# Risk Thinking for Cloud-Based Application Services

Eric Bauer

CRC Press
Taylor & Francis Group
AN AUERBACH BOOK

# Risk Thinking for Cloud-Based Application Services

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

# Risk Thinking for Cloud-Based Application Services

Eric Bauer


CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
AN AUERBACH BOOK

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742



Printed on acid-free paper
Version Date: 20161206

International Standard Book Number-13: 978-1-138-03524-9 (Hardback)



**Library of Congress Cataloging-in-Publication Data**

Names: Bauer, Eric, author.
Title: Risk thinking for cloud-based application services / author, Eric Bauer.
Description: Boca Raton : Taylor & Francis, a CRC title, part of the Taylor & Francis imprint, a member of the Taylor & Francis Group, the academic division of T&F Informa, plc, [2017] | Includes bibliographical references and index.
Identifiers: LCCN 2016045236| ISBN 9781138035249 (hb : alk. paper) | ISBN 9781315268835 (e)
Subjects: LCSH: Cloud computing--Risk assessment. | Business enterprises--Computer networks.
Classification: LCC QA76.585 .B3943 2017 | DDC 004.67/82--dc23
LC record available at https://lccn.loc.gov/2016045236

**Visit the Taylor & Francis Web site at**
**http://www.taylorandfrancis.com**

**and the CRC Press Web site at**
**http://www.crcpress.com**

# Contents

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

# Author

**Eric Bauer** is a Bell Labs Fellow in Nokia's Applications and Analytics business group where he focuses on quality, reliability, availability, and efficiency of cloud-based services. Mr. Bauer has authored three books on cloud computing: *Reliability and Availability of Cloud Computing* (2012), *Service Quality of Cloud-Based Applications* (2013), and *Lean Computing for the Cloud* (2016). Mr. Bauer also wrote *ETSI NFV Quality Accountability Framework* (ETSI, 2016-01), *ETSI NFV Service Quality Metrics* (ETSI, 2014-12), and *Quality Measurements of Automated Lifecycle Management Actions* (QuEST Forum, 2015-08). Before focusing on the cloud, he worked on reliability of software, systems, and network-based solutions and wrote three general reliability engineering books: *Practical System Reliability* (2009), *Design for Reliability: Information and Computer-Based Systems* (2010), and *Beyond Redundancy: How Geographic Redundancy Can Improve Service Availability and Reliability of Computer-Based Systems* (2011). Earlier in his career, Mr. Bauer spent two decades designing and developing embedded firmware, networked operating systems, IP PBXs, Internet platforms, and optical transmission systems. He has been awarded more than twenty US patents and has published several papers in the *Bell Labs Technical Journal*. Mr. Bauer earned a BS in electrical engineering from Cornell University and an MS in electrical engineering from Purdue University. Mr. Bauer lives in Freehold, New Jersey.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

# Abbreviations and Acronyms

| | |
|---|---|
| **ALARP** | as low as reasonably practicable |
| **ASP** | application service provider |
| **BAU** | business as usual |
| **CAS** | Casualty Actuarial Society |
| **CFO** | chief financial officer |
| **COSO** | Committee of Sponsoring Organizations of the Treadway Commission |
| **CRO** | chief risk officer |
| **CSC** | cloud service customer |
| **CSP** | cloud service provider |
| **DOA** | dead on arrival |
| **DSPR** | Dam Safety Priority Ratings |
| **EBITDA** | earnings before interest, dividends, taxes, depreciation, and amortization |
| **ERM** | enterprise risk management |
| **ETA** | event tree analysis |
| **ETSI** | European Telecommunications Standards Institute |
| **FG** | forwarding graph |
| **FMEA** | failure mode effect analysis |
| **FMECA** | failure mode effect and criticality analysis |
| **FMO** | future mode of operation |
| **FTA** | fault tree analysis |
| **HACCP** | hazard analysis and critical control points |
| **HAZOP** | hazard and operability studies |
| **HRA** | human reliability assessment |
| **IEC** | International Electrotechnical Commission |
| **IEEE** | Institute of Electrical and Electronic Engineers |
| **ISO** | International Organization for Standardization |
| **IT** | information technology |
| **ITIL** | Information Technology Infrastructure Library |

| | |
|---|---|
| **ITSCM** | information technology service continuity management |
| **ITSM** | information technology service management |
| **KRI** | key risk indicator |
| **KPI** | key performance indicator |
| **KQI** | key quality indicator |
| **LOPA** | layers-of-protection analysis |
| **MANO** | management and orchestration |
| **MCDA** | multicriteria decision analysis |
| **MOP** | method of procedure |
| **MOS** | mean opinion score |
| **NDI** | nondevelopmental item |
| **NE** | network element |
| **NFV** | network function virtualization |
| **NOAEL** | no-observable-adverse-effect level |
| **NOEL** | no-observable-effect level |
| **OLA** | operations-level agreement |
| **PDCA** | plan–do–check–act |
| **PHA** | preliminary hazard analysis |
| **PMO** | present mode of operation |
| **PNF** | physical network function |
| **QMS** | quality management system |
| **RACI** | responsible–accountable–consulted–informed |
| **RCA** | root cause analysis |
| **RCM** | reliability-centered maintenance |
| **RFP** | request for proposal |
| **ROE** | return on equity |
| **RPO** | recovery point objective |
| **RTO** | recovery time objective |
| **RTP** | real-time transport protocol |
| **SLA** | service-level agreement |
| **SLO** | service-level objective |
| **SLS** | service-level specification |
| **SMS** | service management system |
| **SO** | service outage |
| **SPOF** | single point of failure |
| **TCP** | transmission control protocol |
| **VL** | virtual link |
| **VNF** | virtualized network function (e.g., a software-only application that runs on cloud/NFV) |

# Introduction

Enterprises take risks, like developing new product/service offerings and expanding into new markets and sales channels, in pursuit of reward. Many enterprises are moving their applications and information technology (IT) services to the cloud in order to deliver new services and value faster and to improve their operational efficiency without compromising user service quality. Better risk management results in fewer operational surprises and failures, greater stakeholder confidence, and reduced regulatory concerns; in essence, proactive risk management maximizes the likelihood that an enterprise's objectives will be achieved, thereby enabling organizational success. This work methodically considers the risks and opportunities that an enterprise taking their applications or services onto the cloud must consider to obtain the cost reductions and service velocity improvements they desire without suffering the consequences of unacceptable user service quality. The better the risk management that an enterprise has in place, the more risk the organization can take in pursuit of returns.

## Target Audience of This Book

This book is intended for readers from diverse backgrounds:

- **Operations and maintenance professionals** who are responsible for service management of cloud-based applications
- **Service integration professionals** who integrate applications, infrastructure, management, orchestration, and functional components into compelling services for end users
- **Product and solution engineers** who develop cloud-based applications and services
- **Strategy professionals** and consultants who plan an organization's evolution from the present mode of traditional operation to the future mode of cloud-based operation
- **Business professionals** who assure that cloud-based service offerings meet the organization's business objectives
- **Quality professionals** who are responsible for the quality management systems supporting cloud-based service offerings

Given this diverse target audience, the book assumes only basic knowledge of cloud computing and expects no prior knowledge of either risk management or quality management. Development, integration, operations, and maintenance professionals will benefit from the entire work. Readers with limited time can start with Chapter 32, "Connecting the Dots," to see the arc of the cloud risk management story, and then follow cross-references back into earlier sections of the book that are most relevant to them. Strategy and business professionals will find Chapter 31, "Cloud and Creative Destruction," particularly interesting.

## A Story Told in Five Sections

This cloud risk management story is told in five sections:

- Section I, "Framing the Cloud Service Customer's Problem"—Chapter 1, "Cloud Computing Fundamentals," lays out the standard definition of cloud computing, highlighting the roles of both the cloud service customer, who operates cloud-based applications and user-facing services, and the cloud service provider, who offers infrastructure, management, orchestration, and service components as-a-service to cloud service customers. This work focuses on the cloud service customer organization's risks. Chapter 2, "Desired Cloud Service Customer Benefits," derives the three canonical business goals for a cloud service customer's investment in the cloud:
  1. Deliver new service and value faster
  2. Improve operational efficiency
  3. Deliver acceptable service quality to users

  Chapter 3 introduces risk and risk management, and Chapter 4 discusses cloud service qualities. Note that cloud security risks are not considered in this work.
- Section II, "Analyzing the Cloud Service Customer's Problem," considers the changes that a cloud service customer must make to sustainably improve their operating efficiency and accelerate their pace of service innovation to deliver new services and value faster. The IT service life cycle model is used to frame the Section II analysis. Chapter 5, "Application Service Life Cycle" considers exactly what can change in the application service life cycle to reduce operating expenses and accelerate the pace of service innovation. Chapter 6, "Lean Application Capacity Management," considers how the key cloud characteristic of rapid elasticity and scalability (Section 1.4.5) can reduce the cloud service customer's opex and accelerate their pace of service innovation. Chapter 7,

"Testing Cloud-Based Application Services," considers how application and service testing changes in the cloud, especially automated validation testing. Chapter 8, "Service Design, Transition, and Operations Processes," considers how the key characteristics of cloud computing enable streamlining of service design, transition, and operations processes to improve operational efficiencies. Chapter 9, "Continual Service Improvement," interlocks the plan–do–check–act cycles of IT service management, quality management, risk management, and lean computing. Chapter 10 considers improving operational efficiency of cloud-based applications. Chapter 11, "Service Strategy," considers how a cloud service customer's service strategy can evolve to accelerate and maximize the benefits of improvements in the service design, transition, operation, and continual service improvement processes.

