Cloud Computing Law

Edited by
CHRISTOPHER MILLARD

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

First Edition published in 2013
Impression: 1

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data
Data available

Library of Congress Control Number: 2013940565

ISBN 978–0–19–967167–0 (hbk.)
ISBN 978–0–19–967168–7 (pbk.)

Printed in Great Britain by
CPI Group (UK) Ltd, Croydon, CR0 4YY

Contents

IV. CLOUD REGULATON AND GOVERNANCE

1

Cloud Technologies and Services

W Kuan Hon and Christopher Millard

1. What Is Cloud Computing?

At its simplest, cloud computing is a way of delivering computing resources as a utility service via a network, typically the Internet, scalable up and down according to user requirements. As such, the cloud may prove to be as disruptive an innovation as was the emergence of cheap electricity on demand a century or so ago.[1] Such computing resources may range from raw processing power and storage, such as servers or storage equipment, to full software applications. Users can 'rent' IT resources from third parties when needed, instead of purchasing their own, thus 'turning capex to opex' (capital expenditure into operating expenditure).

The scale, and scalability, of cloud computing make it ideally suited for environments where demand for IT resources may fluctuate widely and rapidly. Cloud may also prove to be a key enabling technology for releasing the latent power of 'Big Data', and for supporting the deployment of mobile devices and applications on a very large scale.[2] Market surveys and statistics abound but in terms of adoption to date, recent research indicates that more than half of US businesses already use cloud computing services.[3] In the EU, the cumulative economic impact of cloud computing has been predicted to be €940 billion and 3.8 million jobs for the period 2015–20.[4]

In slightly more technical terms, cloud computing is an arrangement whereby computing resources are provided on a flexible, location-independent basis that allows for rapid and seamless[5] allocation of resources on demand. Typically, cloud resources are provided to specific users from a pool shared with other customers with pricing, if any,[6] often proportional

[1] Nicholas Carr, *The Big Switch: Rewiring the World, from Edison to Google* (New York/London: WW Norton, 2009).

[2] For a discussion of the relationship between cloud, Big Data, and mobile, see David Linthicum, 'Understanding the Symbiosis of Cloud Computing, Big Data, and Mobile', <http://pro.gigaom.com/blog/understanding-the-symbiosis-of-cloud-computing-big-data-and-mobile/> (accessed 2 April 2013). For a brief introduction to Big Data, see Christopher Kuner, Fred Cate, Christopher Millard, and Dan Svantesson, 'The Challenge of "Big Data" for Data Protection' (2012) 2(2) *International Data Privacy Law* 47.

[3] Reuven Cohen, 'The Cloud Hits the Mainstream: More than Half of U.S. Businesses Now Use Cloud Computing' (*Forbes*, 16 April 2013), <http://www.forbes.com/sites/reuvencohen/2013/04/16/the-cloud-hits-the-mainstream-more-than-half-of-u-s-businesses-now-use-cloud-computing/> (accessed 1 June 2013).

[4] IDC, *Quantitative Estimates of the Demand for Cloud Computing in Europe and the Likely Barriers to Uptake*, 13 July 2012 Final Report for the European Commission, <http://ec.europa.eu/information_society/activities/cloudcomputing/docs/quantitative_estimates.pdf> (accessed 1 June 2013).

[5] 'Seamless' here means adding or removing capacity without specific action by the cloud user, or even awareness of such adjustment. Providers may offer users the option of requesting capacity manually when needed, or setting resources to vary automatically with demand.

[6] Many consumer services are funded by advertising. See Chapter 13 for a discussion of the status of such 'free' services.

to the resources used.[7] The delivery of cloud services often depends on complex, multilayered arrangements between various providers.[8]

Many permutations are possible, but cloud computing activities are often described as falling into one or more of the following three service categories:[9]

- Infrastructure as a Service ('IaaS'): raw computing resources, such as processing power ('compute') and storage.[10]
- Platform as a Service ('PaaS'): platforms for developing and deploying software applications.[11]
- Software as a Service ('SaaS'): end-user applications.[12]

Cloud users may run, typically via web browsers,[13] application software installed on remote servers which sends results to users over the Internet. This means that relatively simple devices, such as mobile phones or tablets, may be used to obtain access to vast computational resources.

The use of 'as a Service' emphasizes a change in focus, from obtaining products or licences, to renting the use of resources as services. These service models sit on a spectrum from IaaS to PaaS to SaaS, rather than being separate or discrete types of cloud computing. Generally, IaaS involves relatively low-level functionality for users, requiring greater user sophistication and expertise, including more hands-on, micromanagement of resources. However, it affords the user more flexibility and fine control. SaaS provides high-level functionality, and generally requires less user technical expertise, but offers less user control. PaaS sits in the middle. Users are spared the need to manage raw

[7] As outlined in Chapter 4, service charges may be calculated per user, per month. There are also other models. For example, PaaS provider AppFog charges according to the amount of RAM used. Alex Williams, 'AppFog Wants to Do for Developer Platforms What Google Did for Email' (*TechCrunch*, 25 July 2012), <http://techcrunch.com/2012/07/25/appfog-wants-to-do-for-developer-platforms-what-google-did-for-email> (accessed 1 March 2013).

[8] See Chapters 2, 3, and 4.

[9] Originally promulgated by the US National Institute of Standards and Technology: Peter Mell and Tim Grance, *The NIST Definition of Cloud Computing, Special Publication 800-145* (Gaithersberg, MD: US National Institute of Standards and Technology, 2011).

[10] The IT resources used over a network here essentially consist of computing hardware infrastructure (servers, storage, etc.) and tools to help users manage those resources. IaaS services include Rackspace, Amazon Web Services' EC2, and Google Compute Engine. Instead of investing in their own data centres or servers, users may install their operating systems and applications of choice on the provider's infrastructure. IaaS, as a data centre/computing hardware substitute, may be used, for example, by start-ups who avoid upfront capital expenditure on physical infrastructure by using IaaS to 'rent' use of a third party's computing and storage infrastructure.

[11] Accordingly, the IT resources used over a network in PaaS comprise platforms for programming, deploying, and hosting applications with application management tools. PaaS services include Google's App Engine, Microsoft's Windows Azure, and Heroku.

