Designing Networks and Services for the Cloud

Delivering business-grade
cloud applications and services

Huseni Saboowala
Muhammad Abid
Sudhir Modali

ciscopress.com

Designing Networks and Services for the Cloud

Huseni Saboowala

Muhammad Abid

Sudhir Modali

Cisco Press

800 East 96th Street

Indianapolis, IN 46240

Designing Networks and Services for the Cloud

Delivering business-grade cloud applications and services

Huseni Saboowala

Muhammad Abid

Sudhir Modali

Published by:
Cisco Press
800 East 96th Street
Indianapolis, IN 46240 USA

Printed in the United States of America

First Printing April 2013

Library of Congress Control Number: 2013938238

ISBN-13: 978-1-58714-294-9

ISBN-10: 1-58714-294-5

Warning and Disclaimer

This book is designed to provide information about designing networks and network services for the cloud. Every effort has been made to make this book as complete and as accurate as possible, but no warranty or fitness is implied.

The information is provided on an “as is” basis. The author, Cisco Press, and Cisco Systems, Inc., shall have neither liability nor responsibility to any person or entity with respect to any loss or damages arising from the information contained in this book or from the use of the discs or programs that may accompany it.

The opinions expressed in this book belong to the author and are not necessarily those of Cisco Systems, Inc.

Trademark Acknowledgments

All terms mentioned in this book that are known to be trademarks or service marks have been appropriately capitalized. Cisco Press or Cisco Systems, Inc. cannot attest to the accuracy of this information. Use of a term in this book should not be regarded as affecting the validity of any trademark or service mark.

Corporate and Government Sales

The publisher offers excellent discounts on this book when ordered in quantity for bulk purchases or special sales, which may include electronic versions and/or custom covers and content particular to your business, training goals, marketing focus, and branding interests. For more information, please contact: **U.S. Corporate and Government Sales** 1-800-382-3419
corpsales@pearsontechgroup.com

For sales outside of the U.S. please contact: **International Sales**
international@pearsoned.com

Feedback Information

At Cisco Press, our goal is to create in-depth technical books of the highest quality and value. Each book is crafted with care and precision, undergoing rigorous development that involves the unique expertise of members from the professional technical community.

Readers' feedback is a natural continuation of this process. If you have any comments regarding how we could improve the quality of this book, or otherwise alter it to better suit your needs, you can contact us through e-mail at feedback@ciscopress.com. Please make sure to include the book title and ISBN in your message.

We greatly appreciate your assistance.

Publisher: Paul Boger

Executive Editor: Brett Bartow

Senior Project Editor: Tonya Simpson

Cover Designer: Mark Shirar

Business Operation Manager, Cisco Press: Jan Cornelssen

Senior Development Editor: Christopher Cleveland

Technical Editors: Sunil Cherukuri, Dave Lively, Ravi Varanasi

Proofreader: Debbie Williams

Associate Publisher: Dave Dusthimer

Managing Editor: Sandra Schroeder

Editorial Assistant: Vanessa Evans

Composition: Bumpy Design

Manager Global Certification: Erik Ullanderson

Copy Editor: Keith Cline

Indexer: Tim Wright

About the Authors

Huseni Saboowala is engaged at Cisco as a senior technical leader in the areas of Software Defined Networking, Cloud, and Unified Communications. He currently focuses on evangelizing the role of the network and network services in accelerating the adoption of cloud services by enterprises. His proposals have won Cisco-wide recognition and awards, and he continues to cultivate business-driven innovations that further enrich application-network interactions. Within Cisco SRG, he leads the architecture and deployment of a custom private cloud, driving his concept to reality across several groups. He has filed patents and spoken on Cloud and UC to large audiences on several occasions. Before joining Cisco, Huseni held several positions, including at Nortel, TTI (acquired by Sonus Networks), and dynamicsoft (acquired by Cisco). Over the past 18 years, his activities have ranged from solution architecture, design, validation, and deployment to leadership of global teams, innovation coaching, and developing technology strategies. He holds a Bachelor's degree in Electronics Engineering from the University of Bombay, and a Master's degree in Software Engineering from Kansas State University.

Muhammad Abid is an innovator who is currently working as a senior product manager in the Services and Routing Group at Cisco. He is engaged in developing the next generation of platforms and innovative technologies that will support data center interconnect and software defined network (SDN) frameworks and play a key role in enabling cloud-based services and applications. Prior to this, he was focused on architecting solutions and driving technology roadmaps across multiple business units for unified communications, collaboration, and threat defense. Before joining Cisco, Muhammad held several positions, including at T-Systems, Padcom, Telcordia, and Latham & Watkins. Over the past 18 years, he has been involved in building innovative products and designing and architecting networks for service providers and enterprises. He has also validated mobility solutions for enterprises and performed technical audits on service provider equipment and networks. He holds a Bachelor's degree in Electrical Engineering from the City University of New York and an Executive Master in Technology Management degree from the Stevens Institute of Technology in New Jersey.

Sudhir Modali is a thinker and innovator currently putting his creative mind to work as a product manager at Cisco, working on products that fuel data center and cloud architectures. He currently focuses on the evolving application requirements and the corresponding network architectures that enable some of the biggest cloud services in the world. His expertise comes from multiple positions he has held at Cisco over the past 13 years, including customer support (TAC); QA lead; technical marketing in areas such as service provider, enterprise, and data center networks; and technologies such as data, voice, and video. He has developed and is a major contributor to several certification courses for data center and cloud fields within Cisco. He holds a Bachelor's degree in Industrial Electronics from Shivaji University (Solapur, India).

About the Technical Reviewers

Dave Lively is currently a director in the Cloud Architectures/Sales team at Cisco. His team focuses on driving the architectures for early/emerging opportunities to leverage cloud computing in various markets/verticals in the service provider space. He has also worked extensively on the engineering side, focusing on system architecture, strategy, and validation for the cloud computing and data center markets. His teams have developed the end-to-end system architecture for the data center and *Next Generation Network* (NGN), including both network and compute infrastructure and management/orchestration. Before working in cloud computing, Dave led the systems efforts for Cisco's multiscreen IP video strategy, enabling service providers to deliver video to the television, the PC, and the mobile phone. In addition, he has served in multiple companies in various product management, marketing, and engineering capacities, working across various technologies, including optical networking, WAN optimization, content networking, VoIP, VoATM, and broadband. He also worked in software and hardware engineering and holds a Bachelor's degree in Computer Engineering from Virginia Tech.

Sunil Cherukuri is a senior technical lead at the Cisco Systems Development Unit (SDU), focusing on Cloud solutions since 2009. He has more than 14 years of experience with design, validation, and deployment of end-to-end networking solutions spanning VoIP, cable, MPLS, security, and cloud architectures. He holds a Master's degree in computer engineering from North Carolina State University. He has presented to both internal and external audiences at various conferences and has represented Cisco in a number of customer events.

Sunil currently works on designing and validating the Cisco Architectures and Orchestration systems for Cloud Computing solutions, for end-to-end functionality, scalability, resiliency, automation, and service delivery, and for delivering the CVDs for the VMDC cloud architectures and related cloud orchestration and cloud assurance systems. He also assists service providers and major enterprises in the design and deployment of such services. He previously worked on designing and validating scalability and performance of large-scale network-based security services, including IPsec and SSL VPN, firewall, IPS, and DDoS.

Dedications

Huseni Saboowala:

I dedicate this book to my wife, Insiya, our daughter, Alisha, and my parents, Ruby and Hakim. They are my inspiration in everything I do. Writing this book wouldn't have been possible without their patience, encouragement, and unyielding support. I love you all!

Muhammad Abid:

This book is dedicated first to my family, my wife, Nadia, and my two awesome children, Zayd and Khadeeja. Without their love, encouragement, support, and patience, I would not have been able to achieve my dream of writing this book. Second, to my mom, Safia, and dad, Gulzar, who instilled in me a strong work ethic, persistence, and a will to never give up. Lastly, to my brothers, sisters, and in-laws, who have always been there full of encouragement. I know they will be very delighted by what follows.

