Parallel & Distributed Computing (WQD7008)

Week 8

Cloud Computing Platforms

2019/2020 Semester 1

Dr. Hamid Tahaei

Cloud Computing and Service Models

- Over the past two decades, the world economy has rapidly moved from manufacturing to more service-oriented.
- In 2010, 80 percent of the U.S. economy was driven by the service industry, leaving only 15 percent in manufacturing and 5 percent in agriculture and other areas.
- In 2009, the global cloud service marketplace reached $17.4 billion. IDC predicted in 2010 that the cloudbased economy may increase to $44.2 billion by 2013. Developers of innovative cloud applications no longer acquire large capital equipment in advance. They just rent the resources from some large data centers that have been automated for this purpose.
- Cloud → the notion of pay per use.
- Cloud computing leverages dynamic resources to deliver large numbers of services to end users.
- Cloud computing is a high-throughput computing (HTC) paradigm whereby the infrastructure provides the services through a large data center or server farms.
- The cloud computing model enables users to share access to resources from anywhere at any time through their connected devices.

Cloud Computing and Service Models

- Cloud computing avoids large data movement, resulting in much better network bandwidth utilization.
- Machine virtualization has enhanced resource utilization, increased application flexibility, and reduced the total cost of using virtualized data-center resources.
- The cloud offers significant benefit to IT companies by freeing them from the low-level task of setting up the hardware (servers) and managing the system software.
- The main idea is to move desktop computing to a service-oriented platform using server clusters and huge databases at data centers.
- Cloud computing leverages its low cost and simplicity to both providers and users.
- Two opposite views of cloud computing → Centralized OR Distributed?
  - All computations in cloud applications are distributed to servers in a data center.
  - These are mainly virtual machines (VMs) in virtual clusters created out of data-center resources.
  - In this sense, cloud platforms are systems distributed through virtualization.

