Data Center Services System

Creating a Tight Linkage Between the Data Center Cloud and Core Network

Executive Summary

The convergence of video, mobile services, and cloud technologies is impacting IP next-generation networks in unprecedented ways. Consumers now expect personalized video content on the medianet delivered when, where, and how they want it. Mobile devices are exploding in both number and capabilities, as are the applications that they access. And business customers now look to the “cloud” for functionality that they used to provide internally from their own data centers. The result is a significant shift in network traffic patterns and tremendous strains on providers’ networks.

Traffic no longer moves simply from the subscriber into the network and then out again. New and emerging services are causing substantial traffic **between** data centers within the network in addition to flows to and from the subscriber - multidirectional traffic. Network operators need to recognize and respond to these changing dynamics, and optimize their networks to best support these new traffic patterns.

This white paper outlines Cisco’s Data Center Services System (DCSS), which can help service providers address these needs. DCSS brings two main benefits:

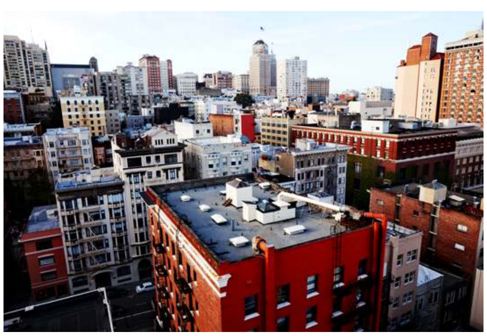
- It makes the network infrastructure more operationally effective
- It allows providers to capitalize on new revenue opportunities

The paper describes how DCSS delivers a set of technologies that combines the strengths of both the next-generation IP network and the virtualized data center.

Initial DCSS technologies include:

- The Network Positioning System (NPS)
- Cloud VPNs (CVPNs)

These technologies and their benefits are described in the pages that follow.



© 2012 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public Information.

© 2012 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public Information.

The Cloud, Video, and Mobility: Data Centers Change the IP NGN

Services in the 21st Century

The last several years have yielded tremendous changes in the types and nature of services offered by providers in virtually every market.

The business services market, from enterprises to small businesses, has seen the emergence of cloud computing as a viable revenue opportunity. Now providers from all sectors are rushing to stake their claim in the cloud services market. Providers with heritages as varied as e-commerce companies and incumbent wireline carriers all have legitimate and compelling offers in this space.

Video services are also changing dramatically. This was once the domain of producers and distributors focusing exclusively on entertainment services, such as major studios, broadcasters, and cable operators. Now, new content forms such as user-generated content and Internet video, and distribution mechanisms such as streaming and mobile, have opened the door for all types of providers to compete for viewership.

And the mobile market has evolved from offering basic voice communication services to supporting a robust ecosystem of devices, applications, and experiences that have become an indispensible part of day-to-day life in the 21st century. Again, all types of providers are vying for relevance in this rapidly evolving ecosystem.

There is one common thread through all three of these trends: the emergence of the data center as an essential element in delivering new services.