Sudhir Modali:

I dedicate this book to my family, to my wife, Subha, and our son, Ananth, my source of inspiration and drive to work on this book; my mom and dad, who have shaped my thinking and are a guide at all times; my sister and brother, who are a source of encouragement in all my endeavors. I know they are proud of my achievements and are close to me in this moment of elation.

Acknowledgments

Dave Lively, director SP Cloud Architecture Technical Sales at Cisco, for his outstanding contributions to this book. Along with deep technical and strategic inputs, his review comments included valuable pointers on enhancing the flow and readability of the book.

Kiran Rane, senior director Cisco SRG Engineering, for his unwavering support throughout this journey and his encouragement during some of the challenging times.

Ravi Chandra, VP/GM, SRG Operations, for sharing his deep insights on the cloud market and needs of Cisco SRG's enterprise and service provider customers.

Pravin Akkiraju, chief executive officer VCE, for spending time and inspiring us with his vision and strategy for the role of networks in solving some of the biggest challenges facing enterprise cloud adopters.

Kelly Ahuja, SVP/GM, SP Mobility Group, for sharing deep insights about how the service providers are looking to evolve their networks to monetize cloud-generated opportunities.

Suresh Thirunavukkarasu, director, product management, Rony Gotesdyner, senior product line manager, Mani Ramasamy, principal engineer, and Chandrodaya Prasad, manager product management at Cisco SRG, for their solid feedback and suggestions on Cisco Cloud Connectors.

Monique Morrow, distinguished consulting engineer at Cisco, for sharing her insights on the future of cloud and networks.

Mike Sullenberger, distinguished engineer and Nic Doyle, principal engineer at Cisco, for their valuable contributions to the NGN section of the book, with a focus on cloud security.

Mostafa Mansour, technical marketing engineer at Cisco SRG, for sharing his expertise on NGN infrastructure and data center interconnect technologies.

The technical reviewers, Ravi Varanasi, engineering director for Cloud Security, and Sunil Cherukuri, senior technical leader at Cisco, who provided excellent technical coverage and kept this book accurate and easy to navigate. Despite their busy travel schedule and ongoing engagements, both Ravi and Sunil stepped up and delivered high-quality input, which enabled the book to benefit from their expertise.

We also want to extend our sincerest gratitude to the following managers and colleagues for their support in this project: Mark Lohmeyer, Vinod Peris, Deependra Vaidya, Jason Rolleston, Marcelo Magno, Afaq Khan, Padmini Sridhar, Ashok Ganesan, Charles Yager, Jeff Raymond, Lakshmi Sharma, Ritch Dusome, Matthias Falkner, and Scott Yow.

Finally, the Cisco Press team: Brett Bartow, the executive editor, for seeing the value and vision provided in the original proposal and believing enough to provide us the opportunity to build this book. In addition, Christopher Cleveland, senior development editor, for his relentless push to develop our rough manuscript into a fine piece of technical literature and pushing the entire team to meet our deadlines. Lastly, Tonya Simpson and everyone else in the Cisco Press team who spent countless hours normalizing the manuscript and its technical drawings and content; their effort can be seen throughout the book pertaining to our ideas, words, and pictures, presented in ways that we ourselves could never have imagined.

Contents at a Glance

Contents

This page intentionally left blank

Introduction

The cloud and the services it has to offer have garnered significant interest worldwide. The cloud offers an elastic model that allows infrastructure capacity to be increased and decreased on demand. The cloud's usage-based model helps governments, educational institutions, and enterprises to increase business agility and reduce costs by seamlessly moving applications and consuming infrastructure resources from the cloud. The cloud's role as an enabler of newer economics for IT is now widely understood.

Despite all the benefits, enterprises have been cautious to adopt the cloud because of concerns around *availability*, *security*, and *application performance*. Lack of visibility and control combined with the need to maintain compliance with regulatory requirements are cited as other reasons that have thus far inhibited the adoption of the cloud.

Business-grade cloud services aim to address these concerns and enable these organizations to adopt the cloud with confidence. These advanced cloud services require that the cloud data centers, networks, applications, and services be tightly integrated. The network is the only entity that interacts with all the elements of a cloud service and is ideally positioned to address the barriers to cloud adoption.

Evolved networks and network services enable the provider to offer cloud services with security, performance and availability *service level agreements* (SLA). These advanced networks provide appropriate levels of visibility and insight that can help businesses with performance and compliance verification. In addition to boosting cloud adoption, such capabilities fuel premium cloud service offerings and enable competitive differentiation.

These cloud-aware networks have additional intelligence—service, location, and cost awareness—that facilitate the seamless extension of IT resources, delivered as an optimized cloud service that can scale rapidly and cost-effectively. The rich set of *application programming interfaces* (APIs) available for automated provisioning of these networks and network services facilitate simplified management and zero-touch operations, which help in driving down costs further.

Networks inherently carry tons of information, including user location, device capabilities, topology, and end-to-end performance characteristics. When exposed appropriately through well-defined APIs, such information can be consumed by cloud applications to fine-tune and customize their efficient delivery. The future holds the promise of increasingly rich application-network interactions.

Cisco, with an industry-leading portfolio of cloud-ready networking products and services, is in a unique position to provide end-to-end architectures for differentiated cloud services. Cisco's innovative platforms extend from the *customer premise equipment* (CPE) at the enterprise branch, to the service provider IP NGN, to the service-rich network fabric in the data centers. Large sets of documentation from various business units are available on these cloud products and solutions from Cisco. In addition, select cloud solutions in the form of CVDs (*Cisco Validated Designs*) are available, as well.

These product documentations and CVDs are implementation heavy, and usually do not address the design choices, application needs, end-to-end cloud service delivery, or business aspects of cloud services. For those seeking to understand the design and architecture of networks and network services pertaining to the delivery of business-grade cloud services, there is no single source of reference available today.

This book provides a concise and easy-to-understand view of how evolved networks and network services can be designed to enable a secure, resilient, and SLA-driven cloud experience. In addition, the book explains how intelligent networks can help providers simplify the complexity of managing cloud services and reduce costs through efficient scaling and improved capacity utilization. The end-to-end service delivery concepts are reinforced with illustrative examples. The goal is to boil down and simplify the design and architectural details and present them in one reference, augmenting the existing installation and configuration guides of the various cloud-related products and solutions already available from Cisco.

The book does not attempt to be prescriptive about how these network services can be put together into a particular cloud solution and dive into the detailed configurations/ CLIs needed to implement the cloud services, because these are tied to the specific requirements of that deployment. The book provides the architectural knowledge that will help you understand the role and capabilities of these advanced networks and network services, along with the design factors to consider for their insertion into a cloud service. For the next steps, CVDs are recommended for obtaining detailed design information on specific cloud solutions that have been qualified by Cisco, and consultative engagements with Cisco Advanced Services are recommended for customized cloud solutions.

Objectives of This Book

The book can help you understand the role of networks—encompassing data center networks, service provider IP NGNs, and the customer premise equipment—in the delivery of business-grade cloud services. The architecture of networks and network services is discussed in context with the underlying trends shaping the technical and business landscape of these cloud services and applications. A major focus is the evolution of today's networks and network services—new technologies and platforms—and how they can be designed to ensure the accelerated adoption of the cloud by addressing the primary inhibitors: availability, security, visibility, and application performance.

The book is organized into four parts: Part I discusses the basics of virtualization and the cloud and the role of networks in clouds. Part II focuses on virtualization-aware data center networks that enable flexible virtual network services for the cloud. Part III covers the evolution of IP NGNs and services for the cloud. Finally, Part IV explores the critical role of the CPE as a control point in accessing cloud services, and then delves into end-to-end cloud SLAs that enable guarantees in the delivery of premium cloud services.

