Distributed Computer Systems

Cloud Computing – Introduction

Dr. Taieb Znati
Computer Science Department
University of Pittsburgh

Fall 2016

Thoughts on Cloud Computing

"A way to increase capacity or add capabilities on the fly without investing in new infrastructure, training new personnel, or licensing new software." "The idea of loosely coupled services running on an agile, scalable infrastructure should eventually make every enterprise a node in the cloud."

Galen Gruman, InfoWorld Executive Editor, and Eric Knorr, InfoWorld Editor in Chief

Thoughts on Cloud Computing

"It's a trap"

"It's worse than stupidity: it's marketing hype. Somebody is saying this is inevitable—and whenever you hear that, it's very likely to be a set of businesses campaigning to make it true."

Richard Stallman, Founder, Free Software Foundation (The Guardian, Sept. 29, 2008)

Thoughts on Cloud Computing

It's nothing new"

"...we've redefined Cloud Computing to include everything that we already do... I don't understand what we would do differently ... other than change the wording of some of our ads."

Larry Ellison, CEO, Oracle (Wall Street Journal, Sept. 26, 2008)

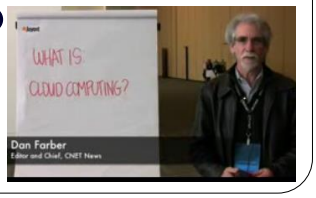
Thoughts on Cloud Computing

"I think it is one of the foundations of the next generation of computing"
"The network of networks is the platform for all computing"

Tim O'Reilly, CEO
O'Reilly Media

Thoughts on Cloud Computing

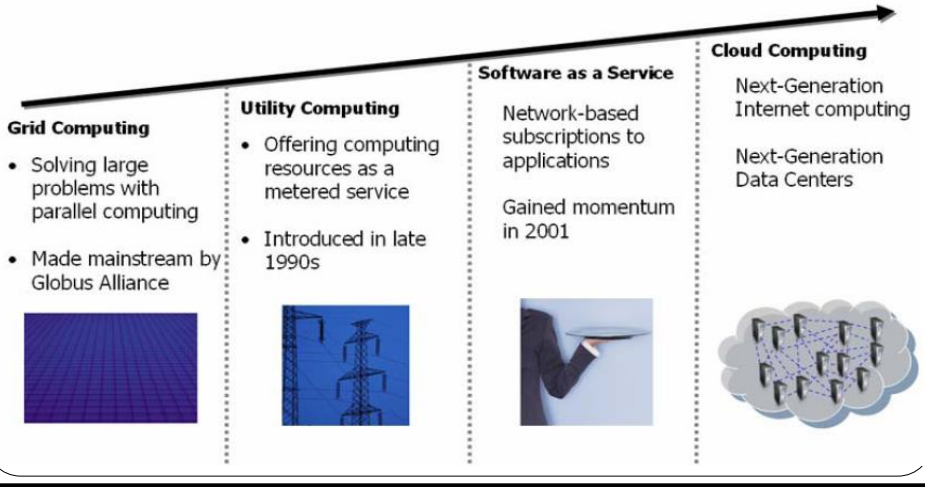
"We are at the beginning of the age of planetary computing. Billions of people will be wirelessly interconnected, and the only way to achieve that kind of massive scale usage is by massive scale, **brutally efficient** cloud-based infrastructure."
Dan Farber, Editor in Chief CNET News



What is Cloud Computing – A Definition

- A model for enabling convenient, on-demand network access to a shared pool of configurable computing and networking resources that can be rapidly provisioned and released with minimal management effort or service provider interaction
  - It provides the means through which all resources – from power to computing infrastructure, applications, business processes to personal collaboration – can be delivered as a service anywhere and anytime.