Traffic Patterns Become Multidirectional

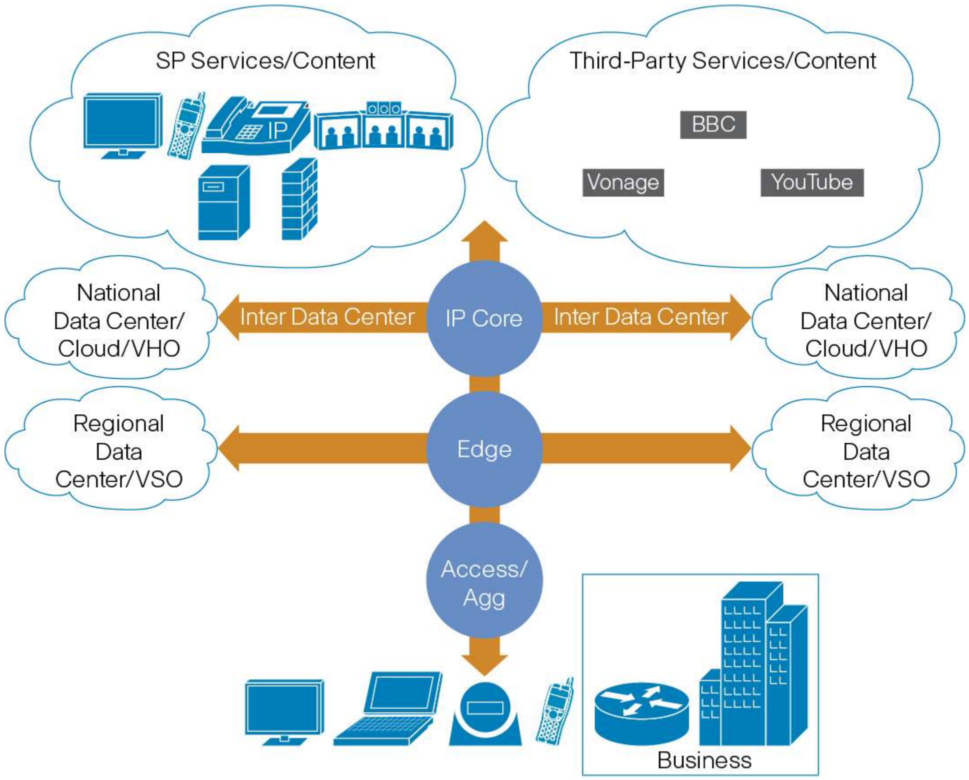
Whether it’s providing a place for enterprise customers to run virtualized workloads in the cloud, the caching and distribution of video to a multitude of possible endpoints, or the “app store” that offers mobile consumers new functionality for the latest handheld devices, the role of data centers is as much a part of these services as is the network.

Of critical importance for network operators is the fact that these data centers - and the services that they help to create - are causing big changes in the network. These changes are characterized by new, multidirectional traffic patterns that differ substantially from the one-dimensional, network-to-subscriber flows seen in years past.
The following are a few examples of these new patterns.

With cloud computing services, providers or their partners are now connecting subscribers to data centers that are “in the cloud.” Workloads and data now move not only from one subscriber endpoint to another, but also to these “cloud” data centers. More importantly, they may also move among data centers in the cloud - perhaps completely transparently to the subscriber.

© 2012 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public Information.

Figure 1. Next-Generation Internet Multidirectional Networking



Video content now comes from more sources than ever before, and the rate of content growth is staggering. Viewers have more choice in what they watch and how: on the TV in the living room, streamed to the laptop or game console, or on a handheld device. As a result, providers are looking to build the medianet, an intelligent network optimized for rich media. On it they need to move a tremendous amount of content among their data centers: from national media centers for content acquisition and long-term storage, to regional data centers for local ad insertion and distribution, and ultimately to points of presence where frequently-accessed content is cached to provide the best viewer experience. Cisco innovated with the ASR 9000 incorporating the Advanced Video Services Module (AVSM) that enables terabytes of streaming capacity at the aggregation edge while simultaneously offering content caching, ad insertion, fast channel change and error correction.

Mobile devices are growing increasingly more powerful, but the true source of their power comes from their connection to data centers. As mobile internet grows the volume of data traffic swells too. Whether it’s turn-by-turn directions, push-email from the enterprise, or a favorite set of applications from the “app store,” the subscriber’s experience stems from the operator’s ability to move information among data centers throughout its footprint, not just out to the handset.

These are all examples of the new dynamic of traffic flows **between** data centers - in addition to the traffic from the network to the subscriber. This is the new twist with which operators now have to contend: the multidirectional nature of network traffic patterns.

© 2012 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public Information.

IP NGN 2.0 and Unified Service Delivery: Dynamic Response to Changing Traffic Patterns

Over the course of the past decade, Cisco has been helping service providers transform their businesses and operations through our innovative IP Next-Generation Network (IP NGN) architecture. The original IP NGN architecture, or IP NGN 1.0, was characterized by the transformation of dedicated networks (TDM voice, ATM transport, Frame Relay data, etc.) into next-generation, all-IP networks at the core, aggregation, and access tiers.

