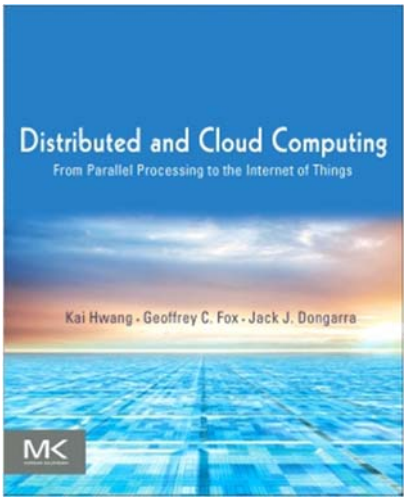
Cloud Computing Overview

Tien-Fu Chen

Dept. of Computer Science
National Chiao Tung Univ.

Reading

- **"Distributed and Cloud Computing,"** by K. Hwang, G. Fox and J. Dongarra, 2011.

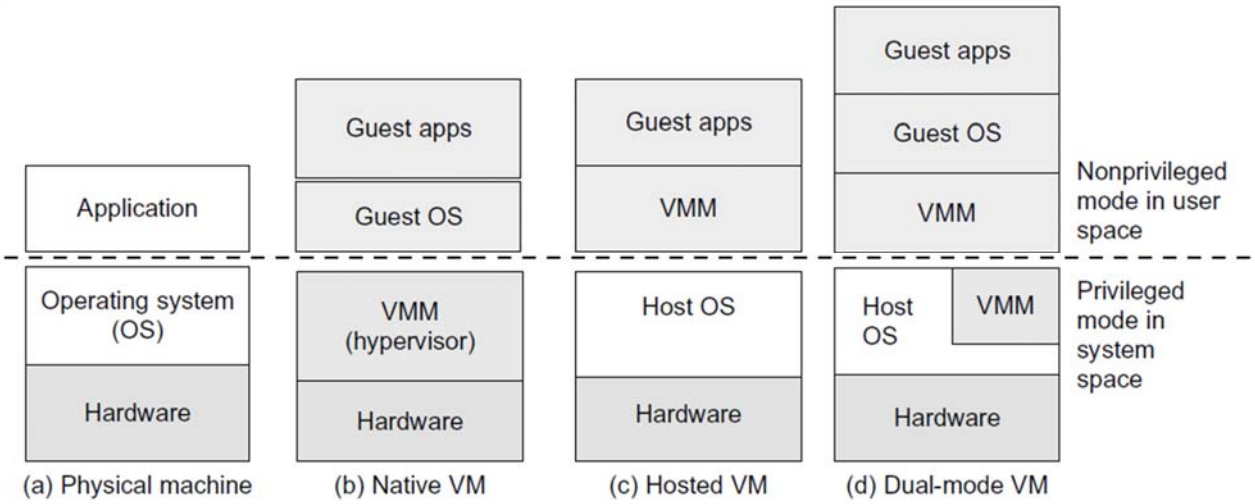


Types of Virtualization

(b) Bare-metal VM:

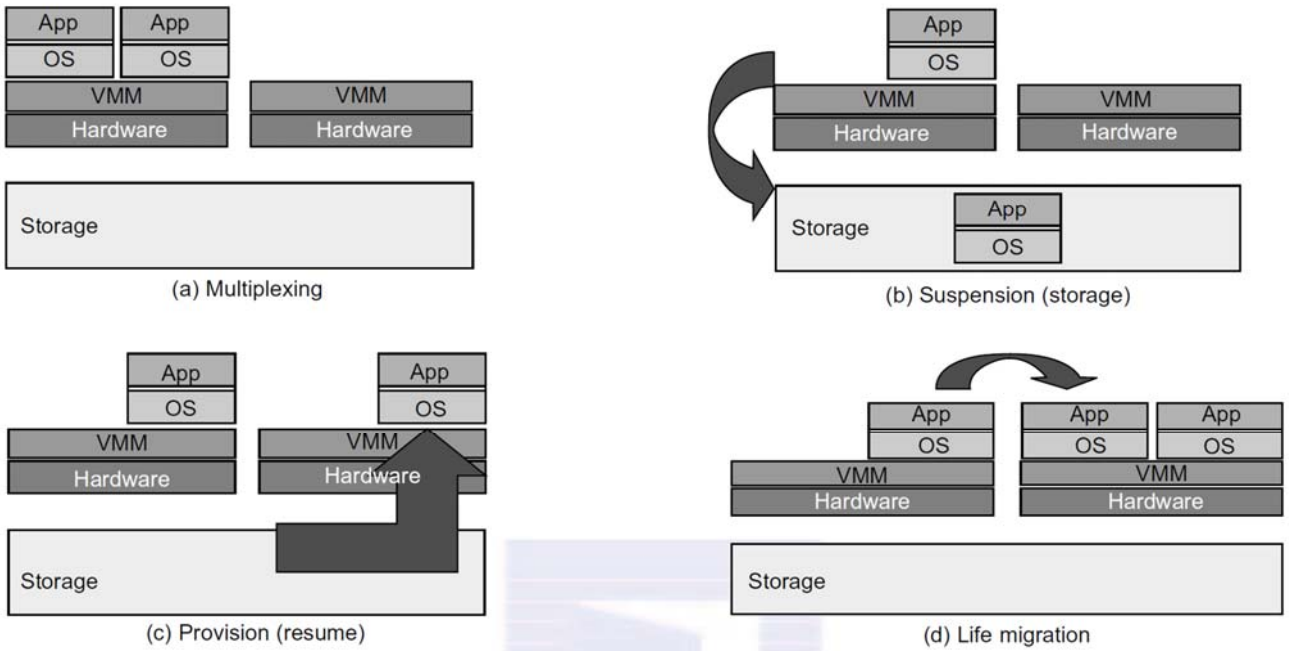
(c). VMM runs in nonprivileged mode. The host OS need not be modified.

(d) Dual mode: VMM runs at the user level and another part runs at the supervisor level. In this case, the host OS may have to be modified.