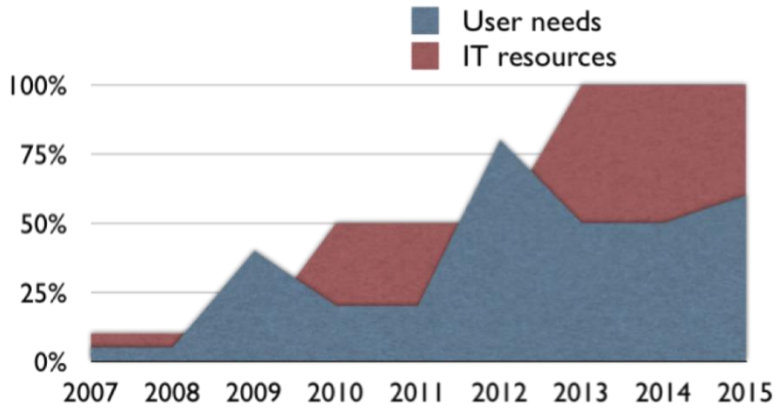
Evolution to Cloud Computing



Elasticity and Scalability

- It defines the ability to scale up and down
- The cloud service provider cannot anticipate when or how customers will use the service
  - Customers may use the cloud during their peak operations and rely on their own infrastructure the rest of the time
  - Customers my use the cloud as a primary development platform for their applications
    - Availability must be anytime anywhere
    - Cloud must provide the ability to scale to accommodate new users and new resources on the fly

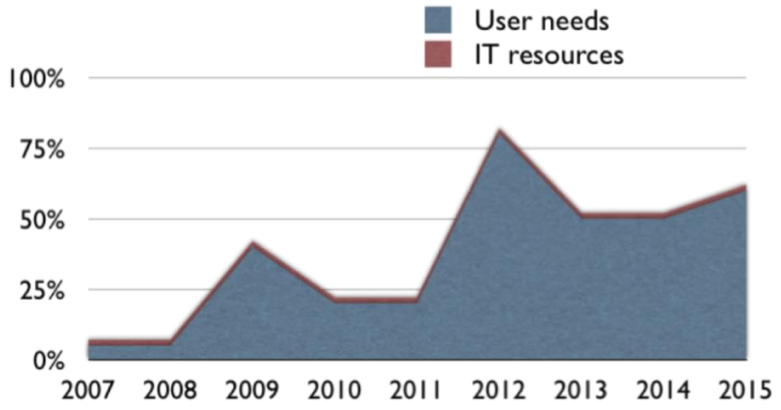
Conventional Computing Infrastructure – Resource Demands



Self-Service Provisioning and Deprovisionning

- Ability of a customer to acquire and relinquish cloud service without going through a lengthy process
  - Conventional data-center process typically goes through multiple steps from decision makers to procurement and configuration before operations
    - The process can take considerably amount of time, depending on organization's policy

Cloud Computing Infrastructure – Resource Demands



Cloud Computing Enablers and Motivators

Approaches to Virtualization

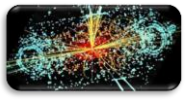
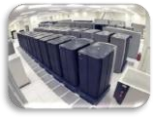
- Full virtualization
  - Sensitive instructions (discovered statically or dynamically at run-time) are replaced by binary translation or trapped by hardware into VMM for SW emulation
  - Any OS software can run in the VM
    - Used by IBM's CP/CMS, Oracle (Sun) VirtualBox, VMware Workstation
- Hardware-assisted virtualization (IBM S/370, Intel VT, or AMD-V)
  - CPU traps sensitive instructions – runs unmodified guest OS
  - Examples: VMware Workstation, Linux Xen, Linux KVM, Microsoft Hyper-V
- Para-virtualization
  - Presents SW interface to virtual machines similar to but not identical to that of the underlying HW, requiring guest operating systems to be adapted
  - Examples: early versions of Xen
- Operating System virtualization
  - Operating system kernel allows for multiple isolated user-space instances, instead of just one
  - Instances look and feel like a real server
  - Used in Solaris Zones, QEMU, BSD Jails, OpenVZ

