KHAI THÁC DỊCH VỤ ĐIỆN TOÁN ĐÁM MÂY

Nguyen Hong Son,PhD

MỤC TIÊU MÔN HỌC

Cung cấp kiến thức cơ bản về nguyên lý của điện toán đám mây và làm quen với khai thác các dịch vụ điện toán đám mây.

Kiến thức

-Phân biệt được các thành phần thiết yếu trong hệ thống hệ thống điện toán đám mây

-Phân biệt được các mô hình dịch vụ điện toán đám mây gồm SaaS, PaaS, IaaS và các xu hướng mở rộng.

Kỹ năng

-Khai thác được các nền tảng đám mây khác nhau như Amazon EC2, Google AppEngine, Windows Azure, IBM Blumix

TÀI LIỆU THAM KHẢO

[1] Rajkumar Buyya, James Broberg, Andrzej M. Goscinski, Cloud Computing: Principles and Paradigms, Wiley Publishing, 2011. (Lý thuyết)

[2] Các user guides từ nhà cung cấp dịch vụ đám mây như Amazon EC2, Microsoft Azure...(Thực hành)

INTRODUCTION TO CLOUD COMPUTING

Ideas

- From providing electric power to cloud computing
- Cluster computing, grid computing: allowing access to large amounts of computing power in a fully virtualized manner, by aggregating resources and offering a single system view
- Main goal: delivering computing as a utility
- Utility computing describes a business model for on-demand delivery of computing power; consumers pay providers based on usage ("pay-as-you-go").

What is “cloud computing”?

“Cloud is a parallel and distributed computing system consisting of a collection of inter-connected and virtualised computers that are dynamically provisioned and presented as one or more unified computing resources based on service-level agreements (SLA) established through negotiation between the service provider and consumers.”

Buyya et al

What is “cloud computing”?

“Clouds are a large pool of easily usable and accessible virtualized resources (such as hardware, development platforms and/or services). These resources can be dynamically reconfigured to adjust to a variable load (scale), allowing also for an optimum resource utilization. This pool of resources is typically exploited by a pay-per-use model in which guarantees are offered by the Infrastructure Provider by means of customized Service Level Agreements.”

Vaquero et al

“Clouds are hardware-based services offering compute, network, and storage capacity where: Hardware management is highly abstracted from the buyer, buyers incur infrastructure costs as variable OPEX, and infrastructure capacity is highly Elastic.”

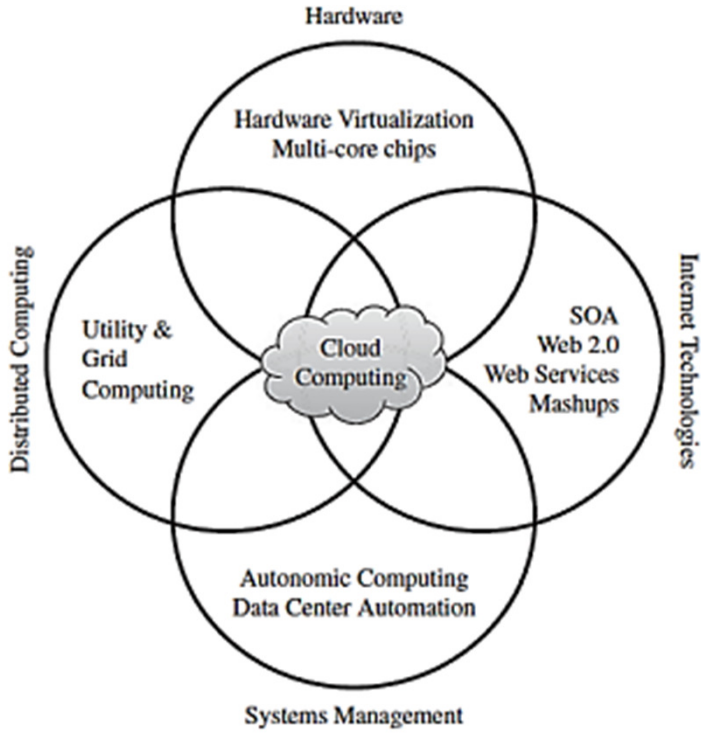
McKinsey and Co

The NIST Definition of Cloud Computing

- NIST: National Institute of Standards and Technology (USA)
- "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."

