

Cloud Computing

DECAP470

Edited by
Ajay Kumar Bansal



LOVELY
PROFESSIONAL
UNIVERSITY



L O V E L Y
P R O F E S S I O N A L
U N I V E R S I T Y

Cloud Computing
Edited By:
Ajay Kumar Bansal

Title: CLOUD COMPUTING

Author's Name: Dr. Tarandeep Kaur

Published By : Lovely Professional University

Publisher Address: Lovely Professional University, Jalandhar Delhi GT road, Phagwara - 144411

Printer Detail: Lovely Professional University

Edition Detail: (I)

ISBN: 978-93-94068-33-9



Copyrights@ Lovely Professional University

CONTENT

Unit 1:	Cloud Computing Introduction	1
	<i>Dr. Tarandeep Kaur, Lovely Professional University</i>	
Unit 2:	Cloud Computing Architecture and Models	20
	<i>Dr. Tarandeep Kaur, Lovely Professional University</i>	
Unit 3:	Cloud Services	46
	<i>Dr. Tarandeep Kaur, Lovely Professional University</i>	
Unit 4:	Introduction to Big Data	74
	<i>Dr. Tarandeep Kaur, Lovely Professional University</i>	
Unit 5:	File system in Cloud	100
	<i>Dr. Tarandeep Kaur, Lovely Professional University</i>	
Unit 6:	Collaborating using Google Cloud	117
	<i>Dr. Tarandeep Kaur, Lovely Professional University</i>	
Unit 7:	Collaborating on Event Management	146
	<i>Dr. Tarandeep Kaur, Lovely Professional University</i>	
Unit 8:	Collaborating on Project Management	172
	<i>Dr. Tarandeep Kaur, Lovely Professional University</i>	
Unit 9:	Collaborating on Databases and Web-based Communication	191
	<i>Dr. Tarandeep Kaur, Lovely Professional University</i>	
Unit 10:	Virtualization Concepts	215
	<i>Dr. Tarandeep Kaur, Lovely Professional University</i>	
Unit 11:	Virtual Machine	233
	<i>Dr. Tarandeep Kaur, Lovely Professional University</i>	
Unit 12:	Security and standards in Cloud	254
	<i>Dr. Tarandeep Kaur, Lovely Professional University</i>	
Unit 13:	Application of Cloud Computing and Related Technologies	273
	<i>Dr. Tarandeep Kaur, Lovely Professional University</i>	
Unit 14:	Additional Cloud Concepts	305
	<i>Dr. Tarandeep Kaur, Lovely Professional University</i>	

Unit 01: Cloud Computing Introduction

CONTENTS

Objectives

Introduction

1.1 What is Cloud Computing?

1.2 Cloud Properties: Google's Perspective

1.3 History of Cloud Computing

1.4 Evolution of Cloud Computing

1.5 Essential Cloud Computing Concepts

1.6 Components of Cloud Computing

1.7 Characteristics of Cloud Computing

1.8 Issues of Cloud Computing

1.9 Challenges in Cloud Computing

1.10 Applications of Cloud Computing

Summary

Keywords

Self Assessment

Answers for Self Assessment

Review Questions

Further Readings

Objectives

After studying this unit, you will be able to:

- review the fundamentals of cloud computing and related concepts.
- identify the history and evolution of cloud computing.
- understand the characteristics and advantages of cloud computing.
- explore and learn about the different cloud components.
- discover the challenges faced in cloud computing.
- investigate the applications of cloud computing.

Introduction

Cloud refers to a Network or Internet. Cloud is something, which is present at remote location. Cloud can provide services over network, that is, on public networks or on private networks, that is, Wide Area Networks (WANs), Local Area Networks (LANs), or Virtual Private Networks (VPNs). Applications such as e-mail, web conferencing, customer relationship management (CRM), all run in cloud.

Cloud Computing

Figure 1: Examples of Cloud Computing

1.1 What is Cloud Computing?

Cloud computing has embarked a revolution in accessing, provisioning and consumption of the information and computing in the ICT industry. It has emerged as a novel paradigm of high-performance and large-scale computing that actuates relocation of computing and data from desktops and personal computers to big data centers. Cloud is a construct (infrastructure) that allows to access application that actually resides at a remote location of another internet connected device, most often, this will be a distant datacenter. Cloud computing takes the technology, services, and applications that are similar to those on the Internet and turns them into a self-service utility (Figure 1).

Cloud provides an abstraction based on the notion of pooling physical resources and presenting them as a virtual resource. It is a new model for provisioning resources, for staging applications, and for platform-independent user access to services.

"Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computer resources (networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction" (Figure 2).

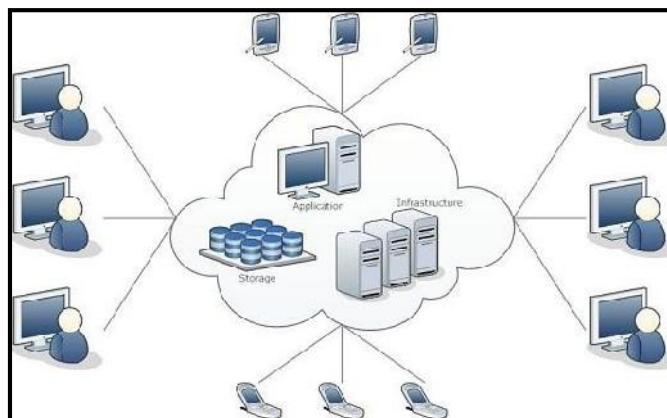


Figure 2: Cloud Scenario

Cloud computing refers to manipulating, configuring, and accessing the applications online. It offers online data storage, infrastructure and application and involves both a combination of software and hardware-based computing resources delivered as a network service. Example: Suppose we want to install MS-Word in our organization's computer. We have to buy the CD/DVD of it and install it or can setup a S/W distribution server to automatically install this application on your machine. Every time Microsoft issues a new version, we have to perform the same task. If some other company hosts your application, that is, they handle the cost of servers and manage the software updates. The customers are charged as per their utilization, that is, as per

the usage (Figure 3). It reduces the cost of using that software along with the reduction in the cost of installation of heavy servers. Additionally, cloud aids in reducing the cost of electricity bills.

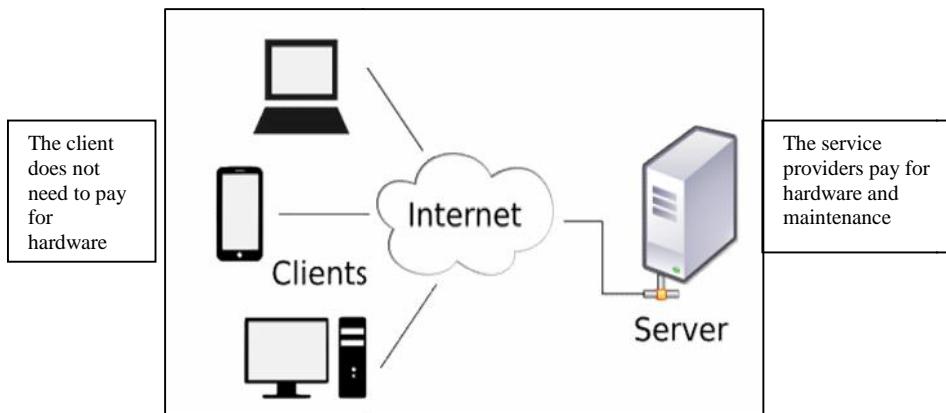


Figure 3: Cloud Client-Server Perspective

Clouds can come in many different types, and services and applications that run on clouds may or may not be delivered by a cloud service provider. The different types and levels of cloud services mean that it is important to define what type of cloud computing system you are working with.

Cloud Computing Collaboration

With the growth of the Internet, there was no need to limit group collaboration to a single enterprise's network environment. The users from multiple locations within a corporation, and from multiple organizations, desired to collaborate on projects that crossed company and geographic boundaries. 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. Cloud-collaboration is also termed as Internet-based group collaboration.

1.2 Cloud Properties: Google's Perspective

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. In effect, any device that accesses your data in the cloud also becomes yours.

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.



Task: Explore the different cloud services offered by Google.

Cloud Computing

1.3 History of Cloud Computing

Before emerging the cloud computing, there was Client/Server computing which is basically a centralized storage in which all the software applications, all the data and all the controls are resided on the server side. If a single user wants to access specific data or run a program, he/she need to connect to the server and then gain appropriate access, and then he/she can do his/her business (Figure 4).

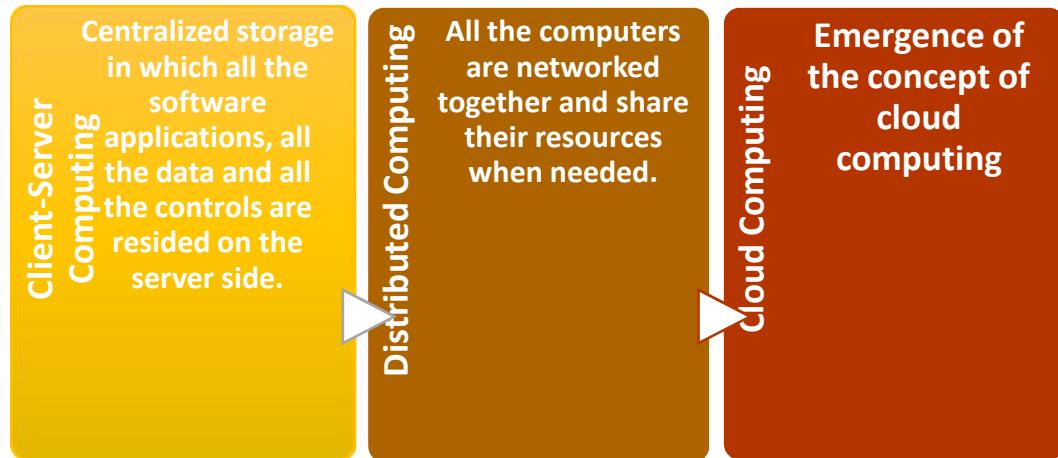


Figure 4: Convergence to Cloud Computing

Then after, distributed computing came into picture, where all the computers are networked together and share their resources when needed. On the basis of above computing, there was emergence of the concept of cloud computing.

Cloud computing was invented in the early 1960s by J.C.R Licklider (Joseph Carl Robnett Licklider), an American Psychologist and Computer Scientist. During his network research work on ARPANet (Advanced Research Project Agency Network), trying to connect people and data all around the world, he gave an introduction to cloud computing technique which we all know today.

Born on March 11th, 1915 in St. Louis, Missouri, US, J.C.R Licklider pursued his initial studies from Washington University in 1937 and received a BA Degree with three specializations including physics, maths, psychology. Later in the year 1938, Licklider completed his MA in psychology and received his Ph.D. from the University of Rochester in the year 1942. His interest in Information Technology and looking at his years of service in different areas and achievements, made him appointed as Head of IPTO at ARPA (US Department of Defense Advanced Research Project Agency) in the Year 1962. His aim led to ARPANet, a forerunner of today's Internet.

At around in 1961, John MacChart suggested in a speech at MIT that computing can be sold like a utility, just like a water or electricity. In 1999, Salesforce.com started delivering of applications to users using a simple website. The applications were delivered to enterprises over the Internet, and this way the dream of computing sold as utility were true.

The beauty of the cloud computing phase went on running throughout the era of the 21st Century. In 2002, Amazon started Amazon Web Services, providing services like storage, computation and even human intelligence. However, only starting with the launch of the Elastic Compute Cloud in 2006 a truly commercial service open to everybody existed.

By 2008, Google too introduced its beta version of the search engine. In 2009, Google Apps also started to provide cloud computing enterprise applications. Earlier announced by Microsoft in the year 2008, it released its cloud computing service named Microsoft Azure for testing, deployment and managing applications and services.

In the year 2012, Google Compute Engine was released but was rolled to the public. By the end of Dec 2013, Oracle introduced Oracle Cloud with three primary services for business (IaaS, PaaS and SaaS). Currently, as per records, Linux and Microsoft Azure share most of their work parallel.

1.4 Evolution of Cloud Computing

Growth of cloud computing is not an instantaneous task but has transited from several intermediate stages. Beginning from an era of mainframe computing where huge and powerful mainframe systems supported many users connected through dummy terminals and running data management applications. There exist five intermediary stages from mainframe computing to the use of personal stand-alone computers running desktop applications deriving personal computing to the influx of interconnected computers converging to networking computing. This stage saw the growth of local networks connected to other networks creating a globally interconnected network such as the Internet for utilizing remote applications and resources. The networked computers usually functioned in autonomic fashion resulting in autonomic computing or followed client-server architectures resulting in client-server computing. Figure 5 lists the evolution of cloud computing. The development of grid computing followed by the rise of cloud computing and is characterized as:

- Development of grid computing offered sharing of computing power and resources spread across multiple geographical domains.
- The recent stage involves rise of cloud computing where service-oriented, market-based computing applications are predominant.
- Virtualization meets the Internet.

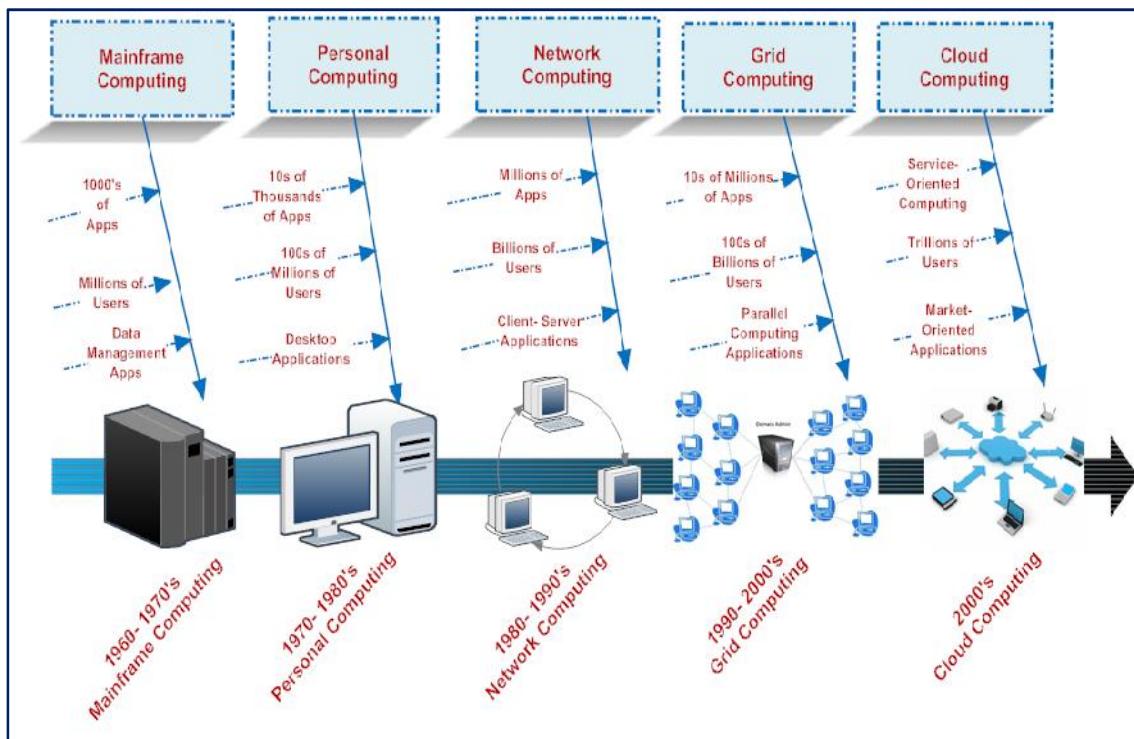


Figure 5: Evolution of Cloud Computing

1.5 Essential Cloud Computing Concepts

"Cloud" makes reference to the two essential concepts:

1. **Abstraction:** Cloud computing abstracts the details of system implementation from users and developers. Applications run on physical systems that aren't specified, data is stored in locations that are unknown, administration of systems is outsourced to others, and access by users is ubiquitous (Present or found everywhere).
2. **Virtualization:** Cloud computing virtualizes systems by pooling and sharing resources. Systems and storage can be provisioned as needed from a centralized infrastructure, costs

Cloud Computing

are assessed on a metered basis, multi-tenancy is enabled, and resources are scalable with agility.

Who Benefits from Cloud Computing?

- **Collaborators:** If you often collaborate with others on group projects. The ability to share and edit documents in real time between multiple users is one of the primary benefits of web-based applications. Example, that you're in charge of an upcoming presentation to the senior management of your company. You need to work with the heads of your company's various departments, which happen to be based in a half-dozen locations. Given everyone's busy schedules, it's tough enough to schedule a group conference call. How in the world can all of you get together to create a cohesive presentation? Just like Google Presentations!
- **Road Warriors:** When you work at one office today, at home the next day, and in another city the next, it's tough to keep track of all your documents and applications. You may end up with one version of a document on your work PC, another on your laptop, and a third on your home PC—and that's if you remember to copy that document and take it with you from one location to the next. When you're in the office, you log in to your web-based app and access your stored document. Even if you travel to another city and the same application and document are still available to you.
- **Cost-Conscious Users:** Another group of users who should gravitate to cloud computing are those who are cost conscious. With cloud computing you can save money on both your hardware and software. Hardware-wise, there's no need to invest in large hard disks or super-fast CPUs. Because everything is stored and run from the web, you can cut costs by buying a less fully.
- **Cost-Conscious IT Departments:** Many corporate IT departments are also becoming enamored of the cloud computing model. Although they might appreciate the software savings we just discussed, for them bigger savings result from having to buy fewer central servers.
- **Users with Increasing Needs:** Hardware-based cost savings also apply to individual computer users. Do you need more hard-disk space to store all your digital photos and MP3 files? You could purchase a new external hard drive, or you could utilize lower-cost (or free) cloud storage instead.

Advantages of Cloud Computing

Cloud computing is an emerging technology that almost every company is being switched to from its on-premises technologies. Whether it is public, private, or hybrid, cloud computing has become an essential factor for companies to rise up to the competition. Let us find out why the cloud is so much preferred over the on-premises technologies.

Cost Efficiency: The biggest reason behind companies shifting to cloud computing is that it takes considerably lesser cost than any on-premise technology. Now, companies need not store data in disks anymore as the cloud offers enormous storage space, saving money and resources. CapEx and OpEx costs are reduced because resources are only acquired when needed and are only paid for when used.

High speed: Cloud computing lets us deploy the service quickly in fewer clicks. This quick deployment lets us get the resources required for our system within minutes.

Excellent Accessibility: Storing information in the cloud allows us to access it anywhere and anytime regardless of the machine making it a highly accessible and flexible technology of the present times.

Back-up and Restore data: Once data is stored in the cloud, it is easier to get its back-up and recovery, which is quite a time-consuming process in on-premise technology.

Manageability: Cloud computing eliminates the need for IT infrastructure updates and maintenance since the service provider ensures timely, guaranteed, and seamless delivery of our services and also takes care of all the maintenance and management of our IT services according to the service-level agreement (SLA).

Sporadic Batch Processing: Cloud computing lets us add or subtract resources and services according to our needs. So, if the workload is not 24/7, we need not worry about the resources and services getting wasted and we won't end up stuck with unused services.

Strategic Edge: Cloud computing provides a company with a competitive edge over its competitors when it comes to accessing the latest and mission-critical applications that it needs without having to invest its time and money on their installations.

Disadvantages of Cloud Computing

Every technology has both positive and negative aspects that are highly important to be discussed before implementing it.

Vulnerability to Attacks: Storing data in the cloud may pose serious challenges of information theft since in the cloud every data of a company is online. Security breach is something that even the best organizations have suffered from and it's a potential risk in the cloud as well. Although advanced security measures are deployed on the cloud, still storing confidential data in the cloud can be a risky affair.

Network Connectivity Dependency: Cloud computing is entirely dependent on the Internet. This direct tie-up with the Internet means that a company needs to have reliable and consistent Internet service as well as a fast connection and bandwidth to reap the benefits of cloud computing.

Downtime: Downtime is considered as one of the biggest potential downsides of using cloud computing. The cloud providers may sometimes face technical outages that can happen due to various reasons, such as loss of power, low Internet connectivity, data centers going out of service for maintenance, etc. This can lead to a temporary downtime in the cloud service.

Vendor Lock-In: When in need to migrate from one cloud platform to another, a company might face some serious challenges because of the differences between vendor platforms. Hosting and running the applications of the current cloud platform on some other platform may cause support issues, configuration complexities, and additional expenses. The company data might also be left vulnerable to security attacks due to compromises that might have been made during migrations.

Limited Control: Cloud customers may face limited control over their deployments. Cloud services run on remote servers that are completely owned and managed by service providers, which makes it hard for the companies to have the level of control that they would want over their back-end infrastructure.

1.6 Components of Cloud Computing

Cloud computing solution is made up of several elements and these elements make up the three components of a cloud computing solution (Figure 6).

- a) Clients
- b) The data center, and
- c) Distributed servers.

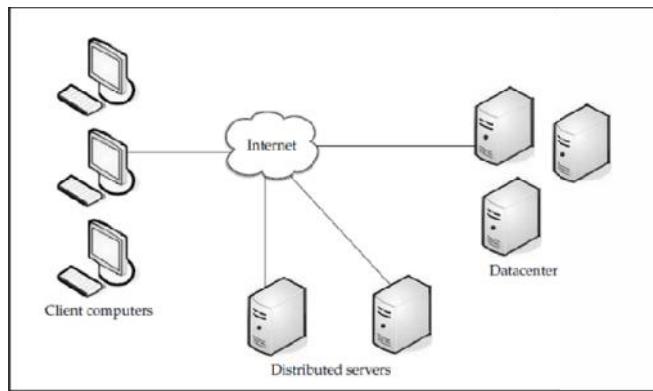
Cloud Computing

Figure 6: Cloud Computing Components

A. Clients: Devices that end users interact with to manage their information on cloud. There can be different types of clients such as:

- **Mobile Clients:** Includes PDAs or smartphones, like a Blackberry, Windows Mobile Smartphone, or an iPhone.
- **Thin Clients:** Computers that do not have internal hard drives, but rather let the server do all the work, but then display the information.
- **Thick Clients:** Thick clients are regular computer, using a web browser like Firefox or Internet Explorer to connect to the cloud.

A thin client is a computing device that's connected to a network. Unlike a typical PC or "fat client," that has the memory, storage and computing power to run applications and perform computing tasks on its own, a thin client functions as a virtual desktop, using the computing power residing on the networked servers.

Advantages of Using Thin Clients

Thin clients are becoming an increasingly popular solution, because of their price and effect on the environment.

- **Lower hardware costs:** Thin clients are cheaper than thick clients because they do not contain as much hardware. They also last longer before they need to be upgraded or become obsolete.
- **Lower IT costs:** Thin clients are managed at the server and there are fewer points of failure.
- **Security:** Since the processing takes place on the server and there is no hard drive, there's less chance of malware invading the device. Also, since thin clients don't work without a server, there's less chance of them being physically stolen.
- **Data security:** Since data is stored on the server, there's less chance for data to be lost if the client computer crashes or is stolen.
- **Less power consumption:** Thin clients consume less power than thick clients. This means you'll pay less to power them, and you'll also pay less to air-condition the office.
- **Ease of repair or replacement:** If a thin client dies, it's easy to replace. The box is simply swapped out and the user's desktop returns exactly as it was before failure.
- **Less noise:** Without a spinning hard drive, less heat is generated and quieter fans can be used on the thin client.

B. Datacenter: Datacenter has a collection of servers where the application to which you subscribe is housed. It is a large room in the basement of your building or a room full of servers on the other side of the world that you access via the Internet. There is a growing trend in the IT world of virtualizing servers. The software can be installed allowing multiple instances of virtual servers to be used. There can be half a dozen virtual servers running on one physical server.

C. Distributed Servers: The distributed servers are in geographically disparate locations. They give the service provider more flexibility in options and security. For instance, Amazon has their cloud solution in servers all over the world. If something were to happen at one site, causing a failure, the service would still be accessed through another site.

Other Components of Cloud Computing

There are other components of cloud computing such as cloud services; platforms; applications; storage; and infrastructure.

Cloud Services: Cloud services, products and solutions that are used and delivered real-time via internet media.Example:

- Identity - OpenID, OAuth, etc.
- Integration - Amazon Simple Queue Service.
- Payments - PayPal, Google Checkout.
- Mapping - Google Maps, Yahoo! Maps.

Cloud Applications: Applications that use cloud computing in software architecture so that users don't need to install but they can use the application using a computer.Example:

- Peer-to-peer - BitTorrent, SETI, and others.
- Web Application - Facebook.
- SaaS - Google Apps, SalesForce.com, and others

Cloud Platform: A service in the form of a computing platform consisting of hardware and infrastructure software. Service in the form of a computing platform which contains infrastructure hardware and software. Example:

- Web Application Frameworks - Python Django, Ruby on Rails, .NET
- Web Hosting
- Proprietary- Force.com

Cloud Storage: Cloud storage involves the process of storing data as a service.Example:

- Database- Google Big Table, Amazon SimpleDB.
- Network Attached Storage- NirvanixCloudNAS, MobileMe iDisk

Cloud Infrastructure: Cloud infrastructure involves the delivery of computing infrastructure as a service.Example:

- Grid Computing- Sun Grid.
- Full Virtualization- GoGrid, Skytap.
- Compute- Amazon Elastic Compute Cloud (Amazon EC2)

1.7 Characteristics of Cloud Computing

Cloud computing harnesses the power of the Internet to allow organizations to remain productive despite the COVID-19 pandemic and work from home arrangements. The technology helps businesses maximize their resources because they don't need to buy their physical servers. Everything is online. Everything is in the cloud.

On-demand self-service- A user can provision computing capabilities, such as server time and storage, as needed without requiring human interaction.

Broader network access- Capabilities are available over a network and typically accessed by the users' mobile phones, tablets, laptops, and workstations.

Shared resource pooling- The provider's computing resources are pooled to serve multiple users using a multi-tenant model, with different physical and virtual resources dynamically assigned and

Cloud Computing

reassigned according to consumer demand. Examples of resources include storage, processing, memory, and network bandwidth.

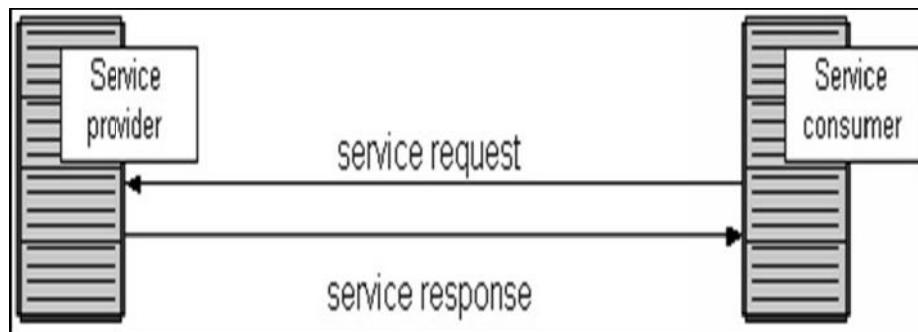
Rapid elasticity- Capabilities can be elastically provisioned and released, in some cases automatically, to scale rapidly outward and inward as needed. For the user, the capabilities available for provisioning often appear to be unlimited and can be appropriated in any quantity at any time.

Measured service- Cloud systems automatically control and optimize resource use by leveraging a metering capability appropriate to the type of service (e.g., storage, processing, bandwidth, and active user accounts). The resource usage can be monitored, controlled, and reported, providing transparency for both the provider and user of the service. It follows a "Pay as you grow" model or for internal IT departments to provide IT chargeback capabilities. The usage of cloud resources is measured and user is charged based on some metrics such as amount of CPU cycles used, amount of storage space used, number of network I/O requests etc. are used to calculate the usage charges for the cloud resources.

Performance- Dynamic allocation of resources as per the application workloads helps to easily scale up or down and maintain performance.

Reduced costs- Cost benefits for applications as only as much computing and storage resources are required can be provisioned dynamically and upfront investment in purchase of computing assets to cover worst case requirements is avoided.

Outsourced Management- Cloud computing allows the users to outsource the IT infrastructure requirements to external cloud providers and save upfront capital investments. This helps in easiness of setting IT infrastructure and pay only for the operational expenses for the cloud resources used.



Cost Management and Containment: Cloud computing can be expensive if you don't know how to manage your computing resources and take maximum advantage of them. Many times, the organizations dwell in a mindset of pay-as-you-go and spend more on cloud than they would have on on-premise infrastructure. One should always optimize the cost by financial analytics and reporting the usage for better monitoring of cost. For the most part, cloud computing can save businesses money. In the cloud, an organization can easily ramp up its processing capabilities without making large investments in new hardware. Businesses can instead access extra processing through pay-as-you-go models from public cloud providers. However, the on-demand and scalable nature of cloud computing services make it sometimes difficult to define and predict quantities and costs.

Lack of Resources/ Expertise: Cloud challenges companies and enterprises. As the usage of cloud technologies is increasing, tools to manage it are getting sophisticated, finding experts on top of this in cloud computing is becoming a bottleneck to many organizations. The organizations are increasingly placing more workloads in the cloud while cloud technologies continue to rapidly advance. Due to these factors, organizations are having a tough time keeping up with the tools. Also, the need for expertise continues to grow. Such challenges can be minimized through additional training of IT and development staff. Many companies are adopting automated cloud management technologies but it's always better to train individuals to satisfy the need of time. Presently, DevOps tools like Chef and Puppet are heavily used in the IT industry.

Governance/ Control: In cloud computing, infrastructure resources are under CSP's control and end-users or companies have to abide by the governance policies from CSP. The traditional IT teams have no control over how and where their data is stored and processed. IT governance should assure how infrastructure assets from CSP are being used. To overcome the downfalls and challenges, onboarding to Cloud, IT must adapt its orthodox way of governance and process control to the induct cloud. Now, IT is playing an important role in benchmarking cloud services requirements and policies. Thus, the proper IT governance should ensure IT assets are implemented and used according to agreed-upon policies and procedures; ensure that these assets are properly controlled and maintained; and ensure that these assets are supporting your organization's strategy and business goals.

Compliance: When organizations are moving their native data to a cloud, they need to comply with particular general body policies if the data is from public sources. Although, finding a cloud provider who will comply with these policies is difficult to find, or one needs to negotiate on that front. Many CSPs are coming with flexible compliance policies for data acquisition and cloud infrastructure. An issue for anyone using backup services or cloud storage. Every time a company moves data from the internal storage to a cloud, it is faced with being compliant with industry regulations and laws. Depending on the industry and requirements, every organization must ensure these standards are respected and carried out. This is one of the many challenges facing cloud computing, and although the procedure can take a certain amount of time, the data must be properly stored.

Managing Multiple Clouds: The challenges facing cloud computing haven't just been concentrated in one, single cloud. The state of multi-cloud has grown exponentially in recent years. But managing multi-cloud infrastructure contrary to a single cloud is very challenging given all the above data-driven challenges. The companies are shifting or combining public and private clouds and, as mentioned earlier, tech giants like Alibaba and Amazon are leading the way. Approximately, 81% of companies are having multi-cloud strategies and have a hybrid cloud structure (public and private clouds). Companies are opting for a multi-cloud scenario because some of the services are cost-effective in public and to manage cost-effectively this cloud model has been very successful in recent years. However, managing such highly networked architecture is a difficult task.

Performance: When a business moves to the cloud it becomes dependent on the service providers. The next prominent challenges of moving to cloud computing expand on this partnership. The performance of the organization's BI and other cloud-based systems is also tied to the performance of the cloud provider when it falters. When your provider is down, you are also down. Cloud computing is on-demand compute service and supports multitenancy, thus performance should not suffer over the acquisition of new users. The CSP should maintain enough resources to serve all the users and any ad-hoc requests.

Building a Private Cloud: Creating an internal or private cloud will cause a significant benefit: having all the data in-house. But IT managers and departments will need to face building and gluing it all together by themselves, which can cause one of the challenges of moving to cloud

Cloud Computing

computing extremely difficult. Many tasks such as grabbing an IP address cloud software layer, setting up a virtual local area network (VLAN), load balancing, firewall rule-setting for the IP address, server software patch, arranging nightly backup queue are quite complex associated tasks for a private cloud. Although building a private cloud isn't a top priority for many organizations, for those who are likely to implement such a solution, it quickly becomes one of the main challenges facing cloud computing - private solutions should be carefully addressed. Many companies are planning to do so because the cloud will on-premise and they will have all the data authority over shared cloud resources.

Segmented Usage and Adoption: Most organizations did not have a robust cloud adoption strategy in place when they started to move to the cloud. Instead, ad-hoc strategies sprouted, fueled by several components. One of them was the speed of cloud adoption. Another one was the staggered expiration of data center contracts/equipment, which led to intermittent cloud migration. Finally, there also were individual development teams using the public cloud for specific applications or projects.

Migration: One of the main cloud computing industry challenges in recent years concentrates on migration. This is a process of moving an application to a cloud. An although moving a new application is a straightforward process, when it comes to moving an existing application to a cloud environment, many cloud challenges arise.

1.9 Challenges in Cloud Computing

Hosting and running the applications of the current cloud platform on some other platform may cause support issues, configuration complexities, and additional expenses. The company data might also be left vulnerable to security attacks due to compromises that might have been made during migrations. The various challenges being faced include:

Service Quality: Service quality should be good and is a major concern of the end-user. The whole ecosystem of cloud computing is presented in virtual environments and thus the CSP should give what is promised in terms of service, be it compute resources or customer satisfaction.

Interoperability: CSP's services should be flexible enough to integrate itself into other platforms and services provided by other CSPs. The data pipeline should be easy to integrate and should drive improved performance. There are a lot of challenges in cloud computing like Big data analysis, long haul transfer, transferring data problems but still, it is the best computing resource available to date.

Availability and Reliability: Data and service from CSP should be available at all times irrespective of the external condition or the ideal condition. Computing resources should be available for the users and their operability should be reliable.

Portability: If the users want to migrate from one CSP to others, the vendor should not lock-in customer data or services and the migration should be ease. There are different laws over data in different countries.

Cloud Integration: Several companies, especially those with hybrid cloud environments report issues associated with having their on-premise apps and tools and public cloud for working together. According to survey, 62% of respondents said integration of legacy systems as their biggest challenge in multi-cloud. Although, combining new cloud-based apps and legacy systems needs resources, expertise, and time but still several companies are considering that the perks of cloud computing dominate the backlogs of this technology.

Vendor Lock-in: Entering a cloud computing agreement is easier than leaving it. "Vendor lock-in" happens when altering providers is either excessively expensive or just not possible. It could be that the service is nonstandard or that there is no viable vendor substitute. It is important to guarantee the services you involve are typical and transportable to other providers, and above all, understand the requirements. When in need to migrate from one cloud platform to another, a company might face some serious challenges because of the differences between vendor platforms.

1.10 Applications of Cloud Computing

Cloud Service Providers (CSP) are providing many types of cloud services and now if we say that cloud computing has touched every sector by providing various cloud applications. The sharing

Unit 01: Cloud Computing Introduction

and management of resources is easy in cloud computing that's why it is one of the dominant fields of computing. Many properties have made it an active component in different fields.

1. **Online Data Storage:** Cloud computing allows storing data like files, images, audios, and videos, etc. on the cloud storage. The organization need not set physical storage systems to store a huge volume of business data which costs so high nowadays. As they are growing technologically, data generation is also growing with respect to time, and storing that becoming problem. In that situation, Cloud storage is providing this service to store and access data any time as per requirement. Example: Google Drive, DropBox, iCloud etc.
2. **Backup and Recovery:** Cloud vendors provide security from their side by storing safe to the data as well as providing a backup facility to the data. They offer various recovery application for retrieving the lost data. In the traditional way, backup of data is a very complex problem and also it is very difficult sometimes impossible to recover the lost data. But cloud computing has made backup and recovery applications very easy where there is no fear of running out of backup media or loss of data.
3. **Bigdata Analysis:** We know the volume of big data is so high, such that, storing that in the traditional data management system for an organization is impossible. Cloud computing has resolved that problem by allowing the organizations to store their large volume of data in cloud storage without worrying about physical storage. Next comes analyzing the raw data and finding out insights or useful information from it is a big challenge as it requires high-quality tools for data analytics. Cloud computing provides the biggest facility to organizations in terms of storing and analyzing big data.
4. **Anti-Virus Applications:** Previously, organizations were installing antivirus software within their system even if we will see we personally also keep antivirus software in our system for safety from outside cyber threats. But, nowadays, cloud computing provides cloud antivirus software which means the software is stored in the cloud and monitors your system/organization's system remotely. This antivirus software identifies the security risks and fixes them. Sometimes also they give a feature to download the software.
5. **E-commerce Application:** Cloud-based e-commerce allows responding quickly to the opportunities which are emerging. Users respond quickly to the market opportunities as well as the traditional e-commerce responds to the challenges quickly. Cloud-based e-commerce gives a new approach to doing business with the minimum amount as well as minimum time possible. Customer data, product data, and other operational systems are managed in cloud environments.
6. **Cloud computing in Education:** Cloud computing in the education sector brings an unbelievable change in learning by providing e-learning, online distance learning platforms, and student information portals to the students. It is a new trend in education that provides an attractive environment for learning, teaching, experimenting, etc. to students, faculty members, and researchers. Everyone associated with the field can connect to the cloud of their organization and access data and information from there.
7. **Technology-enhanced Learning or Education as a Service (EaaS):** There are the following education applications offered by the cloud-



Example:

- **Google Apps for Education:** Google Apps for Education is the most widely used platform for free web-based email, calendar, documents, and collaborative study.
- **Chromebooks for Education:** Chromebook for Education is one of the most important Google's projects. It is designed for the purpose that it enhances education innovation.

Cloud Computing

- Tablets with Google Play for Education: It allows educators to quickly implement the latest technology solutions into the classroom and make it available to their students.

8. **Testing and development:** Setting up the platform for development and finally performing different types of testing to check the readiness of the product before delivery requires different types of IT resources and infrastructure. But Cloud computing provides the easiest approach for development as well as testing even if deployment by using their IT resources with minimal expenses. Organizations find it more helpful as they got scalable and flexible cloud services for product development, testing, and deployment.

9. **E-Governance Applications:** Cloud computing can provide its services to multiple activities conducted by the government. It can support the government to move from the traditional ways of management and service providers to an advanced way of everything by expanding the availability of the environment, making the environment more scalable and customized. It can help the government to reduce the unnecessary cost in managing, installing, and upgrading applications and doing all these with help of could computing and utilizing that money public service.

10. **Cloud Computing in Medical Fields:** In the medical field also nowadays cloud computing is used for storing and accessing the data as it allows to store data and access it through the internet without worrying about any physical setup. It facilitates easier access and distribution of information among the various medical professional and the individual patients. Similarly, with help of cloud computing offsite buildings and treatment facilities like labs, doctors making emergency house calls and ambulances information, etc can be easily accessed and updated remotely instead of having to wait until they can access a hospital computer.



Example: Telehealth, Medi-Cloud, Technology-enhanced Healthcare

11. **Entertainment Applications:** Many people get entertainment from the internet, in that case, cloud computing is the perfect place for reaching to a varied consumer base. Therefore, different types of entertainment industries reach near the target audience by adopting a multi-cloud strategy. Cloud-based entertainment provides various entertainment applications such as online music/video, online games and video conferencing, streaming services, etc and it can reach any device be it TV, mobile, set-top box, or any other form. It is a new form of entertainment called On-Demand Entertainment (ODE). Entertainment industries use a multi-cloud strategy to interact with the target audience. Cloud computing offers various entertainment applications such as online games and video conferencing.

- **Online games:** Today, cloud gaming becomes one of the most important entertainment media. It offers various online games that run remotely from the cloud. The best cloud gaming services are Shaow, GeForce Now, Vortex, Project xCloud, and PlayStation Now.
- **Video conferencing apps:** Video conferencing apps provides a simple and instant connected experience. It allows us to communicate with our business partners, friends, and relatives using a cloud-based video conferencing. The benefits of using video conferencing are that it reduces cost, increases efficiency, and removes interoperability.

12. **Art Applications:** Cloud computing offers various art applications for quickly and easily design attractive cards, booklets, and images. Some most commonly used cloud art applications are given below:

Unit 01: Cloud Computing Introduction

- Moo: One of the best cloud art applications. It is used for designing & printing business cards, postcards, & mini cards.
- Vistaprint: Vistaprint allows us to easily design various printed marketing products such as business cards, Postcards, Booklets, and wedding invitations cards.
- Adobe Creative Cloud: Adobe creative cloud is made for designers, artists, filmmakers, and other creative professionals. It is a suite of apps which includes PhotoShop image editing programming, Illustrator, InDesign, TypeKit, Dreamweaver.

13. **Management Applications:** Cloud computing offers various cloud management tools which help admins to manage all types of cloud activities, such as resource deployment, data integration, and disaster recovery. These management tools also provide administrative control over the platforms, applications, and infrastructure. Some important management applications are:

- Toggl: Toggl helps users to track allocated time period for a particular project.
- Evernote: Evernote allows you to sync and save your recorded notes, typed notes, and other notes in one convenient place. It is available for both free as well as a paid version. It uses platforms like Windows, macOS, Android, iOS, Browser, and Unix.
- Outright: Outright is used by management users for the purpose of accounts. It helps to track income, expenses, profits, and losses in real-time environment.
- GoToMeeting: GoToMeeting provides Video Conferencing and online meeting apps, which allows you to start a meeting with your business partners from anytime, anywhere using mobile phones or tablets. Using GoToMeeting app, you can perform the tasks related to the management such as join meetings in seconds, view presentations on the shared screen, get alerts for upcoming meetings, etc.

14. **Social Applications:** Social cloud applications allow a large number of users to connect with each other using social networking applications such as Facebook, Twitter, LinkedIn, etc. There are the following cloud based social applications-

- Facebook: Facebook is a social networking website which allows active users to share files, photos, videos, status, more to their friends, relatives, and business partners using the cloud storage system. On Facebook, we will always get notifications when our friends like and comment on the posts.
- Twitter: Twitter is a social networking site. It is a microblogging system. It allows users to follow high profile celebrities, friends, relatives, and receive news. It sends and receives short posts called tweets.
- LinkedIn: LinkedIn is a social network for students, freshers, and professionals.

With respect to this as a cloud, the market is growing rapidly and it is providing various services day by day. So, in the future cloud computing is going to touch many more sectors by providing more applications and services.



Lab Exercise: Explore the different service providers of Cloud computing?

Summary

- Cloud computing offers various cloud management tools which help admins to manage all types of cloud activities, such as resource deployment, data integration, and disaster recovery.
- Cloud computing refers to manipulating, configuring, and accessing the applications online.

Cloud Computing

- Cloud computing virtualizes systems by pooling and sharing resources. Systems and storage can be provisioned as needed from a centralized infrastructure, costs are assessed on a metered basis, multi-tenancy is enabled, and resources are scalable with agility.
- Cloud computing eliminates the need for IT infrastructure updates and maintenance since the service provider ensures timely, guaranteed, and seamless delivery of our services and also takes care of all the maintenance and management of our IT services according to the service-level agreement (SLA).
- Cloud computing can be expensive if you don't know how to manage your computing resources and take maximum advantage of them.
- Cloud computing lets us deploy the service quickly in fewer clicks. This quick deployment lets us get the resources required for our system within minutes.

Keywords

- **Service Oriented Architecture (SOA):** SOA is essentially a collection of services which communicate with each other. SOA provides a loosely-integrated suite of services that can be used within multiple business domains.
- **Abstraction:** Cloud computing abstracts the details of system implementation from users and developers. Applications run on physical systems that aren't specified, data is stored in locations that are unknown, administration of systems is outsourced to others, and access by users is ubiquitous.
- **Cloud:** Cloud refers to a Network or Internet. A cloud is usually defined as a large group of interconnected computers. These computers include network servers or personal computers.
- **Cloud computing:** Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computer resources (networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.
- **Cloud computing collaboration:** The users from multiple locations within a corporation, and from multiple organizations, desired to collaborate on projects that crossed company and geographic boundaries. Projects had to be housed in the "cloud" of the Internet, and accessed from any Internet-enabled location. Cloud-collaboration is also termed as Internet-based group collaboration.
- **Multitenancy:** In cloud computing, multitenancy means that multiple customers of a cloud vendor are using the same computing resources. Despite the fact that they share resources, cloud customers aren't aware of each other, and their data is kept totally separate.
- **Thick clients:** Thick clients are regular computers, using a web browser like Firefox or Internet Explorer to connect to the cloud.

Self Assessment

1. Which of the following is an important component of Cloud computing?
 - A. Fat Clients
 - B. Virtualization
 - C. Abstraction
 - D. CapEx

Unit 01: Cloud Computing Introduction

2. The computers that do not have internal hard drives, but rather let the server do all the work, but then display the information are _____.
- A. Fat Client
 - B. Thick Client
 - C. Thin Client
 - D. Internal HDD Computers
3. _____ is a construct (infrastructure) that allows to access application that actually resides at a remote location of the other internet connected device, most often, this will be a distant datacenter.
- A. Grid computing
 - B. Client-server computing
 - C. Cloud computing
 - D. Utility computing
4. The scenario in where the applications running on the physical systems aren't specified; data is stored in locations that are unknown represents _____ in cloud computing.
- A. Virtualization
 - B. Abstraction
 - C. Collaboration
 - D. Vendor Lock In
5. SLA in cloud computing stands for
- A. Service-Low Agreements
 - B. Software-Level Approach
 - C. Software-Low Approach
 - D. Service-Level Agreements
6. _____ as a utility is a dream from the beginning of the computing industry itself.
- A. Software
 - B. Hardware
 - C. Computing
 - D. Model
7. Cloud computing refers to which of the following:
- A. Application and services that run on a distributed network using virtualized resource
 - B. Application that runs on a distributed network using virtualized resource
 - C. Services that run on a distributed network using virtualized resource
 - D. None of the above
8. Which of the following is the main concern related to cloud computing?
- A. Security
 - B. Cost
 - C. Space
 - D. Platforms
9. The capability of cloud computing where a user can provision computing capabilities, such as server time and storage, as needed without requiring human interaction corresponds to

Cloud Computing

- A. Broad Network Access
 - B. On-demand Self-service
 - C. Elasticity
 - D. Resource Pooling
10. The outsourced cloud management aids in
- A. Saving the upfront capital investments
 - B. Easiness of setting IT infrastructure
 - C. Paying only for operational expenses
 - D. All of the above
11. _____ allows multiple users to make use of the same shared resources such as Banking, Financial, Social networking, e-commerce, B2B etc.
- A. Multitenancy
 - B. Distribution
 - C. Virtualization
 - D. Abstraction
12. SOA in cloud computing stands for
- A. Software oriented applications
 - B. Server oriented approach
 - C. Service optimal approach
 - D. Service Oriented Architecture
13. CSPs stands for
- A. Cloud Service Provisioners
 - B. Cloud Service Providers
 - C. Cloud Software Providers
 - D. Cloud Software Provisioners
14. _____ allow a large number of users to connect with each other using social networking applications such as Facebook, Twitter, LinkedIn etc.
- A. Social cloud applications
 - B. E-Commerce applications
 - C. Telehealth applications
 - D. Tele-Learning applications
15. Which one of the following is cloud platform by Amazon?
- A. Azure
 - B. AWS
 - C. Cloudera
 - D. All of the mentioned

Answers for Self Assessment

- | | | | | |
|-------|-------|-------|-------|-------|
| 1. A | 2. C | 3. C | 4. B | 5. D |
| 6. C | 7. A | 8. A | 9. B | 10. D |
| 11. A | 12. D | 13. B | 14. A | 15. B |

Review Questions

1. What is cloud computing? What are its characteristics?
2. Discuss the various components of cloud computing?
3. List the different applications of cloud computing?
4. What different challenges are faced in cloud computing?
5. Discuss in detail the evolution of cloud computing and its history?
6. What are the benefits and disadvantages of cloud computing?
7. How cloud computing has supported collaboration? Explain with examples.
8. Compare thin and thick clients? List the advantages of using the thin clients?
9. Elaborate on the following cloud challenges:
 - a) Compliance
 - b) Security and privacy
 - c) Migration
10. What does the Google's perspective on cloud computing says? Elaborate.

**Further Readings**

Cloud Computing: Concepts, Technology and Architecture by Erl, Pearson Education.

Cloud Computing Black Book by Kailash Jayaswal, Jagannath Kallakurchi, Donald J. Houde, Deven Shah, Kogent Learning Solutions, DreamTech Press.

**Web Links**

<https://www.ibm.com/cloud/learn/cloud-computing>

[What is cloud computing? A beginner's guide | Microsoft Azure](#)

[Cloud Computing \(w3schools.in\)](#)

[What is Cloud Computing? - Cloud Computing Fundamentals Training - Bing video](#)

[What is Cloud Computing? | Cloud Computing Fundamentals | AWS Training | Edureka | AWS Rewind - 1 - Bing video](#)

Unit 02: Cloud Computing Architecture and Models

CONTENTS

Objectives

Introduction

2.1 Why Cloud Computing Matters

2.2 Benefit of Group Collaboration for Both Individuals and Organizations

2.3 Issues in Cloud Computing

2.4 Considerations for Getting into the Cloud

2.5 Cloud Architecture

2.6 Cloud Storage

2.7 Cloud Business Models

Summary

Keywords

Self Assessment

Answers for Self Assessment

Review Questions

Further Readings

Objectives

After studying this unit, you will be able to:

- understand why cloud computing matters.
- know about the issues in cloud.
- explore the cloud architecture and cloud storage.
- discuss the NIST cloud computing reference model and cloud cube model.

Introduction

Cloud refers to a Network or Internet. Cloud is something, which is present at remote location. Cloud can provide services over network, that is, on public networks or on private networks, that is, Wide Area Networks (WANs), Local Area Networks (LANs), or Virtual Private Networks (VPNs). Applications such as e-mail, web conferencing, customer relationship management (CRM), all run in cloud.

Cloud computing is a subscription-based delivery model that provides scalability, fast delivery and greater IT efficiencies. Cloud has removed many physical and financial barriers to aligning IT needs with evolving business goals. With a promise to deliver better applications, platforms and infrastructure quickly and cheaply, cloud computing has become a major force for business innovation across all industries.

Cloud Computing in Today's World

Cloud computing's ability to provide elastic scalability, faster service delivery, greater IT efficiency and a subscription-based accounting model has broken down many of the physical and financial barriers to aligning IT with evolving business goals. With the promise to deliver better business models and services quickly and cheaply, cloud computing has become a major driver of business innovation across all industries. Today, cloud computing has moved to the center of many organizations' technology strategies.

Cloud Computing

Cloud computing's technological advantages with its ability to scale computing resources up and down, more reliable network connections and the ubiquity of big data- makes it appealing to organizations of all sizes has led it to greatest technological innovation. The cloud provides not only delivery of software services, but data storage as well. The growth and flexibility offered by cloud infrastructure has enabled organizations to explore the full potential of data assets in a fast and cost-effective manner. Cloud platforms are now an integral part of many organizations' data strategies.

However, the shift to cloud has not been without problems. The migration of critical business data from on-site, secured data centers to storage on public cloud platforms has raised concerns over data security. This has been the case specifically around personal, customer data storage.

2.1 Why Cloud Computing Matters

Technology has evolved fast over the last few years. People used to run businesses in different ways before the invention of cloud computing. Like in other industries, technology advancement has an impact on the business as well. Before businesses started relying on it, managers had to run applications from servers on their premises. Some even had to hire extra IT experts to help them create their data centres. This reduced their profit margins because they had to pay them a lot of cash. However, things have changed nowadays.

A major revolution has been brought to the business world by cloud technology. This technology has made business to run smoothly plus experience growth. Cloud technology leverages virtualization. Cloud services are maintained at a remote data centre. You can also find them in Amazon web services. This technology can be applied to all companies. So, how does cloud technology trigger business development? Let us find out here below:

Cloud technology has changed the business world a lot. It has improved efficiency, productivity, and overall performance. There are a few reasons why cloud technology matters for business growth:

Cost-Effectiveness:This technology is a game-changer because you do not have to hire extra IT experts or resources to help you in data management. You will only pay for the services you need to remain competitive or enjoy growth in your business. You do not pay for the services you no longer need. Businesses can grow fast when using this technology because there will be no maintenance, labor, or purchasing costs. You only pay a monthly subscription fee for the services your business needs from your cloud technology service provider.

Digitalizes Business:These days, businesses are switching to digital marketing strategies. There are many businesses around the world. And for you to remain competitive, you have to embrace digital transformation. But this does not mean going paperless only. Do you want your business to grow fast? Then, why don't you make your business operations digital? If you want to avoid losses during the digital migration, consider hiring a reliable cloud technology company. Outsourcing the entire digital transformation process will cut the risk of improper migration. Do stick to the traditional ways of operating because they are costly, slow, and outdated.

Offers Better Data Backups and Recovery: Cloud technology offers a lot of benefits, but cloud storage tends to be the greatest deal. Cloud storage can be accessed from anywhere, anytime, and with any device. Thus, businesses with remote workers or offices can benefit a lot from this technology. Are you looking for a reliable business backup and recovery process? Cloud technology is the ideal choice for you because it has an integrated process that keeps your relevant business data safe all the time. You no longer worry about cyber-attacks, natural disasters, or physical thefts as your data is secured and protected. Your data is not stored in one place because the service provider will distribute it to remote data centres. Even if a disaster hits your business premises, you will retrieve the necessary data stored in the clouds. Thus, resuming your operations normally as if nothing happened.

Seamless Scalability:This is another reason why you need cloud technology in your business. If your business uses the IaaS cloud, it will be easier to subscribe to the whole IT infrastructure. Hence, you will not need to buy hardware during the process. If there is a need to add or remove servers in your infrastructure, scaling up and down will be quick and seamless.

Flexibility:Cloud-based services are ideal for businesses with growing or fluctuating bandwidth demands. If your needs increase, it's easy to scale up your cloud capacity, drawing on the service's remote servers. Likewise, if you need to scale down again, the flexibility is baked into the

Unit 02: Cloud Computing Architecture and Models

service. This level of agility can give businesses using cloud computing a real advantage over competitors- it's not surprising that businesses identify operational agility as a key reason for cloud adoption.

Increasing Business Competitiveness: Many businesses are forced to be quick and efficient when adapting to the marketplace changes because competition keeps rising up. They can benefit a lot from the flexible and customizable cloud technology solutions. This may help them increase their agility in their operations.

Geographical Dispersion: Cloud computing allows you to work anytime and from anywhere, provided you have an internet connection. Since most established cloud services also offer mobile apps, you're not even restricted by which device you've got to hand. Businesses can offer more flexible working perks to employees, so they can enjoy the work-life balance that suits them-without productivity taking a hit. Home office is an attractive option for many employees and now, thanks to cloud services, an increasingly accessible idea too.

For Developers: Cloud computing provides increased amounts of storage and processing power to run the applications they develop. Cloud computing also enables new ways to access information, process and analyze data, and connect people and resources from any location anywhere in the world. In essence, it takes the lid off the box; with cloud computing, developers are no longer boxed in by physical constraints.

For IT Departments: For IT departments, cloud computing offers more flexibility in computing power, often at lower costs. With cloud computing, IT departments don't have to engineer for peak-load capacity, because the peak load can be spread out among the external assets in the cloud. And, because additional cloud resources are always at the ready, companies no longer have to purchase assets (servers, workstations, and the like) for infrequent intensive computing tasks. If you need more processing power, it's always there in the cloud—and accessible on a cost-efficient basis.

For End-Users: For end users, cloud computing offers all these benefits and more. An individual using a web-based application isn't physically bound to a single PC, location, or network, his applications and documents can be accessed wherever he is, whenever he wants. Gone is the fear of losing data if a computer crashes out. Documents hosted in the cloud always exist, no matter what happens to the user's machine.

2.2 Benefit of Group Collaboration for Both Individuals and Organizations

The users from around the world can collaborate on the same documents, applications, and projects, in real time. A whole new world of collaborative computing, all enabled by the notion of cloud computing.

Streamlined content: The more employees and partners collaborate on documents, the greater the need for watertight document control. Before the cloud, workers had to send files back and forth as email attachments to be worked on by one user at a time. Sooner or later – usually sooner – you end up with a mess of conflicting file content, formats and titles. The greater visibility means improved collaboration, which ultimately means better work and a healthier bottom line. If you're still relying on the old way, it could be time to try something a little more streamlined.

For Everyone: For everyone concerned, cloud computing does all this at lower costs, because the cloud enables more efficient sharing of resources than does traditional network computing. When you tap into the power of the cloud, you get supercomputing power at PC prices – something that offers particular appeal to individuals and small businesses. With cloud computing, hardware doesn't have to be physically adjacent to a firm's office or data center; cloud infrastructure can be located anywhere, including and especially areas with lower real estate and electricity costs. Cloud has transformed our lives as individuals as well. Many of us use cloud services every day. When we update our status on social media, binge a new streaming series, or check our bank accounts we're most likely using applications that are hosted by cloud services. These apps are accessed through an internet connection rather than installed on our hard drives or devices.

For Innovation: Cloud opens up a wide range of opportunities for innovation and creative business models, especially with the growing number of readily available cloud-based services out there for entrepreneurs to use. Cloud is a relatively new and exciting frontier in the tech and

Cloud Computing

business worlds, offering those with great ideas the opportunity to find their own niche in the market with very little competition.

For Improving Service Delivery: While any of the above benefits are a great incentive for moving a business to the cloud, effective implementation of new customer relationship management (CRM) tools through cloud services are the icing on the cake. This is why Salesforce pioneered enterprise cloud computing. From CRM to marketing automation, most of the solutions are entirely cloud-based. They help to streamline both business and the way customers interact with it- all at the same time.

Cloud technology is helping businesses and people weather the ongoing storm caused by the global pandemic. The companies are seizing on cloud as the key enabler to complete their digital transformation, and the COVID pandemic has further accelerated this mandate. Cloud is becoming a top C-suite agenda item as businesses are transitioning from a piece-meal approach to a more holistic end-to-end digital transformation with Cloud at its core. The winners of tomorrow will be the ones that navigate this change rapidly, make the right choices and engage with the appropriate partners to augment their own capabilities. Cloud technology, people and businesses accelerate and requires speed, new thinking, and an entirely different level of skills and investments to achieve end-to-end digital transformation. Now, more than ever, cloud is vital to help businesses reopen, reinvent, and outmaneuver uncertainty.

2.3 Issues in Cloud Computing

Hosting and running the applications of the current cloud platform on some other platform may cause support issues, configuration complexities, and additional expenses. The company data might also be left vulnerable to security attacks due to compromises that might have been made during migrations. The various challenges being faced include:

Operational Complexity: One of the things that makes cloud services more difficult from traditional ones is that there's a lot of operational complexity that comes along with the shift. Building the capabilities, software and systems required to maintain a business in the cloud isn't always easy, which can make it difficult to scale and deliver products and services to an ever-growing market. This may not apply to those who just want to take advantage of cloud-based services, but is a serious concern for those starting a cloud-based business.

Lack of Customization: While cloud-based software can be great for getting things done, it may lack one feature that some businesses demand: customization. Generally, cloud-based software providers deliver a great product, but it's also a generic one meant to serve the needs of a wide range of businesses and individuals. For some companies who need customized software to do business, this may simply not be an option.

Data Security and Privacy: Security has indeed been a primary, and valid, concern from the start of cloud computing technology: you are unable to see the exact location where your data is stored or being processed. This increases the cloud computing risks that can arise during the implementation or management of the cloud. When we say security and privacy, we are talking about the user data that is stored on Cloud Service Providers (CSP) data centers. A CSP should abide by the rules of not sharing confidential data or any data that matters to the users. The data centers must be secure and privacy of the data should be maintained by a CSP. There is always concern about the actual location of your data, where it is stored and processed. Before onboarding, yourself on the cloud computing platform one should always check the data security and data recovery (in case of disaster) policy of the CSP.

The cloud can provide amazing security, if done right, but not every vendor is of the same quality. Some may not offer services that can provide you with the level of security you need for sensitive data, and major leaks can occur anywhere, cloud or otherwise, if proper precautions aren't taken. In terms of security concerns of cloud technology, we don't find answers to some questions. Mysterious threats like website hacking and virus attack are the biggest problems of cloud computing data security. Before utilizing cloud computing technology for a business, entrepreneurs should think about these things. Once you transfer important data of your organization to a third party, you should make sure you have a cloud security and management system. Cybersecurity experts are more aware of cloud security than any other IT professional. According to Crowd Research Partners survey, 9 out of 10 cybersecurity experts are concerned regarding cloud security. Also, they are worried about the violation of confidentiality, data privacy, and data leakage and loss.

Unit 02: Cloud Computing Architecture and Models

Performance: When a business moves to the cloud it becomes dependent on the service providers. The next prominent challenges of moving to cloud computing expand on this partnership. The performance of the organization's BI and other cloud-based systems is also tied to the performance of the cloud provider when it falters. When your provider is down, you are also down. Cloud computing is on-demand compute service and supports multitenancy, thus performance should not suffer over the acquisition of new users. The CSP should maintain enough resources to serve all the users and any ad-hoc requests.

Dealing with Multi-Cloud Environments: These days, maximum companies are not only working on a single cloud. As per the RightScale report revelation, nearly 84% of the companies are following a multi-cloud strategy and 58% already have their hybrid cloud tactic that is combined with the public and private cloud. A long-term prediction on the future of cloud computing technology gives a more difficulty encountered by the teams of IT infrastructure. However, the professionals have also suggested the top practices like re-thinking procedures, training staff, tooling, active vendor relationship management, and doing the study.

Cloud Migration: Although releasing a new app in the cloud is a very simple procedure, transferring an existing application to a cloud computing environment is tougher. According to the report, 62% said that their cloud migration projects were tougher than they anticipated. Alongside this, 64% of migration projects took more time than predicted and 55% went beyond their budgets. Especially, some organizations migrating their apps to the cloud reported downtime during migration (37%), issues syncing data before cutover (40%), the problem having migration tools to work well (40%), slow data migration (44%), configuring security issues (46%), and time-consuming troubleshooting (47%). And to solve over these issues, nearly 42% of the IT experts said they wished they had increased their budgets, around 45% wished to have employed an in-house professional, 50% wanted to set a longer project duration, 56% of them wanted they had performed more pre-migration testing.

Cloud Integration: Finally, several companies, especially those with hybrid cloud environments report issues associated with having their on-premise apps and tools and public cloud for working together. According to survey, 62% of respondents said integration of legacy systems as their biggest challenge in multi-cloud. Likewise, in a Software One report on cloud cost, 39% of those assessed said integrating legacy systems was one of their biggest worries while utilizing the cloud. Also, combining new cloud-based apps and legacy systems needs resources, expertise, and time.

Unauthorized Service Providers: It is a new concept for most of the business organizations. A normal businessman is not able to verify the genuineness of the service provider agency. It's very difficult for them to check the whether the vendors meet the security standards or not. There is need for an ICT consultant to evaluate the vendors against the worldwide criteria. It is necessary to verify that the vendor must be operating this business for a sufficient time without having any negative record in past. Vendor continuing business without any data loss complaint and have a number of satisfied clients. Market reputation of the vendor should be unblemished.

Hacking of Brand: Cloud involves some major risk factors like hacking. Some professional hackers are able to hack the application by breaking the efficient firewalls and steal the sensitive information of the organizations. A cloud provider hosts numerous clients; each can be affected by actions taken against any one of them. When any threat came into the main server it affects all the other clients also. As in distributed denial of service attacks server requests that inundate a provider from widely distributed computers.

Cloud Management: Managing a cloud is not an easy task. It consists of a lot of technical challenges. A lot of dramatic predictions are famous about the impact of cloud computing. People think that traditional IT department will be outdated and research supports the conclusions that cloud impacts are likely to be more gradual and less linear. Cloud services can easily change and update by the business users. It does not involve any direct involvement of IT department. It is a service provider's responsibility to manage the information and spread it across the organization. So, it is difficult to manage all the complex functionality of cloud computing.

Sustainability: Sustainability refers to minimizing the effect of cloud computing on environment. Indeed, citing the server's effects on the environmental effects of cloud computing, in areas where climate favors natural cooling and renewable electricity is readily available. The countries with favorable conditions, such as Finland, Sweden, and Switzerland are trying to attract cloud computing data centers. But other than nature's favors, would these countries have enough technical infrastructure to sustain the high-end clouds.

Cloud Computing

Cost Management and Containment: Cloud computing can be expensive if you don't know how to manage your computing resources and take maximum advantage of them. Many times, organizations dwell in a mindset of pay-as-you-go and spend more on cloud than they would have on on-premise infrastructure. One should always optimize the cost by financial analytics and reporting the usage for better monitoring of cost. For the most part, cloud computing can save businesses money. In the cloud, an organization can easily ramp up its processing capabilities without making large investments in new hardware. Businesses can instead access extra processing through pay-as-you-go models from public cloud providers. However, the on-demand and scalable nature of cloud computing services make it sometimes difficult to define and predict quantities and costs.

Lack of Resources/ Expertise: Cloud challenges companies and enterprises. As the usage of cloud technologies is increasing, tools to manage it are getting sophisticated, finding experts on top of this in cloud computing is becoming a bottleneck to many organizations. The organizations are increasingly placing more workloads in the cloud while cloud technologies continue to rapidly advance. Due to these factors, organizations are having a tough time keeping up with the tools. Also, the need for expertise continues to grow. Such challenges can be minimized through additional training of IT and development staff. Many companies are adopting automated cloud management technologies but it's always better to train individuals to satisfy the need of time. Presently, DevOps tools like Chef and Puppet are heavily used in the IT industry.

Governance/ Control: In cloud computing, infrastructure resources are under CSP's control and end-users or companies have to abide by the governance policies from CSP. The traditional IT teams have no control over how and where their data is stored and processed. IT governance should assure how infrastructure assets from CSP are being used. To overcome the downfalls and challenges, onboarding to Cloud, IT must adapt its orthodox way of governance and process control to the induct cloud. Now, IT is playing an important role in benchmarking cloud services requirements and policies. Thus, the proper IT governance should ensure IT assets are implemented and used according to agreed-upon policies and procedures; ensure that these assets are properly controlled and maintained; and ensure that these assets are supporting your organization's strategy and business goals.

Compliance: When organizations are moving their native data to a cloud, they need to comply with particular general body policies if the data is from public sources. Although, finding a cloud provider who will comply with these policies is difficult to find, or one needs to negotiate on that front. Many CSPs are coming with flexible compliance policies for data acquisition and cloud infrastructure. An issue for anyone using backup services or cloud storage. Every time a company moves data from the internal storage to a cloud, it is faced with being compliant with industry regulations and laws. Depending on the industry and requirements, every organization must ensure these standards are respected and carried out. This is one of the many challenges facing cloud computing, and although the procedure can take a certain amount of time, the data must be properly stored.

Building a Private Cloud: Creating an internal or private cloud will cause a significant benefit: having all the data in-house. But IT managers and departments will need to face building and gluing it all together by themselves, which can cause one of the challenges of moving to cloud computing extremely difficult. Many tasks such as grabbing an IP address cloud software layer, setting up a virtual local area network (VLAN), load balancing, firewall rule-setting for the IP address, server software patch, arranging nightly backup queue are quite complex associated tasks for a private cloud. Although building a private cloud isn't a top priority for many organizations, for those who are likely to implement such a solution, it quickly becomes one of the main challenges facing cloud computing - private solutions should be carefully addressed. Many companies are planning to do so because the cloud will be on-premise and they will have all the data authority over shared cloud resources.

Segmented Usage and Adoption: Most organizations did not have a robust cloud adoption strategy in place when they started to move to the cloud. Instead, ad-hoc strategies sprouted, fueled by several components. One of them was the speed of cloud adoption. Another one was the staggered expiration of data center contracts/equipment, which led to intermittent cloud migration. Finally, there also were individual development teams using the public cloud for specific applications or projects.

2.4 Considerations for Getting into the Cloud

Take it Slow and Prepare for Roadblocks: When getting into cloud computing, it's smart to take things slow and to start small. It's a different way of working than some are used to and can have a bit of a learning curve for less tech-savvy employees.

Understand that Even Tech-savvy Companies Struggle with Cloud Business: According to recent research from Accenture, even high-tech companies are still struggling to find their way using the new business model. While there are many ways that cloud-based business is easier, there are many other aspects that are quite complex and haven't quite been puzzled out yet even by those leading the industry. That's both good and bad. Good, because there's room for new players in the market, and bad because, well, it does make things a bit more difficult for those entering into cloud-based businesses.

Locate the Services for Each Part of your Business: If you're going to go all in on the cloud, you'll need to find services that will work for every part of your business. That means not only getting the basics like word processing and CRM taken care of, but also finding tools that those in HR and accounting will appreciate as well.

Decide the Level of Cloud Services that are Right for your Business: Of course, not every business is cloud-ready. Everyone will have their own level of comfort with working in the cloud and some may just be better off providing what they offer in a more traditional format. There's nothing wrong with that if that's what's working, but make sure to do a careful analysis before deciding.

Have a Plan: While you may only be employing a small number of cloud services now, you want to have a plan for the role cloud-based computing will play in your business in the future. If you're seeking to start your own cloud-based business, then a plan is obviously a must as well.

Be Creative: Look for ways to use free or low-cost cloud programs to cut down on your operational costs. There are a lot of really great services out there that won't cost your company a dime to use. Do give them a try.

Evaluate the Options: Before committing to any cloud-based service, try it out first. Most companies will give you a free trial period, allowing you to give several options a look before committing to the one that works the best.

Just Go for It: Making a big change is always scary, but with cloud-based services appearing to be the next wave of high-tech must-haves, there's no better time than now to make the transition.



Lab Exercise: Explore a case study of an organization that shifted to cloud and discover the challenges faced by it for transforming into cloud?

2.5 Cloud Architecture

Cloud architecture indicates how individual technologies are integrated to create clouds. The IT environments abstract, pool, and share scalable resources across a network. How all the components and capabilities necessary to build a cloud are connected in order to deliver an online platform on which the applications can run.

A. Cloud Technologies

Certain technologies that are working behind the cloud computing platforms making cloud computing flexible, reliable, usable. These technologies are listed below:

- Virtualization
- Service-Oriented Architecture (SOA)
- Grid Computing
- Utility Computing

Virtualization is a technique, which allows to share single physical instance of an application or resource among multiple organizations or tenants (customers). It does so by assigning a logical name to a physical resource and providing a pointer to that physical resource when

Cloud Computing

demanded(Figure 1). The multitenant architectureoffers virtual isolation among the multiple tenants and therefore the organizations can use and customize the application as though they each have its own instance running.

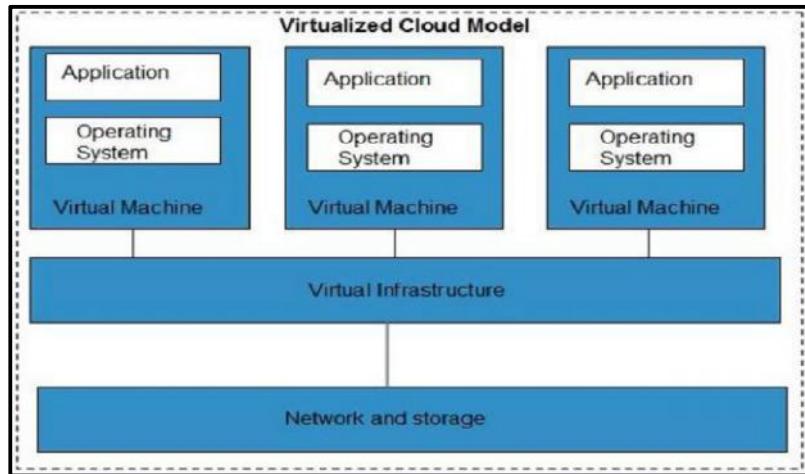


Figure 1: Virtualization Scenario

Service-Oriented Architecture (SOA)helps to use applications as a service for other applications regardless the type of vendor, product or technology (Figure 2). Therefore, it is possible to exchange of data between applications of different vendors without additional programming or making changes to services.

Grid Computingis like distributed computing in which a group of computers from multiple locations are connected with each other to achieve common objective. These computer resources are heterogeneous and geographically dispersed. Grid Computing breaks complex task into smaller pieces. These smaller pieces are distributed to CPUs that reside within the grid.

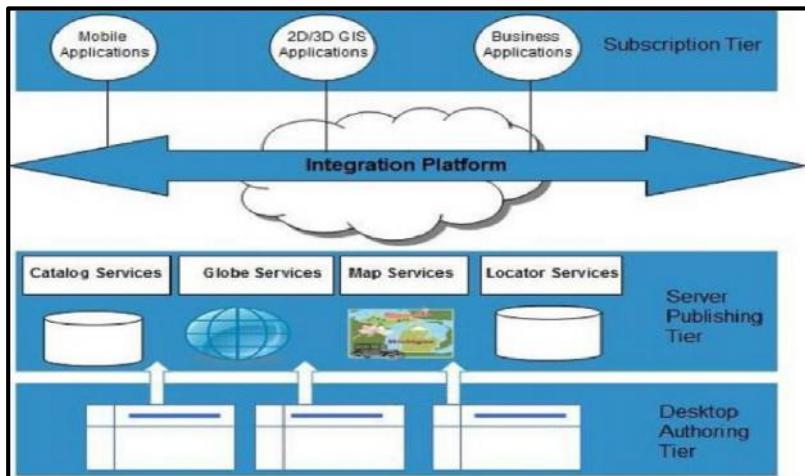


Figure 2: Service-Oriented Architecture (SOA)

Utility computingis based on Pay per Use model. It offers computational resources on demand as a metered service. Cloud computing, grid computing, and managed IT services are based on the concept of utility computing.

B. Cloud Architecture Ends

Cloud computing architecture comprises of many cloud components, each of them are loosely coupled. We can broadly divide the cloud architecture into two parts:

- Front End
- Back End

Each of the ends is connected through a network, usually via Internet.

Unit 02: Cloud Computing Architecture and Models

Front end refers to the client part of cloud computing system. It consists of interfaces and applications that are required to access the cloud computing platforms, e.g., Web Browser.

Back end refers to the cloud itself. It consists of all the resources required to provide cloud computing services. It comprises of huge data storage, virtual machines, security mechanism, services, deployment models, servers, etc.



Task: Consider the example of a cloud service provider and conduct a study of their back-end and front-end technologies, tools and platforms.

C. Cloud Layered Architecture

Cloud architecture describes its working mechanism. It includes the dependencies on which it works and the components that work over it. Cloud is a recent technology that is completely dependent on the Internet for its functioning. Cloud architecture can be divided into four layers based on the access of the cloud by the user (Figure 3).

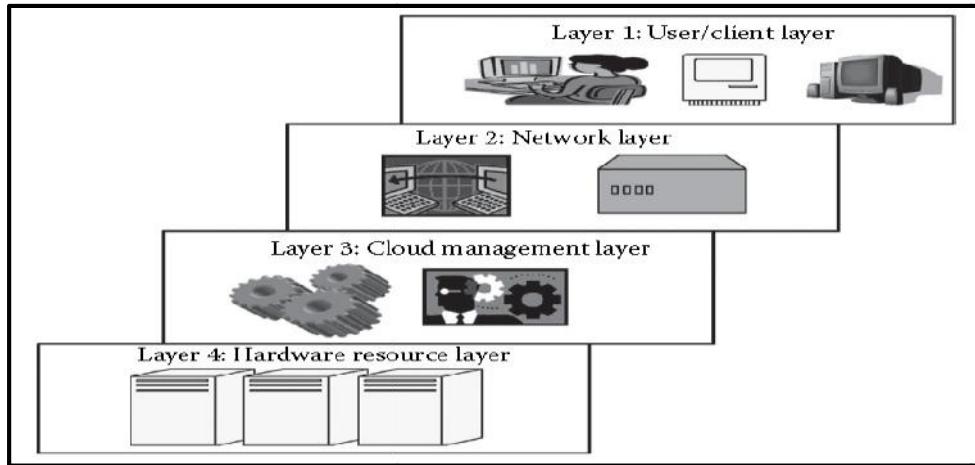


Figure 3: Layered Cloud Architecture

a) Layer 1 (User/Client Layer)

- This layer is the lowest layer in the cloud architecture. All the users or client belong to this layer. This is the place where the client/user initiates the connection to the cloud. The client can be any device such as a thin client, thick client, or mobile or any handheld device that would support basic functionalities to access a web application.
- Thin client here refers to a device that is completely dependent on some other system for its complete functionality. In simple terms, they have very low processing capability.
- Similarly, thick clients are general computers that have adequate processing capability. They have sufficient capability for independent work.
- Usually, a cloud application can be accessed in the same way as a web application. But internally, the properties of cloud applications are significantly different. Thus, this layer consists of client devices.

b) Layer 2 (Network Layer)

- This layer allows the users to connect to the cloud. The whole cloud infrastructure is dependent on this connection where the services are offered to the customers. This is primarily the Internet in the case of a public cloud. The public cloud usually exists in a specific location and the user would not know the location as it is abstract. Public cloud can be accessed all over the world.
- In the case of a private cloud, the connectivity may be provided by a local area network (LAN).

Cloud Computing

- Even in this case, the cloud completely depends on the network that is used.
- Usually, when accessing the public or private cloud, the users require minimum bandwidth, which is sometimes defined by the cloud providers.
- This layer does not come under the purview of Service Level Agreements (SLAs), that is, SLAs do not take into account the Internet connection between the user and cloud for quality of service (QoS).

Layer 3 (Cloud Management Layer)

- Layer 3 consists of software that are used in managing the cloud. The software can be a cloud OS, a software that acts as an interface between the data center (actual resources) and the user, or a management software that allows managing resources. This software usually allow resource management (scheduling, provisioning, etc.), optimization (server consolidation, storage workload consolidation), and internal cloud governance.
- This layer comes under the purview of SLAs, that is, the operations taking place in this layer would affect the SLAs that are being decided upon between the users and the service providers.
- Any delay in processing or any discrepancy in service provisioning may lead to an SLA violation.
- As per rules, any SLA violation would result in a penalty to be given by the service provider.