Cloud Computing and Service Models

Public, Private, and Hybrid Clouds

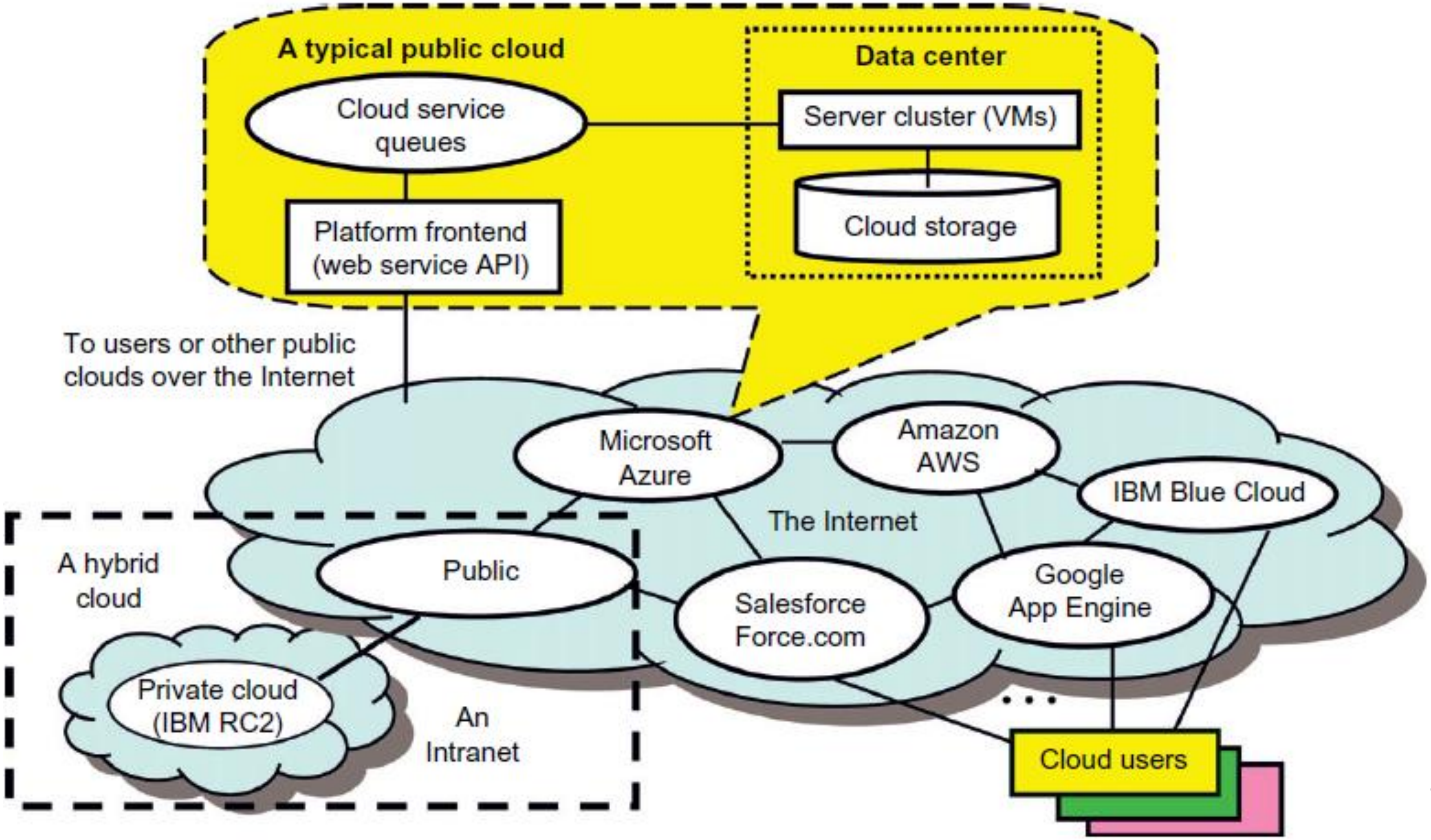


FIGURE 4.1

Public, private, and hybrid clouds illustrated by functional architecture and connectivity of representative clouds available by 2011.

Cloud Computing and Service Models

Public Clouds

- Built over the Internet and can be accessed by any user who has paid for the service.
- Owned by service providers and are accessible through a subscription.
- Many public clouds are available, including Google App Engine (GAE), Amazon Web Services (AWS), Microsoft Azure, IBM Blue Cloud, and Salesforce.com’s Force.com.
- ~~The providers of the aforementioned clouds are commercial providers that offer a publicly accessible remote interface for creating and managing VM instances within their proprietary infrastructure.~~
- A public cloud delivers a selected set of business processes.
- The application and infrastructure services are offered on a flexible price-per-use basis.

Cloud Computing and Service Models

Private Clouds

- Built within the domain of an intranet owned by a single organization.
- Therefore, it is client owned and managed, and its access is limited to the owning clients and their partners.
- Its deployment was not meant to sell capacity over the Internet through publicly accessible interfaces.
- Private clouds give local users a flexible and agile private infrastructure to run service workloads within their administrative domains.
- A private cloud is supposed to deliver more efficient and convenient cloud services.
- It may impact the cloud standardization, while retaining greater customization and organizational control.

