GUDLAVALLERU ENGINEERING COLLEGE

(An Autonomous Institute with Permanent Affiliation to JNTUK, Kakinada)

Seshadrirao Knowledge Village, Gudlavalleru – 521 356

Department of Computer Science and Engineering

HANDOUT

on

CLOUD COMPUTING

(ELECTIVE V)

Vision

To be a Centre of Excellence in computer science and engineering education and training to meet the challenging needs of the industry and society

Mission

- To impart quality education through well-designed curriculum in tune with the growing software needs of the industry.
- To be a Centre of Excellence in computer science and engineering education and training to meet the challenging needs of the industry and society.
- To serve our students by inculcating in them problem solving, leadership, teamwork skills and the value of commitment to quality, ethical behavior & respect for others.
- To foster industry-academia relationship for mutual benefit and growth.

Program Educational Objectives

- Identify, analyze, formulate and solve Computer Science and Engineering problems both independently and in a team environment by using the appropriate modern tools.
- Manage software projects with significant technical, legal, ethical, social, environmental and economic considerations
- Demonstrate commitment and progress in lifelong learning, professional development, leadership and Communicate effectively with professional clients and the public.

HANDOUT ON CLOUD COMPUTING

Class & Sem. :IV B.Tech – II Semester Year: 2019-20
Branch : CSE Credits : 3

==

1. Brief History and Scope of the Subject

Cloud computing is an information technology (IT) paradigm that enables ubiquitous access to shared pools of configurable system resources and higher-level services that can be rapidly provisioned with minimal management effort, often over the Internet. Cloud computing relies on sharing of resources to achieve coherence and economy of scale, similar to a utility.

Third-party clouds enable organizations to focus on their core businesses instead of expending resources on computer infrastructure and maintenance. Advocates note that cloud computing allows companies to avoid or minimize up-front IT infrastructure costs. Proponents also claim that cloud computing allows enterprises to get their applications up and running faster, with improved manageability and less maintenance, and that it enables IT teams to more rapidly adjust resources to meet fluctuating and unpredictable business demand. Cloud providers typically use a "pay-as-you-go" model, which can lead to unexpected operating expenses if administrators are not familiarized with cloud-pricing models

Since the launch of Amazon EC2 in 2006, the availability of high-capacity networks, low-cost computers and storage devices as well as the widespread adoption of hardware virtualization, service-oriented architecture, and autonomic and utility computing has led to growth in cloud computing.

2. Pre-Requisites

- Computer Networks
- Network Security
- Distributed Computing

3. Course Objectives:

- To understand Virtualization, Virtual Machine and different models of VM.
- To familiarize Cloud computing architecture and its security aspects.

4. Course Outcomes:

At the end of the course, students will be able to

CO1: Know about basics of cloud computing.
CO2: Cloud computing and its services available today.
CO3: Distinguish Virtualization and Virtual Machine and its need, Types of Virtualization.
CO4: Understand how to provide security for the cloud.
CO5: Understand disaster recovery and disaster management.
CO6: Design a Cloud for an Enterprise.

5. Program Outcomes:

Graduates of the Computer Science and Engineering Program will have ability to

a. apply knowledge of computing, mathematics, science and engineering fundamentals to solve complex engineering problems.
b. formulate and analyze a problem, and define the computing requirements appropriate to its solution using basic principles of mathematics, science and computer engineering.
c. design, implement, and evaluate a computer based system, process, component, or software to meet the desired needs.
d. design and conduct experiments, perform analysis and interpretation of data and provide valid conclusions.
e. use current techniques, skills, and tools necessary for computing practice.
f. understand legal, health, security and social issues in Professional Engineering practice.
g. understand the impact of professional engineering solutions on environmental context and the need for sustainable development.

h. understand the professional and ethical responsibilities of an engineer.

i. function effectively as an individual, and as a team member/ leader in accomplishing a common goal.

j. communicate effectively, make effective presentations and write and comprehend technical reports and publications.

k. learn and adopt new technologies, and use them effectively towards continued professional development throughout the life.

l. understand engineering and management principles and their application to manage projects in the software industry.

6. Mapping of Course Outcomes with Program Outcomes:

	a	b	c	d	e	f	g	h	i	j	k	l
CO1											H	
CO2		M									L	
CO3	M				M						M	
CO4						M						
CO5					M							
CO6	H	H	M									

7. Prescribed Text Books

1. Michael Miller, Cloud Computing – Web Based Applications That change the way you work and Collaborate Online –Person Education.
2. George Reese Cloud Application Architectures, Ist Edition O'Reilly Media.

8. Reference Text Books

1. David S. Linthicum, Cloud Computing and SOA Convergence in your Enterprise : A Step-by-Step Guide- Addison-Wesley Professional.

2. Kai Hwang, GeofferyC.Fox, Jack J, Dongarra, Distributed & Cloud Computing From Parallel Processing to the Internet of Things.

9. URLs and Other E-Learning Resources

URLs:

- https://www.edureka.co/cloud-computing-certification-courses
- https://www.getmeacourse.com/?query=Cloud%20Computing

- https://www.coursera.org/courses?query=cloud%20computing
- https://onlinecourses.nptel.ac.in/noc17_cs23/preview

10. Lecture Schedule / Lesson Plan

Topic	No. of Periods	
	Theory	**Tutorial**
UNIT - I: Cloud computing		
Introduction	2	2
what it is and what it isn't		
from collaborations to cloud- a short history of cloud computing	1	
Client/Server, P2P, Distributed computing, Collaborative computing, Cloud computing	1	
the network is the computer- How cloud computing works	1	
Cloud Architecture, Cloud storage, Cloud Services	1	
companies in the cloud- Cloud computing today	1	
	7	
UNIT - II: Ready for Computing in the cloud		
The pros and cons of Cloud Computing	1	2
Developing Cloud Services- Why Develop Web-Based Applications	1	
The Pros and Cons of Cloud Service Development	1	
Types of Cloud Service Development	1	
SaaS, PaaS, web services, On-demand computing	1	
Discovering Cloud Services Development services and Tools	1	
Amazon, Google App Engine, IBM, Salesforce.com	1	
	7	
UNIT - III: Virtualization		
Virtualization for cloud	1	2
Need for Virtualization – Pros and cons of Virtualization		
Types of Virtualization	1	
System VM, Process VM	1	
Virtual machine properties	1	
Interpretation and binary translation	1	
HLL VM Hypervisors – Xen	1	
KVM , VMWare	1	
Virtual Box, Hyper-V	1	
	8	
UNIT - IV: Security		
Data Security	1	2
Data Control Encrypt Everything		
Regulatory and Standards compliances	1	

Network Security, Firewall rules, Network Intrusion detection	2	
Host Security, System Hardening	1	
Antivirus Protection, Host Intrusion detection	1	
Data segmentation, Credential Management	1	
	7	
UNIT - V:Disaster		
What is Disaster	1	2
Disaster Recovery Planning		
The Recovery Point objective, The Recovery Time Objective	1	
Disasters in the Cloud		
Backups and data retention	1	
Geographic redundancy, Organizational redundancy	1	
Disaster Management	1	
Monitoring		
Load Balancer Recovery, Application server recovery	1	
Database Recovery	1	
	7	
UNIT - VI: Defining Clouds for the Enterprise		
Storage-as-a-Service, Database-as-a- Service	2	2
Information-as-a-Service, Process as-a-Service	1	
Application-as-a- Service, Platform-as-a-Service	1	
Integration-as-a Service, Security-as-a-Service	1	
Management/Governance-as-a-Service, Testing as-a-Service	2	
Infrastructure-as- a-Service	1	
	8	
Total No.of Periods:	**44**	**12**

UNIT-I
CLOUD COMPUTING

1. CLOUD COMPUTING

Cloud computing portends a major change in how we store information and run applications. Instead of running programs and data on an individual desktop computer, everything is hosted in the “cloud”—a nebulous assemblage of computers and servers accessed via the Internet. Cloud computing lets you access all your applications and documents from anywhere in the world, freeing you from the confines of the desktop and making it easier for group members in different locations to collaborate.

1.1 INTRODUCTION:

- Traditional desktop computing, you run copies of software programs on each computer you own. The documents you create are stored on the computer on which they were created. Although documents can be accessed from other computers on the network, they can’t be accessed by computers outside the network.

- The whole scene is PC-centric. With cloud computing, the software programs you use aren’t run from your personal computer, but are rather stored on servers accessed via the Internet. If your computer crashes, the software is still available for others to use. Same goes for the documents you create; they’re stored on a collection of servers accessed via the Internet.

1.2 WHAT IT IS AND WHAT IT ISN'T:

What Cloud Computing Is

- Key to the definition of cloud computing is the “cloud” itself. For our purposes, the cloud is a large group of interconnected computers. These computers can be personal computers or network servers; they can be public or private.

- This cloud of computers extends beyond a single company or enterprise. The applications and data served by the cloud are available to broad group of users, cross-enterprise and cross-platform. Access is via the Internet. Any authorized user can access these docs and apps from any computer over any Internet connection

  - **Cloud computing is user-centric**: Once you as a user are connected to the cloud, whatever is stored there—documents, messages, images, applications, whatever—becomes yours. In addition, not only is the data yours, but you can also share it with others.

  - **Cloud computing is task-centric**. Instead of focusing on the application and what it can do, the focus is on what you need done and how the application can do it for you., Traditional applications—word processing, spreadsheets, email, and so on—are becoming less important than the documents they create.

  - **Cloud computing is powerful**. Connecting hundreds or thousands of computers together in a cloud creates a wealth of computing power impossible with a single desktop PC.

  - **Cloud computing is accessible**. Because data is stored in the cloud, users can instantly retrieve more information from multiple repositories. You're not limited to a single source of data, as you are with a desktop PC.

  - **Cloud computing is intelligent**. With all the various data stored on the computers in a cloud, data mining and analysis are necessary to access that information in an intelligent manner.

- **Cloud computing is programmable**. Many of the tasks necessary with cloud computing must be automated. For example, to protect the integrity of the data, information stored on a single computer in the cloud must be replicated on other computers in the cloud. If that one computer goes offline, the cloud's programming automatically redistributes that computer's data to a new computer in the cloud.

What Cloud Computing Isn't

- **Cloud computing isn't network computing.** With network computing, applications/documents are hosted on a single company's server and accessed over the company's network. Cloud computing is a lot bigger than that of network computing.

- Cloud computing also isn't traditional outsourcing, where a company farms out (subcontracts) its computing services to an outside firm. While an outsourcing firm might host a company's data or applications, those documents and programs are only accessible to the company's employees via the company's network, not to the entire world via the Internet.

1.3 FROM COLLABORATIONS TO CLOUD- A SHORT HISTORY OF CLOUD COMPUTING

Cloud computing has as its antecedents both client/server computing and peer-to-peer distributed computing.

Client/Server Computing: Centralized Applications and Storage:

- All the software applications, all the data, and all the control resided on huge mainframe computers, otherwise known as servers. If a user wanted to access specific data or run a program, he had to connect to the mainframe, gain appropriate access, and then do his business while essentially "renting" the program or data from the server

- Users connected to the server via a computer terminal, sometimes called a workstation or client. This computer was sometimes called a dumb terminal because it didn't have a lot (if any!) memory, storage space, or processing power.

- the client/server model, while providing similar centralized storage, differed from cloud computing in that it did not have a user-centric focus; with client/server computing, all the control rested with the mainframe—and with the guardians of that single computer. It was not a user-enabling environment

Peer-to-Peer Computing

- The server part of the system also created a huge bottleneck. All communications between computers had to go through the server first, however inefficient that might be. The obvious need to connect one computer to another without first hitting the server led to the development of peer-to-peer (P2P) computing.

- P2P computing defines a network architecture in which each computer has equivalent capabilities and responsibilities. This is in contrast to the traditional client/server network architecture, in which one or more computers are dedicated to serving the others.

- P2P was an equalizing concept. In the P2P environment, every computer is a client and a server; there are no masters and slaves. By recognizing all computers on the network as peers, P2P enables direct exchange of resources and services.

- There is no need for a central server, because any computer can function in that capacity when called on to do so. P2P was also a decentralizing concept. Control is decentralized, with all computers functioning as equals. No centralized server is assigned to host the available resources and services.
- The users' connection to the Usenet server was of the traditional client/server nature, the relationship between the Usenet servers was definitely P2P—and presaged the cloud computing of today.

Distributed Computing: Providing More Computing Power

- When a computer is enlisted for a distributed computing project, software is installed on the machine to run various processing activities during those periods when the PC is typically unused.
- Many distributed computing projects are conducted within large enterprises, using traditional network connections to form the distributed computing network. Other, larger, projects utilize the computers of everyday Internet users, with the computing typically taking place offline, and then uploaded once a day via traditional consumer Internet connections.

Collaborative Computing: Working as a Group

- Early group collaboration was enabled by the combination of several different P2P technologies. The goal was (and is) to enable multiple users to collaborate on group projects online, in real time.

- To collaborate on any project, users must first be able to talk to one another. In today's environment, this means instant messaging for text-based communication, with optional audio/telephony and video capabilities for voice and picture communication. Most collaboration systems offer the complete range of audio/video options, for full-featured multiple-user video conferencing.

Cloud Computing: The Next Step in Collaboration

- With the growth of the Internet, there was no need to limit group collaboration to a single enterprise's network environment. Users from multiple locations within a corporation, and from multiple organizations, desired to collaborate on projects that crossed company and geographic boundaries. To do this, projects had to be housed in the "cloud" of the Internet, and accessed from any Internet-enabled location.

- The concept of cloud-based documents and services took wing with the development of large server farms, such as those run by Google and other search companies. Google already had a collection of servers that it used to power its massive search engine; why not use that same computing power to drive a collection of web-based applications—and, in the process, provide a new level of Internet-based group collaboration.

1.4 THE NETWORK IS THE COMPUTER:HOW CLOUD COMPUTING WORKS

- Network of computers functions as a single computer to serve data and applications to users over the Internet. The network exists in the “cloud” of IP addresses that we know as the Internet, offers massive computing power and storage capability, and enables wide scale group collaboration.

1.5 Understanding Cloud Architecture:

- “cloud”—a massive network of servers or even individual PCs interconnected in a grid. These computers run in parallel, combining the resources of each to generate supercomputing-like power.
- The cloud is a collection of computers and servers that are publicly accessible via the Internet. This hardware is typically owned and operated by a third party on a consolidated basis in one or more data center locations.
- As shown in Figure 1.1, individual users connect to the cloud from their own personal computers or portable devices, over the Internet. To these individual users, the cloud is seen as a single application, device, or document. The hardware in the cloud is invisible.

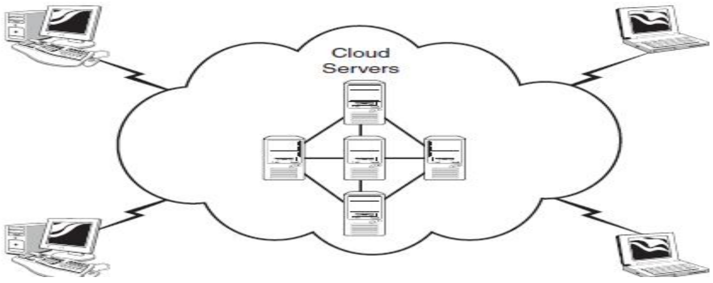


fig1.1: How user connect to the cloud

- This cloud architecture is deceptively simple, although it does require someintelligent management to connect all those computers together and assigntask processing to multitudes of users.

- In Figure 1.2, it all startswith the front-end interface seen by individual users. This is how users select atask or service. The user's request then gets passed to the system management, which finds the correct resources and then calls the system's appropriate provisioning services. These services carve out the necessary resources in the cloud, launch the appropriate web application, and either creates or opens the requested document. After the web application is launched, the system's monitoring and metering functions track the usage of the cloud so that resources are apportioned and attributed to the proper users.

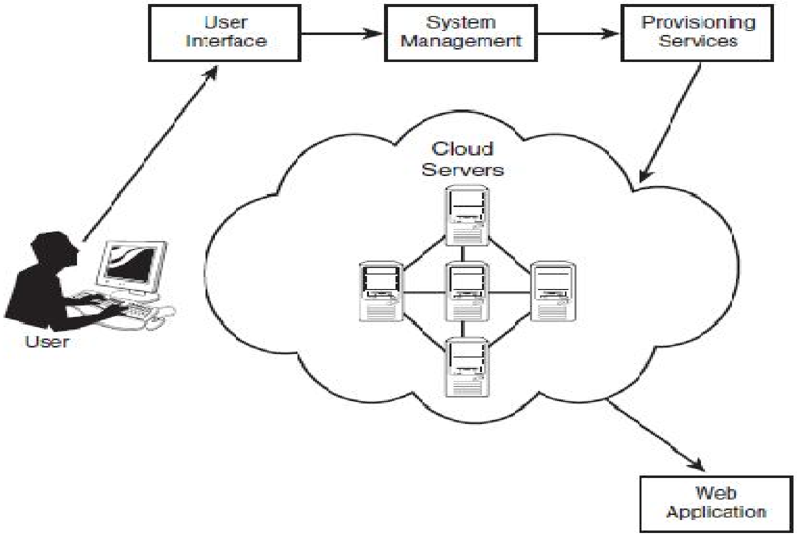


Fig 1.2: The architecture behind a cloud computing system.

Understanding Cloud Storage

- When storing data, the user sees a virtual server—that is, it appears as if the data is stored in a particular place with a specific name. But that place doesn't exist in reality. It's just a pseudonym used to reference virtual space carved out of the cloud. In reality, the user's data could be stored on any one or more of the computers used to create the cloud. The actual storage location may even differ from day to day or even minute to minute, as the cloud dynamically manages available storage space

- Cloud storage has both financial and security-associated advantages. Financially, virtual resources in the cloud are typically cheaper than dedicated physical resources connected to a personal computer or network. As for security, data stored in the cloud is secure from accidental erasure or hardware crashes, because it is duplicated across multiple physical machines; since multiple copies of the data are kept continually, the cloud continues to function as normal even if one or more machines go offline.

Understanding Cloud Services

- Cloud services can include anything from calendar and contact applications to word processing and presentations. Almost all large computing companies today, from Google to Amazon to Microsoft, are developing various types of cloud services.

- Cloud services offer many advantages. If the user's PC crashes, it doesn't affect either the host application or the open document; both remain unaffected in the cloud. In addition, an individual user can access his applications and documents from any location on any PC.

1.6 COMPANIES IN THE CLOUD: CLOUD COMPUTING TODAY

- Cloud computing today is attracting the best and biggest companies from across the computing industry, all of whom hope to establish profitable business models based in the cloud.

- The most noticeable company currently embracing the cloud computing model is Google. As you'll see throughout this book, Google offers a powerful collection of web-based applications, all served via its cloud architecture. Whether you want cloud-based word processing (Google Docs), presentation software (Google Presentations), email (Gmail), or calendar/scheduling functionality (Google Calendar), Google has an offering.

- Other major companies are also involved in the development of cloud services. Microsoft, for example, offers its Windows Live suite of web-based applications, as well as the Live Mesh initiative that promises to link together all types of devices, data, and applications in a common cloud-based platform. Amazon has its Elastic Compute Cloud (EC2), a web service that provides cloud-based resizable computing capacity for application developers.

- IBM has established a Cloud Computing Center to deliver cloud services and research to clients. And numerous smaller companies have launched their own web based applications, primarily (but not exclusively) to exploit the collaborative nature of cloud services.

UNIT-I

Assignment-Cum-Tutorial Questions

SECTION-A

Objective Questions

1. With cloud computing the software programs run on ________ accessed via internet. [ ]
   a. Serversb. Private computers c. Network servers d. all the above
2. Cloud computing is PC-centric. [TRUE/FALSE]
3. Networking computing and outsourcing are not cloud computing

   [TRUE/FALSE]

4. The cloud is a large group of interconnected computers. These computers are_______. [ ]
   a. Personal b. Network Servers c. Public or Private d. All the above
5. From Google's perspective the key properties of cloud computing are ______. [ ]
   i. cloud computing is user-centric.
   ii. cloud computing is task-centric.
   iii. cloud computing is powerful.
   iv. cloud computing is accessible.
   v. cloud computing is intelligent.
   vi. cloud computing is programmable.
   a. both i & ii b. both iii & iv c. both v & vi d. all the above
6. The Google applications that are popular today are__________ . [ ]
   a. Google docs b. Google calendar c. Gmail d. All the above
7.In P2P computing each computer has equivalent capabilities and responsibilities [TRUE/FALSE]
8.Distributed computing is all about ___________ between multiple computers [ ]
   a. Cycle sharing b. File sharing c. Providing internet d. None of the above
9. Cloud in cloud computing represents _____________? [ ]
   a) Wireless b) Hard drives c) People d) Internet
10.Which of these is not a cloud computing pricing model. [ ]
   a)Free b)Pay per use c) Subscription d) Ladder
11. What is/are the key characteristics of cloud computing? [ ]

a) Service offering b) Reliability c) Scalability d) ALL

12. The term __________ has been used historically as a Metaphor for the internet? [ ]

a) Cloud b) Intranet c) grid computing d) None of the above

13.Which one is delivering software services to end users and running code?

a) SOA b) Grid c) Cloud d) None [ ]

14. Which of the following is an example of cloud computing application

a) Facebook Apps b) Twitter or RSS c) Salesforce.com d) Skype[ ]

15. What is Grid computing? [ ]

a) It is a network of computers that share resources – the network can be local or distributed across the internet. Hardware as a service

b) It is a physical arrangement of computer terminals that optimizes computing power – the computers in the center are more powerful.

c) It is a temporary clod computer network that only exists as long as single project is active.

d) All the above.

16. What is an important benefit of cloud. [ ]

a) Highly protected data b) Independent from Internet

c) Reduced cost d) Small bandwidth

17. What is not a valid reason for customer asking a clod provider where there servers are loacated? [ ]

a) Geographical location may tell something about network latency.

b) Geographical location may tell something about network legislation.

c) The number of sites tells you something about disaster recovery possibilities.

d) When a server breaks down, the customer wants to send a technician to fix the problem as soon as possible

18.Which cloud deployment model is operated solely for a single organization and its authorized users. [ ]

a) Community cloud b) Hybrid cloud

c) Public cloud d) Private cloud

19. Which cloud deployment model is managed by a cloud provider, has an infrastructure that is offsite, and is accessible to general public [ ]

a) Community cloud b) Hybrid cloud

c) Public cloud d) Private cloud

SECTION-B

SUBJECTIVE QUESTIONS

1. Define Cloud Computing? Enlist and explain essential characteristics of Cloud Computing?
2. Explain how cloud computing works?
3. Differentiate peer to peer computing and distributed computing?
4. Explain Collaboration to cloud?
5. Explain about cloud application architectures?
5. Enlist various companies in providing cloud computing services.
6. Write a short note on the next step in collaboration?
7. Explain how cloud computing is different from cloud computing?
8. Write a short note on Cloud Storage?
9. Write a short note on cloud services?
10. Explain why cloud computing is important?
11. Explain the architecture behind a cloud computing system?

UNIT-II

CLOUD COMPUTING

2.1 The Pros and Cons of Cloud Computing

Any serious analysis of cloud computing must address the advantages and disadvantages offered by this burgeoning technology.

Cloud Computing: Advantages

- ❖ **Lower-Cost Computers for Users**
  - ✓ Here's a quantitative financial advantage: You don't need a high-powered computer to run cloud computing web-based applications. Because the application runs in the cloud, not on the desktop PC, that desktop PC doesn't need the processing power or hard disk space demanded by traditional desktop software.
  - ✓ Hence the client computers in cloud computing can be lower priced, with smaller hard disks, less memory, more efficient processors, and the like.
  - ✓ A client computer in this scenario wouldn't even need a CD or DVD drive, because no software programs have to be loaded and no document files need to be saved.
- ❖ **Improved Performance**
  - ✓ When a desktop PC doesn't have to store and run a ton of software-based applications, with fewer bloated programs hogging the computer's memory, users will see better performance from their PCs.
  - ✓ Computers in a cloud computing system will boot up faster and run faster, because they'll have fewer programs and processes loaded into memory.

- ❖ **Lower IT Infrastructure Costs**
  - ✓ In a larger organization, the IT department could also see lower costs from the adoption of the cloud computing paradigm. Instead of investing in larger numbers of more powerful servers, The IT staff can use the computing power of the cloud to supplement or replace internal computing resources.
  - ✓ Those companies that have peak needs no longer have to purchase equipment to handle the peaks (and then lay fallow the rest of the time);
  - ✓ Peak computing needs are easily handled by computers and servers in the cloud.
- ❖ **Fewer Maintenance Issues**
  - ✓ Speaking of maintenance costs, cloud computing greatly reduces both hardware and software maintenance for organizations of all sizes.
  - ✓ First, the hardware with less hardware (fewer servers) necessary in the organization, maintenance costs are immediately lowered.
  - ✓ All cloud apps are based elsewhere, so there's no Software on the organization's computers for the IT staff to maintain.
- ❖ **Lower Software Costs**
  - ✓ Then there's the issue of software cost. Instead of purchasing separate software packages for each computer in the organization, only those employees actually using an application need access to that application in the cloud.
  - ✓ Even if it costs the same to use web-based applications as it does similar desktop software (which it probably won't), IT staffs are saved the cost of installing and maintaining those programs on every desktop in the organization.
  - ✓ As to the cost of that software, it's possible that some cloud computing companies will charge as much to "rent" their apps

as traditional software companies charge for software purchases.

