Cloud Computing

II Workshop di Project Management

Ordine degli Ingegneri di Salerno

Vittorio Scarano

Università di Salerno

27 Gennaio 2012

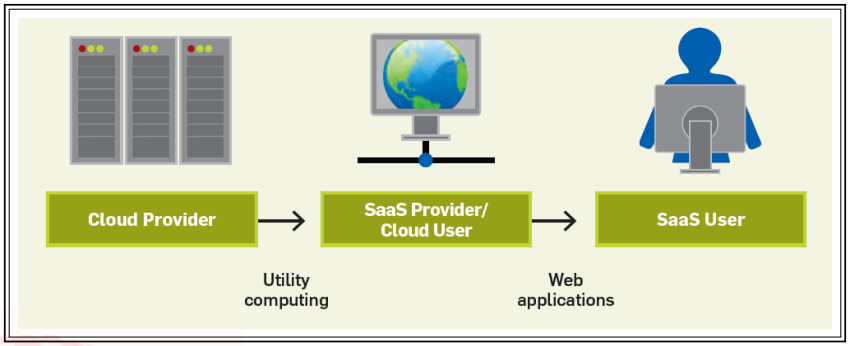
Plan

Plan

Cloud Computing: a definition

- Cloud computing refers to both:
  1. applications delivered as service over the Internet (also called *Software as a Service* (SaaS))
  2. hardware and systems software in the data centers providing the services
- Some other names: Infrastructure/Platform as a Service (IaaS/PaaS)
- H/S in Data Centers is the *cloud*
- *Public* cloud: utility computing
- *Private* cloud: "large-enough" data centers of private organizations

The roles in the cloud: users and providers



A closer look

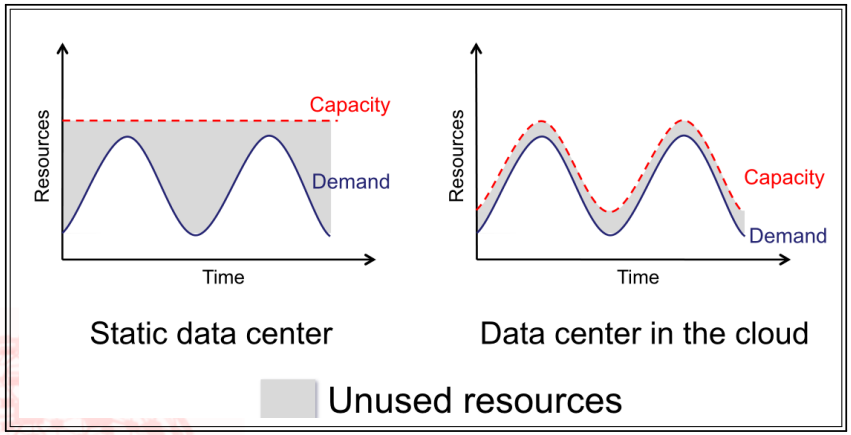
Cloud computing can be divided into three areas (or layers):

- SaaS (software-as-a-service). WAN-enabled application services (e.g., Google Apps, Salesforce.com, WebEx)
- PaaS (platform-as-a-service). Foundational elements to develop new applications (e.g., Coghead, Google Application Engine)
- IaaS (infrastructure-as-a-service). Providing computational and storage infrastructure in a centralized, location-transparent service (e.g., Amazon)

