Chapter 2
A Taxonomy, Survey, and Issues of Cloud Computing Ecosystems

Bhaskar Prasad Rimal, Eunmi Choi, and Ian Lumb

Abstract Cloud computing has emerged as a popular computing model to support processing of volumetric data using clusters of commodity computers. Nowadays, the computational world is opting for pay-for-use models. Hype and discussion aside, there remains no concrete definition of cloud computing. This chapter describes a comprehensive taxonomy for cloud computing architecture, aiming at a better understanding of the categories of applications that could benefit from cloudification and that will address the landscape of enterprise IT, management services, data governance, and many more. Then, this taxonomy is used to survey several cloud computing services such as Google, Force.com, and Amazon. The usages of taxonomy and survey results are not only to identify similarities and differences of the architectural approaches of cloud computing, but also to identify the areas requiring further research.

2.1 Introduction

Cloud computing appears to be a highly disruptive technology, which is gaining momentum. It has inherited legacy technology as well as new ideas on large-scale distributed systems. The concept of cloud computing addresses the next evolutionary step of distributed computing. The goal of this computing model is to make a better use of distributed resources, put them together in order to achieve higher throughput, and be able to tackle large-scale computation problems. The computing power nowadays is easily available for massive computational processing. For example, image processing on Amazon Elastic Cloud Computing (EC2) [20] for New York Times is a great success story for Amazon. The input of 11 million articles (4-terabytes of TIFF data) was processed successfully using Amazon Simple Storage Service

E. Choi (✉)
School of Business IT, Kookmin University, Jeongneung-Dong,
Seongbuk-Gu, Seoul, 136-702, Korea
e-mail: emchoi@kookmin.ac.kr

N. Antonopoulos and L. Gillam (eds.), *Cloud Computing: Principles, Systems and Applications*, Computer Communications and Networks,
DOI 10.1007/978-1-84996-241-4_2,

(Amazon S3), EC2 as hardware, and Hadoop [19] with MapReduce as software framework [29]. The output data was 1.5 terabytes of PDF format, processed within 24 h at a computation cost of just $240. Google has used MapReduce to process 20 petabytes[1] of data a day [1]. Similarly, Google used MapReduce running on 1,000 servers to sort 1 terabyte of data in just 68 s [40]. Hive/Hadoop [51] cluster at Facebook stores more than 2 Petabytes of uncompressed data and routinely loads 15 Terabytes of data daily [50]. Such scenarios prove that cloud computing is becoming cheaper, faster, and easy for massive distributed processing and scalable storage.

Cloud computing is not a completely new concept for the development and operation of web applications. It allows for the most cost-effective development of scalable web portals on highly available and fail-safe infrastructures. In the cloud computing system, we have to address different fundamentals like virtualization, scalability, interoperability, quality of service, failover mechanism, and the cloud delivery models (private, public, hybrid) within the context of the taxonomy. The taxonomy of cloud includes the different participants involved in the cloud along with the attributes and technologies that are coupled to address their needs and the different types of services like "XaaS" offerings where X is software, hardware, platform, infrastructure, data, and business.

The taxonomy is more than defining the fundamentals that provides a framework for understanding the current cloud computing offerings and suggests what is to come. It is provoking those who would seek a single, canonical definition of the term cloud computing. The main idea behind this taxonomy is to find out the technical strength, weakness, and challenges in the current cloud systems and we suggest what should be done in future to strengthen the systems. The emergence of cloud fabrics will enable new insights into challenging engineering, medical, and social problems. The cloud taxonomy should not be a gigantic construct that muddies the water of service development. It should be consistent with a set of principles that provides architectural and design guidance on the usage and crafting of services. It should provide understandable and consistent guidelines that provide clarity and reusability. For that reason, this taxonomy is intentionally small and fuzzy, i.e. the boundaries of the service layers are not rigid with regard to an emerging service.

Taxonomic information is essential for cloud service providers, enterprise firms, and border authorities to detect, manage, and control invasive alien components. Taxonomy identifies and enumerates the components of cloud computing that are providing basic knowledge underpinning management and implementation of the cloud spectrum. Taxonomy is more than defining the fundamentals that provides a framework for understanding the current cloud computing offerings and suggests what is to come. The criteria for defining the taxonomy is based on the core ideas of distributed systems for massive data processing. The criteria focus on cloud architecture, virtualization management, services, fault tolerance, and we analyze mechanisms like load balancing, interoperability, and scalable data storage.

---

[1]Disk Storage: 1,000 Megabytes = 1 Gigabytes, 1,000 Gigabytes = 1 Terabytes, 1,000 Terabytes = 1 Petabytes

This chapter tries to define taxonomy and survey of "Cloud Computing" based on recent advances from academia and industry as well as our experience. This chapter also describes the comparative study of different cloud service providers and their systems. The chapter is organized as follows. Section 2 introduces the background and related work. Section 3 defines the taxonomy of cloud computing. Section 4 describes the classification and comparative study of cloud computing ecosystems. Findings are discussed in Section 5. Some of the issues and opportunities are explained in Section 6. Finally, Section 7 concludes the chapter.

2.2 Background and Related Work

XaaS implies everything as a service [17] like SaaS (Software as a Service), PaaS (Platform as a Service), HaaS (Hardware as a Service), DaaS ([Data center, Database, Desktop] as a Service), IaaS (Infrastructure as a Service), BaaS (Business as a Service), FaaS (Framework as a Service), OaaS (Organization as a Service), etc. There are many cloud computing systems like Amazon EC2, Google App Engine (GAE), Microsoft Azure, IBM Blue Cloud, Nimbus, 3 Tera, etc. There is, however, no standard taxonomy, as everyone tries to define cloud computing and its services in their own way. There has been prior work reflecting the taxonomy of cloud computing. The taxonomy described by the Cloud Computing Use Case Discussion Group [23] is categorized into three views: service developer, service provider, and service consumers. This taxonomy does not cover the data holding governance structure. Crandell [21] defines a taxonomy based on product offerings. He divided the product offerings into three layers, namely Application in the cloud (Salesforce and other SaaS vendors), Platform in the cloud (GAE, Moso, Heroku), and Infrastructure in the cloud (Amazon Web Services, Flexiscale). This taxonomy is attractive for any company with an application that runs in a data center or with a hosted provider, that does not want to reinvent the wheel or pay a premium. Laird's [22] Cloud Vendor Taxonomy gives the classifications and vendors with that related group. This taxonomy divides the cloud vendors into Infrastructure (Public Cloud, Private Cloud), platform (Biz User Platforms, Dev Platform), services (Billing, Security, Fabric Mgmt, System Integrators), and applications. This taxonomy gives a visual map of the SaaS, PaaS, and cloud computing industries. Forrester's Cloud Taxonomy [24] is categorizing cloud services by IT-Infrastructure vs. Business value and by the level of privacy. This taxonomy focuses on the dimensions of privacy and business value. It focuses on the modes of cloud computing (Public Scale-Out Clouds, Public Server Cloud, Virtual Private Scale-Out Clouds, Virtual Private Server Clouds, Private Clouds, Virtual Private SaaS, Public SaaS, PaaS, On-Premises, ASP Concepts, etc.) To provide an even clearer and more explicit view over cloud computing applications, we propose several incremental enhancements of those taxonomies. In this paper, we will adjust, refine, and extend those taxonomies, making them even more suitable and flexible for cloud computing.