Layer 4 (Hardware Resource Layer)

- Layer 4 consists of provisions for actual hardware resources. Usually, in the case of a public cloud, a data center is used in the back end.
- Similarly, in a private cloud, it can be a data center, which is a huge collection of hardware resources interconnected to each other that is present in a specific location or a high configuration system.
- This layer comes under the purview of SLAs. This is the most important layer that governs the SLAs. This layer affects the SLAs most in the case of data centers.
- Whenever a user accesses the cloud, it should be available to the users as quickly as possible and should be within the time that is defined by the SLAs.
- If there is any discrepancy in provisioning the resources or application, the service provider has to pay the penalty. Hence, the datacenter consists of a high-speed network connection and a highly efficient algorithm to transfer the data from the datacenter to the manager.
- There can be a number of datacenters for a cloud, and similarly, a number of clouds can share a datacenter.

D. Cloud Deployment Models

A cloud infrastructure may be operated in one of the following deployment models:

- public cloud,
- private cloud,
- community cloud, or
- hybrid cloud.

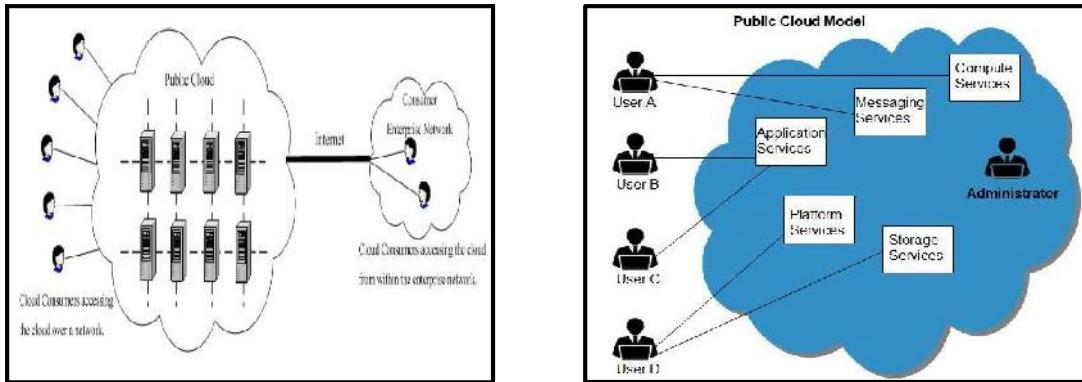
Unit 02: Cloud Computing Architecture and Models

Figure 4: Public Cloud Scenarios

The differences are based on how exclusive the computing resources are made to a Cloud Consumer.

- **Public Cloud:** Public cloud is one in which the cloud infrastructure and computing resources are made available to the general public over a public network (Figure 4). A public cloud is owned by an organization selling cloud services, and serves a diverse pool of clients.
- **Private Cloud:** Private cloud gives a single Cloud Consumer's organization the exclusive access to and usage of the infrastructure and computational resources. It may be managed either by:cloud consumer organization and may be hosted on the organization's premises (that is, on-site private clouds depicted in Figure 5), ora third party, outsourced to a hosting company (that is, outsourced private clouds depicted in Figure 6).

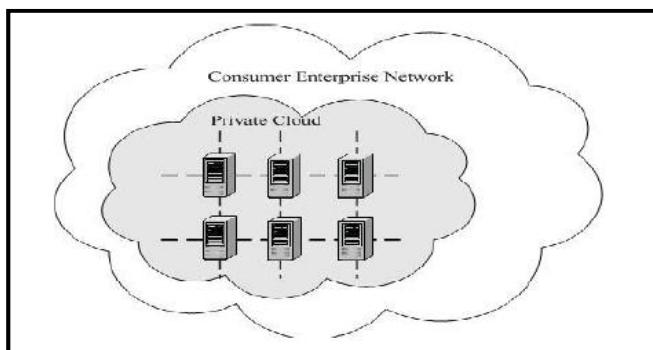


Figure 5: On-site Private Cloud

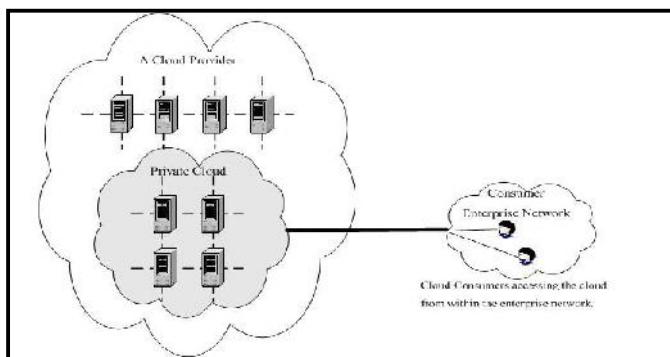


Figure 6: Outsourced Private Cloud

- **Hybrid Cloud:** Hybrid cloud(Figure 7) is a composition of two or more clouds (on-site private, on-site community, off-site private, off-site community or public) that remain as distinct entities

but are bound together by standardized or proprietary technology that enables data and application portability.

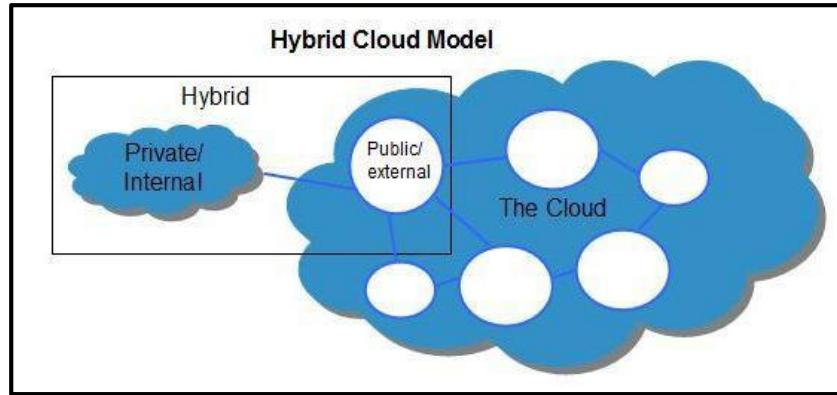


Figure 7: Hybrid Cloud Model Scenario 1

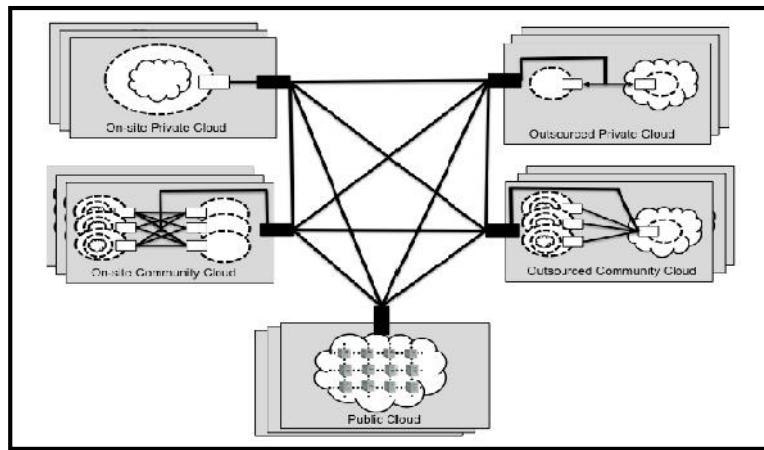


Figure 8: Hybrid Cloud Model Scenario 2

- **Community Cloud:** Community clouds serve a group of cloud consumers which have shared concerns such as mission objectives, security, privacy and compliance policy, rather than serving a single organization as does a private cloud (Figure 9). Similar to private clouds, a community cloud may be managed by: organizations and may be implemented on customer premise (that is, *on-site community cloud*), or, a third party, outsourced to a hosting company (that is, *outsourced community cloud*).

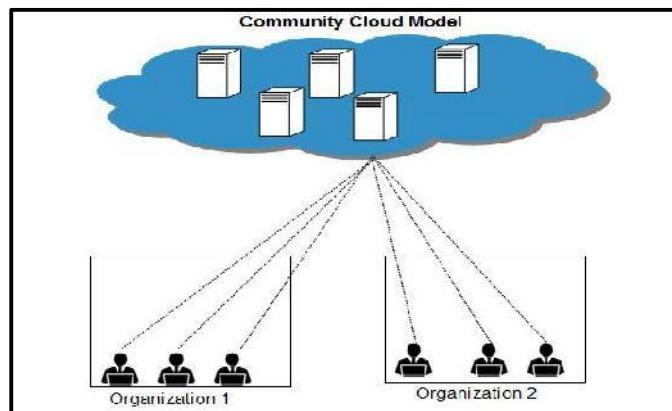


Figure 9: Community Cloud Model

An *on-site community cloud* comprised of a number of participant organizations (Figure 10 and Figure 11). A cloud consumer can access the local cloud resources, and also the resources of other participating organizations through the connections between the associated organizations.

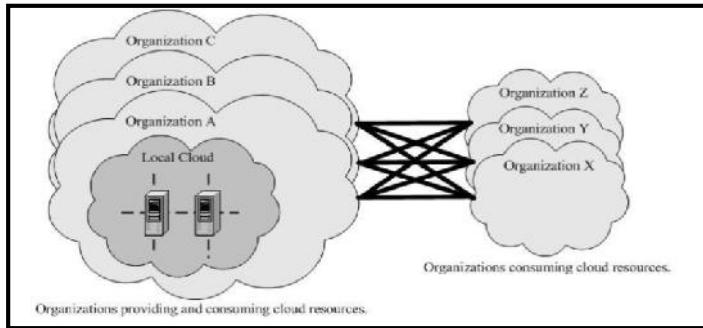


Figure 10: On-site Community Cloud Scenario

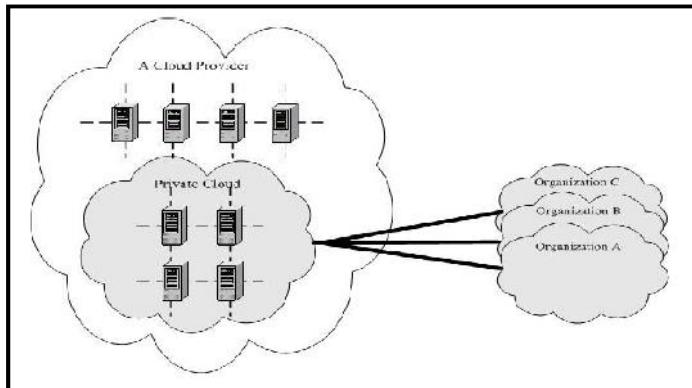


Figure 11: Outsourced Community Cloud

2.6 Cloud Storage

Most innovative technology to store, access, and collaborate through scalable cloud technology. Cloud computing model that stores valuable data through the web and operates it through a storage as a device service. Cloud storage can be delivered on-demand through a cloud computing provider that has the capacity to create a data-storage ecosystem that can eliminate costs through global durability with anytime and anywhere access. It is based on pay-as-you-go model. It involves third-party vendors take care of all the resources viz. capacity, bandwidth, and flexibility to make data accessible on a global platform.

Traditional Storage

The traditional storage option in which we use local physical drives to store the data at the primary location of the client. User generally uses the disk-based hardware to store data and these are used for copying, managing, and integrating the data to software. It has the following features:

- Fast, as they do not rely on internet speeds.
- Security can be manually set up by the user.
- Users can recover the data anytime without having accessibility issues.
- On-site backup and modification are easy.

Cloud Storage

Cloud Storage can be broadly classified into two categories:

- **Unmanaged Cloud Storage:** Unmanaged cloud storage means that the storage is preconfigured for the consumer. The consumer cannot format nor the consumer can install own file system or change drive properties.

- **Managed Cloud Storage:** Managed cloud storage offers online storage space on demand. Managed cloud storage system presents what appears to the user to be a raw disk that the user can partition and format.

Cloud storage is a service that allows saving data on offsite storage system managed by third-party and is made accessible by a web services API. The storage devices can be broadly classified into two categories:

- **Block Storage Devices:** Block storage devices offer raw storage to the clients. This raw storage can be partitioned to create volumes.
- **File Storage Devices:** File storage devices offers storage to clients in form of files, maintaining its own file system. This storage is in the form of Network Attached Storage (NAS).

Cloud Storage vs Traditional Storage

In the cloud storage, data (or files) are saved on a remote server, which is easily accessible from anywhere with internet access. Cloud storage allows access to the data from any device connected to the internet, including computers, tablets and smartphones. In contrast to local data storage, where data is stored on the hard drive of a local desktop or a laptop. With traditional alternatives, what you pay for is what you get. If one has invested in a lot of storage to support a new project, then one is still left paying the same amount once it's over- even if one doesn't need the storage anymore. Businesses are bound to the traditional office-based 9-5 unlike cloud in which, "teams can access everything from wherever they are".Traditional storage solutions involve physical devices where the monitoring, maintaining and patching these devices is up to users which is quite overwhelming. The better the backup plan, the faster the business can recover from disaster and be up and running as normal. Also, depending on the cause of the problem, it takes longer for traditional storage solutions to recover.Cloud storage is more flexible than traditional on-premise alternatives. It is easy to create a tailored solution that suits a user's specific requirements. The users have freedom to choose a course of action based on a user's current setup and what servers are chosen.

Creating Cloud Storage System

Cloud storage system stores multiple copies of data on multiple servers and in multiple locations. If one system fails, then it only requires changing the pointer to stored object's location. To aggregate storage assets into cloud storage systems, the cloud provider can use storage virtualization software, StorageGRID. It creates a virtualization layer that fetches storage from different storage devices into a single management system. The following are the features of cloud storage:

- All platforms can easily be accessed via a web browser.
- Offer apps for ease of access from a smartphone or tablet.
- Feature a directory structure similar to that of a computer drive; this facilitates navigation and organisation.
- Ease of Access: The access to the personal folders is perceived to be more cumbersome (involves 'more clicks').Many people are always logged into Google (and hence Google Drive) in the background, both at home and at school.
- Online Editing:OneDrive and Google Drive offer the possibility of editing documents inside a web browser. There is no additional software needed.The folders or specific files can be shared with others; this facilitates collaboration.
- Online Collaboration: Documents and folders can be shared with colleagues.Editing is possible without downloading the document, eliminating the need to email and save multiple versions of the same documents.

Considerations for Storing Data to Cloud

Various parameters that need to be taken into account whilst saving data on the cloud include:

- Security-** As data is managed by third-party vendors; it is heavily encrypted when data is stored and moving through various points. These security systems work well with permissions and access controls that are pre-defined on the web.
- Availability-** Data can be accessed whenever required. Mandatorily requires internet connectivity.

Examples of Cloud Storage

The examples of cloud storage and their characteristics are listed below (Figure 12):

Cloud Storage- Google Drive

- ‘Pure’ cloud computing service, with all the apps & storage found online.
- Can be used via desktop top computers, tablets like iPad or on smartphones.
- All of Google’s services can be considered cloud -based: Gmail, Google Calendar, Google Voice etc.
- Microsoft’s OneDrive: Similar to Google Drive.



Figure 12: Different Cloud Storage Examples

Cloud Storage- Dropbox

- Commonly used to store documents and images.
- One can set his/her phone to automatically send all pictures taken with it into their Dropbox account, so that even if one loses their phone, the pictures will still be available to him/her up in space.
- One can use it to access documents at home, and then save changes to it.
- Sugarsync is another example.

Cloud Storage- Apple iCloud

- Apple’s cloud service is primarily used by Apple users for online storage and synchronization of user’s mail, contacts, calendar, and more.
- All the data needed is available to a user on whichever device he/she seeks to access it from, iOS, Mac OS, or Windows device.
- If a user makes a change to a document, say, on one of their devices, it will automatically update it so that when next access is made to the account, the amended version will be available on whatever device you use.
- If a user has loads of data up there (perhaps pictures or films have made) then one will need to pay for extra storage.

2.7 Cloud Business Models

Cloud business models are all built on top of cloud computing, a concept that took over around 2006 when former Google’s CEO Eric Schmit mentioned it. It’s often not clear just what “the cloud” actually is, how it helps existing businesses or how entrepreneurs can use it to start or augment a

new business. Most cloud-based business models can be classified as cloud services delivery. While the models are primarily monetized via subscriptions, they are monetized via pay-as-you-go revenue models and hybrid models (subscriptions + pay-as-you-go).

I. NIST Cloud Computing Reference Model

NIST's long-term goal is to provide leadership and guidance around the cloud computing paradigm to catalyze its use within industry and government. NIST aims to shorten the adoption cycle, which will enable near-term cost savings and increased ability to quickly create and deploy safe and secure enterprise solutions. NIST aims to foster cloud computing practices that support interoperability, portability, and security requirements that are appropriate and achievable for important usage scenarios.

Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model promotes availability and is composed of five essential characteristics, three service models, and four deployment models.

Five Essential Characteristics

- On-demand Self-service: A consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service's provider.
- Broad Network Access: Capabilities are available over the network and accessed through standard mechanisms that promote use by heterogeneous thin or thick client platforms (e.g., mobile phones, laptops, and personal digital assistants [PDAs]).
- Resource Pooling: The provider's computing resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand. There is a sense of location independence in that the customer generally has no control or knowledge over the exact location of the provided resources but may be able to specify location at a higher level of abstraction (e.g., country, state, or datacenter). Examples of resources include storage, processing, memory, network bandwidth, and virtual machines.
- Rapid Elasticity: Capabilities can be rapidly and elastically provisioned, in some cases automatically, to quickly scale out and rapidly released to quickly scale in. To the consumer, the capabilities available for provisioning often appear to be unlimited and can be purchased in any quantity at any time.
- Measured Service: Cloud systems automatically control and optimize resource use by leveraging a metering capability³ at some level of abstraction appropriate to the type of service (e.g., storage, processing, bandwidth, and active user accounts). Resource usage can be monitored, controlled, and reported, providing transparency for both the provider and consumer of the utilized service.

A. Cloud Service Models

- **Software as a Service (SaaS):** The capability provided to the consumer is to use the provider's applications running on a cloud infrastructure. The applications are accessible from various client devices through a thin client interface such as a Web browser (e.g., Web-based email). The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user-specific application configuration settings.
- **Platform as a Service (PaaS):** The capability provided to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages and tools supported by the provider. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage, but

has control over the deployed applications and possibly application hosting environment configurations.

- **Infrastructure as a Service (IaaS):** The capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications. The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, deployed applications, and possibly limited control of select networking components (e.g., host firewalls).

B. Cloud Deployment Models

- **Private Cloud:** The cloud infrastructure is operated solely for an organization. It may be managed by the organization or a third party and may exist on premise or off premise.
- **Community Cloud:** The cloud infrastructure is shared by several organizations and supports a specific community that has shared concerns (example: mission, security requirements, policy, and compliance considerations). It may be managed by the organizations or a third-party and may exist on premise or off premise.
- **Public Cloud:** The cloud infrastructure is made available to the general public or a large industry group and is owned by an organization selling cloud services.
- **Hybrid Cloud:** The cloud infrastructure is a composition of two or more clouds (private, community, or public) that remain unique entities but are bound together by standardized or proprietary technology that enables data and application portability (e.g., cloud bursting for load balancing between clouds).

C. Actors in Cloud Computing Reference Model

There are certain actors that form important part of the cloud computing reference model (Figure 13 and Figure 14). Figure 15 depicts the interactions among the actors of cloud computing.

Actor	Definition
Cloud Consumer	A person or organization that maintains a business relationship with, and uses service from, <i>Cloud Providers</i> .
Cloud Provider	A person, organization, or entity responsible for making a service available to interested parties.
Cloud Auditor	A party that can conduct independent assessment of cloud services, information system operations, performance and security of the cloud implementation.
Cloud Broker	An entity that manages the use, performance and delivery of cloud services, and negotiates relationships between <i>Cloud Providers</i> and <i>Cloud Consumers</i> .
Cloud Carrier	An intermediary that provides connectivity and transport of cloud services from <i>Cloud Providers</i> to <i>Cloud Consumers</i> .

Figure 13: Actors in Cloud Computing Reference Model

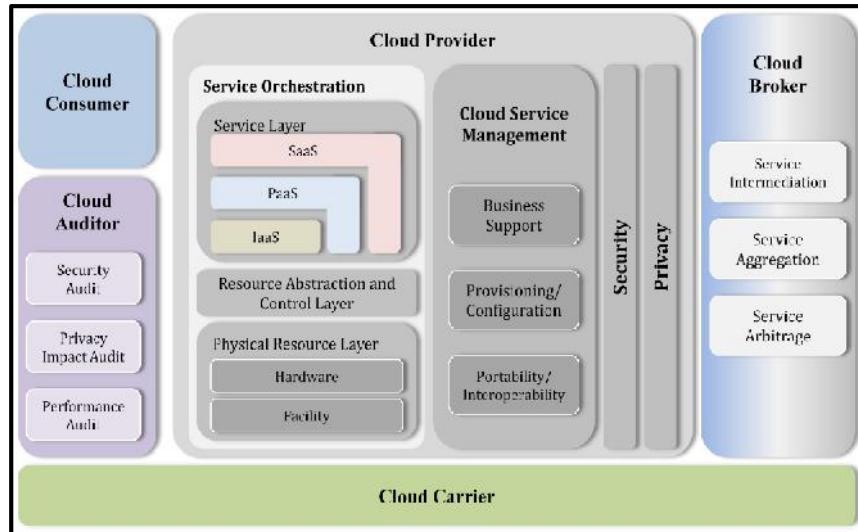


Figure 14: NIST Cloud Computing Reference Model

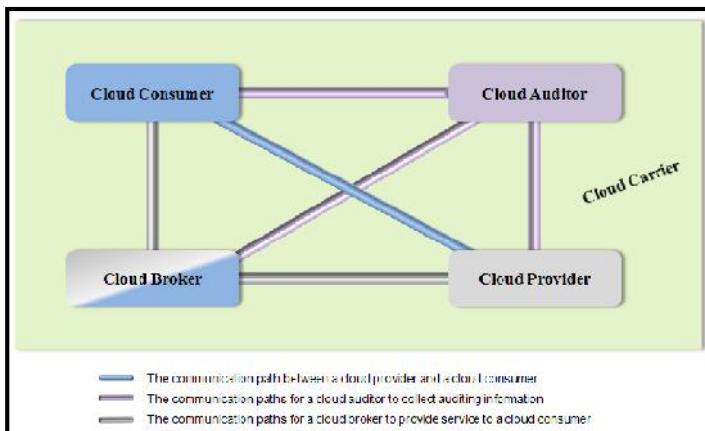


Figure 15: Interactions Among Actors in Cloud Computing

Cloud consumer may request service from a cloud broker instead of contacting a cloud provider directly. Cloud broker may create a new service by combining multiple services or by enhancing an existing service. The actual cloud providers are invisible to the cloud consumer and the cloud consumer interacts directly with the cloud broker. Cloud carriers provide the connectivity and transport of cloud services from cloud providers to cloud consumers. As illustrated in Figure, a cloud provider participates in and arranges for two unique service level agreements (SLAs), one with a cloud carrier (e.g. SLA2) and one with a cloud consumer (e.g. SLA1) (Figure 16).

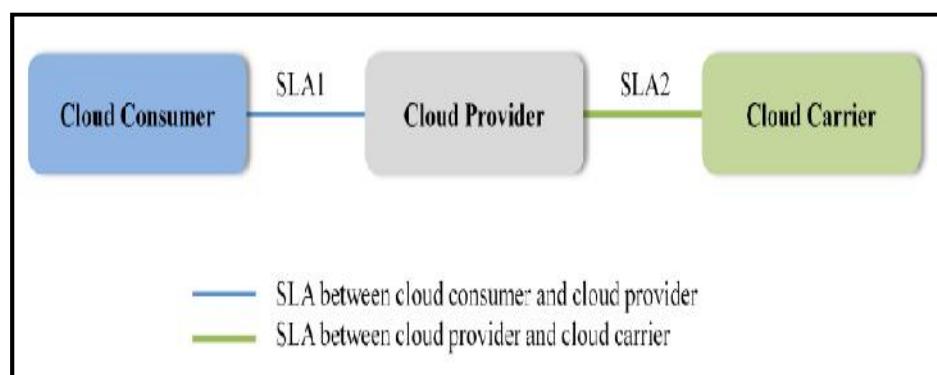


Figure 16: SLA Management Between Cloud Consumer and Cloud Carrier

A cloud provider arranges service level agreements (SLAs) with a cloud carrier and may request dedicated and encrypted connections to ensure the cloud services are consumed at a consistent level according to the contractual obligations with the cloud consumers. In this case, the provider may specify its requirements on capability, flexibility and functionality in SLA2 in order to provide essential requirements in SLA1. For a cloud service, a cloud auditor conducts independent assessments of the operation and security of the cloud service implementation. The audit may involve interactions with both the cloud consumer and the cloud provider.

Cloud consumer is a principal stakeholder for cloud computing service. It can be a person or organization that maintains a business relationship with, and uses the service from a cloud provider. Cloud consumer browses the service catalogue from a cloud provider, requests the appropriate service, sets up service contracts with the cloud provider, and uses the service. Cloud consumer may be billed for the service provisioned, and needs to arrange payments accordingly.

Cloud provider can be a person, or an organization. It is an entity responsible for making a service available to interested parties. A cloud provider can acquire and manage the computing infrastructure required for providing the services, run the cloud software that provides the services; and make arrangement to deliver the cloud services to the Cloud Consumers through network access. A cloud provider's activities can be described in five major areas:

- service deployment,
- service orchestration,
- cloud service management,
- security and privacy

Service orchestration refers to the composition of system components to support the cloud providers activities in arrangement, coordination and management of computing resources in order to provide cloud services to cloud consumers. Cloud service management includes all of the service-related functions that are necessary for the management and operation of those services required by or proposed to cloud consumers.

Cloud auditor is a party that can perform an independent examination of cloud service controls with the intent to express an opinion thereon. Audits are performed to verify conformance to standards through review of objective evidence. Cloud auditor can evaluate the services provided by a cloud provider in terms of security controls, privacy impact, performance, etc. An auditor may ensure that fixed content has not been modified and that the legal and business data archival requirements have been satisfied. As cloud computing evolves, the integration of cloud services can be too complex for cloud consumers to manage. Cloud consumer may request cloud services from a cloud broker, instead of contacting a cloud provider directly.

Cloud broker is an entity that manages the use, performance and delivery of cloud services and negotiates relationships between cloud providers and cloud consumers. A cloud broker can provide services in three categories:

- Service Intermediation: A cloud broker enhances a given service by improving some specific capability and providing value-added services to cloud consumers. The improvement can be managing access to cloud services, identity management, performance reporting, enhanced security, etc.
- Service Aggregation: A cloud broker combines and integrates multiple services into one or more new services. The broker provides data integration and ensures the secure data movement between the cloud consumer and multiple cloud providers.
- Service Arbitrage: Service arbitrage is similar to service aggregation except that the services being aggregated are not fixed. Service arbitrage means a broker has the flexibility to choose services from multiple agencies. The cloud broker, for example, can use a credit-scoring service to measure and select an agency with the best score.

Cloud carrier acts as an intermediary that provides connectivity and transport of cloud services between cloud consumers and cloud providers. Cloud carriers provide access to consumers through network, telecommunication and other access devices. For example, cloud consumers can obtain cloud services through n/w access devices, such as computers, laptops, mobile phones,

mobile Internet devices (MIDs), etc. The distribution of cloud services is normally provided by network and telecommunication carriers or a transport agent, where a transport agent refers to a business organization that provides physical transport of storage media such as high-capacity hard drives.



Notes: A cloud provider will set up SLAs with a cloud carrier to provide services consistent with the level of SLAs offered to cloud consumers, and may require the cloud carrier to provide dedicated and secure connections between cloud consumers and cloud providers.

II. Cloud Cube Model

Jericho Forum has designed the cloud cube model to help select cloud formations for secure collaboration. There are several "cloud formations- or forms of cloud computing. Each offer:

- different characteristics,
- varying degrees of flexibility,
- different collaborative opportunities, and
- different risks

One of the key challenges that businesses face when considering cloud computing as an option is to determine how to choose the cloud formation best suited to their various types of business operations.

Jericho Forum's developed cloud cube model with objectives related to cloud computing to distinctive enabling secure collaboration in the appropriate cloud formations best suited to the business needs. In Feb 2009, they delivered a practical framework geared to showing how to create the right Collaboration Oriented Architecture (COA) to assure secure business collaboration in de-perimeterised environments. For the Jericho Forum, the natural evolution from this is to address how to follow a well-structured path towards enabling secure business collaboration without becoming vulnerable to issues which may put at risk your data, or your ability to work with your chosen business parties, or your regulatory compliance.

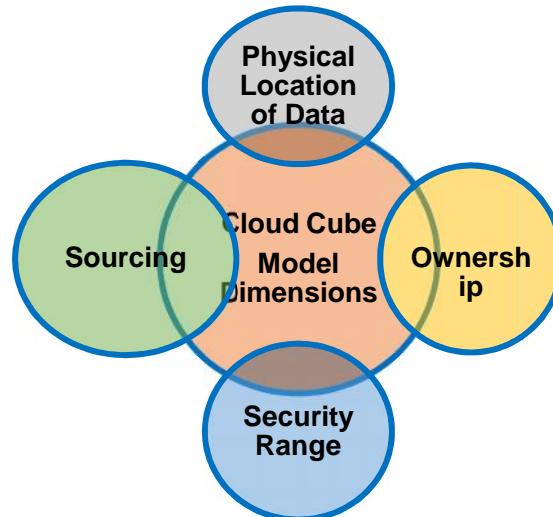


Figure 17: Dimensions of Cloud Cube Model

There are 4 major dimensions of cloud cube model (Figure 17):

- (a) **Physical Location of Data:** The location of data may be internally or externally which ultimately defines the organization's boundary.
- (b) **Ownership:** Ownership is proprietary or open; is a measurement for not only ownership of technology but also its interoperability, use of data & ease of data-transfer & degree of vendor's application's lock-in.

(c) **Security Range:** is parameterized or de-parameterized; which measures whether the operations are inside or outside the security boundary, firewall, etc.

(d) **Sourcing:** In-sourcing or out-sourcing; which defines whether the customer or the service provider provides the service.

Cloud cube model effectively summarizes four detailed dimensions (Figure 18):

- Internal/External
- Proprietary/Open
- Parameterized/De-parameterized Architectures
- Insourced/Outsourced

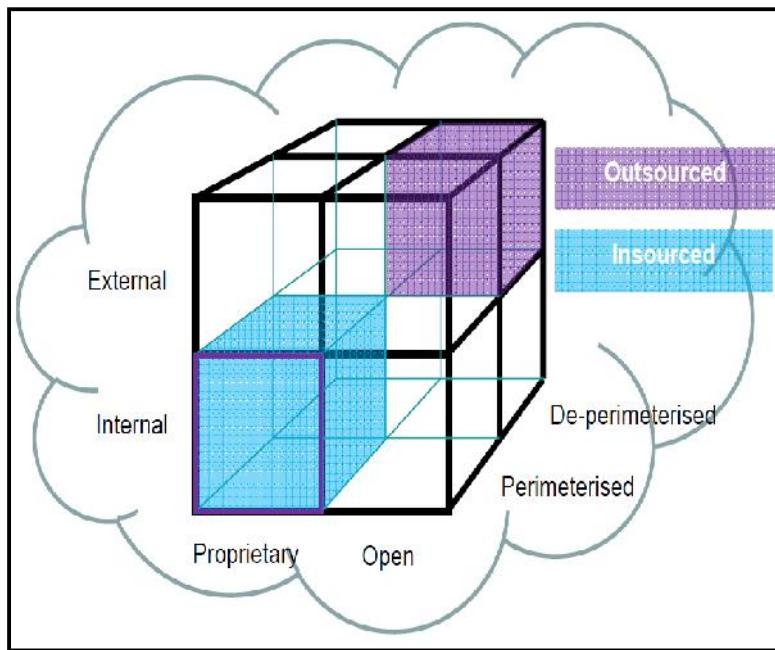


Figure 18: Sub-dimensions of Cloud Cube Model

Internal (I)/ External (E): It defines the physical location of the data: where does the cloud form you want to use exist inside or outside your organization's boundaries. If it is within your own physical boundary then it is internal. If it is not within your own physical boundary then it is external. For example, virtualized hard disks in an organization's data center would be internal, while Amazon S3 would be external at some location "off-site" (remote).

Proprietary (P)/ Open (O): This is the dimension that defines the state of ownership of the cloud technology, services, interfaces, etc. It indicates the degree of interoperability, as well as enabling "data/application transportability" between your own systems and other cloud forms. It indicates the ability to withdraw your data from a cloud form or to move it to another without constraint. Also, it indicates any constraints on being able to share applications. Proprietary means that the organization providing the service is keeping the means of provision under their ownership. As a result, when operating in clouds that are proprietary, you may not be able to move to another cloud supplier without significant effort or investment. Often the more innovative technology advances occur in the proprietary domain. As such the proprietor may choose to enforce restrictions through patents and by keeping the technology involved a trade secret.

Clouds that are open are using technology that is not proprietary, meaning that there are likely to be more suppliers, and you are not as constrained in being able to share your data and collaborate with selected parties using the same open technology. Open services tend to be those that are widespread and consumerised, and most likely a published open standard, for example email (SMTP).

Parameterized (Per) /De-parameterized (D-p) Architectures: It is the third dimension represents the "architectural mindset". It discusses: Are you operating inside your traditional IT perimeter or outside it? De-parameterization has always related to the gradual failure/removal

Cloud Computing

/shrinking / collapse of the traditional silo-based IT perimeter. Parameterized implies continuing to operate within the traditional IT perimeter, often signaled by "network firewalls".

When operating in the parameterized areas, you may simply extend your own organization's perimeter into the external cloud computing domain using a VPN and operating the virtual server in your own IP domain, making use of your own directory services to control access. Then, when the computing task is completed you can withdraw your perimeter back to its original traditional position.

De-parameterized, assumes that the system perimeter is architected following the principles outlined in the Jericho Forum's Commandments and Collaboration Oriented Architectures Framework. In a de-parameterized environment an organization can collaborate securely with selected parties (business partner, customer, supplier, outworker) globally over any COA capable network. De-parameterized areas in our Cloud Cube Model use both internal and external domains but the collaboration or sharing of data should not be seen as internal or external. It is controlled by and limited to the parties that the using organizations select.

Insourced/ Outsourced: The fourth dimension that has 2 states in each of the 8 cloud forms: Per(IP,IO,EP,EO) and D-p(IP,IO,EP,EO), that responds to the question:

"Who do you want running your Clouds?"

Outsourced: the service is provided by a 3rd party

Insourced: the service is provided by your own staff under your control

These 2 states describe who is managing delivery of the cloud service(s) that you use. This is primarily a policy issue (that is, a business decision, not a technical or architectural decision) which must be embodied in a contract with the cloud provider.

Summary

- Cloud computing signifies a major change in the way we run various applications and store our information. Everything is hosted in the "cloud", a vague assemblage of computers and servers accessed via the Internet, instead of the method of running programs and data on a single desktop computer.
- With cloud computing, the software programs you use are stored on servers accessed via the Internet and are not run from your personal computer. Hence, even if your computer stops working, the software is still available for use.
- The "cloud" itself is the key to the definition of cloud computing. The cloud is usually defined as a large group of interconnected computers. These computers include network servers or personal computers.
- Cloud computing has its ancestors both as client/server computing and peer-to-peer distributed computing. It is all about how centralized storage of data and content facilitates collaborations, associations and partnerships.
- With cloud storage, data is stored on multiple third-party servers, rather than on the dedicated servers used in traditional networked data storage.
- Cloud storage system stores multiple copies of data on multiple servers and in multiple locations. If one system fails, then it only requires changing the pointer to stored object's location.
- Jericho Forum has designed the Cloud Cube Model to help select cloud formations for secure collaboration.

Keywords

Cloud: The cloud is usually defined as a large group of interconnected computers. These computers include network servers or personal computers.

Distributed Computing: Distributed computing is any computing that involves multiple computers remote from each other that each have a role in a computation problem or information processing.

Group collaboration software: It provides tools for groups of people or organizations to share information and coordinate activities.

Infrastructure as a Service (IaaS): The capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications.

Private Cloud: The cloud infrastructure is operated solely for an organization. It may be managed by the organization or a third party and may exist on premise or off premise.

Self Assessment

1. Private cloud is a..... .
 - A. standard cloud service offered via the Internet
 - B. a cloud service inaccessible to anyone but the cultural elite
 - C. cloud architecture maintained within an enterprise data center.
 - D. None of the above

2. "Cloud" in cloud computing represents?
 - A. Hard drives
 - B. Floppy Disk
 - C. External drive
 - D. Internet

3. Which of the following offer online editing and storage service using cloud capabilities?
 - A. Google Drive
 - B. Zoho
 - C. Zoom
 - D. None of the above

4. Cloud computing is a _____ delivery model that provides scalability, fast delivery and greater IT efficiencies.
 - A. authentication-based
 - B. payment-based
 - C. subscription-based
 - D. hybrid-based

5. _____ refers to minimizing the effect of cloud computing on environment.
 - A. Adaptability
 - B. Feasibility
 - C. Durability
 - D. Sustainability

6. Which of the following consideration(s) need to be taken when getting into the cloud?
 - A. Locate the services for each part of your business
 - B. Understand that even tech-savvy companies struggle with cloud business
 - C. Decide the level of cloud services that are right for your business

- D. All of the above
7. _____ is a technique that allows to share single physical instance of an application or resource among multiple organizations or tenants (customers).
- A. Consolidation
 - B. Anonymization
 - C. Virtualization
 - D. Pooling
8. Virtualization allows sharing of physical instances among multiple organizations by assigning a _____ name to a physical resource and providing a pointer to that physical resource when demanded.
- A. physical
 - B. identification
 - C. portrayal
 - D. logical
9. Which of the following architecture offers virtual isolation among the multiple tenants and therefore the organizations can use and customize the application as though they each have its own instance running?
- A. Multitenant
 - B. Multicore
 - C. Dual core
 - D. Instant
10. Utility computing is based on _____ model.
- A. Pay per use
 - B. Pay as you go
 - C. Both above options
 - D. None of the above
11. _____ architecture helps to use applications as a service for other applications regardless the type of vendor, product or technology.
- A. Application-oriented
 - B. Service-oriented
 - C. Product-oriented
 - D. Delivery-oriented
12. Which of the following consists of all the resources required to provide cloud computing services (huge data storage, virtual machines, security mechanism, services, deployment models, servers, etc.)?
- A. Front end
 - B. Back end
 - C. Middleware
 - D. None of the above

13. _____ refers to a device that is completely dependent on some other system for its complete functionality.
- Thick client
 - Fat client
 - Hypervisor
 - Thin client
14. SLA stands for
- Service Lowering Associations
 - Software Level Agents
 - Service Level Agreements
 - Software Level Associations
15. The following is an example of online collaboration service(s) based on cloud computing:
- One Drive
 - Google Drive
 - Zoho tools
 - All of the above

Answers for Self Assessment

- | | | | | |
|-------|-------|-------|-------|-------|
| 1. C | 2. D | 3. A | 4. C | 5. D |
| 6. D | 7. C | 8. D | 9. A | 10. C |
| 11. B | 12. B | 13. D | 14. C | 15. D |

Review Questions

- Explain different models for deployment in cloud computing?
- Compare cloud storage and traditional storage?
- How cloud storage works? What are different examples of cloud storage currently?
- How to create cloud storage system?
- What are the considerations for getting into the cloud?
- Which different technologies formulate the cloud computing architecture?
- Explore the layered cloud architecture with elaborative explanation of each layer?
- Discuss the different issues faced in cloud environment?
- Explain the cloud cube model?
- Discuss the cloud business models?



Further Readings

Cloud Computing: Concepts, Technology and Architecture by Erl, Pearson Education.

Cloud Computing Black Book by Kailash Jayaswal, Jagannath Kallakurchi, Donald J.

Houde, Deven Shah, Kogent Learning Solutions, DreamTech Press



Web Links

[Cloud Computing Overview - Tutorialspoint](#)

[Cloud Computing \(w3schools.in\)](#)

[Elastic Sky Labs: Cloud Cube Model \(jasongaudreau.com\)](#)

[https://profsandhu.com/cs6393_s20/NIST_SP-500-291_Version-2_2013_June18_extract.pdf#:~:text=The NIST cloud computing reference architecture is a,the requirements%2C structures%2C and operations of cloud computing.](https://profsandhu.com/cs6393_s20/NIST_SP-500-291_Version-2_2013_June18_extract.pdf#:~:text=The%20NIST%20cloud%20computing%20reference%20architecture%20is,a,the%20requirements%2C%20structures%2C%20and%20operations%20of%20cloud%20computing)

[Cloud Business Models - FourWeekMBA](#)

Unit 03: Cloud Services

CONTENTS

- Objectives
- Introduction
- 3.1 Classification of Cloud Services
- 3.2 Cloud Service Providers
- 3.3 CSP Platforms and Technologies
- 3.4 Criteria for Choosing the Best Cloud Service Provider
- 3.5 Other Cloud Services
- Summary
- Keywords
- Self Assessment
- Answers for Self Assessment
- Review Questions
- Further Readings

Objectives

- Learn about the cloud services and their types.
- Understand Software as a Service, Platform as a Service, Infrastructure as a Service.
- Analyze cloud service providers in terms of their services, platforms and technology offerings.
- Explore the different cloud services such as Storage-as-a-Service, Data-as-a-Service, Database-as-a-Service, Communication-as-a-Service, Monitoring-as-a-Service, Network-as-a-Service, Healthcare-as-a-Service, Education-as-a-Service etc.

Introduction

Cloud services refer to any IT services that are provisioned and accessed from a cloud computing provider. Cloud computing is a broad term that incorporates all delivery and service models of cloud computing and related solutions. Cloud services are delivered over the internet and accessible globally from the internet. Cloud computing services and deployment models describe how the services delivery is carried out in cloud computing. These indicate the topological layouts for the cloud computing. The entities basically correspond to the operational components in cloud computing.

Cloud services provide many IT services traditionally hosted in-house, including provisioning an application/database server from the cloud, replacing in-house storage/backup with cloud storage and accessing software and applications directly from a web browser without prior installation. Cloud services provide great flexibility in provisioning, duplicating and scaling resources to balance the requirements of users, hosted applications and solutions. Cloud services are built, operated and managed by a cloud service provider, which works to ensure end-to-end availability, reliability and security of the cloud.

3.1 Classification of Cloud Services

Cloud services are provisioned through the cloud service models. Currently, cloud service models can be abstracted into two high levels of classifications, namely, Archetypal and Hybrid cloud service models.

*Cloud Computing***a) Classic or Archetypal Category**

The classic or archetypal category pertains to the three original cloud service models (Figure 1):

- Software-as-a-Service (SaaS): SaaS is a software delivery model that helps users to access applications through a simple interface over the Internet. The providers of SaaS possess total control on the applications and they enable the users to access them. The users have an illusion as if their applications are locally hosted without being bothered of the application background details. The typical SaaS examples are social media platforms, email boxes, Facebook, Google Apps etc.
- Platform-as-a-Service (PaaS): PaaS model is more of an urbane model that offers building, testing, deployment, and hosting environments for applications created by users or otherwise acquired from them. The prominent platforms of Microsoft Azure, Google App Engine are perfect PaaS cloud models.
- Infrastructure-as-a-Service (IaaS): IaaS is an undemanding model for delivering the cloud service. It provides an actual physical infrastructure support that includes computing, storing, networking, and other primary resources to the users. The users benefit by renting resources from the IaaS providers and using them on demand instead of incurring their own infrastructure. Examples- Amazon EC2, Nimbus etc.

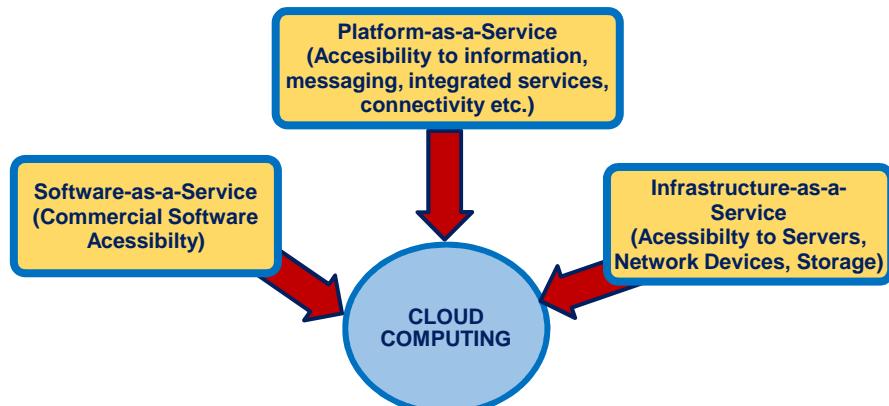


Figure 1: Cloud Classic Service Models

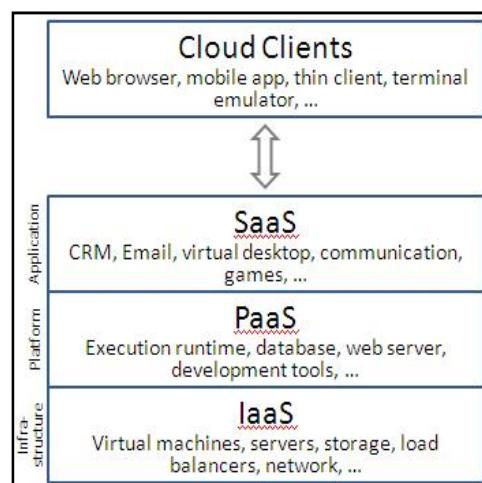


Figure 2: Cloud Service Models

Infrastructure-as-a-Service (IaaS)

- Provides virtual machines, virtual storage, virtual infrastructure, and other hardware assets as resources that clients can provision.

Unit 03: Cloud Services

- Service provider manages all infrastructure, while the client is responsible for all other aspects of the deployment.
- Can include the operating system, applications, and user interactions with the system. Figure 3 shows the different concepts associated with IaaS.



Figure 3: Different Concepts Associated with IaaS

Enterprise Infrastructure: Such infrastructure is offered by internal business networks, such as private clouds and virtual local area networks, which utilize pooled server and networking resources and in which a business can store their data and run the applications they need to operate day-to-day. The expanding businesses can scale their infrastructure in accordance with their growth whilst private clouds (accessible only by the business itself) can protect the storage and transfer of the sensitive data that some businesses are required to handle.

Cloud Hosting: The hosting of websites on virtual servers which are founded upon pooled resources from underlying physical servers. A website hosted in the cloud, for example, can benefit from the redundancy provided by a vast network of physical servers and on demand scalability to deal with unexpected demands placed on the website.

Virtual Data Centers (VDC): A virtualized network of interconnected virtual servers which can be used to offer enhanced cloud hosting capabilities, enterprise IT infrastructure or to integrate all of these operations within either a private or public cloud implementation.

Advantages of IaaS: Figure 4 lists the different advantages of IaaS.

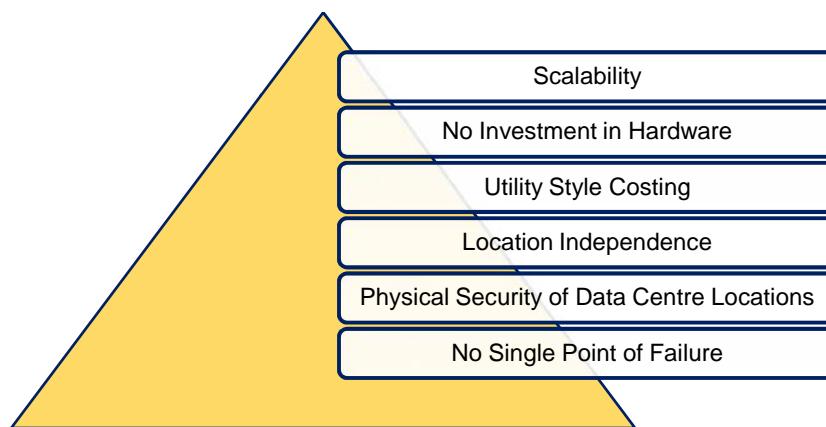


Figure 4: Advantages of IaaS

- **Scalability:** Resource is available as and when the client needs it and, therefore, there are no delays in expanding capacity or the wastage of unused capacity
- **No Investment in Hardware:** The underlying physical hardware that supports an IaaS service is set up and maintained by the cloud provider, saving the time and cost of doing so on the client side
- **Utility Style Costing:** The service can be accessed on demand and the client only pays for the resource that they actually use.
- **Location Independence:** The service can usually be accessed from any location as long as there is an internet connection and the security protocol of the cloud allows it.

Cloud Computing

- Physical Security of Data Centre Locations: The services available through a public cloud, or private clouds hosted externally with the cloud provider, benefit from the physical security afforded to the servers which are hosted within a data center.
- No Single Point of Failure: If one server or network switch, for example, were to fail, the broader service would be unaffected due to the remaining multitude of hardware resources and redundancy configurations. For many services if one entire data center were to go offline, nevermind one server, the IaaS service could still run successfully.

Uses of IaaS: IaaS is useful in the following situations:

- Where Demand is Very Volatile- any time there are significant spikes and troughs in terms of demand on the infrastructure amazon.in, Snapdeal, Flipkart- during festival season.
- For new enterprise without capital to invest in hardware. Example- entrepreneurs starting on a shoestring budget.
- Where the enterprise is growing rapidly and scaling hardware would be problematic. Example- A company that experience huge success immediately.
- Where the enterprise is growing rapidly and scaling hardware would be problematic. Example- A company that experience huge success immediately - animato, Pinterest.
- For specific line of business, trial or temporary infrastructural needs.

Examples of IaaS

- Amazon Web Services: A public cloud that offers subscribers access to virtual servers for product deployment, Cloud storage, tools for development, testing, and analytics. The application provides a ready-to-use environment to develop and test the product and offers the full cloud infrastructure for its deployment and maintenance.
- Microsoft Azure: Combination of IaaS and platform as a service, the software offers 100+ services for software development, administration, and deployment, provides tools for working with innovative technologies (big data, machine learning, Internet of Things), etc.
- IBM Infrastructure: IBM uses its in-house services to store the data of infrastructure users, enabling remote data access via Cloud computing. IBM servers support AI, blockchain, and the Internet of Things. The infrastructure also provides Cloud storage and virtual development environments, enabled on the subscription basis.
- Google Cloud Infrastructure: The large network of international servers that provides users access to remote Cloud data centers. Companies can store their information in Asia, Europe, Latin America, which minimizes the risk of a security breach.



Task: Explore how IaaS services are facilitated by Amazon, Google, Microsoft.

Platform-as-a-Service (PaaS)

PaaS is a category of cloud computing service that provides a platform and environment to allow developers to build applications and services over the internet (Figure 5). PaaS services are hosted in the cloud and accessed by users simply via their web browser. PaaS allows the users to create software applications using tools supplied by the provider. PaaS services can consist of preconfigured features that customers can subscribe to; they can choose to include the features that meet their requirements while discarding those that do not.

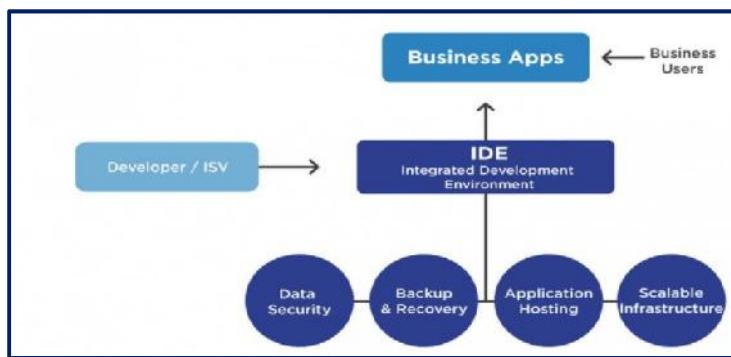


Figure 5: PaaS Offerings

Cloud consumers do not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage, but has control over the deployed applications and possibly configuration settings for the application-hosting environment. PaaS is expected to grow more than 3,000% by 2026. From \$1.78b to \$68.38b, more than double SaaS, the expected growth during the same period. Other PaaS characteristics are:

PaaS provides virtual machines, operating systems, applications, services, development frameworks, transactions, and control structures. The clients can deploy their applications on the cloud infrastructure or use applications that were programmed using languages and tools that are supported by the PaaS service provider. The service provider manages the cloud infrastructure, the operating systems, and the enabling software. The client is responsible for installing and managing the application that it is deploying. PaaS provides a Programming IDE in order to develop their service among PaaS. It integrates the full functionalities which are supported from the underlying runtime environment. It offers some development tools, such as profiler, debugger and testing environment. Examples of PaaS Service Providers: Microsoft Windows Azure, Google App Engine, Hadoop etc.

PaaS providers can assist developers from the conception of their original ideas to the creation of applications, and through to testing and deployment. This is all achieved in a managed mechanism. As with most cloud offerings, PaaS services are generally paid for on a subscription basis with clients ultimately paying just for what they use. Following are the PaaS offerings:

- Operating System
- Coding & Server-side Scripting Environment
- Database Management System
- Server Software
- Support and Hosting
- Storage & Network Access
- Tools for Design and Development

How PaaS Works

PaaS allows users to create software applications using tools supplied by the provider. PaaS services can consist of preconfigured features that customers can subscribe to; they can choose to include the features that meet their requirements while discarding those that do not. The infrastructure and applications are managed for customers and support is available. The services are constantly updated, with existing features upgraded and additional features added.

Advantages of PaaS: PaaS helps to create an abstracted environment that supports an efficient, cost-effective, and repeatable process for the creation and deployment of high-quality applications. The focus is on development, not Ops such as,

- Programmers' development environment
- Presentation layer: HTML, CSS, JavaScript
- Control layer: Web Server code
- Data layer: Data Model
- Optionally, analytics

Cloud Computing

Uses of PaaS: PaaS has the following uses:

- Users don't Need to Invest in Physical Infrastructure: Being able to 'rent' virtual infrastructure has both cost benefits and practical benefits. They don't need to purchase hardware themselves or employ the expertise to manage it. This leaves them free to focus on the development of applications.
- Makes development possible for 'non-experts': with some PaaS offerings anyone can develop an application. They can simply do this through their web browser utilizing one-click functionality, example, WordPress.
- Flexibility: customers can have control over the tools that are installed within their platforms and can create a platform that suits their specific requirements. They can 'pick and choose' the features they feel are necessary.
- Adaptability: Features can be changed if circumstances dictate that they should.
- Teams in various locations can work together: As an internet connection and web browser are all that is required, developers spread across several locations can work together on the same application build.
- Security: Security is provided, including data security and backup and recovery.

Examples of PaaS (Figure 6)



Figure 6: Examples of PaaS

- **AWS Elastic Beanstalk:** A web platform for software deployment and management, powered by the AWS Cloud. Users upload their applications to the service, and it automatically monitors the performance, load capacity, and checks for deployment errors.
- **Apache Stratos:** Cloud computing platform for arranging PHP and MySQL. The PaaS provides users with ready-to-use tools for database development and testing, performance monitoring, integration, and billing.
- **Magento Commerce Cloud:** Magento cloud offers tools for e-commerce development, testing, deployment, and maintenance. The Cloud environment allows accessing the store settings anytime and anywhere as well as automates the key processes.

Software-as-a-Service (SaaS)

SaaS facilitates complete operating environment with applications, management, and the user interface. The applications are accessible from various client devices through a thin client interface such as a web browser (e.g., web-based email). The consumer does not manage or control the underlying cloud infrastructure including network, servers, OSs, storage, or even individual application capabilities. Examples: Google Apps, SalesForce.com, EyeOS etc.

SaaS describes any cloud service where consumers are able to access software applications over the internet. The applications are hosted in "the cloud" and can be used for a wide range of tasks for both individuals and organizations. Google, Twitter, Facebook and Flickr are all examples of SaaS, with users able to access the services via any internet-enabled device (Figure 7). The enterprise users are able to use applications for a range of needs, including accounting and invoicing, tracking sales, planning, performance monitoring and communications (including webmail and instant messaging).



Figure 7: SaaS Offerings

SaaS is often referred to as software-on-demand and utilizing it is akin to renting software rather than buying it. The SaaS users, however, subscribe to the software rather than purchase it, usually on a monthly basis. The applications are purchased and used online with files saved in the cloud rather than on individual computers.

Advantages of SaaS

- No Additional Hardware Costs: Processing power required to run the applications is supplied by the cloud provider.
- No Initial Setup Costs: Applications are ready to use once the user subscribes.
- Pay for What You Use: If a piece of software is only needed for a limited period then it is only paid for over that period and subscriptions can usually be halted at any time.
- Usage is Scalable: If a user decides they need more storage or additional services, for example, then they can access these on demand without needing to install new software or hardware.
- Updates are Automated: Whenever there is an update it is available online to existing customers, often free of charge. No new software will be required as it often is with other types of applications and the updates will usually be deployed automatically by the cloud provider.
- Cross-Device Compatibility: SaaS applications can be accessed via any internet enabled device, which makes it ideal for those who use a number of different devices, such as internet enabled phones and tablets, and those who don't always use the same computer.
- Accessible from Any Location: Rather than being restricted to installations on individual computers, an application can be accessed from anywhere with an internet enabled device.
- Applications can be Customized and White-labelled: With some software, customization is available meaning it can be altered to suit the needs and branding of a particular customer.

Examples of SaaS (Error! Reference source not found.)



Figure 8: Examples of SaaS

- **Google's G Suite:** Top cloud service provides businesses with access to management, communication, and organization tools and uses cloud for data computing. Gmail, Google Drive,

Cloud Computing

Google Docs, Google Planner, Hangouts—these are all SaaS tools that can be accessed anytime and anywhere.

- **Microsoft Office 365:** The series of web services that provide business owners and individuals with access to Microsoft Office main tools directly from their browsers. Users can access Microsoft editing tools, business email, communication instruments, and documentation software.
- **Salesforce:** The most popular CRM on the market that unites marketing, communication, e-commerce. Salesforce uses cloud computing benefits to provide access to its services and internal data. Business owners can keep track of their sales, client relations, communications, and relevant tasks from any device. Salesforce can be integrated into the website — the information about incoming leads will be sent to the platform automatically.

Difference Between IaaS, PaaS, and SaaS

Table 1 lists the key differences between IaaS, PaaS and SaaS. Figure 9 depicts the summarization of the three cloud service models.

Table 1: Key Differences between IaaS, PaaS and SaaS

IaaS	PaaS	SaaS
Provides a virtual data center to store information and create platforms for app development, testing, & deployment	It provides virtual platforms and tools to create, test, and deploy apps	Provides web software and apps to complete business tasks
Provides access to resources such as virtual machines, virtual storage etc.	Provides runtime environments & deployment tools for applications	Provides software as a service to the end-users
Used by the network architects	Used by developers	Used by the end users
Provides only Infrastructure	Provides Infrastructure + Platform	Provides Infrastructure + Platform +Software

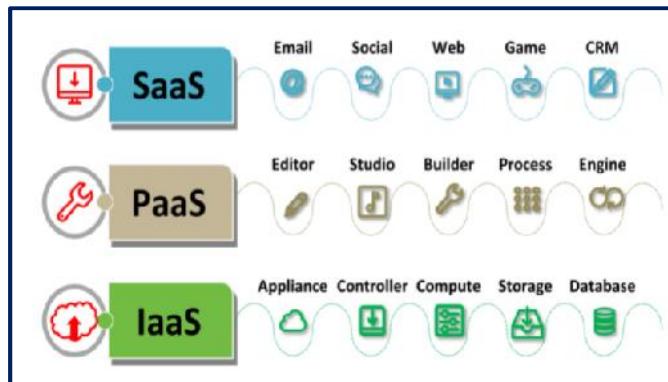


Figure 9: Summarization of Three Cloud Service Models

b) Hybrid Cloud Service Models

Driven by the need for more flexibility and control on the deployment of their applications, hybrid models emerged. Although, the archetypal service models, SaaS, PaaS, and IaaS models have contributed significantly towards delivering cloud services but with the gradual progress of cloud and the users demanding more flexibility and control on the deployment of their applications, these models have often been speculated and considered for improvements. This resulted in consolidation of PaaS and IaaS models paving development and nurturing of new hybrid cloud service delivery models.

Hybrid models have enabled the users to create and experience a single gamut of cloud services. Big IT tycoons such as Google, Microsoft and Amazon have initiated efforts to concurrently blend IaaS and PaaS. Kubernetes, a Google initiative is a hybrid cloud-based solution merging IaaS and PaaS. Also, with the advent of new and innovative computing strategies, the researchers have further classified the cloud services into Hardware-as-a-Service (HaaS) and Data-as-a-Service (DaaS). A recent of new cloud delivery model is Metal-as-a-Service model (MaaS). It is provisioning assemble of Ubuntu Linux-based OS designed to support the deployment and active provisioning of frenzied, high scale computing such as processing of big data applications and services.

3.2 Cloud Service Providers

The cloud service providers are also called as service vendors. A vendor that provides IT solutions and/or services to end users and organizations. This broad term incorporates all IT businesses that provide products and solutions through services that are on-demand, pay per use or a hybrid delivery model. Service vendors incorporate all the IT businesses that provide products and solutions through services that are on-demand, pay-per-use or a hybrid delivery model. A service provider's delivery model generally differs from conventional IT product manufacturers or developers. Typically, a service provider does not require purchase of an IT product by a user or organization. A service provider builds, operates and manages these IT products, which are bundled and delivered as a service/solution. In turn, a customer accesses this type of solution from a service provider via several different sourcing models, such as a monthly or annual subscription fee.

Different Service Providers (Figure 10)

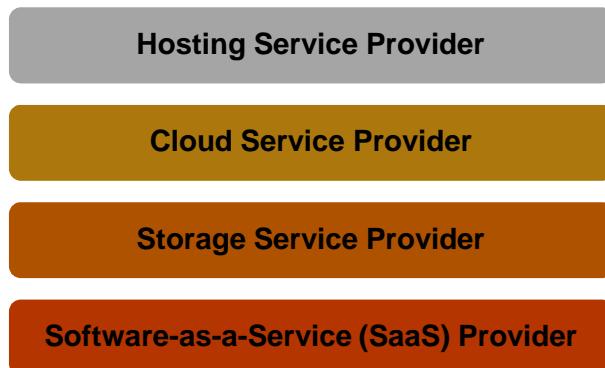


Figure 10: Different Types of Service Providers

- Hosting Service Provider- A type of Internet hosting service that allows individuals and organizations to make their website accessible via the World Wide Web. Web host companies provide space on a server owned or leased for use by clients, as well as providing Internet connectivity, typically in a data center.
- Cloud Service Provider- Offer cloud-based services
- Storage Service Provider- A Storage service provider (SSP) is any company that provides computer storage space and related management services. SSPs also offer periodic backup and archiving.
- Software-as-a-Service (SaaS) Provider- SaaS providers allow the users to connect to and use cloud-based apps over the Internet. Common examples are email, calendaring and office tools (such as Microsoft Office 365).

Cloud service provider can be a third-party company offering a cloud-based platform, infrastructure, application, or storage services. Much like a homeowner would pay for a utility such as electricity or gas, companies typically have to pay only for the amount of cloud services they use, as business demands require. The cloud services can reduce business process costs when compared to on-premise IT. Such services are managed by the Cloud Service Provider (CSPs). CSP provide all the resources needed for the application and hence the company needs not to worry about resource

Cloud Computing

allocation. Cloud services can dynamically scale up based on users' needs. CSP companies establish public clouds, manage private clouds, or offer on-demand cloud computing components (also known as cloud computing services) like IaaS, PaaS, and SaaS.

Cloud Service Providers are helpful way to access computing services that you would otherwise have to provide on your own, such as:

- Infrastructure: The foundation of every computing environment. This infrastructure could include networks, database services, data management, data storage (known in this context as cloud storage), servers (cloud is the basis for serverless computing), and virtualization.
- Platforms: Tools needed to create and deploy applications. These platforms could include operating systems, middleware, and runtime environments.
- Software: Ready-to-use applications. This software could be custom or standard applications provided by independent service providers.

3.3 CSP Platforms and Technologies

The development of a cloud computing application happens by leveraging platforms and frameworks of the CSPs (Figure 11). CSPs provide different types of services, from the bare-metal infrastructure to customizable applications serving specific purposes.

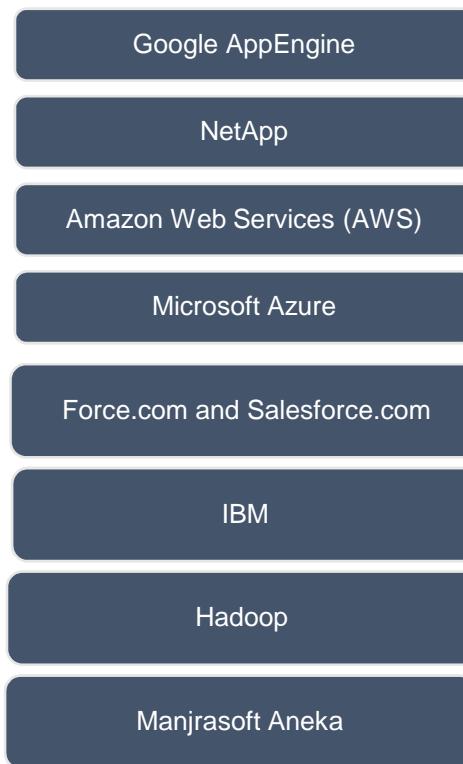


Figure 11: Different Types of Cloud Services Offered by Service Providers

Google App Engine: Often referred to as GAE or simply App Engine. GAE is a cloud-based PaaS for developing and hosting web applications in Google-managed data centers. The applications are sandboxed and run across multiple servers. GAE offers automatic scaling for web applications – as the number of requests increases for an application, App Engine automatically allocates more resources for the web application to handle the additional demand. It primarily supports Go, PHP, Java, Python, Node.js, .NET, and Ruby applications, although it can also support other languages via "custom runtimes". The service is free up to a certain level of consumed resources and only in standard environment but not in flexible environment. The fees are charged for additional storage, bandwidth, or instance hours required by the application.

GAE was first released as a preview version in April 2008 and came out of preview in September 2011. It offers:

- Write code once and deploy
- Absorb spikes in traffic
- Easily integrate with other Google services
- Scale your applications from zero to planet scale without having to manage infrastructure.
- Free up your developers with zero server management and zero configuration deployments.
- Stay agile with support for popular development languages and a range of developer tools.

Google Web Toolkit (GWT): An open-source set of tools that allows web developers to create and maintain JavaScript front-end applications in Java. Other than a few native libraries, everything is Java source that can be built on any supported platform with the included GWT Ant build files. It is licensed under the Apache License 2.0 (Figure 12).

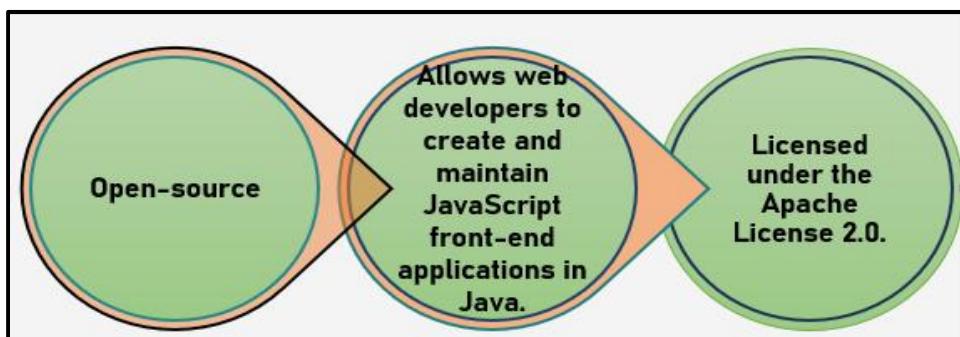


Figure 12: Characteristics of Google Web Toolkit

NetApp: NetApp, Inc. is an American hybrid cloud data services and data management company headquartered in Sunnyvale, California. It has ranked in the Fortune 500 since 2012. It was founded in 1992 with an IPO in 1995, NetApp offers cloud data services for management of applications and data both online and physically. An organization that creates storage and data management solutions for their customers. NetApp was one of the first companies in the cloud, offering data center consolidation and storage services, as well as virtualization. The products include a platform OS, storage services, storage security, software management, and protection software. NetApp competes in the computer data storage hardware industry. In 2009, NetApp ranked second in market capitalization in its industry behind EMC Corporation, now Dell EMC, and ahead of Seagate Technology, Western Digital, Brocade, Imation, and Quantum. In the total revenue of 2009, NetApp ranked behind EMC, Seagate, Western Digital, and Brocade, Xyratex, and Hutchinson Technology. According to a 2014 IDC report, NetApp ranked second in the network storage industry "Big 5's list", behind EMC (DELL), and ahead of IBM, HP and Hitachi. According to Gartner's 2018 Magic Quadrant for Solid-State Arrays, NetApp was named a leader, behind Pure Storage Systems. In 2019, Gartner named NetApp as #1 in Primary Storage.

NetApp goal is to deliver cost efficiency and accelerate business breakthroughs. NetApp products could be integrated with a variety of software products, mostly for ONTAP systems. Other provisions from NetApp include:

- Automation- NetApp provides a variety of automation services directly to its products with HTTP protocol or through middle-ware software.
- Docker- NetApp Trident software provides a persistent volume plugin for Docker containers with both orchestrators Kubernetes and Swarm and supports ONTAP, Azure NetApp Files (ANF), Cloud Volumes and NetApp Kubernetes Service in cloud.
- Backup and Recovery- Cloud Backup integrates with nearly all Backup & Recovery products for archiving capabilities since it is represented as ordinary NAS share for B&R software. The backup and recovery software from competitor vendors like IBM Spectrum Protect, EMC NetWorker, HP Data Protector, Dell vRanger, and others also have some level of integrations with NetApp storage systems.

Cloud Computing

Amazon Web Services (AWS): Amazon Web Services (AWS) is a subsidiary of Amazon providing on-demand cloud computing platforms and APIs to individuals, companies, and governments, on a metered pay-as-you-go basis. AWS provides a variety of basic abstract technical infrastructure and distributed computing building blocks and tools. It offers services on many different fronts, from storage to platform to databases. As of 2021, AWS comprises over 200 products and services including computing, storage, networking, database, analytics, application services, deployment, management, machine learning, mobile, developer tools, and tools for the Internet of Things (IoT).



Case Study: The most popular cloud services from AWS include:

- Amazon Elastic Compute Cloud (Amazon EC2): Amazon EC2 allows the users to rent virtual computers on which to run their own computer applications. EC2 encourages scalable deployment of applications by providing a web service through which a user can boot an Amazon Machine Image (AMI) to configure a virtual machine, which Amazon calls an "instance", containing any software desired. A user can create, launch, and terminate server-instances as needed, paying by the second for active servers- hence the term "elastic". EC2 provides users with control over the geographical location of instances that allows for latency optimization and high levels of redundancy. In November 2010, Amazon switched its own retail website platform to EC2 and AWS.
- Amazon SimpleDB (Simple Database Service): Amazon SimpleDB is a distributed database written in Erlang by Amazon.com. It is used as a web service in concert with Amazon Elastic Compute Cloud (EC2) and Amazon S3 and is part of Amazon Web Services. It was announced on December 13, 2007.
- Amazon Simple Storage Service (Amazon S3): Amazon S3 is a service offered by Amazon Web Services (AWS) that provides object storage through a web service interface. Amazon S3 uses the same scalable storage infrastructure that Amazon.com uses to run its global e-commerce network. Amazon S3 can be employed to store any type of object, which allows for uses like storage for Internet applications, backup and recovery, disaster recovery, data archives, data lakes for analytics, and hybrid cloud storage.
- Amazon CloudFront: Amazon CloudFront is a content delivery network (CDN) operated by Amazon Web Services. Content delivery networks provide a globally-distributed network of proxy servers that cache content, such as web videos or other bulky media, more locally to consumers, thus improving access speed for downloading the content.
- Amazon Simple Queue Service (Amazon SQS): Amazon SQS is a distributed message queuing service introduced by Amazon.com in late 2004. It supports programmatic sending of messages via web service applications as a way to communicate over the Internet. SQS is intended to provide a highly scalable hosted message queue that resolves issues arising from the common producer-consumer problem or connectivity between producer and consumer.
- Amazon Elastic Block Store (Amazon EBS): Amazon Elastic Block Store (EBS) provides raw block-level storage that can be attached to Amazon EC2 instances and is used by Amazon Relational Database Service (RDS). Amazon EBS provides a range of options for storage performance and cost. These options are divided into two major categories: SSD-backed storage for transactional workloads, such as databases and boot volumes (performance depends primarily on IOPS), and disk-backed storage for throughput intensive workloads, such as MapReduce and log processing (performance depends primarily on MB/s).

Microsoft: Microsoft offers a number of cloud services for organizations of any size (Figure 13):

- Azure Services Platform- Windows Azure/ Microsoft Azure, commonly referred to as Azure is a cloud computing service created by Microsoft for building, testing, deploying, and managing

applications and services through Microsoft-managed data centers. It provides software as a service (SaaS), platform as a service (PaaS) and infrastructure as a service (IaaS) and supports many different programming languages, tools, and frameworks, including both Microsoft-specific and third-party software and systems.

- SQL Services- Microsoft SQL Server is a relational database management system developed by Microsoft. As a database server, it is a software product with the primary function of storing and retrieving data as requested by other software applications – which may run either on the same computer or on another computer across a network. Microsoft markets at least a dozen different editions of Microsoft SQL Server, aimed at different audiences and for workloads ranging from small single-machine applications to large Internet-facing applications with many concurrent users.



Figure 13: Cloud Services Offered by Microsoft

- .NET Services- The .NET Framework is a software framework developed by Microsoft that runs primarily on Microsoft Windows. It includes a large class library called Framework Class Library and provides language interoperability across several programming languages. Programs written for .NET Framework execute in a software environment named the Common Language Runtime (CLR). The CLR is an application virtual machine that provides services such as security, memory management, and exception handling. As such, computer code written using .NET Framework is called "managed code".
- Exchange Online- Work smarter, anywhere, with hosted email for business.
- SharePoint Services- SharePoint is a web-based collaborative platform that integrates with Microsoft Office. Launched in 2001, SharePoint is primarily sold as a document management and storage system.
- Microsoft Dynamics CRM- Microsoft Dynamics is a line of enterprise resource planning and customer relationship management software applications. Microsoft Dynamics forms part of "Microsoft Business Solutions". Dynamics can be used with other Microsoft programs and services, such as SharePoint, Yammer, Office 365, Azure and Outlook. The Microsoft Dynamics focus-industries are retail, services, manufacturing, financial services, and the public sector. Microsoft Dynamics offers services for small, medium, and large businesses.

Salesforce.com: Salesforce works on the three primary areas, sales cloud, service cloud and your cloud. It has 3 primary offerings, Force.com, Salesforce.com CRM, AppExchange.

- Sales Cloud- The popular cloud computing sales application.

Cloud Computing

- Service Cloud- The platform for customer service that lets companies tap into the power of customer conversations no matter where they take place.
- Your Cloud- Powerful capabilities to develop custom applications on its cloud computing platform.

IBM: IBM offers cloud computing services to help businesses of all sizes take advantage of this increasingly attractive computing model. IBM is applying its industry-specific consulting expertise and established technology record to offer secure services to companies in public, private, and hybrid cloud models. Some of their services include:

- Industry-specific business consulting services for cloud computing
- Technology consulting, design, and implementation services
- Cloud security

Hadoop: Apache Hadoop is a collection of open-source software utilities that facilitates using a network of many computers to solve problems involving massive amounts of data and computation. It provides a software framework for distributed storage and processing of big data using the MapReduce programming model. All the modules in Hadoop are designed with a fundamental assumption that hardware failures are common occurrences and should be automatically handled by the framework. Apache Hadoop software allows for the distributed storage and processing of large datasets across clusters of computers using simple programming models. Hadoop is designed to scale up from a single computer to thousands of clustered computers, with each machine offering local computation and storage. In this way, Hadoop can efficiently store and process large datasets ranging in size from gigabytes to petabytes of data.

Aneka: ANEKA cloud platform is a software platform and a framework for developing distributed applications on the cloud. It harnesses the computing resources of a heterogeneous network of workstations and servers or data centers on demand. Aneka provides developers with a rich set of APIs for transparently exploiting these resources. The system administrators can leverage on a collection of tools to monitor and control the deployed infrastructure. This can be a public cloud available to anyone through the Internet, or a private cloud constituted by a set of nodes with restricted access. All in all, Aneka-based computing cloud is a collection of physical and virtualized resources connected through a network, which are either the Internet or a private intranet.

3.4 Criteria for Choosing the Best Cloud Service Provider

The first step in switching to cloud computing is determining what kind of cloud services you could be interested in. Then, choosing a cloud computing service is a long-term investment. Your application will heavily rely on third-party capacities, and you need to make sure that the provider is legitimate and fits your needs. Also, check for the latest cloud service provider trends (Figure 14).

