Part I

Introducing the Windows Azure Platform

1

Surveying the Role of Cloud Computing

The term *cloud computing* implies access to remote computing services offered by third parties via a TCP/IP connection to the public Internet. The cloud symbol in a network diagram, which initially represented any type of multiuser network, came to be associated specifically with the public Internet in the mid-1990s. As an example, the following is the first paragraph of Wikipedia's definition of cloud computing as of mid-January 2009:

> Cloud computing is Internet (''cloud'')-based development and use of computer technology (''computing''). It is a style of computing in which resources are provided ''as a service'' over the Internet to users who need not have knowledge of, expertise in, or control over the technology infrastructure (''in the cloud'') that supports them.

Gartner defines cloud computing as

> Scalable, IT-related capabilities provided as a service on the Internet.

The preceding definitions encompass almost all common Internet-based activities, ranging from individuals sending e-mail messages and viewing Web pages to retailers processing credit and debit card charges for online purchases. Google CEO Eric Schmidt narrowed the definition a bit in an August 9, 2006 interview by Danny Sullivan at the Search Engine Strategies Conference (transcribed at `http://bit.ly/wday4`, `www.google.com/press/podium/ses2006.html`):

> What's interesting [now] is that there is an emergent new model, and you all are here because you are part of that new model. I don't think people have really understood how big this opportunity really is. It starts with the premise that the data services and architecture should be on servers. We call it *cloud computing* — they should be in a ''cloud'' somewhere. And that if you have the right kind of browser or the right kind of access, it doesn't matter whether you have a PC or a Mac or a mobile phone or a BlackBerry or what have you — or new devices still to be developed — you can get

> access to the cloud. There are a number of companies that have benefited from that. Obviously, Google, Yahoo!, eBay, Amazon come to mind. The computation and the data and so forth are in the servers. [Emphasis added.]

Mr. Schmidt is considered by many to be the first user of the term *cloud computing* in the context of its embodiment in 2008 and later, but the term didn't reach the threshold for inclusion in Google's Trends service until about September 2007 (see Figure 1-1). Mr. Schmidt makes the assumption in the preceding quotation that data services provided by the cloud-computing servers were defined by the organizations that owned the servers, specifically Google, Yahoo!, eBay, and Amazon.

Figure 1-1: Worldwide traffic for the terms *cloud computing*, *Windows Azure*, *Amazon EC2*, and *Google App Engine* for the years 2000 through 2008 as reported by the Google Trends service.

Amazon released its Elastic Compute Cloud (EC2) web service, which was the first service to permit users to run their own custom programs — rather than host web sites only — in the Internet cloud, on August 23, 2006, just two weeks after the Schmidt interview.

IDC, a well-regarded technology market analysis firm, forecasted in late October 2008 that IT spending on cloud services will grow by a factor of almost three and reach $42 billion by 2012, at which time it would account for about nine percent of total software sales. IDC expects that spending on cloud computing will accelerate during the forecast period, ending up by capturing 25 percent of IT spending growth in 2012 and gaining nearly a third of that growth the following year.

Cloud Computing with the Windows Azure Platform covers the enterprise-oriented cloud computing services offered by Windows Azure Platform as illustrated by the logo of Figure 1-2, which introduced Azure-related technical sessions at the Professional Developers Conference (PDC) 2008, held in Los Angeles October 27–30, 2008. Microsoft released the first Community Technical Preview (CTP) of Azure, formerly known as ''Project RedDog'' and occasionally called ''Stratus,'' at PDC 2008.

Figure 1-2: The Windows Azure Platform was called the Azure Services Platform until July 2009.

Specifically, this book covers

- *Windows Azure*, the operating system which implements the Windows Azure Fabric's production version in virtualized Windows Server 2008 clusters.
- *Azure Storage Services*, which provides scalable persistent storage of structured tables, arbitrary blobs, and queues.
- *SQL Services*: SQL Azure Database implements Microsoft SQL Server in the cloud with features commonly offered by enterprise-scale relational database management systems. SQL Reporting and SQL Analysis services are expected as future data-related SQL Services.
- *.NET Services*: Access Control, Service Bus, and Workflow services, as well as Server Bus Queues and Routers.
- *Windows Azure Software Development Kit* (*SDK*), which implements the Azure Development fabric and Azure Storage Services on local development PCs.
- *Windows Azure Tools for Microsoft Visual Studio*, which provide Visual Studio 2008 and 2010 project templates and other support for developing applications that run on the Windows Azure Development and Production fabrics.

The book does not cover the Live Operating Environment (LOE, formerly Mesh Operating Environment, MOE) and its Live Services because these are consumer-oriented features. Nor does it dig into Microsoft SharePoint Services, Microsoft Dynamics CRM Services, or Office Business Applications (OBAs) because they are Microsoft proprietary applications that have been modified to run on the Azure Production Fabric and use Azure Storage Services to persist state. This book's content is directed to the Azure services that are not crossed out in Figure 1-3.