Roots of Cloud Computing

- Technologies,
  - Hardware (virtualization, multi-core chips),
  - Internet technologies (Web services, service-oriented architectures, Web 2.0),
  - Distributed computing (clusters, grids),
  - Systems management (autonomic computing, data center automation)



From Mainframes to Clouds

- 1970s time-shared mainframe as utilities
- Expensive
- The advent of fast and inexpensive microprocessors, IT data centers moved to collection commodity servers→the mainframe era collapsed
- The advent of increasingly fast fiber-optics networks and new technologies for enabling sharing of computing power over great distances→the potential of delivering computing services with the speed and reliability

SOA, Web Services, Web 2.0, and Mashups

- Web services can glue together applications running on different messaging product platforms, enabling information from one application to be made available to others, and enabling internal applications to be made available over the Internet
- In a SOA, software resources are packaged as "services," which are well-defined, self-contained modules that provide standard business functionality and are independent of the state or context of other services
- In Web 2.0, the consumer Web, information and services may be programmatically aggregated, acting as building blocks of complex compositions, called service ***mashups***

➔***Software as a Service (SaaS) domain***

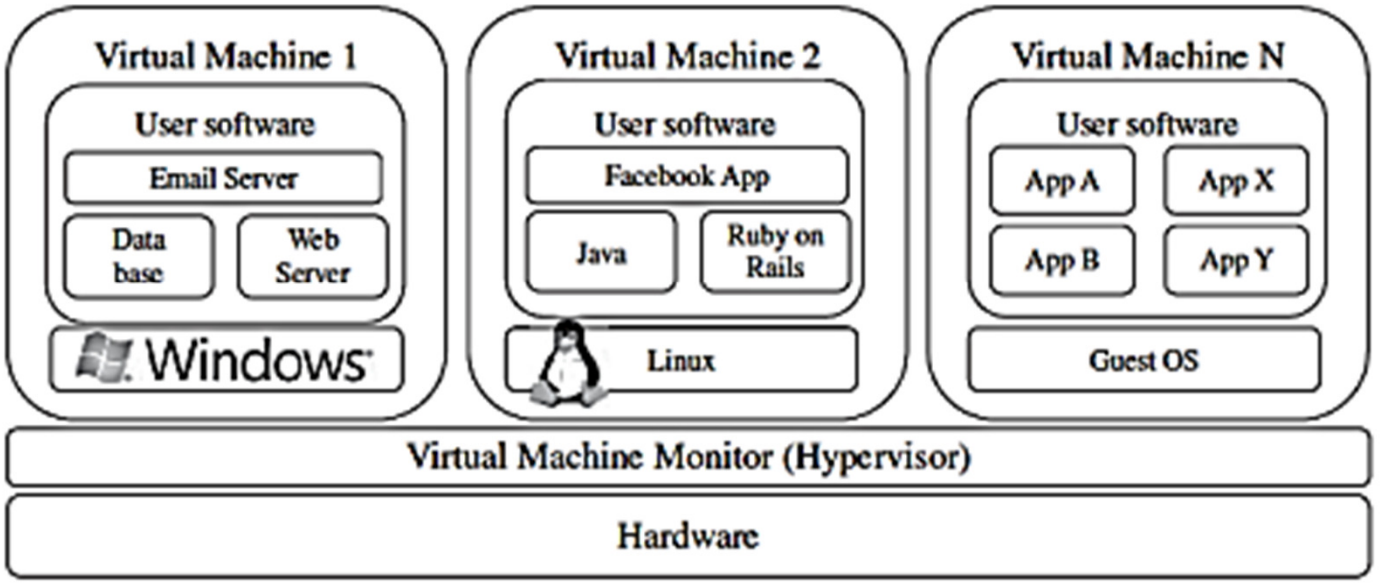
Grid Computing

- Grid computing enables aggregation of distributed resources and transparently access to them
- The development of standardized protocols for several grid computing activities has contributed, theoretically, to allow delivery of on-demand computing services over the Internet

Utility Computing

- In utility computing environments, users assign a “utility” value to their jobs.
- Utility is a fixed or time-varying valuation that captures various QoS constraints
- The valuation is the amount they are willing to pay a service provider to satisfy their demands.

