Cloud Computing: Introduzione

Vittorio Scarano
Corso di Programmazione Distribuita
Laurea in Informatica
Università di Salerno

Indice

- La conclusione di un (lungo) viaggio
- Calcolo scalabile su internet
- Modelli per il calcolo distribuito e su cloud
- Cloud Computing e modelli di servizio

Indice

- **La conclusione di un (lungo) viaggio**
- Calcolo scalabile su internet
- Modelli per il calcolo distribuito e su cloud
- Cloud Computing e modelli di servizio

La struttura del corso di PD

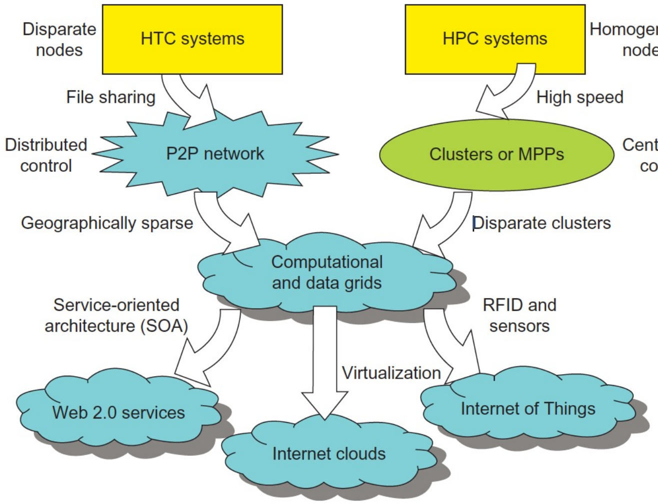
- Partendo dalla concorrenza…
  - necessaria per l'overlapping di comunicazione e computazione
- .. si esaminano i socket
  - soluzione efficiente, ma poco flessibile e con pochi servizi di supporto
- .. per poi passare a Remote Method Invocation
  - astrazione maggiore, familiare ai programmatori
- .. per esaminare le architetture Enterprise
  - che offrono integrazione di layer diversi (data, business, presentation, services)
- .. studiando la comunicazione orientata a messaggi (MOM)
- .. e le Architetture Orientate a Servizi
- Ed arrivare (alla fine!) a qualche cenno su Cloud e Microservices

Indice

- La conclusione di un (lungo) viaggio
- **Calcolo scalabile su internet**
- Modelli per il calcolo distribuito e su cloud
- Cloud Computing e modelli di servizio

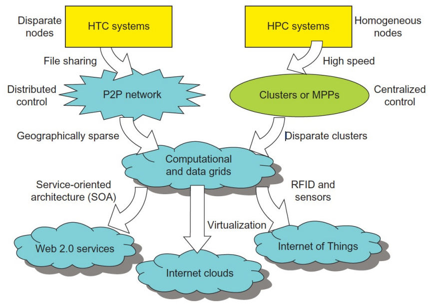
Computazioni scalabili

- Cambiamenti nel paradigma di computazione: da calcolo monolitico a calcolo distribuito
- Data-intensive e network centric
- Scalabilità il punto centrale:
  - Da high performance computing
  - A high-throughput computing



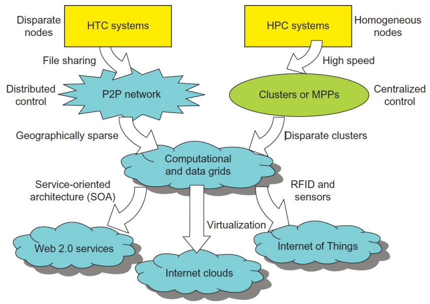
High performance computing

- Guidato dalle richieste di calcolo scientifico (fisica, chimica, manufacturing)
- Focus sulle prestazioni pure: operazioni in floating point
- Linpack benchmarking
- Top 500 list: una sfida annuale



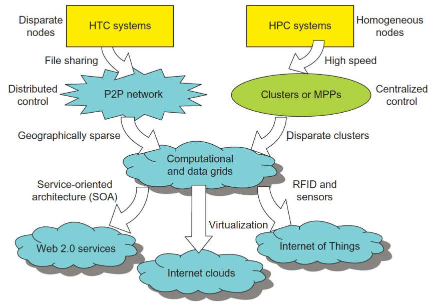
Piattaforme di calcolo

- Reti P2P: file sharing
  - Non strutturate
  - Calcolo cooperative
  - Best effort (at best!)
- Calcolo massivo
  - Prima macchine distribuite strettamente accoppiate
  - Poi, cluster di nodi semi-eterogenei
- Computational grid
  - Il calcolo come utility (elettricità, acqua)



