Business Intelligence and the Cloud

Michael S. Gendron

WILEY

Business Intelligence and the Cloud

Wiley & SAS Business Series

The Wiley & SAS Business Series presents books that help senior-level managers with their critical management decisions.

Titles in the Wiley & SAS Business Series include:

Activity-Based Management for Financial Institutions: Driving Bottom-Line Results by Brent Bahnub

Bank Fraud: Using Technology to Combat Losses by Revathi Subramanian

Big Data Analytics: Turning Big Data into Big Money by Frank Ohlhorst

Branded! How Retailers Engage Consumers with Social Media and Mobility by Bernie Brennan and Lori Schafer

Business Analytics for Customer Intelligence by Gert Laursen

Business Analytics for Managers: Taking Business Intelligence beyond Reporting by Gert Laursen and Jesper Thorlund

The Business Forecasting Deal: Exposing Bad Practices and Providing Practical Solutions by Michael Gilliland

Business Intelligence Applied: Implementing an Effective Information and Communications Technology Infrastructure by Michael S. Gendron

Business Intelligence in the Cloud: Strategic Implementation Guide by Michael S. Gendron

Business Intelligence Success Factors: Tools for Aligning Your Business in the Global Economy by Olivia Parr Rud

CIO Best Practices: Enabling Strategic Value with Information Technology, second edition, by Joe Stenzel

Connecting Organizational Silos: Taking Knowledge Flow Management to the Next Level with Social Media by Frank Leistner

Credit Risk Assessment: The New Lending System for Borrowers, Lenders, and Investors by Clark Abrahams and Mingyuan Zhang

Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring by Naeem Siddiqi

The Data Asset: How Smart Companies Govern Their Data for Business Success by Tony Fisher

Delivering Business Analytics: Practical Guidelines for Best Practice by Evan Stubbs

Demand-Driven Forecasting: A Structured Approach to Forecasting, second edition, by Charles Chase

Demand-Driven Inventory Optimization and Replenishment: Creating a More Efficient Supply Chain by Robert A. Davis

The Executive's Guide to Enterprise Social Media Strategy: How Social Networks Are Radically Transforming Your Business by David Thomas and Mike Barlow

Economic and Business Forecasting: Analyzing and Interpreting Econometric Results by John Silvia, Azhar Iqbal, Kaylyn Swankoski, Sarah Watt, and Sam Bullard

Executive's Guide to Solvency II by David Buckham, Jason Wahl, and Stuart Rose

Fair Lending Compliance: Intelligence and Implications for Credit Risk Management by Clark R. Abrahams and Mingyuan Zhang

Foreign Currency Financial Reporting from Euros to Yen to Yuan: A Guide to Fundamental Concepts and Practical Applications by Robert Rowan

Health Analytics: Gaining the Insights to Transform Health Care by Jason Burke

Heuristics in Analytics: A Practical Perspective of What Influences Our Analytical World by Carlos Andre Reis Pinheiro and Fiona McNeill

Human Capital Analytics: How to Harness the Potential of Your Organization's Greatest Asset by Gene Pease, Boyce Byerly, and Jac Fitz-enz

Implement, Improve and Expand Your Statewide Longitudinal Data System: Creating a Culture of Data in Education by Jamie McQuiggan and Armistead Sapp

Information Revolution: Using the Information Evolution Model to Grow Your Business by Jim Davis, Gloria J. Miller, and Allan Russell

Killer Analytics: Top 20 Metrics Missing from your Balance Sheet by Mark Brown

Manufacturing Best Practices: Optimizing Productivity and Product Quality by Bobby Hull

Marketing Automation: Practical Steps to More Effective Direct Marketing by Jeff LeSueur

Mastering Organizational Knowledge Flow: How to Make Knowledge Sharing Work by Frank Leistner

The New Know: Innovation Powered by Analytics by Thornton May

Performance Management: Integrating Strategy Execution, Methodologies, Risk, and Analytics by Gary Cokins

Predictive Business Analytics: Forward-Looking Capabilities to Improve Business Performance by Lawrence Maisel and Gary Cokins

Retail Analytics: The Secret Weapon by Emmett Cox

Social Network Analysis in Telecommunications by Carlos Andre Reis Pinheiro

Statistical Thinking: Improving Business Performance, second edition, by Roger W. Hoerl and Ronald D. Snee

Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics by Bill Franks

Too Big to Ignore: The Business Case for Big Data by Phil Simon

The Value of Business Analytics: Identifying the Path to Profitability by Evan Stubbs

Visual Six Sigma: Making Data Analysis Lean by Ian Cox, Marie A. Gaudard, Philip J. Ramsey, Mia L. Stephens, and Leo Wright

Win with Advanced Business Analytics: Creating Business Value from Your Data by Jean Paul Isson and Jesse Harriott

For more information on any of the above titles, please visit www.wiley.com.

Business Intelligence and the Cloud

Strategic Implementation Guide

Michael S. Gendron

WILEY

Cover image: © iStockphoto.com/teekid
Cover design: Michael Rutkowski

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

Library of Congress Cataloging-in-Publication Data:

Gendron, Michael S., 1957–
Business intelligence and the cloud : strategic implementation guide / Michael S. Gendron.
1 online resource. — (Wiley & SAS business series)
Includes index.
Description based on print version record and CIP data provided by publisher; resource not viewed.
ISBN 978-1-118-85974-2 (pdf) — ISBN 978-1-118-85984-1 (epub) — ISBN 978-1-118-63172-0 (cloth) — ISBN 978-1-118-91524-0 (obook)
1. Information technology—Management. 2. Cloud computing. 3. Business intelligence. I. Title.
HD30.2
658.4'72028546782—dc23

2013047042

Printed in the United States of America
10 9 8 7 6 5 4 3 2 1

This book is dedicated to the reviewers who added important insights, making each chapter more relevant. I am forever grateful for their perseverance and unwavering support during this project. Their knowledge and background added much that will assist the readers in maximizing their understanding of business intelligence, Big Data, mobile, and the cloud.

Contents

Foreword

In 1991 ADNET Technologies was cofounded with a vision of bridging the gap between information technology (IT) and business—essentially, making IT real for businesses. Michael Gendron and I first met several years ago when he was working on his book *Business Intelligence Applied: Implementing an Effective Information and Communications Technology Infrastructure*. He was (and still is) a professor of information systems at Central Connecticut State University. It was immediately apparent that we shared a common vision to bridge the gap between IT and business.

In the past 30 years we have seen the rise and fall of an array of technologies. Are we now in the process of witnessing the fall of the personal computer as tablets and smartphones take over? I will leave that for history to tell, but what is certain is that technology changes at a rapid pace. As the speed of technology changes, so does the speed of business and competition. As technology has become part of the DNA of business, the speed of technological change has naturally affected the rate at which businesses change. The effects have been widespread, including leveling the playing field between large and small companies. If technology is changing at such a rapid pace that even large companies have difficulty staying abreast of the change, how can small companies possibly compete? What chance do they have in this ever-changing world of technology? The answer is the cloud. The book you are reading makes that point.

While ever more powerful devices with easier-to-use operating systems and applications have made their way from the corporate workplace into our homes, the real game changer has been the rapid adoption of the Internet and the widespread availability of high-speed broadband service. What was not realistically possible (i.e., affordable) back in 2000 is suddenly common. The planets aligned, so to speak, and the "cloud" was born.

Having maintained responsibilities in both the technical and the business worlds, I have seen firsthand the disconnection between the two. My career has been dedicated to bridging the gap between IT (what is possible) and business (the reason IT exists). The author and I share the same drive and vision, as is evident in this book.

Over time, IT has shifted from merely automating manual processes and being viewed as an "expense to be managed" to a business driver. The chief information officer has morphed from a technical guru with minimal business knowledge to a technically savvy business strategist with more of a direct line to becoming the chief executive officer. In my experience consulting with organizations on technology strategy, it is generally the case that the gap between IT and business needs has not been fulfilled. As a result, we spend a lot of time attempting to educate our clients on the very topics covered in this book.

Michael Gendron has done an excellent job of providing the reader with a solid foundation of what the cloud is and how it came about. He provides the reader with the background of the cloud as well as an overview of some of the more important underlying technologies that drive the cloud, such as virtualization. This discussion provides the reader with the necessary foundation to begin to fully engage in discussions on cloud computing.

Once the foundation has been set, the three dimensions of cloud computing are explored, starting with the essential characteristics. What is the cloud? What are the service models? What are the deployment models? Why do I care? These questions are answered in depth and with an eye toward the business reader to begin to bridge the gap between IT and business. Michael continually builds on the foundation set earlier, weaving through the multitude of terms and concepts that face the reader in the real world—cutting through the hype and myths to simplify what may at first seem complicated.

The old adage "you can't manage what you don't measure" still applies and has certainly not been forgotten here. As the reader will discover, the decision to move to cloud computing is not always clear-cut; it depends on a multitude of factors. Measurement of the financial effects of such a move, both short-term and long-term, must be considered. How do we measure the financial impact of the cloud? How do we balance the financial consequences with performance (service level agreements)? How does our staff fit into the overall equation? Total cost of ownership and the measurement of results are given due importance in making decisions. Examples are used to help put the various concepts into context so that the reader can gain a better understanding of how the pieces of the puzzle fit together.

The puzzle is completed in this book with a solid discussion of current trends and how businesses can benefit from those trends. The Big Data, mobile computing, and business intelligence discussions not only educate the reader on each of these trends but also highlight the intersection of these trends with cloud computing.

IT must be an integral part of an organization's strategy and strategic planning process. This book meets the challenge by presenting the material from a strategic implementation perspective. Michael Gendron has a unique talent for maintaining the technological strength of a topic while making it clearly understandable to the business reader—successfully bridging the gap between IT and business. I highly recommend this book for any manager looking to gain a competitive advantage through the strategic application of technology.

EDWARD D. LAPRADE
President and CEO
ADNET Technologies
Farmington, CT
www.thinkADNET.com

Preface

Here is a not-too-unreal scenario in today's C-suite meeting: Sally, the chief information officer of a large company, was at a meeting the other day and was asked by the president of the company, "I keep reading about business intelligence, business analytics, Big Data, and mobile and cloud computing in the trade journals. But it is all so confusing. Can you give us 10 minutes right now and explain how these things can help our business?"

Sally realizes that she is not prepared to answer this question and turns to Sue, the manager of technical services, and says, "Sue, you've recently been to a conference in Atlanta on Big Data and mobile. Can you shed some light on this?" Sue does a great job of fielding the question, but a realization that more background is needed slowly sinks in during the question-and-answer session after her impromptu presentation.

Talking after the meeting, Sue and Sally realize they need to explore the intersection among Big Data, mobile and cloud, and business intelligence (BI) so they can propose a solution that will engage the organization in these technologies. They ponder this intersection knowing that the budget process is about to be upon them. They need to jump-start their understanding so they can justify their budget requests.

Organizational managers are called upon to understand Big Data, BI, mobile and cloud computing, and to craft business and information technology (IT) strategies based on these technologies. The entire IT ecosystem is being disrupted with new offerings and lower technology prices that affect virtually every business decision. Managers are required to justify their projects based on total cost of ownership, return on investment, and capital versus operational expense. Books, trade journals, web sites, and consultants abound that will tell you how to use these technologies, and these are a great resource. However, it is helpful to have a book that provides explanations in an

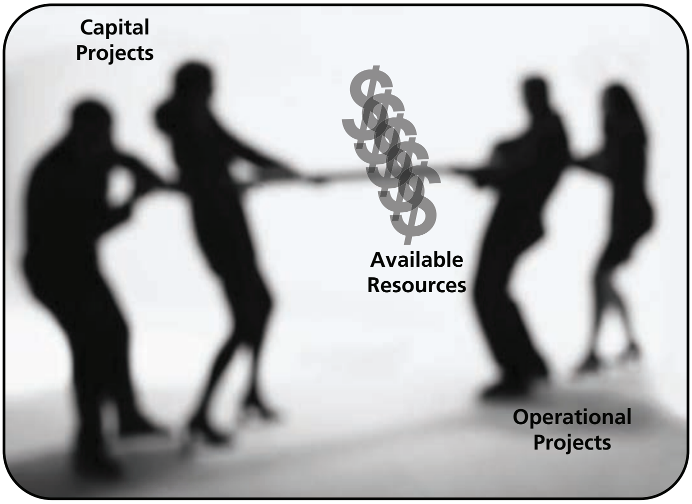


Figure P.1 Projects Must Compete for Funding

easily digestible way so that C-suite members, managers, and staff can obtain the background necessary to deploy the technologies that will bring the greatest competitive return. This book is just such a resource.

Before we delve into the history of how we evolved to cloud computing, it will be helpful if we discuss this book's strategic orientation. This book takes the position that all projects considered for funding by an organization must be weighed against one another before funding decisions are made (see Figure P.1). It does not matter whether they are capital projects (e.g., building a new building or getting new computers for all the employees) or operational projects (e.g., repairing the parking lot or maintaining a software application written in-house). All projects must be considered based on the value each brings to the organization and its customers; the organization must ask what value each project can reasonably expect to yield. For example, will an information and communication technology (ICT) project bring better market share and thus greater profit, or will a building renovation have a greater effect? All projects' return on investment and their alignment with the business strategy must be considered much as you would when purchasing stocks for an investment portfolio.

Implementing a cloud computing strategy at its core is an ICT project. Even though cloud computing is primarily a technology-focused business orientation, at its core it is ICT. Therefore, organizations must consider the ICT value proposition when considering whether to fund a cloud computing project. The *ICT value proposition* is the value that any technology implementation (including a cloud computing implementation) will bring to the organization or its customer.

Here are some basic terms to aid in the discussion:

Information technology (IT). The hardware used to process data and to enable business processes. This is really just the technology, but it is often used to mean the information systems.

Information systems (IS). IT plus software, people, and the actual business processes. This is everything IT, plus the soft side. The terms *IT* and *IS* are often used synonymously, but they really are different.

Information and communications technology (ICT). This is IT, IS, and the communications technologies (e.g., leased lines, routers, switches, etc.) that are used to connect them. This is a new term being used in both business and academia.

To maximize the benefit of ICT, your organization must believe that ICT resources are not only being used to support the business's internal processes but are also being dedicated to projects that create value (e.g., to increase market share, enhance customer relationships, add to the customer's perception of value). Gone are the days when the IT department can ask, "What is my budget this year?" Rather, the question must be, "How can we use technology to support the organization's value proposition(s), and how much is the organization willing to spend to do so?"

Things have changed; we can no longer build ICT using a "field of dreams" approach—if we built it, they will use it. Rather, we must use a decided approach to expending an organization's scarce resources on ICT projects. Otherwise, we will continue to see business units develop "shadow IT" to solve their problems.

The projects selected for funding must support the organization's value proposition, and shadow IT must be avoided. Fuller treatment of the topic of ICT value can be found in the author's previous book, *Business Intelligence Applied: Implementing an Effective Information and Communications Technology Infrastructure*.

Acknowledgments

I want to acknowledge those who helped me create this book. First, I want to thank those who have written about the topics covered in this book before I did; there are too many to list here, but they created the intellectual capital that formed the foundation for this work. I also want to thank my family for its support and understanding while I created this work. Last, I want to thank the reviewers who spent countless hours reading everything written in these pages. The reviewers and their comments about this work are as follows:

> Providing innovative services to customers in an era when technology is being redefined with a new paradigm of IT economics is a game changer for every business. Reviewing this book helped clarify the message on the emerging trends of analytics and driving maximum business value.
>
> **—Larry Carvalho, Principal Consultant, Robust Cloud**

> Reviewing this book was not a chore; it was a fun learning experience. The book is a significant addition to our understanding of the present and future impact of cloud computing.
>
> **—William Holstein, Professor of Strategy and Information Technology, Lorange Institute of Business, Zurich; D. Hollins Ryan Professor, Retired, College of William and Mary; Distinguished Service Professor Emeritus, SUNY Albany**

> In providing IT advisory services to our clients we are often required to go deep into specific areas of technology

in seeking solutions to business problems. This book was a good reminder that we must not lose focus on how all these technologies interact.

—Edward Laprade, President and CEO, ADNET Technologies

Simply stated, my thoughts while reviewing this book were simply to show, and add, additional data to the myriad of examples of cloud, mobility, and BI. The ranges of ideas are seemingly infinite—and it was a pleasure to be a sounding board!

—Christopher Luise, Executive Vice President, ADNET Technologies

Michael Gendron presents a well-written path to managing the complexity of organizations understanding and implementing true cloud strategies. Beyond the marketing hype, this book is a must read for the practitioner and helps set the right framework.

—Chad Mattix, Principal Consultant, Transformation and Global Solutions

This book is invaluable to segments of the IT and non-IT communities that are struggling with the implications of the explosion of corporate Big Data.

—Fred Wergeles, Competitive Intelligence Consultant, Fred Wergeles and Associates

The following companies were involved in the creation of this book:

Adnet Technologies

Intel Strategy

Fred Wegeles & Associates

CompuCom

RobustCloud

PART ONE

The Foundation

*The convergence of cloud and the exponential rise in data are [*sic*] creating unprecedented opportunities for IT professionals, but [*sic*] it also is leading to an ever-increasing knowledge and skills gap in the IT industry.*

—Tim Horton*

There is much confusion over what cloud computing is. In Part One we review the technological history that brought us to cloud computing; then attributes and definitions of cloud computing are brought together to create a unified definition. We conclude by defining cloud delivery mechanisms and public, private, hybrid, and community clouds. In Parts Two and Three we will move into a discussion of how cloud computing can be used to deliver business intelligence.

*Tim Horton is the head of the Cloud Computing Centre of Excellence at Cork Institute of Technology in Ireland; this quote is from the biographical information on his LinkedIn profile.

CHAPTER 1

A History of How We Got to Cloud Computing

Cloud computing is not just a technical orientation, a set of computer architectures, or computing standards developed by the technocrats of the twenty-first century. It is a new way to deliver information systems (IS) that require new business policies; it is a strategic orientation driven by changes in technology that enable rapid business innovation. Cloud computing is enabled by many technologies, but more important than the technologies, cloud computing is a new way to deliver IS and information technology (IT). Many of the old ways (if you can call five years ago "old") are still around and are needed. However, the organization that wants to be competitive has to closely examine cloud computing as a strategic initiative.

The decision to build and fund a cloud computing strategy is one that must be driven by the highest levels of your organization. Making the decision to move from a traditional in-house technology strategy (possibly a "bare metal server approach") to a cloud computing strategy represents substantial exposure for the organization in the form of risk and expense, but the benefits of a successful cloud computing strategy will far outweigh that risk and expense. Those benefits include the potential for greater innovation, faster time to market, greater market share, and greater customer loyalty.

In this chapter, a brief history is given to introduce how the evolution of technology has brought us to cloud computing. Special attention is given to service-oriented architecture and why it is important, and to virtualization and why the reader should care about it. This brief history will set the stage for Chapter 2, which will define cloud computing and its delivery models and describe how those models fit into an organization. Cases are given for many of the difficult concepts so the models of delivery for cloud computing services can be understood. These topics provide the foundation for understanding how cloud computing can help an organization attain its strategic orientation.

THE RISE OF CLOUD COMPUTING

As already said, cloud computing is a business model and a new way to deliver IS and IT. That new delivery is through a "pay as you go" model. To understand how we got to cloud computing, it is helpful to understand the historical evolution of these foundational technologies. Understanding the history informs us about the pitfalls made by others and gives us a foundation on which to base our forward-looking strategies. Reviewing this history will help us understand how cloud computing creates business strategy. (This history will be an overview of some of the major technology antecedents to cloud computing, not a detailed technical review.)

Computing Hardware

Changes in computing hardware have been a large driving force in how IS and IT are delivered to the user. In early computing, the mainframe was largely the "computer in the backroom" approach, in which the users did not have any direct access. Today, the users have direct access and have more processing power on their desks than what was in those early mainframes. In this section, we will examine some of the trends in the last 60 years.

Early Computing and the Mainframe

Through the 1950s and 1960s, large corporations adopted mainframe computers to process their data. The integrated circuit (IC) was

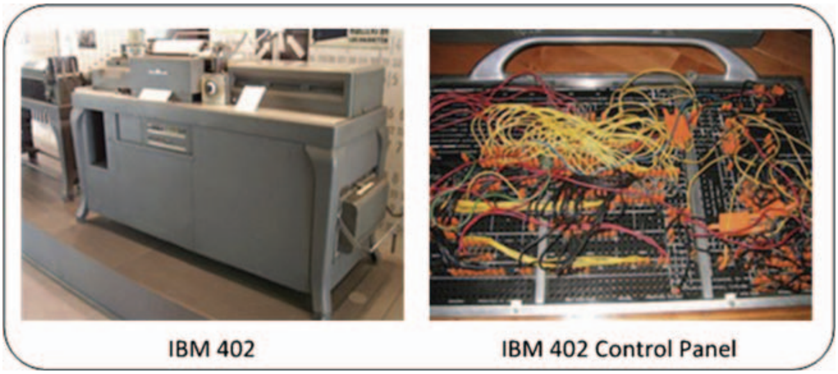


Figure 1.1 IBM 402 *Source*: IBM 402 photo courtesy of Stahlkocher. Control panel photo taken by Chris Shrigley in May 2003.

developed in 1958 by John Kirby and Robert Noyce.[1] In the ensuing years, the IC was reduced in size and cost, making it the technology that would eventually enable the creation of the microcomputer and minicomputer. Before the IC, discrete transistors were the building blocks for mainframes. Terms such as *business intelligence* (BI) and *Big Data* had not been conceived yet, nor was there the need for technology to support them. The mainframe was created to support large organizations in their need to process their data.

By today's standards, these early systems were difficult to use. They had no direct human interface (i.e., no terminal or keyboard) but were given data and programs through hard wiring, punch cards, or paper tape. The IBM 402 (see Figure 1.1) was such a machine, but many other examples also existed.

In these early days of modern computing, "IBM and the seven dwarfs" dominated the computer industry; the "seven dwarfs" were Burroughs, Control Data Corporation, General Electric, Honeywell, National Cash Register (NCR), RCA, and UNIVAC. In the early 1970s, General Electric sold its computer business to Honeywell, and RCA sold to Sperry. The remaining companies became known as "IBM and the BUNCH"; BUNCH was an acronym for Burroughs, UNIVAC NCR, Control Data Corporation, and Honeywell. These companies formed the foundation of modern mainframe computing.

Mainframes remain in use today; they provide *big iron* processing power in organizations that process large amounts of data. These

Table 1.1 Mainframes and Supercomputer Attributes

Mainframe	Supercomputer
■ Handles large amounts of data quickly ■ Good at processing a large number of business transactions quickly	■ Handles large amounts of data quickly ■ Good at processing data and performing scientific number crunching

machines are powerful and able to handle large amounts of data quickly. This makes them especially suitable for things like back-end processing to support a government need to handle its tax returns quickly or an organization's need to process a large inventory in an enterprise resource planning system. Mainframes are good at moving large amounts of data and processing them quickly so other users and other systems (perhaps personal computer–based systems) can further process the information. The mainframe exists as the right tool to do the right job. It's a matter of engineering and architecture: A personal computer (PC) using a high-end graphics card is better at displaying images and streaming video than a mainframe, but a mainframe is better at processing large amounts of transaction-oriented data.

It is important not to confuse mainframes with supercomputers. Both are good at handling a large amount of data quickly, but they have different purposes. Supercomputers are fast at processing data and performing scientific number crunching. Mainframes are good at moving large amounts of data and processing large numbers of business transactions quickly. This is summarized in Table 1.1. With the emergence of cloud computing, multiple servers can be used in unison to approach or even exceed the processing power of a mainframe, but it is a matter of architecture, cost, and business benefit.

Minicomputers

Through the mid-1960s, a class of computers emerged called *minicomputers*. In 1970, the *New York Times* proposed the definition of a minicomputer as a machine costing less than $25,000. Those machines had to have both input and output devices, at least 4,000 words of memory, and be capable of running programs in a computer language such as Fortran or Basic. The minicomputer was built differently from the mainframe and brought computing power to smaller organizations that could not make the investment for a big iron mainframe, but that still had the need to process more data than a human could do in a reasonable amount of time. These machines did not normally have input and output devices for each staff member in the organization and were mainly used for batch processing. As the minicomputer and its software evolved, and as ICs made devices more accessible, some small and medium-size businesses were able to deploy the minicomputer with individual terminals for staff members, but this was expensive and often not cost-effective.

Microprocessor-Based Computers and the PC

As the use of mainframes and minicomputers became common, users began to develop the need to process their own information—especially considering the large amount of information they needed to handle to do their jobs effectively. Mainframes were too unwieldy and not accessible to the normal staff member in an organization, and minicomputers were too expensive for each staff member to have one.

The development and pricing of the IC, along with other technology trends, made PCs accessible to more people and made small scalable servers a reality. The first PCs were introduced in the 1970s by manufacturers like Coleco, Radio Shack, Commodore, and Atari. For several years, PCs from these and other manufacturers were largely seen as "hobbyist" computers. There was also a doubling of the number of transistors on a processor chip approximately every two years that caused evolutionary developments in processing power and capacity; this phenomenon was known as Moore's Law.[2]

By the late 1970s, the need for all staff members to have access to computing power was better understood. In 1981, IBM announced the PC, and in the same year Microsoft introduced MS-DOS under contract with IBM. There are varied accounts of how MS-DOS was created, ranging from it being a spin-off of the earlier Quick and Dirty Operating System (QDOS) to it being copied from Gary Kildall's CP/M operating system. Microsoft founder Bill Gates convinced IBM to let Microsoft retain the rights to MS-DOS and to market it separately from the IBM PC. Subsequently, in 1985, Microsoft introduced the first version of Windows. Coincident with these events, Steve Jobs introduced the Apple Macintosh (Mac) computer in 1984. The Mac, too, evolved through multiple hardware and software versions. This series of events began the revolution toward desktop computing.

Networking and HTML

While mainframes, minicomputers, and desktop PCs were coming into vogue, another phenomenon was occurring: People were developing ways to connect computers so they could transfer messages to one another. In 1969, the development of ARPANET (which some say is the grandparent of the Internet) began. ARPANET is the network of the Defense Advanced Research Projects Agency (DARPA). DARPA is a branch of the federal government that specializes in "the technological superiority of the U.S. military and preventing technological surprise from harming our national security by sponsoring revolutionary, high-payoff research bridging the gap between fundamental discoveries and their military use."[3] ARPANET was built primarily to give researchers access to mainframes that were too far away from their work locations to make frequent access convenient. ARPANET was the first large-scale packet-switched network, and it eventually gave way to the TCP/IP protocol and the Internet.

ARPANET was built to connect four locations: (1) the University of California, Los Angeles (UCLA), (2) the Stanford Research Institute, (3) the University of California, Santa Barbara, and (4) the University of Utah. ARPANET carried its first message on October 29, 1969, from UCLA to the Stanford Research Institute. ARPANET grew, and in 1970 it reached the East Coast by including the technology company Bolt, Beranek and Newman. By 1981, ARPANET had grown to 213 locations, and it began to acquire approximately one new location every 20 days.

In the meantime (around 1975), a transatlantic satellite link was added to ARPANET, and the Norway Seismic Array was brought online. In 1983 ARPANET was split and the military maintained its own network, MILNET, for unclassified communications. The combination of networks for unclassified military and civilian traffic was called the Defense Data Network.

In 1985, the National Science Foundation (NSF) began a program to coordinate its projects over the National Science Foundation

Network (NSFNET). This name was also given to a group of nationwide networks that were constructed from 1985 to 1995. NSFNET was initially created to link the nation's NSF-funded supercomputers, but through collaboration between the NSF and industry partners, the NSFNET became a major part of the Internet. This backbone was connected to a number of regional networks that enabled smaller regional and campus networks to connect to one another through the NSFNET.

The federal law that appropriated the funds for the NSFNET was interpreted to mean that commercial traffic was not allowed. In order to make the appropriate use of the NSFNET known to those connecting to it, the NSF developed an acceptable use policy (AUP). The AUP restricted commercial use even though several commercial Internet service providers (ISPs) were connected to the NSFNET. A number of the early NSFNET industry partners (notably, Merit, IBM, and MCI) created the Commercial Internet Exchange to connect many of the private ISPs, relieving the restrictions imposed by the NSF AUP. The NSFNET became a transitional network between ARPANET and the Internet as we know it today.

Simultaneous to the ARPANET, NSFNET, and Internet transition, standards for local area networks (LANs) were developed. LANs were built to allow computers within organizations to share data and devices such as printers. In the 1970s, there were a number of competing standards—Ethernet, Token Ring, and ARCNET, to name a few of the major ones—and all of these technologies had proponents and a place in the market. With the proliferation of desktop PCs, easy-to-use (compared to the mainframe and minicomputer) operating systems like CP/M and MS-DOS bringing computing power to the desktop, and the user's need to share devices and data, the implementation of LANs grew.

Parc is a Xerox company that describes itself as a technology platform pioneer.[4] Robert Metcalfe and others at Xerox Parc developed Ethernet in the early 1970s, and it was patented in 1975. Because of its ability to adapt to the PC market and its low cost, Ethernet has become the de facto standard for desktop PC and server interconnectivity. Ethernet has been standardized through the Institute of Electrical and Electronics Engineers 802 working group and has evolved from the slow (by today's standards) form of connectivity it initially allowed to the advanced networking we use today.

Another development was the creation of hypertext transfer protocol (HTTP) and hypertext markup language (HTML) by Tim Berners-Lee. Berners-Lee saw the ability to use the newly formed Internet and its supporting protocols and technologies to communicate marked-up text. The proposal and development of HTTP and HTML occurred in the late 1980s and early 1990s, with the first web site ever deployed being that of the European Organization for Nuclear Research in 1991. Berners-Lee went on to create the World Wide Web Foundation (later to be called the World Wide Web Consortium [W3C]) to launch transformative programs that build local capacity to utilize the Web for positive change.[5] The evolution of HTTP and HTML directly contributed to the ability to deliver cloud computing services.

All of these technologies—and many others—have converged to provide the Internet communications infrastructure that is in place today. Companies like Level3.com provide Internet connectivity; AT&T, Verizon, Clear.com, and many others provide mobile connectivity; hardware manufacturers like Cisco provide WiFi and other communication devices; the plain old telephone system has evolved into a digital communication medium; and numerous other companies and providers deliver technologies that converge today to create a global infrastructure that enables us to work (and play) virtually anywhere in the developed world.

Bandwidth

Bandwidth is the rate at which a computer network can transfer information between devices. In this case we are specifically discussing bandwidth sold to businesses that allows them to make connections in their information and communication technology (ICT) infrastructures. Bandwidth is often measured in bits per second or some denomination thereof (e.g., megabits per second, gigabits per second).

As computing platforms developed and the Internet emerged, the cost of bandwidth declined dramatically. The bandwidth marketplace has been inundated with ambitious (some say overambitious) network construction, bankruptcies, and dramatic price declines through 2004. By 2005, the wholesale bandwidth market had stabilized, with prices largely holding at the 2004 level.[6] The prices on the more popular transmission routes experienced the greatest decline and subsequent stabilization. Today there is still constant economic and marketplace pressure on bandwidth price structure.

The decreased cost of communication has enabled cheap telephone calls and inexpensive data transmission. This has fueled the cloud computing revolution and is a large reason that cloud computing has become ubiquitous.

Computing Platforms

With this brief review of the evolution of computers, minicomputer operating systems, and networks completed, we can move on to the effect they had on computing platforms and information systems. It is probably fair to say that these technologies created a "perfect storm" so that current computing platforms and information systems could exist.[7] Three computing platforms have significantly contributed to moving us toward cloud computing: (1) clustering and grid computing, (2) service-oriented architecture, and (3) virtualization.

Clustering and Grid Computing

For a long time, humanity has understood the advantage of clustering. An example of this is the way we process language and how that is used in natural language processing (NLP) by computers to make interactions between humans and machines more humanlike. NLP—a field of computer science, artificial intelligence, and linguistics—refers to software that allows a computer to interact with another computer using spoken human language. NLP creates clusters of sound in human language that can be easily understood, analyzed, and processed by computers.

Clustering can also be seen in the way an Internet search engine delivers its content, grouped by category or cluster. Clustering is a way to group things to make classification easier as well as a way to group computer resources together to maximize their functioning. Clustering is therefore useful in many venues.

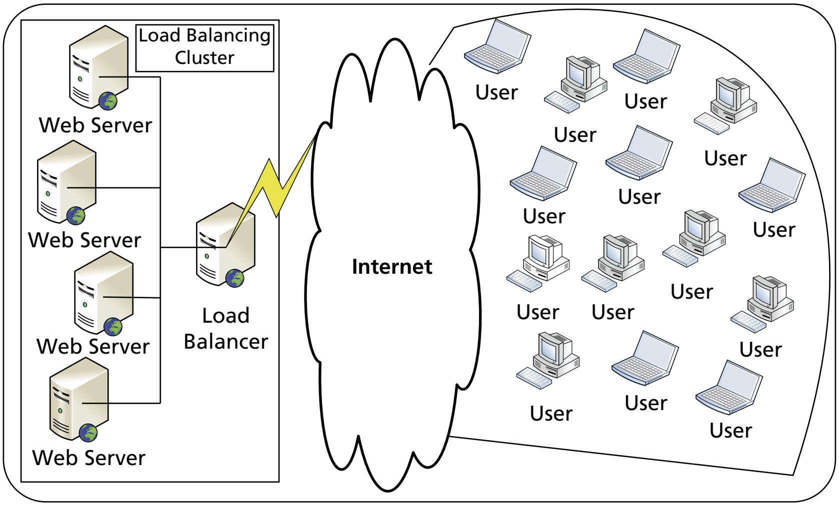


Figure 1.2 Load-Balancing Cluster

Computer clusters are architecture rather than a type of computer. Many types of computers can be clustered together to maximize their use and productivity—think of the adage "the whole is greater than the sum of its parts" applied to computer technology. These clusters can be as simple as two PCs loosely connected so they can share a task or provide backup for each other should one go down. Alternatively, it can be as elegant as a supercomputer composed of a linked set of smaller computers. Clusters are also used to group computer memory and disk drives, which is often referred to as *pooling*.

All computer clusters have one thing in common: They all appear to the user to be one device rather than a number of interconnected devices. Clusters are created with software or hardware that orchestrates the functioning of the cluster to make it appear as one system. A common use of clustering (still very much in use today) is the load-balancing cluster used to deliver web services to a group of users (see Figure 1.2).

The cluster consists of a load balancer, which acts as the orchestrator between the user and the servers of the web site. In this example, behind the load balancer are four servers that each offer up the same web site when requested by a user. The load balancer monitors and distributes the workload of each server. In this way, the load balancer

Table 1.2 Grid and Cluster Computing Characteristics

Cluster Computing Platforms	Grid Computing Platform
Interconnected, and look like one system to the user	Loosely interconnected, and resources do not appear to be one system
Administered using middleware to appear as one system	Administered using middleware that distributes the workload across systems connected to the grid
May be proprietary or have open standards	Often has open standards
Normally has a local or proprietary connection	Can exist on a local or proprietary connection or can interconnect using the Internet

acts as *middleware*, or the orchestrator that manages the servers so their functioning can be maximized. To the user, it appears that he or she is interacting with one web site when in fact there may be a number of web servers that stand ready to deliver the requested web site. It's up to the load balancer to determine which available web server actually does the task.

An architecture that seems similar is grid computing. *Grid computing* platforms have a number of distinguishing characteristics and are compared to cluster computing platforms in Table 1.2. What largely distinguishes a grid from a cluster is that a cluster appears as one system to the user, whereas a grid is true distributed processing. Both clusters and grids use middleware to orchestrate a common task. In our example of cluster computing, the common task was seamlessly delivering a web site to the user. A grid may be used to work on parts of the same problem. A typical grid is shown in Figure 1.3, with one computer that distributes a workload to a number of other computers attached to the Internet.

Service-Oriented Architecture

When service-oriented architecture (SOA) entered the scene, we had already seen all the network standards, grid and cluster computing architectures, and other changes to business and technology that we have discussed thus far. SOA is a software engineering technique

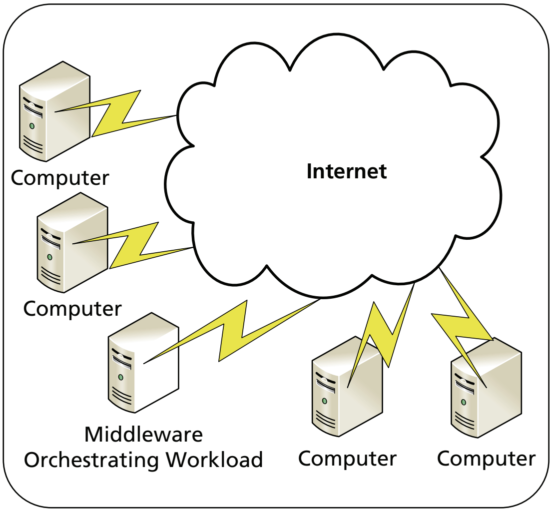


Figure 1.3 A Typical Internet-Based Grid

directed toward designing and building computer software services that interoperate. It was a new way to build computer software that allowed IS developers to build software from components of a business process rather than needing to "reinvent the wheel" each time software was created. This let each organization that provided a service do what it did best, and most cost-effectively.

SOA is a complex but extremely useful and efficient architecture for software development, and many people believe that it was the direct predecessor of cloud computing. Although it is a somewhat technical discussion, we will review some of the concepts that underlie SOA:

- SOA is based on a set of loosely coupled services. An example of this is a web site that uses a service from one vendor to provide shopping cart functionality and uses a service from another vendor (e.g., a bank) to provide credit card validation.
- Each service in an SOA application communicates and exchanges data in well-defined formats or protocols, often XML. These well-defined formats and protocols allow for standardization and interoperability.
- SOA can be seen as orchestration of services from different organizations or vendors. In the example of a web site with

a shopping cart and credit card validation, there would be an orchestrating layer (perhaps the user interface of the web site) that calls and interacts with the services (i.e., the shopping cart and credit card validation).

- Two standards, both based on XML, were initially used to describe the services. Web services description language (WSDL) was frequently used to describe the service itself, and simple object access protocol (SOAP) was frequently used to exchange information between the orchestration layer and the services it calls. Other standards exist (i.e., REST) today.
- A hallmark of SOA is the ability to reuse services from one application in another. In our example, any web site that wanted to use the shopping cart functionality could do so, with proper authorization from the service provider.
- A common implementation of SOA is the web services approach. In that approach there is a service provider that exposes a service so the service consumer can access it.
- Service consumers consume the offerings of a service provider through a software wrapper, which can be a web site, an application, or other software that orchestrates the consumption of the services.
- Services are often provided on a pay-as-you-go basis (e.g., each credit card validation transaction is charged a fee).

IBM provides a convenient conceptual model that describes the SOA architecture as the interaction of three primary parties:

1. The service provider, who publishes a service description and provides implementation for it.
2. The service consumer, who consumes (uses) the service.
3. The service broker, who, if present, maintains a service registry describing the service. Today, the broker is often referred to as the *managed service provider*.

SOA provides key elements that we see in today's cloud computing environments, including software reuse, a pay-as-you-go (or metered use) model, and orchestration. These are very similar to the cloud computing software-as-a-service (SaaS) model, which we will discuss in Chapter 2.

Virtualization

In the earlier stages of computing discussed in this brief historical review, hardware was a single resource—one mainframe meant one mainframe, one server meant one server, and one storage device meant one storage device. There was time slicing (i.e., multiple concurrent operations) on mainframes and minicomputers, but that did not carry forward to the PC. Eventually, we understood that treating each resource individually did not use it effectively; each device had excess capacity that could not be used.