2.3 Taxonomy of Cloud Computing

Several taxonomies [21,22,24] of the cloud computing blueprint can be found, but most were created from the perspective of vendors that are part of the landscape and not from the perspective of enterprise IT, the consumers of cloud services, and software. This taxonomy is split into seven major sections as shown in Fig. 2.1.

This includes architecture, virtualization management, core services, security, data governance, and management services. The subtaxonomy core services include replication, discovery, and load balancing. A scalable, robust, and intelligent replication mechanism is crucial to the smooth operation of cloud computing. Another subtaxonomy security includes encryption/decryption, privacy, federated identification, authorization, and authentication. Ultimately, the cloud computing taxonomy describes certain patterns in *how to understand the cloud components and how to do things*. At the same time, it needs to provide some specific grounding to address the complex issues of integration of services within cloud computing that focus on providing computable semantic interoperability.

2.3.1 Cloud Architecture

Cloud architecture is the design of software applications that use Internet accessible and on-demand service. Cloud architectures are underlying an infrastructure that is

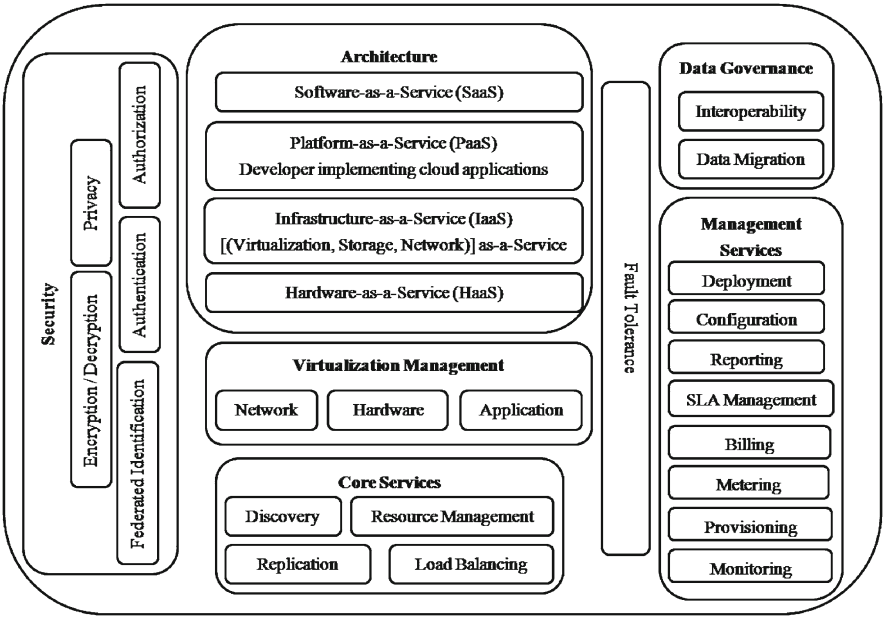

Fig. 2.1 A taxonomy of cloud computing

used only when it is needed to draw the necessary resources on-demand and perform a specific job, then relinquish unneeded resources, and dispose of them after the job is done. The services are accessible anywhere in the world, with the cloud appearing as a single point of access for all the computing needs of consumers. Cloud architectures address the key difficulties surrounding large-scale data processing.

2.3.1.1 Services and Modes of Cloud Computing

There are different categories of cloud services such as infrastructure, platform, and applications. These services are delivered and consumed in real time over the Internet. We discuss these services in the broader view.

Software-as-a-Service (SaaS)

Software as a Service is a multitenant platform. It uses common resources and a single instance of both the object code of an application as well as the underlying database to support multiple customers simultaneously. SaaS [3,4], commonly referred to as the Application Service Provider model, is heralded by many as the new wave in application software distribution. Examples of the key providers are SalesForce.com (SFDC), NetSuite, Oracle, IBM, and Microsoft, etc.

Platform-as-a-Service (PaaS)

Platform-as-a-Service provides developers with a platform, including all the systems and environments, comprising the end-to-end lifecycle of developing, testing, deploying, and hosting of sophisticated web applications as a service delivered by a cloud base. It provides an easier way to develop business applications and various services over the internet (e.g. a real state service provider). Creating and maintaining an infrastructure is the most time-consuming work in the on-premises systems. PaaS was invented to solve exactly this problem. Key examples are Google AppEngine, Microsoft's Azure, Heroku.com, etc. Compared with conventional applications development, this strategy can slash development time, offer hundreds of readily available tools and services, and quickly scale.

Hardware-as-a-Service (HaaS)

In HaaS model, the vendor allows customers to license the hardware directly. According to Nicholas Carr [18], "the idea of buying IT hardware or even an entire data center as a pay-as-you-go subscription service that scales up or down to meet your needs. But as a result of rapid advances in hardware virtualization, IT automation, and usage metering and pricing, I think the concept of hardware-as-a-service – let's call it HaaS, may at last be ready for prime time." This model is advantageous to the enterprise users, since they do not need to invest in building and managing data centers.

Infrastructure-as-a-Service (IaaS)

Infrastructure-as-a-Service is the delivery of computer infrastructure as a service. Aside from higher flexibility, a key benefit of IaaS is the usage-based payment scheme. This allows customers to pay as they grow. Another important advantage is that of always using the latest technology. Customers can achieve a much faster service delivery and time to market. Key examples are GoGrid, Flexiscale, Layered Technologies, AppNexeus, Joyent, and Mosso/Rackspace, etc. Basically, cloud mode can be defined by three types (1) Private Cloud: Data and processes are managed within the organization without the restrictions of network bandwidth, security exposures, and legal requirements that using public cloud services across open, public networks might entail. Some examples are Amazon VPC, Eucalyptus, Enomaly, VMWare, Redplaid, Platform computing, and Intalio. (2) Public Cloud: Describes cloud computing in the traditional mainstream sense, whereby resources are dynamically provisioned on a fine-grained, self-service basis over the Internet, via web applications/web services, from an off-site third-party provider who shares resources. Some examples are Zimory, Azure, SunCloud, Amazon EC2, SymetriQ, GigaSpaces, Rackspace, and Flexiscale. (3) Hybrid Cloud: The environment is consisting of multiple internal and/or external providers. Some example are RightScale, Asigra Hybrid Cloud Backup, Carpathia, Skytap, and Elastra.