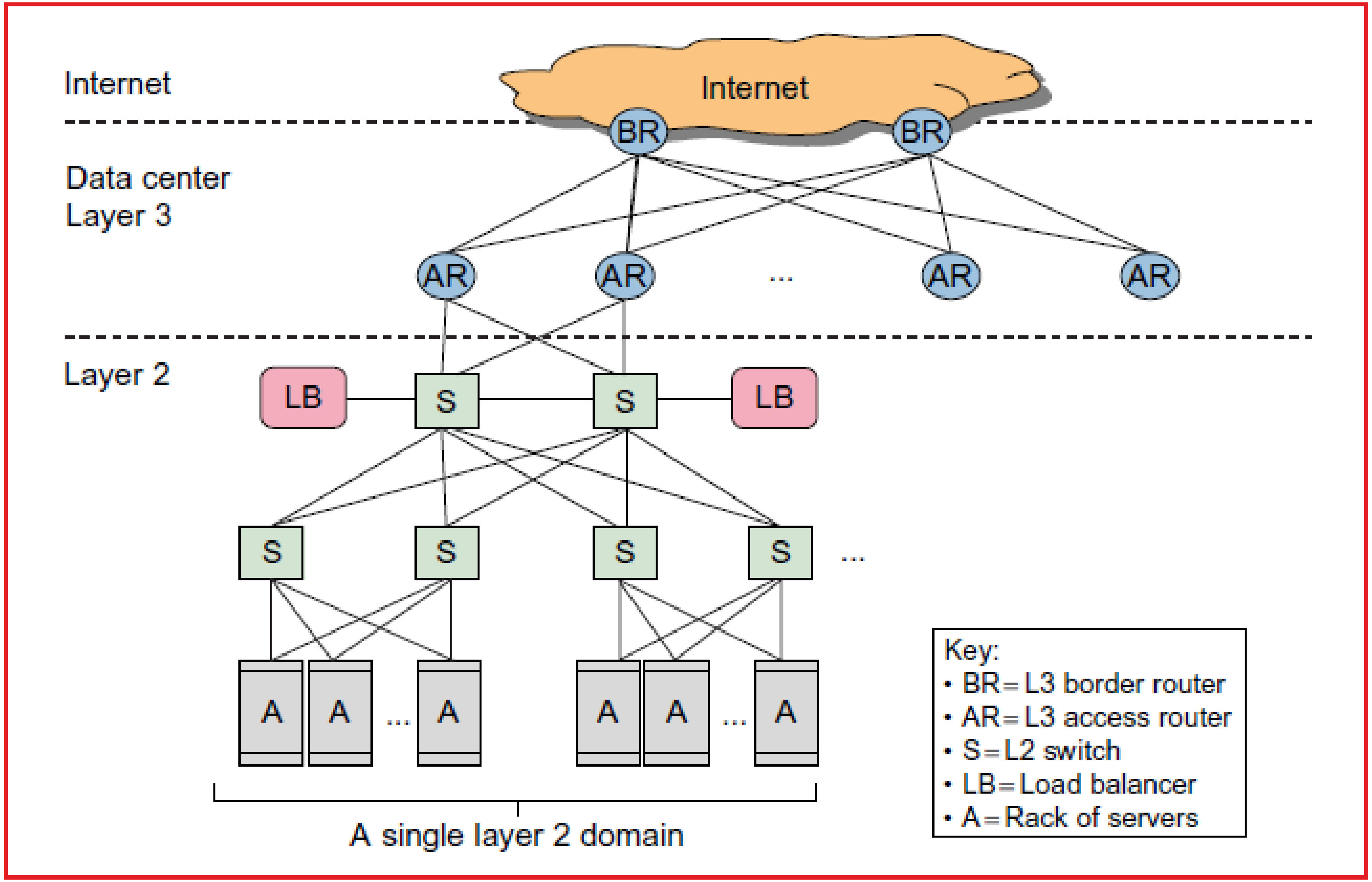
Cloud Computing and Service Models

Hybrid Clouds

- Built with both public and private clouds
- Private clouds can also support a hybrid cloud model by supplementing local infrastructure with computing capacity from an external public cloud.
- A hybrid cloud provides access to clients, the partner network, and third parties.
- In summary, public clouds promote standardization, preserve capital investment, and offer application flexibility.
- Private clouds attempt to achieve customization and offer higher efficiency, resiliency, security, and privacy.
- Hybrid clouds operate in the middle, with many compromises in terms of resource sharing.

Cloud Computing and Service Models

Data-Center Networking Structure



Cloud Computing and Service Models

Data-Center Networking Structure

- Data centers and supercomputers also differ in networking requirements.
- Supercomputers use custom-designed high-bandwidth networks such as fat trees or 3D torus networks.
- Data-center networks are mostly IP-based commodity networks, such as the 10 Gbps Ethernet network, which is optimized for Internet access.
- The server racks are at the bottom Layer 2, and they are connected through fast switches (S) as the hardware core. The data center is connected to the Internet at Layer 3 with many access routers (ARs) and border routers (BRs).
- Grid computing is the backbone of cloud computing in that the grid has the same goals of resource sharing with better utilization of research facilities.
- Grids are more focused on delivering storage and computing resources while cloud computing aims to achieve economies of scale with abstracted services and resources.

Cloud Computing and Service Models

Cloud Ecosystem and Enabling Technologies

- Cloud computing platforms differ from conventional computing platforms in many aspects.
- The traditional computing model, involves buying the hardware, acquiring the necessary system software, installing the system, testing the configuration, and executing the application code and management of resources. What is even worse is that this cycle repeats itself in about every 18 months, meaning the machine we bought becomes obsolete every 18 months.
- The cloud computing model follows a pay-as-you-go model. Therefore the cost is significantly reduced, because we simply rent computer resources without buying the computer in advance. All hardware and software resources are leased from the cloud provider without capital investment on the part of the users. Only the execution phase costs some money.
- The experts at IBM have estimated that an 80 percent to 95 percent saving results from cloud computing, compared with the conventional computing paradigm. This is very much desired, especially for small businesses, which requires limited computing power and thus avoid the purchase of expensive computers or servers repeatedly every few years.