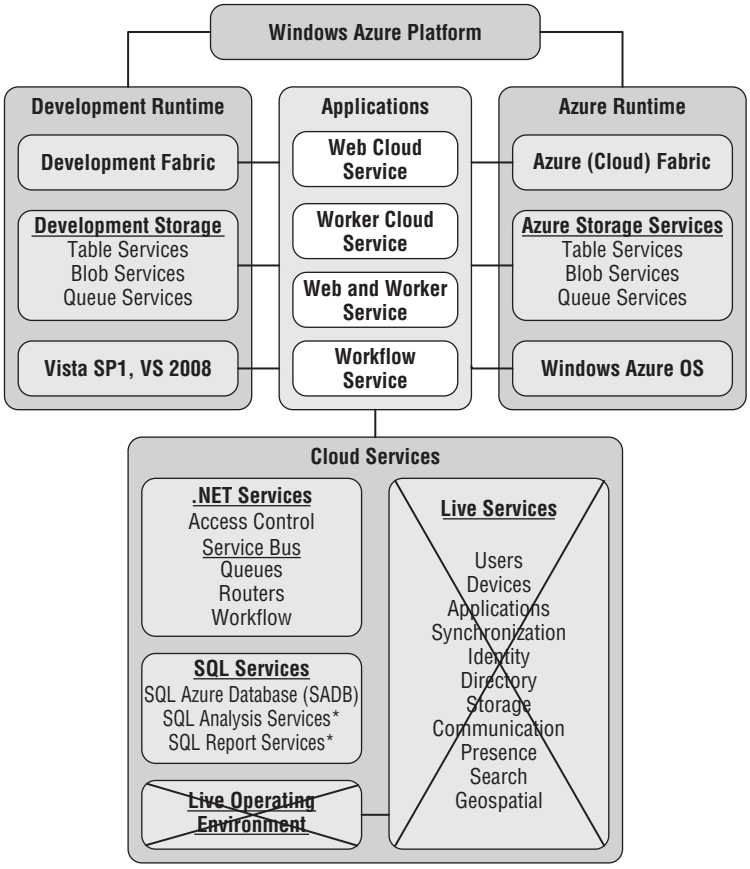

Figure 1-3: Enterprise-oriented Windows Azure Platform and SDK features. Features not covered in this book are crossed out.

This book was written with the fourth (May 2009) and later CTPs of the Windows Azure SDK and Windows Azure Tools for Microsoft Visual Studio.

Chapter 2, ''Understanding Windows Azure Platform Architecture'' and the remaining chapters of Part I, ''Introducing the Windows Azure Platform,'' describe the underlying architecture and implementation of Windows Azure and its repertoire of enterprise-oriented features.

Why Migrate Applications and Services to the Cloud?

Cloud computing is receiving massive press coverage, generating an unending series of conferences, increasing IT management mindshare and substantial software developer resources because it enables small, medium, and large businesses to

- Get new products or services to market faster by minimizing time to deploy fixed IT assets, such as servers, switches, and routers, and by eliminating related incremental capital investment in these assets.
- Conduct market tests quickly and constrain losses by failing fast if the market, product, or service doesn't meet expectations.
- Defer long-term planning until results of initial market tests are known.
- Replace capital expenditures for unneeded capacity to accommodate periodic usage spikes, such as those that occur after announcing seasonal discounts or a new software version, with usage-based monthly payments.

If initial market tests succeed, serving software applications or services from the cloud lets business units deploy new products quickly and scale applications or services almost instantly to meet customer demands. For top management, the key to adopting cloud computing is its ability to trade IT capital investment for usage-based operating expenditures.

Cloud Computing's Ancestry

On the client side, many computer hardware and software suppliers took up the challenge of breaking the Microsoft/Intel hegemony in the PC market by designing and marketing networked diskless workstations, also known as *thin clients*. Microsoft offered its own thin Internet clients as Web TV set-tops and connected to intranets with Zero-Administration Windows (ZAW) for NetPC clients. These client designs reduced cost by eliminating local fixed disks and relied on networked servers to load applications and store user files. However, thin-client prices weren't low enough to capture significant market share from the ubiquitous Windows PC.

The new netbook platform, which appeared in the laptop PC market just as cloud computing gained widespread attention, appears to offer sufficient cost incentive to achieve volume manufacture. Netbooks usually offer conventional hard disks with less capacity than mainstream laptops or solid-state disks (SSDs).

Application service providers (ASPs) and web hosting firms were the first to rent server CPU cycles and storage space on an as-needed basis. The larger of these organizations are expected to participate in the cloud computing market.

Diskless Workstations and Thin Clients

Oracle's Network Computer (NC) concept of the mid-1990s probably is cloud computing's most direct ancestor. Oracle trademarked the *Network Computer* term in 1996 for a diskless network client for business

use and established a Network Computer Reference Profile. The profile required all NC appliances to support HTML, HTTP, Java, and other Internet-related standards. The price advantage of NCs over PCs, if any, wasn't sufficient to create a significant market among businesses and the poor connectivity of dial-up connections discouraged consumer NetPC usage.

> *Microsoft and Intel produced a competing standard called ''NetPC'' to compete with the NC profile. In 1997 Dell Computer introduced a sealed-case PC with no floppy disk, CD drive, or other optional components, that ran Windows NT 4.0 Workstation. Compaq and HP introduced similar NetPC workstations that ran ZAW in mid-1997.*

Sun Microsystems' trademarked ''The Network is the Computer'' motto led to its initiative to replace PCs with JavaStations, which used the Java operating system running on SPARC processors. IBM dipped its toe in the diskless workstation market with Network Stations. JavaStations and Network Stations had the same technology problems as NCs. Wyse Technology, Inc., originally a manufacturer of terminals for mainframes and minicomputers, entered the PC market and then branched into NCs in the 1990s.