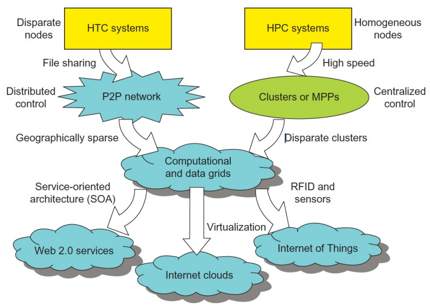
High Throughput computing

- Focus sulla quantità di dati che possono essere calcolati
- Ricerche su Internet e web services
  - Audience e dimensione è quella di Internet
- Scalabilità estrema
- Goal: non i FLOPS ma il numero di tasks completati per unità di tempo



Nuovi paradigmi di calcolo

- Service Oriented Architecture
  - Disaccoppiamento e eterogeneità
- RFID, GPS, sensori
  - Internet of Things (IoT)
- Internet cloud
  - 1984: The network is the computer (Sun)
  - 2008: The data center is the computer (D. Patterson)
  - Now: The cloud is the computer



Classificazioni del calcolo

- Calcolo monolitico,
  - centralizzato
- Calcolo parallelo
  - strettamente accoppiato
- Calcolo distribuito
  - debolmente accoppiato,
  - con nodi autonomi
- Calcolo su cloud
  - "utility/service computing"

Calcolo concorrente

Famiglie di sistemi distribuiti

- Computational grids
- Tipicamente per applicazioni scientifiche
- Strettamente accoppiate (grid) o no (P2P)
- Obiettivi di design di un sistema distribuito:
  - Efficienza (uso del parallelismo, FLOPS o Job throughput)
  - Affidabilità: QoS
  - Adattabilità a diversi workload di dimensioni variegate, e a diversi modelli di servizio
  - Flessibilità nella realizzazione di applicazioni in HPC (science) e HTC (business)

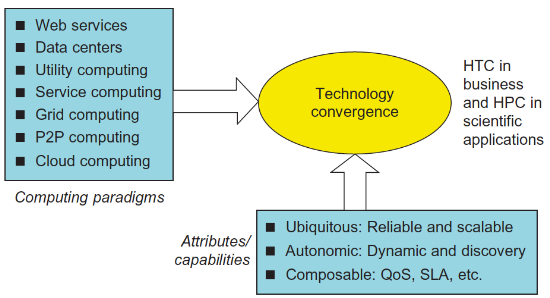
Applicazioni di HPC e HTC

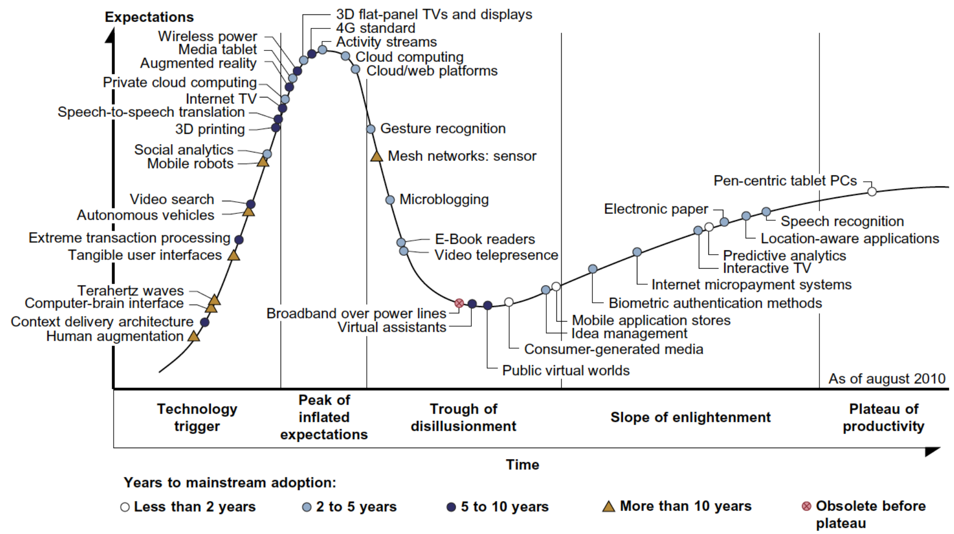
Table 1.1 Applications of High-Performance and High-Throughput Systems

Domain	Specific Applications
Science and engineering	Scientific simulations, genomic analysis, etc. Earthquake prediction, global warming, weather forecasting, etc.
Business, education, services industry, and health care	Telecommunication, content delivery, e-commerce, etc. Banking, stock exchanges, transaction processing, etc. Air traffic control, electric power grids, distance education, etc. Health care, hospital automation, telemedicine, etc.
Internet and web services, and government applications	Internet search, data centers, decision-making systems, etc. Traffic monitoring, worm containment, cyber security, etc. Digital government, online tax return processing, social networking, etc.
Mission-critical applications	Military command and control, intelligent systems, crisis management, etc.