- ✓ Cloud services will be priced substantially lower than similar desktop software. In fact, many companies (such as Google) are offering their web-based applications for free—which to both individuals and large organizations is much more attractive than the high costs charged by Microsoft and similar desktop software suppliers.

- ❖ **Instant Software Updates**
  - ✓ Another software-related advantage to cloud computing is that users are no longer faced with the choice between obsolete software and high upgrade costs.
  - ✓ When the app is web-based, updates happen automatically and are available the next time the user logs in to the cloud.
  - ✓ Whenever you access a web-based application, you're getting the latest version—without needing to pay for or download an upgrade.

- ❖ **Increased Computing Power**
  - ✓ This is an obvious one. When you're tied into a cloud computing system, you have the power of the entire cloud at your disposal.
  - ✓ You're no longer limited to what a single desktop PC can do, but can now perform supercomputing-like tasks utilizing the power of thousands of computers and servers.
  - ✓ We can attempt greater tasks in the cloud than we can on our desktop.

- ❖ **Unlimited Storage Capacity**
  - ✓ The cloud offers virtually limitless storage capacity.
  - ✓ Consider that when your desktop or laptop PC is running out of storage space. Your computer's 200GB hard drive is peanuts

compared to the hundreds of petabytes (a million gigabytes) available in the cloud.

- ✓ Whatever you need to store, you can.

❖ **Increased Data Safety**

- ✓ And all that data you store in the cloud?
- ✓ In desktop computing, where a hard disk crash can destroy all your valuable data, a computer crashing in the cloud doesn’t affect the storage of your data.
- ✓ That’s because data in the cloud is automatically duplicated, so nothing is ever lost.
- ✓ That also means if your personal computer crashes, all your data is still out there in the cloud, still accessible. Cloud computing can keep data safe.

❖ **Improved Compatibility Between Operating Systems**

- ✓ Ever try to get a Windows-based computer to talk to a Mac? Or a Linux machine to share data with a Windows PC? It can be frustrating. Not so with cloud computing.
- ✓ In the cloud, operating systems simply don’t matter. You can connect your Windows computer to the cloud and share documents with computers running Apple’s Mac OS, Linux, or UNIX. In the cloud, the data matters, not the operating system.

❖ **Improved Document Format Compatibility**

- ✓ You also don’t have to worry about the documents you create on your machine being compatible with other users’ applications or operating systems.
- ✓ In a world where Word 2007 documents can’t be opened on a computer running Word 2003, all documents created by web-based applications can be read by any other user accessing that application.
- ✓ There are no format incompatibilities when everyone is sharing docs and apps in the cloud.

- **Easier Group Collaboration**
  - ✓ Sharing documents leads directly to collaborating on documents. To many users, this is one of the most important advantages of cloud computing
  - ✓ The ability for multiple users to easily collaborate on documents and projects.
  - ✓ Imagine that you, a colleague in your West Coast office, and a consultant in Europe all need to work together on an important project.
  - ✓ Before cloud computing, you had to email or snail mail the relevant documents from one user to another, and work on them sequentially.
  - ✓ In cloud computing, now each of you can access the project's documents simultaneously; the edits one user makes are automatically reflected in what the other users see onscreen. It's all possible, of course, because the documents are hosted in the cloud, not on any of your individual computers.
  - ✓ We need is a computer with an Internet connection, and we're collaborating. Of course, easier group collaboration means faster completion of most group projects, with full participation from all involved.
  - ✓ It also enables group projects across different geographic locations. No longer does the group have to reside in a single office for best effect.
  - ✓ With cloud computing, anyone anywhere can collaborate in real time. It's an enabling technology.
- **Universal Access to Documents**
  - ✓ And here's another document-related advantage of cloud computing.
  - ✓ When you edit a document at home, that edited version is what you see when you access the document at work.

- ✓ The cloud always hosts the latest version of your documents; you're never in danger of having an outdated version on the computer you're working on.

❖ **Removes the Tether to Specific Devices**

- ✓ Finally, the cloud computing advantage—you're no longer tethered to a single computer or network. Change computers, and your existing applications and documents follow you through the cloud.
- ✓ Move to a portable device, and your apps and docs are still available. There's no need to buy a special version of a program for a particular device, or save your document in a device-specific format.
- ✓ Your documents and the programs that created them are the same no matter what computer you're using.

Cloud Computing: Disadvantages

There are a number of reasons why you might not want to adopt cloud computing for your particular needs. Let's examine a few of the risks related to cloud computing.

➢ **Requires a Constant Internet Connection**

- ✓ Cloud computing is, quite simply, impossible if you can't connect to the Internet. Because you use the Internet to connect to both your applications and documents, if you don't have an Internet connection, you can't access anything, even your own documents.
- ✓ A dead Internet connection means no work, period—and in areas where Internet connections are few or inherently unreliable.
- ✓ When you're offline, cloud computing just doesn't work. This might be a more significant disadvantage.

- ✓ Sure, you’re used to a relatively consistent Internet connection both at home and at work, but where else do you like to use your computer? If you’re used to working on documents on your deck, or while you’re at a restaurant for lunch, or in your car, you won’t be able to access your cloud based documents and applications—unless you have a strong Internet connection at all those locations.
- ✓ A lot of what’s nice about portable computing becomes problematic when you’re depending on web-based applications.

➢ **Doesn’t Work Well with Low-Speed Connections**

- ✓ Similarly, a low-speed Internet connection, such as that found with dial-up services, makes cloud computing painful at best and often impossible.
- ✓ Web based apps often require a lot of bandwidth to download, as do large documents.
- ✓ If you’re laboring with a low-speed dial-up connection, it might take seemingly forever just to change from page to page in a document, let alone launch a feature-rich cloud service. In other words, cloud computing isn’t for the slow or broadband-impaired.

➢ **Can Be Slow**

- ✓ Even on a fast connection, web-based applications can sometimes be slower than accessing a similar software program on your desktop PC.
- ✓ That’s because everything about the program, from the interface to the document you’re working on, has to be sent back and forth from your computer to the computers in the cloud.
- ✓ If the cloud servers happen to be backed up at that moment, or if the Internet is having a slow day, you won’t

get the instantaneous access you're used to with desktop apps.

- **Features Might Be Limited**
  - ✓ This particular disadvantage is bound to change, but today many web-based applications simply aren't as full-featured as their desktop-based brethren.
  - ✓ For example, the feature set of Google Presentations with that of Microsoft PowerPoint; there's just a lot more you can do with PowerPoint than you can with Google's web-based offering.
  - ✓ The basics are similar, but the cloud application lacks many of PowerPoint's advanced features.
  - ✓ So if you're an advanced user, you might not want to leap into the cloud computing waters just yet.
  - ✓ That said, many web-based apps add more advanced features over time. This has certainly been the case with Google Docs and Spreadsheets, both of which started out somewhat crippled but later added Many of the more niche functions found on Microsoft Word and Excel.
  - ✓ Still, you need to look at the features before you make the move. Make sure that the cloud-based application can do everything you need it to do before you give up on your traditional software.
- **Stored Data Might Not Be Secure**
  - ✓ With cloud computing, all your data is stored on the cloud. That's all well and good, but how secure is the cloud?
  - ✓ Can other, unauthorized users gain access to your confidential data? These are all important questions, and well worth further examination.
  - ✓ We examine just how safe our data is in the cloud.

- **If the Cloud Loses Your Data, You’re screwed**
  - ✓ Data stored in the cloud is unusually safe, replicated across multiple machines.
  - ✓ But on the off chance that your data does go missing, you have no physical or local backup. (Unless you methodically download all your cloud documents to your own desktop, of course which few users do.) Put simply, relaying the cloud puts you at risk if the cloud lets you down.
- **Developing Cloud Service**
  - ✓ Cloud computing from a user’s perspective, focusing on those web-based applications that owe their existence to the cloud.
  - ✓ But cloud computing also offers a lot to software developers, who can now develop web-based applications that take advantage of the power and reach of cloud computing.
  - ✓

2.2 Why Develop Web-Based Applications?

- ✓ The needs of a typical IT department are daunting: They must deliver adequate computing power and data storage to all users within the company. This must be done, of course, within a set budget, to meet peak needs or to add capacity for new users can often send an IT budget soaring.
- ✓ For most companies, it is not financially prudent to add capacity that will be used only a small percentage of the time.
- ✓ What the IT department needs is a way to increase capacity or add capabilities without investing in new servers and networking gear, or licensing new software.

- ✓ Cloud services, in the form of centralized web-based applications, also appeal to the IT professional. One instance of an application hosted in the cloud is cheaper and easier to manage than individual copies of similar software installed on each user's desktop PC.
- ✓ Upgrading a cloud app only has to be done one time, where upgrading traditional software has to be done for each PC on which that software is installed.
- ✓ The advantages of cloud services development are particularly notable to smaller businesses, which otherwise wouldn't have the budget or resources to develop large-scale applications. By hosting locally developed web applications within the cloud, the small business avoids the cost of purchasing expensive hardware to host similar sofware.
- ✓ Most small companies don't have the staff, resources, hardware, or budget to develop and maintain their own applications, or to deal with the rigors of maintaining secure environments. Although they could outsource their software development and hosting, moving those applications to the cloud, companies don't have to invest in locally hosted systems, freeing up their staff and resources to focus on the day-to-day running of their own businesses.

- ✓ A company that develops its own web-based applications gains functionality while reducing expenses. The combined power of the cloud is accompanied by lower software purchase and management costs.

2.3 The Pros and Cons of Cloud Service Development

Why would you choose to develop new applications using the cloud services model? There are several good reasons to do—and a few reasons.

- ➢ **Advantages of Cloud Development**

- One of the underlying advantages of cloud development is that of economy of scale. By taking advantage of the infrastructure provided by a cloud computing vendor, a developer can offer better, cheaper, and more reliable applications than is possible within a single enterprise.
- Speaking of cost, because cloud services follow the one-to-many model, cost is significantly reduced over individual desktop program deployment. Instead of purchasing or licensing physical copies of software programs, cloud applications are typically "rented," priced on a per-user basis.
- It's more of a subscription model than an asset purchase model, which means there is less up-front investment and a more predictable monthly expense stream.
- IT departments like cloud applications because all management activities are managed from a central location rather than from individual sites or workstations.
- This enables IT staff to access applications remotely via the web. There's also the advantage of quickly outfitting users with the software they need (known as "rapid provisioning), and adding more computing resources as more users tax the system (automatic scaling).
- When you need more storage space or bandwidth, companies can just add another virtual server from the cloud. It's a lot easier than purchasing, installing, and configuring a new server in their data center.
- For developers, it's also easier to upgrade a cloud application than with traditional desktop software. Application features can be quickly and easily updated by upgrading the

centralized application, instead of manually upgrading individual applications located on each and every desktop PC in the organization. With a cloud service, a single change affects every user running the application, which greatly reduces the developer’s workload.

- **Disadvantages of Cloud Development**

  - Perhaps the biggest perceived disadvantage of cloud development is the same one that plagues all web-based applications: Is it secure? Web-based applications have long been considered potential security risks.

  - There have been few instances of data loss with cloud-hosted applications and storage. It could even be argued that a large cloud hosting operation is likely to have better data security and redundancy tools than the average enterprise. In any case, however, even the perceived security danger from hosting critical data and services offsite might discourage some companies from going this route.

  - Another potential disadvantage is what happens if the cloud computing host goes offline. Although most companies say this isn’t possible, it has happened.

  - Amazon’s EC2 service suffered a massive outage on February 15, 2008, that wiped out some customer application data. (The outage was caused by a software deployment that erroneously terminated an unknown number of user instances.) For clients expecting a safe and secure platform, having that platform go down and your data disappear is a somewhat rude awakening. And, if a company relies on a third-party cloud platform to host all of its data with no other physical backup, that data can be at risk.

2.4 Types of Cloud Service Development

The concept of cloud services development encompasses several different types of development. Let's look at the different ways a company can use cloud computing to develop its own business applications.

➢ **Software as a Service**

- Software as a service, or SaaS, is probably the most common type of cloud service development. With SaaS, a single application is delivered to thousands of users from the vendor's servers. Customers don't pay for owning the software; rather, they pay for using it. Users access an application via an API accessible over the web.
- Each organization served by the vendor is called a tenant, and this type of arrangement is called a multitenant architecture. The vendor's servers are *virtually partitioned* so that each organization works with a customized virtual application instance.
- For customers, SaaS requires no upfront investment in servers or software licensing. For the application developer, there is only one application to maintain for multiple clients.
- Many different types of companies are developing applications using the SaaS model.The best-known SaaS applications are those offered by Google to its consumer base.

➢ **Platform as a Service**

✓ In this variation of SaaS, the development environment is offered as a service. The developer uses the "building blocks" of the vendor's development environment to create his own custom application.

- ✓ It's kind of like creating an application using Legos; building the app is made easier by use of these redefined blocks of code, even if the resulting app is somewhat constrained by the types of code blocks available.

➢ **Web Services**

- A web service is an application that operates over a network—typically, over the Internet. A web service is an API that can be accessed over the Internet. The service is then executed on a remote system that hosts the requested services.

- This type of web API the developers exploit shared functionality over the Internet, rather than deliver their own full-blown applications. The result is a customized web-based application where a large hunk of that application is delivered by a third party, thus easing development and bandwidth demands for the custom program.

- Good example of web services are the "mashups" created by users of the Google Maps API.

- With these custom apps, the data that feeds the map is provided by the developer, where the engine that creates the map itself is provided by Google.

- The developer doesn't have to code or serve a map application; all he has to do is hook into Google's web API.

- The advantages of web services include faster (and lower-cost) application development, leaner applications, and reduced storage and bandwidth demands.

- Web services keep developers from having to reinvent the wheel every time they develop a new application. By reusing code from the web services provider, they get a jump-start on the development of their own applications.

➢ **On-Demand Computing**

- As the name implies, on-demand computing packages computer resources (processing, storage, and so forth) as a metered service similar to that of a public utility.
- Here, the customers pay for as much or as little processing and storage as they need.
- Companies that have large demand peaks followed by much lower normal usage periods particularly benefit from utility computing. The company pays more for their peak usage, of course, but their bills rapidly decline when the peak ends and normal usage patterns resume.
- Clients of on-demand computing services essentially use these services as offsite virtual servers. Instead of investing in their own physical infrastructure, a company operates on a pay-as-you-go plan with a cloud services provider.
- On-demand computing itself is not a new concept, but has acquired new life thanks to cloud computing.
- In previous years, on-demand computing was provided from a single server via some sort of time-sharing arrangement.
- Today, the service is based on large grids of computers operating as a single cloud.

2.5 Discovering Cloud Services Development Services and Tools

- As we are aware, cloud computing is at an early stage of its development. This can be seen by observing the large number of small and start-up companies offering cloud development tools. In an established industry, the smaller players eventually fall by the wayside as larger companies take center stage.
- Cloud services development services and tools are offered by a variety of companies, both large and small.
- The more fully featured offerings include development tools and pre-built applications that developers can use as the building blocks for their own unique web-based applications.

➢ **Companies and services**.

- **Amazon**

Amazon, one of the largest retailers on the Internet, is also one of the primary providers of cloud development services.

Amazon has spent a lot of time and money setting up a multitude of servers to service its popular website, and is making those vast hardware resources available for all developers to use.

The service in question is called the Elastic Compute Cloud, also known as EC2. This is a commercial web service that allows developers and companies to rent capacity on Amazon's proprietary cloud of server, which happens to be one of the biggest server farms in the world.

EC2 enables scalable deployment of applications by letting customers request a set number of virtual machines, onto which they can load any application of their choice. so, customers can create, launch, and terminate server instances on demand, creating a truly "elastic" operation.

Amazon's service lets customers choose from three sizes of virtual servers:

Small: This offers the equivalent of a system with 1.7GB of memory, 160GB of storage, and one virtual 32-bit core processor.

Large: This offers the equivalent of a system with 7.5GB of memory, 850GB of storage, and two 64-bit virtual core processors

Extra large: This offers the equivalent of a system with 15GB of memory, 1.7TB of storage, and four virtual 64-bit core processor.

EC2 is just part of Amazon's Web Services (AWS) set of offerings, which provides developers with direct access to Amazon's software and machines. By tapping into the computing power that Amazon has already constructed, developers can build reliable, powerful, and low-cost web-based applications.

Amazon provides the cloud and developers provide the rest. They pay only for the computing power that they use. AWS is perhaps the most popular cloud computing service to date. Amazon claims a market of more than 330,000 customers—a combination of developers, start-ups, and established companies.

- **Google App Engine**
  - Google is a leader in web-based applications, so it's not surprising that the company also offers cloud development services. These services come in the form of the Google App Engine, which enables developers to build their own web applications utilizing the same infrastructure that powers Google's powerful applications.
  - The Google App Engine provides a fully integrated application environment. Using Google's development tools and computing cloud, App Engine applications are easy to build, easy to maintain, and easy to scale.

- All you have to do is develop your application (using Google's APIs and the Python programming language) and upload it to the App Engine cloud; from there, it's ready to serve your users.
- Google offers a robust cloud development environment.
    - It includes the following features:
        - Dynamic web serving.
        - Full support for all common web technologies
        - Persistent storage with queries, sorting, and transactions
        - Automatic scaling and load balancing
        - APIs for authenticating users and sending email using Google Accounts
- In addition, Google provides a fully featured local development environment that simulates the Google App Engine on any desktop computer. And here's one of the best things about Google's offering, unlike most other cloud hosting solutions, Google App Engine is completely free to use.
- A free App Engine account gets up to 500MB of storage and enough CPU strength and bandwidth for about 5 million page views a month. If you need more storage, power, or capacity, Google intends to offer additional resources in the future.

➢ **IBM**

- IBM is offering a cloud computing solution. The company is targeting small- and medium-sized businesses with a suite of cloud-based on demand services via its Blue Cloud initiative.
- Blue Cloud is a series of cloud computing offerings that enables enterprises to distribute their computing needs across a globally accessible resource grid. One such offering is the Express Advantage suite, which includes data backup and recovery, email continuity and archiving, and data security functionality,

some of the more data-intensive processes handled by a typical IT department.

- To manage its cloud hardware, IBM provides open source workload-scheduling software called Hadoop, which is based on the MapReduce software used by Google in its offerings. Also included are PowerVM and Xen virtualization tools, along with IBM's Tivoli data center management software.

- **Salesforce.com**

- Salesforce.com is probably best known for its sales management SaaS, but it's also a leader in cloud computing development. The company's cloud computing architecture is dubbed Force.com.
- The platform as a service is entirely on-demand, running across the Internet. Sales force provides its own Force.com API and developer's toolkit. Pricing is on a per log-in basis. Supplementing Force.com is AppExchange, a directory of web-based applications.
- Developers can use AppExchange applications uploaded by others, share their own applications in the directory, or publish private applications accessible only by authorized companies or clients.
- Many applications in the AppExchange library are free, and others can be purchased or licensed from the original developers. Not unexpectedly, most existing AppExchange applications are sales related sales analysis tools, email marketing systems, financial analysis apps, and so forth.
- But companies can use the Force.com platform to develop any type of application. In fact, many small businesses have already jumped on the Force.com bandwagon.
- For example, an April 2008 article in *PC World* magazine quoted Jonathan Snyder, CTO of Dreambuilder Investments, a 10-

person mortgage investment company in New York."We're a small company," Snyder said, "We don't have the resources to focus on buying servers and developing from scratch. For us, Force.com was really a jump-start."

- **Other Cloud Services Development Tools**

Amazon, Google, IBM, and Salesforce.com aren't the only companies offering tools for cloud services developers. There are also a number of smaller companies working in this space that developers should evaluate, and that end users may eventually become familiar with.

These companies include the following:

- ✓ _ 3tera (www.3tera.com), which offers the AppLogic grid operating system and Cloudware architecture for on-demand computing.
- ✓ _ 10gen (www.10gen.com), which provides a platform for developers to build scalable web-based applications.
- ✓ Cohesive Flexible Technologies (www.cohesiveft.com), which offers the Elastic Server On-Demand virtual server platform.
- ✓ Joyent (www.joyent.com), which delivers the Accelerator scalable on demand infrastructure for web application developers, as well as the Connector suite of easy-to-use web applications for small businesses.
- ✓ Mosso (www.mosso.com), which provides an enterprise-level cloud hosting service with automatic scaling.
- ✓ _ Nirvanix (www.nirvanix.com), which offers a cloud storage platform for developers, as well as Nirvanix Web Services, which provides file management and other common operations via a standards-based API.

- ✓ Skytap (www.skytap.com), which provides the Virtual Lab on-demand web-based automation solution that enables developers to build and configure lab environments using pre-configured virtual machines.

- ✓ _ StrikeIron (www.strikeiron.com), which offers the IronCloud cloud based platform for the delivery of web services, along with various Live Data services that developers can integrate into their own applications.

- ✓ In addition, Sun Microsystems has an R&D project, dubbed Project Caroline (www.projectcaroline.net), that provides an open source hosting platform for the development and delivery of web-based applications. Access to Project Caroline's grid is free to the general public.

UNIT-II

Assignment-Cum-Tutorial Questions

SECTION-A

Objective Questions

1.Cloud computing offers __________ for users. [ ]
(a) Lower-software Costs (b) Fewer Maintenance issues
(c) Improved Performance (d) All the above

2.Cloud computing allows to migrate to a portable device, and your apps and docs are still available. [TRUE/ FALSE]

3.Some web based applications are now being designed to work on your desktop when not connected to internet. [TRUE/FALSE]

4. _______ is a web-based technology that turns Google's applications into locally run applications.
(a) Google gears (b) Google Drive (c) Google Maps (d) Picasa [ ]

5. In computing, a web application or web app is a client server application in which the user interface runs in _____________. [ ]

(a) Web browser (b) server (c) cloud servers (d) local machines

6.It is easier to upgrade a cloud application than with traditional desktop software because, ___________. [ ]

(a) it is easy and quick (b). it is cheap (c) reduces the cost (d) All the above

7.If a company relies on a third-party cloud platform to host all of its data with no physical backup, the data can be at risk because _______. [ ]

(a) Providing safe and secure platform is a challenging task.

(b) Cloud might go offline

(c) Maintaining critical data is very difficult

(d) All the above

8.With ________servers, a single application is delivered to thousands of users from the vendor. [ ]

(a) SaaS (b) Paas (c) Iaas (d). None of the above

9.With ______ servers, the development environment is offered as a service.

(a) SaaS (b) Paas (c) Iaas (d). None of the above. [ ]

10. __________ is an API that can be accessed over the Internet. [ ]

(a) web services

(b) remote services

(c) on-demand services

(d) None of the above

11.________ is also known as utility computing. [ ]
(a) on-demand computing (b) parallel computing
(c)distributed computing (d). none

12.On-demand computing and storage are offered by _______ . [ ]
(a) Amazon (b) IBM (c) Sun (d) All the above

13.______ is probably best known for its sales management SaaS. [ ]
(a) salesforce.com (b) Amazon (c) IBM (d) All the above

14.Most existing AppExchange applications are _______ related. [ ]
(a) sales (b) manufacturing (c) business (d) service

15._________ is/ are the companies offering tools for cloud services developers. [ ]
(a) Amazon (b) Google (c) IBM (d) All the above

16.10gen provides platform for developers to build __________________applications . [ ]
(a) AppLogic (b) infrastructure-based
(c) scalable web-based (d) None

17.StrikeIron offers Iron Cloud based platform for the delivery of ________
(a) web services (b). client services [ ]
(c) storage services (d). enterprise services

18.Which is not a type of cloud service development ? [ ]
(a) Software as a Service (b) Platform as a Service
(c) Web Services (d) Compatible Service

19.Which is not considered as one of the three main categories of cloud services ? [ ]
(a) Software as a service (b) Database as a service
(c) Platform as a service (d) Infrastructure as a service

20.Which cloud service is also known as hardware as a service? [ ]
(a) Software as a service (b) Desktop as a service
(c) Platform as a service (d) Infrastructure as a service

SECTION-B

SUBJECTIVE QUESTIONS

1.Describe the pros and cons of cloud computing?

2.What are the advantages of cloud computing?

3.What are the disadvantages of cloud computing?

4.What are benefits of SaaS over Traditional Applications?

5.Explain in brief Software as a service and Platform as a service?

6.Write a short note on the Cloud Services Development Services and Tools provided by Amazon.

7.Why develop web-based applications?

8.What are the advantages of cloud development?

9.What are the disadvantages of cloud development?

10.What are the types of cloud service development?

11.Write a short note on the variety of companies that offers cloud services development services and tools.

12.Explain the layered cloud service model.

