Chapter 8

Cloud Computing

8.1 What Is Cloud Computing?

"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." This is the vision of Google chief executive officer Eric Schmidt, speaking at a search engine conference in 2006. Since then, *cloud computing* has been a buzzword worldwide [277]. This was the first high-profile usage of the term; however, the first mention of cloud computing was in a 1997 paper entitled "Intermediaries in Cloud-Computing: A New Computing Paradigm" by R. Chellappa [276].

The term *cloud* was used as a metaphor for the Internet, based on the cloud drawing used in the past to represent the telephone network, and later to depict the Internet in computer network diagrams as an abstraction of the underlying infrastructure it represents.

Much like the Internet of Things (IoT), cloud computing is a natural evolution of related, existing, and new concepts in the information and communications technologies (ICT) arena, based on the widespread adoption of virtualization (first paper published in 1959 by C. Strachey [138]), cluster computing [139,140,141], grid computing, service-oriented architecture (proposed by Gartner in 1996), web services, parallel and distributed file systems [150], load balance and batch scheduling [142], autonomic, and utility computing technologies. In fact, the cloud computing term collided with many other terms that were already catchphrases in the ICT industry, such as SaaS (software as a service), grid computing, utility computing, PaaS (platform as a service), on-demand services, pervasive computing, and so on. Cloud computing provides computation, software, data access, and storage services that do not require end-user knowledge of the physical location and configuration of the system that delivers the services. Details are abstracted from end-users, who no longer have the need for expertise in, or control over, the technology infrastructure "in the cloud" that supports them.

The underlying concept of cloud computing dates back to 1961, when John McCarthy proposed the concept of utility computing and IBM started to rent its mainframe computing resources as "electric utility" to Wall Street via remotely connected dumb terminals.

Amazon played a key role in the development of cloud computing. After the dot-com bubble, most data centers were using as little as 10 percent of their capacity at any one time, just to leave room for occasional spikes. Having found that the new Amazon infrastructure resulted in significant internal efficiency improvements, Jeff Bezos initiated a new product development effort to provide cloud computing to external customers, and launched Amazon Web Service on a utility computing basis in 2006, which is now categorized as IaaS (infrastructure as a service).

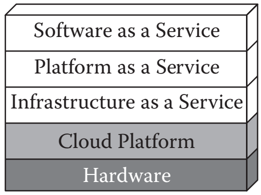

Figure 8.1 Cloud hierarchy. (From Ralf Teckelmann, Anthony Sulistio, and Christoph Reich, "A Taxonomy of Interoperability for IaaS," in Lizhe Wang, Rajiv Ranjan, Jinjun Chen, and Boualem Benatallah (Eds.), *Cloud Computing: Methodology, Systems, and Applications*, Boca Raton, FL: CRC Press, 2011.)

Salesforce was founded in 1999 by Marc Benioff, who is regarded as the leader of what he has termed the "no software" movement. In 2001, salesforce.com pioneered the multitenant SaaS model, a new application-delivering mechanism, a step beyond the application service provider model that made companies such as Exodus a great success in dot-com times. SaaS provides immediate benefits at reduced risks and costs, thanks to the rapid development and maturity of the Internet infrastructure, among other factors (as shown in Figure 8.1). In 2008, the force.com PaaS on-demand development platform was launched and became a new pillar of cloud computing, which has three pillars: IaaS, PaaS, and SaaS. All three pillars can provide services.

8.2 Grid/SOA and Cloud Computing

Much like the Internet of Things, the technological foundation of cloud computing is distributed computing based on communication networks. One of the most directly related works is the use of cluster of workstations (COWs) and networks of workstations.

Google's server farms were based on the same philosophy of COW. In fact, the MapReduce system is a batch processing extension of the scatter–barrier–reduce primitives of MPI/PVM