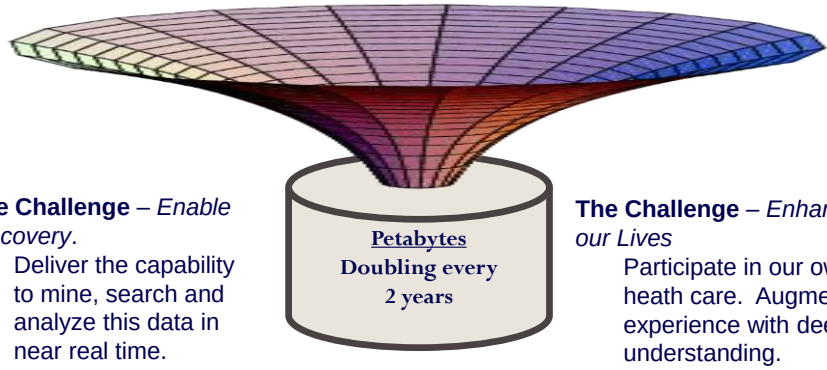
Modern OS Virtualization

- **Hardware-assisted virtualization is a key technological enabler for Cloud Computing**
  - Provides complete isolation on low-cost commodity platforms
  - Enables multiplexing of many users onto single server
- Key contribution is minimal performance overhead (few percent) versus non-virtualized
  - However, high I/O applications incur many VM traps (high CPU overhead), limiting scalability and efficiency
- True performance isolation for multiple applications remains a serious challenge
  - A multi-faceted problem

Data Explosion – From Data To Knowledge

Experiments | Simulations | Archives | Literature | Instruments

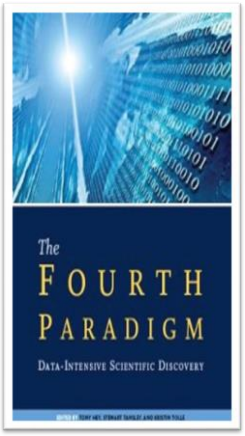



The Challenge – *Enable Discovery*.
Deliver the capability to mine, search and analyze this data in near real time.

The Challenge – *Enhance our Lives*
Participate in our own heath care. Augment experience with deeper understanding.

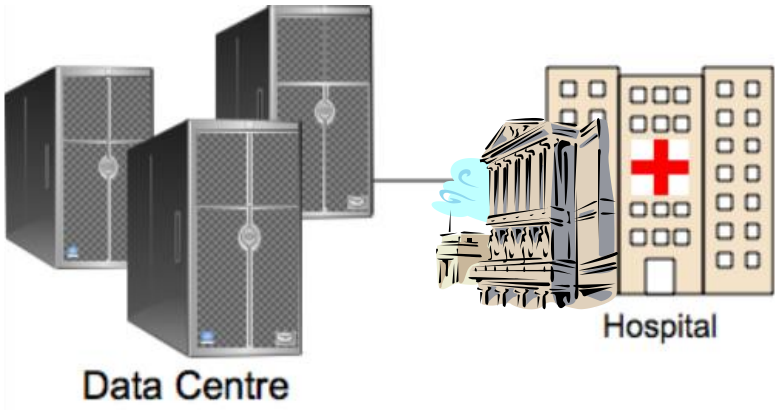
Changing Nature of Discovery

- Complex models
  - Multidisciplinary interactions
  - Wide temporal and spatial scales
- Large multidisciplinary data
  - Real-time steams
  - Structured and unstructured
- Distributed communities
  - Virtual organizations
  - Socialization and management
- Diverse expectations
  - Client-centric and infrastructure-centric