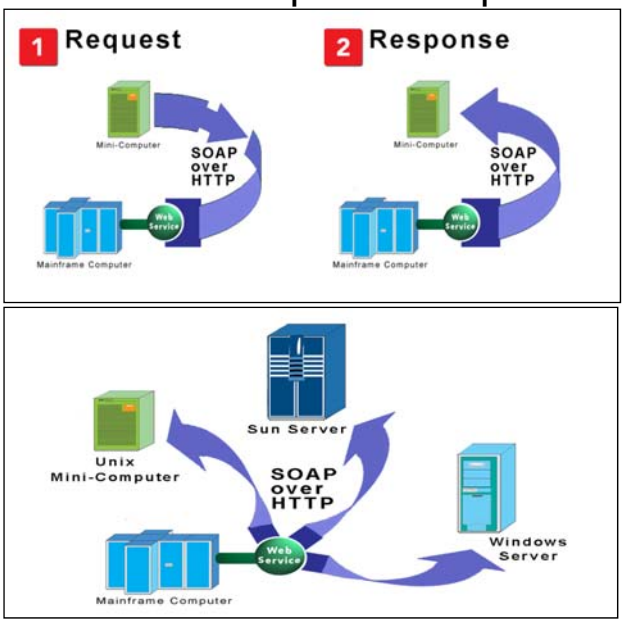
VM primitives

- VM multiplexing, suspension, provision, and migration in a distributed computing environment



Service Oriented Architecture (SOA)

- Allow standard ways for clients and servers to exchange data structure over the network
- An Ideal of Open Interoperability (Loose Coupling)



SOA – A Definition

- ***An IT architecture composed of software that has been exposed as "Services"*** – i.e. invoked on demand using a **standard** protocol.
- **"Web Services"** – software available as a "**service**" using **Internet** protocols.
- **One software application talking to another using a standards-based** (i.e. **non-proprietary**) language over a standards-based communication protocol.
- An **IT architecture** that enables "**loose coupling**" of applications

Cloud Computing

- **Cloud computing** – "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."*

Came about really as business model to allow business to out-source their IT software to a third party Cloud provider.

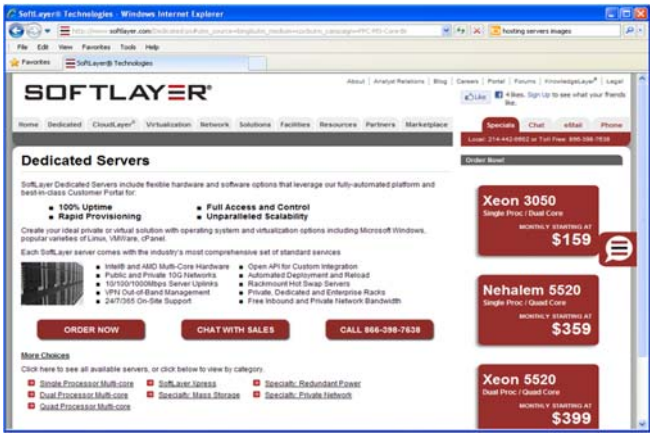
Driven by economics, the Internet and existence of large server farms.

The word "cloud" comes from drawing cloud shapes to represent a network.

* Wikipedia: http://en.wikipedia.org/wiki/Cloud_computing

Server Hosting
Renting Remote Servers

- Around for many years and predates cloud computing – 1990s (?) to present
- Companies provide servers through Internet that users can rent time on.
- Typically done to host web sites.
- Get whole server for your use (dedicated server).
- Pay a monthly fee.
- Generally you load whatever software you want.
- Company only responsible for hardware, OS.
- Still exists although many companies have moved into cloud computing also



Some key aspects of cloud computing

1. Computing resources available on demand, thereby eliminating the need to plan far ahead for provisioning.
2. Elimination of up-front commitment by Cloud users, thereby allowing companies to start small and increase hardware resources only when there is an increase in their needs.
3. Ability to pay for use of computing resources on a short-term basis as needed (e.g., processors by the hour and storage by the day) and release them as needed, thereby rewarding conservation by letting machines and storage go when they are no longer useful.*

* "Above the Clouds: A Berkeley View of Cloud Computing," University of California at Berkeley Technical Report No. UCB/EECS-2009-28.
http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.html

Difference between renting physical servers remotely and cloud computing

- In cloud computing, you get a virtual machine running on servers with your selected OS running on top of virtualization software.
- There could be other users on servers.
- You access servers through a web service/web site.
- You pay for specific time used on processor, storage devices and bandwidth/network.
- Cloud computing focuses on virtualization and service orient approach and making it economical fro companies to use a third party cloud provider to maintain hardware and software on a on-demand basis.

Relationship to Grid computing

- Grid computing – using geographically distributed computing resources collaboratively began as a concept in the mid 1990's with the growth of high speed networks and the Internet.
- Began in the 1990s as a research concept to provide collaborative computing
- The word "grid" came from the idea that grid computing would provide computing power on demand through the Internet in the same way as electrical power come from a distributed electrical Grid utility.
- Cost of usage was not a driving force and usually no costs charged.

Grid and Cloud Computing

- Both Grid computing and Cloud computing take advantage of the Internet.
- One angle of Grid computing was “utility computing” from the original “grid” term.
- Some companies, notably IBM, saw commercial possibilities in the early 2000’s – “on-demand computing” but it did not take off then commercially.

Utility computing resources

Utility computing suggested by John McCarthy in 1960s:

“computation may someday be organized as a public utility." (Wikipedia).

Grid took it up idea in on-demand computing

Cloud computing followed through with:

1. Maturing of virtualization and service-oriented technologies
2. The growth of large underutilized data centers.

Cloud Computing

- IT resources provided as a service
  - Compute, storage, databases, queues
- Clouds leverage economies of scale of commodity hardware
  - Cheap storage, high bandwidth networks & multicore processors
  - Geographically distributed data centers
- Offerings from Microsoft, Amazon, Google

What is the purpose?

- Cloud computing enables companies and applications, which are system infrastructure dependent, to be infrastructure-less.
- By using the Cloud infrastructure on “pay as used and on demand”, all of us can save in capital and operational investment!
- Clients can:
  - Put their data on the platform instead of on their own desktop PCs and/or on their own servers.
  - They can put their applications on the cloud and use the servers within the cloud to do processing and data manipulations etc.