Unit- III

Virtualization

3.1 Virtualization for Cloud

- Virtualization was first introduced in the 1960s by IBM to boost utilization of large, expensive mainframe systems by partitioning them into logical, separate virtual machines that could run multiple applications and processes at the same time. In the 1980s and 1990s, this centrally shared mainframe model gave way to a distributed, client-server computing model, in which many low-cost x86 servers and desktops independently run specific applications.
- When people talk about virtualization, they usually refer to server virtualization, which means partitioning one physical server into several virtual servers, or machines. Each virtual machine can interact independently with other devices, applications, data and users as though it were a separate physical resource.
- Different virtual machines can run different operating systems and multiple applications while sharing the resources of a single physical computer. And, because each virtual machine is isolated from other virtualized machines, if one crashes, it doesn't affect the others.
- Hypervisor software is the secret sauce that makes virtualization possible. This software, also known as a virtualization manager, sits between the hardware and the operating system, and decouples the operating system and applications from the hardware. The hypervisor assigns the amount of access that the operating systems and applications have with the processor and other hardware resources, such as memory and disk input/output.
- In addition to using virtualization technology to partition one machine into several virtual machines, you can also use virtualization solutions to combine multiple physical resources into a single virtual resource. A good example of this is storage virtualization, where multiple network storage resources are pooled into what appears as a single

storage device for easier and more efficient management of these resources. Other types of virtualization you may hear about include:

- Network virtualization splits available bandwidth in a network into independent channels that can be assigned to specific servers or devices.
- Application virtualization separates applications from the hardware and the operating system, putting them in a container that can be relocated without disrupting other systems.
- Desktop virtualization enables a centralized server to deliver and manage individualized desktops remotely. This gives users a full client experience, but lets IT staff provision, manage, upgrade and patch them virtually, instead of physically.
- While virtualization faded from the limelight for a while, it is now one of the hottest trends in the industry again, as organizations aim to increase the utilization, flexibility and cost-effectiveness in a distributed computing environment. VMWare, Citrix, Microsoft, IBM, RedHat and many other vendors offer virtualization solutions.

3.2 Need for Virtualization

- Virtualization can help you shift your IT focus from managing boxes to improving the services you provide to the organization. If you are managing multiple servers and desktops, virtualization can help you to:
- Save money. Companies often run just one application per server because they don’t want to risk the possibility that one application will crash and bring down another on the same machine. Estimates indicate that most x86 servers are running at an average of only 10 to 15 percent of total capacity. With virtualization, you can turn a single purpose server into a multi-tasking one, and turn multiple servers into a computing pool that can adapt more flexibly to changing workloads.

- Save energy. Businesses spend a lot of money powering unused server capacity. Virtualization reduces the number of physical servers, reducing the energy required to power and cool them.
- Save time. With fewer servers, you can spend less time on the manual tasks required for server maintenance. On the flip side, pooling many storage devices into a single virtual storage device, you can perform tasks such as backup, archiving and recovery more easily and more quickly. It's also much faster to deploy a virtual machine than it is to deploy a new physical server.
- Reduce desktop management headaches. Managing, securing and upgrading desktops and notebooks can be a hassle. Desktop virtualization solutions let you manage user desktops centrally, making it easier to keep desktops updated and secure.

What to Consider

- Since virtualization makes it easy to set up new virtual servers, you may end up with a lot of servers to manage. Each server needs to be managed just as if it was a physical server. Keeping track of where everything and how your virtual resources are using physical resources is vital, so shop for solutions that have easy-to-use tools that help you monitor and measure use.
- Virtualization isn't a magic bullet for everything. While many solutions are great candidates for running virtually, applications that need a lot of memory, processing power or input/output may be best left on a dedicated server.
- For all of the upside virtualization isn't magic, and it can introduce some new challenges. But in most cases the many cost and efficiency advantages will outweigh any issues, and virtualization will continue to grow gain popularity.

Advantages of Virtualization

- **Better utilization of existing resources**

  Physical computer resources have become advanced and powerful with time in traditional computing. One machine instance runs on a physical server with hardly utilizes the whole power of the system and does most of the processing power simply remain unutilized for most of the computer system. Running multiple virtual machines on a physical server makes better utilization of the resources and this is known as server consolidation.

- **Reduction in hardware cost**

  Virtualization makes better use of physical resources by running multiple virtual machine on single physical resources.Automatically cost of computing comes down if server consolidation can be combined with capacity planning and management of hardware resources reduced drastically as well.

- **Reduction in computing infrastructure costs**

  Reduced physical computing resource requirements introduces any other associated assets like physical floor space power requirement cooling system and Hardware resources to administrate the system increase in the number of virtual machines over existing physical resources do not add to any of these loads.

- **Improved fault tolerance or zero downtime maintenance**

  The decoupling of virtual machines from specific hardware resources increases the portability of system. In cases of any hardware failure the virtual system can be migrated to another physical setup. This helps to build fault tolerant system by creating scope of a zero downtime maintenance.

- **Simplified system administration**

  Virtualization segments the management of the systems into two groups as physical resources management and virtual system management. Centralized monitoring package can be employed to keep track of health of the systems and raise alert in case of need. As managing virtual computing resources is less critical than physical resources and they are less physical machines, the system administration tasks become easier.

- **Simplified capacity expansion**

  Capacities of virtual resources are easier to increase than expanding and then synchronizing physical computing resources. This also becomes possible due to the decoupling of physical resources from virtual systems.

- **Simplified system installation**

  Simplified system installation of a new system has become easier cost effective and Hassle-free in virtual environment. A new system can be installed almost within no time by cloning a virtual machine instance fresh installation is much easier too than physical machine installation.

- **Support for legacy systems and applications**

  Research institutes or other organizations sometimes need to run legacy software packages which can no longer be run on physically available operating system or hardware platforms. Virtualization is the only solution to tackle this kind of scenarios. As any kind of VM's can easily be created to deploy required Legacy application over them.

- **Simplified system-level development**

  System software development and testing require frequent rebooting of the system this is easier and faster in virtual environment since VM rebooting does not require physical machine to restart and can be performed with some clicks of mouse

- **Simplified system and application testing**

  Performance testing of system software before its release is a rigorous job and requires to test the software on all supported platforms this is a difficult situation since it requires all of those hardware platforms and operating systems to test the software. Virtualization simplifies this process by eliminating much of the time and effort required for system installation and configuration application or system software testing is one of the biggest beneficiaries of virtualization

- **Security**

  Virtualization as a layer of abstraction or physical hardware virtual machines cannot directly access physical resources anymore this can restrict the amount of destruction that might occur when some malicious software attempts to damage the system or corrupt data for example if an entire virtual hard disk gets damaged or corrupted the actual physical disk remains unaffected

Downsides of Virtualization

- **Single point of failure problem**

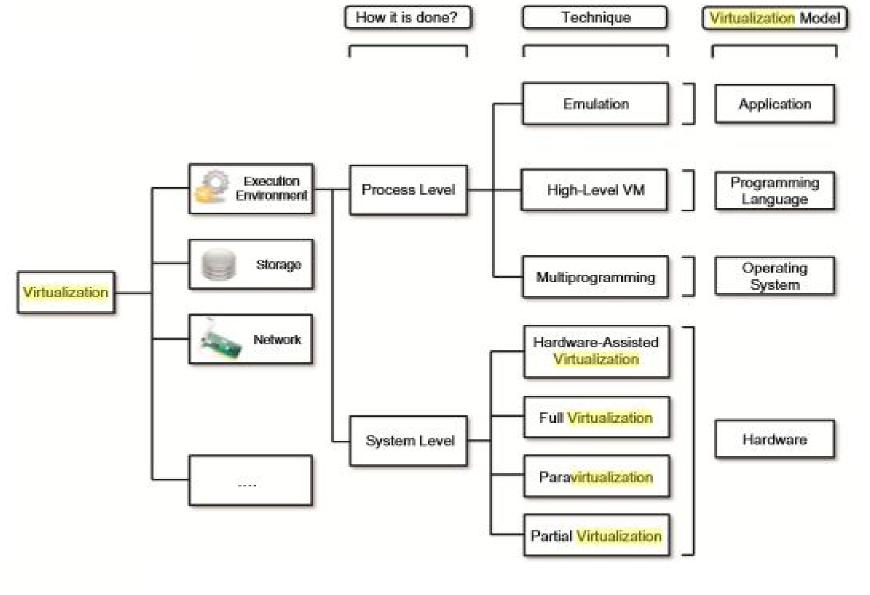
  The major benefit of virtualization is resource sharing multiple virtual machines can run over a physical machine but this has a downside it increases the probability of failure of a number of virtual servers in case of failure of single physical machine although this situation can be handled easily by keeping backup resources and putting those virtual server on the backup set top physical resources voting is not a difficult task as virtualization decouples virtual systems from physical resources
- **Lower performance issue**

  There is a concern weather virtual environment have the capacity to accomplish the full performance of the actual physical system it has been seen that virtual servers can achieve up to 85% to 90% of the performance of the actual physical server as VMS cannot get Direct Access to the hardware
- **Difficulty in root cause analysis**

  With virtualization a new layer of complexity is added which can cause new problems the main difficulty is that if something does not work as it is supposed to it may require considerable extra efforts to find the cause of the problem.

3.3 Types of Virtualization

A virtual machine (VM) is an emulation of a computer system. Virtual machines are based on computer architectures and provide functionality of a physical computer. Their implementations may involve specialized hardware, software, or a combination.



A taxonomy of virtualization techniques.

There are different kinds of virtual machines, each with different functions:

i. System virtual machines (also termed full virtualization VMs) provide a substitute for a real machine. They provide functionality needed to execute entire operating systems. A hypervisor uses native execution to share and manage hardware, allowing for multiple environments which are isolated from one another, yet exist on the same physical machine. Modern hypervisors use hardware-assisted virtualization, virtualization-specific hardware, primarily from the host CPUs.

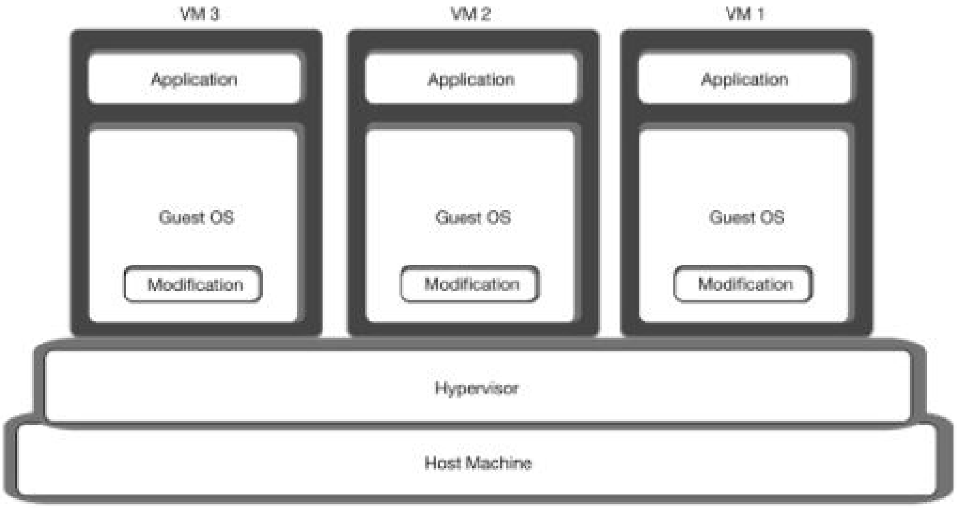
ii. Process virtual machines are designed to execute computer programs in a platform-independent environment.

System virtual machines

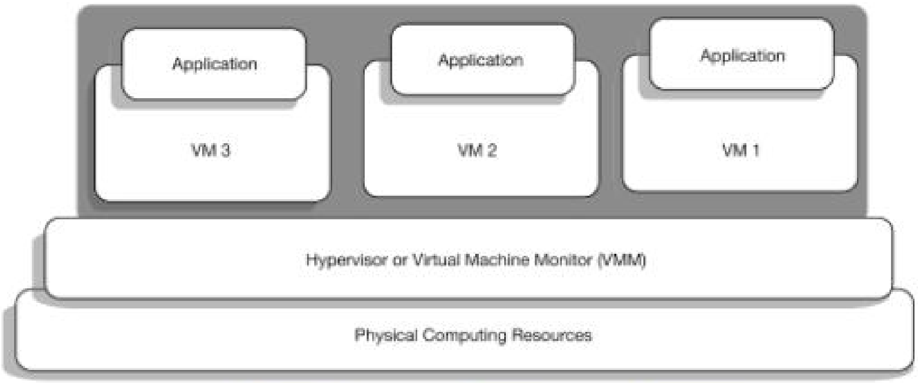
- The desire to run multiple operating systems was the initial motive for virtual machines, so as to allow time-sharing among several single-tasking operating systems.
- In some respects, a system virtual machine can be considered a generalization of the concept of virtual memory that historically preceded it.
- IBM's CP/CMS, the first systems to allow full virtualization, implemented time sharing by providing each user with a single-user operating system, the Conversational Monitor System (CMS).
- Unlike virtual memory, a system virtual machine entitled the user to write privileged instructions in their code.
- This approach had certain advantages, such as adding input/output devices not allowed by the standard system.
- As technology evolves virtual memory for purposes of virtualization, new systems of memory over commitment may be applied to manage memory sharing among multiple virtual machines on one computer operating system.
- It may be possible to share memory pages that have identical contents among multiple virtual machines that run on the same physical machine, what may result in mapping them to the same physical page by a technique termed Kernel Same Page Merging.
- This is especially useful for read-only pages, such as those holding code segments, which is the case for multiple virtual machines running the same or similar software, software libraries, web servers, middleware components, etc.
- The guest operating systems do not need to be compliant with the host hardware, thus making it possible to run different operating systems on the same computer (e.g., Windows, Linux,

or prior versions of an operating system) to support future software.

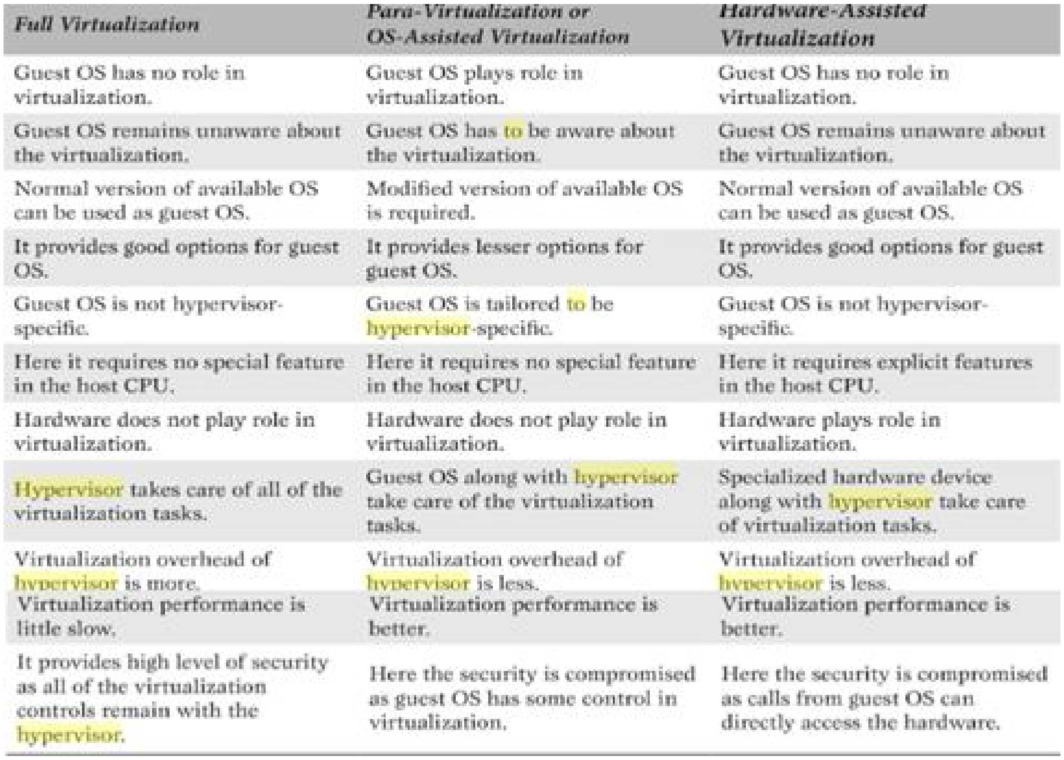
- The use of virtual machines to support separate guest operating systems is popular in regard to embedded systems. A typical use would be to run a real-time operating system simultaneously with a preferred complex operating system, such as Linux or Windows. Another use would be for novel and unproven software still in the developmental stage, so it runs inside a sandbox.
- Virtual machines have other advantages for operating system development, and may include improved debugging access and faster reboots.
- Multiple VMs running their own guest operating system are frequently engaged for server consolidation.



A model of para-virtualization


Application
Application
Application
VM 3
VM 2
VM 1
Hypervisor or Virtual Machine Monitor (VMM)
Physical Computing Resources

A model for the bare metal approach of machine virtualization



Full Virtualization	*Para-Virtualization or OS-Assisted Virtualization*	*Hardware-Assisted Virtualization*
Guest OS has no role in virtualization.	Guest OS plays role in virtualization.	Guest OS has no role in virtualization.
Guest OS remains unaware about the virtualization.	Guest OS has to be aware about the virtualization.	Guest OS remains unaware about the virtualization.
Normal version of available OS can be used as guest OS.	Modified version of available OS is required.	Normal version of available OS can be used as guest OS.
It provides good options for guest OS.	It provides lesser options for guest OS.	It provides good options for guest OS.
Guest OS is not hypervisor-specific.	Guest OS is tailored to be hypervisor-specific.	Guest OS is not hypervisor-specific.
Here it requires no special feature in the host CPU.	Here it requires no special feature in the host CPU.	Here it requires explicit features in the host CPU.
Hardware does not play role in virtualization.	Hardware does not play role in virtualization.	Hardware plays role in virtualization.
Hypervisor takes care of all of the virtualization tasks.	Guest OS along with hypervisor take care of the virtualization tasks.	Specialized hardware device along with hypervisor take care of virtualization tasks.
Virtualization overhead of hypervisor is more.	Virtualization overhead of hypervisor is less.	Virtualization overhead of hypervisor is less.
Virtualization performance is little slow.	Virtualization performance is better.	Virtualization performance is better.
It provides high level of security as all of the virtualization controls remain with the hypervisor.	Here the security is compromised as guest OS has some control in virtualization.	Here the security is compromised as calls from guest OS can directly access the hardware.

Partial virtualization:

- Partial virtualization provides a partial emulation of the underlying hardware, thus not allowing the complete execution of the guest operating system in complete isolation.
- Partial virtualization allows many applications to run transparently, but not all the features of the operating system can be supported, as happens with full virtualization.

- An example of partial virtualization is address space virtualization used in time-sharing systems; this allows multiple applications and users to run concurrently in a separate memory space, but they still share the same hardware resources.

Process virtual machines

- A process VM, sometimes called an application virtual machine, or Managed Runtime Environment (MRE), runs as a normal application inside a host OS and supports a single process. It is created when that process is started and destroyed when it exits. Its purpose is to provide a platform-independent programming environment that abstracts away details of the underlying hardware or operating system, and allows a program to execute in the same way on any platform.

Operating system-level virtualization

- Operating system-level virtualization offers the opportunity to create different and separated execution environments for applications that are managed concurrently.
- Differently from hardware virtualization, there is no virtual machine manager or hypervisor, and the virtualization is done within a single operating system, where the OS kernel allows for multiple isolated user space instances.
- The kernel is also responsible for sharing the system resources among instances and for limiting the impact of instances on each other.
- A user space instance in general contains a proper view of the file system, which is completely isolated, and separate IP addresses, software configurations, and access to devices.
- Operating systems supporting this type of virtualization are general-purpose, timeshared operating systems with the capability to provide stronger namespace and resource isolation.

- This virtualization technique can be considered an evolution of the chroot mechanism in Unix systems.
- The chroot operation changes the file system root directory for a process and its children to a specific directory. As a result, the process and its children cannot have access to other portions of the file system than those accessible under the new root directory.
- Because Unix systems also expose devices as parts of the file system, by using this method it is possible to completely isolate a set of processes.
- Following the same principle, operating system-level virtualization aims to provide separated and multiple execution containers for running applications.
- Compared to hardware virtualization, this strategy imposes little or no overhead because applications directly use OS system calls and there is no need for emulation.
- There is no need to modify applications to run them, nor to modify any specific hardware, as in the case of hardware-assisted virtualization.
- On the other hand, operating system-level virtualization does not expose the same flexibility of hardware virtualization, since all the user space instances must share the same operating system.

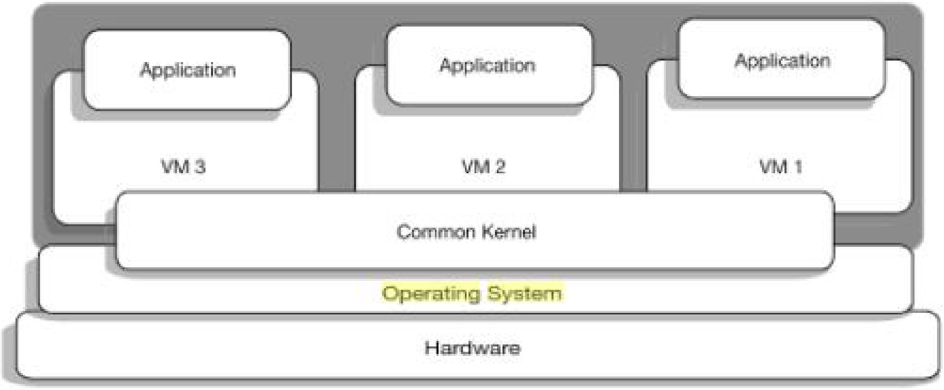


FIG 7.8. A model of operating system-level virtualization approach

Programming language-level virtualization

- Programming language-level virtualization is mostly used to achieve ease of deployment of applications, managed execution, and portability across different platforms and operating systems.
- It consists of a virtual machine executing the byte code of a program, which is the result of the compilation process. Compilers implemented and used this technology to produce a binary format representing the machine code for an abstract architecture.
- The characteristics of this architecture vary from implementation to implementation. Generally, these virtual machines constitute a simplification of the underlying hardware instruction set and provide some high-level instructions that map some of the features of the languages compiled for them.
- At runtime, the byte code can be either interpreted or compiled on the underlying hardware instruction set.
- Virtual machine programming languages become popular again with Sun’s introduction of the Java platform in 1996. Originally created as a platform for developing Internet applications, Java became one of the technologies of choice for enterprise applications, and a large community of developers formed around it.

- The Java virtual machine was originally designed for the execution of programs written in the Java language, but other languages such as Python, Pascal, Groovy, and Ruby were made available.
- The ability to support multiple programming languages has been one of the key elements of the Common Language Infrastructure (CLI), which is the specification behind .NET Framework.
- Currently, the Java platform and .NET Framework represent the most popular technologies for enterprise application development.
- The main advantage of programming-level virtual machines, also called process virtual machines, is the ability to provide a uniform execution environment across different platforms.
- Programs compiled into byte code can be executed on any operating system and platform for which a virtual machine able to execute that code has been provided.

Application-level virtualization

- Application-level virtualization is a technique allowing applications to be run in runtime environments that do not natively support all the features required by such applications. In this scenario, applications are not installed in the expected runtime environment but are run as though they were. In general, these techniques are mostly concerned with partial file systems, libraries, and operating system component emulation. Such emulation is performed by a thin layer—a program or an operating system component—that is in charge of executing the application. Emulation can also be used to execute program binaries compiled for different hardware architectures. In this case, one of the following strategies can be implemented:
    - **Interpretation.** In this technique every source instruction is interpreted by an emulator for executing native ISA instructions, leading to poor performance.

  Interpretation has a minimal startup cost but a huge overhead, since each instruction is emulated.

  - **Binary translation**. In this technique every source instruction is converted to native instructions with equivalent functions. After a block of instructions is translated, it is cached and reused. Binary translation has a large initial overhead cost, but over time it is subject to better performance, since previously translated instruction blocks are directly executed.

- Emulation, as described, is different from hardware-level virtualization. The former simply allows the execution of a program compiled against a different hardware, whereas the latter emulates a complete hardware environment where an entire operating system can be installed.

- Application virtualization is a good solution in the case of missing libraries in the host operating system; in this case a replacement library can be linked with the application, or library calls can be remapped to existing functions available in the host system.

- Another advantage is that in this case the virtual machine manager is much lighter since it provides a partial emulation of the runtime environment compared to hardware virtualization. Moreover, this technique allows incompatible applications to run together.

- Compared to programming-level virtualization, which works across all the applications developed for that virtual machine, application-level virtualization works for a specific environment: It supports all the applications that run on top of a specific environment.

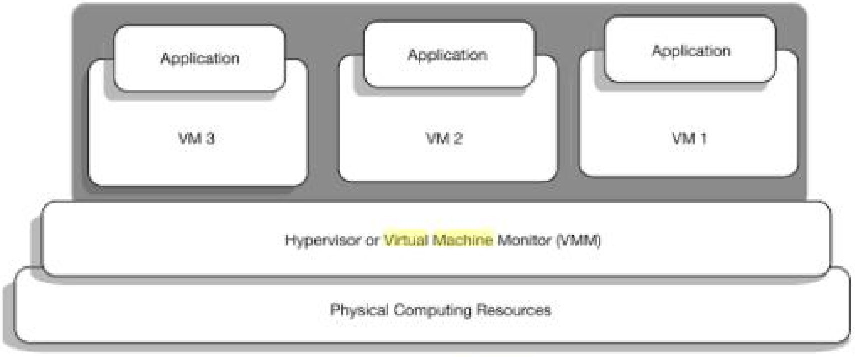
- One of the most popular solutions implementing application virtualization is Wine, which is a software application allowing Unix-like operating systems to execute programs written for the Microsoft Windows platform.