[12] With SaaS, the IT resource used over a network is application software, hence 'Software as a Service'. For example, instead of installing word-processing software to run on each user's local computer, Google Apps or Microsoft Office 365 SaaS may be used for online word processing. Other examples are webmail services such as Yahoo! Mail (instead of desktop email applications) and social networking services like Facebook. Salesforce's online customer relationship management service is an example of SaaS for enterprise cloud users. A subclass of SaaS is 'Storage as a Service', which provides the infrastructure and means for users to manage data storage, organization, and retrieval from any location over the Internet, for example Dropbox or Rackspace's Cloud Files.

[13] Although some services involve or allow installation of separate software on user devices to automate and synchronize storage of local files to the cloud, for example Dropbox, and other storage services, such as Google Drive, SugarSync, and SkyDrive. It is also very common for local applications ('apps') to be installed on mobile devices to facilitate access to specific cloud services.

processing/storage resources actively, and may focus on programming applications to be hosted via the service. Boundaries between them, particularly IaaS and PaaS, may blur; IaaS providers are increasingly offering higher-level functionality,[14] while PaaS providers may offer lower-level detailed control.[15]

SaaS is the most commonly used type of cloud service, particularly among consumers, which is unsurprising as it generally requires the least technical know-how on the part of users, and enables users to procure use of application software quickly without installing any specific software. According to a survey of 300 UK-based organizations in late 2011,[16] the SaaS applications most used were email, backup/disaster recovery, storage, and web hosting services.

Cloud deployment models can also be viewed in various ways, but a widely used classification is:

- Private cloud: where relevant infrastructure is owned by, or operated for, the benefit of a single large customer or a group of related entities.[17]
- Community cloud: where infrastructure is owned by, or operated for, and shared among a specific group of users with common interests, such as US government bodies,[18] UK local government, or the financial services industry.[19]
- Public cloud: where infrastructure is shared among multiple users using the same hardware and/or software.
- Hybrid cloud: involving a mixture of the above, for example, an organization with a private cloud may 'cloud burst' processing activities to a public cloud for 'load balancing' purposes during times of high demand.[20]

[14] For example, Amazon offers resources such as software development kits for Java, mobile (Android, iOS), PHP, Python, Ruby, and ASP.NET programmers. Amazon, 'Java Developer Center', available at <http://aws.amazon.com/java> (accessed 1 March 2013).

[15] For example, Azure PaaS allows use of VMs within cloud services. Windows Azure, 'How to Connect Virtual Machines in a Cloud Service', available at <http://www.windowsazure.com/en-us/manage/windows/how-to-guides/connect-to-a-cloud-service> (accessed 1 March 2013).

[16] Cloud Industry Forum, *Cloud UK: Paper Four Cloud Adoption and Trends for 2012* (2011).

[17] For example, the US state of Alaska. Cisco, 'State Government Deploys Private Cloud to Provide Services to Agencies' (2011), <http://www.cisco.com/en/US/solutions/collateral/ns340/ns517/ns224/state_of_alaska_cs.pdf> (accessed 1 March 2013).

[18] For example, Microsoft Office 365 SaaS as a 'multi-tenant service that stores US government data in a segregated community cloud'. Kirk Koenigsbauer, 'Announcing Office 365 for Government: A US Government Community Cloud' (30 May 2012), <http://blogs.office.com/b/microsoft_office_365_blog/archive/2012/05/30/announcing-office-365-for-government-a-us-government-community-cloud.aspx> (accessed 1 March 2013).

[19] For example, NYSE Euronext's Capital Markets Community cloud for financial services firms (launched in 2011 in partnership with storage provider EMC and virtualization firm VMware), offers applications and services to customers via its own 'app store', computing-on-demand services, and connections to NYSE Euronext's global trading network, including a market data feed. Tom Steinert-Threlkeld, 'Cloud Fundamental to NYSE IT Strategy' (*Information Management*, 28 August 2012), <http://www.information-management.com/news/cloud-fundamental-to-nyse-it-strategy-10023088-1.html?zkPrintable=true> (accessed 8 March 2013).

A 'Trusted German Insurance Cloud' has also been launched, by the German Insurance Association: Federal Office for Information Security 'The BSI', available at <https://www.bsi.bund.de/ContentBSI/Presse/Pressemitteilungen/Presse2012/Trusted-German-Insurance-Cloud_08032012.html> (accessed 8 March 2013).

[20] For an explanation of how cloud bursting works, see Nati Shalom, 'What Is Cloud Bursting?' (*CloudCow.com*, 12 May 2012), <http://www.cloudcow.com/content/what-cloud-bursting> (accessed 1 March 2013).

Private/public should not be equated with on/off-premise. Infrastructure for cloud services may be located on users' premises, or at one or more external locations. Private clouds are not necessarily on-premise; the infrastructure/resources used could be managed, even owned, by a third party, but dedicated to the user concerned. Public clouds, however, are generally off-premise.

2. Cloud Computing Resources and Technologies

2.1 IaaS

Pooled computing resources in IaaS are used mainly for data processing ('compute' capability), storage, and networking or other connectivity services. These are outlined below because the application of certain laws relating to data location, data security, and handling of 'personal data' may be affected by the specific arrangements for processing, storing (including replicating and deleting), and transmitting data.[21]

2.1.1 Compute

For 'compute' a key enabler was the development of virtualization technologies. This facilitated the conceptual separation of different computing elements. Server hardware virtualization is the most common type of virtualization in cloud computing.

IaaS typically involves using 'virtual machines' (VMs). Via virtualization technology[22] a physical server may 'host' multiple VMs. Each VM operates as a virtual server, running independently with its own operating system[23] within which applications, or other software, may also be installed and run.[24]

Different users ('tenants', hence the term 'multi-tenancy') may thus independently create ('instantiate') and run their own VMs (and applications) within one physical server. They can terminate their VMs when no longer required, install their own firewalls, and manage their own virtual networks. Multiple users may share use of common physical infrastructure; this enables resource consolidation, economies of scale, and involves more efficient resource utilization than dedicating separate physical machines to different users (avoiding possible periods of non-use that may waste the machine's peak capacity).

VMs on one physical machine may share use of the same *physical* resources, including processors, hard disks, memory, and network interfaces. Users and their VMs are segregated or isolated from each other by virtualization software only, rather than through physical separation.[25] While shared use enables efficiencies, it also raises security

[21] See Chapters 8, 9, and 10.

