Computing without Space: Cloud Computing

Distributed Systems

Sistemi Distribuiti

Andrea Omicini
andrea.omicini@unibo.it

Dipartimento di Informatica – Scienza e Ingegneria (DISI)
Alma Mater Studiorum – Università di Bologna a Cesena

Academic Year 2016/2017

Material from the Books

- these slides contain material from
  [Coulouris et al., 2012, Baun et al., 2011, Hill et al., 2013]
- some text and some pictures, too, were extracted from the books
- every problem or mistake contained in these slides, however, should be attributed to the sole responsibility of the teacher of this course

Next in Line. . .

Towards Cloud Computing [Coulouris et al., 2012] I

Resources as commodities

- with the increasing maturity of distributed systems infrastructure, a number of companies are promoting the view of distributed resources as a commodity or utility, drawing the analogy between distributed resources and other utilities such as water or electricity.
- with this model, resources are provided by appropriate service suppliers and effectively *rented rather than owned* by the end user.
- this model applies to both physical resources and more logical services

Towards Cloud Computing [Coulouris et al., 2012] II

Physical resources

- *physical resources* such as storage and processing can be made available to networked computers, removing the need to own such resources on their own
- at one end of the spectrum, a user may opt for a remote storage facility for file storage requirements
- at the other end of the spectrum, users can access sophisticated data centres or computational infrastructure using the sort of services now provided by companies such as Amazon and Google
- operating system virtualisation is a key enabling technology for this approach, implying that users may actually be provided with services by a virtual rather than a physical node, thus offering greater flexibility to the service supplier in terms of resource management

Towards Cloud Computing [Coulouris et al., 2012] III

Logical resources

- *software services* can also be made available across the global Internet using this approach
- many companies offer a comprehensive range of services for effective rental, including services such as email and distributed calendars—e.g., Google Apps
- this development is enabled by agreed standards for software services, for example as provided by *web services*

Towards Cloud Computing [Coulouris et al., 2012] IV

Cloud computing

- the term cloud computing is used to capture this vision of *computing as a utility*
- a cloud is defined as a set of Internet-based application, storage and computing services sufficient to support most users' needs, thus enabling them to largely or totally dispense with local data storage and application software
- the term also promotes a view of *everything as a service*, from physical or virtual infrastructure through to software, often paid for on a per-usage basis rather than purchased
- note that cloud computing reduces requirements on users' devices, allowing very simple desktop or portable devices to access a potentially wide range of resources and services

Next in Line. . .

Definitions & Remarks I

Definition [Baun et al., 2011]

By using virtualized computing and storage resources and modern Web technologies, cloud computing provides scalable, network-centric, abstracted IT infrastructures, platforms, and applications as on-demand services. These services are billed on a usage basis.

Definitions & Remarks II

Remarks

- even though cloud services usually rely on *distributed systems*, the definition do not specify this
  - in principle, it might be provided also by a single, high-performance server, such as a mainframe
  - also, cloud management is typically determined in a central (and proprietary) manner by a single provider
  - unlike grid computing, where distributed nodes are usually autonomous
- also, a fundamental issue for cloud computing is its *economic impact*
  - being strictly service-oriented and Web-standard-based, cloud computing is an amenable choice fur a huge number of application scenarios
  - many IT approaches and solutions might exploit cloud services and get economical benefits

Definitions & Remarks III

NIST

Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model is composed of five essential characteristics, three service models, and four deployment models.

NIST, US Department of Commerce
`http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf`

Definitions & Remarks IV

Five essential features according to NIST

on-demand self-service — Services can be provided on demand to consumers with no humans in the loop

broad network access — Services are available over the network in real-time through standard mechanisms

resource pooling — Resources are pooled to enable concurrent service provision to multiple users (multi-tenant model), while being adjusted to the actual demand of each user

(rapid) elasticity — Requests for extra resource are self-managed and automatic in relation to demand; from the consumer's perspective, the supply of resources has no limit

measured (quality of) service — Services leverage a quantitative and qualitative metering capability making usage-based billing and validation of the service quality available

Next in Line. . .

Background

Technologies behind the Cloud

There are some fundamental technologies the Cloud depends upon

- virtualisation
- service-oriented architectures (SOA)
- Web Services

Virtualisation I

Resources in the cloud

- resources can be detached from their physical nature by means of *virtualisation*
- virtualised resources for the cloud include servers, data stores, networks, and software
- the basic idea is to *pool* physical resources, make them transparent, and manage them as a whole
- pools provide transparent access to resources

Virtualisation II

Benefits for providers

resource usage — Load balancing towards full capacity of physical servers

management — Resource pool management can be automated

consolidation — Reducing the number of physical servers, increasing efficiency and cost effectiveness

energy consumption — From consolidation

less space required — From consolidation

emergency planning — Moving virtual machines from one resource pool to another

Virtualisation III

Benefits for customers

dynamic behaviour — Resource pools can easily satisfy almost all requests dynamically—even when requests are peaking

availability — Automated fault tolerance procedures make services highly available

access — Isolation of virtual machines make it possible to safely delegate management functionality to the customer

emancipation — Customers can purchase IT capabilities from a self-service portal, without humans in the loop