Virtual Machine Monitor

- Hypervisor producers all of the virtual computing resources. Virtual machines are created over the layer of hypervisor using this virtual resources it is the responsibility of the hypervisor to manage activities, monitor functionalities and provide support to virtual machines. This is why hypervisor is also called as Virtual Machine monitor or VMM.
- Hypervisors A fundamental element of hardware virtualization is the hypervisor, or virtual machine manager (VMM). It recreates a hardware environment in which guest operating systems are installed. There are two major types of hypervisor:

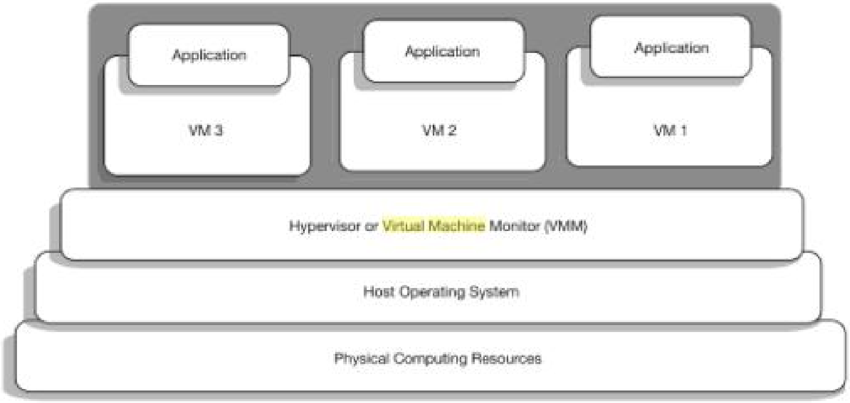
Type I and Type II

• Type I hypervisors run directly on top of the hardware. Therefore, they take the place of the operating systems and interact directly with the ISA interface exposed by the underlying hardware, and they emulate this interface in order to allow the management of guest operating systems. This type of hypervisor is also called a native virtual machine since it runs natively on hardware.


Application
Application
Application
VM 3
VM 2
VM 1
Hypervisor or Virtual Machine Monitor (VMM)
Physical Computing Resources

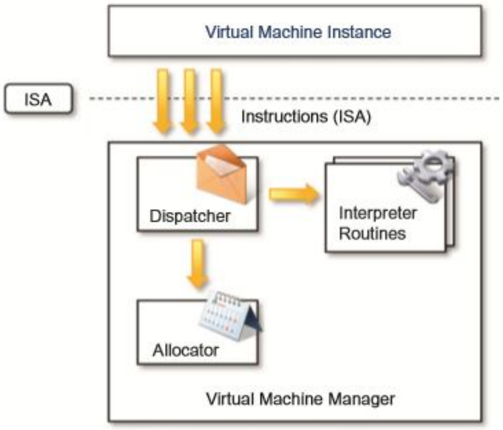
A model for the bare metal approach of machine virtualization

• Type II hypervisors require the support of an operating system to provide virtualization services. This means that they are programs managed by the operating system, which interact with it through the ABI and emulate the ISA of virtual hardware for guest operating systems. This type of hypervisor is also called a hosted virtual machine since it is hosted within an operating system.


Application
Application
Application
VM 3
VM 2
VM 1
Hypervisor or Virtual Machine Monitor (VMM)
Host Operating System
Physical Computing Resources

A model of hosted machine virtualization approach

- Three main modules, dispatcher, allocator, and interpreter, coordinate their activity in order to emulate the underlying hardware.
- The dispatcher constitutes the entry point of the monitor and reroutes the instructions issued by the virtual machine instance to one of the two other modules.
- The allocator is responsible for deciding the system resources to be provided to the VM: whenever a virtual machine tries to execute an instruction that results in changing the machine resources associated with that VM, the allocator is invoked by the dispatcher.
- The interpreter module consists of interpreter routines. These are executed whenever a virtual machine executes a privileged instruction: a trap is triggered and the corresponding routine is executed.
- The design and architecture of a virtual machine manager, together with the underlying hardware design of the host machine, determine the full realization of hardware virtualization, where a guest operating system can be transparently executed on top of a VMM as though it were run on the underlying hardware.



A hypervisor reference architecture.

Three properties have to be satisfied:

• Equivalence. A guest running under the control of a virtual machine manager should exhibit the same behavior as when it is executed directly on the physical host.

• Resource control. The virtual machine manager should be in complete control of virtualized resources.

• Efficiency. A statistically dominant fraction of the machine instructions should be executed without intervention from the virtual machine manager.

The major factor that determines whether these properties are satisfied is represented by the layout of the ISA of the host running a virtual machine manager. Popek and Goldberg provided a classification of the instruction set and proposed three theorems that define the properties that hardware instructions need to satisfy in order to efficiently support virtualization.

HLL VM-Hypervisors

- The high level language is a concept that goes against the idea of conventional computing environment where a compiled application family tied to a particular OS and ISA (Instruction Set Architecture) that in conventional computing environment reporting of application to different computing platform requires Re compilation of the application code for the targeted platform moreover coating of application needs coating of the underlying compiler for the target platform first which is a tedious technical task

- The high level language eases the porting of compiler by rendering HLL to intermediate representation targeted towards abstract machines. The abstract machine then translates the intermediate code to physical machine instruction set. This concept of abstract machine makes the coating of compilers less complex has only the back and part of compiler needs to be posted since the intermediate code representation is same for compilers of HLL on any platform
- Does the intermediate representation of a shell program having compiled over 1 physical computing architecture can be posted on abstract machine running over other architecture search abstract machine is referred as high level language VM or HLL VM java Virtual Machine JVM and Microsoft common line runtime CLR are two examples of high level language VMs
- High level language VM is also known as application VM or process VM application process virtualization can be considered as the smaller version of machine virtualization instead of virtual machines based Technology decouples application software from the underlying platform on which it is executed
- The best known exams this type of virtualization is Java Virtual Machine (JVM), JVM makes the execution of Java applications machine independent as the applications do not interact with underline OS and Hardware platform directly rather than JVM creates a uniform Virtual Machine environment and makes Java applications portable
- Microsoft has even adopted a similar approach in common language runtime CLR used by .NET applications the concept of process virtualization was actually pioneered the UCSDP-system during late

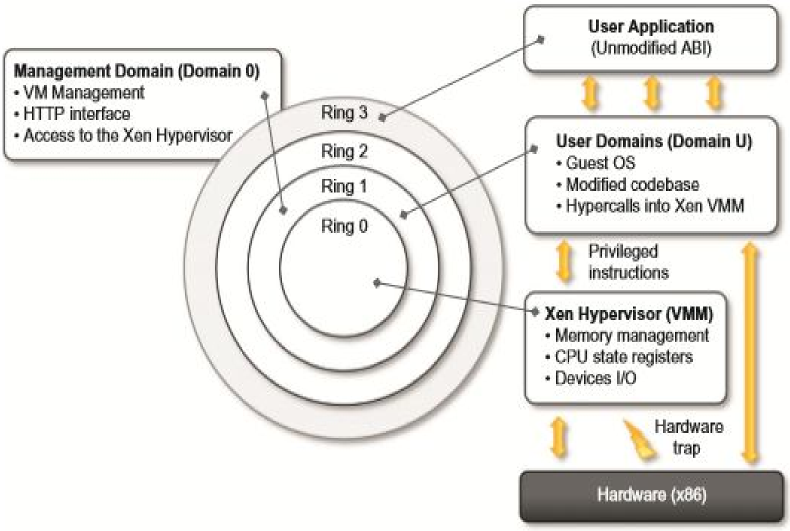
1970 which was developed by a team at the University of California at San Diego, United States.

- Application Virtual Machine places a sort of software wrapper around the application running over it and this rapper restrict the application from directly interacting with system this is referred to as sandbox the sandbox technology is often used to test unverified programs which may contain malicious code in order to prevent them from damaging the underlying hardware devices

Xen:para-virtualization

- Xen is an open-source initiative implementing a virtualization platform based on paravirtualization. Initially developed by a group of researchers at the University of Cambridge in the United Kingdom, Xen now has a large open-source community backing it.
- Citrix also offers it as a commercial solution, XenSource. Xen-based technology is used for either desktop virtualization or server virtualization, and recently it has also been used to provide cloud computing solutions by means of Xen Cloud Platform (XCP).
- At the basis of all these solutions is the Xen Hypervisor, which constitutes the core technology of Xen.
- Recently Xen has been advanced to support full virtualization using hardware-assisted virtualization. Xen is the most popular implementation of paravirtualization, which, in contrast with full virtualization, allows high-performance execution of guest operating systems.
- This is made possible by eliminating the performance loss while executing instructions that require special management.
- This is done by modifying portions of the guest operating systems run by Xen with reference to the execution of such instructions. Therefore, it is not a transparent solution for implementing virtualization.

- A Xen-based system is managed by the Xen hypervisor, which runs in the highest privileged mode and controls the access of guest operating system to the underlying hardware.



- Guest operating system is executed within domains, which represent virtual machine instances. Moreover, specific control software, which has privileged access to the host and controls all the other guest operating systems, is executed in a special domain called Domain 0.
- This is the first one that is loaded once the virtual machine manager has completely booted, and it hosts a Hyper Text Transfer Protocol (HTTP) server that serves requests for virtual machine creation, configuration, and termination.
- This component constitutes the embryonic version of a distributed virtual machine manager, which is an essential component of cloud computing systems providing Infrastructure-as-a-Service (IaaS) solutions.
- Many of the x86 implementations support four different security levels, called rings, where Ring 0 represent the level with the highest privileges and Ring 3 the level with the lowest ones.

- Almost all the most popular operating systems, except OS/2, utilize only two levels: Ring 0 for the kernel code, and Ring 3 for user application and no privileged OS code.
- This provides the opportunity for Xen to implement virtualization by executing the hypervisor in Ring 0, Domain 0, and all the other domains running guest operating systems—generally referred to as Domain U—in Ring 1, while the user applications are run in Ring 3.
- This allows Xen to maintain the ABI unchanged, thus allowing an easy switch to Xen-virtualized solutions from an application point of view.
- Because of the structure of the x86 instruction set, some instructions allow code executing in Ring 3 to jump into Ring 0 (kernel mode).
- Such operation is performed at the hardware level and therefore within a virtualized environment will result in a trap or silent fault, thus preventing the normal operations of the guest operating system, since this is now running in Ring 1.
- This condition is generally triggered by a subset of the system calls. To avoid this situation, operating systems need to be changed in their implementation, and the sensitive system calls need to be reimplemented with hypercalls, which are specific calls exposed by the virtual machine interface of Xen.
- With the use of hypercalls, the Xen hypervisor is able to catch the execution of all the sensitive instructions, manage them, and return the control to the guest operating system by means of a supplied handler.
- Paravirtualization needs the operating system codebase to be modified, and hence not all operating systems can be used as guests in a Xen-based environment.

- More precisely, this condition holds in a scenario where it is not possible to leverage hardware-assisted virtualization, which allows running the hypervisor in Ring -1 and the guest operating system in Ring 0.
- Therefore, Xen exhibits some limitations in the case of legacy hardware and legacy operating systems.
- In fact, these cannot be modified to be run in Ring 1 safely since their codebase is not accessible and, at the same time, the underlying hardware does not provide any support to run the hypervisor in a more privileged mode than Ring 0.
- Open-source operating systems such as Linux can be easily modified, since their code is publicly available and Xen provides full support for their virtualization, whereas components of the Windows family are generally not supported by Xen unless hardware-assisted virtualization is available.
- It can be observed that the problem is now becoming less and less crucial since both new releases of operating systems are designed to be virtualization aware and the new hardware supports x86 virtualization.

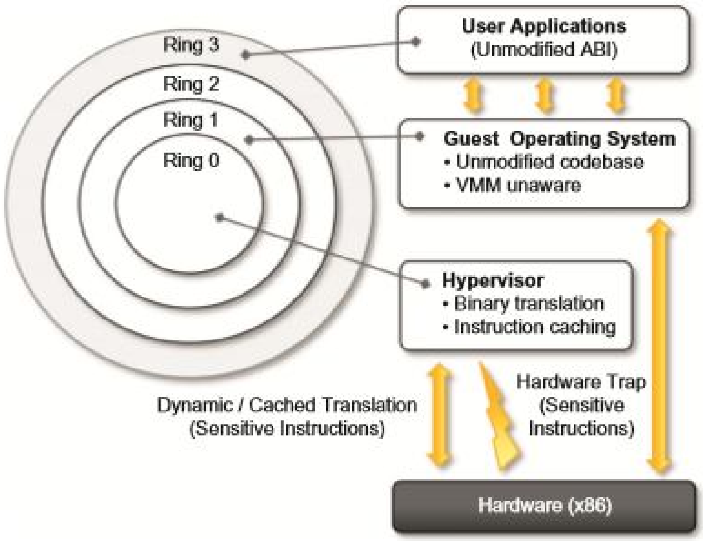
VMware:full virtualization

- VMware's technology is based on the concept of full virtualization, where the underlying hardware is replicated and made available to the guest operating system, which runs unaware of such abstraction layers and does not need to be modified.
- VMware implements full virtualization either in the desktop environment, by means of Type II hypervisors, or in the server environment, by means of Type I hypervisors.
- In both cases, full virtualization is made possible by means of direct execution (for non-sensitive instructions) and binary translation (for

sensitive instructions), thus allowing the virtualization of architecture such as x86.

- Besides these two core solutions, VMware provides additional tools and software that simplify the use of virtualization technology either in a desktop environment, with tools enhancing the integration of virtual guests with the host, or in a server environment, with solutions for building and managing virtual computing infrastructures.
- Full virtualization and binary translation VMware is well known for the capability to virtualize x86 architectures, which runs unmodified on top of their hypervisors.
- With the new generation of hardware architectures and the introduction of hardware-assisted virtualization (Intel VT-x and AMD V) in 2006, full virtualization is made possible with hardware support, but before that date, the use of dynamic binary translation was the only solution that allowed running x86 guest operating systems unmodified in a virtualized environment.
- In the case of dynamic binary translation, the trap triggers the translation of the offending instructions into an equivalent set of instructions that achieves the same goal without generating exceptions. Moreover, to improve performance, the equivalent set of instruction is cached so that translation is no longer necessary for further occurrences of the same instructions.
- This approach has both advantages and disadvantages. The major advantage is that guests can run unmodified in a virtualized environment, which is a crucial feature for operating systems for which source code is not available.
- Binary translation is a more portable solution for full virtualization.
- But, translating instructions at runtime introduces an additional overhead that is not present in other approaches (para-virtualization or hardware-assisted virtualization).

- Even though such disadvantage exists, binary translation is applied to only a subset of the instruction set, whereas the others are managed through direct execution on the underlying hardware.
- This somehow reduces the impact on performance of binary translation. CPU virtualization is only a component of a fully virtualized hardware environment.
- VMware achieves full virtualization by providing virtual representation of memory and I/O devices. Memory virtualization constitutes another challenge of virtualized environments and can deeply impact performance without the appropriate hardware support.
- The main reason is the presence of a memory management unit (MMU), which needs to be emulated as part of the virtual hardware. Especially in the case of hosted hypervisors (Type II), where the virtual MMU and the host-OS MMU are traversed sequentially before getting to the physical memory page, the impact on performance can be significant.
- To avoid nested translation, the translation look-aside buffer (TLB) in the virtual MMU directly maps physical pages, and the performance slowdown only occurs in case of a TLB miss

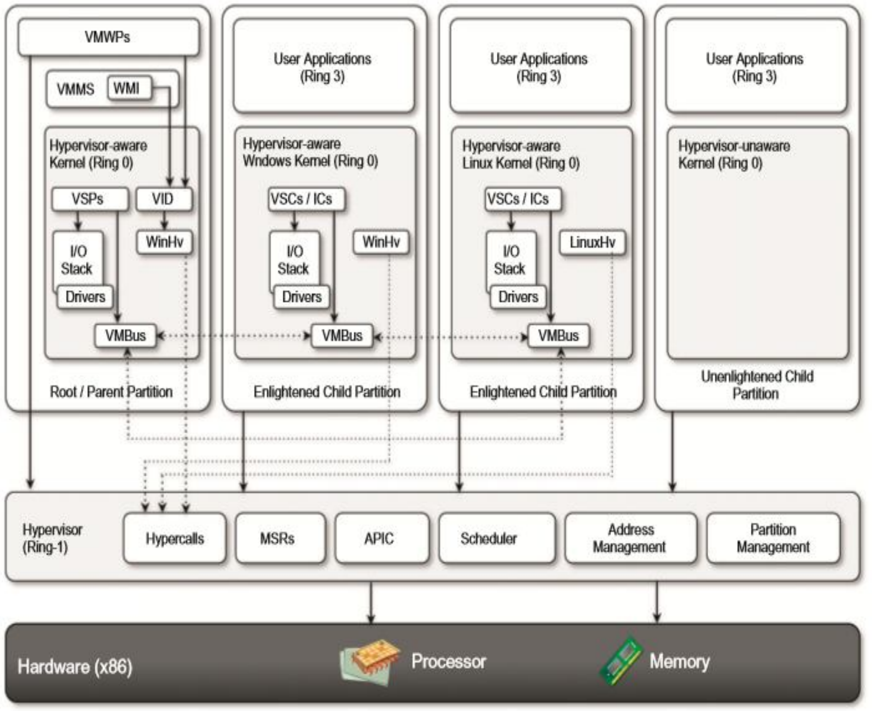


A full virtualization reference model.

Finally, VMware also provides full virtualization of I/O devices such as network controllers and other peripherals such as keyboard, mouse, disks, and universal serial bus (USB) controllers.

- **Microsoft Hyper-V**
- Hyper-V is an infrastructure virtualization solution developed by Microsoft for server virtualization.
- As the name recalls, it uses a hypervisor-based approach to hardware virtualization, which leverages several techniques to support a variety of guest operating systems. Hyper-V is currently shipped as a component of Windows Server 2008 R2 that installs the hypervisor as a role within the server.
- Architecture Hyper-V supports multiple and concurrent execution of guest operating systems by means of partitions. A partition is a completely isolated environment in which an operating system is installed and run.
- Despite its straightforward installation as a component of the host operating system, Hyper-V takes control of the hardware, and the host

operating system becomes a virtual machine instance with special privileges, called the parent partition.


VMWPs
VMMS
WMI
Hypervisor-aware Kernel (Ring 0)
VSPs
VID
I/O Stack
WinHv
Drivers
VMBus
Root / Parent Partition
User Applications (Ring 3)
Hypervisor-aware Wndows Kernel (Ring 0)
VSCs / ICs
I/O Stack
WinHv
Drivers
VMBus
Enlightened Child Partition
User Applications (Ring 3)
Hypervisor-aware Linux Kernel (Ring 0)
VSCs / ICs
I/O Stack
LinuxHv
Drivers
VMBus
Enlightened Child Partition
User Applications (Ring 3)
Hypervisor-unaware Kernel (Ring 0)
Unenlightened Child Partition
Hypervisor (Ring-1)
Hypercalls
MSRs
APIC
Scheduler
Address Management
Partition Management
Hardware (x86)
Processor
Memory

Microsoft Hyper-V architecture.

The parent partition (also called the root partition) is the only one that has direct access to the hardware.

- It runs the virtualization stack, hosts all the drivers required to configure guest operating systems, and creates child partitions through the hypervisor.
- Child partitions are used to host guest operating systems and do not have access to the underlying hardware, but their interaction with it is controlled by either the parent partition or the hypervisor itself.

KVM:

- The kernel-based virtual machine(KVM) is a hypervisor built into Linux kernel. This open-source solution was developed by RedHat corporation to provide virtualization services on the Linux operating system platforms. It has been part of Linux kernel since version2.6.20

and currently being supported by several distributions. A wide variety of guest operating systems work with KVM including several versions of windows. Linux and UNIX.

Oracle VM VirtualBox:

- Oracle VM VirtualBox is the virtualization software package from oracle corporation. This Xen hypervisor based open-source product runs on a wide variety of host operating systems and supports a large number of guest operating systems too.

UNIT-III

Assignment-Cum-Tutorial Questions

SECTION-A

Objective Questions

1._______ is/are the most important advantages of virtualization. [ ]

a. Managed execution b. isolation c. Security d. both a&b

2.The most popular open-source hypervisor available in the market is________. [ ]

a. ESX b. ESXi c. Hyper-V d. Xen

3. Process virtual machines are made to run _________. [ ]

a. Operating system b. Operating system and applications

c. Some specific application d. Any application

4.The allocation of resources and their partitioning among different guests is simplified, because, _________. [ ]

a. The virtual host is controlled by program

b. host is controlled by administrator

c. cycle sharing among user instances

d. performance is not a major issue

5.______ simplifies the administration of virtual machine instances. [ ]

a.portability b.self-containment c. para-virtualization d. both a&b

6.The causes of performance degradation can be traced back by the overhead introduced by the following activities__________. [ ]

i.Maintaining status of virtual processor

ii.Support of privileged instructions

iii.Support of paging within VM

iv.console functions.

a. only i&ii b. only i,ii,&iii c. only ii&iii d. All the above

7.The major source of performance degradation is ____________. [ ]

a.the VMM is executed scheduled together with other applications

b.VMM runs on the user system

c.para-virtualization d.VMware

8.________ and ___________ can slow down the execution of managed applications .

9.The following is/are the disadvantages of virtualization. [ ]

a. performance degradation b. degraded user experience

c. security d. All the above

10.Combining network resources and network functionality into a single, software- based administrative entity is called as__________. [ ]

a. virtual network b. storage virtualization

c. Desktop virtualization d. None of the above

11. A Xen-based system is managed by ____________. [ ]

a. University of Cambridge b. full virtualization

c. Xen-hypervisor d. ALL

12. In a Xen-based system specific control software,which has privileged access to host and controls all the other guest operating systems is executed in special domain called ________ . [ ]

a. Domain 0 b. Domain X c. Domain 1 d. None of the above

13.VM ware technology is based on _____________ [ ]

a.Hardware assisted virtualization b. para virtualization

c. full virtualization d. partial virtualization

14. VMware implements full virtualization either in desktop environment by means of_____ hypervisors, or in server environment, by means of ________ [ ]

a. type I, type II b. type II ,type I c. type I, type 0 d. type 0, type I

15.VMware is well-known for the capability of virtualizing ______ architectures. [ ]

a. x86 b. x85 c. 885 d. 8088

16.The following are the components of hypervisor. [ ]

a. Hyper calls Interface b. MSR c. APIC d. All the above

17.Virtualization overhead of hypervisor is maximum in case of____. [ ]

a. Full virtualization b. Para-virtualization

c. Hardware assisted virtualization d.Equal for all

18.Virtual Machine monitor is the other name of __________. [ ]

a. Guest system b. host system

c. host operating system d.Hypervisor

19.The most popular open source hypervisor available in market is__ [ ]

a. ESX b. ESXi c. Hyper-V d. Xen

20.The single point in the single point of failure problem of virtualization is_______. [ ]

a. Virtual machine b. Guest OS

c. Host machine d. VMM

SECTION-B

SUBJECTIVE QUESTIONS

1.What is virtualization? What is the need for virtualization?

2.What are the advantages of virtualization?

3.Write a short note on the downsides of virtualization.

4.What are the types of virtualization?

5.Briefly explain the role of virtual machine monitor?

6. Why is hypervisor also called as virtual machine monitor?

7. Write a short note on interpretation and binary translation?

8. Enlist the major server virtualization products and vendors?

9. Write the merits and demerits of Virtual Box?

10. Briefly explain the properties of virtual machine?

11. What is the difference between system VM and process VM?

12.Write a short note on Citrix XenServer?

UNIT-IV

Learning Material

4.1 Data security:

- Physical security defines how you control physical access to the servers that support your infrastructure. The cloud still has physical security constraints. After all, there are actual servers running somewhere.
- When selecting a cloud provider, you should understand their physical security protocols and the things you need to do on your end to secure your systems against physical vulnerabilities
- Companies who have outsourced their data centers to a managed services provider may have crossed part of that chasm; what cloud services add is the inability to see or touch the servers on which their data is hosted.
- The main practical problem is that factors that have nothing to do with your business can compromise your operations and your data.
- For example, any of the following events could create trouble for your infrastructure:
  - The cloud provider declares bankruptcy and its servers are seized or it ceases operations.
  - A third party with no relationship to you sues your cloud provider and obtains a blanket subpoena granting access to all servers owned by the cloud provider.
  - Failure of your cloud provider to properly secure portions of its infrastructure—especially in the maintenance of physical access controls—results in the compromise of your systems. The solution is to do two things you should be doing anyway, but

likely are pretty lax about: encrypt everything and keep off-site backups.