[22] Virtualization software may be proprietary, such as Microsoft Hyper-V and VMWare Workstation, or open source, such as Xen, KVM, and Oracle VM VirtualBox. Amazon Web Services' set of (mainly IaaS) cloud computing services uses Amazon's modified version of Xen. Google Compute Engine IaaS is based on KVM.

[23] These may be proprietary (eg Windows), or open source (eg Linux).

[24] Assuming the operating system or other software's licence allows it to be installed and run in VMs instead of separate physical machines. See Chapter 4.

[25] As some users are concerned about this for compliance or security reasons, some providers (eg Amazon) offer services on separate, dedicated physical hardware. Amazon Web Services, 'Amazon EC2 Dedicated Instances', available at <http://aws.amazon.com/dedicated-instances> (accessed 1 March 2013).

concerns, including that the provider or another tenant may be able to access or interfere with the data or processing of another tenant using the same hardware.[26]

VMs are created and may be booted from stored files called virtual machine images ('VMIs') or virtual machine files.[27] Different virtualization technologies use or support different image formats.[28] VMIs allow quick deployment of multiple pre-configured VM instances from a single 'template' file.[29] For customers' convenience, some providers offer various pre-packaged template images with different operating systems pre-installed. Users may configure their own VMs with libraries or applications and so on, then capture or save 'snapshots' of their customized images,[30] and may even make images available publicly. VMs may also be cloned or 'live migrated' while running. If data operated on within a VM are not saved to persistent storage[31] before the VM's termination or failure, generally the data are lost. Depending on the service, even VM instances appearing to have attached storage may lose 'stored' data on 'taking down' the instance, unless actively saved to persistent storage first. The format, storage, and precise arrangements for using VMIs may have an impact on security, data location, and competition law.[32] Thus, several components may be involved in IaaS, from virtualization software and VMI files, to operating systems installed within VMs, and application or other software that users install and run on such operating systems.

Not all cloud computing requires the use of VMs, although it may still involve virtualization in the broad 'abstraction' sense. Some cloud computing involves the opposite: instead of one physical server running multiple VMs, multiple physical computers are harnessed to work together simultaneously, in parallel, on a single processing operation. Workloads are divided into sub-operations and distributed or 'mapped' among different physical servers, with the results of the distributed processing sub-operations being collected or 'reduced'.[33]

[26] Security is discussed further in Chapter 2. IEEEXplore Digital Library, 'Side-Channel Leaks in Web Applications: A Reality Today, a Challenge Tomorrow', available at <http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=5504714&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D5504714>; Thomas Ristenpart et al., 'Hey, You, Get off of My Cloud: Exploring Information Leakage in Third-party Compute Cloud' (*ACM Digital Library*, 2009), <http://dl.acm.org/citation.cfm?id=1653687>; and Yinqian Zhang et al., 'Cross-VM Side Channels and Their Use to Extract Private Keys' (*ACM Digital Library*, 2012), <http://dl.acm.org/citation.cfm?id=2382230> (all accessed 1 March 2013).

[27] Virtual disk files (virtual disk images), which may be stored on physical disks, are 'images' of virtual hard disks (ie the contents of a VM's virtual hard disk drive). VMs may use virtual disks as physical hard disks. Virtual disks may store operating systems, data, and applications that may be run from the disk within a VM.

[28] For example, VMWare, Workstation, Microsoft Hyper-V, and VirtualBox use different formats (but VirtualBox supports several formats).

[29] Amazon's VMIs use Amazon's own proprietary format and are called Amazon Machine Images ('AMIs'). Amazon Web Services, 'Amazon Machine Images (AMIs)', available at <https://aws.amazon.com/amis> (accessed 1 March 2013).

[30] 'Virtual appliances' are software applications ('software appliances') that run within VMs, usually one application per appliance. A virtual appliance *image* incorporates VMI plus 'pre-installed' software appliances. An open standard for packaging and distributing virtual appliances, Open Virtualization Format ('OVF'), has been accepted by standards organizations ANSI and ISO. OVF has seen some, but not universal, support (eg by IBM and VMWare). Distributed Management Task Force, 'Open Virtualization Format Specification' (12 January 2010), <http://dmtf.org/sites/default/files/standards/documents/DSP0243_1.1.0.pdf> (accessed 1 March 2013).

[31] Usually one or more hard disk drives or SSD flash storage.

[32] See Chapters 7, 8, and 12.

[33] A popular open-source framework for scalable distributed applications handling large quantities of unstructured data is the Apache Hadoop project. The project is based on Google's MapReduce framework for massive distributed processing using low-cost commodity systems. Hadoop uses 'clusters' of 'nodes' or servers fragmenting data across nodes automatically, and also replicating data automatically,

Other types of virtualization feature in cloud. Just as operating systems may be made independent of hardware through hardware virtualization, applications may be decoupled from operating systems through application virtualization. Virtualization methods vary, but the end result is the logical 'abstraction' of computing from physical resources, a broader concept which runs through cloud computing and enables users to focus on desired functions and actions at a higher, simpler, and more general level, rather than having to micromanage underlying resources.

2.1.2 Storage

Storage 'virtualization' makes it possible for a user to have access to what appears to be a single file or document while the relevant data are stored in a physically distributed fashion across different pieces of hardware (eg a Storage Area Network or SAN), with specialized software managing storage and retrieval. The intention is that users need not be concerned with the details of physical storage as this is handled automatically for them. RAID[34] is a storage virtualization technology resilient to individual physical drive failure. Multiple drives, a RAID array, behave as one logical drive. Data fragments are 'striped' across those drives automatically and, possibly, also replicated or 'mirrored' in different ways, depending on the RAID scheme or 'level' used. RAIDs may be combined into larger RAIDs.