Virtualisation IV

Drawbacks

cost of abstraction — the abstraction layer requires resources; however, modern virtualisation techniques make such a cost more and more negligible

system management — the physical infrastructure grows in complexity; automation of procedures and availability of management tools requires however much less staff in the overall

Virtualisation V

Concepts

OS virtualisation — applications run within *containers* layered upon the host OS

platform virtualisation — OS & applications run in a virtual environment, based on a VM monitor or *hypervisor*

storage virtualisation — dynamically scalable storage space as a service

network virtualisation — virtual IP addresses, virtual LAN (VLAN)

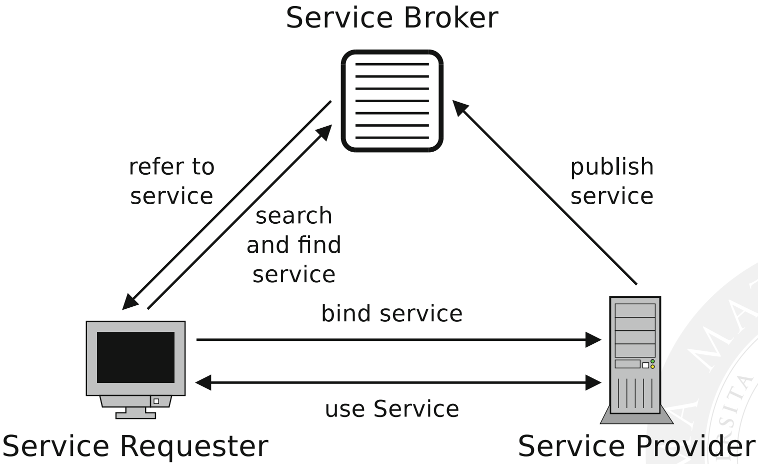
application virtualisation — software sales model for centralised management and distribution of applications through the network

Service-Oriented Architectures (SOA) I

Another prerequisite for cloud computing

- service-oriented architectures (SOA) are architectures whose components are implemented as independent services
- services interact via messages, and coordinate together (orchestration)
- cloud makes virtualised IT infrastructures, platforms, and whole applications available for use

Service-Oriented Architectures (SOA) II



[Baun et al., 2011]

Web Services I

Internet resources as WS

- distributed cloud computing systems integrate heterogeneous resources
- unpredicability of Internet connection demands for loosely coupled, asynchronous, message-based communication

$\rightarrow$ Web Services

Web Services II

Web Services Architecture Working Group

A Web Service is a software application identified by a URI, whose interfaces and binding are capable of being defined, described and discovered by XML artifacts and supports direct interactions with other software applications using XML based messages via internet-based protocols.

`http://www.w3.org/TR/2002/WD-wsa-reqs-20020429`

Next in Line. . .

Perspectives on Cloud

Technical vs. organisational viewpoints

- a *technical* viewpoint focuses on the functional features of cloud architectures
- however, this is not the only one available and meaningful perspective: the way in which the cloud is deployed is also relevant
- in particular, the different roles played by providers and customers is essential for the *organisational* viewpoint

Focus on. . .

Public Cloud

- a *public cloud* (or, *external* cloud) is offered by a provider to users outside its organisation boundaries
- the cloud service is made publicly available through an automated service (typically, a Web portal) with no humans in the loop
- obligations are pre-defined in terms of the service, and automatically negotiated
- billing based on actual usage

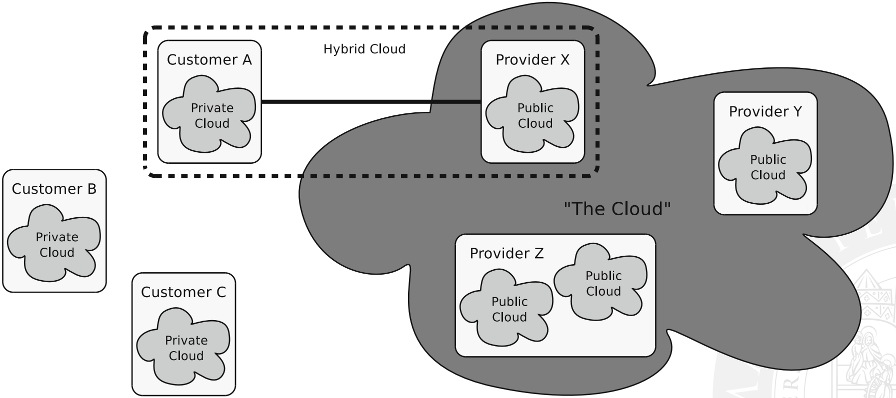
Private Cloud

- providers of a *private cloud* (or, *internal* cloud) belong to the same organisation as the users
- the cloud service is made publicly available through an automated service (typically, a Web portal) with no humans in the loop
- the main aim of private cloud is security, by keeping control of organisation data

Hybrid Cloud

- in order to exploit the same tech & tools of public clouds
- and to scale automatically when private cloud is not enough
- *hybrid clouds* are built by bringing together resources from both private and public clouds
- there, normal operation is handled by the private cloud, while load peaks are typically transferred to the public cloud, obviously excluding data and resources that are critical for the organisation

Public / Private / Hybrid Clouds



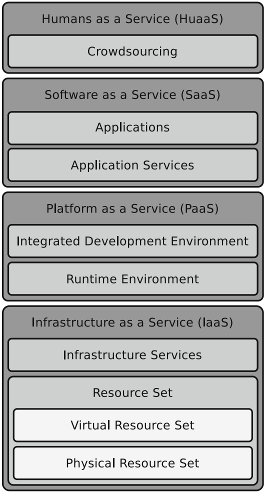
[Baun et al., 2011]

Focus on. . .