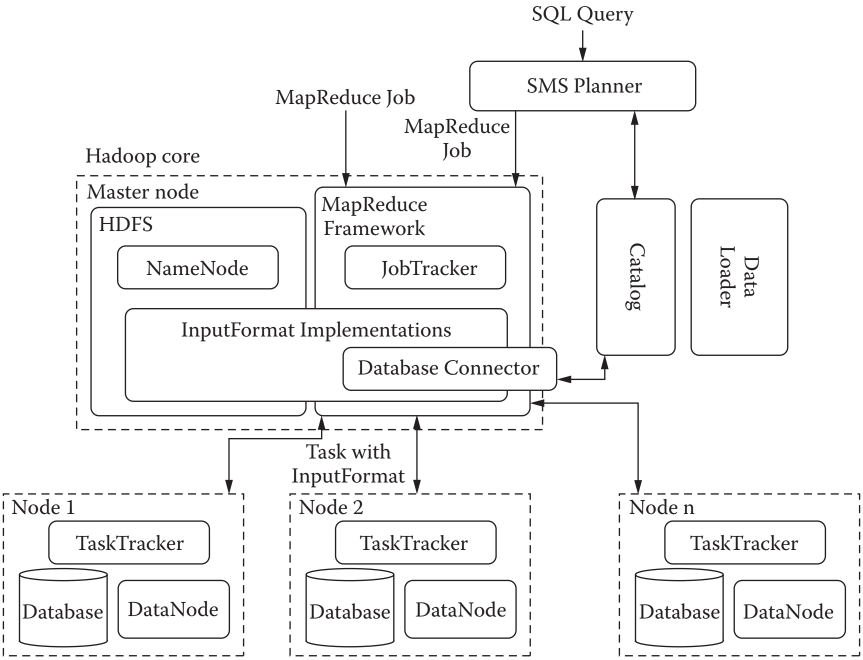

Figure 8.2 Hadoop is a batch system for embarrassingly parallel niche applications.

[143,144]. The well-known Hadoop is an open-source implementation of Google's Bigtable, GFS, and MapReduce [145,146,147] by Doug Cutting et al. in 2004. Hadoop is a high-throughput computing batch processing system, a niche application customized for embarrassingly parallel Internet-based massive data processing, as shown in Figure 8.2. However, as the Internet-related data and users increase rapidly, Hadoop has been widely used and become almost a nickname of cloud computing.

Grid computing is the direct technological ancestor of cloud computing, which also has roots in the COW technology; some of the well-known cloud systems such as Eucalyptus and OpenNebula are directly transformed from earlier grid computing research and development systems. (Much like *cloud*, the term *grid* is chosen as an analogy to the electrical power grid that provides consistent, pervasive, dependable,

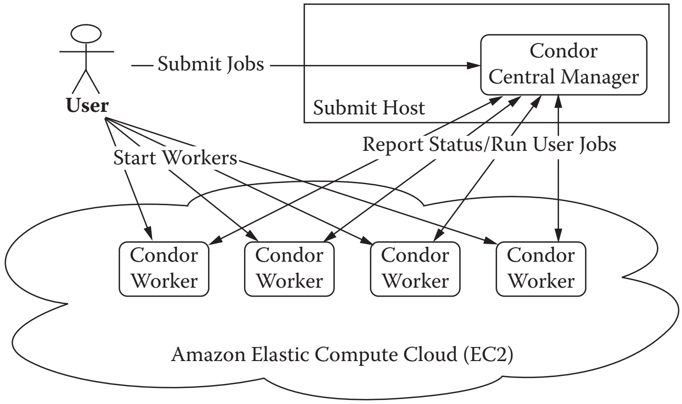

Figure 8.3 Condor scheduler in Amazon Web Services.

transparent access to electric power, irrespective of its source.) The grid computing concept was credited to Ian Foster of Argonne National Laboratory when he initiated the Globus project in 1994 based on the works of PVM/MPI, PBS/Condor [148,149] (as job schedulers of high-performance computing or parallel supercomputing systems), and so on. All of those technologies are generally referred to as cluster computing (other examples include Beowulf, Linux Virtual Server, MOSIX, BONIC) in a nutshell. Besides the basic parallel and distributed computing environments provided by middleware such as PVM and MPI, the job scheduler plays an important role as workload and resource management systems in building the grid and cloud computing/clustering infrastructure. For example, the Condor scheduler can be used in the Amazon system as shown in Figure 8.3. (The author worked in the LoadLeveler team, which was a job scheduler based on Condor [142], and participated in the ASCI–Blue Pacific project to build the then-world's-fastest massively parallel processing supercomputer, as the job scheduling system coordinator in 1996.)