Providers often offer persistent non-volatile data storage, enabling data retrieval, for example after a VM instance terminates.[35] Cloud service providers employ various, in some cases proprietary, systems to manage large-scale distributed data storage and retrieval across different hardware, including distributed file systems and distributed databases.[36] Database management systems, operating at a higher level than file

attempting to use different racks for fault-tolerance. Organizations such as eBay, Facebook, and Yahoo! use Hadoop heavily to provide cloud services to their external users, as well as for internal purposes, including private cloud computing to support public-facing services. Dhruba Borthakur, 'Facebook Has the World's Largest Hadoop Cluster!' (*HDFS*, 9 May 2010), <http://hadoopblog.blogspot.co.uk/2010/05/facebook-has-worlds-largest-hadoop.html>; Ashish Thusoo et al., 'Data Warehousing and Analytics Infrastructure at Facebook' (*ACM Digital Library*, 2010), <http://delivery.acm.org/10.1145/1810000/1807278/p1013-thusoo.pdf>; Yahoo, 'Hadoop at Yahoo!', available at <http://developer.yahoo.com/hadoop/>; and Anil Madan, 'Hadoop—The Power of the Elephant' (*Ebay Tech Blog*, 29 October 2010), <http://www.ebaytechblog.com/2010/10/29/hadoop-the-power-of-the-elephant> (all accessed 1 March 2013). Although Hadoop does not use VMs, there is increasing interest in running Hadoop on VMs. See also Chris Brenton, 'The Basics of Virtualization Security', <https://cloudsecurityalliance.org/wp-content/uploads/2011/11/virtualization-security.pdf> (accessed 1 March 2012).

[34] Originally an acronym for Redundant Array of Inexpensive Disks, RAID is now used to mean Redundant Array of Independent Disks.

[35] Storage systems include Amazon's S3, SimpleDB, or Elastic Block Storage, and Windows Azure's SQL Azure, blob, table storage, or XDrive.

[36] For example, Amazon's Dynamo for structured data storage underlies many services, including its S3 storage. Giuseppe Decandia et al., 'Dynamo: Amazon's Highly Available Key-value Store' (*ACM Digital Library*, 2007), <http://dl.acm.org/citation.cfm?id=1294281> (accessed 1 March 2012).

Google's proprietary software developed for fault-tolerant distributed storage and processing using cheap commodity hardware includes BigTable, for structured data, and Google File System, a distributed file system. Fay Chang et al., 'Bigtable: A Distributed Storage System for Structured Data' (*ACM Digital Library*, 2 June 2008), <http://dl.acm.org/citation.cfm?id=1365816> (accessed 1 March 2012).

Many cloud databases are 'NoSQL' databases, for example the open-source MongoDB used by Twitter and (on Amazon EC2) location-based social media service FourSquare, 'Show and Tell: MongoDB at Foursquare', available at <http://engineering.foursquare.com/2011/12/21/show-and-tell-mongodb-at-foursquare>. Some consider that NoSQL databases handle huge quantities of distributed data better than traditional relational SQL databases. To prioritize high availability, cloud services usually tolerate weaker data consistency, hence the term 'eventual consistency': on updating data, replicas are updated to be consistent—but only eventually, not instantly. For example, Amazon systems, are effectively applications enabling easier and more structured logical management of data than file systems (although ultimately database files are stored in file systems). However, the same database, or even database table, may hold more than one user's data, so it is not just physical storage, but logical storage that may be shared among different users, again, with software handling the segregation.[37] Thus, users rely on such software for security, to ensure one user cannot access another user's data deliberately or inadvertently.

Cloud computing uses the RAID concept at larger scale: one data set may be dispersed in fragments (chunks or 'shards' in database terms) among servers or other storage equipment,[38] to be reunited and delivered to a user logging in with the correct credentials. Sharding is based on the provider's sharding policies, which vary with space constraints and performance considerations.[39] Applications' requests for data are automatically sent to some or all servers hosting relevant fragments and results are coalesced by the application. Sharding assists availability, as smaller fragments are retrieved faster and response times improved. While 'sharding', or 'partitioning', most commonly refers to fragmenting databases,[40] data not within a structured database may also be fragmented for storage or operations.[41] Data fragmentation is relevant to data location and may give rise to questions as to whether data fragments are intelligible or constitute 'personal data'.[42]

In summary, multiple users' data may be stored on common physical equipment, with software handling the storage and retrieval of a particular user's data. IaaS users may choose among available data-storage mechanisms, which may include database and caching tools, and may even use their own applications to manage data storage. Therefore, different degrees of control over data storage and data location are possible.

SimpleDB and Google BigTable use eventual consistency. Werner Vogels, 'Eventually Consistent' (*Practice*, January 2009), <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.187.8545&rep=rep1&type=pdf>; and David Bermbach et al., 'Eventual Consistency: How Soon Is Eventual? An Evaluation of Amazon S3's Consistency Behavior' (*ACM Digital Library*, 2011), <http://dl.acm.org/citation.cfm?id=2093186> (all accessed 1 March 2013).

[37] Ways to segregate users' data exist, even so: 'Technische und organisatorische Anforderungen an die Trennung con automatisierten Verfahren bei der Benutzung einer gemeinsamen IT—Infrastruktur' (10 November 2012), <http://www.lda.brandenburg.de/sixcms/media.php/4055/TOP08_20121011_OH_Mandantenfaehigkeit_v10_b.pdf> (accessed 1 March 2013: regarding technical and organizational separation requirements for automated data processing on shared IT systems, by a working group on technical and organizational data-protection matters, part of the Conference of the German Data Protection Commissioners).

[38] While servers' physical hard drives can store data, storage arrays attached to servers may also handle persistent storage of data.

[39] For example, to equalize workload across different servers and/or data centres. LA Barroso and U Hölzle, 'The Datacenter as a Computer: An Introduction to the Design of Warehouse-scale Machines' in Mark D Hill (ed.), *Synthesis Lectures on Computer Architecture* (Morgan & Claypool, 2009) at 1.

[40] Users may partition or 'shard' cloud databases logically; we use 'sharding' to mean only automatic data fragmentation by providers' systems. Users have no say in such automated sharding, although some providers, for example Amazon and Microsoft (Azure), allow users to confine to broad geographically circumscribed regions, for example the EU or Europe, the storage (and presumably other processing) of the resulting shards. See Chapters 3 and 10.

[41] For example, Amazon's Elastic Block Store ('EBS') appears to users as physical hard drives, but data are stored as EBS snapshots on Amazon's S3 storage service, first broken into chunks whose size depends on Amazon's optimizations. Amazon, 'Amazon Elastic Block Store (EBS)' (2011), <http://aws.amazon.com/ebs> (accessed 1 March 2013).

[42] This is discussed further in Chapter 7.