- Section III, "Cloud Service Quality Risk Inventory," methodically considers risks that must be managed, controlled, and treated by cloud service customers to assure that acceptable service reliability, service latency, service quality, and service outage downtime is delivered to users. These risks are organized into 14 vectors; each risk vector is considered in a separate chapter (Figure 0.1).
- Section IV, "Cloud Service Quality Risk Assessment and Management," applies ISO 31000 risk management to the risk causes detailed in Section III, "Cloud Service Quality Risk Inventory," that confront cloud service customers.
- Section V, "Discussion"—Chapter 31, "Cloud and Creative Destruction," considers how cloud technology will fundamentally disrupt the application service business by dramatically lowering the cost and increasing the flexibility of the computing resources that application services rely upon. Chapter 32, "Connecting the Dots," reviews a holistic risk management framework and recommendations for cloud service customers to reduce the uncertainty that the organization achieves the cost reduction and acceleration in the pace of service innovation that they desire without delivering unacceptable service quality to users.

Extensive cross-references are used, and each section is relatively self-contained, so readers can dive directly into whatever section interests them most.

## Standard, Authoritative Concepts and Terminology

This book uses industry standard terminology and concepts to bypass minor inconsistencies across the industry, such as the definition of cloud computing or the differences between test, verify, and validate. Specifically, this work builds on

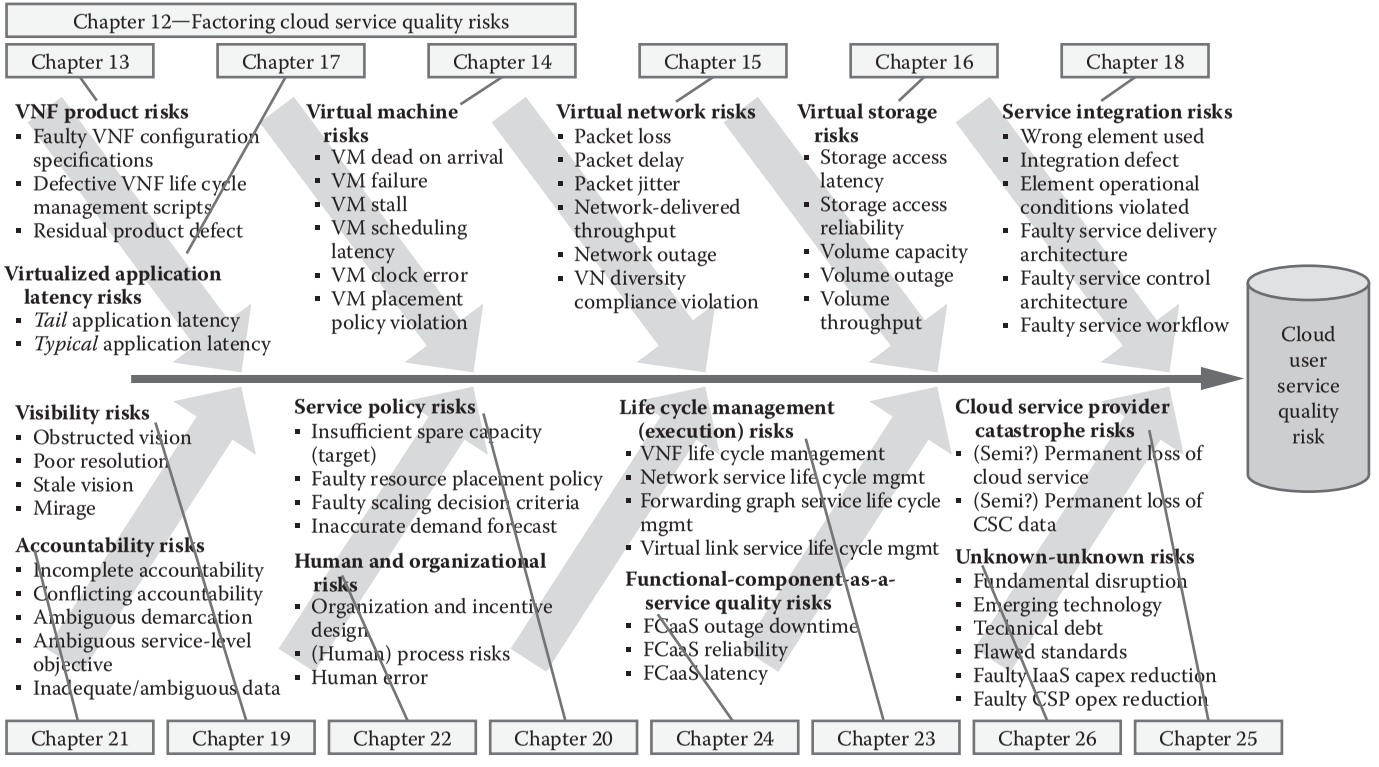


**Figure 0.1 Organization of Section III, "Cloud Service Quality Risk Inventory." CSC, cloud service customer; CSP, cloud service provider; FCaaS, functional-component-as-a-service; IaaS, infrastructure-as-a-service; VM, virtual machine; VN, virtual network; VNF, virtualized network function.**

the most authoritative references available,* especially the following international standards from the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC):

- ISO/IEC 17788, *Cloud Computing Vocabulary and Concepts*
- ISO/IEC 17789, *Cloud Computing Reference Architecture*
- ISO 31000 family of risk management standards
- ISO/IEC 20000 family of IT service management standards
- ISO 9000 family of quality management standards
- ISO/IEC 25000 family of software and system quality standards
- ISO/IEC/IEEE 15288, *System Life Cycle Processes*
- ISO/IEC/IEEE 29119 family of software and system testing standards
- ISO/IEC/IEEE 24765, *Systems and Software Engineering Vocabulary*

As ISO/IEC 17789, *Cloud Computing Reference Architecture*, is not detailed enough to permit rigorous analysis, this work leverages the authoritative European Telecommunications Standardization Institute's (ETSI's) Network Functions Virtualization (NFV, http://www.etsi.org/technologies-clusters/technologies/nfv) suite of standards as a target cloud computing architecture. The book also leverages the IT Infrastructure Library (ITIL®), ISACA (isaca.org) Risk IT, TM Forum (tmforum .org) service-level agreement (SLA) management principles, and QuEST Forum (tl9000.org) quality measurements.

## Reliability, Availability, Capacity Management, and Risk of Cloud-Based Application Services

Industry experts were initially concerned that cloud infrastructure would not be a suitable platform for hosting highly available "five 9s" applications. *Reliability and Availability of Cloud Computing* (Bauer and Adams, 2012) argued that traditional high availability and georedundancy mechanisms should satisfactorily mitigate likely cloud infrastructure failure modes. *Service Quality of Cloud-Based Applications*

* Officially sanctioned international standards development organizations like the International Organization for Standardization (ISO, http://www.iso.org) and International Electrotechnical Commission (IEC, http://www.iec.ch) are deemed the most authoritative references available. References from industry bodies like the Institute for Electrical and Electronic Engineers (IEEE, http://www.ieee.org), Telemanagement Forum (TM Forum, http://www.tmforum .org), and Quality Excellence for Suppliers of Telecommunications Forum (QuEST Forum, http://www.tl9000.org) are authoritative, but less authoritative than officially sanctioned international standards development organizations. References from a national government like the US National Institute for Standards and Technology, the US Department of Defense, or the US Federal Aviation Administration are also authoritative. Peer-reviewed scholarly works are less authoritative because of less rigorous review and approval prior to publication.

(Bauer and Adams, 2013) methodically considered how subcritical and critical failures of cloud infrastructure, management, and orchestration were likely to impact the user service reliability and latency of cloud-based applications. Key insights of that work were captured in ETSI's *NFV Service Quality Metrics* (ETSI, 2014-12), QuEST Forum's *Quality Measurement of Automated Lifecycle Management Actions* (QuEST Forum, 2015-08), and *ETSI NFV Quality Accountability Framework* (ETSI, 2016-01), which your author developed.