Who Should Read This Book?

The book is intended primarily for a technical audience involved in designing, architecting, deploying, and delivering cloud services. Cloud and network visionaries, architects, and engineers at cloud service providers, network service providers, managed service providers, or even enterprises looking to build their own cloud, stand to benefit from the wide range of topics covered by the book.

The book would also prove valuable to cloud consumers, both businesses and individuals, who want to better understand the technical and business landscape surrounding premium cloud services. It can help them make informed choices and enable them to have an engaging discussion with their provider on how they can achieve their security and performance goals while reaping the benefits of the cloud.

How This Book Is Organized

This book is organized into 13 chapters distributed across 4 parts, and although it can be read cover to cover, it does allow for readers to move between chapters and parts, covering only the content that interests them. The four parts of the book are described as follows.

Part I introduces virtualization concepts across compute, network, and storage domains and how virtualization proved to be the cloud harbinger. Part I then covers basics of cloud (the characteristics, the deployment and service models, and the benefits and cloud service management) before diving into the critical role of the network in enabling business-grade cloud services.

Part II covers the all-important data center networks, underlining the importance of a virtualization-aware network fabric and the flexibility provided by virtual network services. It also discusses the concept of network containers and how security and optimization can be designed in this dynamic multitenant environment.

Part III examines the role of the service provider IP NGN in enabling the flexible and highly available extension of resource pools across geographically dispersed data centers. How can network intelligence be leveraged to optimize the placement of cloud services? This section then delves into designing secure access to the cloud and protecting the cloud edge from various attacks. Aspects of application performance are also examined to ensure that the cloud services and applications deliver an enhanced user experience that is expected from business-grade cloud services.

Part IV discusses the critical role of the CPE as a control point in accessing hybrid cloud services. It builds on the material covered earlier and breaks down the complexity of end-to-end SLA guarantees. This part then rounds off the book, with a peek into the future of cloud and networks.

An overview on each of the 13 chapters follows.

- **Chapter 1, "Virtualization":** Provides a brief history of virtualization, before discussing the core concepts for virtualizing the three pillars of the data center: server, network, and storage. Zooming into the server space, the chapter explores compute, memory, and I/O virtualization. Network virtualization concepts are examined with illustrative examples, and the chapter wraps up with a discussion on storage virtualization and the synergies from combining compute, network, and storage virtualization.
- **Chapter 2, "Arrival of the Cloud":** Describes how virtualization enables the transition to the cloud, followed by its definition and key attributes. It then delves into the underlying trends driving the adoption of cloud and also examines the key inhibitors. Finally, the chapter explores the game-changing benefits and impact of the cloud.
- **Chapter 3, "Cloud Taxonomy and Service Management":** Covers the classification of cloud services into the *software/platform/infrastructure* (as a service) SPI model and examines various cloud deployment models, including the evolution toward the intercloud. The chapter then explores a cloud ecosystem before concluding with an overview of cloud service management.
- **Chapter 4, "Networks and Services in the Cloud":** This key chapter explains how networks can help overcome the barriers that inhibit the CIOs from wholeheartedly adopting the cloud. How can these network services be monetized? And how are these networks and network services poised to play an increasingly critical role in the next stage of the cloud journey? The chapter ends with a discussion on the evolution of today's networks to meet the challenges of the cloud.
- **Chapter 5, "Role of the Network Infrastructure in a Virtualized Environment":** Discusses the factors influencing evolution of the network fabric due to virtualization and defines the critical components required of the network infrastructure in the virtual environment.
- **Chapter 6, "Securing and Optimizing Cloud Services":** Security is one of the most important services that is part of any data center architecture. An understanding of business and application workflow is key in designing a security framework. In a cloud-enabled data center, predefined instances can be used to provision security compliant (PCI-DSS, HIPAA, GLBA, SOX, and so on) frameworks. Virtualizing the services enables multitenant-capable security deployment models while retaining the characteristics of a virtual machine such as mobility, elasticity, and manageability.
- **Chapter 7, "Application Performance Optimization":** This chapter focuses on delivering a seamless and persistent cloud experience irrespective of the location and mode of connectivity.

- **Chapter 8, "IP NGN Infrastructure That Supports Cloud Services":** Delving deeper and showing how the IP NGNs are evolving to accommodate the transition to the cloud, this chapter describes various data center interconnect technologies, which enable the flexible, high-availability extension of resource pools across geographically separated data centers. The chapter also focuses on various route optimization techniques
- **Chapter 9, "Securing Cloud Transport and Edge Using NGN Technologies":** Focuses on protecting the cloud edge from various attacks and providing secure access to the cloud to place and consume cloud services and applications.
- **Chapter 10, "Optimizing and Accelerating Cloud Services":** Explains how the network infrastructure needs to become more intelligent; that is, it has to become service, location, and cost aware and enable optimal placement and accelerated delivery of cloud services and applications.
- **Chapter 11, "Connecting Enterprises to the Cloud":** Focuses on the need for enterprises to connect to multiple cloud providers, along with their own data centers. Various cloud connect examples illustrate how these organizations are able to leverage the CPE as a control point toward achieving secure, optimized, and cost-effective access to cloud services.
- **Chapter 12, "End-to-End Cloud SLAs":** This chapter deals with the complexity of cloud SLAs and elaborates on the models that you can use to simplify the delivery of these SLAs. The chapter then delves into end-to-end SLAs and how they can be enabled through a service overlay approach.
- **Chapter 13, "Peeking into the Future":** The final chapter explores two major phenomenon poised to change the future of cloud (the intercloud and the Internet of Things) and the critical role of the network in enabling their success. The chapter then delves into emerging network trends and innovations around application-network interactions and *software-defined networking* (SDN).

This page intentionally left blank

Chapter 4

Networks and Services in the Cloud

In this chapter, you learn about the following:

- Networks helping to overcome barriers to cloud adoption
- Increased relevance of the network and network services
- Monetization of network services
- Evolution of networks toward meeting the challenges of the cloud model
- Map of the subsequent sections of the book

The CIO's Dilemma

The cloud has created a paradigm shift in the way IT resources are provided and consumed. The previous chapters discussed how virtualization has proven to be the disrupter that has accelerated the journey to cloud. Cloud deployments have brought about game-changing benefits for both the providers and the consumers but continue to be challenged by certain inhibitors to adoption. Consider the case of an enterprise's *chief information officer* (CIO) contemplating a move to the cloud. The cost and agility benefits offered by cloud deployments make it an attractive option for the organization. It allows the IT group to focus its limited resources on the core business of the company, enabling it to fund and undertake new projects with business impact. Figure 4-1 illustrates how the majority of IT budgets are spent on maintenance, resulting in unfunded new projects, which ultimately result in missed business opportunities.

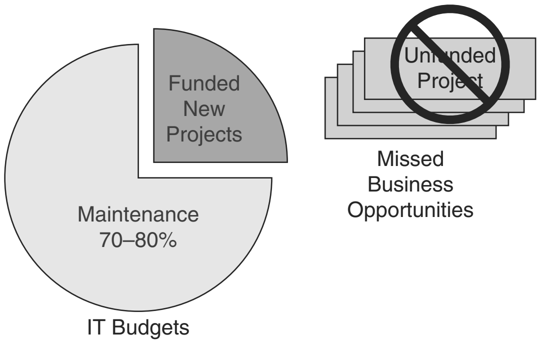

Figure 4-1 *CIO's Dilemma*

The elastic nature of the cloud allows IT to rapidly respond to changing business conditions, scaling up and down on demand. The cloud can help the IT department to cater to demand elasticity and avoid outages/unavailability of business-critical resources such as the company's e-commerce website during the crucial holiday shopping season, for example. The resulting loss of revenue and negative customer sentiment could be avoided by leveraging the nearly unlimited scale offered by the cloud. Clearly, CIOs have a lot to gain by moving workloads to the cloud and enabling IT to focus on providing competitive differentiation for the organization.