- Encrypt sensitive data in your database and in memory. Decrypt it only in memory for the duration of the need for the data. Encrypt your backups and encrypt all network communications.
- Choose a second provider and use automated, regular backups to make sure any current and historical data can be recovered even if your cloud provider were to disappear from the face of the earth. Let's examine how these measures deal with each scenario, one by one.

➢ When the cloud provider goes down This scenario has a number of variants: bankruptcy, deciding to take the business in another direction, or a widespread and extended outage.

➢ The subpoena will compel your cloud provider to turn over your data and any access it might have to that data, but your cloud provider won't have your access or decryption keys.

➢ To get at the data, the court will have to come to you and subpoena you. As a result, you will end up with the same level of control you have in your private data center.

➢ When your cloud provider fails to adequately protect their network When you select a cloud provider, you absolutely must understand how they treat physical, network, and host security.

4.2 Encrypt Everything

➢ In the cloud, your data is stored somewhere; you just don't know exactly where. However, you know some basic parameters:

- Your data lies within a virtual machine guest operating system, and you control the mechanisms for access to that data.
- Network traffic exchanging data between instances is not visible to other virtual hosts.
- For most cloud storage services, access to data is private by default. Many, including Amazon S3, nevertheless allow you to make that data public.

Encrypt your network traffic

- No matter how lax your current security practices, you probably have network traffic encrypted—at least for the most part. A nice feature of the Amazon cloud is that virtual servers cannot sniff the traffic of other virtual servers.
- I still recommend against relying on this feature, since it may not be true of other providers. Furthermore, Amazon might roll out a future feature that renders this protection measure obsolete. You should therefore encrypt all network traffic, not just web traffic.

Encrypt your backups

When you bundle your data for backups, you should be encrypting it using some kind of strong cryptography, such as PGP. You can then safely store it in a moderately secure cloud storage environment like Amazon S3, or even in a completely insecure environment. Encryption eats up CPU. As a result, I recommend first copying your files in plain text over to a temporary backup server whose job it is to perform encryption, and then uploading the backups into your cloud storage system. Not only does the use of a backup server avoid taxing your application server and database server CPUs, it also enables you to have a single highersecurity system holding your cloud storage access credentials rather than giving those credentials to every system that needs to perform a backup.

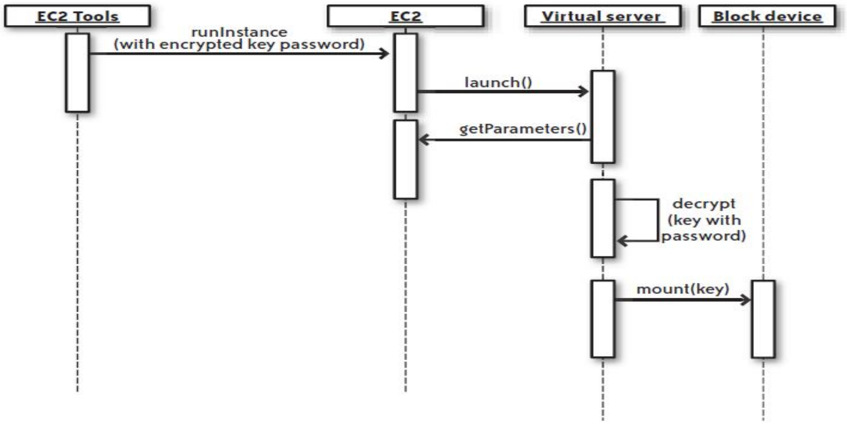
Encrypt your file systems

- Each virtual server you manage will mount ephemeral storage devices or block storage devices. The failure to encrypt ephemeral devices poses only a very moderate risk in an EC2 environment because the EC2 Xen system zeros out that storage when your instance terminates.
- Snapshots for block storage devices, however, sit in Amazon S3 unencrypted unless you take special action to encrypt them.
- The most secure approach to both scenarios is to mount ephemeral and block storage devices using an encrypted file system. Managing the startup of a virtual server using encrypted file systems ultimately ends up being easier in the cloud and offers more security.
- You can add an extra layer of security into the mix by encrypting the password and storing the key for decrypting the password in the machine image. Figure 5-1 illustrates the process of starting up a virtual server that mounts an encrypted file system using an encrypted password

4.3 Regulatory and Standards Compliance

- ➢ Most problems with regulatory and standards compliance lie not with the cloud, but in the fact that the regulations and standards written for Internet applications predate the acceptance of virtualization technologies. In other words, chances are you can meet the spirit of a particular specification, but you may not be able to meet the letter of the specification.
- ➢ For example, if your target standard requires certain data to be stored on a different server than other system logic, can a virtualized server ever meet that requirement? I would certainly argue that it should be able to meet that requirement, but the interpretation as to whether it does may be left up to lawyers, judges, or other non technologists who don’t appreciate the nature of virtualization.

- It does not help that some regulations such as Sarbanes-Oxley (SOX) do not really provide any specific information security requirements, and seem to exist mostly for consultants to make a buck spreading fear among top-level management.



- ***Directive 95/46/EC****:* EC Directive on Data Protection. A 1995 directive for European Union nations relating to the protection of private data and where it can be shared.
- **HIPAA***:* Health Insurance Portability and Accountability Act. A comprehensive law relating to a number of health care issues. Of particular concern to technologists are the privacy and security regulations around the handling of health care data.
- **PCI or PCI DSS***:* Payment Card Industry Data Security Standard. A standard that defines the information security processes and procedures to which an organization must adhere when handling credit card transactions.
- **SOX:** Sarbanes-Oxley Act. Establishes legal requirements around the reporting of publicly held
- Companies to their shareholders. From a security perspective, you'll encounter three kinds of issues in standards and regulations:

- **“How” issues**
  - These result from a standard such as PCI or regulations such as HIPAA or SOX, which govern how an application of a specific type should operate in order to protect certain concerns specific to its problem domain. For example, HIPAA defines how you should handle personally identifying health care data.
- **“Where” issues**
  - These result from a directive such as Directive 95/46/EC that governs where you can store certain information. One key impact of this particular directive is that the private data on EU citizens may not be stored in the United States (or any other country that does not treat private data in the same way as the EU).
- **“What” issues**
  - These result from standards prescribing very specific components to your infrastructure.
  - For example, PCI prescribes the use of antivirus software on all servers processing credit card data.
- The bottom line today is that a cloud-deployed system may or may not be able to meet the letter of the law for any given specification.
- For certain specifications, you may be able to meet the letter of the specification by implementing a mixed architecture that includes some physical elements and some virtual elements.
- Cloud infrastructures that specialize in hybrid solutions may ultimately be a better solution. Alternatively, it may make sense to look at vendors who provide as a service the part of your system that has specific regulatory needs.
- For example, you can use an e-commerce vendor to handle the e-commerce part of your website and manage the PCI compliance issues. In a mixed environment, you don’t host any sensitive data in the cloud. Instead, you offload processing onto privacy servers in a

physical data centre in which the hosts are entirely under your control.

- For example, you might have a credit card processing server at your manage services provider accepting requests from the cloud to save credit card numbers or charge specific cards. With respect to "where" data is stored, Amazon provides S3 storage in the EU.
- Through the Amazon cloud and S3 data storage, you do have the ability to achieve Directive 95/46/EC compliance with respect to storing data in the EU without building out a data centre located in the EU.

4.3 Network security

- ➢ Amazon's cloud has no perimeter. Instead, EC2 provides security groups that define firewall traffic rules governing what traffic can reach virtual servers in that group. Although I often speak of security groups as if they were virtual network segments protected by a firewall, they most definitely are not virtual network segments, due to the following:

- ➢ Two servers in two different Amazon EC2 availability zones can operate in the same

Security Group.

- A server may belong to more than one security group.

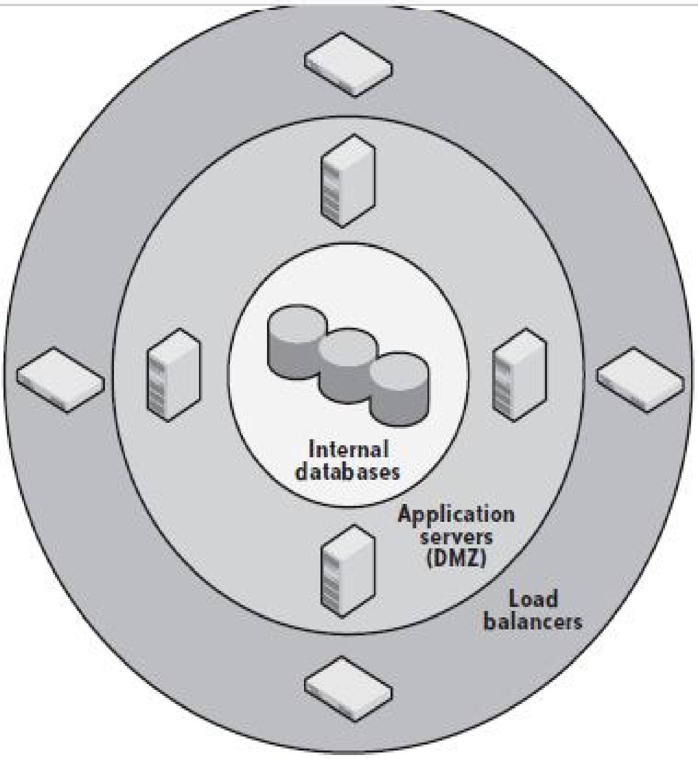
- Servers in the same security group may not be able to talk to each other at all.

- Servers in the same network segment may not share any IP characteristics—they may even be in different class address spaces.

- No server in EC2 can see the network traffic bound for other servers (this is not necessarily true for other cloud systems). If you try placing your virtual Linux server in promiscuous mode, the only network traffic you will see is traffic originating from or destined for your server.

Firewall Rules

- Typically, a firewall protects the perimeter of one or more network segments. Figure 5-2 illustrates how a firewall protects the perimeter.
- A main firewall protects the outermost perimeter, allowing in only HTTP, HTTPS, and (sometimes) FTP* traffic. Within that network segment are border systems, such as load balancers, that route traffic into a DMZ protected by another firewall.
- Finally, within the DMZ are application servers that make database and other requests across a third firewall into protected systems on a highly sensitive internal network. This structure requires you to move through several layers—or perimeters—of network protection in the form of firewalls to gain access to increasingly sensitive data.
- The perimeter architecture's chief advantage is that a poorly structured firewall rule on the inner perimeter does not accidentally expose the internal network to the Internet unless the DMZ is already compromised. In addition, outer layer services tend to be more hardened against Internet vulnerabilities, whereas interior services tend to be less Internet-aware. The weakness of this infrastructure is that a compromise of any individual server inside any given segment provides full access to all servers in that network segment.



- ➢ Figure 5-3 provides a visual look at how the concept of a firewall rule in the Amazon cloud is different from that in a traditional data centre.
- ➢ Each virtual server occupies the same level in the network, with its traffic managed through a security group definition. There are no network segments, and there is no perimeter.
- ➢ Membership in the same group does not provide any privileged access to other servers in that security group, unless you define rules that provide privileged access. Finally, an individual server can be a member of multiple security groups. The rules for a given server are simply the union of the rules assigned to all groups of which the server is a member. You can set up security groups to help you mimic traditional perimeter security.
- ➢ For example, you can create the following:
  - A border security group that listens to all traffic on ports 80 and 443.
  - A DMZ security group that listens to traffic from the border group on ports 80 and 44

- An internal security group that listens to traffic on port 3306 from the DMZ security group

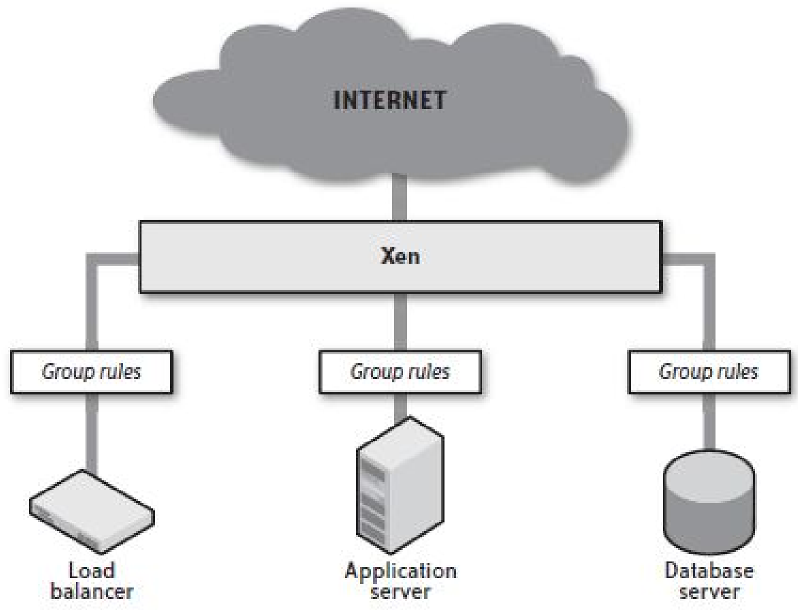


FIGURE 5-3. There are no network segments or perimeters in the cloud

- As with traditional perimeter security, access to the servers in your internal security group requires first compromising the outer group, then the DMZ, and then finally one of the internal servers. Unlike traditional perimeter security, there is the possibility for you to accidentally grant global access into the internal zone and thus expose the zone.
- However, an intruder who compromises a single server within any given zone gains no ability to reach any other server in that zone except through leveraging the original exploit. In other words, access to the zone itself does not necessarily provide access to the other servers in that zone.
- The Amazon approach also enables functionality that used to be out of the question in a traditional infrastructure. For example, you can more easily provide for direct SSH access into each virtual server in

your cloud infrastructure from your corporate IT network without relying on a VPN.

- ➢ You still have the security advantages of a traditional perimeter approach when it comes to the open Internet, but you can get quick access to your servers to manage them from critical locations.
- ➢ Two other advantages of this security architecture are the following:
  - Because you control your firewall rules remotely, an intruder does not have a single target to attack, as he does with a physical firewall.
  - You don't have the opportunity to accidentally destroy your network rules and thus permanently remove everyone's access to a given network segment.
- ➢ I recommend the approach of mimicking traditional perimeter security because it is a well understood approach to managing network traffic and it works. If you take that approach, it's important to understand that you are creating physical counterparts to the network segments of a traditional setup. You don't really have the layers of network security that come with a traditional configuration.
- ➢ A few best practices for your network security include:
  - Run only one network service (plus necessary administrative services) on each virtual server.
  - Every network service on a system presents an attack vector. When you stick multiple services on a server, you create multiple attack vectors for accessing the data on that server or leveraging that server's network access rights.
- ➢ Do not open up direct access to your most sensitive data If getting access to your customer database requires compromising a load balancer, an application server, and a database server (and you're running only one service per server), an attacker needs to exploit three different attack vectors before he can get to that data.
- ➢ Open only the ports absolutely necessary to support a server's service and nothing more Of course your server should be hardened so it is

running only the one service you intend to run on it. But sometimes you inadvertently end up with services running that you did not intend, or there is a non root exploit in the service you are running that enables an attacker to start up another service with a root exploit.

- By blocking access to everything except your intended service, you prevent these kinds of exploits. Limit access to your services to clients who need to access them Your load balancers naturally need to open the web ports 80 and 443 to all traffic.
- Those two protocols and that particular server, however, are the only situations that require open access. For every other service, traffic should be limited to specific source addresses or security groups.
- Even if you are not doing load balancing, use a reverse proxy A reverse proxy is a web server such as Apache that proxies traffic from a client to a server. By using a proxy server, you make it much harder to attack your infrastructure.
- First of all, Apache and IIS are much more battle-hardened than any of the application server options you will be using.
- As a result, an exploit is both less likely and almost certain to be patched more quickly. Second, an exploit of a proxy provides an attacker with access to nothing at all. They must subsequently find an additional vulnerability in your application server itself.
- Use the dynamic nature of the cloud to automate your security embarrassments Admit it. You have opened up ports in your firewall to accomplish some critical business task even though you know better.
- Perhaps you opened an FTP port to a web server because some client absolutely had to use anonymous FTP for their batch file uploads. Instead of leaving that port open 24/7, you could open the port only for the batch window and then shut it down.
- You could even bring up a temporary server to act as the FTP server for the batch window, process the file, and then shut down the server.

The recommendations in the preceding list are not novel; they are standard security precautions.

- The cloud makes them relatively easy to implement, and they are important to your security there.

4.4 Network Intrusion Detection

- Perimeter security often involves network intrusion detection systems (NIDS), such as Snort, which monitor local traffic for anything that looks irregular. Examples of irregular traffic include:
  - Port scans
  - Denial-of-service attacks
  - Known vulnerability exploit attempts
- You perform network intrusion detection either by routing all traffic through a system that analyzes it or by doing passive monitoring from one box on local traffic on your network. In the Amazon cloud, only the former is possible; the latter is meaningless since an EC2 instance can see only its own traffic.

The purpose of a network intrusion detection system

- Network intrusion detection exists to alert you of attacks before they happen and, in some cases, foil attacks as they happen. Because of the way the Amazon cloud is set up, however, many of the things you look for in a NIDS are meaningless. For example, a NIDS typically alerts you to port scans as evidence of a precursor to a potential future attack.
- In the Amazon cloud, however, you are not likely to notice a port scan because your NIDS will be aware only of requests coming in on the ports allowed by your security group rules.
- All other traffic will be invisible to the NIDS and thus are not likely to be perceived as a port scan. PORT SCANS AND THE AMAZON CLOUD.
- When an attacker is looking for vulnerabilities against a particular target, one of the first things they do is execute a port scan against a known server and then examine servers with nearby IP addresses.

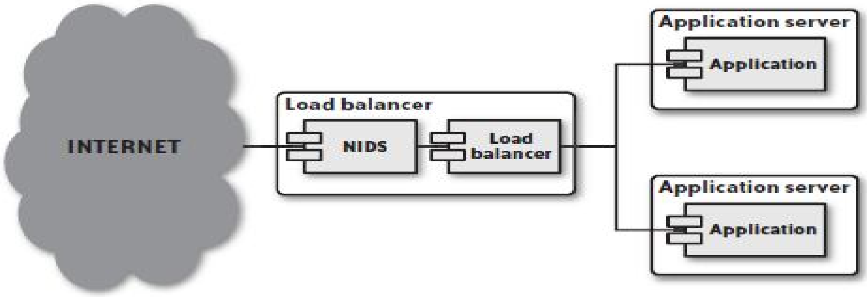
- This approach does not provide terribly useful data when executed against the cloud for a number of reasons: Nodes with proximate IP addresses are almost always unrelated. As a result, you cannot learn anything about the network architecture of a particular organization by executing a port scan.
    - Amazon security groups deny all incoming traffic by default, and requests for ports that have not been opened simply do not respond. As a result, very few ports for any particular server will actually be open. Furthermore, scanning across all ports is a very slow process because each closed port times out instead of actively denying the traffic.
    - Amazon has its own intrusion detection systems in place and does not allow its customers to Execute port scans against their own servers. As a result, an active port scan is likely to be blocked before any real information can be gathered.
- As with port scans, Amazon network intrusion systems are actively looking for denial-of-service attacks and would likely identify any such attempts long before your own intrusion detection software.
- One place in which an additional network intrusion detection system is useful is its ability to detect malicious payloads coming into your network. When the NIDS sees traffic that contains malicious payload, it can either block the traffic or send out an alert that enables you to react.
- Even if the payload is delivered and compromises a server, you should be able to respond quickly and contain the damage.

Implementing network intrusion detection in the cloud

- As I mentioned in the previous section, you simply cannot implement a network intrusion detection system in the Amazon cloud (or any other cloud that does not expose LAN traffic) that passively listens to local network traffic.
- Instead, you must run the NIDS on your load balancer or on each server in your infrastructure. There are advantages and disadvantages

to each approach, but I am not generally a fan of NIDS in the cloud unless required by a standard or regulation.

- The simplest approach is to have a dedicated NIDS server in front of the network as a whole that watches all incoming traffic and acts accordingly.
- Figure 5-4 illustrates this architecture. Because the only software running on the load balancer is the NIDS software and Apache, it maintains a very low attack profile. Compromising the NIDS server requires vulnerability in the NIDS software or Apache—assuming the rest of the system is properly hardened and no actual services are listening to any other ports open to the Web as a whole.



- The load balancer approach creates a single point of failure for your network intrusion detection system because, in general, the load balancer is the most exposed component in your infrastructure.
- By finding a way to compromise your load balancer, the intruder not only takes control of the load balancer, but also has the ability to silence detection of further attacks against your cloud environment.
- You can alternately implement intrusion detection on a server behind the load balancer that acts as an intermediate point between the load balancer and the rest of the system.
- This design is generally superior to the previously described design, except that it leaves the load balancer exposed (only traffic passed by

the load balancer is examined) and reduces the overall availability of the system.

- Another approach is to implement network intrusion detection on each server in the network. This approach creates a very slight increase in the attack profile of the system as a whole because you end up with common software on all servers.
- Vulnerability in your NIDS would result in vulnerability on each server in your cloud architecture. On a positive note, you make it much more difficult for an intruder to hide his footprints.
- As I mentioned earlier, I am not a huge fan of network intrusion detection in the Amazon cloud. Unlike a traditional infrastructure, there just is no meaningful way for a NIDS to serve its purpose. You simply cannot devise any NIDS architecture that will give your NIDS visibility to all traffic attempting to reach your instances.
- The best you can do is creating an implementation in which the NIDS is deployed on each server in your infrastructure with visibility to the traffic that Amazon allows into the security group in which the instance is deployed.
- You would minimally valid proactive alerting, and the main benefit would be protection against malicious payloads. But, if you are encrypting all your traffic, even that benefit is minimal. On the other hand, the presence of a NIDS will greatly reduce the performance of those servers and create a single attack vector for all hosts in your infrastructure.

4.5 Host Security

- Host security describes how your server is set up for the following tasks:
  - Preventing attacks.
  - Minimizing the impact of a successful attack on the overall system.
  - Responding to attacks when they occur.

- It always helps to have software with no security holes. Good luck with that! In the real world, the best approach for preventing attacks is to assume your software has security holes.
- As I noted earlier in this chapter, each service you run on a host presents a distinct attack vector into the host.
- The more attack vectors, the more likely an attacker will find one with a security exploit. You must therefore minimize the different kinds of software running on a server.
- Given the assumption that your services are vulnerable, your most significant tool in preventing attackers from exploiting vulnerability once it becomes known is the rapid rollout of security patches.
- Here's where the dynamic nature of the cloud really alters what you can do from a security perspective. In a traditional data centre, rolling out security patches across an entire infrastructure is time-consuming and risky.
- In the cloud, rolling out a patch across the infrastructure takes three simple steps:
    - Patch your AMI with the new security fixes.
    - Test the results.
    - Relaunch your virtual servers.
- Here a tool such as enStratus or Right Scale for managing your infrastructure becomes absolutely critical.
- If you have to manually perform these three steps, the cloud can become a horrible maintenance headache. Management tools, however, can automatically roll out the security fixes and minimize human involvement, downtime, and the potential for human-error-induced downtime.

4.6 System Hardening

- Prevention begins when you set up your machine image. As you get going, you will experiment with different configurations and constantly rebuild images.

- Once you have found a configuration that works for a particular service profile, you should harden the system before creating your image.
- Server hardening is the process of disabling or removing unnecessary services and eliminating unimportant user accounts. Tools such as Bastille Linux can make the process of hardening your machine images much more efficient. Once you install Bastille Linux, you execute the interactive scripts that ask you questions about your server. It then proceeds to disable services and accounts. In particular, it makes sure that your hardened system meets the following criteria:
  - No network services are running except those necessary to support the server's function.
  - No user accounts are enabled on the server except those necessary to support the services running on the server or to provide access for users who need it.
  - All configuration files for common server software are configured to the most secure settings.
  - All necessary services run under a non privileged role user account (e.g., run My SQL as the my sql user, not root).
  - When possible, run services in a restricted file system, such as a cheroot jail. Before bundling your machine image, you should remove all interactive user accounts and passwords stored in configuration files. Although the machine image will be stored in an encrypted format, Amazon holds the encryption keys and thus can be compelled to provide a third party with access through a court subpoena.

4.7 Antivirus Protection

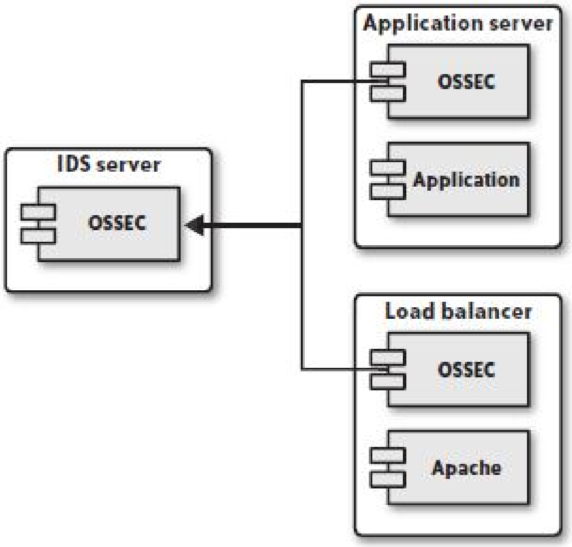
- Some regulations and standards require the implementation of an antivirus (AV) system on your servers. It's definitely a controversial issue, since an AV system with an exploit is itself an attack vector and, on some operating systems, the percentage of AV exploits to known viruses is relatively high. Personally, I have mixed feelings about AV systems.