Responding to the dramatic recent changes in the market, Cisco has evolved the IP NGN architecture into its next iteration: IP NGN 2.0. This new architecture, instead of focusing exclusively on the network, focuses more broadly on transforming the experiences that providers are able to deliver to their subscribers. Rather than simply delivering a few or several relatively disconnected services (voice, video, data, mobility), providers can now deliver rich and fully integrated experiences across their service portfolios.

Whether supporting mobility services infused with video, cloud-based managed services integrating computing and communications, or entertainment services that bring together television and online experiences, IP NGN 2.0 uses an all-IP foundation with innovation at the infrastructure, service, and application layers.

A key component of the IP NGN 2.0 architecture is the Cisco® Unified Service Delivery (USD) solution. USD combines the strengths of the service providers' two greatest assets - the network and the data center - to enable the experience transformation. USD unites the data center and the network so that the whole is greater than the sum of the parts.

The Cisco USD solution provides a common baseline infrastructure for delivery of services from the application, through the data center, across the network, and out to the subscriber. Much like the common IP network did in the original IP NGN architecture, this common USD baseline enables providers adopting the IP NGN 2.0 architecture to deploy different services on a single, horizontal platform that spans their service portfolio.

This single platform for service delivery brings providers three main benefits:

- New revenue opportunities through unique, differentiated services created by integrating products across lines of service, now possible with a single technology foundation
- Quicker time to market with new services through reuse of existing equipment and operational best practices
- Reduced operating costs and capital expenditure by elimination of duplicate infrastructure

There are many discrete elements involved in integrating service delivery through the data center and across the network. Making these elements work in harmony is a fundamental tenet of Cisco's USD solution.

The Data Center Services System brings new capabilities to the IP NGN 2.0 architecture and the USD solution. It tightly integrates capabilities of the network and the data center to bring unique value to an end-to-end Cisco infrastructure.

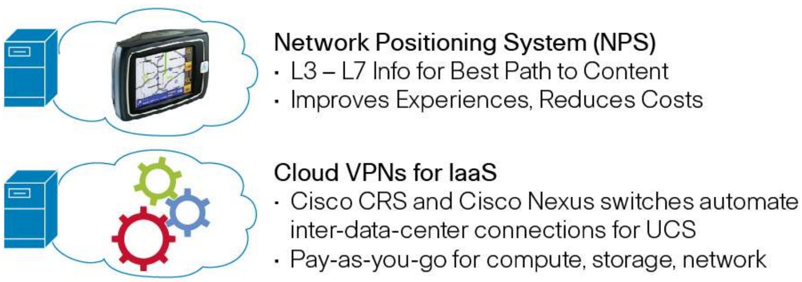
© 2012 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public Information.

Data Center Services System

As traffic becomes multidirectional across the next-generation IP network, the network layers need to provide enhanced visibility into the data-center-based services flowing across them. With extensive experience across video, mobile, and cloud technologies, Cisco is augmenting its IP network infrastructure platforms with the Data Center Services System (DCSS). There are two key technologies initially included in DCSS:

- **The Network Positioning System (NPS):** Location-based network intelligence; enhanced proximity services
- **Cloud VPNs:** Automated VPN connectivity, especially tuned for cloud services

Figure 2. Data Center Services System



Network Positioning System: GPS for the Network

Locating applications, services, and content (files, movies, VoIP gateways/bridges, peers, servers, etc.) is a critical aspect of using cloud computing for service delivery. Different methods and technologies have been proposed but lack efficiency. The ability to dynamically add and move resources in the cloud requires enhancements to the traditional IP connectivity model. With the Network Positioning System, end-points are located in the network infrastructure beyond just physical closeness, which is usually dictated only by coarse metrics such as hops.