*Lean Computing for the Cloud* (Bauer, 2016) methodically considers how application capacity management can fully leverage the key cloud characteristic of rapid elasticity and scalability. However, as cloud service customers operate their application services leaner to reduce excessive online application capacity, they increase the risk that insufficient online capacity will be available to serve user demand with acceptable service quality. This book, *Risk Thinking for Cloud-Based Application Services*, methodically considers how cloud service customers manage the risk of achieving their objective of delivering new services and value faster and with improved operational efficiency against the downside consequences of failing to serve all user demand with acceptable service quality.

## Acknowledgments

The author gratefully acknowledges Mark Clougherty and Randee Adams for their frequent reviews and insightful feedback. Don Fendrick, Brian McCann, Gary McElvany, Barry Hill, Chris Miller, and Dan Johnson provided invaluable support for the work. Renee Miller, Bob Domino, Jean-Marie Calmel, Narayan Raman, and Enrique Hernandez-Valencia provided business and practical insights into the topic. The book benefited greatly from keen feedback provided by Tim Coote, Steve Woodward, Dave Milham, and Paul Franklin.

# FRAMING THE CLOUD SERVICE CUSTOMER'S PROBLEM

# I

The focus of this work is on reducing the uncertainty (i.e., risk) of a cloud service customer achieving their goals of reduced operating expense, accelerated pace of service innovation, and acceptable user service quality. Figure I.1 visualizes how this part explains this focus in the following chapters.

- **Cloud Computing Fundamentals** (Chapter 1)—reviews cloud computing basics, roles in cloud computing, input–output model of cloud computing, key characteristics of cloud computing, and cloud service management fundamentals
- **Desired Cloud Service Customer Benefits** (Chapter 2)—reviews cloud infrastructure service provider business models, cloud service customer business model, factoring benefits of cloud/virtualized network function (NFV), information technology (IT) service management objectives, and the focus of this work
- **Risk and Risk Management** (Chapter 3)—considers risk, simplified risk impact model, risk treatment options, risk appetite, risk management methodologies, and cloud service customer risk management
- **Cloud Service Qualities** (Chapter 4)—considers fundamental quality concepts; user service quality; ISO/IEC 25010 product quality; ISO/IEC 25012 data quality; data, information, knowledge and wisdom; and quality model for cloud-based applications

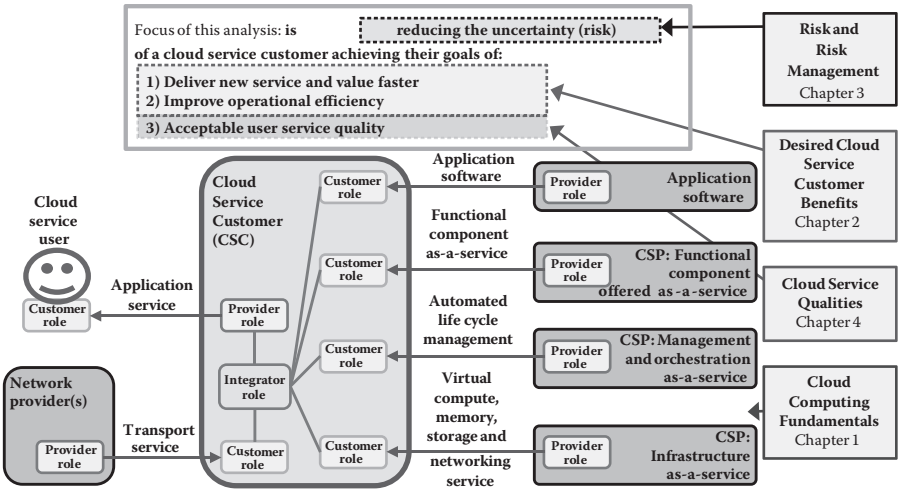


**Figure I.1 How Section I frames the cloud service customer's problem.**

# *Chapter 1*

# Cloud Computing Fundamentals

This section reviews key terminology and concepts that are used in this analysis via the following sections:

- Cloud computing basics (Section 1.1)
- Roles in cloud computing (Section 1.2)
- Input–output model of cloud computing (Section 1.3)
- Key characteristics of cloud computing (Section 1.4)
- Cloud service management fundamentals (Section 1.5)

## 1.1 Cloud Computing Basics

The standard* definition of cloud computing is a "paradigm for enabling network access to a scalable and elastic pool of shareable physical or virtual resources with self-service provisioning and administration on-demand" (ISO/IEC, 2014-10-15). As the standard ISO/IEC 17789 (ISO/IEC, 2014-10-15) cloud computing reference architecture is not detailed enough to support the rigorous analysis of Section II, "Analyzing the Cloud Service Customer's Problem," and Section III, "Cloud Service Quality Risk Inventory," this work considers the network function

* The most authoritative reference on cloud computing is ISO/IEC 17788:2014, *Cloud Computing Overview and Vocabulary* (ISO/IEC, 2014-10-15). The ISO/IEC reference is inspired by the seminal US Government National Institute of Standards and Technology's (NIST's) *NIST Definition of Cloud Computing* (Mell & Grance, September 2011).

virtualization (NFV) cloud architecture. The NFV suite of cloud standards* was developed by the world's leading telecommunications service providers and suppliers via the European Telecommunications Standards Institute (ETSI) as the standard framework for deploying risk communications applications onto cloud computing platforms.

Figure 1.1 gives a high-level architecture framework for cloud-based services to enable users (called *cloud service users*) to access *applications* executing on shared, *virtualized infrastructure* via *access and wide area networking services*. *Management and orchestration systems*, driven by configuration and policy data, automate application and resource life cycle management functions to enable greater service agility with lower operating expenses. Consider each major architectural component separately:

- **Virtualized infrastructure** (or network function virtualization infrastructure, NFVI)—A cloud infrastructure service provider organization makes virtualized compute, storage, and networking resources available to organizations [called *cloud service customers (CSCs)*] to host their application software instances. This architectural component is primarily responsible for fulfilling two key cloud computing characteristics: multitenancy (Section 1.4.3) and resource pooling (Section 1.4.6). Virtualized infrastructure is considered further in Section 1.1.1, "Shared Cloud Infrastructure."
- **Applications**—Application services are implemented as service chains of software components, application instances, and functional components (e.g., database) offered as-a-service. Software applications that execute on cloud infrastructure are called virtualized network functions (VNFs), and cloud-based application services are often composed of multiple VNFs along with functional components offered as-a-service (e.g., database-as-a-service, load-balancing-as-a-service), which together offer valuable services to cloud service users. VNFs are contrasted with physical network functions (PNFs), for which application software is bundled with dedicated physical compute, memory, and storage.
- **Management and orchestration**, along with operations support systems (OSSs), business support systems (BSSs), and management systems together with descriptors and other information elements, enable many aspects of service life cycle management to be automated to shorten fulfillment times, improve quality, and reduce operating expenses. Together, these components are primarily responsible for fulfilling two key cloud computing characteristics: on-demand

* ETSI network function virtualization is reviewed at http://www.etsi.org/technologies-clusters/technologies/nfv. Documents most relevant to this work:
  · Introductory NFV white paper (ETSI, 2012-10-22)
  · Updated NFV white paper (ETSI, 2012-10-22)
  · NFV-MAN 001, *Management and Orchestration* (ETSI, 2014-12)
  · NFV-INF 010, *Service Quality Metrics* (ETSI, 2014-12)
  · NFV-REL 005, *Report on Quality Accountability Framework* (ETSI, 2016-01)

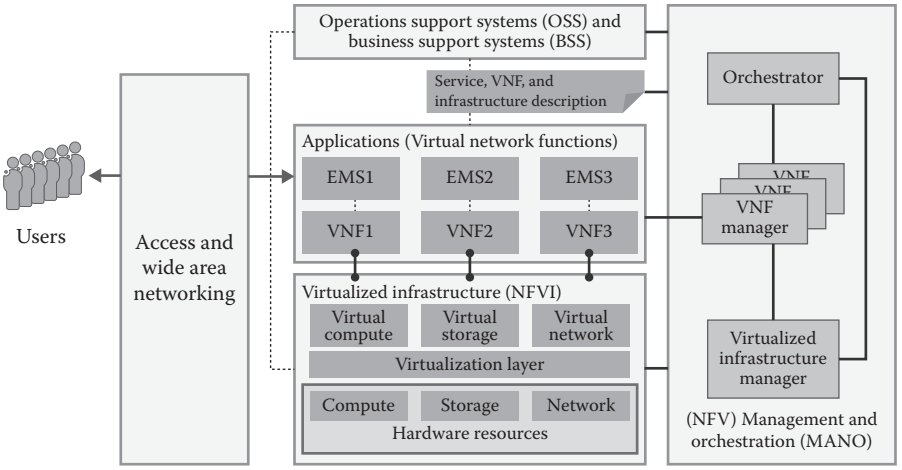


**Figure 1.1 Canonical cloud architectural framework. EMS, element management system. (Based on ETSI, *NFV-MAN 001 Management and Orchestration*. Sophia Antipolis, France: European Telecommunications Standardization Institute: http://www.etsi.org/deliver/etsi_gs/NFV-MAN/001_099/001/01.01.01_60/gs_NFV-MAN001v010101p.pdf, 2014-12.)**

self service (Section 1.4.4) and rapid elasticity and scalability (Section 1.4.5). Automated life cycle management is discussed further in Section 1.1.2.