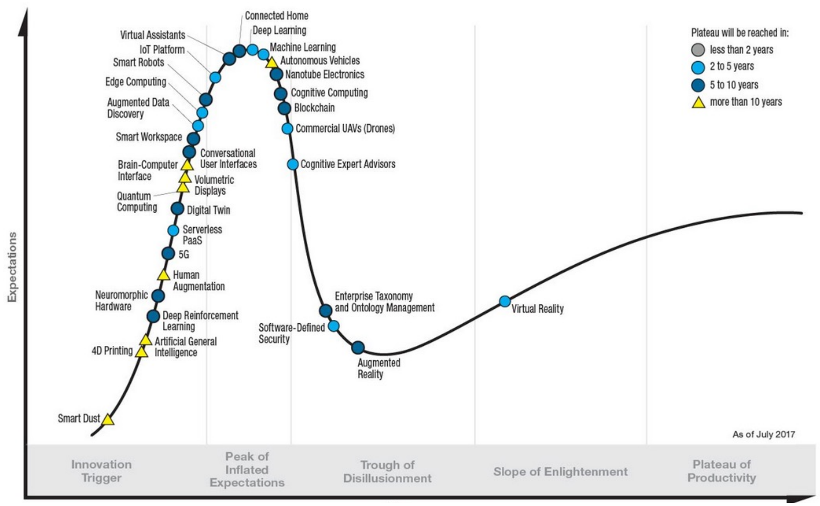
Utility computing

- Modello di business in cui il client riceve calcolo per un canone
- Cloud e grid
- Convergenza tecnologica





Gartner Hype Cycle for Emerging Technologies, 2017



Indice

- La conclusione di un (lungo) viaggio
- Calcolo scalabile su internet
- **Modelli per il calcolo distribuito e su cloud**
- Cloud Computing e modelli di servizio

Table 1.2 Classification of Parallel and Distributed Computing Systems

Functionality, Applications	Computer Clusters [10,28,38]	Peer-to-Peer Networks [34,46]	Data/ Computational Grids [6,18,51]	Cloud Platforms [1,9,11,12,30]
Architecture, Network Connectivity, and Size	Network of compute nodes interconnected by SAN, LAN, or WAN hierarchically	Flexible network of client machines logically connected by an overlay network	Heterogeneous clusters interconnected by high-speed network links over selected resource sites	Virtualized cluster of servers over data centers via SLA
Control and Resources Management	Homogeneous nodes with distributed control, running UNIX or Linux	Autonomous client nodes, free in and out, with self-organization	Centralized control, server-oriented with authenticated security	Dynamic resource provisioning of servers, storage, and networks
Applications and Network-centric Services	High-performance computing, search engines, and web services, etc.	Most appealing to business file sharing, content delivery, and social networking	Distributed supercomputing, global problem solving, and data center services	Upgraded web search, utility computing, and outsourced computing services
Representative Operational Systems	Google search engine, SunBlade, IBM Road Runner, Cray XT4, etc.	Gnutella, eMule, BitTorrent, Napster, KaZaA, Skype, JXTA	TeraGrid, GriPhyN, UK EGEE, D-Grid, ChinaGrid, etc.	Google App Engine, IBM Bluecloud, AWS, and Microsoft Azure

Table 1.2 Classification of Parallel and Distributed Computing Systems

Functionality, Applications	Computer Clusters [10,28,38]	Peer-to-Peer Networks [34,46]	Data/ Computational Grids [6,18,51]	Cloud Platforms [1,9,11,12,30]
Architecture, Network Connectivity, and Size	Network of compute nodes interconnected by SAN, LAN, or WAN hierarchically	Flexible network of client machines logically connected by an overlay network	Heterogeneous clusters interconnected by high-speed network links over selected resource sites	Virtualized cluster of servers over data centers via SLA
Control and Resources Management	Homogeneous nodes with distributed control, running UNIX or Linux	Autonomous client nodes, free in and out, with self-organization	Centralized control, server-oriented with authenticated security	Dynamic resource provisioning of servers, storage, and networks
Applications and Network-centric Services	High-performance computing, search engines, and web services, etc.	Most appealing to business file sharing, content delivery, and social networking	Distributed supercomputing, global problem solving, and data center services	Upgraded web search, utility computing, and outsourced computing services
Representative Operational Systems	Google search engine, SunBlade, IBM Road Runner, Cray XT4, etc.	Gnutella, eMule, BitTorrent, Napster, KaZaA, Skype, JXTA	TeraGrid, GriPhyN, UK EGEE, D-Grid, ChinaGrid, etc.	Google App Engine, IBM Bluecloud, AWS, and Microsoft Azure