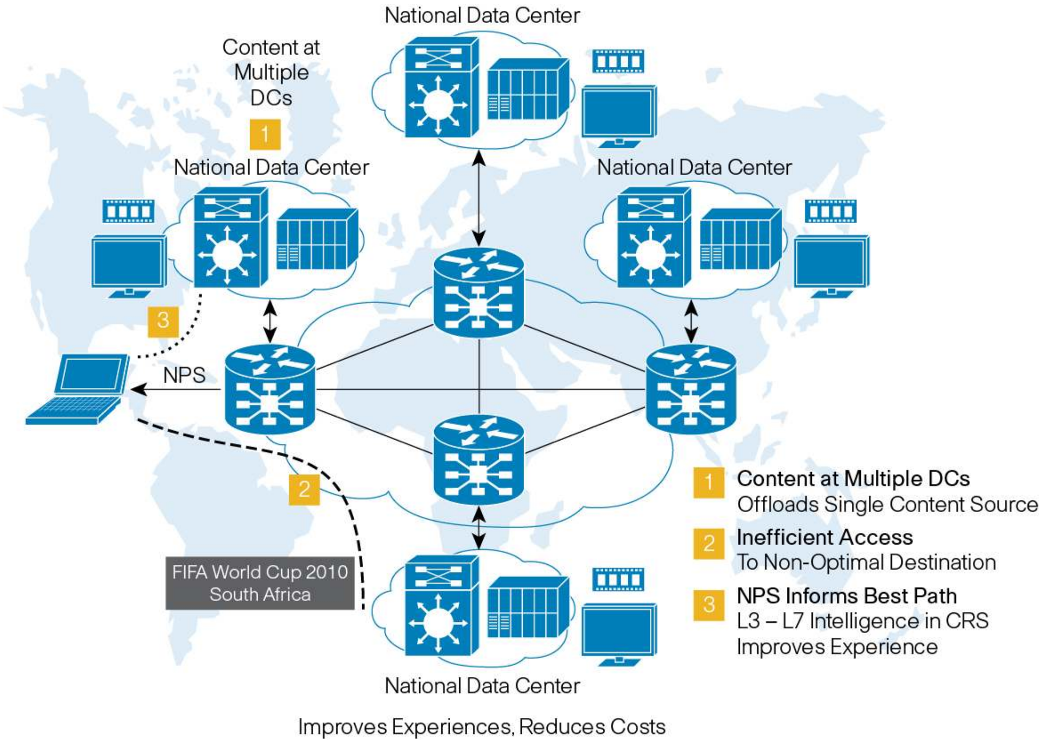
What It Is

The Network Positioning System essentially implements enhanced proximity services in the network. Proximity in the most general form answers the question: “Which destination targets in a given list are closest to the particular requester?” Various techniques such as Domain Name System (DNS), Round-Trip Time (RTT) measurements, and manual configuration have been used to determine proximity. However some of these approaches do not integrate with network routing or applications, while others have inflexible management or are simply unreliable.

A better strategy is to have the **network infrastructure provide recommendations for application layers** to optimize application traffic. These recommendations are based on accurate information like routing protocols (IGP/BGP), management statistics, and policy databases. In this regard, the Network Positioning System navigates the network similarly to how GPS, the Global Positioning System, navigates on the highway.

© 2012 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public Information.

Figure 3. Network Positioning System Use Case



As shown in Figure 3, NPS can improve the user experience while reducing costs. Consider the following scenario: Some compelling sports content originates in South Africa and resides in a data center there. A TV studio in the United States needs to rebroadcast the content after suitable editing. It's quite inefficient for the U.S. studio to access the content located in South Africa. Rather the same content is now available at multiple data centers around the world, including the United States. The core network aggregates all Layer 3-7 intelligence and learns the location of the content nearest the U.S. studio from the NPS. It notifies the requester when asked for this information. The U.S. studio now redirects its content requests to the national data center, thus improving overall infrastructure efficiency.

In essence:

- While routing computes the shortest path to destination at the source, NPS proximity computes the shortest path to the source from multiple destinations at the server. In short routing is contextualized.
- The server at the service provider network derives information based on topology, routing, state, performance, policy, and geographic information. As NPS' proximity functions work at Layer 3 and above, they can span multiple autonomous systems as well as MPLS VPNs.
- Proximity algorithms used in NPS aggregate and correlate/combine information from different sources.
- A standards-based proximity protocol is used between clients and proximity servers delivering location-based ranking services tailored by application-specific requirements.