Hardware Virtualization



Management of workload in a virtualized system

Basic capabilities regarding management of workload in a virtualized system:

- Isolation: all program instructions are fully confined inside a VM
- Consolidation: The consolidation of several individual and heterogeneous workloads onto a single physical platform leads to better system utilization.
- Migration: facilitating hardware maintenance, load balancing, and disaster recovery.

VMM platforms

- VMWare ESXi: It is a bare-metal hypervisor, meaning that it installs directly on the physical server
- Xen: has served as a base to other virtualization products, both commercial and open-source.
- KVM (kernel-based virtual machine): a Linux virtualization subsystem

Virtual Appliances

- An application combined with the environment needed to run it (operating system, libraries, compilers, databases, application containers, and so forth) is referred to as a "virtual appliance."
- An appliance is shaped as a VM disk image associated with hardware requirements, and it can be readily deployed in a hypervisor.
- The VMWare virtual appliance marketplace allows users to deploy appliances on VMWare hypervisors or on partners public clouds
- Amazon allows developers to share specialized Amazon Machine Images (AMI) and monetize their usage on Amazon EC2

Open Virtualization Format

- Each Hypervisor supports a different VM image format and the formats are incompatible with one another
- In order to facilitate packing and distribution of software to be run on VMs several vendors, including VMware, IBM, Citrix, Cisco, Microsoft, Dell, and HP, have devised the Open Virtualization Format (OVF)
- An OVF package consists of a file, or set of files, describing the VM hardware characteristics (e.g.,memory, network cards, and disks), operating system details, startup, and shutdown actions, the virtual disks themselves, and other metadata containing product and licensing information

Autonomic Computing

- Autonomic systems rely on monitoring probes and gauges (sensors), on an adaptation engine (autonomic manager) for computing optimizations based on monitoring data, and on effectors to carry out changes on the system.
- The large data centers of cloud computing need autonomic computing → inspire software technologies for data center automation, perform tasks such as:
  - ✓ management of service levels of running applications;
  - ✓ management of data center capacity;
  - ✓ proactive disaster recovery;
  - ✓ automation of VM provisioning

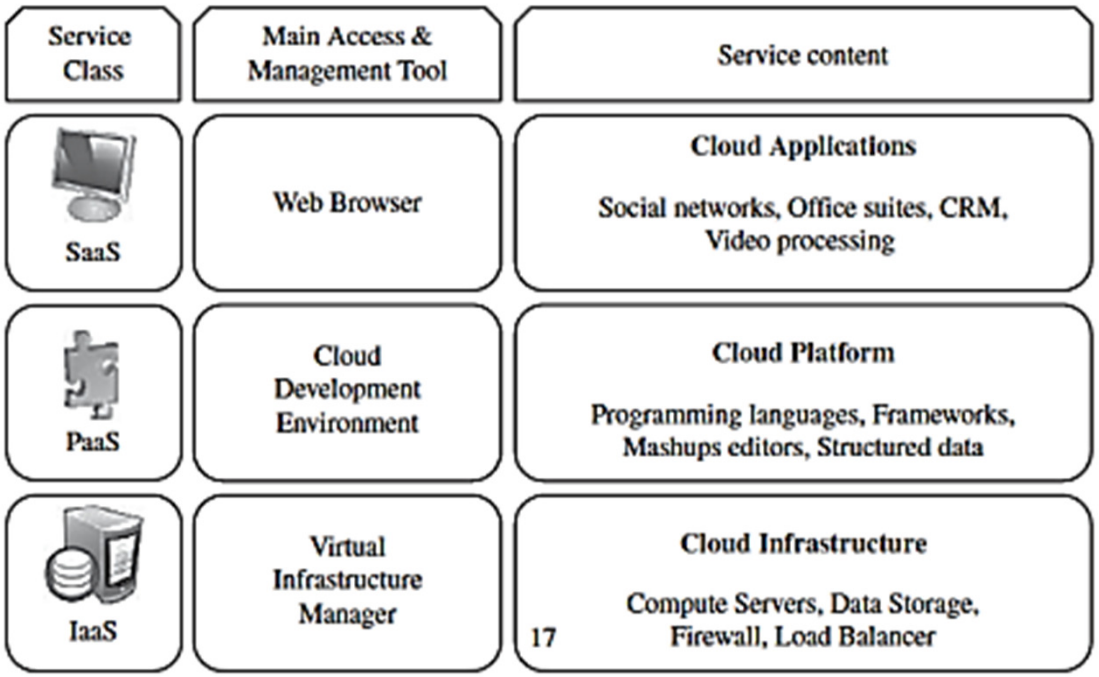
Layers and Types of Clouds

- Cloud computing services are divided into three classes:

  (1) Infrastructure as a Service,

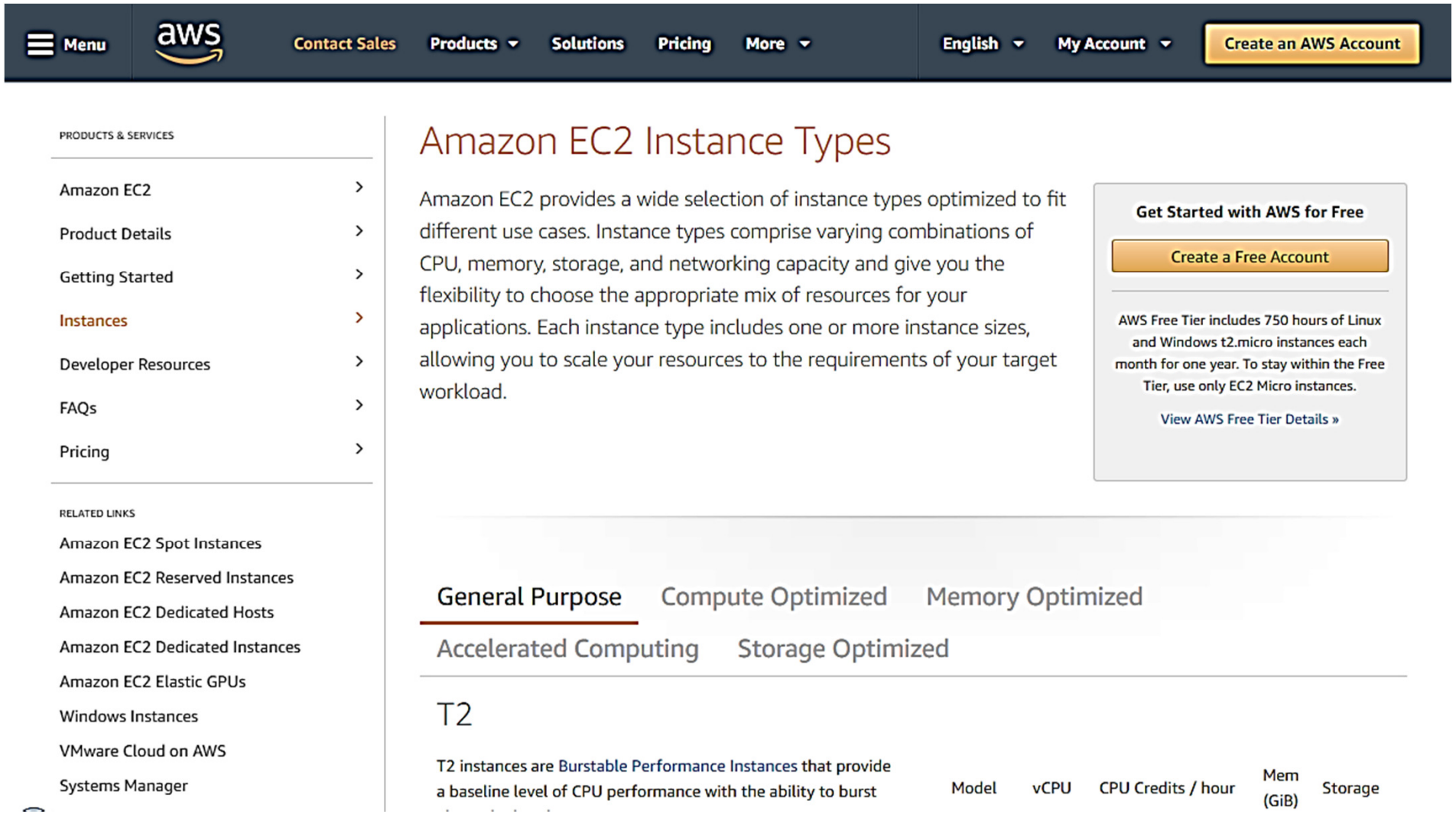
  (2) Platform as a Service,

  (3) Software as a Service



Infrastructure as a Service

- Offering virtualized resources (computation, storage, and communication) on demand
- Amazon Web Service mainly offers IaaS. EC2 service means offering VMs.
- OpenStack is a free and open-source software platform for cloud computing, mostly deployed as infrastructure-as-a-service (IaaS)