Cluster di computer cooperativi

- Integrati e strettamente accoppiati
- Gestione unica

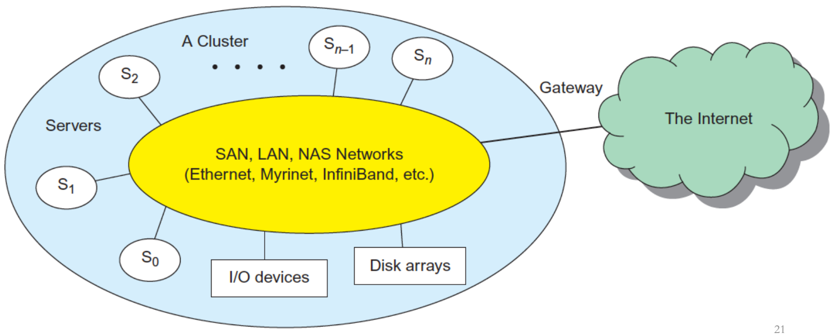
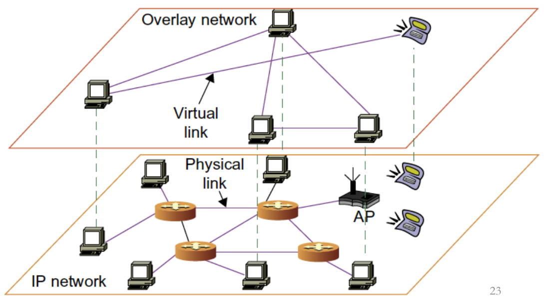


Table 1.2 Classification of Parallel and Distributed Computing Systems

Functionality, Applications	Computer Clusters [10,28,38]	Peer-to-Peer Networks [34,46]	Data/ Computational Grids [6,18,51]	Cloud Platforms [1,9,11,12,30]
Architecture, Network Connectivity, and Size	Network of compute nodes interconnected by SAN, LAN, or WAN hierarchically	Flexible network of client machines logically connected by an overlay network	Heterogeneous clusters interconnected by high-speed network links over selected resource sites	Virtualized cluster of servers over data centers via SLA
Control and Resources Management	Homogeneous nodes with distributed control, running UNIX or Linux	Autonomous client nodes, free in and out, with self-organization	Centralized control, server-oriented with authenticated security	Dynamic resource provisioning of servers, storage, and networks
Applications and Network-centric Services	High-performance computing, search engines, and web services, etc.	Most appealing to business file sharing, content delivery, and social networking	Distributed supercomputing, global problem solving, and data center services	Upgraded web search, utility computing, and outsourced computing services
Representative Operational Systems	Google search engine, SunBlade, IBM Road Runner, Cray XT4, etc.	Gnutella, eMule, BitTorrent, Napster, KaZaA, Skype, JXTA	TeraGrid, GriPhyN, UK EGEE, D-Grid, ChinaGrid, etc.	Google App Engine, IBM Bluecloud, AWS, and Microsoft Azure

Peer 2 Peer

- Rete di sistemi molto debolmente connessi
- Calcolo tra peer, assolutamente collaborativo
- Basato su un overlay network



Cosa si fa con il P2P

Table 1.5 Major Categories of P2P Network Families [46]

System Features	Distributed File Sharing	Collaborative Platform	Distributed P2P Computing	P2P Platform
Attractive Applications	Content distribution of MP3 music, video, open software, etc.	Instant messaging, collaborative design and gaming	Scientific exploration and social networking	Open networks for public resources
Operational Problems	Loose security and serious online copyright violations	Lack of trust, disturbed by spam, privacy, and peer collusion	Security holes, selfish partners, and peer collusion	Lack of standards or protection protocols
Example Systems	Gnutella, Napster, eMule, BitTorrent, Aimster, KaZaA, etc.	ICQ, AIM, Groove, Magi, Multiplayer Games, Skype, etc.	SETI@home, Geonome@home, etc.	JXTA, .NET, FightingAid@home, etc.

Table 1.2 Classification of Parallel and Distributed Computing Systems

Functionality, Applications	Computer Clusters [10,28,38]	Peer-to-Peer Networks [34,46]	Data/ Computational Grids [6,18,51]	Cloud Platforms [1,9,11,12,30]
Architecture, Network Connectivity, and Size	Network of compute nodes interconnected by SAN, LAN, or WAN hierarchically	Flexible network of client machines logically connected by an overlay network	Heterogeneous clusters interconnected by high-speed network links over selected resource sites	Virtualized cluster of servers over data centers via SLA
Control and Resources Management	Homogeneous nodes with distributed control, running UNIX or Linux	Autonomous client nodes, free in and out, with self-organization	Centralized control, server-oriented with authenticated security	Dynamic resource provisioning of servers, storage, and networks
Applications and Network-centric Services	High-performance computing, search engines, and web services, etc.	Most appealing to business file sharing, content delivery, and social networking	Distributed supercomputing, global problem solving, and data center services	Upgraded web search, utility computing, and outsourced computing services
Representative Operational Systems	Google search engine, SunBlade, IBM Road Runner, Cray XT4, etc.	Gnutella, eMule, BitTorrent, Napster, KaZaA, Skype, JXTA	TeraGrid, GriPhyN, UK EGEE, D-Grid, ChinaGrid, etc.	Google App Engine, IBM Bluecloud, AWS, and Microsoft Azure

Utility computing: the grid

- Infrastruttura che connette, computer, software, middleware, strumenti e utenti
- Piattaforme virtuali