Most of the standardization efforts for proximity are underway in IETF under the Application Layer Traffic Optimization Working Group (ALTO WG), of which Cisco is an active member.

© 2012 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public Information.

How It Helps

As network traffic requirements evolve, resulting in network architectures that relax the hierarchical constraints of reachability, the infrastructure requires the capability to locate content and users. This evolution is similar to the one in the peer-to-peer (P2P) space. For some time now, providers have been looking to actively mitigate the enormous bandwidth impact of these trends. With NPS, best paths are computed based on multiple criteria while being standards-compliant. Thus the overall efficiency of the network and data center or cloud infrastructure improves significantly. Additionally, now the service provider can use NPS as a generic service to a content delivery network or application layers, as well as to any other third-party or over-the-top (OTT) providers, opening up new revenue opportunities.

Cloud VPNs: Automating Network Connectivity

As workloads increase, say due to the financial year-end, businesses look to their providers for additional resources such as compute, storage, and network. Also, consumers may reach out to the latest rich-media content on the Internet, resulting in the "flash crowd" phenomenon. Providers may need to quickly replicate content across multiple data centers to avoid online gridlocks. These surges in demand could be temporary, urgent, or both. But they require the provider to rapidly fulfill resource requests across shared infrastructure. At the same time, the provider needs to be cognizant of security and quality of service (QoS) concerns.

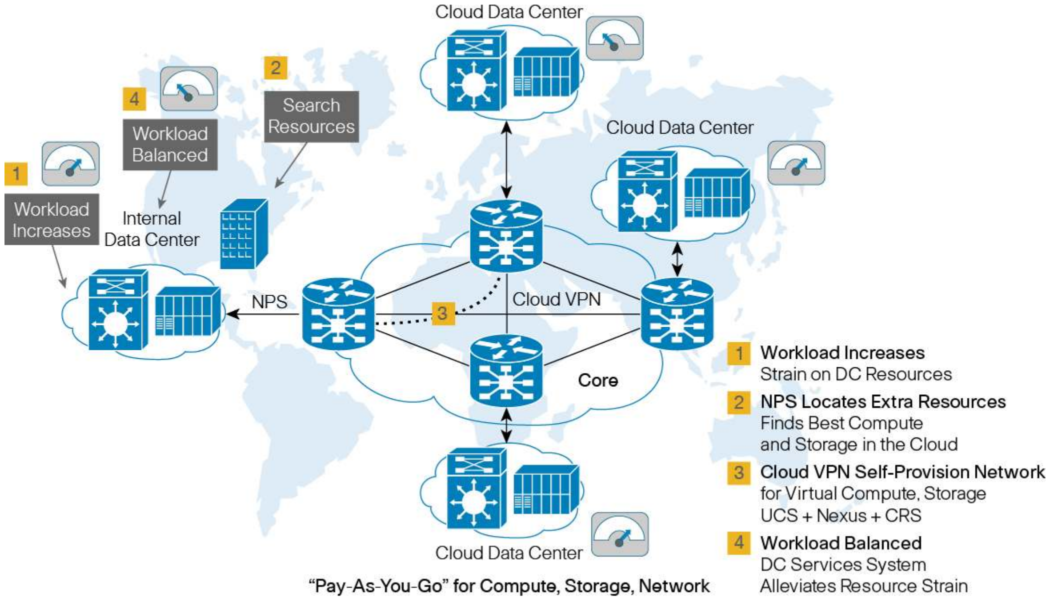
What It Is

Service providers can tune their networks to handle dynamic traffic patterns with Cloud VPN connectivity. As additional resources are demanded by the data-center-based or cloud-based services to handle increased workloads or "flash crowd" events, the network infrastructure interconnecting the data centers is capable of automatically providing connectivity on demand. The connectivity is based on existing technologies in the network such as Layer 3 or Layer 2 VPNs using IP/MPLS. This capability does not need any new connectivity protocols for data-plane traffic; rather it is only a control-plane mechanism using existing protocols. Thus the benefits of using well-established, deployed technologies in the core and edge of networks can continue to be accrued in this scheme. These benefits, derived by using Layer 3 IP/MPLS, include:

- Proven security
- Differentiated QoS
- Efficient traffic replication with multicast
- Rich traffic-flow accounting mechanisms (NetFlow)

© 2012 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public Information.