- **Access and wide area networking**—Wireless and wireline networking carries internet protocol (IP) packets from the user's smartphone, tablet, laptop, or other device across access and wide area networks to the cloud service provider's (CSP's) data center where the virtual infrastructure hosting the application software instance serving the user is located. This architectural component fulfills the key cloud computing characteristic of broad network access (Section 1.4.1).

Two aspects of the canonical cloud architectural framework of Figure 1.1 are particularly important to this analysis:

- Shared cloud infrastructure (Section 1.1.1)
- Automated life cycle management (Section 1.1.2)

### *1.1.1 Shared Cloud Infrastructure*

Resource utilization and operational efficiency are improved via resource pooling (Section 1.4.6) and multitenancy (Section 1.4.3), both of which are implemented via virtualization of shared cloud infrastructure equipment. Virtual compute, memory, storage, and networking can be offered by infrastructure service providers to application service providers via technologies such as hypervisors, Linux containers,

and other virtualization mechanisms. For convenience, this book will refer simply to virtual machines (VMs) as the primary unit of infrastructure capacity, so this should be understood to cover Linux containers and other implementation options as well.

Virtualized infrastructure operated by CSPs has three logical layers:

- **Hardware resources**—physical compute and storage servers, Ethernet switches, along with cabling, power, cooling, and supporting equipment
- **Virtualization layer**—hypervisors and other software that enables multitenancy so multiple application software instances can efficiently and effectively share hardware resources
- **Virtual compute, storage, and network service layer** that provides virtual compute, storage, and networking service as resources to application instances operated by CSCs.

### *1.1.2 Automated Life Cycle Management*

CSP management and orchestration systems, driven by configuration data provided by a CSC, can automate execution of service life cycle management actions. For instance, automated life cycle management actions can include the following:

- Check VNF instantiation feasibility
- Instantiate VNF
- Update or upgrade VNF software
- Modify VNF software
- Terminate VNF instance
- Scale VNF software instance up/out or down/in
- Heal a VNF instance

Management or orchestration systems monitoring application, system, or infrastructure performance; fault and alarms status; demand patterns; and other factors can even automatically apply business logic to trigger appropriate automated life cycle management actions, further reducing the need for manual actions.

## 1.2 Roles in Cloud Computing

Figure 1.2 offers one mapping of standard, primary roles onto the canonical cloud architectural framework of Figure 1.1.

- **Cloud service user** is defined as "natural person, or entity acting on their behalf, associated with a CSC that uses cloud services" (ISO/IEC, 2014-10-15). In other words, a cloud service user is an end user, or an application operating

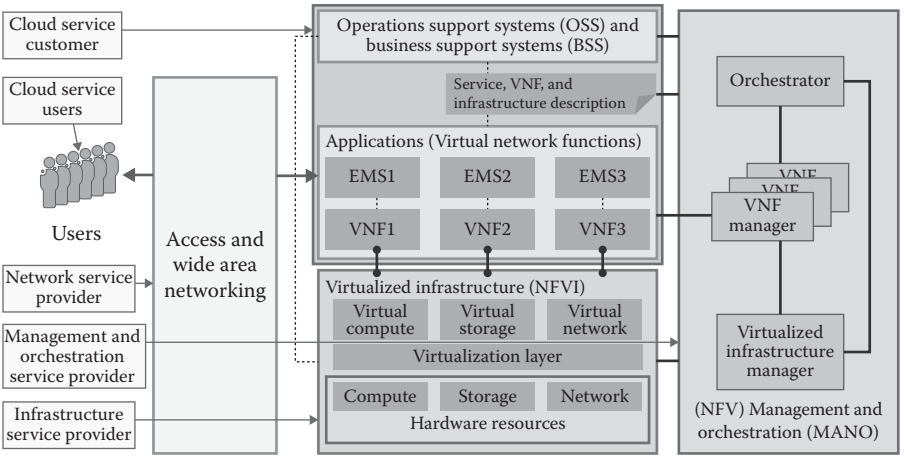


**Figure 1.2 Generic cloud computing roles on the canonical cloud architectural framework.**

on his/her behalf, who enjoys cloud-based application services such as social networking applications or watching streaming movies.

- **Cloud service customers (CSCs)** are organizations that operate cloud services for cloud service users, such as an organization that offers streaming entertainment or real-time communications services to end users via cloud-based applications. For example, a company that offers streaming movie services to end users by deploying their application software onto some other organization's infrastructure-as-a-service offering is a CSC. Application services offered to cloud service users are composed from a chain of application software instances (VNFs) hosted by virtual compute, memory, and storage instances offered by an infrastructure CSP and zero or more functional components offered as-a-service by some CSP. A management and orchestration CSP provides automated life cycle management for the service chain and included service components.
- **Cloud service provider (CSP)** is broadly defined as a "party which makes cloud services available" (ISO/IEC, 2014-10-15). In this work, we focus on three fundamental types of CSPs:
  - Infrastructure CSPs (aka infrastructure-as-a-service providers)—Organizations that offer virtual compute, memory, and storage services to CSCs.
  - Management and orchestration CSPs—Organizations that offer automated life cycle management services to CSCs.
  - Functional component CSPs—*Functional component* is defined as "a functional building block needed to engage in an activity, backed by an implementation" (ISO/IEC, 2014-10-15); some CSPs will offer functional components such as databases or load balancers to consumers via

platform-as-a-service offerings. While there is some disagreement across the industry regarding exactly what functionality offered as-a-service should be classified as platform-as-a-service versus software-as-a-service (or even infrastructure-as-a-service), the exact taxonomy is unimportant for this analysis.

Note that a single CSP organization can offer infrastructure, management and orchestration, and functional components as-a-service. Likewise, a particular CSC organization can elect to buy all of their cloud services from a single organization, like their organization's private CSP or a public CSP, or they can stitch together a service from cloud service offerings from several different providers. CSCs may elect to operate the VNF managers that directly support their VNFs, or they may allow the CSP's generic VNF managers to perform that service. Some customers will select a single CSP organization for simplicity, business considerations, or other reasons; other CSCs will select different CSPs based on pricing, performance, functionality, or other considerations.

- **Network provider**—at the highest level, there are two classes of networking beyond the cloud data center:
  - **User access networking**—for access and wide area networking to carry traffic between cloud service users' service access devices and the point of presence for the CSC's infrastructure-as-a-service provider
  - **Intraservice networking**—formally defined as the CSP network provider role that "may provide network connectivity between systems within the cloud service provider's data centre, or provide network connectivity between the cloud service provider's systems and systems outside the provider's data centre, for example, cloud service customer systems or systems belonging to other cloud service providers" (ISO/IEC, 2014-10-15)

This work does not explicitly consider the differences between (wide area) intraservice networking provided by a CSP network provider and (local area) intraservice networking provided by the infrastructure-as-a-service provider; the CSP network provider role will not be explicitly considered in this work. Instead, the general network provider role will be used to capture network providers who both carry traffic between cloud data centers in the service delivery chain as well as haul traffic to and from cloud service users.

Two other roles relevant to our analysis are as follows:

- **VNF software supplier**—organizations that supply applications and component software to CSCs
- **Service integrator**—organizations that integrate application and component software with functional components offered as-a-service and management information to create offerings that deliver value to cloud service users

Real services offered by CSCs will rely on the offerings of several different software suppliers, integrators, and service providers, as well as CSC staff. Section III, "Cloud Service Quality Risk Inventory," in general, and Chapter 21, Accountability Risks, in particular, considers risks associated with complex cloud service delivery chains.

Figure 1.3 illustrates these cloud computing roles in a useful accountability framework (ETSI, 2016-01). Note that the accountability framework highlights the following:

1. Which party is the provider
2. Which party is the customer
3. What product or service is offered by the provider to the customer

For example, infrastructure-as-a-service CSPs are providers of virtual compute, memory, storage, and networking service to CSCs. The framework of Figure 1.3 will be used throughout this book.

Two fundamental points regarding accountability from the framework of Figure 1.3 are highlighted in Figure 1.4:

1. Cloud service users hold CSCs primarily accountable for delivering valuable services with acceptable quality.
2. CSCs have accountability for arranging, managing, and integrating the set of software suppliers, CSPs, service integrators, and others who directly and indirectly support the CSC's service delivery chain.