These thin clients had sufficient computing power to run a web browser and a few simple applications downloaded from the Web on demand but relied on networked disc storage. Oracle CEO Larry Ellison abandoned the NC project and Sun gave up on the JavaStation in about 2000. NetPCs and ZAW fared no better; of these U.S. thin-client pilgrims, only Wyse was producing significant quantities of dedicated thin-client workstations for business use in 2009.

Thin clients might make a comeback with VMware Inc.'s release of VMware View Open Client, a recently open-sourced desktop infrastructure client that lets you connect a Linux desktop or laptop to hosted virtual Windows desktops managed by VMware View. Gartner predicts that

- Approximately 50 million user licenses for hosted virtual desktops will be purchased by 2013.
- Thin-client terminals will account for about 40 percent of user devices for hosted virtual desktop deployment.

Web TV and Its Clones or Descendants

Microsoft acquired WebTV Networks, which operated an online consumer web service and licensed the design of a diskless workstation that used a conventional TV set as the display, in August 1997. At the time, WebTV Networks had about 150,000 subscribers; both Sony and Philips were producing WebTV set-top boxes under license. Microsoft purchased WebTV Networks' subscribers in 2001 for the Microsoft Network (MSN), terminated Sony Electronics' and Philips Consumer Electronics' licenses, and rebranded WebTV as MSN TV. Thomson remains the sole U.S. set-top box licensee under the RCA brand.

America Online introduced AOL-TV, a WebTV lookalike, in 2000. In 1999 AOL teamed with Liberate Technologies, formerly known as NCI or Network Computers, Inc., a creator of thin-client systems such as the NetChannel, to write software for its set-top box. AOL reportedly had offered $65 million for NetChannel in December 1997, but negotiations broke down and AOL ceased financial support for NetChannel. AOL finally paid $29 million for NetChannel after it shut down service to its 10,000 subscribers on May 3, 1998. Thomson was the producer of NetChannel's set-top box but Philips made AOL-TV's set-top boxes, which sold for $249.95. The AOL-TV subscription cost $14.95 per month on top of AOL's then $21.95 per month PC service charge.

In 2004, when MSN TV 2 launched with set-top boxes that ran the Windows CE operating system and offered broadband access as well as dial-up Internet connectivity, analysts estimated that MSN TV had about one million subscribers.

As of early 2009, RCA MSN TV 2 Internet and Media Players had an MSRP of $199.95 and were available online through Amazon.com and a few other retailers but were on backorder from Microsoft. (Circuit City, the sole in-store MSN TV 2 box retailer, voluntarily liquidated in January 2009.) Microsoft's ''MSN TV Services Fact Sheet'' page on PressPass hasn't changed since May 2006, which might indicate a lack of Microsoft's interest in continuing to devote resources to MSN TV 2.

Netbook Clients

Netbooks are small laptop PCs that are designed for wireless networking and access to the Internet, long battery life, and physical robustness. The netbook platform grew out of Nicholas Negroponte's One Laptop per Child (OLPC) program whose mission is, ''To create educational opportunities for the world's poorest children by providing each child with a rugged, low-cost, low-power, connected laptop with content and software designed for collaborative, joyful, self-empowered learning.'' The original OX-1 model, which went into large-scale production in late 2007, targeted a $100 cost to third-world governments by 2008. The OX-1 features an AMD CPU, 1200 x 900-pixel, 7.5-in. (diagonal) LCD display, 256MB DRAM, 1GB ROM for the Linux operating system and ''Open Firmware,'' 1GB flash memory, a Secure Digital (SD) card slot, and 802.11b/g and 802.11s (mesh) wireless communication. The price in early 2009 for substantial quantities turned out to be about US$219 for the 50 least-developed countries and US$259 for other jurisdictions.

Intel's Classmate PC design, which like the OLPC OX-1 is designed for emerging markets, provides street cred to the almost US$300 actual selling price category. Acer Aspire One, Asus Eee PC, Dell Inspiron Mini, and HP Mini models offer prices ranging from about US$300 to US$400, depending on display size, SSD capacity and other specifications. In early 2009, AT&T offered a US$349 mail-in rebate to Dell Inspiron Mini 9 purchasers who sign up for an AT&T data plan, which reduces the cost of the netbook to US$99. Other carriers probably will join AT&T with iPhone-like hardware subsidies to gain cellular data subscribers.

Netbooks powered by Atom CPUs from Intel running Google's Chrome OS operating system are expected by 2010.

Other assemblers add ''Cloud'' to their model names; for example, Everex introduced its US$399 CloudBook computer in early 2008. The *New York Times* writers Brad Stone and Ashlee Vance point out in their ''$200 Laptops Break a Business Model'' story of January 25, 2009:

> [M]ore experimental but lower-cost technologies like netbooks, Internet-based software services (called cloud computing) and virtualization, which lets companies run more software on each physical server, are on the rise . . .
>
> The only bright spot in the PC industry is netbooks. Analysts at the Gartner research company said shipments rose to 4.4 million devices in the third quarter of 2008, from 500,000 units in the first quarter of last year. Analysts say sales could double this year (2009) despite a deep worldwide recession.