The solution to that problem was virtualization. In computing, *virtualization* means creating copies of something that all share the same physical resource. For example, suppose you have a server (a single resource) and know that that server is being used to only 20 percent of its capacity. It is possible, through virtualization software like VMware or Microsoft Hyper-V, to take the single resource and make it appear like many copies of the same or similar resource.[8] Each copy will execute on the single server, but to the user each copy will appear to be its own dedicated resource.

Three types of virtualization are common: (1) hardware virtualization, (2) desktop virtualization, and (3) application virtualization. These share the following attributes:

- Each virtual instance shares the same hardware. For example, one physical server can be made to look like many physical servers (e.g., a VMware solution); one physical server can be made to provide computing power for many desktops or application (e.g., Microsoft Hyper-V, Citrix, XenDesktop, and XenApp).
- Virtualization consolidates servers to improve utilization and cost-effectiveness; by maximizing the use of the existing resources, the existing hardware can be used more cost-effectively. This also enables an organization to decrease its carbon footprint. The environment and the organization both benefit from lower energy consumption.

- A major advantage of virtualization is that the infrastructure becomes more scalable. The existing hardware can have more virtualized instances added and dropped as needed. This is especially helpful when multiple users are sharing the same physical resource; this is discussed more in the next point.
- Resource sharing often drives virtualization and makes it effective. The fact that multiple users can access the same physical resource and have it appear that each has his or her own resource makes virtualization attractive. Add to this the ability to charge a user by the amount of each resource consumed, and you have an effective business model.
- Virtualization is normally autonomous and based on demand. A virtualized infrastructure in which you purchase the computing resource (e.g., disk or compute power) will sense your usage and scale those resources as needed. As those resources are added or removed, the amount charged is adjusted. This represents a utility model, not unlike the local electric or telephone utility that charges you for what you consume.
- There are unique security issues in a virtualized environment that do not exist in a nonvirtualized infrastructure. For example, in a nonvirtualized environment, each resource may map to one IP address; moving this same environment to a virtualized infrastructure means that each physical resource may have many virtualized resources that it supports, each with its own IP address.

Today we see a mix of nonvirtualized environments, virtualized hardware, and virtualized desktop or application infrastructures implemented. To summarize: hardware virtualization mainly involves making one device appear like many copies of that same (or a similar) device. Desktop virtualization involves making one device (i.e., a server) deliver the entire desktop to multiple client devices simultaneously; application virtualization involves delivering a specific application or set of applications (over a web browser or through another client) to users. An example of each is shown in Figure 1.4. Desktop hardware in a virtualized environment can be less expensive, and centralizing

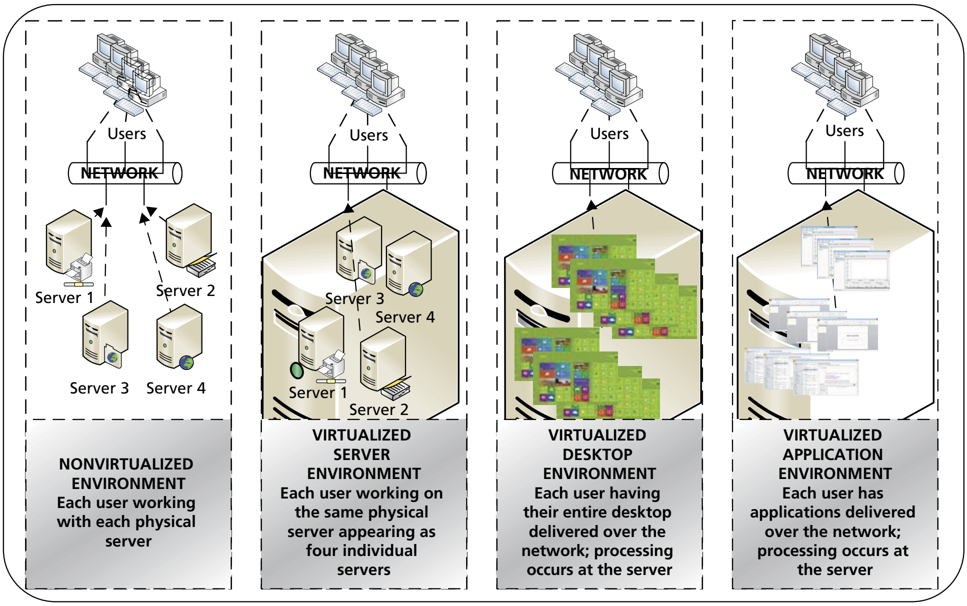


Figure 1.4 Virtualized Server, Desktops, Application

desktop and application delivery can make management of these resources easier and cheaper.

Cloud Computing

Cloud computing evolved from, and has been built on, the technologies discussed in this chapter. At the start of this chapter, a case was made that cloud computing is a new way to deliver IT and IS, and to do business; it is not just a technical orientation. Moving applications to a cloud rather than having them deployed on bare metal servers takes a decided approach that is fundamentally different and can require a change in an organization's strategic orientation. This is especially true when an organization decides to move from delivering all information systems from its in-house technologies and staff to an external cloud vendor. The recent history of technological innovation has prepared us to deliver IT via the cloud through infrastructure-as-a-service (IaaS), platform-as-a-service (PaaS), and software-as-a-service (SaaS). These are discussed in Chapter 2.

CONCLUSION

Looking back over the changes in computing technologies during the last 60 or so years, it is easy to see how we arrived at cloud computing. Functionalities that we think of today as commonplace—like OneClick shopping at Amazon.com, MyLowes at Lowes.com, and online credit authorization—were unheard of before the Internet, HTTP, the World Wide Web, and now cloud computing.

There are many versions of history, and the one presented here is based on the author's own experience in the computing industry as well as some reference materials to assist with dates and other specifics when the timeline might be blurred in his memory. This history is an accurate representation of how we got to cloud computing and what it took to get us there. Hardware, networking, data communications, and software architecture have evolved, and they have come together to bring us to where we are today: the delivery of IT and IS using a cloud computing approach. We are standing at a precipice:

cloud computing has emerged as the dominate way to deliver IT and IS when resource utilization structures must be managed.

Gone are the "field of dreams" days when we built ICT with the hope that the users would find a way for it to add value to the organization and its customers. Today we must ask what the objectives of an ICT project are, how the project can add value, and what resources the organization has to apply to that project.

Before we leave the discussion of how we evolved to cloud computing, it is appropriate to talk about how our current global economic crisis has affected that evolution. The recent fiscal crisis is a major socioeconomic force on technical developments that have pushed us to a greater source of elasticized value with new IT and IS delivery models like cloud computing. It has created a realization that businesses must be more aware of their own budgets and the effect of technology on them—in short, businesses need to find more economical ways to deliver IT and IS. Cloud computing is the next step in computing evolution and will be here to stay. As organizations move from traditional data centers to consuming cloud services, this new way to deliver IT will only become more prevalent.

The historical overview given in this chapter sets the stage for understanding how we got to where we are today, with cloud computing as the inevitable outgrowth of many factors. However, Chapter 1 has only set the stage. That stage is necessary for understanding how and why cloud computing makes sense. The rest of this book will define cloud computing, deal with specific issues affected by cloud computing, and discuss the strategic implementation of cloud-based solutions. As pressure increases to deliver IT and IS in a way that supports business agility, it is more important than ever that IT and IS staff, managers, and the C-suite have the tools to understand and build what is expected. In addition, the growth of social and mobile technology adoption is adding information overload and creating a need for more IT and IS resources to run enterprises and build business intelligence. This book introduces the strategy-oriented concepts and tools required to create cloud-based infrastructure that fulfills the ICT value proposition and builds business intelligence.

NOTES

1. M. Wright, "The Planar IC: Revolution Underestimated," *Edn* 51, no. 9 (2006): 38; and M. Wright, "Miniaturization Enables Innovation—Past, Present, and Future," *Edn* 51, no. 20 (2006): 172–86, http://0-search.proquest.com.www.consuls.org/docview/222434808?accountid=9970.
2. "Moore's Law Inspires Intel Innovation," Intel, www.intel.com/content/www/us/en/silicon-innovations/moores-law-technology.html.
3. "About," DARPA, www.darpa.mil/About.aspx.
4. "Parc Today," Parc, www.parc.com/about.
5. "Frequently Asked Questions," World Wide Web Foundation, www.webfoundation.org/faq.
6. E. Schoonover, "Wholesale Bandwidth Prices Stabilize," *Business Communications Review* 36, no. 7 (2006): 14–17.
7. The *perfect storm*, in colloquial terms, is a set of circumstances that when brought together create the environment for a specific effect. The term is often used to describe weather patterns that come together to create the environment for a particularly strong storm. A business or financial example is the use of the term for the recent global financial crisis.
8. VMware, www.vmware.com/; and "Microsoft Hyper-V Server 2012," Microsoft, www.microsoft.com/en-us/server-cloud/hyper-v-server/default.aspx.

CHAPTER 2

Characteristics and Service Models*

Cloud computing means many things to many people. The Internet, its protocols, and its standards are the technical basis for cloud computing, but the discussion must include the entire organization. In many ways, cloud computing is a technological occurrence, but as emphasized in Chapter 1, cloud computing is a new way of delivering IS and IT. Our historical review showed that cloud computing is the evolutionary product and convergence of many technologies, including the Internet and its growth. In this chapter we will discuss an overview of the first two dimensions of cloud computing: the essential characteristics and the service models. The third dimension, the deployment models, will be covered in Chapter 3.

The plain old telephone system (POTS) has been depicted as a cloud for many years because a person would pick up the phone and

*Some of the material presented in this chapter has also appeared in M. Gendron, *Business Driven Data Communications*, ed. R. Horan (Upper Saddle River, NJ: Pearson Education, 2013); and M. Gendron, "Defining Cloud Architectures: What Attributes Should an Application or Service Have from the Business and Consumer Perspectives?" presented at the Athens Institute for Education and Research, Athens, Greece, 2012. These materials are reproduced by permission of the respective copyright holders.

dial but not know what route the call would take to its destination. The Internet has the same quality: a user enters a web address without knowing what route the data will follow from the user's desktop computer to the server being contacted. This phenomenon is the basis for representing the Internet as a cloud symbol. The next evolutionary step, cloud computing, moves services from the data center, desktop, or other on-premises computers to a cloud. As in POTS, when things are moved to a cloud, where things reside and how connections are handled are unknown to the user.

The basic architecture and other characteristics of the cloud have evolved from a confusing mix of definitions and technical approaches to a current view that is much more stable and clear. Businesses can plan their infrastructures with more confidence now that the different cloud computing approaches are reasonably sound. Organizations can focus on a detailed analysis of alternatives, economic pros and cons, and issues of the fit between technology and business needs rather than on uncertainties about whether cloud computing is just a fad. Major industry and government organizations are settling on approaches and definitions that place cloud computing in a position to offer real solutions to real business problems. Part of that solution will be technology that will assist organizations of all kinds to deal with the explosive growth in IT infrastructure and data.

INTRODUCTORY CONCEPTS

In order to start the discussion of cloud computing dimensions, the stage needs to be set by reviewing some introductory concepts, including abstraction, hosted versus cloud applications, and LAN versus cloud applications. Understanding these things will help as we move into a discussion of the dimensions of cloud computing.

Layers of Abstraction

Cloud computing is defined by layers of abstraction; businesses may not know the specific technologies that are employed in the cloud, but they know the service they want performed. An analogy would be a homeowner contracting for furnace repair; the homeowner may not know or care what brand of tools the repair person uses, but he or she needs to feel comfortable that the repairperson knows how to fix the particular brand of furnace. The furnace repair is the service being purchased, but the tools and other methods are abstract from the person who is requesting the repair service. The same is true of cloud computing. An organization may purchase a service from a cloud-based vendor, but the organization will probably not know much about the technology used by the vendor. All the organization really cares about is whether the cloud vendor can deliver the requested service. (Some employees might want or need to know the technical details, but they are probably in the minority in your organization and among the technical staff.)

When purchasing services from a cloud vendor, the user does not necessarily know what technologies are driving the service that he or she is consuming. An organization signs a contract to purchase a service, such as shopping cart and credit card processing, but normally that contract does not include the specific technical details of the cloud infrastructure. The user consumes a service, possibly through some type of application program interface (API), and the service vendor communicates with the necessary underlying systems and databases to obtain the desired results—for example, a filled shopping cart and an authorized or declined credit card. The services are said to be abstract, as shown in Figure 2.1.

Hosted Applications versus Cloud Computing

What is the difference between hosted applications and cloud computing? Hosted applications have been around since the beginning of the Internet and are referred to as *client-server computing*. A client (e.g., a web browser or an e-mail client) is used to access the hosted application, giving the users access to remote servers. These hosted applications were relatively easy for business and consumers to understand. As software clients evolved, the concepts of object-oriented programming—service-oriented architecture (SOA) and software-as-a-service (SaaS)—also evolved, thus making cloud computing applications and services more abstract. These concepts introduced a level of abstraction that has capitalized on Internet

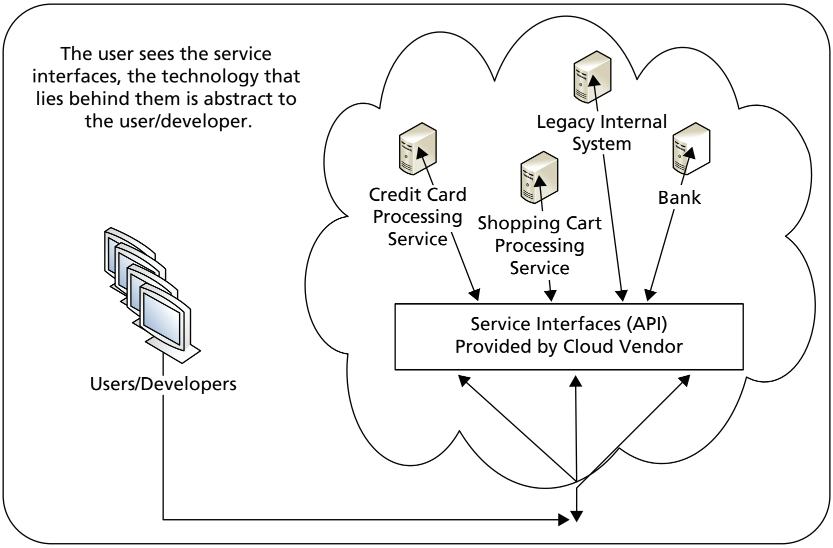


Figure 2.1 Cloud Computing Abstraction—Shopping Cart and Credit Card Processing Example

protocols, allowing them to be combined in new and unique ways to offer many new services. This has also led to a lack of definition of cloud computing and some confusion in the marketplace.

Steve Ballmer, the chief executive officer (CEO) of Microsoft, gave perhaps the best description of the different between hosted servers and cloud computing when he answered this question: Why have hosted servers not historically been called cloud computing? He postulated that hosted servers are not called cloud computing because they have not been restructured to include scaling, fault tolerance, geo-replication, and security in an environment where information is being shared across the cloud. He implied that the restructuring must occur because things are happening outside the firewall.[1]

Since Ballmer gave that explanation in 2008, it is safe to say that today we have seen much of this restructuring and the further definition of cloud computing to include a number of essential characteristics and service and deployment models.

Corporate LAN-Based versus Cloud-Based Applications

Let's take the example of an application offered by a company's data center to remote users over the Internet. It does not matter what the application is; it could be a web site, a payroll application, or something that supports another business process. In outward-facing corporate LAN-based applications offered from the organization's data center, the company must maintain an internal infrastructure with various servers, firewalls, and other technologies to create, enable, and support the application. The company must also have a connection to the Internet that allows remote users to access the application. The organization must maintain the technology and the staff that supports it. This approach to offering the application to remote users is resource-intensive (i.e., it costs money).

What happens when this application is moved into the cloud? Can resource requirements be reduced and application uptime increased? Can the attributes of cloud computing cause the deployment and continued operation of that application to be cheaper, thus increasing return on investment (ROI) and decreasing total cost of ownership (TCO)? Answering those questions is the substance of this book.

Moving the corporate LAN-based application to the cloud yields both advantages and disadvantages. One advantage worth mentioning here (because it is the substance of much of the discussion of cloud computing) is that by purchasing the necessary services from one or more cloud vendors, an organization can reduce the local resources necessary for the deployment of that application. This decrease in resources (i.e., it costs less, at least initially, therefore there is less risk to the organization) potentially decreases the development or deployment time for new applications while also increasing innovation and experimentation. The organization can consume the resources it needs as the demand for the application changes and pay for only the resources it actually uses, potentially reducing the costs of application deployment. The change to ROI and TCO are so important that they are discussed in greater detail elsewhere in this book. Here a simplified example is given.

Examples of a corporate LAN-based deployment and a cloud-based deployment are shown in Figure 2.2. Two different designs are

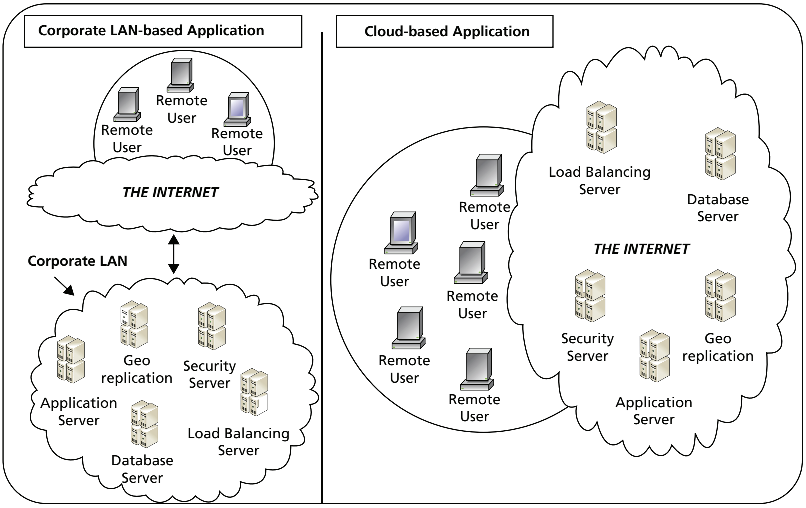


Figure 2.2 Cloud versus Corporate LAN Applications

shown for an application. The corporate LAN-based application runs on the corporate LAN and is offered to remote users over the Internet. The cloud-based application resides on the Internet (or on a public, private, or hybrid cloud), and the user accesses it directly over the Internet. Typically, the corporate LAN is greatly scaled back when applications are deployed to a cloud. The infrastructure necessary to maintain the application and associated services is obtained through cloud service providers, affecting ROI and TCO.

In any discussion of corporate-LAN installations and their connection to cloud-based applications, response time should be considered. "With rising data volumes," Larry Carvalho, principal consultant at Robust Cloud, explains, "Internet connectivity is not always adequate for meeting the service level agreements that users have come to expect based on their experiences with in-house data centers. While the benefits of using cloud infrastructure are very high, it [a cloud infrastructure] does not automatically come with a network capable of delivering the entire spectrum of advantages."[2] In those cases, an organization can avoid Internet connectivity and directly connect (e.g., with a T1 or T3 leased line purchased from a telecommunications carrier) to the cloud provider.

There is crossover among corporate LAN, hosted, and cloud applications. In reality, these are not as discrete as presented here, but their presentation here allows the distinction among them to be understood. When implemented, the corporate LAN, hosted, and cloud applications are often used in concert to support the business processes of the enterprise.

A CLOUD IMPLEMENTATION

Before defining cloud computing dimensions, we will look at 6waves's implementation of cloud-based games on Facebook. This quote gives a perspective on cloud-based applications, and 6waves cloud implementation serves as a good reference example:

> 6waves, a venture-backed company, is the leading international publisher of gaming applications on the Facebook platform. Through partnerships with top developers, 6waves aims to publish the next generation of social games to cover every genre, language, and platform. To date, we [6waves] have over 30 million monthly active players playing our games and applications, a number that continues to grow daily.[3]

This quote shows why many organizations move to the cloud. Global partnerships, a large elastic user base, and the need to cover every platform (some of which are still unknown), language, and genre are a few of the reasons that organizations move to cloud computing—and they are all business reasons, not technology reasons.

6waves uses the following technologies to support its ICT value propositions and enable its gaming applications: Amazon's Elastic Compute Cloud, Simple Storage Service, Elastic Block Store, and Elastic Load Balancing. Cosmos Tong, the system administrator for 6waves, described in an interview how the services fit into the company's architecture. "As the games we publish are relatively independent," he said, "we group servers of different games into separate AWS [Amazon Web Services] accounts and use [Elastic Load Balancing] to set up load balancing and AWS security groups to control access." He further stated that using AWS simplifies the deployment of new

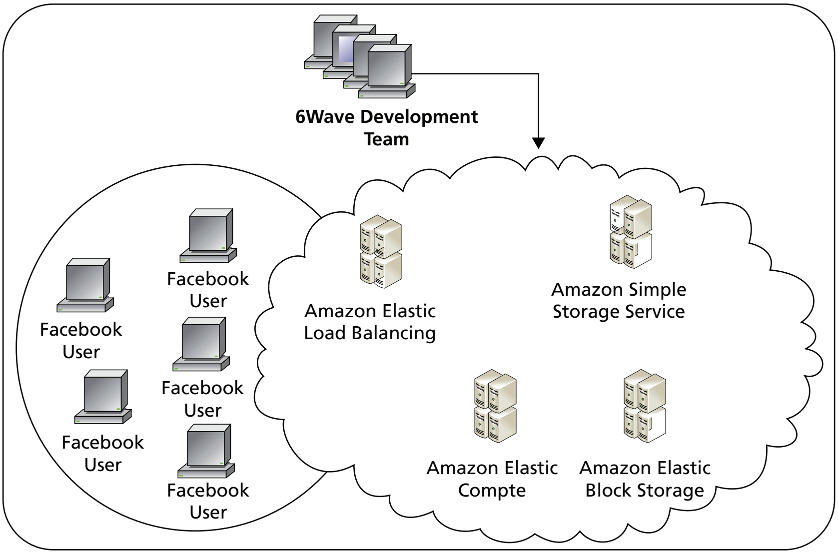


Figure 2.3 6waves Use of Cloud Computing

games. This saves 6waves time and money. An overview of the architecture is shown in Figure 2.3.

DEFINING CLOUD COMPUTING

Cloud computing can be defined along three dimensions: (1) essential characteristics, (2) service models, and (3) deployment models. Dimensions one and two are discussed in this chapter, and the third dimension is addressed in Chapter 3.

Dimension One: Essential Characteristics

In 2008, Larry Ellison, the CEO of Oracle, said that cloud computing can be defined as "everything we already do." He added that cloud computing will just "change the wording on some of our ads."[4] Others have made similar comments, but the understanding and implementation of cloud computing has grown past the sound bites and buzzwords. As we have proceeded to better understand cloud computing and provide definitions that businesses can appreciate, standardized definitions have emerged from organizations like the U.S. National Institute of Standards and Technology.

The substance of cloud computing has gone beyond the media hype. *Cloud computing* may be a term that repackages what the Internet has already enabled, but in reality the term has come to have an additional, and somewhat specific, meaning.

It is possible to look at cloud computing from a technical perspective. In 2009, the Object Modeling Group announced a collaboration with leading technology standards development organizations to coordinate and communicate standards for cloud computing. The organizations that participated in this collaboration are the following: the Distributed Management Task Force, the Open Grid Forum, the Storage Networking Industry Association, Open Cloud Consortium, and the Cloud Security Alliance."[5]

As a result, the Cloud Standards Wiki (www.cloud-standards.org) was formed. This group has proposed a framework for the technical details of cloud computing. This is important work that creates technical standards to support the business model of cloud computing. However, beyond the technical frameworks, there are business and consumer dimensions. The definition given here includes those dimensions and is an amalgamation of the information proposed by Gartner, Microsoft's Steve Ballmer, the National Institute of Standards and Technology, and others.

Cloud computing solutions are offered by many vendors; in fact, there is a large amount of confusion, or at least a lack of understanding, about what the term *cloud computing* really means. Even though the technologies that underlie cloud computing are normally standards-based, cloud computing itself does not have any industry-ratified standards (e.g., by the Institute of Electrical and Electronics Engineers). The Object Modeling Group and others are working on this.

Gartner has highlighted the following five characteristics of cloud computing:

- **Service-based.** Being service-based means that consumer concerns are abstracted from service provider concerns through interfaces that are well-defined. In other words, the person using or buying the service is concerned not with how the service is delivered (i.e., what specific technologies are used to deliver the service) but rather with the correct functioning of the service. The interface to the service hides the implementation details and enables a completely automated response by the service provider to the consumer of the service. The service could be considered ready to use or "off the shelf" because the service is designed to serve the specific needs of a set of consumers, and the technologies are tailored to that need rather than the service being tailored to how the technology works. The articulation of the service feature is based on service levels and ICT outcomes (e.g., availability, response time, performance versus price, clear and predefined operational processes) rather than on technology and its capabilities. In summary, handling a business need (e.g., contract management solutions from companies like Sopima) is more important than how the technologies are used to fill the need.

- **Scalable and elastic.** The service can scale capacity up or down as consumer demands change and can do it at the maximum speed, which may be seconds for some services and hours for others. *Scaling* refers to a system or network ability to adapt and continue to provide services as demand changes. This change may be measured as an increase or decrease in the number of users, as calls to the service, or as some other change in demand. Scaling should occur gracefully, if not automatically. In other words, the service should respond to changing demand by changing available resources. In an ideal situation, the user will not experience any slowdown as the service responds to the changing demand.

  Elasticity is a trait of shared pools of resources, whereas scalability is a feature of the underlying infrastructure and software platforms. Elasticity is associated not only with scale but also with an economic model that enables scaling in both directions—adding or removing resources as needed—in an automated fashion.

- **Shared resources.** Services share a pool of resources (i.e., hardware, software, and communications equipment) to build economies of scale, maximizing efficient deployment. The underlying infrastructure, software, and platforms are shared among the consumers of the service (usually unknown to the consumers), enabling unused resources to serve multiple needs for multiple consumers, with all working at the same time.

- **Metered by use.** Service utilization metrics must be collected to enable resource allocation. Metered use can also enable multiple payment models. Those models can be used to create different pricing plans including pay as you go, subscriptions, fixed plans, and even free plans. Payment plans must be based on usage, not on the cost of the equipment. The amount of the service used by the consumers can be measured in terms of hours, data transfers, or other use-based attributes.
- **Internet technologies.** The service is delivered using Internet identifiers, formats, and protocols, such as URLs, HTTP, ISPs, and web-oriented architectures. Many examples of web technology exist as the foundation for Internet-based services. Some cloud-based examples include Google's Gmail, Amazon's book sales, eBay's auctions, and Lolcats's picture sharing. All of these examples use Internet and web technologies and protocols.[6]

Ballmer of Microsoft has indicated that in order to be considered cloud computing, vendors' services should include three additional attributes:

- **Fault tolerance.** Hardware or software failure is probably inevitable. A fault-tolerant system will adapt to failure and have sufficient capacity and redundancy to respond to faults. The user will experience minimal interruption in service. Cloud-based services should offer fault tolerance to its users and developers so that (1) when hardware failure occurs, the service is automatically moved to other hardware, and (2) when software errors occur, the system either automatically corrects the error or does extensive reporting so the user can correct the error.
- **Geo-replication.** Geo-replication is a way to improve response time and the user experience by distributing servers across a network geographically so the content is closer to the user.

It can also be used to provide backup and to allow organizations to run replicated data centers so the organization can choose where to run applications based on the cost of doing so (e.g., the cost of electrical power). In a geo-replicated system, content is replicated and kept up-to-date across a number of distributed servers. In cloud computing, it might mean having geo-replicated services that appear in every major country. However, it should be noted that with today's fast network and Internet access (normally measured in milliseconds), geo-replication to attain a speed advantage is probably necessary only in the most speed-sensitive applications (e.g., time-critical calculations that allow access to the International Space Station, or some types of controlled nuclear experiments).

- **Security.** Security exists at the physical and IS/IT levels. At the physical level, the data center and the assets used within it must be protected from intrusion and attempts at destruction. IS/IT technologies, policies, and safeguards must exist that prevent unauthorized and/or malicious attempts to change data or interrupt the functioning of the system. IS/IT security includes computing and network security. Cloud computing must encompass both physical and IS security.[7]

These essential characteristics have been proposed by industry leaders concerned with defining cloud computing and are an important part of dimension one. In addition to the essential characteristics from noted industry experts, two more have been offered by the National Institute of Standards and Technology:

- **Broad network access.** Services offered from a cloud vendor are available over a network and accessed through standard mechanisms (e.g., application program interfaces) that promote access through a thin or thick desktop client and/or a mobile device (e.g., tablet, mobile phone, laptop).
- **On-demand self-service.** The service provider must allow the users to set up and change the services they purchase as they wish, without needing to know the underlying technologies and without the direct intervention of the service provider.[8] In addition to

the industry experts and government agency mentioned so far, there are also many standards bodies—such as the Institute of Electrical and Electronics Engineers, the Object Modeling Group, Cloud-Standards, and the International Telecommunication Union—involved in the discussion of the essential characteristics of cloud computing. All of these organizations have their place in creating a marketplace for cloud technology and are involved in setting standards that define both the underlying technologies of cloud computing and the business characteristics. Currently, cloud computing is rife with standards (e.g., TCP/IP, UDDI, REST, SOAP, and XML) that define the underlying technologies, but what is important is bringing to business some definitions that all can agree on.

The essential characteristics discussed previously are divided into those characteristics that are required and those that are optional or desired within a cloud implementation (see Figure 2.4):

- Required for cloud computing
  - **Internet technologies use.** Technologies such as TCP/IP, HTTP, UDDI, REST, and SOAP are used.
  - **Services-based.** A service provider offers a set of services to be consumed.

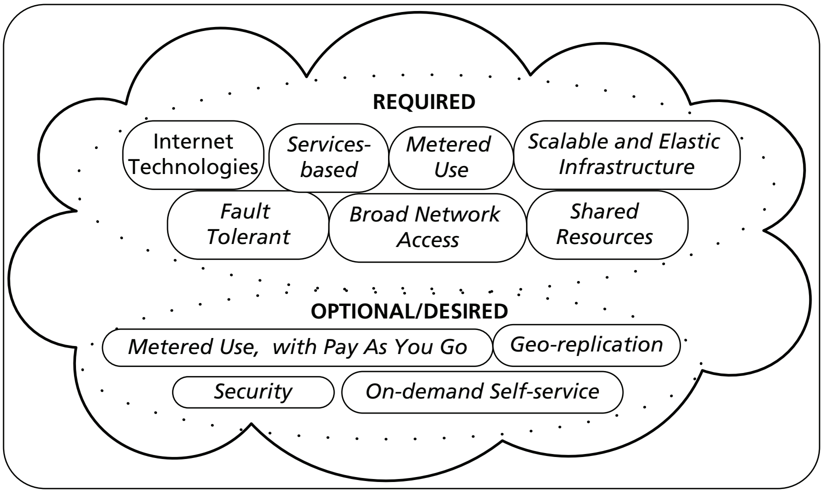


Figure 2.4 Cloud Essential Characteristics

- **Scalable and elastic.** The service offerings scale to the consumer's demand and are flexible.
- **Shared resources.** The service provider pools its resources, making that pooling transparent to the consumer of the services.
- **Fault tolerance.** The infrastructure can detect and react to faults, both on a hardware and a software level.
- **Broad network access.** Services are available over the network and are accessible through a standardized interface (e.g., API).
- **Metered use.** Utilization must be monitored in a cloud environment so that resource pools can be adjusted to accommodate user demand. That adjustment can be done automatically via the cloud itself (i.e., autonomously), or it can be done manually.

- Optional/desired characteristics:
  - **Metered use, with pay as you go.** A private cloud may offer services to its local users without charge, but a public cloud service provider will provide scalable and elastic services with metered use, which will be based on the resources consumed, and then charge based on that metered use. Metered use exists in the private cloud setting, but primarily to monitor utilization and adjust available resources; a charge is not common in the private cloud.
  - **Geo-replication.** Because of today's fast network access, geo-replication is not needed very often, but when it is, the cloud infrastructure should provide it.
  - **On-demand self-service.** In many cases, the service provider may offer consumers the ability to set up and change service contracts themselves, but in other cases this may not be appropriate (e.g., a long-term contract for service provisioning, or cloud implementations in which resource allocation is under the control of someone else, like the IT department).
  - **Security.** Strong security is expected and essential in all cases, but in some settings (e.g., a LAN), security may not be

as strong or as necessary. Responsibility for all security is not automatically passed to cloud providers, and the consumer still must maintain an appropriate level of security (e.g., if a consumer wants to secure files placed on Dropbox, he or she should probably encrypt them before transfer).

The breakdown of essential characteristics into those that are required and those that are optional or desired is used as a framework so that the characteristics can be easily understood. Any particular organization might break these down differently—that is okay, since the industry has not yet settled on what is required and what is optional/desired.

Dimension Two: Service Models

There is a considerable agreement about the service models used by cloud computing. Cloud computing service models refer to how services are delivered to the user. Each delivery system has a number of underlying services that it offers, and these are made available by service providers in three ways: (1) infrastructure-as-a-service (IaaS),

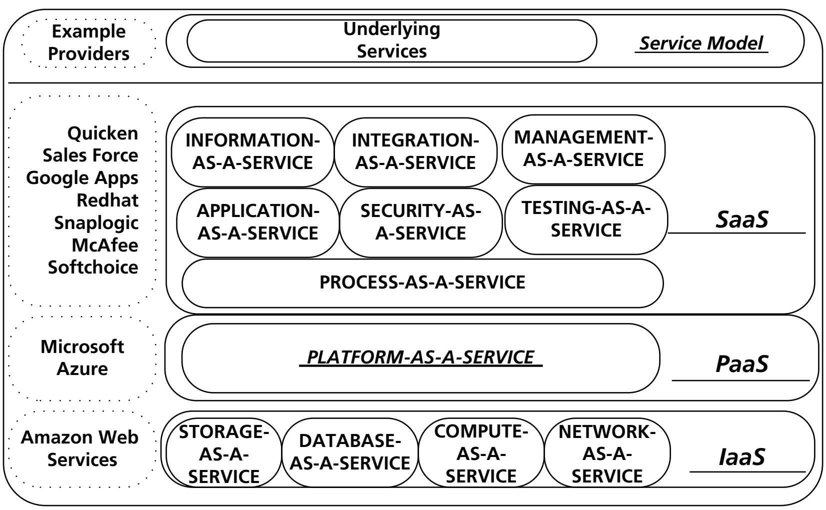


Figure 2.5 Service Delivery Models and Underlying Services

(2) platform-as-a-service (PaaS), and (3) software-as-a-service (SaaS). Not all service providers offer all underlying services, but they do offer a combination of services to fill a business need. Figure 2.5 shows the services that exist within the cloud grouped by service model. Note that this diagram is extensible as necessary to adapt to new service offerings.

Services are used by developers to create applications that solve business problems, and they are consumed by the users of these applications. These services might be used independently as a cloud-based application (e.g., Salesforce) or together to create several cloud-based applications. The user is normally unaware of where these services exist, but rather just knows that they are accessible. The list of services given next is not exhaustive but rather representative of what is currently available. Undoubtedly, as soon as this service list is published, some clever developer will envision new and innovative ways to implement technology that solves unrealized business problems.

The National Institute of Standards and Technology provides good definitions for service models.[9] To those definitions we add a representative list of services.[10] The list of service models and their underlying services are as follows:

- **Software-as-a-service.** A SaaS provider offers the consumer the ability to use the provider's applications running on the service provider's cloud infrastructure. The applications are accessible from various client devices through either a thin interface (e.g., a web browser) or a program interface. The consumer does not manage or control the underlying cloud infrastructure (e.g., network, servers, operating systems, storage, or individual application capabilities), with the possible exception of limited user-specific application configuration settings. Some representative underlying services are the following:
  - **Application-as-a-service.** Any application that is delivered through a browser or specialized client to an end user's PC. Salesforce falls into this category, and so do Google Apps, Google Docs, Google Calendar, and Gmail. Applications delivered as a service may use other services (e.g., storage, or computing) from their own private cloud or from another service provider.

- **Integration-as-a-service.** The features traditionally found in enterprise application integration but delivered as a service. Cloud integration services are offered by companies such as SnapLogic, Boomi, and Cast Iron Systems. Integration-as-a-service offers solutions and services that assist companies in their quest to integrate in-house and cloud-based solutions and may be referred to as a *hybrid cloud*. Cloud integrators provide consulting, training, and service solutions within vertical markets or horizontally across business processes.
- **Security-as-a-service.** "Rather than acquiring your own security software tools and the technical expertise to administer them internally, you contract with security vendors to have a turnkey service of virus defense, firewall management, and e-mail filtering. Outsourcing cyber security eliminates much of the labor and infrastructure, while still giving you the state of the art in anti-virus, firewall, and spam-fighting technologies."[11] Virus defense, firewall, management, and e-mail filtering is offered by vendors such as McAfee and Symantec, and services like identity management are offered through directory services such as Accenture (www.accenture.com).
- **Management-as-a-service.** An on-demand service allows the management of other cloud services. Some examples include asset, topology, resource utilization, virtualization, uptime management, and governance systems. The company Softchoice offers IT asset management as a service.
- **Testing-as-a-service.** The ability to test applications through cloud-delivered testing algorithms that can test web sites, applications, and other software that do not require an on-premises footprint to carry out the testing. Hewlett-Packard offers quality management solutions that, according to the company, uses "a service model that accelerates the implementation of your quality center of excellence."[12]
- **Information-as-a-service.** Any cloud-based service that provides an application programming interface (API) or any other similar method that allows an application to consume information. Examples include providers that offer credit

card verification. These providers will allow you to pass credit card details to them for verification of the card number and the amount of purchase. Some service providers will also process the bank transactions necessary to secure funds from the bank issuing the credit card. Quicken is an example of this. In addition to many other credit card processing providers, numerous information providers offer information or services such as stock prices and address verification.

  - **Process-as-a-service.** This supports business processes through combining other services to create meta-applications. The application at Salesforce (www.salesforce.com) delivers services over the cloud that supports the customer relationship management process of an organization. Salesforce offers the following applications: Sales Cloud, to support sales processes; Service Cloud, to support customer service processes; Chatter, to support collaboration among colleagues in real time; and Force, as a cloud platform for custom application development.

- **Platform-as-a-service.** A PaaS service provider offers the consumer the capability to deploy consumer-created or acquired applications onto the service provider's cloud infrastructure. The consumer does not manage or control the underlying cloud infrastructure—including network, servers, operating systems, or storage—but has control over the deployed applications and possibly configuration settings for the application-hosting environment. PaaS is a subscriber-based service delivered to customers. It typically includes application, interface, and database development, along with storage, testing, and other technologies. PaaS delivers its services from cloud-based hosted data centers, offering a cloud-based infrastructure to organizations so they can deploy their own applications. Those applications may consume the underlying services of other service providers. A few examples are:

  - **Red Hat.** Red Hat delivers PaaS solutions "that can be deployed or offered through a public or private cloud to build, deploy and manage applications across their lifecycle."[13]

- **Microsoft Azure.** The Microsoft Azure platform offers Windows-based tools and resources to an organization so it can deploy its applications on the Azure cloud infrastructure and deliver them over the Internet.
- **Heroku.** Heroku provides PaaS for Ruby on Rails, Node.JS, Clojure, Java, Python, and Scala.
- **Engine Yard.** Engine Yard delivers PaaS that supports Ruby on Rails, PHP, and Node.JS.