However, the CIO has several concerns that impede the migration to cloud. Security and privacy of the organization's data in the cloud is a primary concern for CIOs. The multitenant nature of cloud deployments come with intricate concerns about competitors running workloads on the same shared infrastructure and potentially gaining access to proprietary applications or sensitive data belonging to other tenants. Organizations have different regulatory requirements to comply with depending on their industry and the jurisdictions under which they perform business. CIOs need assurance of compliance to internal and external regulations as they move workloads into the cloud. Can the IT organization still have the ability to run audit reports on their cloud assets? In addition, CIOs need visibility of their workloads in the cloud. Can the cloud resources consumed by the organization be accurately measured?

Service level agreements (SLA) are another key area of concern for the CIO, who is responsible for ensuring a certain level of performance and availability for the organization's service consumers.[1] IT organizations in over two thirds of enterprises provide some form of internal SLA to their customers (the various business and functional units within the company). Can the CIO continue to offer equivalent SLAs after migrating to the cloud? What is the impact to the uptime metrics for the organization's mission-critical applications? And how will performance metrics such as latency, jitter, and loss affect the end user's quality of experience when delivered from the cloud?

How can cloud providers assuage such concerns from the CIO and enable the organization to migrate to the cloud with confidence? The network is uniquely positioned to help address these challenges and accelerate the adoption of cloud services toward fulfilling the mission-critical needs of the organization. Let's take a look at how networks and network services enable the cloud provider to address each of these inhibitors:

- **Security:** Today's multitenant cloud deployments leverage shared infrastructure, causing most organizations to have concerns about securing their data and isolating it from other tenants of the cloud. The ubiquitous nature of the network and its role in connecting physical and virtual cloud resources—inside data centers and beyond—positions it appropriately for providing comprehensive security, from the infrastructure all the way to the application. The network provides an ideal platform to consistently enforce security policies from physical to virtual stacks, from local data center to remote virtual data centers.
- **Visibility and compliance:** The network is inherently aware of user interactions, connected resources, and data traversing service provider networks or the Internet, and even inside and between cloud data centers. This awareness, combined with the powerful capabilities of network analytics, positions the network as an ideal platform for monitoring and providing visibility into the cloud service and infrastructure. Cloud providers could then make relevant pieces of information available to the tenants, allowing them visibility into their current share of cloud resources. Cloud consumers gain deep insights into their services, such as performance statistics, accurate resource use, and location information.

  In addition, this tenant-level visibility enables the generation of event logs and the production of audit reports. This is particularly useful toward verifying compliance with regulations such as *Health Insurance Portability and Accountability Act* (HIPPA), *Payment Card Industry* (PCI), and others, which still need to be adhered to as organizations move to the cloud.
- **User experience/SLAs:** Cloud consumers, especially enterprises, are looking for cloud providers to offer certain levels of availability and performance SLAs. As described previously, the network is naturally suited to monitoring cloud services and infrastructure. This allows the network to intelligently re-orchestrate resources and redirect workloads in the event of failure or performance degradation. Such actions based on policy-driven automation allow the network to improve the resiliency and as a result the availability of the cloud service.

  It also allows protecting the user experience with the cloud service, which is dependent on the latency, jitter, and packet loss that the distributed cloud service is subjected to. In this regard, cloud service providers who also own or have access to IP *Next Generation Network* (NGN) assets are uniquely positioned to offer end-to-end cloud SLAs to their customers, providing them significant differentiation. (Chapter 12, "End-to-End Cloud SLAs," explores these end-to-end SLAs in detail.)

Increasing Relevance of the Network

The network provides the capabilities and analytics that allow the cloud provider to allay the fears of the CIO. So far, this chapter explored the network's pivotal role in spurring the adoption of the cloud, enabling organizations to migrate more and more of their core workloads to the cloud today. And as we look ahead, the network is poised to play an even more crucial role in future clouds.

World of Many Clouds

A variety of clouds exist today: public, private, and hybrid clouds, along with community and specialty clouds to address the needs of different business verticals such as healthcare, media, finance, or government. As illustrated in Figure 4-2, we are moving toward a world of many interconnected clouds, serving the needs of users who want to experience cloud services anywhere, at any time, and on any device, and of businesses, which want IT to be delivered as a service.

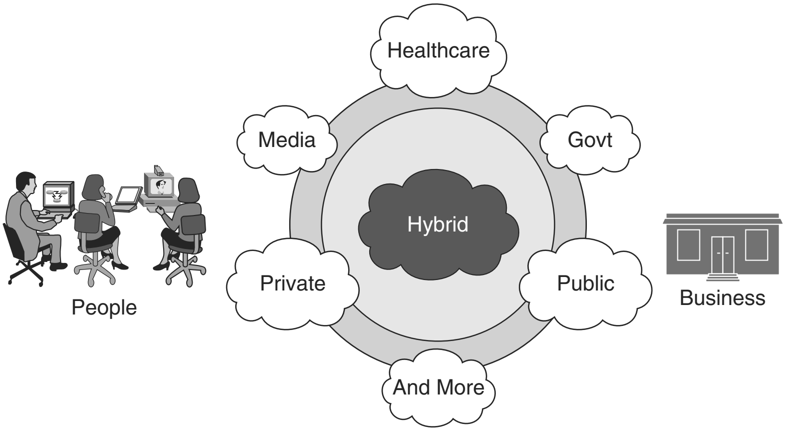

Figure 4-2 *World of Many Clouds (Source: Cisco)*

In this multicloud world, the network's role is significantly expanded because these clouds need to securely connect to each other. In addition, massive amounts of infrastructure resources, along with applications and content, need to be combined and delivered on demand, to provide a secure and consistent user experience regardless of the user location and number of cloud platforms involved. The network fabric enables bringing together these capabilities dynamically, virtualizing connections within the cloud, between clouds, and beyond the clouds to the consumers.

An Even Larger Cloud

Over the past few years, there has been an explosion in the number and types of consumer and business mobile devices, sensors, and actuators, many of which are now connected to the network. Although we tend to think so, clouds are not limited to the servers in data centers. In fact, the cloud extends out to all these network-connected electronic devices, smart meters, and other sensors, as illustrated in Figure 4-3. When you put it all together, it is easy to see that this is an even larger cloud on the horizon, with billions of network-connected components.

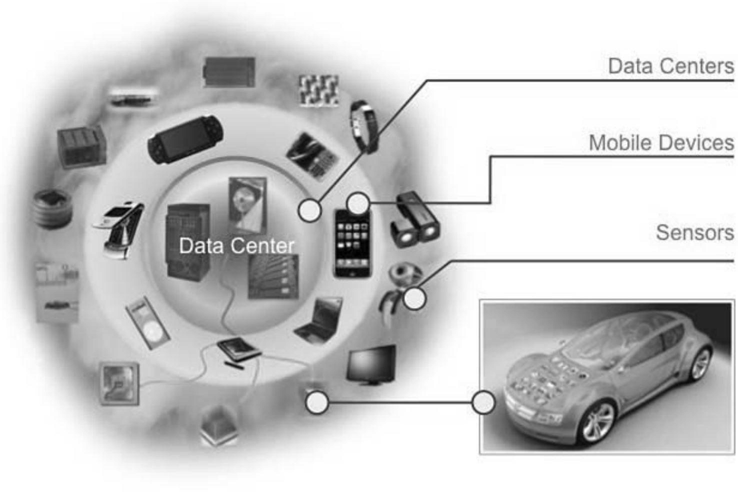

Figure 4-3 *Cloud of Mobile Consumer Devices and Sensor Devices (Source: J. Rabaey, "A Brand New Wireless Day")*

Consider the dozens of sensor devices running inside modern cars today. With 3G/4G mobile data connectivity enabling *machine-to-machine* (M2M) communications, sensor devices can monitor and share vehicle performance data with the car manufacturer, who can then use it to suggest appropriate maintenance or repairs. Or consumers might want their car to communicate with other cars around them, over an ad hoc local network, and learn about road and traffic conditions up ahead. Security is obviously critical here. After all, we would not like untrusted parties gaining access to these devices, with perhaps the ability to start interfering with brakes or other vehicle safety features. The possibilities are endless, and as you can see, dynamic, scalable, and secure networks have an increasingly vital role to play in the cloud in the years ahead. These futuristic clouds are further explored in Chapter 13, "Peeking into the Future."

