that are often retrieved from the cache in applications like Facebook.com. In addition, as other layers require scale out capacity for both front end and storage connectivity, Arista's industry leading density and latency provides a cost-effective solution.

Design Pattern #4: Analytics, Hadoop and Large Scale Data Processing

Key Challenges

The growth of the web has resulted in an explosion of data. Whether for creating search indexes, matching personalized ads to videos, analysis of genomics data, computation of credit card scores or processing of call data records in telecom networks, large-scale data processing has become central to most businesses. Traditional on-premise data warehouses such as Teradata or business intelligence systems such as Business objects are now faced with the challenge of processing more data faster. They are also faced with emerging competition from newer commercial offerings such as Greenplum, Aster Data and others. Indeed, speed of job execution is intimately tied to user experience. Large-scale data processing systems often leverage newer paradigms such as Map/Reduce and Hadoop to process data effectively. Hadoop and Map/Reduce share a common architectural approach of dividing the processing task into several data processing clusters and co-locating the computing with the associated storage or file system. The results of the data processing are then merged to provide the results. Central to the success of such an approach is a fast, symmetrical network that can keep up with data processing volumes and provide high throughput within a data processing cluster and across clusters.



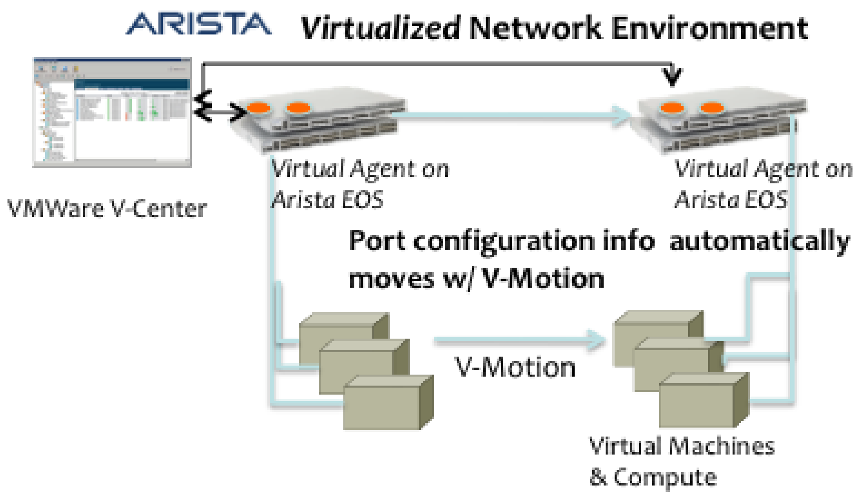
Solution

A high throughput 10Gig network is essential to the success of an analytics or Hadoop-style deployment. The typical "data processing clusters" architecture of a Hadoop environment requires co-location of processing and data as well as a fast intra-rack interconnect. Arista's top-of-the rack switches provides industry-leading throughput while maintaining a best in class price/performance offering.

Design Pattern #5: Virtualized Environments

Key Challenges

Virtualization is at the heart of most clouds. Virtualization provides clouds with a number of advantages – consolidation within servers enables higher asset utilization, faster start/restore times to bring up applications more quickly, and a programmatic means of controlling resources in a manner that is decoupled from the hardware. VMWare's "V-Motion" capability to move virtual machines from one node to another provides an important means of dynamically rebalancing compute resources. However, the use of "V-Motion" does not automatically rebalance or configure the network. Indeed, port configurations are physically tied to a switch and require manual action to re-bind to a new switch in the event of a VM migration. Equally, management of vswitches in a manner that integrates with the physical network is a challenge. From a bandwidth perspective, classic hierarchical network architectures of access to aggregation can bottleneck at certain points as compute resources move around dynamically. Large Layer 2 networks are not possible with traditional 3 tier network architectures. Essentially, the network remains the central bottleneck with VM migration both from a run-time and a management-time perspective. With the emergence of "hypervisor bypass" NICs, one can expect the challenges to compound.



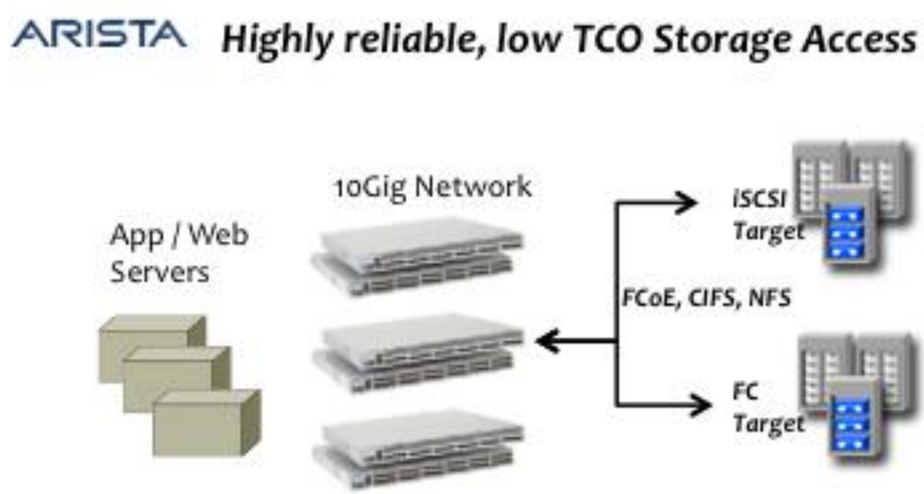
Solution

Arista's EOS software provides an extensible platform for the development of tools to automate V-Motion deployments that integrate seamlessly with V-Center. A virtual agent on the Arista system communicates with V-Center to automatically move network configurations between switches in tandem with a V-Motion event. Equally, the Arista solution permits network administrators to control vswitches via the same CLI that controls the physical network. In addition, Arista's dual 1Gig/10Gig switches enable a balanced network that maximizes resource availability across a large layer 2 environment in a manner that is optimized for buffer size, latency, interconnect speeds and fabric capacity. Finally, 10Gig capabilities provide "future-proofing" for impending traffic explosions that are likely from hypervisor bypass technologies.