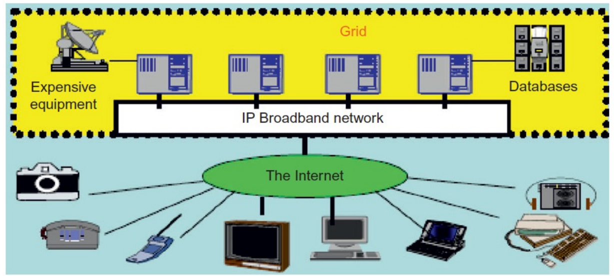


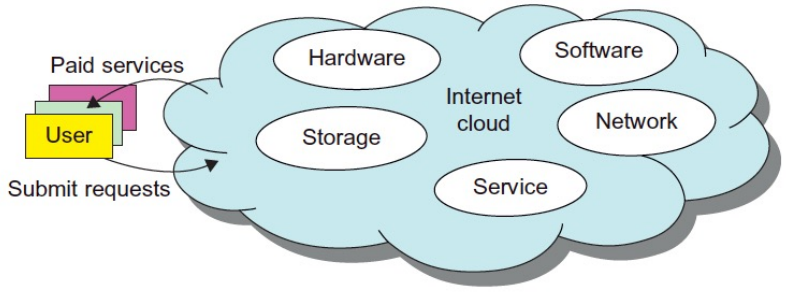
Table 1.2 Classification of Parallel and Distributed Computing Systems

Functionality, Applications	Computer Clusters [10,28,38]	Peer-to-Peer Networks [34,46]	Data/ Computational Grids [6,18,51]	Cloud Platforms [1,9,11,12,30]
Architecture, Network Connectivity, and Size	Network of compute nodes interconnected by SAN, LAN, or WAN hierarchically	Flexible network of client machines logically connected by an overlay network	Heterogeneous clusters interconnected by high-speed network links over selected resource sites	Virtualized cluster of servers over data centers via SLA
Control and Resources Management	Homogeneous nodes with distributed control, running UNIX or Linux	Autonomous client nodes, free in and out, with self-organization	Centralized control, server-oriented with authenticated security	Dynamic resource provisioning of servers, storage, and networks
Applications and Network-centric Services	High-performance computing, search engines, and web services, etc.	Most appealing to business file sharing, content delivery, and social networking	Distributed supercomputing, global problem solving, and data center services	Upgraded web search, utility computing, and outsourced computing services
Representative Operational Systems	Google search engine, SunBlade, IBM Road Runner, Cray XT4, etc.	Gnutella, eMule, BitTorrent, Napster, KaZaA, Skype, JXTA	TeraGrid, GriPhyN, UK EGEE, D-Grid, ChinaGrid, etc.	Google App Engine, IBM Bluecloud, AWS, and Microsoft Azure

Cloud computing

- Cloud computing è realizzato da un pool di risorse offerte da computer virtualizzati
- Una cloud può offrire supporto a workload di tipo diverso
  - Da batch style, backend
  - A workload interattivi e che gestiscono l'utente
- Caratteristiche:
  - Calcolo ridondante
  - Self-recovering
  - Modelli di calcolo (e framework) altamente scalabili
  - Tolleranti ai malfunzionamenti