- **Infrastructure-as-a-service.** An IaaS service provider offers the consumer the capability of provision processing, storage, networks, and other fundamental computing resources in which the consumer is able to deploy and run software, which can include operating systems and applications. The consumer does not manage or control the underlying cloud infrastructure but has control over the operating systems and deployed applications and possibly limited control of select networking components (e.g., host firewalls). IaaS is delivered in two ways:
  - As an off-site hosting arrangement. However, some people do not see that as true cloud computing, since it is really just housing the equipment in someone else's data center. This arrangement refers to actually providing data-center-as-a-service. The difference between this approach and the one described next is that rather than having a service exposed for the consumer to use, the consumer will place the servers in someone else's data center and connect to the infrastructure. The data center's infrastructure will then attach to the Internet. This service is often referred to as *dedicated hosting*, and companies such as Lunarpages and Go Daddy offer these dedicated solutions.
  - In a service model, like other services described in this chapter, in which the user is not familiar with the underlying technologies, but rather just the interfaces to the services being offered. Some examples are:
    - **Storage-as-a-service.** Also known as *disk on demand*, this is the ability to utilize storage that physically exists

at remote locations but logically appears as local storage to any application that requires it. Storage-as-a-service is a low-level cloud computing service often consumed by other services in this list. Amazon S3 is a cloud-based storage service that is available to web sites and other applications to provide remote storage. Storage-as-a-service is economical for the consumer because the initial capital expenses and the ongoing operational expenses are borne by the provider. The consumer pays for the storage used and the transactions that store and retrieve data. For example, the products Dropbox and AllwaySync use Amazon S3 to allow consumers to store data in the cloud and retrieve it at will.

- **Database-as-a-service.** This allows developers and consumers to use the database services of a provider on a pay-as-you-go basis rather than incurring up-front capital expenses. Database-as-a-service providers frequently offer a number of different database management systems (DBMS) so the application can utilize the one that best fits. The same service provider might offer MS-SQL, MYSQL, and Oracle. It is possible that database-as-a-service providers will offer the DBMS and either offer the storage themselves or use a storage-as-a-service provider. For example, Amazon's offering of its relational database service provides a MYSQL DBMS stored in the Amazon cloud using Amazon S3 (a storage-as-a-service offering).
- **Compute-as-a-service.** In this service, the provider allows the user to have computing power on demand within the service provider's infrastructure. The key to successful computing-as-a-service is scalability (the ability to change computing power as the need for it changes) and elasticity (scalability with an economic model that enables an increase or a decrease of computer services). In addition, the service should scale autonomously by changing computing power as an application or an organization requires. Many service providers sell this service,

including major providers like Amazon Web Services and AT&T.

- **Network-as-a-service.** This provides functionalities to network users or the network itself. The functionalities help the network work well and provide things like domain name resolution (e.g., changing a name like www.gendron.info to 205.144.171.13) and Dynamic Host Configuration Protocol (DHCP), which allows client computers in a LAN to request certain protocol parameters that allow them to connect. Other things on a LAN—like e-mail, printing, and file sharing—can also be network-as-a-service.

These service models (IaaS, Paas, and SaaS) and their underlying services are used to develop or deploy applications. This list of services is a good starting place for businesses and consumers to understand what is offered within the service models. The discussion in this section has given examples of the types, breadth, and depth of services that are offered. By using cloud-based services individually or combining them in unique and innovative ways, the organization can build robust applications at relatively low initial cost to enhance an organization's competitive positioning and deliver value.

Dimension Three: Deployment Models

The National Institute of Standards and Technology has developed a widely accepted set of deployment models for cloud computing environments. These are:

- Private cloud
- Public cloud
- Hybrid cloud
- Community cloud

Each cloud has a number of services that are available to users through a service contract over the Internet. These deployment models are covered in Chapter 3.

CONCLUSION

In this chapter we reviewed the dimensions of cloud computing:

- Dimension one: essential characteristics
- Dimension two: service models
- Dimension three: deployment models

A thorough treatment of essential characteristics and service models was given in this chapter, and so was an introduction to deployment models. Essential characteristics and service models form the basis for understanding cloud computing deployment model, which are more fully discussed in the next chapter.

The essential characteristics and services models are what differentiate true cloud computing environments from the commonplace usage of calling the Internet the cloud. In common parlance, "the cloud" has become anything that can be accessed remotely via the Internet. However, true cloud computing deployments must have the essential characteristics and service models discussed in this chapter and the deployment models discussed in the next one.

It gets confusing. We call the Internet the cloud, and vendors like Western Digital have their new My Cloud[14] product. Many products and services are accessible over the Internet, and they all appear as cloud computing. However, for those in business, it is important to understand and distinguish between "cloud" and "cloud computing." Devices and services available as "cloud" are often useful to solve business problems, but they could be different than "cloud computing." This chapter and Chapter 3 make the definition of *cloud computing* clear. Undoubtedly, we will continue to call many things *cloud*. That is okay, but if you are managing a technical project or making decisions about cloud resource expenditures, it is important to know the difference between *cloud* and *cloud computing* because of the cost, resources, and skills necessary to maximize your organization's investment in these technologies.

NOTES

1. Steve Ballmer, *Steve Ballmer on Cloud Computing* [video], September 29, 2009, http://www.youtube.com/watch?v=ODE8-D-ABb0 (video is no longer available).

2. Larry Carvalho, *How Direct-Access Solutions Can Speed Up Cloud Adoption* (Gigaom Pro and Equinix, 2012).
3. Amazon Web Services, "AWS Case Study: 6 Waves Limited," 2011, http://aws.amazon.com/solutions/case-studies/6waves.
4. Larry Ellison, *Larry Ellison—What The Hell Is Cloud Computing?* [video], Septembet 25, 2008, www.youtube.com/watch?v=0FacYAI6DY0.
5. Object Modeling Group, "Major Standards Organizations Collaborate to Further Adoption of Cloud Standards," 2009, www.omg.org/news/releases/pr2009/07-13-09.htm.
6. Gartner, "Gartner Highlights Five Attributes of Cloud Computing," June 23, 2009, http://www.gartner.com/it/page.jsp?id=1035013.
7. Ballmer, *Steve Ballmer on Cloud Computing* [video].
8. National Institute of Standards and Technology, *The NIST Definition of Cloud Computing*, http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf.
9. Ibid.
10. David Linthicum, *Cloud Computing and SOA Convergence in Your Enterprise: A Step-by-Step Guide*. (Boston: Addison-Wesley, 2010). Many of the underlying services were originally identified by Linthicum and have been expanded or adapted for use in this list.
11. McAfee, "Security-as-a-Service," 2012, www.mcafee.com/us/products/security-as-a-service/index.aspx.
12. Hewlett-Packard Development Company, "HP Testing-as-a-Service," 2013, www8.hp.com/us/en/software-solutions/software.html?compURI=1173731#.Un_GxOKmbQM.
13. Red Hat, "Platform-as-a-Service Solutions," 2013, www.redhat.com/open-hybrid-cloud/paas.
14. WD, "My Cloud," 2013, www.wd.com/en/products/products.aspx?id=1140.

CHAPTER 3

Deployment Models

The IT/IS landscape is changing. For many years we deployed IT/IS in the local data center and on the user desktop. However, we have seen vast evolutionary changes in how IT is delivered to users. Some say that we are returning to the way of the mainframe, where all of the processing is occurring in the data center and the user has a dumb terminal on his or her desk that does not do much more than display the results that a mainframe spews out. That is simply not true. We are doing much of the processing in locations other than the desktop; large transaction processing systems like book buying at Amazon or large database systems like the management of customer relationships at Salesforce are doing the processing, and they are delivering their results to the user to display. Following are some major differences between mainframe and cloud processing:

- The digital connection that organizations have to other users and services from cloud providers that is enabled by the Internet.
- The shared processing that occurs between the desktop and the cloud provider.
- Changes in deployment of IT that affect the cost structure of operational and capital expenses (this might have the greatest effect on innovation and deployment).

Throughout this book, the terms *a cloud* and *the cloud* are used somewhat interchangeably. That is a bit at odds to what is often discussed in marketing literature as "moving to *the* cloud." That phrase is a bit of a misnomer, but it sometimes makes sense in context. There is really not one cloud, although the Internet has been envisioned as such; organizations create their own public, private or hybrid clouds in a SaaS, a PaaS, or an IaaS service model, or they buy services from a cloud computing provider.

Another major change largely enabled by cloud computing is that applications are being designed and delivered at a pace that was unknown in the past. Time to market is critical for technology at the forefront of innovations and entrepreneurial efforts. All of these issues are occurring in the context of cloud computing. Clouds enable the developer to capitalize on connection, shared processing, accessibility to an array of cloud service providers, and the ability to develop and deploy applications rapidly. Those applications are deployed in clouds in these ways:

- **Public cloud.** The public cloud infrastructure is provisioned for open use by the general public. It may be owned, managed, or operated by a business, academic, or government organization or a combination of them. It exists on the premises of the cloud provider and is accessible over the Internet. Usage will be monitored by the cloud provider (metered use), and that will drive an economic model that ranges from free to pay as you go.
- **Private cloud.** In a private cloud, infrastructure is provisioned for exclusive use by a single organization comprising multiple consumers (e.g., business units). It may be owned, managed, or operated by the organization, a third party, or a combination of them, and it may exist on- or off-premises.
- **Hybrid cloud.** The hybrid cloud infrastructure is a combination of public and private clouds. It is composed of two or more distinct cloud infrastructures (private or public) that remain unique entities but are bound together by standardized or proprietary technology that enables data and application portability (e.g., cloud bursting for load balancing between clouds).

These three deployment models form the basis of dimension three. There is a fair amount of overlap among these definitions, and

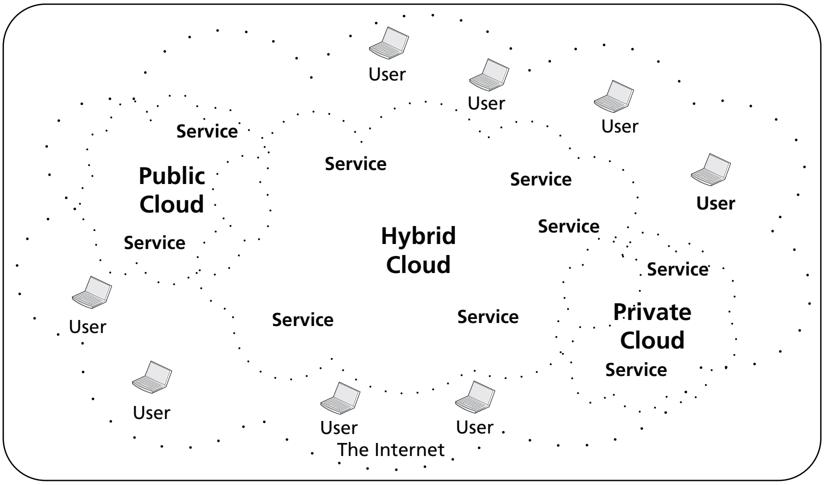


Figure 3.1 Cloud Computing Core Deployment Models

the differences are largely a matter of purpose (deployed for one organization or shared among many) rather than any other substantive variability. The interplay of public, private, and hybrid deployment models are shown in Figure 3.1.

Another type of cloud is the *community cloud*. This is provisioned for exclusive use by a specific community of consumers from organizations that have shared interests or concerns (e.g., mission, security requirements, policy, or compliance considerations). A community cloud is built using public, private, or hybrid clouds. It may be owned, managed, or operated by one or more of the organizations in the community, a third party, or a combination of them, and it may exist on- or off-premises. Public, private, hybrid, and community clouds are discussed in more detail later.

IMPORTANT CONSIDERATIONS

The shift from traditional IT deployment to a cloud-based infrastructure requires substantial thought. To make the switch, the organization's IT department must shift its focus from assembling hardware components and building and installing software to establishing a scalable infrastructure of resources that internal users can consume. The effect this will have on your IT department and other staff must be considered as your organization moves from a traditional LAN infrastructure to a cloud-based infrastructure. Some of the major considerations are as follows:

- **OPEX vs. CAPEX.** From an accounting perspective, when technology is purchased, it is normally capitalized, then expensed across its lifetime. This is referred to as a *capital expense*, or CAPEX. When services are purchased by an organization, they are normally considered an *operational expense*, or OPEX. Moving to the cloud will affect OPEX and CAPEX as well as the organization's balance sheet and income statement. An example of this is covered in Chapter 4, along with other measures of financial effects that must be considered when moving to a cloud-based infrastructure.
- **Personnel readiness in the IT department.** Moving to a cloud-based infrastructure that delivers services to users may require a new skill set in the IT department. This may mean additional IT staff training and staff unavailability for other projects during that training. The existing infrastructure is often bare metal and often does not support the cloud attributes discussed in Chapter 2. (*Bare metal* refers to operating systems being installed directly on the hardware; one instance of the operating system is installed per device [e.g., server or PC]. Compare this to virtualized environments. Moving from a bare metal environment to one in which multiple instances of operating systems, desktops, or applications share the same device may require the development of new skills within the IT department. The cloud approach can deliver economies of scale for your organization, but only with properly trained staff. This topic is more thoroughly covered in Chapter 6.)
- **Infrastructure adequacy.** The current infrastructure of your organization may not be robust enough to handle the increased demand that cloud computing brings. That demand may come in the form of increased utilization of IT resources caused by organizational (or even the IT) staff finding new and innovative ways to use technology, or an increased need for bandwidth. Whatever the reason, the existing infrastructure will undoubtedly have to be enhanced to support the increased demand. With proper implementation, moving to a cloud-based

infrastructure will enhance the distribution of IT services, innovation, time to market, and responsiveness to organizational and customer technology needs.

- **General staff knowledge.** Your organization's staff has probably been accustomed to requesting services from the current IT department. When requesting those services, the staff is also probably accustomed to it taking a long time and requiring substantial justification for new infrastructure. Moving to the cloud changes much of that, and with proper implementation and training, the staff can self-serve new IT services on demand. That change means that the staff can be more agile in its use of technology and potentially respond to customer demand more quickly. This new ability may result in substantial innovation by the staff, but it may also mean that the staff needs training on the new abilities. General staff knowledge may have to be enhanced so the organization can maximally benefit from a cloud computing implementation.
- **Service level considerations.** When using external cloud providers, it is customary to have a service level agreement (SLA). It may be negotiated (i.e., a custom agreement between your organization and the provider that states what happens when a service is not provided as agreed) or standardized (e.g., you just signed up for an account on a cloud provider, such as Amazon Web Services, and you click okay to its terms and conditions). Whatever the agreement, your organization has expectations in order to meet the needs of its customers, but the reality is that your organization has relinquished control to another company—the cloud provider. It is important that your organization have an alternate plan in case the cloud provider does not deliver. This topic is so important that it is covered in detail in Chapter 5.
- **Core business functions versus innovation.** When deploying a cloud infrastructure, an organization should consider two types of implementations: business IT cloud and user IT cloud. This can be by policy or actually separate clouds, but is often built from shared resources (see Figure 3.2).[1] Business IT cloud supports the organization's core business functions, and user

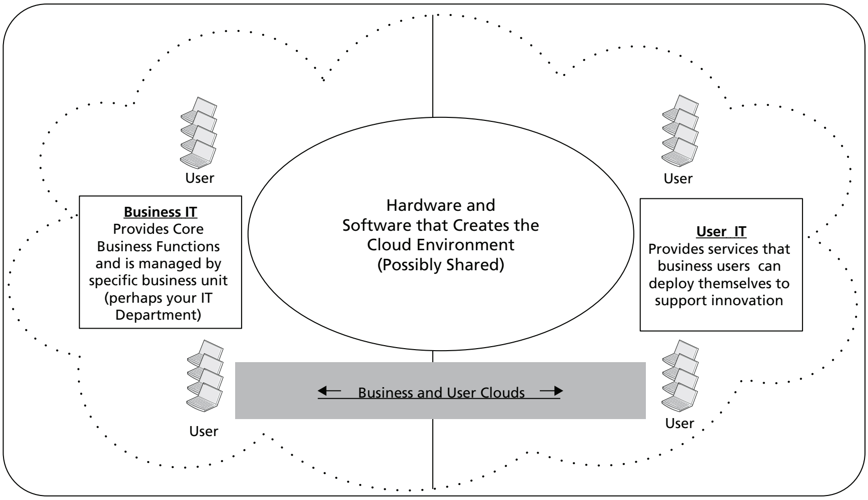


Figure 3.2 Business and User IT Clouds

Table 3.1 Integrated Clouds: The Business Cloud and the User Cloud

Business Cloud	User Cloud
Supports core business functions	Supports innovation and rapid deployment of applications
Is maintained by business unit to ensure that services are available to support core business functions	Is maintained by business unit to ensure that services are available to support innovation by users and rapid deployment of new applications
Is under business unit control	Is under user control

Based on http://assets.cio.com/documents/cache/pdfs/privatecloud.pdf.

IT cloud provides cloud services to users that they can control and deploy IT as needed (see Table 3.1). These can be divided as follows:

- **Business IT cloud—core business functions.** There should be a business unit within your organization (e.g., the IT department) that is responsible for the organization's core IT business services. The monitoring and support for these services will occur with the responsible business unit. The necessary services may include the organization's web site, an enterprise resource planning system, or other software to support the organization's core business functions.
- **User IT cloud—where innovation occurs.** The other type of implementation that should be available will support users in their creative processes and innovation. This cloud will be maintained by a business unit within the organization (probably the IT department) but should be under the user's control. It may take a more sophisticated user than the traditional office worker to utilize these cloud services, but the user cloud should be in place to support increased innovation. This is a major consideration when hiring new employees; their ability to innovate using available technologies within your organization should be considered, for both technical and nontechnical positions.

PUBLIC AND PRIVATE CLOUDS

Public and private clouds form the basis for cloud computing. Hybrid and community clouds (discussed later in this chapter) are built on private or public clouds. Public and private clouds are technically identical, except that the public cloud is provisioned for multiple organizations to share infrastructure, whereas a private cloud is deployed for one organization. Both types of clouds share the same essential characteristics and service models. A major difference is that unless there is a charge-back arrangement in your organization, private clouds do not normally enforce pay-as-you-go billing.

Clouds are accessed much like a utility (e.g., electricity or telephone). Consider what happens when placing a phone call outside your organization: You typically need to dial a number (e.g., 9) to access the public phone utility, and then you dial the phone number of the person you are trying to reach. Compare that with placing a phone call within your own organization: You dial an extension and are connected to a coworker using your own private phone system. Both connections allow a conversation, but one uses a public utility and the other uses a private utility.

Cloud computing is much like the phone system example. Accessing the public phone utility is much like the public cloud, and accessing the private phone system is much like the private cloud. Users can consume services in the public cloud and be billed for what they use, or they can operate a private cloud that offers services to its business units. Where they are located (on-premises or off-premises) is irrelevant.

Although this may be redundant, it is important that we summarize the differences between public and private clouds. Public clouds operate in a service delivery fashion like private clouds, but there are a few big differences:

- The public cloud will be a shared resource in which multiple users deploy their applications on the same infrastructure (i.e., multitenant). The applications are separate (i.e., secure from one another), but they share the same infrastructure, promoting economies of scale.

- The public cloud will have metered use and payment by an economic model tied to that metered use. The economic model extends from "freeconomics" to premium. The private cloud will have metered use so that utilization can be monitored and infrastructure adjusted to support that utilization, but it will not normally have an associated payment model.
- Location is irrelevant. It is possible to have a private cloud that is located on someone else's infrastructure and have a dedicated carrier circuit (e.g., a leased line) to that private cloud.

Amazon Web Services is a good example of a public cloud. It offers IaaS that allows the users to consume services-on-demand like storage, database, or computing. The consumer is billed based on the amount of resources consumed. Amazon's economic model requires the users to pay for all the services they consume after a brief free period.

Sharing infrastructure gives economies of scale that allow cloud providers to charge less for the same usage than what a bare metal approach might cost. Normally, you can see dramatic changes to CAPEX and OPEX when moving an application from a non-cloud infrastructure to a public cloud. This might result in it being less expensive (and therefore less risky) for an organization to deploy new applications, and it therefore might increase innovation.

EXAMPLES OF CLOUD COMPUTING

Public and private clouds provide all of the service models: IaaS, PaaS, and SaaS. Following are examples of organizations that have deployed each.

An IaaS Example: Netflix Moves to Amazon Web Services

IaaS provides a set of services on which users can run their operating systems, development environments, and applications. It can be a public or private cloud. What defines it is its ability to provide services that can be configured to develop or deploy applications.

Netflix was founded in 1997 to provide movies on DVD through the regular mail to customers' homes. The customers are allowed to use Netflix's web site to browse the selection of movies and set up a personal queue for the movies they desire to have shipped to them. The movies in a customer's personal queue are automatically shipped according to a schedule and corresponding fee chosen by the customer.

In January 2007, Netflix began offering subscribers the option of instantly watching movies on their PC. According to a press release by Netflix, "The introduction marks an important enhancement to the Netflix service. Subscribers will continue to receive DVDs by mail from the company's market-leading catalog of over 70,000 titles and will have the additional option of instantly watching about 1,000 movies and TV series on their PCs."[2] Other press releases announced that Netflix streaming services would be available on the Xbox 360 (November 2008), Sony Bravia TV and Blu-Ray Disc Player (November 2009), Playstation3 (November 2009), the iPad (April 2010), and the iPhone and iPod (August 2010). As device manufacturers adopted the Netflix streaming service, Netflix saw the need for a more agile and resilient infrastructure.

Netflix's move to cloud computing was largely motivated by Netflix wanting to do what it does best, which it realized was not running data centers. Yuri Izrailevsky, the vice president of Cloud and Platform Engineering for Netflix, put it this way: Netflix must be good at many things, including content, personalization, device penetration, customer service, and user interfaces. Netflix understood that operating a data center is not something it had to be good at. Izrailevsky added that "in fact, [running data centers is] a distraction for us."[3]

Some of the other attributes of cloud computing that made it attractive for Netflix are the following:

- **Fault tolerance.** Netflix realized that it needed a more fault-tolerant system than it currently owned. It saw Amazon Web Services as a clear leader in the cloud space and believed that Amazon had the capacity to support Netflix now and as it grew.
- **Increased innovation.** In 2009, some of the Netflix software moved to the Amazon cloud, and in 2010 the streaming

services moved there. This allowed Netflix to be innovative while attempting development in the cloud environment before making a full commitment to it.

- **Metered use, economics, and scalability.** A critical component of the decision to move to the Amazon cloud was the scalability it offered: As the demand for resources changed (i.e., the demand for streaming movies, causing more or less infrastructure to be needed), the Amazon cloud could respond to those changes. Netflix was responsible only for the marginal cost of consumption and did not need to deploy or pay for infrastructure that might not be used.
- **Focus on core competencies.** Netflix could concentrate on its core competencies and leave the running of the ICT infrastructure, including data centers, to the experts in that area, Amazon.

The balance between CAPEX and OPEX was affected by the Netflix move to the Amazon cloud. Netflix was left with application development on an IaaS platform, but it did not have to manage the data centers that supported those applications. The move from an on-premises infrastructure to a public cloud seems to be serving Netflix well. The Amazon Web Services cloud provides multiple services that allow Netflix to concentrate on its core competencies rather than managing data centers.

A PaaS Example: Sopima Moves to Microsoft Azure

PaaS provides a development and operating environment built around a specific operating platform. It can be a public or private cloud. What defines it is its ability to provide a unified development or deployment platform that can be configured to meet the user's needs.

There are many PaaS providers. This example is based on Microsoft Azure. Here are some of the major considerations when deciding whether to move to an external PaaS platform:

- The technologies supported by the PaaS provider.
- How a PaaS implementation will affect time to market and innovation.

- Whether developing an application totally in-house or using cloud services makes sense (e.g., in terms of economics, staff, or training).
- The effect on CAPEX and OPEX.
- The type, sophistication, and service level required to support the application, and whether the PaaS provider can meet those needs.

Along with these considerations, the decision must be driven by organizational needs and the needs of the customers.

Sopima, a Finnish software company, used Azure to create an online contract bank. Sopima provides online cloud-based contract management solutions throughout the contract life cycle. Sopima's support for the contract life cycle includes business development, administrative, production, and customer contracts. Sopima provides the following services:

- Drafting the initial contract
- Handling the negotiation of contract terms
- Managing the acceptance and signing of contract
- Handling the initial execution and follow-up
- Providing ongoing analysis, reporting, and alerts
- Managing renewal and termination of contracts

Sopima wanted to provide global online-accessible contract management tools but did not have an IT staff or infrastructure to provide this service, so it turned to the Microsoft Azure platform. Markus Mikola, a partner in Sopima, has stated that "as a small company in a time of recession, we [Sopima] have built a global solution with minimal investment. If we hadn't used the Windows Azure platform, we wouldn't have been able to launch this service [Sopima's contract solution] at all."[4] In other words, Sopima might not have been able to surmount the barriers to entry (e.g., the necessary talent or the cost of building the necessary infrastructure). Those barriers might have prevented Sopima from building and launching a globally accessible service.

Sopima evaluated a number of cloud service vendors and in March 2009 selected the Microsoft Windows Azure platform. Sopima did not want to administer individual parts of service-based products purchased from different vendors but rather wanted a PaaS approach that moved all ICT administration to a single vendor on a pay-as-you-go basis. Sopima could have used an IaaS approach and implemented in the Amazon Web Services cloud or a SaaS approach like Google Apps, but in Sopima's opinion, "neither provided the full-scale platform and service management capabilities that the company [Sopima] was looking for."[5] After analyzing its needs and the services offered by various cloud vendors, Sopima chose a PaaS approach and decided to develop its application on Microsoft Azure.

Sopima uses Microsoft SQL Azure, a cloud-based relational database platform built on Microsoft SQL server technologies, to manage contracts and contract information. Using SQL Azure, Sopima stores all deadline and project milestone reminders for users involved in the contract creation process. As part of the Windows Azure platform, SQL Azure provides automated management capabilities, including built-in data protection, self-healing, and disaster recovery—features that Sopima's management found crucial for safeguarding customers' data. "SQL Azure relieves Sopima of the database management aspect, which would otherwise require a huge investment of resources. Antti Makkonen, research and development leader at Sopima, estimates that they would have had to spend approximately $20,000 annually in server licensing and pay several support salaries amounting to roughly $300,000 in savings."[6] The selected PaaS approach thus saved Sopima substantial money.

A SaaS Example: Dunkin' Donuts Uses Salesforce

SaaS provides the greatest amount of abstraction for the user because the applications are exposed to the user through a web site or a set of application program interfaces (API). The user can configure and combine the services to meet business needs but cannot directly deploy applications on the SaaS provider infrastructure.

In its most direct form, a SaaS application is delivered to the user via a web browser. There are many examples of SaaS applications—e-mail, social networking, picture sharing, and even business applications like customer relationship management—that can be consumed with a normal web browser. SaaS services were discussed in Chapter 2. Our example here reviews the American coffee and doughnut chain Dunkin' Donuts and its SaaS usage.

Dunkin' Donuts has a Facebook fan base that is roughly the size of Hong Kong. It uses social media like Facebook and Twitter to connect to its customers and to become part of their everyday lives. John Costello, the chief marketing officer of Dunkin' Donuts, explained that "being social is not just about winning fans and influencing followers. It's about sharing content that inspires our customers and their friends, giving them the opportunity to engage with us."[7] Dunkin' Donuts uses social media and Salesforce, both cloud services, to manage customer information and become part of social network conversations about its products. Being involved in these conversations helps Dunkin' Donuts build brand loyalty.

With an estimated 7 million followers subscribed on Facebook alone, Dunkin' Donuts decided to utilize Salesforce to provide a way to manage its customer relationships. Dunkin' Donuts uses the following Salesforce products:

- **Sales cloud.** This is a tool kit that allows an organization to manage social accounts like Facebook, Twitter, YouTube, and LinkedIn within the Salesforce service. Using this tool, the organization can manage information from fans who follow it on its social sites as well as manage the information the organization drives to social media sites.
- **Marketing cloud.** This service aids the organization in its social marketing and provides Radian6, a social listening platform.
- **Salesforce platform.** This platform brings together the various services of Salesforce so the organization can build applications, share and store data, and deliver web sites quickly.

Dunkin' Donuts uses the SaaS offerings of Salesforce to support its unified social media approach.

PRIVATE CLOUDS IN CONTEXT

For many organizations, the journey into cloud computing begins with a private cloud. That infrastructure provides a way for an organization to standardize its business processes and deliver them as a service to their users. It also provides a new way to manage infrastructure so that resources are pooled, allowing resource allocation in a matter of hours rather than days or weeks and making the delivery of information systems better able to meet changing business needs. This shifts the focus from build-to-order solutions to the supply of applications and service on demand.

For other organizations, the foray into cloud computing begins with a public cloud deployment, which can allow an organization to test cloud computing without making a substantial investment in infrastructure. Whichever is chosen, the benefits of cloud computing can be realized.

Using a private cloud computing implementation, let's discuss private cloud computing benefits. A private cloud can be built in-house, or it can be hosted by a provider; location is irrelevant. What defines it as a cloud is that it has some combination of dimension one, essential characteristics, and dimension two, service models. The defining characteristic of a private cloud is that it is deployed and intended to be used by one organization.

Resource allocation (i.e., based on metered usage and the infrastructure automatically adjusting resources to changes in demand) is paramount to defining a cloud infrastructure. Both public and private clouds will meter use and adjust resources; what is different is that in the public cloud, resource changes are billed to the customer, whereas in the private cloud, billing based on resource consumption is unusual.

In 2009, *CIO* magazine cited RADLab of the University of California at Berkley as saying that on-premises infrastructure configured as a cloud cannot really be called cloud computing because it cannot provide unlimited resources, "since there is inevitably some point at which additional resources are not available, as when an individual company runs out of room in its data center."[8] The ability to run out of resources exists in a private cloud as well as a public cloud; no organization has unlimited resources. Resource usage must be monitored, and additional resources must be added to the existing resource pools to ensure that the expected user demand can be met.

Puma and Eucalyptus: A Private Cloud Solution

Eucalyptus is a private cloud solution that was adopted by Puma shoes to deliver its web sites and online marketing campaigns. Jay Basnight, the head of digital strategy at Puma, stated that "Puma .com must be [as] reliable as our shoes."[9] Puma selected a cloud solution from Eucalyptus Systems to provide an on-premises infrastructure that would meet its business needs. Because of the retail nature of Puma, it required a solution that would maximize response time for consumers; Puma realized that slow response time at its web site meant that consumers would abandon the Puma web site in favor of sites that were more responsive.

Puma decided that a private cloud, with dynamic resource allocation, would meet the need for an agile, cost-effective, extensible, and manageable IT infrastructure. This was especially true considering the increased demand for collaborative web sites and because of Puma's increased digital presence. Puma needed to roll out web-based applications quickly to support these business demands, and a private cloud was the chosen solution.

Before moving to a private cloud, Puma's in-house IT infrastructure included various web environments and a number of dedicated servers that created a complex hosting environment. That environment was composed of multiple operating systems, application program interfaces, and applications. "The result was a graveyard of apps and sites, some with administrative control and some without," said Basnight. "Adding to the complexity, the relatively small group supporting interactive projects had to be ready at a moment's notice for events that could require significant IT support."[10] He added that Puma needs dynamic content management and IaaS. These things motivated the move to an on-site private cloud–based solution that automatically provisioned resources, scaled to meet the traffic requirements at Puma's web sites and allowing for more rapid deployment of marketing campaigns delivered through those web sites.

Today, Puma's infrastructure is a scalable and agile private cloud that meets the company's needs for rapid deployment of new applications and provides suitable response time to visitors at the web site. As you can see from the Puma private cloud example, there is little distinction between a public and private cloud implementation.

HYBRID CLOUDS

Hybrid clouds are built using both public and private cloud deployments. With a hybrid arrangement, an organization has decided that its requirements will be best met by deploying both private and public clouds to support its business needs. Perhaps a private cloud is used for application development, but the actual deployment occurs on a public cloud. Or an organization could run its web site on a private cloud, but when that private cloud cannot handle the increase in demand (i.e., too many simultaneous visitors to the web site), the users can be automatically rerouted to the same web site on a public cloud. Sometimes organizations start as a pure public cloud but then decide that a hybrid cloud will serve its needs better, as in the case of Zynga, the company discussed next.

Zynga: A Hybrid Cloud Example

The editors of Wikibon, a professional community that attempts to solve business and technology problems through open-source sharing of free advisory ideas, have stated that Zynga, a company that creates and operates social Internet games (e.g., Farmville or Words with Friends), is the "poster child for a company that moved from a pure public cloud model, using AWS [Amazon Web Services], to a hybrid model, choosing to complement Amazon with its own internal data center."[11] These statistics show the substantial scale of Zynga's operation:

- Zynga has more than 250 million active users globally.
- Five new players join Zynga every second.
- About 10 percent of the world's Internet population plays Zynga games every month.
- One in five Americans plays social games.

Zynga started operating with a dedicated hosted IT infrastructure, but its social games soon outpaced Zynga's ability to add infrastructure fast enough to support user demand. Zynga found it necessary to move operations to Amazon Web Services. According to the Wikibon case study cited earlier, Zynga moved to Amazon "to accommodate the rapid scale demands in the vertical growth phase of the company." That vertical growth phase is the steep part of the traditional S-curve (i.e., the S-curve is a graphic tool that describes how technology payback changes as it matures) used to measure technology deployment, adoption, and timing. For Zynga, this steep part came early in the deployment of its products. Zynga is deployed on Amazon because Amazon could provide the technology to keep up with user demand. Amazon Web Services specializes in building cloud infrastructure, whereas Zynga needed to concentrate on building games and fulfilling user needs. Early in Zynga's history, this was a perfect match, with a balance between user demand and available IT infrastructure.

After its initial growth stage, Zynga began to level out. Once the Zynga staff had a better understanding of user demand, the company decided to build a private cloud to complement its Amazon cloud deployment. With a better understanding of the organization's IT needs and with more available financial resources, there was impetus to move from a public cloud to a hybrid public and private model to change the mix of CAPEX and OPEX. This presents an interesting business life-cycle question: Should the mix of public and private cloud utilization change as the company matures? This is explored in Chapter 4.

This leaves another question: How does the Zynga hybrid cloud deliver success for the company and its customers? The answers can be found in a *Networkworld* interview with Allan Leinwand, the chief technical officer for infrastructure at Zynga. He stated the following:

- The Zynga hybrid cloud "works by combining the capacity of Amazon Web Service's public cloud with the company's custom-built private cloud."
- Amazon handles spike demand when user requirement exceeds Zynga's internal capacity.
- The private cloud is optimized for Zynga's business, and the public portion is more of a generic, scalable, and homogenous infrastructure.
- The public and private portions are treated as one hybrid cloud and managed through the same software dashboard (i.e., a "single pane of glass" interface).

- Interoperability at Zynga is handled through the private cloud having similar equipment as the public cloud (technically, they use the same hypervisor technologies). In addition, virtual machines are built so they can move between the private and public clouds using the common software dashboard.[12]

Zynga's move from a pure public cloud implementation to hybrid one teaches us many lessons. Among them is that despite the increased innovation a public cloud can bring early in the business life cycle, there may come a time when an organization and its customers can be best served by the organization building its own infrastructure.

COMMUNITY CLOUDS

Community clouds are built around the other cloud deployment models (i.e., public, private, and hybrid) and represent affinity groups. Community clouds provide ways for related groups of people or organizations to share technologies and data. The organizations have a common concern related to a specific issue: information sharing, regulatory compliance, or something else. Brandon Butler, an author for *Networkworld*, says, "Fundamentally, community clouds are not necessarily a new technology . . . rather, they're a different way of congregating users under an umbrella of services."[13]

Here are some examples of community clouds:

- **Information sharing.** The "Chesapeake Regional Information System for our Patients" allows Maryland's physicians and nurses to share information about their patients.
- **Gaming industry.** IGT delivers a cloud service to game providers globally that provides a seamless experience on an individual's land-based, mobile, and online devices.
- **Regulatory compliance.** In the United States, the Health Insurance Portability and Accountability Act regulates the sharing of health-care information along with a number of other things. Optum Health Cloud provides cloud services to the health-care industry with an eye on strict compliance with the act.

- **Government.** According to the federal government, "The Interior Business Center (formerly the National Business Center) is a federal shared services provider that offers business solutions to create efficiencies and economies of scale for the Department of the Interior as well as other federal agencies."[14]
- **Financial services industry.** SAVVIS (www.savvis.com/financial-services) provides cloud-based IaaS offerings to the financial services industry.

Community clouds present opportunities for organizations to cooperate with one another through resource sharing while potentially also competing in the marketplace. This requires a large amount of trust and organizational acceptance. Building that trust and getting that acceptance is perhaps the most difficult part of creating community clouds.

CONCLUSION

Cloud computing has a number of characteristics, some required and some optional. It also has a number of service models, which were covered in Chapter 2 and used in this chapter to discuss cloud computing deployment models.

Cloud computing is built upon public and private clouds. The major difference between the public and private cloud is who the cloud is deployed for; the private cloud is deployed for a single organization, whereas the public cloud is built to be shared by multiple organizations. The hybrid cloud is built using both public and private clouds and is most often deployed to manage excess utilization of private cloud resources. The community cloud is deployed to serve the needs of a specific group of organizations that share a common interest or requirement, such as compliance or common industry needs.

Cloud computing deployment models offer organizations the opportunity to implement technologies, increase innovation, decrease time to market, and thus gain market share. These opportunities are offered in an environment in which the balance between CAPEX and OPEX is tilted in favor of an organization's need to support a greater IT value proposition.

NOTES

1. The notion of two major focuses of a private cloud was introduced and developed by Bernard Golden through a series of blog posts and articles (2009–2013) in *CIO*, an industry trade journal. That idea has been expanded in this book.
2. Netflix press release, January 16, 2007, https://signup.netflix.com/MediaCenter/Press.
3. *AWS Case Study: Netflix* [video], Amazon Web Services, http://aws.amazon.com/solutions/case-studies.
4. Microsoft, "Sopima: Software Firm Launches Business Contract Service with Lean Staff, Low Investment," November 17, 2009, www.microsoft.com/casestudies/Case_Study_Detail.aspx?CaseStudyID=4000005881.
5. Ibid.
6. Ibid.
7. "America Runs on Dunkin'. Dunkin Runs on 7+ Million," Dunkin' Donuts, www.salesforce.com/customers/stories/dunkin.jsp.
8. Bernard Golden, "Defining Private Clouds," *CIO*, May 14, 2009, www.cio.com/article/492695/Defining_Private_Clouds_Part_One. RADLab was a demonstration project hosted at the University of California at Berkley that ended in 2001. For information on this project, see http://radlab.cs.berkeley.edu.
9. Eucalyptus Systems, "Case Study of Puma," http://go.eucalyptus.com. Eucalyptus sells private and hybrid cloud solutions that are compatible with Amazon Web Services.
10. Ibid.
11. David Cahill, "Zynga Cloud Case Study: The Journey to a Real Private Cloud," Wikibon, February 12, 2013, http://wikibon.org/wiki/v/Zynga_Cloud_Case_Study:_The_journey_to_a_real_private_cloud.
12. Ibid.
13. Brandon Butler, "Are Community Cloud Services the Next Hot Thing," *Networkworld*, March 1, 2012, www.networkworld.com/news/2012/030112-are-community-cloud-services-the-256869.html.
14. U.S. Department of the Interior, "Interior Business Center," www.doi.gov/ibc/index.cfm.

PART TWO

Cloud Economics

Part One introduced concepts related to cloud computing. In this part we now begin the discussion of cloud economics and strategic measurement of cloud computing implementations. Measurement of cloud computing implementations is necessary for at least two reasons: (1) all types of implementations, including business intelligence implementations, need to be justified before scarce IT resources are expended, and (2) once a cloud strategy is adopted and an infrastructure is implemented, the implementation must be continually monitored so the organization can be sure that it continues to deliver an optimal return on investment. Gaining maximal return on the implementation of a cloud computing strategy is predicated on the ability to understand these metrics. The introductory concepts found in Part One are necessary so we have a common set of terms and language to support our discussion of strategic measurement.