Market researcher DisplaySearch projects sub-US$300 netbooks to increase from worldwide sales of one million units in 2007 to 14 million in 2009. Netbooks and smartphones probably will constitute the majority of clients connected to cloud-computing virtual servers by 2010.

Application Service Providers and Software as a Service

The ASP market fueled the late 1990s dot-com bubble but ASPs also were one of the largest market segments to survive the early 2000s burst. As Service-Oriented Architecture (SOA) gained traction with software developers and enterprise IT departments, ASPs gradually became known as *Software as a Service* (*SaaS*) providers. There are five generally accepted ASP market segments:

- *Specialty ASPs* usually deliver a single application, such as credit card or other payment processing, customer relationship management (CRM), human resources management system (HRMS), word processing, spreadsheet, database or timesheet services. Google Apps provide web-based email, calendar, word-processing, spreadsheet and presentation modules to business users for a fixed charge per user per year, while Salesforce.com rents CRM capabilities and Intuit provides its QuickBase RDBMS with per subscriber per month billing.
- *Enterprise ASPs* deliver a broad spectrum of specialty ASP solutions. For example, Microsoft rents Microsoft SharePoint Services, Microsoft Dynamics CRM Services, and Office Business Applications (OBAs), as well as Windows Live services online.
- *Vertical-market ASPs* deliver multiple software solutions for a specific customer category, such as medical or dental practice, insurance brokerage, church congregation, residential or commercial construction, or personal finance management.
- *Local-market ASPs* deliver geocoded marketing services to small service businesses, such as restaurants, pubs and bars, within a limited geographic region.

ASPs usually charge fixed monthly fees per subscriber, which include software license fees. ''Excessive usage'' surcharges aren't common, but providers often add disproportionate fees for ancillary ''a la carte'' services. Applications that require the provider to train customers' users commonly involve setup fees, yearly commitments, minimum payments, and the like.

Web Hosting Services

Web hosting services, which have been available since about 1991, are the most prolific of all cloud-computing forebears; it's estimated about 50,000 services in the U.S. host 100 or more web sites. Web hosting services provide operating systems, web server implementations, e-mail processing, content storage, high-speed Internet connectivity, and related services at monthly charges ranging from free to thousands of dollars, depending on resources consumed. Web hosting services fall into the following categories:

- *Shared server hosting* runs multiple sites from a single physical server and operating system instance. Relatively little protection exists for an individual web site's intellectual property with shared server hosting because several services run on shared resources, including the same operating system instance. Most free and low-cost (US$30.00 per month and lower) services use shared server hosting. It's common to include content storage up to about 1GB and Internet traffic to 1TB or so per month in the basic monthly charge with surcharges for added storage and traffic. Setup fees are uncommon.

- *Virtual Private Server* (VPS), also called *dedicated virtual server* hosting, isolates the operating system and web server in a virtualized instance, which allows a site to be logically partitioned from other sites on one or a cluster of physical machines. VPS hosting provides additional security and costs from about US$40 or more per month with increased storage and traffic limits. Small-scale e-commerce sites commonly use VPS hosting. Some firms charge small setup fees for VPS hosting.
- *Dedicated server hosting* leases a physical web server to the operator for increased security by content isolation at a cost of from about US$200 per month and up, with the monthly charge dependent on resources provided. Setup fees are common for dedicated server hosting.
- *Colocation facilities* house the web site operator's server and storage hardware in a data center building, often inside a fenced enclosure with restricted access. This is the only web hosting category in which the hosting firm doesn't own the Web and application servers. The colocation provider supplies Internet connectivity, power, cooling, fire protection, data backup, and other security services. Colocation commonly is used for large content-oriented web site and medium-size or larger e-commerce sites. Setup and monthly charges are based on floor area, power consumption, and Internet traffic.

Colocation facilities suffered mightily when the dot-com bubble burst and several such organizations declared bankruptcy. Exodus Communications, one of the early large dedicated server hosting and collocation facilities, captured a NASDAQ record for 13 consecutive quarters of more than 40 percent growth and then opted for Chapter 11 bankruptcy in September 2001 during the demise of the dot-com bubble.

Rackspace Hosting, Inc. is a large web hosting firm that offers VPS hosting (which it calls *cloud hosting*) and specializes in *managed hosting*, which includes dedicated server hosting and collocation, targeting small and medium-sized businesses (SMBs). Rackspace launched its Mosso division in February 2008 to compete in the cloud computing market. The company acquired in October 2008 JungleDisk, an online backup service, and Slicehost, a virtualized server provider, to enhance its competitive stance against Amazon Web Services' EC2, Simple Storage Services, and Elastic Block Storage. By early 2009, Rackspace was managing more than 40,000 servers and devices for customers around the globe.

Cloud Computing and Everything as a Service

Cloud computing services, like many other SOA implementations, are *composable*. Wikipedia defines a *highly composable system* as a system that ''provides recombinant components that can be selected and assembled in various combinations to satisfy specific user requirements. The essential attributes that make a component composable are that it be: self-contained (modular), that is, it can be deployed independently . . .; it may cooperate with other components, but dependent components are replaceable. It must also be stateless, which means it treats each request as an independent transaction, unrelated to any previous request.''