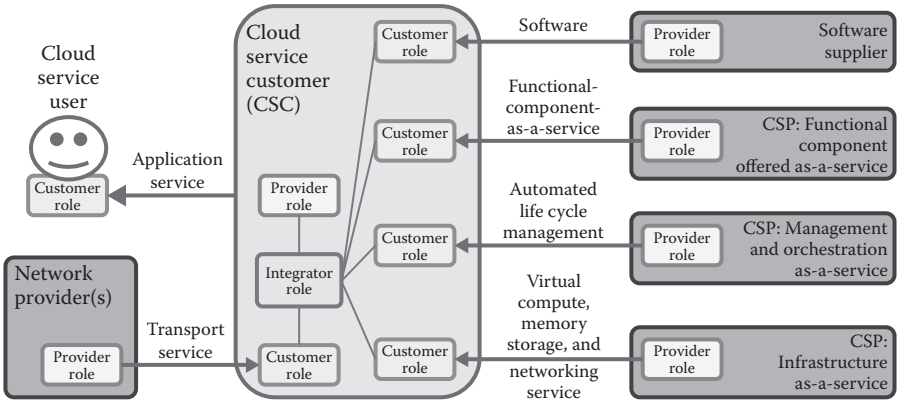


**Figure 1.3 Canonical cloud accountability framework.**

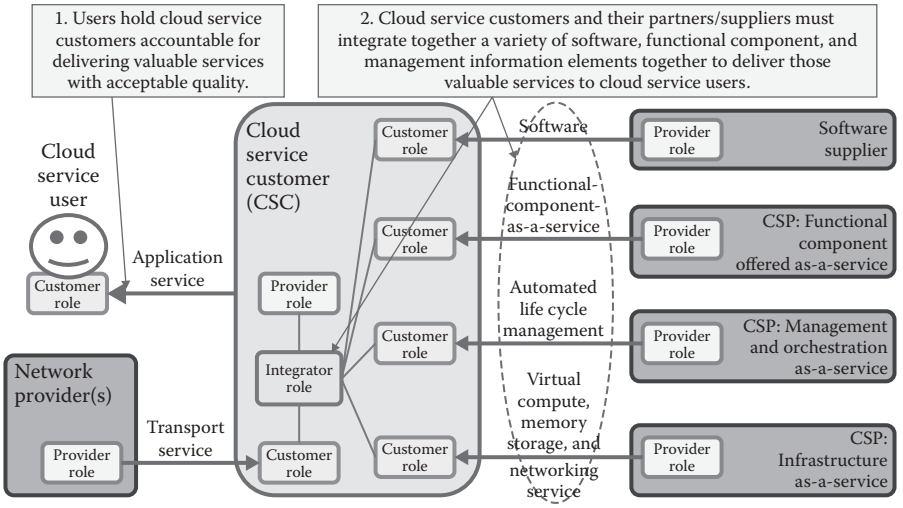


**Figure 1.4 Primary accountabilities for cloud service.**

## 1.3 Input–Output Model of Cloud Computing

Cloud service delivery can usefully be viewed as the input–output model shown in Figure 1.5. Fundamentally, cloud service users consume application services from various CSC organizations, and those CSC organizations rely primarily on both CSP organizations and software suppliers. The CSPs rely primarily on both electricity providers and infrastructure hardware and software suppliers. Cash flows

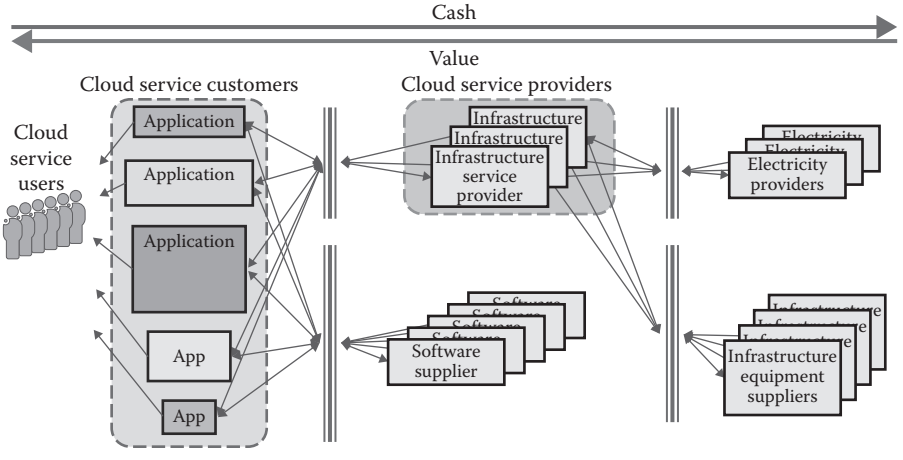


**Figure 1.5 Input–output model of cloud computing (general view).**

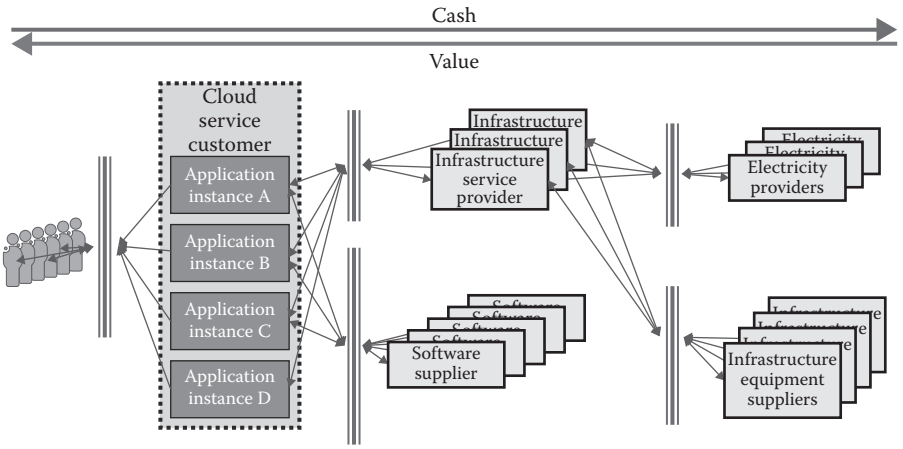


**Figure 1.6 Input–output model of cloud computing (cloud service customer view).**

from left (cloud service users) to right (hardware, software, and electricity suppliers), and value increases from right to left (e.g., suppliers to CSPs to CSCs).

As this analysis focuses on the CSC, the CSC's view of the input–output model shown in Figure 1.6 is useful. A single CSC will often have application instances deployed on CSP infrastructure in several geographic regions, which gives the CSC flexibility to

1. Place an individual user's workload onto a particular application instance based on the user's location, workload patterns, application performance characteristics, maintenance schedules, and myriad other factors
2. Scale application capacity in different infrastructures based on resource pricing and other considerations
3. Rapidly mitigate service impact of disaster scenarios that impact a single cloud data center or region

## 1.4 Key Characteristics of Cloud Computing

ISO/IEC (ISO/IEC, 2014-10-15) stipulates that cloud computing has six key characteristics*:

- Broad network access (Section 1.4.1)
- Measured service (Section 1.4.2)

* The six (ISO/IEC, 2014-10-15) key characteristics of cloud computing are fundamentally the five essential characteristics of cloud computing offered by NIST in Mell and Grance (September 2011) plus *multitenancy* (Section 1.4.3).

- Multitenancy (Section 1.4.3)
- On-demand self-service (Section 1.4.4)
- Rapid elasticity and scalability (Section 1.4.5)
- Resource pooling (Section 1.4.6)

### 1.4.1 Broad Network Access

Broad network access is defined as "a feature where the physical and virtual resources are available over a network and accessed through standard mechanisms that promote use by heterogeneous client platforms. The focus of this key characteristic is that cloud computing offers an increased level of convenience in that *users can access physical and virtual resources from wherever they need to work*, as long as it is network accessible, using a wide variety of clients including devices such as mobile phones, tablets, laptops, and workstations" (ISO/IEC, 2014-10-15). Operationally, this means that end users can access cloud-based application services via generally available wireless and wireline IP networks.

### 1.4.2 Measured Service

Measured service is defined as "a feature where the metered delivery of cloud services is such that usage can be monitored, controlled, reported, and billed…. The focus of this key characteristic is that *the customer may only pay for the resources that they use*. From the customers' perspective, cloud computing offers the users value by enabling a switch from a low efficiency and asset utilization business model to a high efficiency one" (ISO/IEC, 2014-10-15). When CSCs pay only for resources that are used, application services that are engineered so cloud resource usage tracks with application service usage, which tracks with application revenue, can reduce business risk by better linking the application service provider's costs with application service revenues.

### 1.4.3 Multitenancy

Multitenancy is defined as "a feature where physical or virtual resources are allocated in such a way that multiple tenants and their computations and data are isolated from and inaccessible to one another" (ISO/IEC, 2014-10-15). Multitenancy enables CSPs to share virtualized resources across many applications with different patterns of demand and boost the utilization of their physical infrastructure equipment.