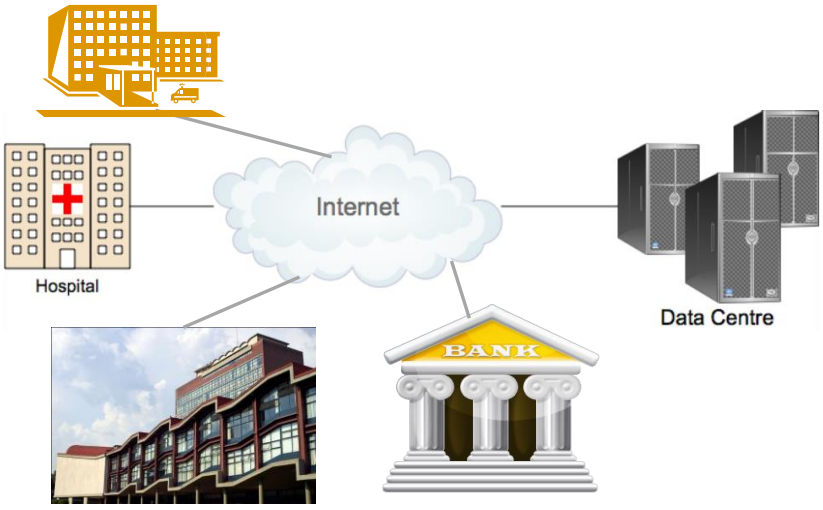


Cloud Computing Types

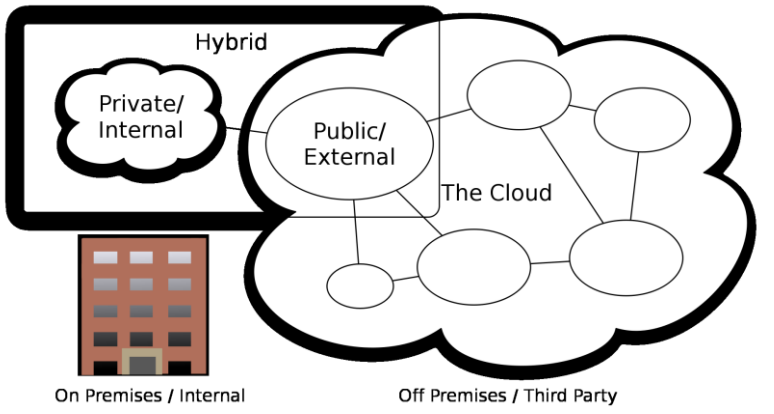
Private Cloud



Public Cloud



Hybrid Cloud

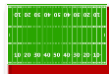


Cloud Physical Architecture

Clouds are built on Data Centers

- Range in size from "edge" facilities to megascale.
- Economies of scale
  - Approximate costs for a small size center (1000 servers) and a larger, 100K server center.

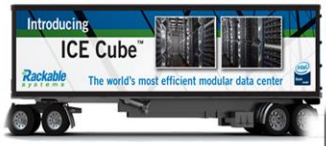
Technology	Cost in small-sized Data Center	Cost in Large Data Center	Ratio
Network	$95 per Mbps/ month	$13 perMbps/ month	7.1
Storage	$2.20 per GB/ month	$0.40 per GB/ month	5.7
Administration	~140 servers/ Administrator	>1000 Servers/ Administrator	7.1

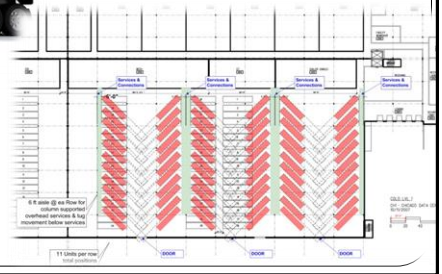


Each data center is 11.5 times the size of a football field

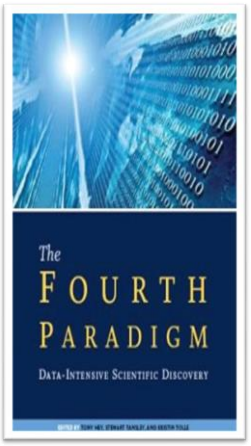
Advances in DC Deployment

- Conquering complexity.
  - Building racks of servers & complex cooling systems all separately is not efficient.
  - Package and deploy into bigger units