Strategic measurement refers to establishing the metrics necessary to set, accomplish, and monitor a strategic orientation. These measurements are as follows:

- Understanding the financial effect of implementing cloud infrastructure (including business intelligence implementations).

- Knowing how total cost of ownership (or operation), return on investment, and CAPEX/OPEX change when discussed in the context of cloud computing.
- Measuring the performance and effects of cloud computing.
- Understanding the effects of cloud computing on stakeholders and how to measure quality of services provided by cloud suppliers (service level agreements).
- Determining current technology and staff readiness for the move to cloud computing.

These types of metrics, which will be covered in this section, are necessary for an organization's C-suite to understand so that cloud computing deployments support enhanced business intelligence. These types of metrics can also be used during cloud implementations to ensure that an implementation is on track, enabling alignment with business and technology strategies so that economic benefit can be maximized.

CHAPTER 4

Strategic Measurement

TCO, ROI, OPEX/CAPEX

Organizations are bombarded with large amounts of data and need a way to convert them into actionable information to aid in decision making. That process is referred to as creating *business intelligence* (BI). Whether it is a decision to build a new parking lot or a decision to move the organization's application to the cloud, it requires actionable information.

Commitment to a cloud strategy to support BI requires that a large amount of information be considered before the commitment of resources can be made. In this chapter, concepts and measures that provide effective ways to organize and review the multitude of data and information are discussed. Each of the concepts and indicators should be considered when determining whether a cloud infrastructure is the best choice. Specifically, we will examine these primary financial concepts:

- **Total cost of ownership (TCO).** This is a financial estimate of the direct and indirect costs of owning something like cloud infrastructure. Understanding the TCO of existing infrastructure aids in knowing how businesses benefit from that infrastructure. It is also very useful when comparing various options being considered for funding. In the cloud space, we should consider the total cost of ownership as well as another, newer meaning of TCO: total cost of operation.

- **Capital expense (CAPEX) and operational expense (OPEX).** CAPEX refers to the way in which large purchases with an expected life of greater than one year are handled from an accounting perspective. The goal is to align the use of that large purchase with the income it generates. OPEX refers to items that are purchased by an organization and consumed during the current year.
- **Return on investment (ROI).** This measure shows how well something performed financially. It is especially useful when comparing options that each has an estimated ROI.

A *cloud computing strategy* is defined as the use of technologies with the essential characteristics, service models, and deployment models discussed in Part One. A cloud strategy changes the job of the IT department from building solutions to providing services. It also promotes innovation, since the time that solution development takes and its cost can change dramatically when applications are moved to the cloud. This is especially true when an organization uses external service providers instead of building applications in-house.

TOTAL COST OF OWNERSHIP

Total cost of ownership (TCO) refers to determining all the costs involved in owning something. The costs will be specific to the project being considered and will vary among projects. A useful (non-ICT) TCO example is purchasing a car. This example will help explain what is meant by the total cost of owning something. At a minimum, the TCO for a car purchase should include the purchase of the car, the cost of insurance and maintenance, the probable cost of an accident (e.g., the rental of a car while yours is being repaired as well as the insurance deductible), and the eventual sale (or amount lost through depreciation) of the car. Together these constitute the TCO. Similar estimates must be made for the TCO of an ICT infrastructure investment.

It is possible to take this TCO example and morph it into the total cost of operation—the new TCO. The reality is that some cars are cheaper to own and run than others. Consider owning a high-efficiency hybrid car versus a high-end luxury car. The cost of fuel is cheaper for the hybrid car, and the high-end car may require specialized maintenance. These factors make the high-end car more expensive both to own and to operate. The notion of total cost of operation extends the basic TCO model to include operating costs. Throughout this book, TCO is assumed to be the total cost of both ownership and operation.

In any BI cloud implementation, TCO must be a financial estimate that includes the direct and indirect costs related to the purchase of any capital item and the ongoing operation of that item. These consist of the following:

- **Direct costs.** When ICT is purchased, direct costs include the purchase of the technology, installation, and maintenance. Depending on the implementation, direct costs may also include testing, the cost of facilities, and environmental (e.g., air-conditioning) and power conditioning.
- **Indirect costs.** Indirect costs are not directly associated with the purchase of the asset but are real costs that are necessary for the implementation to be successful. These costs must be included in any financial analysis of technology purchases. These indirect costs include training, an outage (either planned or unplanned), a security breach (e.g., physical breach, virus or malware breach), disaster preparedness and recovery, other development expenses, and eventual decommissioning.
- **Operating costs.** The cost of operating a BI cloud-based infrastructure must include the operational costs of purchasing services from cloud vendors. If your organization requires additional staff because of a cloud implementation, you should include that in the operation cost as well.

Table 4.1 lists basic examples of things to be considered during a TCO analysis. It is often easy to determine direct and operational costs; it is much harder to determine indirect costs. However, it is essential that any necessary and probable indirect costs be included in all financial estimates.

Table 4.1 Total Cost of Ownership Examples

Direct Costs	Indirect Costs	Operating Costs
■ Purchase price ■ Maintenance ■ Installation ■ Upgrades	■ Training ■ Outage ■ Breach ■ Disaster preparedness ■ Disaster recovery ■ Floor space ■ Testing ■ Development ■ Decommissioning	■ Additional staff needed to support cloud implementation ■ Cost of services from cloud service and data providers

How TCO Changes in the Cloud

TCO changes when considering cloud infrastructure. The major changes to consider are listed here. There may be other considerations based on your organization's needs, but this is a good starting place:

- **Cloud infrastructure should enable increased innovation.** Increased innovation may mean an increase in demand for ICT infrastructure and thus an increase in the overall resources necessary to support the increased demand. These increases lead to extra personnel costs to support the infrastructure and/or to increased technology costs.
- **A greater telecommuting experience might be enabled by cloud infrastructure because of the accessibility of applications.** A cloud implementation can mean that anywhere there is an Internet connection and a suitable device (e.g., tablet, notebook, or smartphone), an employee can work. This too can lead to a need for greater resources. It can also mean that a different management culture has to be adopted in your organization.
- **Software licensing, hardware, and facility costs change when you move to a cloud infrastructure.** The cost of

these items will be lower initially, but TCO may be higher for the cloud implementation over time. The benefits of a cloud implementation should be factored into the decision to move to a cloud-based infrastructure so the new ROI can be realized.

Build versus Buy

TCO changes in different scenarios. An organization may be interested in the TCO of building an in-house application versus buying the necessary services from external cloud providers. When infrastructure and software are built in-house, there are extensive development costs, time is lost to the development effort, there is a need for internal testing, and so on. When services are purchased, you may save many of these costs but gain the cost of integrating the external services with your internal processes and infrastructure. The decision to build or buy may be determined by the time value of having a working system now from a cloud provider versus having to wait for an internal development team to build the system. This example oversimplifies the build-versus-buy decision, but it does point out some of the differences.

TCO is important and useful when comparing options, because some options cost more than others over their lifetime. If we had three options, each with different ownership and operating costs, we could have three very different multiyear TCOs. Although TCO is only one of the financial measures that must be considered, it is very important. Some of the other major considerations are the ROI, time value of money, source of capital investment resources, effect on the company's financial statements, tax implications, internal rate of return, and net present value.

It is important that all financial measures take the TCO perspective and not just analyze the start-up costs. Using TCO will give your organization a reasonable picture of how infrastructure purchases will affect the company's financial position over time. You must look at not just the first-year costs but also the cost of operating the solution over its expected useful life.

FINANCIAL MEASUREMENTS

Taking the TCO perspective just discussed, we will now discuss two major financial measures that must be considered when adopting a cloud strategy: (1) operating expense (OPEX) versus capital expense (CAPEX) and (2) return on investment (ROI). If you use a total cost of ownership/operation perspective with the CAPEX/OPEX and ROI measures, you will have sufficient background to understand measures not discussed in this chapter. CAPEX/OPEX and ROI will be discussed next. We will use two scenarios that will help explain these measurements and how a cloud computing strategy affects the organization.

- **Scenario 1.** An organization purchases $200,000 in ICT and the necessary applications; both have an expected life of five years. If you were using straight-line depreciation for the infrastructure and application purchase, the expenditure would be $40,000 per year, reducing the organization's net income by that amount annually. This should match the expenditure to any revenue generated during the same period. Assume that there is $15,000 in start-up costs in year one, a yearly maintenance expenditure of $36,000, and $175,000 per year in personal costs to maintain the equipment (personal and maintenance are trended at 5 percent per year).
- **Scenario 2.** If we move IS to an external cloud vendor, the expenses will differ. Assume that in order to develop the application, we would need initial ICT that costs $30,000, incur setup up costs of $20,000, and have personnel costs of $75,000, ICT maintenance costs of $5,400, and cloud computing service costs of $110,000 (personnel, maintenance, and cloud services are trended at 5 percent per year).

Capital Expense (CAPEX) versus Operational Expense (OPEX)

A major consideration when deciding to move IS to an external cloud provider is the trade-off that occurs between capital expense (CAPEX) and operational expense (OPEX). When operating an on-premises cloud, the same dramatic shift from CAPEX to OPEX may not be seen. The CAPEX/OPEX change is situational and must be computed for each scenario under consideration.

When large purchases like IT infrastructure are acquired, they are traditionally treated as CAPEX. This means that they are recorded as an asset on the balance sheet of the organization and depreciated over time. Conversely, when an organization is spending money to operate its business, it is normally expensed in the year that it is actually spent. Buying infrastructure for in-house installation is CAPEX, and buying services from a cloud vendor is OPEX.

There is a trade-off between CAPEX and OPEX that occurs when moving applications to an external cloud. This can be looked at as two models of funding an ICT implementation, each with different effects on the organization's balance sheet and income statements:

- In the CAPEX funding model, the organization needs cash (this can be from its own funds or borrowed) to purchase capital assets. A long-term commitment to that asset is created because the organization owns it. When the organization is done with the depreciated asset, it must decide what to do with it. During the useful life of the asset, the organization's staff is responsible for managing it. This creates several financial requirements (some would say disadvantages):
  - The cost of buying, maintaining, and disposing of the asset.
  - The cost of staff to run the asset.
  - A determination of how to keep the asset generating value because the organization has an investment in it.

- In the OPEX funding model, there is no long-term commitment, and start-up cash will be substantially less. Your organization pays a fee to use the asset of an external cloud provider that owns it. This fee is normally based on a metered use basis (see Part One). It is up to the external cloud provider that owns the asset to determine how to keep it generating value. When the organization is done using the asset, it can just walk away from it and not be concerned with disposal. There is a shift of responsibility and staffing in a purely OPEX model. The organization's staff becomes concerned with managing applications; the cloud service provider is concerned with managing the IT asset. This relationship creates the following financial changes:

- The asset is owned by someone else, and the organization can pay for what it consumes (i.e., through metered use) rather than making a long-term commitment to the asset.
- The organization does not need staff to run the infrastructure, but instead can be concerned with operating applications.
- The upfront capital needed to start a venture will be lower, but the TCO over the life of the project will be higher. This must be analyzed for each project to determine what makes sense. You should consider that the lower upfront cost could be a major driver of entrepreneurship and innovation for your organization.

The value of not having to make long-term financial commitments is important but can cost more than owning an asset. A short-term rental car is a good example. Suppose you want to rent a car for a short period while on vacation. It is generally understood that the price paid to rent that car far exceeds what you could purchase it for, but the benefit of having that car for a short time motivates you to pay a higher price. Consider how often you could go on vacation if every time you traveled you needed to purchase a car at your destination rather than rent it. That oversimplifies the CAPEX/OPEX discussion, but there are a number of useful parallels. In the case of car rental and consuming cloud services, both are cheaper than purchasing the asset each time you need it, but over time the rental of the asset is usually more expensive than the purchase of it.

When an organization rents an application from an external cloud provider, there is a shift that occurs between CAPEX and OPEX. Few implementations are purely one or the other. For example, you do not sell the car you own so you can go on vacation and rent one. The organization that relies heavily on external cloud providers will probably still have some (albeit smaller) internal infrastructure that has to be managed. Nevertheless, a cost shift occurs that affects an organization's IT strategy and financial position.

Generally, an analysis of CAPEX and OPEX is required when deciding whether to use an external cloud vendor on a metered use basis or build a solution in-house. An external cloud vendor's services are normally able to be expensed in the current accounting period, whereas infrastructure that is purchased is normally capitalized.[1] From a financial perspective, analyzing ICT expenditures to determine whether building an in-house solution or consuming the services of a cloud vendor is better for your organization requires that you look at the effects of those decisions on your organization's balance sheet, income statement, and tax situation. According to generally accepted accounting principles, when a business incurs an expenditure, it should be allocated over the entire period in which the expense will benefit the company.[2] It will be helpful to look at an example. The two scenarios given earlier are shown in Figure 4.1.

The example in Figure 4.1 shows the change is CAPEX and OPEX and how this affects the organization's profit-and-loss statement. In this example, we see the following changes when moving to an external cloud infrastructure:

- CAPEX and OPEX are very different in the two scenarios. This will affect both the organization's balance sheet and its income statements.
- The cash needs at start-up and throughout the life of the infrastructure (five years) are different. The external cloud infrastructure will cost less at start-up, but the five-year expenses will be more.

These changes can drive innovation and make the move to cloud services very attractive financially. In addition, as the OPEX and CAPEX emphasis changes, so does the accumulation of assets on the company's balance sheet. In scenario 2, the overall expenses will change, most things will be expensed, and few assets will be accumulated, potentially making the balance sheet weaker.

The business life cycle shown in Figure 4.2 refers to what occurs to a business as it progresses from start-up through the growth and

Scenario One: In-House Infrastructure Budget						
		Year				
	Start-Up Costs	One	Two	Three	Four	Five
Capital Expenses						
Initial In-House ICT	150,000	30,000	30,000	30,000	30,000	30,000
In-House Setup	65,000	13,000	13,000	13,000	13,000	13,000
Annual Capital Expense (CAPEX)		**43,000**	**43,000**	**43,000**	**43,000**	**43,000**
Total 5-Year CAPEX						**215,000**
Operating Expenses						
In-House Personnel		160,000	168,000	176,400	185,220	194,481
In-House ICT Maintenance		59,500	62,475	65,599	68,879	72,323
Annual Operating Expense (OPEX)		**219,500**	**230,475**	**241,999**	**254,099**	**266,804**
Total 5-Year OPEX						**1,212,876**
Expenses to the Profit and Loss Statement						
Annual Expenses		262,500	273,475	284,999	297,099	309,804
Total 5-Year Expenses →						**1,427,876**

Scenario Two: External Cloud						
		Year				
	Start-Up Costs	One	Two	Three	Four	Five
Capital Expenses						
Initial In-House ICT	75,000	15,000	15,000	15,000	15,000	15,000

Figure 4.1 CAPEX versus OPEX

Scenario Two: External Cloud						
	Start-Up Costs	Year				
		One	Two	Three	Four	Five
In-House Setup	15,000	3,000	3,000	3,000	3,000	3,000
Annual Capital Expense (CAPEX)		18,000	18,000	18,000	18,000	18,000
Total 5-Year CAPEX						**90,000**
OPERATING EXPENSES						
In-House Personnel		75,000	78,750	82,688	86,822	91,163
In-House ICT Maintenance		5,400	5,670	5,954	6,251	6,564
Cloud Computing Service		210,000	220,500	231,525	243,101	255,256
Annual Operating Expense (OPEX)		**290,400**	**304,920**	**320,166**	**336,174**	**352,983**
Total 5-Year OPEX						**1,604,643**
Expenses to the Profit and Loss Statement						
Annual Expenses		**308,400**	**322,920**	**338,166**	**354,174**	**370,983**
Total 5-Year Expenses →						**1,694,643**

All Capital Expense Items Use 5-Year Straight-Line Depreciation

Summary	Scenario One	Scenario Two
Total 5-Year CAPEX	215,000	90,000
Total 5-Year Opex	1,025,005	1,604,643
Total 5-Year Expenses to P&L	1,240,005	1,694,643
Year 1 Cash Needs	400,500	380,400
5-Year Expenses	1,240,005	1,694,643

Figure 4.1 *Continued*

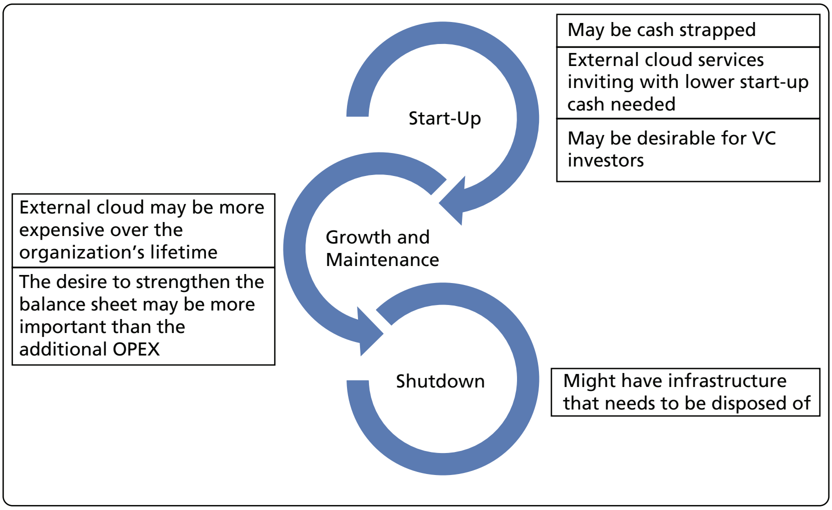


Figure 4.2 The Business Life Cycle

maintenance stages and through shutdown. The business life cycle must be considered when implementing cloud infrastructure. Early in the cycle, many organizations are strapped for cash, and an external cloud implementation may be inviting, with its lower requirements for start-up cash and its heavy emphasis on OPEX. This also may be desirable for investors (especially venture capitalists), who can be concerned that if a business does not succeed, the investors will have a costly infrastructure to dispose of.

As a business progresses into the growth and maintenance phases, it may have more cash, and from an accounting perspective may not want the emphasis on an OPEX funding model. A later-stage business may want to develop internal assets and desire a greater emphasis on CAPEX. It could still implement a cloud infrastructure, but that might be an in-house private cloud or a hybrid cloud with an emphasis on CAPEX. Whatever the deployment model, the balance between CAPEX and OPEX should be considered, along with the financial needs of the organization. Zynga, a company discussed in Chapter 3, is a good example of an organization that started out with a public cloud infrastructure and then moved to a hybrid cloud.

Return on Investment (ROI)

Return on investment (ROI) is an accounting valuation method. It is useful to compare the rate of ROI for various options, including ICT investments.[3] In its simplest form:

ROI = Net income ÷ Book value of assets

However, ROI can be calculated in a number of ways, including the following:

ROI = (Net income + Interest) ÷ Book value of assets

It is possible to modify the formula to take into account many factors that influence cost and return. When undertaking a project, the ROI formula might be as follows:

ROI = (Gain from investment − Cost of investment) ÷ Cost of investment

The preferred calculation includes those values that are most meaningful to the organization. Someone in marketing may compare the ROI of various products by dividing the revenue that each product generates by its expenses. Similarly, an ICT manager choosing among systems may select different parameters.

Using our two scenarios, let's assume that we can make estimates for expected revenue; these are shown in Figure 4.3.

In this example, ROI takes into account the costs associated with the proposed solutions and the expected gain from each solution (revenue is the same for each scenario, but it is possible that it might be different in each). If you are able to quantify TCO and the expected gains, ROI is a good tool to assist with selecting the choice that provides the highest return. When using ROI this way, you would interpret it as follows:

- If ROI is negative, then the investment would cost more than it would generate. The organization would probably not undertake the project unless it was required to meet regulatory requirements and it were choosing among solutions that had the least effect on ROI.
- If ROI is being used to compare alternatives, the alternative with the highest ROI should be selected.

Expected Revenue from New Infrastructure	260,000	286,000	314,600	346,060	380,666	1,587,326

Scenario One: In-House Infrastructure Budget						
	Year					
	One	Two	Three	Four	Five	Total
Annual Capital Expense (CAPEX)	**43,000**	**43,000**	**43,000**	**43,000**	**43,000**	**215,000**
Annual Operating Expense (OPEX)	**234,500**	**246,225**	**258,536**	**271,463**	**285,036**	**1,295,761**
Total Annual Expenses	277,500	289,225	301,536	314,463	328,036	**1,510,761**
Return on Investment	−6.31%	−1.12%	4.33%	10.05%	16.04%	**5.07%**

Scenario Two: External Cloud						
	Year					
	One	Two	Three	Four	Five	Total
Annual Capital Expense (CAPEX)	18,000	18,000	18,000	18,000	18,000	90,000
Annual Operating Expense (OPEX)	**290,400**	**304,920**	**320,166**	**336,174**	**352,983**	**1,604,643**
Total Annual Expenses	308,400	322,920	338,166	354,174	370,983	**1,694,643**
Return on Investment	−15.69%	−11.43%	−6.97%	−2.29%	2.61%	**−6.33%**

The ROI formula used here is ROI = (Gain from investment − Cost of investment) ÷ Cost of investment and is expressed as a percentage increase from the expense of the infrastructure.

Figure 4.3 Return on Investment

Calculating ROI allows management to determine whether the percentage of return is acceptable before a project is undertaken. Suppose that an organization wanted to track its investment in an infrastructure upgrade. The business analyst could use the costs of the upgrade and estimate the additional gains obtained from the upgrade.

The ROI accounting valuation method has several limitations and overstates ROI for the following reasons:

- Projects that take a long time to implement will overstate ROI. In fact, the longer the project, the larger the overstatement.
- The lag between investment outlay and the realization of expected gains will cause estimates to be inaccurate because the timing of the investment and the timing of any gains are not considered in the calculation.
- Organizations whose market share, income, and/or expenses grow faster will experience a lower ROI than slower-growing companies.[4]

Despite these limitations, ROI provides a good rough estimate and is a helpful tool to assist in understanding the financial effect of various options. In our two scenarios we saw a very different ROI, CAPEX/OPEX, start-up cash needs, and total cost of ownership (or operation).

CONCLUSION

In this chapter, we introduced three important tools (TCO, ROI, CAPEX/OPEX) that can be used to determine the true cost and effect of using an external cloud provider. It is important to use these three tools so that your organization will understand the real financial effects of moving to an external cloud provider for its ICT infrastructure needs.

A TCO perspective must be taken when considering CAPEX/OPEX, ROI, and any other appropriate analysis for your infrastructure choices. CAPEX/OPEX analysis is important because it emphasizes the different expenses that are part of building internal infrastructure and using an external cloud provider. The ROI metric is important because it shows how much revenue an investment will return.

Taken together, TCO, CAPEX/OPEX, and ROI form the basis for analyzing options and for determining how well your infrastructure is operating (both for internal infrastructure and for external cloud infrastructure). There are changes to these metrics that can be generalized as follows:

- TCO really becomes a mixture of ownership and operating costs when using an external cloud provider. This means that extra diligence is necessary to ensure that all expenses are considered and used in any analysis.
- In terms of CAPEX/OPEX, an internal infrastructure normally means that something is capitalized over a number of years—this is the CAPEX model. Using an external cloud provider means that your organization may have lower initial start-up costs but will probably have greater operating costs over the life of an application. This must be analyzed.
- ROI provides an estimate of the return on an investment. ROI can be used to select among options when the expenses are known and revenue can be estimated for each option. ROI can also be used as a hindsight metric that uses known financial results.

All the measurements discussed in this chapter and all the implementations discussed throughout this book should be undertaken to make a business more agile. *Business agility* is the ability of an enterprise to sense and react quickly to changes in its environment; those changes can be a competitor going out of business or introducing a new product, the loss or gain of significant customers, new regulatory requirements, or something else.

Business intelligence is important in sensing these things, and cloud computing is important in responding to them. Cloud computing can enable an organization to reduce cycle time, increase innovation, and react to changes more quickly. The organization that accepts the shift to greater agility is more likely to succeed in the marketplace.

NOTES

1. The actual analysis of CAPEX versus OPEX requires an accountant. This book is not making specific recommendations but rather is suggesting that the benefits of CAPEX versus OPEX must be considered when moving IS to a cloud.
2. I want to thank Harlan Shakun, CPA, for reviewing the accounting language and the two scenarios in this section to ensure that they are correct from an accounting perspective.
3. Steven M. Bragg, *Business Ratios and Formulas: A Comprehensive Guide* (Hoboken, NJ: John Wiley & Sons, 2002); and Ciaran Walsh, *Key Management Ratios: Master the Management Metrics That Drive and Control Your Business*, 3rd ed. (Harlow, UK: Financial Times Prentice Hall, 2002).
4. "Management Methods, Management Models, Management Theories," Value-Based Management, June 10, 2006, www.valuebasedmanagement.net/index.html.

CHAPTER 5

Cloud Adoption

Are Your Organization and Its Stakeholders Ready to Adopt Cloud Computing?

Specifying the parameters of deployment and service type is clearly the purview of designers and those who specify services. However, before deployment and services are specified, the need must be defined and the effects of the delivery mechanisms must be understood. These are business issues, and the decisions involved must be made in cooperation with managers who must identify the business needs. Decisions about cloud services affect your organization and the linkage between your organization and its stakeholders.

Successful usage of a cloud service requires that you understand its effects on your organization and its stakeholders. In this chapter, the major areas to be considered include the effects on the following:

- Current IT staff
- Current technology
- Internal stakeholders (e.g., managers, users, and other organizational staff)
- External stakeholders (e.g., customers or suppliers)

Table 5.1 The Effects of Using a Cloud Strategy

Effects on Non-IT Staff	Effects on IT Staff	Effects on External Stakeholders
■ Regulatory influence ■ Increased expectations in technical abilities of both current employees and the newly hired ■ Organizational cultural transformation—potential for increased innovation ■ Increased organizational agility and its consequences ■ Increased expectations of IT staff	■ Regulatory influence ■ Different focus for the CIO ■ Changes in governance structure to support a cloud strategy ■ Changes in management of IT resources ■ Need to assess staff readiness	■ Regulatory influence ■ New infrastructure expenses ■ Increased innovation and reduction in product cycle times ■ Changes in payment methods ■ Staff training

An overview of these is shown in Table 5.1. Each of these effects will be discussed in turn in this chapter, and the next chapter will be dedicated to a discussion of how service level agreements affect cloud strategies.

REGULATORY INFLUENCE

Before proceeding to the specific issues that affect your organization and its stakeholders, the regulatory influence must be discussed since it will affect all entities in the cloud arrangement. The type of business your organization is in and the regulations that govern that business will determine the regulatory influence on your organization and its stakeholders. An example of this is the Health Insurance Portability and Accountability Act (HIPAA) in the United States. If your organization provides services that fall under this act, it is your responsibility to ensure the security and privacy of the health-care data you maintain. It is fairly obvious, but this also means that any cloud provider your organization selects for applications or content must also be compliant with HIPAA. What may be less obvious is that your suppliers who deal with health-care information must also be compliant. This may present a new burden to those suppliers when they are suddenly required to participate in a new cloud arrangement.

As an example, let's say your organization is a health maintenance organization and you desire to move to a cloud provider for your IT. Here are a few considerations:

- Your cloud must be compliant with HIPAA, which will entail an extra cost to audit cloud providers (e.g., your organization must audit all its external cloud providers to ensure compliance).
- Your organization must meet HIPAA guidelines for patient security and privacy, and that may mean a new expense (e.g., encrypted high-speed connectivity to the cloud provider so that patient information can be securely transmitted).
- If your cloud provides for patients (arguably, your customers) access to their health-care data, your organization must take reasonable precautions to ensure that the data is protected (e.g., authentication to your cloud, encryption during transmissions, secure e-mail).
- Your suppliers (e.g., hospitals or doctors' offices) who are connected to the cloud solution will need to employ appropriate measures to ensure HIPAA compliance and your organization may need to audit them. This may be a new expense.

Moving to the cloud and deploying to your suppliers and customers is not a zero-sum game. There are expenses and other consequences that must be considered, since they affect the relationship among you, your staff, and your external stakeholders. These issues are not limited to health care in the United States; similar issues emerge whenever an industry (regulated or not) moves to the cloud.

THE EFFECTS ON YOUR ORGANIZATION

Moving any content or application to a cloud provider will affect your organization, from the management of IT to the operations staff. The focus of this book is BI, but the reality is that moving any content or application to a cloud provider has consequences. The effects on the non-IT staff are considered first in this section, and then the existing technology and the IT staff are discussed.

The Non-IT Staff

Cloud computing will affect your organizational staff. There are at least four areas that must be considered when you are determining whether moving to a cloud platform is appropriate for your organization. Specifically, a cloud environment will cause the following:

- New expectations in technical expertise of the organizational staff and the technologies it manages.
- A transformed organizational culture.
- Increased organizational agility that will speed up the innovation cycle.
- Expectations by the organizational staff that the IT department will be more responsive to its needs.

Cloud computing offers the potential for a new organizational culture and architecture that fosters innovation.[1] That innovation architecture brings with it the hope of increased competitive advantage as well as the concern that the existing staff might not be able to exploit these new technologies. The backbone of enhanced competitive advantage is strategic in nature; thus, when an organization chooses the IT strategy of cloud computing, it must consider whether the existing organizational staff is ready for this challenge. It could be necessary to hire new staff or get the existing staff trained so that it can use the technologies.

New technologies, faster cycle times for innovations, and more rapid deployment of services call into question the readiness of the existing staff and its sheer sufficiency. Staffing considerations become paramount to attain the maximal ROI of a cloud strategy.

Hiring new staff can cause changes in an organization that has adopted a cloud strategy. This hiring can be disruptive to the organization and its existing staff. If the new staff members are expected to deploy innovations, they must possess both the business and the technical skills that enable them to do so. That could effectively change the competencies that you look for in high-value new employees.

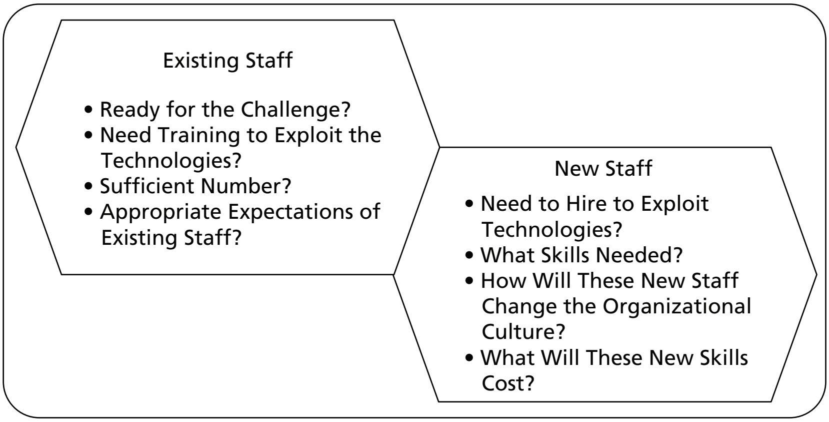


Figure 5.1 Staffing Considerations

This is especially important because these new staff members might come at a premium—for instance, they might demand higher salaries. The effect is that the budgeting and hiring processes might have to change. Staffing considerations are summarized in Figure 5.1.

Organizational Transformation

Organizational transformation can be brought about through a cloud strategy, but this is a "chicken or egg" problem. Does the organization first change its culture to one with a different risk profile that fosters innovation, or does it first implement a cloud strategy and assume that the cultural transformation will follow? The truth lies somewhere in between. It is probably not possible for your organization to stop operations and change to a culture that is more technically agile; rather, the transformation will occur over time, and it requires planning.

Organizations need to continue operations so they can make money to pay their expenses. But at the same time, they need to invest in their future. BI solutions and cloud computing can be a way to make that investment. Based on whether you see BI and cloud computing as "bleeding edge" or leading-edge technologies, your organization should expect to spend either 10 to 25 percent or 50 to 75 percent of its IT budget on these expenses.[2] These choices affect

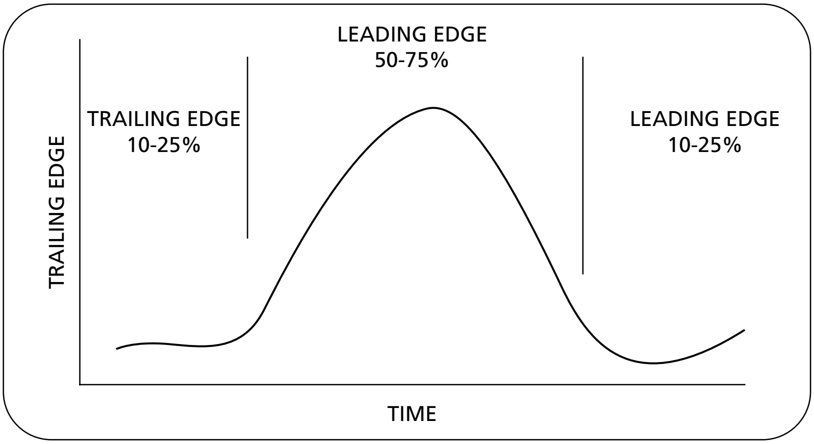


Figure 5.2 Budgeting for IT

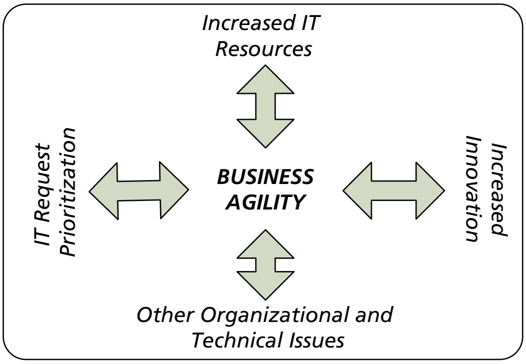


Figure 5.3 The Effect of Organizational Agility on IT

your organization and its stakeholders. Recommendations for budgeting them are shown in Figure 5.2.

Increased Organizational Agility

In Chapter 4, we covered the increased organizational agility enabled by cloud computing. This brings about several issues that should be considered (see Figure 5.3):

- **Increased innovation.** Innovation within the organization should increase as the staff realizes that IT services are available more quickly. This means that customer and staff requests for new services must be completed in a shorter time frame.

- **Increased IT resources.** An increase in IT resources (e.g., staff, hardware, or software) may be required to meet the new requests to support innovations. As more new services are requested more quickly, an increase in the resources required to fulfill those requests must be available.
- **IT request prioritization.** IT requests must be prioritized. Those that generate greater profitability for the organization should be considered for implementation first. This will affect the allocation of IT resources and may call for a more robust IT governance structure with a stronger portfolio management of projects.
- **Other organizational and technical issues.** The organization will be faced with a number of structural, cultural, and technical issues that must be considered. Many of these are covered in this book, but every organization is unique, and you must consider how a new technical and business strategy will affect your enterprise.[3]

An agile organization can adapt to changes in the market, including customer demand and the emergence or elimination of competitors, and can possibly develop new industry-changing technologies. Cloud computing enables these adaptations, but they may require changes in the organizational culture. This is really somewhat circular: As these changes occur, the organization becomes more agile, and as the organization becomes more agile, more changes in IT may be necessary.

The IT Department

When moving to a cloud strategy, management must consider the following effects on the IT department: (1) the job of the chief information officer (CIO), (2) the readiness the current IT staff, and (3) the ability of the current technology to handle the new requirements of the cloud infrastructure. These are shown in Figure 5.4 and are discussed next.

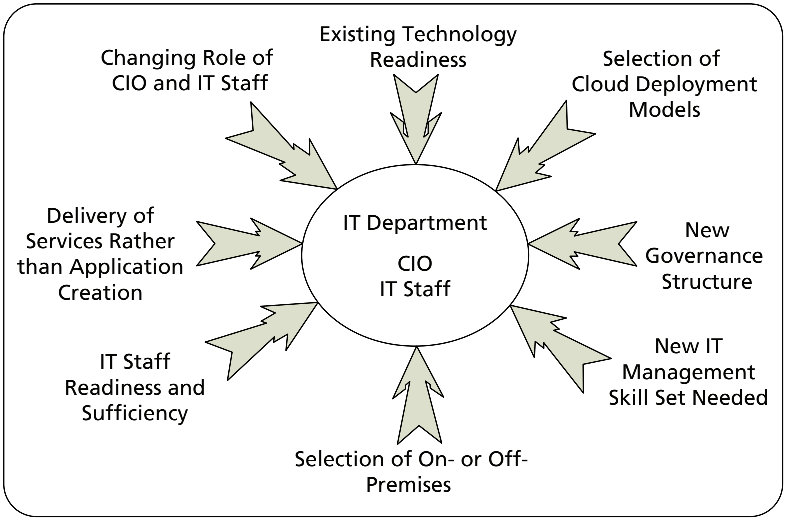


Figure 5.4 Pressures on the Existing IT Department

The CIO

The CIO historically has a short life span at any one organization. A number of reasons for this have been posited, but what is constant is that the job of the CIO has changed substantially. At one time, the focus of the IT staff in most organizations was to build systems and solutions that support business processes. These solutions often took a long time to build, and the CIO was largely the senior technocrat of the organization.

In the past, CIOs needed business knowledge, but they needed technical knowledge more. These senior technocrats were concerned with "keeping the trains running on time" by making sure that e-mail flow or the customer relationship management (CRM) system was functioning smoothly. However, that has changed. "I don't think about that anymore," said Stephen Potter, the CIO at World-Check. "I have virtually no one on my team thinking about that anymore. I spend much more of my time thinking about IT that can drive the business forward, which is really what I should be doing."[4] The CIOs' position, which was largely reactive to meet organizational needs, has changed to a more proactive role that consists of looking for ways to deploy technology to support the competitive positioning of the company.

Successful implementation of a cloud strategy requires that the CIO be retooled to become the lead innovator in an organization on the forefront of new product development and managing change. This means that the CIO must have greater business knowledge than ever before and be less of the lead technocrat. CIOs must spend much more of their time considering how to implement IT solutions (including BI solutions) that move the organization forward and assist it in its attempts to reach an enhanced competitive position.

The CIO must now make IT service delivery, rather than solution development, a priority. The job is no longer just about keeping the existing systems running and providing new solutions; rather, the CIO must now ensure that users can innovate and create their own solutions by providing services ready to support enterprise-wide innovation. Innovation can be supported by an internal cloud infrastructure or by an external cloud provider.

Governance Structure and IT Staff Skills

A strong governance structure must be put in place to ensure the integration of internal business processes and cloud services. That structure must foster innovation while also making sure that shadow (or unsanctioned) IT does not proliferate. This is not easy, especially in an organization that has a preexisting culture that has allowed shadow IT. The concomitant change in organizational culture, the new expectations of the operational staff, the changed role of the CIO, and a host of other factors requires that the emphasis of the IT staff changes and that a governance structure (policy and procedures) is put into place that recognizes the shift caused by cloud computing. The CIO and the IT staff must realize that a governance structure that enables the empowerment of all staff is a substantial part of their job in any cloud deployment and that they must foster a climate that creates that empowerment.

The governance structure will have a different emphasis depending on whether an organization has significant technology presence on-premises or uses cloud providers. In the on-premises governance structure, there is much more emphasis on the local deployment of technology, including increased innovation using that technology. The IT staff with an on-premises infrastructure will be expected to manage the infrastructure and foster innovation at the same time. The staff must have intimate knowledge of the technologies it maintains along with the business knowledge that creates competitive advantage. The key to the success of this governance structure is a collaborative environment that welcomes all stakeholder input into the procurement, deployment, and management of IT resources.