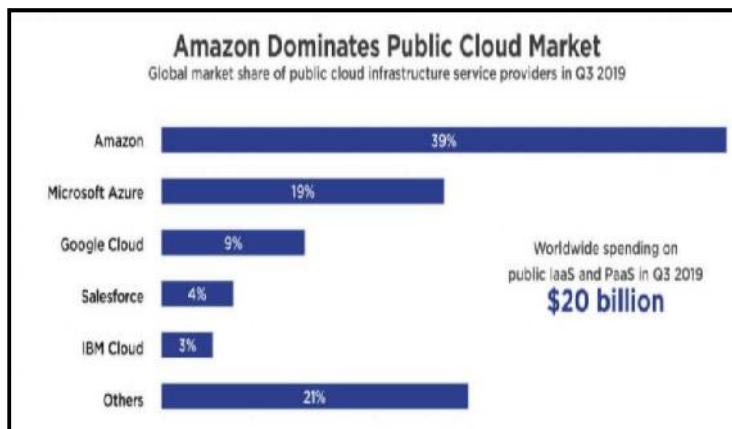


Figure 14: Global Cloud Market Trends

Here's a short criteria checklist on choosing a trustworthy cloud provider:

- Financial Stability: Your cloud provider should be well-financed and receive steady profits from the infrastructure. If the company shuts down due to monetary issues, your solutions will be in jeopardy, too. In the worst-case scenario, you will have to cease the support of your solutions, or, in a better case, migrate to a new provider, which is an expensive and time-consuming process.
- Industries that Prefer the Solution: Before committing to a cloud services company, take a look at its existing clients and examine their markets. Ideally, the provider should be popular among companies in your niche, or at least in the neighbouring ones. Another road to take is asking competitors and partners about their favourite choices.
- Datacenter Locations: To avoid safety risks, make sure that cloud providers can enable geographical distribution for your data. Ideally, you want to locate your data on servers in Asia, Europe, America, without betting on a single region. Also, pay attention to countries – some, like Japan or Germany, are known to be more secure, whereas Russia, for instance, is not the safest option.
- Security Programs: Take a look at the security programs of your favourite cloud providers. The majority of companies have dedicated papers and e-books that discuss this matter in detail – take your time to go through them. Start with taking a look at security documentation of the top cloud providers- AWS, G Suite, Microsoft Azure, Salesforce. You can use these pages as references during your safety research.
- Encryption Standards: Make sure the cloud provider specifies the use of encryption. The provider should encrypt the data both when it's being transferred to the cloud and during the storage itself. No matter what is the stage of data storage, the information should be secured end-to-end, so there is no way even for developers of the service to access the file contents.
- Check Accreditation and Auditing: The most common online auditing standard is SSAE – the procedure that verifies that the online service checked the safety of its data-storing practices. ISO 27001 certificate verified that a cloud provider complies with international safety standards for data storage.
- Look for solutions that offer Free Cloud Backup: OneDrive, Google Drive, Dropbox, and Box offer free space to create cloud backup copies, both manually and automatically.

Cloud Service Providers Prospective Trends

Cloud services have been among the most popular platforms lately, with companies like Microsoft, Amazon, Google leading the way for technology growth. Instead of relying on their servers, companies prefer outsourcing their storage to trusted providers, passing over the responsibility for supporting the infrastructure and assuring security.

Let's take a brief look at the statistics of the cloud computing market to see its latest trends:

- In 2020, the market is expected to demonstrate the growth rate of 17%;
- In 2019, Cloud infrastructure accounted for about 3% of overall IT infrastructure;
- IP traffic in Cloud services is expected to grow up to 19.5 zettabytes in 2021.
- The trust of the number of businesses in cloud service is rapidly increasing.

3.5 Other Cloud Services

Cloud computing offers large number of services (Figure 15). The cloud service offerings evolve almost daily, as the technology migrates from the traditional on-premise model to the new cloud model:

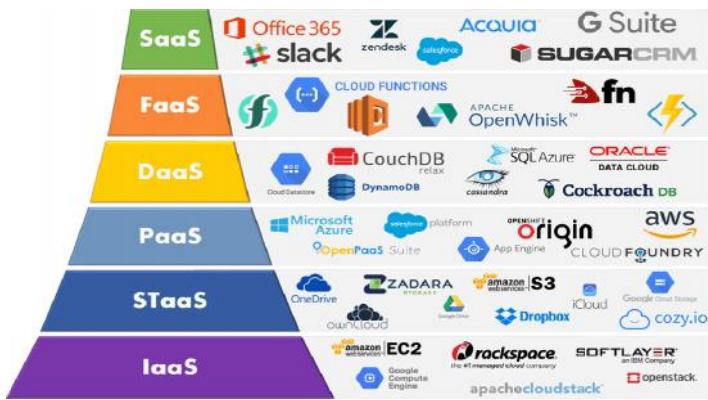
Cloud Computing

Figure 15: Other Cloud Services

Storage-as-a-Service (STaaS)

- Cloud service model in which a company leases or rents its storage infrastructure to another company or individuals to store either files or objects.
- Economy of scale in the service provider's infrastructure theoretically allows them to provide storage much more cost-effectively than most individuals or corporations can provide their own storage when the total cost of ownership is considered.
- STaaS is generally seen as a good alternative for a small or mid-sized business that lacks the capital budget and/or technical personnel to implement and maintain their own storage infrastructure.
- Small companies and individuals often find this to be a convenient methodology for managing backups, and providing cost savings in personnel, hardware and physical space.
- STaaS operates through a web-based API that is remotely implemented through its interaction with the client application's in-house cloud storage infrastructure for input/output (I/O) and read/write (R/W) operations.
- If the company ever loses its local data, the network administrator could contact the STaaS provider and request a copy of the data.
- For an end-user-level cloud storage, Dropbox, Google Drive, Apple's iCloud and Microsoft OneDrive are among the leading end-user-level cloud storage providers.
- For enterprise-level cloud storage, Amazon S3, Zadara, IBM's SoftLayer and Google Cloud Storage are some of the more popular providers.

Data-as-a-Service (DaaS)

- In the DaaS computing model (a more advanced, fine-grained form of STaaS), data (as opposed to files) is readily accessible through a Cloud-based platform.
- Data (either from databases or object containers) is supplied "on-demand" via cloud platforms (as opposed to the traditional, on-premise models in which the data remains in the customer's hands) and the vendor provides the tools that make it easier to access and explore.
- Based on Web Services standards and Service-oriented Architecture (SOA), DaaS provides a dynamic infrastructure for delivering information on demand to users, regardless of their geographical location or organizational separation- and, in the process, presents solution providers with a number of significant opportunities.
- DaaS eliminates redundancy and reduces associated expenditures by accommodating vital data in a single location, allowing data use and/or modification by multiple users via a single update point.

- Typical business applications include customer relationship management (CRM), enterprise resource planning (ERP), e-commerce and supply chain systems and, more recently, Big Data analytics.
- Some of the best-known enterprise-level DaaS providers are Oracle's Data Cloud, Amazon DynamoDB, Microsoft SQL Database (formerly known as SQL Azure) and Google Cloud's Datastore.
- For an open-source projects, Apache Cassandra, CockroachDB or CouchDB will almost certainly catch your eye.

Communication-as-a-Service (CaaS)

Communications as a Service (CaaS) is an outsourced enterprise communications solution that can be leased from a single vendor (Figure 16). CaaS vendor is responsible for all hardware and software management and offers guaranteed Quality of Service (QoS). CaaS allows businesses to selectively deploy communications devices and modes on a pay-as-you-go, as-needed basis. Such communications can include: Voice over IP (VoIP or Internet telephony); Instant messaging (IM); and Collaboration and Video conference applications using the fixed and mobile devices.

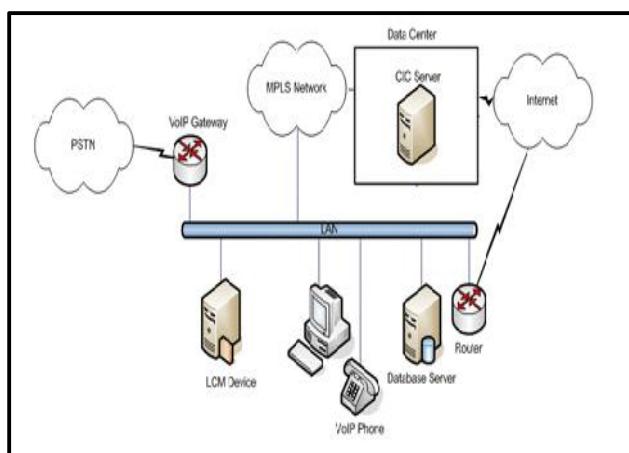


Figure 16: Communication-as-a-Service Scenario

Advantages of CaaS

- Hosted and Managed Solutions
- Fully Integrated, Enterprise - Class Unified Communications
- No Capital Expenses Needed
- Flexible Capacity and Feature Set
- No Risk of Obsolescence
- No Facilities and Engineering Costs Incurred
- Guaranteed Business Continuity

Monitoring-as-a-Service (MaaS)

A concept that combines the benefits of cloud computing technology and traditional on-premise IT infrastructure monitoring solutions (Figure 17). MaaS is a new delivery model that is suited for organizations looking to adopt a monitoring framework quickly with minimal investments. MaaS is a framework that facilitates the deployment of monitoring functionalities for various other services and applications within the cloud. The most common application for MaaS is online state monitoring, which continuously tracks certain states of applications, networks, systems, instances or any element that may be deployable within the cloud. MaaS makes it easier for users to deploy state monitoring at different levels of cloud services.

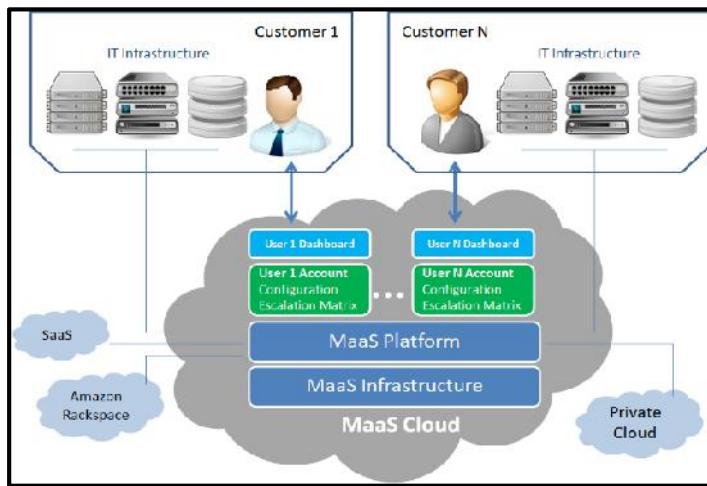


Figure 17: Monitoring-as-a-Service Scenario

Advantages of MaaS

- Ready to Use Monitoring Tool Login: The vendor takes care of setting up the hardware infrastructure, monitoring tool, configuration and alert settings on behalf of the customer. The customer gets a ready to use login to the monitoring dashboard that is accessible using an internet browser. A mobile client is also available for the MaaS dashboard for IT administrators.
- Inherently Available 24x7x365: Since MaaS is deployed in the cloud, the monitoring dashboard itself is available 24x7x365 that can be accessed anytime from anywhere. There are no downtimes associated with the monitoring tool.
- Easy Integration with Business Processes: MaaS can generate alert based on specific business conditions. MaaS also supports multiple levels of escalation so that different user groups can get different levels of alerts.
- Cloud Aware and Cloud Ready: Since MaaS is already in the cloud, MaaS works well with other cloud-based products such as PaaS and SaaS. MaaS can monitor Amazon and Rackspace cloud infrastructure. MaaS can monitor any private cloud deployments that a customer might have.
- Zero Maintenance Overheads: As a MaaS, customer, you don't need to invest in a network operations centre. Neither do you need to invest an in-house team of qualified IT engineers to run the monitoring desk since the MaaS vendor is doing that on behalf of the customer.

Assets Monitored by MaaS

- Servers and Systems Monitoring: Server Monitoring provides insights into the reliability of the server hardware such as Uptime, CPU, Memory and Storage. Server monitoring is an essential tool in determining functional and performance failures in the infrastructure assets.
- Database Monitoring: Database monitoring on a proactive basis is necessary to ensure that databases are available for supporting business processes and functions. Database monitoring also provides performance analysis and trends which in turn can be used for fine tuning the database architecture and queries, thereby optimizing the database for your business requirements.
- Network Monitoring: Network availability and network performance are two critical parameters that determine the successful utilization of any network – be it a LAN, MAN or WAN network. Disruptions in the network affect business productivity adversely and can bring regular operations to a standstill. Network monitoring provides pro-active information about network performance bottlenecks and source of network disruption.

- Storage Monitoring: A reliable storage solution in your network ensures anytime availability of business-critical data. Storage monitoring for SAN, NAS and RAID storage devices ensures that your storage solution are performing at the highest levels. Storage monitoring reduces downtime of storage devices and hence improves availability of business data.
- Applications Monitoring: Applications Monitoring provides insight into resource usage, application availability and critical process usage for different Windows, Linux and other open-source operating systems-based applications. Applications Monitoring is essential for mission critical applications that cannot afford to have even a few minutes of downtime. With Application Monitoring, you can prevent application failures before they occur and ensure smooth operations.
- Cloud Monitoring: Cloud Monitoring for any cloud infrastructure such as Amazon or Rackspace gives information about resource utilization and performance in the cloud. While cloud infrastructure is expected to have higher reliability than on-premise infrastructure, quite often resource utilization and performance metrics are not well understood in the cloud. Cloud monitoring provides insight into exact resource usage and performance metrics that can be used for optimizing the cloud infrastructure.
- Virtual Infrastructure Monitoring: Virtual Infrastructure based on common hypervisors such as ESX, Xen or Hyper-V provides flexibility to the infrastructure deployment and provides increased reliability against hardware failures. Monitoring virtual machines and related infrastructure gives information around resource usage such as memory, processor and storage.

Database-as-a-Service (DBaaS)

Database as a Service (DBaaS) is an architectural and operational approach enabling DBAs to deliver database functionality as a service to internal and/or external customers. DBaaS architectures support following required capabilities: customer side provisioning and management of database instances using on-demand, self-service mechanisms; automation of monitoring with provider-defined service definitions, attributes and quality SLAs; and fine-grained metering of database usage enabling show-back reporting or charge-back for both internal and external functionality for each individual consumer.



Notes: Why DBaaS?

DBaaS standardizes and optimizes the platform requirements which eliminates the need to deploy, manage and support dedicated database hardware and software for each project's multiple development, testing, production, and failover environments. DBaaS architectures are inherently designed for elasticity and resource pooling. DBaaS providers deliver production and non-production database services that support average daily workload requirements & are not impacted by: resource limitations, time-sensitive projects; and hardware limitations and budgets.

Setting up DBaaS

In order to set-up DBaaS, a cloud administrator will need to:

- Define roles and users in the self-service portal.
- Installing the agent to manage all “unmanaged hosts” so self-discovery of any DBaaS environments that are created.
- Set quotas, privileges.
- Software library will allow automation.
- Provisioning will set who will be granted and how much allocated to each customer, administrator and/or business unit.

Network-as-a-Service (NaaS)

In NaaS, the users who do not want to use their own networks take help from service providers to host the network infrastructure. The connectivity and bandwidth are provided by the service provider for the contracted period. NaaS represents the network as transport connectivity. The network virtualization is done in this service.

NaaS is “an emerging procurement model to consume network infrastructure via a flexible operating expense (OpEx) subscription inclusive of hardware, software, management tools, licenses, and lifecycle services.”

What's Driving the Trend Toward NaaS?

Traditional network model requires capital expenses (CapEx) for physical networks with switches, routers, and licensing. The do-it-yourself IT model requires time for planning and deployment as well as expertise to install and configure infrastructure and to ensure security access policies are in place. This model involves the following:

- Diligent monitoring for updates and security patches is essential due to rapid changes in technology and security threats.
- Provisioning a new service is a manual process that requires a technician to deploy and configure equipment at various locations.
- Service provisioning and issue resolution have historically been lengthy processes.
- As networks have grown in complexity—with more mobile users connecting from everywhere and with the expansion to cloud—IT teams have been challenged to keep pace.

NaaS Service Models

- Connectivity Cloud: A model in which a private fiber fabric or wireline "Middle Mile" network is used to bypass often less-optimal public (internet) routing and congestion to provide connectivity for critical Enterprise resource and services access. It is controlled via a distributed software platform, the model supports "cloud-aligned" elastic consumption including on-demand provisioning, any-to-any connectivity, and flexible bandwidth deployment through both portal and programmable API operation and introspection. By integrating the platform API with provisioning and application deployment playbooks, the resulting WAN can realize an infrastructure as code paradigm for Wide Area Networks - "network-as-code". They resulting services include custom WAN interconnectivity, hybrid cloud and multi-cloud connectivity.
- Virtual Private Network (VPN): A tunnel overlay that extends a private network and the resources contained in the network across networks like the public Internet. It enables a host computer to send and receive data across shared or public networks as if it were a private network with the functionality and policies of the private network.
- Virtual Network Operation: Model common in mobile networks in which a telecommunications manufacturer or independent network operator builds and operates a network (wireless, or transport connectivity) and sells its communication access capabilities to third parties (commonly mobile phone operators) charging by capacity utilization. A mobile virtual network operator (MVNO), is a mobile communications services provider that does not own the radio spectrum or wireless network infrastructure over which it provides services. Commonly a MVNO offers its communication services using the network infrastructure of an established mobile network operator.

Benefits of NaaS

NaaS is a cloud model that enables users to easily operate the network and achieve the outcomes they expect from it without owning, building, or maintaining their own infrastructure. NaaS can

replace hardware-centric VPNs, load balancers, firewall appliances, and Multiprotocol Label Switching (MPLS) connections. The users can scale up and down as demand changes, rapidly deploy services, and eliminate hardware costs. NaaS offers ROI (return on investment), enabling customers to trade CapEx for OpEx and refocus person hours on other priorities. Figure 18 depicts the different benefits of NaaS.

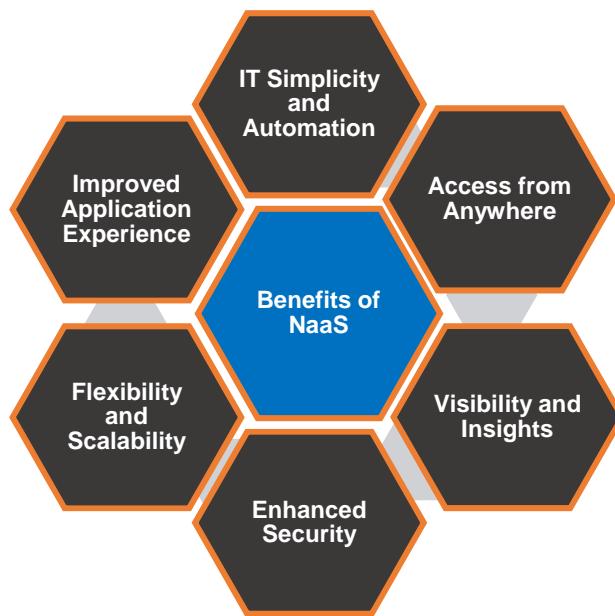


Figure 18: Benefits of NaaS

- **IT Simplicity and Automation-** Businesses benefit when they align their costs with actual usage. They don't need to pay for surplus capacity that goes unused, and they can dynamically add capacity as demands increase. Businesses that own their own infrastructure must implement upgrades, bug fixes, and security patches in a timely manner. Often, IT staff may have to travel to various locations to implement changes. NaaS enables the continuous delivery of new fixes, features and capabilities. It automates multiple processes such as onboarding new users and provides orchestration and optimization for maximum performance. This can help to eliminate the time and money spent on these processes.
- **Access from Anywhere-** Today's workers may require access to the network from anywhere—home or office—on any devices and without relying on VPNs. NaaS can provide enterprises with global coverage, low-latency connectivity enabled by a worldwide POP backbone, and negligible packet loss when connecting to SaaS applications, platform-as-a-service (PaaS)/infrastructure-as-a-service (IaaS) platforms, or branch offices
- **Visibility and Insights-** NaaS provides proactive network monitoring, security policy enforcement, advanced firewall and packet inspection capabilities, and modeling of the performance of applications and the underlying infrastructure over time. Customers may also have an option to co-manage the NaaS.
- **Enhanced Security-** NaaS results in tighter integration between the network and the network security. Some vendors may "piece together" network security. By contrast, NaaS solutions need to provide on-premise and cloud-based security to meet today's business needs.
- **Flexibility-** NaaS services are delivered through a cloud model to offer greater flexibility and customization than conventional infrastructure. Changes are implemented through software, not hardware. This is typically provided through a self-service model. IT teams can, for example, reconfigure their corporate networks on demand and add new branch locations in a fraction of the time. NaaS often provides term-based subscription with usage billing and multiple payment options to support various consumption requirements.
- **Scalability-** NaaS is inherently more scalable than traditional, hardware-based networks. NaaS customers simply purchase more capacity instead of purchasing, deploying, configuring, and securing additional hardware. This means they can scale up or down quickly as needs change.

Cloud Computing

- Improved Application Experience- NaaS provides AI-driven capabilities to help ensure SLAs and SLOs for capacity are met or exceeded. NaaS provides the ability to route application traffic to help ensure outstanding user experience and to proactively address issues that occur.

Healthcare-as-a-Service (HaaS)

Gone are days when healthcare organizations used to store patient data in piles of papers and files. Not only was that inconvenient and time-consuming, but also expensive in terms of both money and resources. With the exponential growth in technology, more and more healthcare businesses are moving to the cloud. Cloud computing has impacted the essential divisions of society, especially the healthcare industry.

Technology-enabled Healthcare includes telehealth, telecare, telemedicine, tele-coaching, mHealth and self-care services that can put people in control of their own health, wellbeing and support, keeping them safe, well and independent and offering them and their family's peace of mind.

Cloud computing in healthcare market can be segmented as:

A. Global Cloud Computing in Healthcare Market, By Type

- Clinical Information Systems
- Non-Clinical Information Systems

B. Global Cloud Computing in Healthcare Market, By Pricing Model

- Pay-As-You-Go
- Spot Pricing Model

C. Global Cloud Computing in Healthcare Market, By Service Model

- Software-as-a-Service
- Infrastructure-as-a-Service
- Platform-as-a-Service

Benefits of Cloud-based HaaS: In the healthcare industry, efficiency is essential. The doctors, drug manufacturers, and illness prevention departments should always stay on schedule or go ahead in time if possible. Why? Since many lives and the public's health are at stake, every second counts and can potentially change things. While local computing works, it's not as efficient as the newer cloud standard.

- Faster Services: With cloud, the healthcare companies can process data and transfer information instantly. The workers can complete their tasks remotely, removing the need to wait for slow servers or travel to a specific office to access files or information. This is all possible thanks to the cloud's powerful processing, faster servers, and ease of access.
- Improved Collaboration: Collaboration is also a vital part of every healthcare company. In hospitals, doctors, nurses, and other staff need to accurately transfer and receive information like patients' data and other related things. However, local servers and programs are slow to process data, get easily congested with multiple user access, or don't have effective remote and collaboration functions.
- Supports Both Consumer and Specialized Programs: Cloud computing supports both consumer and specialized programs. A healthcare company can choose to deploy its applications to a cloud provider or subscribe to a commercial service. Accordingly, healthcare workers can now send, upload, and access files in varying formats and sizes remotely, converse via chat, call, or video feed, and use more collaboration methods.
- Cost-Efficient Operations: The organizations can save money when choosing cloud computing over local processing. The need to hire people who can maintain and solve issues of local servers is less as cloud providers ensure their platforms are always up and running. The cost of repairs and maintenance most of the time outweighs the expense of paying for cloud subscriptions monthly. Now, it is more practical to get a cloud platform package than build an entire server system in an organization. Most favour the faster cloud deployment than waiting for weeks or months to complete and potentially troubleshoot a local system. Cloud computing is excellent at

Unit 03: Cloud Services

scaling, a healthcare company will only need to buy the space or computing power it requires. Additionally, Cloud offers easy downgrade plans with cost reductions.

- Enhanced Patient Care Efficiency: Cloud computing can enhance patient care in many ways. When accepting new patients, doctors and staff can quickly check the online database to check a person's potential existing medical records. As a result, they can spend more time on actual consultations and not on the paperwork. Next, a hospital can efficiently distribute patient information like condition, status, schedules, and medication to nurses and doctors. The workers can also avoid information inaccuracy when data mixes with other groups or new entries overlap existing ones.
- Better Data Management: If a healthcare organization uses local computing, chances are, it only has limited methods of storing and accessing data. The offline databases are more constrained, they can restrict users to access faster or do things they can typically perform on commercial programs. Cloud providers employ powerful technologies, it's now possible to store more complex data and file types without worrying about slow down or errors. Additionally, organizing massive data collections won't be as hard compared to doing it on a local server. Above all, healthcare workers can instantly upload/access information and files remotely without messing up the system or waiting for queues and slow loading times.
- Improved Privacy: Unlike offline processing, cloud computing provides more privacy and security for the healthcare sector. Cloud platforms offer high-level encryption, multiple stacks of protection, and superior threat detection methods, it's harder for criminals to infiltrate a healthcare company's system.

Education as a Service (EaaS)

EaaS involves providing learning and development to learners/teachers through various electronic media such as the internet, audio, video, etc. EaaS learning is education done through Internet. It is often called "e-learning". EaaS model provides learners with another or enhancement to more expensive four-year degree programs and delivers customized learning opportunities for learners. It facilitates academic organizations and businesses to utilize courses that align with their offering without having to get excessive learning materials that are not being executed in its program. Its offerings range from intangible benefits (Increment in knowledge, aptitude, professional expertise, skill); produced with the help of a set of tangible (infrastructure); and Intangible (faculty expertise and learning) aids.

Need for Education-as-a-Service (EaaS)

- Affordable computers, faster Broadband connectivity, and rich educational content has made it possible for ICT to revolutionise education.
- These advances allow training providers to improve the delivery of their content, and there's a fundamental need for this change.
- The current educational paradigm sees students paying large sums of money for courses that take several years to complete.
- Graduates are left with considerable amounts of debt and the rising costs make it even more difficult for mature students to return to education.
- Returns on Investment (ROI) with regard to tuition fees is a key consideration for many learners entering any type of training programme, and the current model is not always providing value for money.
- Education-as-a-Service (EaaS) affords students the opportunity to gain accreditation by choosing modules that are relevant to their career goals and only paying for what they need

Benefits of EaaS

- Learners pay for the Education They Want/ Need: Mostly, the degrees and training courses can be costly and often require the students to follow a dictated set of modules. EaaS gives students the option to pick and choose the modules they want to purchase according to their needs. The training structure is tailored to the student, by the student, and their time and money are not wasted on irrelevant learning.
- Advocates Flexible Learning: The flexibility of EaaS allows students to learn at a time, place, and pace that they choose. This method of learning often goes hand-in-hand with blended learning, as both models give the student control and responsibility of their own learning.

Cloud Computing

- Learner-centric: The traditional education sees the teacher making all decisions regarding curriculum. They impose the place, time and pace of content delivery. This old-fashioned approach forces students to take a passive role in their education. In contrast, learner-centric courses encourage students to be active in designing and executing their own educational journey. This manner of learning is supported by constructivist theory - the idea that humans generate knowledge and meaning from interactions between their experiences and ideas. This theory is key to corporate learning practices and is commonly used to inform adult educational programmes. By getting students to use previous experiences and existing knowledge in their learning, a deeper understanding of the content can be achieved.
- Encourages Agile Content Development by Course Designers: Almost continuous access to the Internet and the popularity of social media has taught many of us to expect dynamic content - content that is constantly being changed or updated, based on new information as it is made available. As such, developing content for an online or blended learning platform requires new ideas and the continual update of learning materials.

Function-as-a-Service (FaaS)

FaaS is a concept of serverless computing via serverless architectures where developers can leverage this to deploy an individual "function", action, or piece of business logic.

Principles of FaaS:

- Complete abstraction of servers away from the developer.
- Billing based on consumption and executions, not server instance sizes.
- Services that are event-driven and instantaneously scalable.
- FaaS services provide a platform allowing customers to develop, run, and manage application functionalities without the complexity of building and maintaining the infrastructure typically associated with developing and launching a complete application.



Example:

- AWS Lambda: The service allows accessing software code without server setting and management. Developers need only to upload the code, and the solution will automatically connect the app to servers, language runtimes, OS, and highlight the functional code fragments. From that point, developers only choose features for editing.
- Azure Functions: The platform uses trigger mechanisms to highlight functions. Developers can set events that will lead to changes in code - for instance, a particular user input (interaction with an app or provided data) can turn on a function (like showing a pop-up or opening a page). The developers set up these triggers and responses without building the software infrastructure.
- IBM Open Whisk: Similar to Lambda and Azure, IBM Open Whisk reacts to trigger effects and produces a series of organized outputs. Developers only have to set up action sequences and describe possible trigger events. The action itself will be enabled by IBM's infrastructure - the users don't have to control these aspects.

Summary

- Cloud computing signifies a major change in the way we run various applications and store our information. Everything is hosted in the "cloud", a vague assemblage of computers and servers accessed via the Internet, instead of the method of running programs and data on a single desktop computer.
- Technology-enabled Healthcare includes telehealth, telecare, telemedicine, tele-coaching, mHealth and self-care services that can put people in control of their own health, wellbeing and support, keeping them safe, well and independent and offering them and their family's peace of mind.

Unit 03: Cloud Services

- The first step in switching to cloud computing is determining what kind of cloud services you could be interested in. Then, choosing a cloud computing service is a long-term investment. Your application will heavily rely on third-party capacities, and you need to make sure that the provider is legitimate and fits your needs.
- IBM offers cloud computing services to help businesses of all sizes take advantage of this increasingly attractive computing model. IBM is applying its industry-specific consulting expertise and established technology record to offer secure services to companies in public, private, and hybrid cloud models.
- NaaS is “an emerging procurement model to consume network infrastructure via a flexible operating expense (OpEx) subscription inclusive of hardware, software, management tools, licenses, and lifecycle services”.
- SaaS describes any cloud service where consumers are able to access software applications over the internet. The applications are hosted in “the cloud” and can be used for a wide range of tasks for both individuals and organizations.

Keywords

Amazon Simple Storage Service (Amazon S3): Amazon S3 is a service offered by Amazon Web Services (AWS) that provides object storage through a web service interface. Amazon S3 uses the same scalable storage infrastructure that Amazon.com uses to run its global e-commerce network.

Function-as-a-Service: FaaS is a concept of serverless computing via serverless architectures where developers can leverage this to deploy an individual “function”, action, or piece of business logic.

Database as a Service: Database as a Service (DBaaS) is an architectural and operational approach enabling DBAs to deliver database functionality as a service to internal and/or external customers.

Cloud Hosting: The hosting of websites on virtual servers which are founded upon pooled resources from underlying physical servers.

Virtual Data Centers (VDC): A virtualized network of interconnected virtual servers which can be used to offer enhanced cloud hosting capabilities, enterprise IT infrastructure or to integrate all of these operations within either a private or public cloud implementation.

Communications as a Service: Communications as a Service (CaaS) is an outsourced enterprise communications solution that can be leased from a single vendor. CaaS vendor is responsible for all hardware and software management and offers guaranteed Quality of Service (QoS).

Self Assessment

1. As per a survey, IP traffic in cloud services is expected to grow up to _____ in 2021.
 - A. 19.5 zettabytes
 - B. 20.5 zettabytes
 - C. 1000 Exabytes
 - D. 20.5 Exabytes
2. STaaS stands for
 - A. Software Technology as a Service
 - B. Storage-as-a-Service
 - C. Software Team as a Service
 - D. Structure as a Service
3. Which of the following comes under monitoring components under Monitoring as a Service (MaaS)?
 - A. Virtual Infrastructure Monitoring
 - B. Storage Monitoring

Cloud Computing

- C. Applications Monitoring
 - D. All of the above
4. _____ is a concept of serverless computing via serverless architectures where developers can leverage this to deploy an individual “function”, action, or piece of business logic.
- A. Serverless as a Service
 - B. Business as a Service
 - C. Function as a Service
 - D. None of the above
5. _____ provides IT solutions and/or services to end users and organizations in cloud environment.
- A. Service vendor
 - B. Service organizer
 - C. Service broker
 - D. Service auditor
6. Service provider that provides computer storage space and related management services is termed as:
- A. Storage hosting service provider
 - B. SaaS provider
 - C. PaaS provider
 - D. Storage service provider
7. Google App Engine (GAE) is a cloud-based _____ for developing and hosting web applications in Google-managed data centers.
- A. SaaS
 - B. PaaS
 - C. FaaS
 - D. MaaS
8. Which of the following is not a cloud service offered by Amazon Web Services?
- A. Elastic Compute Cloud
 - B. App Engine
 - C. SimpleDB
 - D. CloudFront
9. _____ is a construct that allows to access application that actually resides at a remote location of other internet connected device, most often, this will be a distant data center.
- A. Cloud computing
 - B. Batch computing
 - C. Data computing
 - D. Construction computing
10. IaaS is also known as _____.
- A. Hoard as a Service

- B. Hardware as a Service
 C. Infra as a Service
 D. Physical as a Service
11. _____ model consists of the particular types of services that you can access on a cloud computing platform.
 A. Application
 B. Service
 C. Virtualization
 D. Deployment
12. Cloud providers provide cloud services to the cloud users.
 A. True
 B. False
13. Which of the following cloud service model provides virtual platforms and tools to create, test, and deploy applications?
 A. IaaS
 B. SaaS
 C. PaaS
 D. MaaS
14. SaaS supports multiple users and provides a shared data model through _____ model.
 A. single-tenancy
 B. multiple-instance
 C. multi-tenancy
 D. all of the mentioned
15. Which of the following are offerings from Infrastructure as a Service?
 A. Virtual Server Space
 B. Network Connections
 C. Load Balancers
 D. All of the above

Answers for Self Assessment

- | | | | | |
|-------|-------|-------|-------|-------|
| 1. A | 2. B | 3. D | 4. C | 5. A |
| 6. D | 7. B | 8. A | 9. A | 10. B |
| 11. B | 12. A | 13. C | 14. C | 15. D |

Review Questions

1. Write a short note on Healthcare-as-a-Service and Education-as-a-Service?
2. Compare classic and hybrid cloud service models?
3. Elaborate Communication-as-a-Service?

Cloud Computing

4. How Database-as-a-Service works?
5. What different cloud services are offered by Amazon Web Services (AWS)?
6. Discuss in detail cloud classic service models?
7. Who are service providers? Classify.
8. Discuss about different cloud service providers and their services?
9. What are the different benefits of Network-as-a-Service?
10. List the different service providers for cloud computing services? Explain each in detail.



Further Readings

Cloud Computing: Concepts, Technology and Architecture by Erl, Pearson Education.

Cloud Computing Black Book by Kailash Jayaswal, Jagannath Kallakurchi, Donald J. Houde, Deven Shah, Kogent Learning Solutions, DreamTech Press



Web Link

[Cloud Service Models SaaS, IaaS, PaaS - Choose the Right One for Your Business - Fingent Technology](#)

[Cloud Service Models - javatpoint](#)

[Infrastructure as a service - Wikipedia](#)

[Monitoring as a Service - simple, flexible, comprehensive \(paessler.com\)](#)

[What is Database-as-a-Service \(DBaaS\)? | Stratoscale](#)

[What is DBaaS \(Database-as-a-Service\) | IBM](#)

[What is NaaS \(network-as-a-service\)? | Cloudflare](#)

[What is Communications as a Service \(CaaS\)? - Definition from Techopedia](#)

[How Healthcare-As-a-Service Can Improve the Health System \(businessinsider.com\)](#)

Unit 04: Introduction to Big Data

CONTENTS

Objectives
Introduction
4.1 Facts and Figures Pertaining to Big Data
4.2 Traditional Database Methods vs Big Data
4.3 Benefits of Big Data Processing
4.4 Types of Big Data
4.5 Challenges in Handling Big Data
4.6 Big Data Analytics
4.7 Big Data Analytics Cycle
4.8 Big Data Analytics- Data Analysis Tools
4.9 Parallel Data Processing
4.10 Distributed Data Processing and Cluster Processing
4.11 Processing Workloads in Big Data
4.12 Hadoop Framework
4.13 Hadoop Modules
4.14 Hadoop MapReduce
Summary
Keywords
Self Assessment
Answers for Self Assessment
Review Questions
Further Readings

Objectives

After this lecture, you will be able to,

- Learn about Big Data and its characteristics
- Understand the different V's of Big Data.
- Discover Big Data Analytics and its role.
- Explore Hadoop Framework and its modules.
- Discuss MapReduce, its phases and algorithm.

Introduction

Big data is a field that treats ways to analyse, systematically extract information from, or otherwise deal with data sets that are too large or complex to be dealt with by traditional data-processing application software. Data with many fields (columns) offer greater statistical power, while data with higher complexity (more attributes or columns) may lead to a higher false discovery rate. Big data analysis challenges include capturing data, data storage, data analysis, search, sharing, transfer, visualization, querying, updating, information privacy, and data source. The analysis of big data presents challenges in sampling, and thus previously allowing for only observations and

Cloud Computing

sampling. Therefore, big data often includes data with sizes that exceed the capacity of traditional software to process within an acceptable time and value.

Big data usually includes data sets with sizes beyond the ability of commonly used software tools to capture, curate, manage, and process data within a tolerable elapsed time. Big data philosophy encompasses unstructured, semi-structured and structured data. However, the main focus is on unstructured data. Big data "size" is a constantly moving target; as of 2012 ranging from a few dozen terabytes to many zettabytes of data. Big data requires a set of techniques and technologies with new forms of integration to reveal insights from datasets that are diverse, complex, and of a massive scale.

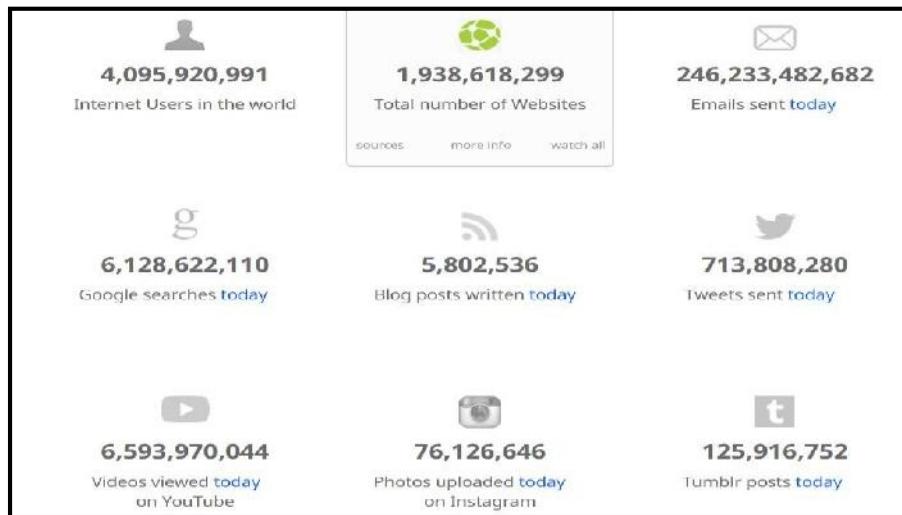
Big Data is a collection of data that is huge in volume yet growing exponentially with time. It is a data with so large size and complexity that none of traditional data management tools can store it or process it efficiently. Big data is also a data but with huge size.

Big data manages the collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications. Data whose scale, diversity, and complexity require new architecture, techniques, algorithms, and analytics to manage it and extract value and hidden knowledge from it. There can be several correlations within data such as spotting business trends, determining quality of research, preventing diseases, linking legal citations, combating crime, and determining real-time roadway traffic conditions.

Big data forms an important aspect of data science. Data science is the study of data analyzing by advance technology (Machine Learning, Artificial Intelligence, Big data). It processes a huge amount of structured, semi-structured, unstructured data to extract insight meaning, from which one pattern can be designed that will be useful to take a decision for grabbing the new business opportunity, the betterment of product/service, ultimately business growth.

4.1 Facts and Figures Pertaining to Big Data

Big data is becoming extensively important owing to the large amounts of data generated. Walmart handles 1 million customer transactions/hour. Facebook handles 40 billion photos from its user base and inserts 500 terabytes of new data every day. Facebook stores, accesses, and analyses 30+ petabytes of generated data. Even a flight generates 240 terabytes of flight data in 6-8 hours of flight. Not only this, more than 5 billion people are calling, texting, tweeting & browsing on mobile phones worldwide. The decoding of human genome originally took 10 years to process; now it can be achieved in one week. The largest AT&T database boasts titles including largest volume of data in one unique database (312 terabytes) and the second largest number of rows in a unique database (1.9 trillion), which comprises AT&T's extensive calling records.



From Where Does So Much Data Come Up

There are several sources of big data such as people, machine, organization, ubiquitous computing. Currently, there are more people carrying data-generating devices (such as mobile phones running Facebook, GPS, Cameras etc.).

4.2 Traditional Database Methods vs Big Data

There are certain drawbacks associated with traditional RDBMS in processing big data. The traditional RDBMS queries aren't sufficient to get useful information out of the huge volumes of data. Traditional tools are used to search for it. However, in order to find out if a particular topic was trending it would take so long that the result would be meaningless by the time it was computed. Bigdata offers solution to store this data in novel ways in order to make it more accessible, and also to come up with the methods of performing analysis on it. It involves capturing, storing, searching, sharing, analyzing and visualization of data. The traditional model pertains to data processing where few companies are generating data, all others are consuming data (Figure 1) whilst the novel data model involves all are generating as well as consuming data (Figure 2).



Figure 1: Traditional Model



Figure 2: Novel Model

H

Although the concept of big data itself is relatively new, the origins of large data sets go back to the 1960s and '70s when the world of data was just getting started with the first data centers and the development of the relational database.

Around 2005, people began to realize just how much data users generated through Facebook, YouTube, and other online services. Hadoop (an open-source framework created specifically to store and analyze big data sets) was developed that same year. NoSQL also began to gain popularity during this time.

The development of open-source frameworks, such as Hadoop (and more recently, Spark) was essential for the growth of big data because they make big data easier to work with and cheaper to store. In the years since then, the volume of big data has skyrocketed. Users are still generating huge amounts of data—but it's not just humans who are doing it.

With the advent of the Internet of Things (IoT), more objects and devices are connected to the internet, gathering data on customer usage patterns and product performance. The emergence of machine learning has produced still more data.

While big data has come far, its usefulness is only just beginning. Cloud computing has expanded big data possibilities even further. The cloud offers truly elastic scalability, where developers can simply spin up ad hoc clusters to test a subset of data. And graph databases are becoming increasingly important as well, with their ability to display massive amounts of data in a way that makes analytics fast and comprehensive.

4.3 Benefits of Big Data Processing

The ability to process big data brings multiple benefits, such as-

- Businesses can utilize outside intelligence while taking decisions
- Access to social data from search engines and sites like Facebook, Twitter etc. are enabling organizations to fine tune their business strategies.
- Improved customer service
- Traditional customer feedback systems are getting replaced by new systems designed with Big Data technologies. In these new systems, Big Data and natural language processing technologies are being used to read and evaluate consumer responses.
- Early identification of risk to the product/services, if any
- Better operational efficiency

Big data technologies can be used for creating a staging area or landing zone for new data before identifying what data should be moved to the data warehouse. In addition, such integration of Big Data technologies and data warehouse helps an organization to offload infrequently accessed data.

Characteristics of Big Data

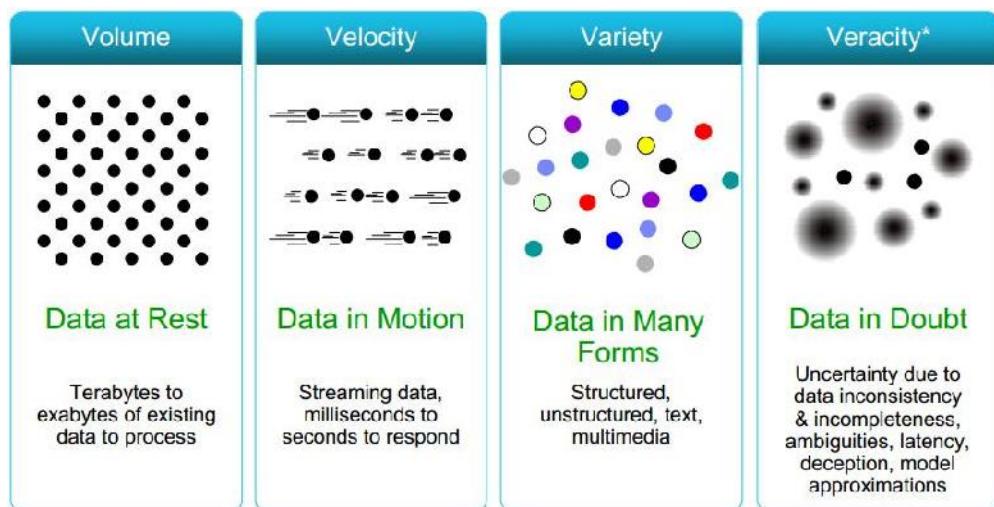


Figure 3: Big Data Characteristics

Big data can be described by the following characteristics (Figure 3):

- **Volume**- The name big data itself is related to a size which is enormous. Size of data plays a very crucial role in determining value out of data. Also, whether a particular data can actually be considered as a Big Data or not, is dependent upon the volume of data. Hence, 'volume' is one characteristic which needs to be considered while dealing with big data.

The amount of data matters. With big data, you'll have to process high volumes of low-density, unstructured data. This can be data of unknown value, such as Twitter data feeds, clickstreams on a web page or a mobile app, or sensor-enabled equipment. For some organizations, this might be tens of terabytes of data. For others, it may be hundreds of petabytes.

- **Variety**- The next aspect of big data is its variety. Variety refers to heterogeneous sources and the nature of data, both structured and unstructured. During earlier days, spreadsheets and databases were the only sources of data considered by most of the applications. Nowadays, data in the form of emails, photos, videos, monitoring devices, PDFs, audio, etc. are also being considered in the analysis applications. This variety of unstructured data poses certain issues for storage, mining and analyzing data.

Unit 04: Introduction to Big Data

- **Velocity-** The term 'velocity' refers to the speed of generation of data. How fast the data is generated and processed to meet the demands determines real potential in the data. Velocity deals with the speed at which data flows in from sources like business processes, application logs, networks, and social media sites, sensors, mobile devices, etc.

Sometimes even 2 minutes is too late. For time-sensitive processes such as catching fraud, big data must be used as it streams into an enterprise in order to maximize its value. It scrutinizes 5 million trade events created each day to identify potential fraud. It analyzes 500 million daily call detail records in real-time to predict customer churn faster.

Presently, data is being generated fast and needs to be processed fast. Velocity plays an important role in online data analytics such as e-promotions (based on user's current location and purchase history); healthcare monitoring where the sensors monitoring your activities and body measuring any abnormal measurements require immediate reaction.

- **Variability-** This refers to the inconsistency which can be shown by the data at times, thus hampering the process of being able to handle and manage the data effectively.
- **Value-** This attribute pertains to the integration of data while reducing data complexity; increasing data availability. It helps to unify the data systems.
- **Veracity-** Refers to the biases, noise & abnormality in data, trustworthiness of data. 1 in 3 business leaders don't trust the information they use to make decisions. How can you act upon information if you don't trust it? Establishing trust in big data presents a huge challenge as the variety and number of sources grows.
- **Valence-** This refers to the connectedness of big data such as in graph networks.
- **Validity-** It pertains to the accuracy and correctness of the data relative to a particular use. Example: Gauging storm intensity.
- **Viscosity and Volatility-** Both are related to the velocity. Viscosity pertains to the data velocity relative to timescale of event being studied. Volatility refers to the rate of data loss and stable lifetime of data.
- **Visualization-** Another characteristic of big data is how challenging it is to visualize. The current big data visualization tools face technical challenges due to limitations of in-memory technology and poor scalability, functionality, and response time. You can't rely on traditional graphs when trying to plot a billion data points, so you need different ways of representing data such as data clustering or using tree maps, sunbursts, parallel coordinates, circular network diagrams, or cone trees.
- **Viability-** Which data has meaningful relations to the questions of interest? (Figure 4)
- **Venue-** Where does the data live and how do you get it? Various types of data arrived from different sources via different platforms like personnel system and private & public cloud.



Figure 4: Additional V's in Big Data

- **Vocabulary**-This pertains to the metadata describing structure, content, and provenance. It includes schemas, semantics, ontologies, taxonomies, vocabularies.
- **Vagueness**-Vagueness concern the reality in information that suggested little or no thought about what each might convey.

4.4 Types of Big Data

Following are the different types of big data:

1. Structured
2. Unstructured
3. Semi-structured

Structured Data: Any data that can be stored, accessed and processed in the form of fixed format is termed as a 'structured' data. Over the period of time, talent in computer science has achieved greater success in developing techniques for working with such kind of data (where the format is well known in advance) and also deriving value out of it. However, nowadays, we are foreseeing issues when a size of such data grows to a huge extent, typical sizes are being in the rage of multiple zettabytes.



Do you know?

10^{21} bytes equal to 1 zettabyte or one billion terabytes form a zettabyte.

Unstructured Data: Any data with unknown form or structure is classified as unstructured data. In addition to the size being huge, un-structured data poses multiple challenges in terms of its processing for deriving value out of it. A typical example of unstructured data is a heterogeneous data source containing a combination of simple text files, images, videos etc. Nowadays, the organizations have wealth of data available with them but unfortunately, they don't know how to derive value out of it since this data is in its raw form or unstructured format. An example of Unstructured data is the output returned by 'Google Search'.

Semi-structured Data: Semi-structured data can contain both forms of data. We can see semi-structured data as a structured inform, but it is actually not defined with an example, a table definition in relational DBMS. Example of semi-structured data is a data represented in an XML file.

How Big Data Works?

Big data gives you new insights that open up new opportunities and business models. Getting started involves three key actions:

Integrate: Big data brings together data from many disparate sources and applications. Traditional data integration mechanisms, such as extract, transform, and load (ETL) generally aren't up to the task. It requires new strategies and technologies to analyze big data sets at terabyte, or even petabyte, scale. During integration, you need to bring in the data, process it, and make sure it's formatted and available in a form that your business analysts can get started with.

Manage: Big data requires storage. Your storage solution can be in the cloud, on premises, or both. You can store your data in any form you want and bring your desired processing requirements and necessary process engines to those data sets on an on-demand basis. Many people choose their storage solution according to where their data is currently residing. The cloud is gradually gaining popularity because it supports your current compute requirements and enables you to spin up resources as needed.

Analyze: Your investment in big data pays off when you analyze and act on your data. Get new clarity with a visual analysis of your varied data sets. Explore the data further to make new discoveries. Share your findings with others. Build data models with machine learning and artificial intelligence. Put your data to work.

A step-by-step and detailed description for the same is:

Unit 04: Introduction to Big Data

- Stage 1: Business case evaluation- Big Data analytics lifecycle begins with business case, which defines reason & goal behind the analysis.
- Stage 2: Identification of data- Here, the broad variety of data sources is identified.
- Stage 3: Data filtering- All the identified data from the previous stage is filtered here to remove corrupt data.
- Stage 4: Data extraction- Data that is not compatible with the tool is extracted & then transformed into a compatible form.
- Stage 5: Data aggregation- In this stage, data with the same fields across different datasets is integrated.
- Stage 6: Data analysis- Data is evaluated using analytical and statistical tools to discover useful information.
- Stage 7: Visualization of data- With tools like Tableau, Power BI, & QlikView, Big Data analysts can produce graphic visualizations of the analysis.
- Stage 8: Final analysis result- Last step of Big Data analytics lifecycle, where the final results of analysis are made available to business stakeholders who will take action.

4.5 Challenges in Handling Big Data

Various challenges faced in Big data have been listed below. Figure 5 depicts different big data challenges versus growth of data.

- Bottlenecks in technology
 - New architecture, algorithms, techniques are needed.
- Technical skills
 - Experts in using the new technology and dealing with big data.
 - Other Challenges: Capture, curation, storage, search, sharing, transfer, analysis, and visualization of larger data sets due to additional information derivable from analysis of a single large set of related data, as compared to separate smaller sets with the same total amount of data.

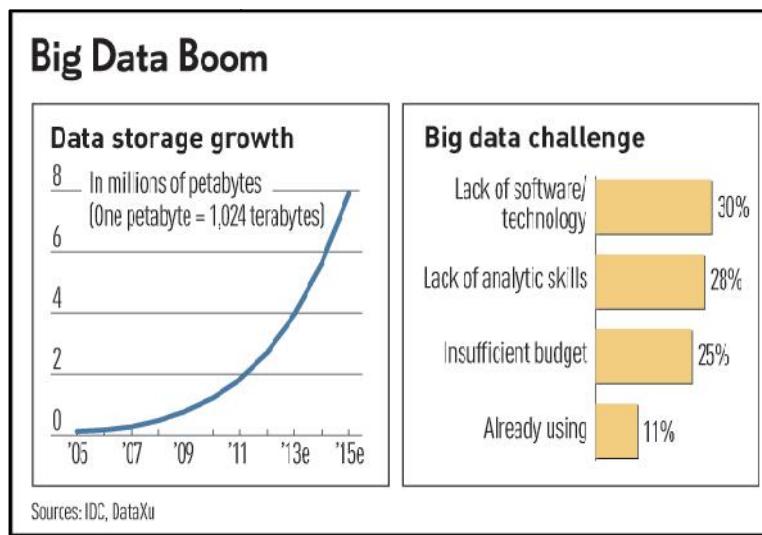


Figure 5: Big Data Challenges

4.6 Big Data Analytics

The global big data market revenues for software and services are expected to increase from \$42 billion to \$103 billion by year 2027. Every day, 2.5 quintillion bytes of data are created, and it's only

Cloud Computing

in the last two years that 90% of the world's data has been generated. In fact, if we predict correctly, there's likely much more to come. The world is driven by data, and it's being analyzed every second, whether it's through a phone's Google Maps, Netflix browsing habits, or someone's reserved items in online shopping cart. Data is unavoidable and disrupting almost every known market. The business world looks to data for market insights and ultimately, to generate growth and revenue. Data is becoming a game changer within the business arena, it's important to note that data is also being utilized by small businesses, corporate and creative alike. Big data analytics is expanding in such scenarios.

Big data analytics is the use of advanced analytic techniques against very large, diverse data sets that include structured, semi-structured and unstructured data, from different sources, and in different sizes from terabytes to zettabytes.

The analysis of big data allows the analysts, researchers and business users to make better and faster decisions using data that was previously inaccessible or unusable. Businesses can use advanced analytics techniques such as text analytics, machine learning, predictive analytics, data mining, statistics and natural language processing to gain new insights from previously untapped data sources independently or together with an existing enterprise data.

Where Role of Big Data Analytics Comes into Play

The volume of data that one has to deal with has exploded to unimaginable levels in the past decade, and at the same time, price of data storage has systematically reduced. Private companies and research institutions capture terabytes of data about their users' interactions, business, social media, and also sensors from devices such as mobile phones and automobiles. This is where big data analytics comes into picture.

Data analytics is the process of examining data sets (within the form of text, audio and video), & drawing conclusions about the information they contain, more commonly through specific systems, software, and methods. It is the process of converting large amounts of unstructured raw data, retrieved from different sources to a data product useful for organizations. Several data analytics technologies are used on an industrial scale, across commercial business industries, as they enable organizations to make calculated, informed business decisions. Globally, enterprises are harnessing the power of various data analysis techniques & using it to reshape their business models. As technology develops, new analysis software emerges, and as the Internet of Things (IoT) grows, the amount of data increases. Big data has evolved as a product of our increasing expansion and connection, and with it, new forms of extracting, or rather "mining", data.

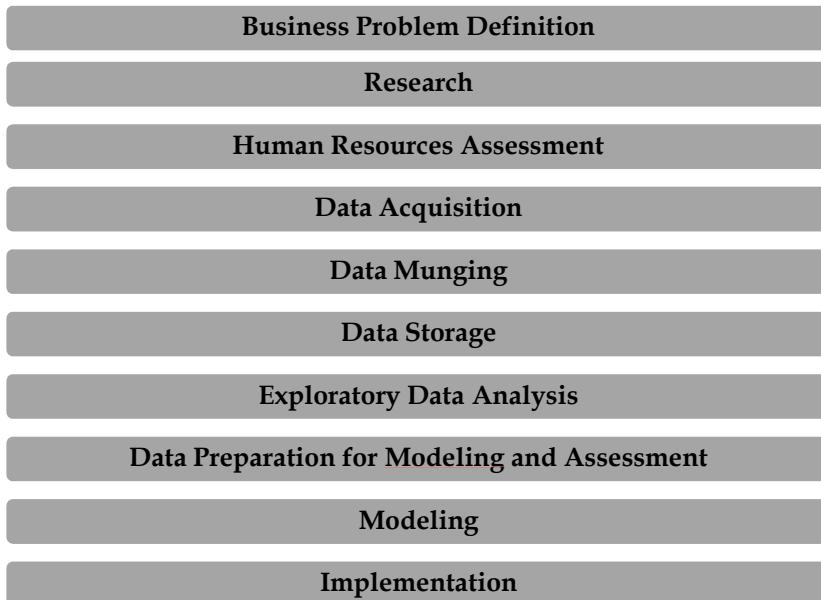


Figure 6: Big Data Analytics Cycle

Big Data Analytics largely involves collecting data from different sources, munge it in a way that it becomes available to be consumed by analysts and finally deliver data products useful to the organization business. The global survey from McKinsey revealed that when organizations use

Unit 04: Introduction to Big Data

data, it benefits the customer and the business by generating new data-driven services, developing new business models and strategies, and selling data-based products and utilities. The incentive for investing and implementing data analysis tools and techniques is huge, and businesses will need to adapt, innovate, and strategize for evolving digital marketplace.

4.7 Big Data Analytics Cycle

It consists of the following phases and their key features:

1. Business Problem Definition

This phase is the common point in traditional Business Intelligence (BI) and big data analytics life cycle. A non-trivial stage of a big data project to define the problem & evaluate correctly how much potential gain it may have for an organization. It evaluates what are expected gains and costs of the project.

2. Research

This phase analyzes what other companies have done in the same situation. It involves looking for solutions that are reasonable for your company, even though it involves adapting other solutions to resources and requirements that your company has. In this stage, a methodology for the future stages should be defined.

3. Human Resources Assessment

Once the problem is defined, it's reasonable to continue analyzing if the current staff is able to complete the project successfully. Traditional BI teams might not be capable of delivering an optimal solution to all the stages, so it should be considered before starting the project if there is a need to outsource a part of the project or hire more people.

4. Data Acquisition

Data acquisition is a key stage in a big data cycle and defines which type of profiles would be needed to deliver resultant data product. Data gathering is a non-trivial step of the process; it normally involves gathering unstructured data from different sources. Example: Writing a crawler to retrieve reviews from a website. It involves dealing with text, perhaps in different languages normally requiring a significant amount of time to be completed.

5. Data Munging

Once the data is retrieved, say, from the web, it needs to be stored in an easy-to-use format. Example: Let's assume data is retrieved from different sites where each has a different display of the data. Suppose one data source gives reviews in terms of rating in stars, therefore it is possible to read this as a mapping for the response variable $y \in \{1, 2, 3, 4, 5\}$. Another data source gives reviews using two arrows system, one for up voting and the other for down voting. This would imply a response variable of the form $y \in \{\text{positive, negative}\}$. In order to combine both the data sources, a decision has to be made in order to make these two response representations equivalent. This can involve converting first data source response representation to the second form, considering one star as negative and five stars as positive. This process often requires a large time allocation to be delivered with good quality.

6. Data Storage

Once the data is processed, it sometimes needs to be stored in a database. Big data technologies offer plenty of alternatives such as using Hadoop File System for storage that provides users a limited version of SQL, known as HIVE Query Language. From the user perspective, data analytics allows most analytics tasks to be done in similar ways as in traditional BI data warehouses. The other storage options can be MongoDB, Redis, and SPARK. The modified versions of traditional data warehouses are still being used in large scale applications. Example: Teradata and IBM offer SQL databases that can handle terabytes of data; open-source solutions such as PostgreSQL and MySQL are still being used for large scale applications. It is important to implement a big data solution that would be working with real-time data, so in this case, we only need to gather data to

Cloud Computing

develop the model & then implement it in real time. So, there would not be a need to formally store the data at all.

7. Exploratory Data Analysis

Once the data has been cleaned and stored in a way that insights can be retrieved from it, the data exploration phase is mandatory. The objective of this stage is to understand the data, this is normally done with statistical techniques and also plotting the data. It is a good stage to evaluate whether the problem definition makes sense or is feasible.

8. Data Preparation for Modeling and Assessment

Data preparation stage involves reshaping the cleaned data retrieved previously and using statistical pre-processing for missing values imputation, outlier detection, normalization, feature extraction and feature selection.

9. Modelling

The prior stage in analytics cycle produces several datasets for training and testing, for example, a predictive model. It involves trying different models and looking forward to solving the business problem at hand. The expectation is the model would give some insight into business. Finally, the best model or combination of models is selected evaluating its performance on a left-out dataset.

10. Implementation

In this stage, the data product developed is implemented in the data pipeline of the company. It involves setting up a validation scheme while a data product is working, in order to track its performance. Example: Implementing a predictive model. It involves applying the model to new data and once the response is available, evaluate the model.

Big Data Analysis Techniques

Big data is characterized by 3 V's: major volume of data, the velocity at which it's processed, and the wide variety of data. Velocity pertains that data analytics has expanded into technological fields of machine learning and AI. Alongside evolving computer-based analysis techniques & data harnesses, analysis also relies on the traditional statistical methods. Ultimately, how data analysis techniques function within an organization is twofold; big data analysis is processed through streaming of data as it emerges, and then performing batch analyses of data as it builds – to look for behavioral patterns and trends. As generation of data increases, so various techniques manage it. As data becomes more insightful in its speed, scale, and depth, more it fuels innovation. There are 6 big data analysis techniques:

1. A/B Testing

- Involves comparing a control group with a variety of test groups, to discern what treatments or changes will improve a given objective variable.
- McKinsey gives example of analysing what copy, text, images, or layout will improve conversion rates on an e-commerce site.
- Big data once again fits into this model as it can test huge numbers; however, it can only be achieved if the groups are of a big enough size to gain meaningful differences.

2. Data Fusion and Data Integration

By combining a set of techniques that analyze and integrate data from multiple sources and solutions, the insights are more efficient and potentially more accurate than if developed through a single source of data.

3. Data Mining

Data mining is a common tool used within big data analytics. Data mining extracts patterns from large data sets by combining methods from statistics & machine learning, within database management. Example: When customer data is mined to determine which segments are most likely to react to an offer.

4. Machine Learning

Machine learning is well-known within the field of artificial intelligence and is also used for data analysis. Emerging from computer science, it works with computer algorithms to produce assumptions based on data. It provides predictions that would be impossible for human analysts.

5. Natural Language Processing (NLP)

NLP is known as a subspecialty of computer science, artificial intelligence, and linguistics. This data analysis tool uses algorithms to analyze human (natural) language.

6. Statistics

This technique works to collect, organize, and interpret data, within surveys and experiments.

Other data analysis techniques include spatial analysis, predictive modelling, association rule learning, network analysis and many, many more.

4.8 Big Data Analytics- Data Analysis Tools

R Programming Language: R is an open-source programming language with a focus on statistical analysis. It is competitive with commercial tools such as SAS, SPSS in terms of statistical capabilities and is thought to be an interface to other programming languages such as C, C++ or Fortran. Another advantage of R is the large number of open-source libraries that are available. In CRAN, there are more than 6000 packages that can be downloaded for free and in Github, there are wide variety of R packages available.

Python for Data Analysis: Python is a general-pura wide programming language. It contains a significant number of libraries devoted to data analysis (pandas, scikit-learn, theano, numpy and scipy). Python can be used quite effectively to clean and process the data line by line. For machine learning, scikit-learn is a nice environment that has a largenumber of algorithms that can handle medium-sized datasets without a problem.

Julia: It is ahigh-level, high-performance dynamic programming language for technical computing. The syntax is quite similar to R or Python, so if you are already working with R or Python, it should be quite simple to write the same code in Julia.

SAS: It is a commercial language that is still being used for business intelligence. It has a base language that allows the user to program a wide variety of applications and contains quite a few commercial products that give non-experts users the ability to use complex tools such as a neural network library without the need of programming.

SPSS: SPSS, is currently a product of IBM for statistical analysis and is used to analyze survey data and for users that are not able to program, it is a decent alternative. It provides a SQL code to score a model. This code is normally not efficient, but it's a start whereas SAS sells the product that scores models for each database separately. For small data and an unexperienced team, SPSS is an option as good as SAS is.

Matlab, Octave: Matlab or its open-source version (Octave) are mostly used for research.R or Python can do all that's available in Matlab or Octave.

4.9 Parallel Data Processing

Parallel data processing involves the simultaneous execution of multiple sub-tasks that collectively comprise a larger task. The goal is to reduce the execution time by dividing a single larger task into multiple smaller tasks that run concurrently. Although parallel data processing can be achieved through multiple networked machines, it is more typically achieved within the confines of a single machine with multiple processors or cores, as shown in figure below. Figure 7 shows that A task is divided into three sub-tasks that are executed in parallel on three different processors within the same machine.

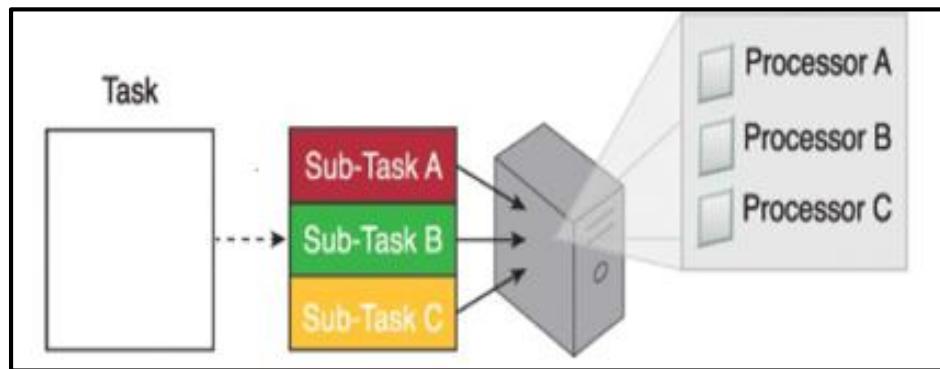


Figure 7: Parallel Data Processing Scenario

4.10 Distributed Data Processing and Cluster Processing

Distributed data processing is closely related to parallel data processing. It follows the principle of “divide-and-conquer”. It is always achieved through physically separate machines that are networked together as a cluster. As illustrated in the Figure 8, a task is divided into three sub-tasks that are then executed on three different machines sharing one physical switch.

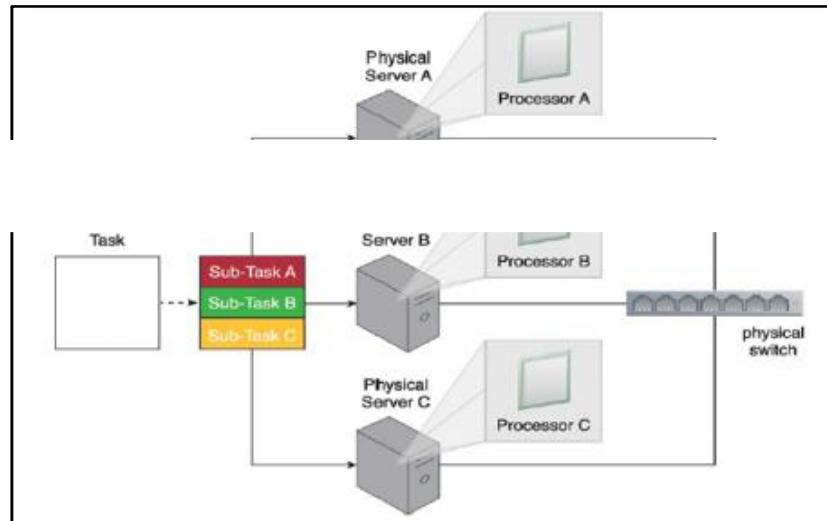


Figure 8: Distributed Data Processing Scenario

Cluster processing provides necessary support to create horizontally scalable storage solutions. It also provides the mechanism to enable distributed data processing with linear scalability. Since clusters are highly scalable, they provide an ideal environment for Big data processing as large datasets can be divided into smaller datasets and then processed in parallel in a distributed manner. When leveraging a cluster, Big Data datasets can either be processed in batch mode or real-time mode.

4.11 Processing Workloads in Big Data

The amount and nature of data that is processed within a certain amount of time. The workloads are usually divided into two types:

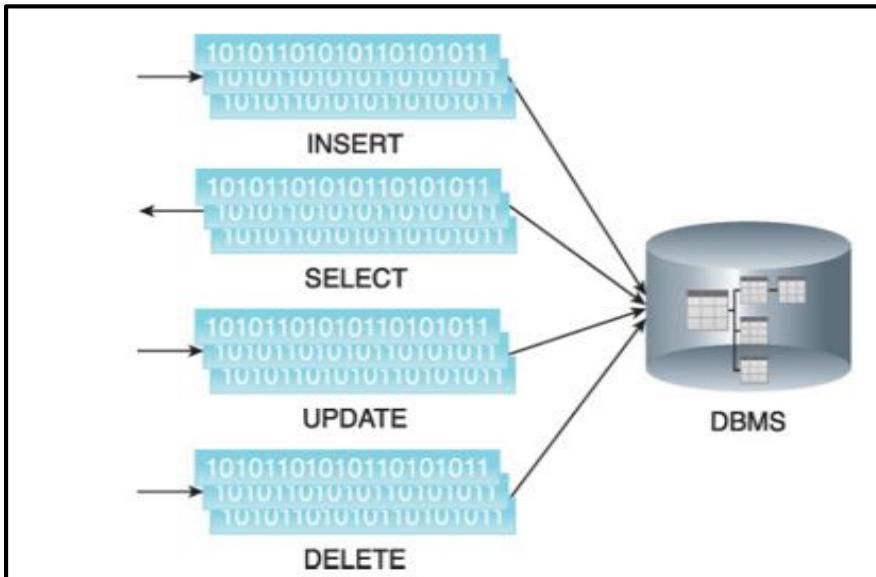


Figure 9: Batch Processing Scenario

(a) Batch Processing Workload: It is also known as offline processing. In batch mode, data is processed offline in batches and the response time could vary from minutes to hours (Figure 9). As well, data must be persisted to the disk before it can be processed. It involves processing data in batches and usually imposes delays, which in turn results in high-latency responses. Batch mode generally involves processing a range of large datasets, either on their own or joined together, essentially addressing the volume and variety characteristics of big data datasets. It is relatively simple, easy to set up and low in cost compared to real-time mode. Batch workloads typically involve large quantities of data with sequential read/writes and comprise of groups of read or write queries. The queries can be complex and involve multiple joins. Batch workload comprises grouped read/writes that have a large data footprint and may contain complex joins and provide high-latency responses. The majority of big data processing occurs in batch mode. Example:

- o OLAP systems commonly process workloads in batches.
- o Strategic BI and analytics are batch-oriented as they are highly read-intensive tasks involving large volumes of data.

(b) Transactional Processing Workloads: It is also known as online processing and follows an approach whereby data is processed interactively without delay, resulting in low-latency responses (Figure 10). The transaction workloads involve small amounts of data with random reads and writes. The transactional workloads comprise random reads/writes that involve fewer joins than business intelligence and reporting workloads.

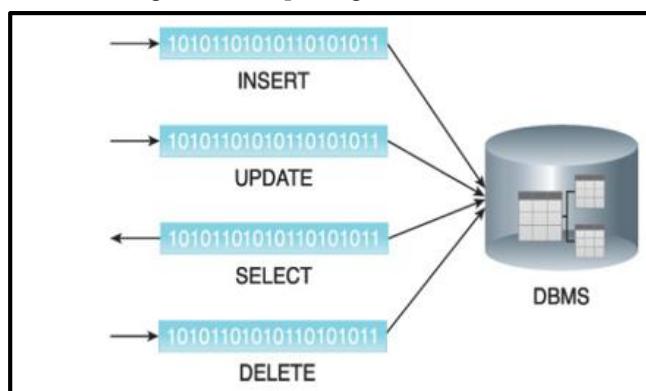


Figure 10: Transactional Processing Scenario

4.12 Hadoop Framework

Hadoop is an open-source framework for large-scale data storage and data processing that is compatible with commodity hardware. It is a software framework for storage and large-scale processing of datasets on clusters of commodity hardware (Figure 11).

Hadoop framework has established itself as a de facto industry platform for contemporary big data solutions. It was created by Doug Cutting and Mike Cafarella in 2005, named after his son's toy elephant.

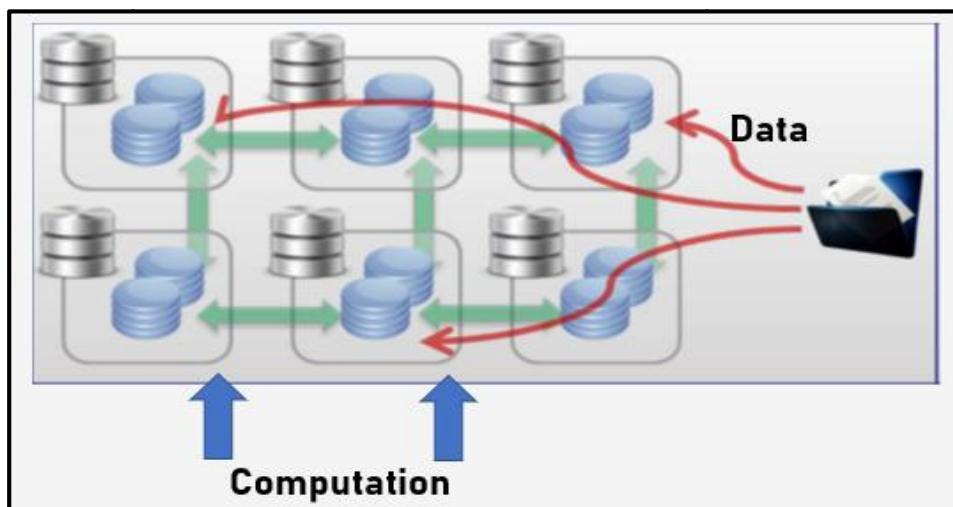
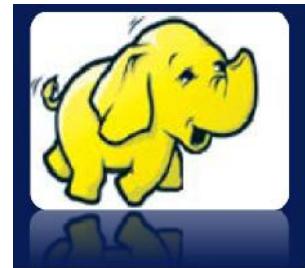


Figure 11: Hadoop Offers Moving Computation to Data- Computation is Moved to Data

Hadoop Offerings

- New Approach to Data: Hadoop offers a different approach to data.



- New Kinds of Analysis: Hadoop can be used as an ETL engine or as an analytics engine for processing large amounts of structured, semi structured and unstructured data. From an analysis perspective, Hadoop implements the MapReduce processing framework.



4.13 Hadoop Modules

Hadoop has certain modules consisting of (as shown in Figure 12):

Hadoop Common: Hadoop Common refers to the collection of common utilities and libraries that support other Hadoop modules. It is an essential part or module of the Apache Hadoop Framework, along with the Hadoop Distributed File System (HDFS), Hadoop YARN and Hadoop

Unit 04: Introduction to Big Data

MapReduce. It assumes that hardware failures are common and that these should be automatically handled in software by the Hadoop Framework. It is also known as Hadoop core.

The Hadoop common package is considered as the base/core of the framework as it provides essential services and basic processes such as abstraction of the underlying operating system and its file system. It contains the necessary Java Archive (JAR) files and scripts required to start Hadoop. It provides source code and documentation, as well as a contribution section that includes different projects from the Hadoop community.

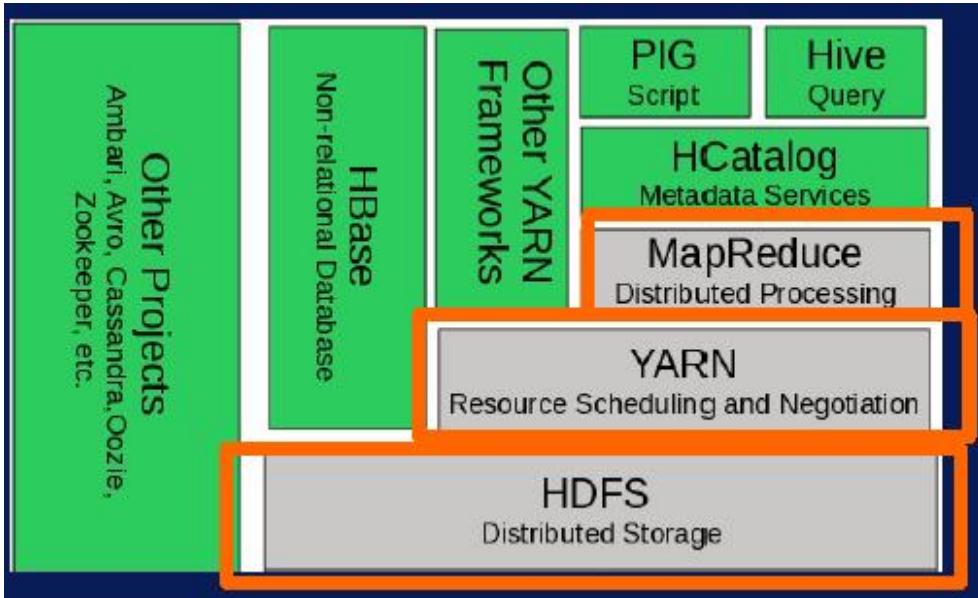


Figure 12: Hadoop Modules

Hadoop YARN: One of the main components in Hadoop. It stands for Yet Another Resource Negotiator though it is called Yarn by the developers. It supports a resource management platform that is responsible for managing compute resources in the cluster and using them in order to schedule users and applications. It is used for job scheduling and resource management. Yarn is the parallel processing framework for implementing distributed computing clusters that processes huge amounts of data over multiple compute nodes. Hadoop Yarn allows a compute job to be segmented into hundreds and thousands of tasks.

Yarn was introduced as a layer that separates the resource management layer and the processing layer. It supports a broader range of different applications. It combines central resource manager with different containers. Can combine the resources dynamically to different applications and the operations are monitored well. It was previously called MapReduce2 and Nextgen MapReduce.

Yarn was initially named MapReduce 2 since it powered up the MapReduce of Hadoop 1.0 by addressing its downsides and enabling the Hadoop ecosystem to perform well for modern challenges. It enables Hadoop to support different processing types. It runs interactive queries, streaming data and real time applications.