Cloud Computing and Service Models

Cloud Design Objectives

- **Shifting computing from desktops to data centers** Computer processing, storage, and software delivery is shifted away from desktops and local servers and toward data centers over the Internet.
- **Service provisioning and cloud economics** Providers supply cloud services by signing SLAs with consumers and end users. The services must be efficient in terms of computing, storage, and power consumption. Pricing is based on a pay-as-you-go policy.
- **Scalability in performance** The cloud platforms and software and infrastructure services must be able to scale in performance as the number of users increases.
- **Data privacy protection** Can you trust data centers to handle your private data and records? This concern must be addressed to make clouds successful as trusted services.
- **High quality of cloud services** The QoS of cloud computing must be standardized to make clouds interoperable among multiple providers.
- **New standards and interfaces** This refers to solving the data lock-in problem associated with data centers or cloud providers. Universally accepted APIs and access protocols are needed to provide high portability and flexibility of virtualized applications.

Cloud Computing and Service Models

Cost Model

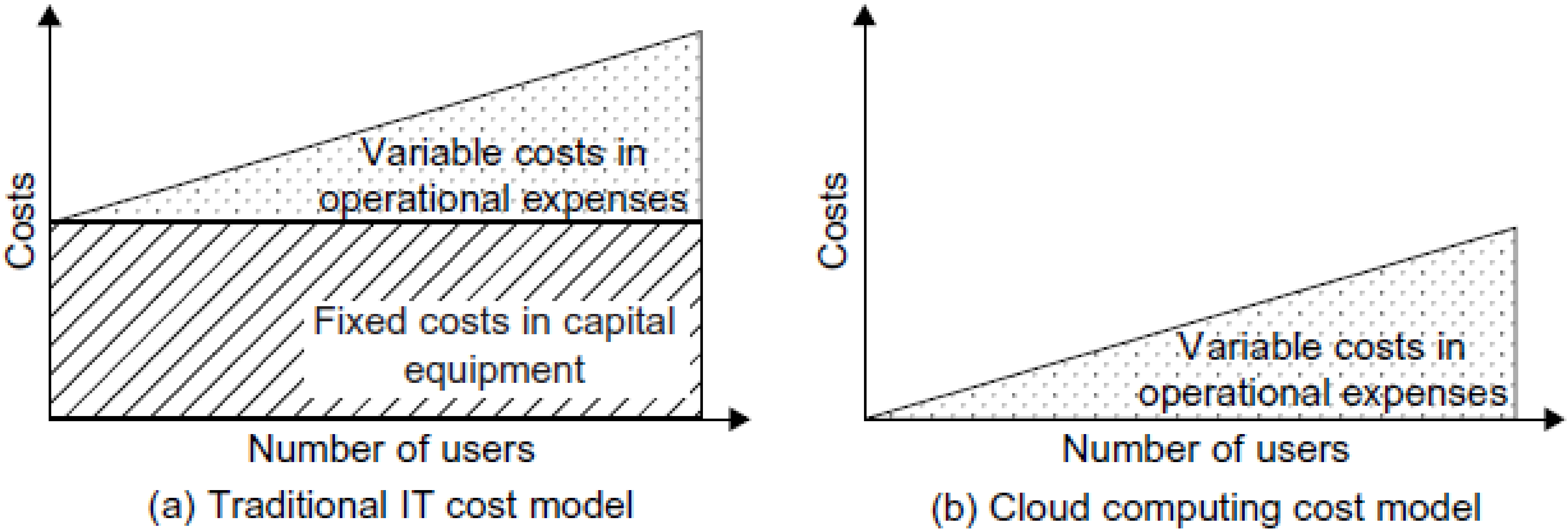


FIGURE 4.3

Computing economics between traditional IT users and cloud users.

Cloud Computing and Service Models

Cloud Ecosystems

An ecosystem (four levels) was suggested for building private clouds.

- At the user end, consumers demand a flexible platform.
- At the cloud management level, the cloud manager provides virtualized resources over an IaaS platform.
- At the virtual infrastructure (VI) management level, the manager allocates VMs over multiple server clusters.
- Finally, at the VM management level, the VM managers handle VMs installed on individual host machines.