One of the key features or functionalities of grid and cloud computing is providing a single system image (or a single parallel virtual machine) that hides the underlying scalable,

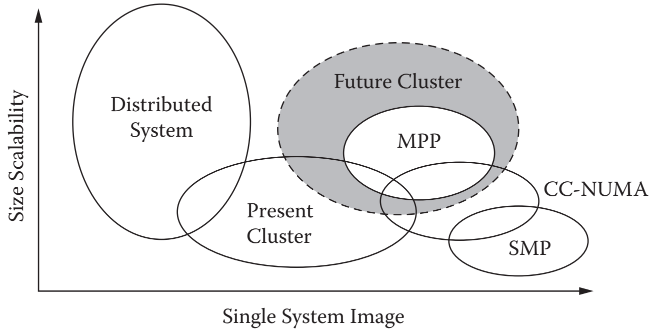

Figure 8.4 Single system image is the core.

elastic infrastructure (such as the Amazon backend server and storage farms) based on clustering technologies (as shown in Figure 8.4) and provisions a unified user interface via web services (such as the Amazon Web Services) and SOA over the Internet.

Virtualization is another important concept often mentioned in the cloud computing context. There are two sides of the virtualization coin: single system virtualization (SSV, i.e., one-to-many virtualization) and multisystem virtualization (MSV, i.e., many-to-one virtualization). This categorization of SSV and MSV for cloud computing was proposed by the author [75]. Utility computing started with SSV when IBM provided computing resources for rent via networked dumb terminals to Wall Street in the 1970s as mentioned before. A mainframe computer was virtualized into multiple virtual computers (as shown in Figure 8.5) via *hypervisor* technology so that

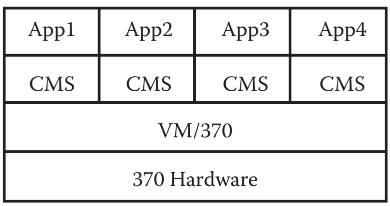

Figure 8.5 Earliest hypervisor.

enterprise users feel like a dedicated computer is providing services to them.

Modern SSV technologies are similar to the hypervisor technologies that IBM used decades ago. The purpose is to simulate multiple computers on top of one computer—run multiple operating systems on one computer hardware to enable maximum usage of the ever-increasing power of a single computer such as a PC and increase efficiency of overall resources. For example, one of the most important uses of SSV in earlier times was to simulate all the operating systems on a few servers so that a startup company in Silicon Valley could test their software products on all operating systems without having to buy all kinds of computers. SSV can be further categorized as three types of virtual machine monitors: Type 1 (hypervisor), Type 2, and hybrid [240].

The virtualization (currently almost a synonym of VMWare) talked about in the context of cloud computing currently is SSV, which makes many believe that SSV is a must for cloud computing. In fact, one-to-many virtualization is not required to build a cloud computing system, although it enables new platforms to run on legacy environments, and it helps to consolidate and simplify the management of the system by making the nodes of the system homogenous, thus simplifying the handling of issues such as (fault tolerance) check-pointing and migration (e.g., VMWare vMotion) when some of the nodes run into failures.

On the other hand, many-to-one virtualization is the foundation of cloud computing. MSV refers aggregately to the use of the aforementioned distributed and parallel clustering technologies such as COW, high-performance computing (HPC), grid computing, high-throughput computing, high-availability computing, and so forth to build a single, gigantic, parallel virtual computer or a single centralized service-providing virtual resource that serves many users for a plethora of applications. Some sample MSV architectures are shown in Figure 8.6. SSV