2.3.2 Virtualization Management

Virtualization Management is the technology that abstracts the coupling between the hardware and operating system. It refers to the abstraction of logical resources away from their underlying physical resources in order to improve agility, flexibility, reduce costs, and thus enhance business value. Basically, virtualizations in cloud are of different types such as server virtualization, storage virtualization, and network virtualization. A common interpretation of server virtualization is the mapping of single physical resources to multiple logical representations or partitions. In a virtualized environment, computing environments can be dynamically created, expanded, shrunk, or moved as the demand varies. Virtualization [2] is therefore extremely well suited to a dynamic cloud infrastructure, because it provides important advantages in sharing, manageability, and isolation. Different solutions are available to manage virtual machines such as XEN, VMWare, KVM, VirtualBox, Microsoft Virtual Machine Manager and many more. Virtualization management deals with the different types of virtualizations in the context of cloud computing such as Desktop Virtualization (Virtual PC), Network Virtualization, Storage Virtualization, Server Virtualization (Virtual Server), Application Virtualization (SoftGrid Application Virtualization), and Presentation Virtualization (Terminal Server). Storage capacity and performance are scalable because there is no central bottleneck. When expected demand exceeds higher server utilization, the storage can be scaled (horizontal scalability or vertically scalability) to meet them. One study from Gartner [25] shows that fewer than five million PCs were "virtualized" in 2006; by 2011, that figure will rise to between 480 million and 846 million. In another study, Gartner also estimated [26] that revenue from hosted virtual desktop will more than triple from $74.1 million to $298.6 million in 2009, while revenue from server virtualization management software will increase 42% from $913.9 million in 2008 to $1.3 billion in 2009. Revenue from server virtualization infrastructure will grow 22.5% from $917 million in 2008 to $1.1 billion in 2009. These data give a direction that is the major infrastructure for cloud computing. Therefore, it is the essential component for the cloud taxonomy. It has several benefits such as test and development optimization, resource maximization, business cost reduction, and much more.

2.3.3 Core Services

This section focuses on the core services of cloud computing. In core services, we will discuss discovery, replication, load balancing, and resource management in details.

2.3.3.1 Discovery and Replication

Service discovery promotes reusability by allowing service consumers to find the existing services. RESTful services [48] support discovery and reuse at design time. Replication can be used to create and maintain copies of an enterprise's data at these sites. When events affecting an enterprise's primary location occur, key application services can effectively be restarted and run at the remote location incurring no capital expenditure, only operational expenditure, until such time as the primary site is brought back online. Replication (Eager and Lazy) [54] keeps all replicas as a part of one atomic transaction. Replication technology is available in storage arrays, network-based appliances, and through host-based software.

2.3.3.2 Load Balancing

Load balancing prevents system bottlenecks due to unbalanced loads. It also considers implementing failover for the continuation of a service after failure of one or more of its components. This means that a load balancer provides a mechanism by which instances of applications can be provisioned and deprovisioned automatically without changing network configuration. This is an inherited feature from grid-based computing for cloud-based platforms. Energy conservation and resource consumption are not always a focal point when discussing cloud computing; however, with proper load balancing in place, resource consumption can be kept to a minimum. This not only serves to keep costs low and enterprises "greener," it also puts less stress on the hardware infrastructure of each individual component, making them potentially last longer. Load balancing also enables other important features such as scalability.

2.3.3.3 Resource Management

Cloud computing provides a way of deploying and accessing massively scalable shared resources on demand, in real time, and at affordable cost. Cloud resource management protocols deal with all kinds of homogeneous and heterogeneous resource environment. Management of virtualized resources, Workload and resource scheduling, Cloud resource provisioning with QoS, Scalable resource management solutions are the concerning points. Dynamic resource scheduling across a virtualized infrastructure for those environments is another issue for cloud.

2.3.4 Data Governance

When data begins to move out of organizations, it is vulnerable to disclosure or loss. The act of moving sensitive data outside the organizational boundary may also violate national regulations for privacy. In Germany, passing data across national boundaries can be a federal offence. Governance in the cloud "who and how" is the big challenge for enterprise clouds. Governance places a layer of processes and technology around services (location of services, service dependencies, service monitoring, service security, and so on) so that anything occurring will be quickly known [45]. There are some questions that need to be solved before mission critical data and functionality can be moved outside a controllable environment.

2.3.4.1 Interoperability

Interoperability means easy migration and integration of applications and data between different vendors' clouds. Owing to different hypervisors (KVM, Hyper-V, ESX, ESXi), VM technologies, storage, configuring operating systems, various security standards and management interfaces, many cloud systems are not interoperable. However, many enterprises want interoperability between their in-house infrastructure and the cloud. The issue of interoperability needs to be addressed to allow applications to be ported between clouds, or to use multiple cloud infrastructures, before critical business applications are delivered from the cloud. Most clouds are completely opaque to their users. Most of the time, users are fine with this until there is an access issue. In such situations, frustration increases exponentially with time, partly because of the opacity. Is a mechanism like a network weather map required? In other words, some form of monitoring solution like autonomous agents.

2.3.4.2 Data Migration

Data migration between data centers or cloud systems are important concerns of taxonomy. While migrating data, some considerations should be taken into account like no data loss, availability, scalability, cost–efficiency, and load balancing. User should be able to move their data and applications any time from one to another seamlessly, without any one vendor controlling it. Seamless transfer, as in mobile communication, is required for cloud computing to work. Many enterprises do not move their mission critical data and applications to the cloud because of vendor lock-in, security, governance, and many more complications.

2.3.5 Management Services

The management services contain deployment, monitoring, reporting, service-level agreement, and metering billing. We discuss these in detail.

2.3.5.1 Deployment and Configuration

To reduce the complexity and administrative burden across the cloud deployment, we need the automation process life cycle. RightScale Cloud Management Platform addresses three stages of the cloud application deployment lifecycle, namely design, manage, and deploy. Automated configuration and maintenance of individual or networked computers, from the policy specification, is very important in the computing arena; it improves robustness and functionality without sacrificing the basic freedoms and self-repairing concepts. That is why, to handle complex systems like cloud environment and data center, we need such configuration management. Tools such as cfengine [35], Chef from Opscode-chef [42], rPath [41], and Puppet are available as configuration management frameworks. These tools help software developers and engineers to manage server and application configuration by writing code, rather than running commands by hand.