- They are definitely necessary in some circumstances, but a risk in others. For example, if you are accepting the upload of photos or other files that could be used to deliver viruses that are then served to the public, you have a obligation to use some kind of antivirus software in order to protect your site from becoming a mechanism for spreading the virus.
- Unfortunately, not all AV systems are created equally. Some are written better than others, and some protect you much better than others. Finally, some servers simply don't have an operational profile that makes viruses, worms, and Trojans viable attack vectors.
- I am therefore bothered by standards, regulations, and requirements that demand blanket AV coverage. When looking at the AV question, you first should understand what your requirements are. If you are required to implement AV, then you should definitely do it.
- Look for two critical features in your AV software:
    - How wide is the protection it provides? In other words, what percentage of known exploits does it cover?‡
    - What is the median delta between the time when a virus is released into the wild and the time your AV product of choice provides protection against it? Once you have selected an AV vendor and implemented it on your servers, you absolutely must keep your signatures up to date. You are probably better off with no AV system than one with outdated versions or protections.

4.8 Host Intrusion Detection

- Whereas a network intrusion detection system monitors network traffic for suspicious activity, a host intrusion detection system (HIDS) such as OSSEC monitors the state of your server for anything unusual.

- An HIDS is in some ways similar to an AV system, except it examines the system for all signs of compromise and notifies you when any core operating system or service file changes.
- In my Linux deployments, I use OSSEC (http://www.ossec.net) for host-based intrusion detection. OSSEC has two configuration profiles:
    - Standalone, in which each server scans itself and sends you alerts.
    - Centralized, in which you create a centralized HIDS server to which each of the other servers sends reports.
- In the cloud, you should always opt for the centralized configuration. It centralizes your rules and analysis so that it is much easier to keep your HIDS infrastructure up to date.
- Furthermore, it enables you to craft a higher security profile for your HIDS processing than the individual services might allow for. Figure 5-5 illustrates a cloud network using centralized HIDS.
- As with an AV solution, you must keep your HIDS servers up to date constantly, but you do not need to update your individual servers as often. The downside of an HIDS is that it requires CPU power to operate, and thus can eat up resources on your server.
- By going with a centralized deployment model, however, you can push a lot of that processing onto a specialized intrusion detection server.



4.9 Data Segmentation

- In addition to assuming that the services on your servers have security exploits, you should further assume that eventually one of them will be compromised. Obviously, you never want any server to be compromised.
- The best infrastructure, however, is tolerant of—in fact, it assumes—the compromise of any individual node. This tolerance is not meant to encourage lax security for individual servers, but is meant to minimize the impact of the compromise of specific nodes.
- Making this assumption provides you with a system that has the following advantages:
  - Access to your most sensitive data requires a full system breach.
  - The compromise of the entire system requires multiple attack vectors with potentially different skill sets.
  - The downtime associated with the compromise of an individual node is negligible or nonexistent.
- The segmentation of data based on differing levels of sensitivity is your first tool in minimizing the impact of a successful attack.

- We examined a form of data segmentation in Chapter 4 when we separated credit card data from customer data. In that example, an attacker who accesses your customer database has found some important information, but that attacker still lacks access to the credit card data.
- To be able to access credit card data, decrypt it, and associate it with a specific individual, the attacker must compromise both the e-commerce application server and the credit card processor. Here again the approach of one server/one service helps out.
- Because each type of server in the chain offers a different attack vector, an attacker will need to exploit multiple attack vectors to compromise the system as a whole.

4.10 Credential Management

- Your machine images OSSEC profile should have no user accounts embedded in them. In fact, you should never allow password-based shell access to your virtual servers.
- The most secure approach to providing access to virtual servers is the dynamic delivery of public SSH keys to target servers.
- In other words, if someone needs access to a server, you should provide her credentials to the server when it starts up or via an administrative interface instead of embedding that information in the machine image.
- Of course, it is perfectly secure to embed public SSH keys in a machine image, and it makes life a lot easier. Unfortunately, it makes it harder to build the general-purpose machine images I described in Chapter 4. Specifically, if you embed the public key credentials in a machine image, the user behind those credentials will have access to every machine built on that image.
- To remove her access or add access for another individual, you subsequently have to build a new machine image reflecting the changed dynamics. Therefore, you should keep things simple and

maintainable by passing in user credentials as part of the process of launching your virtual server.

- At boot time, the virtual server has access to all of the parameters you pass in and can thus set up user accounts for each user you specify. It's simple because it requires no tools other than those that Amazon already provides. On the other hand, adding and removing access after the system boots up becomes a manual task.
- Another approach is to use existing cloud infrastructure management tools or build your own that enable you to store user credentials outside the cloud and dynamically add and remove users to your cloud servers at runtime.
- This approach, however, requires an administrative service running on each host and thus represents an extra attack vector against your server.

Compromise Response

- Because you should be running an intrusion detection system, you should know very quickly if and when an actual compromise occurs. If you respond rapidly, you can take advantage of the cloud to eliminate exploit-based downtime in your infrastructure.
- When you detect a compromise on a physical server, the standard operating procedure is a painful, manual process:
    - Remove intruder access to the system, typically by cutting the server off from the rest of the network.
    - Identify the attack vector. You don't want to simply shut down and start over, because the vulnerability in question could be on any number of servers. Furthermore, the intruder very likely left a root kit or other software to permit a renewed intrusion after you remove the original problem that let him in. It is therefore critical to identify how the intruder compromised the system, if that compromise gave him the ability to compromise other systems, and if other systems have the same vulnerability.

- Wipe the server clean and start over. This step includes patching the original vulnerability and rebuilding the system from the most recent uncompromised backup.
- Launch the server back into service and repeat the process for any server that has the same attack vector.

- ➢ This process is very labour intensive and can take a long time. In the cloud, the response is much simpler. First of all, the forensic element can happen after you are operating. You simply copy the root file system over to one of your block volumes, snapshot your block volumes, shut the server down, and bring up a replacement.
- ➢ Once the replacement is up (still certainly suffering from the underlying vulnerability, but at least currently uncompromised), you can bring up a server in a dedicated security group that mounts the compromised volumes. Because this server has a different root file system and no services running on it, it is not compromised. You nevertheless have full access to the underlying compromised data, so you can identify the attack vector. With the attack vector identified, you can apply patches to the machine images. Once the machine images are patched, simply re launch all your instances. The end result is a quicker response to a vulnerability with little (if any) downtime.

UNIT-IV

Assignment-Cum-Tutorial Questions

SECTION-A

Objective Questions

1. The big chasm between traditional data centers and the cloud is __.[ ]

   (A) location of data on someone else's computer

   (B) locations of data on personal computer

   (C) encrypted data on servers

   (D) None of the above

2. The following events could create trouble for your infrastructure. [ ]

   (A) The cloud provider declares bankruptcy

   (B) Third party sues your cloud provider

   (C) Failure of cloud provider to secure portions of its infrastructure

   (D) All the above.

3. Which of the following is/are the solutions to tackle practical problems that arise for a cloud user? [ ]

   (A) Encrypt everything (B) keep offsite backup

   (C) Both A & B (D) None of the Above

4. ______ is a feature of Amazon cloud. [ ]

   (A) virtual servers cannot sniff the traffic of other virtual servers.

   (B) data centers are known to the user

   (C) virtual servers can sniff the traffic of other virtual servers

   (D) users need not worry about the network

5. When you bundle your data for backups, you should be encrypting it using some kind of strong cryptography, such as ______. [ ]

   (A) EC2 (B) Amazon S3 (C) PGP (D) None

6. Amazon's cloud has no perimeter. Instead,___________ provides security groups that define traffic rules. [ ]

   (A) Amazon S3 (B) EC2 (C) PGP (D) None of the above

7. Servers in EC2 can see the network traffic bound for other servers in EC2. [TRUE/FALSE]

8. Two servers in two different Amazon EC2 availability zones can operate in the same security group. [TRUE/FALSE]

9. Maintaining off-site backup can help to recover when the cloud provider goes off. [TRUE/FALSE]

10. Network traffic exchanging between instances is visible to other hosts.

[TRUE/FALSE]

11. Amazon publishes its security standards and processes at_____________. [ ]

(A) aws.amazon.com (B) amazoncloud.com

(C) amazonsecuiry.com (D) a2zamazon.com

12. Why is it recommended to copy your files in plain text over to a temporary backup server whose job is to perform encryption and then upload backups to the cloud . [ ]

(A) encryption eats up CPU (B) ISP monitors host traffic

(C) data is stored in plain text (D) None of the above

13. From a security perspective, you'll encounter the following issues in standards and regulation.

i. How issues ii. Where issues iii. What issues [ ]

(A) both i & ii (B) both i & iii (C) both ii&iii (D) All i,ii & iii

14. Placing your virtual Linux server in _______ mode, the only network traffic you will see is the traffic originating from or destined for your server. [ ]

(A) promiscous (B) server centric (C) cloud centric (D) kernel

15. Using SCP is more secure than FTP because:

i. FTP transmits passwords in plain text

ii. SCP uses SSH protocol for authentication [ ]

(A) only I (B) only ii (C) both i&ii(D) None of the above

16. The weakness of perimeter security infrastructure is_______ [ ]

(A) A compromise of any individual server inside any given segment provides full access to all servers in that segment

(B) Interior services tend to be less internet aware

(C) Outer layer services tend to be more hardened against internet

(D) DMZ is poorly structured

17. ___________ is an open source, free and light weight network intrusion detection system .

(A) snort (B) snoop (C) DMZ (D) Amazon EC2 [ ]

18. Examples of irregular traffic include

i. Port scans

ii. Denial-of-service attacks

iii. Known vulnerability exploit attempts [ ]

(A) both i & ii (B) both i & iii (C) both ii & iii (D) All i, ii & iii

19. _______ monitors the state of your server for anything unusual .[ ]

(A) HIDS (B) NIDS (C) OSSC (D) snort

20.Each virtual server you manage will mount______ storage devices.[ ]

(A) ephermeral (B) long lasting (C) Secondary (D) No specific location

SECTION-B

SUBJECTIVE QUESTIONS

1. What is the standard operating procedure when you detect a compromise on a physical server?
2. Explain in detail about data segmentation.
3. Briefly describe about Host security.
4. Write a short note on system hardening.
5. Explain the process of starting a virtual server with encrypted file system
6. What is the purpose of a network intrusion detection system?
7. What are the few best practices for network security?
8. Explain firewall rules?
9. Discuss the events that could create trouble for infrastructure?

10.Write a short note on network intrusion detection?

11.Describe how your server is setup for

a. presenting attacks

b. minimizing the impact of a successful attack on the overall system

c. responding to attacks when they occur

12. Write a short note on host intrusion detection.

UNIT-V
CLOUD COMPUTING
Learning Material

5. Disaster

Disaster recovery is the practice of making a system capable of surviving unexpected or extraordinary failures. A disaster recovery plan, for example, will help your IT systems survive a fire in your data center that destroys all of the servers in that data center and the systems they support.

5.1 Disaster Recovery Planning

- ✓ Disaster recovery deals with catastrophic failures that are extremely unlikely to occur during the lifetime of a system. If they are reasonably expected failures, they fall under the auspices of traditional availability planning. Although each single disaster is unexpected over the lifetime of a system.
- ✓ Through disaster recovery planning, you identify an acceptable recovery state and develop processes and procedures to achieve the recovery state in the event of a disaster..

5.2 Defining a disaster recovery plan involves two key metrics:

Recovery Point Objective (RPO)

- ✓ The recovery point objective identifies how much data you are willing to lose in the event of a disaster. This value is typically specified in a number of hours or days of data.
- ✓ For example, if you determine that it is OK to lose 24 hours of data, you must make sure that the backups you'll use for your disaster recovery plan are never more than 24 hours old.

Recovery Time Objective (RTO)

- ✓ The recovery time objective identifies how much downtime is acceptable in the event of a disaster.
- ✓ If your RTO is 24 hours, you are saying that up to 24 hours may elapse between the point when your system first goes offline and the point at which you are fully operational again.
- ✓ In addition, the team putting together a disaster recovery plan should define the criteria that would trigger invocation of the plan. In general, invocation of any plan that results in accepting a loss of data should involve the heads of the business organization—even if the execution of the plan is automated.
- ✓ The nature of a disaster, however, generally requires you to accept some level of loss; anything else will come with a significant price tag.
- ✓ In a citywide disaster like Hurricane Katrina, the cost of surviving with zero downtime and zero data loss could have been having multiple data centers in different geographic locations that were constantly synchronized.
- ✓ In other words, you would need two distinct data centers from different infrastructure providers with dedicated, high-bandwidth connections between the two. Accomplishing that level of redundancy is expensive. It would also come with a nontrivial performance penalty.
- ✓ The cold reality for most businesses is likely that the cost of losing 24 hours of data is less than the cost of maintaining a zero downtime/zero loss of data infrastructure.
- ✓ Determining an appropriate RPO and RTO is ultimately a financial calculation: at what point does the cost of data

loss and downtime exceed the cost of a backup strategy that will prevent that level of data loss and downtime?

- ✓ The final element of disaster recovery planning understands the catastrophic scenario. There's ultimately some level of disaster your IT systems will not survive no matter how much planning and spending you do. A good disaster recovery plan can describe that scenario so that all stakeholders can understand and accept the risk.

The Recovery Point Objective

- ✓ The Armageddon scenario results in total loss of all system data and the binaries of all applications required to run the system.
- ✓ Your RPO is somewhere between the application state when you first deployed it and the state at the time of the disaster.
- ✓ You may even define multiple disaster levels with different RPOs.* Just about any software system should be able to attain an RPO between 24 hours for a simple disaster to one week for a significant disaster without incurring absurd costs.
- ✓ Of course, losing 24 hours of banking transactions would never be acceptable, much less one week.

RPO is typically governed by the way in which you save and back up data:

- ✓ Weekly off-site backups will survive the loss of your data center with a week of data loss. Daily off-site backups are even better.
- ✓ Daily on-site backups will survive the loss of your production environment with a day of data loss plus replicating transactions during the recovery period after

the loss of the system. Hourly on-site backups are even better.

- ✓ A NAS/SAN will survive the loss of any individual server, except for instances of data corruption with no data loss.
- ✓ A clustered database will survive the loss of any individual data storage device or database node with no data loss.
- ✓ A clustered database across multiple data centers will survive the loss of any individual data center with no data loss.

The Recovery Time Objective

Having up-to-the-second off-site backups does you no good if you have no environment to which you can restore them in the event of failure. The ability to assemble a replacement infrastructure for your disasters is:

The data restore time—governs the RTO.

- What would happen if your managed services provider closed its doors tomorrow? If you have a number of dedicated servers, it can be days or weeks. In a traditional infrastructure, a rapid RTO is very expensive.
- Have an agreement in place with another managed services provider to provide either a backup infrastructure or an SLA for setting up a replacement infrastructure in the event your provider goes out of business.
- Depending on the nature of that agreement, it can nearly double the costs of your IT infrastructure.

5.3 Disasters in the Cloud

Assuming unlimited budget and capabilities, we focus on three key things in disaster recovery planning:

1. Backups and data retention

2. Geographic redundancy
3. Organizational redundancy

- ✓ If we can take care of those three items, it’s can meet most RPO and RTO needs.
- ✓ But we have never been in a situation in which we had an unlimited budget and capabilities, so always had to compromise.
- ✓ As a result, the order of the three items matters. In addition, if hosting provider is a less-proven organization, organizational redundancy may be more important than geographic redundancy. Fortunately, the structure of the Amazon cloud makes it very easy to take care of the first and second items.

➢ **Backup Management**

Ability to recover from a disaster is limited by the quality and frequency of backups. In a traditional IT infrastructure, companies often make full weekly backups to tape with nightly differentials and then ship the weekly backups off-site.

➢ **BACKUPS, BUSINESS CONTINUITY, AND AWS**

- ✓ Here a number of technologies that Amazon Web Services provide to help to manage backups effectively.
- ✓ If we are using a different cloud, they likely have some similar tools as well as some that are completely different.
- ✓ A critical part of any backup strategy, however, is the concept of off-site backups. Whatever the backup strategy, must not only have a mechanism for moving all data critical for achieving the Recovery Point Objectives out of the cloud, but we must also store that data in a portable format
- ✓ So it can recover into an environment that might be radically different from the current cloud provider.

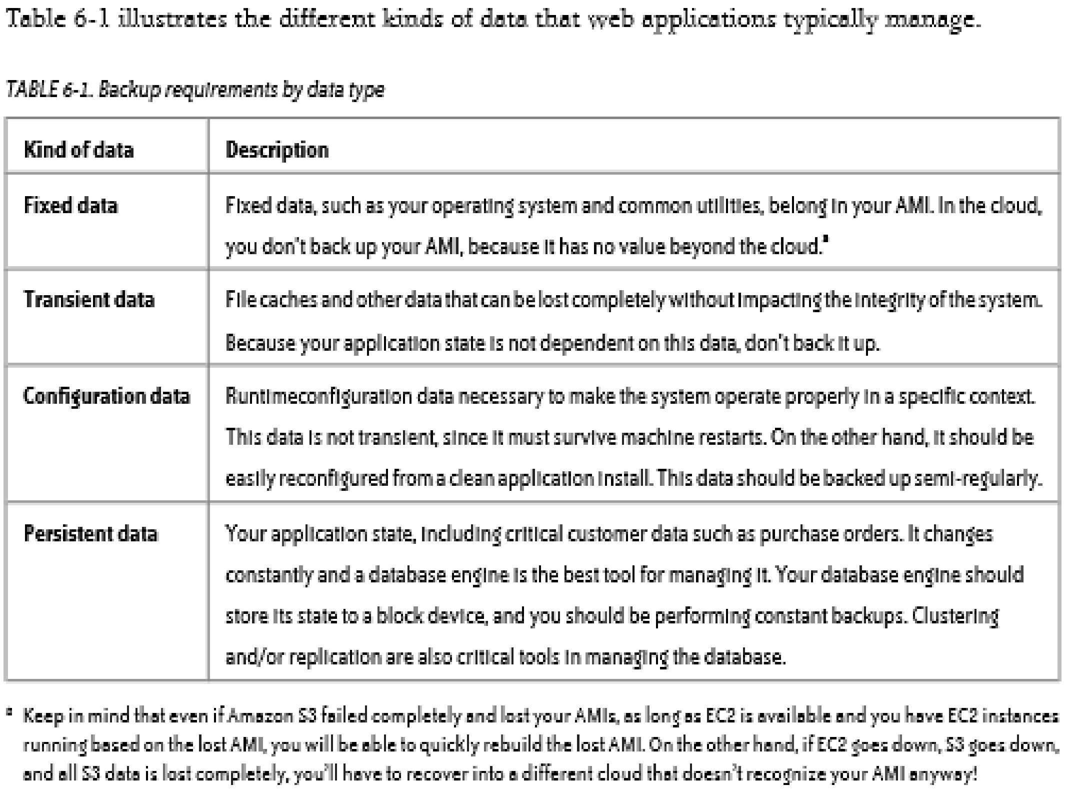


Table 6-1 illustrates the different kinds of data that web applications typically manage.

TABLE 6-1. Backup requirements by data type

Kind of data	Description
Fixed data	Fixed data, such as your operating system and common utilities, belong in your AMI. In the cloud, you don't back up your AMI, because it has no value beyond the cloud.[a]
Transient data	File caches and other data that can be lost completely without impacting the integrity of the system. Because your application state is not dependent on this data, don't back it up.
Configuration data	Runtimeconfiguration data necessary to make the system operate properly in a specific context. This data is not transient, since it must survive machine restarts. On the other hand, it should be easily reconfigured from a clean application install. This data should be backed up semi-regularly.
Persistent data	Your application state, including critical customer data such as purchase orders. It changes constantly and a database engine is the best tool for managing it. Your database engine should store its state to a block device, and you should be performing constant backups. Clustering and/or replication are also critical tools in managing the database.

[a] Keep in mind that even if Amazon S3 failed completely and lost your AMIs, as long as EC2 is available and you have EC2 instances running based on the lost AMI, you will be able to quickly rebuild the lost AMI. On the other hand, if EC2 goes down, S3 goes down, and all S3 data is lost completely, you'll have to recover into a different cloud that doesn't recognize your AMI anyway!

- ✓ In disaster recovery, persistent data is generally the data of greatest concern. It can always rebuild the operating system, install all the software, and reconfigure it, but we have no way of manually rebuilding the persistent data.
- ❖ **Fixed data strategy**
- ✓ If we are fixated on the idea of backing up our machine images, can download the images out of S3 and store them outside of the Amazon cloud.
- ✓ If S3 were to go down and incur data loss or corruption that had an impact on our AMIs, then would be able to upload the images from our off-site backups and reregister them.
- ✓ It's not a bad idea and it is not a lot of trouble, but the utility is limited.

- ❖ **Configuration data strategy**
- ✓ A good backup strategy for configuration information comprises two levels.
- ✓ The first level can be either a regular file system dump to your cloud storage or a file system snapshot. For most applications, back up our configuration data once a day or even once a week
- ✓ If the configuration data changes twice a day and we have a two-hour RPO, it will need to back up configuration data twice a day.
- ✓ If configuration data changes irregularly, it may be necessary to make hourly backups or specifically tie backups to changes in application configuration.
- ✓ An alternate approach is to check the application configuration into a source code repository outside of the cloud and leverage that repository for recovery from even minor losses.
- ✓ Snapshots tend to be the most efficient and least intrusive mechanism for performing backups, but they are also the least portable.
- ✓ direct access to the EC2 snapshots not possible, and even if did, they would not be usable outside of the Amazon cloud.
- ✓ At some point, you do need to get that data out of the cloud so that you have off-site backups in a portable format.

Here's what It recommend:

- **<u>Create regular</u>**—at a minimum, daily—snapshots of the configuration data.
- **<u>Create semi-regular</u>**—at least less than your RPO—file system archives in the form of ZIP or TAR files and move those archives into Amazon S3.
- **<u>On a semi-regular basis</u>**—again, at least less than the RPO—copy the file system archives out of the Amazon cloud into another cloud or physical hosting facility.

Let's say we had a one-day RPO with an application housing less than 10 GB of data whose configuration data could change on a whim:

- It would make hourly snapshots of the file system with the configuration data and archive the whole thing in a portable format to S3 at least once a day.
- It would retain a week's worth of full backups in S3 in case of a corruption issue during backup.
- Finally, It would keep a week's worth of backups off site for the last week along with one backup a week for the prior month and one backup a month for the prior year.
    - ✓ Depending on the amount of data in question and the data retention issues/requirements, It might do more or less in the archiving of old backups.
    - ✓ Different RPOs and application behaviors mandate different strategies.

❖ **Persistent data strategy (aka database backups)**

Using a relational database to store customer information and other persistent data.

- ✓ The purpose of a relational database is to maintain the consistency of complex transactional data.
- ✓ The challenge of setting up database backups is doing them regularly in a manner that does not impact operations while retaining database integrity.
- ✓ MySQL, like all database engines, provides several convenient tools for backups, but it must use them carefully to avoid data corruption.
- ✓ The techniques are well-documented in the database literature, but they're important to summarize them and as well as apply them to an Amazon EC2 environment.
- ✓ With the configuration data, it will be making a backup in between the writing of two different files that must remain consistent or in the middle of writing out a file to the file system.
- ✓ With database storage, it is a near certainty that every time we try to copy those files, the database will be in the middle of doing something with them.

- ✓ As a result, we need to get clever with our database backup strategy. **The first line of defense is either multi master replication or clustering** .

  <u>A multi master database</u> is one in which two master servers execute write transactions independently and replicate the transactions to the other master.

  <u>A clustered database</u> environment contains multiple servers that act as a single logical server. Under both scenarios, when one goes down, the system remains operational and consistent. Instead, you can perform master-slave replication .
- ✓ Master-slave replication involves setting up a master server that handles the write operations and replicating transactions over to a slave server . Each time something happens on the master, it replicates to the slave
- ✓ Replication in itself is not a “first line of defense,” since replication is not atomic with respect to the transactions that take place on the master.
- ✓ In other words, a master can crash after a transaction has completed but before it has had time to replicate to the slave.

To get around this problem, generally do the following:

  - o Set up a master with its data files stored on a block storage device.
  - o Set up a replication slave, storing its data files on a block storage device.
  - o Take regular snapshots of the master block storage device based on my RPO.
  - o Create regular database dumps of the slave database and store them in S3.
  - o Copy the database dumps on a semi-regular basis from S3 to a location outside the Amazon cloud.