Cloud Computing and Service Models

Cloud Ecosystems

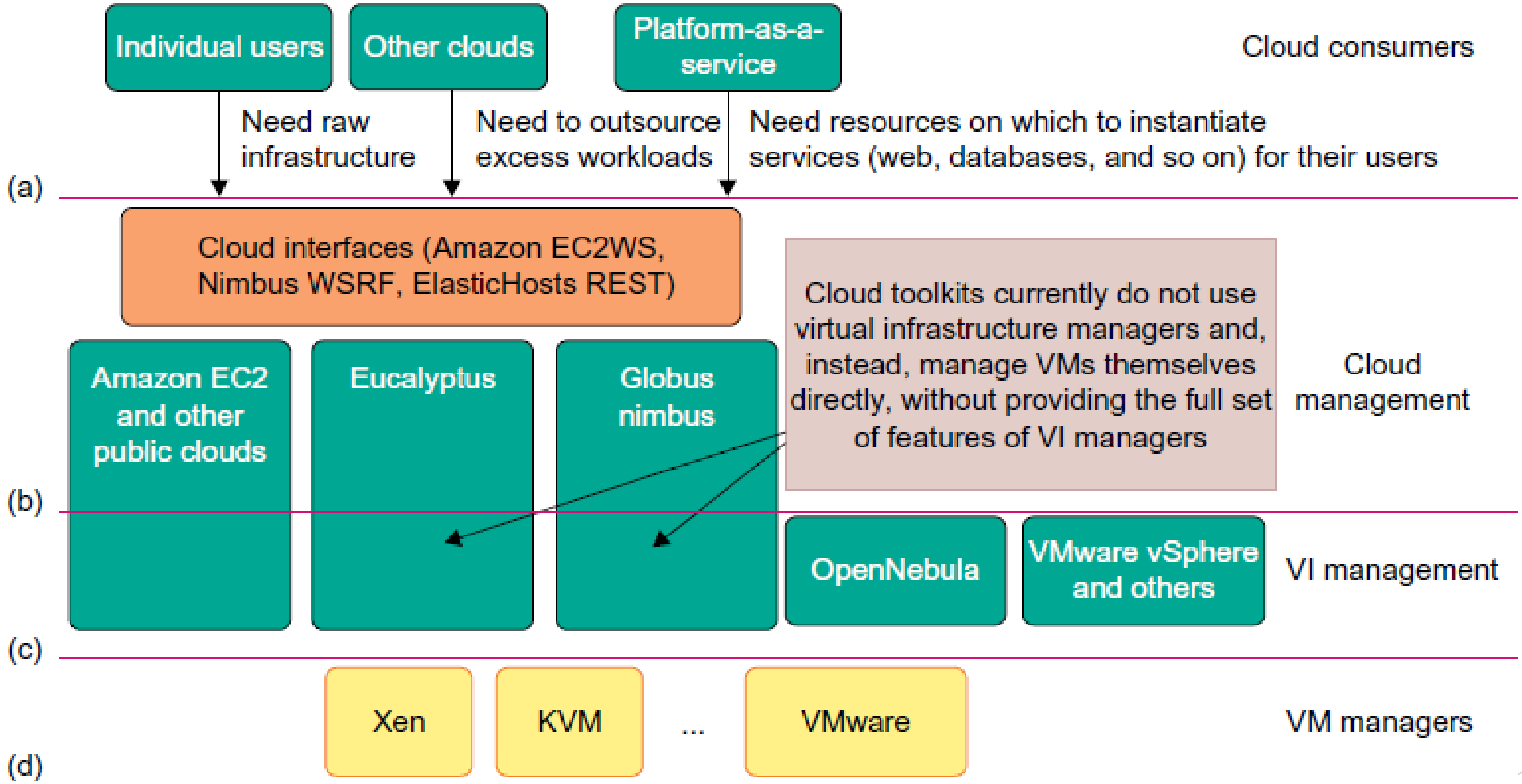


FIGURE 4.4

Cloud ecosystem for building private clouds: (a) Consumers demand a flexible platform; (b) Cloud manager provides virtualized resources over an IaaS platform; (c) VI manager allocates VMs; (d) VM managers handle VMs installed on servers.