2.3.5.2 Monitoring and Reporting

Developing, testing, debugging, and studying the performance of cloud systems is quite complex. Management cost increases significantly as the number of sites increases.

To address such problems, we need monitoring and reporting mechanisms. Monitoring basically monitors the SLA lifecycle. It also determines when an SLA completes and reports to the billing services. There are some services that monitor the cloud systems and produce health reports such as Hyperic HQ [32], which monitors SimpleDB, SimpleQueue Service, and Flexible Payment Service, all offered by Amazon. It collects the matrix and provides a rich analysis and reporting.

2.3.5.3 Service-Level Agreements (SLAs) Management

Users always want stable and reliable system service. Cloud architecture is considered to be highly available, up and running 24 h × 7 days. Many cloud service providers have made huge investments to make their system reliable. However, most cloud vendors today do not provide high availability assurances. If a service goes down, for whatever reason, what can a user do? How can users access their documents stored in the cloud? In such a case, the provider should pay a fine to the consumer as compensation to meet SLAs. An SLA specifies the measurement, evaluation, and reporting of the agreed services level standards such as [39]:

1. How raw quality measures will be used to evaluate agreed service component.
2. How the raw quality measures will be qualified as a service quality measure.
3. How the qualified quality measures will be used to estimate the service quality levels.
4. How the results of service evaluation will be reported.
5. How disputes on service-level evaluation will be resolved.

Currently, Amazon offers a "99.9% Monthly Uptime Percentage" SLA for Simple Storage Service (Amazon S3) and credit is limited to 10% [38]. Amazon credits 25% of charges if the availability drops below 99.0%, whereas 3Tera Virtual Private Data Center (VPDC) service will include a 99.999% availability SLA that is supposed to help assure customers about putting mission-critical apps and services in the cloud.

2.3.5.4 Metering and Billing

Transparent metering and billing will increase the trust level of users towards cloud services. Pay-as-you-go subscription or pay-as-you-consume model of billing and metering are popular for cloud. This service gets the status of the SLA, and invokes the credit service, which debits the user credit card or account and informs the user. There are many pricing strategies such as RAM hours, CPU Capacity, Bandwidth (Inbound/Outbound Data Transfer), Storage Space (gigabytes of data), Software License Fee), and Subscription-Based Pricing. There are some interesting new billing models such as GoGrid prepaid cloud hosting plan [33] and IDC cloud billing research [34], which are great examples of moving cloud pricing models towards telecom models.

2.3.5.5 Provisioning

Self-service application provisioning enables application developers to set up application infrastructure, such as Java application servers, databases, and messaging servers, without any help or assistance from infrastructure teams. Self-service application provisioning hides the complexity of the enterprise cloud from application developers and empowers them to set up and configure complex application infrastructure with the click of a button. By building a self-service portal for on-demand provisioning, we can reduce process overheads. Provisioning can help to manage the resource management, workload management, and autorecovery and task and process automation.

2.3.6 Security

Security is one of the main hurdles for enterprises to move their in-house data to public cloud. Most public cloud providers do not guarantee the security of the data while being transported to the public cloud. Many discussions around cloud computing are centered around this topic. In June-August 2009, several social networking sites such as Twitter, Facebook, Livejournal, and Google blogging pages were hit by DDoS attacks [43], [44]. DDoS are more robust and potentially simpler to implement in noisy environments such as EC2. Corporate information is not only a competitive asset, but it often contains information of customers, consumers, and employees that, in the wrong hands, could create a civil liability and possibly criminal charges. The key challenges of cloud security are performance, risk management, governance, design, and deployability. Building trust between various cloud stakeholders (users, corporations, network, and service providers) is a major consideration [49]. Establishing best practices of security in cloud computing for an end-user could be a good idea; for example, customers should talk with software vendors about licensing, or should know the network scheme.

2.3.6.1 Encryption/Decryption

Customers who worry about the privacy of their information should encrypt their data before moving it to the cloud. The provider should provide the utilities to simplify the process of encrypting the files and storing them in the cloud; similarly, for retrieval decryption will need. Cloud provider can use an Advanced Encryption Standard (AES) that may be AES-128, AES-192, or AES-256.

2.3.6.2 Privacy and Federated Identity

In cloud computing, a data center holds information that would more traditionally have been stored on the end-user's computer. This raises concerns regarding user privacy protection, since the users do not "own" their data. Also, the move to centralized services may affect the privacy and security of users' interactions. Federation is the act of combining data or identities across multiple systems. Federation can be done by a cloud provider or by a cloud broker. Each user can subscribe to a portal and be given an access card, which will be used to identify the subscriber at this particular portal or other portals in collaboration.

2.3.6.3 Authorization and Authentication

In public clouds, safeguards must be placed on machines to ensure proper authentication and authorization. Within the private cloud environment, one can track, pinpoint, control, and manage users who try to access machines with improper credentials. Single sign-on is the basic requirement for a customer who accesses multiple cloud services.

2.3.7 Fault Tolerance

In case of failure, there will be a hot backup instance of the application, which is ready to take over without disruption. Cloud computing outages extend into the more refined version of cloud service platforms. Some outages have been quite lengthy. For example, Microsoft Azure had an outage that lasted 22 h on March 13–14, 2008. Cloud reliance can cause significant problems if downtime and outages are removed from your control. Table 2.1 shows failover records from some cloud service provider systems. These are the significant downtime incidents. Reliance on the cloud can cause real problems when time is money.

Google has also had numerous difficulties with its Gmail and application services. These difficulties have generated significant interest in both traditional media and the blogosphere owing to deep-seated concerns regarding service reliability. The incidents mentioned here are just the tip of the iceberg. Every year, thousands of websites struggle with unexpected downtime, and hundreds of networks break

Table 2.1 Outages in different cloud services

Services and outage	Duration	Date
Microsoft Azure: malfunction in Windows Azure [5]	22 h	Mar 13–14, 2008
Gmail and Google Apps engine [6]	2.5 h	Feb 24, 2009
Google search outage: programming error [7]	40 m	Jan 31, 2009
Gmail: site unavailable due to outage in contacts system [8]	1.5 h	Aug 11, 2008
Google AppEngine partial outage: programming error [9]	5 h	Jun 17, 2008
S3 outage: authentication service overload leading to unavailability [10]	2 h	Feb 15, 2008
S3 outage: single bit error leading to gossip protocol blowup [11]	6–8 h	Jul 20, 2008
FlexiScale: core network failure [12]	18 h	Oct 31, 2008

or have other issues. So, the major problem for cloud computing is how to minimize outage/failover to provide reliable services. It is important to adopt the well-known Recovery-Oriented Computing (ROC) paradigm [46] in large data centers. Google uses Google File System (GFS) [47] or distributed disk storage; every piece of data is replicated three times. If one machine dies, a master redistributes the data to a new server.