Whatever the service type, providers may back up data to multiple locations to protect against data corruption or loss due to hardware or software weaknesses or failure, and to preserve data and service availability. Some providers may, for business continuity reasons, replicate stored data to separate equipment within the *same* data centre. Replication may even be 'synchronous', that is in real time or near real time. Providers may or may not back up further, for instance to tapes (which may be stored in the same or different physical locations). Some providers also automatically replicate data to other, geographically separated, data centres.[43] If irrecoverable data loss occurs, for example due to a natural disaster affecting one data centre, copies may be retrievable from data centres holding replicas (disaster recovery). Similarly, upon any failure, users may be switched over ('failover') to using another data centre, maintaining availability and giving the illusion of instantaneous data 'migration' to another location.[44] However, not all providers back up data, particularly across different data centres (rather than just within the same one). Large providers may have sufficient infrastructure, whereas smaller providers might not.

Providers are increasingly allowing users to choose a processing region, for instance 'Europe'. However, it is not always clear whether all stored data and replicas will also be restricted to the chosen region, and it may be difficult as a technical matter for users, or for example their auditors, to verify that their data and processing are indeed confined to the specified physical location.

Finally, in terms of data location and security, data may also be stored, albeit often only temporarily, on a hard disk or memory[45] during active processing, and deletion methods of such temporary copies may vary. If a user runs processing operations on data using cloud resources, generally the data exist only ephemerally unless, and until, saved to persistent storage. During operations, data retrieved from and saved to persistent storage may be temporarily cached automatically, perhaps in distributed fragments, to speed up operations.[46] Data may also be cached in particular locations to speed up data delivery to those geographical areas (content delivery networks or CDNs). Therefore, data may also be located in any server operating on the data (usually in the same data centre from which such data were retrieved, but technically it could differ), and any temporary caches holding the data.

[43] Google and Microsoft maintain two replicas of each data set, typically in different data centres. Rajen Sheth, 'Disaster Recovery by Google' (*Official Google Enterprise Blog*, 4 March 2010), <http://googleenterprise.blogspot.com/2010/03/disaster-recovery-by-google.html> and Brad Calder et al., 'Windows Azure Storage: A Highly Available Cloud Storage Service with Strong Consistency' (2011), <http://sigops.org/sosp/sosp11/current/2011-Cascais/printable/11-calder.pdf> (both accessed 1 March 2013). Replication strategies vary. 'Facebook replicates data with all writes going through a single master data center . . . Yahoo! mail partitions data across [data centres] based on user . . . Facebook design has a single master coordinate replication—this speeds up lookups but concentrates load on the master for update operations', Albert Greenberg et al., 'The Cost of a Cloud: Research Problems in Data Center Networks', <http://research.microsoft.com/en-us/um/people/dmaltz/papers/dc-costs-ccr-editorial.pdf>. Google claims 'live' synchronous (real-time) replication, for example for Google Apps SaaS: Sheth, ibid.

[44] This is the main circumstance when cloud data might be perceived to 'move' from one data centre to another, but clearly does not involve deletion of data from one data centre and recreation in another.

[45] Increasingly, to enhance performance, data may be stored simply in memory and backed up to persistent storage: Amazon Web Services, 'SAP HANA One—Now Available for Production Use on AWS', available at <http://aws.typepad.com/aws/2012/10/sap-hana-now-available-for-production-use.html> (accessed 1 March 2013).

[46] Just as, when using application software on a desktop computer to operate on data, some data may be cached temporarily.

2.1.3 Networking

With network virtualization, multiple—often distributed—networking hardware and software resources may be combined into a single logical unit, for instance, to increase network capacity or resilience; however, users perceive one virtual network.[47] Conversely, multiple virtualized networks, isolated from each other, may use shared physical infrastructure simultaneously. Thus, as above, software decouples virtual networks from hardware, and multiple resources may be pooled. As with physical networks, virtual networks may also involve virtual Internet Protocol (IP) addresses, virtual routers, virtual switches, virtual firewalls, and virtual links. Again, IaaS users may manage their own virtual networks, such as virtual private networks (VPNs) connecting VMs of their choice. Similar risks arise as above regarding the security of shared infrastructure, for example whether one user can ever 'see' another user's network traffic, and the adequacy of the software to enforce segregation between users.

2.1.4 Resource management

To share hardware among different users it is necessary to manage or orchestrate the automated use and release of processing, storage, and networking resources to distribute and balance different users' workloads across the available resource pool, scaling up or down as needed across different equipment. The aim is to provide the on-demand, seamlessly scalable, efficient use of resources characteristic of cloud. For example, if a physical server is 'full' to capacity, and hosting the maximum number of VMs it can hold, then, unless a VM running in the server has terminated making room in the server's memory to host another VM, any new VM that subsequently needs to be created must be instantiated on a different physical server. Similarly, different users' data may be stored not only on the same physical system (the same hardware), but also on the same logical system (the same database or even database table), again separated only by software. Storage and retrieval of users' data on a common infrastructure will also need to be managed.

Software 'cloud platforms',[48] 'cloud fabric', or 'cloud operating systems'[49] enable such automated resource management across pools of shared, distributed hardware and software infrastructure at scale.[50] They automate 'bringing up' (or 'spinning up', or 'provisioning') VM instances, spinning them down as needed in different physical machines, or related clusters of machines, according to available capacity and VM 'health', including

[47] Many organizations have long used virtual private networks (VPNs) to provide private networks over the Internet.

[48] Not to be confused with the specialized use of 'platform' in PaaS (an application development and hosting 'platform'), or for example Google's use of 'cloud platform'. Here, we consider 'platform' to be the software infrastructure of cloud computing.

[49] 'Cloud fabric' and 'cloud operating system' are not (yet) terms of art and no common definition has emerged to describe them.

[50] Such 'cloud platforms' may be proprietary, for example IBM SmartCloudEnterprise, VMware vSphere, and what underlies Amazon Web Services; or, they may be open source, such as the Eucalyptus cloud computing platform, OpenNebula, CloudStack (acquired by Citrix in 2011), and OpenStack (developed more recently by US agency NASA and Internet hosting organization Rackspace, and used in Citrix's Webex web conferencing service <http://www.openstack.org/user-stories/cisco-webex/>). Open Stack, 'A Collaborative Cloud', <http://www.openstack.org/user-stories/cisco-webex/a-collaborative-cloud> (accessed 1 March 2013). The Hadoop framework (see n. 33) may also be considered such a platform.