### 1.4.4 On-Demand Self-Service

On-demand self-service is defined as "a feature where a cloud service customer can provision computing capabilities, as needed, automatically or with minimal

interaction with the cloud service provider. The focus of this key characteristic is that *cloud computing offers users a relative reduction in costs, time, and effort needed to take an action, since it grants the user the ability to do what they need, when they need it, without requiring additional human user interactions or overhead*" (ISO/IEC, 2014-10-15). This means that CSCs (or automated systems working on their behalf) can install, configure, and provision cloud resources to serve their applications in real time. On-demand self-service of capacity planning and fulfillment actions, coupled with rapid elasticity, enables significant reductions in fulfillment times for capacity change actions compared to traditional deployments.

### *1.4.5 Rapid Elasticity and Scalability*

Rapid elasticity and scalability is defined as "a feature where physical or virtual resources can be rapidly and elastically adjusted, in some cases automatically, to quickly increase or decrease resources. For the cloud service customer, the physical or virtual resources available for provisioning often appear to be unlimited and can be purchased in any quantity at any time automatically, subject to constraints of service agreements. Therefore, the focus of this key characteristic is that cloud computing means that the *customers no longer need to worry about limited resources and might not need to worry about capacity planning*" (ISO/IEC, 2014-10-15). Application service providers (or automated systems working on their behalf) can allocate and release infrastructure resources on the fly, thereby enabling application service providers to transform from allocating and configuring capacity based on peak forecast demand (which may never even be approached) to just-in-time, demand-driven capacity configuration.

### *1.4.6 Resource Pooling*

Resource pooling is defined as "a feature where a cloud service provider's physical or virtual resources can be aggregated in order to serve one or more cloud service customers…. From the customer's perspective, all they know is that the service works, while they generally have no control or knowledge over how the resources are being provided or where the resources are located. This offloads some of the customer's original workload, such as maintenance requirements, to the provider" (ISO/IEC, 2014-10-15). Resource pooling, coupled with multitenancy, enables CSPs to leverage economies of scale to boost operational efficiencies beyond what has traditionally been feasible.

## 1.5 Cloud Service Management Fundamentals

CSCs are fundamentally executing information technology (IT) service management processes to support the design, transition, delivery, and improvement of some

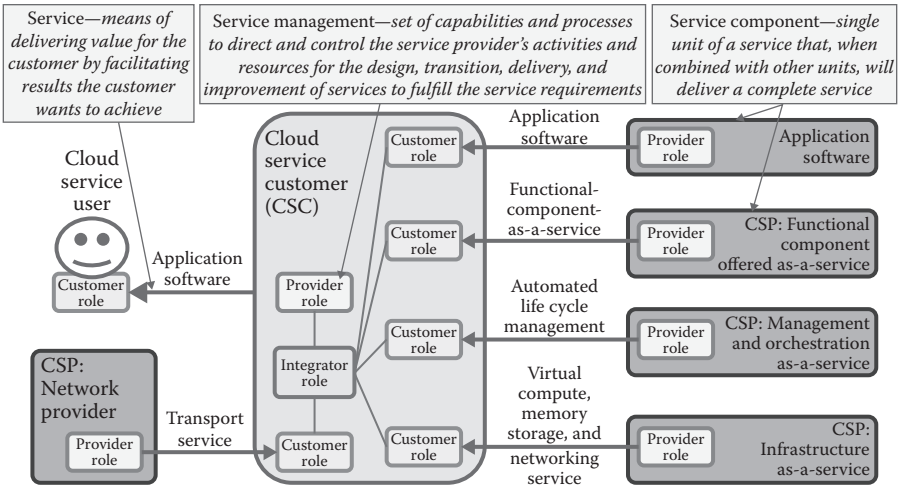


**Figure 1.7 Fundamental IT service management concepts.**

information technology service offered to cloud service users. The ISO/IEC 20000 family of standards is the most authoritative reference for IT service management* for the design, transition, delivery, and improvement of information technology services. Figure 1.7 overlays standard IT service management concepts relevant to cloud user service quality onto the canonical cloud accountability framework of Figure 1.5. A CSC delivers an application *service* to end users that is produced by a chain of *service components*. The CSC's IT *service management* processes support the design, transition, delivery, and improvement of that service.

ISO/IEC 20000-1, *IT Service Management System Requirements* (ISO/IEC, 2011-04-15), formally defines the concepts of Figure 1.7 as follows:

> **Service—**means of delivering value for the customer by facilitating results the customer wants to achieve

* The ISO/IEC 20000 series of IT service management standards includes the following:
- ISO/IEC 20000-1:2011, *Service Management System Requirements*
- ISO/IEC 20000-2:2012, *Guidance on the Application of Service Management Systems*
- ISO/IEC 20000-3:2012, *Guidance on Scope Definition and Applicability of ISO/IEC 20000-1*
- ISO/IEC Technical Report 20000-4:2010, *Process Reference Model*
- ISO/IEC Technical Report 20000-5:2013, *Exemplar Implementation Plan for ISO/IEC 20000-1*
- ISO/IEC Technical Report 20000-9:2015, *Guidance on the Application of ISO/IEC 20000-1 to Cloud Services*
- ISO/IEC Technical Report 20000-10:2015, *Concepts and Terminology*
- ISO/IEC Technical Report 20000-11:2015, *Guidance on the Relationship Between ISO/IEC 20000-1:2011 and Service Management Frameworks: ITIL®*.

**Service management**—set of capabilities and processes to direct and control the service provider's activities and resources for the design, transition, delivery and improvement of services to fulfill the service requirements

**Service component—**single unit of a service that when combined with other units will deliver a complete service

*EXAMPLE*: hardware, software, tools, applications, documentation, information, processes or supporting services.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

# *Chapter 2*

# Desired Cloud Service Customer Benefits

Cloud computing is a disruptive technology that can unlock business value for organizations that effectively exploit it. Cloud service providers (CSPs) offer shared, elastic resources on demand in a pay-as-you-go business model to cloud service customer (CSC) organizations. CSC organizations leverage the elastic, on-demand resources via agile techniques to build application services and value faster than with traditional deployment models. In addition, CSCs can offer service to users with materially lower business risk because the pay-as-you-go resource pricing dramatically lowers the cost of an unsuccessful (think "fail fast, fail cheap") service offering; elastic scalability coupled with pay-as-you-go resource pricing enables the CSC to efficiently cover the upside of surging user demand.

Section 2.1 reviews the standard Cloud Infrastructure Service Provider Business Models, and Section 2.2 considers the Cloud Service Customer Business Model. Section 2.3, "Factoring Benefits of Cloud/Network Functions Virtualization," considers how the canonical benefits of cloud computing are likely to accrue to both CSPs and customers. Section 2.4 reviews the standard IT Service Management Objectives, and Section 2.5 reviews the Focus of this Work, including what topics are out of scope.

## 2.1 Cloud Infrastructure Service Provider Business Models

Sustainable business models that enable both CSPs and CSCs to thrive are materially different from traditional business models. In particular, cloud infrastructure

service provider organizations take on the capital expense of physical compute, memory, storage, and network infrastructure equipment, as well as physical data centers to house the equipment, along with providing electric power, operations, administration, maintenance, and support to offer on-demand resources to CSC organizations. No successful business would invest to build and offer cloud infrastructure without a clear model for obtaining a business value for both the up-front capital investment and the ongoing costs of operating the physical equipment to deliver infrastructure-as-a-service.

The commercial (aka public) CSP business model is fairly straightforward: the aspiring cloud infrastructure service provider organization constructs a business model that covers the capital costs of the equipping and operating their cloud data centers, and then recovers their costs plus a return via fees paid by their CSCs for usage of their elastic, on-demand cloud infrastructure resources. To survive, the public CSP strives to maximize the service revenue returned on their infrastructure investment while minimizing their operating expenses. The business model is not unlike that of a commercial airline: having sunk a large investment in aircraft capacity, the enterprise strives to maximizes return on that investment by filling their service capacity at the highest price the market will bear. As readers are undoubtedly well aware, clever pricing (aka yield management) of airline tickets, hotel rooms, and myriad other services is a best practice for managing (aka shaping) service demand to maximize a service provider's return on finite physical assets. Fundamentally, charging CSCs on a fee-for-service basis simplifies the CSP's business model because of the virtuous cycle in which increasing service demand leads to greater revenues, which justify greater investments in service delivery capacity.

The CSP has to solve two fundamental business problems:

1. Deploy the right amount of physical infrastructure capacity in the right geographies to serve CSC demand. The right amount of physical infrastructure capacity for a CSP is driven by their CSCs' needs, business model, capital resources, appetite for risk, and other factors.
2. Shape CSC demand for virtual resource services to maximize the CSP's operational efficiencies so the business case works. For example, resource pricing might vary based on the following: time of day; grade of service such as whether resource service can be curtailed or preempted; whether capacity was reserved or purchased on the spot market; and so on.