Following are the generally accepted recombinant components that contribute to delivering cloud computing:

- *Files [storage] as a Service*: FaaS, often called *Data Storage as a Service* (DaaS), lets users store files of various data types in a highly scalable hierarchical file system and retrieve them over the Internet as various Multipurpose Internet Mail Extension (MIME) types. FaaS was one of the first cloud-based services. Several Internet start-ups, such as SmugMug, DropBox,

Ozmo, and HolaServers, use Amazon Web Services' Simple Storage Service (S3) to hold graphic images and other files, charging users a small or no access fee. Microsoft Live SkyDrive is a FaaS provider that gives users up to 25GB of free file storage at no charge.

The term *Data Storage* or *Database as a Service* implies structured storage with at least some relational database management system (RDBMS) features, such as query capabilities, primary and foreign key indexes, and entity associations through simulated JOINs. Commercial cloud services, such as Amazon Web Services (AWS), Google App Engine (GAE), and Windows Azure, offer indexed Entity-Attribute-Value (EAV) tables and query languages having some relationship to SQL. Microsoft says SQL Azure Database (SADB) ''offer highly scalable and Internet-facing distributed database services in the cloud for storing and processing relational queries.'' SADB, Amazon SimpleDB, and GAE's DataStore offer advanced features that qualify them as Databases as a Service (DBaaS).

- *Software as a Service*: SaaS delivers a packaged or equivalent commercial software application to end users over the Internet with a subscription or usage-based pricing model, as opposed to a traditional lifetime license for a particular version. Examples include Microsoft Office Live, Microsoft Exchange Online, Microsoft SharePoint Online, Microsoft Dynamics CRM Online, and Salesforce.com. Microsoft was an early SaaS supporter with SOAP-based web services but has gradually migrated to promoting Software plus Services (S+S). *Application as a Service* is a synonym for SaaS.
- *Software plus Services*: S+S is Microsoft's marketing terminology for traditional licensed on-premises software offered as a hosted service by Microsoft or hosting partners. Hosting partners can offer virtualized private-labeled Microsoft server applications, such as Exchange or SQL Server, or value-added services to Microsoft-hosted applications, such as Dynamics CRM. The feature that distinguishes S+S is the ability for customers to run the equivalent services on premises. The most interesting example of S+S is Amazon Web Service's EC2 running Windows Server 2003 and SQL Server [Express] 2005 with Elastic Block Store data storage and S3 storage for Amazon Machine Images (AMIs) and EBS snapshot backups.
- *Infrastructure as a Service*: IaaS provides traditional data center resources, such as highly scalable virtualized computing power, memory and storage, over a network (typically, but not necessarily, the Internet) and usually with a subscription or per usage pricing model. IaaS is also called *utility computing*. Internet-delivered cloud examples include Amazon Web Services, GoGrid, and Flexiscale. IaaS or PaaS delivered over an intranet is called a *private cloud*.
- *Communication as a Service*: CaaS provides communication capability that is service-oriented, configurable, schedulable, predictable, and reliable, as well as network security, dynamic provisioning of virtual overlays for traffic isolation or dedicated bandwidth, guaranteed message delay, communication encryption, and network monitoring. CaaS is critical to meeting Service Level Agreements (SLAs) but usually is considered to be a component of SaaS, S+S, or IaaS.
- *Monitoring as a Service*: MaaS notifies the user of cloud computing or network outages, errors, or slowdowns. For example, Cloud Status is a simple iPhone application that monitors the status of Amazon Web Services, Google App Engine, and Twitter and reports whether service is normal, has problems, or is down. MaaS can contain auditing components for network vulnerability assessment or to verify SLA conformance and the accuracy of monthly usage charges. Some suppliers of MaaS services, such as RightScale, also provide instance deployment automation for increasing the number of running AMI instances during demand peaks and reducing the number as demand subsides.

- *Platform as a Service*: PaaS usually comprises at least these three distinct elements:
  - *Tools as a Service* (TaaS), which provides Web-based development tools and languages, such as Microsoft Visual Studio (for Visual C#, Visual Basic, IronPython, and IronRuby) or open-source Eclipse (primarily for Java). The Windows Azure Tools for VS 2008 include templates for creating Web, Worker, Web and Worker, and Cloud Sequential Workflow Services that can run under a local (developer) or cloud (production) Windows Azure instance (fabric). Google App Engine offers a hosted Python variant as well as webapp and Django frameworks.
  - *A virtualized runtime application platform* that enables running applications in the cloud, typically on top of an IaaS and delivered as SaaS. Amazon EC2 has pre-built AMIs for 32-bit and 64-bit Linux distributions, Windows Server 2003 R2 with SQL Server 2005, and Oracle databases, as well as 64-bit OpenSolaris. Windows Azure runs on Windows Server 2008 with a custom version of Microsoft's Hyper-V hypervisor. Google App Engine offers Python.
  - *FaaS* to persist the state of the runtime application in Amazon's Elastic Block Store, SimpleDB or S3, Google's BigTable, or Windows Azure Storage Services' tables and blobs.
- *Everything as a Service*: EaaS, XaaS, or *aaS is a subset of cloud computing, according to Wikipedia, which calls EaaS ''a concept of being able to call up re-usable, fine-grained software components across a network.'' What's missing in this definition is orchestrated interaction between the components to solve a business problem, which is often called *Integration as a Service*.