Yarn uses master servers and data servers. There is only one master server per cluster. It runs the resource manager daemon. There are many data servers in the cluster, each one runs on its own Node Manager daemon and the application master manager as required.

Components of YARN

The major components responsible for all the YARN operations are as follows:

- Resource Manager: Resource Manager allocates the cluster resources. This is made possible by a scheduler for scheduling the required jobs and an Application Manager for accepting the job submissions and executing the necessary Application Master. Its functionalities include:
 - ✓ Resource Management: The Resource Manager interacts with the Node Manager to keep an inventory of cluster-wide resources such as RAM, CPU and Network usage. It manages the availability and allocation of the resources thus gaining the ultimate authority of managing the resources.

Cloud Computing

- ✓ Scheduling Management: This is responsible for only allocating resources as requested by various Application Masters. It does not monitor or track the status of the job.
- ✓ Application Management: This is the most important responsibility of the Resource Manager. It ensures the start-up and job completion. It accepts the job and assigns it to an Application Master on a data server. In addition to this, it also monitors the activity and status of the Application Master and provides a restart in case of failure of Application Master.
- ✓ Containers in Hadoop: Hadoop v2.0 has enhanced parallel processing with the addition of containers. Containers are the abstract notion that supports multi-tenancy on a data node. It is a way to define requirements for memory, CPU and network allocation by dividing the resources on the data server into a container.
- ✓ Resource Containers: Resource Manager is responsible for scheduling resources by allocating containers. This is done in accordance with the input provided by client, cluster capacity, queues & overall prioritization of resources on the cluster. Resource Manager uses an algorithm to allocate the containers.
- Node Manager: The primary goal of the Node Manager is memory management. Node Manager tracks the usage and status of the cluster inventories such as CPU, memory, and network on the local data server and reports the status regularly to the Resource Manager.
- Application Master - Application Master is responsible for execution in parallel computing jobs. Its daemon is accountable for executing the job, monitoring the job for error, and completing the computer jobs. These daemons are started by the resource manager at the start of a job. Each compute job has an Application Master running on one of the data servers. The Application Master requests the data locality from the name node of the master server. It then negotiates with the scheduler function in the Resource Manager for the containers of resources throughout the cluster.

Key Features of Yarn

- Multi-Tenancy: Yarn allows multiple engine access and fulfills the requirement for a real-time system for knowing where the data is located and for managing the movement of that data within the framework for dispatching compute jobs to run on the right data at the right location.
- Sharing Resources: Yarn ensures there is no dependency between the compute jobs. Each compute job is run on its own node and does not share its allocated resources. Each job is responsible for its own assigned work.
- Cluster Utilization: Yarn optimizes the cluster by utilizing and allocating its resources in a dynamic manner.
- Fault Tolerance: Yarn is highly fault tolerant. It allows the rescheduling of the failed to compute jobs without any implications for the final output.
- Scalability: Yarn focuses mainly on scheduling the resources. This creates a pathway for expanding the data nodes and increasing the processing capacity.
- Compatibility: The jobs that were working in MapReduce v1 can be easily migrated to higher versions of Hadoop with ease ensuring high compatibility of Yarn.

Hadoop Distributed File System (HDFS)

Hadoop Distributed File System is the major component of Apache Hadoop. It handles large data sets running on commodity hardware. It allows multiple files to be stored and retrieved at the same

Unit 04: Introduction to Big Data

time at an unprecedented speed. HDFS is used to scale a single Apache Hadoop cluster to hundreds (and even thousands) of nodes. HDFS should not be confused with or replaced by Apache HBase, which is a column-oriented non-relational database management system that sits on top of HDFS and can better support real-time data needs with its in-memory processing engine. The servers in HDFS are fully connected, and they communicate through TCP-based protocols. Though designed for huge databases, normal file systems (FAT, NTFS, etc.) can also be viewed. The current status of a node is obtained through the checkpoint node. Table lists the different features of HDFS.

Goals of HDFS

- Fast Recovery from Hardware Failures- Because one HDFS instance may consist of thousands of servers, failure of at least one server is inevitable. HDFS has been built to detect faults and automatically recover quickly.
- Access to Streaming Data- HDFS is intended more for batch processing versus interactive use, so the emphasis in the design is for high data throughput rates, which accommodate streaming access to data sets.
- Accommodation of Large Data Sets- HDFS accommodates applications that have data sets typically gigabytes to terabytes in size. HDFS provides high aggregate data bandwidth and can scale to hundreds of nodes in a single cluster.
- Portability- To facilitate adoption, HDFS is designed to be portable across multiple hardware platforms and to be compatible with a variety of underlying operating systems.

Table 1: Key Features of HDFS

HDFS Key Features	Description
Storing bulks of data	HDFS is capable of storing terabytes and petabytes of data.
Minimum intervention	It manages thousands of nodes without operators' intervention.
Computing	HDFS provides the benefits of distributed and parallel computing at once.
Scaling out	It works on scaling out, rather than on scaling up, without a single downtime.
Rollback	HDFS allows returning to its previous version post an upgrade.
Data integrity	It deals with corrupted data by replicating it several times.

An Example of HDFS

Consider a file that includes the phone numbers for everyone in the United States; the numbers for people with a last name starting with A might be stored on server 1, B on server 2, and so on. With Hadoop, pieces of this phonebook would be stored across the cluster, and to reconstruct the entire phonebook, your program would need the blocks from every server in the cluster. In order to ensure availability if and when a server fails, HDFS replicates these smaller pieces onto two additional servers by default. (The redundancy can be increased or decreased on a per-file basis or for a whole environment; for example, a development Hadoop cluster typically doesn't need any data redundancy). This redundancy offers multiple benefits, the most obvious being higher availability. The redundancy also allows the Hadoop cluster to break up work into smaller chunks and run those jobs on all the servers in the cluster for better scalability. Finally, you gain the benefit of data locality, which is critical when working with large data sets.

IBM and Cloudera have partnered to offer an industry-leading, enterprise-grade Hadoop distribution, including an integrated ecosystem of products and services to support faster analytics at scale.

**Notes:****Why Should You Use HDFS?**

- ✓ HDFS stores the same data in multiple sets-
- ✓ Distributed across hundreds or even thousands of servers with each node storing a part of the file system. Since the storage is done on commodity hardware, there are more chances of the node failing and, with that, the data can be lost. HDFS gets over that problem by storing the same data in multiple sets.
- ✓ HDFS works quite well for data loads that come in a streaming format. So, it is more suited for batch processing applications rather than for interactive use. It is important to note that HDFS works for high throughput rather than low latency.
- ✓ HDFS works exclusively well for large datasets, and the standard size of datasets could be anywhere between gigabytes and terabytes. It provides high-aggregate data bandwidth, and it is possible to scale hundreds of nodes in a single cluster. Hence, millions of files are supported in a single instance.
- ✓ HDFS works on the assumption that moving of computation is much easier, faster, and cheaper than moving of data of humongous size, which can create network congestion and lead to longer overall turnaround times.
- ✓ HDFS provides the facility to let applications access data at the place where they are located.
- ✓ HDFS is highly profitable in the sense that it can easily work on commodity hardware that is of different types without any issue of compatibility. Hence, it is very well suited for taking advantage of cheaply and readily available commodity hardware components.

Hadoop MapReduce

Hadoop MapReduce is a programming model that scales data across a lot of different processes. MapReduce is defined as the framework of Hadoop, which is used to process a huge amount of data parallelly on large clusters of commodity hardware in a reliable manner. It allows the application to store the data in the distributed form and process large dataset across groups of computers using simple programming models, so that's why we can call MapReduce as a programming model used for processing huge amount of data distributed over the number of clusters using the different steps like Input splits, Map, Shuffle and Reduce.

How does MapReduce in Hadoop make working so easy?

- MapReduce makes it easy to scale up data processing over hundreds or thousands of cluster machines.
- The MapReduce model actually works in two steps called map and reduce, and the processing called mapper and reducer, respectively.
- Once we write MapReduce for an application, scaling up to run over multiples or even multiple of thousand clusters is merely a configuration change.
- This feature of the MapReduce model has attracted many programmers to use it.

4.14 Hadoop MapReduce

Hadoop MapReduce is a framework of Hadoop used to process a huge amount of data parallelly on large clusters of commodity hardware in a reliable manner. It allows the application to store the data in the distributed form and process large dataset across groups of computers using simple programming models. A programming model used for processing huge amounts of data distributed over the number of clusters using the different steps like Input splits, Map, Shuffle and Reduce.

MapReduce Architecture

Hadoop cluster stores a large set of data which is parallelly processed mainly by MapReduce. It provides parallelism, fault-tolerance, and data distribution. Map Reduce provides API with features such as parallel processing of huge amounts of data, batch processing, and high availability. The Map Reduce programs are written by programmers when there is a need for an application for business scenarios. The development of applications and deployment across Hadoop clusters is done by the programmers when they understand the flow pattern of MapReduce. Hadoop can be developed in programming languages like Python and C++. MapReduce Hadoop is a software framework for ease in writing applications of software processing huge amounts of data.

MapReduce is a framework which splits the chunk of data, sorts the map outputs and input to reduce tasks. The file-system stores the work and input of jobs. Re-execution of failed tasks, scheduling them, and monitoring them is the task of the framework. The architecture of MapReduce basically has two main processing stages, Map and Reduce. The MapReduce happens in Job tracker while the intermediate processes will take place in between the Map and Reduce phases. The sort and shuffle are the tasks taken up by Map and Reduce, which are done intermediately. The local file system stores the intermediate data.

Map() Function: Create and process the import data. Takes in data, converts it into a set of other data where the breakdown of individual elements into tuples is done—no API contract requiring a certain number of outputs.

Reduce() Function: Mappers output is passed into the reduction. Processes the data into something usable. Every single mapper is passed into the reduced function. The new output values are saved into HDFS.

How MapReduce in Hadoop Works?

MapReduce program executes mainly in 4 steps:

- Input splits
- Map
- Shuffle
- Reduce

(a) Map Step- It involves a combination of the input splits step and the Map step. In the Map step, the source file is passed as line by line. Before input passes to the Map function job, the input is divided into the small fixed-size called Input splits. The Input split is a chunk of the information which a single map could consume. In the Map step, each split data is passed to the mapper function, then the mapper function processes the data and then output values. Generally, the map or mapper's job input data is in the form of a file or directory stored in the Hadoop file system (HDFS).

(b) Reduce Step- This step is the combination of the Shuffle step and the Reduce. The reduce function or reducer's job takes the data, which results from the map function, after processing by reducing the role new set of effect produces, which again store back into the HDFS.

Hadoop framework is not sure that each cluster performs which job, either Map or Reduce or both Map and Reduce. So, the Map and Reduce tasks' request should be sent to the appropriate servers in the cluster. Hadoop framework itself manages all the tasks of issuing, verifying completion of work, fetching data from HDFS, copying data to the nodes' group, and so all. In Hadoop, mostly the computing takes place on nodes and data in nodes itself which reduces the network traffic.

Working of MapReduce Organizers

The job is divided into two components: Map tasks (Splits and mapping) and Reduce tasks (Reducing and shuffling) (Figure 13). The job/task tracker is associated with complete execution of a given job, by behaving like a master. Whereas the multiple task trackers act like slaves by performing the job each.

Map Phase- The map phase splits input data into two parts- Keys and Values (Figure 14). The writable and comparable is the key in the processing stage where only in the processing stage, Value is writable. Let's say a client gives input data to a Hadoop system; task tracker is assigned tasks by job tracker. Splitting of input is done into several inputs. Key-value pair conversion is done

Cloud Computing

with the input data by the record reader. This is the actual data input for Map as mapped information for further processing. The format type varies, so the coder has to look into each piece of data format and code accordingly. The mini reducer, which is commonly called a combiner, the reducer code places input as the combiner. A combiner is a type of local Reducer that groups similar data from the map phase into identifiable sets. It takes the intermediate keys from the mapper as input and applies a user-defined code to aggregate the values in a small scope of one mapper. It is not a part of the main MapReduce algorithm; it is optional. The network bandwidth is high when a huge amount of data is required. Hash is the default partition used. The partition module plays a key role in Hadoop. More performance is given by reducing the pressure by petitioner on the reducer.

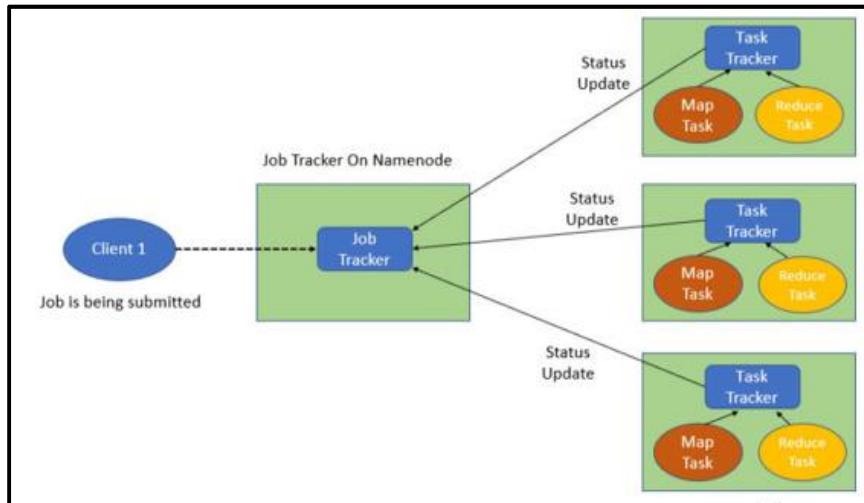


Figure 13: MapReduce Organizer Working

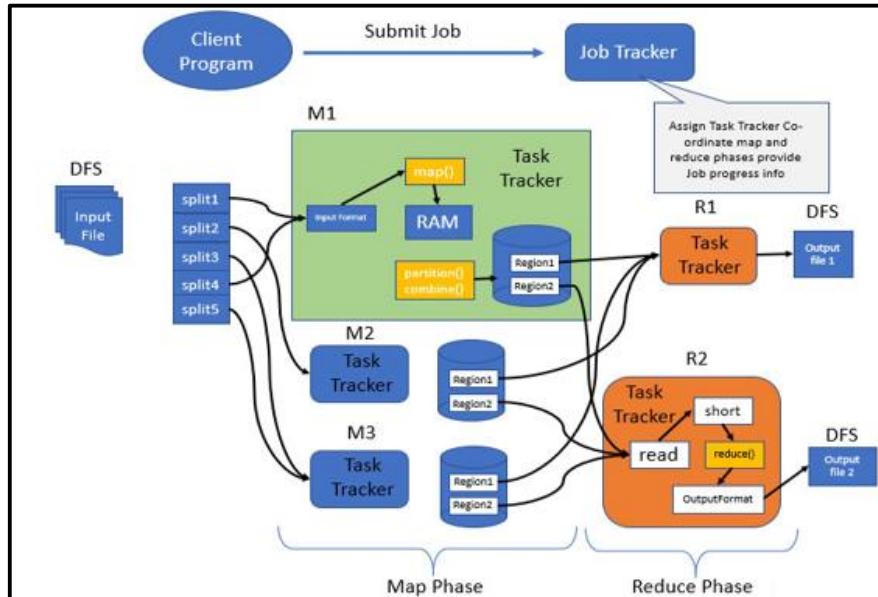


Figure 14: MapReduce Architecture & Workflow

Processing in Intermediate- In the intermediate phase, the map input gets into the sort and shuffle phase. Hadoop nodes do not have replications where all the intermediate data is stored in a local file system. Round-robin data is used by Hadoop to write to local disk, the intermediate data. There are other shuffles and sort factors to be considered to reach the condition of writing the data to local disks. The Reducer task starts with the Shuffle and Sort step. It downloads the grouped key-value pairs onto the local machine, where the Reducer is running. The individual key-value pairs are sorted by key into a larger data list. The data list groups the equivalent keys together so that their values can be iterated easily in the Reducer task.

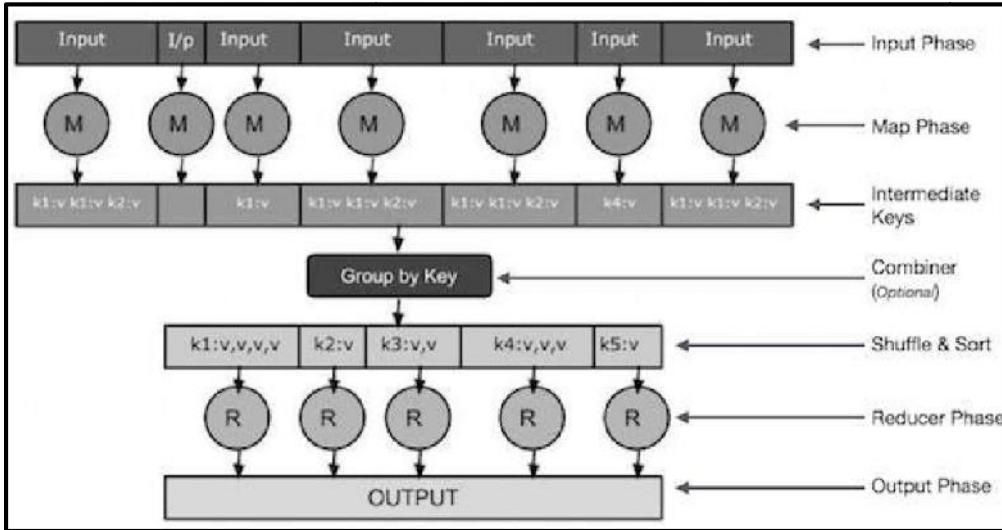


Figure 15: MapReduce Processing

Reducer Phase- The reducer takes in the data input that is sorted and shuffled. All the input data will be combined, and similar key-value pairs are to be written to the HDFS system. For searching and mapping purposes, a reducer is not always necessary. Setting some properties for enabling to develop of the number of reducers for each task. During job processing, speculative execution plays a prominent role. The performance First In First Out (FIFO), and if more than one mapper is working on similar data, and if one is running slow, then the tasks are assigned to the next mapper for a fast program run by the job tracker. The Reducer takes the grouped key-value paired data as input and runs a Reducer function on each one of them. Here, the data can be aggregated, filtered, and combined in a number of ways, and it requires a wide range of processing. Once the execution is over, it gives zero or more key-value pairs to the final step.

Output Phase- In the output phase, we have an output formatter that translates the final key-value pairs from the Reducer function and writes them onto a file using a record writer.

MapReduce Example

Traditional enterprise approach: In this approach, an enterprise will have a computer to store and process big data (Figure 16). For storage purposes, the programmers will take the help of their choice of database vendors such as Oracle, IBM, etc. In this approach, the user interacts with the application, which in turn handles the part of data storage and analysis.

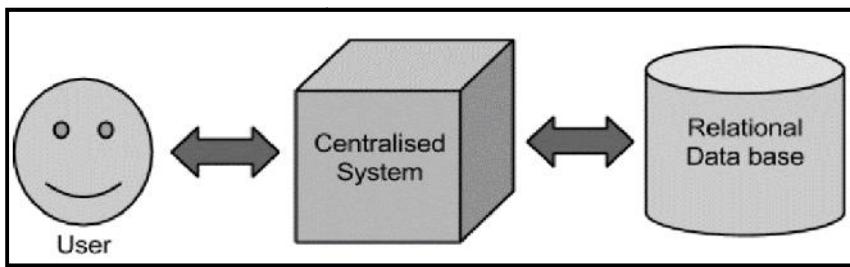


Figure 16: Traditional Enterprise Approach

Google's Solution- MapReduce Algorithm: The traditional approach works fine with those applications that process less voluminous data that can be accommodated by standard database servers, or up to the limit of the processor that is processing the data. But when it comes to dealing with huge amounts of scalable data, it is a hectic task to process such data through a single database bottleneck.

Google solved this problem using an algorithm called MapReduce. This algorithm divides the task into small parts and assigns them to many computers, and collects the results from them which when integrated, form the result dataset (as shown in the Figure 17).

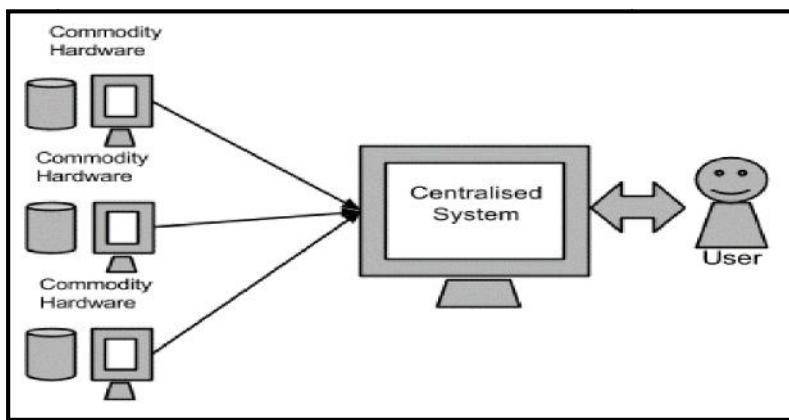


Figure 17: Google's MapReduce Approach

Let us take a real-world example to comprehend the power of MapReduce. Twitter receives around 500 million tweets per day, which is nearly 3000 tweets per second. The following illustration shows how Twitter manages its tweets with the help of MapReduce (Figure 18).

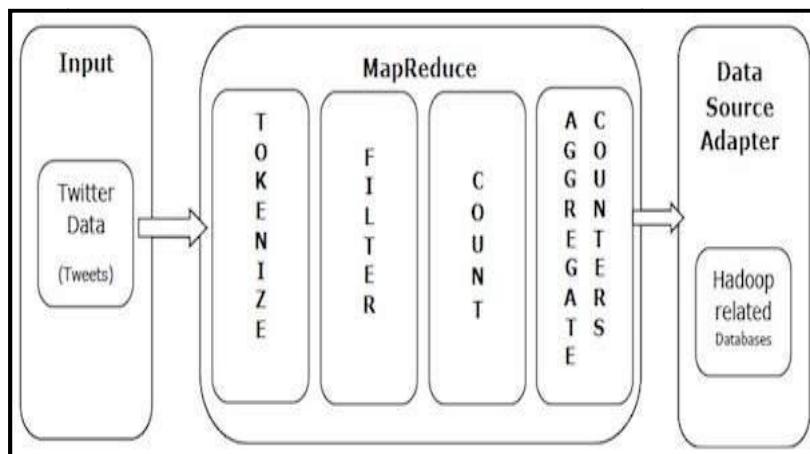


Figure 18: Twitter MapReduce Approach

As shown in the illustration, the MapReduce algorithm performs the following actions-

- Tokenize– Tokenizes the tweets into maps of tokens and writes them as key-value pairs.
- Filter– Filters unwanted words from the maps of tokens and writes the filtered maps as key-value pairs.
- Count– Generates a token counter per word.
- Aggregate Counters– Prepares an aggregate of similar counter values into small manageable units.

Summary

- Big data is a great quantity of diverse information that arrives in increasing volumes and with ever-higher velocity.
- Big data is related to extracting meaningful data by analyzing the huge amount of complex, variously formatted data generated at high speed, that cannot be handled, processed by the traditional system.
- Big data can be structured (often numeric, easily formatted and stored) or unstructured (more free-form, less quantifiable).
- Any data that can be stored, accessed and processed in the form of a fixed format is termed as a 'structured' data.

Unit 04: Introduction to Big Data

- Any data with unknown form or structure is classified as unstructured data. In addition to the size being huge, un-structured data poses multiple challenges in terms of its processing for deriving value out of it.
- Big data can be collected from publicly shared comments on social networks and websites, voluntarily gathered from personal electronics and apps, through questionnaires, product purchases, and electronic check-ins.
- Big data is most often stored in computer databases and is analyzed using software specifically designed to handle large, complex data sets.
- Big data analytics largely involves collecting data from different sources, munge it in a way that it becomes available to be consumed by analysts and finally deliver data products useful to the organization business.
- Hadoop Distributed File System is the major component of Apache Hadoop. It handles large data sets running on commodity hardware. It allows multiple files to be stored and retrieved at the same time at an unprecedented speed.
- MapReduce is a framework which splits the chunk of data, sorts the map outputs and input to reduce tasks.

Keywords

Data Mining: It is a process of extracting insight meaning, hidden pattern from collected data that is useful to take a business decision in the purpose of decreasing expenditure and increasing revenue.

Big Data: This is a term related to extracting meaningful data by analyzing the huge amount of complex, variously formatted data generated at high speed, that cannot be handled, processed by the traditional system.

Unstructured Data: The data for which structure can't be defined is known as unstructured data. It becomes difficult to process and manage unstructured data. The common examples of unstructured data are the text entered in email messages and data sources with texts, images, and videos.

Value: This big data term basically defines the value of the available data. The collected and stored data may be valuable for societies, customers, and organizations. It is one of the important big data terms as big data is meant for big businesses and the businesses will get some value i.e., benefits from the big data.

Volume: This big data term is related to the total available amount of the data. The data may range from megabytes to brontobytes.

Semi-Structured Data: The data not represented in the traditional manner with the application of regular methods is known as semi-structured data. This data is neither totally structured nor unstructured but contains some tags, data tables, and structural elements. Few examples of semi-structured data are XML documents, emails, tables, and graphs.

MapReduce: MapReduce is a processing technique to process large datasets with the parallel distributed algorithm on the cluster. MapReduce jobs are of two types. "Map" function is used to divide the query into multiple parts and then process the data at the node level. "Reduce" function collects the result of "Map" function and then find the answer to the query. MapReduce is used to handle big data when coupled with HDFS. This coupling of HDFS and MapReduce is referred to as Hadoop.

Self Assessment

1. Which is the process of examining large and varied data sets?
 - A. Big data analytics
 - B. Cloud computing
 - C. Machine learning

Cloud Computing

- D. None of the above
2. Big Data is used to uncover?
- A. Hidden patterns and unknown correlations
 - B. Market trends and customer preferences
 - C. Other useful information
 - D. All of the above
3. The composition of data deals with?
- A. Structure of data
 - B. State of data
 - C. Sensitivity of data
 - D. None
4. The important 3V's in big data are?
- A. Volume, Vulnerability, Variety
 - B. Volume, Velocity, Variety
 - C. Velocity, Vulnerability, Variety
 - D. Volume, Vulnerability, Velocity
5. Big data is an evolving term that describes any voluminous amount of data that has the potential to be mined for information?
- A. Structured
 - B. Unstructured
 - C. Semi- Structured
 - D. All of the above
6. Any data that can be stored, accessed and processed in the form of fixed format is termed as a?
- A. Structured
 - B. Unstructured
 - C. Semi- Structured
 - D. All of the above
7. What does velocity in big data mean?
- A. Speed of input data generation
 - B. Speed of individual machine processors
 - C. Speed of only storing data.
 - D. Speed of storing and processing of data
8. Which one is false about big data analytics?
- A. It collects data

Unit 04: Introduction to Big Data

- B. It looks for pattern
C. It analyses data
D. It does not organize data
9. In MapReduce architecture, _____ involves taking in data, converting it into a set of other data where the breakdown of individual elements into tuples is done.
A. Hadoop() function
B. Shuffle() function
C. Map() function
D. Reduce() function
10. _____ is a resource management platform responsible for managing compute resources in the cluster and using them in order to schedule users and applications in Hadoop.
A. Hadoop Common
B. HDFS
C. Hadoop YARN
D. Hadoop Mapper
11. _____ facilitates ability to manage, analyze, summarize, visualize & discover knowledge from the collected data in a timely manner & in a scalable fashion at a faster pace.
A. Big Data
B. Data Summarization
C. Text Summarization
D. Knowledge Summarization
12. Which of the following is/are example(s) of unstructured data?
A. Sensor data
B. Log file
C. Web data
D. None of the above
13. Veracity in big data is concerned with:
A. Accuracy and correctness of the data relative to a particular use
B. Schemas, semantics, ontologies, taxonomies, vocabularies
C. Data has meaningful relations to questions of interest
D. Biases, noise and abnormality in data, trustworthiness of data
14. _____ refers to the connectedness of big data.
A. Variety
B. Valence
C. Velocity
D. Volume
15. Which of the following stage in big data cycle defines which type of profiles would be needed to deliver resultant data product?

Cloud Computing

- A. Data Munging
- B. Data Acquisition
- C. Data Operations
- D. Resource Assessments

Answers for Self Assessment

- | | | | | |
|-------|-------|-------|-------|-------|
| 1. A | 2. D | 3. A | 4. B | 5. D |
| 6. A | 7. D | 8. C | 9. C | 10. C |
| 11. A | 12. A | 13. D | 14. B | 15. B |

Review Questions

1. Explain the data analysis techniques in Big data?
2. What are the different data analysis tools in Big data?
3. Explore the different phases in the Big data analytics cycle?
4. What is Big data? Explain its characteristics?
5. Discuss the different V's in Big data?
6. How big data differs from the traditional database methods?
7. Distinguish Parallel, Distributed and Cluster processing?
8. Distinguish between Structured, Unstructured and Semi-structured data?
9. Discuss how MapReduce in Hadoop works?
10. Explain the MapReduce architecture?



Further Readings

Big Data and Analytics, Subhashini Chellappan Seema Acharya, Wiley Publications

Big Data: A Revolution That Will Transform How We Live, Work, and Think by Viktor Mayer-Schonberger and Kenneth Cukier. ISBN-10: 9780544002692



Web Links

What is BIG DATA? Introduction, Types, Characteristics, Example (guru99.com)

Big Data Definition (investopedia.com)

What is Big Data? - GeeksforGeeks

What is MapReduce in Hadoop? Architecture | Example (guru99.com)

MapReduce Architecture - GeeksforGeeks

Unit 05: File system in Cloud

CONTENTS

Objectives

Introduction

5.1 GFS Assumptions

5.2 GFS Interface

5.3 GFS Architecture

5.4 GFS Operation

5.5 Distributed File System (DFS)

5.6 Hadoop Distributed File System (HDFS)

5.7 The File System Namespace

5.8 HDFS Operations

5.9 Comparison- GFS v/s HDFS

Summary

Keywords

Self Assessment

Answers for Self Assessment

Review Questions

Further Readings

Objectives

After this lecture, you will be able to,

- learn about Google File System (GFS).
- know the architecture and operations of GFS.
- explore the Hadoop Distributed File System (HDFS).
- understand the architecture and operations of HDFS.
- discover the comparison of GFS and HDFS.

Introduction

Google File System (GFS) is a scalable distributed file system for large distributed data-intensive applications that provides fault tolerance while running on inexpensive commodity hardware, and it delivers high aggregate performance to a large number of clients. In order to meet the rapidly growing demands of Google's data processing needs, GFS has successfully met the present storage needs. It is widely deployed within Google as the storage platform for the generation and processing of data used by the service as well as research and development efforts that require large data sets.

GFS is the largest cluster to date, provides hundreds of terabytes of storage across thousands of disks on over a thousand machines, and it is concurrently accessed by hundreds of clients. GFS shares many of the same goals as previous distributed file systems such as performance, scalability, reliability, and availability. However, its design has been driven by key observations of the application workloads and technological environment, both current and anticipated, that reflect a marked departure from some earlier file system design assumptions. The initial exploration on design space before implementing GFS include: First, component failures are the norm rather than the exception. The file system consists of hundreds or even thousands of storage machines built

from inexpensive commodity parts and is accessed by a comparable number of client machines. The quantity and quality of the components virtually guarantee that some are not functional at any given time and some will not recover from their current failures. There exist problems caused by application bugs, OS bugs, human errors, and the failures of disks, memory, connectors, networking, and power supplies. Therefore, constant monitoring, error detection, fault tolerance, and automatic recovery must be integral to the system.

Second, the files are huge by traditional standards. Multi-GB files are common. Each file typically contains many application objects such as web documents. When we are regularly working with fast growing data sets of many TBs comprising billions of objects, it is unwieldy to manage billions of approximately KB-sized files even when the file system could support it. As a result, design assumptions and parameters such as I/O operation and block sizes have to be revisited.

Third, most files are mutated by appending new data rather than overwriting existing data. Random writes within a file are practically non-existent. Once written, the files are only read, and often only sequentially. A variety of data share these characteristics. Some may constitute large repositories that data analysis programs scan through. Some may be data streams continuously generated by running applications.

Some may be archival data. Some may be intermediate results produced on one machine and processed on another, whether simultaneously or later in time. Given this access pattern on huge files, appending becomes the focus of performance optimization and atomicity guarantees, while caching data blocks in the client loses its appeal.

Fourth, co-designing the applications and the file system API benefits the overall system by increasing the flexibility. For example, the relaxed GFS's consistency model to vastly simplify the file system without imposing an onerous burden on the applications is available. Additionally, it supports an atomic append operation so that multiple clients can append concurrently to a file without extra synchronization between them.

The multiple GFS clusters are currently deployed for different purposes. The largest ones have over 1000 storage nodes, over 300 TB of disk storage, and are heavily accessed by hundreds of clients on distinct machines on a continuous basis.

5.1 GFS Assumptions

- **Hardware failures are common (commodity hardware)**-The system is built from many inexpensive commodity components that often fail. It must constantly monitor itself and detect, tolerate, and recover promptly from component failures on a routine basis.
- **Files are large (GB/TB) and their number is limited (millions, not billions)**-The system stores a modest number of large files. Multi-GB files are the common case and should be managed efficiently. Small files must be supported, but there is no need to optimize for them.
- **Two main types of reads: large streaming reads and small random reads!**-The workloads primarily consist of two kinds of reads: large streaming reads and small random reads. In large streaming reads, individual operations typically read hundreds of KBs, more commonly 1 MB or more.
- **Workloads with sequential writes that append data to files**- The successive operations from the same client often read through a contiguous region of a file. A small random read typically reads a few KBs at some arbitrary offset. Performance-conscious applications often batch and sort their small reads to advance steadily through the file rather than go back and forth.
- **Once written, files are seldom modified. (!=append) again**-The workloads have many large, sequential writes that append data to files. Typical operation sizes are similar to those for reads. Once written, files are seldom modified again. Small writes at arbitrary positions in a file are supported but do not have to be efficient.
- **Random modification in files possible, but not efficient in GFS**-The system must efficiently implement well-defined semantics for multiple clients that concurrently append to the same file. Our files are often used as producer-consumer queues or for many-way merging. Hundreds

of producers, running one per machine, will concurrently append to a file. Atomicity with minimal synchronization overhead is essential. The file may be read later, or a consumer may be reading through the file simultaneously.

- o **High sustained bandwidth trumps low latency**-The high sustained bandwidth is more important than low latency. Most of our target applications place a premium on processing data in bulk at a high rate, while few have stringent response time requirements for an individual read or write.

5.2 GFS Interface

GFS provides a familiar file system interface, though it does not implement a standard API such as POSIX. Files are organized hierarchically in directories and identified by pathnames. It supports usual operations to create, delete, open, close, read, and write files. GFS has snapshot and record append operations. The snapshot creates a copy of a file or a directory tree at low cost. Record append allows multiple clients to append data to the same file concurrently while guaranteeing the atomicity of each individual client's append. It is useful for implementing multi-way merge results and producer-consumer queues that many clients can simultaneously append to without additional locking. Such types of files are invaluable in building large distributed applications.

5.3 GFS Architecture

GFS cluster consists of a single master and multiple chunkservers and is accessed by multiple clients. Each of these is typically a commodity Linux machine running a user-level server process. It is easy to run both a chunkserver and a client on the same machine, as long as machine resources permit and the lower reliability caused by running possibly flaky application code is acceptable. The files are divided into fixed-size chunks. Each chunk is identified by an immutable and globally unique 64-bit chunk handle assigned by the master at the time of chunk creation. The chunkservers store chunks on local disks as Linux files and read or write chunk data specified by a chunk handle and byte range. For reliability, each chunk is replicated on multiple chunkservers. By default, three replicas are stored, though users can designate different replication levels for different regions of the file namespace (Figure 1).

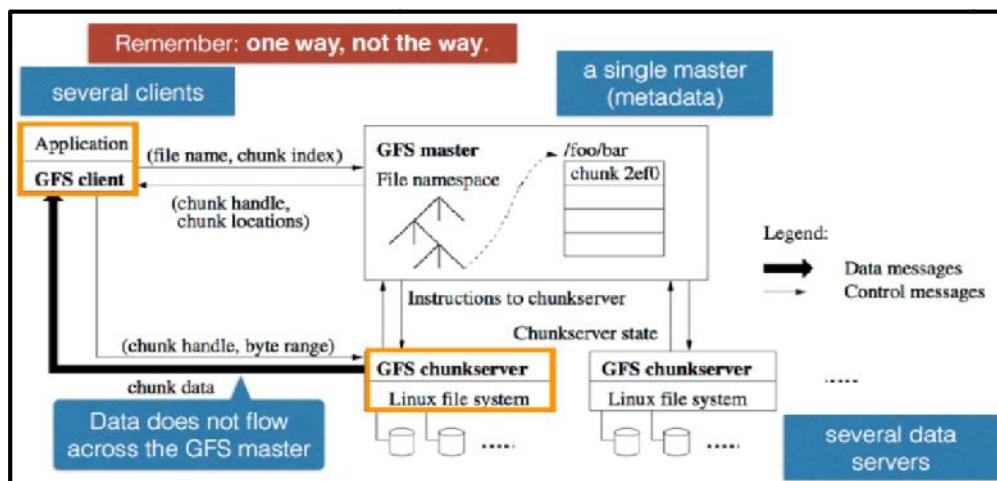


Figure 1: GFS Architecture

Single Master-Having a single master vastly simplifies GFS design and enables the master to make sophisticated chunk placement and replication decisions using global knowledge. However, it is important to minimize its involvement in reads and writes so that it does not become a bottleneck. Clients never read and write file data through the master. Instead, a client asks the master which chunkservers it should contact. It caches this information for a limited time and interacts with the chunkservers directly for many subsequent operations. It maintains all file system metadata. This includes the namespace, access control information, the mapping from files to chunks, and the current locations of chunks. It controls system-wide activities such as chunk lease management, garbage collection of orphaned chunks, and chunk migration between

chunkservers. Periodically communicates with each chunkserver in HeartBeat messages to give it instructions and collect its state.

Client Code-GFS client code linked into each application implements the file system API and communicates with the master and chunkservers to read or write data on behalf of the application. Clients interact with the master for metadata operations, but all data-bearing communication goes directly to the chunkservers. We do not provide the POSIX API and therefore need not hook into the Linux vnode layer.

Neither the client nor the chunkserver caches file data. Client caches offer little benefit because most applications stream through huge files or have working sets too large to be cached. Not having them simplifies the client and the overall system by eliminating cache coherence issues. (Clients do cache metadata, however.) Chunkservers need not cache file data because chunks are stored as local files and so Linux's buffer cache already keeps frequently accessed data in memory.

Chunk Size- Chunk size is one of the key design parameters. Each chunk replica is stored as a plain Linux file on a chunkserver and is extended only as needed. Lazy space allocation avoids wasting space due to internal fragmentation, perhaps the greatest objection against such a large chunk size.

Metadata- The master stores three major types of metadata: the file and chunk namespaces, the mapping from files to chunks, and the locations of each chunk's replicas. All metadata is kept in the master's memory. The first two types (namespaces and file-to-chunk mapping) are also kept persistent by logging mutations to an operation log stored on the master's local disk and replicated on remote machines. Using a log allows us to update the master state simply, reliably, and without risking inconsistencies in the event of a master crash. The master does not store chunk location information persistently. Instead, it asks each chunkserver about its chunks at master startup and whenever a chunkserver joins the cluster.

In-Memory Data Structures-Since metadata is stored in memory, master operations are fast. Furthermore, it is easy and efficient for the master to periodically scan through its entire state in the background. This periodic scanning is used to implement chunk garbage collection, re-replication in the presence of chunkserver failures, and chunk migration to balance load and disk space usage across chunkservers. One potential concern for this memory-only approach is that the number of chunks and hence the capacity of the whole system is limited by how much memory the master has. This is not a serious limitation in practice. The master maintains less than 64 bytes of metadata for each 64 MB chunk. Most chunks are full because most files contain many chunks, only the last of which may be partially filled. Similarly, the file namespace data typically requires less than 64 bytes per file because it stores file names compactly using prefix compression. If necessary, to support even larger file systems, the cost of adding extra memory to the master is a small price to pay for the simplicity, reliability, performance, and flexibility gained by storing the metadata in memory.

Chunk Locations- The master does not keep a persistent record of which chunkservers have a replica of a given chunk. It simply polls chunkservers for that information at startup. The master can keep itself up-to-date thereafter because it controls all chunk placement and monitors chunkserver status with regular HeartBeat messages.

Operation Log-Operational log is central to GFS and contains a historical record of critical metadata changes. Not only is it the only persistent record of metadata, but it also serves as a logical time line that defines the order of concurrent operations. Files and chunks, as well as their versions, are all uniquely and eternally identified by the logical times at which they were created.

5.4 GFS Operation

Firstly, using the fixed chunk size, the client translates the file name and byte offset specified by the application into a chunk index within the file. Then, it sends the master a request containing the file name and chunk index. The master replies with the corresponding chunk handle and locations of the replicas. The client caches this information using the file name and chunk index as the key. The client then sends a request to one of the replicas, most likely the closest one. The request specifies the chunk handle and a byte range within that chunk. Further reads of the same chunk require no more client-master interaction until the cached information expires or the file is reopened. In fact, the client typically asks for multiple chunks in the same request and the master can also include the

information for chunks immediately following those requested. This extra information sidesteps several future client-master interactions at practically no extra cost.

GFS has a relaxed consistency model that supports the highly distributed applications well but remains relatively simple and efficient to implement.

Detailed GFS Master Operation

The master executes all namespace operations. In addition, it manages chunk replicas throughout the system. It makes placement decisions, creates new chunks and hence replicas, and coordinates various system-wide activities to keep chunks fully replicated, to balance load across all the chunkservers, and to reclaim unused storage.

Namespace Management and Locking: Many master operations can take a long time: for example, a snapshot operation has to revoke chunkserver leases on all chunks covered by the snapshot. If you don't want to delay other master operations while they are running, then allow multiple operations to be active and use locks over regions of the namespace to ensure proper serialization. Unlike many traditional file systems, GFS does not have a per-directory data structure that lists all the files in that directory. Nor does it support aliases for the same file or directory (that is, hard or symbolic links in Unix terms). GFS logically represents its namespace as a lookup table mapping full pathnames to metadata. With prefix compression, this table can be efficiently represented in memory. Each node in the namespace tree (either an absolute file name or an absolute directory name) has an associated read-write lock. Each master operation acquires a set of locks before it runs.

Replica Placement: The GFS cluster is highly distributed at more levels than one. It typically has hundreds of chunkservers spread across many machine racks. These chunkservers, in turn, may be accessed from hundreds of clients from the same or different racks. Communication between two machines on different racks may cross one or more network switches. Additionally, bandwidth into or out of a rack may be less than the aggregate bandwidth of all machines within rack. The multi-level distribution presents a unique challenge to distribute data for scalability, reliability, and availability. The chunk replica placement policy serves two purposes: maximize data reliability and availability, and maximize network bandwidth utilization.

Creation, Re-replication, Rebalancing: Chunk replicas are created for three reasons: chunk creation, re-replication, and rebalancing. When the master creates a chunk, it chooses where to place the initially empty replicas. It considers several factors:

- You want to place new replicas on chunkservers with below-average disk space utilization. Over time this will equalize disk utilization across chunkservers.
- In case you want to limit the number of “recent” creations on each chunkserver, although creation itself is cheap, it reliably predicts imminent heavy write traffic because chunks are created when demanded by writes, and in our append-once-read-many workload they typically become practically read-only once they have been completely written. You want to spread replicas of a chunk across racks.

The master re-replicates a chunk as soon as the number of available replicas falls below a user-specified goal. This could happen for various reasons: a chunkserver becomes unavailable, it reports that its replica may be corrupted, one of its disks is disabled because of errors, or the replication goal is increased. Finally, in order to minimize the impact of failures on running applications, the boosting of the priority of any chunk that is blocking client progress is done.

5.5 Distributed File System (DFS)

A set of client and server services that allow an organization using Microsoft Windows servers to organize many distributed SMB file (Server Message Block (SMB) is a networking file share protocol included in Windows 10 that provides the ability to read and write files and perform other service requests to network devices.) shares into a distributed file system.

The two components to its service: Location transparency (via the namespace component) and Redundancy (via the file replication component). These components improve data availability in the case of failure or heavy load by allowing shares in multiple different locations to be logically

grouped under one folder, the "DFS root". Microsoft's DFS is referred to interchangeably as 'DFS' and 'Dfs' by Microsoft and is unrelated to the DCE Distributed File System, which held the 'DFS' trademark but was discontinued in 2005. It is also called "MS-DFS" or "MSDFS" in some contexts, example in the Samba user space project. The DFS root can only exist on a server version of Windows (from Windows NT 4.0 and up) and OpenSolaris (in kernel space) or a computer running Samba (in user space).

The Enterprise and Datacenter Editions of Windows Server can host multiple DFS roots on the same server. OpenSolaris intends on supporting multiple DFS roots in "a future project based on Active Directory (AD) domain-based DFS namespaces".

Two ways of implementing DFS on a server:

- Standalone DFS Namespace- Allows for a DFS root that exists only on the local computer, and thus does not use Active Directory. A Standalone DFS can only be accessed on the computer on which it is created. It does not offer any fault tolerance and cannot be linked to any other DFS. This is the only option available on Windows NT 4.0 Server systems. Standalone DFS roots are rarely encountered because of their limited utility.
- Domain-based DFS Namespace- Stores the DFS configuration in Active Directory, making the DFS namespace root accessible at

`\<domainname>\<dfsroot>`

or

`\<FQDN>\<dfsroot>`

5.6 Hadoop Distributed File System (HDFS)

Hadoop Distributed File System (HDFS) is the major component of Apache Hadoop. It handles large data sets running on commodity hardware. It allows multiple files to be stored and retrieved at the same time at an unprecedented speed. HDFS is used to scale a single Apache Hadoop cluster to hundreds (and even thousands) of nodes. HDFS should not be confused with or replaced by Apache HBase, which is a column-oriented non-relational database management system that sits on top of HDFS and can better support real-time data needs with its in-memory processing engine. The servers in HDFS are fully connected, and they communicate through TCP-based protocols. Though designed for huge databases, normal file systems (FAT, NTFS, etc.) can also be viewed. The current status of a node is obtained through the checkpoint node.

Table 1 lists the different features of HDFS.

Goals of HDFS

- Fast Recovery from Hardware Failures- Because one HDFS instance may consist of thousands of servers, failure of at least one server is inevitable. HDFS has been built to detect faults and automatically recover quickly.
- Access to Streaming Data- HDFS is intended more for batch processing versus interactive use, so the emphasis in the design is for high data throughput rates, which accommodate streaming access to data sets.
- Accommodation of Large Data Sets- HDFS accommodates applications that have data sets typically gigabytes to terabytes in size. HDFS provides high aggregate data bandwidth and can scale to hundreds of nodes in a single cluster.
- Portability- To facilitate adoption, HDFS is designed to be portable across multiple hardware platforms and to be compatible with a variety of underlying operating systems.
- Simple Coherency Model: HDFS applications need a write-once-read-many access model for files. A file once created, written, and closed need not be changed. This assumption simplifies data coherency issues and enables high throughput data access. A MapReduce application or a web crawler application fits perfectly with this model. There is a plan to support appending-writes to files in the future.



Task: Explore Hadoop Distributed File System and various assumptions for the same.

Table 1: Key Features of HDFS

HDFS Key Features	Description
Storing bulks of data	HDFS is capable of storing terabytes and petabytes of data.
Minimum intervention	It manages thousands of nodes without operators' intervention.
Computing	HDFS provides the benefits of distributed and parallel computing at once.
Scaling out	It works on scaling out, rather than on scaling up, without a single downtime.
Rollback	HDFS allows returning to its previous version post an upgrade.
Data integrity	It deals with corrupted data by replicating it several times.

"Moving Computation is Cheaper than Moving Data"

A computation requested by an application is much more efficient if it is executed near the data it operates on. This is especially true when the size of the data set is huge. This minimizes network congestion and increases the overall throughput of the system. The assumption is that it is often better to migrate the computation closer to where the data is located rather than moving the data to where the application is running. HDFS provides interfaces for applications to move themselves closer to where the data is located.

An Example of HDFS

Consider a file that includes the phone numbers for everyone in the United States; the numbers for people with a last name starting with A might be stored on server 1, B on server 2, and so on. With Hadoop, pieces of this phonebook would be stored across the cluster, and to reconstruct the entire phonebook, your program would need the blocks from every server in the cluster. In order to ensure availability if and when a server fails, HDFS replicates these smaller pieces onto two additional servers by default. (The redundancy can be increased or decreased on a per-file basis or for a whole environment; for example, a development Hadoop cluster typically doesn't need any data redundancy). This redundancy offers multiple benefits, the most obvious being higher availability. The redundancy also allows the Hadoop cluster to break up work into smaller chunks and run those jobs on all the servers in the cluster for better scalability. Finally, you gain the benefit of data locality, which is critical when working with large data sets.

IBM and Cloudera have partnered to offer an industry-leading, enterprise-grade Hadoop distribution, including an integrated ecosystem of products and services to support faster analytics at scale.



Notes: Why Should You Use HDFS?

- ✓ HDFS stores the same data in multiple sets-
- ✓ Distributed across hundreds or even thousands of servers with each node storing a part of the file system. Since the storage is done on commodity hardware, there are more chances of the node failing and, with that, the data can be lost. HDFS gets over that problem by storing the same data in multiple sets.

- ✓ HDFS works quite well for data loads that come in a streaming format. So, it is more suited for batch processing applications rather than for interactive use. It is important to note that HDFS works for high throughput rather than low latency.
- ✓ HDFS works exclusively well for large datasets, and the standard size of datasets could be anywhere between gigabytes and terabytes. It provides high-aggregate data bandwidth, and it is possible to scale hundreds of nodes in a single cluster. Hence, millions of files are supported in a single instance.
- ✓ HDFS works on the assumption that moving of computation is much easier, faster, and cheaper than moving of data of humongous size, which can create network congestion and lead to longer overall turnaround times.
- ✓ HDFS provides the facility to let applications access data at the place where they are located.
- ✓ HDFS is highly profitable in the sense that it can easily work on commodity hardware that are of different types without any issue of compatibility. Hence, it is very well suited for taking advantage of cheaply and readily available commodity hardware components.

HDFS Architecture

HDFS follows the master-slave data architecture (Figure 2). Each cluster comprises a single Namenode that acts as the master server in order to manage the file system namespace and provide the right access to clients. Another component in the HDFS cluster is- Datanode, usually one per node in the HDFS cluster. The Datanode is assigned with the task of managing the storage attached to the node that it runs on. HDFS also includes a file system namespace that is being executed by the Namenode for general operations like file opening, closing, and renaming, and even for directories. The Namenode also maps the blocks to Datanodes.

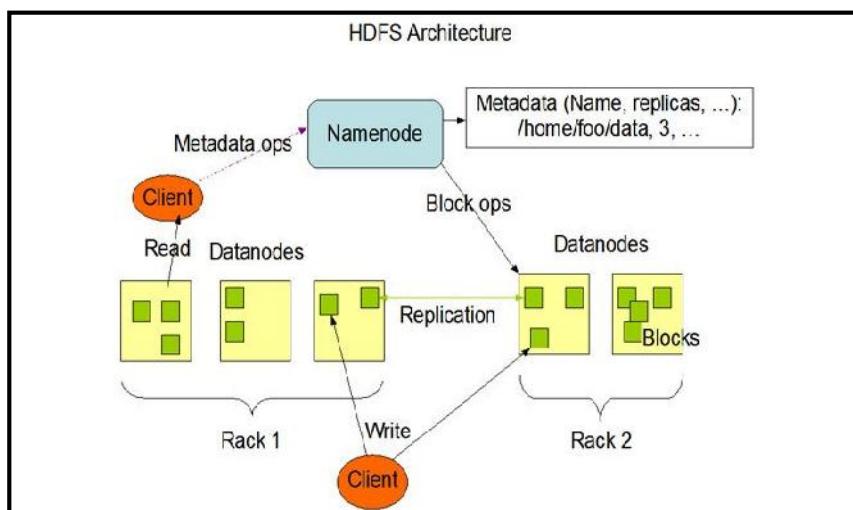


Figure 2: HDFS Architecture

Namenode and Datanode are actually Java programming codes that can be extensively run on the commodity hardware machines. Entire HDFS is based on the Java programming language. These machines could most probably be running on Linux OS or GNU OS. The single Namenode on a Hadoop cluster centralizes all the arbitration and repository-related issues without creating any ambiguity.

- a) Namenode:** Namenode is the commodity hardware that contains the GNU/Linux operating system and the namenode software. It is a software that can be run on commodity hardware. The system having the namenode acts as the master server and it does the following tasks: manages the file system namespace; regulates client's access to files; and executes file system operations such as renaming, closing, and opening files and directories.

- b) **Datanode:** Datanode is a commodity hardware having the GNU/Linux OS and datanode software. For every node (Commodity hardware/System) in a cluster, there will be a datanode that manages the data storage of their system. Datanodes perform read-write operations on the file systems, as per client request. It performs the operations such as block creation, deletion, and replication according to the instructions of the namenode.
- c) **Blocks:** HDFS stores a file in a sequence of blocks. It is easy to configure the block size and the replication factor. The blocks of files are replicated in order to ensure that there is always fault tolerance. It is possible to specify the required number of replicas of a certain file. All the files have write-once and read-multiple-times format. Namenode is the deciding authority when it comes to issuing the number of replication blocks needed.

Data Organization in HDFS

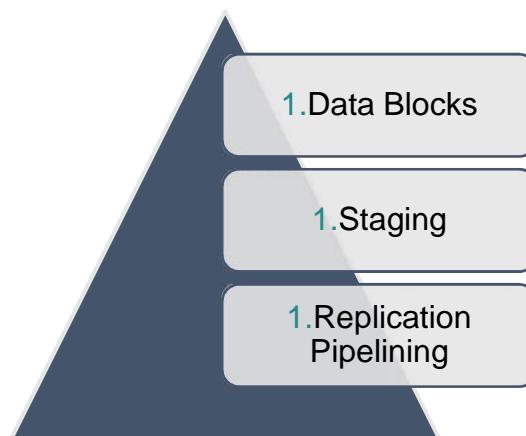


Figure 3: Data Organization in HDFS

HDFS data organization has been depicted in the Figure 3.

- **Data Blocks:** HDFS is designed to support very large files. Applications that are compatible with HDFS are those that deal with large data sets. These applications write their data only once but they read it one or more times and require these reads to be satisfied at streaming speeds. HDFS supports write-once-read-many semantics on files. A typical block size used by HDFS is 64 MB. Thus, an HDFS file is chopped up into 64 MB chunks, and if possible, each chunk will reside on a different DataNode.
- **Staging:** A client request to create a file does not reach the NameNode immediately. Initially, HDFS client caches the file data into a temporary local file. Application writes are transparently redirected to this temporary local file. When the local file accumulates data worth over one HDFS block size, the client contacts the NameNode. The NameNode inserts the file name into the file system hierarchy and allocates a data block for it. NameNode responds to the client request with the identity of the DataNode and the destination data block. Then the client flushes the block of data from the local temporary file to the specified DataNode. When a file is closed, the remaining un-flushed data in the temporary local file is transferred to the DataNode. The client then tells the NameNode that the file is closed. At this point, the NameNode commits the file creation operation into a persistent store. If the NameNode dies before the file is closed, the file is lost.
- **Replication Pipelining-** When a client is writing data to an HDFS file, its data is first written to a local file. Suppose the HDFS file has a replication factor of three. When the local file accumulates a full block of user data, the client retrieves a list of DataNodes from the

NameNode. This list contains the DataNodes that will host a replica of that block. The client then flushes the data block to the first DataNode. The first DataNode starts receiving the data in small portions (4 KB), writes each portion to its local repository and transfers that portion to the second DataNode in the list. The second DataNode, in turn starts receiving each portion of the data block, writes that portion to its repository and then flushes that portion to the third DataNode. Finally, the third DataNode writes the data to its local repository. Thus, a DataNode can be receiving data from the previous one in the pipeline and at the same time forwarding data to the next one in the pipeline. Thus, the data is pipelined from one DataNode to the next.

5.7 The File System Namespace

HDFS data platform format follows a strictly hierarchical file system. An application or a user first creates a directory, and there will be files within this directory. The file system hierarchy is identical to other file systems. It is possible to add or remove a file and also move files from a directory to another. A user can even rename files.

The data replication is an essential part of the HDFS format. Since it is hosted on the commodity hardware, it is expected that nodes can go down without any warning as such. So, the data is stored in a redundant manner in order to access it at any times.

HDFS does not yet implement user quotas. HDFS does not support hard links or soft links. The NameNode maintains the file system namespace. Any change to the file system namespace or its properties is recorded by the NameNode. An application can specify the number of replicas of a file that should be maintained by HDFS. The number of copies of a file is called the replication factor of that file. This information is stored by the NameNode.

- **Data Replication:** HDFS is designed to reliably store very large files across machines in a large cluster. It stores each file as a sequence of blocks; all blocks in a file except the last block are the same size (Figure 4). The blocks of a file are replicated for fault tolerance. The block size and replication factor are configurable per file. An application can specify the number of replicas of a file. The replication factor can be specified at file creation time and can be changed later. The files in HDFS are write-once and have strictly one writer at any time.

The NameNode makes all decisions regarding replication of blocks. It periodically receives a Heartbeat and a Blockreport from each of the DataNodes in the cluster. The receipt of a Heartbeat implies that the DataNode is functioning properly. A Blockreport contains a list of all blocks on a DataNode.

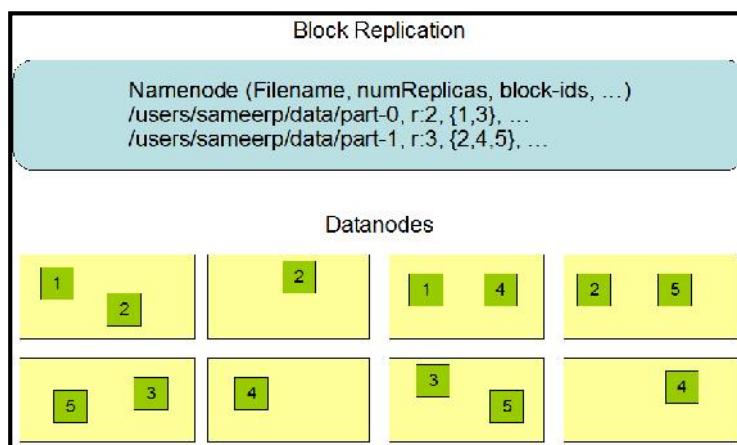


Figure 4: Data Replication in HDFS

- **Replica Placement:** The placement of replicas is critical to HDFS reliability and performance. Optimizing replica placement distinguishes HDFS from most other distributed file

systems. The purpose of a rack-aware replica placement policy is to improve data reliability, availability, and network bandwidth utilization and to validate it on production systems. Large HDFS instances run on a cluster of computers that commonly spread across many racks. The communication between two nodes in different racks has to go through switches. In most cases, network bandwidth between machines in the same rack is greater than network bandwidth between machines in different racks. The NameNode determines the rack id each DataNode belongs to via the process outlined in Hadoop Rack Awareness.

- **Replica Selection:** To minimize global bandwidth consumption and read latency, HDFS tries to satisfy a read request from a replica that is closest to the reader. If there exists a replica on the same rack as the reader node, then that replica is preferred to satisfy the read request. If the HDFS cluster spans multiple data centers, then a replica that is resident in the local data center is preferred over any remote replica.
- **Safemode:** On startup, the NameNode enters a special state called Safemode. The replication of data blocks does not occur when the NameNode is in the Safemode state. NameNode receives Heartbeat and Blockreport messages from the DataNodes. A Blockreport contains the list of data blocks that a DataNode is hosting. Each block has a specified minimum number of replicas. A block is considered safely replicated when the minimum number of replicas of that data block has checked in with the NameNode.

5.8 HDFS Operations

Starting HDFS

- Initially you have to format the configured HDFS file system, open namenode (HDFS server), and execute the following command:

```
$ hadoop namenode -format
```

- After formatting the HDFS, start the distributed file system. The following command will start the namenode as well as the data nodes as cluster:

```
$ start-dfs.sh
```

Listing Files in HDFS- Given below is the syntax of ls that you can pass to a directory or a filename as an argument.

```
$ $HADOOP_HOME/bin/hadoop fs -ls <args>
```

Inserting Data into HDFS- Follow the steps given below to insert the required file in the Hadoop file system:

- Step 1- You have to create an input directory.

```
$ $HADOOP_HOME/bin/hadoop fs -mkdir /user/input
```

- Step 2- Transfer and store a data file from local systems to the Hadoop file system using the put command.

```
$ $HADOOP_HOME/bin/hadoop fs -put /home/file.txt /user/input
```

- Step 3- You can verify the file using ls command.

```
$ $HADOOP_HOME/bin/hadoop fs -ls /user/input
```

Retrieving Data from HDFS- Assume we have a file in HDFS called outfile. Given below is a simple demonstration for retrieving the required file from the Hadoop file system.

- Step 1- Initially, view the data from HDFS using cat command.

```
$ $HADOOP_HOME/bin/hadoop fs -cat /user/output/outfile
```

- Step 2- Get the file from HDFS to the local file system using get command.

```
$ $HADOOP_HOME/bin/hadoop fs -get /user/output/ /home/hadoop_tp/
```

Shutting down HDFS- You can shut down the HDFS by using the following command:

```
$ stop-dfs.sh
```

HDFS communication protocols are layered on top of the TCP/IP protocol. The client establishes a connection to a configurable TCP port on the NameNode machine. It talks the ClientProtocol with the NameNode. DataNodes talk to the NameNode using the DataNodeProtocol. A Remote Procedure Call (RPC) abstraction wraps both the Client Protocol and the DataNode Protocol. The primary objective of HDFS is to store data reliably even in the presence of failures. The three common types of failures are-

- NameNode failures
- DataNode failures and
- Network partitions



Lab Exercise: Install Hadoop and implement it.

Advantages of Using HDFS

- HDFS is by far the most resilient and fault-tolerant technology that is available as an open-source platform, which can be scaled up or scaled down depending on the needs, making it really hard for finding an HDFS replacement for Big Data Hadoop storage needs.
- Some of the biggest enterprises on earth are deploying Hadoop in unprecedented scales, and things can only get better in the future. Companies like Amazon, Facebook, Google, Microsoft, Yahoo, General Electrics, and IBM run massive Hadoop clusters in order to parse their inordinate amounts of data.
- Distributed across hundreds or even thousands of servers with each node storing a part of the file system. Since the storage is done on commodity hardware, there are more chances of the node failing and, with that, the data can be lost. HDFS gets over that problem by storing the same data in multiple sets.
- HDFS works quite well for data loads that come in a streaming format. So, it is more suited for batch processing applications rather than for interactive use. It is important to note that HDFS works for high throughput rather than low latency.
- HDFS works exclusively well for large datasets, and the standard size of datasets could be anywhere between gigabytes and terabytes. It provides high-aggregate data bandwidth, and it is possible to scale hundreds of nodes in a single cluster. Hence, millions of files are supported in a single instance.
- HDFS works on the assumption that moving of computation is much easier, faster, and cheaper than moving of data of humongous size, which can create network congestion and lead to longer overall turnaround times.
- Accessibility- HDFS provides the facility to let applications access data at the place where they are located.
- HDFS is highly profitable in the sense that it can easily work on commodity hardware that are of different types without any issue of compatibility. Hence, it is very well suited for taking advantage of cheaply and readily available commodity hardware components.

- As a forward-thinking IT professional, this technology can help you leave your competitors way behind and make a big leap in your career.

5.9 Comparison- GFS v/s HDFS

HDFS is a simplified version of GFS.

Similarities-

- Master and Slaves: Both GFS and HDFS use single master + multiple slaves mode. The master node maintains the check-point, data migration, log.
- Data blocks and replication: Both maintain multiple copies (usually 3) to support better reliability and performance.
- Tree structure: Maintains a tree-structure file system, and allows operations like those under Linux system such as copy, rename, move, copy, delete etc.

Dissimilarities

- File Appends: GFS allows multiple appends and allows multiple clients to append simultaneously. If every append will visit the master node, it will be of low efficiency. GFS uses "Leasing Mechanism" to deliver the write permission of Chunk to Chunk Server. Check server can write the chunks within the lease. Since multiple servers may write simultaneously, and the API is asynchronous, the records might be in different order. This makes the system design very complicated.

HDFS only allows one open and data append. The client will first write the data in local tmp file, and when the size of tmp data reaches the size of a chunk (64M), then it will ask the HDFS master to assign a machine and chunk number to write the Chunk data.

- Master Failure: In GFS, a backup master node exists. When the main master node fails, a new master node will be voted from the backup nodes. It supports snapshot by using "copy on write" approach. However, HDFS needs human-interactions in terms of failure. It does not support snapshot.
- Garbage Collection (GC): In GFS, a lazy GC exists. It will mark the files to be deleted (example- rename the file to one contains time information), thus the files will not be able to be visited by normal users. The master node will periodically check files & delete the outdated ones.

HDFS- HDFS uses simple and direct delete mechanism.

Summary

- GFS is the largest cluster to date, provides hundreds of terabytes of storage across thousands of disks on over a thousand machines, and it is concurrently accessed by hundreds of clients.
- GFS cluster consists of a single master and multiple chunk servers and is accessed by multiple clients. Each of these is typically a commodity Linux machine running a user-level server process.
- GFS has a relaxed consistency model that supports the highly distributed applications well but remains relatively simple and efficient to implement.
- The Enterprise and Datacenter Editions of Windows Server can host multiple DFS roots on the same server. OpenSolaris intends on supporting multiple DFS roots in "a future project based on Active Directory (AD) domain-based DFS namespaces".

- IBM and Cloudera have partnered to offer an industry-leading, enterprise-grade Hadoop distribution, including an integrated ecosystem of products and services to support faster analytics at scale.
- HDFS follows the master-slave data architecture. Each cluster comprises a single Namenode that acts as the master server in order to manage the file system namespace and provide the right access to clients.
- Namenode and Datanode are actually Java programming codes that can be extensively run on the commodity hardware machines. Entire HDFS is based on the Java programming language.
- HDFS is by far the most resilient and fault-tolerant technology that is available as an open-source platform, which can be scaled up or scaled down depending on the needs, making it really hard for finding an HDFS replacement for Big Data Hadoop storage needs.

Keywords

- ***Google File System (GFS)***: Google File System (GFS) is a scalable distributed file system for large distributed data-intensive applications that provides fault tolerance while running on inexpensive commodity hardware, and it delivers high aggregate performance to a large number of clients.
- ***GFS Client Code***: GFS client code linked into each application implements the file system API and communicates with the master and chunkservers to read or write data on behalf of the application
- ***Chunk Size***: Chunk size is one of the key design parameters. Each chunk replica is stored as a plain Linux file on a chunkserver and is extended only as needed.
- ***Operation Log***: Operational log is central to GFS and contains a historical record of critical metadata changes.
- ***Distributed File System (DFS)***: Distributed File System (DFS) is a set of client and server services that allow an organization using Microsoft Windows servers to organize many distributed SMB file shares into a distributed file system.
- ***Hadoop Distributed File System (HDFS)***: Hadoop Distributed File System (HDFS) is the major component of Apache Hadoop. It handles large data sets running on commodity hardware. It allows multiple files to be stored and retrieved at the same time at an unprecedented speed.
- ***Datanode in HDFS***: Datanode is a commodity hardware having the GNU/Linux OS and datanode software. For every node (Commodity hardware/System) in a cluster, there will be a datanode that manages the data storage of their system.

Self Assessment

1. Which of the following is an important characteristic of Google File System?
 - Scalable distributed file system for large distributed data-intensive applications.
 - Provides fault tolerance while running on inexpensive commodity hardware.
 - To meet the rapidly growing demands of Google's data processing needs.
 - All of the above
2. "System is built from many inexpensive commodity components that often fail" pertains which of the assumption for GFS?
 - Files are large (GB/TB) and their number is limited
 - Hardware failures are common

- C. Workloads with sequential writes that append data to files
D. None of the above
3. As per the Google file system assumption, the high sustained bandwidth is more important than _____ latency.
A. high
B. low
C. dual
D. tri
4. Which type of metadata is stored by the Master in GFS architecture?
A. the file and chunk namespaces
B. the mapping from files to chunks
C. the locations of each chunk's replicas
D. all of the above
5. In GFS interface, _____ creates a copy of a file or a directory tree at low cost.
A. snapshot
B. gutter
C. fieldname
D. directorylist
6. In GFS architecture, the files are divided into _____ chunks.
A. variable-size
B. double-size
C. single-size
D. fixed-size
7. Which of the following component in GFS architecture maintains all file system metadata?
A. Chunk locations
B. Master
C. Client code
D. Operation log
8. The GFS architecture component, the _____ contains a historical record of critical metadata changes.
A. chunk location
B. client code
C. operation log
D. in-memory data structures
9. GFS is not implemented in the kernel of an operating system, but is instead provided as a user space library.
A. True
B. False

10. GFS provides capability for _____ by constant monitoring, replicating crucial data, and fast and automatic recovery.
- A. scalability
 - B. elasticity
 - C. virtualization
 - D. fault tolerance
11. SMB stands for
- A. Software Message Block
 - B. Software Master Block
 - C. Server Master Block
 - D. Server Message Block
12. A _____ can only be accessed on the computer on which it is created. It does not offer any fault tolerance and cannot be linked to any other DFS.
- A. standalone DFS
 - B. domain-based DFS
 - C. namespace DFS
 - D. fault-tolerant DFS
13. Servers in HDFS are _____ connected, and they communicate through TCP-based protocols.
- A. loosely
 - B. partially
 - C. semi-loosely
 - D. fully
14. HDFS applications need a _____ access model for files.
- A. write-once-read-once
 - B. write-once-read-many
 - C. write-many-read-many
 - D. write-many-read-once
15. Namenode and Datanode in HDFS architecture are _____ programming codes that can be extensively run on the commodity hardware machines.
- A. Python
 - B. C++
 - C. ASP.Net
 - D. Java

Answers for Self Assessment

- | | | | | |
|------|------|------|------|-------|
| 1. D | 2. B | 3. B | 4. D | 5. A |
| 6. D | 7. B | 8. C | 9. A | 10. D |

11. D 12. A 13. D 14. B 15. D

Review Questions

1. Discuss the architecture of Google File System.
2. Compare GFS and HDFS?
3. How distributed file system works?
4. Discuss the operations of HDFS in detail.
5. Describe the data organization in HDFS.
6. What is Google File System? What assumptions it is based upon?
7. Elaborate the architecture of HDFS and role of each component?



Further Readings

Distributed Storage: Concepts, Algorithms, and Implementations, Yaniv Pessach, 1st edition, 2013.

Hadoop Succinctly, Elton Stoneman, 2016.

Hadoop- The Definitive Guide, Tom White, O'Reilly, Yahoo! Press, 2010.

Google File System- The Cookbook, Gerard Blokdyk, 2017 (Paperback).

Handbook of Cloud Computing, Borko Furht, Armando Escalante, Springer, 2010.



Web Links

[paper.dvi \(googleusercontent.com\)](#)

[What is Google File System \(GFS\)? - Definition from Techopedia](#)

[The Google File System - Google Research](#)

[Distributed File System \(Microsoft\) - Wikipedia](#)

[HDFS Architecture Guide \(apache.org\)](#)

[Hadoop - HDFS \(Hadoop Distributed File System\) - GeeksforGeeks](#)

[Hadoop - HDFS Overview \(tutorialspoint.com\)](#)

Unit 06: Collaborating using Google Cloud

CONTENTS

Objectives

Introduction

6.1 Cloud as Community Benefit

6.2 Cloud Collaboration Tools

6.3 Creating Word Documents in Collaboration

6.4 Google Docs

6.5 Getting Started with Google Sheets

6.6 Working with Google Forms

Summary

Keywords

Self Assessment

Answers for Self Assessment

Review Questions

Further Readings

Objectives

After this lecture, you will be able to,

Know about cloud collaboration.

Understand G-Suite and Google Drive Storage capabilities.

Learn and practically work on the cloud-based Word documents.

Empirically explore collaborating on sheets in cloud.

Learn about Google forms and practically work on collaborating forms in cloud.

Introduction

Cloud collaboration refers to the ability of workers to work together simultaneously on a particular task. It is a way of sharing and co-authoring computer files through the use of cloud computing. The cloud collaboration technologies allow users to upload, comment and collaborate on documents and even amend the document itself, evolving the document.

Businesses in the last few years have increasingly been switching to use of cloud collaboration. The document collaboration can be completed face to face. However, collaboration has become more complex, with the need to work with people all over the world in real time on a variety of different types of documents, using different devices. The documents are uploaded to a central "cloud" for storage, where they can then be accessed by others.

Cloud collaboration services include Google, Microsoft, WebEx, Jive Software, eXo Platform, Synaptop and Salesforce.com. A report mapped out five reasons why workers are reluctant to collaborate more:

- People resist sharing their knowledge.
- Safety issues
- Users are most comfortable using e-mail as their primary electronic collaboration tool.
- People do not have incentive to change their behavior.

Cloud Computing

- Teams that want to or are selected to use the software do not have strong team leaders who push for more collaboration.
- Senior management is not actively involved in or does not support the team collaboration initiative.

With the growth of the Internet, there was no need to limit group collaboration to a single enterprise's network environment. The users from multiple locations within a corporation, and from multiple organizations, desired to collaborate on projects that crossed company and geographic boundaries. 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. Cloud-collaboration is also termed as Internet-based group collaboration.

6.1 Cloud as Community Benefit

Cloud computing isn't just for home users. Cloud has tremendous benefits for the entire community, from neighborhood groups to sports teams to school organizations. Any time any group of people in the community need to communicate and collaborate, web-based applications are the way to go.

Communicating Across the Community

One of the key components of any community collaboration is communication. The communication on community issues is to use a web-based email program, such as Gmail (mail.google.com), Microsoft Windows Live Hotmail (mail.live.com), or Yahoo! Mail (mail.yahoo.com). These programs can be accessed from any computer connected to the Internet. You use your web browser to send and view email messages hosted on the web. You can send and receive messages at work, at home, or from wherever you happen to be. Everything you send and receive is stored in the cloud, accessible from anywhere at any time.

6.2 Cloud Collaboration Tools

Many cloud service providers created cloud collaboration tools-

- As a tool for collaboration internally between different departments within a firm, but also externally as a means for sharing documents with end-clients as receiving feedback.
- Includes the integration of email alerts into collaboration software and the ability to see who is viewing the document at any time.
- Tools group up- All the tools a team could need are put into one piece of software so workers no longer have to rely on email.
- Makes cloud computing a very versatile tool for firms with many different applications in a business environment.

The following are the features of cloud collaboration tools and services:

- All platforms can easily be accessed via a web browser.
- Offer apps for ease of access from a smartphone or tablet.
- Feature a directory structure similar to that of a computer drive; this facilitates navigation and organisation.
- Ease of Access: The access to the personal folders is perceived to be more cumbersome (involves 'more clicks'). Many people are always logged into Google (and hence Google Drive) in the background, both at home and at school.
- Online Editing: OneDrive and Google Drive offer the possibility of editing documents inside a web browser. There is no additional software needed. The folders or specific files can be shared with others; this facilitates collaboration.

Unit 06: Collaborating using Google Cloud

- Online Collaboration: Documents and folders can be shared with colleagues. Editing is possible without downloading the document, eliminating the need to email and save multiple versions of the same documents.
- Cloud Storage: The examples of cloud storage and their characteristics are depicted below (Figure 1). Example: Google Drive. The different features of Google Drive include (Figure 2):
 - ‘Pure’ cloud computing service, with all the apps & storage found online.
 - Can be used via desktop top computers, tablets like iPad or on smartphones.
 - All of Google’s services can be considered cloud -based: Gmail, Google Calendar, Google Voice etc.
 - Microsoft’s OneDrive: Similar to Google Drive

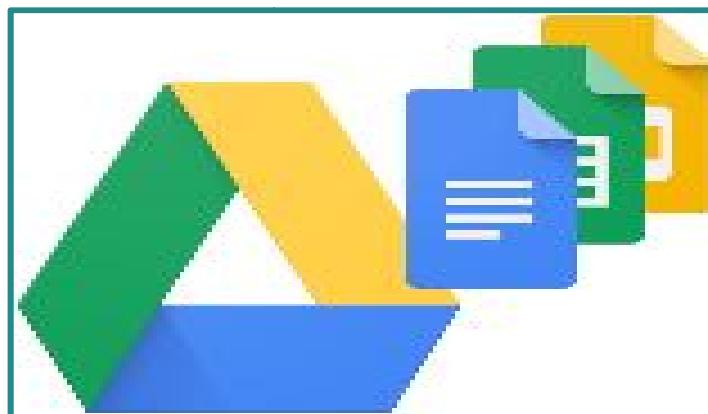


Figure 1: Examples of Cloud Storage

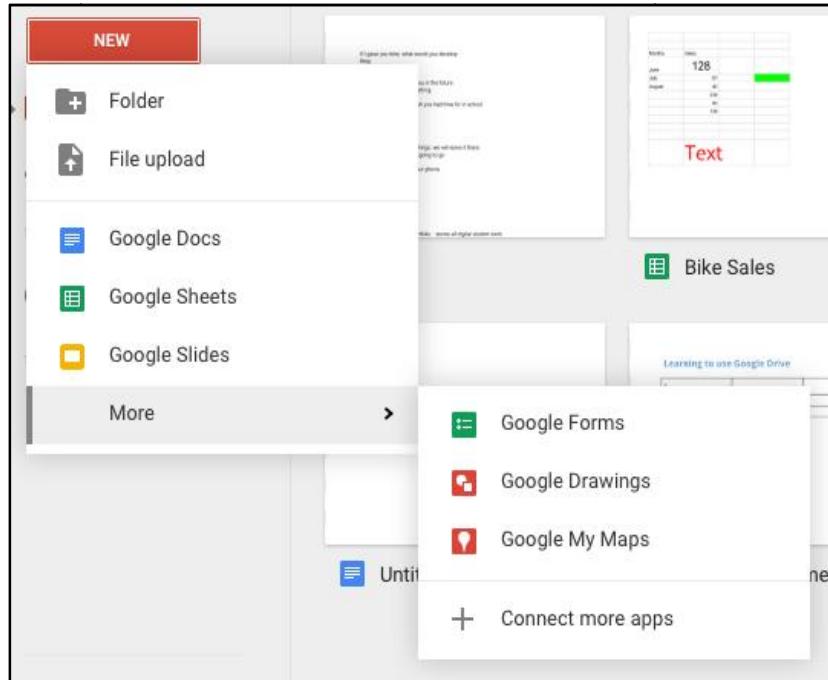
Cloud Computing

Figure 2: Google Drive Scenario

Google Drive is the new home for Google Docs. The setting up a free account and creating and editing documents using Google drive can be done. The users can create and share their work online and access their documents from anywhere. It is easy to manage documents, spreadsheets, presentations, surveys, and more all in one location. In order to access all of their files, folders, and Google Docs, a user has three options:

1. Sign into Google Drive on the web at drive.google.com.
2. Open the Google Drive folder on a computer (created when installing Google Drive for Mac or PC).
3. Go to the Google Drive mobile app on the Android or iOS device.