Growth of Cloud Data Traffic

Consumer and business cloud services, including rich-media services, keep growing in popularity, leading to an explosion in data center traffic. According to Cisco's Global Cloud Index, cloud IP traffic is expected to grow at 66 percent *compound annual growth rate* (CAGR) from 2010 to 2015, which is twice the 33 percent CAGR expected for overall data center IP traffic during the same period. As illustrated in Figure 4-4, overall data center traffic volume is expected to reach 4.8 zettabytes in 2015. And cloud traffic is expected to be over a third of that pie (1.6 zettabytes). (A *zettabyte* is a billion terabytes; the number 1 followed by 21 zeros!)

Note Cisco's Global Cloud Index considers all provider and enterprise data centers, and includes the following traffic categories:

1. Traffic that remains inside the data center
2. Traffic between data centers
3. Traffic from data center to end users over the Internet or IP WAN

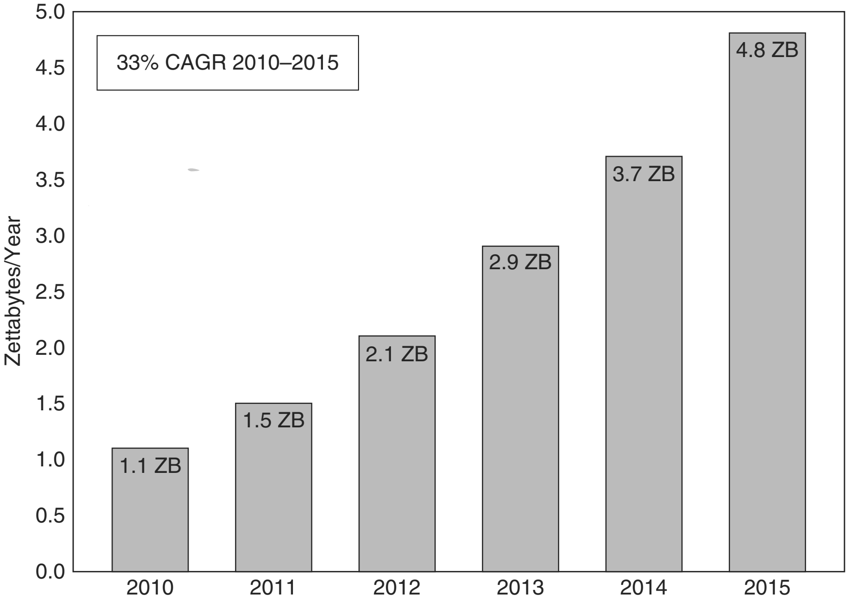

Figure 4-4 *Data Center Traffic Quadruples from 2010 to 2015. Cloud Traffic Is Expected to Be Just over One Third of the Data Center Traffic in 2015. (Source: Cisco Cloud Index)*

Let's try to put 1.6 zettabytes in perspective. This is the equivalent of 5 trillion hours of business web conferencing or 1.6 trillion hours of HD video streaming. Another interesting comparison is with the overall global Internet traffic, which in 2015 is expected to be just under 1 zettabyte, according to the Cisco *Visual Networking Index* (VNI).

In addition to the mind-boggling growth in traffic volumes, cloud applications, services, and infrastructure are responsible for transforming the pattern of data center traffic flows. Cloud-ready networks inside data centers, between data centers, and from data center to users will play an increasingly crucial role in terms of scaling efficiently to handle this growth in cloud data traffic and maintain profitability for the cloud providers without compromising the end-user experience.

Monetization

Earlier in this chapter, we discussed the role of the network in speeding up adoption of cloud services, providing solutions to the fundamental concerns that businesses have about wholeheartedly embracing the cloud. Cloud providers can leverage their network assets to enable their customers to confidently start moving more and more of their critical workloads to the cloud. On top of this, what if cloud providers could also directly monetize their network assets? What if networks and network services could be offered by the provider as a service; that is, *network-as-a-service* (NaaS)?

Along with compute and storage, networks and network services can be offered as a service, to be consumed, metered, and billed, based on usage. The economics of this model provide network vendors and cloud providers with strong incentives to innovate on compelling network services that add significant value for their customers.

The following are methods to offer networks and services for consumption.

Service Catalog

The discussion on cloud service management in Chapter 3, "Cloud Taxonomy and Service Management," explained how cloud services, defined in the service catalog, are offered to customers through self-service portals or via *application programming interface* (API) access. In addition to including various predefined cloud services, the service catalog enables the flexibility to add or modify optional features for those services. The same service catalog provides a means to define and offer networking for consumption (ranging from a basic VLAN service to a complex network service that provides security across multiple data centers).

To include network services in the service catalog, they need to be abstracted and presented in a simplified manner to the customer who may not be a networking expert. The intricacies and complex operations involved in enabling the network service must be hidden from the customer. Simplification is key, and ordering NaaS should be as easy as a few clicks on the cloud portal or a small number of intuitive API calls.

Here are a few examples of data center networking services, both basic and premium, that a provider could offer in their service catalog:

- Traffic isolation between tenants
- Access control between *virtual machines* (VM) of three-tier apps
- Load balancing across tiers of the three-tier apps
- *Virtual private network* (VPN) termination to isolated segments
- *Quality of service* (QoS) inside the data center fabric

The service catalog does not need to be restricted to network services inside the data center. After all, the end user consumes the cloud service from across the WAN (Provider IP NGN) or Internet. Cases where the cloud provider owns or controls network assets in the IP NGN present an opportunity to abstract network services available in the IP NGN bring it up to the service catalog. Examples of such services include the following:

- *Virtual Private LAN Service/Multiprotocol Label Switching* (VPLS/MPLS) VPN for private access to cloud
- WebVPNs for public access to cloud
- App performance enhancement with WAN acceleration, web caching
- Security through firewall, *deep packet inspection* (DPI), and distributed threat detection services in the NGN
- Optimal cloud services placement based on network proximity and performance

Not only do these NGN services open up additional revenue streams for the cloud provider, they also enable the provider to offer end-to-end security and performance capabilities. Certain network services such as firewall, QoS, and WAN application acceleration could potentially be distributed across the NGN and data center networks.

Network Services à la Carte

One option for monetization is to offer network services à la carte. Here network connectivity and services can be individually ordered by the consumer. The exact needs are conveyed as part of the API call or via a portal. For instance, if the developer needs to simply connect the database VM to an isolated virtual network segment that is not routable from the Internet but reachable from the web servers, those network attributes would be specified as part of the API invocation, as shown in the following pseudo API example:

1. Create a DB network, specifying the following address range:

```
create_network(name="db-net", cidr="10.0.1.0/24")
```

2. Attach the DB VM to the network created in Step 1:

```
attach_vm(vm=vm_uuid, network="db-net")
```

3. Create a route to allow web servers to access the DB servers:

```
create_route("web-net","db-net", "local")
```

A well-designed API enables the users to easily describe what they want out of the network: for example, a network that supports a certain amount of bandwidth, a network with QoS, or perhaps a network with monitoring services. The APIs represent a contract to provide a certain service. While the underlying networking devices may differ, the functionality delivered by the API call is expected to be the same. In essence, a network hypervisor is needed. Analogous to the compute hypervisor, the network hypervisor would provide the ability to abstract the underlying networking hardware into services that can then be consumed by the user.