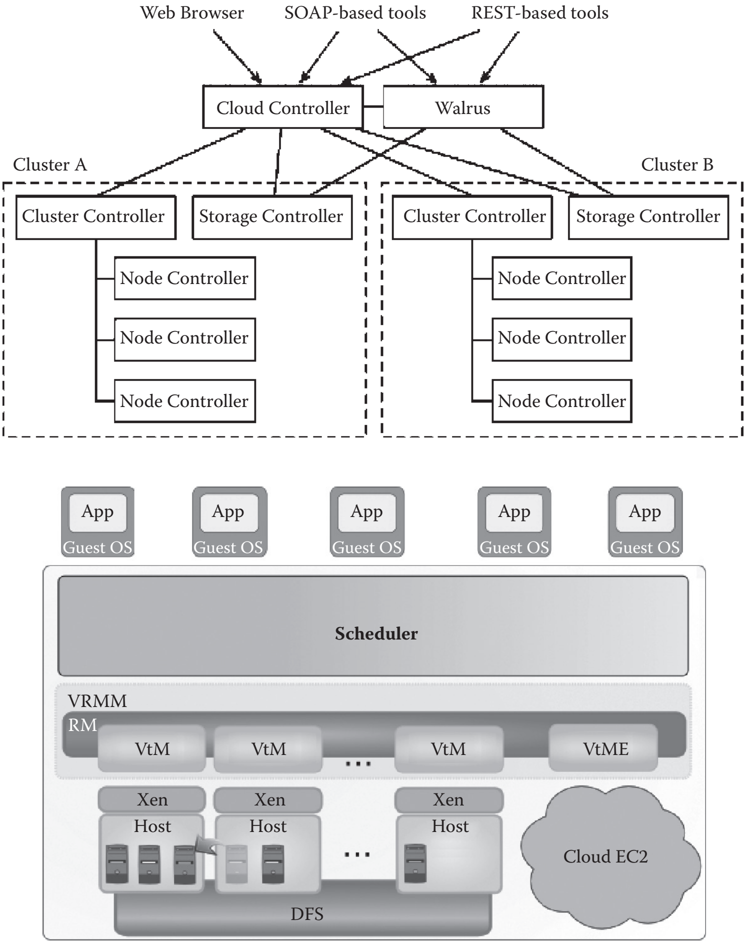

Figure 8.6 Many-to-one virtualization is the foundation of cloud computing.

technologies can be used at the node level of MSV but are not required.

The computing and storage resource are delivered to the end users using SOA (including SOAP or REST-based web services, SaaS, EAI, etc.) via the Internet, sometimes via intranet and extranet for private cloud applications. Many of the technologies and protocols of the SOA standard stack [241] can be used in all of the three layers: IaaS, PaaS, and SaaS.

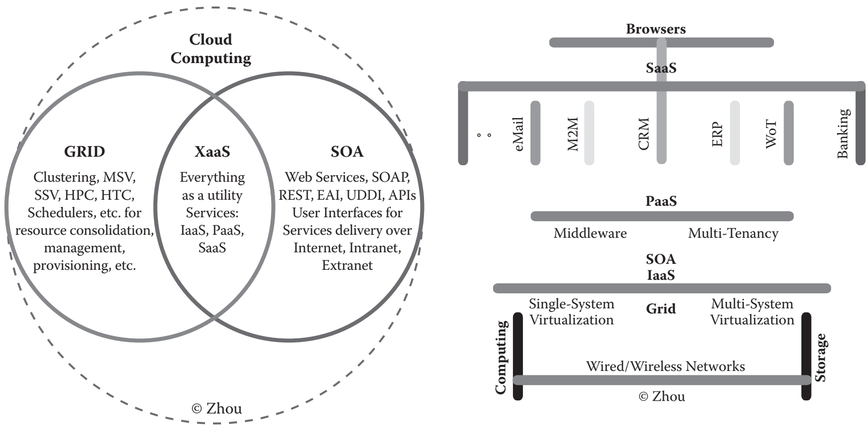

Figure 8.7 Cloud computing is the fusion of grid and SOA.

To summarize, cloud computing is the fusion of grid computing and SOA technologies to provide everything as utility-style services, as shown in Figure 8.7 [75]. It is "a large-scale distributed computing paradigm that is driven by economies of scale, in which a pool of abstracted, virtualized, dynamically-scalable, managed computing power, storage, platforms, and services are delivered on demand to external customers over the Internet," as defined by Ian Foster [278]. The graphic on the right of Figure 8.7 is based on the Chinese word for *cloud* that depicts the technologies, resources used, and application models.

8.3 Cloud Middleware

Again, much like the Internet of Things, the cloud computing system is also a multitiered architecture built on a middleware stack as shown in Figure 8.8.