Menu
aws
Contact Sales
Products
Solutions
Pricing
More
English
My Account
Create an AWS Account
PRODUCTS & SERVICES
Amazon EC2
Product Details
Getting Started
Instances
Developer Resources
FAQs
Pricing
RELATED LINKS
Amazon EC2 Spot Instances
Amazon EC2 Reserved Instances
Amazon EC2 Dedicated Hosts
Amazon EC2 Dedicated Instances
Amazon EC2 Elastic GPUs
Windows Instances
VMware Cloud on AWS
Systems Manager
Amazon EC2 Instance Types
Amazon EC2 provides a wide selection of instance types optimized to fit different use cases. Instance types comprise varying combinations of CPU, memory, storage, and networking capacity and give you the flexibility to choose the appropriate mix of resources for your applications. Each instance type includes one or more instance sizes, allowing you to scale your resources to the requirements of your target workload.
Get Started with AWS for Free
Create a Free Account
AWS Free Tier includes 750 hours of Linux and Windows t2.micro instances each month for one year. To stay within the Free Tier, use only EC2 Micro instances.
View AWS Free Tier Details »
General Purpose
Compute Optimized
Memory Optimized
Accelerated Computing
Storage Optimized
T2
T2 instances are Burstable Performance Instances that provide a baseline level of CPU performance with the ability to burst
Model
vCPU
CPU Credits / hour
Mem (GiB)
Storage

Platform as a Service

- Offering a high level of abstraction to make a cloud easily programmable.
- Google AppEngine offers a scalable evironment for developing and hosting Web applications

Google Cloud

Tìm kiếm

BẢNG ĐIỀU KHIỂN

Why Google | Products | Solutions | Launcher | Pricing | Security | Customers | Documentation | Supp

TRY IT FREE

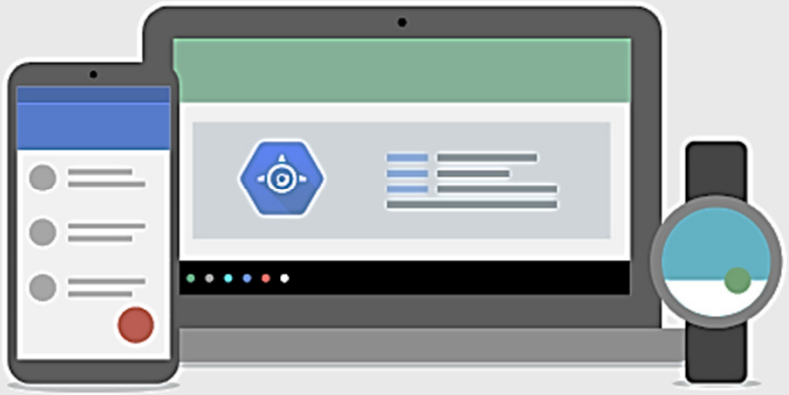
GOOGLE APP ENGINE

Build scalable web and mobile backends in any language on Google's infrastructure