Intelligent pricing models for infrastructure service enable CSCs to shape their individual infrastructure demand so that the aggregate infrastructure demand that the CSP serves is high enough to maximize the CSP's operational efficiencies. For instance, applications with high resource demands during the day (e.g., when their human users are awake) should be financially incented to release excess resources in off-peak periods, and applications that can run anytime (e.g., remote software updates) should be financially incented to shift their demand to off-peak periods

to level the CSP's workload by increasing usage in less popular periods. Proper resource pricing is a win for all stakeholders:

- CSCs are charged a fair pay-as-you-go price that permits them to acquire resources on demand and to release those resources when they are no longer needed to reduce their costs.
- CSCs further reduce their costs by shaping their cloud service demand to increase the CSP's operational efficiencies (e.g., discounts for off-peak periods, discounts for accepting voluntary demand management actions by the CSP).
- CSPs invest and deploy sufficient physical capacity to serve all demand, shape demand to maximize utilization and operational efficiencies of their finite physical capacity, and share enough of the savings with their customers that both the CSP and their CSCs win.

There are four standard deployment models for cloud infrastructure:

- **Public cloud**—"cloud services are potentially available to any cloud service customer and resources are controlled by the cloud service provider…" (ISO/IEC, 2014-10-15)
- **Private cloud**—"cloud services are used exclusively by a single cloud service customer and resources are controlled by that cloud service customer…" (ISO/IEC, 2014-10-15)
- **Community cloud**—"cloud services exclusively support and are shared by a specific collection of cloud service customers who have shared requirements and a relationship with one another, and where resources are controlled by at least one member of this collection…" (ISO/IEC, 2014-10-15)
- **Hybrid cloud**—"Cloud deployment model using at least two different cloud deployment models…" (ISO/IEC, 2014-10-15)

Public cloud infrastructure-as-a-service providers will use the aforementioned market-based pricing model for balancing CSP infrastructure capacity supply with aggregate CSC infrastructure service demand. Some private cloud, community cloud, and hybrid cloud infrastructure-as-a-service providers will also use market-based pricing of virtual resources to optimally address their two fundamental business problems of deploying the right amount of physical infrastructure and shaping demand to maximize their operational efficiencies.

Private and community cloud infrastructure-as-a-service provider arrangements that do *not* rely on market-based pricing of virtualized infrastructure resources must find some other mechanism (e.g., centralized planning) to solve the fundamental problems of deploying sufficient capacity to serve aggregate CSC demand with acceptable quality and with sufficient efficiency that operating expenses are

controlled. The history of centralized planning in the last century does not suggest that a utopian model of planned resource allocation will work as well (Hayek, 1944) for private, community, or hybrid clouds as market-based pricing mechanisms might; this work focuses on the CSC's business. Fortunately, bursting CSC demand that exceeds a private or community CSP's capacity to an appropriate public CSP (along with appropriate demand management techniques) may provide a practical alternative to the traditional capacity management practice of deploying materially more hardware capacity than is necessary to serve the peak forecast demand several years into the future.

## 2.2 Cloud Service Customer Business Model

CSCs have three fundamental business problems:

- **Develop and deploy compelling application services to users faster than competitors**—Delivering new services and value to users faster than one's competitors is a significant business advantage. The essential cloud characteristic of broad network access (Section 1.4.1) means that cloud service users can be served by myriad CSC organizations, rather than merely CSC organizations in the same city, state, or region as the user. Successful CSCs must bring compelling services to market faster than their competitors to gain critical mass and market share. Thus, velocity of service innovation and agility are critical success factors for CSCs.
- **Scale up capacity for successful services on a pay-as-you-go basis**—It is fundamentally difficult to predict if and when application services will become popular (e.g., go viral). The essential cloud characteristics of rapid elasticity and scalability (Section 1.4.5) coupled with measured service (Section 1.4.2) and usage-based pricing enable CSCs to deploy modest online application capacity initially with modest pay-as-you-go costs and rapidly scale capacity with user demand. Thus, more of the CSC's operating expenses—and hopefully revenue—track with actual service demand. Coupling more of the CSC's operating expenses for pay-as-you-go virtual resources with actual application service demand derisks the CSC's application business case by reducing frontend costs that must be sunk regardless of whether the application is successful or not.
- **Retire unsuccessful services quickly and inexpensively (aka fail fast, fail cheap)**—Inevitably, many service prototypes, trials and deployments will not be commercially successful, so they should be retired expeditiously to control costs. On-demand self-service (Section 1.4.4), Rapid elasticity and scalability (Section 1.4.5), and measured service (Section 1.4.2) enable the CSC to expeditiously retire unsuccessful or obsolete services and immediately stop paying for them, thereby *failing fast* and *failing cheap*.

Solving these three fundamental CSC business problems enables organizations to rapidly and inexpensively prototype and trial services until they hit upon the right offering that becomes popular with users.

## 2.3 Factoring Benefits of Cloud/Network Functions Virtualization

Cloud computing offers a range of exciting benefits; however, the upside rewards and downside risks are not evenly distributed to CSPs and CSCs. This section considers how the nine "benefits of network functions virtualisation" expected by the world's largest telecommunications service providers for their massive planned investment in cloud/network functions virtualization (NFV) in the European Telecommunications Standards Institute (ETSI) NFV white paper (ETSI, 2012-10-22) apply to both CSCs and cloud infrastructure service providers. Benefits of cloud computing suggested by other sources are broadly similar to these benefits. None of the nine benefits in the white paper are specific to the telecommunications industry; enterprises in other businesses are likely to be seeking most or all of these benefits when they invest in cloud. Section 2.3.10 gives a Benefit Summary.

### *2.3.1 Reduced Equipment Costs*

The first benefit given in the NFV white paper (ETSI, 2012-10-22) is as follows:

> Reduced equipment costs and reduced power consumption through consolidating equipment and exploiting the economies of scale of the IT industry.

The primary implications of this benefit:

- Suppliers of open, commodity compute, memory, storage, and networking infrastructure for the Information Technology (IT) industry are in fierce competition to maximize throughput and power efficiency.
- CSP organizations who purchase this infrastructure equipment directly enjoy the benefits of this competition among suppliers.

The benefit of reduced equipment costs is captured by infrastructure-as-a-service cloud providers who own and operate infrastructure equipment. The CSP may, or may not, decide to pass along some of those cost savings to their CSCs. There is a small faulty infrastructure capex reduction risk (Table 26.5) to user service quality in that some infrastructure equipment cost reduction feature may compromise compatibility in a way that creates user service impact.

### 2.3.2 Increased Velocity of Time to Market

The second benefit given in the NFV white paper (ETSI, 2012-10-22) is as follows:

> Increased velocity of Time to Market by minimising the typical network operator cycle of innovation. Economies of scale required to cover investments in hardware-based functionalities are no longer applicable for software-based development, making feasible other modes of feature evolution.

This enables the following benefits for CSCs:

- **Leverage software-, platform-, and infrastructure-as-a-service** to both derisk development (because offered services are demonstrated to be stable and mature) and shorten time to market (by selecting only as-a-service offerings that are generally available).
- **Leverage the open and diverse ecosystem** of software and integration suppliers. Rather than having to develop all software from scratch, CSCs can source application components from industry and open-source projects. Suppliers with expertise in integration, testing, and other specialties can be contracted to further accelerate time to market.
- **Apply modern development practices** like Agile and DevOps. While Agile and DevOps can be applied to applications hosted on traditional hardware platforms, the cloud characteristics of rapid elasticity and scalability (Section 1.4.5) and on-demand self-service (Section 1.4.4) make it easier to fully apply Agile and DevOps principles and practices.
- **Leverage rapid elasticity and scalability** to shorten test intervals by executing more test cases in parallel on elastically scaled test-bed capacity.

CSPs can also accelerate their pace of innovation compared to traditional deployments. Shortening a CSC's service design and transition intervals to increase velocity inevitably carries a modest risk that flawed process and tool changes will allow user service–impacting defects to escape into the service operation phase.

### 2.3.3 Reduced Development Costs and Intervals

The third benefit given in the NFV white paper (ETSI, 2012-10-22) is as follows:

> The possibility of running production, test and reference facilities on the same infrastructure provides much more efficient test and integration, reducing development costs and time to market.

The fulfillment steps described in Section 2.3.2 largely apply to reduced development costs and intervals as well. In addition,

- Leveraging off-the-shelf functional components offered as-a-service by CSPs or software from commercial suppliers or open-source projects is generally faster and cheaper than developing bespoke service components.
- Leveraging standardized interfaces and automated life cycle management mechanisms can accelerate service integration activities.
- Leveraging on-demand resource capacity can eliminate bottlenecks associated with scheduling development, integration, and testing activities onto a finite pool of target compute, memory, storage, and networking resources. For instance, as cloud capacity is elastic and scalable, CSCs can order sufficient test-bed capacity to potentially execute all tests in parallel to shorten test pass intervals rather than having to serialize test case execution across a limited number of test beds.
- Improving test effectiveness by reducing the difference between test configurations and production configurations is enabled by cloud. In addition, organizations can use on-demand elastic resource capacity to create huge test configurations that are exercised by huge fleets of test clients to verify at-scale performance that traditionally was impractical or infeasible.