Cloud Computing and Service Models

IaaS, PasS, SaaS

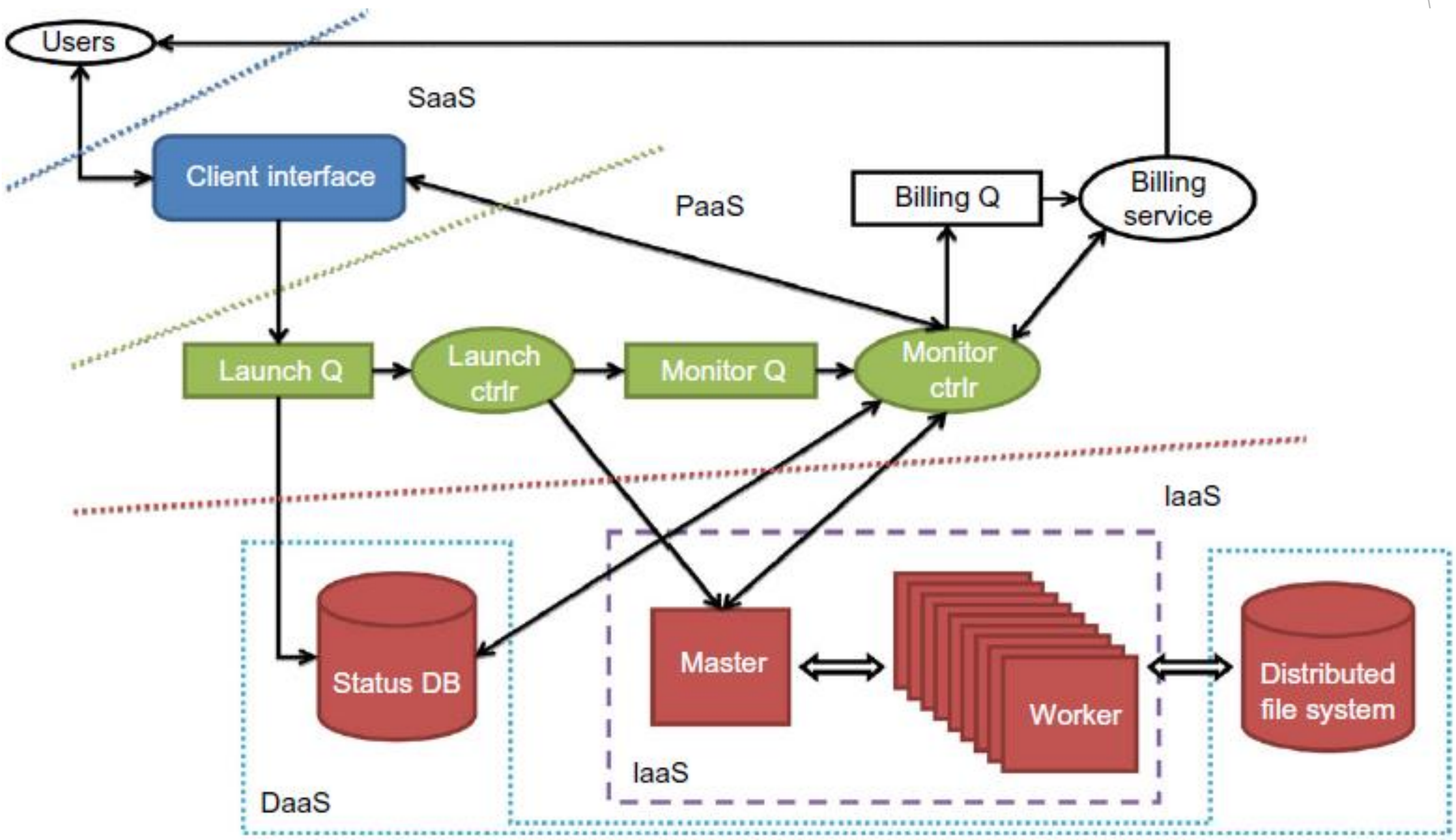


FIGURE 4.5

The IaaS, PaaS, and SaaS cloud service models at different service levels..

Cloud Computing and Service Models

Infrastructure-as-a-Service (IaaS)

- This model allows users to use virtualized IT resources for computing, storage, and networking. In short, the service is performed by rented cloud infrastructure. The user can deploy and run his applications over his chosen OS environment.
- The user does not manage or control the underlying cloud infrastructure, but has control over the OS, storage, deployed applications, and possibly select networking components. This IaaS model encompasses storage as a service, compute instances as a service, and communication as a service.

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

Cloud Computing and Service Models

Platform as a Service (PasS)

- To be able to develop, deploy, and manage the execution of applications using provisioned resources demands a cloud platform with the proper software environment. Such a platform includes operating system and runtime library support. This has triggered the creation of the PaaS model to enable users to develop and deploy their user applications.

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

Cloud Computing and Service Models

Software as a Service (SaaS)

- This refers to browser-initiated application software over thousands of cloud customers. Services and tools offered by PaaS are utilized in construction of applications and management of their deployment on resources offered by IaaS providers.
- The SaaS model provides software applications as a service. As a result, on the customer side, there is no upfront investment in servers or software licensing.
- On the provider side, costs are kept rather low, compared with conventional hosting of user applications.
- Customer data is stored in the cloud that is either vendor proprietary or publicly hosted to support PaaS and IaaS.
- The best examples of SaaS services include Google Gmail and docs, Microsoft SharePoint, and the CRM software from Salesforce.com.