Figure 4. Cloud VPNs Use Case

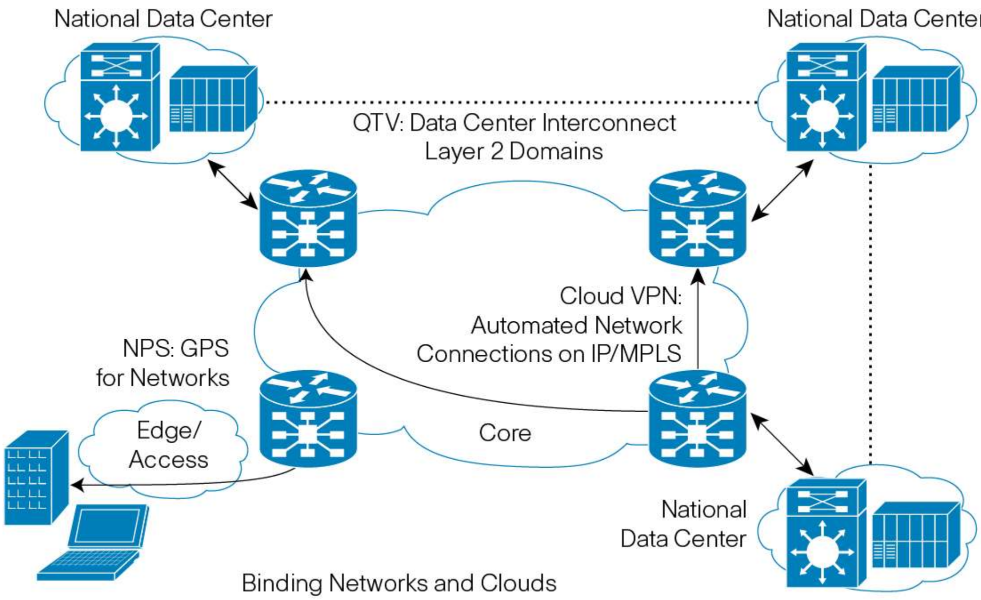


As shown in Figure 4, Cloud VPNs allow a "pay-as-you-go" business model for networks supporting compute and storage resources available in a cloud-based or on-demand environment. Consider, for example, how an enterprise's workload increases at the financial quarter-end. This places higher load on its internal data center resources such as storage and compute. The enterprise can initiate a request from the network, and the service provider can locate the data center within the cloud that can best meet the needs of the enterprise workload. The provider informs the enterprise of the best match via NPS. The enterprise requests that the provider supply connectivity to the best-match data center, and the provider automatically instantiates a Cloud VPN to the destination. This permits the cloud data center to fulfill the additional workload needs from its resources, relieving the enterprise infrastructure strain. This enables the network to provide secure, quality-assured connectivity to resources on an on-demand basis. Previously, cloud services were only available with security and QoS through static, pre-provisioned network connectivity, or via best effort through the public Internet. Cloud VPNs complete the on-demand model of emerging cloud-based services, permitting secure, quality-assured network connectivity to be delivered instantly only when needed and billed on a usage basis.

In this new cloud-services model, resources such as compute and storage that typically reside within a data center are now distributed across multiple geographical locations. Unfortunately, many of the applications used in the data center have not been adapted to a geographically distributed architecture. They still often require Layer 2 adjacency between nodes for functionality such as virtualization, workload portability, clustering, and storage connectivity. Layer 2 adjacency is achieved with various techniques under the Data Center Interconnect (DCI) realm. Numerous methods are candidates for DCI such as virtual private LAN service (VPLS), but only the Cisco innovation of Overlay Transport Virtualization (OTV) is custom-built to meet the data center challenges in a scalable, resilient, and transparent manner.

© 2012 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public Information.