  HP is one of the major proponents of Everything as a Service. ''Topic 22: Creating a Business Operating Environment in the Global Services Ecosystem,'' one of HP Labs' 2008 Research Topics in its Innovative Research Programs, starts with these two paragraphs:

  > In this applied research project, HP Labs is investigating what customer service lifecycles and experiences are possible in an ''Everything as a Service'' model and prototyping underlying intellectual property to enable them. HP Labs' goal in this research area is to address the technical challenges that must be overcome to move a business task to services over the Internet.
  >
  > Shane Robison, HP's Chief Strategy and Technology Officer, has detailed a set of ''Everything as a Service'' predictions that he believes will shape the IT industry in years to come. One of his predictions is that ''by 2012, a Fortune 50 company will research, develop, and launch a major product using only Internet-based services.'' This opinion is supported by information available from industry analysts, such as Gartner and IDC. In this project, we ask: ''What would a corporation wishing to move to an 'Everything as a Service' model need to do?''

The preceding component definitions incorporate concepts and content from Wikipedia, blog posts by Geva Perry, David Linthicum, and James Urquhart, as well as the ''Toward a Unified Ontology of Cloud Computing'' research paper (`http://bit.ly/12BPZD`*,* `www.cs.ucsb.edu/~lyouseff/CCOntology/CloudOntology.pdf`*) by Lamia Youseff (University of California, Santa Barbara, California), and Maria Butrico and Dilma Da Silva (IBM T.J. Watson Research Center, Yorktown, New York).*

Cloud Computing Ontologies

The term *cloud computing* has yet to gain a meaning, set of technologies, or level of abstraction upon which all participants — observers, suppliers, and consumers — can agree. Catchall terminology, such as SOA, *utility computing,* or *open services,* isn't precise enough to identify the cloud-computing model accurately. This situation invites information scientists to attempt creating a cloud computing *ontology*. According to Wikipedia, ''ontology deals with questions concerning what entities exist or can be said to exist, and how such entities can be grouped, related within a hierarchy, and subdivided according to similarities and differences.''

Rising to the occasion, Lamia Youseff, a Ph.D candidate at the University of California, Santa Barbara, California, and Maria Butrico and Dilma Da Silva, researchers at the IBM T.J. Watson Research Center, Yorktown, New York, published in 2008 a ''Toward a Unified Ontology of Cloud Computing'' research paper that establishes the five-layer model shown in Figure 1-4 to define the relationships between SaaS, PaaS, IaaS, DaaS, CaaS, and HaaS.

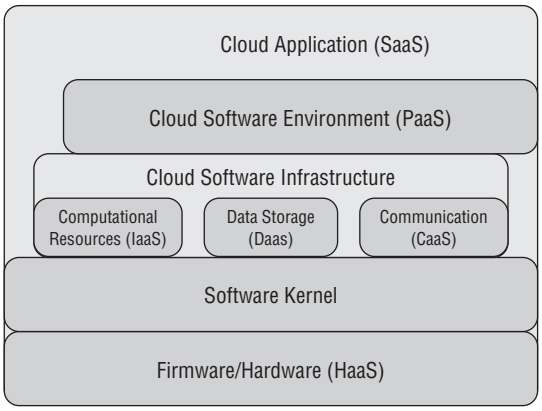

Figure 1-4: The five-layer structure of cloud computing as described in the ''Toward a Unified Ontology of Cloud Computing'' research paper.

Following is a high-level overview of Youseff's five-layer ontological model as used in this book:

- Youseff and her colleagues designate the top-level *Cloud Application Layer* as the access point for SaaS applications, such as Salesforce CRM and GAE, through Web portals.
- Cloud application developers use the *Cloud Software Environment Layer*, which provides support for a programming language and a set of application programming interfaces (APIs) ''to facilitate the interaction between the environments and the cloud applications,'' which leads to the Platform as a Service moniker. The Cloud Software Environment Layer is built on the Software Kernel and Firmware/Hardware layers and provides Computational Services (IaaS), Data Storage (DaaS), and Communication (CaaS) services. Virtual machines (VMs) commonly deliver IaaS, although Windows Azure offers the option of a dedicated server running Windows Server 2008. However, it's arguable that CaaS capabilities belong at the lower Firmware/Hardware (HaaS) level because off-premises HaaS isn't practical without CaaS.

  The authors classify Salesforce CRM with its Apex ''on-demand'' programming language and GAE, which supports Python, in the top two layers. Salesforce.com designates Force.com as a PaaS offering that supports 800+ applications from independent software vendors (ISVs)

and 80,000+ custom applications as of early 2009. GAE requires at least some familiarity with Python programming to provide useful services, but promises to support other languages in the future. The Windows Azure Platform's name and its dependence on Visual Studio 2008 place Microsoft's cloud offering squarely in the PaaS category. Windows Azure supports any programming language that conforms to the Common Language Runtime (CLR). The Youseff research paper didn't include a reference to Azure as of early 2009.

- The *Software Kernel* can be implemented as an OS kernel, hypervisor, virtual machine monitor and/or clustering middleware, or various combinations of these systems. Although grid applications played a significant role in early cloud computing implementations, the grid has given way to the hypervisor as the preferred software kernel for cloud computing because the latter abstracts hardware idiosyncrasies from the service. Adding CaaS makes this layer equivalent to traditional VPS Web hosting.
- The *Firmware/Hardware* layer is the physical computing, switching, and routing hardware that forms the cloud's backbone. The HaaS provider operates, manages, and upgrades the hardware on behalf of its lessees, who supply their own operating system and application software, and charges by the GB for data ingress and egress, similar to web server colocation. Leasing eliminates users' need to invest in building and managing data centers and might reduce the cost of power and insurance.