Benefits

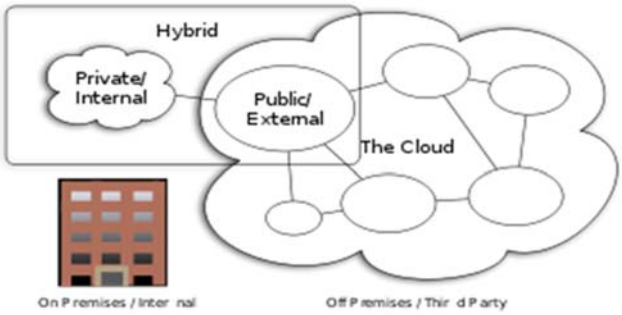
- Cost & management
  - Economies of scale, “out-sourced” resource management
- Reduced Time to deployment
  - Ease of assembly, works “out of the box”
- Scaling
  - On demand provisioning, co-locate data and compute
- Reliability
  - Massive, redundant, shared resources
- Sustainability
  - Hardware not owned

Cloud-Sourcing

- Why is it becoming a Big Deal:
  - Using high-scale/low-cost providers,
  - Any time/place access via web browser,
  - Rapid scalability; incremental cost and load sharing,
  - Can forget need to focus on local IT.
- Concerns:
  - Performance, reliability, and SLAs,
  - Control of data, and service parameters,
  - Application features and choices,
  - Interaction between Cloud providers,
  - No standard API – mix of SOAP and REST!
  - Privacy, security, compliance, trust…

Types of Cloud Computing

- **Public Cloud**: Computing infrastructure is hosted at the vendor's premises.
- **Private Cloud**: Computing architecture is dedicated to the customer and is not shared with other organisations.
- **Hybrid Cloud**: Organisations host some critical, secure applications in private clouds. The not so critical applications are hosted in the public cloud
  - **Cloud bursting**: the organisation uses its own infrastructure for normal usage, but cloud is used for peak loads.
- **Community Cloud**

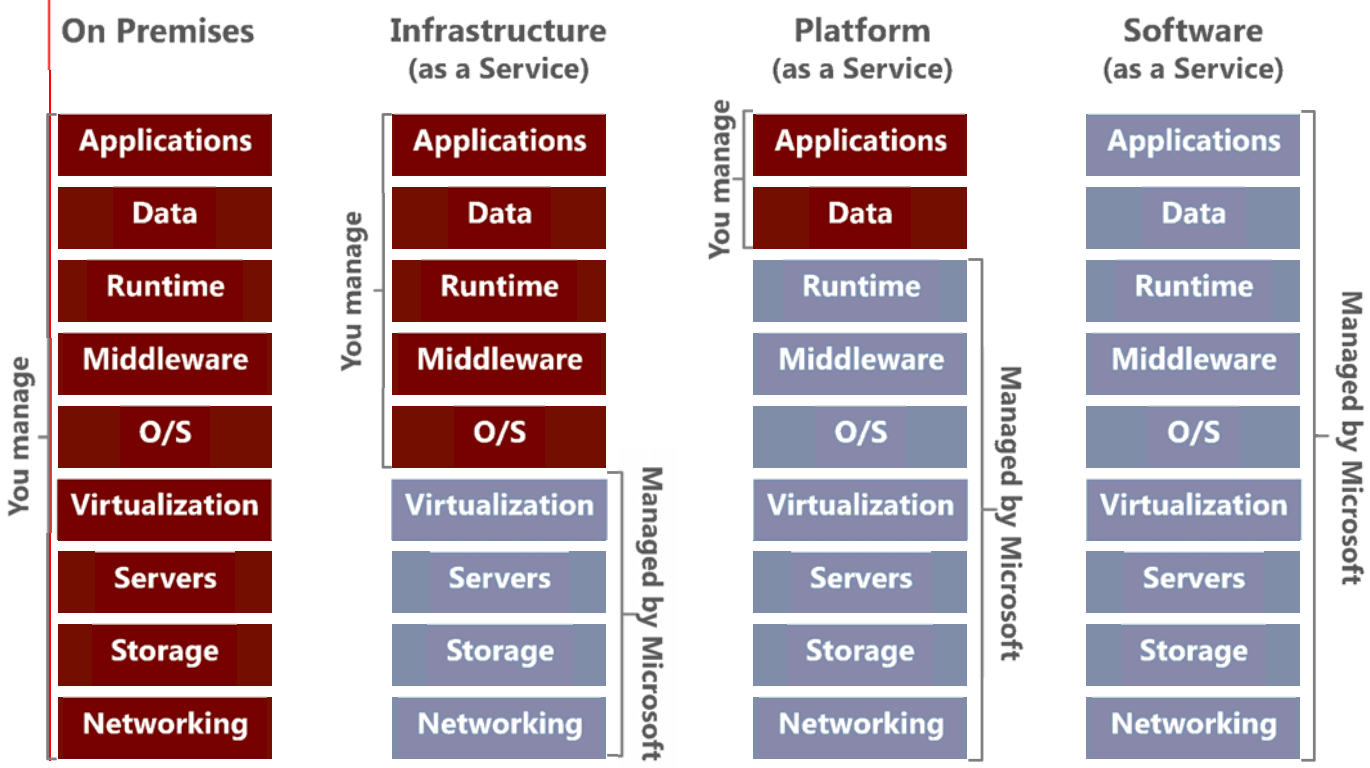


Cloud Computing Types

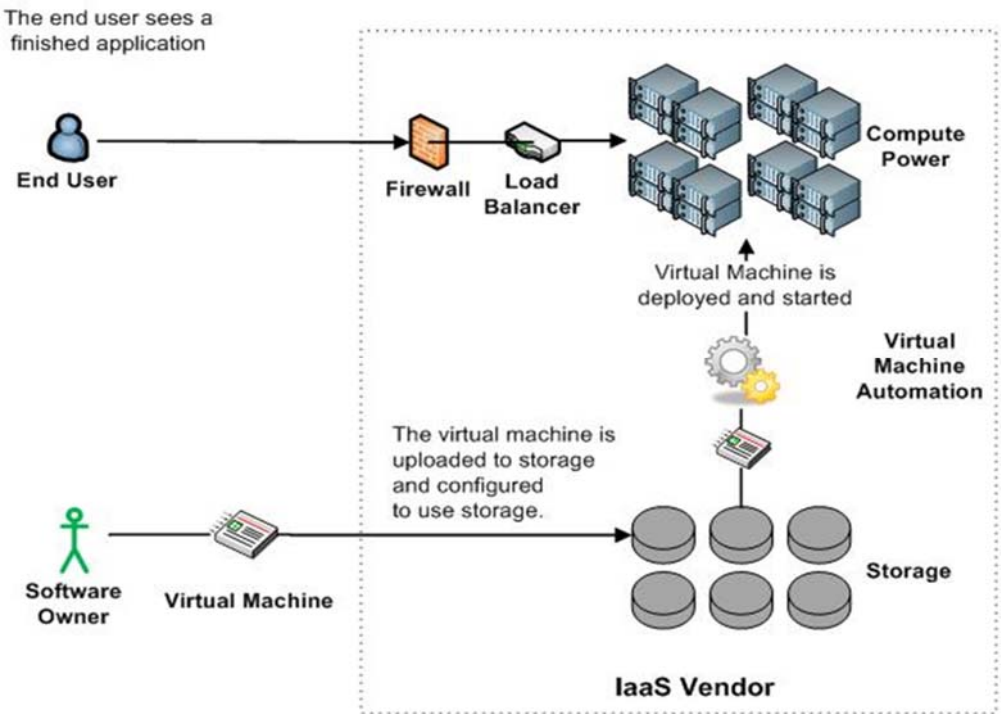
Classification of Cloud Computing based on Service Provided