Figure 5. Cloud VPNs and Data Center Interconnect



While Cloud VPNs are mechanisms to automatically interconnect data centers using deployed connectivity technologies in core and edge networks, DCI is a mechanism to extend intra-data-center connectivity over a wide area. DCI uses any deployed connectivity technologies in the core and edge, but maximum benefits arise from using automated mechanisms such as Cloud VPNs.

How It Helps

Automation vastly improves the ability of a service provider to reduce manual touch-points in the network. It also increases the speed of fulfilling resource requests across the infrastructure, improving operational agility. Cloud VPNs allow elastic capacity in the IP network to which the data center or cloud service is connected, reducing the need for pre-provisioned bandwidth capacity. At the same time security, QoS, and multicast benefits of an IP/MPLS network continue. Thus the overall infrastructure becomes more efficient. Furthermore, the service provider can offer this capability as a service to the third-party data center/OTT providers (for example, Infrastructure as a Service, or IaaS). With a pay-as-you-go model, providers can optimize their infrastructure use for customers while opening up new revenue opportunities.

Components

The Cisco Data Center Services System (DCSS) comprises various components that work in unison across the product portfolio. The exceptional scalability and feature richness of these components give providers a comprehensive solution across the network and data center cloud infrastructure to address the changing nature of traffic.

© 2012 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public Information.

IP NGN Routers

As part of the medianet, a media-aware Cisco IP NGN, the **Cisco CRS-3 Carrier Routing System** delivers continuous, always-on operations and scales easily from numerous single-chassis form factors to a massive multi-chassis system. The Cisco CRS-3 Family more than triples the performance of the earlier Cisco CRS-1 models. It comprises a 4-slot, 8-slot, and 16-slot single chassis as well as a multi-chassis system scaling to a massive 322 Tbps. The Cisco CRS-3 continues to feature the mid-plane design pioneered in the Cisco CRS-1 with a three-stage switch fabric based on a Benes architecture. It is now powered by the QuantumFlow Array, the chipset architecture specifically engineered to provide the industry's highest performance for Core and DCSS technologies. **QuantumFlow Array** is based on multiple components engineered to work in tandem while providing the benefits of single-flow performance at high bandwidths.

Figure 6. Cisco CRS-3 Family

Data Center 3.0 Switches and Unified Computing

Cisco's Data Center 3.0 portfolio consists of switching/routing and computing systems specifically designed for the data center and for cloud services. The **Cisco Unified Computing System** (UCS) is the next-generation compute platform that unites network, processing, storage, and virtualization resources in a single system. **Unified Fabric** plays a central role in the design of the platform. Unified Fabric consolidates different types of traffic onto a single, general-purpose, high-performance, highly available network, greatly simplifying the network infrastructure and reducing costs. The Cisco **Nexus**® Family was designed to support Unified Fabric and the Data Center Services System, meeting the stringent requirements of the next-generation data center.

Figure 7. Data Center 3.0 (Cisco Nexus switches, Unified Computing System)

© 2012 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public Information.

For More Information

For more information about the Cisco Data Center Services System, visit: http://www.cisco.com/go/crs.

For more information about the Cisco powered network cloud, visit: http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/white_paper_c11-532553.html.

For more information about the Cisco CRS Family, visit: http://www.cisco.com/go/crs.

For more information about Cisco solutions for service providers, visit: http://www.cisco.com/go/sp.

For more information about Cisco solutions for core networks, visit: http://www.cisco.com/en/US/netsol/ns573/networking_solutions_solution.html.

For more information about Cisco Nexus switches, visit: http://www.cisco.com/go/nexus.

For more information about data center solutions, visit: http://www.cisco.com/go/datacenter.

For more information about Overlay Transport Virtualization (OTV), visit: http://www.cisco.com/en/US/netsol/ns1153/index.html.

Additionally, you may also contact your local Cisco account representative.

cisco

Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV Amsterdam,
The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at **www.cisco.com/go/offices.**

Cisco and the Cisco logo are trademarks or registered trademarks of Cisco and/or its affiliates in the U.S. and other countries. To view a list of Cisco trademarks, go to this URL: www.cisco.com/go/trademarks. Third party trademarks mentioned are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (1110R)

Printed in USA C11-587663-01 01/12

© 2012 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public Information.