2.4 Classification and Comparison between Cloud Computing Ecosystems

Even though there has been some comparative research on cloud computing from academia and enterprise perspectives, there remains an absence of a comprehensive technical study. We study cloud computing systems in terms of various classifications such as infrastructure technology, and solutions, PaaS provider, and open source. This section provides a technical comparison of several technologies and cloud providers. Tables 2.2–2.3 compare between different infrastructure technologies and solution providers such as Amazon Web Service (AWS), GoGrid, Flexiscale, and Moso. Tables 2.4–2.6 compares different SaaS and PaaS service providers such as Google AppEngine (GAE), GigaSpaces, Azure, RightScale, SunCloud, and Salesforce.com (SFDC). Similarly, Tables 2.7–2.8 compare open source cloud-based services like Eucalyptus, Open Nebula, Nimbus, and Enomaly.

2.5 Findings

Based on the proposed taxonomy, comprehensive technical studies, and survey, we notice some of the findings from different cloud computing systems that may help in future for new development and improvement on the existing systems.

2.5.1 Cloud Computing Infrastructure Technology and Solution Provider

In EC2 architecture, users are able to monitor and control their applications as an instance but not as a service. To achieve service manageability, the following capabilities are required: application-defined SLAs, such as workload capacity and concurrent computational tasks, dynamic provision of additional services to handle additional workload, and "Focal Server" approach. AWS is becoming popular as a de facto standard; many cloud systems are using a similar API. Eucalyptus is an open-source implementation of the AWS APIs. The biggest concern of current cloud computing system is auditing of the security controls and mechanism in terms of

Table 2.2 Cloud computing infrastructure technology and solution provider(1\2)

Features	AWS	GoGrid	Flexiscale	Rackspace cloud
Computing architecture	EC2 allows uploading XEN virtual machine images to the infrastructure and gives client APIs to instantiate and manages them	Dedicated computer resources on grid architecture	– Data center architecture – Autonomically reconfiguring for infrastructure to cater to fluctuations in the demand	– Merge the idea of cloud computing with the traditional managed/shared server environment – Private Cloud's single-tenant architecture
Virtualization management	Xen hypervisor	Xen hypervisor	XEN-based hypervisor to provide hardware virtualization on Intel VT	VMware ESX Server
Service	IaaS, Xen images	IaaS	IaaS	IaaS
Load balancing	Balance incoming requests and traffic across multiple EC2 instances by using Round-Robin algorithm	F5 load balancing, Round-Robin algorithm	Uses migration of virtual servers between physical nodes. It supports both horizontal and vertical scaling	By request balancing algorithm -Simple software Load Balancer using a Cloud-Server-Scale CloudServer horizontally or vertically
Fault tolerance	System should automatically alert, failover, and re-sync back to the "last known state" as if nothing had failed	Instantly scalable and reliable file-level backup service	It provides full self-service for start/stop/delete, and changes memory/CPU/ storage/IPs of virtual dedicated servers	Share an IP between two servers. Heartbeat application runs on both Master and Slave

Table 2.3 Cloud computing infrastructure technology and solution provider (2\2)

Features	AWS	GoGrid	Flexiscale	Rackspace cloud
Interoperability	Support horizontal Interoperability, e.g. interoperability among EC2, Eucalyptus, etc.	Interoperable with other clouds such as GigaSpaces	Applications can be deployed once and managed transclouds to run on Amazon, GoGrid, and Mosso	– Open Cloud manifesto – Provides open specs for Cloud Servers APIs and Cloud Files APIs
Storage	– Amazon Simple Storage Service (S3) – Amazon SimpleDB	– Connecting each server to Private Network – Transfer protocols (RSYNC, FTP, SAMBA, SCP) to transfer data to and from Cloud Storage	Fully virtualized high-end SAN/NAS back-end and uses a NetApp FAS3050 (hybrid SAN/NAS device, maximum storage capacity of 168TB spread over 336 drives)	– Storage is based on Rackspace Cloud Files – Uses limelight network
Security	AWS Secret Access Key, Type II (SAS70 Type II) certification – firewall, X.509 certificate, SSL-protected API	– Secure VLAN management – PrimeCloud service for hosted private cloud with no resources shared with other customers	Provides Virtual Private Servers, which gives privacy of a dedicated server	Encrypted communication channel, API Access Key, session authentication token
Programming framework	Amazon Elastic MapReduce framework. Supports Java, Ruby, PHP, etc.	Supports languages: Java, Python , Ruby, PHP	Flexiscale API support C, C #, C++, Java, PHP, Perl, and Ruby	Supports .NET, Java, Python, Ruby, PHP

Table 2.4 Cloud computing PaaS and SaaS provider(1\3)

Features	GAE	GigaSpaces	Azure	RightScale	SunCloud	Salesforce.com
Computing architecture	Google geo-distributed architecture	Space base architecture	An internet scale cloud services platform hosted in Microsoft data centers, which provides an OS and a set of developer services	– Multiserver clusters – Gives virtual private Servers monitoring system – Cloud management platform – Provides Elastic IPs	– Solaris OS, and Zetta-byte File System (ZFS) – Q-layer enabled for Data Warehouse and enterprise resource planning – Open dynamic infrastructure management strategy	Multitenant architecture with metadata-driven model
Virtualization management	Multitenancy	GigaSpace Service Virtualization Framework	Hypervisor (based on Hyper-V)	Xen hypervisor	– Hypervisor (Sun xVM Server) – OS (Solaris Containers) – Network (crossbow) – Storage (COMSTAR, ZFS) and applications (Glassfish and Java CAPS	Multitenancy architecture. It improves separation between shared and private data and logic.
Service	PaaS	PaaS	PaaS	PaaS	PaaS	SaaS Confined to API

Table 2.5 Cloud computing PaaS and SaaS provider(2\3)

Features	GAE	GigaSpaces	Azure	RightScale	SunCloud	Salesforce.com
Load balancing	Automatic scaling and load balancing	Performed through GigaSpaces high-performance communication protocol over EC2	Built-in hardware load balancing	High Availability Proxy load balancing in the cloud	Horizontal scalability, Vertical scalability	Load balancing among tenants
Fault tolerance	– Automatically pushed to a number of fault-tolerant servers – App Engine Cron Service	Uses OpenSpaces Service Virtualization Framework (SVF)'s failover capabilities	If a failure occurs, SQL data services will automatically begin using another replica of the container	Basic, intermediate, and advance Failover Architectures for using Elastic IPs	Resource based scheduling of service request	Self-management and self-tuning
Storage	Bigable distributed storage	In-memory data grid technique uses for front-end to the database. MySQL acts as in-sync persistence storage in the background	– SQL Server Data Services (SSDS) – Allows storing binary large objects (blobs) and can be geo-located	Open storage model, MySQL backups are Elastic Block Store (EBS) are saved to S3	Sun cloud storage WebDAV API, and Sun Cloud storage object API	Force.com database, which is tightly integrated with Apex programming language