ARCHITECTURAL DESIGN OF COMPUTE AND STORAGE CLOUDS

A Generic Cloud Architecture Design

- Cloud Platform Design Goals:
  - Scalability, virtualization, efficiency, and reliability

Enabling Technologies for Clouds

- Clouds are enabled by the progress in hardware, software, and networking technologies:

Table 4.3 Cloud-Enabling Technologies in Hardware, Software, and Networking

Technology	Requirements and Benefits
Fast platform deployment	Fast, efficient, and flexible deployment of cloud resources to provide dynamic computing environment to users
Virtual clusters on demand	Virtualized cluster of VMs provisioned to satisfy user demand and virtual cluster reconfigured as workload changes
Multitenant techniques	SaaS for distributing software to a large number of users for their simultaneous use and resource sharing if so desired
Massive data processing	Internet search and web services which often require massive data processing, especially to support personalized services
Web-scale communication	Support for e-commerce, distance education, telemedicine, social networking, digital government, and digital entertainment applications
Distributed storage	Large-scale storage of personal records and public archive information which demands distributed storage over the clouds
Licensing and billing services	License management and billing services which greatly benefit all types of cloud services in utility computing

ARCHITECTURAL DESIGN OF COMPUTE AND STORAGE CLOUDS

Virtualization Support and Disaster Recovery - Virtualization for IaaS

Use of VMs in clouds has the following distinct benefits:

- (1) System administrators consolidate workloads of underutilized servers in fewer servers;
- (2) VMs have the ability to run legacy code without interfering with other APIs;
- (3) VMs can be used to improve security through creation of sandboxes for running applications with questionable reliability;
- (4) virtualized cloud platforms can apply performance isolation, letting providers offer some guarantees and better QoS to customer applications.

ARCHITECTURAL DESIGN OF COMPUTE AND STORAGE CLOUDS

Virtualization Support and Disaster Recovery - VM Cloning for Disaster Recovery

- VM technology requires an advanced disaster recovery scheme.
  - One scheme is to recover one physical machine by another physical machine.
  - The second scheme is to recover one VM by another VM.
  - Traditional disaster recovery from one physical machine to another is rather slow, complex, and expensive. A much shorter disaster recovery time, about 40 percent of that to recover the physical machines. Virtualization aids in fast disaster recovery by VM encapsulation.
- The idea is to make a clone VM on a remote server for every running VM on a local server. The remote VM should be in a suspended mode. A cloud control center should be able to activate this clone VM in case of failure of the original VM, taking a snapshot of the VM to enable live migration in a minimal amount of time. The migrated VM can run on a shared Internet connection. Only updated data and modified states are sent to the suspended VM to update its state. The Recovery Property Objective (RPO) and Recovery Time Objective (RTO) are affected by the number of snapshots taken. Security of the VMs should be enforced during live migration of VMs.

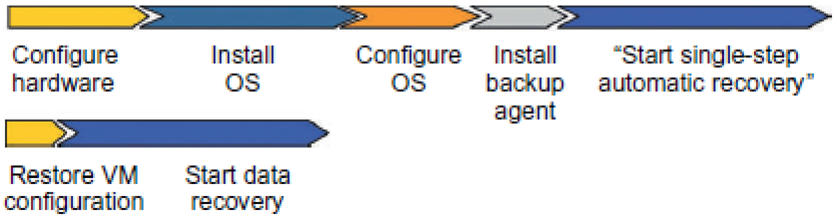


FIGURE 4.18

Recovery overhead of a conventional disaster recovery scheme, compared with that required to recover from live migration of VMs.

ARCHITECTURAL DESIGN OF COMPUTE AND STORAGE CLOUDS

Architectural Design Challenges

Challenge 1—Service Availability and Data Lock-in Problem

- The management of a cloud service by a single company is often the source of single points of failure. To achieve HA, one can consider using multiple cloud providers.
- Even if a company has multiple data centers located in different geographic regions, it may have common software infrastructure and accounting systems.
- Therefore, using multiple cloud providers may provide more protection from failures.
- Another availability obstacle is distributed denial of service (DDoS) attacks.
- Criminals threaten to cut off the incomes of SaaS providers by making their services unavailable.
- Some utility computing services offer SaaS providers the opportunity to defend against DDoS attacks by using quick scale-ups.