In situations in which the cloud infrastructure is off-premises, the IT staff members will probably need less technical background, since they are not responsible for managing the technical infrastructure of the cloud. They will still need to manage any local infrastructure, which could be as simple as Internet connectivity, a local area network, and/or PCs and printers. The role of the IT staff will be focused less on technology and more on business. A collaborative governance structure is also essential in off-premise deployments. When there is minimal local infrastructure, the IT staff (or possibly the informatics staff) will be expected to have greater business knowledge so it can support innovation in the off-premises efforts of the enterprise.

The type of organization and whether implementations are on- or off-premises mandate different governance structures and skill sets in the IT staff. These are important considerations when moving to a cloud infrastructure.

IT Resource Management

Adopting a cloud strategy means that IT resource management changes. A new IT management skill set is needed. In a non-cloud environment, an organization requires appropriate technologies and an IT staff that can create and manage them. It also needs an operations staff that understands the business. To be successful, a business must be good at its core activities, but in today's technological world it also needs to have an IT staff who understands the business and can manage technology. That begs the question, what are the necessary skills of the operations and IT staff, and are both needed?

Netflix has grappled with the question of having both an operations and an IT staff. The company decided that it wanted to do what it does best—managing customers and movies—instead of trying to do something it knows very little about; it left the IT infrastructure and associated staff to another company (Amazon Web Services). The question of whether it is more beneficial to host your infrastructure in-house or to use a cloud provider must be considered. Your organization must take into account several important issues: (1) where the organization is in the business cycle, (2) the cost-effectiveness of running its own infrastructure and staff, versus using a cloud provider, and (3) how an external cloud provider will affect its stakeholders.

Existing Technology Readiness

Before you proceed on a cloud strategy, it is important that your organization determine which applications or content will be moved to the cloud. Once that is done, you can determine which cloud deployment model(s) best fits and whether the existing technology is up to the challenge. It is necessary that your organization thoroughly examine itself to determine the appropriate content or applications to move to the cloud and what new services or technologies will have to be purchased or deployed. That is not trivial, and it requires an honest look at your organization and its staffing, existing technologies/applications, and resources.[5]

On-Premises versus Off-Premises and Cloud Deployments

If your organization is embarking on building a cloud environment, it needs to determine whether it will be done in-house or purchased from a cloud provider. An in-house cloud environment requires that the IT staff have the appropriate technical knowledge and skills. The IT staff also needs to empower the non-IT staff to access and deploy applications and content. A cloud environment calls into question the technical skills of your existing staff and the need for new staff. The staff skill set must be appropriate to the on-premises or off-premises deployment.

If your organization is building an external cloud environment, it has made the commitment to purchase services from a cloud provider. In that case the IT staff will not need the proficiency to run an in-house cloud, but it will need the business knowledge to maximize

ROI. The external cloud provider has become a partner in delivering services to your organization, and therefore you must make sure that the cloud provider is able and willing to provide them. This is especially true for any mission-critical services. Ways to cope with this new relationship are discussed in Chapter 6.

Cloud deployments have specific requirements, and these exist along the dimensions of service model (PaaS, IaaS, or SaaS), location of deployment (on-premises or off-premises), and deployment model (public, private, hybrid, or community). Here are two considerations:

- Any cloud (the deployment and service models do not matter) that is deployed on-premises will require that your staff manage the deployment of the content or applications and the associated technology.
- Any cloud (the deployment and service models do not matter) that is off-premises will require that your staff manage the content or applications while the cloud provider manages the technology.

The requirements for IaaS, PaaS, and SaaS vary by where the cloud is deployed (on-premises or off-premises). These are shown in Table 5.2, which lists the differences between on-premises and off-premises across deployment models.

An example will serve to make the point. Where do you want to deploy your applications and content? Suppose your organization wants to create a CRM solution. It has many choices, but in our example let's say the organization has narrowed down its options to the following three:

1. **Build the application**. Build the CRM application internally and deploy it on the organization's on-premises cloud or on an external cloud provider (e.g., Microsoft Azure or Amazon Web Services).
2. **Buy the application**. Buy a CRM solution and deploy it on the organization's on-premises cloud or on an external cloud provider (e.g., Microsoft Azure or Amazon Web Services).
3. **Use a cloud provider for the application**. Use a cloud provider for the CRM solution (e.g., Salesforce) and either use the

Table 5.2 Cloud Deployment Responsibility

	Private Cloud Deployment	Public Cloud Deployment	Hybrid Cloud Deployment	Community Cloud Deployment
Deployed by your IT staff on-premises	■ Can be IaaS, PaaS, or SaaS. ■ Organization responsible for creation, deployment, and operation of the cloud infrastructure. ■ Organization responsible for delivery of appropriate cloud services.	By definition, you would not have a public cloud internal to your organization unless you were a cloud provider. The discussion in this chapter is from the perspective of your organization.	Has features of on- and off-premises deployments.	Could have features of on- and off-premises deployments based on whether your organization is just using a public cloud provider or you are integrating your organization's cloud with the community cloud.
Deployed off-premises using cloud provider(s)	■ Organization responsible for creation and deployment of applications and content. ■ Cloud provider responsible for technology and its operation. ■ Single tenant: Your organization has dedicated technology at the cloud provider's location.	■ In an IaaS or a PaaS model (e.g., Amazon Web Services or Microsoft Azure), the organization is responsible for content or application creation and deployment; the cloud provider is responsible for the technology and operation. ■ In a SaaS model (e.g. Salesforce or Dropbox), your organization is responsible for content; the cloud provider is responsible for the application and operation. ■ Multitenant: Your organization shares technology with other users at the cloud provider's location.		

provider's web portal to access the services (i.e., a totally off-premises deployment) or build your own web portal and run it on your own cloud, which accesses the cloud provider's application program interface to process and store the CRM data (i.e., on-premises for the web portal but off-premises for the application and data).

Before a decision can be made, you must answer this question for each option: Where do you deploy the solution or content? On-premises has the advantages of control, but off-premises also has advantages, associated with your organization being able to concentrate on the CRM application and leave much of the technical management to the cloud provider. Each also has risks associated with the need for staffing, control, and technical management. This must be assessed when deciding where and how to deploy solutions.

When an off-premises cloud is adopted, your organization transfers responsibility for operation of the technology to the cloud provider. The off-premises cloud provider becomes a partner in the delivery of IT services; significant issues about this are discussed in Chapter 6. In addition to the obvious technical issues, where the cloud is deployed changes the staffing model (i.e., the type of staff and its technical competencies). An in-house cloud requires that the technical staff is able to manage it, whereas the off-premises cloud requires that the organization understand and accept the risk of having an external business partner (the cloud provider).

Before we leave the topic of where the cloud is deployed, it should be noted that even when an external cloud solution is chosen, there will always be a need for local IT support. Whether that staff is the "lone ranger" that builds and operates its own IT or is the mega-conglomerate that has a large IT staff, there will always be a need for someone to know and understand the in-house IT infrastructure. Cloud computing requires connectivity to the Internet, and at a minimum someone needs to understand how that connection works.

IT Staff Readiness

The readiness of the existing IT staff to adopt a cloud strategy is largely determined by whether it can adapt to a culture of service delivery rather than application creation and whether it has the skill set necessary to manage a cloud infrastructure. Understanding the cultural readiness of the IT staff is tantamount to realizing the ROI of a cloud initiative. This is a difficult issue: Does the culture exist to maximize the investment in cloud computing? We have all been there—we have had an IT staff member who believes that IT should be operated by IT staff and that all other staff should be beholden to the IT staff to get things done. Such a difficult IT staff member can be a problem. In a cloud environment, this attitude simply does not support a culture that maximizes ROI. Rather, the attitude must be one of empowering the operational staff to create content and deploy applications. If you are reading this book, you probably understand the difference and will at some point have to deal with a difficult IT staff member.

The other important issue is IT staff skills. Does the IT staff possess the necessary technical skills to deploy a cloud solution? Training encompasses everything from technical knowledge about cloud deployment to special security requirements. There are many training outlets (e.g., the Cloud Credential Council at www.cloudcredential.org), making it largely a matter of resources. In addition to considering the direct cost of training, you need to decide if your organization can spare the existing staff so that it can be trained.

THE EFFECTS ON EXTERNAL STAKEHOLDERS

The deployment of a cloud computing solution that requires external stakeholders to interact with it will place a burden on those stakeholders. This includes the need for increased technology, changes in payment models, the training of a stakeholder's staff, regulatory influence, and an increased demand for innovation (see Figure 5.5). These are major effects on all external stakeholders, including customers and suppliers, which must be considered.

The Need for Increased Technology

When cloud technologies necessitate that external stakeholders' infrastructure is able to integrate with them, the external stakeholders must be considered. Take the case of a company that deploys a supply

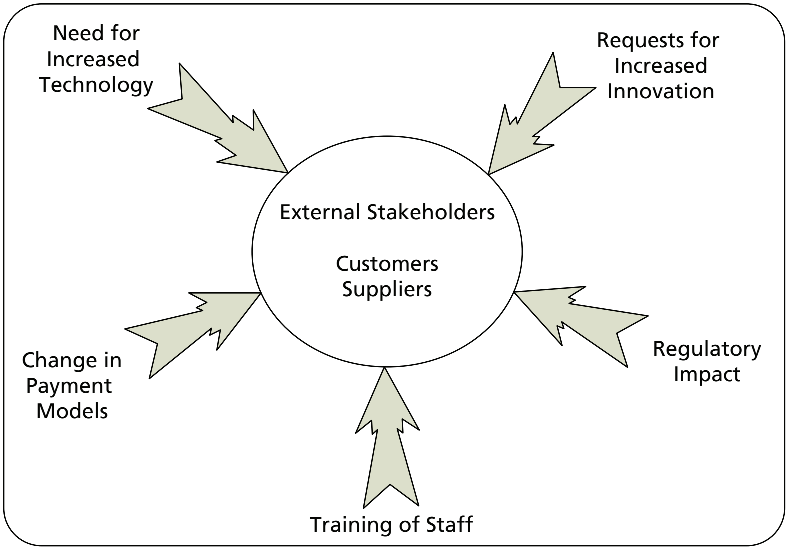


Figure 5.5 Pressure on External Stakeholders

chain management solution in the cloud that requires its vendors to have technologies in-house (e.g., radio-frequency identification and its associated technology) to allow the vendor to fulfill orders electronically. What happens to the vendor that provides a rare item but does not have the resources to interact with the cloud solution cost-effectively? Do you stop doing business with that vendor? Do you provide it with some incentive to obtain the necessary technology? Do you source the item elsewhere, and if so, what effect does that have on your production?

Changes in Payment Models—Changing Costs

If your organization deploys content or applications to an external cloud provider, you will undoubtedly take on new expenses, probably in a metered use arrangement. This means that you will be paying to use someone else's IT infrastructure. There will be benefits from the cloud computing deployment (e.g., increased innovation in your organization, faster time to market, a shift in the burden for maintaining the technology to the cloud provider), but they will come with increased cost. There are at least two issues that must be dealt with:

1. If your IT expenses increase, those resources must come from somewhere. Do you "take the hit to the bottom line," pass the increase along to customers, negotiate with your suppliers for better rates, or charge suppliers to access your new cloud?
2. Suppliers will need to decide whether selling to your organization is cost-effective. If not, they might need to either stop providing goods to you or increase the price of their goods.

Increased expenses and costs affect your organization as well as your customers and suppliers. Cost shifting will occur and must be managed.

The Training of Staff

As with any new deployment, whether it is new content or new applications, training is a large expense. It is no different when deploying to the cloud. Just as the staff of your organization will need training, so will the staff of your suppliers. Who pays for that training—your organization or the supplier?

Requests for Increased Innovation

The cloud can enable increased innovation along with a reduction in cycle times for new products. This means that external stakeholders may come to expect product changes, including innovations, at a quicker pace than your organization previously expected to deliver them. These new expectations must be managed.

CONCLUSION

There are a myriad of stakeholders who must be considered when deploying all IT. In this chapter, we have mentioned a few of the major ones. It is important to realize that all new initiatives affect all stakeholders, both internal and external. It does not matter what you are providing to the stakeholders; it is important to manage expectations. One way to do this is to implement a service level agreement. The next chapter reviews the service level agreement with a focus on managing the relationship between your organization and the cloud provider, and how that relationship affects all stakeholders.

NOTES

1. T. Clohessy, T. Acton, and C. Coughlan, "Innovating in the Cloud," *International Journal of Innovations in Business* 2, no. 1 (2013): 29–41, http://0-search.proquest.com.www.consuls.org/docview/1316057717?accountid=9970.
2. Michael S. Gendron, *Business Intelligence Applied: Implementing an Effective Information and Communications Technology Infrastructure* (Hoboken, NJ: John Wiley & Sons, 2013), 49–50.
3. Ibid.
4. Martin Tantow, "Cloud Computing: Impact on CIOs and the IT Team," *Cloud Times*, April 19, 2011, http://cloudtimes.org/2011/04/19/cloud-computing-impact-on-cios-and-the-it-team.
5. Gendron, *Business Intelligence Applied*.

CHAPTER 6

Service Level Agreements

Traditionally, the service level agreement (SLA) has been thought of as pertaining to technologies that require a particular quality of service (QOS) in order to ensure availability and response time on a network. QOS also applies to cloud computing but in that context must be expanded past availability and response time to include other business parameters, like data retention and drivers of business continuity. The cloud SLA is essential so organizations know the minimum level of service they can expect and the providers know the minimum level of service they must provide. The SLA must also specify what occurs when a provider does not meet the SLA's business parameters. The overarching purposes of the SLA are as follows:

- Identify and define the customer's needs.
- Provide a framework for understanding.
- Simplify complex issues.
- Reduce areas of conflict.
- Encourage dialogue in the event of disputes.
- Eliminate unrealistic expectations.[1]

A typical SLA includes the following specifications: (1) service availability, (2) the length of time a provider has to rectify any issue with the service, and (3) service monitoring. These specifications are set out in a number of measurable objectives and consequences in case they are not met. Typically, there are two types of QOS objectives: (1) operational (e.g., mean time between failures), and (2) service specific (e.g., availability or delay).[2] Operational and service-specific objectives are discussed next to describe a typical telecommunication's SLA.

THE TRADITIONAL OR TYPICAL TELECOM SLA QOS

Asynchronous transfer mode (ATM) is a communication technology that often uses an SLA because ATM provides a strong set of QOS measures and is a good example for illustrating QOS. ATM QOS measurable objectives include the following:

- Operational objectives
    - **Mean time between failures (MTBF).** The average time allowed or expected between equipment failures.
    - **Mean time to repair (MTTR).** The average length of time it takes to repair equipment that has failed.
    - **Mean provisioning time.** The average time it takes to make new services available.
- Service objectives
    - **Availability.** The amount of time that a service is available (or unavailable).
    - **Traffic parameters, delay, and output.** The type of traffic (e.g., constant, real-time), the speed of that traffic, and the acceptable amount of delay during transmission.
    - **Transmission errors.** The acceptable number of errors during transmission.

This is an oversimplification of an ATM QOS, but it makes the point that there are different types of objectives, and they affect your organization and its stakeholders. For example, the mean time to repair a down service will indicate the provider's ability to respond to a need quickly, whereas traffic parameters may indicate whether a provided service can meet the need for the service (e.g., speed, delay, output). These objectives affect the organization, and if the ATM service involves your customers or suppliers, they are affected, too.

ATM QOS gives a basis for understanding why we enter an SLA and how you could use an SLA in a cloud environment. Purchasing services from a cloud provider means that your organization is moving at least partial responsibility for IT or IS services to a cloud service provider. The cloud service provider becomes an extension of your organization and must be treated as such.

INTRODUCING THE CLOUD SLA

In order to meet the needs and expectations of your stakeholders in a cloud ecosystem, you must consider the cloud SLA. The cloud SLA determines how cloud services will be delivered and defines the services that are provided. This is true whether your organization's IT department is providing cloud service to your staff or you are purchasing cloud services from an external provider. In the remaining sections of this chapter, we will focus on purchasing cloud services from an external provider, but the concepts can easily be used in an SLA internal to your organization.

If your organization is going to use an external cloud service provider, then the SLA must be considered. It is probably obvious, but a cloud provider will not be anxious to negotiate an SLA with an organization that signs up for an account unless that account is important to the provider (e.g., a large customer that is purchasing a substantial amount). Typically, when an organization signs up to use a cloud provider's services, it will agree to a standardized usage agreement and an SLA. However, if the services are important to your organization, it is important that you understand the limitations of a standard usage agreement and negotiate an SLA specific to your organization, if possible. In order to understand what is necessary when purchasing cloud services and the SLA, the following will be covered first:

- Cloud SLA types
- Cloud use cases
- SLA by type of service and deployment
- Anatomy of a standardized SLA

After covering these introductory topics, we will discuss how to negotiate the cloud SLA.

SLA TYPES

There are two major categories of SLAs:

- **Standardized.** The standardized SLA (sometimes called a usage agreement, a customer agreement, or service terms) is the more frequent. It consists of standard language that is agreed to when you create a new account with a cloud provider. It is normally written in favor of the cloud provider.
- **Negotiated.** The negotiated SLA is infrequent. It is negotiated between your organization and cloud provider, and uses metrics to determine QOS. This is not typical unless your organization is purchasing a large amount of services from a cloud provider.

Small and medium-size businesses (SMBs) will typically use the standardized SLA, whereas a large business has more influence to negotiate. The reality is that SMBs have less clout because of their lower buying power in the cloud marketplace. Nevertheless, it is important that SMBs understand the limitations of the standardized SLA so they are cognizant of the cloud provider's impact on the SMBs' operations and their stakeholders.

CLOUD USE CASES

There are many use cases of cloud-based services. Although the overriding sentiment is that individual or singular organizational accounts with a cloud provider will normally be met with a standardized SLA, the SLA must be understood in the context of use cases. If your cloud use case involves a mission-critical application, you must consider the SLA and its parameters. Typical use cases are shown in Figure 6.1 and detailed next:

- **End user → public cloud provider.** Dropbox is a good SaaS example of an end user directly accessing a cloud application.

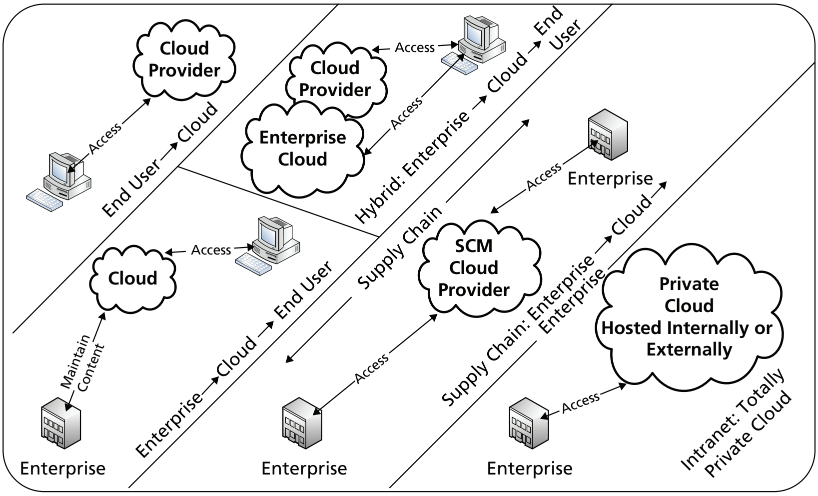


Figure 6.1 Cloud Use Cases

The end user can set up an account, store files on Dropbox, and access them from anywhere.

- **Enterprise → public cloud → end user.** Amazon Web Services provides IaaS, and Microsoft Azure provides PaaS. Your organization can deploy its applications and content to one of these services and then make them directly available to end users.
- **Hybrid: enterprise (private cloud) → public cloud provider → end user.** Zynga and its relationship with Amazon Web Services was discussed in Chapter 3; this provides a good example of a hybrid cloud arrangement. Zynga deploys it applications on its own private cloud, and in a hybrid arrangement moves users to Amazon when Zynga's cloud is at capacity, thus optimizing the user experience.
- **Supply chain: enterprise → public cloud provider → enterprise.** AXIT AG, an IT solutions company in Germany, and is an example of supply chain cloud services. It provides a cross-company platform for the management of logistics

services through SaaS offerings that connects the enterprises in a supply chain.

- **Intranet: private cloud.** An organization can have a cloud deployed only for the use of its employees. It could be implemented behind the firewall within the organization, or it could be a private cloud hosted by a cloud provider. What defines the intranet private cloud is not its location (i.e., on-premises or off-premises) but that it is available only to your organization's staff.

These typical use cases provide a basis for understanding the various types of SLAs. Some considerations for SLA implementation are shown by use case in Table 6.1.

Whether the use of a standardized or negotiated SLA is important is largely based on the effect that moving your organization's content or applications might have. That effect must be known and managed when you are choosing a cloud provider.

Table 6.1 Cloud Use Cases and SLA Considerations

Typical Use Cases	SLA Considerations
End user → public cloud provider	Normally this is a standardized SLA with individual user accounts.
Enterprise → public cloud → end user	An enterprise might negotiate an SLA for cloud services. However, the standardized SLA will be offered most frequently.
Hybrid: enterprise → public cloud provider → end user	Same as above.
Supply chain: enterprise → public cloud provider → enterprise	An enterprise will use the standardized SLA, or, if the enterprise is large enough, it may negotiate an individualized SLA with the cloud provider.
Intranet: totally private cloud	When an enterprise moves its intranet to a cloud provider, it will be presented with a standardized SLA. The enterprise may want to negotiate something stronger in order to protect its operation.

Based on: Cloud Computing Use Case Discussion Group, *Cloud Computing Use Cases*, July 2, 2010.

Table 6.2 SLA by Service and Deployment Models

		Service Models		
		IaaS	PaaS	SaaS
Deployment Models	Private	Should consider negotiating a strong SLA, but standardized is usually offered.		Usually standardized.
	Public	SLA is usually standardized and largely in favor of the cloud provider, but for a very large purchaser, a negotiated SLA is usually possible.		
	Hybrid	SLA is dependent on type of hybrid cloud.		
	Community	SLA is normally negotiated between the consortium that provides the community cloud and the cloud provider.		

SLA BY TYPE OF SERVICE AND DEPLOYMENT

Use cases employ the various deployment models (private, public, hybrid, and community). Each service model (IaaS, PaaS, and SaaS) and each deployment model has various SLA characteristics. This occurs because each provides different capabilities to your organization, and the scope and scale of the services changes. Recommendations for an SLA are shown in Table 6.2. However, objectives, metrics, and other SLA parameters must be considered as unique to your business and its needs.

ANATOMY OF A STANDARDIZED SLA

The Amazon Web Services customer agreement provides a good example of a standardized SLA. The concepts discussed in this section are generally applicable to all cloud service models (IaaS, PaaS, and SaaS) and deployment types (private, public, hybrid, and community).

When you sign up for an account with Amazon Web Services, you are required to consent to a standardized agreement between your organization and Amazon.[3] The customer agreement contains the following sections that are pertinent to our discussion:

- Use of service offerings
  - You may access and use the various cloud services specified in accordance with this customer agreement.
  - Your organization will adhere to all applicable laws, rules, and regulations in your legal jurisdiction and will adhere to Amazon's acceptable use policy.
  - To get support beyond what is normally provided by Amazon free of charge, your organization must enroll and pay for extended support in accordance with Amazon's support guidelines.
- Security and data policy
  - Amazon will implement reasonable and appropriate measures to secure your organization's content from accidental or unlawful access or disclosure.
  - Amazon adheres to a safe harbor program, as described in its privacy policy.
- Your organization's responsibilities
  - Development, content, operation, maintenance, and use of any content you deploy on the Amazon cloud service.
  - Technical operation of your content, including any calls that your applications make to Amazon's application program interfaces (APIs).
  - Compliance with all Amazon policies, including the acceptable use policy, and any applicable laws.
  - Any claims relating to your content.
  - Proper handling of any notices sent to you or your affiliates that your content violates a person's rights, including notices pursuant to the Digital Millennium Copyright Act.[4]
  - Proper configuration and use of the Amazon Web Services offerings; maintenance of your own appropriate security, protection, and content backup.

    - All actions that you permit, assist, or facilitate, including actions that you allow end users to take; assurance that end users' actions as they pertain to your content on Amazon are consistent with this agreement.
    - All support to end users, unless a separate support agreement is in place.
- Conditions for temporary suspension
    - You pose a security risk to Amazon's service offerings.
    - Your content or operation may adversely affect any Amazon service offering.
    - You subject Amazon, its affiliates, or any third party to any possible liability.
    - Your account with Amazon is fraudulent.
    - If your organization or your end user is temporarily suspended, you will still be responsible for any accrued and ongoing fees, and your content will not be erased solely because of termination.
- Termination
    - Either your organization or Amazon can terminate for its convenience with a 30-day advance notice.
    - Amazon can terminate your organization's service immediately upon notice under the following circumstances:
        - Any act or omission by your organization or your end user can result in temporary suspension (described earlier).
        - Amazon's relationship with a third-party partner that provides technology used by Amazon's services expires or somehow materially changes.
        - Amazon believes that providing a service could create substantial economic or technical hardship or create a material security risk.
- Proprietary rights
    - Your organization is responsible for and retains all licensing of software and content deployed on Amazon's systems. Licensing does not transfer to Amazon just because software or content is deployed on its hardware.

- All submissions to Amazon are governed by the terms of the Apache software license, unless another Amazon-supported license is specified when you submit your content.

A separate agreement specifies the service terms for the various cloud services offered by Amazon Web Services. These include the following universal terms:

- All content deployed on Amazon must be owned, licensed, or lawfully obtained by your organization.
- It is your organization's responsibility to comply with all technical documentation applicable to the Amazon services, including the appropriate developer guides.
- Your organization will make client information and other material available to Amazon so it can verify your organization's compliance with all applicable agreements and documentation.

In addition to the universal service terms, each of the Amazon cloud services (e.g., compute, database) has specific service terms that must be followed. The content and terms differ for each service, but the intent in all of them is that your organization will comply with all Amazon Web Services agreements and technical documentation. There are also some application-specific terms (e.g., how Microsoft software can be used).

Amazon also has an acceptable use policy and a set of privacy policies that must be followed. This presents a complicated set of documents that should be understood if your organization is going to deploy using the Amazon platform; the Amazon SLA is a typical standardized SLA for cloud services and provides a good example. Other providers' standardized SLAs are similar across deployment and delivery models. The question that your organization must answer is whether the standardized SLA (which largely protects the cloud provider and sets responsibility largely with your organization) is adequate for your deployment.

NEGOTIATING THE CLOUD SLA

Whether your IT department is providing services to your staff or you are purchasing cloud services from someone else, there are 10 steps that you should follow to create a relevant SLA:

1. Understand roles and responsibilities.
2. Evaluate data policies and business policies.
3. Understand service and deployment model differences.
4. Identify critical performance objectives.
5. Evaluate security and privacy requirements.
6. Identify service management requirements.
7. Prepare for service failure management.
8. Understand the disaster recovery plan.
9. Develop an effective management process.
10. Understand the exit process.[5]

Each of these steps is described next, along with its effect on the organization.

Step 1: Understand Roles and Responsibilities

It is necessary to understand whether cloud services are being provided to internal and/or external stakeholders. Internal stakeholders are individuals who are directly employed by your organization, whereas external stakeholders are individuals or organizations that provide or obtain services/goods to/from you. They can be categorized as cloud consumers or cloud providers.[6] These two types of entities are the major actors in the cloud-computing SLA, but there are others, such as the following:

- **Cloud carrier.** The intermediary that provides connectivity between the cloud consumer and the cloud provider.
- **Cloud broker.** When present, the organization that manages the use, performance, and delivery of cloud services. Accenture is an example of a cloud broker; it provides certification of and access to cloud providers, along with service integration through a single pane of glass interface.[7] This entity is sometimes referred to as a *managed service provider*.
- **Cloud auditor.** A third party employed to conduct individual assessments of the cloud services used to audit the performance of cloud offerings.

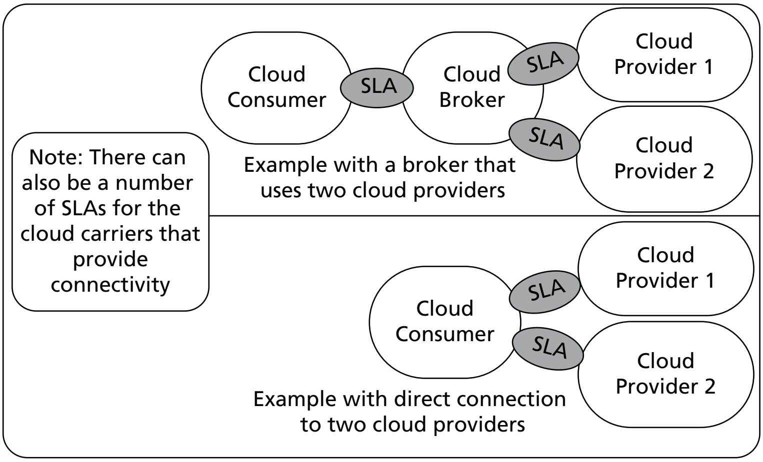


Figure 6.2 The Cloud Ecosystem and Its SLAs

It is important that the roles of all entities in an SLA be determined. Various relationships exist among the entities, and there could be a number of SLAs. These create a set of direct and indirect relationships within a cloud ecosystem that must be understood when you consume cloud services—see Figure 6.2 for examples. Your organization's ability to manage expectations among the stakeholders will depend on the SLA.

Step 2: Evaluate Data Policies and Business Policies

Two types of SLA policies are necessary for managing cloud arrangements: data and business. Some of these policies may not be germane to your specific cloud ecosystem, but they are representative of the following issues that should be considered:

SLA Data Policies

- **Data preservation and redundancy.** How any data in the cloud will be protected and maintained.
- **Data location.** Where data are stored and how that storage location is made user-accessible; when data are moved, how your organization can verify data consistency. It may be illegal to have some data located in a different country; that could necessitate compliance with another country's laws (e.g., the European Union's privacy laws).

- **Data seizure.** Data protection from illegal search and seizure; access to data if the cloud provider goes out of business.
- **Data privacy.** Data protection in compliance with the appropriate laws to protect consumer privacy.

SLA Business Policies

- **Guarantees.** Operational, service, and business objectives or measures with associated penalties when they are not met.
- **Acceptable use policy.** Ways in which the services of the cloud provider can be used.
- **List of services not covered.** The services covered by the SLA must be listed, along with the services offered by the cloud provider but not covered (i.e., the services contracted for must be specifically listed).
- **Excess usage.** The amount of allowed usage and what occurs when that is exceeded (e.g., extra cost).
- **Activation.** How a service is activated or triggered and when metered use charges are incurred (e.g., when an API is triggered or a user logs in).
- **Payment and penalty models.** How any payment is made and how you get credit when the service is unavailable or does not meet the agreed-upon objectives.
- **Governance or versioning.** How services changes are made by the provider, what your involvement in those decisions is, and how you are informed of service changes so your organization can adapt to them.
- **Renewals.** How the contract is renewed so that both parties can negotiate better or more appropriate terms.
- **Transferability.** If your organization sells its business, or for some other reason, how the current contract can be transferred.
- **Support.** How support is accessed, how long it takes to respond to support requests, how support calls are prioritized, what is provided, and what it costs.
- **Planned maintenance.** When the provider's services will be unavailable because of maintenance and whether that

downtime will count against any service guarantees (e.g., 99 percent uptime).

- **Subcontracted services.** Details about subcontracted services (e.g., whether allowed at all or the details of an allowed subcontract). If the cloud provider is to act as a broker (i.e., use upstream providers), that should be specified along with the appropriate SLA guarantees.
- **Licensed software.** If a third-party license is required (e.g., a Windows server license), that should be specified, along with who is responsible to purchase and patch that software.
- **Industry-specific standards.** Especially in regulated industries like health care and financial services, but also in general, industry standards and the responsibilities that lie with your organization and the cloud provider should be detailed.
- **Additional terms for different geographic regions.** Any change of SLA terms caused by changes in geographic region or country.

Step 3: Understand Service and Deployment Model Differences

As discussed in Part One, cloud services are delivered through service models. These service models have effects that must be considered, as follows:

- IaaS delivers an infrastructure on which consumers can deploy their own applications and content. It provides relatively low-level services, each with a specific set of metrics, including compute, network, and storage. These metrics will affect many areas of the SLA.
- PaaS delivers a specific platform (e.g., Microsoft Azure) that can be used to develop applications and deliver content. PaaS metrics are not well standardized and vary among cloud PaaS providers. Care should be taken to review the available metrics, and negotiate/incorporate them into your cloud SLA.

- SaaS delivers a software solution (e.g., Salesforce). There is a wide range of types of services and SLA guarantees offered by SaaS providers. You should expect guarantees like service availability (e.g., uptime expressed as a percentage) and application response time. There should also be language that outlines how data can be moved to another provider, if necessary.

Cloud deployment models also have certain SLA requirements:

- The private cloud delivers services to your organization only, regardless of whether the cloud is on-premises or off-premises. The SLA requirements are similar to the infrastructure being provided by your own IT department.
- The public cloud consists of services being provided on hardware shared by multiple organizations. The SLA should be carefully reviewed to ensure that adequate measures are taken to protect the cloud provider's perimeter (e.g., firewalls) and communication links (e.g., leased lines). Security, reliability, and performance risks posed by the multi-tenancy arrangement should also be addressed.
- The hybrid cloud provides for a combination of public and/or private clouds under one deployment model. The SLA requirements are similar to the private and public clouds, with the additional concern of integration between the various cloud models and providers. The SLA should cover service, data, and business integration.
- The community cloud provides services to an affinity group. SLAs must deal with the specific needs of the community that is being served (e.g., a health-care community cloud in the United States must comply with HIPAA).

Step 4: Identify Critical Performance Objectives

It is essential that your organization understand what it wants to accomplish with a cloud deployment and what SLA objectives it needs. An organization might initially consider availability, response time, and other similar parameters, and it is true that these must be included. However, the parameters that will most affect your business competitive positioning must be considered first. The business processes that are core to your organization, and their need for IT services, must drive the parameters of the SLA.[8] These can be understood by reviewing the business processes of your organization and determining how a cloud environment could affect them.

The common performance objectives of availability and response time and their effects on competitive positioning might be as follows:

- **Availability** is the percentage of uptime for a service. Service availability's effect on customer retention is straightforward, considering the number of providers for virtually every technology; if a service is not available consistently, the customers may move to another provider. That is true even for the megaconglomerate Google; if it could not consistently provide search services, it would not be a market leader for very long.
- **Response time** is the elapsed time between a service request and the completion of that request. If customers experience slow response times after requesting services from a public cloud they are likely to move to another provider; in the case of an on-premises cloud they might build shadow IT to solve their business problems.

These are just examples, but what is important is that your organization consider its business processes, the consequences of deploying those processes in a cloud ecosystem, and the effect of that ecosystem on competitive positioning.

Step 5: Evaluate Security and Privacy Requirements

Security in the cloud is basically the same as in any other IT infrastructure. Security involves protecting digital assets (e.g., data, content, applications). When digital assets are implemented in an external cloud, there is an additional risk because your organization is depending on the cloud provider to deliver the required security.

The harm your organization would suffer if your data, content, or applications were compromised must be considered. This must drive the requirements (e.g., type of security, actions in the case of a breach, steps to detect data compromise) imposed on the cloud provider and your organization by an SLA.

It is necessary to consider any regulatory requirements when you negotiate security-related SLA parameters. For example, if your organization is a university that receives funds for an applicable program of the U.S. Department of Education, you are required to comply with the Family Educational Rights and Privacy Act (FERPA), which protects the privacy of students' educational records. Ellucian is a company that provides educational management solutions in the cloud and must meet the FERPA guidelines. Your organization must consider regulatory compliance appropriate to your industry.

Step 6: Identify Service Management Requirements

Cloud services must be managed in order to meet the organization's goals. At a minimum, the SLA should incorporate parameters that ensure these management requirements:

- Cloud services that are auditable so your organization can be sure it is meeting the SLA objectives.
- Clear and concise reporting of metered use, performance, and problem notification.
- Change management, provisioning of new services, and appropriate testing.
- Upgrade and licensing responsibilities.

This list is representative of the management requirements that should be included in an SLA to ensure that your organization can reach its stated goals, meet the needs of your customers, and maintain its competitive position.

Step 7: Prepare for Service Failure Management

Sometimes services are not, or cannot be, delivered as promised in an SLA. In order to avoid misunderstandings and to detail what occurs when services are not delivered as promised, it is necessary to outline the remedies up front in the SLA. It is also common to have exclusions in an SLA in order to limit cloud provider liability in certain situations (e.g., scheduled or emergency outages, acts of force majeure, or Internet access issues outside the cloud provider's control). Recognizing service failure occurs in one of two ways: (1) your organization sees the failure because of its own monitoring, or (2) the cloud provider monitors and reports it.

Step 8: Understand the Disaster Recovery Plan

In order to ensure business continuity, your SLA must codify disaster recovery procedures. Deploying to the cloud does not relieve your organization from having to implement disaster recovery locally. The cloud provider may take care of infrastructure and platform issues, but your organization is responsible for data and content.

Your organization, not the cloud provider, will generally interface with the customers. The organization's disaster recovery procedures must specify what will occur in the event of a disaster and a related service outage. That includes how data and services are restored and how customers are dealt with. All of that must be part of the SLA so that you and the cloud provider understand how to proceed if a disaster occurs.

Step 9: Develop an Effective Management Process

The SLA relationship must be managed. The SLA is not an agreement that you can execute and then just assume that services will be delivered as expected. The metrics that monitor your organization's objectives and service requirements must be reviewed periodically. The status of those objectives should be discussed at periodic meetings (perhaps monthly or more frequently for mission-critical objectives), and actions should be taken if those objective or service requirements are not met. That way, the effect on your stakeholders can be realized, and any necessary corrective actions can be taken.

You should consider having the cloud provider be part of all status meetings, since it is essential to understanding the SLA parameters and metrics and to developing appropriate responses to unmet metrics. The cloud provider can make its internal resources available (e.g., elevating a problem to a higher support team).

Step 10: Understand the Exit Process

Every SLA should have an exit clause detailing the responsibilities of the cloud provider and your organization at the termination of the service (whether that is prematurely or at the end of the agreement). This exit clause must include the following procedures necessary to ensure business continuity for your organization:

- Transmission and preservation of data and services held by the cloud provider.
- Format for data transmitted to your organization (e.g., digital or hardcopy).
- Associated fees for termination and data transmission.
- Length of time the cloud provider will maintain your organization's data after termination.
- Business continuity options during transition of services and data to a new cloud provider.

SLA Expectations

The SLA must be written to ensure that the expectations of your organization's IT department, your operational staff, and the cloud provider are understood and will be met. A small organization may be forced into a standardized SLA, but whenever a negotiated SLA is possible it should be created from the perspective that all parties are on the same team. Whether the SLA is standardized or negotiated, the steps given above should be followed to increase the likelihood of success.

CONCLUSION

A strong business case is needed before committing to a cloud computing project, and it must be supported by an equally strong SLA to ensure that your organization gets acceptable returns (e.g., ROI, customer satisfaction, business continuity). In most cases, a standardized SLA will be presented to your organization. The standardized SLA should be reviewed so your organization understands its responsibilities and those of the cloud provider. When a negotiated SLA is executed, your organization should closely follow the 10 steps detailed earlier. The needs of all the stakeholders will influence the decisions about the type of cloud deployment, service types, and SLA parameters. Once an SLA is executed, your organization must monitor the agreed-to performance metrics. In that way, ROI can be maximized.

Note: Cloud services offered via a negotiated SLA may cost more than a standardized SLA. You should compare the additional costs and consider how they will affect ROI.