Cloud Data Architectures

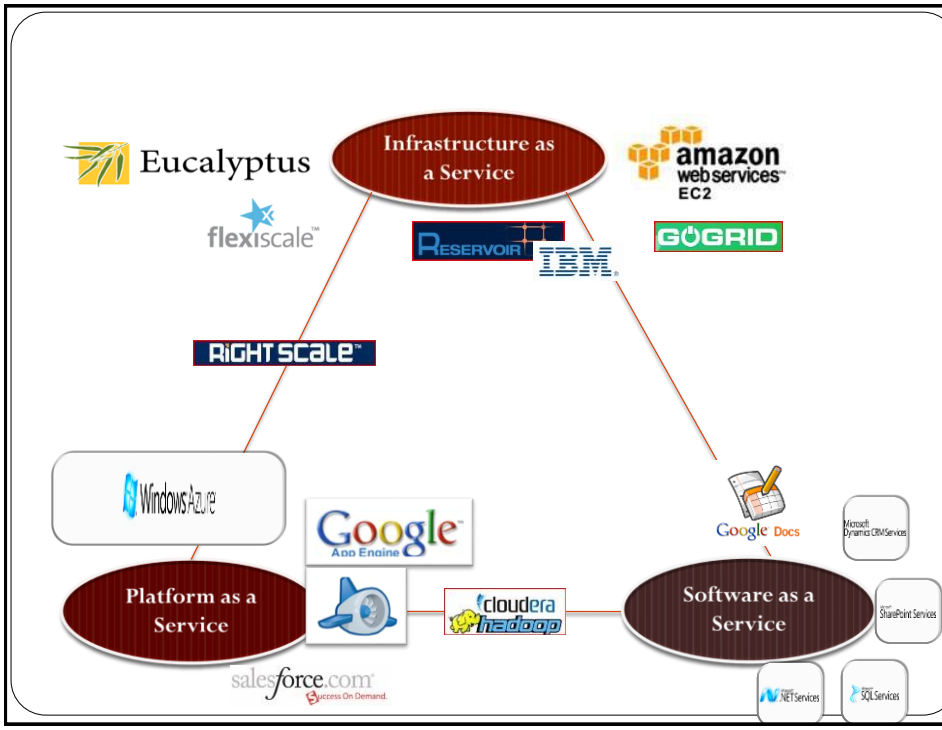


- A close integration of data with computation
  - "Move the computation to the data" – *Jim Gray*
  - Data is stored on server disks
  - Optimized more for reads than writes
- Data replication
  - Multiple copies of each data object
  - Copies are distributed
- Unstructured data
  - "Blob" storage- basic metadata + binary object
  - Streaming data from instruments
- Structured data
  - Tables – billions of rows and columns
    - Table partitioned into blocks of rows and blocks are distributed and replicated.
  - Databases – replicated relational databases

Cloud Software Models

The Cloud Landscape

- Infrastructure as a Service (IaaS)
  - IaaS offers storage and computing resources that developers and IT organization use to deliver custom business solutions
  - Platform as a Service (PaaS)
    - PaaS offers a development and programming environment to build cloud application s
    - The cloud deploys and manages the application for the client
- Software as a Service (SaaS)
  - Delivery of software and purpose-built applications from the cloud to the desktop