- ➢ **Amazon's Elastic Block Storage** (EBS)
- ✓ Offering database corruption in the event of MySQL server failing. However, that corruption is extremely likely—as it is with many file systems—and have adjusted my backup strategy accordingly.
- ✓ Actually taking snapshots or creating database dumps for some database engines is very tricky in a runtime environment, especially if want to do it hourly or even more frequently.
- ✓ The challenge in creating the backups for these database engines is the need to stop processing all transactions while the backup is taking place.
- ✓ To complicate the situation, database dumps can take a long time to complete. As a result, applications will grind to a halt while the make any database dumps.
- ✓ Snapshots are available in most cloud environments and provide an important approach for maintaining database integrity without completely shutting down application processing— even with large data sets in databases such as MySQL.

We need to freeze the database only for an instant to create the snapshot.

The process follows these steps:

1. Lock the database.
2. Sync the file system (this procedure is file system-dependent).
3. Take a snapshot.
4. Unlock the database

- ✓ The whole process should take about one second. On Amazon EC2, it will store the snapshots directly onto block storage.
- ✓ Unfortunately, the snapshots are not portable, so can't use them for off-site storage. Therefore will need to do database dumps, no matter how much it would rather avoid doing them.
- ✓ Because of this need, It run my backups against a database slave. The slave can afford to be locked for a period of time while a database dump completes.

- ✓ The fact that the slave may be a transaction or two (or ten) behind the master is unimportant for the purposes of making a backup.

The steps for creating the database dump are:

1. Execute the database dump.
2. When complete, encrypt the dump and break it into small, manageable chunks.
3. Move the dump over to S3.

- ✓ Amazon S3 limits the file size to 5 GB.
- ✓ As a result, probably need to break the database into chunks, and you should definitely encrypt it and anything else you send into Amazon S3.
- ✓ Now that have a portable backup of database server, it can copy that backup out of the Amazon cloud and be protected from the loss of S3 backups.
- ➢ **Backup security**
- ✓ The file systems are encrypted to protect the snapshots that are making for backups from prying eyes. The harder part is securing portable backups as they stored in S3 and move them off site. Here we use PGP-compatible encryption for portable backups.
- ✓ Need to worry about two issues:
    - o Keeping the private decryption key out of the cloud.
    - o Keeping the private decryption key some place that it will never, ever get lost.
- ✓ The cloud needs only the public encryption key so it can encrypt the portable backups.
- ✓ can't store the decryption key with our backups. Doing so will defeat the purpose of encrypting the backups in the first place.
- ✓ Because we will store our decryption key somewhere else, run the risk of losing our decryption key independent of our backups.
- ✓ On the other hand, keeping a bunch of copies of our decryption key will make it more likely it will fall into the wrong hands.

The best approach Keep two copies:

- One copy stored securely on a highly secure server in our internal network.
- One copy printed out on a piece of paper and stored in a safety deposit box nowhere near the same building in which the house where highly secure server.

- ✓ More than one person should know the locations of these copies. A true disaster can unfortunately result in the loss of personnel, so personnel redundancy is also important for a disaster recovery plan.
- ✓ If we are automating the recovery from portable backups, it will also need to keep a copy of the private decryption key on the server that orchestrates the automated recovery efforts.

- ➢ **Geographic Redundancy**

- ✓ The virtualization technologies behind the cloud simply make it a lot easier to automate the processes and have a relatively inexpensive mechanism for off-site backups. Turning now to Recovery Time Objective, the key is redundancy in infrastructure.
- ✓ If we develop geographical redundancy, it can survive just about any physical disaster that might happen. With a physical infrastructure, geographical redundancy is expensive.
- ✓ In the cloud, however, it is relatively cheap. It don’t need to have the application running actively in all locations, but it need the ability to bring the application up from the redundant location in a state that meets Recovery Point Objective within a timeframe.
- ✓ If you have a 2-hour RTO with a 24-hour RPO, geographical redundancy means that the second location can be operational within two hours of the complete loss of the primary location using data that is no older than 24 hours.
- ✓ Amazon provides built-in geographic redundancy in the form of regions and availability zones.

- ✓ If we have the instances running in a given availability zone, we can get them started back up in another availability zone in the same region without any effort.

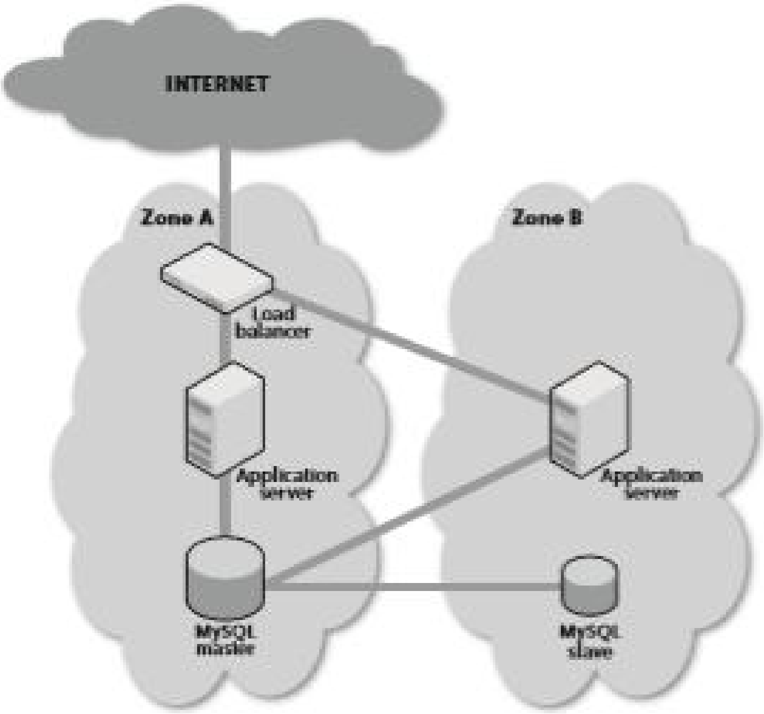


FIGURE 6-1. By spanning multiple availability zones, you can achieve geographic redundancy

Spanning availability zones

Amazon infrastructure except block storage devices is available across all availability zones in a given region. Although there is a charge for network traffic that crosses availability zones, that charge is generally well worth the price for the leveraging ability to create redundancy across availability zones.

- ✓ Figure 6-1 illustrates an application environment that can easily tolerate the loss of an entire availability zone. If it lose the entire availability zone B, nothing happens.
- ✓ The application continues to operate normally, although perhaps with degraded performance levels. If it lose availability zone A, it will need to bring up a new load balancer in availability zone B and promote the

slave in that availability zone to master. The system can return to operation in a few minutes with little or no data loss.

- ✓ If the database server were clustered and had a spare load balancer running silently in the background, the could reassign the IPaddress from the old load balancer to the spare and see only a few seconds of downtime with no data loss.
- ✓ The Amazon SLA provides for a 99.95% uptime of at least two availability zones in each region. The U.S. East Coast, for example, has three availability zones.§ As a result, you have only a 33% chance of any given failure of two availability zones being exactly the two zones you are using.
- ✓ Even in the event that you are unfortunate enough to be operating in exactly the two zones that fail, it can still exceed Amazon's SLA as long as the region that is operating in has more than two availability zones.
- ✓ The trick is to execute your disaster recovery procedures and bring your infrastructure back up in the remaining availability zone. As a result, it can be operational again while the other two availability zones are still down

- ➢ **Operating across regions**
  - ✓ Amazon supports two regions: us-east-1 (Eastern United States) and eu-west-1 (Western Europe). These regions share little or no meaningful infrastructure.
  - ✓ The advantage of this structure is that the application can basically survive a nuclear attack on the U.S. or EU (but not on both!) if it operate across regions.
  - ✓ On the other hand, the lack of common infrastructure makes the task of replicating the environments across regions more difficult.
  - ✓ Each region has its own associated Amazon S3 region. Therefore, it cannot launch EC2 instances in the EU using AMIs

from the U.S., and cannot use IP addresses formerly associated with a load balancer in the EU with a replacement in the U.S.

How we manage operations across regions depends on the nature of our web application and our redundancy needs.

- ➢ **DNS management**
- ✓ We can use round-robin DNS to work around the fact that IP addresses are not portable across regions, but will end up sending European visitors to the U.S. and vice versa (very inefficient network traffic management) and lose half our traffic when one of the regions goes down.
- ✓ We can leverage a dynamic DNS system such as UltraDNS that will offer up the right DNS resolution based on source and availability.
- ➢ **Database management**
- ✓ Clustering across regions is likely not practical .We can set up a master in one region with a slave in the other. Then it performs write operations against the master, but read against the slave for traffic from the region with the slave.
- ✓ Another option is to segment the database so that the European region has "European data" and the U.S. region has "American data." Each region also has a slave in the other region to act as a recovery point from the full loss of a region.
- ➢ **Regulatory issues**
- ✓ The EU does not allow the storage of certain data outside of the EU. As a result, legally it may not be allowed to operate across regions, no matter what clever technical solutions it devises.
- ✓ In reality, an Amazon+GoGrid or Amazon+Rackspace approach to redundancy may be more effective than trying to use Amazon's two cross-jurisdictional regions.

For most purposes, It recommend a process for regularly copying infrastructure elements (AMIs, backups, and configuration) over into the other region and then having the ability to rapidly start that infrastructure in the event of a total, prolonged failure of your core zone.

- **Organizational Redundancy**
  - ✓ If we have an infrastructure that does everything that have recommended so far, we are pretty well protected against everything physical that can happen.
  - ✓ We are still exposed at the business level, however. Specifically, if Amazon or Rackspace or GoGrid or whoever are using goes out of business or decides it is bored with cloud computing, might find our self in trouble.
  - ✓ Physical disasters are a relatively rare thing, but companies go out of business everywhere every day—even big companies like Amazon and Rackspace.
  - ✓ Even if a company goes into bankruptcy restructuring, there's no telling what will happen to the hardware assets that run their cloud infrastructure.
  - ✓ The disaster recovery plan should therefore have contingencies that assume that cloud provider simply disappears from the face of the earth.
  - ✓ Probably won't run concurrent environments across multiple clouds unless it provides some level of geographic advantage.
  - ✓ Even in that case, our environments are not likely to be redundant so much as segmented for the geographies they are serving.
  - ✓ Instead, the best approach to organizational redundancy is to identify another cloud provider and establish a backup environment with that provider in the event our first provider fails.

In particular, it must consider all of the following concerns:

- o Storing the portable backups at our secondary cloud provider.
- o Creating machine images that can operate our applications in the secondary provider's virtualized environment.

  - o Keeping the machine images up to date with respect to their counterparts with the primary provider.
  - o Not all cloud providers and managed service providers support the same operation systems or file systems.

- ✓ If the application is dependent on either, we need to make sure you select a cloud provider that can support your needs.

- ➢ **Disaster Management**
- ✓ When we are performing the backups and have an infrastructure in place with all of the appropriate redundancies. To complete the disaster recovery scenario, it needs to recognize when a disaster has happened and have the tools and processes in place to execute our recovery plan.
- ✓ One of the coolest things about the cloud is that all of this can be automated. we can recover from the loss of Amazon's U.S. data centers while we sleep
- ➢ **Monitoring**
- ✓ Monitoring the cloud infrastructure is extremely important. we cannot replace a failing server or execute the disaster recovery plan if we don't know that there has been a failure.
- ✓ The trick is that our monitoring systems cannot live in either our primary or secondary cloud provider's infrastructure. They must be independent of our clouds. If we want to enable automated disaster recovery, they also need the ability to manage the EC2 infrastructure from the monitoring site.
- ✓ Our primary monitoring objective should be to figure out what is going to fail before it actually fails. The most common problem have encountered in EC2 is servers that gradually decrease in local file I/O throughput until they become unusable. This problem is something can easily watch for and fix before users even notice it.
- ✓ On the other hand, if we wait for our application to fail, chances are users have had to put up with poor performance for some

period of time before it failed completely. It may also prove to be a precursor to a larger cloud failure event.

- ✓ There are many other more mundane things that you should check on in a regular environment.
- ✓ In particular, it should be checking capacity issues such as disk usage, RAM, and CPU.

We need to monitor for failure at three levels:

  - o Through the provisioning API (for Amazon, the EC2 web services API)
  - o Through our own instance state monitoring tools
  - o Through our application health monitoring tools

- ✓ The cloud provider's provisioning API will tell about the health of our instances, any volumes they are mounting, and the data centers in which they are operating. When we detect a failure at this level, it likely means something has gone wrong with the cloud itself.
- ✓ Before engaging in any disaster recovery, it will need to determine whether the outage is limited to one server or affects indeterminate servers, impacting an entire availability zone or an entire region.
- ✓ Monitoring is not simply about checking for disasters; mostly it is checking on the mundane.
- ✓ With enStratus, a Python service on each server that checks for a variety of server health indicators—mostly related to capacity management.
- ✓ The service will notify the monitoring system if there is a problem with the server or its configuration and allow the monitoring system to take appropriate action.
- ✓ It also checks for the health of the applications running on the instance.

Load Balancer Recovery

- ✓ One of the reasons companies pay absurd amounts of money for physical load balancers is to greatly reduce the likelihood of load balancer failure. With cloud vendors such as GoGrid— and in the future, Amazon—we can realize the benefits of hardware load balancers without incurring the costs.
- ✓ Under the current AWS offering, we have to use less-reliable EC2 instances. Recovering a load balancer in the cloud, however, is lightning fast.
- ✓ As a result, the downside of a failure in cloud-based load balancer is minor.
- ✓ Recovering a load balancer is simply a matter of launching a new load balancer instance from the AMI and notifying it of the IP addresses of its application servers. And then further reduce any downtime by keeping a load balancer running in an alternative availability zone and then remapping the static IP address upon the failure of the main load balancer.

➢ **Application Server Recovery**

- ✓ If we operating multiple application servers in multiple availability zones, the system as a whole will survive the failure of any one instance—or even an entire availability zone. Then still need to recover that server so that future failures don't affect the infrastructure.
- ✓ The recovery of a failed application server is only slightly more complex than the recovery of a failed load balancer.
- ✓ Like the failed load balancer, start up a new instance from the application server machine image. Then pass it configuration information, including where the database is.
- ✓ Once the server is operational, must notify the load balancer of the existence of the new server (as well as deactivate its knowledge of the old one) so that the new server enters the load-balancing rotation.

- **Database Recovery**
- ✓ Database recovery is the hardest part of disaster recovery in the cloud. The disaster recovery algorithm has to identify where an uncorrupted copy of the database exists.
- ✓ This process may involve promoting slaves into masters, rearranging the backup management, and reconfiguring application servers.
- ✓ The best solution is a clustered database that can survive the loss of an individual database server without the need to execute a complex recovery procedure.
- ✓ Absent clustering, the best recovery plan is one that simply launches a new database instance and mounts the still functional EC2 volume formerly in use by the failed instance.

When an instance goes down, any number of related issues may also have an impact on that strategy:

- The database could be irreparably corrupted by whatever caused the instance to crash.
- The volume could have gone down with the instance.
- The instance's availability zone (and thus the volume as well) could be unavailable.
- You could find yourself unable to launch new instances in the volume's availability zone.

As a result, we need a fallback plan for our recovery plan.

The following process will typically cover all levels of database failure:

- Launch a replacement instance in the old instance's availability zone and mount its old volume.
- If the launch fails but the volume is still running, snapshot the volume and launch a new instance in any zone, and then create a volume in that zone based on the snapshot.

- If the volume from step 1 or the snapshot from step 2 is corrupt, the need to fall back to the replication slave and promote it to database master.
- If the database slave is not running or is somehow corrupted, the next step is to launch a replacement volume from the most recent database snapshot.
- If the snapshot is corrupt, go further back in time until you find a backup that is not corrupt.
- Step 4 typically represents your worst-case scenario. If we get to 5, there is something wrong with the way you are doing backup.

UNIT-V

Assignment-Cum-Tutorial Questions

SECTION-A

Objective Questions

1. Which of the following are key elements in disaster recovery planning?

i. backup & data retention
ii. geographic redundancy
iii. organizational redundancy [ ]

A) i & ii B) ii & iii C) i & iii D) All the above

2. Ability to recover from a disaster is limited by _______ of backups.[ ]

A) quality B) frequency C)both A&B D) none

3. In disaster recovery ________ data is generally the data of greatest concern. [ ]

A)persistent B) short term C) meta (D) none of the above

4. __________ of your file system tend to be most efficient. [ ]

A) snapshots B) zipped file system C) centralized backup D) none

5. _________ involves setting up a master server that handles your write operations and replicating transactions over to a slave server. [ ]

A) Master slave replication B) Multi-master replication
C) Clustering D) Master server

6. A__________database is one in which two master servers execute write transactions independently and replicate the transactions to the other master. [ ]

A)Master slave replication B)Multi master replication
C)Master server D)none of the above

7. The correct sequence of steps for creating database dump are___.[ ]

i. encrypt the dump and break it in to small , manageable chunks
ii. execute the database dump
iii. move the dump over to S3

A)i-ii-iii B)ii-i-iii C)iii-i-ii D)None of the above

8. Amazon S3 limits your file size to be _______GB. [ ]

A)2 B) 5 C) 10 D) 20

9. ___________ need the ability to manage your EC2 infrastructure fromthe monitoring site.

A) automated disaster recovery B) disaster management [ ]
C) database recovery D) application server recovery

10. __________ is the art of being able to resume normal systems operations when faced with a disaster scenario. [ ]
A) disaster recovery B) database backup
C) accepting disaster D) None of the above

11. A __________ will help your IT systems survive a fire in your data center thatdestroy all of the servers in the data center and the systems they support. [ ]
A) Virtualization B)Data center C) Cloud computing D)None

12. __________ lets you automate disaster recovery. [ ]
A) virtualization B) data center
C) cloud computing D)cloud infrastructure

13. Disaster recovery plan involves two key metrics____ and _____ .[ ]
A) Recovery point objective & Recovery time objective
B) disaster point objective & disaster time objective
C) Disaster plan virtualization & data center
D) None of the above

14. _________ objective identifies how much data you are willing to loose in the event of a disaster. [ ]
A) Recovery point B) Recovery time C) disaster point D) disaster time

15. ________ objective identifies how much down time is acceptable in the event of disaster. [ ]
A) Recovery point B) Recovery time C) disaster point D) disaster time

16. An ideal disaster recovery scenario is which has_______. [ ]
A) no down time B) no loss of data
C) both A&B D) depends on nature of disaster

17. A _______ will survive the loss of any individual data storage or database node with no data loss. [ ]
A) Clustered database B) Distributed database C) Both A&B D) None

18. In traditional infrastructure, a rapid RTO is very expensive. [TRUE/FALSE]

19. A _______ will survive the loss of any individual server, except for instances data corruption with no data loss. [ ]
A) NAS B) SAN C) both A&B D) None

SECTION-B

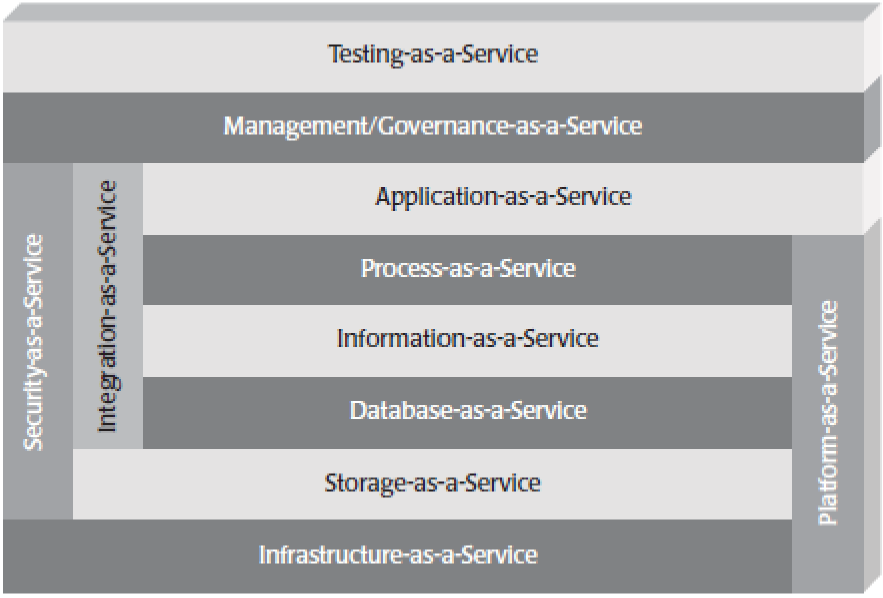
SUBJECTIVE QUESTIONS

1. Explain Disaster Recovery Planning.
2. Illustrate how RPO is typically governed by the way in which you save and back up data.
3. Explain the metrics of Disaster Recovery Plan.
4. Explain about the key disasters in cloud.
5. Explain different kinds of data that web applications typically manage?
6. Explain about Amazon's Elastic Block Storage.
7. Explain the process will typically cover all levels of database failure.

UNIT-6
Cloud Computing

6. Components of Cloud Computing

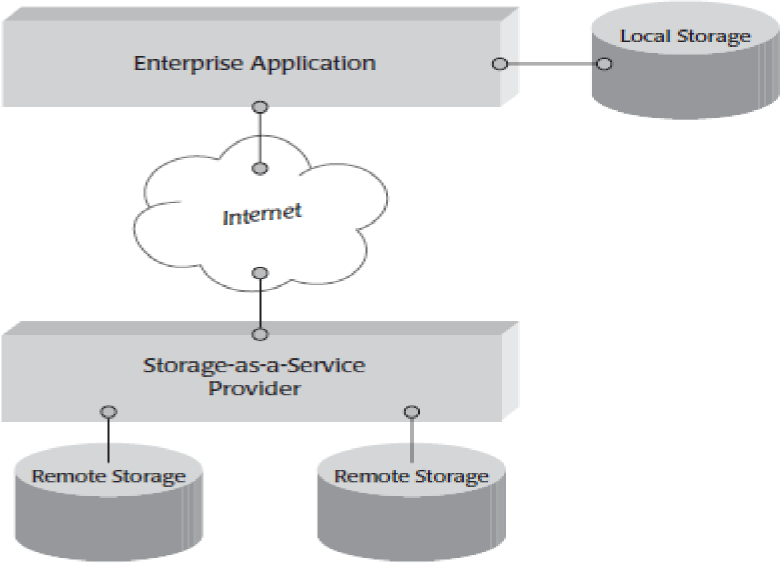
- Cloud computing has the following components:
  1. Storage-as-a-service
  2. Database-as-a-service
  3. Information-as-a-service
  4. Process-as-a-service
  5. Application-as-a-service
  6. Platform-as-a-service
  7. Integration-as-a-service
  8. Security-as-a-service
  9. Management/governance-as-a-service
  10. Testing-as-a-service
  11. Infrastructure-as-a-service



6.1 Storage-as-a-Service

- Storage-as-a-service is the ability to leverage storage that physically exists remotely but is logically a local storage resource to any application that requires storage.

- This is the most primitive component of cloud computing and is leveraged by most of the other cloud computing components.
- Using a disk that we access over the Internet is a bit illogical at first.
- Why would some enterprise leverage storage that exists thousands of miles away, as a service, when disk space is so cheap and getting cheaper?
- Storage-as-a-service allows you to store information on a remote disk drive as if it were local.



Benefits of Storage-as-a-service:

- We can expand the amount of disk space available as we need it and pay only for what you use. We can reduce the amount of disk space and thereby cost. This makes storage-as-a-service solutions cost effective only for larger volumes of data, typically more than 500 gigabytes, either through direct access or by using the disk as if it were local to your client computer.
- We can also use the storage-as-a-service provider as a redundant backup for critical files.
- We do not have to maintain the hardware. Drives can go down and you do not have to replace them; it is all a part of the service. When compared with an on-premise solution where we have to physically

repair the drive, storage -as-a-service removes us from having to deal with that issue.

- Finally, the storage-as-a-service provider provides the disaster recovery system for us, and getting back deleted files or entire directories is part of the service. The provider backs up and restores the file system as you require.
- We do not have to pay someone to handle that task within the data center, and local staff will not have the responsibility of maintaining the storage systems properly.

Drawbacks to Storage-as-a-service:

- We are dependent on the Internet as the mechanism to connect to storage as-a-service provider, and if the network goes down, you lose that connection.
- If a mission-critical need is compromised by a rare and temporary loss of access to your storage, then perhaps storage-as-a-service is not something that makes sense.
- In many instances, those who leverage storage-as-a-service are surprised to find that they cannot access their shared disk space when not connected to the Internet, such as when on a plane.
- Performance can be an issue. When compared to on-premise storage, where the disks are physically located near the applications that leverage them, storage-as-a-service does not provide the same performance.
- If performance is a critical success factor, storage-as-a-service may not be the approach we want to leverage. Performance is usually about half the speed on a typical Internet connection when compared with a local network.
- We can use faster connections, but the cost of implementing a higher speed network connection quickly diminishes the cost savings of storage-as-a-service.
- The cost of the storage-as-a-service provider can be prohibitive when compared with an on-premise solution. While SOA using

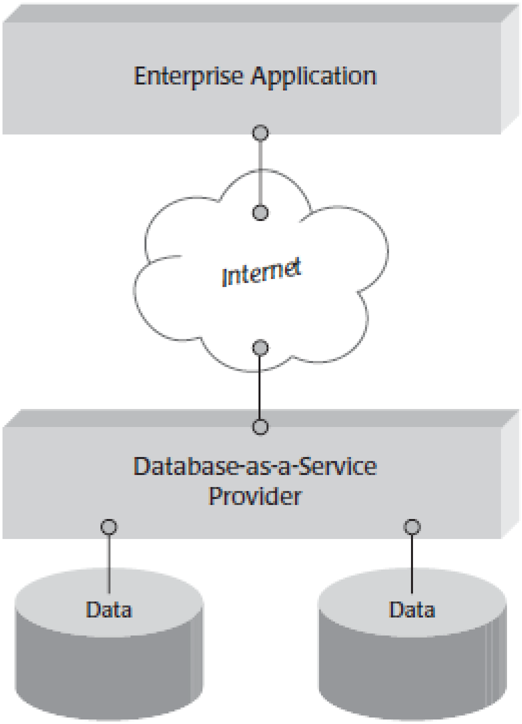
cloud computing is cost effective in some instances, in many instances it is not.