Table 2.6 Cloud computing PaaS and SaaS provider(3\3)

Features	GAE	GigaSpaces	Azure	RightScale	SunCloud	Salesforce.com
Interoperability	Interoperability between platforms of different vendors and programming languages	Interoperability between different programming languages such as Java, .NET, and C++	Interoperable platform can be used to build new applications to run from the cloud or enhance the existing applications	Integrated management dashboard, application can be deployed once and managed across clouds	– Open source philosophy and java principles – Interoperability for large-scale computing resources across multiple clouds	Application level integration between different clouds
Security	– Google Secure Data Connector – SDC uses TLS-based server authentication – SDC uses RSA/128-bit or higher AES CBC/SHA	Support Amazon Security Groups, built-in SSH tunneling	– Security token service (STS) creates Security Assertion Markup – Language token according to rule	Private VLANs – Assign Multiple Security Groups	– User-provisioning and meta directory solution – Process and user rights management trusted extensions	– SysTrust SAS 70 Type II – Users and security, programmatic and platform security framework
Programming framework	MapReduce programming framework that support Python, Java as Java Servlet API, JDO, and JPA	Supports for Spring/Java, .NET, C++	Microsoft .NET, PHP	Ruby, PHP, Amazon's Simple Queue Service	Solaris OS, Java, C, C++, FORTRAN, RESTful, Java, Python, Ruby	Supports for .NET, C # Apache Axis (Java and C++)

Table 2.7 Open source based cloud computing services (1\2)

Features	Eucalyptus	OpenNebula	Nimbus	Enomaly elastic computing platform
Computing architecture	Ability to configure multiple clusters, each with private internal network addresses, into a single cloud	– Focused on the efficient, dynamic, and scalable management of VMs within data centers – Based on Haizea scheduling	– Client-side cloud computing interface to Globus-enabled TeraPort cluster – Context Broker combines several deployed virtual machines into "turnkey" virtual clusters	– A clustered virtual server hosting platform; ElasticDrive, a distributed remote storage system; and GeoStratus, a private content delivery network – Uses GlusterFS for scaling to several petabytes
Virtualization management	Xen hypervisor	Xen, KVM, and on-demand access to Amazon EC2	Xen Virtualization	KVM supports Xen OpenVZ and Sun's Virtual Box, Xen hypervisor
Service	IaaS, Xen images	IaaS	IaaS	IaaS
Load balancing	Simple load balancing cloud controller	Nginx Server configured as load balancer, used round-robin or weighted selection mechanism	Launches self-configuring virtual clusters, i.e. the context broker	– Uses user-mode load-balancing software with its own network stacks that runs over Linux and Solaris in the form of a virtual server – Supports different load-balancing methods, including round-robin, random, hash, and least resource
Programming framework	Hibernate, Axis2, and Java	Java, Ruby	Python, Java	Ruby on rails, PHP, Python

Table 2.8 Open source based cloud computing services (2\2)

Features	Eucalyptus	OpenNebula	Nimbus	Enomaly elastic computing platform
Fault tolerance	Separate cluster within the Eucalyptus cloud reduces the chance of correlated failure	– The daemon can be restarted and all the running VMs recovered – Persistent database backend to store host and VM information	Checking worker nodes periodically and recovery	Overflow, disaster, and failover services
Interoperability	Multiple cloud computing interfaces using the same "back-end" infrastructure	Interoperable between intracloud services such as access Amazon EC2 and elastic hosts cloud via plug-in	Standards: "rough consensus and working code"	Cloud portability and interoperability to cross cloud vendors
Storage	Walrus (the front end for the storage subsystem)	– Database, persistent storage for ONE data structures – SQLite3 backend is the core component of the OpenNebula internal data structures	Provides secure management of cloud disk space giving each user a repository view of VM images and works with globus GridFTP	Multiple remote cloud storage services (S3, Nirvanix, and CloudFS), uses MySQL for data sharing
Security	WS-security for authentication, Cloud controller generates the public/private key	Firewall, virtual private network tunnel	PKI credential required Works with Grid proxies VOMS, Shibboleth (via GridShib), custom PDPs	"Clustered" handling of security

user level. Amazon S3 lacks in access control that supports delegation and auditing, and makes implicit trust assumptions between S3 and clients [52]. Amazon's work [13] towards Statement on Auditing Standards No.70: Service Organizations, Type II (SAS70 type II) certification may be helpful for those concerned with widely varying levels of security competency. Generally, this is better than no certification whatsoever. Some of the important security aspects of cloud-centric computing are secure cloud resource virtualization, security for cloud programming models, binary analysis of software for remote attestation and cloud protection, cloud-centric regulatory compliance issues, and mechanisms and foundations of cloud-centric threat models that need to be considered for future cloud work.

2.5.2 Cloud Computing PaaS and SaaS Provider

Google App Engine (GAE) provides a useful basis for people and companies to make web applications from scratch without needing to worry about infrastructure. GAE provides for automatic scaling and load balancing. This alone will be worth while for a certain class of application developers. GAE has some clear advantages and lowers the barriers to entry for startups and independent developers. The potential problem is lock-in that creates risk and more cost for long term. The lock-in is caused by custom APIs such as BigTable, Python launcher, accounts and transparent scaling for both Python scripts and database. Google App Engine uses master/slave replication between data centers. They chose this approach to provide low latency writes, data center failure survival, and strong consistency guarantees.

GigaSpaces use an In-Memory Data-Grid (IMDG) technique to manage state data in a database, which bridges the bottleneck of scalability. It provides all the basic features of a high-end Data Grid as well as unique features, such as continuous query and seamless integration with external data sources, and makes it extremely easy to deploy, modify, and ensure high availability for applications running on Amazon EC2.

GigaSpaces's Space-Based Architecture (SBA) approaches are based on the Tuple Space model [53] that can meet the challenge of running low-latency transactional applications in a highly distributed environment such as Amazon EC2.

Security isolation is managed via virtualization in Azure. The Azure Fabric Controller is a service that monitors, maintains, and provisions machines to host the application that the developer creates and stores in the Microsoft cloud. Azure storage provides persistent, redundant storage in the cloud. It can store data in three different ways such as Blobs (large binary data), Queues (service communication abstraction), and Tables (service state and user data). Storage can be geo-located, which means you can choose in which region it can be hosted.