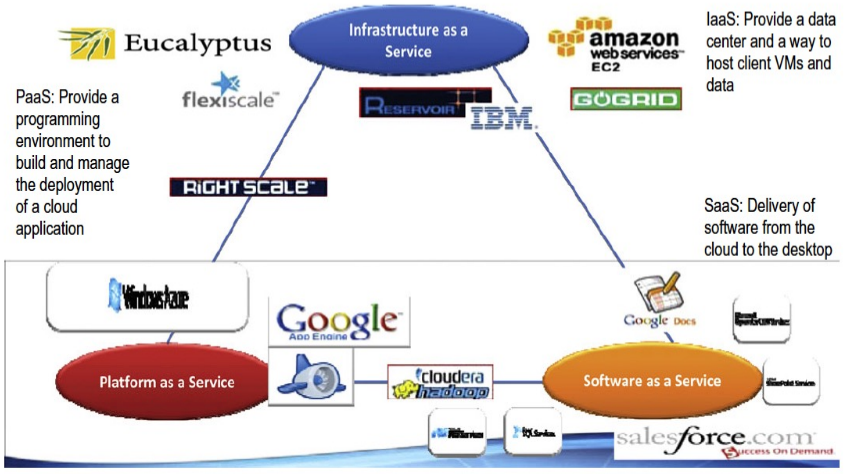
Pool di risorse

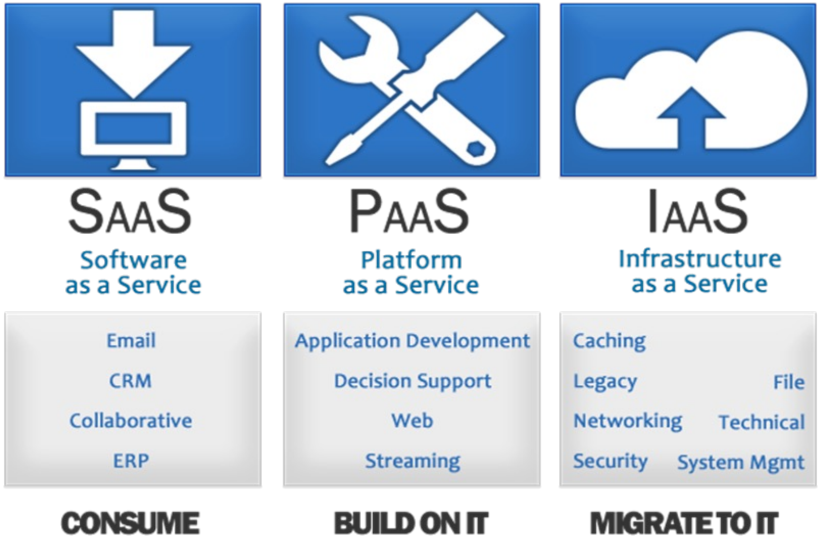


Question time!

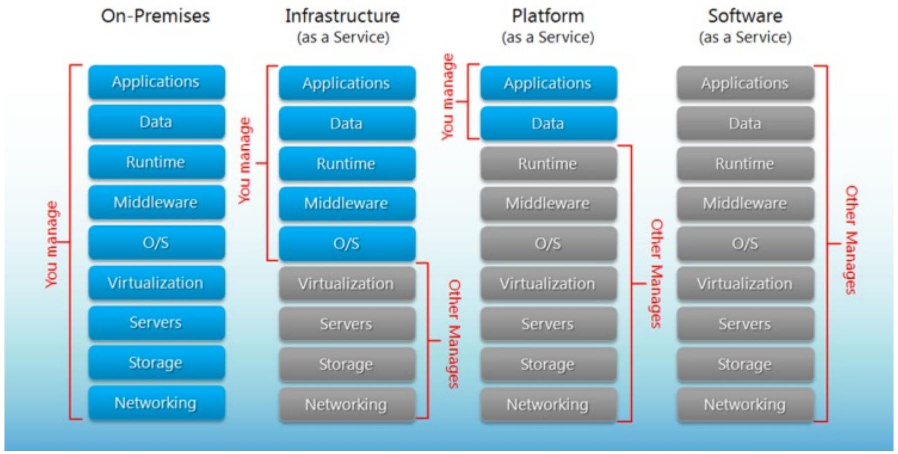
I modelli del cloud: IaaS, Paas, SaaS

- Infrastructure as a Service (IaaS):
  - l'utente noleggia una infrastruttura
- Platform as a Service (PaaS):
  - l'utente noleggia un ambiente per poter sviluppare e eseguire applicazioni su cloud
- Software as a Service (SaaS):
  - l'utente noleggia l'uso di software fornito (via web) dal cloud





Separation of Responsibilities



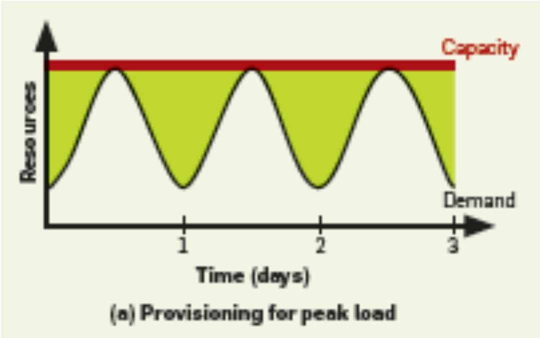
Perché il cloud? 8 ragioni

- Spazi specificamente disegnati per il calcolo con protezione e efficienza energetica
- Condivisione di capacità di calcolo tra numerosi utenti, migliorando la utilizzazione
- Separazione del costo di manutenzione dal costo di sviluppo applicazioni
- Riduzione nel costo di calcolo, notevole rispetto al paradigma tradizionale
- Ambienti di programmazione scalabili (big data)
- Service discovery e content distribution
- Privacy e security
- Modelli di costo e di business ritagliati su necessità (on-demand)

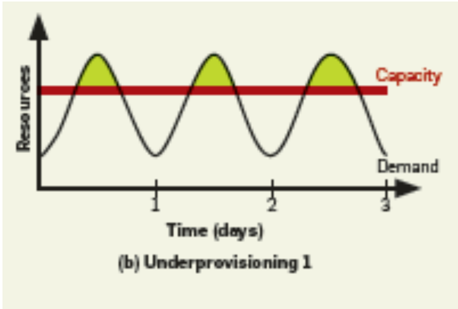
Un approfondimento: Elasticity

- Fornire nuove risorse di calcolo da poter impiegare, in tempo brevissimo (minuti) invece che di settimane
- L'utilizzo delle macchine nei centri di calcolo nel mondo è tra il 5% e il 20%
  - Progettate per picchi di carico (fattore da 2 a 10 volte superiore al carico normale)
- Un esempio: un servizio che richiede 500 server di picco (mezzogiorno) ma 100 a mezzanotte, con un carico normale di 300 server.
- Se paghiamo per poter gestire i picchi, paghiamo 500 x 24h al giorno, sottostimando l'uso delle risorse
  - Normalmente di un fattore stimato intorno a 1.7 volte il costo pay-as-you-go