Other recognized and self-anointed cloud computing ''thought leaders'' offer numerous cloud computing definitions and ontologies. For example, David Linthicum of Blue Mountain Labs proposes and briefly describes the following 10 major cloud computing components in his ''Defining the Cloud Computing Framework'' blog post of January 18, 2009 (`http://bit.ly/iYgXc`, `http://cloudcomputing.sys-con.com/node/811519`):

- Storage-as-a-Service
- Platform-as-a-Service
- Database-as-a-Service
- Integration-as-a-Service
- Information-as-a-Service
- Security-as-a-Service
- Process-as-a-Service
- Management/Governance-as-a-Service
- Application-as-a-Service
- Testing-as-a-Service

According to its CCIF Mission & Goals Web page, the Cloud Computing Interoperability Forum (`http://bit.ly/YAmDP`, `http://groups.google.com/group/cloudforum/web/ccif-mission-goals`) ''was formed in order to enable a global cloud computing ecosystem whereby organizations are able to seamlessly work together for the purposes for wider industry adoption of cloud computing technology and related services. A key focus will be placed on the creation of a common agreed-upon framework/ontology that enables the ability of two or more cloud platforms to exchange information in a unified manner.'' The CCIF's Google Group had about 550 members in early 2009.

Attempts to create detailed taxonomies for cloud computing inevitably result in re-running the Red Queen's race: ''It takes all the running you can do to keep in the same place.'' Youseff's five-level ontology is likely to suffice until cloud computing reaches adolescence or the CCIF produces an alternative.

The Red Queen in Lewis Carroll's Through the Looking-Glass and What Alice Found There said, ''It takes all the running you can do to keep in the same place. If you want to get somewhere else, you must run at least twice as fast as that.''

According to Wikipedia, the Red Queen Principle can be restated as: ''For an evolutionary system, continuing development is needed just in order to maintain its fitness relative to the systems it is co-evolving with.''

Cloud Computing Concerns

Privacy and security are the two primary governance issues that IT managers face when attempting to reduce project budgets and improve scalability with PaaS, IaaS, SaaS, or any combination of cloud computing services. An *InformationWeek* magazine poll of 456 ''business technology professionals'' conducted in late 2008 found only 18 percent were using cloud services and 34 percent had no interest in doing so (`http://bit.ly/SqcTY`, `www.internetevolution.com/document.asp?doc_id=170782`). More than half of the respondents expressed concern about security; performance, control, vendor lock-in, and support were the next four most-expressed doubts about cloud computing. SaaS has been subject to the same litany of doubt, but the successes of Salesforce.com and AWS prove that governance issues can be overcome.

Following are the ''Five Fast Fixes'' to secure data in the cloud recommended by Mike Fratto, the author of *InformationWeek*'s ''Cloud Control'' article of January 26, 2009 that delivered the poll's conclusions:

1. *Define Your Governance Needs*: Are they internal, external, legal? List the requirements and how they're satisfied.
2. *Classify Your Data*: Before you can determine what data you can safely put in the cloud, you first have to classify and label it according to sensitivity and type.
3. *Choose Wisely*: Identify cloud vendors that can satisfy your processing and governance needs. Direct business leaders to walk away from the rest, no matter how attractive pricing is.
4. *Set Limits*: Define what the service provider can do with your data. Prohibiting the outsourcing of processing to a third party without your consent is basic.
5. *Put Rules in Writing*: Publish policies and procedures stating which cloud vendors can receive which types of data.

One of the most important elements of cloud governance is ascertaining where data is located when it's in the cloud. As Fratto observes in his article, it's possible for SaaS and other cloud providers to store data on servers that are under the control of another organization. Outsourcing data storage or backup by the computational services vendor can lead to two or more degrees of separation between your organization and its original data or backups.

Industry groups, such as the Payment Card Industry (PCI) require banks, online merchants and Member Service Providers (MSPs) to protect cardholder information by compliance with a set of security standards, which include MasterCard's Site Data Protection (SDP) program and Visa's Cardholder Information Security Program (CISP). The United States Health Insurance Portability and Accountability Act of 1996 (HIPAA) establishes standardized mechanisms for electronic data interchange (EDI), security, and confidentiality of all healthcare-related data, as well as security mechanisms to ensure confidentiality and data integrity for any health information that identifies an individual. Not knowing who had physical possession of your charge card data would surely fail a PCI audit, which might preclude a merchant from accepting credit or debit card payment; it would certainly violate HIPAA confidentiality

regulations, which can result in fines or other sanctions. However, these security and privacy issues also apply to outsourcing conventional data entry and processing operations, which is becoming increasingly commonplace, and aren't specific to cloud computing.

The Information Technology Laboratory of the U.S. National Institute of Standards and Technology (NIST) is contemplating the identification of minimal standards and architecture to enable federal agencies to create or purchase interoperable cloud computing capabilities. The ITL's Computer Security Division has the mission ''to provide standards and technology to protect information systems against threats to the confidentiality of information, integrity of information and processes, and availability of information and services in order to build trust and confidence in Information Technology (IT) systems.''