TRY IT FREE

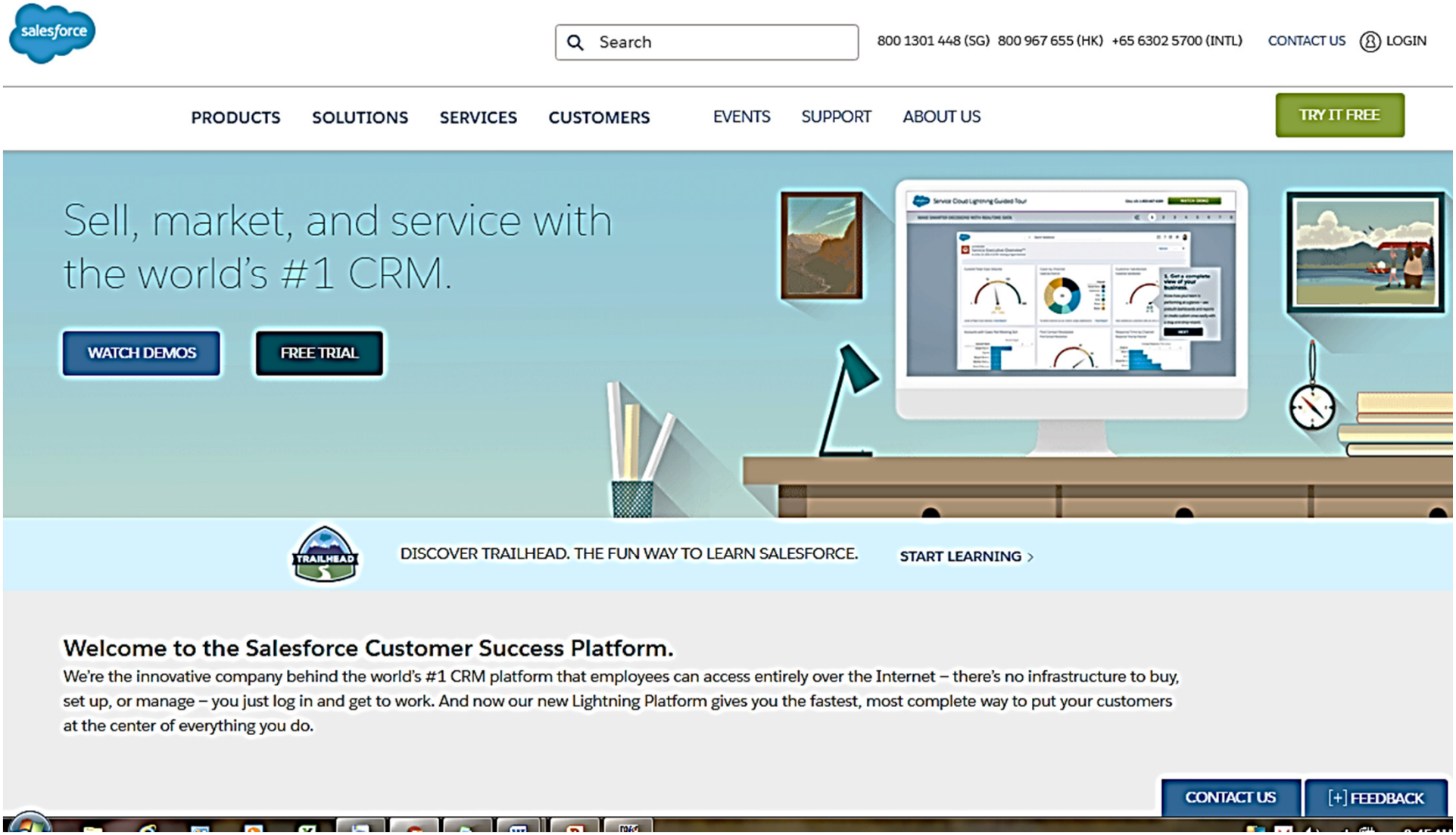
App Engine for All

Build modern web and mobile applications on an open cloud platform: bring your own language runtimes, frameworks, and third party libraries. Google App Engine is a fully managed platform that completely abstracts away infrastructure so you focus only on code. Go from zero to planet-scale and see why some of today's most successful companies power their applications on App Engine.