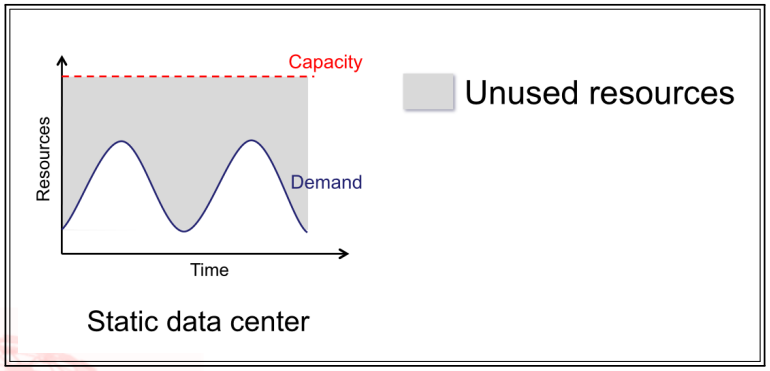
What is new in Cloud Computing

- *Infinite* computing resources available on demand, quickly (no "plan ahead" in designing large data centers)
- No up-front committment from the users: one can start a small company and grow easily and cheaply as new needs appear
- Pay per use on a short term basis (computation by the hour, storage by the day)

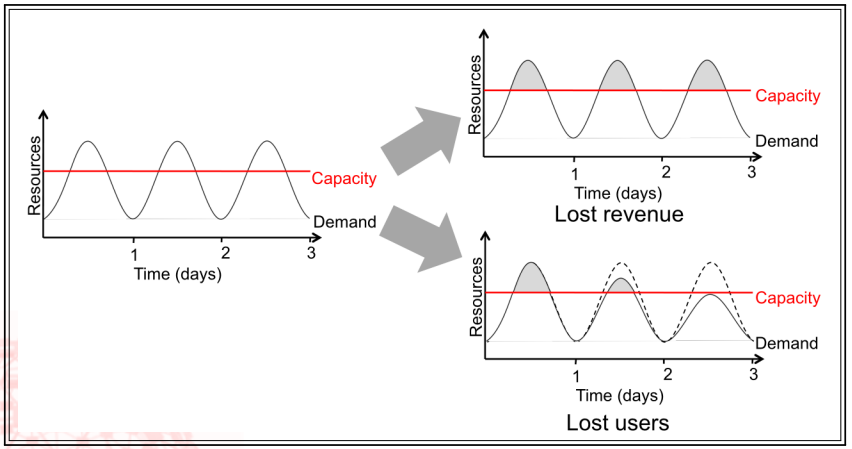
Pay-per-use vs. Provisioning for Peaks



Penalty for Over-provisioning



Heavier Penalty for Under-provisioning



The "Kiss of death" of success in the first Internet Bubble!

Plan

Larry Ellison (Oracle)

The interesting thing about Cloud Computing is that we've redefined Cloud Computing to include everything that we already do...

I don't understand what we would do differently in the light of Cloud Computing other than change the wording of some of our ads.

Andy Isherwood (HP VP)

A lot of people are jumping on the cloud bandwagon, but I have not heard two people say the same thing about it.

There are multiple definitions out there of "the cloud."

Richard Stallman (Free software guru)

It's stupidity.

It's worse than stupidity: it's a marketing hype campaign.

Somebody is saying this is inevitable - and whenever you hear somebody saying that, it's very likely to be a set of businesses campaigning to make it true.

Plan

"Large" data centers

How "large" should a large-datacenter be?

- Comparing medium-sized data centers ($\sim$ 1000 servers) with large data centers ($\sim$ 50000 servers), the savings in hardware, network and workforce are rather significant
- An estimate in 2006:

Technology	Medium size	Large size	Ratio
Network	\$95 per Mbit/sec/-month	\$13 per Mbit/sec/-month	7:1
Storage	\$2.20 per Gbyte/-month	\$0.40 per Gbyte/-month	5:1
Administration	$\sim$ 140 servers/administrator	$\sim$ 1000 servers/administrator	7:1

Key enabling technologies

- Inexpensive storage
- Inexpensive and plentiful client CPU bandwidth to support significant client computation
- Sophisticated client algorithms, including HTML, CSS, AJAX, REST
- Client broadband
- SOA (service-oriented architectures)
- Large infrastructure implementations (like Google, Yahoo, Amazon), intended originally only for internal usage that are now able to provide real-world, massively scalable, distributed computing
- Commercial virtualization
- Web 2.0 vs. Web 1.0: easy, pay-as-you-go services (like PayPal for payment, AdSense for ads, Amazon CloudFront vs. Akamai for content distribution)