A user may install proprietary or open-source cloud platforms on its own infrastructure for private cloud computing, for example OpenStack. Turn-key enterprise Hadoop platforms for private cloud include Cloudera's, IBM's, and Nimbula/MapR.

monitoring, tracking, and migrating VMs as needed. They can manage networks and virtual networks, not just monitoring in-coming traffic and handling security, but also 'load balancing' to distribute in-coming traffic to suitable available instances,[51] typically through automated software with algorithms to optimize resource allocation and usage based on defined policies and parameters. They can also manage storage and retrieval of data (including structured data in databases) across different equipment. They include tools to enable users to manage their own quota of resources.

Cloud platform software may be hosted, for example, installed on a provider's infrastructure and offered as a service to users (including 'data centre as a service'), or may be installed on user-controlled infrastructure for use as private cloud.[52] Service providers can install applications on IaaS to offer PaaS or SaaS to their own users.[53] IaaS users may install applications for internal use. Installed software could be developed by whoever installs it, or a third party.[54] Thus, IaaS offers users considerable control and flexibility. IaaS may (subject to any technical or contractual restrictions imposed by providers)[55] be used for any purposes for which a user's own computing infrastructure may be used, whether for internal applications, PaaS platforms, or SaaS applications offered to the user's own customers, or even website hosting. This may be important in terms of liability for content (such as infringement of intellectual property rights), and responsibility for compliance (eg under data protection law).

2.2 PaaS

Whereas IaaS users must manage their own virtual computing resources, PaaS involves what technologists term a higher level of 'abstraction'. PaaS effectively provides an integrated computing infrastructure and programming/hosting platform, usually including database and web server services.

PaaS users need not manage VMs or other computing resources at a low level, but can focus on programming application code. After the code of the user's choice is deployed to the PaaS service as an application, the user's application may be run via the Internet, for example as SaaS. The provider's platform automatically handles management and load balancing of computing resources, virtualized or otherwise, to serve and scale the application as necessary, including data storage and replication.

Users have some choice over data-storage mechanisms, including how their application accesses stored data. However, while freed from detailed resource management, PaaS users are restricted to coding applications using only programming languages, frameworks, libraries, and so on supported by the PaaS provider, and user applications must meet any other provider limitations (imposed for scalability and security or other

[51] Amazon's Elastic Load Balancer redirects traffic among an 'Auto-Scaling' group of multiple VM instances. Amazon Web Services, 'New Features for Amazon EC2: Elastic Load Balancing, Auto Scaling and Amazon CloudWatch', <http://aws.typepad.com/aws/2009/05/new-aws-load-balancing-automatic-scaling-and-cloud-monitoring-services.html>; 'AWS Management Console Now Supports Elastic Load Balancing', <http://aws.typepad.com/aws/2009/12/aws-management-console-now-supports-elastic-load-balancing.html> (both accessed 1 March 2013).

[52] Amazon IaaS is only available as a hosted service; Eucalyptus may be installed on a user's own infrastructure.

[53] For example, Heroku's PaaS service is built on Amazon's IaaS service, as is Dropbox's SaaS storage service.

[54] Third-party software licences for installation on a user's own servers on-premise may not necessarily permit installation on VMs or external clouds, so licence terms may need checking and, if necessary, renegotiation. See Chapter 4.

[55] See Chapters 3 and 4.

reasons).[56] Compared with IaaS, users may also have less visibility into what resources are used to run their application, and how they are provided. PaaS platforms may even be built on other providers' IaaS platforms.[57] Accordingly, PaaS offers users less flexibility and control than IaaS. However, PaaS users do control the code they deploy, and therefore have control over the security of their application, which may affect responsibility and liability.

As with IaaS platforms, PaaS cloud platforms may be used as provider-hosted services, or installed on users' infrastructure for private cloud computing.[58] Also, PaaS applications may be for use by the user's employees only, or offered to the user's own customers as SaaS. For example, a software developer wishing to provide online applications to consumers as SaaS may program and host those applications using PaaS (or IaaS) instead of buying its own servers. Like IaaS, PaaS may be used for highly scalable web hosting.[59]

2.3 SaaS

SaaS sits at an even higher level of abstraction than PaaS. Users need not be concerned even with application code; they simply use the application(s) provided. How underlying resources are deployed to provide the service is managed by providers, and as mentioned above some SaaS providers base their services on other providers' IaaS or PaaS, although others may use their own physical or software infrastructure.

While users can set preferences for some SaaS applications, and control how their own quota is used (for instance, storage space), their ability to customize applications is usually limited, and they cannot control how providers manage underlying resources. Some SaaS services even use a single running application to serve multiple users. Again, different users' data may be stored in the same database, even database table, posing potential security risks. Users must rely on the SaaS software to enforce separation.

While SaaS applications often originate from the SaaS provider, third-party SaaS applications may be installed on a SaaS provider's infrastructure for offering as a service, or on a private cloud for internal use.[60]

3. Cloud Supply Chain: Key Concepts

The cloud supply chain is complex. One cloud service may combine hardware and/or software components from different suppliers or providers. Also, cloud services themselves may be combined or layered.

[56] For example, Google's App Engine supports applications coded using Python and Java, Heroku is for Ruby on Rails programs, VMWare's Cloud Foundry supports Java, Ruby, Node.js, and Scala, while RedHat's OpenShift Origin and Microsoft's Windows Azure support multiple languages.

[57] For example, Engine Yard PaaS uses Amazon Web Services IaaS.

[58] For example, Microsoft's proprietary Windows Azure, or VMWare's open-source Cloud Foundry. Some PaaS platforms are only available as a service; for example, Google's App Engine, hosted by Google, is not available for installation on a user's own infrastructure.

[59] For example, Accenture used Google App Engine to build the 2011 UK Royal Wedding website, available at <http://www.officialroyalwedding2011.org> (accessed 8 March 2013).

[60] For example, Microsoft Office 365 application software may be licensed for internal use in a private cloud, or for offering in a public cloud, such as by telecommunications provider Vodafone UK 'Microsoft Office 365 from Vodafone', available at <http://www.vodafone.com/content/index/about/what/business/productivity_services/office_365.html> (accessed 8 March 2013), or could be used as a full SaaS service on Microsoft-run infrastructure.

3.1 Combining components

Cloud services ultimately employ physical infrastructure: equipment housed in physical locations, typically data centres. The data centre ecosystem may involve different players providing physical space, equipment (whether servers, storage, or networking), software infrastructure and services (including, of course, cloud services), and related ancillary services. Cloud platforms used as software infrastructure for cloud services may be proprietary or open source, hosted-only or available as installable software, and may not necessarily involve virtualization. Cloud service providers need not use their own cloud platforms or application software. Therefore, the owner, operator, manager, and user of a physical or software component may be different entities.