- The cost effectiveness of cloud computing is enterprise and domain dependent. For instance, a shared disk in a storage-as-a-service solution would be of high value for a virtual business with a distributed employee base. It would save on hardware and maintenance and would provide easy sharing of disk space as well. However, if the employees or applications are in the same building, the benefits of storage-as-a-service versus on-premise storage solutions are not as compelling.

6.2 Database-as-a-Service

- Database-as-a-service provides the ability to leverage the services of a remotely hosted database, sharing it with other users and having it logically function as if the database were local.
- We can self-provision a database, create the tables, load the data, and access the data using the interface provided, all on demand and via cloud computing.
- Like storage-as-a-service, database-as-a-service provides access to a resource that we neither own nor host and thus saves you the hardware, software, and maintenance costs.
- Database-as-a-service allows you to access enterprise-grade databases over the Internet.
- With the self-provisioning capabilities, we can think about a new database at 8:00 AM and have it running by noon, without buying hardware or software and without even leaving our office. This makes it incredibly easy to provision a database as needed.
- Database services include everything that you can do with a local database, such as setting up the tables and the relations among them, adding data, extracting data, and deleting data.
- Database-as-a-service providers provide not only basic database functions but also brand-specific services such as Oracle, Sybase, and Microsoft, so you can leverage proprietary features if we need them.

- An effective database-as-a-service provider should be able to offer database services that appear local in terms of performance and functionality.



Benefits of Database-as-a-service:

- The ability to avoid hardware and software costs by leveraging a remote database that we use as we need it and just what we need to use.
- Database licensing costs are a major part of the software budget, and avoiding those costs will go right to the bottom line.
- Database maintenance, including backing up and restoring the database and managing users, can be avoided through the use of database as-a-service.
- We do not have to focus on the maintenance activities required for a database; we can focus instead on its design and use.
- We can avoid the task of doing upgrades and bug fixes to the database.

- Many a DBA (databases administer) has spent a great deal of time applying patches and fixes to enterprise databases. Using database-as-a-service providers, that activity is handled for us and is transparent to us.
- We should always have the most current bug-free version of the database engine, since it is centrally updated on the cloud computing site and nothing needs to be distributed.

Drawbacks of Database-as-a-service

- There are legal, compliancy, and privacy issues around data, and in some instances, leveraging remote databases is illegal and/or not within compliance for some types of data.
- We must check before hosting data remotely, but in most cases, remote hosting is just fine and should meet our security requirements.
- Security can be an issue when using database-as-a-service. When we require complete security, the use of remote databases that we do not control or secure may be contraindicated, depending on the type of data you place in those databases. However, there is no reason you cannot have your data exist securely on a database-as-a-service cloud offering if you leverage the right approach to security for your SOA and the right security technology.
- We need to work closely with our database-as-a-service provider and consider our own requirements to determine the best approach to secure our database.
- Many of the interfaces offered by database-as-a-service providers are proprietary in nature and thus can be difficult to leverage from applications that need to access the data. While many cloud computing providers are moving toward standard interfaces, we need to understand and test their interfaces and/or APIs.
- Some database-as-a-service providers offer only a subset of the capabilities found in traditional on-premise enterprise databases. We may find that you are missing features and functions required by the enterprise applications.

- For example, stored procedures and triggers may not be supported in the same manner as in on-premise databases, or they may be proprietary, and thus difficult to port if you need to move off the database-as-a service provider at some point in the future.

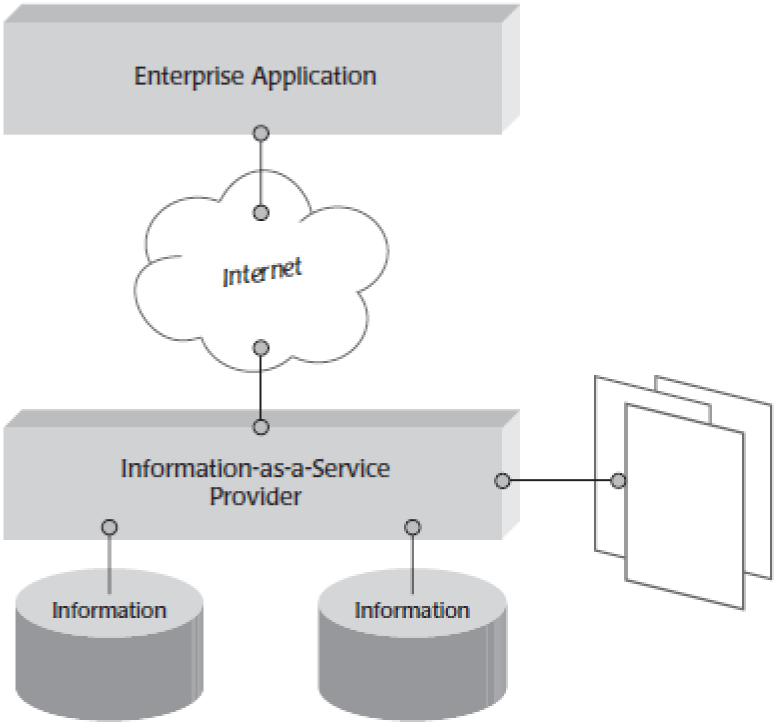
6.4 Information-as-a-Service

- Information-as-a-service refers to the ability to consume any type of remotely hosted information—stock price information, address validation, credit reporting, for example—through a well-defined interface such as an API.
- Over a thousand sources of information can be found these days, most of them listed at www.programmableweb.com.
- While they typically “serve up” information using standard Web Services APIs, some use proprietary interfaces.
- Therefore, as you must for database-as-a-service, you need to consider the interfaces offered by information-as-a-service providers.
- Typically, APIs function like this:

```
GetSSNName(SSN_Number);
        or
GetSSNName(333-33-3333);
with the return of
        "John H. Smith"
```

- Information-as-a-service allows any application to access any type of information using an API.
- You can leverage a wide variety of Web APIs these days, including APIs for social networking sites like Twitter and Facebook, for business statistics, for stock quotes, and the list goes on.
- As far as cloud computing categories go, information-as-a-service is the most eclectic.
- We use information through these APIs for several reasons, including the ability to mix and match a variety of information from many different sources through a single application or mashup.

- The idea is that it is much cheaper to leverage information that other people maintain and host than it is to host it yourself.



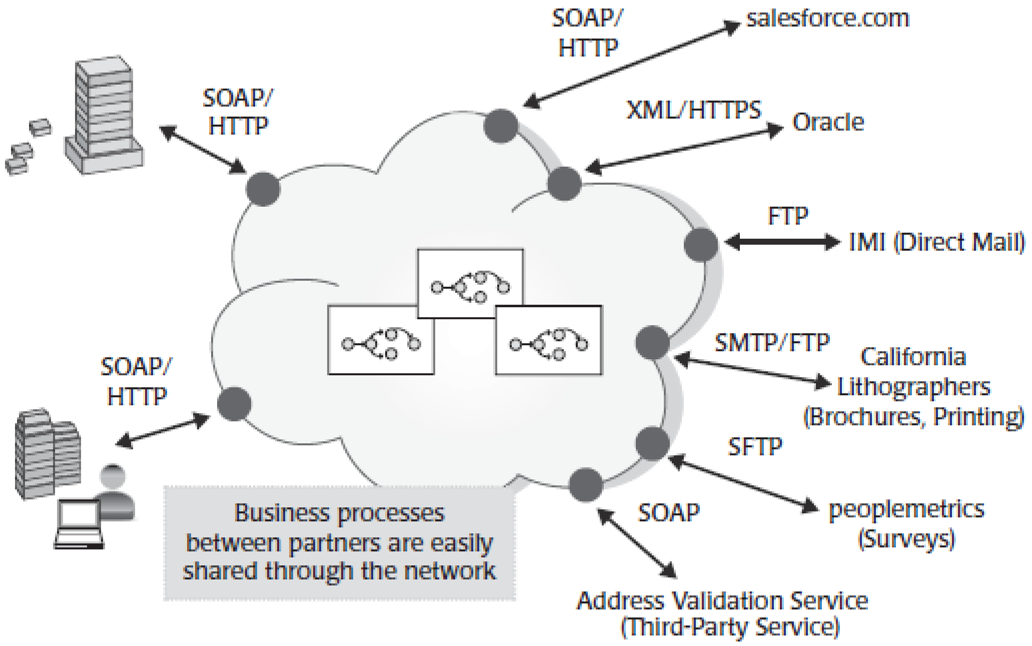
- Let's explore these concepts, focusing on the costs, the benefits, and the business case in the context of leveraging information-as-a-service, or Web APIs.
- The core value of leveraging Web APIs is that you do not have to incur the cost of creating or hosting the API or the information it abstracts. While most hang their value hat on that truth, there is indeed cost to leveraging an outside API:
    - Cost of binding APIs into applications or processes, including abstracting an API to fit an application or process.
    - Cost of inefficiencies brought about by the use of the API, such as downtime, or decreased speed.
    - Cost of ongoing maintenance as APIs and applications change.
    - Cost of the API service itself, typically per use.

- Since you bind an application to a remote resource of a network, we have to account for the remote resource—the network—being down from time to time. Furthermore, there is always ongoing maintenance and the cost of the service itself. So,

  - Onetime cost = cost of binding and abstraction
  - Ongoing cost = cost of downtime + ongoing maintenance + cost of the API service

6.5 Process-as-a-Service

- Process-as-a-service refers to a remote resource that can bind many resources together, either hosted within the same cloud computing resource or remotely, to create business processes.
- Process-as-a-service allows you to bind on-premise or cloud-delivered resources together to form business solutions.


SOAP/ HTTP
salesforce.com
SOAP/ HTTP
XML/HTTPS
Oracle
FTP
IMI (Direct Mail)
SMTP/FTP
California Lithographers (Brochures, Printing)
SOAP/ HTTP
SFTP
peoplemetrics (Surveys)
SOAP
Business processes between partners are easily shared through the network
Address Validation Service (Third-Party Service)

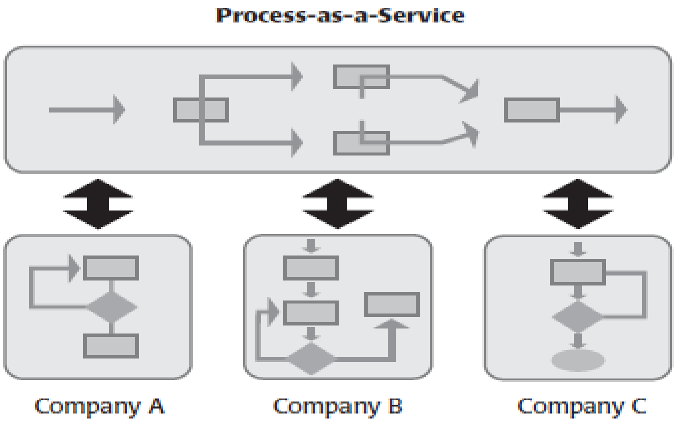
- The processes are meta-applications that bind many services and information together to form a business solution. Because they follow a configuration rather than a programmatic approach, it is often easier to create and change processes using a graphical interface than to write new programs.

- Process-as-a-service provides a mechanism to bind other resources together to form a solution. While our information and APIs may be hosted within a cloud provider, or perhaps on-premise, you would leverage this service to abstract and bind these resources together to form a business solution.
- We can think of processes as a sequence of events that must occur in a certain order, leveraging any number of services and portions of data. For example,

  Process "Ship Product"

  1. Transmit order to warehouse.
  2. Process shipping provider.
  3. Price shipping.
  4. Turn over to shipping provider.
  5. Track shipment.
  6. Report to customer.
- Each step above includes services called by the process, but the services themselves are not processes. Processes provide control instructions about how to do something using many resources that can exist on-premise or in the clouds.
- Processes can span a single enterprise or, more often, many enterprises when dealing with process-as-a-service.
- Process engines are really nothing new, although the existence of process engines on demand is. As we move forward with cloud computing, the use of process engines to leverage and manage any number of local and remote services to form them into business solutions will be an important component to cloud computing and to SOA using cloud computing.

- Process-as-a-service allows you to create common processes that span many



companies, cloud services, and on-premise services.

6.6 Application-as-a-Service, Software-as-a-service

- Application-as-a-service, also known as software-as-a-service, is any application delivered over the platform of the Web to an end user, typically leveraging the application through a browser.
- While many associate application-as a- service with enterprise applications, such as Salesforce SFA, office automation applications are indeed applications-as-a-service as well, including Google Docs, Gmail, and Google Calendar.
- They typically offer

  _ A user interface.

  _ Predefined application behavior.

  _ Predefined data.

  _ Support for any number of client platforms, since they run through the browser.
- Application-as-a-service was really the first drive into modern cloud computing, but it is based on the more traditional time-sharing model from years past whereby many users shared one application and one computer.
- The differences are that we use a Web browser, not a terminal, and the applications are typically sold by subscription, not by time. Some

are free of charge and obtain revenue through advertising or in other ways.

- The advantage of application-as-a-service is the ability to leverage an enterprise-class application without having to buy and install enterprise software.
- Thus, business functionality typically only available to those who could afford SAP, Oracle Financials, and other larger packaged systems are available to any business user for a small subscription fee.
- Indeed, Salesforce.com became a multibillion dollar business using this model, and other application as- a-service providers are catching up quickly, including many providing specialized applications for human resources, logistics management, and trade risk management, to name just a few.
- Many application-as-a-service providers offer API access to their internal behaviours and data. Their customers need programmatic access to the application behavior and the information for any number of purposes, such as integration, or the ability to leverage services from an application-as-a-service provider for their on-premise enterprise applications.
- In addition to the larger business application-as-a-service, there are also the office automation applications-as-a-service, including e-mail, document management, word processing, spreadsheets, and other productivity applications delivered through a browser. Google provides these applications for free, as do a few other providers. Some charge a small subscription fee.

Benefits of Applications-as-a-service:

- **Cost**, because it is typically free. However, you can use Sun's Open Office open source office automation software on your desktop, which is also free, just to be fair.
- **Convenience**, since any computer with a browser can become your personal workspace with access to your documents and e-mail.

6.7 Platform-as-a-Service

- Platform-as-a-service is a complete platform, including application development, interface development, database development, storage, and testing, delivered through a remotely hosted platform to subscribers.
- Based on the traditional time-sharing model, modern platform-as-a-service providers offer the ability to create enterprise-class applications for use locally or on demand for a small subscription price or for free.
- Platform-as-a-service is one-stop shopping for those looking to build and deploy applications.
- Platform-as-a-service provides self-contained platforms with everything you need for application development and operational hosting. Platforms such as Google App Engine and Force.com (part of Salesforce.com) are popular ways to approach application development on the cloud.
- Core to the platform-as-a-service notion are a few **major components**:
  - Design,
  - Development,
  - Deployment,
  - Integration,
  - Storage, and
  - Operations.
- **Design** is the ability to design our application and user interfaces.
- **Development** is the ability to design, develop, and test applications right out of the platform, on demand, using development tools that are delivered on demand.
- **Deployment** is the ability to test, bundle, and deliver the platform-as-a service– created applications. This means hosting the applications, typically accessing them visually, through a browser, or as Web services.

- **Integration** is the ability to integrate the applications developed on our platform-as-a-service provider with software-as-a-service applications or applications that may exist within your enterprise.
- **Storage,** the ability to provide persistence for the application, means an on-demand database or on-demand file storage.
- **Operations** is the ability to run the application over a long period of time, dealing with backup, restore, exception handling, and other things that add value to operations.

Benefits of Platform-as-a-service:

- We can access a complete enterprise-class development environment at a low cost and build complete enterprise applications, from the data to the user interface.

Drawbacks of Platform-as-a-service:

- Many of the platform-as-a-service vendors leverage proprietary programming languages and interfaces; thus, once your application is there, it may be difficult to move it to an on-premise server or another platform-as-a service provider.

6.8 Integration-as-a-Service

- Integration-as-a-service is the ability to deliver a complete integration stack from the cloud, including interfacing with applications, semantic mediation, flow control, and integration design.
- Integration-as-a-service includes most of the features and functions found within traditional EAI (enterprise application integration) technology but delivered as a service.
- Integration is a tough problem to solve, and integration on demand does not make that any easier.
- Any integration engine, on-premise or in the cloud, has to support some basic functions, including:
    - Transformation
    - Routing
    - Interface
    - Logging

- **Transformation** means that we can convert the information semantics from one system to the information semantics of another system, so the target system can receive information in a format it understands.
- **Routing** means that information is routed to the correct systems on the basis of predefined logic (called intelligent routing).
- **Interface** means that we can connect into the source or target systems using whatever interface they expose.
- **Logging** means that we can log all integration activities, such as messages flowing in and out, as well as other events.

Benefits of Integration-as-a-service:

- We can access pretty pricy integration software functionality for the price of a rental agreement.
- Many of the integration-on-demand providers have very sophisticated software delivered through a browser that leverages the new rich Internet application technology such as AJAX.

Drawbacks of Integration -as-a-service:

- There are many firewall mediation issues to deal with.
- Many systems you may want to integrate do not have Port 80–compliant interfaces and protocols, meaning they cannot speak outside of the firewall to the remote, on-demand integration server. Thus, many integration-on demand providers leverage software that has to exist behind the firewall to mediate the differences in the local, native interfaces and turn them into something that can be sent outside of the firewall, typically Web Services that leverage Port 80–compliant Simple Object Access Protocol (SOAP).
- We end up with an on-premise footprint that diminishes the value of an integration-on-demand solution.

6.9 Security-as-a-Service

- Security-as-a-service, is the ability to deliver core security services remotely over the Internet.

- While the security services provided today are often rudimentary, more sophisticated services, such as identity management, are becoming available.
- Security-as-a-service is a tough sell considering that security is typically a weak point of cloud computing.
- Providing security on demand seems like an unnatural act. However, there are times when security delivered out of the cloud makes sense, such as for securing a cluster of cloud resources we are leveraging within your enterprise or even between enterprises.
- Thus, we can enforce security hierarchies between physical organizations out of the cloud or perhaps have cloud-delivered on-demand encryption services or identity management solutions.
- However, as time goes on and security on demand becomes more sophisticated, and as more corporate data and applications reside in the clouds, then there will be an uptake in security-as-a-service.

6.10 Management/Governance-as-a-Service

- Management/governance-as-a-service is any on-demand service that provides the ability to manage one or more cloud services, typically simple things such topology, resource utilization, virtualization, and uptime management.
- Governance systems, such as the ability to enforce defined policies on data and services, are becoming available as well.
- Most enterprises like to control management and governance. However, as more applications and data are outsourced, it may make sense to manage and govern those resources from the clouds as well.

6.11 Testing-as-a-Service

- Testing-as-a-service is the ability to test local or cloud-delivered systems using remotely hosted testing software and services.
- While a cloud service requires testing unto itself, testing-as-a-service systems have the ability to test other cloud applications, Web sites, and internal enterprise systems, and they do not require a hardware or software footprint within the enterprise.

Benefits of Testing-as-a-service:

- The advantages of testing-as-a-service include the ability to avoid purchasing test servers and testing software.
- In many respects, testing, and either on-premise or in the clouds, is better done through a testing service that connects to those applications over the Internet, since many real-life users will do the same thing.
- If we are looking to test a Web site or a Web-delivered application, testing-as-a-service is actually more logical than testing on-premise in many instances.
- The downsides are the ones you might expect. Many of those who build and deploy applications like to control their testing environments and would not dream of leveraging testing servers and software that they do not own or host.
- Again, as more applications are re-hosted in the cloud, testing-as-a-service will become more of an accepted paradigm.

6.12 Infrastructure-as-a-Service

- Infrastructure-as-a-service is a data center-as-a-service, or the ability to access computing resources remotely.
- We lease a physical server that is ours to do with what we will and that for all practical purposes is our data center, or at least part of a data center.
- The difference with this approach versus more mainstream cloud computing is that instead of using an interface and a metered service, we get access to the entire machine and the software on that machine.
- We defined database-as-a-service, storage-as-a-service, and so on, as separate categories of cloud computing. Infrastructure-as-a-service can provide all of them, including database, storage, governance, application development, application processing, security, and more. Anything that can be found in a traditional data center can be delivered as an infrastructure-as-a-service.

Benefits of Infrastucture-as-a-service:

- The advantage of infrastructure-as-a-service is that you can access very expensive data center resources through a rental arrangement and thus preserve capital for the business.
- Somebody is there to manage the physical machines for you, including replacement of downed disk drives and correction of any networking issues.

Drawbacks of Infrastucture-as-a-service:

- The disadvantage is that there is typically less granular on-demand expandability of the resource.
- With database-as-a-service and storage-as-a service, we just purchase additional capability as you need it and as much as we need. However, many infrastructure-as-a-service providers require that you lease an entire server for a defined amount of time. Thus, the whole selling point of adjusting your cloud resources to meet your exact needs kind of goes out the door.

UNIT-VI

Assignment-Cum-Tutorial Questions

SECTION-A

Objective Questions

1. ____________ is the ability to leverage storage that physically exists remotely but is logically a local storage resource to any application that requires storage. [ ]

(a) Storage-as-a-service (b) Database-as-a-service
(c) Information-as-a-service (d) Process-as-a-service

2. The most primitive component of cloud computing is ______. [ ]

(a) Storage-as-a-service (b) Database-as-a-service
(c) Information-as-a-service (d) Process-as-a-service

3. _________ provides the ability to leverage the services of a remotely hosted database, sharing it with other users and having it logically function as if the database were local. [ ]

(a) Storage-as-a-service (b) Database-as-a-service
(c) Information-as-a-service (d) Process-as-a-service

4. ____________ refers to the ability to consume any type of remotely hosted information. [ ]

(a) Storage-as-a-service (b) Database-as-a-service
(c) Information-as-a-service (d) Process-as-a-service

5. ____________ refers to a remote resource that can bind many resources together. [ ]

(a) Storage-as-a-service (b) Database-as-a-service
(c) Information-as-a-service (d) Process-as-a-service

6. ______ was really the first drive into modern cloud computing.[ ]

(a) Storage-as-a-service (b) Database-as-a-service
(c) Information-as-a-service (d) Application-as-a-service

7. ______________________ is any application delivered over the platform of the Web to an end user, typically leveraging the application through a browser. [ ]

(a) Storage-as-a-service (b) Database-as-a-service
(c) Information-as-a-service (d) Application-as-a-service

8. ____________ is the ability to test, bundle, and deliver the platform-as-a service– created applications. [ ]

(a) Design (b) Development (c) Deployment (d) Integration

9. ____________ is the ability to run the application over a long period of time, dealing with backup, restore, exception handling. [ ]

(a) Design (b) Development (c) Deployment (d) Operations

10.Converting the information semantics from one system to the information semantics of another system, so the target system can receive information in a format it understands. [ ]

(a) Transformation (b) Routing (c) Interface (d) Logging

11. SOAP stands for________________________________.

12. __________________ is the ability to deliver core security services remotely over the Internet. [ ]

(a) Storage-as-a-service (b) Database-as-a-service

(c) Security-as-a-service (d) Application-as-a-service

13. _________________ is any on-demand service that provides the ability to manage one or more cloud services. [ ]

(a) Management-as-a-service (b) Database-as-a-service

(c) Security-as-a-service (d) Application-as-a-service

14. Testing-as-a-service is the ability to test local or cloud-delivered systems using remotely hosted testing software and services. [ ]

(a) Management-as-a-service (b) Testing-as-a-service

(c) Security-as-a-service (d) Application-as-a-service

15. ______________________ is a data center-as-a-service and the ability to access computing resources remotely. [ ]

(a) Management-as-a-service (b) Testing-as-a-service

(c) Security-as-a-service (d) Infrastructure-as-a-service

SECTION-B

SUBJECTIVE QUESTIONS

1.List the components of Cloud Computing.

2.Explain how Storage-as-a-service allows us to store information on a remote disk drive as if it were local.

3.Summarize the benefits and drawbacks of Storage-as-a-service.

4.Explain Database-as-a-service providers.

5.Justify how Information-as-a-service has the ability to consume any type of remotely hosted information.

6.Discuss how Process-as-a-service allows us to bind on-premise or cloud-delivered resources together to form business solutions.

7.Explain the following:

(i) Application-as-a-service

(ii) Security-as-a-service

(iii) Infrastructure-as-a-service

8.Explain the major components of Platform-as-a-service?

9.Explain the major functions of an Integration Engine.