Esempio di non-elasticità (over- under-provisioning)



Esempio di non-elasticità (over- under-provisioning)



Esempio di non-elasticità (over- under-provisioning)

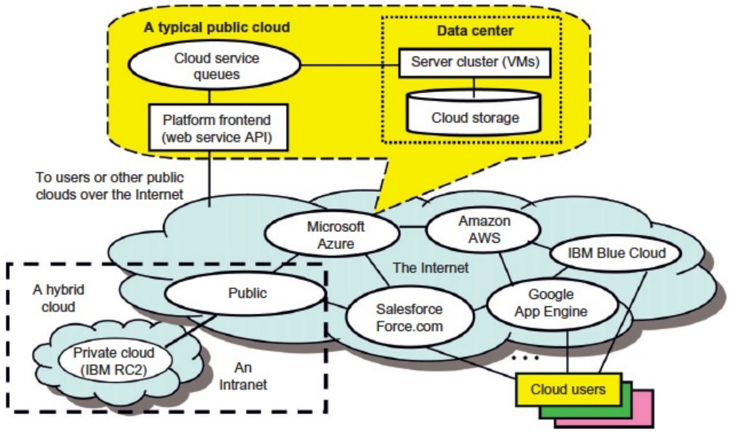
Indice

- La conclusione di un (lungo) viaggio
- Calcolo scalabile su internet
- Modelli per il calcolo distribuito e su cloud
- **Cloud Computing e modelli di servizio**
- Conclusioni

Il Cloud come risorsa per l'economia

- Il Cloud computing porta benefici a tutta la industria dei servizi
- Nuovo paradigma
- Particolare attenzione alla dinamicità della risposta alle necessità (altrettanto dinamiche)
- Nuovi player che entrano in gioco sul panorama mondiale grazie al Cloud
  - Da un lato, Microsoft, Apple, Oracle
  - Dall'altro, Google, Amazon …

Public, Private e Hybrid Cloud



Il modello di utilizzo del Cloud

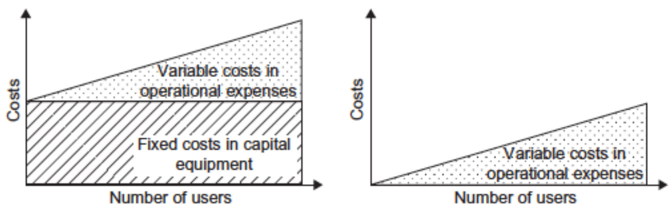
Classical Computing	**Cloud Computing**
(Repeat the following cycle every 18 months)	(Pay as you go per each service provided)
Buy and own	**Subscribe**
Hardware, system software, applications to meet peak needs	- - - -
Install, configure, test, verify, evaluate, manage	**Use** (Save about 80-95% of the total cost)
- - - -	- - - -
Use	(Finally)
- - - -	**$ - Pay for what you use**
Pay $$$$$ (High cost)	based on the QoS

Gli obiettivi progettuali del Cloud

- Spostare la computazione dal desktop e server locali verso data center su Internet
- Fornitura di servizi
  - Service Level Agreement (SLA)
  - Efficienza di calcolo, storage, consumo di corrente
  - Pricing "pay-as-you-go"
- Scalabilità di prestazioni
  - Adatta al mercato dinamico per aziende in crescita velocissima

Modelli di costo

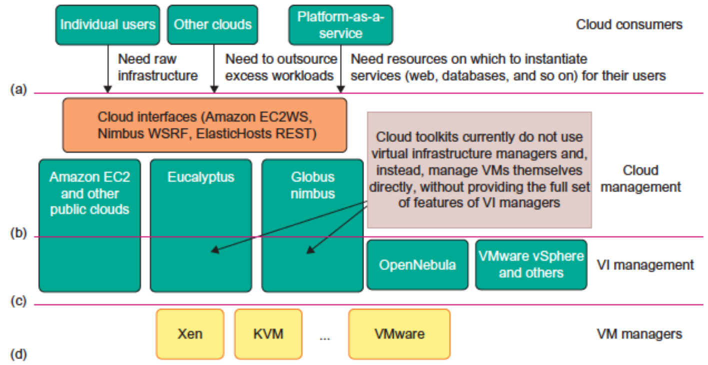
- Sposta da Capital expenses (CAPEX) a Operational Expenses (OPEX)
- CAPEX: costi fissi, investimenti notevoli, immobilizzazione di risorse
  - 1 utente o 100000 non fa (quasi) differenza
- OPEX: costi variabili che dipendono dal numero di utenti