Not too long ago, though, developers did not have any visibility or control over the network, with *infrastructure-as-a-service* (IaaS) offerings focusing primarily on compute and storage, as illustrated in Figure 4-5. The network was there only to provide connectivity. Each VM would have a very flat view of the world, and there would not be any topology at all. Obviously, network services would not be available for consumption in such architectures.

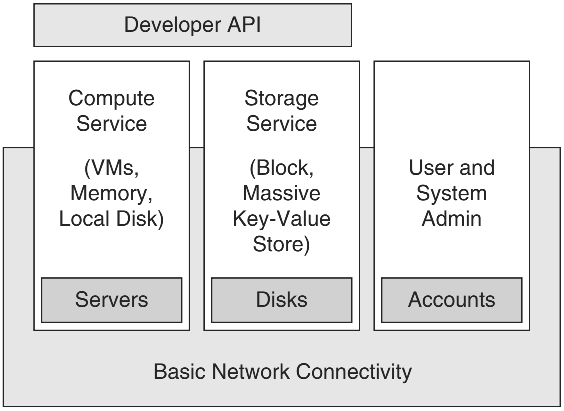

Figure 4-5 *IaaS Offerings Lacking API Access to the Network (Source: Cisco, Lew Tucker)*

OpenStack Quantum

OpenStack is open source software that enables any organization to build their public or private cloud stack. It aims to deliver a massively scalable cloud operating system, along the lines of the software that powers colossal clouds such as Amazon EC2 today. OpenStack has been gaining momentum, with contributions from a growing global community of developers, vendors, and service providers helping it grow in functionality and maturity.

Initially, OpenStack started off as a platform underpinned by three major services: the Nova compute service, the Swift storage service, and the Glance virtual disk image service. The OpenStack development community has been actively engaged in developing additional services, some of which are shown in Figure 4-6. One such service, named Quantum, aims to provide network connectivity as a service. Along with requesting VMs and storage, developers can now request network connectivity, as well, using the Quantum API.

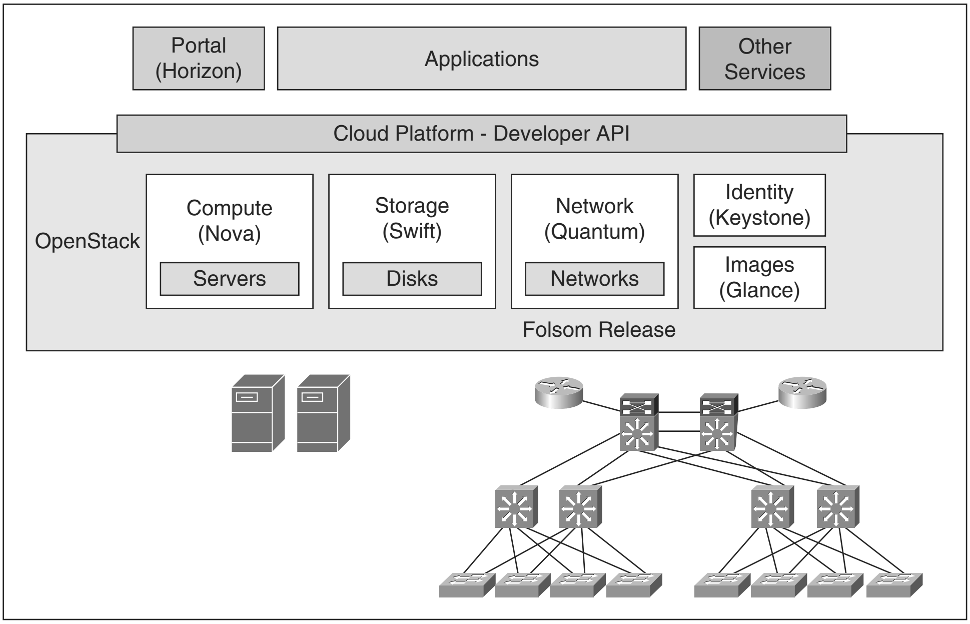

Figure 4-6 *OpenStack Services*

Figure 4-7 shows how Quantum has a pluggable framework with plug-ins offered by multiple networking vendors, including Cisco and Nicira/VMware. This is key to adoption; customers do not have to fear being locked into a particular vendor. The plug-ins map the API abstractions to the actual networking device underneath. In addition to offering basic Layer 2 virtual network segments, the Quantum API has an extensible architecture allowing advanced network services to be offered through the API extensions. And this extensible architecture is important, as the Quantum API is still evolving, and new network features such as firewalls, VPNs, and load balancers can be offered through the extensions first, before they get baked into the core Quantum API over time. Cloud providers have an opportunity to differentiate themselves by offering advanced networking features via the extensions.

Services such as OpenStack Quantum represent a fundamental shift in cloud networking. Networks are no longer hidden beneath the hypervisor, and network services are no longer limited to providing basic connectivity for the VMs. Applications can interact with network services via the API, bypassing the hypervisors.

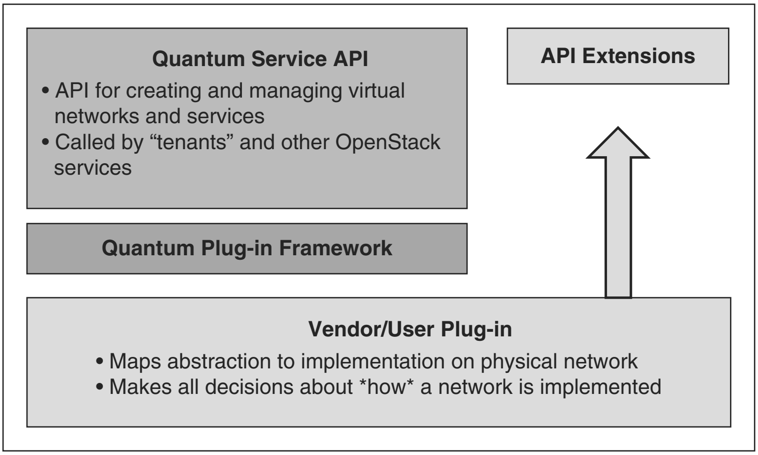
Quantum Service API
• API for creating and managing virtual networks and services
• Called by "tenants" and other OpenStack services
API Extensions
Quantum Plug-in Framework
Vendor/User Plug-in
• Maps abstraction to implementation on physical network
• Makes all decisions about *how* a network is implemented

Figure 4-7 *Quantum API Architecture*

Network Containers

Network containers provide a representation of the data center network infrastructure that is dedicated to a tenant for the provisioned time. As compared to ordering individual network services, containers enable a higher level of abstraction, encompassing the set of network connectivity and network services allocated to a tenant service. Figure 4-8 shows an example of a tenant network container for a three-tier web application. Separate network containers have been created for the Web, App, and DB tiers, nested inside the tenant network container and separated by firewall services. External connectivity is provided for the container to be reachable from the corporate VPN for management purposes, while the Web container is reachable from the Internet through a load balancer.

If the entire topology in Figure 4-8 can be saved as an abstract model, it could be offered through the services catalog for consumption. That would significantly ease the deployment of the tenant's application, freeing the tenant from the lengthy process of individually ordering these network services and managing the interdependencies. A sophisticated network abstraction system such as the Cisco *Network Services Manager* (NSM) enables such use of network container models to define the behavior of the network services as a holistic virtual network infrastructure.