- Infrastructure as a service (IaaS)
  - Offering hardware related services using the principles of cloud computing. These could include storage services (database or disk storage) or virtual servers.
  - Amazon EC2, Amazon S3, Rackspace Cloud Servers and Flexiscale.
- Platform as a Service (PaaS)
  - Offering a development platform on the cloud.
  - Google's Application Engine, Microsofts Azure, Salesforce.com's force.com .
- Software as a service (SaaS)
  - Including a complete software offering on the cloud. Users can access a software application hosted by the cloud vendor on pay-per-use basis. This is a well-established sector.
  - Salesforce.coms' offering in the online Customer Relationship Management (CRM) space, Googles gmail and Microsofts hotmail, Google docs.

Cloud Models



Infrastructure as a Service (IaaS)



Infrastructure as a Service (IaaS)

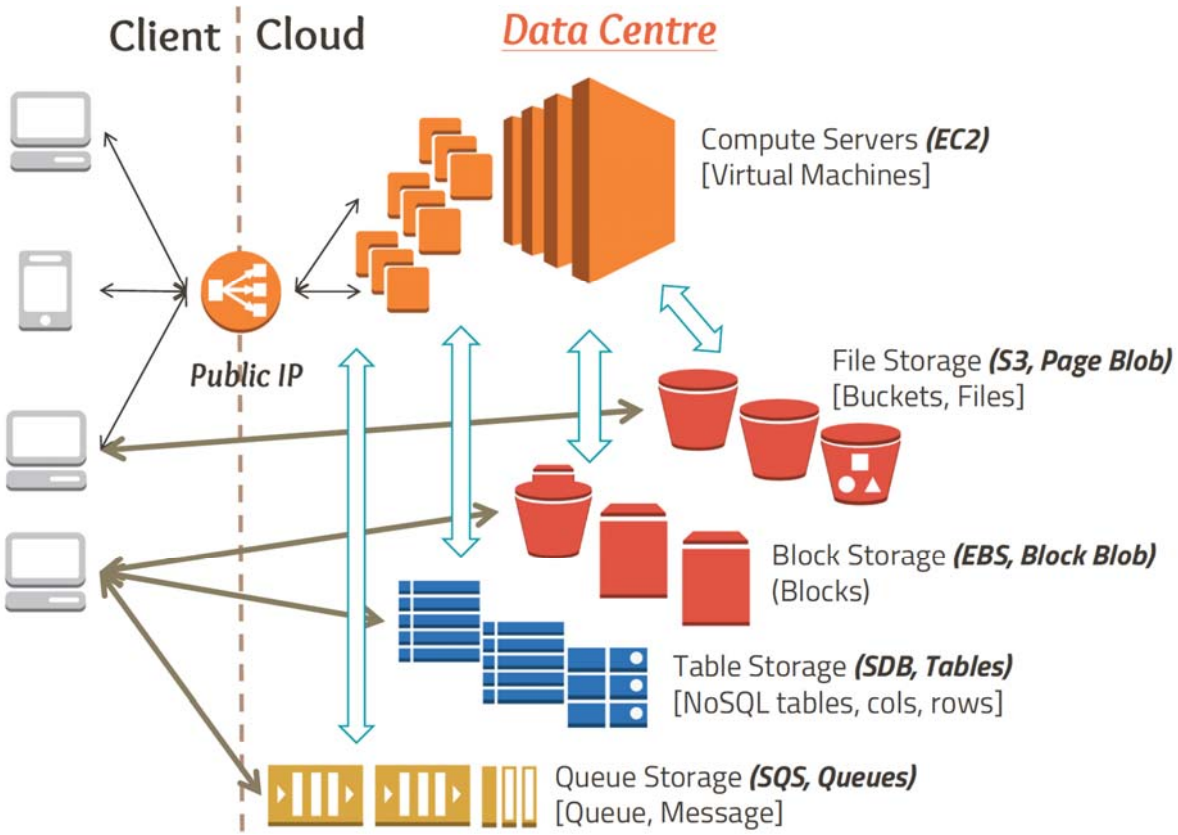
"Deliver computer infrastructure – typically a platform virtualization environment – as a service, along with raw (block) storage and networking." *

In IaaS, customers rent computing resources rather than purchase them and access the resources through a (Web) service infrastructure. Service billed typically monthly on a usage basis.

Example: Amazon EC2 (see later)

* http://en.wikipedia.org/wiki/Cloud_computing

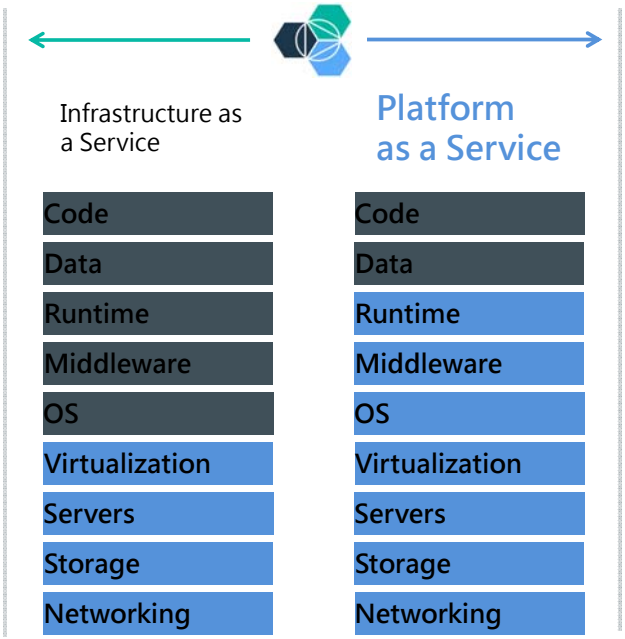
Typical IaaS Architecture



Take IBM Bluemix as an example

Capabilities in Bluemix now span PaaS and IaaS and can be delivered as a public, dedicated, or on-premises* implementation.