**Did you Know?**

Google Docs, Google Sheets, and Google Slides save to a secure, online storage facility. Thus, one can create documents, spreadsheets, presentations, and drawings without the need to save to a local hard drive. A user can also access his/her documents from any computer by signing into Google account. In the event of a local hard drive crash, a user won't lose the saved content.

This is a home for all your files with the Google Drive, you can also sync it to your particular personal computer, you can download all those files onto your particular personal computer as well. If I want to add any new Google Docs to my particular drive if I want to create it, I will simply click on this new on this drive fold part new and Google Docs. If I want to create a Google Sheets document, I can click on this Google Sheet. If I want to create a Google Slides, I can click on this Google Slides, if I want to click on this Google Forms, I can click on this Google form, so you can have additional capabilities that are listed out into this more part like Google drawings or the Google maps. There are so many different options that are available for actually listing out those particular contents that are listed that are given out into a particular drive as well, so we can make use of so many different capabilities that are listed out with the collaboration tools such as those offered through Google.

6.3 Creating Word Documents in Collaboration

A word processor is a software program capable of creating, storing, and printing documents. Unlike the standard typewriter, users using word processors have the ability of creating a document and making any changes anywhere in the document. This document can also be saved for modification at a later time or to be opened on any other computer using the same word processor. Today, the word processor is one of the most frequently used programs or online services used on a computer.

Web-based Word Processing

It involves the preparation of textual documents on computer. A word-processing system typically consists simply of a personal computer linked to a computer printer, but it may instead employ a terminal linked to a mainframe computer.



Caution: Word processing differs from typewriter typing in numerous ways.

The electronic text can be moved around at will; misspelled terms can be corrected throughout the document by means of a single command; spelling and grammar checkers can automatically alert the user to apparent errors of spelling, punctuation, and syntax; and the document's format, layout, and type fonts and sizes can be changed repeatedly until a satisfactory design is achieved.

Since all editing ideally occurs on-screen, word processing can result in decreased paper usage and simplified editing. When the final draft is ready, the document can be printed out (in multiple copies if necessary), sent as an email attachment, shared on a computer network, or simply stored as an electronic file.

Features of Standard Word Processors

File management: Many word processors contain file management capabilities that allow you to create, delete, move, and search for files.

Font specifications: Allows you to change fonts within a document. For example, you can specify bold, italics, and underlining. Most word processors also let you change the font size and even the typeface.

Footnotes and cross-references: Automate the numbering and placement of footnotes and enables you to easily cross-reference other sections of the document.

Graphics: Allows you to embed illustrations and graphs into a document. Some word processors let you create the illustrations within the word processor; others let you insert an illustration produced by a different program.

Headers, footers and page numbering: Allows you to specify customized headers and footers that the word processor will put at the top and bottom of every page.



Notes: The word processor automatically keeps track of page numbers so that the correct number appears on each page.

Layout: Allows you to specify different margins within a single document and to specify various methods for indenting paragraphs.

Macros: A macro is a character or word that represents a series of keystrokes. The keystrokes can represent text or commands. The ability to define macros allows you to save yourself a lot of time by replacing common combinations of keystrokes.

Merges: Allows you to merge text from one file into another file. This is particularly useful for generating many files that have the same format but different data.

Spell checker: A utility that allows you to check the spelling of words. It will highlight any words that it does not recognize.

Cloud Computing

Tables of contents and indexes: Allows you to automatically create a table of contents and index based on special codes that you insert in the document.

Thesaurus: A built-in thesaurus that allows you to search for synonyms without leaving the word processor.

Windows: Allows you to edit two or more documents at the same time. Each document appears in a separate window. This is particularly valuable when working on a large project that consists of several different files.

WYSIWYG (what you see is what you get): With WYSIWYG, a document appears on the display screen exactly as it will look when printed.

6.4 Google Docs

Google Docs is an online word processor included as part of the free, web-based Google Docs Editors suite offered by Google, which also includes Google Sheets, Google Slides, Google Drawings, Google Forms, Google Sites, and Google Keep. Google Docs is accessible via an internet browser as a web-based application and is also available as a mobile app on Android and iOS and as a desktop application on Google's Chrome OS.

Google Docs allows users to create and edit documents online while collaborating with other users in real-time. Edits are tracked by user with a revision history presenting changes. An editor's position is highlighted with an editor-specific color and cursor and a permissions system regulates what users can do. Updates have introduced features using machine learning, including "Explore", offering search results based on the contents of a document, and "Action items", allowing users to assign tasks to other users.

Google Docs and the other apps in the Google Drive suite serve as a collaborative tool for cooperative editing of documents in real-time. Documents can be shared, opened, and edited by multiple users simultaneously and users are able to see character-by-character changes as other collaborators make edits. Changes are automatically saved to Google's servers, and a revision history is automatically kept so past edits may be viewed and reverted to.[18] An editor's current position is represented with an editor-specific color/cursor, so if another editor happens to be viewing that part of the document they can see edits as they occur. A sidebar chat functionality allows collaborators to discuss edits. The revision history allows users to see the additions made to a document, with each author distinguished by color. Only adjacent revisions can be compared, and users cannot control how frequently revisions are saved. Files can be exported to a user's local computer in a variety of formats (ODF, HTML, PDF, RTF, Text, Office Open XML). Files can be tagged and archived for organizational purposes

Working with Google Docs

In order to open up a Google Doc, you need to first login to your Google Drive, and there we get an option of this Google Docs. Google Docs is a Google's word processor. It helps you in real time processing, real time, word processing. It is quite similar to what actually you can do with a Microsoft Word, but the major difference is that it is an internet-enabled real time processing. You can add large number of users who can work concurrently, that we can say we can add concurrent users to our particular work over here.

Google Drive is actually a new home for the Google Docs. Just setting up a free account and creating and editing the documents you can do it using a Google Drive. You can create and share your work online and you can access your documents, apart from just sharing it with others, you can access your particular document from anywhere in the world, irrespective of your devices, irrespective of the location you are and irrespective of the time you are in. In other words, you can access your documents from anywhere on any device such as laptop, smart phone or tablet or iPad as well. The condition provided is that you need to have an internet access, and a login details that you need to log in to a particular Google account in this case.

In the Figure 2, you can see my Google Drive, I can directly open my Docs from here as well or if I want to go through to the Google Docs through my drive as well I can click on this Google Drive. And here I get an option for new, so when you get an option for new click on this new, you get an

Unit 06: Collaborating using Google Cloud

option for file upload, where you can upload a file, you get an option for folder upload where you can upload a particular folder.

Creating a New Document: We also get an option where you can directly open a Google Docs if you want to open a blank document, or you want to open it from a particular template. You can choose according to your requirement. Like if I want to choose this blank document, I will click on this blank document, it will open up a Google document in front of me. Or you can simply log into this particular area over here, <https://docs.google.com>. You can open a particular document. Finally, a new document is open in front of us titled, "Tarandeep SampelDocs" (Figure 3).

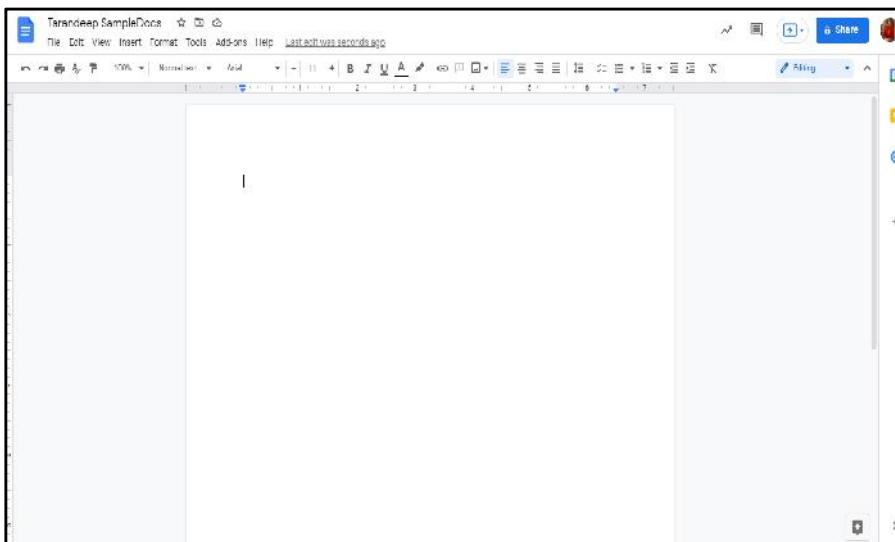


Figure 3: Google Docs Blank Document

Opening a Document: If I want to open another document, instead of going through my particular drive, I can open it through my particular inbox as well, I can directly open it from this G suite, and I can click on this Docs. So here I'm able to open my particular Google Docs over here, I can directly open start a new document, either I can open a blank document, or I can open up a template, what are templates, they are a pre-existing design that have been created for your easiness (Figure 4 and Figure 5).

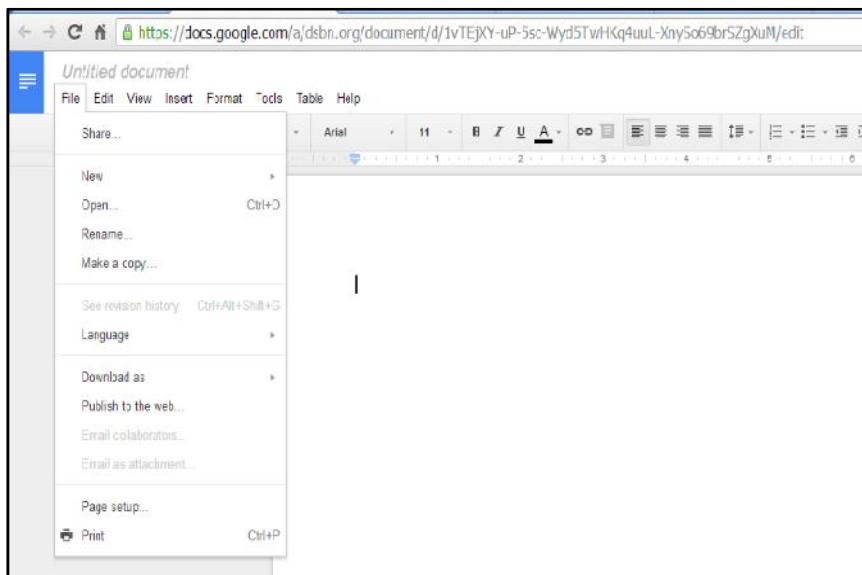


Figure 4: Google Docs File Menu Options

Saving a File: In case of Google Docs or Google Sheets or Google spreadsheet, they are automatically saved. The auto saving facility is already available, you don't need to save unlike Microsoft Word, where you need to save the documents by pressing Ctrl+ S, or saving it to the File menu. Here, we are not concerned with saving it is automatically getting saved after some certain period of time it is getting auto saved.

Cloud Computing

Emailing a File: Next is email options, so if you want to email this particular file to anyone, you can use this option as well.

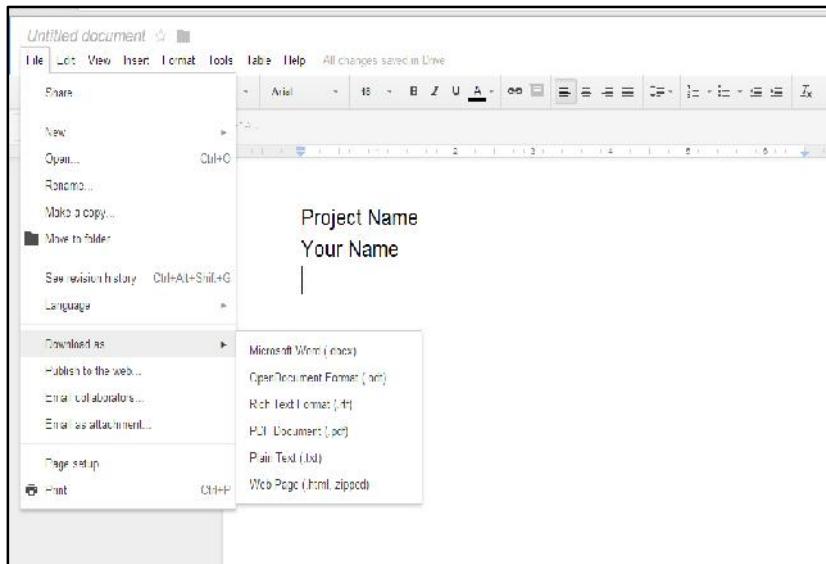


Figure 5: Various Options in File Menu

Page Setup: From Menu bar select File >> Page Setup on the Google Docs screen (Figure 6), you will see a menu bar with options like File, Edit, View, Insert. We will select “File” from the option and from the Drop-Down list select “Page Setup” by scrolling down a little. Set your Page margins for all sides (In inch).



Figure 6: Page Setup Options

Downloading a File: First, open the Google Docs on your PC and then open the document in Google Docs that you want to download and then review it. Next, click on the File tab from the menu bar. After that, click on the Download option from the dropdown menu. And then click on the format in which you want to download it.

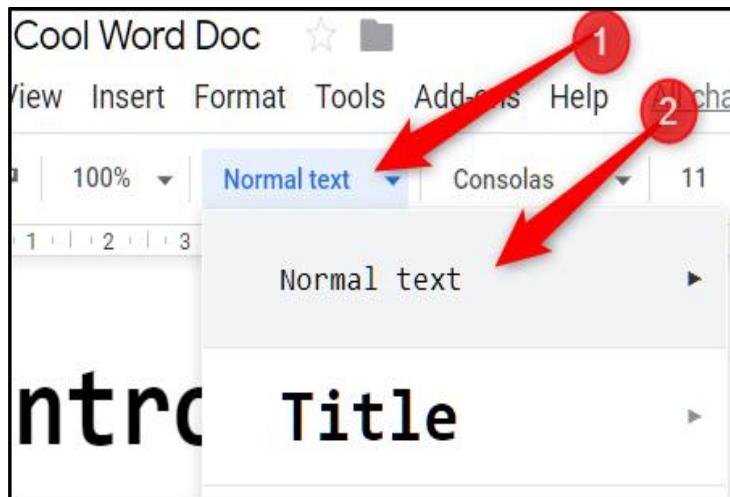
Change How Paragraphs & Fonts Look: You can use the toolbar at the top of a document to:

- o Edit and format the text and paragraph spacing
- o Change the font, and the background color.
- o Bold, italicize, underline, or strikethrough text.

Google Docs uses a standard font when you create a document. If you aren't happy with its appearance, here's how you can change the default font instead of manually adjusting it each time.

Unit 06: Collaborating using Google Cloud

Fire up your browser, open a Google Docs file, and then change the font. Click the drop-down box and make sure the Style is set to “Normal Text,” as well. This refers to the main text style (excluding titles or headings) used throughout your document. Click the drop-down box and make sure normal text is selected.



Notes: Certain formatting can also be applied to the default font, such as size, bold, underline, colour, and highlight colour.

Word Count: Whenever writing an article or an essay or just writing a document for your office or presentation, it is important to have a word count to know how much you are writing and when to wrap up your document. In this article, we are going to help you to find the exact word count in google docs. Google Docs is a widely used application by people for writing their business documents, presentations, product documents, edit existing documents, write an essay, write articles and a lot more things. Google docs is a handy app as it supports Android and iOS Platform or you can even access it using your browser.

To see word count of the entire document then just go to Tools >> Word Count (Figure 7). You can even check Word Count for a single paragraph or multiple paragraphs by selecting the paragraph text and clicking Tools >> Word Count.

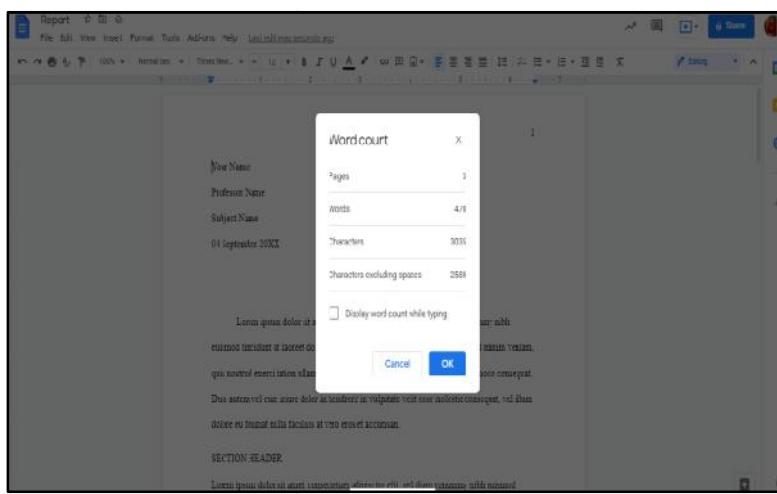


Figure 7: Word Count Options

If you want to see word count in Google docs while typing the document then simply click the checkbox which says “Display Word Count While Typing” and click “OK” to apply the changes (Figure 8).

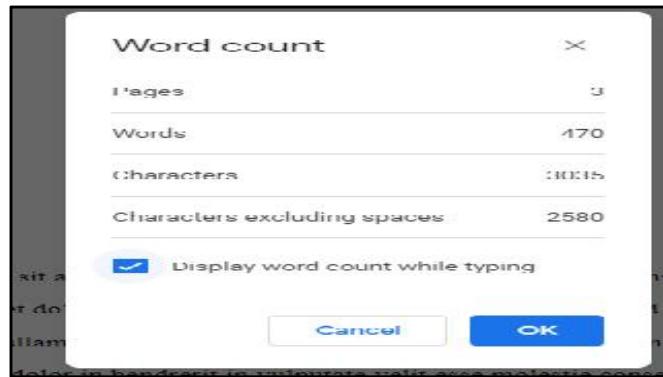
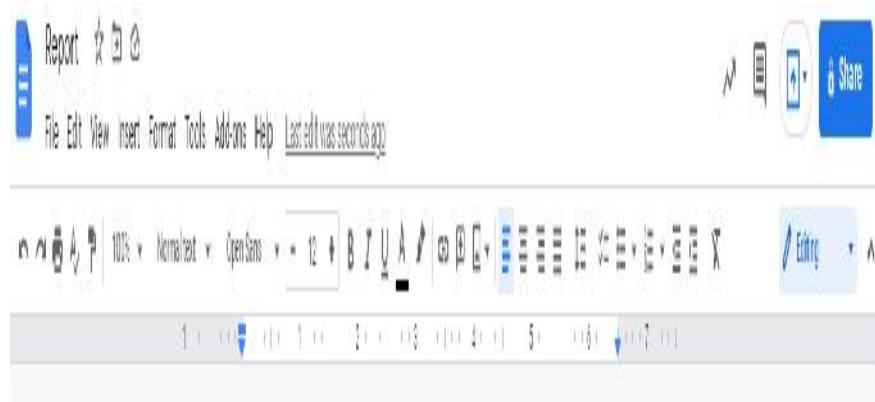
Cloud Computing

Figure 8: Checking the Options on Word Count

Sharing a File: One of the best features of Google Docs is the ability to share documents with others. You can let them just view the document, or you can give them the ability to edit it as well. There are several ways to share a document, but we will look at the two most common: sharing as a link, and sharing with specific people.

How to share a document with specific people: Most often you will want to share a document with specific people or groups of people. Below is how to share a document with people or groups that you specifically choose.

1. Hover your mouse over the file and then click the “Actions” menu to the left of the file name in the file list.
2. Now click “Share” in the drop-down menu and click “Sharing settings...”. This will open the “Sharing settings” window.
3. At the bottom of the window there will be a box labeled “Add people:”
4. Click in that box and type in the people or groups (staff, student, and parent listservs) that you want to share the document with. The address book will find matches based on what you type.
5. Next click the “Can edit” button to pick whether the users can edit or only view the document
6. Optionally click in the “personal message” box to include directions or additional information for the recipients.
7. Finally, click the “Share” button (Figure 9).
8. All of the recipients will now get an email message indicating that the document has been shared with them.



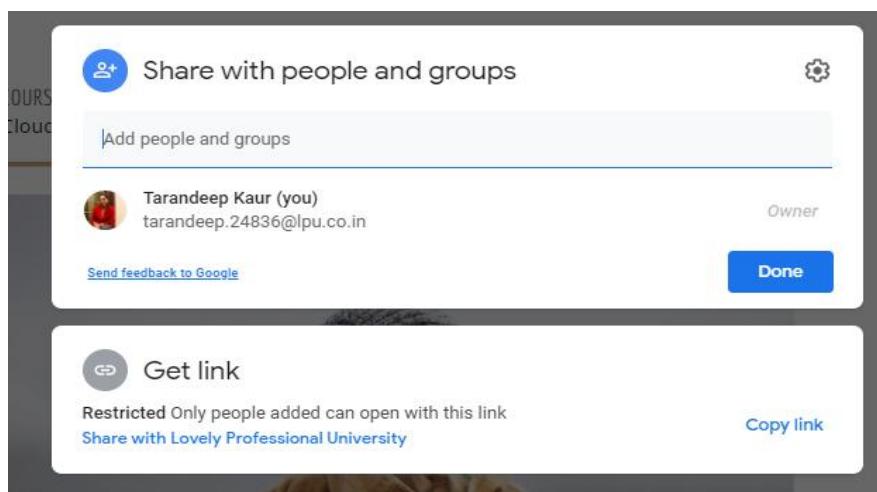


Figure 9: Sharing Option in Google Docs

The document will also automatically show up in their Google Docs list of documents. In the future you can always return to the “Sharing settings” window to change or remove people. Next to each user's name will be a drop-down menu to switch between “Can edit” and “Can view”. Also, there will be an “X” that you can click to remove that person's access entirely (Figure 10).

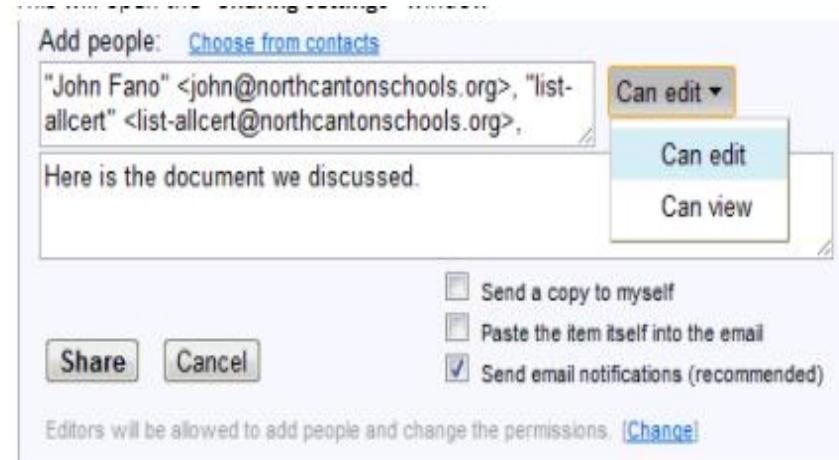
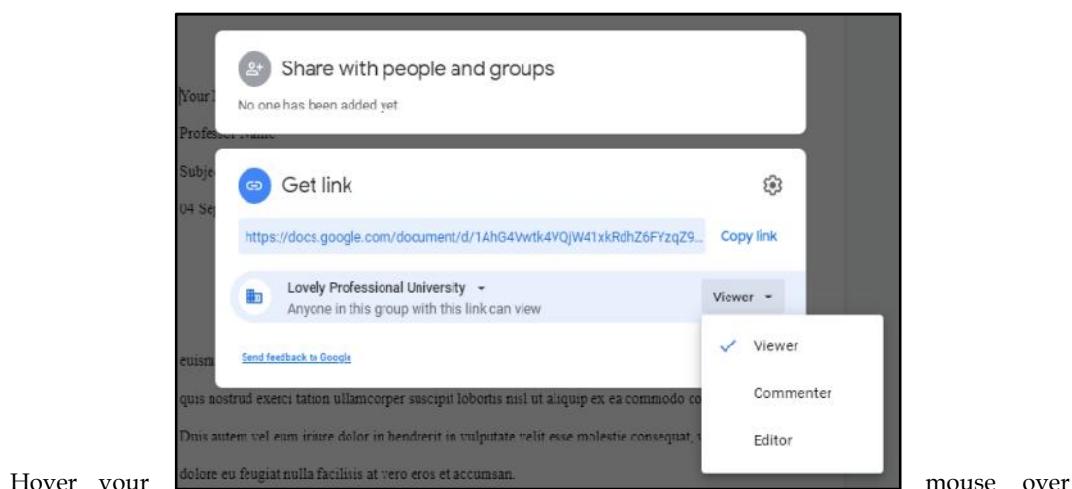


Figure 10: Adding Users in Google Docs

How to share a document as a link: Instead of sharing a document with specific people, you can instead share the document with anyone who has the link to the document. For example, this is useful if you want to share the link with anyone who comes to your website (Figure 11).



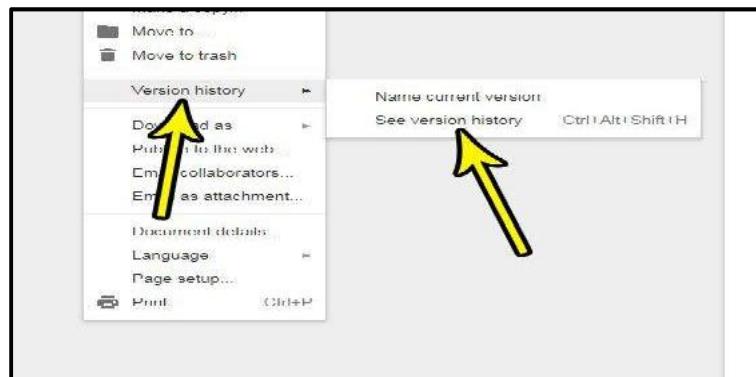
Cloud Computing

the file and then click the “Actions” menu to the left of the file name in the file list.

1. Now click “Share” in the drop-down menu and click “Sharing settings...”.
2. This will open the “Sharing settings” window.
3. At the top you will see “Private - Only the people listed below can access”.
4. Click the “Change” link next to that line.
5. A new window will open up where you can change the privacy settings.
6. If you select “Anyone with the link” then you will get a web link for the document to share with others (through email or posting online). Anyone who has that link can click it and view your document.
7. If you select “People at Lovely professional University” then only people with the link who also have an LPU account can view the document.
8. If you also want to let people edit the document (not just view) then you can select the box for “Allow anyone to edit”.

Version History: Assuming that you have a document in Google Docs that you have edited at least a couple of times. You will be able to view versions of that document, and even restore the document to one of those versions, if you so choose (Figure 12).

1. Sign into your Google Drive and open the document for which you wish to view the version history.
2. Choose the File option at the top of the window.
3. Select the Version history option in the menu, then click the See version history option. As noted on this menu item, you can also press Ctrl + Alt + Shift + H on your keyboard to open this menu.



4. Choose the version of the document that you want to see. If you would like to restore this version of the document (Figure 13), then you can click the Restore this version button at the top of the window.

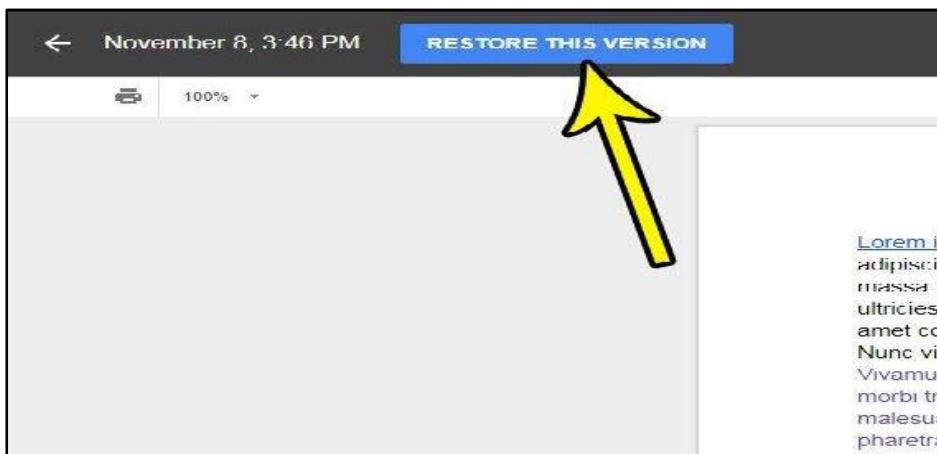


Figure 13: Restoring a Version

6.5 Getting Started with Google Sheets

Google Sheets is a spreadsheet program included as part of the free, web-based Google Docs Editors suite offered by Google. The service also includes Google Docs, Google Slides, Google Drawings, Google Forms, Google Sites, and Google Keep. Google Sheets is available as a web application, mobile app for Android, iOS, Windows, BlackBerry, and as a desktop application on Google's Chrome OS. The app is compatible with Microsoft Excel file formats. The app allows users to create and edit files online while collaborating with other users in real-time. Edits are tracked by user with a revision history presenting changes. An editor's position is highlighted with an editor-specific color and cursor and a permissions system regulates what users can do. Updates have introduced features using machine learning, including "Explore", offering answers based on natural language questions in a spreadsheet.

Google Sheets originated from XL2Web, a web-based spreadsheet application developed by 2Web Technologies, founded by Jonathan Rochelle and Farzad "Fuzzy" Khosrowshahi. XL2Web was acquired by Google in 2006 and turned into Google Labs Spreadsheets. It was launched as a test for a limited number of users, on a first-come, first-served basis on June 6, 2006. The limited test was later replaced with a beta version available to all Google Account holders, around the same time as an official announcement press release was issued. In March 2010, Google acquired the online document collaboration company DocVerse. DocVerse allowed multiple-user online collaboration on Excel-compatible document as well as other Microsoft Office formats such as Word and PowerPoint. Improvements based on DocVerse were announced and deployed in April 2010. In June 2012, Google acquired Quickoffice, a freeware proprietary productivity suite for mobile devices. In October 2012, Google Spreadsheets was renamed Google Sheets.

Google Sheets is available as a web application supported on Google Chrome, Microsoft Edge, Mozilla Firefox, Internet Explorer, and Apple Safari web browsers. The users can access all spreadsheets, among other files, collectively through the Google Drive website. In June 2014, Google rolled out a dedicated website homepage for Sheets that contain only files created with Sheets. In 2014, Google launched a dedicated mobile app for Sheets on the Android and iOS mobile operating systems. In 2015, the mobile website for Sheets was updated with a "simpler, more uniform" interface, and while users can read spreadsheets through the mobile websites, users trying to edit will be redirected towards the mobile app to eliminate editing on the mobile web. The common spreadsheet terms include-

- Cell: Single data point or element in a spreadsheet.
- Column: Vertical set of cells.
- Row: Horizontal set of cells.
- Range: Selection of cells extending across a row, column, or both.
- Function: Built-in operation from the spreadsheet app, which can be used to calculate cell, row, column, or range values, manipulate data, and more.
- Formula: Combination of functions, cells, rows, columns, and ranges used to obtain a specific result.
- Worksheet (Sheet): The named sets of rows and columns making up your spreadsheet; one spreadsheet can have multiple sheets.

Cloud Computing

- o Spreadsheet: Entire document containing your worksheets.

Accessing Google Sheets: Find the application by navigating to Google Sheets (Google Sheets- Figure 14). Alternatively, you may click the grid menu in the upper, right hand corner of your screen while on Google's homepage, your Gmail account or any page with Google search results. Select Docs from the menu, and then click the menu with three horizontal lines on the upper, left-hand corner of the Docs page. Sheets should appear as your second option from the top, and you can click that. Finally, you may access Sheets through Google Drive by clicking the My Drive dropdown menu. Sheets will appear as one of the options.

Creating a Spreadsheet and Fill It with Data: There are 3 ways to create a new spreadsheet in Google Sheets (Figure 15):

- 1) Click the red "NEW" button on your Google Drive dashboard and select "Google Sheets".
- 2) Open the menu from within a spreadsheet and select "File > New Spreadsheet".
- 3) Click "Blank" or select a template on the Google Sheets homepage.

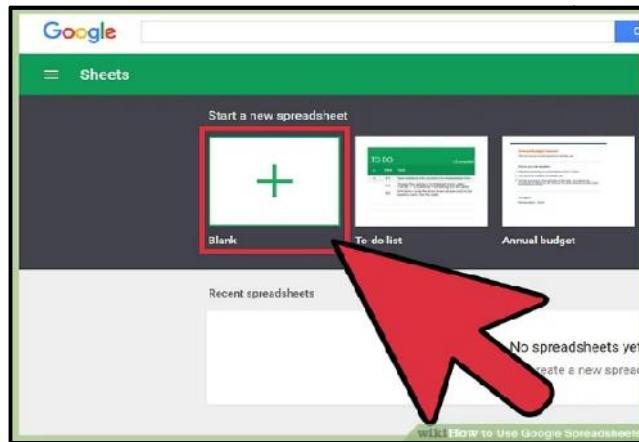
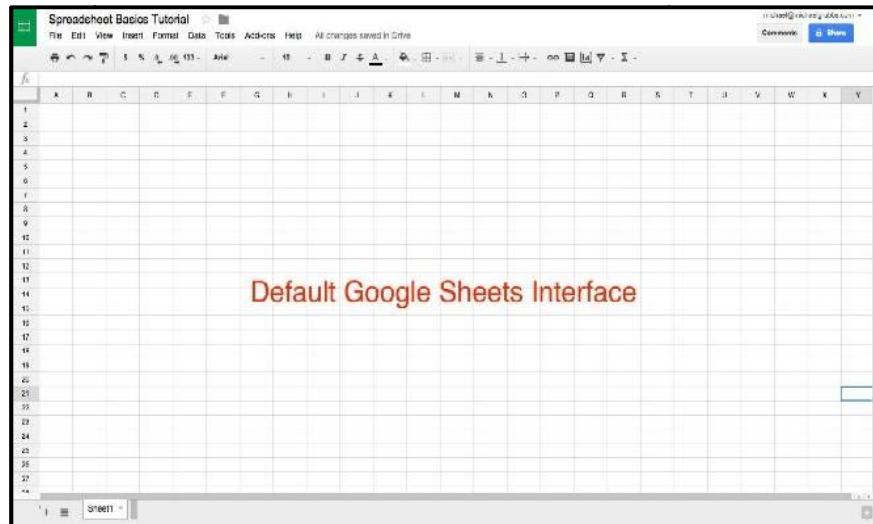
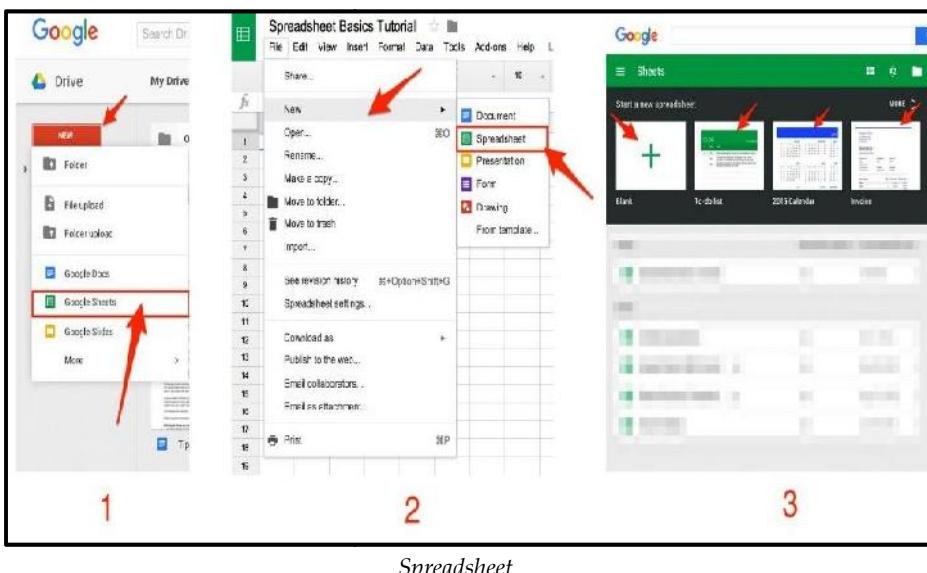


Figure 15: Creating a New



Unit 06: Collaborating using Google Cloud



Spreadsheet

Adding Data to Your Spreadsheet: An individual square in a spreadsheet is called a cell; they're organized into rows and columns with number and letter IDs, respectively (Figure 16). Each cell should contain one value, word, or piece of data.

- 1) Press ENTER to save the data and move to the beginning of the next row.
- 2) Press TAB to save the data and move to the right in the same row.
- 3) Use the ARROW KEYS on your keyboard (up, down, left, and right) to move 1 cell in that direction.
- 4) Click any cell to jump directly to that cell.

Format Data for Easy Viewing: Whether you're tracking expenses, recording students' grades, or keeping track of customers in a homebrew Customer Relationship Management (CRM), you'll want to manipulate and format your data. The basic formatting options in Google Sheets are available above your first cell. They're labeled in the image below (Figure 17), but for quick reference while you're working on a sheet, just hover over an icon to see its description and shortcut key.

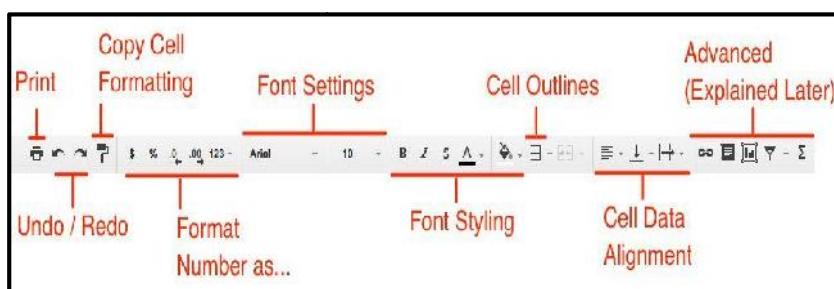


Figure 17: Formatting Options

Print, Undo / Redo, and the Font Settings / Styling function similarly to what you'd expect from your favorite word processor. The shortcut keys are the same as well, so just treat it like you're editing any other document!

Importing a file: Importing a file is simple as well. You can either import directly into the current spreadsheet, create a new spreadsheet, or replace a sheet (that is, an individual tab) with the imported data. The most common files you'll import are CSV (comma separated values) or XLS and XLSX (files from Microsoft Excel). To import a file from outside of your Google Drive, go to the FILE > IMPORT > UPLOAD menu.

Cloud Computing

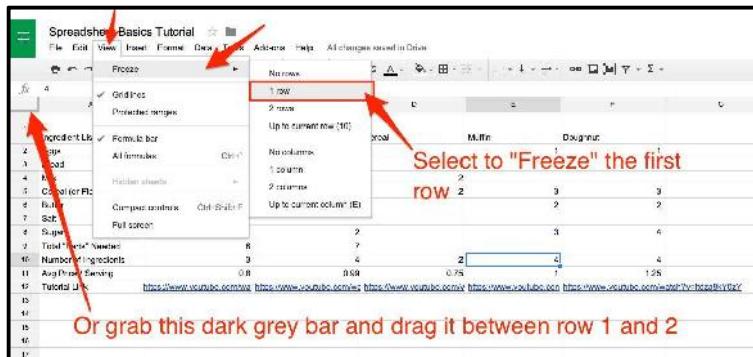
Merging Cells in Google Sheets: Open a Google Sheets spreadsheet on your Computer. In order to combine rows, columns, or cells, select them. Then, select Format from the drop-down menu at the top followed by selecting how you want your cells to be combined before merging them.

Making a Graph/ Chart in Google Sheets: Follow these steps in order to make a graph in Google sheets-

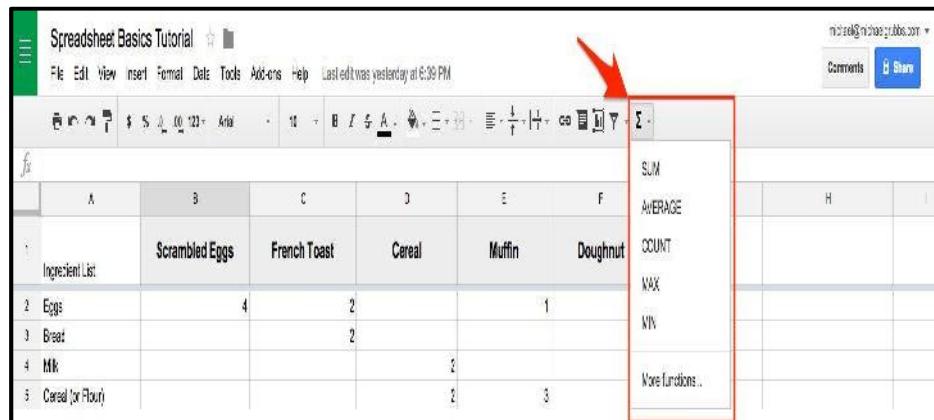
- 1) Open the spreadsheet.
- 2) Select the “Insert” button in the toolbar.
- 3) Select “Graph.” A menu pops up with many different types of graphs that you can insert into your spreadsheet.
- 4) Select one of them and they will be inserted into the spreadsheet.
- 5) Customize the graph by dragging it around, resizing it, or changing its appearance.

Freezing in Google Sheets: "Freeze" the first row in place. That means if we scroll down the spreadsheet, the first row will still be visible, no matter how much data lies below it. This allows you to have a long list and helps to keep tabs on what you're actually looking at. There are two ways to freeze rows (Figure 18):

- 1) Click VIEW > FREEZE > 1 ROW in the navigation bar to lock the first row in place.
- 2) Hover the dark grey bar in the top left of the spreadsheet (until it becomes a hand) and drag between rows 1 and 2.



Adding Formulas in Google Sheets: Google Sheets, like most spreadsheet apps, has a bunch of built-in formulas for accomplishing a number of statistical and data manipulation tasks. You can also combine formulas to create more powerful calculations and string tasks together. And if you're already accustomed to crunching numbers in Excel, the exact same formulas work in Google Sheets most of the time.



You can click a formula to add it to a cell, or you can start typing any formula with a = sign in a cell followed by the formula's name. Sheets will auto-fill or suggest formulas based on what you type, so you don't need to remember every formula. The most basic formulas in Sheets include:

- o SUM: adds up a range cells (e.g. $1+2+3+4+5 = \text{sum of } 15$)
- o AVERAGE: finds the average of a range of cells (e.g. $1,2,3,4,5 = \text{average of } 3$)

Unit 06: Collaborating using Google Cloud

- COUNT: counts the values in a range of cells (ex: 1,blank,3,4,5 = 4 total cells with values)
- MAX: finds the highest value in a range of cells (ex: 1,2,3,4,5 = 5 is the highest)

MIN: finds the lowest value in a range of cells (ex: 1,2,3,4,5 = 1 is the lowest)

Basic Arithmetic: You can also perform functions like addition, subtraction, and multiplication directly in a cell without calling a formula.

Downloading Your Data: If you need to send your files to external collaborators, upload a file into another system, or just like having backups for posterity, then turn towards one of Google Sheets' many data export options (Figure 20).

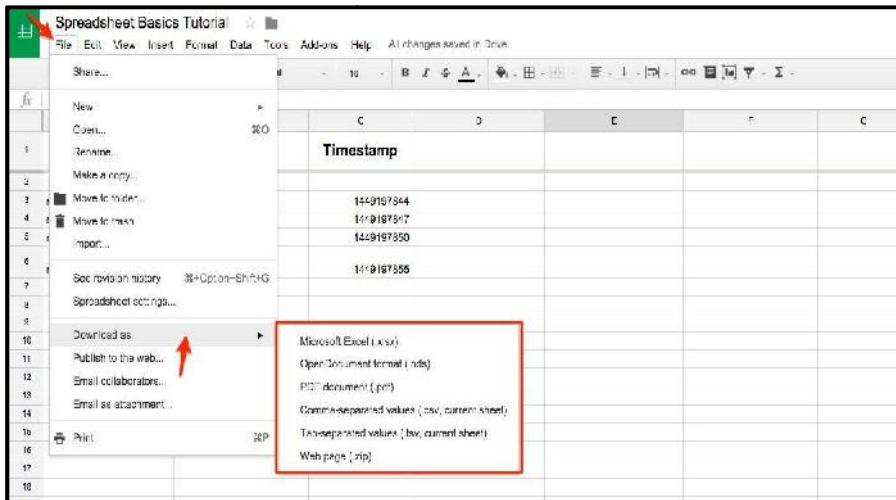


Figure 20: Download Options in Google Sheets

How to use Google Sheets: Killer features

Let's see some of the unique, powerful features that Google Sheets has, as a cloud-based piece of software.

- **Comments (and Notes):** Do you want to add some context to numbers in the cells of your Sheets, without having to add extra columns or mess up your formatting? Simply, add a comment to a cell!.

You can tag people (via their email address) who you want to see the comment too. They can reply and mark it resolved once it's been acted upon. You can also add simple notes to cells as well if you wish.



Notes: The comments and notes can also be deleted when not required anymore.

To add a comment to a cell, first select the cell, then right click to bring up the menu of options. Select "Insert comment" and then simply type in your comment. To tag somebody in your comment, type the plus sign (+) and their name or email address (you'll see auto-complete options from your contacts, so you shouldn't have to type in the whole email address).

You'll notice a small orange triangle in the top right corner of the cell to indicate the comment. The comment will show up when you hover over this cell. If you click on the cell, it'll also add orange shading to the cell background. The comments can be edited, deleted, linked to, replied to and resolved (comment disappears from Sheet and is archived). You can reach and control all the Comments in your Sheet from the big Comments button in the top right of the screen, next to the blue Share button. You can also add a note to cells in the same way (look for it in the menu next to Insert Comment). It's like a pared down version of a comment, intended for your own reference.

- **Share your Sheets:** You can share your Google Sheets with other people. Since it's on the cloud, they can access your Sheet and see the same, live Sheet that you're in. In other words, if you make changes, they will show up automatically and in near real-time for everybody viewing the

Cloud Computing

Sheet. You can have multiple people viewing and working on the same Sheet. Essentially, you have three options to share your sheet with:

- o View-only access, so that person cannot change or comment on any data
- o Comment-only access, so that person can add comments but still not make any changes to the data in the Sheet
- o Editing access, so that person can make changes to the sheet (including comments)

The sharing options are found by clicking on the big blue button in the top right corner, which will open up the Sharing settings: You can grab the link (the URL) to the Sheet, choose the share setting (view/comment/edit) and then share that link with people you want to see the Sheet. Or, you can enter someone's email address directly, choose the share setting (view/comment/edit) and then share the Sheet directly with the person. If you want to review the sharing settings or have even more control, click the Advanced options buttons. Here you can:

- 1) Grab the sharing link.
- 2) Review who has access.
- 3) Change the access rights of anyone listed.
- 4) Invite new people to access the Sheet.
- 5) Change the advanced owner settings, to restrict who can control the sharing settings and specific view/comment rights.
- 6) Confirm when you're finished.

Real-time Collaboration: Ok, so you've shared your Sheet with someone. If they open it whilst you're still working in the sheet, you'll see their cursor show up on whatever cell (or range) they've selected. It'll be a different color, for example green to your blue. If they enter data or delete data, you'll see it happening in real-time!

Gym membership table						
	First Name	Last Name	Full Name	Tier	Member since	Monthly fee
3	Charles	Dickens	Charles Dickens	1	1-Dec-2017	\$199
4	Ernst	Hammingway	Ernst Hammingway	3	1-Jan-2018	\$79
5	Mark	Twain	Mark Twain	1	30-Jun-2017	\$199
6	Jane	Austen	Jane Austen	2	20-Aug-2017	\$129
7				4		\$606
8						\$152
9						
10						
11						
12						
13						
14						
15						
16						

Figure 21: Real-Time Collaboration

In this case, (Figure 21) my active cell is the blue-outlined cell. I see somebody else, denoted by the green-outlined cell, show up in this Sheet and enter data into a few cells before deleting it.

6.6 Working with Google Forms

Google Forms is a survey administration app that is included in the Google Drive office suite along with Google Docs, Google Sheets, and Google Slides. The forms feature all of the collaboration and sharing features found in Docs, Sheets, and Slides. They are easy to create surveys and forms for everyone. The users are able to create custom forms for surveys and questionnaires at no extra cost. Google forms allow collecting information from users via a personalized survey or quiz. The information is then collected and automatically connected to a spreadsheet. The spreadsheet is populated with the survey and quiz responses.

How to Use Google Forms

Step 1: Set up a new form or quiz

- 1) Go to forms.google.com
- 2) Click Blank Plus.
- 3) A new form will open.

Creating a Form in Google Sheets: When you create a form in Google Sheets, the responses will be saved in a new sheet. Learn more about where you can save responses.

- o On a computer, open a spreadsheet at sheets.google.com.
- o Click Insert and then Form.
- o A new sheet will appear in your spreadsheet, and your form will open.

How to Customize Your Form: One of the first things you might want to do after you create a new blank form is give it a bit of personality. Google Forms lets you customize the theme and help it stand out by adding an image, color, and font style (Figure 22). Click the artist's palette at the top of the screen. From here, you can choose a header image from one of the many stock photos provided (or upload one of your own), the primary color of the form, the background color, and font style.

While customization is lacking when it comes to the theme of forms (aside from being able to upload any image for the header), Google Forms makes up for it with everything else it has to offer.

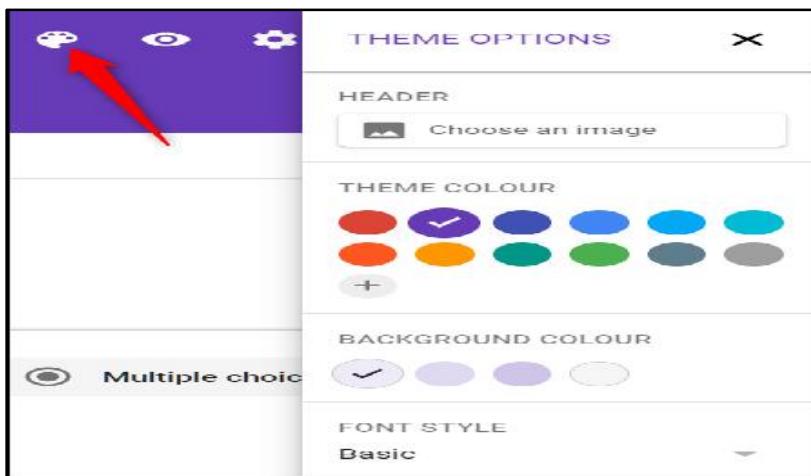
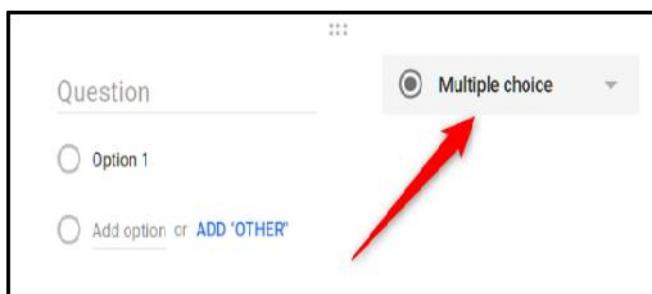


Figure 22: Customizing a Google Form

When you create a Google Form, you can choose the types of questions you want people to answer. Whether you want static answers from a multiple-choice form or essay-length responses, you can create your ideal form in a snap!

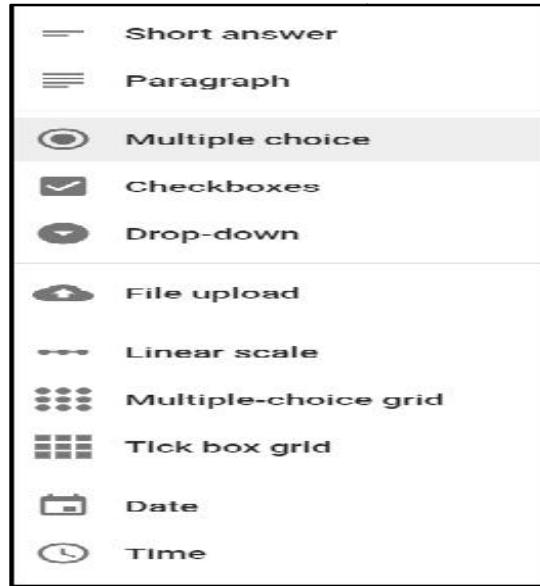
Click the drop-down menu next to the question field (Figure 23). Then, select the type of question you want from the list.



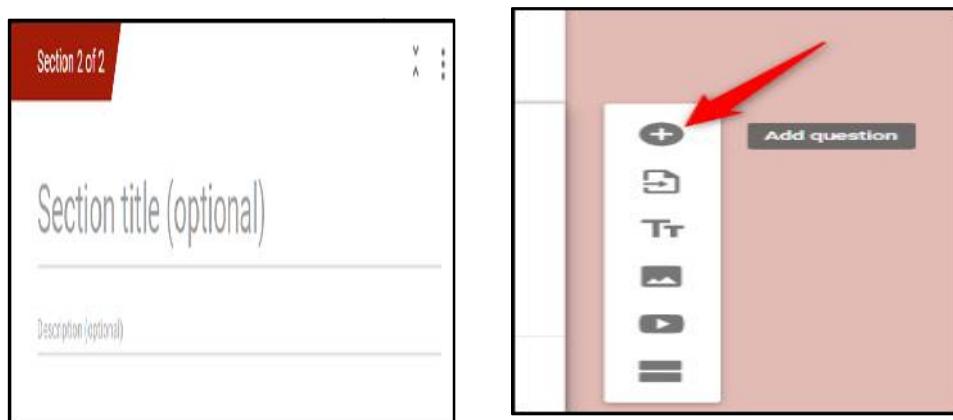
Cloud Computing

Your choices are:

- Short Answer: Responses only require a few words. You can set rules people have to follow in their answer with data input validation. Great for email addresses or URLs.
- Paragraph: Responses require long-form answers of one or more paragraphs. Data input validation is available for this type of response, as well.
- Multiple Choice: People choose between a set of options (one per question). You can include “Other” and an option so people can input a short answer. Depending on a person’s answer, you can also send them to a different section of the form.
- Checkboxes: Responders choose one or more of a set of options, including the “Other” option for a short answer. Depending on a person’s answer, you can send them to a different section of the form.
- Drop-down: People choose their answer from a set of options in a drop-down menu (one per question). Based on the answer, you can, again, send people to another section of the form.
- File Upload: This allows the person to upload a file in response to a question. Uploaded files use Google Drive space for the survey owner. You can specify the size and type of files people can upload.
- Linear Scale: People can rate your question on a scale that starts at 0 or 1, and ends on a whole number from 2 to 10.
- Multiple Choice Grid: This creates a grid from which people can select one answer per row. Optionally, you can limit answers to one choice per column and shuffle the row order.
- Checkbox Grid: This option creates a grid from which people can select one or more answer per row. Optionally, you can limit answers to one choice per column and shuffle the row order.
- Date: Responder must choose the date as an answer to the question. The default is day, month, and year. Optionally, you can include the time in people’s answers.
- Time: Responder must choose the time of day or a duration of time.

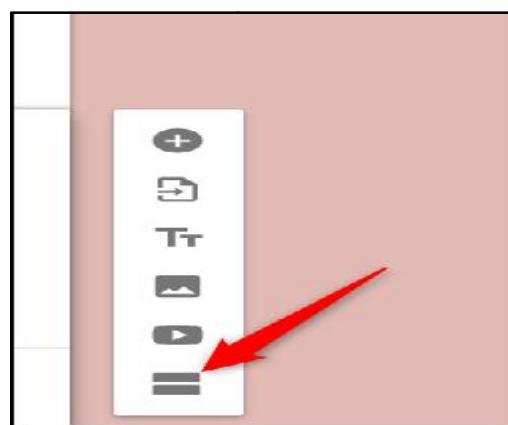


How to Add More Questions: If you’re creating a survey or quiz, you’re likely going to include more than one question on it. Google Forms makes it super-easy to add as many questions as you want, and you can vary the types of questions (Figure 24). You can even separate them into sections, so everything doesn’t appear on a single page. In order to add more questions to your form, click the plus sign (+).

Unit 06: Collaborating using Google Cloud

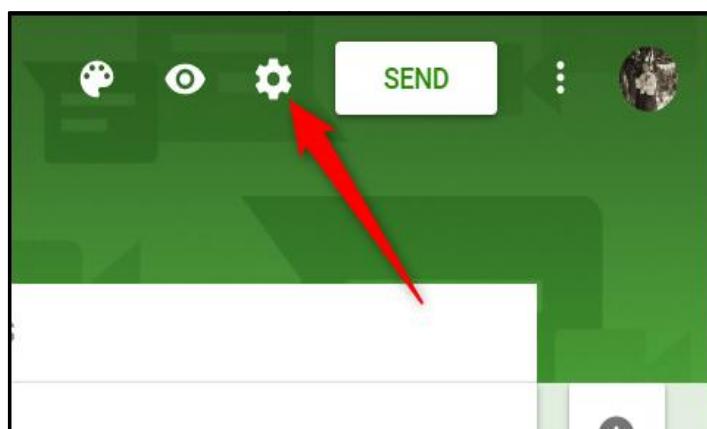
How to Add Multiple Sections: To add another section to separate questions, click the icon that looks like two rectangles (Figure 25). Optionally, you can give the section a name and description to distinguish it from other sections later on. If you want to add any questions to a different section, it's simple!

Just drag and drop them between sections. At the end of the section, click the drop-down menu to choose where the form should direct people next (Figure 26).

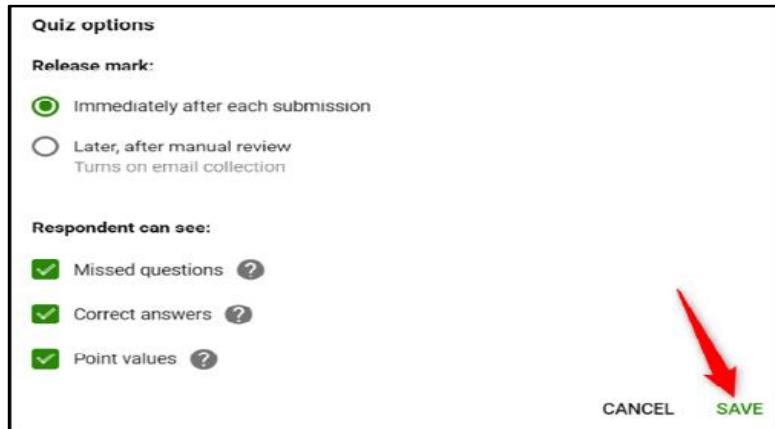
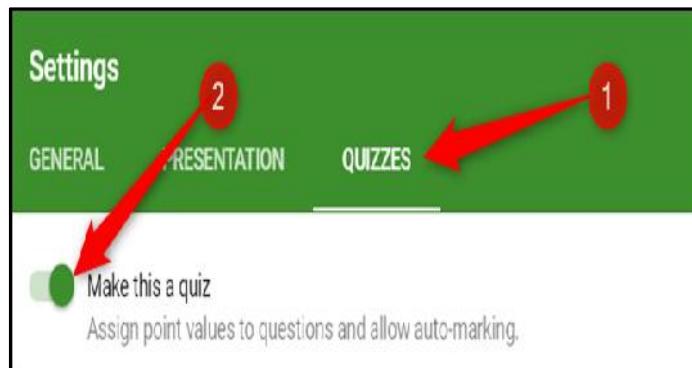


How to Create a Quiz: Google Forms isn't just for surveys or event invites. Teachers can use Forms to create digital quizzes, which automatically grade, send results (if enabled), and collect student responses. It's one of the easiest ways to give students immediate feedback and reduce the amount of time you spend grading quizzes.

Click the Settings cog at the top of the page (Figure 27).



Click the “Quizzes” tab, and then toggle on “Make This a Quiz.”



After you enable quiz mode, you can choose when to release a student's mark, and what information he can see after he's submitted his quiz. When you finish, click "Save" to exit the window (Figure 29).

Once you return to your quiz, select one of the questions, and then click "Answer Key" to edit the correct response, and the weight each question has in the quiz (Figure 30).



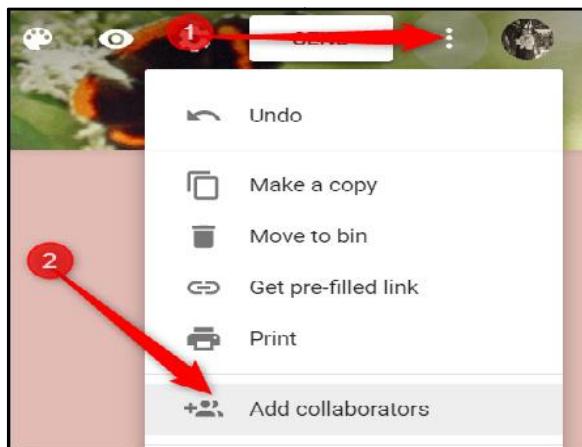
Figure 30: Answer Key Option

Here's where you set the correct answer(s), decide how many points each one is worth, and add answer feedback for the answers to each question (Figure 31).



Figure 31: Quiz Setting Option

How to Collaborate on Forms: Like all Google suite applications, Forms lets you collaborate with others. Anyone you share a unique link with can edit the questions in your form. This makes it a lot easier to work on the same survey with a group. To do this, click the three dots at the top of the page, and then click “Add Collaborators” (Figure 32).



Next, under the heading “Who Has Access,” click “Change” (Figure 33).



Select “On – Anyone with the link” to generate a shareable link (Figure 34 and Figure 35). Anyone with this link can then access and edit your form.

Click “Save”. Now, you can copy and share the link with anyone you want to have access to your form.

You can do a lot more with these shareable links, as they also work with other Drive files and on mobile.

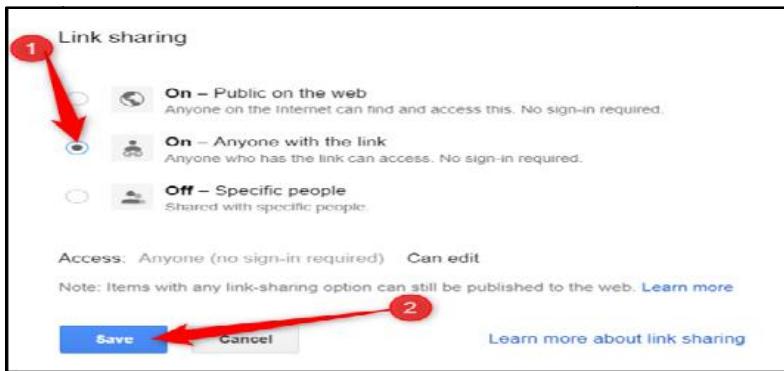
Cloud Computing

Figure 35: Link-based Sharing

How to Store Responses in Google Sheets: Google Forms stores the answers to your Form automatically. It saves each response in the “Responses” tab (Figure 36) at the top of your form and updates in real-time as people answer questions.



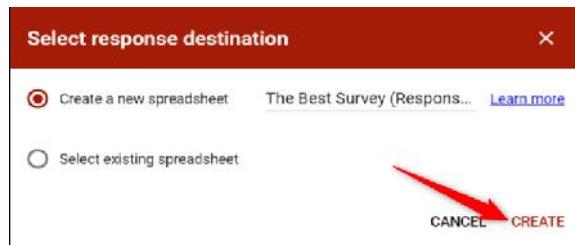
However, if you prefer a more in-depth way to analyse responses from your form, you can generate a new Google Sheet—or link to an existing one—to store and view answers (Figure 37). When viewing data stored in a spreadsheet, you can apply many types of calculations and Google Sheets’ functions to create formulas that manipulate your responses. In order to do this, select the “Responses” tab, and then click the green Sheets icon.

Next, click “Create” to generate a new spreadsheet to store all your answers (Figure 38).



Unit 06: Collaborating using Google Cloud

Click "Create". Each spreadsheet contains all responses, along with a timestamp of when the survey was completed. Each spreadsheet contains all responses, along with a timestamp of when the survey was completed. If you already have a spreadsheet you want to use, you can do that, too! Instead of clicking "Create," click "Select existing spreadsheet," and then click "Select".



How to Use a Form Template: Sometimes, you don't want to think about the format—or questions—of a survey. When that's the case, you can use a template from the Google Forms template gallery (Figure 39). It has templates for everything from party invites to course evaluation forms. To get started, head over to the Google Forms homepage and place the cursor on the multi-coloured plus sign (+) in the bottom-right corner.

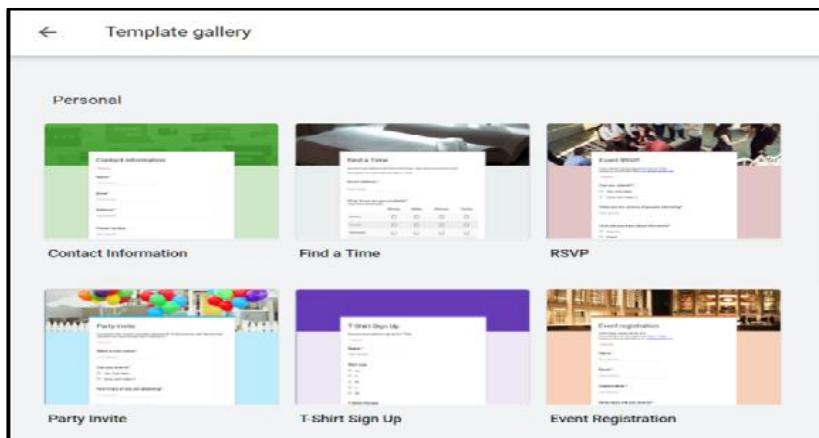
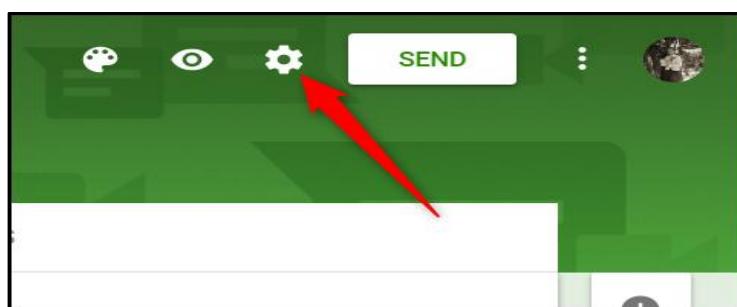


Figure 39: Template Gallery

Click on a template. The form opens in the current tab and saves to your Drive with all your other forms. If you want to add any questions or edit any existing ones, the templates are customizable, just like any other form.

Other Settings: Before you share your form with everyone, be sure to check out the settings. From here, you can collect email addresses, create a confirmation message, limit responses to one per person, and more. Click the Settings cog at the top of the page (Figure 40 and Figure 41).



Cloud Computing

The screenshot shows the 'Settings' screen for a Google Form. At the top, there are three tabs: 'GENERAL', 'PRESENTATION', and 'QUIZZES'. The 'GENERAL' tab is selected. Under 'GENERAL', there are several configuration options:

- Collect email addresses:** A checked checkbox labeled 'Collect email addresses'.
- Response receipts:** An unchecked checkbox labeled 'Response receipts' with a question mark icon.
- Requires sign-in:** A checked checkbox labeled 'Limit to 1 response' with the note 'Respondents will be required to sign in to Google'.
- Respondents can:**
 - An unchecked checkbox labeled 'Edit after submit'.
 - An unchecked checkbox labeled 'See summary charts and text responses'.

The first tab has a few settings you can enable. From here, you can collect email addresses and limit each person to one submission. You can also choose whether respondents can edit their answers after they're submitted or see a summary chart at the end of the survey.

Summary

- Cloud has tremendous benefits for the entire community, from neighbourhood groups to sports teams to school organizations.
- Google Drive is actually a new home for the Google Docs. Just setting up a free account and creating and editing the documents you can do it using a Google Drive.
- Google Docs is an online word processor included as part of the free, web-based Google Docs Editors suite offered by Google, which also includes Google Sheets, Google Slides, Google Drawings, Google Forms, Google Sites, and Google Keep.
- Google Docs is accessible via an internet browser as a web-based application and is also available as a mobile app on Android and iOS and as a desktop application on Google's Chrome OS.
- Google Docs allows users to create and edit documents online while collaborating with other users in real-time.
- Google Sheets is available as a web application supported on Google Chrome, Microsoft Edge, Mozilla Firefox, Internet Explorer, and Apple Safari web browsers.
- Google forms allow collecting information from users via a personalized survey or quiz. The information is then collected and automatically connected to a spreadsheet.

Keywords

Word Processor: A word processor is a software program capable of creating, storing, and printing documents. Unlike the standard typewriter, users using word processors have the ability of creating a document and making any changes anywhere in the document.

Cloud Collaboration Tools: Cloud collaborations act as a tool for collaboration internally between different departments within a firm, but also externally as a means for sharing documents with end-clients as receiving feedback.

Google Docs: Google Docs is an online word processor included as part of the free, web-based Google Docs Editors suite offered by Google, which also includes Google Sheets, Google Slides, Google Drawings, Google Forms, Google Sites, and Google Keep.

Google Sheets: Google Sheets is a spreadsheet program included as part of the free, web-based Google Docs Editors suite offered by Google.

Google Forms: Google Forms is a survey administration app that is included in the Google Drive office suite along with Google Docs, Google Sheets, and Google Slides. The forms feature all of the collaboration and sharing features found in Docs, Sheets, and Slides.

Self Assessment

1. Cloud _____ is a way of sharing and co-authoring computer files through the use of cloud computing.
 - A. Virtualization
 - B. Consolidation
 - C. Collaboration
 - D. SharingIt

2. Google Drive offers users _____ of free storage through Google One.
 - A. 10 gigabytes
 - B. 15 gigabytes
 - C. 20 gigabytes
 - D. 5 gigabytes

3. Which of the following characterises collaboration in cloud computing?
 - A. All platforms can easily be accessed via a web browser.
 - B. Offer apps for ease of access from a smartphone or tablet.
 - C. Feature a directory structure similar to that of a computer drive
 - D. All of the above

4. _____ is a file storage and synchronization service developed by Google.
 - A. Google Drive
 - B. Google Toolkit
 - C. Google App Engine
 - D. Google Coding Kit

5. Google Docs is Google's version of:
 - A. Microsoft Outlook
 - B. Microsoft Word
 - C. Microsoft Excel
 - D. Microsoft Publisher

6. You can share documents you create using Docs by:
 - A. Sending a link
 - B. Sharing directly with a person's email address
 - C. Both
 - D. Neither

7. You can make a shareable link "Public on the web."
 - A. True
 - B. False

8. You can revert back to earlier versions of a document using:
 - A. Sharing permissions
 - B. Version history
 - C. Refresh button
 - D. Helpdesk

Cloud Computing

9. Word documents can be converted to Google Docs.
 - A. True
 - B. False

10. Google Sheets is a _____, cloud-based spreadsheet application.
 - A. Free
 - B. Paid
 - C. Subscription-based
 - D. Pay-per-use model

11. Columns in Google Sheets refer to the _____ set of cells.
 - A. Horizontal
 - B. Vertical
 - C. Upper
 - D. Lower

12. Filtering in Google Sheets can be done in various ways that include:
 - A. Filter by value
 - B. Filter by condition
 - C. Both
 - D. Neither

13. Google Forms is a _____ app that is included in the Google Drive office suite.
 - A. Database administration
 - B. Energy administration
 - C. Network administration
 - D. Survey administration

14. Multi-section option is available in Google Forms.
 - A. True
 - B. False

15. Google Forms is a tool that allows collecting information from users via a personalized _____.
 - A. Table
 - B. Database
 - C. Quiz
 - D. Image

Answers for Self Assessment

- | | | | | |
|-------|-------|-------|-------|-------|
| 1. C | 2. B | 3. D | 4. A | 5. B |
| 6. C | 7. A | 8. B | 9. A | 10. A |
| 11. B | 12. C | 13. D | 14. A | 15. C |

Review Questions

1. What is meant by cloud collaborations?
2. What is web-based word processing?

Unit 06: Collaborating using Google Cloud

3. How can you create Word Documents in Collaboration?
4. What are the features of a standard word processor?
5. What is meant by filtering in Google Sheets? Explain with an example?
6. How does sharing and collaboration occurs in Google Sheets?
7. How we can add or create multiple sections in Google Forms?
8. Discuss and demonstrate the following in Google Docs:
 - (a) Word count
 - (b) Version history
 - (c) Sharing a file
 - (d) Downloading a file
 - (e) Adding comments

**Further Readings**

- The Ridiculously Simple Guide to Google Apps (G Suite): A Practical Guide to Google Drive Google Docs, Google Sheets, Google Slides, and Google Forms by Scott La Counte.
- Google Apps and G-suite: A Complete and Practical Guide on How to Use Google Drive, Google Docs, Google Sheets, Google Slides, Google Forms, Google ... and Google Photos. Tips and Tricks Included by Masha Cromwell, 2021.
- Cloud Computing: Concepts, Technology and Architecture by Thomas Erl, Pearson Education.
- Cloud Computing Black Book by Kailash Jayaswal, Jagannath Kallakurchi, Donald J. Houde, Deven Shah, Kogent Learning Solutions, DreamTech Press.

**Web Links**

1. Google Docs: Free Online Document Editor | Google Workspace
2. Google Docs - Wikipedia
3. Google Forms: Free Online Form Creator | Google Workspace
4. The Ultimate Guide to Google Forms (hubspot.com)
5. Google Sheets: Free Online Spreadsheet Editor | Google Workspace
6. Google Sheets - Wikipedia
7. What is Google Sheets and How is it Used? (techtarget.com)
8. The Ultimate Guide to G Suite (zapier.com)

Unit 07: Collaborating on Event Management

CONTENTS

- Objectives
- Introduction
- 7.1 Implementation of Google Calendar
- 7.2 Collaborating on Schedules
- 7.3 Collaborating on To-Do Lists
- 7.4 Collaborating on Task Management
- 7.5 Collaborating on Event Management
- 7.6 Collaborating on Budgets
- 7.7 Virtual Communities in the Cloud
- 7.8 Collaborating on Contact Management
- Summary
- Keywords
- Self Assessment
- Answers for Self Assessment
- Review Questions
- Further Readings

Objectives

After this lecture, you will be able to,

- Know about collaborating on calendars.
- Practically explore the collaborations using Google Calendar.
- Understand schedules and task management.
- Explore collaborating on schedules and task management.
- Know about creating To-Do lists.
- Explore collaborating on contact management.

Introduction

Most computer users today have embraced keeping their schedules on their PCs. Not that the old-fashioned wall-hanging calendar is dead, it's just that it's a whole lot easier to track appointments and events electronically; the computer does all the busywork for you. The problem, however, with using calendar software (such as Microsoft Outlook or Windows Calendar) is that all your appointments have to reside on a single computer. If you keep a personal calendar on your home PC, you can't reference it from work or when you're traveling. This limits the calculator program's usefulness.

That's why, instead of using a calendar that's wedged to a single computer, many users are moving to web-based calendars. Web-based calendar service stores your calendars on the Internet, where they can be accessed from any computer that has an Internet connection. It lets you check your schedule when you're on the road, even if your assistant in the office or your spouse at home has added new appointments since you left.

Web-based calendars are also extremely easy to share with other users in any location, which make them great for collaborative projects. Yahoo! Calendar does pretty much everything Google Calendar does, and should be another favorite, especially among non-business users. Apple's new

Cloud Computing

MobileMe Calendar, which is already attracting a lot of attention; considered by anyone also looking at the Google and Yahoo! Applications. Let us discuss each in detail.

Google Calendar

Most popular web-based calendar today, no doubt due to association with the web's most-used search engine, is Google Calendar (calendar.google.com). It is free, full featured, and easy to use. It lets you create both personal and shared calendars, which makes it ideal for tracking business group, family, and community schedules. Google Calendar looks pretty much like every other calendar you've ever seen (Figure 1). You enter your appointments (which Google calls "events") directly into the calendar, which you can display in either daily, weekly, or monthly views. You can also, if you like, view your weekly agenda on a single page.

Like all web-based calendars, all your events are stored in the cloud (in this case, the cloud created by Google's own network of servers), not on your own computer. This means that you can access your calendar from any computer anywhere in the world. Just log in to the Google Calendar page and your calendar and all events are there.