At the lowest machine virtualization (SSV) level, there are middleware that help reduce the overhead of virtualization. SSV is useful and widely used, but it does not come cheap. The performance cost of virtualization, for I/O-intensive workloads in particular, can be heavy. Common approaches to

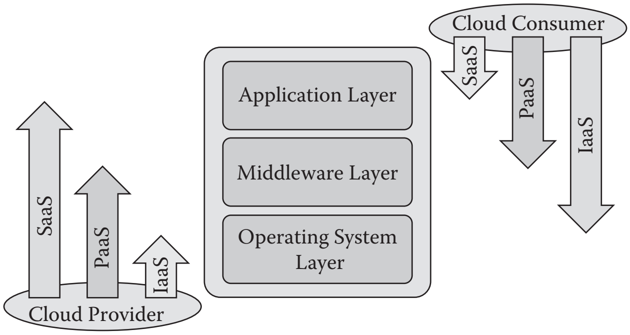

Figure 8.8 Multitiered cloud architecture based on middleware.

solve the I/O virtualization overhead focus on the I/O stack, thereby missing optimization opportunities in the overall stack. As an example, VAMOS [242], built by IBM, is a novel middleware architecture that runs its middleware modules at the hypervisor level. VAMOS reduces I/O virtualization overhead by cutting down on the overall number of guest/hypervisor switches for I/O intensive workloads. Applying VAMOS to a database application improved its performance by up to 32 percent. Here, the middleware concept is extended to include software that does interprocess communication not necessary over a network.

At the cluster computing or grid computing level, many types of work are done by middleware. The parallel computing environments such as PVM and MPI are (HPC) middleware by definition; the Hadoop system and the job scheduler such as Condor, LoadLeveler, and others are all middleware. The HPC middleware fills the gap that the operating systems and the programming languages lack to support parallel computing [151]. A number of grid middleware initiatives (such as http://www.eu-emi.eu/) have been formed by interested members, mostly in the scientific computing community. Some of those middleware are aggregately referred to as grid middleware [152,153] and listed as follows:

- Low-level middleware
- MPI, Open MPI
- PVM (parallel virtual machine)
- POE (parallel operating environment, IBM)
- Middleware for file systems and resources
- MPI-IP
- PVFS/GPFS (parallel virtual file system/general parallel file system IBM)
- Sector-Sphere
- Condor/PBS/LoadLeveler (IBM)
- High-level middleware
- Beowolf
- Globus Toolkit
- Gridbus
- Legion
- Unicore
- OSCAR/CAOS/Rocks
- OpenMosix/NSA/Perceus

Many research works [243] demonstrate a typical grid computing system (other similar systems include Distributed European Infrastructure for Supercomputing Application [DEISA], Teragrid, Enabling Grids for E-Science [EGEE], NorduGrid, SEE-GRID, OSG, etc.) and its components based on grid middleware before cloud computing gained momentum. A grid computing system aims to serve all kinds of applications as a more generic cloud computing system than Hadoop.

As discussed before, grid computing is the foundation of cloud computing infrastructure, so grid middleware is the basis of IaaS middleware. In addition, the IaaS middleware (part of cloud middleware [244]) may include components such as system management, network management, billing and operation support systems, provision, configuration, automation, orchestration, service level agreement (SLA) management, and so on.

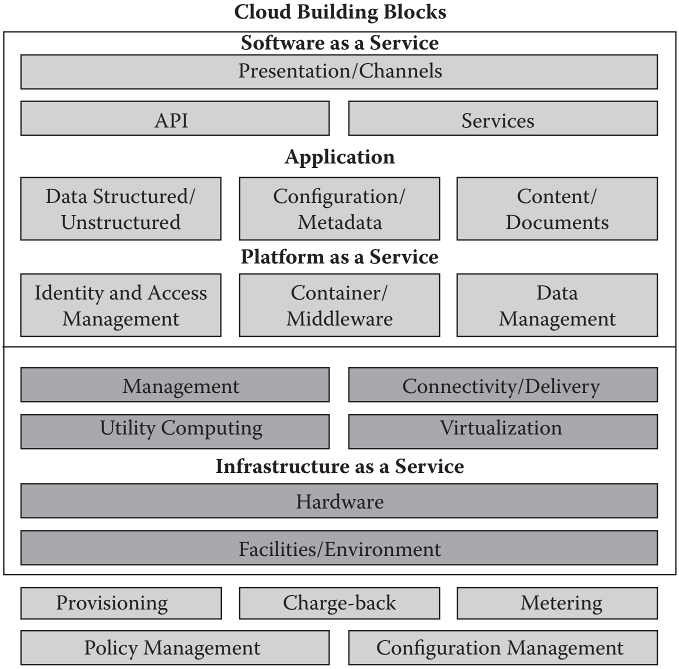

Figure 8.9 PaaS and cloud middleware.