Infrastructure CSPs are likely to remain somewhat constrained by hardware, but functional component as-a-service and management and orchestration CSPs can potentially reduce their development costs and intervals as CSCs can.

Shortening a CSC's service design and transition intervals to increase velocity inevitably carries a modest risk that flawed process and tool changes will allow user service–impacting defects to escape into the service operation phase.

### *2.3.4 Targeted Service Introduction and Rapid Scaling*

The fourth benefit given in the NFV white paper (ETSI, 2012-10-22) is as follows:

> Targeted service introduction based on geography or customer sets is possible. Services can be rapidly scaled up/down as required. In addition, service velocity is improved by provisioning remotely in software without any site visits required to install new hardware.

CSCs leverage rapid elasticity and scalability (Section 1.4.5) and on-demand self-service (Section 1.4.4) to scale online application capacity ahead of demand so that the CSC's opex tracks closer to application demand, which hopefully is tied to revenue or business value; Chapter 6, "Lean Application Capacity Management" considers this topic in detail. For instance, a new service can be deployed with limited capacity (and hence modest cost) and promoted to a limited target market; if the service proves popular, then the CSC can rapidly scale online application to serve that rising demand.

Functional component as-a-service and management and orchestration CSPs can also capture benefits of targeted service introduction and rapid scaling. Note that infrastructure service is fundamentally tied to actual capacity of the underlying physical infrastructure equipment, which is not subject to rapid scaling, and thus, infrastructure CSPs cannot fully capture the benefits of rapid scaling.

Targeted service introduction and rapid scaling inherently carry user service quality risk because if online capacity is not correctly scaled ahead of user demand, then at least some users will not receive acceptable service quality. Specific user service quality risks associated with rapid scaling include the following:

- Faulty scaling decision criteria risk (Table 20.3)
- Inaccurate demand forecast risk (Table 20.4)
- Life cycle management (Execution) risks (Chapter 23)
- Visibility risks (Chapter 19)
- Faulty resource placement policy risk (Table 20.2)

### 2.3.5 Open and Diverse Ecosystem

The fifth benefit given in the NFV white paper (ETSI, 2012-10-22) is as follows:

> Enabling a wide variety of eco-systems and encouraging openness. It opens the virtual appliance market to pure software entrants, small players and academia, encouraging more innovation to bring new services and new revenue streams quickly at much lower risk.

This benefit is primarily captured by CSCs who have ready access to a wide range of VNFs, open-source projects, and functional components to source service components from. As many of these service components will be offered off the shelf from some supplier or service provider's catalog, they will be available faster, cheaper, and probably with higher service quality than bespoke service components.

Software-as-a-service and platform-as-a-service providers who offer functional components can also leverage diverse software products from across the ecosystem. Infrastructure CSPs have less ability to leverage diverse players across the ecosystem because their drive for operational efficiency may push them to aggressively drive commonality and eliminate complexity from their operational environment.

While new suppliers of key service components can bring benefits to CSCs, they also raise risks to user service quality, especially the following:

- VNF product risks (Chapter 13)—Unfamiliar software suppliers, especially start-ups, may have weak development processes, which allow more residual defects to escape into production resulting in VNFs that may be less reliable than those from best-in-class software suppliers.

- Service integration risks (Chapter 18)—Unfamiliar software suppliers, especially start-ups, may have different integration requirements and assumptions compared to familiar and best-in-class software suppliers, thereby increasing the risk of a service integration defect being introduced and escaping into the service operation phase.
- Visibility risks (Chapter 19)—Unfamiliar software suppliers, especially start-ups, may not provide sufficient visibility into operation of their component to enable rapid and reliable localization and root cause analysis of service quality impairments.
- Service policy risks (Chapter 20)—Unfamiliar software suppliers, especially start-ups, may have unusual, and perhaps unstated, operational policy needs for optimal performance.
- Accountability risks (Chapter 21)—Unfamiliar software suppliers, especially start-ups, may not have consistently and clearly articulated their roles, responsibilities, and demarcation points.

Appropriate supplier qualification diligence can treat these risks.

### *2.3.6 Optimized Capacity and Workload Placement*

The sixth benefit given in the NFV white paper (ETSI, 2012-10-22) is as follows:

> Optimizing network configuration and/or topology in near real time based on the actual traffic/mobility patterns and service demand. For example, optimisation of the location & assignment of resources to network functions automatically…

CSCs can leverage rapid elasticity and scalability (Section 1.4.5) and on-demand self-service (Section 1.4.4) to optimally place online application capacity near cloud service users to assure the best quality of user experience. Serving user demand from application instances hosted in a local cloud data center should both reduce transport latency (because photons or electrons don't travel so far) and improve networking quality (because fewer intermediate systems and facilities are in the service delivery path to introduce network impairments). Likewise, software-as-a-service and platform-as-a-service providers can colocate online service capacity with the CSC application instances to optimize service latency, reliability, and availability.

Note that optimized capacity and workload placement has a completely different view from the infrastructure CSP, as their goal is likely to be driving up utilization of their physical infrastructure equipment to maximize their operational efficiency. For example, CSPs can maximize their operational efficiency by placing CSC workloads onto their physical infrastructure equipment to smooth and optimize aggregate demand for their cloud services.

Faulty capacity optimization or workload placement risks delivering unacceptable service quality to some cloud service users.

### 2.3.7 Multitenancy Support

The seventh benefit given in the NFV white paper (ETSI, 2012-10-22) is as follows:

> Supporting multi-tenancy thereby allowing network operators to provide tailored services and connectivity for multiple users, applications or internal systems or other network operators, all co-existing on the same hardware with appropriate secure separation of administrative domains.

Having multiple CSCs all coexisting on the same hardware (i.e., multitenancy [Section 1.4.3]) fundamentally benefits CSPs as it leverages resource pooling (Section 1.4.6) and permits the infrastructure service providers to increase utilization of their physical resources and drive operational efficiency improvements. However, multitenancy support increases the CSC's application user service quality risk due to virtual machine risks (Chapter 14), virtual networking risks (Chapter 15), virtual storage risks (Chapter 16), and virtualized application latency risks (Chapter 17) ultimately caused by resource-sharing policies and multitenancy operations.

### 2.3.8 Reduced Power Consumption

The eighth benefit given in the NFV white paper (ETSI, 2012-10-22) is as follows:

> Reduced energy consumption by exploiting power management features in standard servers and storage, as well as workload consolidation and location optimisation. For example, relying on virtualisation techniques it would be possible to concentrate the workload on a smaller number of servers during off-peak hours (e.g., overnight) so that all the other servers can be switched off or put into an energy saving mode.

Infrastructure CSPs pay for the electricity that powers physical infrastructure equipment and the data centers that house that equipment, so cost savings due to reduced power consumption are captured by the CSP.

How the CSP balances their desire for reduced power consumption against consistently high-quality delivery of virtual infrastructure services to their CSCs determines the level of user service quality risk. For example, "live" virtual machine migration enables a CSP to consolidate workloads so they can power off infrastructure equipment offering capacity that is not needed during low-usage periods to reduce power consumption, but the virtual machine migration event may cause

transient impact to user service quality. Uncoordinated power management actions by CSPs risk compromising the user service quality delivered to CSC users being served by service components hosted in the virtual resources being manipulated. Thus, it is essential that CSPs and CSCs agree on what level of virtual resource service impact is acceptable and what coordination/orchestration will be provided to minimize the risk of user service impact.

### *2.3.9 Improved Operational Efficiency*

The ninth benefit given in the NFV white paper (ETSI, 2012-10-22) is as follows:

> Improved operational efficiency by taking advantage of the higher uniformity of the physical network platform and its homogeneity to other support platforms:
>
> - IT orchestration mechanisms provide automated installation, scaling-up and scaling out of capacity, and re-use of Virtual Machine (VM) builds.
> - Eliminating the need for application-specific hardware. The skills base across the industry for operating standard high volume IT servers is much larger and less fragmented than for today's telecom-specific network equipment.
> - Reduction in variety of equipment for planning & provisioning. Assuming tools are developed for automation and to deal with the increased software complexity of virtualisation.
> - Option to temporarily repair failures by automated re-configuration and moving network workloads onto spare capacity using IT orchestration mechanisms. This could be used to reduce the cost of 24/7 operations by mitigating failures automatically.
> - The potential to gain more efficiency between IT and Network Operations.
> - The potential to support in-service software upgrade (ISSU) with easy reversion by installing the new version of a Virtualised Network Appliance (VNA) as a new Virtual Machine (VM). Assuming traffic can be transferred from the old VM to the new VM without interrupting service. For some applications it may be necessary to synchronise the state of the new VM with the old VM.

The improved operational efficiency scenarios listed associated with *taking advantage of the higher uniformity of the physical network platform and its homogeneity* will primarily be captured by the CSPs who own and operate the uniform and homogeneous physical network platforms. However, some of the automated life