The agile nature of Sun Cloud provides multiple hardware architectures to customize systems for workload, multitenancy, and resource sharing among a large pool of users allowing centralized infrastructure with lower costs. Sun modular data center is flourishing and ten times faster to deploy than a conventional data center.

Sun's open storage provides a unique business model, which provides snapshot, replication, and compression without additional cost for data services. Hybrid cloud architecture is very important. One of the nice mechanisms of it is the open storage model that is provided by Sun Cloud, which is a new and unique business model as well.

SFDC introduces the Force.com metadata-driven, multitenant, internet application platform. In multitenant architecture, a single instance of the hosted application is capable of servicing all customers (tenants). Not all clouds are using virtualization. Clouds like GAE and SFDC use completely different technologies to create multitenancy. From the developer point of view, multitenancy is not the main event.

2.5.3 Open Source Based Cloud Computing Services

The role of open source cloud computing is to build mechanisms around digital identity management [14], and outline technological building blocks that are needed for controllable trust and identity verification. Nimbus supports the OASIS WSRF standard [32] that defines a framework and uses web services to model and access stateful resources. Enomaly cloud is focusing on the issue of interoperability, which is essential for enterprise cloud system. Most of the open source clouds are providing IaaS.

2.6 Comments on Issues and Opportunities

There are some issues related to mechanisms such as security, privacy, (erosion of) data integrity, load balancing, interoperability, and scalable storage. Cloud computing services often provide common business applications online that are accessed from a web browser, while the software and data are stored on the servers. One of the issues is an integration of data and application across clouds. This involves leveraging technology such as EAI (enterprise application integration), EII (enterprise information integration or federated database), and ESB (enterprise service bus). The market prognosis suggests raising the subscription fees as cloud vendors provide higher performance, scalability, availability, better support, and security. Transmitting huge volumes of multimedia data across clouds will continue to be a challenge, and needs further research. Discovery and composition of the services between multiple clouds is also a promising arena for enterprise cloud. Clouds have a different paradigm for resource utilization, so they need a different paradigm for managing these resources. Each previous revolution in computing also revolutionized how resources were managed. Collaborating amongst different technologies, businesses, and people in cloud computing will be an issue that will enable the enterprise to play a role as well. Quality assurance and information security are always challenging. Researchers should leverage identity and security management for business units. Furthermore, there are opportunities for the provision of a new range of privacy services. As the user requirement changes, functionality and privacy requirements may change, and so privacy requirements need to be reassessed at regular intervals [30]. Policy-based dynamic privacy design patterns may be a better technique for cloud computing. Cloud computing brings some novel attacks that have not figured in much of the security discussion to date. We need more research into this. Cloud computing systems for High-Performance Computing (HPC) are also a promising area for future provision. Cloud is not yet mature enough for HPC [31]. However, cloud computing helps save enterprise 30–60% of their technology expenditure, but owing to lack of agreement on common standards, many enterprises are losing opportunities. It is not so easy for cloud computing to achieve its aim of being a universally accessible application that is based on open standards. Amazon AWS Import/Export supports importing and exporting data into and out of Amazon S3 buckets in the USA, but still leaves complications in migration of data between clouds. A major challenge of moving applications to the cloud is the need to master multiple programming languages and operating environments [27]. Special attention is needed for government agencies to integrate their data from traditional to PaaS, a need to learn some new programming models residing in the cloud. Interoperability is another important issue for cloud. There is a need for data access interoperability, which is a unique programming interface to access diverse databases (such as JDBS, ODBC, Ado.NET). There are lots of standardization issues; in the race to standardization, many organizations and forums are working, but need to leverage the collaboration and discussions between them. Cloud Computing Interoperability Forum (CCIF) [16] was formed to define an organization that would enable interoperable enterprise class cloud computing platforms through application integration and stakeholder cooperation. Similarly, Microsoft's approach to interoperability principles [28] is a good starting point. Other organizations such as Open Cloud Consortium (OCC) [36], Open Grid Forum (OGF) [37], and Distributed Management Task Force (DMTF) [38] are also working on interoperability issues and open formats. Armbrust et al. [15] also identified many issues for future research. There are some complications with current programming frameworks and programming languages for cloud computing such as Google AppEngine with its SQL-like syntax called "GQL." Select statements in GQL can be performed on one table only. GQL does not support a join statement. The cloud developers will need more flexible query-oriented and API-oriented programming in future. Automated diagnosis is one of the problems in Hadoop. MapReduce is better for limited tasks like text searching or data mining, the things Google does on an epic scale. For tasks that require relational database capabilities at web scale, database sharing has become a favorite practice. The main problem of why several users do not use cloud computing yet is the lack of trust in the cloud itself (services, providers, etc.) and this lack is based on several issues (no acknowledgement of the policies applied for confidentiality of the user's information, privileges of the users in charge of the data, level of satisfaction in regard to compliance with the contract specifications, if the provider permits audits, technical support offered). The complexity will be there for developers to apply the disciplines of development across multiple platform technologies and computational models. The alignment of user needs with business strategy is also a challenging job for CIOs.

2.7 Conclusions

Cloud computing is a promising paradigm for delivering IT services as computing utilities. Clouds are designed to provide services to external users; providers need to be compensated for sharing their resources and capabilities. There are significant challenges and opportunities behind the ecosystem of cloud computing such as resource management, reliability, fault tolerance, security, SLA, utility model, and performance issues. There are many taxonomies, but they are vendor-concern oriented. The proposed taxonomy focused more on engineering approaches such as functional as well as structured aspects of cloud computing systems. We provided a consistent set of guidelines for clarity, and reusability, which is employed to classify a wide range of cloud computing systems. The value of the offered taxonomy lies in that it captures a given system's scope, scalability, generality, reusability, manageability, and flexibility. This chapter presented a different way of representing a taxonomy to classical approaches. This might be a new way to think about the components of taxonomy as layered services that can give a wide range of spectrum for flexibility and reusability. This taxonomy has been applied to the different cloud systems to find out the technical strengths and weaknesses. A survey of different cloud systems has been presented, and captures the different aspects of the taxonomy that provide an idea about functional and architectural view of the systems that they adopted. We concluded the chapter with a discussion of the considered systems, as well as directions for future research. It is hoped that this can provide stimulus to the researcher and ideas to the developer with respect to current cloud systems, hype, and challenges.

Acknowledgments This research was supported by the MKE (Ministry of Knowledge and Economy), Korea, under the ITRC (Information Technology Research Center) support program supervised by the NIPA (National IT Industry Promotion Agency) (NIPA-2009-C1090-0902-0026), and research program on Kookmin University.