From the distributed enterprise computing standpoint, almost all of the EAI and business-to-business (B2B) middleware described in the previous chapters are needed to build cloud computing systems for enterprise and commercial applications. They all are part of cloud middleware, particularly part of the PaaS middleware. Multitenancy [245] is one of the basic functions of PaaS middleware, evolving from the traditional platform middleware. The multitenant efficiency functionalities of a PaaS platform are often required and implemented in a traditional middleware such as the three-tiered application servers described in the previous chapters and as shown in Figure 8.9. More comprehensive guides on all of the building blocks of cloud computing have been discussed and depicted [267,268].

The PaaS middleware is often referred to as the cloud middleware that underpins and supports the SaaS applications.

The graphic in [246] depicts the different deployment options of a PaaS middleware in cloud systems: the more middleware is shared, the cloud systems scale to larger numbers of tenants and with lower operational costs.

The well-known middleware [247] quadrant from Gartner depicts the market landscape of middleware vendors. Salesforce.com was included for the first time in 2010, most likely because its foundational platform (force.com) is recognized as one of the most important cloud (PaaS) middleware vendors.

To summarize, the cloud middleware consists of two kinds of middleware—IaaS and PaaS middleware—and their relation is shown in Figure 8.10. (Note: SaaS are not middleware, they are applications on top of middleware.)

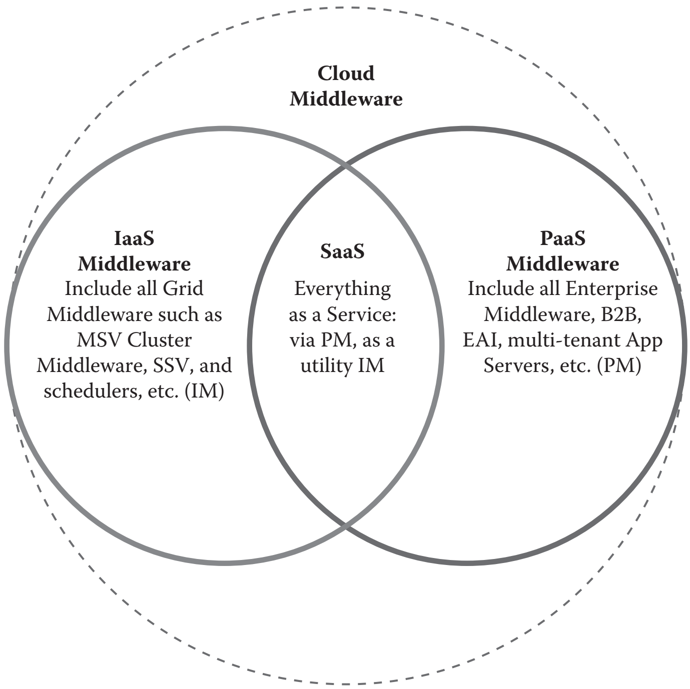

Figure 8.10 Cloud middleware.

8.4 NIST's SPI Architecture and Cloud Standards

The U.S. National Institute of Standards and Technology (NIST) has come up with a widely accepted definition [154] that characterizes important aspects of cloud computing: Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

This cloud model is composed of the following:

- Three service models: IaaS, PaaS, and SaaS
- Four deployment models: private cloud, public cloud, community cloud, and hybrid cloud
- Five essential characteristics: on-demand self-service, broad network access, resource pooling, rapid elasticity, and measured service

An earlier version (Version 14) of the specification also listed 12 foundational elements or enablers.

Cloud computing is an evolving paradigm. The comparative benefits of the different service models of cloud computing are compared in http://itcandor.net/2010/11/22/cloud-computing-benefits-q410/. The NIST specification is a milestone that clarifies and settles most of the confusion and arguments about cloud computing. It can be used as a starting point for standardization. Electronics and Telecommunications Research Institute (ETRI) of Korea proposed to address standards on nine aspects (http://www.etri.re.kr/eng/res/res_0102020301.etri):

- Definition, taxonomy, terminologies
- Provisioning model

- Business process
- Security
- Interoperability
- Legality
- Environmental issues
- Architecture
- Availability

The NIST specification covers a few of the aspects, such as the standardization of definition, taxonomy, and terminologies.