As a concrete illustration, a person X may buy or lease a dedicated data centre, or rent space in a third party's data centre where others also rent space (colocation). X could buy or rent servers or storage devices from other third parties. Servers and other equipment could be dedicated to X, or shared with others. X might manage 'its' servers itself, with only its own employees having access to them, for example in a locked cage or room, perhaps with biometric controls for entry, and so on. Or, X might use a third-party service provider to help run and maintain its servers. On those servers, X could install a proprietary or open-source cloud platform.[61] Some suppliers even sell physical servers with open-source or proprietary cloud software infrastructure pre-installed. X could offer the use of its cloud infrastructure to others as IaaS.[62] Or, X could build its own PaaS platform on this infrastructure, to develop and host its own applications for private cloud, or to offer PaaS services to others. PaaS platforms, whether stand-alone or built on existing IaaS platforms, may also be installable on X's equipment, for X's own use or offering to others as hosted services. Physical and software infrastructure could be managed by X, or a third party on its behalf, such as a systems integrator. X might have a separate consultancy, or other services contract, with an integrator to help it set up, manage, or support its cloud. X could install, on its own or third-party cloud infrastructure, application software it developed internally, or licensed from third parties. It could use these applications internally, as private cloud, or offer them as a service to others, as a SaaS provider. These illustrate that many combinations are possible, and users may not necessarily know how a cloud service has been put together or who supplies, provides, or operates different components.

Users may also combine different cloud providers' services. Ancillary support for primary cloud services includes analytics, monitoring, and cloud-based billing systems. SaaS across different providers is increasingly integrated.[63] Providers may use third-party cloud security providers, and integrate applications with, or support, 'non-cloud' components.

Cloud use is becoming increasingly sophisticated. With traditional IT, organizations may install and operate different applications, while with cloud, customers increasingly integrate different cloud applications and support services, with each other and with legacy internal systems.

[61] For instance, Canonical's Ubuntu Enterprise Cloud, which itself leverages OpenStack. This illustrates that even cloud platform software is not a single concept; there may be different kinds at different levels, for example with more user functionality added, such as with Ubuntu Cloud.

[62] For example, European telecommunications and managed services provider Colt uses VMWare's vCloud platform to offer private and public cloud services as 'virtual data centres' in Colt's physical data centres, and Colt also provides connectivity for those services.

[63] Google Apps MarketPlace enables customers of this SaaS productivity suite to use third-party SaaS integrated with, managed, and accessed through, Google Apps. 'Google Apps Marketplace Now Launched' (*Google Apps*, 10 March 2010), <http://googleappsupdates.blogspot.com/2010/03/google-apps-marketplace-now-launched.html> (accessed 8 March 2013).

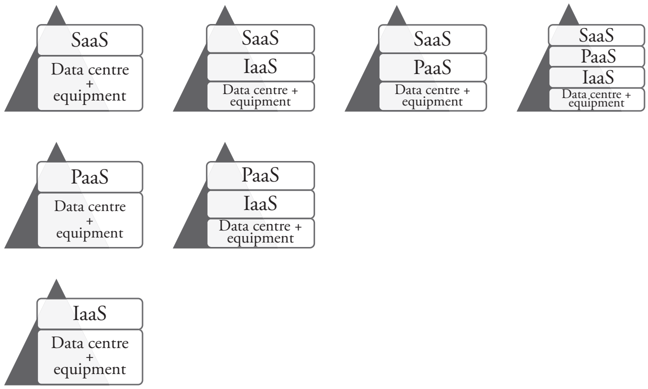

Figure 1.1 Cloud layers—different architectures

3.2 Layers or chains of cloud services

Cloud computing often involves a combination of 'layers'[64] of services and such layering may not be transparent to users. The classification of a service depends on exactly which layers and actors are under consideration. For example, customers of Dropbox may consider that they obtain a SaaS storage service from Dropbox. However, from the perspective of Dropbox, which built its SaaS service on Amazon's IaaS infrastructure, Amazon provides an IaaS service, which Dropbox uses to offer SaaS to its own customers.[65] Thus, Dropbox is both a cloud user (of Amazon IaaS) and cloud provider (of SaaS storage, to its customers).

Furthermore, as already mentioned, PaaS may be layered on IaaS, SaaS on PaaS or IaaS; triple layers are possible. Figure 1.1, albeit in a simplified way, illustrates the main alternatives.

Examples of the different alternatives are:

1. 'Unlayered'[66] IaaS, such as Amazon Web Services, RackSpace, GoGrid, or Google Compute Engine.
2. 'Unlayered' PaaS, such as Google App Engine, Microsoft Windows Azure, or Salesforce's Force.com.
3. PaaS on IaaS, such as dotCloud, Engine Yard, or Heroku (all built on Amazon IaaS).
4. 'Unlayered' SaaS, for example social networking or sharing services such as Facebook[67] and Flickr, webmail services such as Gmail and Outlook.com, and Salesforce's customer relationship management service.

[64] 'Layer' is not a defined term of art in this context.

[65] See Chapter 3 for a discussion of the contractual implications of this type of arrangement.

[66] Here, 'unlayered' refers to the cloud service. As mentioned in Section 3.1, physical infrastructure may be provided by parties other than the cloud service provider. Figure 1.1 shows 'data centre + equipment' as the implicit bottom, physical 'layer'.

[67] Facebook allows developers to create applications operating within Facebook, ie provides a 'platform' for developers to offer applications to Facebook users. However, strictly speaking Facebook does not provide PaaS, but only a way for applications to *interface* with its proprietary social networking platform. Facebook apps are, in fact, hosted elsewhere, such as on Amazon IaaS: Amazon Web Services, 'Facebook Application Hosting', available at <http://aws.amazon.com/facebook-application-hosting> (accessed 8 March 2013).

5. SaaS on IaaS, such as Dropbox, or Mozy (both on Amazon IaaS); indeed, any SaaS service built on Amazon, such as location-based consumer SaaS service Foursquare.
6. SaaS on PaaS, such as any SaaS service built on App Engine or Azure.[68]
7. SaaS on PaaS on IaaS: any SaaS service built on IaaS-based PaaS services, such as dotCloud or Heroku.[69]

This multiplicity of possible architectures for what appears, to the end user, to be a single cloud service, means that users may be dependent on several different providers and sub-providers,[70] including physical infrastructure providers. Different contractual arrangements for supply or provision of different components may also exist between different parties. Despite the potential importance for users of multiple dependencies, it is often difficult for users to know who is involved in 'hidden' service layers behind the direct provider, or to assess the risks of a hidden provider's service or equipment failing.