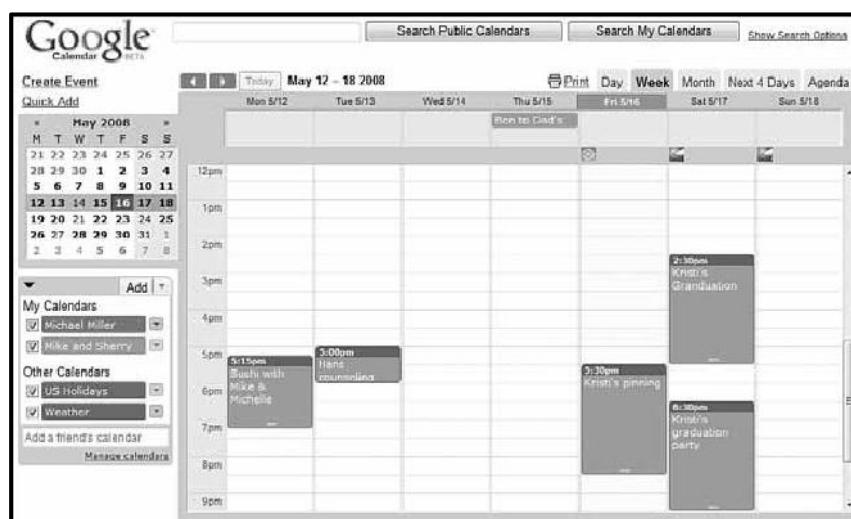


Figure 1: Google Calendar Interface

Yahoo! Calendar

One of Google Calendar's primary competitors is Yahoo! Calendar (calendar.yahoo.com) (Figure 2 and Figure 3), hosted by its search competitor Yahoo! This web-based calendar looks, feels, and functions quite similarly to Google Calendar, and is also free for anyone to use. Most web-based calendars have a similar visual look. (How different can you make a calendar look, anyway?) One subtle difference in Yahoo! Calendar, however, is the presence of an Add Task button. This reflects Yahoo! Calendar's offering of tasks in addition to events. You can still add individual items to your daily schedule, but you can also add longer-term tasks and have their due dates show up on your calendar.

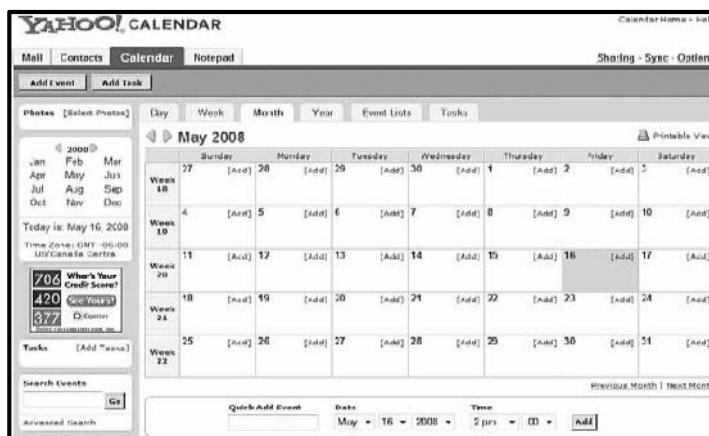


Figure 2: Yahoo Calendar

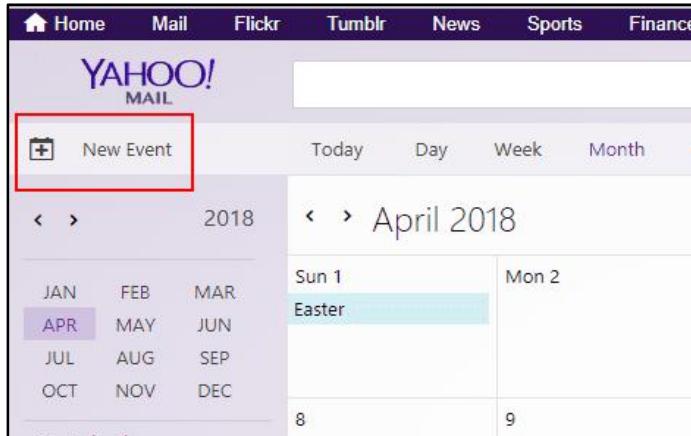


Figure 3: Adding Events to Yahoo Calendar

Just click the Sharing link and indicate how you want to share – no sharing, view-only for friends, view-only for anyone, or view-only with special friends allowed to edit. At present, Yahoo! Calendar only lets you create a single calendar. All the events, public and private, have to be stored on this calendar; you can't create different calendars for different functions. (That's one advantage that Google has over Yahoo! here.).

Windows Live Calendar

Windows Live Calendar (mail.live.com/mail/calendar.aspx) is Microsoft's web-based calendar, actually part of the Windows Live Hotmail email service. It looks a lot like both of its primary competitors. It offers tasks, like Yahoo! Calendar, and also lets you schedule meetings with other calendar and Hotmail users. You can share your calendars with authorized users for group collaboration.

Apple MobileMe Calendar

Apple's MobileMe (www.me.com) is a new competitor in the web-based apps market. It includes online mail, contacts, and calendar, as well as an online photo gallery and file storage. A web-based calendar that can be accessed from any computer connected to the Internet, Mac or Windows. What makes it more unique and potentially more useful is that it can also be accessed from Apple's iPhone, which makes it a truly mobile calendar. As with competing calendars, you can display MobileMe in daily, weekly, or monthly modes. You can also synchronize your MobileMe calendars with Apple's iCal and Microsoft Outlook calendars. Even though MobileMe Calendar doesn't offer much new or innovative (save for the iPhone interoperability, of course), it's bound to be a strong competitor in the online apps market, especially for non-business users. That's partly because of Apple's cachet and partly because Apple does tend to get the details right. MobileMe Calendar looks and feels a little slicker than all its competitors, Google Calendar included.

AOL Calendar

America Online isn't quite the powerhouse that it used to be, but it still has millions of users, both paid subscribers and free web users (Figure 4). Any registered user can access AOL Calendar (calendar.aol.com), which integrates with the AOL Instant Messenger (AIM) service for both instant messaging and email. As with competing calendars, AOL Calendar lets you share calendars with authorized users; your calendars can be either private or public.

Cloud Computing

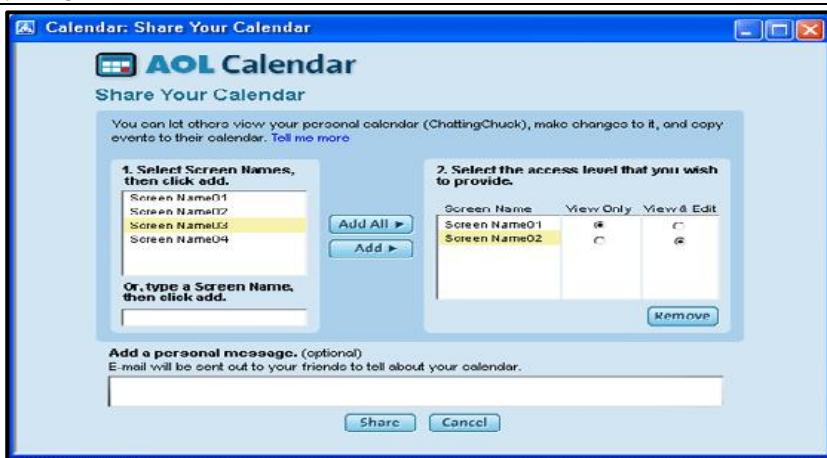


Figure 4: AOL Calendar

CalendarHub

Beyond Google, Yahoo!, Apple, and their ilk, many independent sites offer full-featured web-based calendars. Perhaps, the most notable of these is CalendarHub (www.calendarhub.com) (Figure 5). It offers all the features found in the previously discussed web-based calendars—private/public calendars, sharing/collaboration, multiple calendars, task-based to-do lists, and the like. It lets you publish calendars on your blog or website, which makes it great for creating sites for community groups, sports teams, and the like. The other users can sign up to receive email notification of new events, or subscribe to RSS feeds for any calendar view. It is completely free.

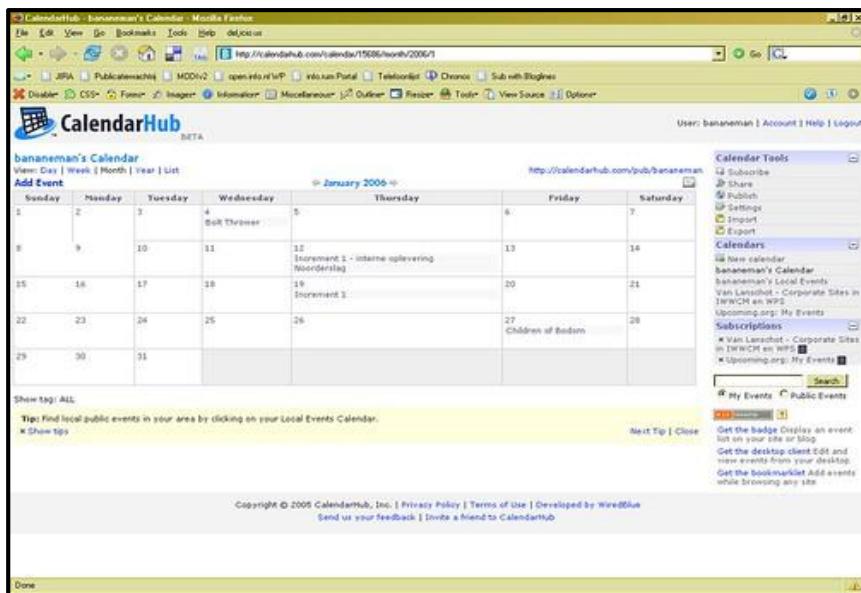


Figure 5: CalendarHub Interface

Hunt Calendars

Hunt calendars (www.huntcal.com) are an event-based web calendars that include useful features like email reminders, notification of event conflicts, notification of new and updated events, and the like. It lets you add web links and images to calendar events, which is fairly unique. Also nice is the ability to customize the color scheme and graphics to reflect your organization's look and feel. They particularly attractive to businesses and community groups.

Famundo

If you keep the schedule for a community group, check out Famundo (www.famundo.com) that offers Famundo for the organizations, a free web-based calendar ideal for schools, churches, sports teams, and the like (Figure 6). After the public calendar has been created, users can subscribe to be

Unit 07: Collaborating on Event Management

notified of new and upcoming events. You can also add message boards, blogs, and other features to your calendar. It also offers Famundo for Families, a personal version of their Organizations calendar. This version includes a family address book and message board, to facilitate family communication.

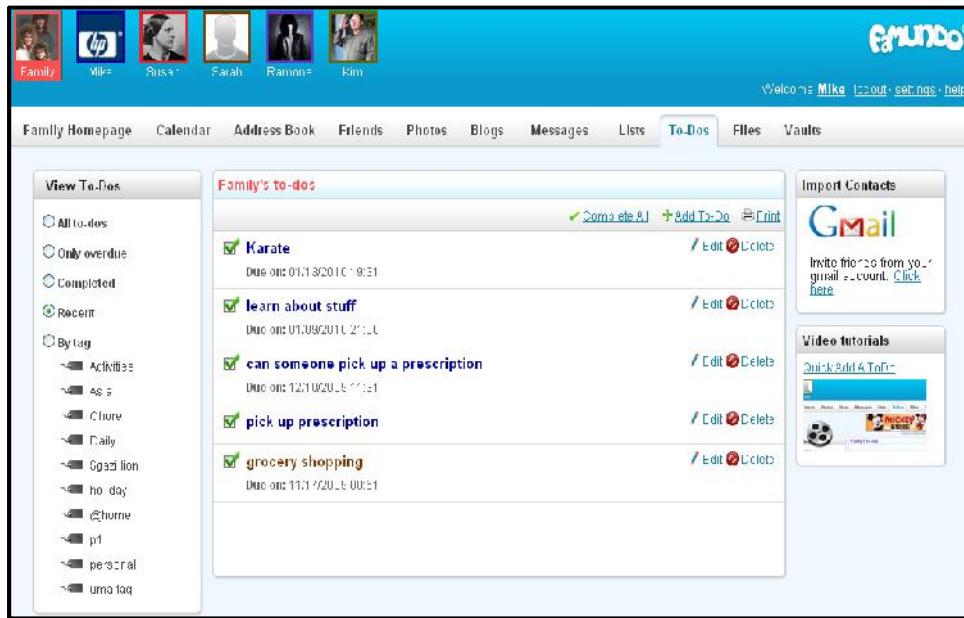


Figure 6: Famundo Interface

eStudio Calendar

eStudio Calendar (www.same-page.com/calendar-software.html) is designed specifically for business use. You get three types of calendars in a single interface:

- Member Event calendar helps users manage their personal time, keep track of meetings with others, and so on.
- Team Event calendar is used to schedule activities for a group, as well as schedule facilities.
- Supervisor calendar provides reports to managers about business activities and schedules.

It can be used to broadcast information about group activities (via email) and to schedule meetings. Information about company events can also be automatically published to your website.

Trumba

Trumba (www.trumba.com) offers web-based calendars ideal for community organizations, schools, and similar public entities. The company lets you embed individualized widgets (dubbed "spuds"). These widgets let users view full calendars, add events to the schedule, receive email notification of events, and such.

Calendars Net

Calendars Net (www.calendars.net) is a free web-based calendar designed for companies or individuals who want to add interactive calendars to their websites. A typical calendar fits into a frame on your website, with little coding required. The site also hosts personal calendars in the cloud. You can employ four different levels of security (so that different users can view the calendar), add events, edit events, and even change universal calendar settings.

Jotlet

Another way to add web-based calendar functionality to your website (www.jotlet.net) (Figure 7). It has a JavaScript API and library that you can use to build rich calendar functionality into any web page. If you're skilled in HTML programming, this is a good way to build a calendar-based

Cloud Computing

page. The JotletAPI is free for non-commercial use, and also available (for a fee) for commercial sites.

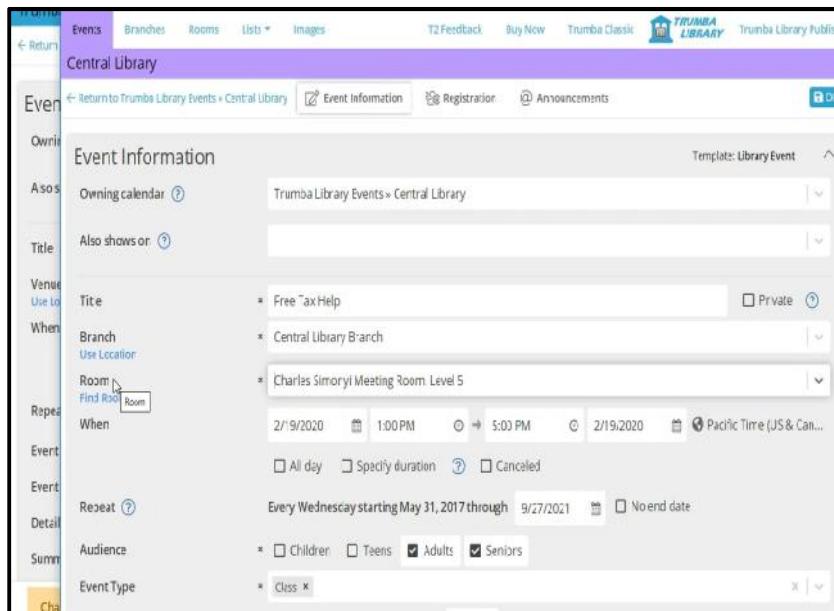


Figure 7: Jotlet Interface

7.1 Implementation of Google Calendar

Google Calendar makes it easy to manage all your significant events in one place. As Google Calendar is connected to your Google account, all of your events and settings will be available anywhere you sign in—whether you're using a computer, phone, or tablet.

Open Google Calendar: One of the best things about Google Calendar is that you can use it on nearly any device (Figure 8). If you're using a computer, visit <https://calendar.google.com> in any web browser. On a phone or tablet, tap the blue-and-white calendar icon labeled "Google Calendar" (iPhone/iPad) or just "Calendar" (Android). If Google Calendar isn't on your Android, download it from the Play Store. If you're using an iPhone or iPad, you can download Google Calendar from the App Store.

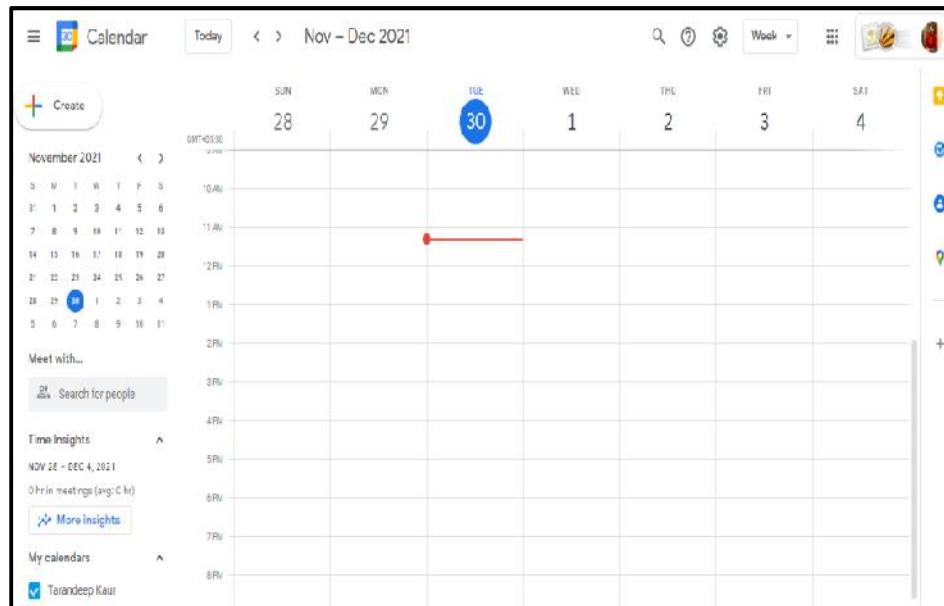


Figure 8: Opening Google Calendar

Log in with your Google account: If you're already signed in, your calendar will appear. Otherwise, you'll be prompted to sign in now. If you use Google Calendar on the web or elsewhere,

Unit 07: Collaborating on Event Management

signing in with the same Google account now will automatically sync any existing calendar events. Visit <https://accounts.google.com/signup> to create a Google account if you don't already have one.

Customize your calendar's look and feel: To open your settings, click the gear icon at the top-right corner (on a computer) or tap the three-line menu at the top-left corner (phone or tablet), and then select Settings. Tap General on your phone or tablet to change the starting day of the week, time zone, default event settings, and more. If you're using a computer, Settings opens to the General tab, on which you can adjust your language, time zone, default event settings, and starting day of the week. Use the menu on the left to see additional settings. When using Google Calendar on a computer, you can customize some color and size options by selecting Density and color from the settings menu.

Choose which dates to see: Your calendar will open to the current day, week, or month, depending on your settings (Figure 9). Here's how you can switch the view as needed:

- o Mobile: Tap the three-line menu at the top-left corner, then select Day, 3 Day, Week, Month, or a different option to switch the view.
- o Computer: Click the drop-down menu at the top-right corner (it should say Month by default) and select Day, Week, 7 days, or a different option.

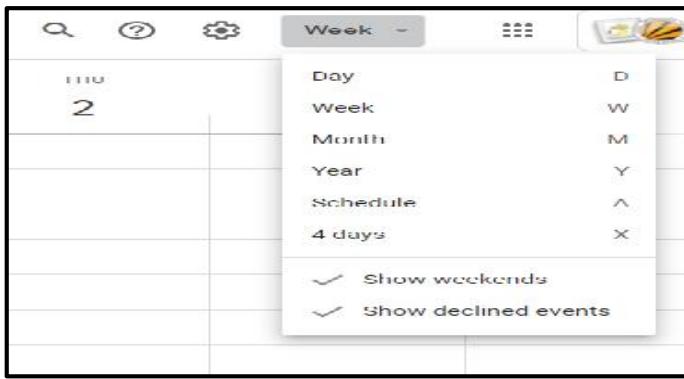


Figure 9: Calendar View Customization

Setting Views of a Calendar: You can see different views of a calendar such as holiday view, that is the number of holidays that are existing for a particular month. So, if you want to see a holiday view, you will see all the holidays in India (Figure 10). For a particular month, you can see the entire holidays in a month. If you want to see other calendars, you want to hide it; you can hide it as well by unchecking the holidays option.

For birthdays view, you can see the birthday of the month. You have an option for adding the reminders for adding the tasks, but you just want to see the calendar. In case you don't want to see any calendar, you can remove or can change the color for it.

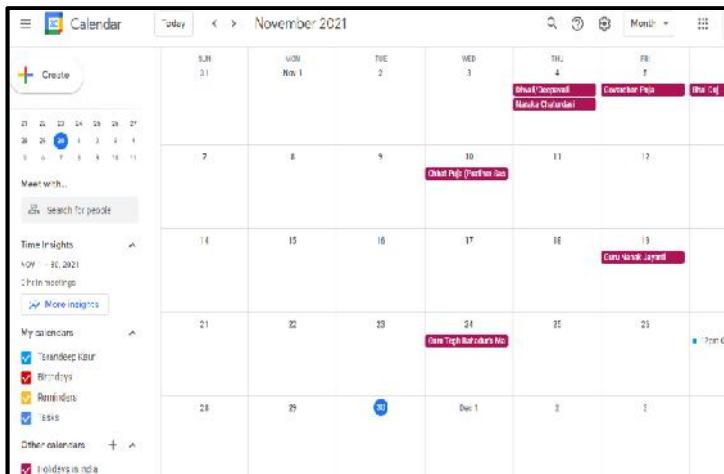


Figure 10: Setting Views of a Calendar

Cloud Computing

Creating an Event: You can also create an event by selecting the date on which the event will occur. You can add the title for the event and set the timings for the same (Figure 11). In order to create an event that's similar to an existing event, you can use the Duplicate feature. Click or tap an event to open it, select the three-dot menu, then select Duplicate.

Click or tap the dates and times to make your selections. The time zone is set automatically based on your computer, phone, or tablet, but you can adjust it if necessary. If the event will take up an entire day (or entire set of days), select "All day" at the top. If the event repeats more than once, you can set a schedule. Click or tap Does not repeat (you may have to tap More options first if you're on a phone or tablet) and choose the desired schedule. Tap Custom if you need to enter a repeat schedule that's more specific.

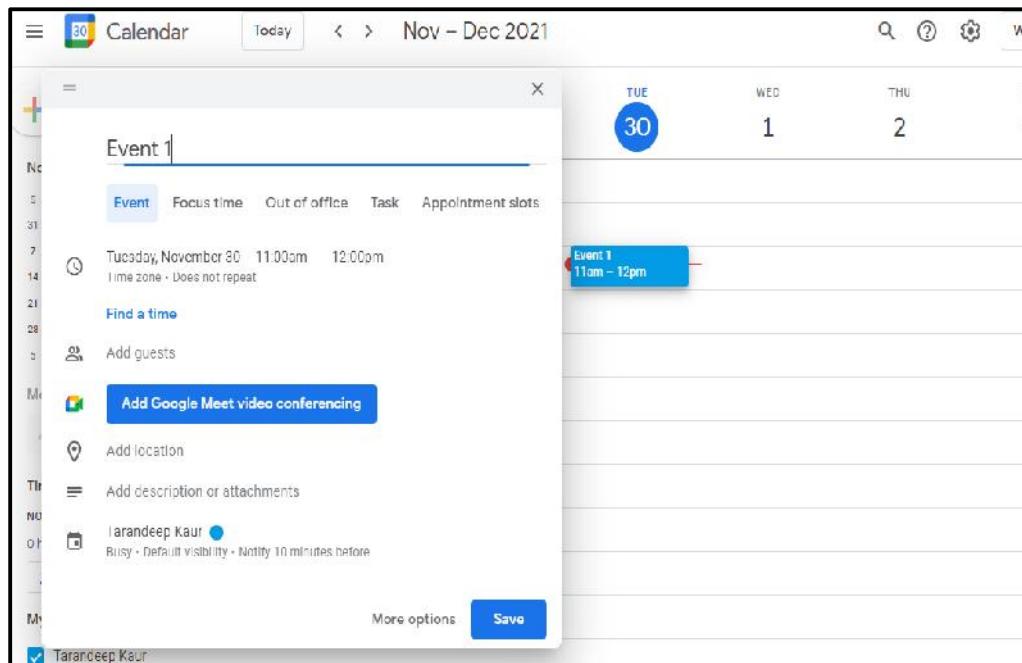


Figure 11: Creating an Event

Adding location to an Event: This is not required, but clicking or tapping Add location lets you enter an address or other location details so you can easily pull up directions in Google Maps if needed.

Create a notification for the event: Google Calendar will automatically send you a notification one day before the event, but you can change that if you wish. Use the menus to choose a notification time on your computer, or tap 1 day before in the mobile app to choose an alternative time.

Adding guests: If you want others to attend this event, you can add them now or share the event with them later. To add guests now:

- Tap Add guests in the mobile app, or click Add guests on the right side of the page on a computer.
- Select or enter contacts to invite. You can also adjust guest permissions, such as whether guests can invite others or view the guest list.

Adding Web-Conferencing to an Event: A user can add a web-conference to an event using Web-Conferencing link (Figure 12).

Unit 07: Collaborating on Event Management

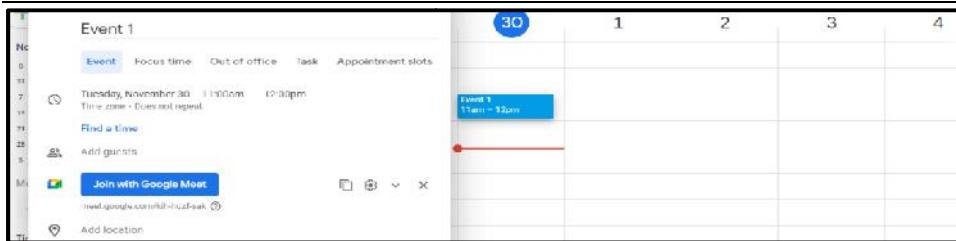


Figure 12: Adding Web-Conferencing to an Event

Add notifications/ Email notifications: To help remind you about upcoming events, you can get notifications on your phone, computer, or by email. You can change your notification settings for one event or multiple events. You can get event notifications in:

- Emails.
- Desktop notifications that appear outside of your web browsers. Your calendar must be open.
- Alerts that appear inside your Google Calendar window.

Setting All Day Event: Sometimes companies will host all day training sessions during the week covering a new implementation in the workplace. Other times departments will be in all day meeting sessions to set certain goals to be met during the quarter. When you are scheduling these events on your calendar, the meeting or training session may begin at 9am and end at 5pm. That's a large chunk of time to be blocked off on your calendar and if your colleagues are subscribed to your Google Calendar. This event may also get in the way of other appointments that you may have scheduled on that day. To ensure that you are not missing any previously scheduled events on that day and to make sure that you aren't 'event bombing' your calendar where others may be distracted, simply create an 'All Day' event. When you arrive at the 'Event Details' screen and you've entered all the information for the event, simply select 'All Day' underneath the title of the event (Figure 13). This option simply removes the start time and end time from the event.

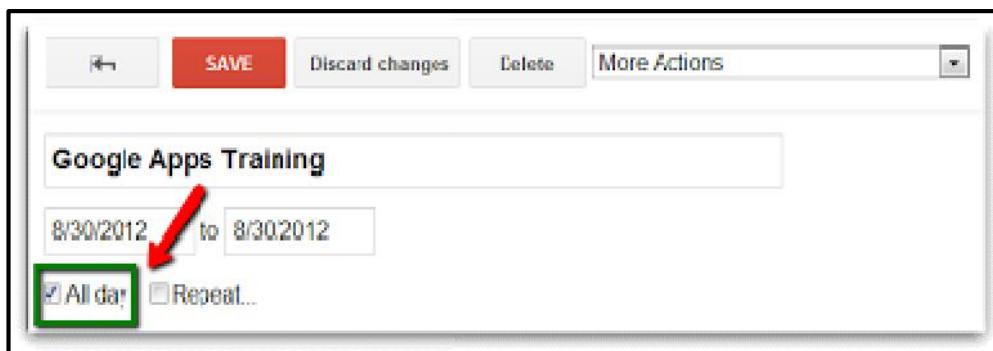


Figure 13: All-Day Event Option

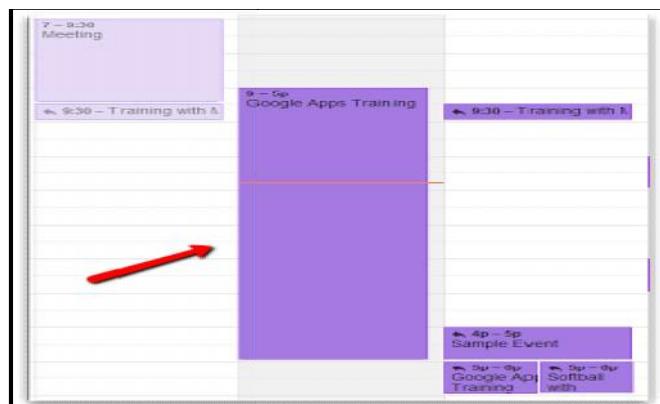


Figure 14: All-Day Event

Cloud Computing

Setting Colors of Events: Google Calendar makes it easy to keep track of all of the important events in your life. From keeping track of your friends' birthdays and anniversaries to ensuring you don't miss dentist appointments; you can create and sync multiple calendars across devices so that you're always on top of your schedule. One of the best ways to keep your calendars organized is to change the color of your Google Calendar, or specific events (Figure 15).

1. Go to Google Calendar homepage on your PC or Mac computer and sign into your Google account. On the left-hand side of the screen, locate the calendar you want to change the color of, and hover your mouse over it.
2. Click on the three vertically stacked dots to reveal a list of options – including a palette of colors.
3. Choose your desired color by clicking on one of the choices available. If you want to create a custom color, you can click on the "+" sign to see additional options.

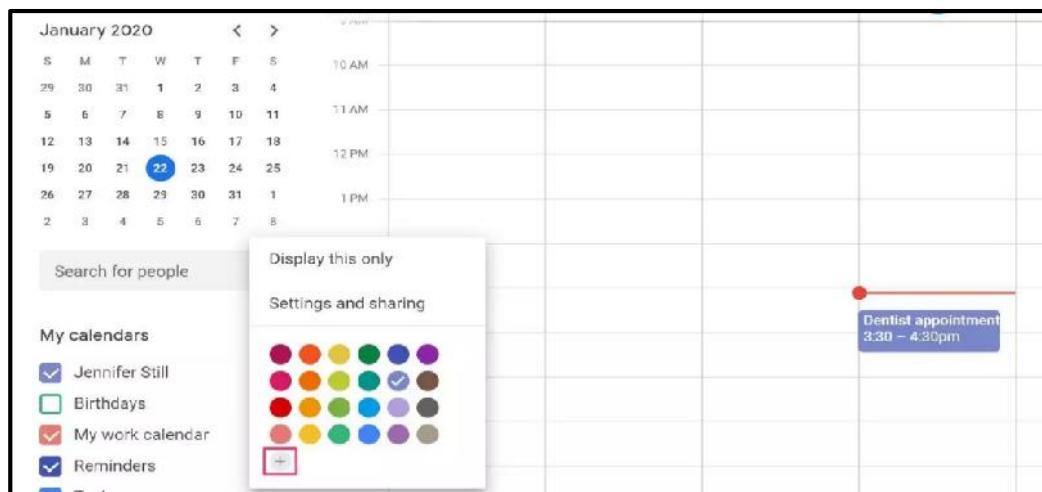


Figure 15: Setting Colour of Events

While changing your Google Calendar color will change all of the events in your calendar to that color, you can choose to change the color of individual events by following these steps.

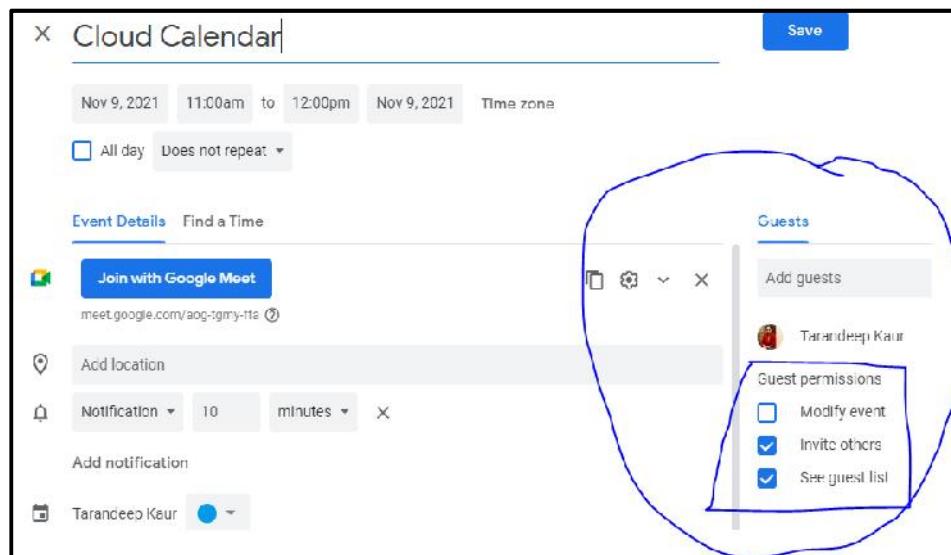


Figure 16: Guest Permissions

Giving Guest Permissions: A guest added in a calendar has certain privileges such as (Figure 16):

- o Invite people to your Calendar event

Unit 07: Collaborating on Event Management

- o Invite people using Google Groups – You can invite a Google Group to your event instead of inviting people one by one. Just follow the steps above and add the email address for the group. Learn more about inviting groups to your event.
- o If someone has shared their calendar with you, or is part of your work, school, or other organization, you can check if they're free or busy when adding them to an event.

If you have permission to add guests to an event, you can forward the invite to new guests by email. When new guests respond to the invite, they'll be added to the guest list. They'll also be able to invite others. If you don't want guests to be able to invite others, edit your event and under "Guests can:" uncheck the box next to "invite others."

7.2 Collaborating on Schedules

When it comes to coordinating multiple individuals or families in a community activity, you have your work cut out for you. Whether it's a youth sports team, community organization, school event, or some community event, trying to line up who's free and who's not on a given duration of time takes a lot of effort – unless, that is, you're using web-based scheduling tools.

Web-Based Scheduling Tools

Web based scheduling software is an online business application specially designed to simplify the internal organization process of all daily business activities within mid to large size companies. Web based scheduling software applications were basically developed for service-based businesses which were required to deal simultaneously with an impressive number of customers. A web-based scheduling software has a positive distinctive feature which permits the appointing service to be externalized so that customers might check online the availability of a particular service and might operate on their own an appointment or a reservation. This scheduling systems allow internal users to keep track of day-to-day activities and to revise in due time schedule as cancellations are easily registered. Examples:

- o Sports Team Schedules
- o School Schedules
- o Community Group Schedules
- o Event Schedules and Management

Sports Team Schedules

One of the most common community activities are youth sports. Maybe your son is on an intramural football team, or your daughter is into youth soccer. Whatever the age, whatever the sport, there's a lot of activities that need to be scheduled – practices, home games, away games, team meetings, you name it. Multiply the number of players (and coaches) on each team times the number of events, and you see the complexity; it gets even worse if you're trying to manage events for an entire league. How best, then, to juggle all the schedules of so many people and events? The best way is to use a web-based calendar tool, such as Google Calendar (calendar.google.com), Yahoo! Calendar (calendar.yahoo.com), or CalendarHub (www.calendarhub.com). Just create a public calendar and provide the URL to all the team members. After you add all your team activities to the calendar, team members simply have to log in to see what's coming up this week and next.

There are also dedicated sports team website builders. These sites offer tools designed specifically for sports teams, including home pages complete with schedule, roster, player profiles, box scores, and the like. Most of these services even design your site using your team colors and logo. Several of these web-based applications, including eteemz (www.eteemz.com), League Athletics (www.leagueathletics.com), League Lineup(www.leaguelineup.com), and TeamSnap (www.teamsnap.com). Typical League Athletics site, for the Lakeville (Minnesota) Baseball Association

School Schedules

Web-based calendars are also ideal for keeping track of various school schedules. Whether it's homework assignments for a particular class or a schoolwide events schedule, it's easy for a teacher or school to post that schedule on a web-based calendar. Make the calendar public (but make sure only authorized personnel can post new events), and then provide the calendar's URL to all

Cloud Computing

students and parents. Assuming that all families have Internet access (it helps to check this first), there should be no excuse for missed homework or absence from key events.

Community Group Schedules

Any community group can benefit from organizing their activities via a web-based calendar. Want to schedule practices for a community theatre production? Announce meetings for your local school board? Organize bingo nights for your church? Any and all of these group activities can easily be managed online, in the cloud, using a web-based calendar.

Event Schedules and Management

You can also use web-based calendars to post dates and schedules for specific public events, such as school plays, or for all events in a given community. Although any web-based calendar program can do this job, as well, some event-specific applications are worth noting. For example, Zvents (zvents.com) is a web search engine for local events. Upload your event schedule into the Zvents database, and then anyone in your area can find out what's happening in the coming days and weeks. Users can also search for events by type, location, and date. Also interesting is the suite of event management software from ServiceU (www.serviceu.com). It is included in this suite is the EventU application, which offers event, resource, and facility scheduling for organizations small or large.

Collaboration on Scheduling Meetings

Scheduling a meeting can be a frustrating experience. Not only do you have to clear time from all the attendees' individual schedules, you also have to make sure that the right-sized meeting room is available at the designated time. The experts claim that it takes seven emails or voice mails to arrange a single meeting; a typical businessperson can spend more than 100 hours each year just scheduling meetings.

If you work in a large company, you know what a hassle it is to schedule even simple meetings. One person is free on Tuesday at 9:00, but another is out of the office, and yet other attendee is already booked for that time slot. Plus, you're not even sure which meeting room is free at that time. You end up sending a flurry of emails back and forth, trying to find the one spot in everybody's schedules that is free. There has to be a better way. That is, Web-based Scheduling.

The web-based scheduling is part of online scheduling applications. The web-based app takes much of the pain out of scheduling meetings, for both large and small groups. Such apps require all users to enter their individual calendars beforehand. When you schedule a meeting, the app checks attendees' schedules for the first available free time for all. Everyone places his or her schedule in the cloud, which then enables the meeting's organizer to easily see who's available when. The cloud-based app finds the best time for all involved and schedules the meeting. No more emails, no more phone calls; it all happens automatically, in the cloud. It generates automated email messages to inform attendees of the meeting request (and the designated time), followed by automatic confirmation emails when attendees accept the invitation.

Example: Let's suppose you need to schedule a meeting sometime next week with a dozen different attendees. You punch in the details of the meeting and the desired attendees, and the scheduling app finds the first available timeslot when all attendees are free.

Alternatively, the app might have to pick a timeslot when the maximum number of people can attend. This kind of "best case scenario" scheduling might be the only way to get your meeting on the calendar in a reasonable period of time.

Web-based scheduling programs let you schedule both in-person meetings and teleconferences with attendees from multiple locations. You're not limited to just those people located in your office; you can work with the schedules of people around the country and even in different firms. Much of this can be accomplished with simple web-based calendar programs, such as Google Calendar (calendar.google.com) and Yahoo! Calendar (calendar.yahoo.com). In order to take the advantage of the more advanced automatic scheduling features, however, you need to use an industrial strength scheduling application, such as AppointmentQuest (www.appointmentquest.com), hitAppoint (www.hitappoint.com), and Schedulebook (www.schedulebook.com). Naturally, these enterprise-level apps cost more to use than the free web-based calendars; expect to pay anywhere from \$20 to \$200 subscription fees per month.

Professionals who schedule appointments with their clients—doctors, lawyers, hairdressers, and the like—face similar scheduling challenges. For this purpose, separate web-based appointment scheduling applications exist. These apps function similarly to traditional meeting schedulers, but

Unit 07: Collaborating on Event Management

with a focus on customer appointments. Let us explore the different online scheduling applications in detail.

Jiffle-Jiffle (www.jifflenow.com) (Figure 17), which schedules meetings, appointments, and the like for the enterprise environment. To track employees' free time, it synchronizes seamlessly with both Microsoft Outlook and Google Calendar. It also offers its own Jiffle Calendar application. Jiffle allows the originating user to mark available time slots on his calendar, and then share them with proposed attendees via a Jiffle-generated email invitation. These attendees view the invitation, log in to the Jiffle website, and then select their preferred time slots from the ones proposed. Based on these responses, Jiffle picks the best time for the meeting and notifies all attendees via an automatic confirmation email. For smaller companies, Jiffle is free for up to 10 meeting confirmations per month. For larger companies, Jiffle Plus, Jiffle Pro, and Jiffle Corporate plans are available.



Figure 17: Jiffle Interface for Scheduling a Meeting

Presdo- Unlike Jiffle, Presdo (www.presdo.com) is a scheduling tool that isn't limited to a single company (Figure 18). Presdo lets you schedule meetings and events with anyone who has an email address. Adding an event is as simple as entering a description into a box. You then enter the email addresses of other participants, and Presdo emails out the appropriate invites. When an attendee responds, he's automatically added to the event's guest list. (And, for the convenience of all guests, it's a one-button process to add an event to a user's Microsoft Outlook, Google Calendar, Yahoo! Calendar, or Apple iCal calendar.)

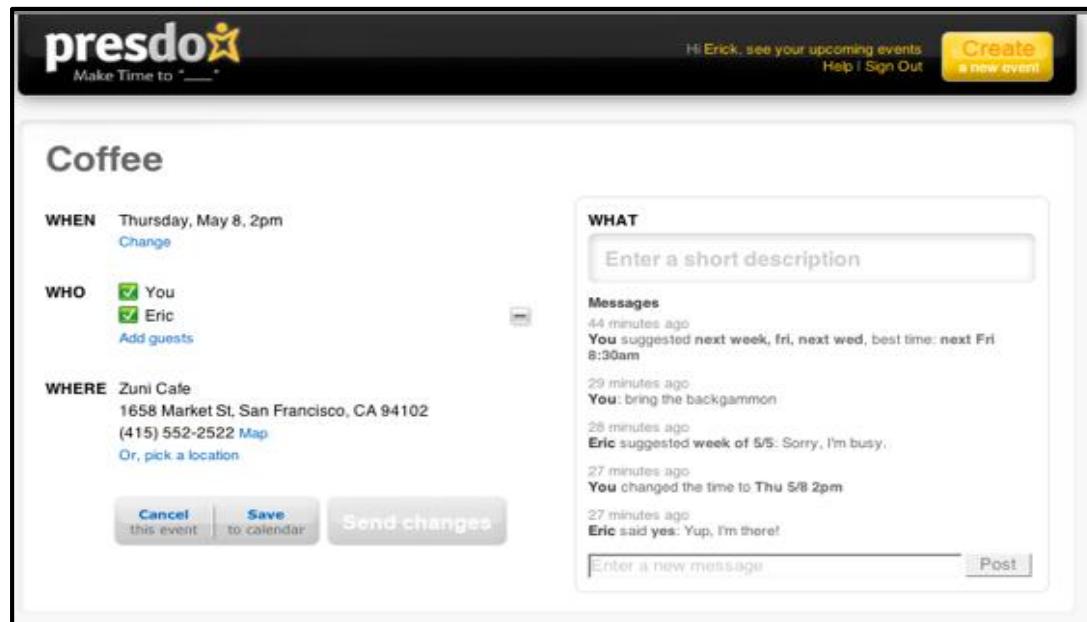


Figure 18: Presdo Interface

Cloud Computing

Diarised-Diarised (www.diarised.com) is, like Presdo, a web-based meeting maker that users across different companies can use. It helps you pick the best time for a meeting by sending out emails to invitees, letting them choose the best times for them, and then sending you a summary of those best dates. You pick the final date, Diarised notifies everyone via email, and your meeting is scheduled.

Windows Live Events- Event scheduling is now part of Microsoft's bag of tricks. Microsoft's Windows Live Events (home.services.spaces.live.com/events/) is a customized version of its Live Spaces offering; it lets Live Spaces users organize events and share activities between participants. To schedule an event, you set up a list of invitees and then send out a mass email with a link back to your Live Event site. (All the event details are also available as an RSS feed.) Information about the event is posted on the site itself, which also serves as a place for attendees to come back after the event and share their photos, videos, and blog posts about the event. With its user-friendly consumer features, Live Events isn't robust enough (or professional enough) for most business users. It is, however, a nice way to plan more personal and informal events.

Schedulebook- Schedulebook (www.schedulebook.com) offers several different types of web-based scheduling services. Depending on the application, you can use Schedulebook to schedule employees, customers, or other interested parties. The company's three offerings are:

- Schedulebook Professionals, which is a business-oriented schedule/calendar/planning application.
- Schedulebook Office, which schedules the use of any shared resource, such as company meeting rooms or even vacation homes
- Schedulebook Aviation, which is used by the aviation industry to schedule aircraft, flight training, and similar services.

Acuity Scheduling- If you run a business that requires scheduling appointments with clients or customers, Acuity Scheduling (www.acuityscheduling.com) (Figure 19) can help ease your scheduling operations. Acuity Scheduling lets you clients schedule their own appointments 24/7 via a web-based interface, you don't have to manually schedule any appointment.

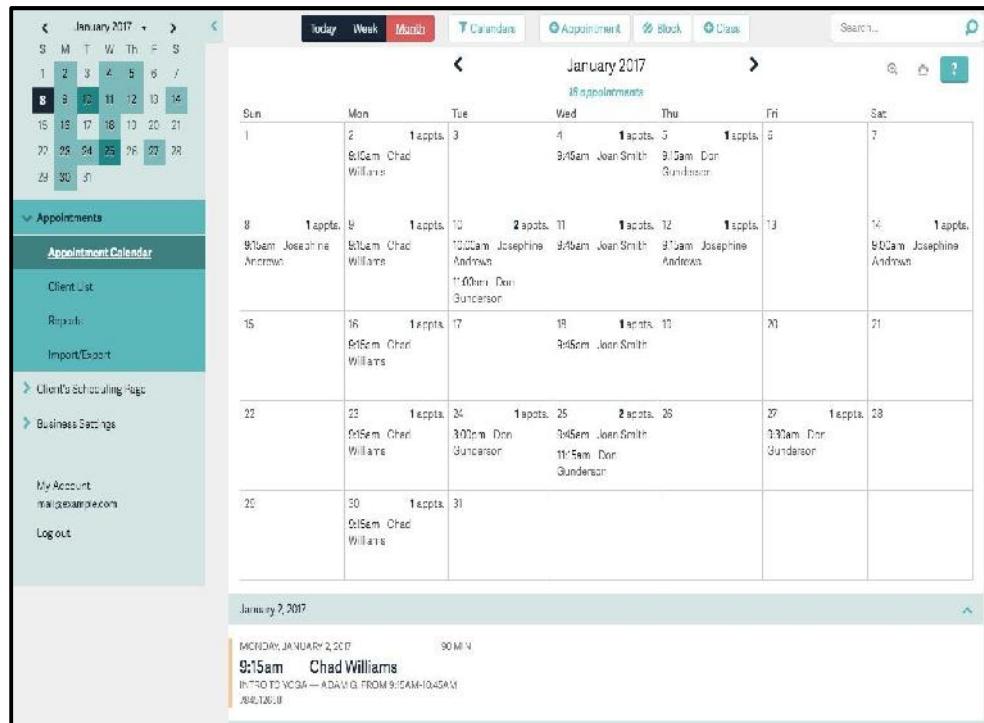


Figure 19: Typical Client Scheduling Screen from Acuity Scheduling

AppointmentQuest-Like Acuity Scheduling, AppointmentQuest (www.appointmentquest.com) is designed to solve the scheduling problems of busy professionals. This application not only

Unit 07: Collaborating on Event Management

enables clients to make and you to accept appointments over the web, it also lets you manage personnel, schedules, and other calendar-related items.

hitAppoint-hitAppoint (www.hitappoint.com) enables online client booking. Like the previous similar application, it's ideal for any business that requires the making of customer appointments – barbershops, hair salons, doctor and dentist offices, consultants, financial advisors, car repair shops, computer technicians, and the like.

7.3 Collaborating on To-Do Lists

Let's start with simple task management, in the form of the old-fashioned todo list. These are web-based lists that multiple group members can access from any web browser. Tasks are entered (complete with due date) and checked off when completed. The different To-do list apps includes:

Taskade:Taskade is a holistic project management platform and the only collaborative to-do software that offers true, real-time (as opposed to push/pull) synchronization! That means every edit and update is instantly reflected in your shared task lists.Taskade is also one of the few project management platforms out there that offer 100% free, integrated video conferencing, calls, and chat (Figure 20).

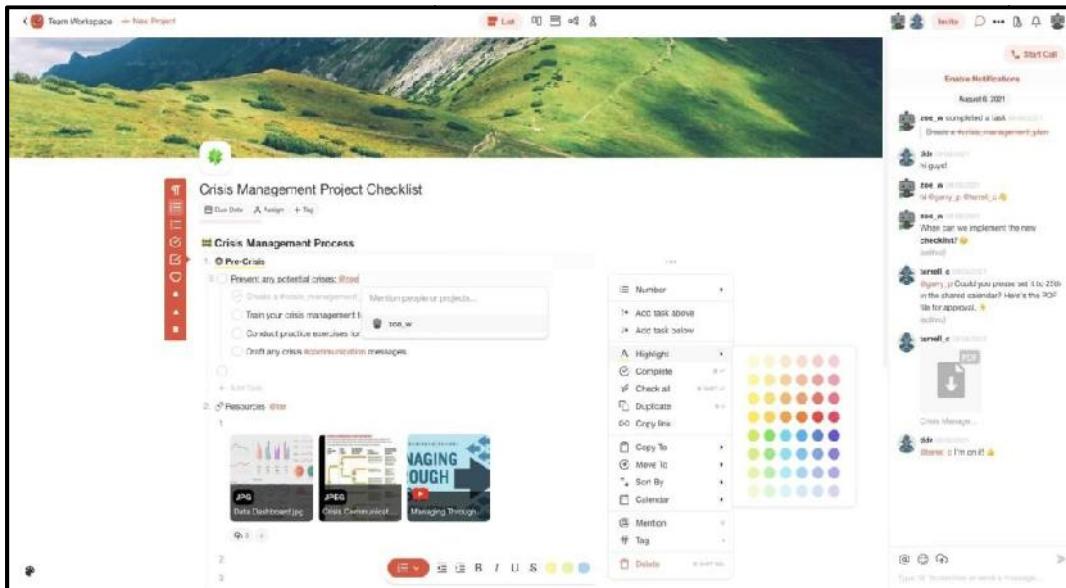


Figure 20: Taskade Interface



Did you know?

The best part with using Taskade is that your to-do list and team conversations happen in the same window, at the same time.

Todoist:Todoist lets you share lists, assign tasks, and integrate with Twist for a convenient team chat. Thanks to a solid set of task and project management features, Todoist remains one of the best collaborative to-do software in the wild.

Asana:Asana is another notable collaborative to-do software that has been around since 2008 (Figure 21). A brainchild of Dustin Moskovitz and Justin Rosenstein, Asana lets you create aesthetic lists topped with the usual set of features you'd expect in a solid to-do app.

Cloud Computing

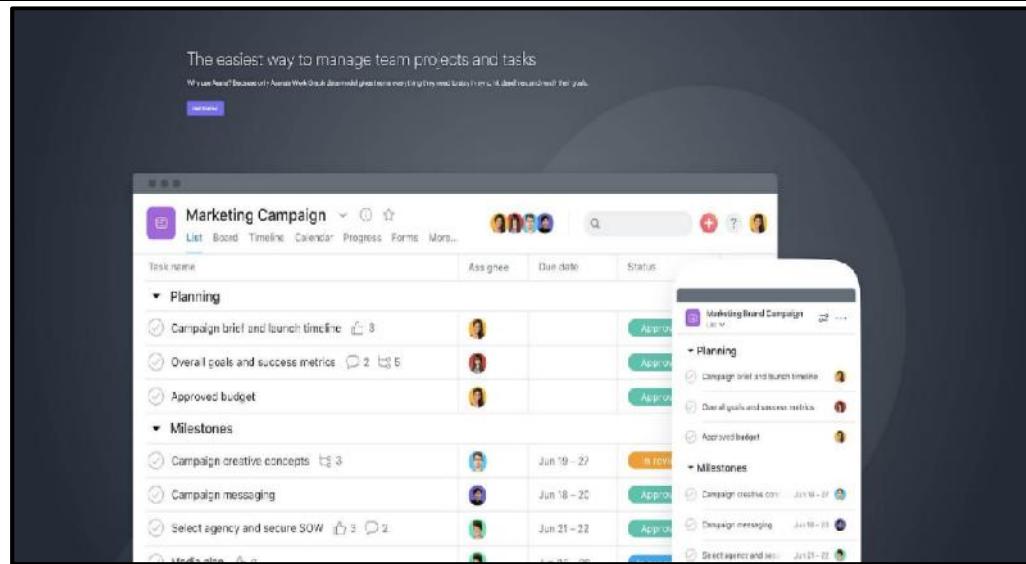


Figure 21: Asana Interface

TickTick: TickTick is a minimalist to-do list app with collaboration features. A successor to 2010 GTasks, the app lets you share lists, assign tasks, and check progress inside a simple and distraction free environment. A good match for those looking for a frictionless to-do experience.

Any.do Teams: Any.do Teams is an offshoot of the popular Any.do to-do list app with a number of collaborative features sprinkled on top. The Teams subscription unlocks advanced list and file-sharing options, integrations, tagging, as well as priority support.

Monday.com: Monday.com is a complete package for advanced project management. Apart from its core to-do list functionality, Monday offers advanced automation and integrations as well as communication, planning, and collaboration tools for all types of projects (Figure 22).

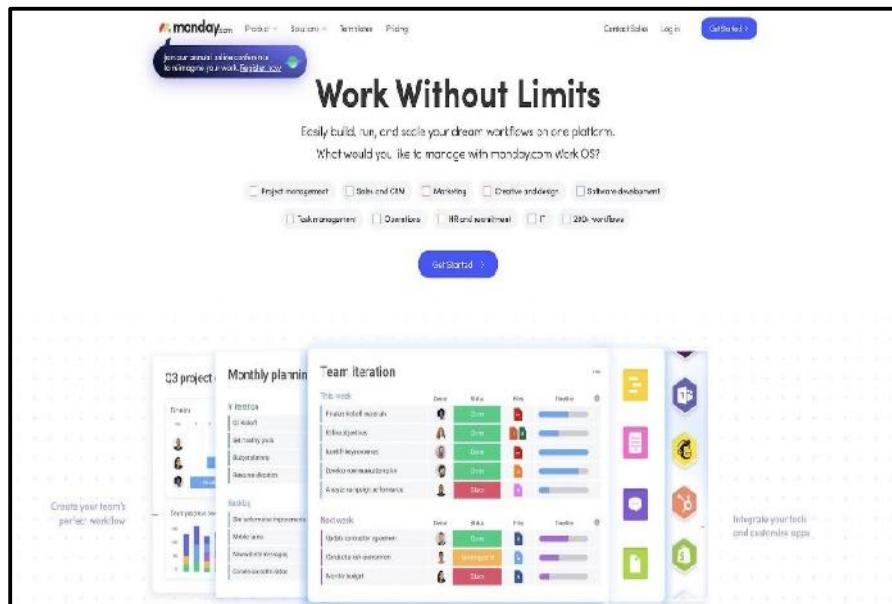


Figure 22: Monday.com Interface

Wrike: Wrike is an enterprise-grade project management tool and collaborative to-do software. The platform can handle anything from simple lists to automated workflows and AI-backed data insights. Wrike is a solid choice for large organizations.

Google Keep: Keep is part of Google's Workspace suite (previously G Suite). The app complements the productivity/collaboration platform with basic note-taking and to-do list capabilities.

Wunderlist/Microsoft To-Do: Microsoft To-Do is a simple app with all the essential to-do features. To-Do allows you to share lists with other users or integrate with Outlook and collaborate within the Microsoft 365 platform.

7.4 Collaborating on Task Management

For managing more complex tasks, a simple to-do list application might not cut the mustard. Instead, using a web-based task management application that lets you manage the multiple pieces and parts of large projects can be considered. The basic task management can be accomplished with applications such as HiTask (www.hitask.com) and Zoho Planner (planner.zoho.com). For the most complicated projects, consider using a dedicated project management application, such as Basecamp (www.basecamphq.com) or Goplan (www.goplan.com).



Notes: Whichever option you choose, you must break your project down into all its component tasks, set a due date for each, and then assign each task to a specific group member.

The group members can then access the application online from their own computers and update the project with their own individual progress. The planning and task applications let you manage everything from simple to-do lists to complex group tasks, all over the Internet and collaboratively with other users.

iPrioritize: The sharing of to-do lists is important for families, community groups, and businesses. Your to-do list might be as simple as a grocery list or as complex as a list of activities for a community program or business project. Whatever the application, iPrioritize (www.iprioritize.com-Figure 23) is a good basic to-do list manager. The authorized users can create a new to-do list, add items to the list, prioritize tasks by dragging them up and down the list, and mark items complete when finished. And, because it's web based, you can access your lists anytime and anywhere. When you have a list, you can print it out, email it to someone else, subscribe to changes in the list via RSS, and even view lists on your mobile phone – which is a great way to consult your grocery list when at the supermarket!

The screenshot shows the iPrioritize web application interface. At the top, there's a header with the logo 'iPrioritize' and the tagline 'The efficient way to get organized'. Below the header, a welcome message 'Welcome, Michael Miller' and a 'log out' link. On the left, a sidebar titled 'Michael's Lists' shows a single 'Grocery list' entry with a link to 'manage lists'. The main content area is titled 'Grocery list' and contains three items: 'Milk' (unchecked), 'Butter' (unchecked), and 'Bread' (unchecked). Each item has an 'edit' button and an 'X' button. At the bottom of this list is a button labeled 'Add New Task'. To the right of the list is a 'Options' sidebar with links for 'Print', 'Email', 'RSS Feed' (with a checked checkbox), 'Mobile', and 'Share'.

Figure 23: iPrioritize Interface

Bla-Bla List: Bla-Blalist (www.blablist.com) is another simple to-do list manager. It's web based, of course, so you can access your lists from any location at any time. You can even publish your lists via RSS so that family and coworkers can get instant updates.

Hiveminder: Hiveminder (www.hiveminder.com) is similar to all the previously discussed to-do list managers. What's nice about Hiveminder is that you can enter list items in a kind of freeform fashion, and it will help you create and prioritize lists based on your "brain dumps."

Remember the Milk: When you need to "remember the milk" at the grocery store, check out the aptly named Remember the Milk (www.rememberthemilk.com) web-based todo list manager. Once you create a list, you can arrange reminders via email, instant messaging, or text messages to your mobile phone.

Cloud Computing

Ta-da List: Another web-based to-do list manager. Ta-da List (www.tadalist.com) lets you make all sorts of lists, share them with friends, family, and coworkers, and then check off items as they're completed.

Tudu List: TuDu List (www.tudulist.com) is a little different from other to-do list managers in that it also includes a web-based calendar. Items are added both to the appropriate to-do list and to your calendar, on the date they're due.

TaskTHIS: TaskTHIS (taskthis.darthapo.com) is similar to most other to-do list managers, but offers the ability to add extended notes to any individual task. You can publish your tasks via RSS or share with others via the web.

Vitalist: Like other to-do list managers, Vitalist (www.vitalist.com) organizes all sorts of tasks and projects. It's unique in that it uses the Getting Things Done (GTD) workflow methodology popularized by management consultant David Allen.

TracksLife: Trackslife (www.trackslife.com) is a database-oriented task manager. Each "track" is a separate database that combines columns of money, numbers, words, paragraphs, and yes/no responses. The application sends out reminders of critical events via email or RSS.

Voo2Do: Voo2Do (www.voo2do.com) moves beyond simple to-do list management into more sophisticated priority management. This web-based application lets you set up different projects, organize tasks by project, track time spent and remaining on a given task or project, publish task lists, and even add tasks via email.

HiTask: More sophisticated task management can be had with HiTask (www.hitask.com), a business-oriented task manager. Tasks are added to your calendar and color tagged for easy viewing. The task manager and scheduler both utilize drag-and-drop editing, and you can share and assign tasks and projects to a group of people via the web.

Zoho Planner: Zoho Planner (planner.zoho.com) is perhaps the most sophisticated task planner evaluated here. Its features and functionality approach those of the project management applications. With Zoho Planner, you create a new page for each project you're working on. To that project, you add lists with individual to-dos within each list. Each list item can include extensive notes as well as images. You can share each project page with users you designate. Each to do item also appears on your central calendar.

7.5 Collaborating on Event Management

When you're putting on a big event such as a concert or conference, you have a whole new set of challenges to face. Not only do you have to manage the tasks involved with putting together the event, you also have to handle attendee registration, event marketing, ticket sales, and the like. It's a massive effort—made somewhat easier by web-based event management tools. With the web-based event management applications, the cloud hosts everything you need to schedule and market your events, as well as handle registration, payment, and other important tasks. For example, you can create an online event calendar so that attendees can learn about and sign up for future events via the web; offer web-based event registration and payment; manage requests for hotel rooms, airline flights, and car rentals; and check in attendees live onsite via a notebook PC with Internet connection. Most of these apps also offer detailed task and budget management functions. These are very robust applications, capable of handling every last detail over the web.

Some of the most popular of these apps include Cvent (www.cvent.com), RegOnline (www.regonline.com), and ViewCentral (www.rmkr.com/viewcentral). Unlike some other web-based apps, these aren't free; you have to pay for the power you need to manage the details of your particular event. Most event management applications also handle basic event marketing. But when it comes to promoting your community events, you want to go beyond the basics to more creative forms of marketing. For example, you may want to create a brochure or flyer to announce your event. Fine and dandy, but everybody in the group (including all the community bigwigs) wants input on the final piece. This may have been difficult in precloud days, but now you can use a web-based application such as Google Docs (docs.google.com) to create your piece and make it available online for everyone to see and comment on. Naturally, you can also use web-based local search sites, such as Zvents (www.zvents.com), to post announcements of your community events. You may even want to use cloud-based social media sites, such as Facebook and MySpace, to promote your event online. And, after the event, you can post pictures of the event on community photo-sharing sites, such as Flickr (www.flickr.com). It's all possible because of the cloud!

7.6 Collaborating on Budgets

Every event, small or large, comes with its own set of costs. With community events, those costs are often managed by a group of people, each responsible for a specific operation or group of operations. For simple events, you can collaborate on your budget using web-based spreadsheet programs, such as Google Spreadsheets (part of the Google Docs suite, at docs.google.com). You just need to create a private spreadsheet, authorize access for each member responsible for the budget, and then start adding data online. Once everyone has finished entering numbers for their line items, the group member responsible for the entire budget can log on and do her thing.

For larger or more complex events, you may want to use the budget function available in most event management programs. You may also want to consider some of the accounting applications in the Salesforce.com AppExchange (www.salesforce.com/appexchange/). Some of these web-based apps are relatively low-priced, which is an attractive asset for most cash-strapped community groups.

7.7 Virtual Communities in the Cloud

The question of concern is: Is a virtual community a real community?

When we think of community organizations, we tend to focus on those groups operating within the confines of a physical community. But not all communities are geographically based; the growth of the Internet has seen the advent of virtual communities, located solely within the cloud. A virtual community is an assemblage of like-minded individuals, from anywhere in the world, online. The pre-Internet, virtual communities sprung up on online bulletin board systems (BBSs) that were accessible via dial-up connections. In the early 1990s, virtual communities migrated to the commercial online services provided by AmericaOnline, CompuServe, and Prodigy. Then, when the Internet first gained public prominence, even more virtual communities found their home in Usenet newsgroups.

Today, the Internet is synonymous with the web, and almost all virtual communities are web based. Many communities are based in web forums and message boards. Still others coalesce around eBay and other hobbyist-friendly sites. But the biggest home of virtual communities today is Facebook, MySpace, and other social networking sites. These Web 2.0 sites let like-minded individuals create free-flowing communities around users' personal pages; musicians, celebrities, and even normal individuals all create communities of friends around their Facebook and MySpace pages. It is when there is a constant interchange of thoughts and ideas between group members—which is what happens on a popular Facebook or MySpace page.

Community communications don't have to happen face to face; a group can have a discussion that's entirely virtual, thanks to today's cloud-based communications tools. And, to be honest, many online communities are more vibrant than some physical communities. After all, the physical proximity doesn't guarantee a commonality in interests or a harmony in dispositions. It might actually be easier to find people you like and have something in common with online than it is in your own local community. That is, perhaps, as good a reason as any to embrace virtual communities—while still supporting your local community, of course.

7.8 Collaborating on Contact Management

Through the use of contact management, a solution for managing a set of collaborative contacts is provided. Each collaborative contact includes one or more collaboration capacities, each of which corresponds to a unique collaboration tool (e.g., voice over IP, instant messaging, email, and/or the like). Identifying text can be used to search the set of collaborative contacts and obtain a set of candidate contacts.

Managing Contact Lists

Salespeople have to deal with lots and lots of contacts. Not only is their address book full, they need to know when to contact certain clients, when follow-up calls are necessary, what the boss needs them to do today, and the like. This is difficult to do from a simple desktop contact management program, such as Microsoft Outlook, which merely acts as a storehouse for names and contact info. It's also difficult to do when you're on the road and need access to all your contacts.

Cloud Computing

Imagine the typical day on the road for a traveling salesperson. You wake up in your hotel room, turn on your notebook computer, and log in to the hotel's free wireless Internet connection. Using your web browser, you access the CRM application's website and look at today's list of tasks. You click a button to launch a list of scheduled emails to be sent to selected clients, and then scroll through the list of phone calls you need to make today. If you need more information about a particular client, just click that contact's link. Everything you've entered about that client is stored online and instantly accessible. You can even click a link to view driving directions and a map to the first client you need to visit. And, to see how you're doing against this month's quota, you end the session by creating a detailed sales tracking report.

Web-based Contact Management or Customer Resource Management (CRM) Application

Contact management is the process of recording contacts' details and tracking their interactions with a business. Such systems have gradually evolved into an aspect of CRM systems, which allow businesses to improve sales and service levels leveraging a wider range of data. CRMs are tailored to the needs of a busy salesperson and come complete with features such as activity scheduling, appointment reminders, email templates, and the like.

Among the most popular of these applications are BigContacts (www.bigcontacts.com), Highrise (www.highrisehq.com), and the market leading Salesforce.com (www.salesforce.com). Many of these apps include additional functions of use to large sales departments, including expense account management, sales activity reports, and various team management features. Using a web-based contact or CRM application can be as simple or as complex as you make it.

You might need nothing more than access to a large list of contacts from any location; that's the simple usage. On the other hand, you might want to customize the program so that you're automatically flagged each day with a list of accounts to contact (and for what purposes). Some communication can even be automated, via the use of scheduled emails. The key is that everything is stored and managed in the cloud so that you—and your sales manager—can access important contact information from anywhere at any time; all you need is an Internet connection.

How contact management evolved into customer relationship management?

Contact management evolved into customer relationship management as a result of two key developments:

- The shift in focus from contact details to relationships.
- The move from individual desktop databases to sharing information business-wide.

The free contact management software, or packages that came bundled with other business utilities such as spreadsheets, made it easy to store and retrieve contact information. However, the contact management tools and programmes have been integrated into and superseded by CRM systems that can track everything from customers and sales leads to marketing campaigns and sales team performance.



Notes: Every interaction with a contact can now be tracked and connected to actions and reminders for follow-ups. And doing this through a single, shareable database cuts confusion and keeps everyone in sync, automatically

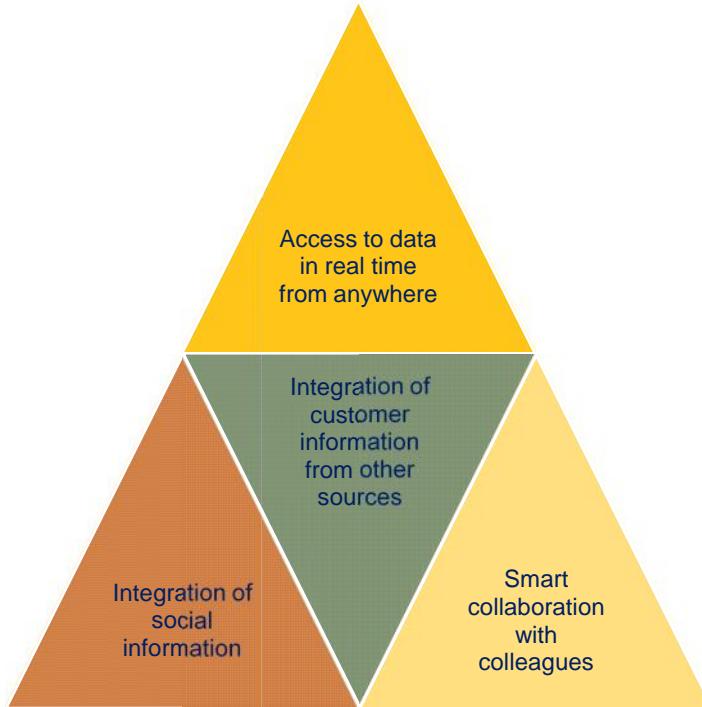
Every interaction with a contact can now be tracked and connected to actions and reminders for follow-ups. And doing this through a single, shareable database cuts confusion and keeps everyone in sync, automatically.

CRM systems make all that contact management information instantly available to others, to the benefit of the whole organization. For example: Salespeople and customer service teams can present a joined-up presence to prospects and customers. Sales managers can monitor the effectiveness of individuals or whole teams. Marketers can measure the effectiveness of their campaigns.

Cloud-based Contact Management and CRM Systems

Unit 07: Collaborating on Event Management

A cloud-based system allows employees to update their system with the latest information about a contact wherever they are, from any web-enabled device. This is possible because all data is stored on a single database based in the cloud, so new and updated information is instantly available to all employees.



Access to Data in Real-Time from Anywhere: With a traditional desktop-based CRM or contact management system, information is only available via a workstation with the software installed – people have to rely on an obsolescent database snapshot. The cloud-based CRM software provides up-to-the-minute information that is available to all instantaneously via any connected device, anywhere. This is because it uses multi-tenancy cloud architecture, which means all users and applications share a common infrastructure that is centrally maintained and updated.

Integration of Customer Information from Other Sources: When a customer is at the centre of a cloud-based contact management system, it's easier to fill in any missing gaps about them, using other, compliant sources. When connected to other cloud-based data sources such as Data.com, the contact management system can automatically fill in the blanks in customer information.

Integration of Social Information-Cloud can also easily be used to analyse the social interactions and conversations that contacts are having on Facebook, Twitter, LinkedIn and other social networks. Is there a problem that a customer is having with a product that they are discussing with others socially? A connected contact management system can give the organizations a real-time view of all those interactions and proactively find a solution. But social networking goes far beyond handling support issues and dealing with problems any contacts may encounter. It is noticeable that by integrating social networking and contact management, it's easy to see who the contacts' contacts are, who they are talking to, who the key players and influencers are who ideally should be contacted, along with an indication of who can provide introductions to them.

Smart Collaboration with Colleagues-While salespeople are often ambitious individuals who've developed a personal style that works for them when it comes to pitching, selling and closing, there are times the business needs to pull together as a team to get a deal over the line. This can especially be the case when:

- a deal is complex
- a deal is in danger of stalling
- The primary account rep isn't an expert in some of the key aspects of the deal.

Cloud Computing**Limitations of Contact Management Systems (CMSs)**

Contact management has quickly evolved from Rolodex and Filofax systems into desktop-based contact management software and email clients with built-in contact management tools. Yet despite these significant technical advances, many businesses found that their CMSs fundamentally amounted to little more than a simple database containing names, phone numbers and notes.

Pobuca Connect-Pobuca offers you a cloud contact management tool to share your contacts with your co-workers and partners. You can convert your multiple business contact lists into one unified address book and access it from everywhere! The Pobuca Relationships plan is a cutting-edge mobile lightweight CRM app that helps you to organize your company's leads, set marketing campaigns and manage your sales pipeline. Meet Pobuca Bot, your virtual assistant that will search and update the stored info of your ad.

SalesRabbit-SalesRabbit's game-changing methodology is based on research they've done with sales pros and the top teams in the industry. It targets the main traction points of every successful organization: rep motivation, team efficiency, and sales mastery. The app is built to help your company make changes that enable peak performance at every door. Their dedicated CSMs and support staff work with you to implement these strategies, ensuring you get the results every door-to-door company hopes for.

Lusha-Lusha helps business professionals establish a fast and true connection with their leads, contacts, and candidates. B2B engagement is based on trust. Lusha helps you build that trust using simple tools which allow you to enrich and verify business profiles. Today, over 250,000 sales representatives, recruitment managers, and marketers transform the way they engage with their leads, contacts, and candidates by using Lusha on a daily basis.

Salesmate-Salesmate helps businesses close more deals faster. It offers an easy-to-use interface and automated workflows let your sales team focus on the right deals at the right time. Never miss a beat and start closing more deals. Some of our hottest features: - Built-in Calling w/ Power Dialer - Email Drip Sequences - Workflow Automation (Auto-assign touchpoints & follow-ups with prospects & clients) Give us a free test ride today and our friendly staff is ready to help you!

Bitrix24-If you are looking for free online contact management software that doesn't limit how many contacts you can have, comes with email marketing, fully featured CRM and contact manager. Bitrix24 is extremely easy to use and comes jam packed with advanced sales tools.

EngageBay-EngageBay is a simple, affordable, all-in-one marketing and sales CRM software built for small businesses and startups to acquire, engage, nurture web visitors and convert them to happy customers. Use EngageBay's simple contact management software to keep all your contacts in one place. Stay organized, make meaningful decisions and close more sales. - all from one easy to use platform.

Salesflare-Salesflare is an intelligent CRM & contact management platform for small and medium sized B2B businesses who sell more with less work. Salesflare automatically fills out your address book and keeps track of all interactions with your contacts. It takes data from email, social media, company databases, phone, calendar and hands it to you in automated customer timelines that tell you everything you need to know. You just have to write the emails, make the calls, and have the coffee.

Salesforce Sales Cloud-Drives growth with Sales Cloud 360, the best-in-class sales solution that drives rep productivity on the world's #1 CRM platform.

Zoho CRM-Zoho CRM empowers organizations with a complete customer relationship lifecycle management solution.

Microsoft Dynamics 365-Integrated business management solution that automates financial, customer relationship and supply chain processes.

Summary

- Lusha helps business professionals establish a fast and true connection with their leads, contacts, and candidates. B2B engagement is based on trust. Lusha helps you build that trust using simple tools which allow you to enrich and verify business profiles.

Unit 07: Collaborating on Event Management

- The free contact management software, or packages that came bundled with other business utilities such as spreadsheets, made it easy to store and retrieve contact information.
- Salesmate helps businesses close more deals faster. It offers an easy-to-use interface and automated workflows let your sales team focus on the right deals at the right time.
- Contact management has quickly evolved from Rolodex and Filofax systems into desktop-based contact management software and email clients with built-in contact management tools.
- Wrike is an enterprise-grade project management tool and collaborative to-do software. The platform can handle anything from simple lists to automated workflows and AI-backed data insights. Wrike is a solid choice for large organizations.
- Google Calendar is a free, full featured, and easy to use that lets you create both personal and shared calendars, which makes it ideal for tracking business group, family, and community schedules.

Keywords

Bla-Bla List: Bla-Bla List is another simple to-do list manager. It is a web based, of course, so you can access your lists from any location at any time. You can even publish your lists via RSS so that family and coworkers can get instant updates.

Cloud-based Contact Management and CRM Systems: A cloud-based system allows employees to update their system with the latest information about a contact wherever they are, from any web-enabled device. This is possible because all data is stored on a single database based in the cloud, so new and updated information is instantly available to all employees.

SalesRabbit-SalesRabbit's game-changing methodology is based on research they've done with sales pros and the top teams in the industry.

Pobuca Connect-Pobuca offers you a cloud contact management tool to share your contacts with your co-workers and partners.

Contact management: Contact management is the process of recording contacts' details and tracking their interactions with a business.

TracksLife: Trackslife is a database-oriented task manager. Each "track" is a separate database that combines columns of money, numbers, words, paragraphs, and yes/no responses.

Diarised-Diarised (www.diarised.com) is, like Presdo, a web-based meeting maker that users across different companies can use. It helps you pick the best time for a meeting by sending out emails to invitees, letting them choose the best times for them, and then sending you a summary of those best dates.

Self Assessment

1. Windows Live Calendar is _____ web-based calendar, actually part of the Windows Live Hotmail email service.
 - A. Microsoft's
 - B. Google's
 - C. Yahoo's
 - D. Apple's

2. Which of the following types of calendars you can create with Google Calendar?
 - A. Personal calendars
 - B. Public calendars
 - C. Holiday calendars
 - D. All of the above

Cloud Computing

3. Users can check their Google Calendars on which device?
 - A. Mobile Device
 - B. Laptop
 - C. Tablet
 - D. All of the above

4. A Google Calendar is used to:
 - A. Collect data from a form
 - B. Send out messages to others
 - C. Coordinate schedules with others
 - D. Collaborate on a shared document

5. A user can allow other people to make changes and manage sharing for Google Calendar.
 - A. True
 - B. False

6. What permissions are available for the guests added in the Google calendar?
 - A. Modify event
 - B. Add more guests
 - C. See guest list
 - D. All of the above

7. A user can invite other people to an event created by him/her on the Google Calendar?
 - A. True
 - B. False

8. Which of the following is/are example(s) of Web-based scheduling tools?
 - A. Sports Team Schedules
 - B. School Schedules
 - C. Community Group Schedules
 - D. All of the above

9. Web-based calendars that are ideal for keeping track of various homework assignments for a particular class or a schoolwide events schedule are
 - A. Homework calendars
 - B. School schedules
 - C. Homework schedules
 - D. None of the above

10. Web-based scheduling programs let you schedule both in-person meetings and _____ with attendees from multiple locations.
 - A. Text conferences
 - B. Teleheating
 - C. Teleconferences
 - D. Physical monitoring

Unit 07: Collaborating on Event Management

11. Which of the following is one of the offering from Schedulebook.com?
- Schedulebook Aviation
 - Schedulebook Transport
 - Schedulebook Cars
 - Schedulebook Calendar
12. Trackslife is a _____ task manager.
- health-oriented
 - priority-set
 - workflow-oriented
 - database-oriented
13. The full form of CRM is:
- Client Reservation Management
 - Client Remote Monitoring
 - Customer Resource Management
 - Customer Reservation Monitoring
14. Contact management evolved into customer relationship management as a result of which of the following developments?
- The shift in focus from contact details to relationships
 - The move from individual desktop databases to sharing information business-wide
 - The need for easy storage and retrieval of contact information
 - All of the above
15. Which of the following is/are benefits of using CRM?
- Access to data in real time from anywhere
 - Integration of social information
 - Smart collaboration with colleagues
 - All of the above

Answers for Self Assessment

- | | | | | |
|-------|-------|-------|-------|-------|
| 1. A | 2. D | 3. D | 4. C | 5. A |
| 6. D | 7. A | 8. D | 9. B | 10. C |
| 11. A | 12. D | 13. C | 14. D | 15. D |

Review Questions

- Discuss the concept of Cloud-based collaboration?
- Explain the Cloud-based Contact Management and CRM Systems?
- What are the different limitations of Contact-based Management Systems?
- Briefly explain the collaboration on Task Management?
- Describe the different tools for collaboration using To-do lists?