Design Pattern #6: Cloud Storage

Key Challenges

As a number of cloud providers provide cloud storage APIs and enterprise clouds aim to expose the data to a range of applications, scalable and accessible storage becomes essential. The explosion in videos, images, documents and compliance requirements compound storage requirements. The growth in unstructured data particularly threatens traditional storage networking and endpoint patterns. The scale of storage deployments and demands for affordable TCO is particularly driving the adoption of iSCSI. In addition, convergence between networks and storage is driving the adoption of FCoE standards. Reliable access to a range of storage media types through a range of access protocols is essential. Network performance continues to be a significant bottleneck in storage throughput and IOPS. The emergence of SSD technologies makes a fast network even more crucial. In addition, a number of open-source and cloud storage solutions go beyond CIFS and NFS and provide a means to scale out storage on raw disks (MogileFS, Lustre, ZFS, GFS). These newer file access and storage techniques rely on stringing together unreliable disks via a reliable network.



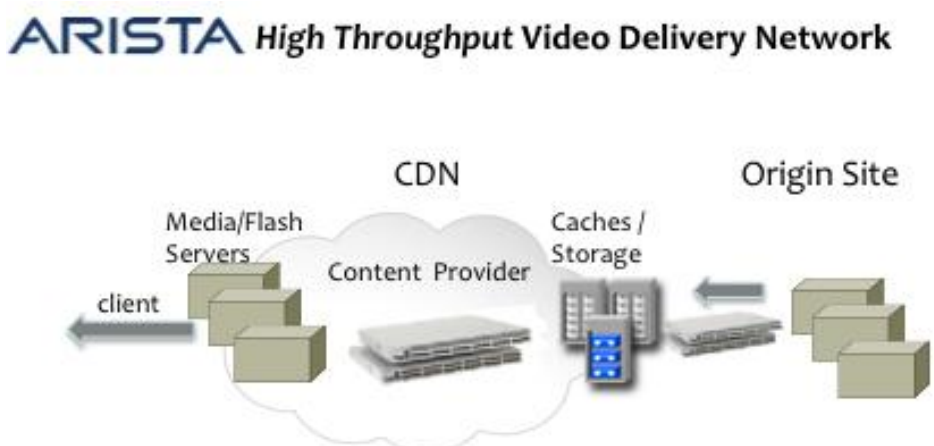
Solution

Arista's 10Gig solution provides the highest available throughput in the market and ensures that storage endpoints are accessed fast in both iSCSI and FCoE environments. It supports reliable block and file storage access and remains consistent with a unified Ethernet network that can serve as a platform for newer file systems to access raw disks and applications to access SSD devices.

Design Pattern #7: Video Content Creation and Delivery

Key Challenges

The growth of online video, both professionally created and user generated, is driving massive growth in traffic at Content Delivery Networks (CDNs) and media properties. The transition from SD to HD video rapidly fills 1 Gigabit infrastructure and ultimately causes choke points in traditional network architectures. Content developers require 10Gig interconnect for digital media creation, rendering and delivery. As CDNs and media properties compete for users, the quality of the video viewing experience and the underlying network has emerged as a central competitive playing field.



Solution

Arista's switches both at the origin media property and the CDN edges provide the fastest throughput available at 960Gbps and 720Mpps in 1 RU. As CDNs offer live streaming products and content providers aim for next generation editing and rendering, Arista's switches will provide a robust network platform for quality video experiences.

Summary

The use cases and design patterns presented in this paper highlight two apparently contradictory facts – on the one hand, Cloud computing provides a dramatically different computing paradigm, on the other hand the use cases that will drive cloud computing are familiar ones.

From a networking perspective, cloud computing does require a re-thinking of data center networks. While the Cloud Network of the future will likely embrace a number of switching concepts of the present and the past – like standardizing on Ethernet, the architecture of those network nodes will be different. Networks will be more balanced and require 1Gig/10Gig capabilities. Throughput and latency, as they always have been, will need to be industry leading. Economic imperatives will require networking vendors to re-think their hardware choices and will favor merchant silicon. The software that runs those switches will need to be extensible to allow for integration into cloud orchestration systems and vendor APIs. The switches will need to separate state from processing. Failure scenarios and manageability will be as central to design patterns as performance.

From a use case perspective, cloud computing will provide choices in application paradigms. Users will have the choice of adopting the newer computing architecture in-house (private clouds), through PAAS or Public Cloud Vendors. Regardless of the choice, the applications that run in the cloud will be more web and user-centric but will remain consistent with the trends we have seen in data centers for several years – more data, more stringent user latency requirements, more video, more virtualization, more personalization and more business on the line to deliver all of these things.

As Cloud Networking gains prominence across the industry, an important pillar of "how" cloud computing will really work will be answered. The good news is that the march towards a network-centric vision of the future will remain in good hands – riding on the twin pillars of a new computing architecture and a new network architecture.