Customer Managed
Service Provider Managed



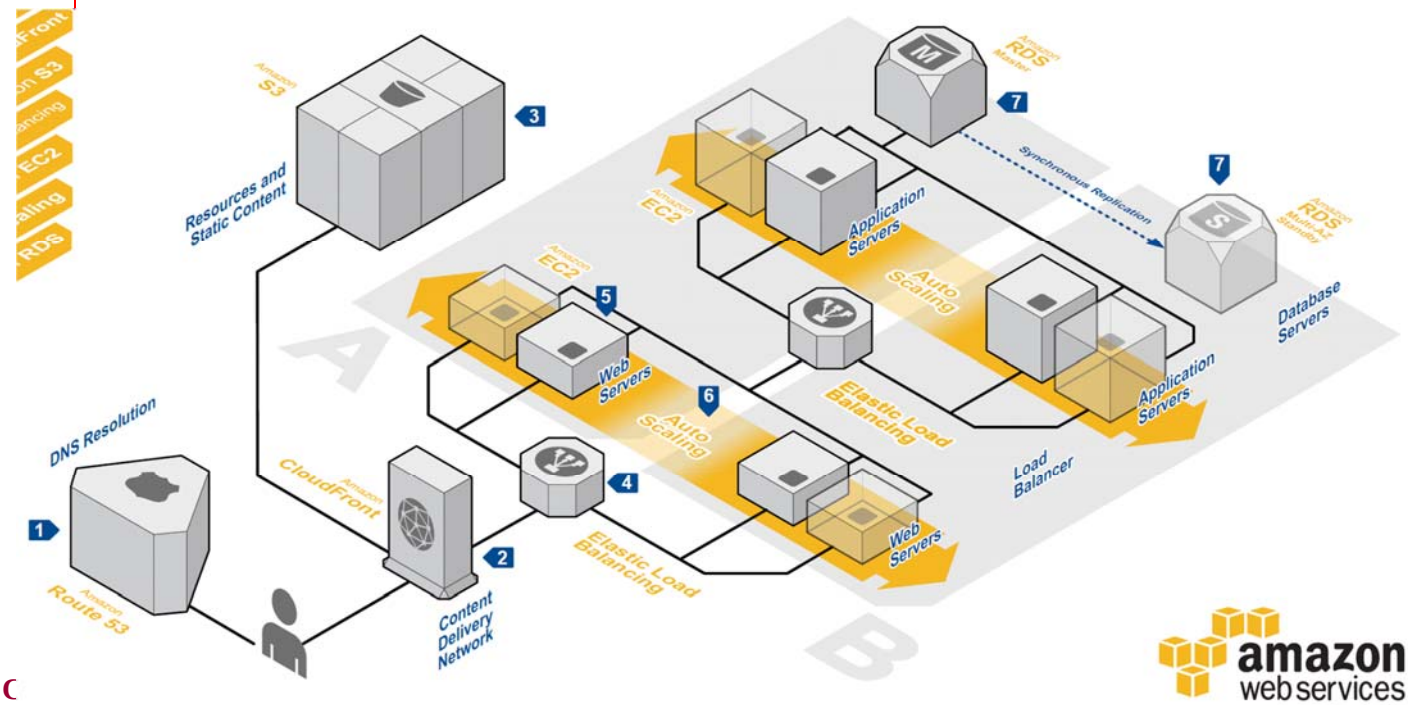
Built on open technologies:

*Bluemix Local coming Summer 2015

Amazon: Web-application Hosting

S3: Simple Storage Service
EC2: Elastic Compute Cloud
AMI: Amazon Machine Image

Amazon RDS: Amazon Relational Database Service



Infrastructure as a Service (IaaS) Advantages

- Access to preconfigured environment
- Use of latest technology
- Reduced cost and risk of having third party maintain resources
- No capital investment
- No IT personal to maintain remote hardware/software
- Able to manage peak demand as needed without having to purchase a larger system that would be underutilized at other times
- Secure – security handed by provider

Disadvantages: Delays in network (Internet), confidential data concerns, Intellectual property

Platform as a service (PaaS)

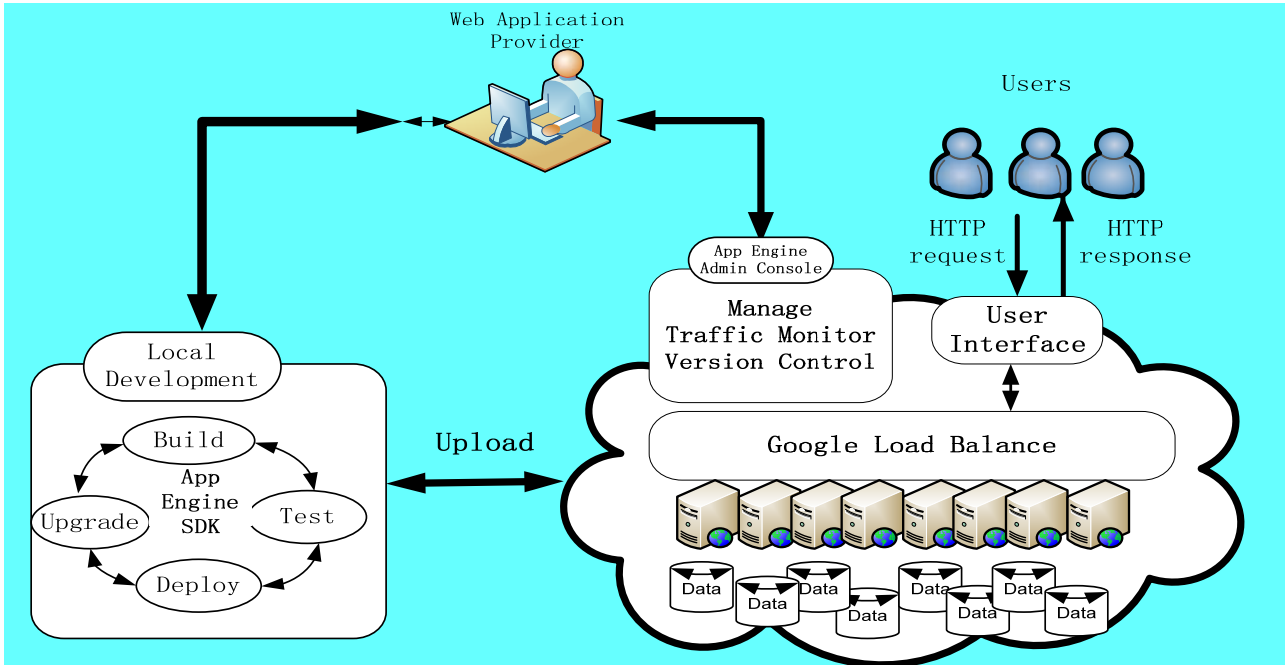
- Delivery of a computing platform and solution stack as a service
- Cloud providers deliver a computing platform typically including OS, programming language execution environment, database, and web server.
- AP developers develop and run their software on a cloud platform without the cost and complexity of buying and managing the underlying hardware and software layers.
- Examples of PaaS include: Amazon Elastic Beanstalk, Cloud Foundry, Heroku, Force.com, EngineYard, Mendix, Google App Engine, Microsoft Azure and OrangeScape.