Cloud Computing

6. What are the major benefits of making use of cloud-based Contact Management and CRM systems?



Further Readings

- Introduction to Google Calendar for beginners, eLearning.
- Chomsky For Beginners, David Cogswell, 2011.
- Mastering Cloud Computing by Rajkumar Buyya, Morgan Kaufmann Publishers, 2013.
- Mastering Cloud Computing by "Rajkumar Buyya, Christian Vecchiola, S ThamaraiSelvi, Tata McGraw-Hill Education, 2021.



Web Links

1. Google Calendar - Wikipedia
2. Introduction To Google Calendar for Beginners | Free eBooks Download - EBOOKEE!
3. Lovely Professional University - Calendar - Week of November 28, 2021 (google.com)
4. Collaborative To-Do List Apps: The Best Way to Get Things Done - Taskade Blog
5. What is contact management? - Salesforce EMEA
6. Cloud-based communication, collaboration and contact center (structured.com)
7. Top 12 Best To Do List Apps to Simplify Your Task Management in 2021 (hubspot.com)

Unit 08: Collaborating on Project Management

CONTENTS

- Objectives
- Introduction
- 8.1 Types of Project Management Software Implementation
- 8.2 Cloud-Based Project Management
- 8.3 Cloud-Based Project Management Tools
- 8.4 Benefits of Cloud-based Project Management
- 8.5 Where Cloud PM Software Works
- 8.6 Steps for Finding Best Cloud-based Project Management Solution
- 8.7 Best Cloud-based Project Management Solutions
- 8.8 Cloud-based Project Management- Zoho Projects
- 8.9 What Does Zoho Projects Offer
- 8.10 What can be done with Zoho?
- 8.11 Management of Project using a Cloud-based Project Management Tool- Zoho Projects
- Summary
- Keywords
- Self Assessment
- Answers for Self Assessment
- Review Questions
- Further Readings

Objectives

After this lecture, you will be able to,

- Learn about project management.
- Explore different project management cloud tools and solutions.
- Practically explore the management of project using a cloud-based project management tool.
- Discuss collaborating on projects.

Introduction

Project management is the practice of initiating, planning, executing, controlling, and closing the work of a team to achieve specific goals and meet specific success criteria at the specified time. The primary challenge of project management is to achieve all of the project goals within the given constraints. This information is usually described in project documentation, created at the beginning of the development process. The project management caters to the process of leading the work of a team to achieve all project goals within the given constraints. The primary constraints in project management are scope, time, budget. The secondary challenge is to optimize the allocation of necessary inputs and apply them to meet pre-defined objectives. Most companies at one point or another have at least one big project going on – the type of project that involves multiple employees from multiple departments and perhaps multiple locations. The projects of this type have tons of individual pieces and parts, each of which dependent on the completion of a previous task. It involves keeping track of all the individual tasks – who's doing what and when – can take a gargantuan effort. However, managing a large project can be an exhaustive task. Even the smallest

Cloud Computing

project has numerous pieces and parts, all of which have to be completed in a precise order and on an exacting timetable for the project to come in on time and on budget. If just one-piece slips, the whole project goes out of whack.

Project management consists of using specific knowledge, skills, tools and techniques to deliver something of value to people. The development of software for an improved business process, the construction of a building, the relief effort after a natural disaster, the expansion of sales into a new geographic market – these are all examples of projects.

All projects are a temporary effort to create value through a unique product, service or result. All projects have a beginning and an end. They have a team, a budget, a schedule and a set of expectations the team needs to meet. Each project is unique and differs from routine operations – the ongoing activities of an organization – because projects reach a conclusion once the goal is achieved. The changing nature of work due to technological advances, globalization and other factors means that, increasingly, work is organized around projects with teams being brought together based on the skills needed for specific tasks.

Leading these projects are Project Professionals – people who either intentionally or by circumstance are asked to ensure that a project team meets its goals. Project professionals use many different tools, techniques and approaches to meet the needs of a project. Some projects are needed to quickly resolve problems, with an understanding that improvements will be made over a period of time. Other projects have a longer duration and/or produce a product or other outcome that will not need major improvements outside of projected maintenance, such as a highway. The project professionals use a variety of skills and knowledge to engage and motivate others to reach a project's goals. Project professionals are critical to the success of projects and are highly sought after to help organizations achieve their goals.

8.1 Types of Project Management Software Implementation

When software became available and affordable to business, companies replaced their pen-and-paper project tracking systems with project management software. But technology continues to evolve, so desktop applications improved in features, scale, and security that allowed multiple users access to the software over a company's own data network. Cloud computing has presented options and advantages, project management software vendors offered their subscription-based products over the internet to businesses of all sizes. It has opened the market to more providers offering a variety of systems.

Web-based Project Management Software (PMS)-A web-based PMS is accessible from any internet-connected web browser through a software-as-a-service (SaaS) subscription. These online platforms cater to businesses of all sizes across industries. The users in various locations can access the tool from different devices such as desktops, tablets, and smartphones to get up-to-date project status and information from a central database. Such PMS offer a rich set of features that can compete with enterprise systems. Many of these apps provide several methods to visualize project tasks and goal progress, collaborate with coworkers, automate workflows, share files, message team members, and integrate with email. By expanding their feature sets, many web-based project management tools also fall into the categories of collaboration software or business process management software.



Examples- Zoho Projects, Microsoft Project, and Basecamp.

On-premise Project Management Software (PMS)- An on-premise PMS resides on the company's own server. The server will be inside their data center and managed by its IT employees. Companies often purchase a single on-premise license and may have to purchase additional upgrades or updates. The companies prefer on-premise software for their concern on security as well as long-term costs. When the company runs the software on their private servers, they have full control over the data storage and security protocols to protect that data. The enterprises with 50, 100, or more users can save from monthly subscription costs if they purchase one-time license fees used in on-premise project management software. This enables control when to schedule software updates to make preparations and not experience any downtime due to an automatic product update inherent in SaaS systems.

8.2 Cloud-Based Project Management

Cloud computing is particularly useful to support a lean and digitize approach to project management. The project management can be made easier with the use of a web-based project management application. The project members can log in from any location to access the project's master file; they can add or delete tasks, mark tasks as complete, enter detailed billing information for individual tasks, and so forth. The cloud-based project management software coordinates the planning, collaborating, monitoring, and delivering of a project. It allows the project managers and teams to get work done using a network of tools available within the software, rather than using a plain old whiteboard and sticky notes. The use of project management software scales from business to business—and team to team. It is designed to make managing projects and deadlines easier and more effective.

When you employ a web-based project management application, you can more easily manage all the pieces and parts, no matter where the players are located. Your project is turned into a single database hosted in the cloud, accessible by all from any Internet-connected computer. As the project is hosted in the cloud, every team member sees the same Gantt or PERT chart and the same list of tasks, instantly updated when any other member makes an edit. Many project management applications include additional functions useful in the management of group projects. These features may include group to-do lists, web-based file sharing, message boards, time and cost tracking, and so on. However, the most robust of these apps lets you manage multiple projects simultaneously; users can schedule their time across multiple projects.

These are not simple applications; they can't be, given the enormity of many enterprise-level projects. Therefore, they're expensive to license and often difficult to learn how to use. The most popular of these apps include AceProject (www.aceproject.com), Basecamp (www.basecamphq.com), onProject (www.onproject.com), and Project Insight (www.projectinsight.com).

8.3 Cloud-Based Project Management Tools

For almost any company, whether large or small, project management is necessary for getting—and remaining—organized. There are multiple project management tools available to help your business manage resources, project and control costs, communicate effectively with multiple team members, and track progress on any given project. While some of these tools can be installed as software on individual network computers, a great many of them are web-based—in other words, they're available in the cloud. There are several offerings from project management that include:

Task Management: The task management software features let the project team members create, update, and manage the progress of tasks throughout a project's lifetime. This feature allows team members to:

- Set dependencies on a task
- Create and manage subtasks from larger tasks
- Set a task to repeat or recur at a specified time or date
- Assign tasks to one or more people
- Import a list of tasks from an external file

Task management is essential in project management software. The best project management tools offer visual task management in Kanban boards, Gantt charts, waterfall dependencies, and burndown charts. These tools give a visual representation of project progress that is often easier to understand than a set of lists or items in a spreadsheet.

Team Collaboration: Collaboration tools help solve complex projects. These features ensure that everyone who can see a task or portion of the project can comment, add assets to, and notate changes in the same interface where all of the project information is stored. Visual task representations, messaging, task descriptions, notation features, asset upload capabilities, tags, email integrations, calendar tools, and more are considered collaboration features.

- Collaboration tools enable project teams & stakeholders to:
- Exchange information
- Hold discussions
- Send comments or approvals

Cloud Computing

- Put context to project files
- Directly message one another

The creative teams may run several smaller projects concurrently, and can make use of Gantt charts and cross-team approvals to keep work flowing to reach deadlines. The large teams should use collaboration tools to keep the project up to date, alert team members to major and minor changes, and centralize assets.

Support Documentation: Software support documentation helps new and existing team members troubleshoot issues within the software and learn to use the tools quickly. Documentation and support tools that shorten the learning curve include:

- Help desk articles
- Support videos
- Periodic educational webinars
- In-software tooltips
- Chatbots
- Email or phone support

Some tools offer email and phone support only to paying subscribers, while others offer only online documentation or customer forums for any plan.

Email Integration: Email features within a project management tool will:

- Send notifications to individual team members regarding the status of their projects
- Send alerts regarding due dates, keywords, or status changes to individuals or the entire team
- Make new or update existing tasks via email
- Add comments, files, or custom labels to an existing task

This feature significantly increases productivity as the user can manage tasks and updates from their inbox without logging in to the tool. Individual members should also be able to set limits on the frequency and types of email notifications they receive.

Document Management: A good project management app provides a space for teams to store and manage files directly in the tool or by using an integrated third-party solution. The document management features that are indispensable in PM software are:

- A centralized document storage
- Ability to attach documents to tasks
- Ability to attach notes and comments to uploaded documents
- Multiple document upload using drag and drop functionality
- Document organization within folders, projects, or custom settings
- Version control
- Security features that can restrict document sharing to a small team, an entire project, or outside parties
- An online PM app also provides cloud storage space or integrates with a third-party service that provides it.

Project Portfolio Management: Large teams that manage many concurrent, complex projects may need to track each of those projects separately while also retaining the ability to report on and analyse their progression. Project portfolio management software gives companies an interface to run several projects within one or more portfolios. The project managers and stakeholders can then take a high-level view of project progress, quickly identify bottlenecks, and drill down into problem areas with each project. These tools are also instrumental in team resource allocation and expense tracking across a number of project timelines, as they can show where slack can be taken up by other individuals.

Mobile App: Today's work has moved from the office desk to the field. Mobile access via native project management apps helps all types of teams, and are especially useful for distributed or remote teams. It provides a familiar interface and functionality so users can catch up on work, track

Unit 08: Collaborating on Project Management

progress, and message their teammates easily while on the go. PM vendors offer mobile app solutions on various platforms and technology such as iOS, Android, and HTML5.

Third-party Integrations or API: The best project management tools have a solid core of PM features that allow users to manage an entire project effectively. It also has multiple integrations directly or via APIs for features and data sources outside of the project management tool. A flexible solution can readily integrate with many popular apps and services in use by companies today such as G Suite, Microsoft 365, Salesforce, or Dropbox. Zoho Projects, Wrike, Asana, and many other project management vendors offer a lot of integrations to connect to many other business software types.

Reporting: PMS has reporting functionality included—but to varying degrees. Consider the depth of your reporting needs and whether the software you're considering meets those needs. Another feature to consider is the ability to export project data to a third-party business intelligence tool if your company has one or is planning to have one. Helpful reporting features might track progress by due date or task type, update stakeholders on resource allocation, give a list of unfinished tasks, use tags to filter data, and estimate on-time delivery.

Scheduling: PMS with scheduling functionality promotes organization in large, multiple, and remote teams. It allows the user to set schedules on tasks, create a project timeline with milestones, and determine dependencies and resources. For some, this is an essential project management function. However, for organizations with simple projects, short-duration projects, recurring tasks, or small teams, this may not be as important.

Time Management: Time management features in a project management app let team members track how much time they spend on independent tasks, let project managers set ideal time goals for the project, and integrate with a timesheet or payroll app. The teams with long-term projects or that regularly employ contractors may find these tools especially useful. In software options built for large teams or industries with contractors, these features may also be called resource management and include expense tracking features.

Visualizations: Task and project visualizations help teams see the progress of multiple projects running at the same time. These tools can assist in managing the progress of the initiative, but they can also be useful for project planning by giving a visual mind map of the project timeline. Whether it's through a Kanban board, Gantt chart, burndown chart, or other custom visualization, these tools give an overview of the project's progress. The small businesses should identify what data or metric is important for them to see and check that their choice of PM solution can display them by default or will allow for customization. There are different visualization tools as depicted in the Figure 1.

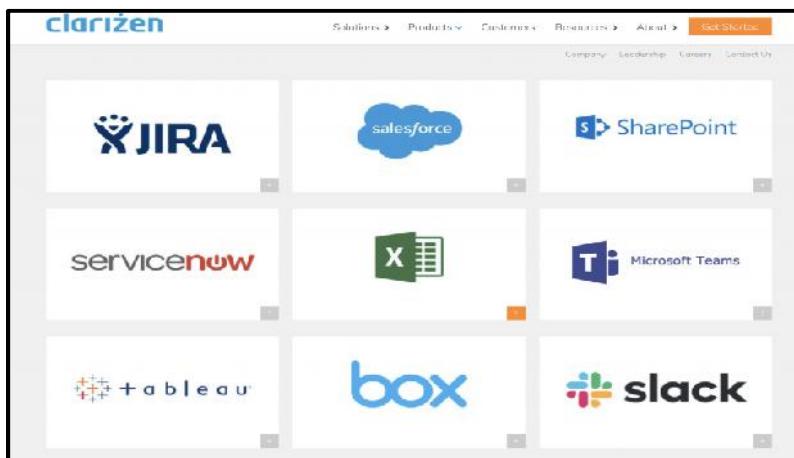


Figure 1: Data Visualization Tools

8.4 Benefits of Cloud-based Project Management

Lower upfront cost. When you use on-premises project management software, you have to pay for a number of items, including software license investments, server hardware, IT staff for support, lengthy deployments and more. Alternately, a cloud-based solution typically only requires a monthly or annual fee.

Cloud Computing

More Secure than Ever: Despite what many business executives have believed in the past, the cloud offers just as much security as on-premises solutions—the company simply does not control the physical servers. As a recent article in Forbes suggests, “cloud computing is no longer an oxymoron.”

Accessible from Anywhere: As long as you have an Internet connection, you can access your project management software in the cloud. Because access is achieved through a web browser, it’s also possible to allow team members to view, edit and manage documents via devices other than computers.

Connected Team: Even when workers are separated by miles, they can feel part of a cohesive unit with a cloud-based solution. Because everyone receives to-the-minute updates, no one relevant to the project is accidentally left out of the conversation. Furthermore, conversations can be conducted in the project space (instead of via email) and stored for future reference.

Increased Productivity: Thanks to a centralized location where all project summaries, tasks, documents, newsfeeds and calendars are kept, cloud-based project management solutions can save time. And when time is saved, productivity can increase. In addition, a centralized location allows team members to access project documents in real time wherever they are located.

Real-time Communication: One of the biggest benefits of using project management software in the cloud is the ability to give quick feedback on information as soon as it’s uploaded to the project workspace. This can also increase productivity because it lowers the instances of work that needs to be redone.

Up-to-date, Data-driven Decisions: As a business looking to become—and stay—profitable, it’s important to base decisions on more than just a whim. Thanks to the fact cloud-based software is instantaneously updated, you have the necessary information at your fingertips. In addition, most tools will have built-in reports to bring you the insights you need to make business-related decisions.

Simplified IT: When you use on-premises project management software, you are typically responsible for server maintenance, which requires the attention of internal IT personnel. Cloud-based software, on the other hand, is maintained for you by the service provider (or another resource)—saving you the time and hassle of dealing with any updates or bugs, and allowing you to focus on running your projects smoothly.

Instant Updates: Unlike on-premises software, which requires you to download updates whenever a new version is released, cloud-based tools are updated automatically. There’s no need to pay to upgrade your entire application suite or wait for your IT team to have the time available to implement the updates.

Low Maintenance Technology: Cloud-based PMSs being highly cost-effective are also low on maintenance. You don’t need an army of IT people running here and there solving tackling software related challenges. Moreover, the vendors take care of any issues or requirements such as version upgrades and anomalies in the online cloud-based project management software.

8.5 Where Cloud PM Software Works

While the large corporations tend to stick to the more traditional installed PM software, cloud versions are being embraced by companies that lack million-dollar IT budgets. The low initial costs, minimal or non-existent IT infrastructure investment, pay-as-you-go rates and anywhere access of cloud software is ideal for:

- Startups and entrepreneurs
- Small and mid-sized businesses
- Virtual companies with geographically diverse teams
- Freelance project managers

Of course, not all project management software is created equal. Companies will need to find software with features to meet their specific needs, and avoid extra bells and whistles that can drive up the subscription price.

8.6 Steps for Finding Best Cloud-based Project Management Solution

Certain steps are followed for finding the best cloud-based PMS as described below (Figure 2):

Step 1: Conduct a Needs Analysis

This won't come as a shock to you, but every company is different—with different sizes, different goals and different needs. That means it's up to your company to decide exactly which solution fits you best. Be sure those who actually will use the product have a voice in the discussion. Consider factors like whether your projects only involve internal employees or external clients as well, or what kinds of reports you will need to run. Once you know your needs, it's easier to evaluate service providers based on the features they offer.

Step 2: Research, Research, Research

When searching for a service provider, be on the lookout for those that offer the following features:

- Enterprise-grade reliability.
- Ability to add teams and projects quickly.
- Reporting features that give quick insights about your portfolio.
- A familiar environment or at least one that's easy to navigate and use.
- Ability to see and act on tasks in one location.
- Scheduling capabilities that allow you to plan and manage tasks effectively.
- Summary dashboards for data-driven decisions.
- 24/7 IT support
- Continuous data backup, disaster recovery and globally redundant data centers.
- Scalable solution that grows with your business.

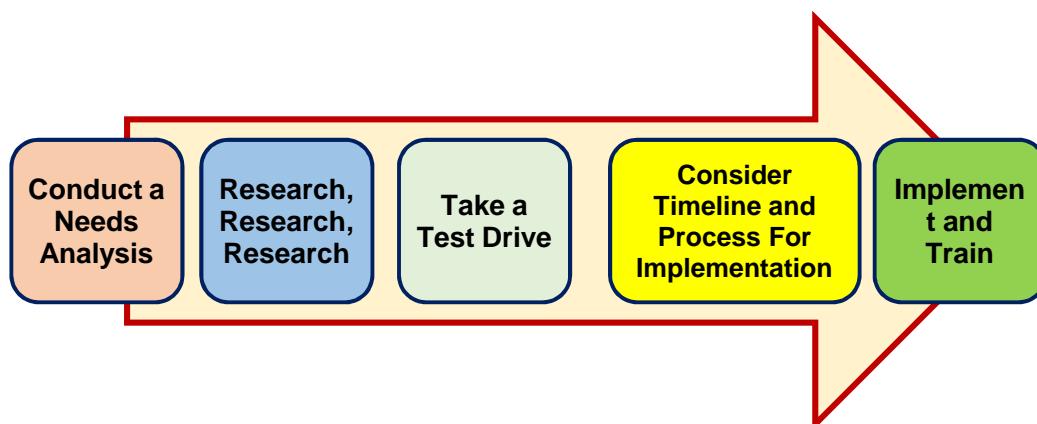


Figure 2: Cycle for Finding Best Cloud-based PMS

Step 3: Take a Test Drive

In the same way you'd test drive a vehicle before purchasing it, you should try out a project management solution before implementation. Test drives allow you to get a closer look at a tool's features and imagine how the tool would work in a situation similar to yours.

Step 4: Consider Timeline and Process for Implementation

If you've decided to go with an online-based tool for project management, it doesn't matter if it matches up well with your current applications—this only applies to on-premises solutions. But you will need to consider when you will have the manpower and budget to make the switch, and then coordinate with your chosen service provider to make it happen.

Step 5: Implement and Train

You've finally implemented a new project management software solution. Congratulations! But unfortunately, your work isn't complete. It's important to make sure your employees have the training they need to use the new tool effectively. Set up multiple sessions so everyone can work one into their schedule. In addition, create practice projects so everyone has a chance to explore and become familiar with the tool's features before they have to use it in the real world.

Cloud Computing

8.7 Best Cloud-based Project Management Solutions

ProofHub-ProofHub is a made-for-all solution. From freelancers to multinational enterprises from different business verticals – ProofHub is a project management and collaboration software that conveniently fits within any corporate setup. The software comes with integrated one-on-one/group chat, quick Discussions, Workflows, and Kanban boards, custom reports, custom roles, centralized files, notes, Gantt charts, proofing, timesheets, timers, and many more powerful features. The purpose of ProofHub is to enable teams to collaborate over projects in an easy, diverse manner and get them delivered as scheduled.

Wrike- Wrike is a project management tool with a load of useful features like task management, task prioritization, real-time newsfeed, interactive timeline (Gantt chart), and workload management. It works for both distributed and co-located teams. With Wrike, teams can easily schedule, discuss, and prioritize their tasks, and track projects as they progress. Wrike's Paid packages start from \$9.80/user/month (annual billing) for the Professional plan.

Clarizen- Clarizen is a project management software with a key focus on social engagement. It is designed to automatically prioritize projects, tasks, resources, budgets, and portfolios so that key resources and information are managed appropriately during the project management process. Other than that, templates, personal calendars, project portfolio management, workflows, dashboards, approval processes, and version control are some important features that make Clarizen a part of our top 5 project management tools. Clarizen offers two priced editions – Clarizen's Enterprise plan starts at \$45/user/month and Clarizen's Unlimited plan starts at \$60/user/month.

ProjectManager-ProjectManager is a tool designed to elevate efficiency in project planning, budgeting, scheduling, execution, and reporting. It is one of the most trusted project management systems in the market today. The platform essentially focuses on the three major parts of a project management process – planning, monitoring, and reporting. Additionally, it offers advanced add-on features like real-time dashboards, automated emails, and quick report generation. If this interests you, then you can easily sign up for ProjectManager's free trial.

Zoho Projects- Zoho Projects is a cloud-based project management platform designed to make your business projects more productive. The platform comes equipped with features that help to improve collaboration, facilitate monitoring, and enhance productivity and overall quality of the output. With Zoho Projects, teams can get work done and delivered on time. Zoho Projects pricing scheme is quite flexible. It comes with five plans, with a free plan for freelancers and small teams.

Mavenlink- Mavenlink is a modern software platform that connects people, projects, and profits. It seamlessly provides project and financial data, so that users always have the needed information and complete view of their operation. Real-time data shows project performance at any given minute, so stakeholders can always make decisions regarding time and budget. They can always check on costs of delivery, to make sure margins and ROI are on target. It provides overview and details whether it is a single project or a portfolio of projects. Resource management enables users to allocate resources where it is needed, with advanced features such as role-based estimating and business intelligence. Reports include up-to-date view of performance trends, problem areas, upcoming needs, and more.

Jira- Jira is an agile project management software used by development teams to plan, track, and release software. It is a popular tool designed specifically and used by agile teams. Aside from creating stories, planning sprints, tracking issues, and shipping up-to-date software, users also generate reports that help improve teams, and create their own workflows. As part of Atlassian, it integrates with many tools that enable teams to manage their projects and products from end to end. Jira Software is built for every member of your software team to plan, track, and release great software. Every team has a unique process for shipping software. Use an out-of-the-box workflow, or create one to match the way your team works.

WorkflowMax- Cloud-based workflow and job management software delivered as Software-as-a-Service. It is an end-to-end PM solution that has tools for leads, quotes, timesheets, job management, and invoicing.

Caspio (Custom Database Apps)- Cloud platform for creating business database applications, forms and reports fast and without coding. It includes everything you need to create and run online database applications on any website and for any number of users.

Cloud-based PM software are ground-breaking with the power to fast-track any organization onto the path of success. In today's fast-changing business environment, cloud-based software

Unit 08: Collaborating on Project Management

is revolutionizing the traditional workplace. Such software stands undeniably tall in fostering world-class collaboration and improved productivity at the workplace. These have become an indispensable prerequisite at workplace to achieve tremendous success.

8.8 Cloud-based Project Management- Zoho Projects

ZOHO offers online business, network and IT infrastructure management applications and software maintenance and support services. Zoho projects is a cloud-based project management solution designed for small and midsize companies. Zoho projects are online project management software that helps you run your business with ease. It stands as a comprehensive solution to your day-to-day problems in project management and adds value to your business by leaps and bounds. It offers project scheduling and budgeting, which allows project managers to define project tasks, assign them to their teams, estimate project costs and follow up on in-progress tasks.

8.9 What Does Zoho Projects Offer

- Helps you to plan, track, collaborate, and achieve your business targets with ease.
- Automates document processes such as revision tracking, access control and search and retrieval.
- Other product features include document management, which provides a portal for sharing documents within the team members.
- Offers issue management features, which help managers resolve errors in the project tasks.
- Dashboards and reporting module help users prepare business reports and management summaries.
- Users can use Gantt charts to get a visual presentation on the progress of tasks in comparison to what was planned.
- Resource allocation chart helps users know how many team members are free and how many are overloaded with tasks.
- Milestones and task lists help users organize complex projects by breaking them into subtasks.

Planning	Milestones, Tasklists, Tasks, and Subtasks help you plan work with ease.
Collaboration	Feeds, Activity stream, Wiki, Forum, and Chat help you to be on the same page with your peers.
Tracking	Reports and Charts help you track work within and across projects easily.
Productivity	Timesheets, Resource utilization chart, and Global Timers help you measure productivity effectively.
Knowledge Management	Documents and Forums help you manage files, discuss within and across teams in your portal with ease.
Effective Billing	Invoices and Expenses help you bill projects by hours and staff and manage project billing meticulously.
Issue management	The Bugs module helps you submit, track, and close bugs in your project.
Integration	Integrate Zoho Projects with Crashlytics, Dropbox, Slack and many other applications and access data with more fluidity.

Figure 3: Zoho Project Offerings

Cloud Computing**8.10 What can be done with Zoho?**

- Create a standalone task or group similar tasks into a task list.
- Raise bugs if a feature doesn't work as per the expected functionality.
- Define and automate any process workflow in your project using the blueprint.
- Track your project's budget and generate invoices with the extensible budget module.
- Track and manage the daily or weekly efforts of resources using a timesheet.
- Handle all your project related updates and the documents in one place.

8.11 Management of Project using a Cloud-based Project Management Tool- Zoho Projects

Zoho Projects, a cloud-based project management software, helps you plan your projects, track work efficiently, and collaborate with your team, wherever they are.

Sign up for Zoho Projects and create your own Portal: In order to create a new portal, follow the steps below (Figure 4):

- 1) Navigate to Zoho Projects Sign Up Page.
- 2) Enter your Company Name. This will also be your portal name. Your portal name is unique. If a portal with this name already exists, you will be prompted to enter another name. An email ID can be used to create only one portal. Once you create a portal with your email ID you cannot use the same ID to create a new Zoho Projects portal.
- 3) Enter your credentials, choose your region, and sign up to create your Zoho Projects portal.

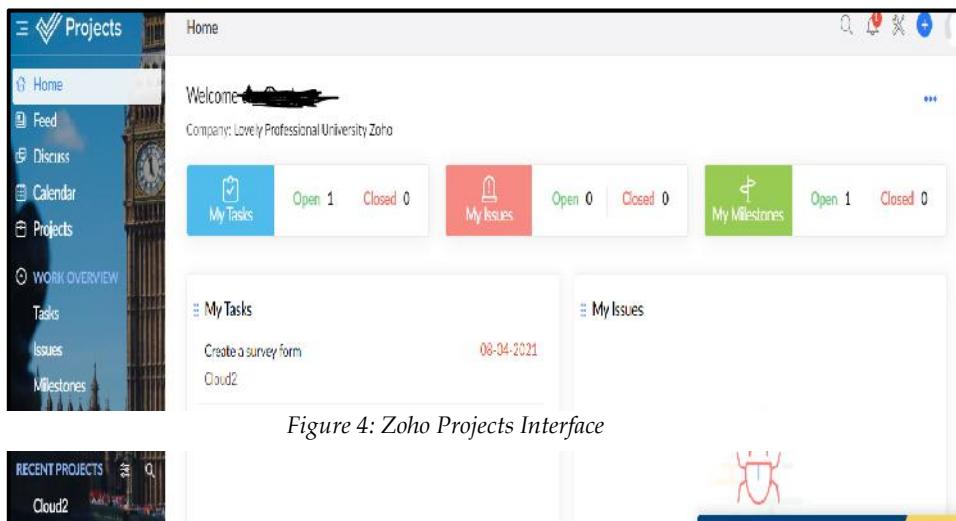


Figure 4: Zoho Projects Interface

Creating Projects: You can initially create a new project by clicking on the "Projects" tab on the dashboard and further clicking on the '+' symbol at the top right corner of the screen (Figure 5). You can also view the recent projects in the dashboard by choosing the "Projects" option.

The screenshot shows the 'Projects' page in Zoho Projects. At the top, it says 'All Projects (Active)'. A message indicates 'You have reached the maximum limit.' with a 'Upgrade Now' button. The main table lists two projects: 'L0-4 Cloud2' (Active, 0% complete, owner chauhan.taran, tasks 0/0%, milestones 0/0%, issues 2, start date 08-04-2021) and 'L0-3 Taran Cloud1' (Delayed, 0% complete, owner chauhan.taran, tasks 0/0%, milestones 0/0%, issues 3, start date 03-24-2021). The table has columns for PROJECT NAME, OWNER, STATUS, TASKS, MILESTONES, ISSUES, and START DATE.

Figure 5: Existing Zoho Projects

Unit 08: Collaborating on Project Management

Adding Portal Users: The following steps are required

The screenshot shows the 'Add User' interface in Zoho Projects. At the top, there's a note about separating multiple IDs by commas. Below that, there are dropdown menus for 'Role' (set to Manager) and 'Profile' (set to Manager). A 'Rate Per Hour' field shows USD 70. Under 'Select Projects', there's a search bar and a list of projects: Advanced Webinar, Avery Dennison Corporation, Brochure design, Budgetting, Business Breakfast meeting, Chennai Pet Expo, and Collaboration Hall Renovation. At the bottom, there are three buttons: 'Add' (highlighted in blue), 'Save & Add New', and 'Cancel'.

Figure 6: Adding Portal Users Interface

to create new portal users in Zoho projects.

- 1) Click in the top band.
- 2) Navigate to the Manage Users and then click Portal Users.
- 3) Click Add User in the upper right corner.
- 4) Enter a UserEmail address. You can add multiple email addresses separated by a comma.
- 5) Select a Role and Profile for the users. If your portal is integrated with Zoho Books or Zoho Invoice you can also add user rates.
- 6) Select the projects from the picklist to add users to them.
- 7) Click Add.

Edit a Portal User: You can edit a user Role, Profile, and User rates from Portal User section. These edits can also be made at an individual project level. You can also apply the user's portal defaults to any project to which they're assigned.

- 1) Click in the top band.
- 2) Navigate to the Manage Users section and click Portal Users.
- 3) Hover over a user to display available actions.
- 4) Click.
- 5) Update the Role, Profile, and Rate as needed.
- 6) Edit the Role and Rate Per Hour for individual projects in the list. (If the Rate Per Hour field for a project shows "NA", it means that project's billing is based on project hours, and you won't be able to set a user rate.)
- 7) To add the user to more projects, click Assign More Projects above the list.
- 8) To apply the user's default values to projects, use the checkboxes to select the projects you want and click Apply Portal Values.
- 9) Click Update.

Delete a Portal User: You can also delete a portal user if required by following the steps below-

- 1) Click in the top band.
- 2) Navigate to the Users section and click Portal Users.
- 3) Hover over a user to display available actions.

Cloud Computing

- 4) Click.
- 5) Check the box if you want to let the user know they've been removed.
- 6) Click Delete User.

Edit a Project User: When you edit a user within a project, the changes you make affect only that project.

- 1) Click Projects in the top band.
- 2) Select a project from the list.
- 3) Select Users from the left-hand panel.
- 4) Hover over a user to display available actions.
- 5) Click.
- 6) Customize user permissions using the toggle switch.
- 7) Click Update.

All Milestones					
MILESTONE	%	FLAG	OWNER	TAGS	
Taran Cloud1					
Implementation Environment	0%	Internal	tarandeep.kaur	Phase 5	
Feasibility Study Document	0%	Internal	chohan.mona	Mile 2	
Software Design	0%	Internal	chauhan.taran		
Cloud2					
Feasibility Document	0%	Internal	nimajit132	Mile 2	
Software Requirement Specification (SRS)	0%	Internal	chohan.mona	Mile 1	

Figure 7: Milestones in a Project

Milestones: Milestones are goal-based targets in a project. Milestones are listed by their status (Active, Completed, and Archived). To access the milestones module in Zoho Projects: Click Projects in the left navigation panel and select the project for which you would like to set milestones.

- 1) Click Projects and in the Recently Accessed Projects section, select the project for which you would like to set milestones.
- 2) Click Milestones in the left menu. You will see a list of milestones shown under following categories:

- o Active Milestones
- o Completed Milestones
- o Archived Milestones

New Milestone

Project*

Select Project

Milestone*

Start Date*

12 02 2021

End Date*

12 02 2021

Add Add More Cancel

Figure 8: Adding a Milestone

Adding a Milestone

- 1) Click New Milestone in the upper-right corner of the page (Figure 8).
- 2) Give a name for your Milestone and set its Start and End Dates.
- 3) Select a project user Who is responsible for the milestone.
- 4) Set the Milestone Flag. If the flag is Internal, only the portal users can access the milestone. If it is External, the client users can also access it.
- 5) Click Save.

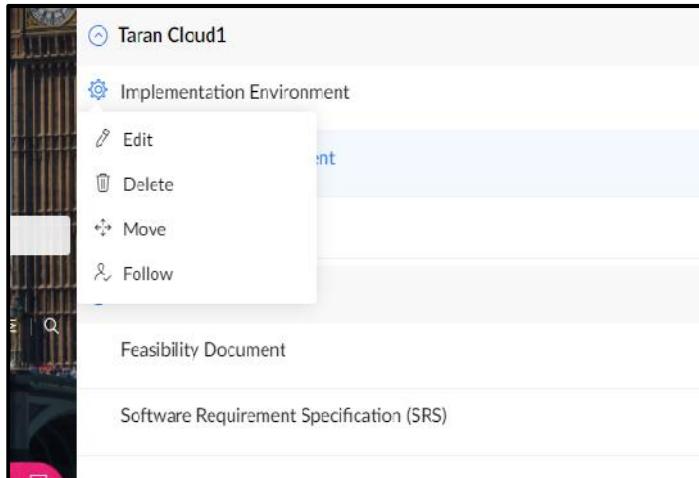


Figure 9: Editing or Deleting a Milestone

Edit or Delete a Milestone (Figure 9)

- 1) Hover over an active milestone.
- 2) Click to Edit and update its attributes; Complete to close it; Delete if you no longer need it; view the Release Notes to get the list of completed tasks and bugs in the milestone and apply blueprint to the milestone. Milestones having tasks with invoiced time logs cannot be deleted.

Adding Task List and Tasks to a Project: Right after creating a project, you break down your plan into workable entities called Tasks (Figure 10). These tasks are assigned to the users in your project and it is accomplished within a defined period of time. You can copy a task, set duration, priority, reminders, set recurring frequency etc. And, when you complete a task it's easy to mark it as complete and store it in the closed tasks. After completing a task, you may want to edit it, if so, you can do it by reopening the task again. You can add task list to a milestone and then tasks to it by clicking on the Tasks tab on a project dashboard. For creating a task, click on "Add Task" option (Figure 11 and Figure 12).

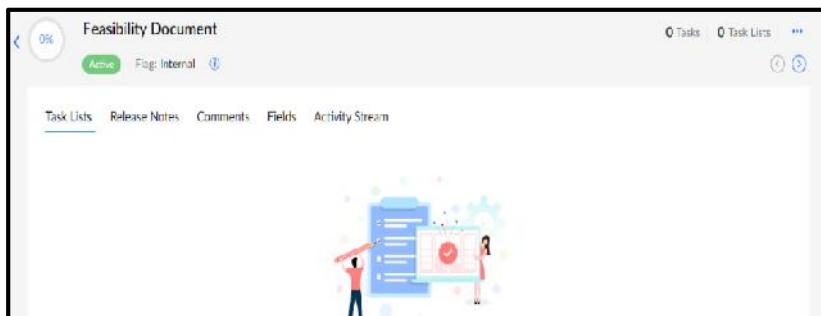


Figure 10: Creating a Task List in a Project

Cloud Computing

Task	Owner	Status	Tags	Start Date	Due Date
CL1-T2 Broadcast the survey form to all the users	chauhan.mona	On Hold			
CL1-T1 Create a survey form	chauhan.taran	Open		05-04-2021	05-04-2021
CL1-T3 Gather information through surveys of the form shared with	chauhan.mona	On Hold		05-06-2021	05-06-2021

Figure 11: List of Existing Tasks in a Project

New Task

Task Name*

Add Description

Task List

T1.4 (Software Requirement Specification (SRS))

Drop files or add attachments here... Maximum 10 files

Add Add More Cancel

Figure 12: Adding a New Task

Adding Client Users: A company or organization for whom the projects are done is called a client company. A user who can view the progress of the client company project, Example: paul@bowman.com is the client user of that project. You can add the client users to a project by clicking on the “Users” tab. Figure 13 shows the existing users in a project. To add a new client user, you can click on the “Add User” option as shown in (Figure 13). You can also view profile permissions for a user (whether as a manager, employee or Administrator) as well. The following steps illustrate of how to add a Client Company in a Portal (across projects).

chauhan.taran chauhan.taran@gmail.com	Manager • Admin	Last accessed on 12-02-2021 0...
chauhan.taran chauhan.taran@gmail.com	Administrator • Manager	Last accessed on -
paul.bowman paul.bowman@gmail.com	Manager • Manager	Last accessed on -

Figure 13: Client Users

Unit 08: Collaborating on Project Management

- 1) Click in the top navigation panel. Navigate to the Manage Users section and click Client Users.
- 2) Hover over the drop-down () on the Add Client User button. Click Add Client Company.
- 3) Enter a Client Company Name and select the projects you want to associate with this company.
- 4) Enter the Contact Email address, the Web Address, and the other postal address details respectively. Click Add.
- 5) Select Add New User to add users who are not part of the portal.
- 6) Enter the email addresses separated by commas.
 - (a) Give each user a Role and assign a Profile. You can add user rates if you have subscribed to Zoho Invoice or Zoho Books and completed the integration.
- 7) Click Add to List. You can now select the users from the pick list and add them to the project.
- 8) An email invite is sent to the added users along with the login credentials.

Documents Uploading to a Project: You can upload different types of documents to a project and share them to be accessible for other project users. Just choose the “Documents” tab on the dashboard and click on the “New” option (Figure 14). You get the further options for choosing the type of document you want to upload.

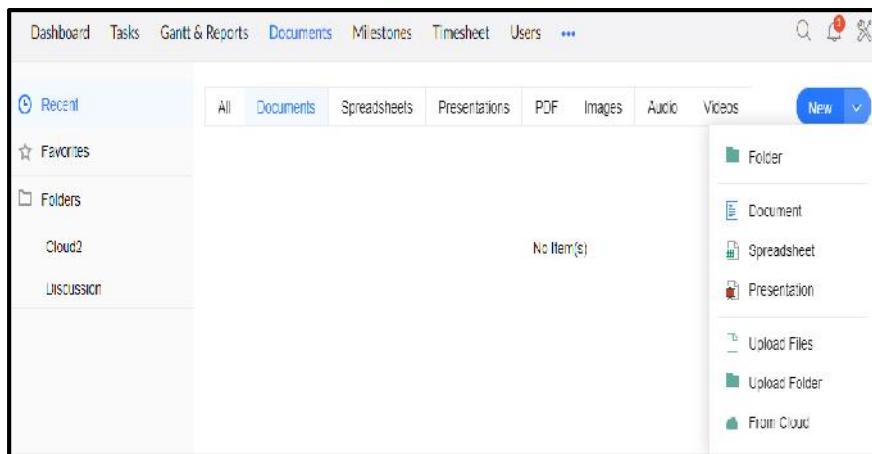


Figure 14: Uploading Documents

Sharing and Adding Collaborators: You can share the documents by choosing the document that you intend to share. Simply click on the option highlighted in the figure below (Figure 15). You need to enter the email ID of the person (s) with whom you want to share. You also get the options to share the documents with variable permissions such as, read-only; Read/ Write, Co-owner etc.



Figure 15: Sharing a Document

Settings Option: You can alter the settings for a project. Simply click on the settings option at the top right corner of the interface (Figure 16). You get multiple options such as Manage users, Data administration, import and export data etc.

Cloud Computing

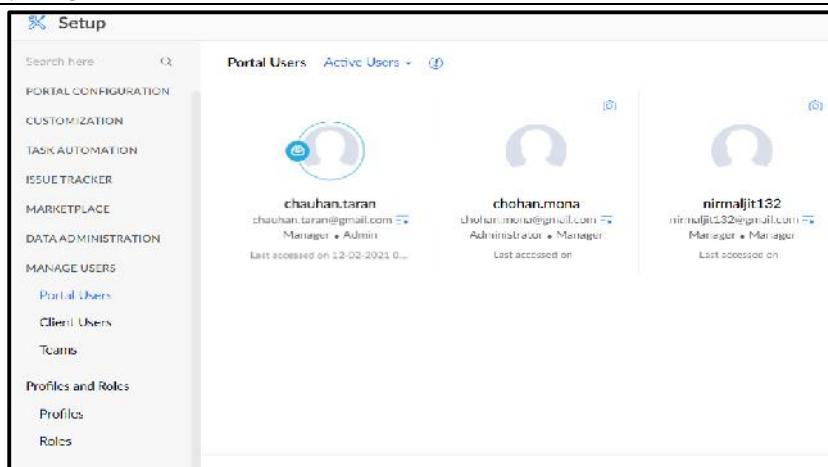


Figure 16: Settings in a Project

Summary

- Project management caters to the process of leading the work of a team to achieve all project goals within the given constraints. The primary constraints in project management are scope, time, budget.
- Project management consists of using specific knowledge, skills, tools and techniques to deliver something of value to people.
- Cloud computing is particularly useful to support a lean and digitize approach to project management. The project management can be made easier with the use of a web-based project management application.
- There are multiple project management tools available to help your business manage resources, project and control costs, communicate effectively with multiple team members, and track progress on any given project.
- Cloud computing is particularly useful to support a lean and digitize approach to project management. The project management can be made easier with the use of a web-based project management application.
- Time management features in a project management app let team members track how much time they spend on independent tasks, let project managers set ideal time goals for the project, and integrate with a timesheet or payroll app.

Keywords

Project Management: Project management is the practice of initiating, planning, executing, controlling, and closing the work of a team to achieve specific goals and meet specific success criteria at the specified time.

Web-based Project Management Software (PMS)- A web-based PMS is accessible from any internet-connected web browser through a software-as-a-service (SaaS) subscription.

On-premise Project Management Software (PMS)- An on-premise PMS resides on the company's own server. The server will be inside their data center and managed by its IT employees.

Task Management: The task management software features let the project team members create, update, and manage the progress of tasks throughout a project's lifetime.

ProofHub-ProofHub is a made-for-all solution. From freelancers to multinational enterprises from different business verticals - ProofHub is a project management and collaboration software that conveniently fits within any corporate setup.

Self Assessment

1. The primary constraints for project management include:

Unit 08: Collaborating on Project Management

- A. scope
B. time
C. budget
D. all of the above
2. _____ software coordinates the planning, collaborating, monitoring, and delivering of a project.
A. Cloud-based textual documents
B. Cloud-based project management
C. Project management
D. Cloud-based project monitor
3. Which of the following is an example of web-based project management software?
A. Basecamp
B. Zoho mail
C. Oracle Suite
D. Google Suite
4. In a project management software implementation, the _____ project management software resides on the company's own server.
A. off-premise
B. on-premise
C. neither
D. both
5. The multiple tools in project management software assist in
A. management of resources, project and cost controls.
B. communicating effectively with multiple team members
C. tracking progress on any given project
D. All of the above
6. Project _____ software gives companies an interface to run several projects within one or more portfolios.
A. portfolio monolithic
B. portfolio reservation
C. portfolio management
D. portfolio tracking
7. The reporting in project management software include(s) task (s) such as:
A. Tracking of progress by due date or task type
B. Updating stakeholders on resource allocation
C. Using tags to filter data, and estimate on-time delivery
D. All of the above
8. Which of the following is a Cloud-based project management solution?
A. Wrike
B. Jira-X

Cloud Computing

- C. Caspian
 - D. Workflow
9. Task management software involves different tools such as visual task management that give a visual representation of project progress that is often easier to understand than a set of lists or items in a spreadsheet. Which of the following is an example of visual task management?
- A. Kanban boards
 - B. Gantt charts
 - C. burndown charts
 - D. All of the above
10. Zoho Projects is a _____ project management solution designed for small and midsize companies and that helps you run your business with ease.
- A. offline
 - B. physical
 - C. cloud-based
 - D. infrastructural
11. _____ module in Zoho projects help users prepare business reports and management summaries.
- A. Dashboards and reporting
 - B. Gantt Charts
 - C. Milestones
 - D. Resource allocation
12. With the use of Zoho projects, which of the following task(s) can be carried out?
- A. Track your project's budget and generate invoices with the extensible budget module
 - B. Track and manage the daily or weekly efforts of resources using a timesheet
 - C. Handle all your project related updates and the documents in one place
 - D. All of the above
13. _____ charts in Zoho projects help users to know how many team members are free and how many are overloaded with tasks.
- A. Base
 - B. Jira
 - C. Jira-X
 - D. Resource allocation
14. A feature in Zoho projects that helps the managers to resolve errors in the project tasks is
- A. resource management
 - B. issue management
 - C. error management
 - D. task management
15. Zoho projects automates document processes such as revision tracking, access control and search and retrieval.
- A. True

B. False

Answers for SelfAssessment

- | | | | | |
|-------|-------|-------|-------|-------|
| 1. D | 2. B | 3. A | 4. B | 5. D |
| 6. C | 7. D | 8. A | 9. D | 10. C |
| 11. A | 12. D | 13. D | 14. B | 15. A |

Review Questions

1. Differentiate Web-based Project Management Software and On-premise project management software?
2. Explain the cloud-based project management and its benefits.
3. Discuss about the different Cloud-Based Project Management Tools?
4. Explain the following with respect to the Cloud-based project management:
 - (a) Document management
 - (b) Task management
 - (c) Team collaboration
5. What are the different steps for finding the best Cloud-based project management solution?
6. What are Zoho projects? How Zoho project help in collaborative work?



Further Readings

- Cloud-based collaboration and project management, Kalyan Vaidyanathan, Koshy Varghese, Ganesh DevkarbyRoutledge 2020.
- Mastering Cloud Computing by Rajkumar Buyya, Morgan Kaufmann Publishers, 2013.
- Mastering Cloud Computing by "Rajkumar Buyya, Christian Vecchiola, S Thamarai Selvi, Tata McGraw-Hill Education, 2021.
- Cloud Computing: Concepts, Technology and Architecture by Thomas Erl, Pearson Education.
- Cloud Computing Black Book by Kailash Jayaswal,Jagannath Kallakurchi, Donald J. Houde, Deven Shah, Kogent Learning Solutions, DreamTech Press.



Web Links

1. Zoho Projects | Online Project Management Software
2. Project management software - Wikipedia
3. Cloud-based Project Management | MS project Server Alternative (kendomanager.com)
4. Cloud-Based Project Management Software Benefits (proofhub.com)
5. 10 Best Web-Based Project Management Tools Of 2021 (thedigitalprojectmanager.com)
6. Zoho Accounts
7. Best Cloud Project Management Software 2021 | Datamation

Unit 09: Collaborating on Databases and Web-based Communication

CONTENTS

- Objectives
- Introduction
- 9.1 Database
- 9.2 Types of Databases
- 9.3 Purpose of Using a Database
- 9.4 Database System Applications
- 9.5 Working of Databases
- 9.6 Collaborating on Databases
- 9.7 Web-Based Communication
- 9.8 Collaborating via Social Networks and Groupware
- 9.9 Collaborating via Blogs and Wikis
- Summary
- Keywords
- Self Assessment
- Answers for Self Assessment
- Review Questions
- Further Readings

Objectives

After studying this unit, you will be able to:

- Learn about basics of Database and its types.
- Explore the online databases.
- Understand the collaboration using databases and cloud databases.
- Classify the traditional database from online database.
- Evaluate the online databases.
- Learn about collaboration via Web-communication tools.

Introduction

Data are units of information, often numeric, that are collected through observation. In a more technical sense, data are a set of values of qualitative or quantitative variables about one or more persons or objects, while a datum (singular of data) is a single value of a single variable. Data is the raw material from which useful information is derived. It is defined as raw facts or observations. Data is a collection of unorganized facts but can be made organized into useful information (**Error! Reference source not found.**). A set of values of qualitative or quantitative variables about one or more persons or objects. Data is employed in scientific research, businesses management (e.g., sales data, revenue, profits, stock price), finance, governance (e.g., crime rates, unemployment rates, literacy rates), human organizational activity (e.g., censuses of the number of homeless people by non-profit organizations).

Data is basically a collection of raw facts and figures we have a rough material that can be possessed by any computing machine that actually pertains to data. Example, we have few

Cloud Computing

numbers with us, but which number it is referring to, it is unknown. We have no idea if it is specifying somebody's age, or it is specifying some currency what it is specifying we don't know. It is just a number. Okay, so actually the collection of facts, which can be in different kind of conclusions can be drawn from it that is actually called as data.

Data is employed in scientific research, businesses management (e.g., sales data, revenue, profits, stock price), finance, governance (e.g., crime rates, unemployment rates, literacy rates), human organizational activity (e.g., censuses of the number of homeless people by non-profit organizations).

Information pertains to a systematic and meaningful form of data. It is the knowledge acquired through study or experience. The information helps human beings in their decision-making. Information can be said to be data that has been processed in such a way as to increase the knowledge of the person who uses the data.

Knowledge is the understanding based on extensive experience dealing with information on a subject. It is a familiarity, awareness, or understanding of someone or something, such as facts (descriptive knowledge), skills (procedural knowledge), or objects (acquaintance knowledge). By most accounts, knowledge can be acquired in many different ways and from many sources, including but not limited to perception, reason, memory, testimony, scientific inquiry, education, and practice. The philosophical study of knowledge is called epistemology.

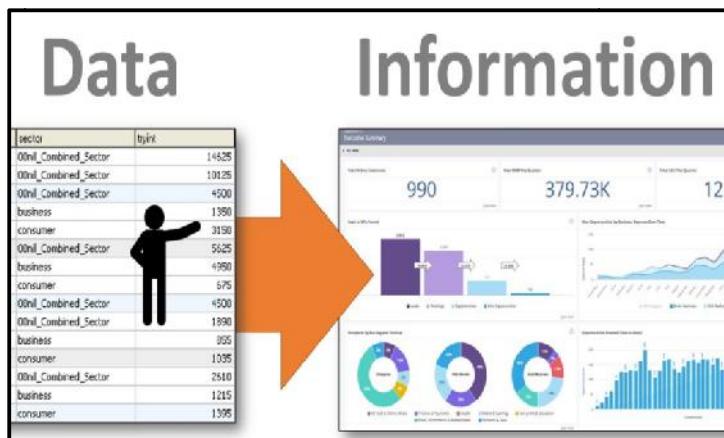


Figure 1: Conversion of Raw Data into Information

Let us consider an example of Mount Everest. Most of the books they site out what is the height of Mount Everest that is just a data. The height can be measured precisely with any altimeter and entered into a particular database. This data can be included in any particular book, along with other data on Mount Everest like how many trees are there or other things related to Mount Everest what is a sea level height. All this data can describe the mountain in a manner that would be useful for those who wish to actually make a decision about whether they should decide out the best manner how they can climb it. Thereby an understanding that is based on experience of climbing the mountains, that could actually be used by some other persons on a way that they can reach the Mount Everest. That is actually the knowledge. However, when we consider that the practical way of actually climbing of Mount Everest by a person climbs, based on his knowledge may be considered as his wisdom. In other words, we can say that wisdom refers to the practical application of a person's ability or a person's knowledge, depending upon the kind of circumstances, which are mostly considered as good.

Data Processing

The process of converting the facts into meaningful information. Data processing refers to the process of performing specific operations on a set of data or a database. As illustrated in Figure 2, Input one input two input three there are so many different kinds of data processing stages that are involved, like pre-processing is, the storage is there, then manipulations, analysis and we have so many different kinds of stages involves sorting is also involved, and then an output is generated so this output is the meaningful information or we can say organized data.

Unit 09: Collaborating on Databases and Web-based Communication

Example, every year about one lakh students take admission in National Open School (NOS). If you are asked to memorise records of date of birth, subjects offered and postal address of all these students, it will not be possible for you. In order to deal with such problems, we construct a database where in such situations, all information about students can be arranged in a tabular form. Several records can be maintained.

A database is an organized collection of facts and information, such as records on employees, inventory, customers, and potential customers.



Figure 2: Data Processing Scenario

Metadata: Metadata is the data that describes the properties or characteristics of other data, like we say we have a cat. We have some data about the cat, the name, we can specify the height, we can specify that as a metadata. Therefore, all such data describes the properties or characteristics of other data.

9.1 Database

A database is a system intended to organize, store, and retrieve large amounts of data easily. It consists of an organized collection of data for one or more uses, typically in digital form. One way of classifying databases involves the type of their contents, for example: bibliographic, document-text, statistical. Digital databases are managed using database management systems, which store database contents, allowing data creation and maintenance, and search and other access.

A database is a collection of information that is organized so that it can easily be accessed, managed, and updated. In one view, databases can be classified according to types of content: bibliographic, full-text, numeric, and images.

In computing, databases are sometimes classified according to their organizational approach. The most prevalent approach is the relational database, a tabular database in which data is defined so that it can be reorganized and accessed in a number of different ways. A distributed database is one that can be dispersed or replicated among different points in a network. An object-oriented programming database is one that is congruent with the data defined in object classes and subclasses.

Computer databases typically contain aggregations of data records or files, such as sales transactions, product catalogs and inventories, and customer profiles. Typically, a database manager provides users the capabilities of controlling read/write access, specifying report generation, and analyzing usage.

9.2 Types of Databases

There are different types of database as discussed below:

Analytical Database: Analysts may do their work directly against a data warehouse or create a separate analytic database for Online Analytical Processing (OLAP). For example, a company might extract sales records for analyzing the effectiveness of advertising and other sales promotions at an aggregate level.

Data Warehouse: Data warehouses archive modern data from operational databases and often from external sources such as market research firms. Often operational data undergoes

Cloud Computing

transformation on its way into the warehouse, getting summarized, anonymized, reclassified, etc. The warehouse becomes the central source of data for use by managers and other end-users who may not have access to the operational data (Figure 3).

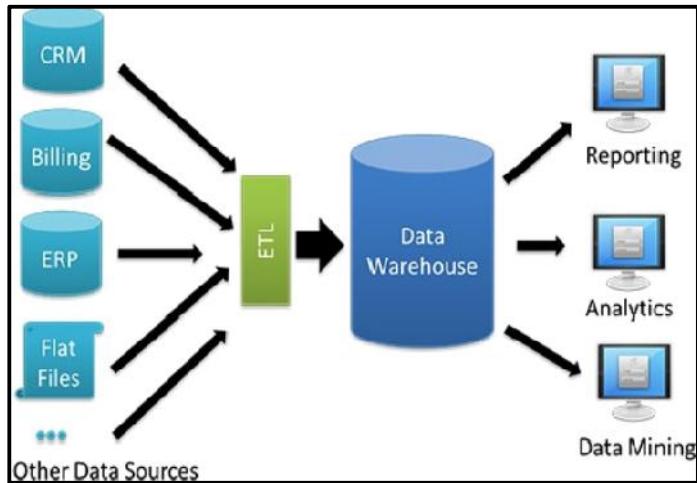


Figure 3: Data Warehouse Architecture

Distributed Database: These are databases of local work-groups and departments at regional offices, branch offices, manufacturing plants and other work sites. These databases can include segments of both common operational and common user databases, as well as data generated and used only at a user's own site.

End User Database: These databases consist of data developed by individual end-users. Examples of these are collections of documents in spread sheets, word processing and downloaded files, even managing their personal baseball card collection.

External Log Database: These databases contain data collected for use across multiple organizations, either freely or via subscription. The Internet movie database is one example.

Hypermedia Log Database: World Wide Web can be thought of as a database, spread across millions of independent computing systems. Web browsers "process" this data one page at a time, while Web crawlers and other software provide the equivalent of database indexes to support search.

Operational Log Database: These databases store detailed data about the operations of an organization. They are typically organized by subject matter, process relatively high volumes of updates using transactions. Essentially every major organization on earth uses such databases. Examples include customer databases that record contact, credit, and demographic information about a business' customers, personnel databases that hold information such as salary, benefits, and skills data about employees, Enterprise resource planning that record details about product components, parts inventory, and financial databases that keep track of the organization's money, accounting and financial dealings.

Customer Database: Such databases record contact, credit, and demographic information about a business' customers.

Personnel Database: It can hold information such as salary, benefits, skills data about employees.

Enterprise resource planning: Such database records details about product components, parts inventory.

Financial Database: Such database tracks an organization's money, accounting and financial dealings.

9.3 Purpose of Using a Database

A database is typically designed so that it is easy to store and access information. A good *database is crucial* to any company or organization. This is because the **database** stores all the important details

Unit 09: Collaborating on Databases and Web-based Communication

about the company such as employee records, transactional records, salary details. Database management systems are essential for businesses because they offer an efficient way of handling large amounts and multiple types of data. The ability to access data efficiently allows companies to make informed decisions quicker. There are certain roles associated with a database such as:

Database Administrator (DBA): A Database Administrator (DBA) in Database Management System (DBMS) is an IT professional who works on creating, maintaining, querying, and tuning the database of the organization. They are also responsible for maintaining data security and integrity.

Database Manager: Database manager provides users the capabilities of controlling read/write access, specifying report generation, and analyzing usage. They are prevalent in large mainframe systems, but are also present in smaller distributed workstation and mid-range systems such as the AS/400 and on personal computers. A database manager (DB manager) is a computer program, or a set of computer programs, that provide basic database management functionalities including creation and maintenance of databases. Database managers have several capabilities including the ability to back up and restore, attach and detach, create, clone, delete and rename the databases.

Database Three-Level Architecture

Database three-level architecture is used for describing the structure of specific database systems. In this architecture, the database schemas can be defined at three levels (Figure 4).

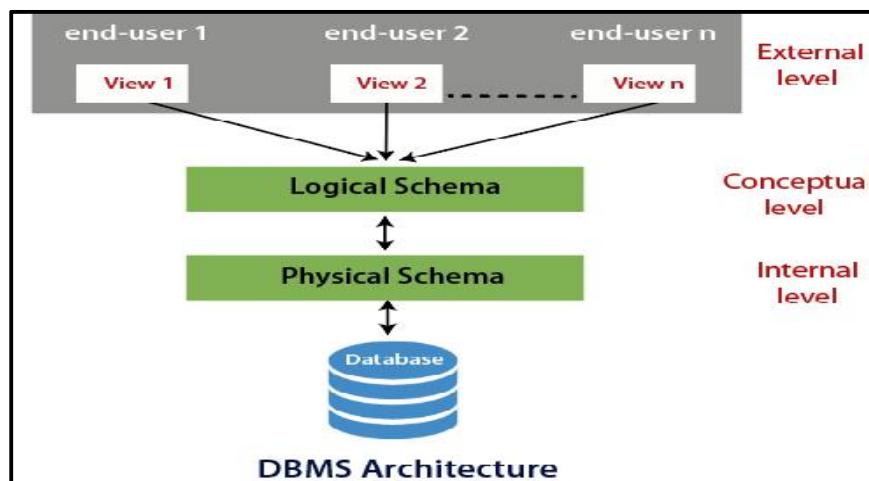


Figure 4: Database- Three Level Architecture

Internal/physical level: Shows how data are stored inside the system. It is the closest level to the physical storage. This level is also known as physical level. This level describes how the data is actually stored in the storage devices. This level is also responsible for allocating space to the data. This is the lowest level of the architecture. The physical level keeps the information about the actual representation of the entire database, that is, the actual storage of the data on the disk in the form of records or blocks.

Conceptual/logical level: It is also called **logical level**. The whole design of the database such as relationship among data, schema of data etc. are described in this level. Database constraints and security are also implemented in this level of architecture. This level is maintained by DBA (database administrator). This level lies between the user level and the physical storage view. Only one conceptual view for single database is available. It hides the details of physical storage structures and concentrates on describing entities, data types, relationships, user operations, and constraints.

External level: It is also called **view level**. The reason this level is called "view" is because several users can view their desired data from this level which is internally fetched from database with the help of conceptual and internal level mapping. This is the highest or top level of data abstraction (No knowledge of DBMS Software and Hardware or physical storage). This level is concerned with the user. Each external schema describes the part of the database that a particular user is interested in and hides the rest of the database from user. There can be n number of external views for

Cloud Computing

database where 'n' is the number of users. Example: A accounts department may only be interested in the student fee details. All the database users work on external level of DBMS.

9.4 Database System Applications

Banking: The banks can use database for storing the customer information, their account details loans that have been sanctioned out, and the different kinds of banking transactions can be also listed out. Thereby, database can be used for storing the customer information, accounts, and loans, and banking transactions.

Airlines: The aviation sector can use database as a basis for reservations and for scheduling different kinds of information.

Universities: A university can maintain the student information and also can maintain the employee database. It can consider out the grade achieved by the students; it can consider out the registrations that are done for a particular course over there.

Credit card Transactions: For purchases on credit cards and generation of monthly statements.

Telecommunication: For keeping records of calls made, generating monthly bills, maintaining balances on prepaid calling cards, and storing information about the communication networks.

Finance: For storing information about holdings, sales, and purchases of financial instruments such as stocks and bonds.

Sales: For customer, product, & purchase information.

Manufacturing: For management of supply chain & for tracking production of items in factories, inventories of items in warehouses/ stores, and orders for items.

Human Resources: For information about employees, salaries, payroll taxes and benefits, and for the generation of paychecks. Mahindra and Mahindra organization are collecting the data from DTO and feeding into database for the purpose of decision making.

9.5 Working of Databases

Databases have been a staple of business computing from the very beginning of the digital era. In fact, the relational database was born in 1970. Originally, databases were flat, means, that the information was stored in one long text file, called a tab delimited file. Each entry in the tab delimited file is separated by a special character, such as a vertical bar (|). Each entry contains multiple pieces of information (fields) about a particular object or person grouped together as a record. The text file makes it difficult to search for specific information or to create reports that include only certain fields from each record. One has to search sequentially through the entire file to gather related information, such as age or salary.

Relational databases have grown in popularity to become the standard. Unlike traditional databases, the relational database allows you to easily find specific information. It allows sorting based on any field and generate reports that contain only certain fields from each record. Relational databases use tables to store information. Relational database model takes advantage of this uniformity to build completely new tables out of required information from existing tables. Each table contains a column or columns that other tables can key on to gather information from that table. A typical large database, like the one a big Web site, such as Amazon would have, will contain hundreds or thousands of tables, all used together to quickly find the exact information needed at any given time.

The major uses and characteristics of relational database are:

- One can quickly compare salaries and ages because of the arrangement of data in columns.
- Uses the relationship of similar data to increase the speed and versatility of the database.
- The "relational" part of the name comes into play because of mathematical relations.
- A typical relational database has anywhere from 10 to more than 1,000 tables.
- Relational databases are created using a special computer language, structured query language (SQL), that is the standard for database interoperability.

Unit 09: Collaborating on Databases and Web-based Communication

-
- o SQL is the foundation for all of the popular database applications available today, from Access to Oracle.

Database Management Systems

DBMS is a collection of interrelated data and a set of programs to access those data. It can be said to be a software that interacts with end users, applications, and the database itself to capture and analyze the data. It encompasses the core facilities provided to administer the database. The sum total of the database, the DBMS and the associated applications can be referred to as a "database system". The term "database" is often used to loosely refer to any of the DBMS, the database system or an application associated with the database. The data in the databases is viewed as rows and columns in a series of tables, and the vast majority use Structured Query Language (SQL) for writing and querying data. The various reasons behind using DBMS are:

- o To avoid data redundancy
- o To avoid data inconsistency
- o Data independence
- o Efficient data access
- o Data integrity and security
- o Data administration
- o Concurrent access and crash recovery
- o Reduced application development time

9.6 Collaborating on Databases

A database does many of the same things that a spreadsheet does, but in a different and often more efficient manner. Many small businesses use spreadsheets for database-like functions.

Working of Online Databases

A local database is one in which all the data is stored on an individual computer. A networked database is one in which the data is stored on a computer or server connected to a network, and accessible by all computers connected to that network. An online or web-based database stores data on a cloud of servers somewhere on the Internet, which is accessible by any authorized user with an Internet connection. The cloud database is also a type of online database.

Advantages of Online Databases

Data can easily be shared with a large number of other users, no matter where they may be located. When your employee database is in the cloud, users sitting at different location can have easy access to the information stored in the database. As the data itself is stored in the cloud, when someone at one location updates a record, everyone accessing the database sees the new data. Synchronization is not an issue. Most online databases are oriented toward quick information sharing among members of workgroups who've assembled to attack a project for a month or two. When accessing data in this manner, ease of use is paramount, which most of these cloud applications address with simple and intuitive interfaces.

Cloud Database

The cloud database is a collection of informational content, either structured or unstructured that resides on a private, public or hybrid cloud computing infrastructure platform. From a structural and design perspective, a cloud database is no different than one that operates on a business's own on-premises servers. However, the critical difference lies in where database resides. "Cloud database" can be one of two distinct things: a traditional or NoSQL database installed and running on a cloud virtual machine, or a cloud provider's fully managed database-as-a-service (DBaaS) offering. It can be said to be running own self-managed database in a cloud environment where an on-premises database is connected to local users through a corporation's internal local

Cloud Computing

area network (LAN), a cloud database resides on servers and storage furnished by a cloud or database as a service (DBaaS) provider and it is accessed solely through the internet. The cloud DBaaS is natural database equivalent of software-as-a-service (SaaS). Cloud database access is based on the pay as you go model.

How Cloud Database Work?

There are two prominent cloud databases: Relational and Non-Relational databases. The relational database is written in Structured Query Language (SQL) and is a set of interrelated tables organized into rows & columns. The relationship between tables and columns (fields) is specified in a schema. Such databases are used in banking transactions or a telephone directory. The popular cloud platforms and cloud providers include MySQL, Oracle, IBM DB2 and Microsoft SQL Server. The non-relational database is sometimes also called as NoSQL as it does not employ a table model. It stores content, regardless of its structure, as a single document and is well-suited for unstructured data, such as social media content, photos and videos.

Benefits of Using a Cloud Database

- Ease of Access: Users can access cloud databases from virtually anywhere, using a vendor's API or web interface.
- Scalability: Cloud databases can expand storage capacities on run-time to accommodate changing needs. Organizations only pay for what they use.
- Disaster Recovery: In the event of a natural disaster, equipment failure or power outage, data is kept secure through backups on remote servers.
- Overall cost: The cost of adopting cloud databases can be considerably less than expanding your existing on-site server capability. There are reduced maintenance costs can also cut administrative costs substantially. Cloud's pay-as-you-go costs only increase if one expands or requires additional services.
- Flexible Solutions; Moving database solutions to the cloud can release a business from the demands and costs of managing own services. Cloud databases are massively efficient, as they have no inherent restrictions on their ability to expand.
- Mobile Access: Cloud databases add easiness in business expansion due to the ability of cloud platforms to be accessed and used from a range of remote devices. The applications can be built with geographically dispersed teams at no loss to efficiency or security.
- Disaster Recovery: The applications require reliable connections to the databases that power them. There is built-in redundancy and 24/7 uptime the norm. Cloud offers reliable platform for application development. Moreover, the robust cloud infrastructures are supported.
- Safe and Secure: Moving sensitive data to a cloud platform outside of a business's firewall could be risky. Cloud offers comprehensive security often more robust than that of on-site servers. The adoption of DBaaS infrastructures delivers world-class security.
- Scaling and Managing Database is Easier: Cloud service providers keep on evolving their services, any business can take instant advantage of these improvements, making both scaling and managing a database easier.



Example of Cloud Database Service- MongoDB Atlas Cloud Database

- MongoDB Atlas, is a part of MongoDB's broader Data-as-a-Service (DaaS) development platform.
- Powerful and compelling alternative to managing own NoSQL, or traditional database, or to using a cloud provider-specific managed offering.

Unit 09: Collaborating on Databases and Web-based Communication

- Offers fully managed database services through user's choice of cloud provider, including AWS, Azure, and GCP.
- Fully-managed database services handle the complexities of maintaining a consistently available, high-performance data cluster that developers can access as a single, globally available resource.

Exploring Web-Based Databases

Let us discuss the different Web-based Databases one by one in detail.

Blist: One of the newest entrants in the web-based database market is Blist (www.blist.com). Blist is a relatively easy-to-use database designed for nontechnical businesspeople; in fact, the company bills it as something of a cross between a spreadsheet and database program. Not surprisingly, the default Blist interface uses a spreadsheet metaphor, complete with rows and columns. That said, you can switch to a forms-based interface, which is perhaps better for entering raw data one record at a time. Blist provides some robust database management and reporting capabilities. You can make your data completely public or share it with designated users. Databases can be read-only, or users can have the option of adding new records or deleting old ones.

Cebase: Cebase (www.cebase.com) lets you create new database applications with a few clicks of your mouse; all you have to do is fill in a few forms and make a few choices from some pull-down lists. Data entry is via web forms, and then your data is displayed in a spreadsheet-like layout,

Dabble DB: Similar to Cebase is Dabble DB (www.dabbledb.com). Like Cebase, Dabble DB makes it easy to create new databases and add new records. Your data can be displayed in a number of different views, including the spreadsheet-like tableview. You can then sort, group, and filter your data; create various types of reports; and use your data to generate graphs, calendars, and maps. Dabble DB offers three ways to share your data. The Pages option enables you to collect data from other users without granting access to the underlying database. The Users option lets other users access the raw data in the database.

Lazybase: Lazybase (www.lazybase.com) is a simple online database, better suited for individuals than for large businesses. Creating a new database is as easy as filling in a few forms. Adding new records is just as easy. Your data is then presented on a clean, easy-to-grasp web page. Although there aren't a lot of fancy reports and such, Lazybase does offer simple data entry and display – easy enough for most home users to grasp.

MyWebDB: The application called myWebDB (hu.oneteamtech.com/mywebdb.html) is billed as a do-it-yourself Web 2.0 database application. Like the other web-based databases, myWebDB lets you create powerful applications by using its Builder Wizard to fill in a few forms, no coding necessary. myWebDB application then designs an application to match your data needs. The applications generated by myWebDB feature good-looking interfaces, complete with navigation menus, editable data grids, and intuitive data entry screens.

QuickBase: QuickBase (quickbase.intuit.com) is one of the oldest and most feature-rich online databases available today. Because it comes from Intuit, the company behind Quicken and Quickbooks, it's also one of the most stable and reliable web-based applications you can find. You can get started with QuickBase by using one of the application's ready-made applications, and then customize that application to suit your own specific needs. If there's no appropriate existing application, you can create a new one from scratch.

TeamDesk: TeamDesk (www.teamdesk.net) is, like QuickBase, a powerful web-based database management application that facilitates advanced application development. You can work from predefined applications for many business functions or create your own custom apps. The TeamDesk Application Library includes applications for project management, marketing, sales, customer support, human resources, billing, and other business functions.

Trackvia: Trackvia (www.trackvia.com) is similar to TeamDesk, in that it lets you create your databases from dozens of sample applications or completely from scratch. You can choose to view several predefined reports for each database application or generate a custom report.

Zoho Creator: Zoho offers two different database products: Zoho Creator and Zoho DB & Reports. The Zoho Creator is the easiest to use and best suited for casual users. Zoho Creator (creator.zoho.com) is a versatile data repository, complete with data-entry forms and spreadsheet-

Cloud Computing

like list views. While easy to use, Zoho Creator is robust enough to let you create your own simple database applications.

Zoho DB & Reports: If your database needs are more robust, turn to Zoho DB & Reports (db.zoho.com). Zoho DB offers more in-depth reporting than is available with Zoho Creator, including charts, pivot tables, and other report types. Zoho DB supports web-based APIs that enable it to serve as a back end for your own hosted applications; it also supports SQL for more robust data queries. Like Zoho Creator, Zoho DB can be accessed from any Internet-connected computer. You can also share your data and reports for collaborative development and analysis and embed your reports into your own website. In short, if your needs are simple, use Zoho Creator. If your needs are more complex, or if you're accustomed to working with Access, SQL Server, and similar relational database management programs, go with Zoho DB & Reports.

Evaluating Online Databases

When you're trying to decide which online database to use, it seems as if there's QuickBase and then there's everything else. That's because QuickBase has been around longer than most of the competition, is offered by an established software company (Intuit), and is targeted at large companies. Not all of QuickBase's competitors have the same lineage, level of technical support, and goals. I certainly recommend that you consider QuickBase, especially if you're working for a large company or organization and have correspondingly sophisticated database needs. QuickBase should also be on your short list if you want to develop specialized applications that use the data from a large database. If your needs are more modest, however, a number of the applications discussed above can be explored. I'm particularly partial to Zoho's two offerings, Creator (for smaller needs) and DB & Reports (for larger databases and more sophisticated applications). Bottom line, there are a lot of differences between the web-based database applications available today. The most basic applications are easy to use but limited in functionality; the more sophisticated applications have a steeper learning curve but offer more sophisticated reporting with more advanced automation. Choose your application wisely—you don't want to end up with an underpowered app or one that's too difficult for your organization to use.

9.7 Web-Based Communication

Web-based communication is also called as Internet Communications. If you use the Internet, then you probably use Internet-based communications to contact family, friends or co-workers. From sending an instant message to a friend, to e-mailing co-workers, to placing phone calls, to conducting video conferences, the Internet offers a number of ways to communicate. There are several advantages of Internet-based communications. Since you're already paying for an Internet account (or your employer is), you can save money on phone calls by sending someone an instant message or by using VoIP instead of standard local telephone services. Like all technologies (and especially technology tied to the Internet), the way we can communicate online is constantly evolving.

Collaborating via Web-Based Communication Tools

Groups located anywhere in the world can use web-based communication tools to communicate with other group members—and further their collaboration on group projects. There are three main categories for the web-based communication. These include:

- Web E-mail services
- Instant messaging services
- Web conferencing tools

a) Web E-mail Services

Web based email is a service that allows to access and use electronic mail via web browser. It contrasts with email service accessible through a specialized email client software. In this case email software is running on a web server allowing client to log-in, read and send mail from any computer with internet connection.

Initially, there used to be traditional email method. Contrarily, the traditional email is anything but cloud based. The type of email program you probably have installed on your PC uses a protocol

Unit 09: Collaborating on Databases and Web-based Communication

called the Post Office Protocol (POP). POP email requires the use of a dedicated email client program, such as Microsoft Outlook or Outlook Express, and at the ISP level email servers are used to send and receive messages. The problem with traditional POP email is that you're tied to the client program installed on your PC. The messages you receive are stored on that PC. You usually can't access them when you're traveling or away from that PC. There are none of the "anytime, anywhere" advantages you're used to with cloud-based services.

The web-based e-mail services is also known as web mail or HTTP email. First Web Mail implementation was developed at CERN in 1993 by Phillip Hallam-Baker as a test of the HTTP protocol stack, but was not developed further. In the next two years, however, several people produced working webmail applications. Web mail can be accessed from any PC using any web browser, and all your messages are stored on the web, not locally. Web mail is more versatile than traditional POP email, it's also easier to set up. All you need to know is your user ID and password, and then you access a page that lets you view the contents of your inbox, read and reply to messages, create new messages, and (in many cases) store messages in folders. You can even, on some services, use your web mail account to access your ISP's POP email. It offers web-based email accounts, free email accounts that are operated from a website. Three largest web mail services today are hosted by Google, Microsoft, and Yahoo! Examples: Hotmail, Gmail and Yahoo Mail. Webmail allows the users to access their emails as long as they have access to an Internet connection and a web browser. This also means that the user cannot read an old email or draft a new email offline.