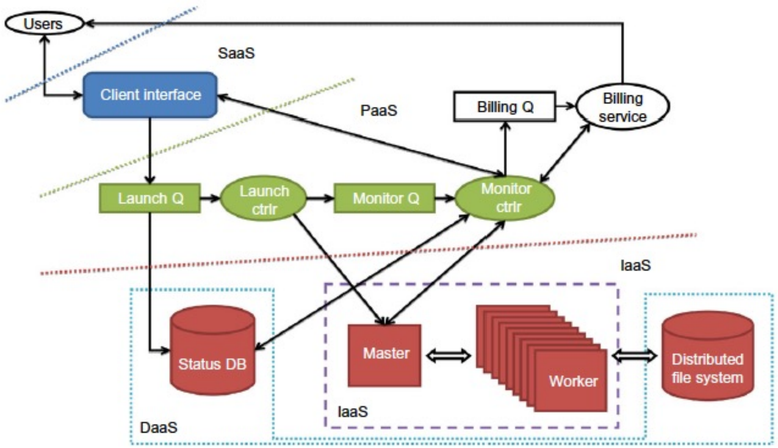
(a) Traditional IT cost model (b) Cloud computing cost model

Ecosystema Cloud

- Un insieme di provider, utenti tecnologie attorno alle cloud pubbliche e private (aziendali)



Modelli di servizio: IaaS, PaaS e SaaS



Esempi di IaaS

Table 4.1 Public Cloud Offerings of IaaS [10,18]

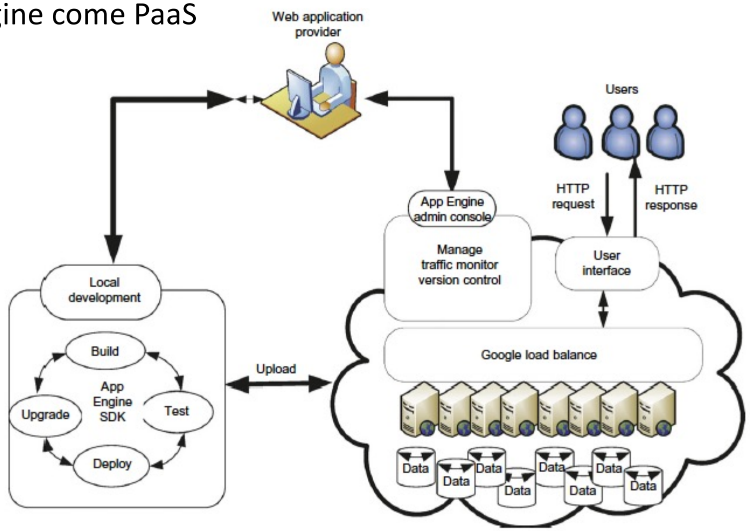
Cloud Name	VM Instance Capacity	API and Access Tools	Hypervisor, Guest OS
Amazon EC2	Each instance has 1–20 EC2 processors, 1.7–15 GB of memory, and 160–1.69 TB of storage.	CLI or web Service (WS) portal	Xen, Linux, Windows
GoGrid	Each instance has 1–6 CPUs, 0.5–8 GB of memory, and 30–480 GB of storage.	REST, Java, PHP, Python, Ruby	Xen, Linux, Windows
Rackspace Cloud	Each instance has a four-core CPU, 0.25–16 GB of memory, and 10–620 GB of storage.	REST, Python, PHP, Java, C#, .NET	Xen, Linux
FlexiScale in the UK	Each instance has 1–4 CPUs, 0.5–16 GB of memory, and 20–270 GB of storage.	web console	Xen, Linux, Windows
Joyent Cloud	Each instance has up to eight CPUs, 0.25–32 GB of memory, and 30–480 GB of storage.	No specific API, SSH, Virtual/Min	OS-level virtualization, OpenSolaris

Esempi di PaaS

Table 4.2 Five Public Cloud Offerings of PaaS [10,18]

Cloud Name	Languages and Developer Tools	Programming Models Supported by Provider	Target Applications and Storage Option
Google App Engine	Python, Java, and Eclipse-based IDE	MapReduce, web programming on demand	Web applications and BigTable storage
Salesforce.com's Force.com	Apex, Eclipse-based IDE, web-based Wizard	Workflow, Excel-like formula, Web programming on demand	Business applications such as CRM
Microsoft Azure	.NET, Azure tools for MS Visual Studio	Unrestricted model	Enterprise and web applications
Amazon Elastic MapReduce	Hive, Pig, Cascading, Java, Ruby, Perl, Python, PHP, R, C++	MapReduce	Data processing and e-commerce
Aneka	.NET, stand-alone SDK	Threads, task, MapReduce	.NET enterprise applications, HPC

Google App Engine come PaaS



Esempi di SaaS

- Software applicativo utilizzato via browser HTML
- Alcuni esempi:
  - Google Gmail, Docs, Photos, etc.
  - Microsoft Office365
  - Customer Relationship Management software da Salesforce.com

Conclusioni

- La conclusione di un (lungo) viaggio
- Calcolo scalabile su internet
- Modelli per il calcolo distribuito e su cloud
- Cloud Computing e modelli di servizio
- **Conclusioni**

Question time!