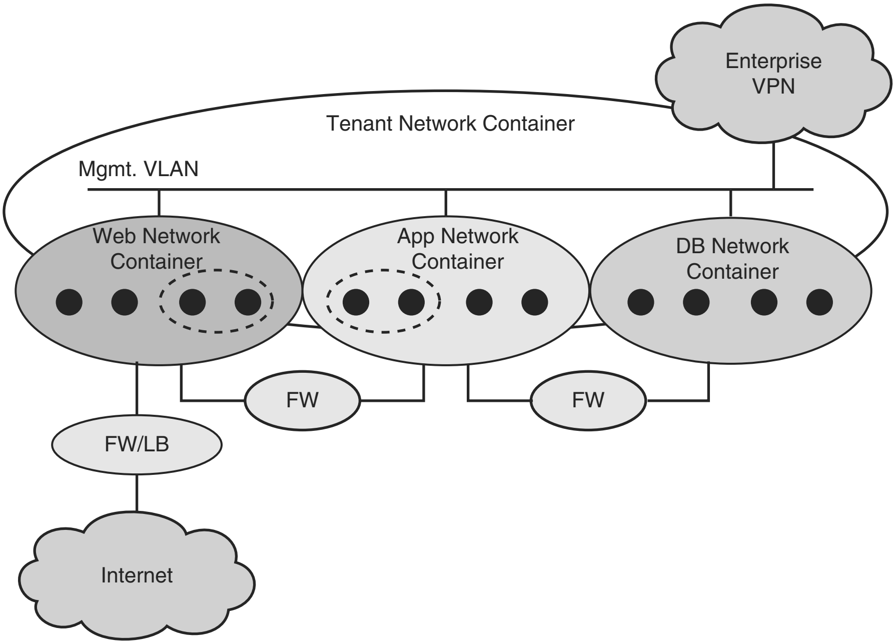

Figure 4-8 *Network Containers with External Connectivity for a Tenant's Three-Tier App*

Cisco Network Services Manager

Cisco NSM provides model-based policy-driven abstraction and orchestration of the cloud network environment, leading to increased flexibility in terms of what can be done in the network, what services/capabilities can be exposed from the network, and what tenant container environments can be provisioned on the network. A REST-based API allows orchestration and other systems to interact with NSM and access the abstractions.

Comprehensive network container models, such as the three-tier web application in Figure 4-8, can be instantiated on diverse cloud network infrastructures, with NSM abstracting away the platform-specific behaviors of the underlying networks. Figure 4-9 shows an NSM system managing three cloud infrastructure stacks or pods. One of the pods could be based on Nexus networking platforms, the other may be leveraging existing Catalyst-based networking, and the third may be based solely on virtual network services. The NSM service controller associated with a pod understands the specific devices and platforms in the pod, and when it receives a directive to instantiate a particular abstract topology model, it interacts with the networking devices in that pod to stitch that topology together.

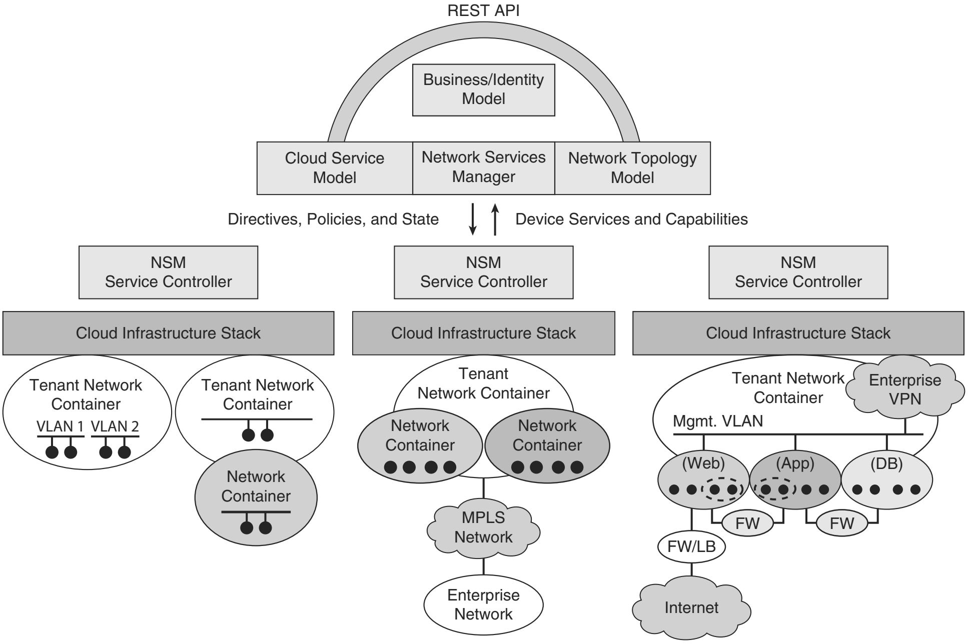

Figure 4-9 *Cisco NSM and Instantiated Network Containers for Multiple Tenants*

In addition to the abstraction, this model enables the mobility of network containers. Instantiated network containers, including the application and data residing in them, can be moved from one cloud pod to another, as needed, without any changes.

Various types or tiers of container model can be included in the service catalog, addressing different requirements such as security, performance, or application delivery. The customer can then pick one or more of these containers, and then select the VMs, which will be placed inside the container. The cloud administrator designs these container models to address the varied network service needs of their customers and enable the provider to offer differentiated pricing on these containers based on the density, complexity, and perceived value of the included network services.

Even though the service catalog allows the tenant to easily pick and choose from a variety of network services and predesigned topologies, tenants might need to customize and fine-tune their logical network in the cloud to meet their goals. Providers that can offer the tenant admin increased flexibility on day 2 operations, such as runtime configuration and modification of network services, will be able to further differentiate their offerings from the competition.

Through our discussion about OpenFlow Quantum service and the Cisco NSM system, you saw how network services can be offered in a simplified manner to spur consumption (either as individual network connectivity services or as network containers). These offerings enable cloud providers to gain access to additional revenue streams, realizing improved returns on their infrastructure investments.

Evolution of Network Services for the Cloud

To fulfill their role in the adoption and monetization of cloud services, networks need to adapt to the cloud environment. The rise of cloud models is changing what is happening on the network:

- Change in traffic patterns caused by increasing server-to-server traffic and the location-independent endpoints at both sides of a service.
- The new infrastructure is highly virtualized and programmable; servers and applications have become increasingly mobile.
- Change in access patterns for applications and services, predominantly through mobile devices.
- New applications are more data intensive, collaborative, and media rich.

These changes are driving the rapid evolution of networks. But not everything about the network has to change. Its foremost purpose still remains the same. The network still has to provide transport for the movement of data between the various components of an application, its storage, and the end user. It still has to provide security for access to applications and data. And it is still responsible for delivering a certain level of application performance to the end user. What changes is how these jobs are to be performed (with automated provisioning and management, with support for virtualization and multitenancy, and with location independence).

Automation

Automation is one of the most important areas of evolution for networks. And APIs are a fundamental means of enabling automation. One of the biggest impacts of the cloud on networks is the sheer scale and the frequency of change. And APIs allow us to address both of them. When network and network services can be provisioned and managed with well-designed APIs, such as those exposed by the network hypervisors discussed earlier in this chapter, the cloud network can scale efficiently from one rack to a whole data center to collections of data centers. At the same time, frequent changes brought about to the network, as tenants allocate and de-allocate cloud services, can be handled without any human touch. The economics of the cloud make such zero-touch operations mandatory.

Virtualization Awareness and Multitenancy

A couple aspects of virtualization are relevant to the evolution of networks. First is the network's awareness of server virtualization, which was introduced in Chapter 1, "Virtualization." Such virtualization-aware networks can identify and treat each VM as a separate networking endpoint. In addition, such networks can attach security and other policy profiles to VMs in a sticky fashion. As VMs migrate from one physical host to another, or one data center to another, these profiles move along with them.