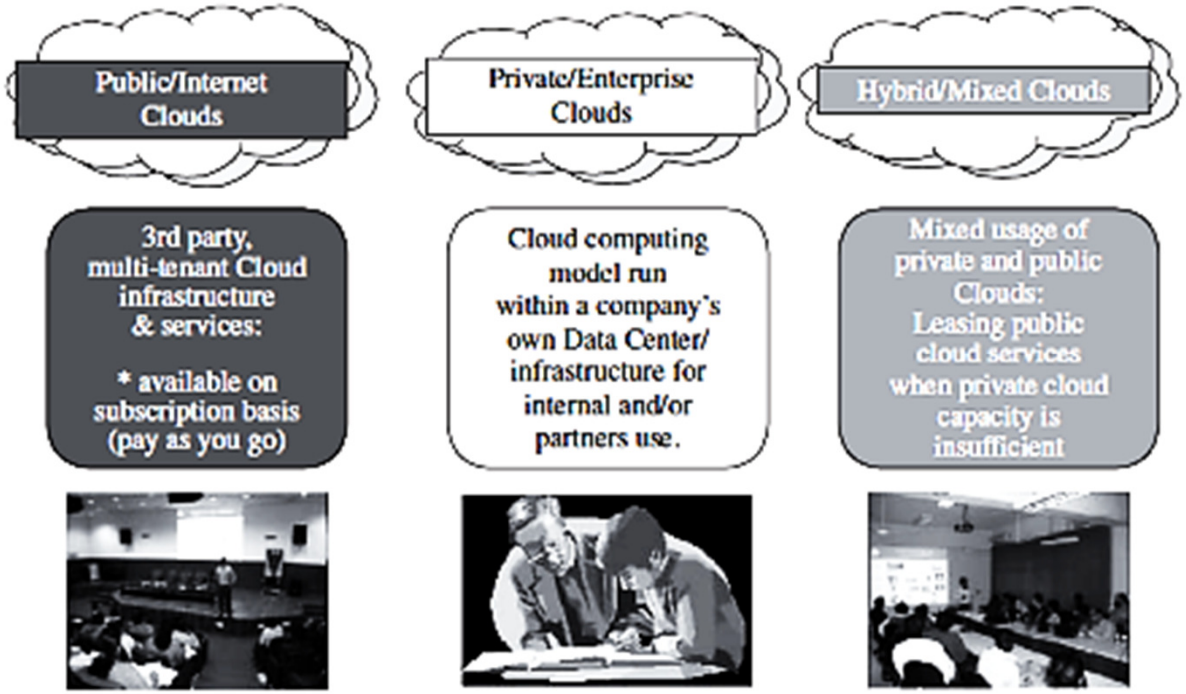
Software as a Service

- Applications reside on the top of the cloud stack.
- End users can access services of the layer through Web portal
- Salesforce.com offers CRM applications that reside completely on their servers, allowing customers to customize and access application on demand


salesforce
Search
800 1301 448 (SG) 800 967 655 (HK) +65 6302 5700 (INTL)
CONTACT US
LOGIN
PRODUCTS
SOLUTIONS
SERVICES
CUSTOMERS
EVENTS
SUPPORT
ABOUT US
TRY IT FREE
Sell, market, and service with the world's #1 CRM.
WATCH DEMOS
FREE TRIAL
TRAILHEAD
DISCOVER TRAILHEAD. THE FUN WAY TO LEARN SALESFORCE.
START LEARNING
Welcome to the Salesforce Customer Success Platform.
We're the innovative company behind the world's #1 CRM platform that employees can access entirely over the Internet – there's no infrastructure to buy, set up, or manage – you just log in and get to work. And now our new Lightning Platform gives you the fastest, most complete way to put your customers at the center of everything you do.
CONTACT US
[+] FEEDBACK

Deployment Models

- Public cloud as a “cloud made available in a pay-as-you-go manner to the general public”
- Private cloud as “internal data center of a business or other organization, not made available to the general public.”
- A community cloud is “shared by several organizations and supports a specific community that has shared concerns (e.g., mission, security requirements, policy, and compliance considerations).”
- A hybrid cloud takes shape when a private cloud is supplemented with computing capacity from public clouds



DESIRED FEATURES OF A CLOUD

- Self-Service
- Per-Usage Metering and Billing
- Elasticity
- Customization

CLOUD INFRASTRUCTURE MANAGEMENT

- The software toolkit responsible for orchestration of resources is called a virtual infrastructure manager (VIM)
- "Cloud Operating System" = VIM
- Two categories of tools used to manage clouds:
  - ✓The cloud toolkits—includes those that "expose a remote and secure interface for creating, controlling and monitoring virtualize resources," .
  - ✓The virtual infrastructure managers—provide advanced features such as automatic load balancing and server consolidation, but do not expose remote cloud-like interfaces.

Features of VIM

- Virtualization Support
- Self-Service, On-Demand Resource Provisioning
- Multiple Backend Hypervisors
- Storage Virtualization
- Interface to Public Clouds
- Virtual Networking
- Dynamic Resource Allocation
- Virtual Clusters.
- Reservation and Negotiation Mechanism
- High Availability and Data Recovery

The most popular VIMs

- Apache VCL
- AppLogic
- Citrix Essentials.
- Enomaly ECP
- Eucalyptus
- Nimbus3
- OpenNebula
- OpenPEX
- oVirt
- Platform ISF
- VMWare vSphere and vCloud

INFRASTRUCTURE as a SERVICE PROVIDERS

- Public Infrastructure as a Service providers commonly offer virtual servers containing one or more CPUs, running several choices of operating systems and a customized software stack.
- In addition, storage space and communication facilities are often provided.