PaaS Offerings from Public Clouds

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 Platform for PaaS Operations



Software as a Service (SaaS)

" Software delivery model in which software and its associated data are hosted centrally (typically in the (Internet) cloud) and are typically accessed by users using a thin client, normally using a web browser over the Internet." *

- Customer pays for access to software that is installed on providers remote computing resources, typically paid for on a subscription licensing model.

- Many business software (accounting, email, management software, …) suitable for SaaS

SaaS example: Google docs

* http://en.wikipedia.org/wiki/Software_as_a_Service

Software as a Service (SaaS) Advantages

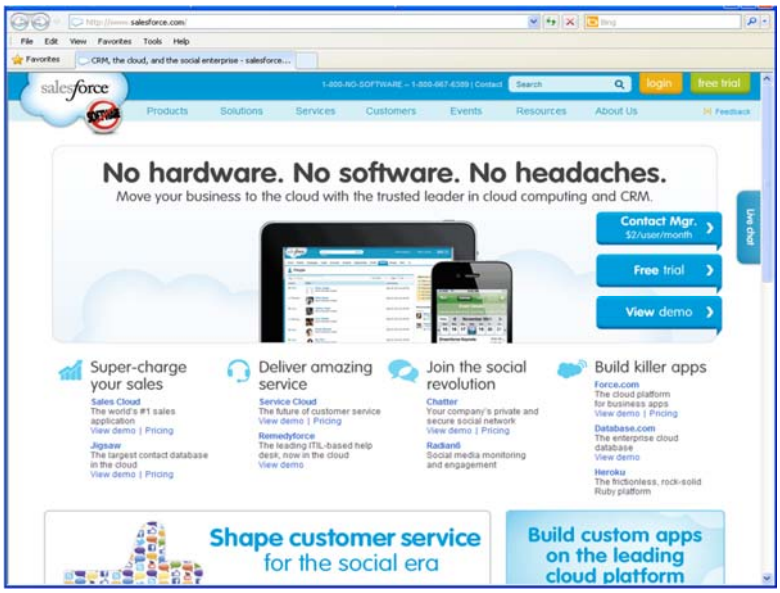
- Relives businesses of maintaining software - updates. licenses, multiple copies being consistent… etc.
- Since access is through a web browser, can access software from anywhere (globally) - mobile device etc.
- Facilitates internal collaboration
- Compatible data - All users use same software version
- No or less dedicated application programming

- Ease of scalability
- Optimize hardware design on the software being rented

Disadvantages of SaaS

- Outsource mission critical functions
- Lack of control and customization
- Integration with other on demand vendor software
- No differentiation with competitors
- Potentially more costly as a function of time

Salesforce.com



Formed in 1999 focusing on Software as a Service (SaaS) and Customer relationship management (CRM)

CRM – “strategy for managing a company’s interactions with customers, clients and sales prospects… technology to organize, automate, and synchronize business processes—principally sales activities, but also … marketing, customer service, and technical support.”

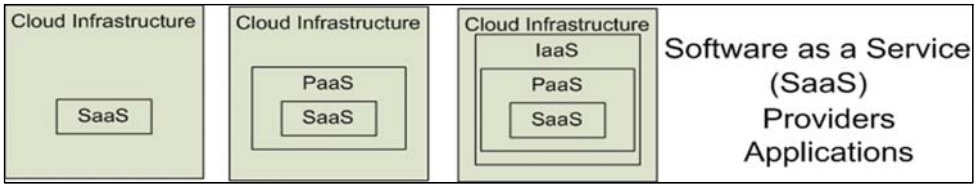
http://en.wikipedia.org/wiki/Customer_Relationship_Management

Cloud Service Models

Software as a Service (SaaS) | **Platform as a Service (PaaS)** | **Infrastructure as a Service (IaaS)**

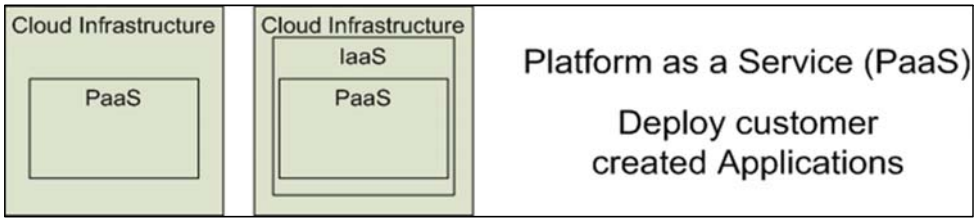
SalesForce CRM

LotusLive



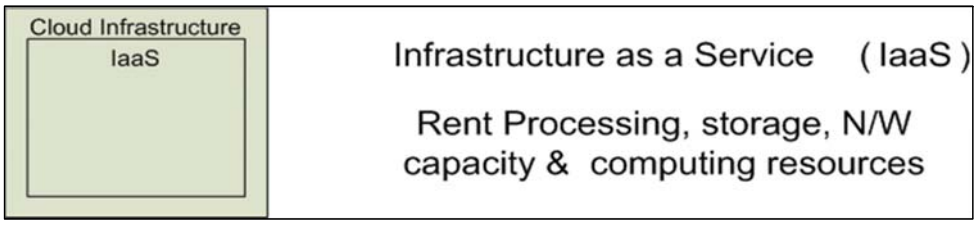
Google App

Windows Azure



amazon web services

rackspace HOSTING



Some Commercial-Oriented Cloud platforms/technologies

System Property	Amazon EC2 & S3	Google App Engine	Microsoft Azure	Manjrasoft Aneka
Focus	IaaS	IaaS/PaaS	IaaS/PaaS	PaaS
Service Type	Compute (EC2), Storage (S3)	Web apps	Web and non-web apps	Compute/Data
Virtualisation	OS Level: Xen	Apps container	OS level/Hyper-V	Resource Manager and Scheduler
Dynamic Negotiation of QoS	None	None	None	SLA-oriented/ Resource Reservation
User Access Interface	EC2 Command-line Tools	Web-based Administration Console	Windows Azure portal	Workbench, Tools
Web APIs	Yes	Yes	Yes	Yes
Value-added Service Providers	Yes	No	Yes	No
Programming Framework	Amazon Machine Image (AMI)	Python	.NET framework	Multiple App models in.NET languages

Everything as a Service

- Utility computing = Infrastructure as a Service (IaaS)