References

1. Dean J, Ghemawat S (January 2008) MapReduce: simplified data processing on large clusters. Commun ACM 51(1):107–113
2. Sun Microsystems (2009). Virtualization for dummies
3. Software and Information Industry Association (2001, February). Software as a service: strategic backgrounder
4. Choudhary V (2009) Software as a service: implications for investment in software development. Proceedings of the 40th Hawaii international conference on system sciences
5. Kolakowski N (2009) Microsoft's cloud azure service suffers outage. Retrieved from, http://www.eweekeurope.co.uk/news/microsoft-s-cloud-azure-service-suffers-outage-396
6. Cruz A (2009) Gmail site reliability manager. Update in Gmail. Retrieved from, http://googleblog.blogspot.com/2009/02/update-on-gmail.html
7. Mayer M (2009) Search Products andUser Experience: This site may harm your computer on every search results. Retrieved from, http://googleblog.blogspot.com/2009/01/this-site-may-harm-your-computer-on.html

8. Jackson T (2008) Gmail product manager: we feel your pain, and we're sorry. Retrieved from, http://gmailblog.blogspot.com/2008/08/we-feel-your-pain-and-were-sorry.html
9. Pete, App Engine Team (2008) App engine outage today. Retrieved from, http://groups.google.com/group/google-appengine/
10. Allen Stern (2008) Update from Amazon Regarding Friday's S3 Downtime. Retrieved from, http://www.centernetworks.com/amazon-s3-downtime-update
11. AWS Service Health Dashboard (2008, July 20) Amazon S3 availability event. Retrieved from, http: //status.aws.amazon.com/s3-20080720.html
12. Tubanos A (2008) FlexiScale suffers 18-hour outage. Retrieved from, http://www.thewhir.com/
13. Amazon Web Services (AWS) (2008, Sept) Amazon web services: overview of security processes
14. Cavoukian A (2008, May 28) Privacy in the clouds: privacy and digital identity – implications for the Internet. Information and privacy commissioner of Ontario
15. Armbrust M et al (2009, February 10) Above the clouds: a berkeley view of cloud computing. EECS department, University of California, Berkeley, Technical Report No. UCB/EECS-2009-28
16. The Cloud Computing Interoperability Forum (CCIF) (2009) http://www.cloudforum.org/
17. Gathering Clouds of XaaS! (2008) Retrieved from, http://www.ibm.com/
18. http://www.roughtype.com (2008)
19. Apache Hadoop project (2009) Available from http://hadoop.apache.org/
20. Amazon Elastic Cloud Computing (EC2) (2009) Available from http://aws.amazon.com/ec2/
21. Crandell M (2008) Defogging cloud computing: a taxonomy. Available from http://gigaom.com/2008/06/16/defogging-cloud-computing-a-taxonomy/
22. Laird P (2009) Different strokes for different folks: a taxonomy of cloud offerings. Enterprise cloud submit, INTEROP
23. Cloud Computing Use Case Discussion Group (2009, August) Cloud computing use case. White Paper version 1.0. 5
24. Ried S (2009) Yet another cloud – how many clouds do we need? Retrieved from Forrester Research, http://www.forrester.com/
25. Gammage B, Shiffler III G (2007, August 8) Report highlight for dataquest insight: PC virtualization forecast scenarios. Gartner
26. Dayley A et al (2009, Jan 5) Dataquest insight: virtualization market size driven by cost reduction, resource utilization and management advantages. Gartner
27. Hayes, B (2008, July) Cloud computing. Commun ACM 51(7)
28. Microsoft's approach to interoperability (2009) Retrieved from, http://www.microsoft.com/interop/principles/default.mspx. Accessed 25 Sept 2009
29. Gottfrid D (2009) Self-service, prorated super computing fun! Retrieved from http://open.blogs.nytimes.com/
30. Pearson S (2009) Taking account of privacy when designing cloud computing services. Proceedings of the 2009 ICSE workshop on software engineering challenges of cloud computing, IEEE Comp Soc, pp 44–52
31. Napper J, Bientinesi P (2009) Can cloud computing reach the TOP500? Proceeding of the combined workshops on unconventional high performance computing workshop plus memory access workshop, ACM, pp 17–20
32. OASIS (Organization for the Advancement of Structured Information Standards) (2009) http://www.oasis-open.org/
33. GoGrid's prepaid cloud hosting plans (2009). http://www.gogrid.com/pricing/plans.php
34. Rainge E (2009, May) Worldwide telecom cloud billing 2009–2013 forecast. IDC Doc #217313
35. CFengine (2009) http://www.cfengine.org/
36. The Open Cloud Consortium (OCC) (2009) http://opencloudconsortium.org/
37. The Open Grid Forum (2009) http://www.ogf.org/
38. The Distributed management Task Force (DMTF) (2009) http://www.dmtf.org/about/

39. Buco MJ et al (2004, Jan) Utility computing SLA management based upon business objectives. IBM Syst J 43(1):159–178
40. The Official Google Blog (2008, Nov) Sorting 1 PB with MapReduce. Retrieved from http://googleblog.blogspot.com/
41. rPath (2009) http://www.rpath.com
42. Opscode (2009) http://www.opscode.com/
43. Sheehan M (2009) Message from GoGrid founders regarding denial of service attack. Retrieved from, GoGrid Official Blog, http://blog.gogrid.com/
44. Ristenpart T et al (2009) Hey, you, get off of my cloud: exploring information leakage in third-party compute clouds. Proceeding of ACM conference on computer and communications security
45. Linthicum DS, Morrison KS (2009) Value of SOA for cloud computing. Layer 7 technologies
46. Patterson D (2009) Recovery oriented computing. http://roc.cs.berkeley.edu
47. Ghemawat S, Gobioff H, Leung ST (2003) The google file system. Proceedings of the nine-teenth ACM symposium on operating systems principles, pp 29–43
48. Fielding RT (2000) Architectural styles and the design of network-based software architectures. Dissertation of doctor of philosophy, University of California, Irvine
49. RSA (2009, March) The role of security in trustworthy cloud computing. Continental Automated Building Association (CABA), Information Series, IS 2009-39
50. Thusoo A (2009, June 11) Hive-A petabyte scale data warehouse using hadoop. Retrieved from Facebook Engineering page, http://www.facebook.com
51. Hadoop/Hive (2009) http://wiki.apache.org/hadoop/Hive
52. Palankar M et al (2008) Amazon S3 for science grids: a viable solution? Proceedings of the 2008 international workshop on data-aware distributed computing workshop (DADC), pp 55–64
53. Carriero N, Gelernter D (1989) Linda in the context. Commun ACM 32(4):444–458
54. Gray J et al (1996) The dangers of replication and solution. Proceedings of the 1996 ACM SIGMOD international conference on management of data, pp 173–182