NOTES

1. Christopher Luise, Adnet Technologies, personal communication, June 2013.
2. Daniel Putka Manoel Camillo Penna, and Vinicisu Prodocimo, "Service Level Management in ATM Networks," Latin American Network Operation and Management Symposium, Quito Ecuador, October 10–11, 2011, www.lanoms.org/2005/anaiscd/1999/5-3.pdf.
3. This section is meant to be a business commentary, not a legal opinion.
4. The Digital Millennium Copyright Act is a federal law that implements various treaties of the World Intellectual Property Organization. The law criminalizes certain acts of production and dissemination of technology, devices, or services. It also makes it illegal to circumvent access controls implemented via digital rights management software.
5. This list represents a synthesis of the steps outlined in Cloud Standards Wiki, "Practical Guide to Cloud Service Level Agreements," 2012.
6. National Institute of Standards and Technology, Cloud Computing, Reference Architecture, www.nist.gov/itl/cloud/refarch.cfm.
7. Pete Swabey, "Accenture's 'Cloud Broker' Bid," *Information Age*, April 29, 2013, http://www.information-age.com/it-management/outsourcing-and-supplier-management/123457010/accenture-s-cloud-broker-bid.
8. Gendron, *Business intelligence Applied*.

PART
THREE

Business Intelligence and the Cloud

In Parts One and Two, the foundation was laid for understanding how to extract the maximum benefit from cloud computing. Part One provided a brief historical review and extensive definitions. Part Two covered the economic and strategic concepts necessary to understand cloud computing and how business intelligence (BI) fits. Now we move on to discussing BI and how it is enhanced by cloud computing. This includes reviewing the topics of Big Data and mobile computing.

Gartner's IT glossary defines *business intelligence* as "an umbrella term that includes the applications, infrastructure and tools, and best practices that enable access to and analysis of information to improve and optimize decisions and performance."[1] This glossary defines *business analytics* as "solutions used to build analysis models and simulations to create scenarios, understand realities and predict future states. Business analytics includes data mining, predictive analytics, applied analytics and statistics, and is delivered as an application suitable for a business user."[2]

These definitions will be incorporated into cloud strategy in Part Three as we deal with these issues:

- What is business intelligence, what is business analytics, and what is the difference?
- How does Big Data affect BI?
- What is the intersection of BI and mobile computing?

NOTES

1. Gartner, "Business Intelligence," IT Glossary, www.gartner.com/it-glossary/business-intelligence-bi.
2. Gartner, "Business Analytics," IT Glossary, www.gartner.com/it-glossary/business-analytics.

CHAPTER 7

Business Intelligence

The Interaction of Business Intelligence and Cloud Computing

Several months ago, the author was discussing BI at a conference when another attendee asked, "Can you use those two terms together—*business* and *intelligence*?" Not being one to be outdone, the author quickly answered, "You cannot have one without the other, at least if you want to be successful." As you might imagine, that led to a long discussion of what BI is (or should be). The author continued to interject the notion that cloud computing must be at the core of BI today and that having BI that is not enabled by cloud resources is sort of like eating apple pie without vanilla ice cream: You are not getting the most out of the experience (at least, not from the author's point of view!). BI needs the cloud to be most successful. This chapter explains why.

Enterprise applications, such as enterprise resource planning (ERP) and customer relationship management (CRM), have been rapidly advancing to include analytics that create BI. Our understanding of descriptive and predictive analysis has been expanded to include cloud-based solutions. What for decades was known as *decision support* has grown to become *business analytics* and *business intelligence* and has become the cornerstone of business strategy. We once primarily discussed the analysis of *databases* and *data warehouses*, but today we discuss *Big Data*. There has been a surge in tools and techniques that have created the new discipline of *data analytics*, with *data scientists* at the center of enabling data-driven knowledge and strategy.

This has been fed by a new level of interest in finding actionable and competitive information through the combination of an organization's internal data with data commoditized by cloud data providers. Online analytical processing (OLAP) solutions that allow inquiry in real time rather than the periodic generation of reports are the norm. OLAP provides instantaneous access to data rather than the hindsight view of organizational performance provided by reports. These rapid changes in the technology have created a number of terms, concepts, and techniques that have to be understood to get the most out of the data. This chapter makes sense of these terms, concepts, and techniques so that actionable information that supports managerial decisions and competitive strategy can be created.

In the past decade the amount of data available to support decision making has exploded. We have all heard about this explosion, but what do we do about it? Organizations often do not have the ability to analyze it efficiently in-house. That is where cloud-based analytic solutions providers and cloud data providers come into the picture. These will be discussed as we move through the analytics cycle. Just like any innovation that is first implemented in the cloud to minimize up-front organizational risk, BI can also be implemented using cloud providers.

BI STRATEGY

Organizations use the term *business intelligence* quite often; in fact, organizations are often abuzz with it. But have they carefully thought about what BI really is or why it is important? Basically, BI is something you create by asking the right questions of the right data using the right analytical tools. This requires knowing where your organization is today, having a solid mission-driven BI plan, and using the BI that is created to improve your organization.

Successful BI requires that it be driven by your organization's strategy. The path to successfully creating BI starts with knowing the questions you want to answer, then understanding how data can be used to answer those questions. The type of questions and desired outcomes are based on your type of organization: A for-profit organization may want to enhance customer service to increase market share, whereas a government agency may want to better manage its budget to improve citizen perception of the agency. The organization's mission, vision, and objectives are of central importance.

Before undertaking a BI project, your organization must ask how well it is currently using analytics to create BI and how well those efforts are aligned with the organization's strategy. If your data are stored in spreadsheets or on hard copy, you may find creating BI difficult. If you are using a database-driven application (e.g., ERP or CRM), it may be a matter of determining what questions you want to ask and what you are going to do with the answers. Successful BI depends on the alignment of your organization's mission, strategy, and objectives with its people, processes, and technology.

BI must take into account the internal value-producing activities of your organization (see Figure 7.1) and the external forces on your organization (see Figure 7.2).[1] Taking stock in the internal valuing producing activities of your organization increases the likelihood that a BI strategy permeates the organization at all levels. Understanding external forces allows you to ask the right questions and share the

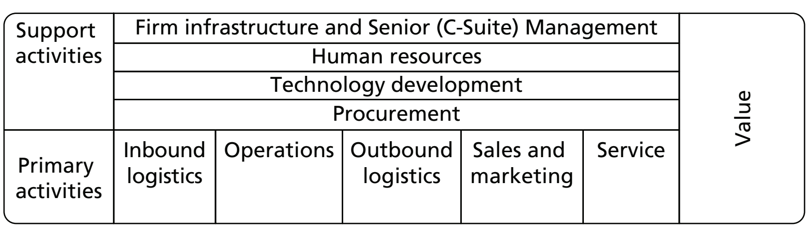


Figure 7.1 Internal Value-Producing Activities Adapted from M. E. Porter, *The Competitive Advantage of Nations* (New York: Free Press, 1990).

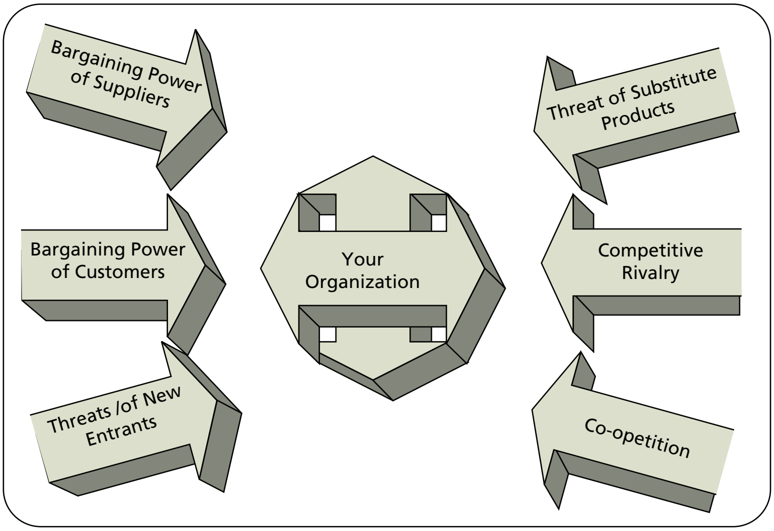


Figure 7.2 External Forces The first five forces were identified by Porter, *The Competitive Advantage of Nations*. The sixth force was identified by D. R. Gnyawali and B. J. Park, "Co-Opetition and Technology Innovations in Small and Medium-Sized Enterprises: A Multilevel Conceptual Model," *Journal of Small Business Management* 47, no. 3 (2009): 308–30.

results with the correct parties so the BI project is based on the correct strategic goals and objectives.

Like any technology project, developing a BI solution involves risk for the organization, which can be in the form of initial resource outlay; effect on customers, suppliers, or staff; or something else. It is important that you know the consequences, benefits, and objectives before taking that risk.

OBJECTIVES FOR A BI PROJECT

The BI project must be planned and based on business and IT strategy. It must also have specific objectives, such as to improve organizational performance, lower cost, produce superior products, or be more competitive in the marketplace. This can be done through analytics like descriptive reports or with more elaborate schemes that include data mining and executive dashboards.

Objectives are often crafted based on the skill set of the existing staff. This might be a mistake. Standardizing objectives and solutions, then hiring the necessary staff or consultants or buying the needed solution, may be the best way to achieve the desired results. According to Ed Laprade, the CEO of ADNET Technologies, the debate of standardization can "sometimes almost be a religious one."[2] These thorny issues must be considered as objectives are set.

For more than 100 years, Graniterock—a Watsonville, California, provider of stone and quarried materials—has held stellar customer service as its biggest competitive differentiator.[3] To further enhance its operations, Graniterock decided to implement a new Radio Frequency Identification (RFID) system with tags and readers. As the RFID infrastructure was being implemented, it became apparent that the existing BI solution had to be retooled because it was not suited to analyze the amount of data that would be generated by the new RFID system. The result was that a new BI infrastructure was created, with the objective of enhancing customer service by allowing customers direct access. Graniterock had to deal with the following issues:

- An existing BI infrastructure that was not scalable or robust enough to handle the new data from the RFID system.
- Increased costs because of the new BI infrastructure.
- Increased data from the RFID solution.
- Whether the BI solution could create ROI (often it does not, at least not directly).
- The effect of the revamping of the BI solution on staff and customers.
- Whether it would be possible to build a new BI solution that would support superior customer service

The Graniterock case highlights some important issues. It would have been better if Graniterock had understood these issues before it embarked on a BI project. A more thorough analysis of the company's existing systems before implementing the RFID solution would have made the entire implementation easier. Graniterock illustrates why it is important to create objectives before embarking on a BI project.

Graniterock's initial BI solution had grown organically alongside its ERP, whereas the new BI solution was an outgrowth of its RFID project. During the RFID project, the Graniterock staff quickly developed the objective of enhancing customer service through the new BI solution, and it was an important guide in the continued resource expenditure. Graniterock employees said it best: "From a customer-service standpoint, more and more we're realizing that we have all of this data that we can get in our customers' hands." That data fed into the BI project objectives. The moral is clear: A stated objective should guide all BI projects and should be decided *before* the sources of data are identified, not along the way.

THE ANALYTICS CYCLE

An analytics cycle describes how we collect, analyze, and distribute/disseminate BI. A well-implemented analytics cycle creates BI that acts as a feedback loop, thus improving the cycle (see Figure 7.3). Each section of the analytics cycle is described later, but first data types will be reviewed.

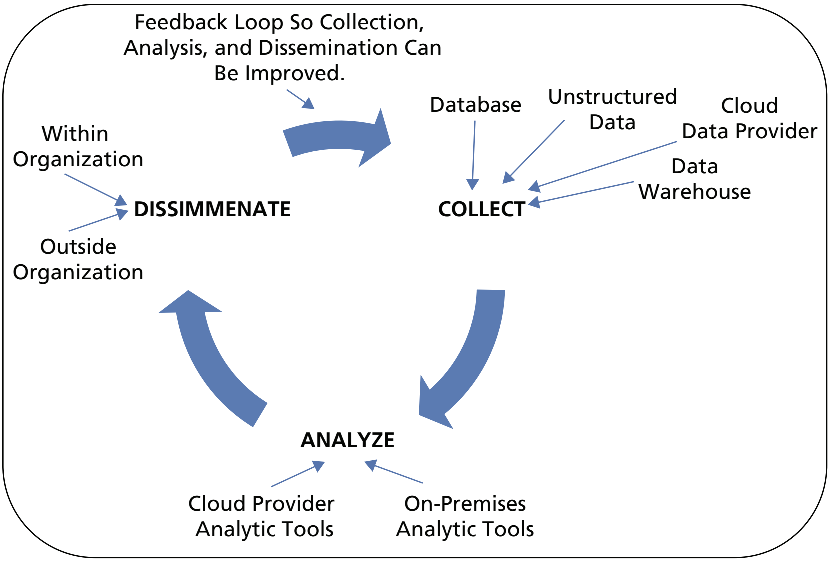


Figure 7.3 Analytics Cycle

Structured Data Example

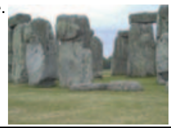
XML, which has a defined structure. This is a sample of XML from the menu system of author's website at www.gendron.info.

```
<?xml version="1.0" encoding="utf-8" ?>
<siteMap xmlns="http://schemas.microsoft.com/AspNet/SiteMap-File-1.0" >
 <siteMapNode url="http://www.gendron.info/default.aspx" title="Home"  description="gendron.info">
  <siteMapNode url="~/BDDC/default.aspx" title="BDDC Textbook"  description="BDDC Textbook">
     <siteMapNode url="" title="Get Involved"  description="Join the Textbook Mailing List">
       <siteMapNode url="~/BDDC/Submit.aspx" title="Submit"  description="Submit Materials />
```

Unstructured Data Example

Email, which is freeform decided by the creator of the email.
This is a message to this book's reader.

Dear Reader: This book provides a wealth of information about cloud compting, business intelligence and strategy. I hope you enjoy it. Also, I like this picture of Stonehenge.



Regards,
Michael Gendron

Figure 7.4 Examples of Structured and Unstructured Data

Organizations hold everything from sales transactions to institutional memory as data, which can be divided into two data types (see Figure 7.4):

1. **Unstructured.** This is a data set that does not have a predefined data model and does not fit well into a relational table. It is usually text-heavy, and text-mining tools are often used to analyze it. It may contain discrete items like numbers, dates, and other facts, but not in a predefined format. The ambiguous nature of unstructured data make them difficult to store and analyze. A large proportion of available data is unstructured. E-mail is a good example of unstructured data.
2. **Structured.** This is data set whose format is predefined and that potentially has a formal data model. Real-world items like product numbers and descriptions, suppliers' addresses, customer information and orders, and even the menu system on a web site are created digitally in predefined formats.

Data Collection

Organizations are faced with an enormous amount of data from internal and external sources. It is important that you understand why you are collecting data and how those data support the objectives of your organization and its BI plan. As part of a BI plan, the first step in data collection is to develop a set of objectives (e.g., what you want to accomplish or the BI you want to create) so you can decide which data will support those objectives. Once that is done, internal and external data sources can be reviewed, and their fit into your objectives can be understood.

Internal Sources of Data

Information exists in many forms. Some organizations still have filing cabinets filled with orders and other information, while other organizations have adopted the use of spreadsheets. These forms of information are difficult to manipulate into usable BI, but with effort it is possible. Organizations commonly have data stored in online transaction processing (OLTP) systems that record accounting, inventory, financial, human resources, and other information. An OLTP often operates on an SQL database that is optimized to support the relationships among the data (e.g., with a common key like a social security number that allows records for the same person to be related).

These data are stored in databases and represent a wealth of information that can be used to create BI. These are the structured data of the organization; they are often the input to BI applications and can be effectively analyzed. Here are some important concepts involved in storing structured internal data:

- **Data model.** Specifies how data are stored and the relationships among those data. A data model is often referred to as a *schema*. An example is the schema that describes how airline reservation data is to be structured and stored.
- **Database.** A repository of data typically organized according to a data model and structured in such a way that it can be queried; for example, a database of all airline reservations that can be queried to determine seat availability.
- **Data management system.** The system's software that allows users and applications to interact with a database (e.g., MY-SQL, MS-SQL).

- **Data warehouse.** Views of data that are often integrated across various databases within an organization and are used for reporting and analysis; for example, an airline reservation database that is combined with a food service database allowing the analysis of trends in food consumption by flight patterns.
- **Online transaction processing (OLTP) systems.** A type of information system that facilitates the processing of transactions immediately and interactively (e.g., sales orders or inventory) and updates the appropriate databases (e.g., www.amazon.com provides a front end to its OLTP that processes sales transaction).

Figure 7.5 shows the interaction of these data components.

It would be great if all information was structured and could be described by a data model. The fact is that most information is unstructured and large. Just think of the number of e-mails, memos, spreadsheets, LinkedIn pages, and Tweets generated worldwide on a daily basis. These hold valuable information but do not have the formatting requirement of a data model. These unstructured data are hard to store and even harder to analyze. These looser consistency data are larger than the structured data your organization

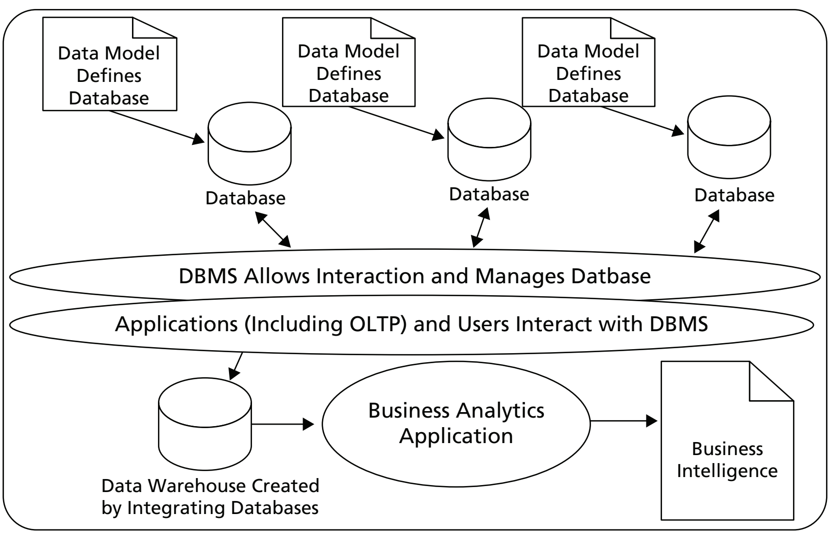


Figure 7.5 Interaction of Data Components

holds, but they can yield a wealth of information, if stored and analyzed properly.

One solution is to store these data in NoSQL ("not only" SQL) databases. NoSQL allows structured and unstructured data to be stored in the same database. The NoSQL database is optimized for record storage and retrieval, whereas the traditional SQL database is optimized to support relationships among data. NoSQL is designed to support large amounts of both structured and unstructured data.

External Sources of Data

Many cloud providers give access to their data via an application program interface (API). These data can be combined with your own internal data or other cloud provider data to allow augmented analysis. The combination of internal and external data creates an explosion of available data for analysis. For example, let's assume your internal human resources database has people's names and addresses, but you want to get more social information about your employees; you could use a LinkedIn or Facebook API to get the data that the employee have made available on those platforms. The idea of extended analysis by combining data from different providers is part of the Big Data phenomenon. There are a number of cloud providers that allow API access to data that they store and make available as the data owner allows, some examples are:

- **LinkedIn, Facebook, Twitter, and Foursquare.** These social media applications provide API access to data that they store, including location dependent data.
- **Salesforce.** A suite of CRM applications provide access to your organization's customer data.
- **Google Apps.** This desktop productivity application provider has calendar, word processing, spreadsheet, and presentation functions.

There are also a number of providers that commoditize their own data and make them available via an API; some examples are:

- **InfoGroup.** This source of business demographic information provides API searches of current local businesses (e.g., restaurants, shops, entertainment, sports, beauty parlors, and fitness centers). InfoGroup provides other companies (e.g., Topix and Google) with business details to be used in their services.

- **National Aeronautics and Space Administration (NASA).** NASA provides API access to data such as global rainfall measurements and solar dynamics.
- **PolicyMap.** This provider has collected and commoditized geographic information system data along with about 15,000 indicators related to demographics, real estate, city crime rates, health, schools, and more. These data are made available via an API to organizations that want to incorporate PolicyMap data with other data and create custom analytics. A group called Citi Community Development created "a custom web-based application that provides on-the-fly congressional level statistics. . . . Citi Community Development and PolicyMap worked together to create a new interactive, web-based tool—My District Data—that forges a critical link between federal decisions and local accountability."[4]

There are numerous internal and external sources of data. It is important that your organization know its objectives for a BI project in order to be able to select the appropriate data. When a cloud data provider is used, the model changes as shown in Figure 7.6.

Data Analysis

With BI objectives in place and data sources selected, you can now decide how you are going to analyze the data. You can store all the data in-house and run the analysis there, or you can send the selected data to a cloud provider and run the analysis in-house or at a cloud provider; it's a matter of meeting your organization's needs and objectives (see Figure 7.7). The choice of where applications and data will reside requires both technical and business decisions:

- Data that have specific requirements (e.g., data governed by European Union privacy laws) may need to be stored on-premises or in a specific country.
- The available technology may dictate that analysis be done off-premises because of the amount of computing power required.

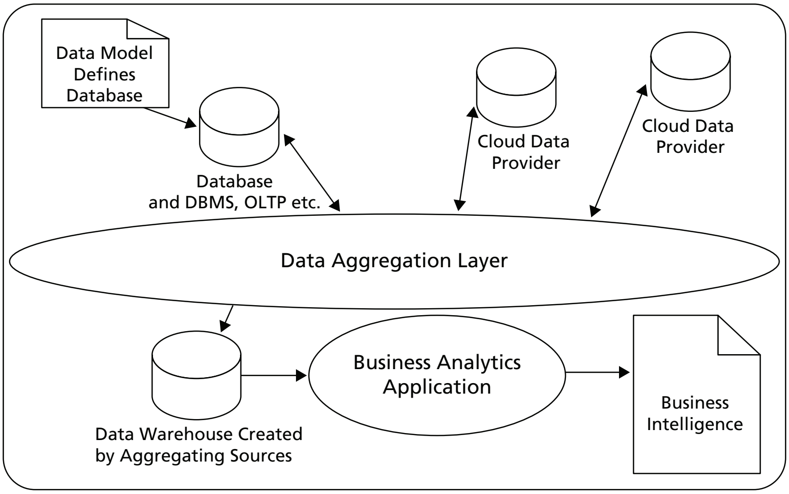


Figure 7.6 Cloud Data Provider BI

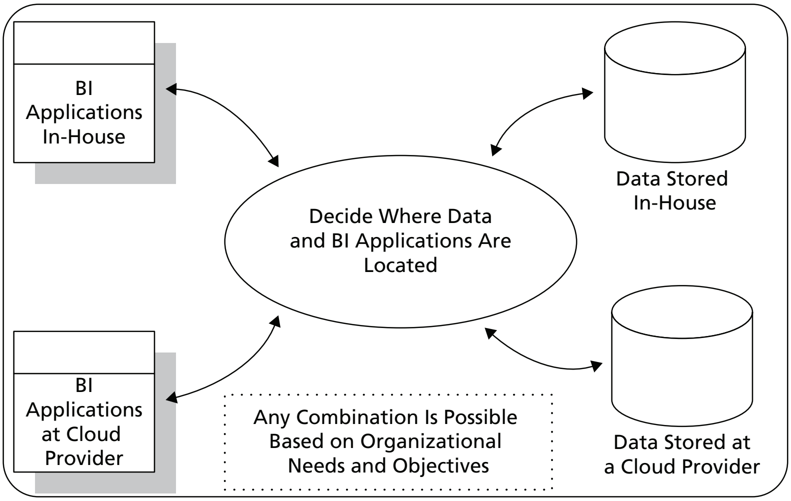


Figure 7.7 BI Data and Application Location

- Data may be purchased from a cloud provider and accessed via an API, necessitating a certain type of analysis.
- Your organization may prefer an approach that lowers the financial risk (i.e., lower upfront resource outlay or metered use) by deploying data and/or applications to a cloud provider.

The analysis to create BI is limited only by the innovation, data, background, and statistical knowledge of the analyst. However, there are several major types of analysis that should be considered for the development of a BI solution. These can be used to create tools that support decision making (i.e., decision support tools or decision support systems). The types of analysis are as follows:

- **Descriptive reporting.** Basic data queries that may involve tabular reports with averages, ratios, and totals. Normally a hindsight approach to analyzing historical data. An example would be a monthly sales report by region.
- **Online analytical processing (OLAP).** The ability to quickly analyze complex data interactively to perform on-demand data modeling. OLAP provides these tools:
  - **Data cube.** A multidimensional arrangement of data that can be used to produce summaries, means, and other statistics across dimensions. A data cube can have any number of dimensions. Each cell in the cube represents the computed statistic. OLAP arranges the data along the specified dimensions (e.g., sales data by sales region, product type, and fiscal quarter; see Figure 7.8, which shows these three dimensions). Statistical packages like SPSS and SAS, along with many other software solutions, support data cube analysis.
  - **Slicing and dicing.** The analysis of data at any specific point in the data cube across any dimensions.
- **Data mining.** Discovering hidden relationships and using them to build predictive models. Data mining complements OLAP. For example, suppose that the sale of disposable baby diapers could predict the sale of beer. Why would that be true, and what should a grocery store do about it? In fact, at least one study shows this relationship to be true; it has been hypothesized that when fathers go out to buy diapers (perhaps on a late-night run), they also pick up beer. That suggests an important question: Should the diapers be moved closer to the beer to make access easier, or should they be moved far apart to force the customers to walk past other products they might buy? Data mining can discover the "beer and diapers" types

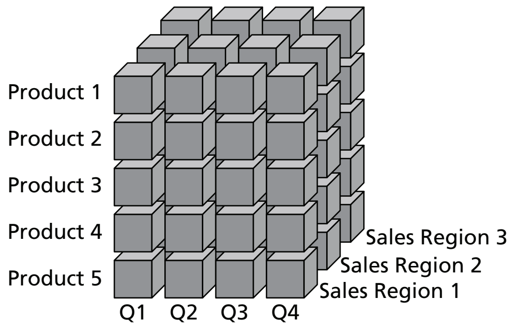


Figure 7.8 Data Cube Example

of relationships. Here are some basic ways that data mining does this:

- **Association discovery.** The connection between diapers and beer is an example of association discovery. An association between these two products has appeared frequently when grocery stores have done market basket analysis for a number of sales.
- **Sequence discovery.** Used to show relationships over time. Suppose your grocery store has a customer loyalty program that requires you to use an ID card to get discounts. Whenever you shop, you are tying each purchase you make to your ID card. Suppose you buy diapers one day and beer four days later. The grocery store now has a buying sequence that it can analyze. If that sequence happens often enough, then perhaps the grocery store should consider how it markets diapers and beer.
- **Clustering.** Grouping records together by a similar attribute. Suppose that to obtain your grocery store ID card, you had to state your gender and income level. The grocery store might find that certain brands of diapers or beer cluster around one gender or the other and/or income levels. This would allow the store to market those products by targeting certain groups of individuals.
- **Classification**. Similar to clustering, but known beforehand. A grocery store could segment sales data by gender to

understand which gender purchases what brand of beer or diapers.

- **Unstructured data analysis.** The analysis of data without a specific data model. It is often useful to uncover customer opinions and preferences that an OLTP-driven analysis might not provide. The following tools are available for analyzing free-form data:
    - **Text mining.** Techniques to extract information from text documents such as letters, e-mails, customer service logs, or feeds from social media services.
        - **Natural language processing (NLP)**. NLP is embedded in many text-mining tools and allows computers to determine meaning from text or human language input. It is an analytic tool that can assess sentiment in text (e.g., Twitter or Facebook posts). NLP should be included in all BI plans that analyze text or other human language input.
    - **Web usage mining.** A way to understand the path that users take through a web site. This can assist in optimizing web site layout, menus, and searches.
- Unstructured data analysis is a large discipline with an expansive number of methods and options available, such as the following:
    - Appliances to make sense out of social, Web, and application data offered by companies such as Precog.
    - Applications and services offered by companies that perform resume scanning, such as Sovren, to assist in matching candidates to particular jobs.
    - IBM's Watson engagement advisor, which is a "combination of natural language processing, hypothesis generation & evaluation, and machine learning."[5]
- **Predictive analysis.** Using data to predict consumer behavior. Suppose you are a satellite TV provider and you want to understand your market penetration in metropolitan and rural areas so you can predict consumer behavior and your target market.

You might create a graph (i.e., based on linear regression and best fit) like the one shown in Figure 7.9. The graph suggests that the higher-income customers in metropolitan areas subscribe longer than lower-income customers in the same areas, whereas lower-income customers in rural areas subscribe for longer periods and leave the service less quickly. Your organization might determine that your marketing resources are best targeted to lower-income rural people.

- **Knowledge management.** Though not a specific mathematical or statistical technique, this is about managing knowledge generated and contained within your organization. In this type of system, knowledge is treated as an asset and is often textual information. It is useful for recall and mining. An example would be:
  - A consulting firm that requires employees to submit a summary of their experiences and solutions into a knowledge management system after each engagement so that future employees can look for solutions when they are working with new customers.
- **Information visualization.** Used to show complex representations of data so they can be more easily understood and analyzed. Some major types of visual techniques are the following:
  - **Geographic information system (GIS).** Data-driven maps used to enhance an understanding of the geography and related data. For example, a map of weather patterns and the proposed locations of cell towers might give more information about optimal locations and be more informative than a graph. Or a map that shows the locations of automated teller machines for cash withdrawals might be more informative than a list by address.
  - **Dashboard.** A display of key indicators required to make a decision. Often includes the ability to click on an indicator and get more details. For example, a manager looking at a map of sales data by geographic region might be able to click on an area that appears to be performing poorly in order to see which store is the issue.

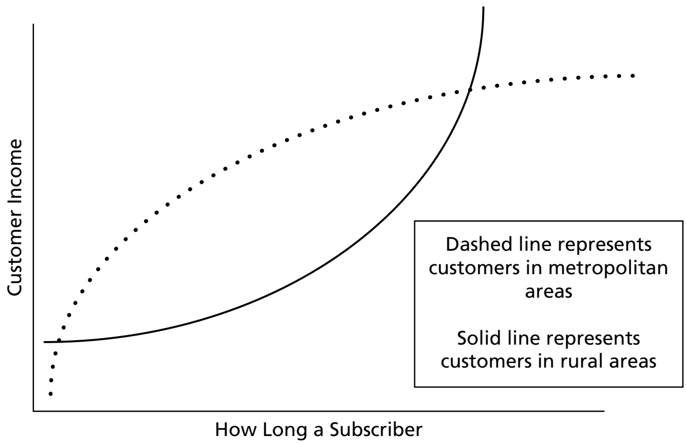


Figure 7.9 Predictive analysis Example

These are the fundamental tools used to create BI. There are a number of vendors that provide tools, including statistical software vendors like SAS and SPSS. BI tools are also built into many applications, like Salesforce. The types of tools you need depend on your BI plan. As a decision support tool, BI aids in the understanding of a large amount of data. The tools can be used by themselves or in combination to enhance decision making, improve customer service, and ultimately improve competitive positioning.

Data Distribution

As part of a BI plan, you must decide on the distribution of the intelligence. This decision must be driven by the goals and objectives of the BI project, but that is just the beginning. Not only should you disseminate the information, you must also provide a way for the users of the information to provide feedback about the quality, content, and format—this is called the *BI feedback loop*. In that way, the BI can improve over time.

Self-Service BI

Many BI tools are not very user-friendly, which can create problems when disseminating tools to end users. Self-service BI is a way to solve that problem—it gives users the tools to ask questions and do analysis on their own terms. "Self-service BI is technology that provides non-technical individuals who need data with the ability to gather and display meaningful information about the desired subject matter."[6] Before self-service BI is implemented, strategic decisions must be made and the BI plan must be created. The technical and business intricacies of cloud providers or on-premises software, identification of data sources, and such must be addressed before a self-service BI application is made available. These significant undertakings before deploying a self-service BI make it possible for the users to access complicated queries on their own terms.

The Primary and Preventative Care GIS is a public-sector self-service BI solution.[7] It allows authorized public health agencies to query health indicators from a statewide database by various geographic areas, to overlay the areas and indicators, and to incorporate local data on a thematic map. For example, if the number of patients with cancer was mapped along with the location of toxic waste sites, a relationship might become evident. If you also overlaid the number of oncologists, you could determine whether there are sufficient numbers of them in the right locations. This mapping might cause a policy shift within the state.

Self-service BI is a way to reduce the effect of BI requests on an organization's IT department. Nike was faced with this problem. According to Jimmy Lee, an architect at the company, Nike is creating self-service BI in response to the constantly changing needs of business users. With each new report "our users say, 'We have more questions that we need answered.' . . . Meanwhile, the complexity of the analysis that the business needs is growing, too. We're having to do more complicated work—more data types, deeper questions—in a shorter amount of time."[8]

Who Gets BI

The bottom line in distribution/dissemination is that BI has to get in the right hands for it to have an effect. It is not enough for BI to be created; we have all seen the big reports that are generated every month and stored in case they are needed. That is not BI. BI creates views of your organization and its customers, suppliers, processes, and staff that allow the organization to be managed better. Well-disseminated BI provides a way for consumers of the information to provide feedback so the BI system can grow more useful.

CONCLUSION

BI is a large topic with many facets that must be driven by an organization's business and IT strategies. A BI plan must be in place to obtain the maximum from the investment in a BI deployment and must be driven by strategic objectives. You must consider the internal value-producing activities of the organization and the external forces that put pressure on your organization in order to obtain maximal outcomes.

The basic analytic cycle is to collect data, analyze it, and then distribute/disseminate it. There are many decisions to make throughout this cycle, including whether to deploy in-house or use a cloud provider for software and data. The objectives of the BI project, the types of analysis, the necessary data elements, and the dissemination of the BI are all important decisions. Within each decision, many issues must be considered.

The final part of the analytic cycle, distributing the BI, is probably one of the trickiest. Who gets the BI, how they get it (self-service or delivered by your BI department), and how they can give feedback to the system are all of paramount importance to the success of a BI plan. A successful BI plan can be instrumental in driving the success of a business.

The cloud is a game changer for BI—you no longer depend on just your own organization's data stores. Data analysis takes on a whole new meaning when you can leave your data in the cloud (or use your local data store) and combine it with data from cloud providers. The cloud brings together storage, compute, and network services in ways that change the creation of BI. With providers offering software like IBM's Business Analytics in the Cloud and Pentaho's Data Integration and Business Analytics for Cloud Applications, your organization no longer needs to deploy a large expensive infrastructure in-house but can use cloud providers on a metered use basis. That allows your organization to manage the risk of BI projects. Cloud providers offer a myriad of data from social media and CRM applications to business demographics and geographic data. Cloud data and business analytics providers allow a whole new generation of BI to be developed that provides your organization with new and better ways to manage its core business and interact with its supply chain.

A final note: BI plans and systems can be resource-intensive (e.g., a manufacturing executive estimated to the author privately that its executive dashboard cost more than $5 million a year to maintain). Using a cloud approach can reduce the overall risk and some of the initial resource outlay. However, as with any project, the scope and scale of the BI project will dictate what longer-term resources are necessary.

NOTES

1. Gendron, *Business Intelligence Applied*.
2. Edward Laprade, ADNET Technologies, personal communication, July 10, 2013.
3. S. Swoyer, "BI Case Study: RFID Keeps Customer Service Stellar, ROI Strong," *Business Intelligence Journal* 10 (Fall 2005): 35–38, http://0-search.proquest.com.www.consuls.org/docview/222643146?accountid=9970.
4. Policy Map, "Client Stories," 2013, www.policymap.com/clientstories.html.
5. "IBM Watson Engagement Advisor," IBM Watson Solution, IBM, May 2013, www-03.ibm.com/innovation/us/watson/pdf/ibm_watson_engagement_brochure.pdf.
6. Ziff Davis, "Self Service Business Intelligence," white paper, 2012, http://hosteddocs.ittoolbox.com/zdwp_self_service_bi_v2.pdf.
7. Michael Gendron, "The PPCGIS: A Technical Overview of the Pilot Implementation," paper presented at the Conference of the American Public Health Association, New York, 1996.
8. Pete Swabey, "Self-Service Business Intelligence," *Information Age*, June 17, 2013, www.information-age.com/technology/information-management/123457131/self-service-business-intelligence.

CHAPTER 8

Big Data's Effects on BI Efforts in the Cloud

Big Data is a phenomenon that is created by the ubiquity of transactions and devices that generate digital information in our society today. It is not just databases or data views created from large data stores. Globally, individuals generate a multitude of tweets, LinkedIn postings, Foursquare location data, Instagram images, and many other social factoids every second of the day. Walmart's data store is estimated to be more than 2.5 petabytes—the result of keeping the details of more than a million transactions per hour.[1] Many web sites contain forums that allow purchasers to rate hotel rooms, airline flights, or other products and services. Organizations have customer service, sales, human resource, and other data in their OLTP stores. An estimated 90 trillion e-mails per year and a wealth of other memos and phone records are available for analysis. These existing sources amount to an almost overwhelming amount of information that can be stored, processed, and understood—that is, the Big Data phenomenon.

The cloud enables the coupling of data stores to create BI that can help organizations competitively. Big Data also affects the individual; for example, the individual can use existing data stores to research companies where they want to be employed or to find out about an organization's customer service record, just as the enterprise can use these sources to uncover customers' perceptions of their products. The uses of these data are limited only by the innovativeness of the person analyzing them and the resources available to do so (e.g., time, money, and computing power).

According to Gartner, Big Data exists along three dimensions (known as the three *V*s):

1. **Volume.** The amount of data available.
2. **Velocity.** The speed at which data change.
3. **Variety.** The types of data available.

Organizations are faced with complexities and costs to extract the maximum BI from Big Data; they also reap benefits by doing so. As the volume, velocity, and variety of data increase, the cost of managing the data also increases. But so does the ability to create more complex and accurate BI. The complexity of suitable analytics goes from basic descriptive statistics (e.g., mean, ratio, and sums) to more predictive ones (e.g., data mining and regression) as there are larger amounts and more varied types of data to analyze (see Figure 8.1).

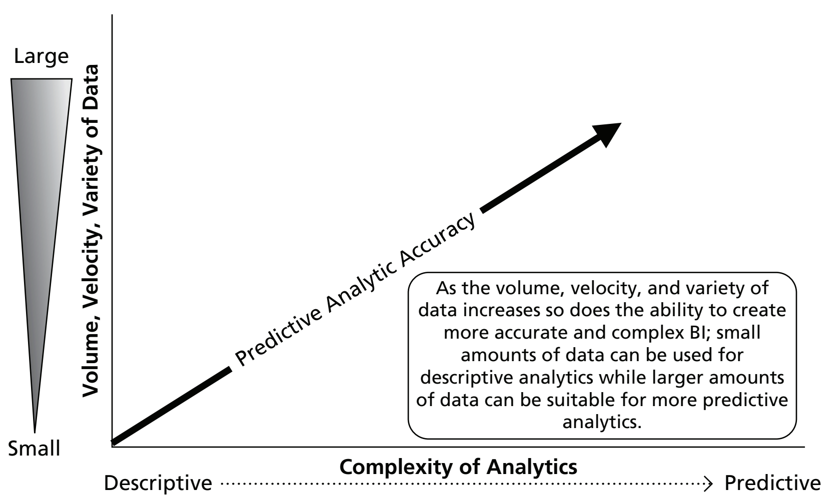


Figure 8.1 Data Characteristics and Type of Analytics

Analyzing an organization's customer service system might yield customer perceptions of the organization's products and services, but analyzing online forums or social factoids is likely to greatly enhance the organization's BI.