- Storage-as-a-service
- Database-as-a-service
- Information-as-a-service
- Process-as-a-service
- Application-as-a-service
- Platform-as-a-service
- Integration-as-a-service
- Security-as-a-service
- Management/ Governance-as-a-service
- Testing-as-a-service
- Infrastructure-as-a-service

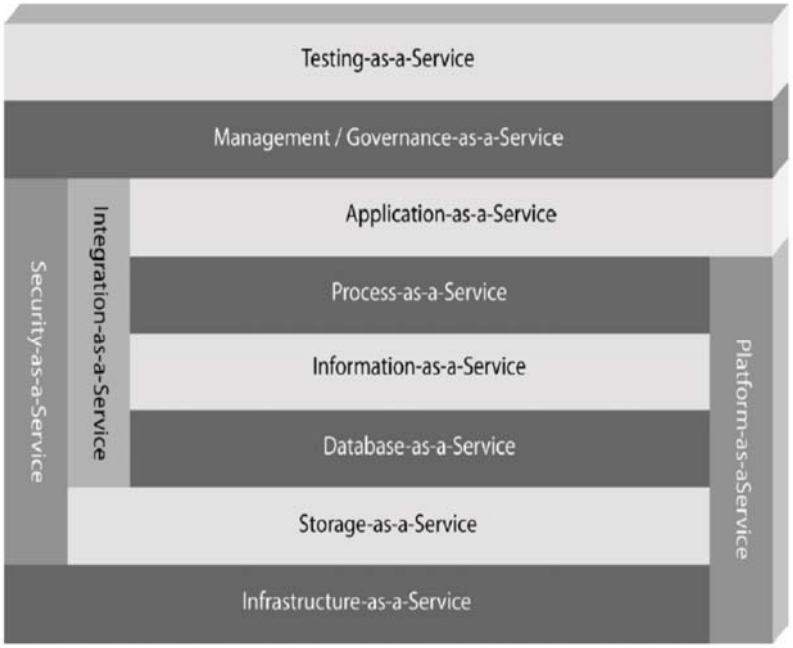


Figure 1: The patterns or categories of cloud computing providers allow you to use a discrete set of services within your architecture.

InfoWorld Cloud Computing Deep Dive

Warehouse-Scale Computer (WSC)

- Provides Internet services
  - Search, social networking, online maps, video sharing, online shopping, email, cloud computing, etc.
- Differences with HPC "clusters":
  - Clusters have higher performance processors and network
  - Clusters emphasize thread-level parallelism, WSCs emphasize request-level parallelism
- Differences with datacenters:
  - Datacenters consolidate different machines and software into one location
  - Datacenters emphasize virtual machines and hardware heterogeneity in order to serve varied customers

Design Considerations for WSC:

- Cost-performance
  - Small savings add up
- Energy efficiency
  - Affects power distribution and cooling
  - Work per joule
- Dependability via redundancy
- Network I/O
- Interactive and batch processing workloads
- Ample computational parallelism is not important
  - Most jobs are totally independent
  - "Request-level parallelism"
- Operational costs count
  - Power consumption is a primary constraint when designing system
- Scale and its opportunities and problems
  - Can afford customized systems since WSC require volume purchase

Typical Datacenter Layout

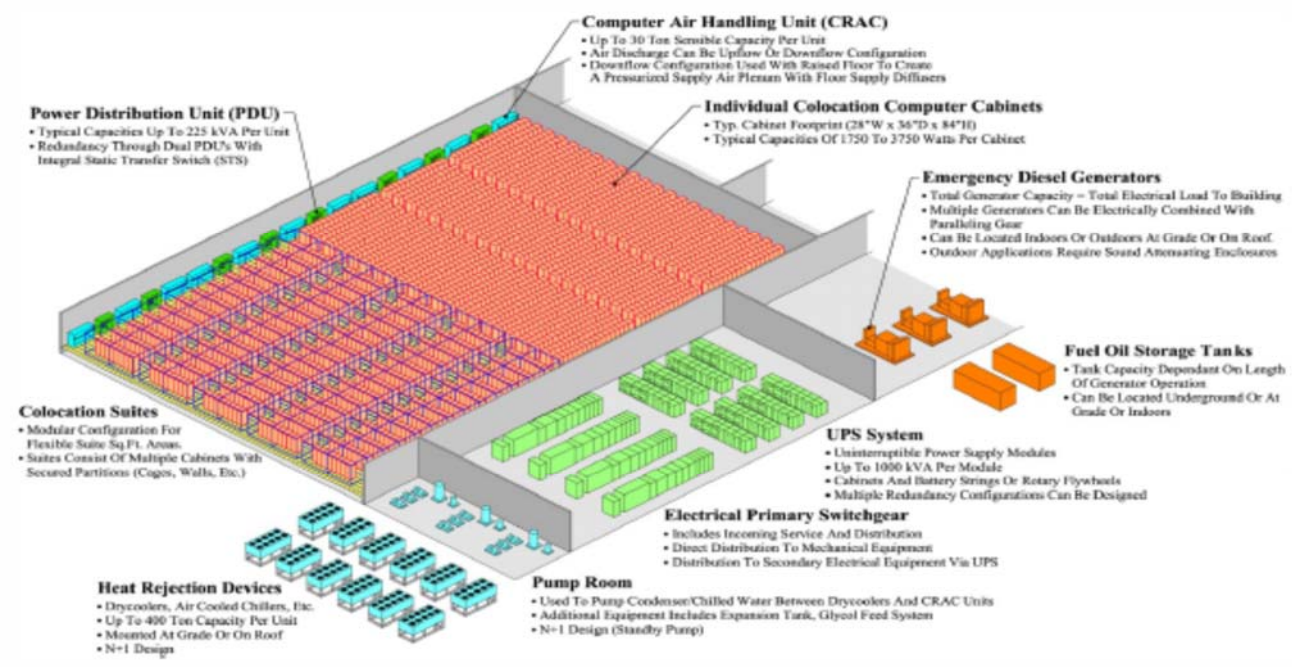


FIGURE 4.1: The main components of a typical datacenter (image courtesy of DLB Associates [23]).