New application opportunities

- Mobile interactive applications, relying on large data hosted on large datacenters
- Parallel batch processing: using hundreds of computer together for a short time costs as much as using few computer for a long time.
  - in this context, motivating factor is the availability of both algorithmic frameworks (like Map-reduce) and open-source frameworks (like Hadoop)
- Business analytics
- Compute-intensive Desktop Applications:
  - virtualized apps over small desktop that scale-up to the cloud
  - data and heavy compute in the cloud and interfaces and GUI on the desktop (image rendering or 3D animation)

The economy

- Developing an application service requires a large CapEx (capital expense)
  - to build infrastructure for peak service demand before deployment.
- Cloud computing allows expenses to be related with resource use, thus following income rather than having to purchase for peak capacity before income is realized.
  - It moves CapEx to OpEx (operational expense).
  - Reduced system-administration head count by avoiding the need for internally purchased servers.
  - It smooths the path to service scaling by not requiring the CapEx-intensive architectural changes needed to scale up service capacity in the event of service success.

Plan

The reference model

- Any application needs:
  1. A model of computation
  2. A model of storage
  3. A model of communication (distributed applications)
- The economics pushes toward the statistical multiplexing of the same hardware by different servers requires virtualization
- Depending on abstraction and management of resources, different services will be offered to the programmer

"Close to hardware"

- Amazon
- At the one end of the spectrum, since it looks like physical hardware
- Users can control CPUs, cores, memory, ... but also system software etc.
- Thin API exposed to control and configure the virtualized hardware

⊕ No apriori limitations: on a generic CPU(s) with memory, disk, etc. anything that can be coded can be executed

⊖ On the downside: difficult to manage automatic scalability and failover: the semantics of replication are highly application-dependant and not "known" by the underlying system that manages the cloud

"Close to the application"

- Google AppEngine, Force.com
- Close to the application-specific platforms
- Targeted at specific domains: Google AppEngine is aimed at web applications
- For Google AppEngine:
  - traditional separation in stateless computation and stateful storage
  - request-reply based

⊕ The approach facilitates automatic management and scaling up

⊖ Abstractions forced onto the programmer

"In the middle"

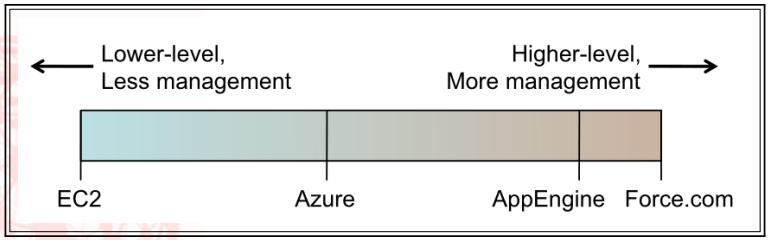
- Microsoft Azure
- In-between the extreme flexibility of Amazon and the programmer convenience of AppEngine
- Azure is based on .Net libraries, executed on the CLR runtime
- Thereby, it supports general purpose computing, with the choice of language (all Microsoft, of course!)

⊕ Some automatic network configuration and management of failover/scalability

⊖ ... with the help of the programmer that must declaratively instrument the applications

A comparison

- In a way, the fight between maximum flexibility and efficiency (but complex and difficult) vs. high-level, domain-specific solutions (immediate but somewhat less flexible and efficient)
- Similarity with the power and expressiveness of C and the immediate results obtained by frameworks like Rubyb on Rails for web-based applications.
- Different problems will result in different kind of solutions



Plan

Utility computing vs. conventional hosting

1. When demand for a service varies with time
   - Provisioning a data center for the peak load ...
   - ... leaves is underutilized for the rest of the times
   - Example: Christmas gifts for e-commerce
2. When demand is unknown in advance
   - Web startups: spike induced by popularity, followed by disappointment by low performances
3. "Cost associativity" to finish faster
   - 1000 EC2 machines for one hour costs as much as 1 machine for 1000 hours