DEFINING BIG DATA

In recent years organizations have had to determine how they are going to respond to the Big Data paradigm to stay competitive. Never before have we had the amount or variety of data streaming into our organizations at such a speed as we do today. These characteristics of Big Data require an organization to develop and implement a BI plan, as was discussed in Chapter 7. For an organization to maintain a competitive position, it is important that sources of data, ways to analyze those data, and dissemination methods (e.g., self-service BI) are identified and put in place. This is all complicated by the Big Data phenomenon. This phenomenon goes beyond database structures, data warehouses, structured data, and unstructured data to the sheer volume, variety, and velocity of data available today.

The three dimensions (volume, velocity, and variety) combine to create Big Data (see Figure 8.2). Organizations face Big Data along any two or all three of these dimensions. Your BI plan must deal with the dimensions that occur in your data stream and put appropriate processes in place. The organization that has a high volume and high variety of data that change slowly (i.e., not high velocity) will deal with this differently from the way that an organization deals with a high volume and high variety of data with a high velocity. Each organization requires a BI plan tailored to the type of data stream it encounters. In order to help make those distinctions, the dimensions of Big Data will be discussed next.

High Volume

The amount of data that organizations create today can be somewhat overwhelming. This is driven by a number of forces. Technology has exploded in the past century, and it has become common to collect everything, from data about our business and who accesses

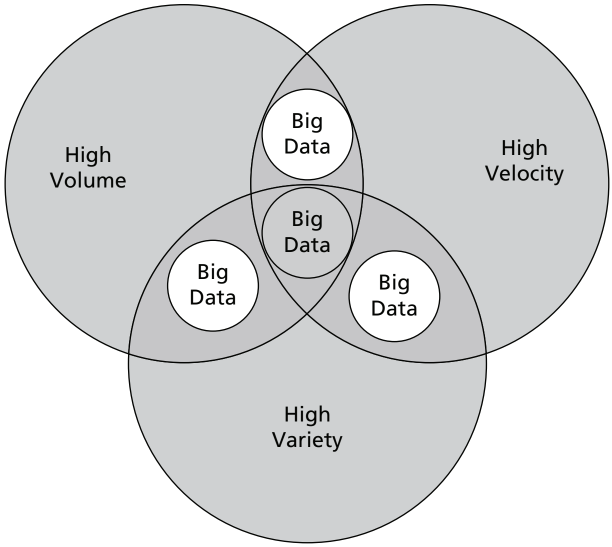


Figure 8.2 Big Data

its data stores to biometric data. Some of the driving forces are the following:

- **Legal requirements.** We have seen countries put laws into place that have the unintended consequence of mandating additional data collection. For example, laws to protect privacy may have the secondary effect of forcing organizations to collect, store, and report information on who accesses their data. These laws create the need for a completely new level of data collection and the staff to manage that data. Examples of such privacy laws are the Health Insurance Portability and Accountability Act and the Fair Credit Reporting Act in the United States; and Article 8 of the European Convention on Human Rights in Europe. These laws were created to protect citizen privacy but also require a new layer of data collection that must be managed.
- **Need to understand customer perceptions.** If a company is interested in knowing how current and potential customers view its products and services, it must be familiar with web sites that provide customer ratings. These are too many to list

here, but some examples are www.bookings.com and www.hotel.com in the hotel industry, www.travelocity.com and www.kayak.com in the travel industry, and www.amazon.com, which sells everything from books to lawn mowers.

- **Need to understand the *social* in social networking.** Besides the web sites that allow you to purchase products and services and then rate them, there are many social networking sites like Facebook, LinkedIn, and Foursquare, where people can express their opinions. Organizations are not exempt from involvement in social networks. That involvement is essential if an organization wants to understand customers' perceptions.

  Social media marketing allows companies to reach customers through online venues such as Facebook and Twitter. This marketing is different from other outlets like billboards, TV, radio, newspapers, and magazines. Online social media marketing allows two-way communication between the organization and its customers. It also adds to the Big Data phenomenon whereby organizations need to analyze customer perceptions while also being part of the ongoing conversations about their products.

  Coca-Cola provides a good example of being involved in social media with its Ambassador Program. The program sent a number of people around the world to places where Coca-Cola products are sold and then had the ambassadors' blog about the experiences. These blogs added a substantial amount of content to social networks about Coca-Cola and allowed the company to both monitor and manage perceptions about its products. Today, Coca-Cola boasts a large social media presence that includes almost 70 million "likes" on Facebook, 287,000 followers on LinkedIn, and 94,000 followers on Twitter. This impressive cloud-based social media presence was expensive to create and is expensive to maintain, but it allows Coca-Cola to be part of discussions about its products and maintain contact with its consumer base. This allows Coca-Cola to utilize social media to maintain its market share and would not be possible without cloud technologies.

- **Maximization of existing data sources.** The amount of data being generated through OLTP systems, e-mail, memos, phone records, and cloud data providers grows annually. Organizations must be aware of which sources of data are important and which should be analyzed.

  Consider the organization that monitors Amazon's Author Central so they understand how their books are selling and how they rank against other books. Author Central provides a dashboard that shows sales by geographic region, using data provided by Nielsen BookScan (the part of the Nielsen Company that tracks print book sales). Now consider what Amazon or Nielsen must do to collect sales data on all books and correlate it to customer reviews—that epitomizes Big Data.

- **New uses for technology.** New uses for technology are found every day. Grocery stores have been tracking our purchases for years with loyalty cards; at the grocery store we often use loyalty cards and scan product bar codes to determine prices. This creates a linkage between us and the products we buy, allowing the grocery store to target market. Today, some grocery stores and even other retail stores are linking those loyalty cards and purchases to social media sites like Facebook or to customers' e-mail addresses. That way the store can create linkages between the data from cloud data providers and their own data. This allows them to mine for customer preferences and social factoids.

  Many other organizations are creating data linkages. For example, when we reserve an airline flight through a carrier that uses the Sabre reservation system, the carrier has the ability to mine our reservations. Since Sabre manages about 40 percent of all flights, that means a large portion of airline reservations can be compared and linked to cloud data sources, including social media sites like Facebook.[2]

- **Effects of mobile computing.** No longer do you need to be tethered to a desktop to create data. Mobile devices allow data generation, storage, and analysis virtually anywhere and anytime. This topic is so important that Chapter 9 is devoted to it.

These are just examples of the many forces that create Big Data. The ubiquity of the Internet, mobile devices, e-commerce, and desktop computing converge with other forces and technologies to yield Big Data. There are potentially billions of people and organizations that can buy and sell books, reserve airline tickets, or post social factoids in real time, generating tremendous numbers of OLTP transactions, customer reviews, author rankings, and other information. The result is high-volume, high-variety, high-velocity data that must be collected, stored, analyzed, turned into BI, and distributed. Organizations have a large number of data sources to select from when they want to create BI, and they must do so carefully to yield an appropriate return.

High Velocity

High-velocity data means that organizations are bombarded with a stream of data at a very fast rate. This relatively new phenomenon presents particular problems. As stated on the SAS web site, "Customer transactions, internal processes, changing market conditions and burgeoning social media add to the volumes of data that organizations must access and deal with each day. Traditional approaches, which apply analytics after data is stored, may provide insights too late."[3]

Unlike never before, organizations have data flowing into their data stores from a variety of sources at an unheard-of speed, and if those data drive critical decision making, you have the Big Data high-velocity problem. Event stream processing, which analyzes data as they are received is one way to deal with this problem. It analyzes high-velocity data in real time to shorten the time for critical decision making. Some examples of event stream processing are the following:

- **Fraud detection**. Banks can scan each transaction (e.g., credit card charges or ATM and checking account withdrawals) as they are received to ensure they are legitimate.
- **Mining of social media for customers' perceptions.** Organizations can perform real-time, on-demand analysis of social factoids as they are posted in order to understand

customer perceptions. The author has experienced this personally: when he posted his perceptions about a recent mortgage closing to Facebook, the mortgage company contacted him within 20 minutes.

- **Personalized marketing.** It is possible to target a market based on customers' web site searches. Your organization could present targeted offers to customers as they search the web for products or services.

There have been many technical methods proposed to handle the high-velocity Big Data problem. One method is Apache Hadoop. Hadoop is an open-source framework that was derived from Google's MapReduce and Google's File System. Hadoop was designed to support running applications on large clusters of commodity hardware in a parallel computing environment. It is designed to manage lots of data, both structured and unstructured, and can be deployed on cloud provider platforms such as Amazon Web Services, Microsoft Azure, and Heroku. Hadoop deployment originally required the local coding of software to interface with data. However, some vendors have adapted their tools to support the Hadoop platform—Informatica, a data integration and services company, is a good example. These tools allow Hadoop to be integrated into the rest of the organization's data management infrastructure so that Hadoop does not become another data silo in the organization. Other software vendors like IBM, Oracle, and Microsoft, offer connectors or have retooled their software to interact with Hadoop.

Hadoop is not the only solution to the Big Data high-velocity problem. Some other examples are SAS, DataFlux, Oracle Fast Data Resources, and IBM InfoSphere. These solutions offer software engines that give the ability to analyze real-time data streaming into your organization. These vendors offer their own data management solutions as well as Hadoop connectors. The issue for your organization is to determine the best way to manage high-velocity Big Data as a competitive asset that can be turned into BI.

High Variety

More varied data types and sources are available today than ever before. As data variety expanded, our understanding of how to use those data has also expanded. Early decision support systems created reports, dashboards, and other visual representations of data from our own systems. Not very many data sources external to organizations were easily available. The hard data often came from our own data stores (most frequently OLTP systems), and the soft data often came from analysis of trade journals, government reports, and news articles. Together these data formed the foundations of *decision support*. With the advent of cloud data and service providers, the necessity for BI infrastructure and the variety of data necessary to attain and maintain competitive position has expanded. Cloud computing, mobile devices, the Internet, and Big Data became a game changer for the way organizations create and use BI.

Throughout this book we have reviewed a number of data sources available to analyze and create BI:

- Social media and social factoids, such as Facebook, LinkedIn, Instagram, and Foursquare
- Internal data stores from your organization (e.g., OLTP, CRM, and ERP systems)
- Legally mandated stores
- Product reviews on various web sites (possibly your own organization's web site)
- E-mail
- Data providers like Nielsen BookScan (book sales data), InfoGroup (business demographics), and PolicyMap (geographic information)
- SaaS providers like Salesforce or Google Apps

And the list goes on. An Internet search will yield others. It would probably take this entire book to list all the data providers available today, and by the time the book is published, the list would be outdated. Clearly, there are many data providers available to create BI. As part of your organization's BI plan, you must determine which data providers can give you the information needed to maintain competitive position in your industry.

Big Data, BI, and the Cloud

As you progress through the bevy of in-house data stores, cloud data, and service providers and contemplate combining them to create BI, you must determine whether the technical and analytical capabilities of your staff are up to the challenge. Cloud service providers may require specific programming or other technical skills, data providers may require specialized knowledge of creating BI from the various data sources, and internal data stores may require a skill set unique to managing those technologies. Internal and external BI solutions all have specialized needs that affect resource allocations, budgets, and timelines. The organizational consequences of these factors and others must be understood as Big Data, BI, mobile, and cloud deployments are considered.

MANAGING BIG DATA

According to Forrester's glossary and their pragmatic definition of Big Data, your organization needs to be able to manage its data stores. You must be able to create actionable and pragmatic BI. This is accomplished by the following:

- **Storing.** The ability to collect and store Big Data.
- **Processing.** Having the staff and other resources to properly clean, enrich, and analyze appropriate data stores to create BI.
- **Accessing.** Knowing how to retrieve, search, and integrate data stores to create meaningful visual representations that are useful and accessible to decision makers.

The organization's ability to store, process, and access Big Data must be as well understood as the variety, velocity, and volume of Big Data so that investments in Big Data and BI can be maximized.

Storing

Big Data involves a stream of information that organizations collect and store. The type, size, and other characteristics of your organization determine how much information is streaming in. That data stream must be stored and made ready for processing so your users can access it. This may sound trivial, but based on your organization's technical abilities and needs, where to store your data can be somewhat complex. For example:

- Do you manage and store your e-mail or web site on-premises, or do you deploy to a SaaS provider like Google, a PaaS provider like Microsoft Azure, or an IaaS provider like Amazon Web Services?
- Storing on-premises may make sense if you need that level of control, have staff with the necessary technical abilities, and have the infrastructure and Internet access to handle the data stream. Conversely, off-premises service providers will eliminate some of the up-front costs and need for on-premises infrastructure and staff, but may yield less long-term operational control of your content at a higher total cost of operation.
- Do you collect data about (name your need: GIS, demographics, even annual rainfall) and build your own data store, or do you purchase the data from a cloud data provider? Which is easier to combine with your on-premises data stores, and which maximizes BI creation? Collecting data in-house may make sense if they are not available cost-effectively from a provider. However, if you do collect data in-house, do you have the staff to maintain them and the means to ensure their veracity and currency? On the other hand, would a service provider be better at that? Would its SLA provide appropriate guarantees?
- Where do you store data in general? Does an IaaS provider like Amazon Web Services make sense? Alternatively, is it more cost-effective, and does it give a higher level of security, processing ability, and access to have the data on-premises? If you are a high-tech laboratory like Sandia, a contractor for the U.S. Department of Energy's National Nuclear Security Administration, you might need on-premises infrastructure to store your information securely, but a business may find it more effective to store its data and files at a cloud provider so the employees can access it from anywhere.

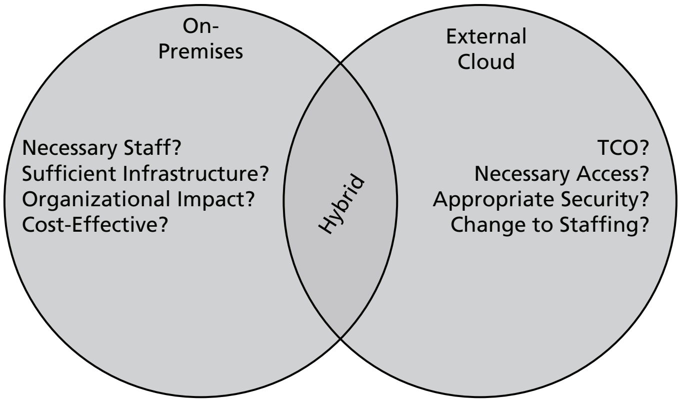


Figure 8.3 Where to Store Big Data

Dropbox and SkyDrive are two good examples of SaaS file storage providers that allow information to be stored/retrieved on any device with the appropriate software installed. However, Dropbox and Skydive store files, and although the files can be stored, processed, and combined with other data to create BI, the question of data stores and their location is a bit different. IaaS providers like Amazon Web Services can store databases, Hadoop file structures, and other formats very cost-effectively, but that leaves local resources and staff to manage the IaaS provider.

Regarding storage to accommodate Big Data, you have the choice to store on-premises, in an external cloud arrangement, or in a hybrid (see Figure 8.3).

There are several questions you must ask before making a decision on how to collect and store data:

- Do you have the necessary staff to build and maintain the data store? Both on-premises Big Data and cloud stores require staff, but of different types. What type of staffing makes most sense for your organization? Do you have or need technical staff or staff familiar with the data who can clean and maintain it? Or both?
- Do you hire staff to build infrastructure, or do you hire less staff and use cloud providers? Do you already have staff in-house with expertise in the technologies or that can be trained?

- What solution is less risky for your organization? Is it better to hire staff and build infrastructure or move to external cloud providers, where you pay on a metered use basis?
- Do you have sufficient ICT infrastructure (e.g., hardware or software) to store locally or put your data with a cloud provider?
- What will the consequences to your organization be, in terms of staffing, resources, and risk, by storing locally or in the cloud?
- What is the TCO for an on-premises infrastructure versus a cloud infrastructure? Does a hybrid infrastructure make sense?
- Do the cloud providers under consideration guarantee required access and security? Can you provide these better on-premises?

This short list of questions covers the basics that must be considered when collecting, storing, and maintaining Big Data. Figure 8.4 gives an example of how you could compare vendors using these questions. These considerations are important for maximizing the investment in Big Data and BI.

- Enter the vendor name
- Modify/add lines as necessary
- Determine how important the attributes are to your organization
- Determine how well vendor meets need at the level of importance indicated
- Talley the number of yes/no in the **Vendor Meets Need** column and use to compare vendors
    - E.G., if an attribute is rated as EXTREMELY IMPORTANT and vendor does an exceptional job of meeting the need then they would be scored YES, if they do not do an exceptional job of meeting that need then they would be scored NO
- Scoring: Organizational Importance → Level At Which Vendor Meets Need

(*Continued*)

Figure 8.4 Cloud Data Storage Adoption Worksheet

Figure 8.4 *(Continued)*

- Extremely Important → Exceptional
- Very Important → Very Good
- Important → Good

- It is assumed that
    - All attributes are important to your organization
    - Vendors that do not score at least "good" on all attributes will not be considered

Vendor Name:________________________________

Attribute		Organizational Importance			Vendor Meets Need	
		Extremely Important	Very Important	Important	Yes	No
1	Necessary and Sufficient Staff					
2	Cloud Provider Support					
3	Risk of Using Vendor					
4	Sufficient ICT Already On-Premises					
5	Organizational Impact					
6	Total Cost of Ownership On-Premises					
7	Total Cost of Ownership in Cloud					
8	Sufficient Security in a Cloud Solution					
9	Sufficient Security On-Premises					
10	Add More Lines as Needed					
				Total Number of Yes and No		

Processing

Choosing where to process Big Data is much like choosing where to store it. Similar to collecting and storing Big Data, processing it can occur on-premises or with an IaaS, a SaaS, or a PaaS cloud provider. Each offers different opportunities and benefits to your organization. A Big Data solution can also be a hybrid of any of these options. The highlights of each are as follows:

- IaaS providers like Amazon Web Services offer an entire infrastructure (i.e., compute, storage, and network) that your organization can use to deploy a Big Data solution. In order to process your data at an IaaS provider, your organization would procure the necessary software licenses, build the necessary software images, and then upload them. The IaaS provider would supply computing and storage on a metered use basis and charge accordingly. Building a Big Data processing solution with an IaaS provider is more technically complex than other cloud service models, but it gives you the most flexibility among the cloud choices. Your staff would remotely access the data stores and associated BI applications processed at a cloud provider.

  An IaaS solution requires the necessary infrastructure and connectivity on-premises that allows access to the IaaS provider, but it will avoid the initial expense of building the entire Big Data infrastructure in-house. In an IaaS solution, the on-premises infrastructure is much smaller, since Big Data stores and processing are at a cloud provider. With an IaaS solution, you can purchase the services necessary to deploy Big Data and BI solutions, but you will still need a technically sophisticated staff to build and maintain them.

- A PaaS provider like Heroku will offer an integrated platform with database, operating environment, and other services on a pay-as-you-go basis.[4] The provider operates within a specific environment, and your organization's staff builds and/or deploys applications on that environment. The applications are accessed remotely over the Internet. The platform is built and maintained by the PaaS provider, alleviating the need for your staff to do so and thus decreasing the upfront cost to deploy a Big Data or BI solution. You may be limited by the updates done to the PaaS environment. For example, if you require an update to Unix or Java because you need new functionality, you may be limited by the revision level offered by the PaaS provider. Overall, a PaaS solution is good for an organization that wants to develop or deploy an application on an infrastructure that it does not want to own or maintain. Your organization will most likely have a larger cost of operation long-term, but it will get the ability to discontinue a Big Data or BI project that is not delivering as promised.

- SaaS providers like GoodData offer Big Data storage and a cloud BI application packaged together. With limited programming, your data scientists can build Big Data stores and process them. Big Data SaaS providers often have connections to other cloud providers, like Twitter, Facebook, and Salesforce. Your organization will need the necessary infrastructure and connectivity on-premises so it can access the SaaS provider's services. The SaaS provider will offer service contracts on a metered use basis, often allowing your organization to use provider services and deploy or cancel at will. That minimizes your organization's risk (i.e., very low upfront costs and the ability to discontinue on short notice), compared to other methods of creating Big Data solutions. However, your organization will be constrained by the type of data and processing supported by the SaaS provider and its BI solution.

On-premises Big Data solutions can be built. They require staff and infrastructure resources to process the data. Upfront, this is the most expensive Big Data or BI solution, but when security, privacy, or other concerns are of paramount importance, it may be the best solution.

The choice of where to process Big Data depends on several factors:

- **Risk.** Does your organization want to minimize risk (i.e., use cloud providers on a pay-as-you-go basis rather than build in-house infrastructure)? Cloud providers allow you to discontinue using and paying for their services, whereas building a Big Data or BI processing solution on-premises means that you have to make a substantial up-front investment and may be stuck with that infrastructure if the project is discontinued.

- **On-premises or off-premises.** Does your organization's Big Data or BI processing needs (e.g., security or access) require an on-premises solution, and does that make the most sense?
- **Features.** A particular cloud provider may give you a limited platform on which to deploy a Big Data or BI solution. The cloud provider service offerings must be totally understood to make sure you can process and create the type of BI your organization needs.
- **Hybrid.** Does a hybrid on-premises and a cloud solution make the most sense?

Organizations must be concerned with these issues when processing Big Data to create useful BI. They must also be concerned with innovation and answer questions like these: Is your organization interested in deploying a new, possibly untested, Big Data or BI solution? Might a cloud solution offer advantages that promote innovation? Is the mix of CAPEX/OPEX important? What is the TCO? Is deploying a solution without substantial up-front investment important? All these issues (and others) must be part of any Big Data or BI plan.

Accessing

Accessing Big Data depends on the following: Have you stored and processed your data so that your data scientists can retrieve them, integrate them with other data stores, and make them useful by applying BI techniques? Here our Big Data definitions from this chapter and our BI definitions from the previous chapter converge. Accessing Big Data is really about being able to make it actionable to support organizational decision making.

It is important that your Big Data efforts do not fall into the monthly reporting rut. Big Data should not be collected, stored, and processed only to have a monthly report created that is looked at when an issue arises. The author has seen this happen in his consulting practice. A health insurance company comes to mind—let's call it Insure-A.

Every month the claims processing system (the OLTP system and internal data store) at Insure-A was used to create a summary report of all claims by diagnosis (i.e., the claims experience report). One month the chief financial officer (CFO) noticed that cash flow was off and requested a special claims experience report. The CFO was told that it would be four days before the report was available. Once the CFO got the report, a few diagnoses that made up the bulk of the claims in the past month were noted, and it was apparent why cash flow was off. However, this was a hindsight approach to monitoring cash and claims; the damage was done to Insure-A's cash position, and it had to dip into its actuarial reserves. The CEO of Insure-A called in a consultant (the author) to determine how this could be avoided in the future. The recommendation was to put in systems that would do the following:

- **Increase access to claims experience.** An OLAP system was developed that allowed the finance department to access information about claims experience in real time and model trends.
- **Add new data to the existing internal data stores.** An early warning system was put into place so health-care providers could notify Insure-A about trends they were seeing in their patient visits (e.g., an increase of asthma or flu). As part of the provider agreement, the existing and new health-care providers were required to enter the number of patient visits by diagnosis weekly. This information was available weeks before the claim was processed and allowed the finance department to estimate cash needs.
- **Shorten the time it took to finalize claims.** By using the cloud-based services of a health information exchange and a claims clearinghouse, health-care providers were able to get their claims into Insure-A substantially faster and more accurately. This reduced the time it took from claims submission to payment.

The on-premises and cloud infrastructure got larger at Insure-A, and it truly had a Big Data infrastructure after these systems were implemented. But by being better able to access its data so it could create effective BI, Insure-A was able to more effectively manage its claims, cash flow, and cash reserves making the company more financially stable. What started as an access problem for the CFO and management team (not being able to access the claims experience report on a timely basis) turned into a Big Data and BI solution.

Big Data and Security

Discussing Big Data and security can cause some confusion. When discussing this topic, we are not just talking about how to secure Big Data. Securing data is an important topic, and there are volumes already written about it. The volumes about security and its associated technologies are very different in content from this book on BI and the cloud. What is important is to discuss how Big Data can be used to enhance security or, perhaps more correctly, to enhance surveillance—the monitoring to ensure that individuals are in compliance with the policies of an organization.[5]

There are numerous examples of how Big Data is used to determine consumer sentiments or track credit card fraud, but with Big Data and security, we mean "matching security data with the right collected data" to create security intelligence—who is doing what to your organization (e.g., external or internal perpetrators attempting to attack your digital assets).[6]

Security intelligence is defined as understanding the following (all may not apply to all security threats):

- The objective of a threat
- How it is propagated or implemented
- Where it originates from
- The action being undertaken
- What software or hardware vulnerability it is exploiting
- The asset that is being attacked
- What the attack is attempting to do
- The effect on performance[7]

Big Data allows security intelligence to be created by linking data stores—for example, relating data from system log files to employee positions. Suppose the log files showed that an employee in the finance department was accessing proprietary process control or patent files more than he or she should. By relating the employee's position in the company with the files this person accesses, a case of corporate espionage might be thwarted.

Big Data security intelligence is another layer of protection from would-be attackers. Adopting a security intelligence model uses Big Data analytics to assess risks and allows organizations to create knowledge about how to defend the organization from attacks. The more risk-aware an organization is the more agile and better able it will be in stopping security threats before they happen.[8]

CONCLUSION

Big Data may seem passé, and cloud providers may seem like little more than Web hosts that offer services. However, the trend is for an increasing number of organizations to put their data in a cloud provider's stores and process it there. This is because the ability to store, process, and access data in the cloud is less financially risky, requires less up-front capital, and is more supportive of innovation than building local infrastructure.

Before the term *Big Data* came into vogue, many organizations already had it—think of the early use of Twitter, Facebook, and LinkedIn by marketing firms. These firms would combine and visualize postings to these social media sites before we realized the effect of automatically mining these web sites to create actionable BI. Companies were utilizing these social media outlets long before we had the term *Big Data* and realized the significance of data volume, velocity, and variety.

The Big Data ecosystem (see Figure 8.5) is defined by the variety, volume, and velocity of data that stream into organizations along with how those organizations store, process, and provide access.

Organizations also need to be concerned with data veracity, since Big Data often forces analysis of data that might be imprecise.[9] At the core is the ability to create BI that supports decision making. By

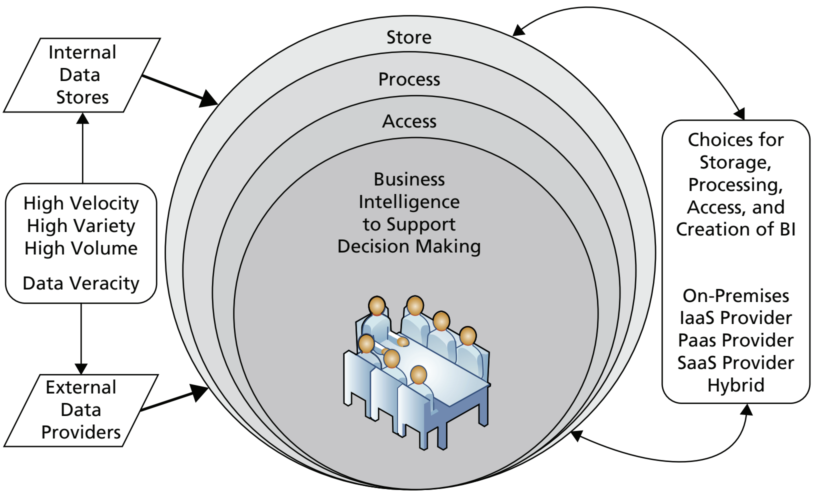


Figure 8.5 The Big Data Ecosystem

combining internal data stores with external data providers and mining for new correlations and interactions, organizations can capitalize on Big Data momentum.

NOTES

1. Erin McCarthy, "The 5Ws and 1H of Big Data (Part 2 of 2)," SourceLink, August 18, 2012, www.sourcelink.com/blog/guest-author-series/2012/08/18/the-5ws-and-1h-of-big-data-%28part-2-of-2%29.
2. Sabre Airline Solutions, "Set Your Business Free with Our Solutions Portfolio," www.sabreairlinesolutions.com/images/uploads/AS_Overview_Brochure.pdf.
3. SAS Event Stream Processing Engine, "Gain Immediate Analytic Insights from Real-Time Data Streaming into Your Organization," SAS, www.sas.com/software/information-management/decision-management/event-stream-processing.html.
4. Heroku provides a Unix operating platform that offers Ruby on Rails, Node.JS, Clojure, Java, Python, and Scala; www.heroku.com.
5. Gendron, *Business-Driven Data Communications*.
6. Richard Stiennon, "Applying Intelligence to Big Data for Security," *Forbes*, May 21, 2013, www.forbes.com/sites/richardstiennon/2013/05/21/applying-intelligence-to-big-data-for-security.
7. N. Ye, C. Newman, and T. Farley, T. "A System-Fault-Risk Framework for Cyber Attack Classification," *Information, Knowledge, Systems*, 5 (2006): 135–151.

8. Sam Curry, Engin Kirda, Eddie Schwarz, William Steward, and Amit Yoran, "Big Data Fuels Intelligence-Driven Security," EMC2, January 2013, www.emc.com/collateral/industry-overview/big-data-fuels-intelligence-driven-security-io.pdf.
9. Michael Walker, "Data Veracity," Data Science Central, November 28, 2012, www.datasciencecentral.com/profiles/blogs/data-veracity.

CHAPTER 9

Mobile Computing Intersection

The Intersection of Mobile, Cloud, Big Data, and Business Intelligence

Mobile computing has brought a whole new level of data into play and is a major driver of Big Data and BI. Big Data comes from applications that companies create to interact with their customers, social media, cloud data providers, and a host of other sources. BI comes from using the data that are stored, processed, and distributed to create intelligence that can be used to innovate the business. All of this is enabled by cloud and mobile computing.

Mobile, cloud, Big Data, and BI all intersect to provide a new frontier of information processing. Mobile computing is composed of many applications and technologies. Mobile devices provide locator applications that know your geographic position and indicate the closest service (e.g., a store or an ATM), money transfer, and package tracking details. These are just a few of the recent mobile innovations.

Organizations must determine how personal, commercial, and business applications of mobile computing can increase employee, customer, and supplier satisfaction and productivity throughout the supply chain. Doing so makes it possible to increase market share and enhance the bottom line.

Annual sales of mobile computing devices—phones, tablets, and other handheld devices—have outstripped the sale of desktop devices; that occurred more than 10 years ago. Driven by IT consumerization, the desktop is a staple at home and in the office, but now we live in a post-PC era, with many mobile devices having more computing power than was available on the desktop 15 years ago. This has ushered in a new paradigm of mobile and cloud computing, creating new ways that organizations can collect, store, process, and distribute Big Data and BI. The challenge is to harness the new opportunities to create competitive position in the marketplace. This includes reckoning with the phenomenon of bring your own device (BYOD).

Mobile computing must be considered from the personal, commercial, and business perspectives to realize its benefits. These perspectives are defined as follows:

- **Personal mobile computing.** Applications that a person interacts with for his or her own use or benefit but that do not include direct interaction with a business. For example:
  - An electronic health record that is enabled by mobile devices and cloud computing. The mobile device can collect data like heart rate and blood pressure (there are even free apps that do these things) that can be uploaded to the electronic personal health record (e.g., Microsoft HealthVault).
  - Various apps like the compass, flashlight, or turn-by-turn directions.
  - The www.youtube.com app that allows the consumer to watch digital clips uploaded by other consumers for free.
- **Commercial mobile computing.** Applications that allow the consumer to interact with a business. For example:
  - Amazon's app allowing the consumer to shop at its web site.

  - Electronic banking apps that allow consumers to access their bank accounts.
  - A Netflix app that allows consumers to watch movies from the Netflix fee-based digital archive.
- **Business mobile computing.** Applications or devices that allow an enterprise's stakeholders (e.g., employees, suppliers, or distributors) to interact with the enterprise. For example:
  - RFID devices (i.e., tags and scanners) that allow automated inventory control.
  - Remote point-of-sale devices that allow sales transactions, updating of inventory, and credit transactions to occur untethered from a cash register.
  - Mobile device apps that allow functions like self-service BI, e-mail, calendars, and a number of other apps to support the organization.

Consumers and employees (who are just consumers at work) drive the demand for mobile computing. Consumers want apps that allow them to be untethered from the desktop at home and at work. Here are some examples:

- The locator that allows the consumer to find the closest store or service (e.g., an ATM affiliated with one's bank so there are no fees).
- The self-service BI app that allows immediate access to critical sales data so a retail district manager can make decisions while on the road.
- Social media apps that allow friends to check in at their current location or take pictures of their lunch.

As market demand changes, so does the pressure exerted by consumers on organizations to provide mobile apps. It is essential for an organization to provide these apps to maintain employee satisfaction, develop innovation through the organization, and maintain its competitive position in the marketplace.

DEFINING MOBILE COMPUTING

Mobile computing is defined by a number of technologies that converge to create untethered access to data assets, including assets owned by the enterprise and the individual. Enterprise assets include on-premises and cloud-based data stores. Individual information assets include files stored at cloud providers like Dropbox and SkyDrive as well as data stored at one's bank, on social media sites, and by a myriad of other online providers.

In this chapter, we will define mobile computing and discuss how to deal with it. We will specifically deal with business mobile concerns.

Mobile computing is defined by its attributes. It is important to understand how cloud computing and mobile computing go hand in hand. It is somewhat of a "chicken or egg" problem. Without the cloud to enable mobile devices and without mobile devices to create the need for data storage and processing, neither cloud nor mobile could have become the force that it is. This symbiotic relationship has allowed both to flourish and will continue to do so.

Mobile computing is defined as the devices and technologies that have or can do the following:

- **Communicate over mobile networks.** This includes devices that can communicate over WiFi or mobile networks.
- **A unique user interface.** Mobile devices typically have an interface that is driven by voice, finger movements, or gestures.
- **Apps**. Lightweight, small-footprint applications that can run on mobile devices. These include personal mobile apps (e.g., Facebook, LinkedIn, and consumer banking) as well as commercial and business applications that allow access to corporate information assets.
- **Recognition of geographic and physical context.** The mobile device knows where it is geographically because of an onboard GPS and can provide that information to a locator app. The mobile device also knows how it is moving in space because of its accelerometer. If you shake your mobile device when the CNN news app is running, it might update the news.

Alternatively, rotate your tablet to see a picture in portrait or landscape.

- **Computing power.** Mobile devices have computing power that allows local processing and mobile connectivity.

MOBILE IN THE ENTERPRISE

The pervasiveness of mobile computing poses new challenges. One of these is whether to allow personal mobile devices to access corporate information assets. Another is that mobile allows access both more quickly and from anywhere, with "anywhere access" having the greatest impact. You must also consider that the forefront of maximizing the investment in mobile access is self-service BI. If mobile access is being considered, then the enterprise has to deal with employees bringing their own devices to the workplace.

The mobile device allows managers and staff to access the digital assets of the enterprise. It is not uncommon for managers and staff to access multiple devices, including a desktop PC, laptop, tablet, and/or mobile phone. This untethered access signals a change in the way employees interact with customers, the enterprise's digital assets, and other stakeholders in the supply chain—in fact, mobile computing has the potential to change practically every interaction and the associated cost of those interactions. Two examples provide insight into those changes:

1. In the health-care setting, patient interactions have changed because of mobile devices. Doctors and nurses initially entered patient information on paper charts and then later into desktop computers. In both cases the entries were made after seeing the patients. Today it is not uncommon for medical providers to enter information into a mobile device in the exam room or at the patient's bedside. That allows the providers to interact with patients differently than they did even five years ago, capturing information in real time while seeing the patient.
2. In the manufacturing, retail, and warehouse setting, we see similar changes. Mobile devices are used to monitor and

collect data from process control devices. It is also common to see mobile devices (e.g., RFID tags and scanners) used for inventory control.

The changes motivated by mobile computing have increased the accuracy of data assets but have also caused new concerns for the enterprise. Some of these concerns are resource-based, whereas others can be addressed by a mobile device management solution.

Mobile Device Management Solutions

According to Gartner, "Enterprise mobile device management (MDM) software is: (1) a policy and configuration management tool . . . and (2) an enterprise mobile solution for securing and enabling enterprise users and content."[1]

Implementing an MDM solution assists an enterprise in its transition to an environment where mobile devices are allowed to access the organization's digital assets. MDM can be a crucial part of a BYOD initiative or be used to manage devices owned by the organization. MDM solutions are put in place to manage all types of mobile devices, including mobile phones, tablet computers, laptops, mobile printers, mobile point-of-sale devices, and others. The three major goals of an MDM solution are as follows:

- **Access management.** MDM solutions implement the policies of the enterprise that control mobile device access to the enterprise's digital assets. When a device attempts access, the MDM solution interacts with the device to register it to the organization's network and to ensure that the device meets the required policies (e.g., software, operating system, and virus updates).
- **Update management.** Updates for a mobile device's operating systems, applications, and utilities (e.g., virus protection) are periodically available. Management of updates enforces your organization's rules. For example, if your organization requires that virus protection updates are no more than 30 days old, mobile devices that are not current will not be allowed access until the required criterion is met. The same type of time-sensitive rules (or version-sensitive rules) can be implemented

for applications, operating systems, and even firmware on the mobile devices.

- **Maintenance of configuration settings.** MDM solutions allow the enterprise to change the configuration setting for devices that are registered with them. This includes settings like ring tones, calendar entries, and service settings (e.g., FTP, telnet, and e-mail). This allows your organization to control how a device behaves while it is on your network.

MDM solutions operate with a server component, a client component, and a management console (see Figure 9.1). Communication between the mobile devices and the MDM service is over the wireless mobile network. The MDM service is configured with your organization's rules via a management console normally located on-premises. The MDM solution can be on-premises or can be purchased from a SaaS cloud vendor like AppSense, AT&T (Toggle), Citrix, Juniper Networks, Microsoft, Novell, or VMware. Mobile devices are required to register with the MDM service before they can access the company's digital assets. An employee will be required to install an MDM client that will enforce the organization's policies before access is allowed.

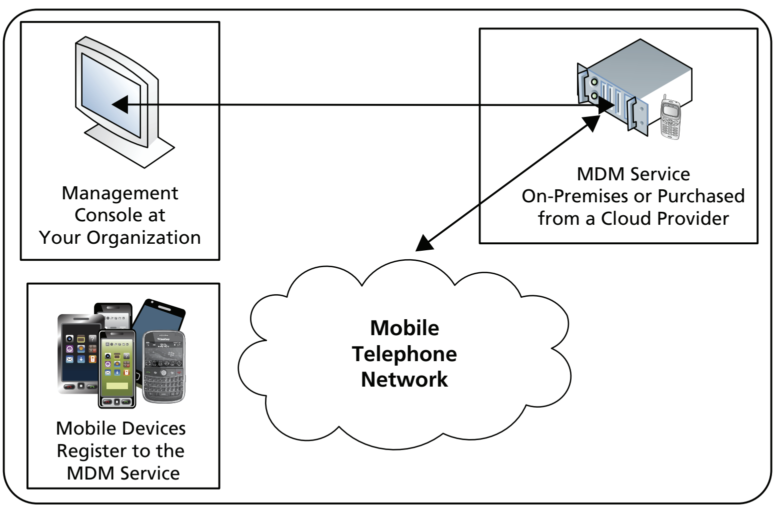


Figure 9.1 MDM Solutions

BYOD Policy

Mobile devices are part of our everyday lives, and organizations must recognize that they are now part of the fabric of business and our society. No longer is anyone surprised to encounter someone using multiple mobile devices simultaneously (e.g., a mobile phone and a tablet). Considering that sales of mobile devices outpace desktop sales and that virtually everyone in the business world deals with mobile devices on a daily basis, it is practical for an organization to have a policy that dictates how these devices can be used in the workplace. It is better to manage mobile devices than to let them proliferate and be uncontrolled, creating another form of shadow IT. Organizations must reckon with personally owned mobile devices and need to create a BYOD policy that accomplishes their goals.