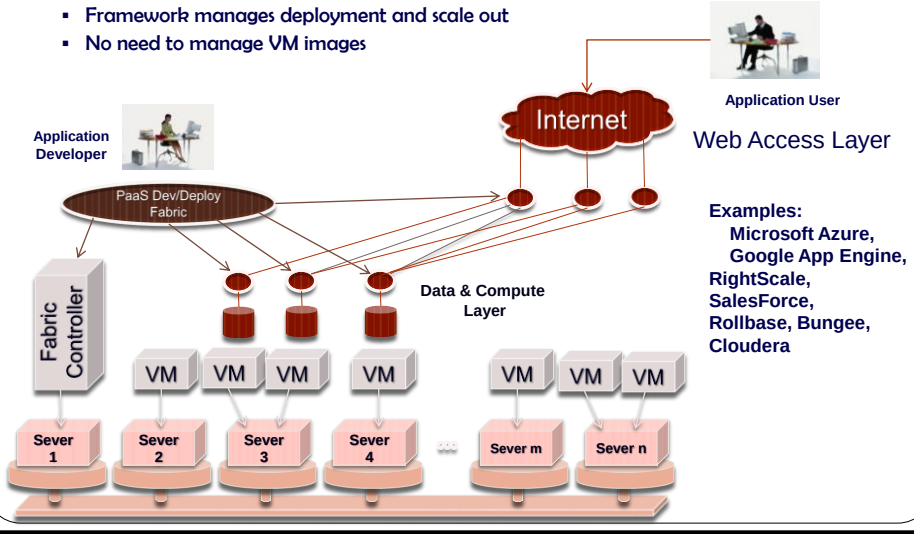


IaaS Characteristics

- IaaS mostly delivers computer hardware – servers, networking technology, storage and data center space, power, etc
  - In some cases, operating systems and virtualization technology can be included
- IaaS is paid for on a "pay-as-you-go" basis
  - The service level agreement (SLA) may include provision for dynamic scaling to accommodate raising customer's needs
    - Scaling up or scaling down
  - SLA may specify a level of availability, i.e., 99.999%
- Amazon Elastic Computing Platform is an example

PaaS Characteristics

- An application development, deployment and management fabric.
  - User programs web service front end and computational & Data Services
  - Framework manages deployment and scale out
  - No need to manage VM images



Examples:
Microsoft Azure, Google App Engine, RightScale, SalesForce, Rollbase, Bungee, Cloudera

PaaS Characteristics

- Paas goes beyond infrastructure and delivers a solution stack – an integrated set of software and development tools typically required for the development of applications
  - Both for software development and for runtime support
- PaaS provides lifecycle management
  - Capabilities needed to manage all software development stages from planning and design, to building and deployment, to testing and maintenance

PaaS Characteristics

- PaaS is inherently multi-tenants and typically support Web service standards
- PaaS delivers dynamic scaling of software – up and down – as well as the need to separate concerns of access and data security to customers
- A major drawback of PaaS is service and data "lock-in"
  - Customers may be locked in to the use of a particular software development and management platform – development tools and libraries, data format, etc.
    - Emergence of Open Platform as a Service (OPaaS)

SaaS Characteristics

- SaaS has its roots in early Application Service Providers (ASP) models
  - Hosting of supply chain applications customer relationship management (CRM)
    - Email hosting ASPs
- CRM is one of the most common categories of SaaS
  - Salesforce.com is a prominent CRM SaaS

SaaS Modes

- Simple Multi-tenancy
  - Each user has its own resources segregated from those of other customers
    - Inefficient, particulalry in a large-scale cloud
- Fine grain multi-tenancy
  - Segregates among users from a software engineering perspective
    - Customer data and access capabilities are segregated within the application
    - Hardware resource are typically shared among users

Massively Scaled SaaS

- This service mode is beneficial in environments where a very large number of users perform exactly the same operation – executing the same applications on their own data
  - Facebbok, eBay, Skype, Google Apps are designed for massive scaling
- Massively Scaled SaaS has the possibility to achieve dramatic economies of scale
  - Cost efficiency results from reducing per-unit costs when more of the same entity is produced or same workload is processed

Conclusion

- Computing On the Cloud
  - The next stage in evolution of the Internet?
- Types of Cloud Computing
- Cloud Computing Service Models
  - IaaS
  - PaaS
  - SaaS
- Economy of Scale
  - Efficient way to provide and acquire the service