Some of the standardization in the grid computing domain provided a foundation for extended work, such as the MPI, openMP standards, as well as job description language standards (such as Job Submission and Description Language [155] and Basic Execution Service [156] of Open Grid Forum—Open Grid Services Architecture) for job scheduling.

Table 8.1 lists some of the cloud computing standardization organizations and their websites. The following are some of the works done by those standards organizations:

- NIST: Working definition of cloud computing
- Distributed Management Task Force: Open Virtualization Format, Open Cloud Standards Incubator, DSP-IS0101 Cloud Interoperability White Paper V1.0.0
- Cloud Management Working Group: DSP-IS0102 Architecture for Managing Clouds White Paper V1.0.0, and DSP-IS0103 Use Cases and Interactions for Managing Clouds White Paper V1.0.0
- European Telecommunications Standards Institute: TC cloud definition
- Standards Acceleration to Jumpstart Adoption of Cloud Computing: 25 use cases
- Open Cloud Consortium: Open Cloud Testbed, Open Science Data Cloud, benchmarks, reference implementation

Table 8.1 List of Standardization Efforts

National Institute of Standards and Technology (http://csrc.nist.gov/groups/SNS/cloud-computing/index.cfm)
Distributed Management Task Force (http://www.dmtf.org)
The European Telecommunications Standards Institute (http://www.etsi.org)
Open Grid Forum (http://www.ogf.org)
Open Cloud Computing Interface Working Group (http://www.occi-wg.org)
Object Management Group (http://www.omg.org)
Storage Networking Industry Association (http://www.snia.org)
Open Cloud Consortium (http://www.opencloudconsortium.org)
Organization for the Advancement of Structured Information Standards (http://www.oasis-open.org)
Association for Retail Technology Standards (http://www.nrf-arts.org)
The Open Group (http://www.opengroup.org)
Cloud Security Alliance (http://www.cloudsecurityalliance.org)

- The Cloud Computing Interoperability Forum: framework/ontology, semantic web/resource description framework, unified cloud interface
- The Open Group: SOA, The Open Group Architecture Framework
- Association for Retail Technology Standards: Cloud Computing White Paper V1.0
- TM Forum: Cloud Services Initiative, Enterprise Cloud Leadership Council Goals, Future Collaborative Programs, BSS/OSS/SLA
- ITU-T FG Cloud: Introduction to the Cloud Ecosystem: Definitions, Taxonomies and Use Cases;
- Global Inter-Cloud Technology Forum: Japan, Interoperability

- Cloud Standards Coordination: Standards Development Organization Collaboration on Networked Resources Management
- Open Cloud Manifesto (http://www.opencloudmanifesto.org/)
- Open Grid Forum: Open Cloud Computing Interface, Open Grid Services Architecture
- Cloud Security Alliance: Security Guidance for Critical Areas of Focus in Cloud Computing, Cloud Controls Matrix, Top Threats to Cloud Computing, CloudAudit
- Storage Networking Industry Association: Cloud Storage Technical Work Group, Cloud Data Management Interface

8.5 Cloud Providers and Systems

In five short years, cloud computing has gone from being a quaint technology to being a major catchphrase. It started in 2006 when Amazon began offering its Simple Storage Service and soon following up with its Elastic Compute service, and Google's CEO Eric Schmidt's speech about cloud computing. Just like the Internet of Things, the market potential is huge. Many vendors, old and new, have joined the gold rush to provide cloud services and products. There are many forecasts about market size of cloud computing. For example, Gartner estimated that, among the three SPI segments, SaaS generates most of the revenue, because it directly creates value for the end users. IaaS helps reduce the costs of organizational users, which has the fastest growth. Gartner predicts the change of revenue on percentage among the three SPI segments between 2010 (SaaS: 72%, PaaS: 26%, IaaS: 2%) and 2014 (SaaS: 61%, PaaS: 36%, IaaS: 3%). However, this prediction may not count the revenue of PaaS as a middleware product sold independently, but only the part of the revenue of PaaS as a hosted service, in which case PaaS is sold as part of SaaS most of the times.