Everything as a Service (XaaS)

- through the cloud, so many diverse infrastructures, services, and applications are made available as Web services
- as a result, cloud services present with diverse functions and goals
- the cloud offering can be then depicted as an architecture layered according to different layers of abstractions
- the three main layers follow the Everything-as-a-Service paradigm (XaaS)

Cloud Stack



[Baun et al., 2011]

Infrastructure as a Service (IaaS)

- the *IaaS layer* provides an abstract view on the physical resources — computers, mass storage systems, networks, etc.
- a user interface is provided to manage resources in the resource set sub-layer (RS), allowing users allocate and use them
- items in the resource set can be subdivided into
  - physical resource set (PRS), representing and offering proprietary physical hardware—e.g., Emulab, iLO
  - virtual resource set (VRS), built on top of virtualisation technologies, and making virtual instances available—e.g., Amazon EC2, Nimbus, OpenNebula
- infrastructure services, too, belong to the IaaS layer—e.g., Hadoop MapReduce, Amazon S3, Dropbox

Platform as a Service (PaaS)

- the *PaaS layer* provides developers with tools for the cloud
- programming environments (PE) and execution environments (EE) for programming the cloud

PE e.g., Django Framework, Sun Caroline

EE e.g., Google App Engine (Python, Java), Azure by Microsoft (multilanguage)

Software as a Service (Saas)

- the *SaaS layer* directly provides users with software via cloud
- applications are no longer installed locally, but are instead stored and managed by the providers, and downloaded when needed
- application services, such as Google Maps
- full-fledged applications, such as Google Docs, Microsoft Windows Live

Humans as a Service (Huaas)

- the *HuaaS layer* is the top layer of the cloud computing stack, showing that the cloud paradigm can be extended to include services provided by *human beings as resources*
- until humans exhibit capabilities that outperform computer systems, their technical integration as resources is potentially useful
- e.g., *crowdsourcing*—such as Amazon Mechanical Turk

Next in Line...

Service Level Agreements (SLAs)

- a service level agreement is an agreement between a service provider and a service consumer related to the service level (quality of service, QoS)
- the SLA implies a mutual agreement with respect to security, priorities, responsibilities, guarantees, and billing modalities
- furthermore, SLA specifies metrics such as availability, throughput, response times, and others
- two phases for service level management
    - agreement on the quality of service
    - service monitoring at runtime

Lifecycle & Automation

- cloud services have a lifecycle: made available by the provider, selected and used by the customer, closed and detached at the end, finally accounted for
- groups of similar resources are managed automatically as an *ensambles*, stepping through *monitoring*, *analysis*, *scheduling*, *execution*

Management Services & Tools

- many tools are made available by providers, either via CLI or via Web
- for
  - *monitoring*
  - *control*
  - *development*

Security Management

- nothing new here, safety goals are the same as a local data center
  - confidentiality
  - integrity
  - availability
  - authenticity
  - accountability
  - pseudonymity

Risk Management

- again, cloud does not add much in the risk management perspective
- risks are more or less the same of any outsourcing process, and linked to data transfer, service disruption, provider bankruptcy, vendor lock-in

Legal Compliance

- compliance ensures that laws are applicable to cloud computing as to renting (in case of paid services) or lending (with respect to unpaid services)
- geographical location of providers affects applicable law
- e.g., according to article 4 of the European Data Protection Directive, the pertaining national data protection legislation of the respective country is applicable—which is why cloud services are typically provided region-based
- particular attention should be devoted to legal aspects related to personal data processing—e.g., in Europe directive 2000/31/EC is applicable
- special legislation takes care of sensitive data in scenarios such as healthcare and finance

References

Baun, C., Kunze, M., Nimis, J., and Tai, S. (2011).
Cloud Computing. Web-Based Dynamic IT Services.
Springer Berlin Heidelberg.

Coulouris, G., Dollimore, J., Kindberg, T., and Blair, G. (2012).
Distributed Systems. Concepts and Design.
Pearson, 5th edition.

Hill, R., Hirsch, L., Lake, P., and Moshiri, S. (2013).
Guide to Cloud Computing. Principles and Practice.
Computer Communications and Networks. Springer London.

Computing without Space: Cloud Computing

Distributed Systems

Sistemi Distribuiti

Andrea Omicini
andrea.omicini@unibo.it

Dipartimento di Informatica – Scienza e Ingegneria (DISI)
Alma Mater Studiorum – Università di Bologna a Cesena

Academic Year 2016/2017