ARCHITECTURAL DESIGN OF COMPUTE AND STORAGE CLOUDS

Architectural Design Challenges

Challenge 2—Data Privacy and Security Concerns

- Current cloud offerings are essentially public (rather than private) networks, exposing the system to more attacks. Many obstacles can be overcome immediately with well-understood technologies such as encrypted storage, virtual LANs, and network middleboxes (e.g., firewalls, packet filters). For example, you could encrypt your data before placing it in a cloud. Many nations have laws requiring SaaS providers to keep customer data and copyrighted material within national boundaries.
  Traditional network attacks include buffer overflows, DoS attacks, spyware, malware, rootkits, Trojan horses, and worms. In a cloud environment, newer attacks may result from hypervisor malware, guest hopping and hijacking, or VM rootkits. Another type of attack is the man-in-the-middle attack for VM migrations. In general, passive attacks steal sensitive data or passwords. Active attacks may manipulate kernel data structures which will cause major damage to cloud servers.

ARCHITECTURAL DESIGN OF COMPUTE AND STORAGE CLOUDS

Architectural Design Challenges

Challenge 3—Unpredictable Performance and Bottlenecks

- Multiple VMs can share CPUs and main memory in cloud computing, but I/O sharing is problematic. For example, to run 75 EC2 instances with the STREAM benchmark requires a mean bandwidth of 1,355 MB/second. However, for each of the 75 EC2 instances to write 1 GB files to the local disk requires a mean disk write bandwidth of only 55 MB/second. This demonstrates the problem of I/O interference between VMs. One solution is to improve I/O architectures and operating systems to efficiently virtualize interrupts and I/O channels.

ARCHITECTURAL DESIGN OF COMPUTE AND STORAGE CLOUDS

Architectural Design Challenges

Challenge 4—Distributed Storage and Widespread Software Bugs

- The database is always growing in cloud applications. The opportunity is to create a storage system that will not only meet this growth, but also combine it with the cloud advantage of scaling arbitrarily up and down on demand. This demands the design of efficient distributed SANs. Data centers must meet programmers' expectations in terms of scalability, data durability, and HA. Data consistence checking in SAN-connected data centers is a major challenge in cloud computing.
  Large-scale distributed bugs cannot be reproduced, so the debugging must occur at a scale in the production data centers. No data center will provide such a convenience. One solution may be a reliance on using VMs in cloud computing. The level of virtualization may make it possible to capture valuable information in ways that are impossible without using VMs. Debugging over simulators is another approach to attacking the problem, if the simulator is well designed.

ARCHITECTURAL DESIGN OF COMPUTE AND STORAGE CLOUDS

Architectural Design Challenges

Challenge 5—Cloud Scalability, Interoperability, and Standardization

- The pay-as-you-go model applies to storage and network bandwidth; both are counted in terms of the number of bytes used. Computation is different depending on virtualization level. Google App Engine (GAE) automatically scales in response to load increases and decreases; users are charged by the cycles used. AWS charges by the hour for the number of VM instances used, even if the machine is idle. The opportunity here is to scale quickly up and down in response to load variation, in order to save money, but without violating SLAs.
- Open Virtualization Format (OVF) describes an open, secure, portable, efficient, and extensible format for the packaging and distribution of VMs. It also defines a format for distributing software to be deployed in VMs. This VM format does not rely on the use of a specific host platform, virtualization platform, or guest operating system. The approach is to address virtual platform-agnostic packaging with certification and integrity of packaged software.
- OVF also defines a transport mechanism for VM templates, and can apply to different virtualization platforms with different levels of virtualization. In terms of cloud standardization, we suggest the ability for virtual appliances to run on any virtual platform. We also need to enable VMs to run on heterogeneous hardware platform hypervisors. This requires hypervisor-agnostic VMs. We also need to realize cross-platform live migration between x86 Intel and AMD technologies and support legacy hardware for load balancing. All these issue are wide open for further research.

ARCHITECTURAL DESIGN OF COMPUTE AND STORAGE CLOUDS

Architectural Design Challenges

Challenge 6—Software Licensing and Reputation Sharing

- Many cloud computing providers originally relied on open source software because the licensing model for commercial software is not ideal for utility computing. The primary opportunity is either for open source to remain popular or simply for commercial software companies to change their licensing structure to better fit cloud computing. One can consider using both pay-for-use and bulk-use licensing schemes to widen the business coverage.
- One customer's bad behavior can affect the reputation of the entire cloud. For instance, blacklisting of EC2 IP addresses by spam-prevention services may limit smooth VM installation. An opportunity would be to create reputation-guarding services similar to the "trusted e-mail" services currently offered (for a fee) to services hosted on smaller ISPs. Another legal issue concerns the transfer of legal liability. Cloud providers want legal liability to remain with the customer, and vice versa. This problem must be solved at the SLA level.