No two BYOD policies are alike. Each reflects the character and culture of an organization and must be driven by well-defined goals that flow from the organization's business strategy. You might be tempted to copy a policy from the Internet, and that is a good place to start, but your policy must reflect your organization. Compare a pharmaceutical company and a marketing company. The BYOD policy for a pharmaceutical company will have different characteristics from the BYOD policy for a marketing company. There will be similarities, but the pharmaceutical policy may reflect stronger concerns about issues like protecting patentable products than the marketing company. A common factor in the pharmaceutical and marketing companies' BYOD policies are that business strategy and enterprise policies must be the driving force.

BYOD policies should be implemented in an iterative fashion in which small incremental changes allow greater access over time. A company may start with communication services like e-mail or other collaboration technologies before moving to delivering self-service BI to their employees. The foundation must be the organization's business and IT strategies, and those strategies must drive which services are delivered and in what order. Of paramount importance to success is realizing that BYOD is about giving the choice to access digital assets on a desktop or via a mobile device. Mobile devices supplement rather than replace the desktop. BYOD increases employee productivity and job satisfaction, and these are both good reasons to implement a BYOD policy. Keep in mind that BYOD devices are not a replacement for the tethered desktop, but rather an enhancement.

The development of a BYOD policy must involve all parts of the organization. It must be an integrative exercise, since BYOD affects all aspects of the organization. For example:

- The human resources department will be helpful in understanding employee training needs and how the telecommuting experience will work.
- IT security will aid the understanding of security polices and how BYOD will affect them.
- Organizational security will shed light on mobile device theft and other physical threats.
- C-suite members will bring a unique perspective that assists in aligning the BYOD policy with the business strategy.
- Accounting or finance department members will aid in analyzing the cost-benefit of a BYOD project.
- IT staff will be able to determine what technologies are necessary to implement a BYOD policy and the consequences of MDM technologies on the existing ICT infrastructure and the IT staff.
- Users must be part of the creation of the BYOD policies since they will be the most affected. The BYOD policy may limit certain applications, but only the users will be able to determine whether those apps are essential to productivity. An old adage applies here: "Managers tell you how a business process should work, but employees tell you how a business process actually works." The difference in this perspective is essential to implementing the correct technologies in the correct order. A good BYOD policy has direct input from both users and managers.

A basic road map to developing a BYOD policy includes the following:

- Before the actual creation of a BYOD policy is undertaken, the following must be considered:

   - The creation of a BYOD policy must be driven by your organization's culture and its business and IT strategies. These strategies must be clearly stated, and the group tasked with creating a BYOD policy must be familiar with them.
   - Users should be consulted to ensure that BYOD is important to them and that the correct applications are available for their mobile devices. The correct implementation of BYOD is essential to increased productivity and employee satisfaction.
   - Human resources, managers, and staff should be queried to determine the effect of BYOD on telecommuting. If a telework policy already exists, the BYOD policy must articulate (and not disagree) with it.
   - A thorough understanding of the challenges posed by BYOD to the organization must be considered. Among them are funding, user readiness, workforce dispersion, current IT staff ability and training, and existing security policies. This will undoubtedly mean involvement by the C-suite.
   - A cost-benefit analysis should be performed.[2] This analysis must take into account employee productivity increases, cost shifting, cost of technologies, and training of IT staff and users. Productivity increases are difficult to quantify, but based on your type of business, metrics can be developed. Cost shifts are easier to quantify; for example, although BYOD may mean that desktop PCs need not be purchased or upgraded as often, potentially saving money, employees may want the organization to provide monthly payments to defray additional voice or data costs.
- Creating a BYOD policy should involve all the stakeholders (human resources, the legal department, security, the accounting and financial department, and IT). With these considerations understood and a cost-benefit analysis in hand, the company can decide to create a BYOD policy. To create a policy, the following steps should be undertaken:
   - The legal department should provide a framework for the types of data that can be accessed with BYOD and whether

that access can be given off-premises—remember, BYOD often allows users to access corporate digital assets while not at work. For example, in certain situations, health-care data cannot be accessed on a mobile device outside the office. The same can be said of other types of data (e.g., personnel records, confidential e-mails).

- Outside legal counsel should be consulted on issues such as the effects of existing labor laws.
- Based on user input, a list of which apps will be supported must be developed (e.g., e-mail, self-service BI, or access to organizational file shares).
- A list of supported devices must be created. Not all services and apps support all devices. An example of this is the original SAS Mobile BI QuickView app which supported the iPhone, iPad, RIM Blackberry, and Android smartphones but not the Windows Phone.[3] Each application, including virtual desktop and virtual application delivery, must be checked for compatibility with supported mobile devices.[4]
- A plan must be developed to train IT staff and users and to assist users with configuring their mobile devices. The IT staff might not be familiar with MDM, and thus training will be necessary. Users will need help configuring their mobile devices to access corporate digital assets.
- Deployment will consist of creating, disseminating, and implementing the policies, services, and apps. IT training will have to occur before the MDM and associated technologies are brought online. User training and configuration support must be in place as well.

Some best practices to consider when developing a BYOD policy and implementing the associated MDM solution are:

- Prevent unauthorized access to corporate digital assets.
- Monitor and update application and operating system version on all connected mobile devices.
- Maintain the appropriate configuration on all registered mobile devices.

- Provide appropriate security and encryption to protect data as they traverse the mobile network and while they are on the mobile device.
- Ensure user privacy and freedom of access.
- Prevent data leakage by keeping corporate and personal apps separate.
- Implement appropriate virtual private network technologies.[5]

The road map and best practices described here embody an ongoing and iterative process. As new applications come online, the legal environment changes, and organizational staff needs change, the BYOD policy will have to be updated. This means revisiting the original considerations (e.g., business and IT strategies, telework needs, cost-benefit analyses) and updating them to reflect the changes. The need for BYOD policies is so pervasive that even national governments have begun to disseminate guidelines reflective of the issues discussed here. For example, the White House has created a tool kit to support federal agencies.[6]

Mobile Innovation

Integrating mobile computing into an enterprise is complex, requiring the input of staff from all segments of the organization. It is complicated to set up a solution that allows access to corporate digital assets. This is largely because of the technical details and the investment of resources required. However, if your organization wants to encourage its employees to be more innovative with mobile technologies, that will require the investment of time and resources too. Good financial and organizational returns gained through mobile computing are predicated on employees having appropriate access to corporate digital assets, external data providers, and BI tools. A mobile deployment should be designed so it yields increased innovation that brings direct gain to customers. As discussed earlier, enterprise mobile deployments will start out with an MDM solution and a few applications like e-mail and calendar. However, to maximize the return on investment of a mobile deployment, Big Data and BI solution should not be far behind.

Big Data

With the right tools, employees are creative. Staff members generally want the organization they work for to succeed. There may be a few who just want a job, but most want to enjoy their work, be creative, and move the company forward. That is a good thing. It means that management, with a mobile computing deployment, must provide the tools necessary for the employees to reach these goals. Big Data and BI are just such tools.

In Chapter 8, we discussed many sources of data accessible over the Internet and from your organization's own data stores, including both structured and unstructured data. These data stores create Big Data and can be combined to create innovative mobile solutions. The intersection of mobile computing, Big Data, and BI create opportunities for innovation. Being able to assemble data stores and cloud services through mobile apps creates a convergence that can greatly benefit the organization and the consumer, and thus your organization's market share. By harnessing this intersection, you can positively affect the bottom line.

An example of the intersection of Big Data, mobile computing, and BI may be found at TripAdvisor (www.tripadvisor.com). TripAdvisor bills itself as the provider of "unbiased hotel reviews, photos and travel advice."[7] To accomplish its goals, TripAdvisor links to the consumer's Facebook account to customize the online experience. The web site uses a cloud mapping and geo-locator service to provide street maps for hotels and other properties. TripAdvisor is monetized through a dedicated division that provides information to the tourism industry. It relies on Big Data, mobile computing, and cloud services to create those innovations, which include the following:

- **Wisdom of Friends.** This allows consumers to get advice from their friends on Facebook as they plan their vacations.
- **City Guides.** This is a smartphone app that allows consumers to get quick snapshots of travelers' reviews of hotels, restaurants, and attractions.
- **Maps.** Using location-based services, this innovation allows consumers to find desired hotels. Hotel popularity, price, and availability are provided in a dynamic mapping interface.

- **Check Rates.** Using cloud services, consumers are allowed to compare hotel pricing and availability across multiple vendor web sites.
- **TripWatch.** Customizable e-mail alerts on specific hotels, attractions, and destinations.
- **Third-party social networking applications.** Innovations are available as applications for use on popular social networking sites like Facebook.

These consumer-oriented innovations provide a treasure trove of structured and unstructured data for use in creating BI. Reservation information can be collected, and consumer trends and sentiments (through unstructured consumer reviews) can be tracked over time. This lets the travel industry follow trends, allowing it to understand customer preferences and behavior.

Business Intelligence

The ability to receive e-mail and webcasts and to perform video conferencing on a mobile device is important, but aligning those communication services with mobile BI is a game changer. This innovative approach allows staff to be able to make decisions untethered from their desktops, and the aligned communication services allow those decisions to be carried out. The BI generated as visual analytics and reports delivered through self-service BI allow users to quickly access the information they need to drive enterprise success.

Self-service mobile BI provides essential capabilities so that decisions can be made and acted on more quickly. SAS mobile BI is an example of analytics that allows the staff to make decisions while untethered. When you select a mobile BI solution, it is important to examine these key attributes:

- **Approachable user interface.** The interface for mobile BI solutions must not require technical skills. Data visualization and reports must be easy to generate, understand, and manipulate.
- **Required data stores.** The mobile solution must make access to data stores straightforward. The data stores necessary to

support decision making must be easily integrated and used. The application must support the integration of the necessary corporate digital assets with cloud-based data.

- **Predefined views and reports.** The ability to create predefined views of the required data along with predefined reports is essential. These reports and views must incorporate dashboards and other visual representations of your organization's key performance indicators (KPIs). However, the predefined items must not limit the user's access to only those reports and views. The user must be able to innovate and create custom views and reports, even combining those that already exist.
- **Manipulation and examination.** BI solutions must allow the user to easily manipulate existing data views and reports. For example, a district sales manager must be able to examine a dashboard of KPIs for his or her region, then be able to penetrate to the sub-regional level and then to the store level. The user must be able to easily manipulate the views, reports, dashboards and other visualizations in order to understand where the higher-level aggregated KPIs come from.
- **Improved productivity.** A mobile BI solution will require that technical staff members deploy it. However, once deployed, that solution must be easy enough to use so it enhances productivity rather than decreasing it. It is important that mobile BI solutions are designed and implemented with user input so that the solution does not become too complex.
- **Untethered and tethered access.** The user must be able to access the same views and reports on their mobile device and desktop. This continuity is an important aspect of increased productivity and accuracy.
- **Deployment and maintenance.** The BI solution must integrate with existing digital resources and be an extension of them. Mobile BI solutions should not replace existing solutions but rather must support them. The mobile solution must also be easy to maintain and not require substantial additional staff and resources to work effectively.

All these attributes must be considered under the spotlight of how the mobile BI solution aligns with existing business and IT strategies. The technological underpinning of these key attributes must recognize, according to SAS, "that there are two opposing forces at work—the need for IT to control the creation and distribution of BI assets and the demand from information workers to have freedom and flexibility without requiring IT help."[8] A balance between these two forces must be reached. Knowledge workers, managers, and other staff must have free access to all BI components while IT has oversight of the self-service BI environment to observe its utilization and manage technology resources.

MOBILE AND THE CONSUMER EXPERIENCE

Looking at mobile computing from the consumer perspective is different than looking at mobile computing as a way to empower employees. Yet getting customers involved in utilizing mobile technologies does have some similarities to employee involvement: There are technologies to be considered, apps that must be deployed or created, and decisions that must be made regarding which digital assets should be available. This is especially important because the consumer mobile market drives much of the Big Data movement. The mobile customer is both a consumer of the apps and a creator of most of the data we use. Consider the example of TripAdvisor that we just discussed. TripAdvisor provides mobile apps that allow consumers to create ratings and comments about tourism locations. Those unstructured and structured data are mined to provide BI to the travel industry about consumer sentiments.

Consumers are involved in and affected by all types of mobile computing. At the personal level, they use apps like HealthVault; at the commercial level, apps allow them to interact with their preferred stores, banks, and cloud providers (e.g. Amazon, Netflix); and at the business level, there is self-service BI. There are a few commonalities in the personal, commercial, and business levels: (1) The individual is a consumer of services, (2) consumers both use and create data (even new BI is really a synthesis of existing data to create new data), (3) the needs of the consumer or employee must drive app creation and deployment, and (4) any new data created as a result of mobile computing add to Big Data stores.

The topic of mobile and the consumer experience is expansive and somewhat beyond the scope of this book, which is to focus on the strategic implementation of cloud-based BI. You should consult other sources for a more thorough treatment of the exciting and complex subject of mobile and the consumer experience.

CONCLUSION

The topic of mobile computing must be examined from the personal, commercial, and business perspectives. These perspectives assist with an organizational understanding of how and why to build mobile solutions. This chapter was largely about integrating these perspectives with the intersection of mobile and cloud computing, BI and Big Data, with an emphasis on how mobile computing is an integrative force to deliver self-service BI.

The core of delivering mobile services to corporate users is the MDM solution and the BYOD policy. The MDM solution allows management of mobile devices and facilitates secure access to digital assets, while the BYOD policy sets out expectations and aligns them with organizational objectives and strategy. Together the MDM solution and a strong BYOD policy create a structure in which employees can be untethered from the desktop, make better decisions faster, and implement them more quickly. By creating this structure, a mobile computing deployment will enhance the organization, improve employee job satisfaction, and lead to an increase in the bottom line.

NOTES

1. Phillip Redman, John Girard, Terrence Cosgrove, Monica Basso, "Magic Quadrant for Mobile Device Management Software," Gartner, May 23, 2013. www.gartner.com/technology/reprints.do?id=1-1FRVS5W&ct=130524&st=sb.
2. Gendron, *Business Intelligence Applied*.
3. Doug Henschen, Mobile BI Apps Target iPad," *Information Week*, April 15, 2010, www.informationweek.com/software/application-optimization/mobile-bi-apps-target-ipad/224400267.
4. Gendron, *Business Intelligence Applied*.

5. Computer Cloud World, “Five Key Issues to Factor into Your MDM Decision,” e-mail alert, July 30, 2013.
6. Bring Your Own Device, “A Toolkit to Support Federal Agencies Implementing Bring Your Own Device (BYOD) Programs,” the White House, August 23, 2012, http://www.whitehouse.gov/digitalgov/bring-your-own-device.
7. TripAdvisor, “Fact Sheet,” www.tripadvisor.com/PressCenter-c4-Fact_Sheet.html.
8. Claudia Imhoff and Colin White, “Self-Service Business Intelligence: Empowering Users to Generate Insights,” SAS, 2011, www.sas.com/resources/asset/TDWI_BestPractices.pdf.

Conclusion

BI software connects people with information when and where they need it, and provides capabilities far beyond spreadsheets to deliver a true picture of the business. For small and mid-size firms, however, finding a BI strategy that matches their resources, expertise and budgets can be particularly challenging.

—Gartner, *A Step-by-Step Approach to Successful Business Intelligence*

Organizations often start small with BI implementations by creating periodic reports or being reactive to enterprise events. That reporting is often based on historical data. However, moving past that is a real possibility.

Over time, organizations can move to using data to make predictive decisions. Big Data, mobile computing, internal data stores, and the cloud combine to create an environment in which "the sky is the limit" when it comes to using data to understand customers' perceptions as well as the general state of the business. Creating useful BI is a moving target that must adapt as available data and organizational needs change.

BI is not just the job of the IT department. It often starts that way, but it must grow. IT may focus on the storage, processing, and dissemination of BI, but for BI to be truly useful to your organization, IT must work alongside the business constituents who generate, understand, and use the intelligence. BI is not just a tool of the IT department—it may be enabled by technology, but it is a tool of business to enhance the understanding of the state of affairs within and outside your organization to enable better decision making.

BI is enhanced through community-enabled solutions and a strong BI feedback loop.[1] A community-enabled solution relies on the users of the solution to provide the feedback necessary for it to improve and mature. An enterprise must create a community to gain insights about its products so it can enrich its offerings. For example:

- SAP HANA is building a community of users that will enable SAP customers to give feedback, derive data from a broad selection of HANA solutions, and thus enrich the HANA platform.
- Microsoft 365 is creating a community by giving one-year free access to its Build Conference attendees. The community is expected to send feedback and increase the adoption of Microsoft 365.

Creating BI through a community approach maximizes investment in a BI solution. This approach maximizes intelligence through the multiple data points that a community creates—this applies to product creation as well as BI for day-to-day decision making. Cloud, mobile, Big Data, and BI converge to create a natural ecosystem for community-enabled solutions to BI problems.

Three forces are essential in finding the BI "sweet spot" (see Figure C.1).

Mobile, cloud, and Big Data intersect to create an environment in which organizations can maximize their investment in BI. Trip Advisor (discussed in Chapter 9) has exploited these forces to monetize its services. Its BI sweet spot employs the following technologies:

- **Mobile.** This is used to collect consumer ratings via a custom app as well as to deliver self-service BI.
- **Cloud.** The ubiquity of travel transactions and consumer ratings are enabled through the cloud along with interactions among business partners.
- **Big Data.** Consumer sentiments, travel sales transactions, and data from business partners are used to create BI that is sold to the travel industry.

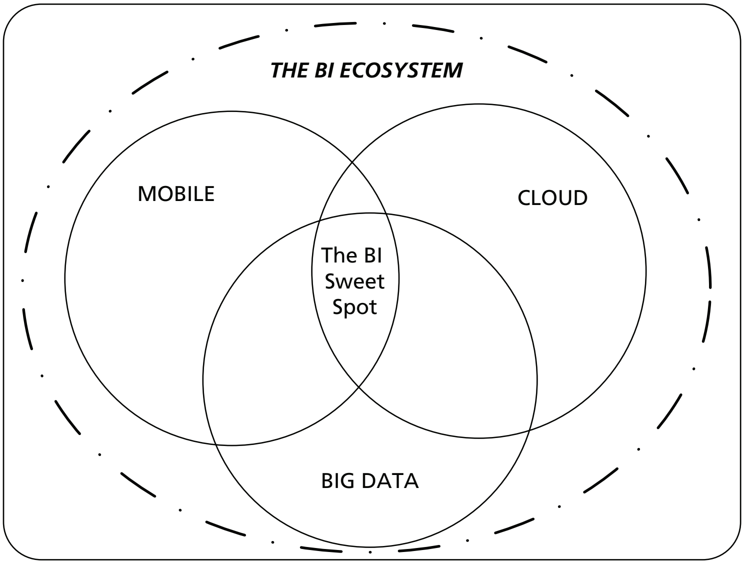


Figure C.1 The BI Sweet Spot

These forces can be exploited by all organizations to maximize the investment in BI and allow them to find the BI sweet spot.

This book started with a brief history of how we got to cloud computing, then proceeded to a detailed description of the characteristics, service models, and deployment models of cloud computing. Next, some basic cloud economics tools were described so the reader had a common basis for understanding cloud strategy and BI in the cloud. The final part of this book defined and discussed BI, Big Data, mobile computing, and their intersection.

The author hopes that this book enables you to understand the importance of BI, Big Data, mobile and cloud computing to the success of an enterprise. Let him know how he can assist moving the BI agenda forward by sending him an e-mail at BIA@Gendron.info. Good luck as you create or update your BI plan based on the strategic objectives of your organization.

NOTE

1. Larry Carvalho, "Community-Enabled Solutions Are the Future," blog, August 7, 2013, http://robustcloud.com/community-enabled-solutions-are-the-future.

Glossary

The number in parentheses after each definition indicates the chapter where more information can be found about the topic.

Application-as-a-service
Any application that is delivered through a browser or specialized client to an end user's PC. SaaS providers offer this service. (2)

Application program interface (API)
A software interface in which cloud providers allow software to access their services. (6)

Association discovery
A data mining technique that analyzes data sets and discovers the association among elements. (7)

Availability (network)
The amount of time a network is accessible for use. Often expressed as a percentage and often required to be 99.999 percent—referred to as *five 9s*. (6)

Big Data
Data with high volume, high velocity, and high variety. (7, 8)

Big iron
Massive computers that provide processing power in organizations that need to manipulate large amounts of data or transactions. (1)

Bring your own device (BYOD)
The practice of employees bringing their own personal digital devices (e.g., phones, smartphones, tablets, laptops) to work with the intent of using them on the employer's network. (9)

Broad network access
A required essential characteristic of cloud computing. Services are available over the network and are accessible via a standardized interface (e.g., application program interface). (2)

Business agility
The ability of an enterprise to sense and react quickly to changes in its environment. (4)

Business analytics
Techniques that create business intelligence, including analysis and data visualization. (7)

Business continuity
The ability of a business to stay open so that customers can access its services. (6)

Business intelligence (BI)
Using business analytic techniques to create reports, dashboards, geographic information systems, or other forms of visualization that allow managers to make better decisions. (7)

Business intelligence feedback loop
A way for the users of business intelligence to provide feedback about quality, content, and format. (7)

Business mobile computing
Applications or devices that allow enterprise stakeholders (e.g., employees, suppliers, distributors) to interact with the enterprise or cloud-based services. (9)

Capital expense (CAPEX) versus operational expense (OPEX)
Accounting and strategic measurement tools. CAPEX refers to how large purchases with an expected life of greater than one year are handled on the balance sheet and profit/loss statement. OPEX refers to items that are purchased by an organization and consumed during the current year. (4)

Classification
A data mining technique that is very similar to clustering, except that a classification is known beforehand. (7)

Client-server computing
A client (e.g., a web browser or an e-mail client) is used to access a hosted application, giving users access to remote servers. (2)

Cloud auditor
A third party employed to conduct an individual assessment of cloud services. (6)

Cloud broker
An organization that manages the use, performance, and delivery of cloud services. (6)

Cloud carrier
The intermediary that provides connectivity and transport between the cloud consumer and the cloud provider. (6)

Cloud computing
A new way to deliver information services that is composed of service models (i.e., SaaS, IaaS, PaaS), services, and essential characteristics. (2)

Cloud computing service model
How services are delivered to the user or developer (i.e., IaaS, PaaS, SaaS). Each delivery system has a number of underlying services that it offers. (2)

Cloud consumer
An individual or an organization that uses or consumes the (free or purchased) services of a cloud provider. (6)

Cloud data provider
An organization that provides data to cloud consumers. (8)

Cloud deployment
The specific requirements for how solutions are implemented. These requirements involve the dimensions of service type (i.e., PaaS, IaaS, SaaS), location of deployment (on-premises versus off-premises), and deployment models (i.e., public, private, hybrid, or community). (5)

Cloud ecosystem
The system composed of cloud consumers, providers, carriers, and brokers. (6)

Cloud service provider
An organization that provides services to cloud consumers. (6)

Clustering
A data mining technique that groups together records with a similar attribute. (7)

Commercial mobile computing
An application that allows the consumer to interact with a business. (9)

Community cloud
Cloud deployments provisioned for exclusive use by a specific community of consumers from organizations that have shared interests or concerns (e.g., mission, security requirements, policy, or compliance considerations). (3)

Compute-as-a-service
An IaaS offering. A service provider allows an individual user or an organization to have computing power on demand within the service provider's infrastructure. (2)

Computer in the back room
An approach in which users do not have direct access to computers. (1)

CP/M
An early PC operating system created by Gary Kildall. (1)

Dashboard
A display of key indicators that is necessary to make decisions. They often have the ability to click on an indicator and get more detail. (7)

Database
A repository of data typically organized according to a data model and structured in such a way that it can be queried. (7)

Database-as-a-service
An IaaS offering that allows developers and consumers to use the database services of a provider on a pay-as-you-go basis. (2)

Data cube
An online analytic processing system technique that is a multidimensional arrangement of data that can be used to produce summaries, means, and other statistics across dimensions. (7)

Data management system
A system's software that allows users and applications to interact with a database. (7)

Data mining
The use of software that discovers unknown predictive relationships and complements an online analytic processing system. (7)

Data model
Specifies how data are stored and the relationships among those data. (7)

Data retention
An organization's policy (often driven by local or state laws) that specifies how long digital assets must be preserved. (6)

Data scientist
The person in an organization who is tasked with managing and modeling data using computer science and statistical, analytical, and/or mathematical techniques to create business intelligence. (7)

Data veracity
An understanding of the quality of Big Data used in decision making so that decisions are given appropriate consideration. Big Data can sometimes

be inconsistent, incomplete, or ambiguous because of its high volume, velocity, and variety. (8)

Data warehouse
A repository of views of data that is often integrated across various databases within an organization and is used for reporting and analysis. (7)

Decision support system (DSS)
Software that analyzes business data and presents them to managers; often referred to as *data visualization*. The presentation of data is designed to assist managers with decision making. (7)

Decision support tool
A software tool that creates visualizations to aid managers in decision making. (7)

Dimension one: Essential characteristics
The characteristics of cloud computing that are either required or optional. Required characteristics must be present, whereas optional characteristics are desired but not normally required. (2)

Dimension three: Deployment models
How a cloud is implemented: public, private, hybrid, and community clouds. (3)

Dimension two: Services models and underlying services
IaaS, SaaS, and PaaS cloud offerings and the software exposed in each of these offerings. (2)

Direct cost
A total cost of ownership metric used when purchasing information and communication technology. Direct costs include the purchase of technology, installation, and maintenance. Depending on the implementation, direct costs may also include testing, the cost of facilities, and environmental (e.g., air-conditioning) and power conditioning. (4)

Fault tolerance
A required essential characteristic of cloud computing. The cloud infrastructure can detect and react to the faults of both hardware and software. (2)

Geographic information system (GIS)
Data-driven maps used to enhance an understanding of the geography and the data that are overlaid with it. (7)

Geo-replication
An optional, desired cloud computing characteristic. Storing the same digital assets in different places that are physically closer (from a geographic perspective) to the user so that access is faster. Because of today's fast network access, geo-replication is not needed very often, but when it *is* needed, the cloud infrastructure should provide it. (2)

High variety (Big Data)
The types of data available. (8)

High velocity (Big Data)
The speed of data availability. (8)

High volume (Big Data)
The amount of data available. (8)

Hosted application
A server used to store and execute the applications (e.g., the Web or e-mail) that are used in client-server computing. (2)

Hybrid cloud
A combination of two or more distinct cloud infrastructures (private or public) that remain unique entities but are bound together by standardized or proprietary technology that enables data and application portability. (3)

Increased information technology resources
An aspect of organizational agility in which more staff, hardware, or software may be required to meet new requests to support innovations. As more services are requested more quickly, the additional resources needed to fulfill those requests must be available. (5)

Increased innovation
In terms of organizational agility and cloud computing, innovation within the organization should increase as the staff realizes that information technology services are available more quickly. (5)

Indirect cost
A total cost of ownership metric used when purchasing information and communication technology. Indirect costs are not directly associated with the purchase of an asset but are real costs that are necessary for an implementation to be successful. These costs must be included in any financial analysis of technology purchases. Indirect costs include training, an outage (either planned or unplanned), a security breach (e.g., physical breach,

virus, or malware breach), disaster preparedness and recovery, other development expenses, and eventual decommissioning. (4)

Information and communication technology (ICT)
Information technology, information systems, and the communications hardware that is used to interconnect them. (1)

Information-as-a-service
Any cloud-based service that provides an application-programming interface or other similar method to allow an application to use or consume information. This service is offered by a SaaS provider. (2)

Information systems (IS)
Information technology plus software, people, and business processes. The terms *information systems* and *information technology* are often used synonymously, but they really are different. (1)

Information technology (IT)
The hardware used to process data and enable business processes. (1)

Information technology governance structure
The organizational information technology controls and the way that information and communication technology resource decisions are made in an organization. (5)

Information technology request prioritization
For organizational agility and cloud computing, information technology requests must be prioritized. The requests that generate greater profitability for the organization should be considered for implementation first. This will affect the allocation of IT resources and may call for a more robust IT governance structure with a stronger portfolio management of projects. (5)

Information visualization
How information is presented. It is used to show complex representations of data so they can be more easily understood and analyzed. (7)

Infrastructure-as-a-service (IaaS)
An IaaS service provider offers the capability for the consumers to provision processing, storage, networks, and other fundamental computing resources so they can deploy and run software. (2)

Integration-as-a-service
Solutions and services that help companies integrate in-house and cloud-based solutions; may be referred to as a *hybrid cloud*. Cloud integrators

provide consulting, training, and service solutions within vertical markets or horizontally across business processes. This service is offered by a SaaS provider. (2)

Internet technology
A required essential characteristic of cloud computing. Technologies such as TCP/IP, HTTP, UDDI, SOAP, and REST must be employed for a service to be considered cloud. (2)

Knowledge management
Though not a specific mathematical or statistical technique, knowledge management is about organizing, measuring, and converting tacit information generated by employees into digital form. (7)

Management-as-a-service
An on-demand service that provides the ability to manage one or more cloud services. These typically include asset, topology, resource utilization, virtualization, and uptime management. This service is offered by a SaaS provider. (2)

Mean provisioning time
The average time it takes to make new services available. (6)

Mean time between failures (MTBF)
The average time allowable or expected between equipment failures. (6)

Mean time to repair (MTTR)
The average length of time it takes to repair equipment that has failed. (6)

Metered use
A required essential characteristic of cloud computing. The monitoring of utilization in the cloud environment so that resource pools can be adjusted to accommodate user demand. (2)

Metered use, with pay as you go
An optional, desired cloud computing characteristic. Although a private cloud may offer services to its local users without charge, a public cloud service provider will provide scalable and elastic services through metered use that will be based on the resources consumed, and charges will be based on that metered use. (2)

Middleware
Software that sits between various servers to provide services like load balancing or data conversion. (1)

Mobile device management (MDM)
Software that normally consists of a server, a client, and a mobile client who implements policy and configuration management on mobile devices. (9)

MS-DOS
An early PC operating system created by Bill Gates in collaboration with IBM. This operating system was command-line driven and did not have a graphical user interface. (1)

Natural language processing (NLP)
An unstructured text analysis technique that allows computers to determine meaning from text or human language input. (7)

Negotiated service level agreement
Negotiated terms between an organization and a cloud provider (as opposed to a standardize service level agreement). (6)

Network-as-a-service
An IaaS offering that provides functionalities to network users or the network itself. These functionalities help the network perform well and provide things like domain name resolution (changing a name like www.gendron.info to 96.31.40.182) and Dynamic Host Configuration Protocol (DHCP) (allowing client computers in a local area network to request certain protocol parameters). Other functions on the local area network, like e-mail, printing, and file sharing, can also be seen as network-as-a-service. (2)

On-demand self-service
An optional, desired cloud computing characteristic. The ability to set up and change service contracts without direct provider action. (2)

Online analytic processing system (OLAP)
The ability to analyze complex data interactively to perform on-demand data modeling. (7)

Online transaction processing system (OLTP)
A type of information system that facilitates the processing of business transactions immediately and interactively. (7)

Operating cost
A total cost of ownership metric used when purchasing information and communication technology. The cost of operating a cloud-based infrastructure. It must include the operational costs of purchasing services from cloud vendors. If an organization requires additional staff because of a cloud implementation, this should be included in the operating cost as well. (4)

Operational objective
In a network service level agreement, this is a technical target that a network must meet, like mean time between failures. (6)

Orchestrator
Software that manages servers so their functioning can be maximized. See *Middleware*. (1)

Organization agility
See *Business agility*. (5)

Organizational transformation
Materially changing an organization by implementing new policies, cloud, or other technology. The change can be in process, organizational culture, or something else, but it must be transformative and substantial. (5)

Pay as you go
The ability to pay for only the portion of information and communication technology services you actually use. (1)

Personal mobile computing
An application with which a person interacts for his or her own use or benefit, but this does not include direct interaction with a business (e.g., a heart rate monitoring app). (9)

Platform-as-a-service (PaaS)
A PaaS service provider offers the capability to deploy consumer-created or acquired applications onto the service provider's cloud infrastructure. (2)

Predictive analysis
In the context of data mining, this refers to using data to forecast consumer behavior. (7)

Private cloud
In a private cloud, infrastructure is provisioned for exclusive use by a single organization comprising multiple consumers (e.g., business units). It may be owned, managed, and operated by the organization, a third party, or some combination of them, and it may exist on- or off-premises. (3)

Process-as-a-service
Supports business processes through combining other services to create meta-applications. This service is offered by a SaaS provider. (2)

Public cloud
An infrastructure provisioned for open use by anyone (i.e., all users share the same infrastructure, potentially resulting in economies of scale). It may be owned, managed, and operated by a business, academic, or government organization or some combination of them. It exists on the premises of the cloud provider and is accessible over the Internet. (3)

Quick and dirty operating system (QDOS)
An early (1960s and 1970s) PC operating system. (1)

Regulatory impact
How laws and regulations effect an organization. (5)

Response time (network)
How long it takes to get a reply from a network after something is submitted—for example, clicking a button on a web page. (6)

REST (REpresentational State Transfer)
A simple stateless architecture often used in mobile applications that emphasize the interactive of clients over http. An alternative to the SOAP protocol. (1)

Return on investment (ROI)
An accounting measurement that shows how well something performed financially. (4)

Scalable and elastic infrastructure
A required essential characteristic of cloud computing. A cloud service offering that adapts to consumer demand. (2)

Security
An optional, desired cloud computing characteristic. The protection of digital assets by physical and software methods. (2)

Security-as-a-service
Rather than acquiring personal security software tools and the technical expertise to administer them, the user contracts with a security vendor to consume turnkey services like virus defense, firewall management, and e-mail filtering. This service is offered by a SaaS provider. (2)

Sequence discovery (data mining)
Sequence discovery is used to show relationships over time when data items are being mined. (7)

Service broker
An entity that maintains a service registry describing the available cloud services. When the broker exists today, it might be referred to as a "managed service provider." (1)

Service consumer
An individual user or an organization that purchases cloud services from a service provider. (1)

Service level agreement (SLA)
A contract between a cloud consumer and a provider that specifies the business, data, and technical objectives that must be met. (4)

Service level agreement business policy
Specifies guarantees, acceptable use, services not covered, excess usage, activation, payment and penalty models, governance and versioning, renewal, transferability, support, planned maintenance, subcontracted services, licensed software, industry-specific standards, and additional terms for different geographic regions. (6)

Service level agreement data policy
Specifies data preservation, redundancy, location, seizure, and privacy. (6)

Service model
The way cloud services are delivered. Includes infrastructure-as-a-service (IaaS), platform-as-a-service (PaaS), and software-as-a-service (SaaS). (2)

Service-oriented architecture (SOA)
A software engineering technique directed toward designing and building computer software services that interoperate. (1)

Service provider
A cloud service provider that exposes (i.e., makes available) services for use or purchase. (1)

Service-specific objectives
In a service level agreement, the software targets that a network must meet (e.g., traffic parameters) that indicate how well a service is performing. (6)

Services
Software applications that are exposed (i.e., made available) by a service provider for use or consumption. (2)

Services-based
A required essential characteristic of cloud computing. Cloud offerings are provided as a set of facilities to be used by consumers. (2)

Shared resource
A required essential characteristic of cloud computing. The service provider combines its resources, making that combination transparent to the consumer of the services. All users employ the same pool of resources simultaneously, but that simultaneous use does not affect the user, nor do the users know they are using pooled resources. (2)

Simple object access protocol (SOAP)
An XML protocol used to exchange information between the orchestration layer of a web service and the services it calls. (1)

Slicing and dicing
An analysis technique in which the data at any specific point in a data cube are measured. (7)

Social factoid
Information of a social nature (e.g., location or birth date) that is posted to social media sites (e.g., Facebook). (8)

Software-as-a-service (SaaS)
A model of software delivery provided from a remote server rather than from direct installation on the user's PC. Software delivery is often accomplished via the web browser. (1, 2)

Staff readiness
Strategic measurements put in place to understand whether the staff is ready or able to implement cloud technologies. (4)

Standardized service level agreement
Standard terms that are agreed to when a user creates a new account relationship with a cloud provider. (6)

Storage-as-a-service
The ability to utilize storage that physically exists at remote locations but logically appears as local storage to any application that requires it. An IaaS service, also known as *disk on demand*. (2)

Strategic measurement
Establishment of the metrics necessary to set, accomplish, and monitor a strategic orientation. (4)

Structured data
Data that follow a data model. (7)

Testing-as-a-service
Providers offer the ability to test applications through cloud-delivered algorithms that can test web sites, applications, and other software that do not require an on-premises footprint to carry out the testing. This service is offered by a SaaS provider. (2)

Text mining
Unstructured data analysis technique to extract information from text documents, including letters, e-mail, customer service logs, and feeds from social media services such as Twitter, LinkedIn, and Foursquare. (7)

Total cost of ownership (or operation) (TCO)
A financial estimate of the direct and indirect costs of owning something like cloud infrastructure. Understanding the total cost of existing infrastructure aids in knowing how businesses benefit from that infrastructure. It is also very useful when comparing various options being considered for funding. (4)

Traffic parameters
The type of traffic (e.g., constant, real-time, non-real-time), the speed of that traffic, and the acceptable amount of delay during transmission. (6)

Transmission errors
The acceptable number of errors during transmission. (6)

Unstructured data
Data such as text files, documents, e-mails, web pages, and social media sites that do not follow a data model and are free-form in nature. (7)

Unstructured data analysis
Analyzing data that do not have a specific data model. (7)

Virtualization
Creating copies of something that all share the same physical resource. Each copy will execute on a single device, but to the user it will appear as if each has its own dedicated resource. (1)

Web services description language (WSDL)
An XML-based language that is used to describe cloud services. (1)

Web usage mining
An unstructured data analysis technique used to understand the path that users take through a web site. (7)

Whole is greater than the sum of its parts
In the information and communication technology context, a computer cluster in which multiple devices are loosely connected via networks and software, resulting in a combined greater processing power than that of any PC alone. (1)

About the Author

Michael Gendron practices what he writes about in this book through his consulting practice and his information systems professoriate where he creates theories used in business. His approach is that theories become most useful only when they have practical application. Some of the recommendations on his LinkedIn profile may describe him best.

From a colleague in the cloud space:

> Michael Gendron is an internationally known thought leader on the intersection of business intelligence, mobile, cloud and Big Data. . . . His experience includes decades of IT management, where he rose through the ranks from programmer to systems analyst to CIO. Currently, he is a professor of Information Systems.

From a former student:

> Dr. G is one of the most genuine professors I have ever come across. His passion for technology and teaching is absolutely contagious. He has been influential in my life as a technology buff, and I can say that it was through his teachings that I developed my ability to research, to always look for creative ways to conquer technological challenges, and to always believe in myself. I owe a lot to Dr. G, and he is truly an asset to have in my corner.

Michael has worked as a consultant, an information technology coordinator, a chief information officer, and in several other IT-related positions in which he created business strategy, provided technical insight, and generally oversaw the implementation of technical solutions. As an academic, his research includes the following:

- Strategic use of information and communications technology
- Value-driven information and communications technology
- Business intelligence and analytics
- Information supply chain
- Health-care information and communication technology (ICT)
- Internet governance
- Psychological impact of information systems
- Data visualization through geographic information systems

Index

Figures and tables are indicated by *f* and *t*, respectively.