Cloud-computing as a commodity

- SpotCloud: the first market for cloud computing
- The idea: firms with excess computing capacity put it up for sale. Others, with short-term needs can bid for it
- The firm that manages the marked (Enomaly) gets it 10%-30% share
- Opaque market: you don't know where you are buying from
- Enomaly has no infrastructure
  - It is on the Cloud itself! (Google App Engine)
- Example: entertainment company offered 4000 servers that were unused (a "between movies" company?)

Cloud computing and economy

- Because the cost of deploying new services is much lower and expenses track real usage, ...
- ... businesses can develop and deploy more services without huge capital investments
- Start-ups are able to go completely without infrastructure, focussing on its core differentiating efforts
- Larger enterprises are equally focused on cost and flexibility (service changes and agility).

Success stories - 1

- An international financial exchange
  - It hosted data in the cloud
  - Application on client's desktop.
  - All operations were on a pay-as-you-go basis: very low initial investment required to make a commercial service operational.
- Shazam:
  - It samples songs being played on the radio,
  - Matches the songs to a library in the cloud and returns a link to purchase that song on the iPod.
  - Smart device coupled with cloud-based computation and storage.
- Animoto, hosted on Amazon
  - FB application: video clip from user selected pictures, video and music
  - able to scale up from 50 instances to 3,500 instances over a three-day period (750.000 new users!)

Success stories - 2

- A national newspaper wanted to place scanned images covering a 60-year period online.
  - After being repeatedly turned down by the CIO for the use of six servers,
  - the newspaper moved four terabytes into S3,
  - ran all the software over a weekend on EC2 for $25, and launched its product.
- A major international auto-race organizer: live streaming video and realtime technical information.
  - It had to retain an ISP, acquire massive server power, and hire 500 engineers to baby-sit the servers at the ISP to institute server failover manually.
  - When on EC2, the savings in server rental were not that big, but it did realize several orders of magnitude in personnel cost savings.

Success stories - 3

- Mogulus streams 120,000 live TV channels over the Internet
  - It owns no hardware except for the laptops it uses.
  - It did all of the election coverage for most of the large media sites.
  - Its CEO states that he could not be in business without IaaS.
- Peter Harkins, a Senior Engineer at The Washington Post:
  - used 200 EC2 instances (1,407 server hours) to convert 17,481 pages of Hillary Clinton's travel documents
  - nine hours after they were released they were on WWW into a friendly format

,

Plan

The rise of a new era...

- The first era: the mainframe (key players: IBM, ... and the others)
- The second era: the PC (key players: IBM, but also Microsoft, Apple, Intel...)
- The third era: the Internet (key players: Cisco, Dell, Sun, EBay, Yahoo, Google)
- The fourth era: Cloud Computing (key players: "*place your favorite, here*")

Some benefits ...

- Allow rapid staging, set-up and take-down of a variety of compute environments as needed to test and validate an application
- Allows management techniques in large projects, previously unfeasible due to limited computing power, storage, network
- Lead to greater economies of scale, facilitating architectural standardization and process optimization, with "natural" reuse of components

... and some risks

- If in a dedicated IT environment, the possibility of failure is high ...
- ...in the cloud, the risk grows exponentially
- The parallel processing and the scalability, besides remoteness and security, require robust project management
- But the concept of "infinite" resources is really, really, appealing...

What we have done today

References

- "A view on Cloud Computing", Armbrust and others (Berkeley view). Comm. of ACM April 2010, vol.-53. No.4
- "Cloud Computing - A classification, Business Models and Research Directions". C. Weinhart and others. Business & Information Systems Engineering, 2009
- "A market for computing power". The Economist. Feb 17th 2011.
- "Cloud Computing meets Project Management". PMWorld Today, June 2010 (Vol. XII, Issue VIII)