It is possible that a particular cloud user might be the only entity involved in a private cloud arrangement and might have direct control over every component of its cloud service 'stack'. In almost all cases, however, cloud computing arrangements are a new way of sourcing different IT resources from multiple providers and it is common for there to be complex relationships between users and providers and between providers and sub-providers.

It is tempting to regard cloud computing as just a new form of outsourcing. Many commercial, legal, and regulatory issues relevant to outsourcing do indeed apply to cloud computing. However, cloud computing has some fundamental characteristics that distinguish it from traditional outsourcing and which may affect the provider's or user's position in relation to risk management, contractual terms, and so on. In particular, there is considerable scope for confusion in dealing with layered services, particularly as regards assurances relating to security and sub-contractors.[71]

[68] Toyota's planned SaaS services for its car owners will run on Azure. Sharon Pian Chan, 'Microsoft and Toyota Bringing Cloud Computing to the Car' (*The Seattle Times*, 6 April 2011), <http://seattletimes.nwsource.com/html/microsoftpri0/2014698772_microsoftwillbuild.html> (accessed 8 March 2013).

[69] Many SaaS apps offered to users of Facebook or the Apple iPhone smartphone are developed and deployed using Heroku. For example, Heroku, 'Facebook Apps that Scale with Heroku', <http://success.heroku.com/cardinal_blue>; and 'Heroku Fuels Social App Development with New Facebook Program' (*Heroku*, 9 November 2010), <http://news.heroku.com/news_releases/heroku-fuels-social-app-development-with-new-facebook-program> (both accessed 8 March 2013).

A concrete example: Chase Jarvis's Best Camera app for the iPhone smartphone was produced and deployed using Heroku. A user of that app may use an iPhone to photograph another person, then store and share that photo via the app, ultimately on Amazon's servers. Here, we have three providers: Chase Jarvis provides the iPhone app to the end user; Heroku (sub-provider) provides the platform used by Chase Jarvis to build and deliver the app; and Amazon (sub-sub-provider) provides the underlying computing infrastructure. The end user may not know, and may not necessarily be concerned to know, that Best Camera uses Heroku/Amazon. Best Camera App, available at <http://thebestcamera.com/app.html> and Heroku, 'Launched Top iPhone App & Social Site with Heroku', <http://success.heroku.com/ubermind> (both accessed 8 March 2013).

[70] For example, when Amazon Web Services suffered an outage in its US East Region in April 2011, SaaS providers who relied on Engine Yard and Heroku were also affected: Derrick Harris, 'Cloud Platforms Heroku, DotCloud & Engine Yard Hit Hard by Amazon Outage' (*Gigaom*, 21 April 2011), <http://gigaom.com/cloud/more-than-100-sites-went-down-with-ec2-including-your-paas-provider> (accessed 8 March 2013).

[71] These issues are explored further in Chapter 2, Section 5. See also Chapter 4 and W Kuan Hon and Christopher Millard, 'Cloud Computing vs Traditional Outsourcing—Key Differences' (*Social Science Research Network*, 12 September 2012), <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2200592> (accessed 8 March 2012).

3.3 Cloud compared with grid and utility computing

Grid computing[72] also involves distributed computing harnessing multiple physical machines. However, a key difference lies in how resources are allocated, and perhaps also in cloud's greater 'on demand' capability.[73] With grid computing, large amounts of data are requested by a small number of users (few but large allocation requests are made), whereas cloud computing better suits environments with many users requesting small amounts of data (many but small allocation requests): ' "Grids are well suited for complex scientific work in virtual organizations," . . . Clouds, on the other hand, are well suited for simple work such as many short-running jobs.'[74] Or, to look at it another way, 'In a computational grid, one large job is divided into many small portions and executed on multiple machines. This characteristic is fundamental to a grid; not so in a cloud'.[75]

Utility computing is used generally to refer to the commoditization of computing, whereby computing services are delivered like water, electricity, gas, and telecommunications utilities. Cloud computing is one way to deliver utility computing, including access on demand and paying only for what is used.[76]

4. Concluding Remarks

To summarize, many components may be involved in one cloud service, with different possible deployment models and layers, and cloud supply chains and contractual relationships may be complex. Chapter 2 explores what this means in terms of control, risk, and security in the cloud.

[72] For a detailed comparison of cloud computing and grid computing, by business model, architecture, resource management and programming, application and security models, see Ian Foster et al., 'Cloud Computing and Grid Computing 360-Degree Compared', available at <http://arxiv.org/ftp/arxiv/papers/0901/0901.0131.pdf> (accessed 8 March 2013).

[73] Judith Myerson, 'Cloud Computing versus Grid Computing' (*IBM*, 3 March 2009), <http://www.ibm.com/developerworks/web/library/wa-cloudgrid> (accessed 8 March 2013).

[74] Jennifer Schiff, 'Grid Computing and the Future of Cloud Computing' (*Enterprise Storage Forum*, 21 January 2010), <http://www.enterprisestorageforum.com/outsourcing/features/article.php/3859956/Grid-Computing-and-the-Future-of-Cloud-Computing.htm>; and Thorsten, 'Cloud Computing vs Grid Computing' (*RightScale Blog*, 7 July 2008), <http://blog.rightscale.com/2008/07/07/cloud-computing-vs-grid-computing> (both accessed 8 March 2013).

[75] Karishma Sundaram, 'Cloud Computing vs Grid Computing' (*Bright Hub*, 20 May 2011), <http://www.brighthub.com/environment/green-computing/articles/68785.aspx> (accessed 8 March 2013)—thus, Hadoop (n. 33) might be considered close to grid computing.

[76] Rajkumar Buyaa et al., 'Cloud Computing and Emerging IT Platforms: Vision, Hype and Reality for Delivering Computing as the 5th Utility' (*Science Direct*, June 2009), <http://www.sciencedirect.com/science/article/pii/S0167739X08001957> (accessed 8 March 2013); although as mentioned earlier in this chapter, some cloud services are free, and others are charged on, for example, a per-month per-user basis rather than on amount used.