Features of IaaS Providers

- Geographic distribution of data centers;
- Variety of user interfaces and APIs to access the system;
- Specialized components and services that aid particular applications (e.g., load balancers, firewalls);
- Choice of virtualization platform and operating systems;
- Different billing methods and period (e.g., prepaid vs. post-paid, hourly vs. monthly).

The most popular public IaaS clouds

- Amazon Web Services.
- Flexiscale
- Joyent
- GoGrid
- Rackspace Cloud Servers

PLATFORM as a SERVICE PROVIDERS

- Public Platform as a Service providers commonly offer a development and deployment environment that allow users to create and run their applications with little or no concern to low-level details of the platform.
- Specific programming languages and frameworks are made available in the platform, as well as other services such as persistent data storage and in memory caches.

Features of PaaS Providers

- Programming Models, Languages, and Frameworks:
  - Programming models define how users can express their applications using higher levels of abstraction and efficiently run them on the cloud platform.
  - Each model aims at efficiently solving a particular problem.
  - The most common activities that require specialized models are: processing of large dataset in clusters of computers (MapReduce model), development of request-based Web services and applications; definition and orchestration of business processes in the form of work flows (Workflow model); and high-performance distributed execution of various computational tasks.
- Persistence Options:
  - Allowing applications to record their state and recover it in case of crashes, as well as to store user data.
  - Amazon SimpleDB and Google AppEngine datastore

The most popular public PaaS clouds

- Aneka platform
- App Engine
- Microsoft Azure
- Force.com
- Heroku

HEROKU Products Marketplace Pricing Documentation Support More Log in or Sign up

MOVE FAST

Unleash your inner startup

Choose Heroku for the same reasons disruptive startups do: it's the best platform for building with modern architectures, innovating quickly, and scaling precisely to meet demand.

SIGN UP FOR FREE

Explore Heroku Customers

CHALLENGES AND RISKS

- Issues to be faced include:
  - user privacy,
  - data security,
  - data lock-in,
  - availability of service,
  - disaster recovery,
  - performance,
  - scalability,
  - energy-efficiency,
  - programmability

Security, Privacy, and Trust

- How to make cloud computing environments as secure as in-house IT systems
- There is a massive use of third-party services and infrastructures that are used to host important data or to perform critical operations
- The trust toward providers is fundamental to ensure the desired level of privacy for applications hosted in the cloud

Data Lock-In and Standardization

- Data locked-in by a certain provider
- Users may want to move data and applications out from a provider that does not meet their requirements
- Cloud computing infrastructures and platforms do not employ standard methods of storing user data and applications→ do not interoperate and user data are not portable
- Cloud Computing Interoperability Forum (CCIF) was formed by organizations such as Intel, Sun, and Cisco
- Unified Cloud Interface (UCI) by CCIF aims at creating a standard programmatic point of access to an entire cloud infrastructure.
- Open Virtual Format (OVF) aims at facilitating packing and distribution of software to be run on VMs so that virtual appliances can be made portable

Availability, Fault-Tolerance, and Disaster Recovery

- Availability of the service
- Overall performance,
- Measures are to be taken when something goes wrong in the system
- SLAs, which include QoS requirements, must be ideally set up between customers and cloud computing providers to act as warranty

Resource Management and Energy-Efficiency

- The efficient management of virtualized resource pools
- Data centers consumer large amounts of electricity.
- According to a data published by HP [4], 100 server racks can consume 1.3 MW of power and another 1.3 MW are required by the cooling system, thus costing USD 2.6 million per year
- Physical resources such as CPU cores, memory, disk space, and network bandwidth must be sliced and shared among virtual machines running potentially heterogeneous workloads.
- Scheduling, load balancing, VM migration.

Programmability

- The ability to control and change the functions of IT infrastructure
- The existing functions of the infrastructure are configured, changes are made, and existing services are used as part of normal operations
- Using software rather than box-by-box manual configuration

The End