NIST's ''Perspectives on Cloud Computing and Standards'' presentation lists the following characteristics of a potential *Federal Cloud Infrastructure*:

- Agencies would own cloud instances or ''nodes.''
- Nodes would provide the same software framework for running cloud applications.
- Nodes would participate in the Federal cloud infrastructure.
- Federal infrastructure would promote and adopt cloud architecture standards (non-proprietary).
- ''Minimal standards'' refers to the need to ensure node interoperability and application portability without inhibiting innovation and adoption, thus limiting the scale of cloud deployments.

Subsequently, NIST issued their Draft NIST Working Definition of Cloud Computing v13 (`http://bit.ly/10TNdu`, `http://csrc.nist.gov/groups/SNS/cloud-computing/cloud-def-v14.doc`) *and Presentation on Effectively and Securely Using the Cloud Computing Paradigm v18* (`http://bit.ly/17PKbM`, `http://csrc.nist.gov/groups/SNS/cloud-computing/cloud-computing-v22.ppt`). *The Obama White House is expected to be the first major federal government user of cloud-computing services.*

When this book was written, there were no non-proprietary cloud architecture standards; node interoperability was the target of a CCIF spinoff called the Unified Cloud Interface (UCI) project (`http://bit.ly/EHRrp`, `http://code.google.com/p/unifiedcloud`) started by Enomaly (`http://bit.ly/16DldY`) founder and chief technologist, Reuven Cohen. According to Cohen, UCI's ''concept is to provide a single interface that can be used to retrieve a unified representation of all multi-cloud resources and to control these resources as needed.'' Cohen writes in the project's UCI_Requirements Wiki:

> The key drivers of a Unified Cloud Interface (UCI) is ''One Abstraction to Rule Them All'' — an API for other APIs. [UCI would be a] singular abstraction that can encompass the entire infrastructure stack as well as emerging cloud centric technologies through a unified interface. What a semantic model enables for UCI is a capability to bridge both cloud-based APIs such as Amazon Web Services with existing protocols and standards, regardless of the level of adoption of the underlying API's or technology. The goal is simple, develop your application once, deploy anywhere at anytime for any reason.

Cohen posits that ''you can't standardize what you can't define,'' so a cloud computing taxonomy/ontology will play an important role in UCI. The final goal appears to be to prevent cloud vendor lock-in by making it possible to migrate a deployed PaaS, IaaS, or SaaS project from an external to a private cloud or to another vendor's cloud infrastructure quickly and with minimal inconvenience or cost.

NIST promises to create these additional ''Special Publications'' in 2009 and later:

- Securing cloud architectures
- Securing cloud applications
- Enabling and performing forensics in the cloud
- Centralizing security monitoring in a cloud architecture
- Obtaining security from third-party cloud architectures through service-level agreements
- Security compliance frameworks and cloud computing (for example, HIPAA, FISMA, SOX)

You can expect to find draft versions of these publications from a link to the Special Publications (800 Series) on the Computer Security Division's Computer Security Resource Center Publications page (`http://bit.ly/17d9qq`, `http://code.google.com/p/unifiedcloud`).

If NIST can come up with a set of non-proprietary security features for an interoperable, non-proprietary cloud architecture, Information Technology Laboratory will make a substantial contribution to cloud computing's future.

Summary

Cloud computing is an emerging technology that threatens to reach ''next best thing'' status in 2009 and 2010 while throwing off the remnants of its Wild West ancestry. Spurred by tightened cost controls on fixed asset purchases, enterprise-scale IT departments will migrate beta deployments to full production on Windows Azure, Amazon EC2/EBS/S3, Google App Engine, GoGrid, or other commercial cloud platforms. IDC found cloud computing in October 2008 to be ''accelerating adoption and poised to capture IT spending growth over the next five years.'' The key to cloud computing's growth is monthly charges based on usage instead of massive investment in on-premises server and networking facilities.

Off-premises web site hosting was an early precursor of cloud computing, but organizations such as Amazon.com and Google were the first purveyors of Platform as a Service with Amazon Web Services EC2 and the Google App Engine. Organizations that specialize in web site hosting, such as Rackspace Hosting, Inc., began to expand their traditional service repertoire to offer Everything as a Service in the last half of 2008. Microsoft was late to the cloud-computing party with its introduction of the Windows Azure Platform in late October 2008.

Defining generic cloud computing is difficult because there's no generally accepted ontology or taxonomy of its services. SOA introduced Software as a Service and Microsoft's Software+Services implementations, but Platform, Infrastructure, Computing, Storage, Communications, and Hardware as services also have their place in cloud computing's attempt to provide Everything as a Service. Three information scientists have proposed a five-layer model that includes SaaS, PaaS, IaaS, DaaS, CaaS, and HaaS.

The cloud symbology implies unlimited or restricted access via the public Internet. Therefore, security for cloud-based applications and data is one of upper management's primary governance concerns. Today's clouds rely primarily on token-based user authentication and authorization schemes, but federated access control integrated with enterprise directories is in most cloud purveyors' plans. Another issue is vendor lock-in because currently no standards exist for cloud interoperability and such standards might never be agreed upon. Organizations that don't trust existing Internet access control protocols can create their own on-premises ''private clouds'' behind corporate firewalls, but doing so requires capital investment in IT infrastructure and negates the primary justification for cloud computing.