The other aspect relates to networks themselves: that is, network virtualization. Also discussed in Chapter 1, virtualization of networks and network services enables the end-to-end isolation required to allow multiple tenants to securely coexist on the same shared underlying infrastructure. Advanced network abstractions such as containers can build on top of this virtualization and provide the flexibility of carving up the infrastructure into network containers. Such containers, described earlier in this chapter, would be completely isolated from the network containers of other tenants, enabling multitenancy.

Location Independence

Networks today support user and device mobility in various ways. With the advent of cloud, network capabilities around mobility need to evolve further. The virtualization and resource pooling aspects of clouds means that servers and applications are no longer tied to physical infrastructure either. In fact, applications can be thought of as floating over a pool of infrastructure resources, seamlessly extended within and between clouds.

With the mobility of applications and data in addition to the users themselves, networks can no longer depend solely on their location to make policy decisions. These modern networks, shown in Figure 4-10, gather and rely on context information in this borderless world, ensuring that users can access only those applications and that data to which they are entitled. In addition, these networks strive to achieve a consistent level of user experience, irrespective of the location of the user, application, and data in the cloud.

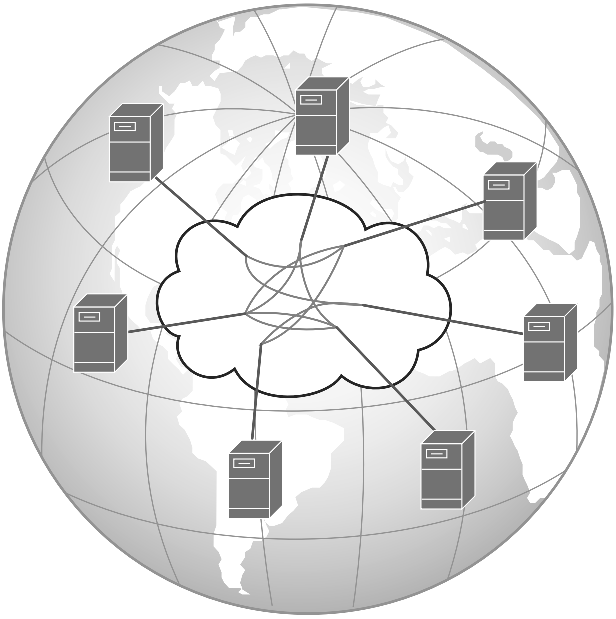

Figure 4-10 *Application/Data Mobility*

Quick Guide to the Rest of This Book

The network fabric is the glue that securely binds together heterogeneous resources inside clouds and between clouds and delivers them beyond the cloud to the end users. Based on requirements, characteristics, and administrative domains, cloud networks can be divided into three distinct entities:

- Data center networks
- WAN/IP NGNs
- Enterprise/consumer networks

How are these networks evolving to support cloud models? What is the role played by these networks in enabling business-grade cloud services? And how do we instantiate these concepts in deployment use cases? What end-to-end considerations apply for the secure delivery of cloud services with an SLA? These are some of the questions we explore in the rest of this book. The three parts of this book that follow are organized along the lines of the network sections listed here. The first one delves into data center networks. The next one explores the network between the data centers and from the data centers to cloud users. And the final one covers cloud consumer/enterprise networks, and then brings it all together with an end-to-end view of cloud service delivery. Here's a reader's map to these three sections.

Part II: Inside the Data Center Networks

We begin in Chapter 5, "Role of the Network Infrastructure in a Virtualized Environment," by examining the changes in networking infrastructure required to adapt to the virtualized environment of today's cloud data centers. What trends are driving the data center network design? How are virtual network services hosted on this network fabric? Next, in Chapter 6, "Securing and Optimizing Cloud Services," we examine the design of secure, multitenant data center networks. How can virtual security services be enabled inside a tenant's network container, and then across tenants? How can predefined instances be used to provision security compliant frameworks for PCI-DSS, HIPAA, and other regulations? Then, Chapter 7, "Application Performance Optimization," delves into optimization of cloud services and enhancing the end user experience. How do virtual application delivery solutions work?

Part III: Inside the SP Next Generation Network (WAN)

Cloud service providers that own or control WAN/IP NGN assets are able to mobilize their cloud resources between data centers and are also able to securely deliver and optimize the cloud service all the way to the customer edge. Chapter 8, "NGN Infrastructure That Supports Cloud Services," discusses *Data Center Interconnect* (DCI), the drivers, and the technologies. We also explore exciting changes that allow the cloud network to automatically adjust and optimize to account for such mobility. Chapter 9, "Securing Cloud Transport and Edge Using NGN Technologies," explores advanced security technologies in the NGN that protect the cloud edge and enable secure access to cloud services and applications. Then, we wrap up this section with acceleration technologies for cloud services over the WAN, in Chapter 10, "Optimizing and Accelerating Cloud Services." In addition, we explore how network intelligence, exposed by innovations such as the Network Positioning System, facilitates the optimal placement and selection of cloud services.

Part IV: Putting It All Together—Cloud Services Delivered

Enterprise networks are adapting to this new world order and playing a critical role as a control point in the consumption of cloud services. Chapter 11, "Connecting Enterprises to the Cloud," covers the architecture of cloud connectors and explains how advanced branch networks enable survivability, optimization, security, and performance of cloud services. We then discuss the all-important topic of cloud SLAs and how distinct networks can be stitched together to enable end-to-end cloud service delivery in Chapter 12, "End-to-End Cloud SLAs." Finally, in Chapter 13, "Peeking into the Future," we look at future trends as related to the cloud and what they mean for networks and network services.

Summary

This chapter began with a discussion about the CIO's dilemma in moving to the cloud and how the network can catalyze the confident adoption of cloud services by enterprises. However, the role of the network does not stop here, and in fact it is poised to become even more critical as we enter the world of many clouds and the bigger cloud brought on by the rapid growth of the Internet of Things. In addition, we explored how providers can monetize their investment in the network and offer network services in the service catalog for consumption through an API or a web portal. These services can be ordered individually or via abstracted network container models. Finally, we discussed how today's networks are evolving to meet the challenges of the cloud model.

This chapter explored the role of the network in enabling the success of business-grade cloud services, which is the central theme of this book. Subsequent chapters in this book build on the concepts discussed here and extend them to the different areas of the network involved with the cloud (DC networks, the WAN/NGN, and the enterprise network) and tie them together end to end, from the production point all the way to the consumption point.

Review Questions

You can find answers to these questions in Appendix A, "Answers to Review Questions."

1. What are the enterprise's areas of concerns about migrating to the cloud that the network helps to address?

   a. Compliance
   b. Security
   c. SLA
   d. All of the above

2. How can cloud providers better monetize their network investments?

   a. Offer advanced network services through a service catalog
   b. Protect network assets by not exposing their services to tenants
   c. Offer basic network connectivity for VMs
   d. Embed network services inside orchestrator

3. Which of the following is an open source cloud platform offering networking as a service?

   a. OpenFlow
   b. Amazon EC2
   c. OpenStack
   d. OpenOffice

4. Which of the following are key areas of evolution for networks in the cloud?

   a. Automation/API
   b. Flexible Multitenancy
   c. Location independence
   d. All of the above

References

1. Cloud Networking Report, Ashton, Metzler and Associates: http://www.webtorials.com/content/2010/12/2010-cloud.html

- OpenStack – Open Source Cloud Operating System: http://www.openstack.org
- Cisco White Paper - Cloud: Powered by the Network: www.cisco.com/en/US/solutions/…/white_paper_c11-609220.pdf
- Cisco CloudVerse: Enabling the World of Many Clouds: http://www.cisco.com/en/US/solutions/collateral/ns341/ns991/solution_overview_c22-693654.html
- Cisco Global Cloud Index: http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns1175/Cloud_Index_White_Paper.html
- Cisco White Paper - Networking and Cloud, An Era of Change: http://www.cisco.com/en/US/solutions/collateral/ns340/ns517/ns224/ns836/ns976/white_paper_c11-677946.html

Index