Power and Cooling Requirements

- Cooling system also uses water (evaporation and spills)
  - E.g. 70,000 to 200,000 gallons per day for an 8 MW facility
- Power cost breakdown:
  - Chillers: 30-50% of the power used by the IT equipment
  - Air conditioning: 10-20% of the IT power, mostly due to fans
- How many servers can a WSC support?
  - Each server:
    - "Nameplate power rating" gives maximum power consumption
    - To get actual, measure power under actual workloads
  - Oversubscribe cumulative server power by 40%, but monitor power closely

Measuring Efficiency of a WSC

- Power Utilization Effectiveness (PEU)
  - = Total facility power / IT equipment power
  - Median PUE on 2006 study was 1.69
- Performance
  - Latency is important metric because it is seen by users
  - Bing study: users will use search less as response time increases
  - Service Level Objectives (SLOs)/Service Level Agreements (SLAs)
    - E.g. 99% of requests be below 100 ms

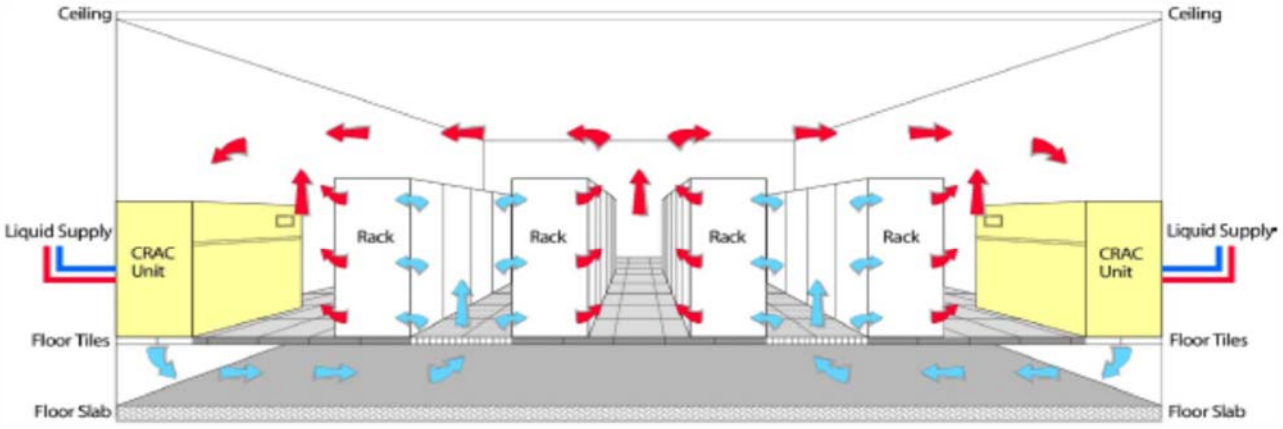
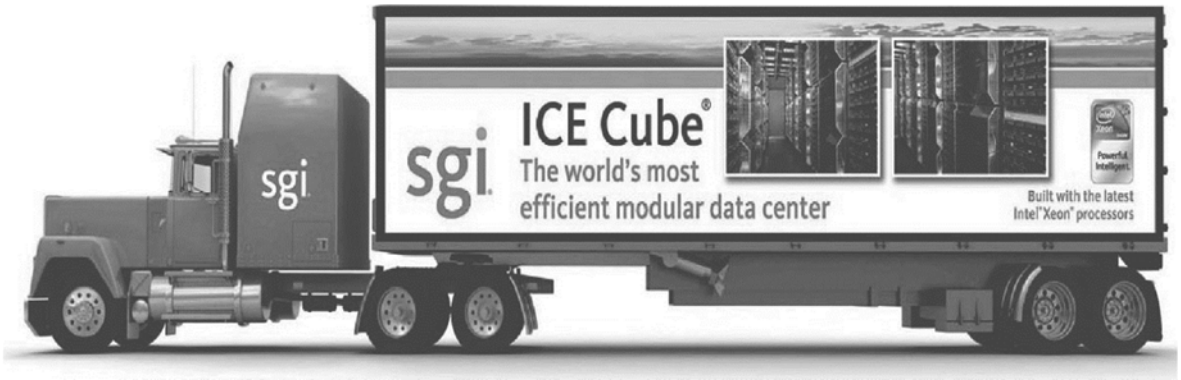


FIGURE 4.2: Datacenter raised floor with hot–cold aisle setup (image courtesy of DLB Associates [23]).

$$\text{Efficiency} = \frac{\text{Computation}}{\text{Total Energy}} = \underbrace{\left(\frac{1}{\text{PUE}}\right)}_{\text{(a)}} \times \underbrace{\left(\frac{1}{\text{SPUE}}\right)}_{\text{(b)}} \times \underbrace{\left(\frac{\text{Computation}}{\text{Total Energy to Electronic Components}}\right)}_{\text{(c)}}$$

(Courtesy of Luiz Andre Barroso and Urs Holzle, Google Inc., 2009)

Modular Data Center



Cloud Computing as A Service

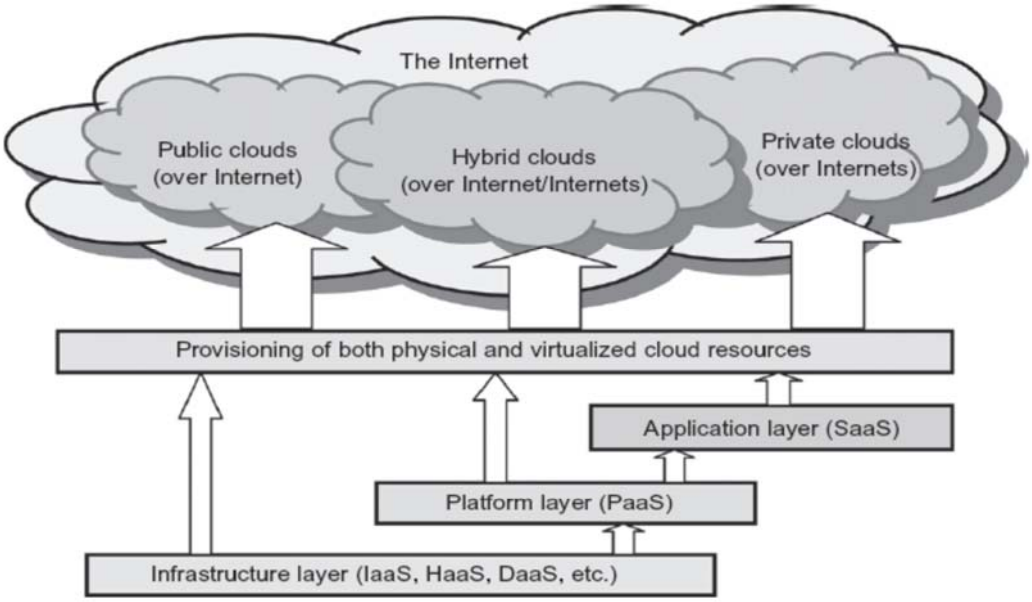


FIGURE 4.15

Layered architectural development of the cloud platform for IaaS, PaaS, and SaaS applications over the Internet.

Virtualized servers, storage , and network for cloud platform construction