Revenue generated by cloud technology companies, excluding the larger, more mature SaaS segment, is forecast to grow from $984 million in 2010 to $4 billion in 2013, according to The 451 Group, representing a compound annual growth rate of 60 percent. Including SaaS, total cloud technology vendor revenue was $8.5 billion in 2010, expected to grow to $16.3 billion in 2013, a compound annual growth rate of 24 percent. Of course, the amount spent by companies on cloud products and services is much larger, with Gartner estimating worldwide cloud services revenue in 2010 of $68.3 billion, an increase of 16.6 percent from $58.6 billion of 2009. Gartner estimates that the cloud services revenue will reach $148.8 billion in 2014.

We will give an overview of the current cloud providers based on their participation in providing the building blocks as depicted in the graphics noted below that include the services and products for the three SPI pillars, as well as additional products such as development tools, security frameworks, system management software, adaptor frameworks, and so on.

There are many top 10, top 20, and top 50 listings of cloud providers on the web that can be easily found with Google search. The graphic at http://www.opencrowd.com/assets/images/views/views_cloud-tax-lrg.png [157] lists some of the top cloud providers in the SPI and general software categories.

Gartner published its Magic Quadrant (http://www.cloudbusinessreview.com/wp-content/uploads/2011/01/magic-quadrant-gartner.png) about the leading IaaS providers and the emphasis (http://blogs.pcmag.com/miller/assets_c/2010/10/Cloud%20Vendor%20Emphais-16535.php) of the best-known cloud providers. Another well-known cloud vendor taxonomy graphic is from Peter Laird of Oracle (BEA) created in 2009, which can be found at http://farm4.static.flickr.com/3312/3597138202_496ae06a68_o.png.

The cloud computing boom has brought a surge of opportunity to the open-source world. Open-source developers and users are taking advantage of these opportunities. Many open-source applications are now available on a SaaS basis. Other open-source projects have taken the steps necessary to make them easy to use in the cloud, for example, by making preconfigured images available through Amazon Web Services or other public clouds. However, most open-source developers are contributing to the growth of cloud computing by creating the tools that make cloud computing feasible. They offer infrastructure, middleware, and other software that make it easier for companies to develop and run their applications in the cloud. The following is a list of open-source projects:

- Open-source IaaS and PaaS projects: OpenStack, cloud.com Cloud Stack, OpenNebula, Eucalyptus, AppScale, Scalr, Traffic Server, RedHat Cloud, Cloudera (Hadoop), Puppet, Enomaly, Joyent, Globus Nimbus, Reservoir, Amanda/Zmanda, XCP, TPlatform, and so forth
- Open-source SaaS projects: Zoho, Phreebooks, Pentaho, Palo BI Suite, Jaspersoft, Processmaker, eyeOS, Alfresco, SugarCRM, SourceTap, KnowledgeTree, OpenKM, Collabtive, Zimbra, Feng Office, Open ERP, Openbravo, Compiere, Orange HRM, JStock, Ubuntu, OpenProj, openSIS, TimeTrex, GlobaSight, and others

To summarize, the author has created a free-style panoramic view graphic [75] of existing cloud providers and their products and services in five layers (including vendors and products in China that are mostly at the PaaS and SaaS layers):

- Chip and hardware supports for virtualization: Intel-VT (VT-x, VT-x2), AMD-V (SVM), SUN/Oracle UltraSPARC T1, T2, T2+, SPARC T3, and others
- Hypervisors (one-to-many SSV virtualization) vendors and products
- IaaS (many-to-one MSV virtualization) grid/cluster computing, web services–based delivery vendors and products

Figure 8.11 Five-layer panoramic view of cloud vendors and products.

- PaaS (multitiered middleware) vendors and products
- SaaS vendors and products (due to the vastly large number of SaaS vendors and products, only some of them are listed; some of the IoT SaaS services such as Pachube are also listed. See Figure 8.11.)

8.6 Summary

In this chapter, we talked about what cloud computing is, its relationship with earlier concepts, and paradigms such as grid computing, cluster computing, SOA, SaaS, and the like. The importance of middleware in cloud computing is described and emphasized. The systematic specification of NIST and many standardization efforts were introduced and discussed. And finally, a comprehensive summarization of the currently existing vendors, service providers, and systems is provided.

Much like cloud computing, the Internet of Things is also about distributed computing. The two have many things in common and many shared underlying technologies and paradigms, which will be discussed in the next chapter.