In Europe, there were three implementations, Søren Vejrum's "WWW Mail", Luca Manunza's "WebMail", and Remy Wetzels' "WebMail". Søren Vejrum's "WWW Mail" was written when he was studying and working at the Copenhagen Business School in Denmark, and was released on February 28, 1995. Luca Manunza's "WebMail" was written while he was working at CRS4 in Sardinia, from an idea of Gianluigi Zanetti, with the first source release on March 30, 1995. Remy Wetzels' "WebMail" was written while he was studying at the Eindhoven University of Technology in the Netherlands for the DSE and was released early January 1995. In the United States, Matt Mankins wrote "Webex", and Bill Fitler, while at Lotus cc:Mail, began working on an implementation which he demonstrated publicly at Lotusphere on January 24, 1995. Matt Mankins, under the supervision of Dr. Burt Rosenberg at the University of Miami, released his "Webex" application source code on August 8, 1995, although it had been in use as the primary email application at the School of Architecture where Mankins worked for some months prior. Bill Fitler's webmail implementation was further developed as a commercial product, which Lotus announced and released in the fall of 1995 as cc:Mail for the WWW 1.0; thereby providing an alternative means of accessing a cc:Mail message store. Early commercialization of webmail was also achieved when "Webex" began to be sold by Mankins' company, DotShop, Inc., at the end of 1995. Within DotShop, "Webex" changed its name to "EMUmail"; which would be sold to companies like UPS and Rackspace until its sale to Accurev in 2001. EMUmail was one of the first applications to feature a free version that included embedded advertising, as well as licensed version. Hotmail and Four11's RocketMail both launched in 1996 as free services and immediately became very popular.

As the 1990s progressed, and into the 2000s, it became more common for the general public to have access to webmail because: many Internet service providers (such as EarthLink) and web hosting providers (such as Verio) began bundling webmail into their service offerings (often in parallel with POP/SMTP services); many other enterprises (such as universities and large corporations) also started offering webmail as a way for their user communities to access their email (either locally managed or outsourced); webmail service providers (such as Hotmail and RocketMail) emerged in 1996 as a free service to the general public, and rapidly gained in popularity. In some cases, webmail application software is developed in-house by the organizations running and managing the application, and in some cases, it is obtained from software companies that develop and sell such applications, usually as part of an integrated mail server package (an early example being Netscape Messaging Server). The market for webmail application software has continued into the 2010s.

Web-based E-mail Service Solutions

Gmail: Google's web mail service is called Gmail (mail.google.com), and at first blush it looks a lot like the other services we discuss in this chapter. Gmail is free, it lets you send and receive email from any web browser, and the interface even looks similar to its competitors. Gmail offers a few unique features that set it apart from the web-based email crowd. First, Gmail doesn't use folders. That's right, with Gmail you can't organize your mail into folders, as you can with the other services. Instead, Gmail pushes the search paradigm as the way to find the messages you want—not a surprise, given Google's search-centric business model. Gmail let you "tag" each message

Cloud Computing

with one or more labels. This has the effect of creating virtual folders, as you can search and sort your messages by any of their labels. In addition, Gmail groups together related email messages in what Google calls conversations. A conversation might be an initial message and all the replies (and replies to replies) to that message; a conversation might also be all the daily emails from a single source that have a common subject, such as messages from subscribed-to mailing lists. Like most of the other services, Gmail is a free service; all you have to do is sign up for an account. Of course, if you already have an account for any other Google service, that account can serve as your Gmail account. When you sign up for your Gmail account, you get assigned your email address (in the form of name@gmail.com) and you get access to the Gmail inbox page.

Yahoo! Mail: Yahoo! Mail (mail.yahoo.com) is another web mail service, provided by the popular Yahoo! search site. The basic Yahoo! Mail is free and can be accessed from any PC, using any web browser. Yahoo! also offers a paid service called Yahoo! Mail Plus that lets you send larger messages and offers offline access to your messages via POP email clients. Whether you use the free or the paid version, Yahoo! Mail gives you unlimited storage—which means you can effectively use Yahoo! Mail as an online backup or file-storage system. All you have to do is email yourself those files you want to store, and then place those messages (with attachments) in your designated storage folder. Yahoo! Mail interface is more functional than that offered by Gmail. It also offers traditional folder-based organization. You get a message pane and a reading pane, just as you do with Microsoft Outlook. Yahoo! also offers users the SpamGuard spam filter and Norton AntiVirus virus scanner.

Windows Live Hotmail: Hotmail was one of the first web-based email services, and it's still one of the largest. But it's not called "Hotmail" anymore; Microsoft has moved it into its Windows Live suite of online services and now calls it Windows Live Hotmail. Like most web mail services, Hotmail (we're going to call it by its old, shorter name) can be accessed from any web browser on any PC anywhere in the world, for free. Microsoft gives you 5GB of storage, not quite as much as you get with Gmail (6GB) or Yahoo! Mail (unlimited). You have your folder pane, message page, and reading pane. The new Hotmail also integrates with your Windows Live contacts and calendar, as well as other Windows Live services.

Apple MobileMe Mail: As part of its MobileMe suite of applications, Apple offers MobileMe Mail (www.me.com). What makes MobileMe Mail unique is that it's not limited to just computer users; you can also send and receive emails from your Apple iPhone or iPod touch, via Wi-Fi Internet or cellular network. MobileMe Mail is a web-based service that can also be accessed with your existing Mac or Windows-based email program, including Outlook, Outlook Express, and Windows Mail. It has its own native interface on the iPhone and iPod touch.

Other Web Mail Services: Gmail, Yahoo! Mail, and Windows Live Hotmail are the three largest web mail services (and MobileMe Mail promises to be a competitor), but there are literally hundreds more. Besides these big providers, there are dozens of independent web mail services, plus a plethora of topic-specific websites that offer (among other content and services) their own branded HTTP email. In addition, just about every cloud service provider, such as Zoho, offers web mail as part of its suite; web mail is also part of most web-based desktops. So, if you're looking for a web mail service and don't want to go with one of the big three, there are some other major providers. There are other web-based mail services:

- AOL Mail
- BigString
- Excite Mail
- FlashMail
- GMX Mail
- Inbox.com
- Lycos Mail
- Mail.com
- Zoho Mail

b) Instant Messaging Services

One of the fastest-growing forms of Internet communications is instant messaging, or IM. It is a text-based computer conference between two or more people. An IM communications service enables you to create a kind of private chat room with another individual in order to communicate

Unit 09: Collaborating on Databases and Web-based Communication

in real-time over the Internet. Typically, the IM system alerts you whenever somebody on your buddy or contact list is online. You can then initiate a chat session with that particular individual. One reason that IM has become so popular is its real-time nature. Unlike e-mail, where you will wait for the recipient to check his or her e-mail and send a reply, if a person you want to reach is online and available in your IM contact list, your message appears instantly in a window on their screen. While IM is used by millions of Internet users to contact family and friends, it's also growing in popularity in the business world. Employees of a company can have instant access to managers and co-workers in different offices and can eliminate the need to place phone calls when information is required immediately. Overall, IM can save time for employees and help decrease the amount of money a business spends on communications.

Instant messaging is a better way to talk. It is more immediate, because you can send text messages in real time to your friends and co-workers. No more waiting for people to respond to your emails—when both parties are online at the same time, it's just like having a one-on-one conversation. Several big players in the IM market today, including America Online (with both AOL Instant Messenger and ICQ), Google (Google Talk), Microsoft (Windows Live Messenger), and Yahoo! (Yahoo! Messenger).

IM, however, doesn't use servers at all. When you send an instant message to another user, that message goes directly to that user's PC; it's not filtered by or stored on any servers. The technical name for this type of connection is peer-to-peer (P2P), because the two computers involved are peers to each other. All IM needs to work is a piece of client software (one for each computer involved, of course) and the IP addresses of each computer. The messages go directly from one IP address to another, with no servers in the middle to slow things down.

Instant Messaging Tools

AOL Instant Messenger: Most-used instant messaging program is AOL Instant Messenger (www.aim.com), also known as AIM. AOL claims more than 60 million users, which makes it the number-two IM service today, second only to Yahoo! Messenger. For whatever reason, AIM is especially popular among the teen and preteen crowd, although people of all ages can and do use it. AIM supports all manner of special features in addition to basic text messaging. You get file sharing, RSS feeds, group chats, ability to text message to and from mobile phones, voice chat, video chat, and even a mobile client. You can also enhance the basic AIM experience with a variety of official and user-created plug-ins.

Google Talk: Google Talk is the name of both Google's instant messaging network and its IM client. You can download the Google Talk client and learn more about the Google Talk network at talk.google.com. You can access Google Talk from a web-based Google Talk gadget, a standalone Google Talk client program (similar to what's offered by both AIM and Yahoo! Messenger), or from your Gmail and iGoogle web pages. As with competing IM systems, Google Talk lets you send and receive both text-based instant messages and Voice over IP (VoIP) Internet phone calls. Most people will use Google Talk via the web-based Google Talk "gadget." You launch the Google Talk gadget by going to talk.google.com and clicking the Launch Google Talk button. With the gadget, there's no software to download. If you want increased functionality, such as file transfer, you can download the Google Talk client, which is a separate software program. The main Google Talk client window looks a lot like the Google Talk gadget window; however, if the person you're chatting with also has the Google Talk client installed, you can send files back and forth between yourselves, using the Send Files button.

ICQ: The granddaddy of all instant messaging programs is ICQ (www.icq.com). ICQ was birthed by a company named Mirabilis back in 1996, but was acquired by America Online in 1998. AOL maintains ICQ and AIM as separate programs—so separate that ICQ users can't talk to AIM users, or vice versa. Like most other IM programs, ICQ is totally free. You also get grouped conversations, voice messaging, photo viewing, and other state-of-the-art features.

Windows Live Messenger: Not surprisingly, Microsoft is a major participant in the instant messaging market. The program currently known as Windows Live Messenger does all the main things AIM and Yahoo! Messenger do, including voice chat and the ability to page a contact's mobile phone. With more than 27 million users, Windows Live Messenger is a solid middle-of-the-pack player.

Yahoo! Messenger: With more than 90 million users, the most popular instant messenger program today is Yahoo! Messenger. In addition to traditional text messaging, Yahoo! Messenger features voice and video messaging, PC-to-phone and PC-to-PC calling, voicemail, file sharing, and chat

Cloud Computing

rooms. It also lets you receive up-to-the minute stock prices, news headlines, sports scores, weather forecasts, and notification of any waiting Yahoo! Mail—all courtesy of the Yahoo! family of services.

c) Web Conferencing Services

When you need to include more than two people in your communications, or when you want to give a presentation to a group of people who aren't all in the same location, a different communications tool is needed. This new tool is called a web conference, and it's a way to conduct live meetings and presentations over the Internet. In a typical web conference, each participant sits at his own computer in his own location. Each participant's computer is connected to the conference via the Internet, and each participant sees the presentation on his or her screen, in real time. Web conference can be one way, as when the presenter delivers some sort of PowerPoint-like presentation, or two way, where each participant can join in and show the contents of their active applications or desktops. The communication between participants can be audio only (via streaming audio, VoIP, or traditional telephony) or include audio and video (typically using webcams). Most web conferencing services are hosted on the vendor's servers. You typically have to arrange a conference in advance, and the hosting service will help you set everything up. However, depending on the vendor, this can be a costly service, viable only for larger organizations. Make sure you check the price before you commit to using a particular service.

Features of Web Conferencing Services (Figure 5)

Application sharing, where the presenter and participants can all access and use the same application in real time. This is useful for smaller group meetings, when all participants are collaborating on a project.

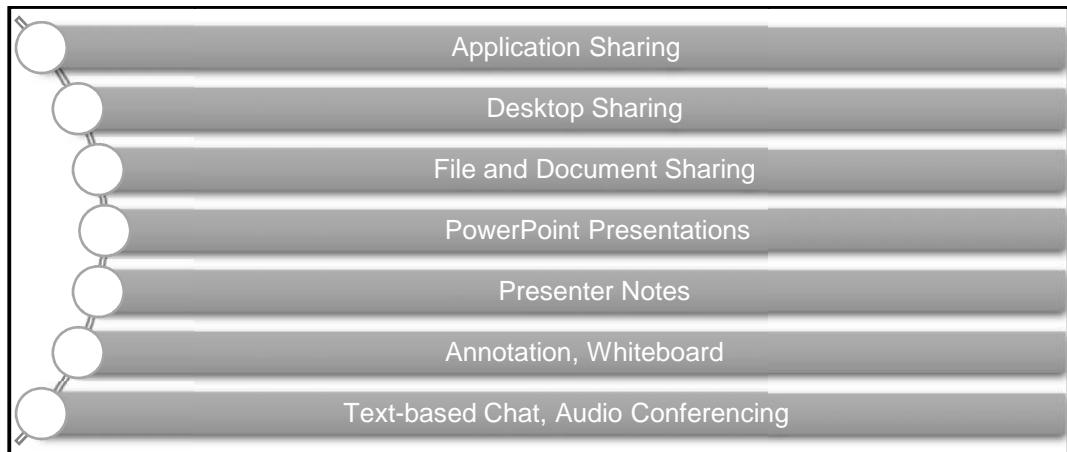


Figure 5: Features of Web Conferencing Services

Desktop sharing, similar to application sharing, but with the presenter's entire desktop visible and accessible to participants.

File and document sharing, with individual files and documents open for all to edit, also useful for group collaboration.

PowerPoint presentations, the core component of large presentations; the presenter gives a PowerPoint presentation in real time, complete with slide transitions and animations, using audio conferencing tools to narrate the presentation.

Presenter notes, which let the presenter take notes during the course of the conference for future action.

Annotation, which lets the presenter mark up the document or presentation being shared or given, typically by drawing or highlighting on the screen.

Whiteboard, which is a blank screen on which the presenter or participants can draw or highlight objects.

Text-based chat, which lets participants discuss the presentation with each other in real time.

Unit 09: Collaborating on Databases and Web-based Communication

Audio conferencing, which adds the spoken words of the presenter to a PowerPoint presentation. With two-way audio, all participants can speak—assuming that they all have microphones, of course.

Web Conferencing Services Solutions

Adobe Acrobat Connect: The Adobe Acrobat Connect software and service offers personal online “meeting rooms” for large organizations. For \$39/month (and up), you get audio/video conferencing, screen sharing, whiteboard, and chat functionality.

Convenos Meeting Center: A web-based conferencing service that starts at \$30/month. For that price, you get online presentations, file and document sharing, whiteboard, polling (the ability to ask questions of your audience), and integration with Skype for conference audio.

Genesys Meeting Center: Similar features to that of the Convenos service. Genesys gives you online PowerPoint presentations, file and document sharing, chat, desktop video, whiteboard, and polling and E-Quizzes. Pricing is by request only.

Glance: Glance is a web-based conferencing service priced from \$49.95/month. Its main focus is easy-to-use screen sharing, with no client software necessary to install.

IBM Lotus Sametime: IBM’s web conferencing service is dubbed Lotus Sametime, and it comes in several different versions: Entry, Standard, Advanced, and Unyte. The web conferencing service comes complete with enterprise instant messaging, multiway chat, VoIP and point-to-point video, and integration with most major desktop applications. Pricing varies by size of company.

Microsoft Office Live Meeting: Microsoft Office Live Meeting is a hosted service available in two versions (Standard and Professional). You get audio/video conferences, a PowerPoint viewer, integration with Microsoft Outlook, application and desktop sharing, and the like. Pricing is on a peruser basis, with volume licensing available.

WebEx: Cisco’s WebEx is perhaps the most-used web conferencing solution. Various solutions and pricing plans are available, for organizations large and small. Features include VoIP support, integrated audio and video, application sharing, on-the-fly annotation, meeting recording and playback, and so on.

Zoom Meeting: It is a simplified video conferencing and messaging across any device that is extensively used today. It includes the expected features, including application/desktop sharing, chat, and Skype integration, as well as remote PC control.

Google Meeting: Google Meet (formerly known as Hangouts Meet) is a video-communication service developed by Google. It is one of two apps that constitute the replacement for Google Hangouts, the other being Google Chat. The other features of the Google meet include:

- Two-way and multi-way audio and video calls with a resolution up to 720p.
- An accompanying chat, Call encryption between all users.
- Ability to join meetings through a web browser or through Android or iOS apps.
- Integration with Google Calendar and Google Contacts for one-click meeting calls.
- Screen-sharing to present documents, spreadsheets, presentations, or (if using a browser) other browser tabs.
- Ability to call into meetings using a dial-in number in the US.
- Hosts being able to deny entry and remove users during a call.
- Ability to raise and lower hand, Video filters, effects and augmented reality masks.

9.8 Collaborating via Social Networks and Groupware

When it comes to collaborating with a group of people who may or may not share the same physical location, one naturally turns to the web. When all team members have access to the Internet, why not use the Internet to connect the members of the group—to enable communication, file sharing, and the like?

*Cloud Computing***Creating Groups on Social Networks**

You're probably already familiar with social networks such as Facebook and MySpace. The typical social network is a hosted site that aims to create a community of users, each of whom posts his or her own personal profile on the site. Each user includes enough person information in her profile to enable other users with similar interests to connect as "friends"; one's collection of friends helps to build a succession of personal communities. Most profile pages include some form of blog, discussion forum, or chat space so that friends can communicate with the person profiled. In many instances, individual users also post a running list of their current activities so that their friends always know what they're up to.

**Notes**

Given that social networks are personal in nature, what value do they hold for businesses, community groups, and families? Lots, if you use them properly.

You see, most social networking sites let you create your own topic-specific groups. In this instance, a group is a collection of users who share the same interest; group members can communicate via discussion boards, share photos and videos, and even upload and download documents and other files. In other words, a social network group is like a virtual meeting or community room. Instead of posting notices on a physical bulletin board, you post notices on a virtual message board. Instead of exchanging brochures and papers by hand, you upload photos, documents, and other files for all to share. And, because most social networks are free for all to use (in exchange for the occasional on-page advertisement), it's a cheap way to keep the members of your group up-to-date and organized.

Social network groups are especially useful for community groups, far-flung friends, and families. You get just enough functionality to keep everyone in touch with each other, at no cost to anyone involved. No IT support is necessary, nor do you have to lease web hosting space; the social network site maintains all the servers and technology. And, of course, all of these sites are easy to join and easy to use, which is nice if your groups include non-tech-savvy members.

There are situations where these social network groups are less useful for larger businesses. In a nutshell, these groups lack the advanced collaboration features that help to keep group projects on track. In addition, the profusion of web page advertising is anathema to many businesspeople. Finally, many businesses aren't comfortable posting their business on a non secure third-party site (nor should they be), especially when more secure options are available. With all this in mind, let's take a quick look at the two most popular social networking sites and what they offer in terms of group collaboration features.

Facebook: Of all the social network sites, I recommend Facebook first and foremost for those serious about group collaboration. Compared to MySpace, Facebook is more of a site for grown-ups; MySpace is more suited for teenagers and preteens. When you create a group on Facebook, you end up with a group page. Facebook has various offerings:

- Recent news
- Discussion board
- Uploaded photos and videos
- Posted web pages
- The Wall – a kind of chat board
- Your group can be Open (public), Closed (description if public, but members have to be approved), or Secret (membership by invitation only).

MySpace: A group on MySpace is even more limited in functionality than what you can find on Facebook; this isn't surprising, given MySpace's typically younger audience. There's no file uploading, although members can upload group photos. There's a facility for posting group bulletins, and the obligatory discussion board, but that's it.



Task: Find the different uses of social networks and various solutions relate to it.

Other Web Groups

The groups on social networking sites aren't the only groups you can create on the web. In fact, they may be some of the less-functional groups out there; other sites do groups better.

Google Groups: When you create a Google Group, you get the obligatory message forum, but you also get to upload and share files, as well as create topic-specific pages within the group; group members can be notified of new posts via email. A Google Group can be Public (anyone can join, but only members can read messages), Announcement- Only (anyone can join, but only moderators can post messages), or Restricted (only the people you invite can join).

Yahoo! Groups: Here you also get a message forum (with email notification of new posts) and file uploading, and also a photo library, group calendar, and polls. You can select whether your group appears in the Yahoo! Groups directory, whether anyone can join or if you have to approve all members, and who can post messages to the group. Because of the file-uploading and -sharing options, either of these two groups might be more useful to you than a Facebook or MySpace group—even though they might not have the cachet of the social networking groups.

Evaluating Online Groupware

For larger businesses, a social network group probably won't suffice. What you need instead is a collection of web-based collaborative tools that help your team members not only communicate with each other but also manage their group projects. This type of solution is commonly known as groupware, and when it's based in the cloud it's called online groupware. In a nutshell, groupware is collaboration software for workgroups. Online groupware does away with the physical constraints of traditional groupware, letting members from throughout an organization, in any location, access group assets. The online groupware typically includes some or all of the following tools:

- File and document uploading and sharing
- Web calendar
- Task/project manager
- Message boards
- Text-based chat rooms / instant messaging
- Wiki-like collaborative pages
- Blogs

Why use Online Groupware?

- It puts all your group communications (and, in some cases, files) all in one place—and that one place is accessible to group members in any location, as long as they have an Internet connection.
- Makes it easier to communicate, thereby reducing the number of meetings & conferences calls, as well as your email traffic.
- Increases your group's collective and members' individual productivity.
- For example, suppose you're managing a community not-for-profit group.
- You can use online groupware to connect other managers and volunteers across the community. You can share plans, proposals, and other documents with all members, and use the groupware to solicit and receive proposals and invoices from suppliers. And, best of all, you can do this from your own computer, which means fewer phone calls, car trips, and unnecessary meetings—all of which translates into less time involved and fewer expenses, both of which are important for charities.

Cloud Computing

Online Groupware Solutions

AirSet: AirSet provides a cloud-based website for your group. Your AirSet site can include group announcements, a web calendar, contact list, task list, instant messaging, wiki for collaborative publishing, blog, file sharing and online storage, photo albums, and music playlists. And with all these tools, when one person in the group makes a change, everyone else sees the updated information.

ContactOffice: ContactOffice is a web-based data management system that lets you share emails, contacts, tasks, appointments, and documents with other group members. You can create internal or intercompany groups; the latter helps you communicate with customers, suppliers, and other people outside your immediate office. You also get a web-based calendar, address book, message forum, and real-time chat.

Google Sites: Google Sites, formerly known as Jotspot, lets you create a group web page (hosted by Google). This page is completely customizable with your choice of file uploads, group announcements, task/project management, mailing lists, group calendar, and the like. Google Sites also integrates with Google's other online apps, including Gmail, Google Calendar, Google Docs, and Google Talk. And, as with most things Google, it's completely free.

Huddle: Huddle is a hosted environment that combines online collaboration, project management, and document sharing, using social networking principles. You create a network of collaborative team workspaces, managed from a central dashboard. You can then take advantage of Huddle's online file storage, project calendar, RSS and email notifications, whiteboard, wiki, and other collaborative tools.

Nexo: Nexo lets you create a free personalized group website. The site can include photos, videos, forums, message boards, interactive calendars, polls, and to-do lists. Nexo targets its service to family, friend, and community groups, although it may also function for some less-demanding business groups.

OpenTeams: OpenTeams is better suited for larger businesses. It offers team folders, blogging, and wiki-like collaborative pages, all monitored via a customizable Navigator page. From here you can keep track of key team members, organize resources with tags, participate in threaded discussions, and monitor new content posted by team members. Pricing is on a per-user, per-use basis.

ProjectSpaces: ProjectSpaces provides an online workspace designed especially for group collaboration. You get an online document library, email discussion lists, task management, group announcements via email and RSS, a shared group calendar, and shared group documents.

TeamSpace: Our final online groupware application is called teamspace, with a lowercase 't'. This application offers task and project management, contact management, an online calendar, message forum, notice board, file sharing, text-based chat, and synchronization with Microsoft Outlook. Pricing is on a per-member basis, with additional fees for storage space used.

9.9 Collaborating via Blogs and Wikis

Group projects are all about collaboration and communication, so it pays to seek out every possible way to communicate with other group members. The blogs and wikis can both be housed in the cloud or on dedicated servers, depending on the service. But the point is that they're both web-based and they both facilitate group collaboration.

Evaluating Blogs for Collaboration

Blogs are also called as "web log", a kind of online journal that its author updates frequently with new musings and information. It is a collection of individual posts or messages. The posts are arranged in reverse chronological order, with the newest posts at the top – which makes it easy to keep track of the latest developments. The older posts are relegated to the blog archives, which are generally accessible via a link in the sidebar column. At the end of each post, you'll find a link to comments; this is where blog readers can register their own personal comments about any given post. A blog doesn't have to be the work of a single author; it can include posts from multiple contributors, as well as comments on each of those posts. This makes a blog ideal for keeping track of progress on a group project.

You create your blog, hosted on your company's servers or on a popular blogging tool such as Blogger or WordPad. You make it a private blog and assign authorship status to all the members of

Unit 09: Collaborating on Databases and Web-based Communication

your team. This means that everyone on your team can initiate new posts, as well as comment on the posts of others. When you have something important to say to the group, you make a blog post. Same with the other members; when they have updated info, they post it. Additionally, other members can comment on your posts—for example, you can create a post to schedule a meeting, and have the other members comment on your post with their replies. The members of your group can access the blog by navigating to its web page to see what's new, or subscribe to an RSS feed that will notify them whenever there's a new post to the blog, so they're never in the dark.

Blogger: Blogger is Google's blog-hosting community, and with more than 8 million individual blogs, the largest blog host on the Internet. All Blogger blogs are free, which contributes to their popularity.

TypePad: TypePad is quite similar to Blogger. You can customize your blog with a number of different designs and widgets, and you can select multiple co-authors for your blog.

WordPress: WordPress is another popular blog-hosting community. It's a lot like both Blogger and TypePad, but perhaps a bit more customizable. You get lots of themes to choose from, sidebar widgets, and a private members-only option. You also can create multiple blogs and assign multiple authors. And, like Blogger, a WordPress blog is completely free.

Wikipedia: Wikipedia is, in essence, a giant online encyclopedia—but with a twist. Wikipedia's content is created solely by the site's users, resulting in the world's largest online collaboration. The wikipedia articles are written, edited, and elaborated on by people of all types, from students, to subject-matter experts and professional researchers, to interested amateurs. It's a true group collaboration. A group wiki can be public (open to all users), as Wikipedia is, or private—which is ideal for project groups, businesses, and other organizations. A private wiki invites all group members to create new pages on the wiki site or to edit any existing page. All writing and editing is done within the web browser, no extra software or tools necessary. In most instances, there is no review of the articles or edits before they're accepted. The result is a collection of articles or documents, written collaboratively. The wiki software organizes the articles behind the scenes and manages the versioning for each article.

Pbwiki: Pbwiki offers various levels of wiki hosting. Small wikis (one to three users) are free; larger ones are priced as low as \$4 per user per month. Wiki creation is easy, thanks to a variety of pre-made templates. You also get online file storage to help you organize your other documents as part of your wiki.

Versionate: Versionate offers hosted wikis designed for group collaboration. A Versionate wiki is business friendly, thanks to SSL-level security and full control over editing privileges; you can also import Word, Excel, PowerPoint, and PDF documents into your wiki. The company offers several different plans.

wikihost.org: Wikihost.org site provides free wiki hosting. Wiki creation is via the GeboBebo engine, which offers a local database structure, user and rights management, RSS feeds and email notification for new and updated articles, and image and file uploading.



Lab Exercise: Create a blog where you will post about cloud computing and recent advancement in it.

Summary

- A database is a system intended to organize, store, and retrieve large amounts of data easily. It consists of an organized collection of data for one or more uses, typically in digital form. A database is a collection of information that is organized so that it can easily be accessed, managed, and updated.
- DBMS is a collection of interrelated data and a set of programs to access those data. It can be said to be a software that interacts with end users, applications, and the database itself to capture and analyze the data.
- The enterprise resource planning holds database records details about product components, parts inventory.
- The financial database tracks an organization's money, accounting and financial dealings.

Cloud Computing

- Unlike traditional databases, the relational database allows you to easily find specific information. It allows sorting based on any field and generate reports that contain only certain fields from each record. Relational databases use tables to store information.
- An Instant Messaging communications service enables you to create a kind of private chat room with another individual in order to communicate in real-time over the Internet.
- Web conference can be one way, as when the presenter delivers some sort of PowerPoint-like presentation, or two way, where each participant can join in and show the contents of their active applications or desktops.
- Social network groups are especially useful for community groups, far-flung friends, and families. You get just enough functionality to keep everyone in touch with each other, at no cost to anyone involved. No IT support is necessary, nor do you have to lease web hosting space; the social network site maintains all the servers and technology.
- Online groupware does away with the physical constraints of traditional groupware, letting members from throughout an organization, in any location, access group assets.
- Metadata is the data that describes the properties or characteristics of other data, like we say we have a cat. We have some data about the cat, the name, we can specify the height, we can specify that as a metadata. Therefore, all such data describes the properties or characteristics of other data.

Keywords

Groupware: The groupware is collaboration software for workgroups.

Local Database: A local database is one in which all the data is stored on an individual computer.

Networked Database: A networked database is one in which the data is stored on a computer or server connected to a network, and accessible by all computers connected to that network.

Distributed Database: Distributed databases are of local work-groups and departments at regional offices, branch offices, manufacturing plants and other work sites. These databases can include segments of both common operational and common user databases, as well as data generated and used only at a user's own site.

End User Database: The end-user databases consist of data developed by individual end-users. Examples of these are collections of documents in spread sheets, word processing and downloaded files, even managing their personal baseball card collection.

Three-level Database Architecture: Database architecture consists of three-level architecture. The internal/physical level shows how data are stored inside the system. The whole design of the database such as relationship among data, schema of data etc. are described in the logical level. An external level also called a view level exists and is called external because several users can view their desired data from this level which is internally fetched from database with the help of conceptual and internal level mapping.

Database Administrator (DBA): A Database Administrator (DBA) in Database Management System (DBMS) is an IT professional who works on creating, maintaining, querying, and tuning the database of the organization. They are also responsible for maintaining data security and integrity.

Database Manager: Database manager provides users the capabilities of controlling read/write access, specifying report generation, and analyzing usage.

Glance: Glance is a web-based conferencing service priced from \$49.95/month. Its main focus is easy-to-use screen sharing, with no client software necessary to install.

Microsoft Office Live Meeting: Microsoft Office Live Meeting is a hosted service available in two versions (Standard and Professional).

Lazybase: Lazybase (www.lazybase.com) is a simple online database, better suited for individuals than for large businesses. Creating a new database is as easy as filling in a few forms. Adding new records is just as easy.

Self Assessment

1. The internal view of the database offers the lowest level of _____.
 - A. Data manipulation
 - B. Data excess
 - C. Data abstraction
 - D. Data mode

2. The process of converting the facts into meaningful information is termed as
 - A. Data processing
 - B. Data converging
 - C. Data analyzing
 - D. Data deduplication

3. A(n) _____ database is one that can be dispersed or replicated among different points in a network.
 - A. Optical
 - B. Analytical
 - C. Distributed
 - D. Back-end

4. Which of the following is characteristic of distributed database?
 - A. It is useful for local work-groups and departments at regional offices, branch offices, manufacturing plants and other work sites
 - B. Can include segments of both common operational and common user databases, as well as data generated and used only at a user's own site
 - C. Can be replicated among different points in a network
 - D. All of the above

5. Which of the following is hypermedia log database?
 - A. World Wide Web
 - B. Web browsers
 - C. Web crawlers
 - D. All of the above

6. Which database level is closest to the users?
 - A. External
 - B. Internal
 - C. Physical
 - D. Conceptual

7. _____ is an archive of modern data from operational databases and often from external sources such as market research firms.
 - A. Mined data house
 - B. Data warehouse
 - C. Archival
 - D. External sources data

Cloud Computing

8. Web-based communication is also called as _____.
 - A. Common communication
 - B. Internet communication
 - C. Internal communication
 - D. Offline communication

9. Web-based communication include(s):
 - A. Web E-mail services
 - B. Instant messaging services
 - C. Web conferencing tools
 - D. All of the above

10. When you send an instant message to another user, that message goes directly to that user's PC; it's not filtered by or stored on any servers. The technical name for this type of connection is _____.
 - A. Peer-to-peer
 - B. Client-server
 - C. Non-filtered
 - D. Non-storage

11. Which of the following is not a feature of web-conferencing services?
 - A. Application sharing
 - B. Desktop sharing
 - C. Annotation
 - D. Physical document sharing

12. A _____ network group is like a virtual meeting or community room.
 - A. Social
 - B. Virtualized
 - C. Zoom
 - D. Zoho

13. The collection of web-based collaborative tools that help your team members not only communicate with each other but also manage their group projects is called as
 - A. Alterware
 - B. Groupware
 - C. Communication ware
 - D. Message ware

14. AirSet provides a cloud-based website for your group. The AirSet site can include
 - A. Group announcements
 - B. Web calendar
 - C. File sharing and online storage
 - D. All of the above

Unit 09: Collaborating on Databases and Web-based Communication

15. A _____ is a discussion or informational website published on the World Wide Web consisting of discrete, often informal diary-style text entries (posts).
- blog
 - hack info
 - wall
 - paragraph info

Answers for Self Assessment

- | | | | | |
|-------|-------|-------|-------|-------|
| 1. C | 2. A | 3. C | 4. D | 5. D |
| 6. A | 7. B | 8. B | 9. D | 10. A |
| 11. D | 12. A | 13. B | 14. D | 15. C |

Review Questions

- How does collaborating via Social Networks and Groupware takes place?
- Discuss the different social networks solutions.
- What are databases and discuss their applications?
- Explore the three-level architecture of database? Discuss each level.
- What are Online Groupware and their uses? Also, discuss different online groupware solutions.
- Write a short note on the following:
 - Web E-mail services
 - Instant messaging services
 - Web conferencing tools
 - Online Databases
- What is a Cloud database? Explain its benefits and working with an example.
- What is meant by Web-based communication? How the Web email services facilitate Web-based communication?
- What are the different features of Web-based Conferencing services?
- Explain any 5Web-based Databases in detail?

**Further Readings**

- Cloud-based collaboration and project management, Kalyan Vaidyanathan, Koshy Varghese, Ganesh Devkar by Routledge 2020.
- Mastering Cloud Computing by "Rajkumar Buyya, Christian Vecchiola, S Thamarai Selvi, Tata McGraw-Hill Education, 2021.
- Cloud Computing: Concepts, Technology and Architecture by Thomas Erl, Pearson Education.
- Cloud Computing Black Book by Kailash Jayaswal, Jagannath Kallakurchi, Donald J. Houde, Deven Shah, Kogent Learning Solutions, DreamTech Press.

**Web Links**

- What is a cloud database? | IBM
- Cloud Databases (tutorialspoint.com)

Cloud Computing

- 3.Types of databases (tutorialspoint.com)
- 4.What is web-based communication? | Study.com
- 5.Online database - Wikipedia
- 6.The 7 Best Online Databases as Simple as Spreadsheets (makeuseof.com)
- 7.10 Best Online Communication Tools For Business In 2021
(thedigitalprojectmanager.com)
- 8.13 Online Collaboration Tools for Productive Web Teams | Pantheon

Unit 10: Virtualization Concepts

CONTENTS

- Objectives
- Introduction
- 10.1 How Does Virtualization Work
- 10.2 Types of Virtualization
- 10.3 Pros of Virtualization
- 10.4 Cons of Virtualization
- Summary
- Keywords
- Self Assessment
- Answers for Self Assessment
- Review Questions
- Further Readings

Objectives

After this lecture, you will be able to,

- Learn about virtualization concepts and the need for virtualization.
- Explore the features of virtualization and the working of virtualization in cloud.
- Know about the different types of virtualization.
- Understand the virtualized environments.
- Analyze the pros and cons of virtualization.

Introduction

In computing, virtualization or virtualisation is the act of creating a virtual (rather than actual) version of something, including virtual computer hardware platforms, storage devices, and computer network resources. Virtualization began in the 1960s, as a method of logically dividing the system resources provided by mainframe computers between different applications. Since then, the meaning of the term has broadened. Virtualization technology has transformed hardware into software. It allows to run multiple Operating Systems (OSs) as virtual machines (Figure 1). Each copy of an operating system is installed in to a virtual machine.



Figure 1: Virtualization Scenario

Cloud Computing

You can see a scenario over here that we have a VMware hypervisor that is also called as a Virtual Machine Manager (VMM). On a physical device, a VMware layer is installed out and, on that layer, we have six OSs that are running multiple applications over there, these can be the same kind of OSs or these can be the different kinds of OSs in it.

Why Virtualize

1. Share same hardware among independent users- Degrees of Hardware parallelism increases.
2. Reduced Hardware footprint through consolidation- Eases management and energy usage.
3. Sandbox/migrate applications- Flexible allocation and utilization.
4. Decouple applications from underlying Hardware- Allows Hardware upgrades without impacting an OS image.

Virtualization enables sharing of resources much easily, it helps in increasing the degree of hardware level parallelism, basically, there is sharing of the same hardware unit among different kinds of independent units, if we say that we have the same physical hardware and on that physical hardware, we have multiple OSs. There can be different users running on different kind of OSs. Therefore, we have a much more processing capability with us. This also helps in increasing the degree of hardware parallelism as well as there is a reduced hardware footprint throughout the VM consolidation. The hardware footprint that is overall hardware consumption also reduces out the amount of hardware that is wasted out that can also be reduced out. This consequently helps in easing out the management process and also to reduce the amount of energy that would have been otherwise consumed out by a particular hardware if we would have invested in large number of hardware machines would have been used otherwise. Virtualization helps in sandboxing capabilities or migrating different kinds of applications that in turn enables flexible allocations and utilization of the resources. Additionally, the decoupling of the applications from the underlying hardware is much easier and further aids in allowing more and more hardware upgrades without actually impacting any particular OS image.

Virtualization raises abstraction. Abstraction pertains to hiding of the inner details from a particular user. Virtualization helps in enhancing or increasing the capability of abstraction. It is very similar to how the virtual memory operates. It helps to access the larger address spaces physical memory mapping is actually hidden by an OS with the help of paging. It can be similar to hardware emulators where codes are allowed on one architecture to run on a different physical device such as virtual devices central processing unit, memory or network interface cards etc. No botheration is actually required out regarding the hardware details of a particular machine. The confinement to the excess of hardware details helps in raising out the abstraction capability through virtualization.

Basically, we have certain requirements for virtualization, first is the efficiency property. Efficiency means that all innocuous instructions are executed by the hardware independently. Then, the resource control property means that it is impossible for the programs to directly affect any kind of system resources. Furthermore, there is an equivalence property that indicates that we have a program which has a virtual machine manager or hypervisor that performs in a particular manner, indistinguishable from another program that is running on it.

Before and After Virtualization

Before virtualization, the single physical infrastructure was used to run a single OS and its applications, which results in underutilization of resources (Figure 2). The nonshared nature of the hardware forces the organizations to buy a new hardware to meet their additional computing needs. For example, if any organization wants to experiment or simulate their new idea, they have to use separate dedicated systems for different experiments. So, to complete their research work successfully, they tend to buy a new hardware that will increase the CapEx and OpEx. Sometimes, if the organization does not have money to invest more on the additional resources, they may not be able to carry out some valuable experiments because of lack of resources. So, people started thinking about sharing a single infrastructure for multiple purposes in the form of virtualization.

Unit 10: Virtualization Concepts

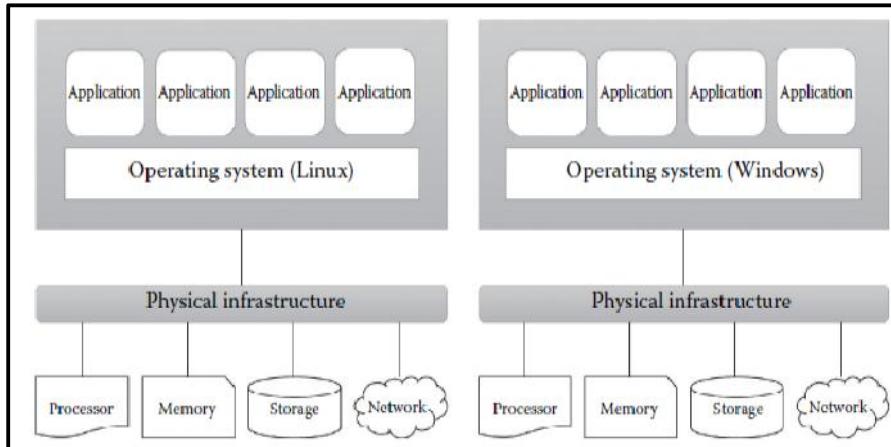


Figure 2: Before Virtualization

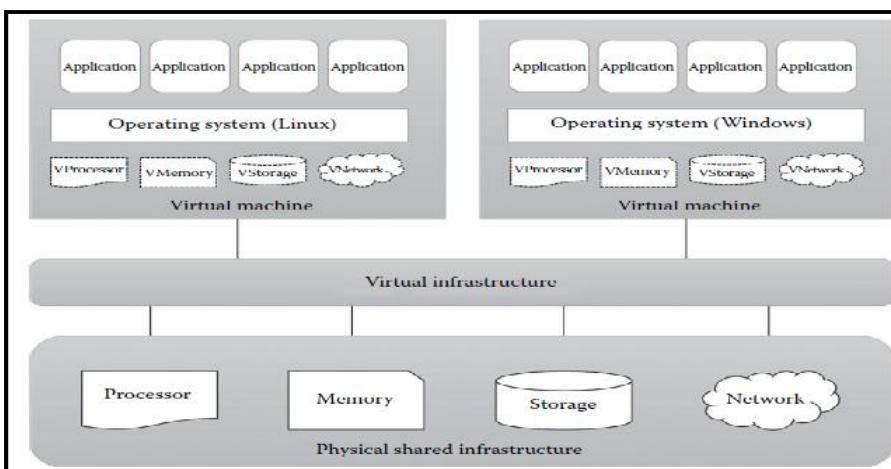


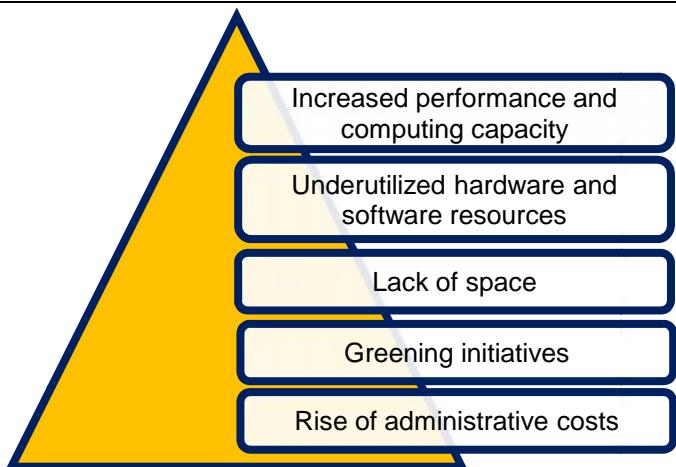
Figure 3: Post Virtualization Scenario

After virtualization was introduced, different OSs and applications were able to share a single physical infrastructure (Figure 3). The virtualization reduces the huge amount invested in buying additional resources. The virtualization becomes a key driver in the IT industry, especially in cloud computing. Generally, the terms cloud computing and virtualization are not same. There are significant differences between these two technologies.

Virtual Machine (VM): A VM involves an isolated guest OS installation within a normal host OS. From the user perspective, VM is software platform like physical computer that runs OSs and apps. VMs possess hardware virtually.

Factors Driving the Need of Virtualization

Increased Performance and Computing Capacity: PCs are having immense computing power. Nowadays, the average end-user desktop PC is powerful enough to meet almost all the needs of everyday computing, with extra capacity that is rarely used. Almost all these PCs share resources enough to host a VMM and execute a VM with by far acceptable performance. The same consideration applies to the high-end side of the PC market, where supercomputers can provide immense compute power that can accommodate the execution of hundreds or thousands of VMs.



Underutilized Hardware and Software Resources- Hardware and software underutilization is occurring due to: increased performance and computing capacity, and the effect of limited or sporadic use of resources. The computers today are so powerful that in most cases only a fraction of their capacity is used by an application or the system. Moreover, if we consider the IT infrastructure of an enterprise, many computers are only partially utilized whereas they could be used without interruption on a 24/7/365 basis. For example, desktop PCs mostly devoted to office automation tasks and used by administrative staff are only used during work hours, remaining completely unused overnight. Using these resources for other purposes after hours could improve the efficiency of the IT infrastructure. To transparently provide such a service, it would be necessary to deploy a completely separate environment, which can be achieved through virtualization.

Lack of Space: The continuous need for additional capacity, whether storage or compute power, makes data centers grow quickly. Companies such as Google and Microsoft expand their infrastructures by building data centers as large as football fields that are able to host thousands of nodes. Although this is viable for IT giants, in most cases enterprises cannot afford to build another data center to accommodate additional resource capacity. This condition, along with hardware under-utilization, has led to the diffusion of a technique called server consolidation, for which virtualization technologies are fundamental.

Greening Initiatives: Recently, companies are increasingly looking for ways to reduce the amount of energy they consume and to reduce their carbon footprint. Data centers are one of the major power consumers; they contribute consistently to the impact that a company has on the environment. Maintaining a data center operation not only involves keeping servers on, but a great deal of energy is also consumed in keeping them cool. Infrastructures for cooling have a significant impact on the carbon footprint of a data center. Hence, reducing the number of servers through server consolidation will definitely reduce the impact of cooling and power consumption of a data center. Virtualization technologies can provide an efficient way of consolidating servers.

Rise of Administrative Costs: The power consumption and cooling costs have now become higher than the cost of IT equipment. Moreover, the increased demand for additional capacity, which translates into more servers in a data center, is also responsible for a significant increment in administrative costs. Computers – in particular, servers – do not operate all on their own, but they require care and feeding from system administrators. Common system administration tasks include hardware monitoring, defective hardware replacement, server setup and updates, server resources monitoring, and backups. These are labor-intensive operations, and the higher the number of servers that have to be managed, the higher the administrative costs. Virtualization can help reduce the number of required servers for a given workload, thus reducing the cost of the administrative personnel.

- Share same hardware among independent users.

Degrees of hardware parallelism increases.

- Reduced hardware footprint through consolidation

Eases management, energy usage.

- Sandbox/migrate applications

Unit 10: Virtualization Concepts

Flexible allocation & utilization.

- Decouple applications from underlying Hardware
 - Allows hardware upgrades without impacting an OS image.

Features of Virtualization

- Virtualization Raises Abstraction
 - Similar to Virtual Memory: To access larger address space, physical memory mapping is hidden by OS using paging.
 - Similar to Hardware Emulators: Allows code on one architecture to run on a different physical device, such as, virtual devices, CPU, memory, NIC etc.
 - No botheration about the physical hardware details.
- Virtualization Requirements
 - Efficiency Property: All innocuous instructions are executed by the hardware.
 - Resource Control Property: It must be impossible for programs to directly affect system resources.
 - Equivalence Property: A program with a VMM performs in a manner indistinguishable from another.Except: Timing & resource availability.

Virtualized Environments

Virtualization is a broad concept that refers to the creation of a virtual version of something, whether hardware, a software environment, storage, or a network. In a virtualized environment, there are three major components (Figure 4):

- Guest: Represents the system component that interacts with the virtualization layer rather than with the host, as would normally happen.
- Host: Represents the original environment where the guest is supposed to be managed.
- Virtualization Layer: Responsible for recreating the same or a different environment where the guest will operate.

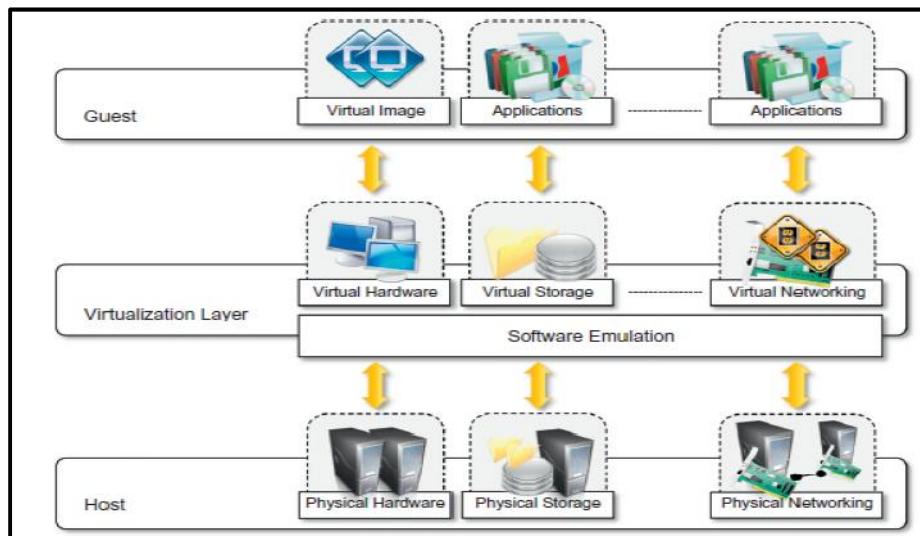


Figure 4:Virtualized Environment

The components of virtualized environments include: In the case of hardware virtualization, the guest is represented by a system image comprising an OS and installed applications. These are installed on top of virtual hardware that is controlled and managed by the virtualization layer, also called the VMM. The host is instead represented by physical hardware, & in some cases OS, that

Cloud Computing

defines an environment where VMM is running. The guest— Applications and users—interacts with a virtual network, such as a virtual private network (VPN), which is managed by specific software (VPN client) using physical network available on the node. VPNs are useful for creating an illusion of being within a different physical network & thus accessing the resources in it, which would otherwise not be available. The virtual environment is created by means of a software program. The ability to use software to emulate a wide variety of environments creates a lot of opportunities, previously less attractive because of excessive overhead introduced by the virtualization layer.

10.1 How Does Virtualization Work

For virtualizing the infrastructure, a virtualization layer is installed. It can involve the use of Bare-metal or Hosted Hypervisor architecture. It is important to understand how virtualization actually works. Firstly, in virtualization a virtual layer is installed on the systems. There are two prominent virtualization architectures, bare-metal and hosted hypervisor.

In a hosted architecture, a host OS is firstly installed and then a piece of software that is called as a hypervisor or it is called as a VM monitor or Virtual Machine Manager (VMM) (Figure 5). The VMM is installed on the top of host OS. The VMM allows the users to run different kinds of guests OSs within their own application window of a particular hypervisor. Different kinds of hypervisors can be Oracle's VirtualBox, Microsoft Virtual PC, VMware Workstation.

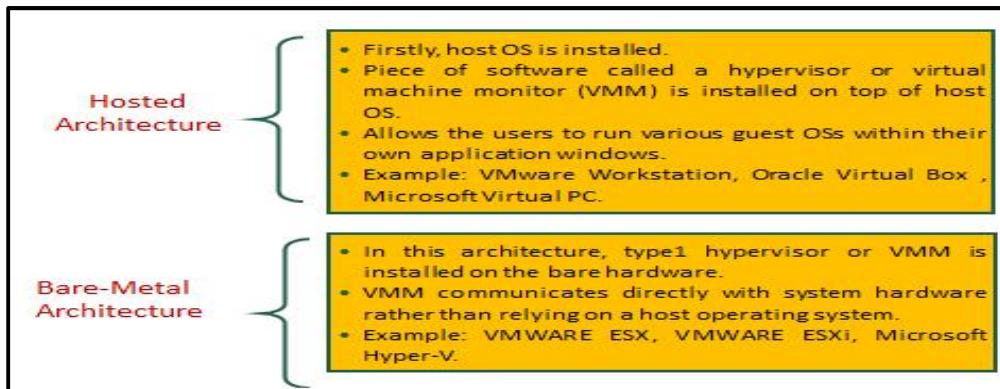


Figure 5: Hosted vs Bare-Metal Virtualization



Did you Know?

VMware server is a free application that is supported by Windows as well as by Linux OSs.

In a bare metal architecture, one hypervisor or VMM is actually installed on the bare metal hardware. There is no intermediate OS existing over here. The VMM communicates directly with the system hardware and there is no need for relying on any host OS. VMware ESXi and Microsoft Hyper-V are different hypervisors that are used for bare-metal virtualization.

A. Hosted Virtualization Architecture

A hosted virtualization architecture requires an OS (Windows or Linux) installed on the computer. The virtualization layer is installed as application on the OS.

Figure 6 illustrates the hosted virtualization architecture. At the lower layer, we have the shared hardware with a host OS running on this shared hardware. Upon the host OS, a VMM is running that and is creating a virtual layer which is enabling different kinds of OSs to run concurrently. So, you can see a scenario we have a hardware then we add an operating system then a hypervisor is added and different kinds of virtual machines can run on that particular virtual layer and each virtual machine can be running same or different kind of OSs.

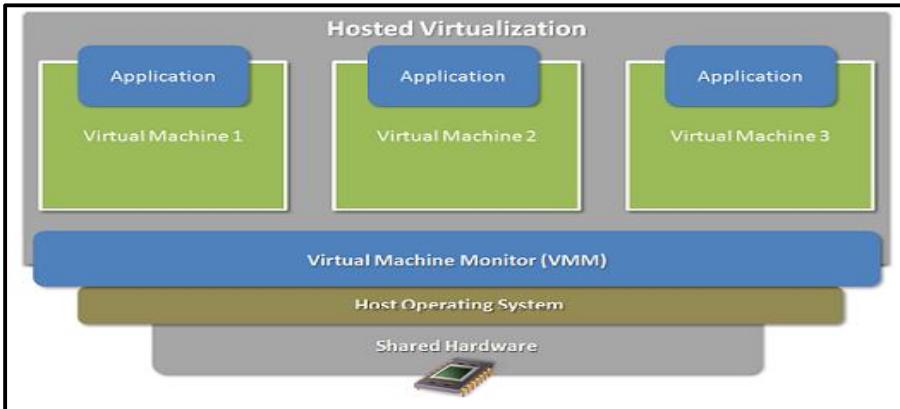
Unit 10: Virtualization Concepts

Figure 6: Hosted Virtualization Architecture

Advantages of Hosted Architecture

- Ease of installation and configuration
- Unmodified Host OS & Guest OS
- Run on a wide variety of PCs

Disadvantages of Hosted Architecture

- Performance degradation
- Lack of support for real-time OSs

B. Bare-Metal Virtualization Architecture

In a bare metal architecture, there is an underlying hardware but no underlying OS. There is just a VMM that is installed on that particular hardware and on that there are multiple VMs that are running on a particular hardware unit. As illustrated in the Figure 7, there is shared hardware that is running a VMM on which multiple VMs are running with simultaneous execution of multiple OSs.

Advantages of Bare-Metal Architecture

- Improved I/O performance
- Supports Real-time OS

Disadvantages of Bare-Metal Architecture

- Difficult to install & configure
- Depends upon hardware platform

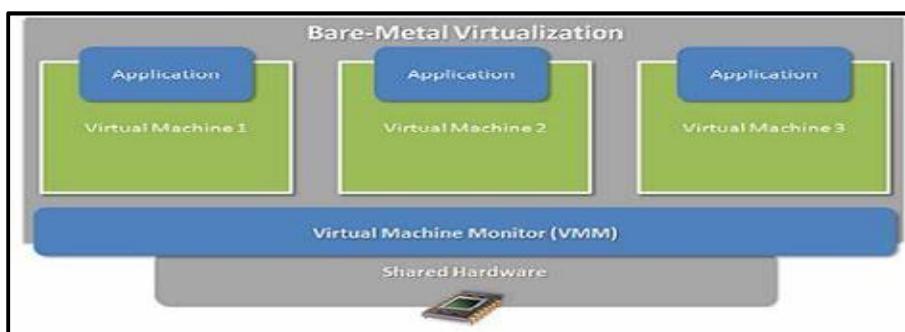


Figure 7: Bare-Metal Virtualization Scenario

10.2 Types of Virtualization

Virtualization covers a wide range of emulation techniques that are applied to different areas of computing. A classification of these techniques helps us better understand their characteristics and

Cloud Computing

use. Before discussing virtualization techniques, it is important to know about protection rings in OSs. The protection rings are used to isolate the OS from untrusted user applications. The OS can be protected with different privilege levels (Figure 8).

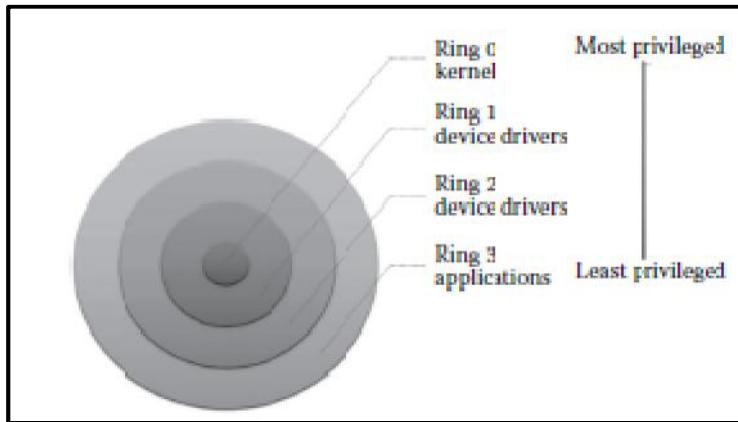


Figure 8: Protection Rings in OSs

Protection Rings in OSs

In protection ring architecture, the rings are arranged in hierarchical order from ring 0 to ring 3. The Ring 0 contains the programs that are most privileged, and ring 3 contains the programs that are least privileged. Normally, the highly trusted OS instructions will run in ring 0, and it has unrestricted access to physical resources. Ring 3 contains the untrusted user applications, and it has restricted access to physical resources. The other two rings (ring 1 & 2) are allotted for device drivers. The protection ring architecture restricts the misuse of resources and malicious behavior of untrusted user-level programs. For example, any user application from ring 3 cannot directly access any physical resources as it is the least privileged level. But the kernel of the OS at ring 0 can directly access the physical resources as it is the most privileged level. Depending on the type of virtualization, the hypervisor and guest OS will run in different privilege levels. Normally, the hypervisor will run with the most privileged level, and the guest OS will run at the least privileged level than the hypervisor. There are 4 virtualization techniques namely,

- Full Virtualization (Hardware Assisted Virtualization/ Binary Translation).
- Para Virtualization or OS assisted Virtualization.
- Hybrid Virtualization
- OS level Virtualization

Full Virtualization: The VM simulates hardware to allow an unmodified guest OS to be run in isolation. There are 2 type of full virtualizations in the enterprise market, software-assisted and hardware-assisted full virtualization. In both the cases, the guest OSs source information is not modified.

The software-assisted full virtualization is also called as Binary Translation (BT) and it completely relies on binary translation to trap and virtualize the execution of sensitive, non-virtualizable instruction sets. It emulates the hardware using the software instruction sets. It is often criticized for performance issue due to binary translation. The software that fall under software-assisted (BT) include:

- VMware workstation (32Bit guests)
- Virtual PC
- VirtualBox (32-bit guests)
- VMware Server

The hardware-assisted full virtualization eliminates the binary translation and directly interacts with hardware using the virtualization technology which has been integrated on X86 processors since 2005 (Intel VT-x and AMD-V). The guest OS's instructions might allow a virtual context execute privileged instructions directly on the processor, even though it is virtualized. There is

Unit 10: Virtualization Concepts

several enterprise software that support hardware-assisted- Full virtualization which falls under hypervisor type 1 (Bare metal) such as:

- VMware ESXi /ESX
- KVM
- Hyper-V
- Xen

Para Virtualization: The para-virtualization works differently from the full virtualization. It doesn't need to simulate the hardware for the VMs. The hypervisor is installed on a physical server (host) and a guest OS is installed into the environment. The virtual guests are aware that it has been virtualized, unlike the full virtualization (where the guest doesn't know that it has been virtualized) to take advantage of the functions. Also, the guest source codes can be modified with sensitive information to communicate with the host. The guest OSs require extensions to make API calls to the hypervisor.

Comparatively, in the full virtualization, guests issue hardware calls but in para virtualization, guests directly communicate with the host (hypervisor) using the drivers. The list of products which supports para virtualization are:

- Xen (Figure 9)
- IBM LPAR
- Oracle VM for SPARC (LDOM)
- Oracle VM for X86 (OVM)

However, due to the architectural difference between windows-based and Linux-based Xen hypervisor, Windows OS can't be para-virtualized. It does for Linux guest by modifying the kernel. VMware ESXi doesn't modify the kernel for both Linux and Windows guests.

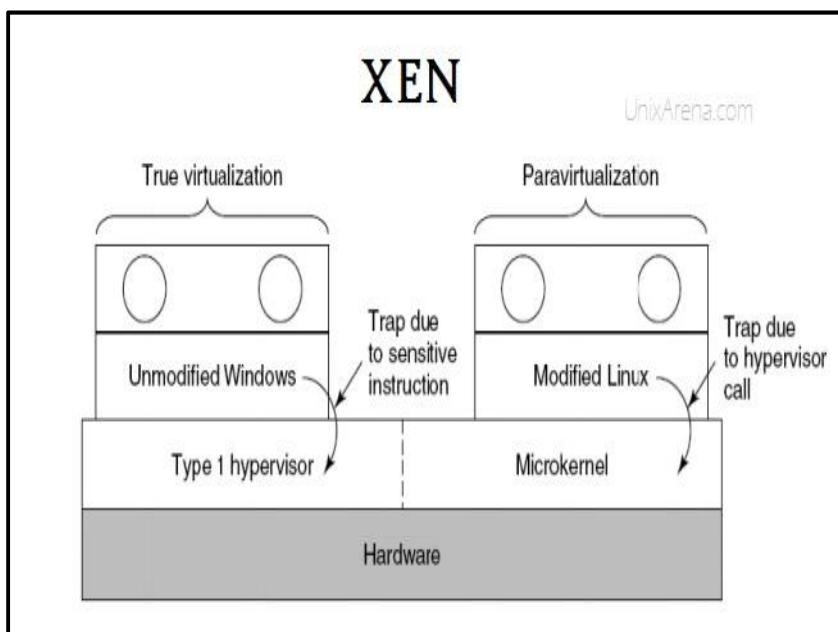


Figure 9: Xen Supports both Full-Virtualization and Para-Virtualization

Hybrid Virtualization (Hardware Virtualized with PV Drivers): In the hardware-assisted full virtualization, the guest OSs are unmodified and many VM traps occur and thus high CPU overheads which limit the scalability. Para virtualization is a complex method where guest kernel needs to be modified to inject the API. Therefore, due to the issues in full- and para- virtualization, engineers came up with hybrid paravirtualization, that is, a combination of both full and paravirtualization. The VM uses paravirtualization for specific hardware drivers (where there is a bottleneck with full virtualization, especially with I/O & memory intense workloads), and host uses full virtualization for other features. The following products support hybrid virtualization:

- Oracle VM for x86

Cloud Computing

- Xen
- VMware ESXi

OS Level Virtualization: It is widely used and is also known as “containerization”. The host OS kernel allows multiple user spaces aka instance. Unlike other virtualization technologies, there is very little or no overhead since it uses the host OS kernel for execution. Oracle Solaris zone is one of the famous containers in the enterprise market. The list of other containers:

- Linux LCX
- Docker
- AIX WPAR

Processor Virtualization: It allows the VMs to share the virtual processors that are abstracted from the physical processors available at the underlying infrastructure (Figure 10). The virtualization layer abstracts the physical processor to the pool of virtual processors that is shared by the VMs. The virtualization layer will be normally any hypervisors. But processor virtualization can also be achieved from distributed servers.

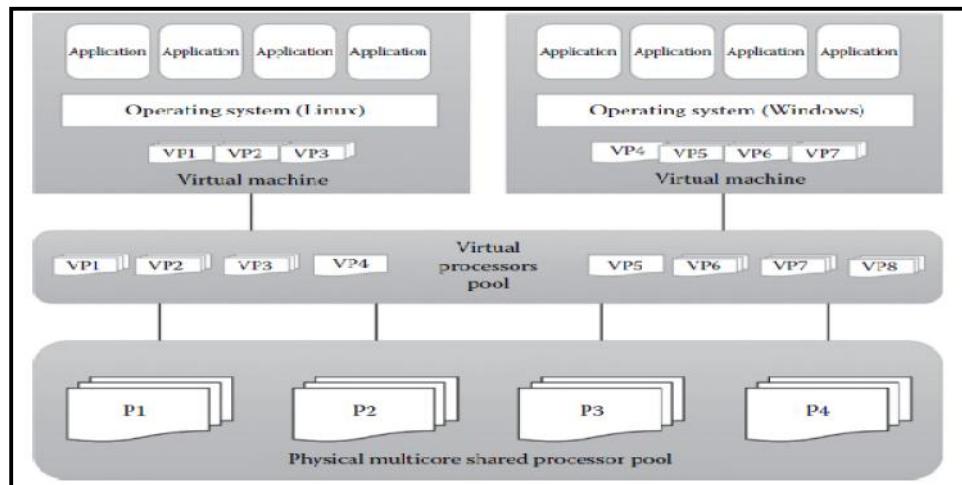


Figure 10: Processor Virtualization

Memory Virtualization: Another important resource virtualization technique is memory virtualization (Figure 11). It involves the process of providing a virtual main memory to the VMs known as memory virtualization or main memory virtualization. In main memory virtualization, the physical main memory is mapped to the virtual main memory as in the virtual memory concepts in most of the OSs.

The main idea of main memory virtualization is to map the virtual page numbers to the physical page numbers. All the modern x86 processors are supporting main memory virtualization. The main memory virtualization can also be achieved by using the hypervisor software. Normally, in the virtualized data centers, the unused main memory of the different servers will consolidate as a virtual main memory pool and can be given to the VMs.

Unit 10: Virtualization Concepts

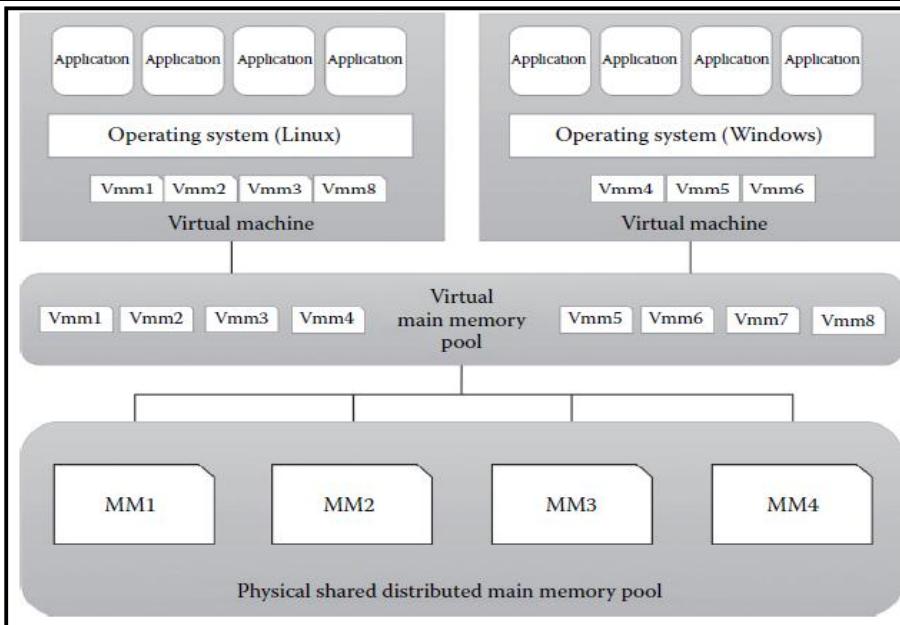


Figure 11: Memory Virtualization

Storage Virtualization: A form of resource virtualization where multiple physical storage disks are abstracted as a pool of virtual storage disks to the VMs (Figure 12). Normally, the virtualized storage will be called a logical storage.

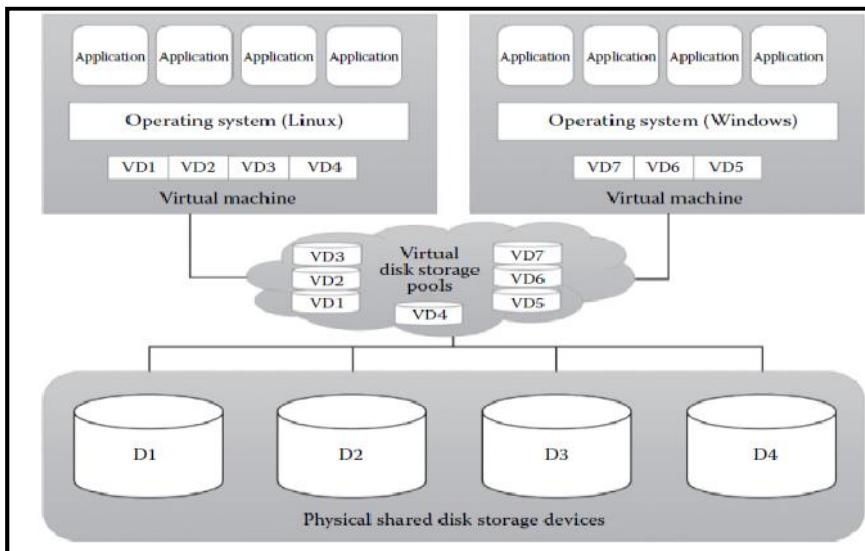


Figure 12: Storage Virtualization

Storage virtualization is mainly used for maintaining a backup or replica of the data that are stored on the VMs. It can be further extended to support the high availability of the data. It efficiently utilizes the underlying physical storage. Other advanced storage virtualization techniques are storage area networks (SAN) and network-attached storage (NAS).

Network Virtualization: It is a type of resource virtualization in which the physical network can be abstracted to create a virtual network (Figure 13). Normally, the physical network components like router, switch, and Network Interface Card (NIC) will be controlled by the virtualization software to provide virtual network components. Virtual network is a single software-based entity that contains the network hardware and software resources. Network virtualization can be achieved from internal network or by combining many external networks. It enables the communication between the VMs that share the physical network. There are different types of network access given to the VMs such as bridged network, network address translation (NAT), and host only.

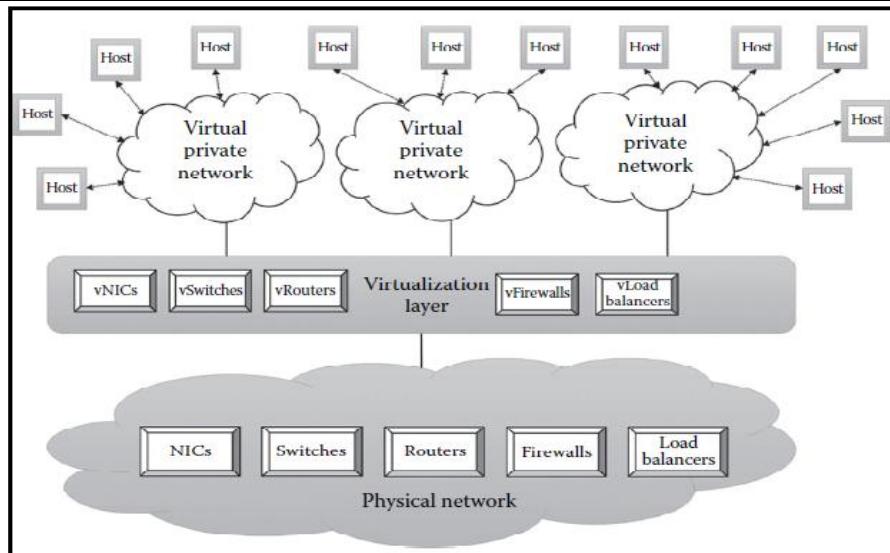
Cloud Computing

Figure 13: Network Virtualization

Data Virtualization: Data virtualization offers the ability to retrieve the data without knowing its type and the physical location where it is stored (Figure 14). It aggregates the heterogeneous data from the different sources to a single logical/virtual volume of data. This logical data can be accessed from any applications such as web services, E-commerce applications, web portals, Software-as-a-Service (SaaS) applications, and mobile application. It hides the type of the data and the location of the data for the application that access it and ensures the single point access to data by aggregating data from different sources. It is mainly used in data integration, business intelligence, and cloud computing.

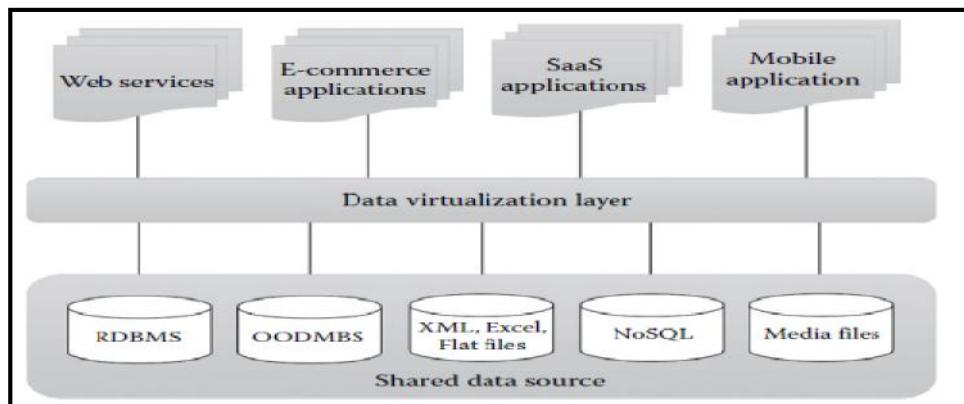


Figure 14: Data Virtualization

Application Virtualization: Application virtualization is the enabling technology for SaaS of cloud computing that offers the ability to the user to use the application without the need to install any software or tools in the machine (Figure 15). The complexity of installing the client tools or other supported software is reduced. Normally, the applications will be developed and hosted in the central server. The hosted application will be again virtualized, and the users will be given the separated/isolated virtual copy to access.

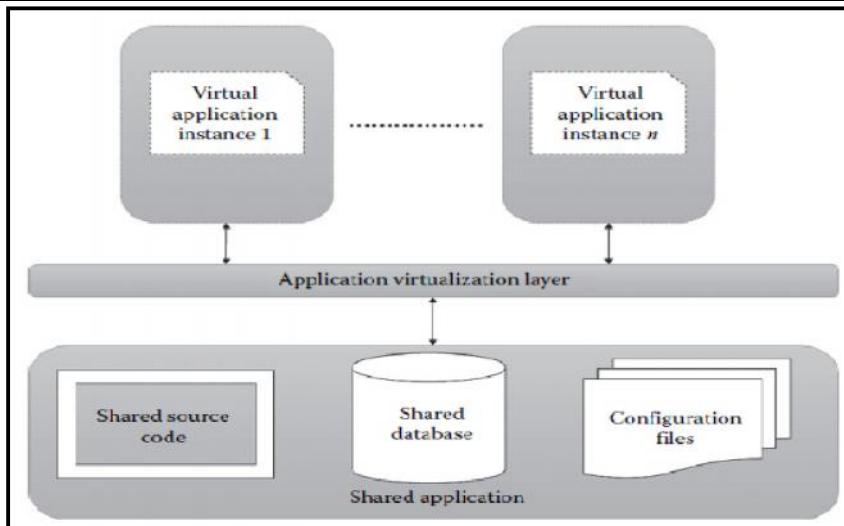


Figure 15: Application Virtualization

10.3 Pros of Virtualization

Increased Security

The ability to control the execution of a guest in a completely transparent manner opens new possibilities for delivering a secure, controlled execution environment. VM represents an emulated environment in which the guest is executed. All the operations of the guest are generally performed against the VM, which then translates & applies them to the host. By default, the file system exposed by the virtual computer is completely separated from the one of the host machines. This becomes the perfect environment for running applications without affecting other users in the environment.

Managed Execution

Virtualization of the execution environment not only allows increased security, but a wider range of features also can be implemented such as:

- o Sharing: Virtualization allows creation of a separate computing environments within same host, thereby, making it possible to fully exploit capabilities of a powerful guest (that would otherwise be underutilized).
- o Aggregation: A group of separate hosts can be tied together and represented to guests as a single virtual host. This function is naturally implemented in middleware for distributed computing, with a classical example represented by cluster management software, which harnesses the physical resources of a homogeneous group of machines and represents them as a single resource.
- o Emulation: Guest programs are executed within an environment that is controlled by the virtualization layer, which ultimately is a program. This allows for controlling and tuning the environment that is exposed to the guests.
- o Isolation: Virtualization allows providing guests – whether they are OSs, applications, or other entities – with a completely separate environment, in which they are executed. The guest program performs its activity by interacting with an abstraction layer, which provides access to the underlying resources.

Portability

Concept of portability applies in different ways according to the specific type of virtualization considered. In the case of a hardware virtualization solution, the guest is packaged into a virtual image that, in most cases, can be safely moved and executed on top of different virtual machines.

Cloud Computing

- In the case of programming-level virtualization, as implemented by the JVM or the .NET runtime, the binary code representing application components (jars or assemblies) can be run without any recompilation on any implementation of the corresponding virtual machine.
- This makes the application development cycle more flexible and application deployment very straight forward: One version of the application, in most cases, is able to run on different platforms with no changes.
- Portability allows having your own system always with you and ready to use as long as the required VMM is available. This requirement is, in general, less stringent than having all the applications and services you need available to you anywhere you go.

More Efficient Use of Resources

Multiple systems can securely coexist and share the resources of the underlying host, without interfering with each other. This is a prerequisite for server consolidation, which allows adjusting the number of active physical resources dynamically according to the current load of the system, thus creating the opportunity to save in terms of energy consumption and to be less impacting on the environment.

10.4 Cons of Virtualization

Virtualization also has downsides. The most evident is represented by a performance decrease of guest systems as a result of the intermediation performed by the virtualization layer. In addition, sub-optimal use of the host because of the abstraction layer introduced by virtualization management software can lead to a very inefficient utilization of the host or a degraded user experience. Less evident, but perhaps more dangerous, are the implications for security, which are mostly due to the ability to emulate a different execution environment.

Performance Degradation- Performance is definitely one of the major concerns in using virtualization technology. Since virtualization interposes an abstraction layer between the guest and the host, the guest can experience increased latencies. For instance, in case of hardware virtualization, where the intermediate emulates a bare machine on top of which an entire system can be installed, the causes of performance degradation can be traced back to the overhead introduced by following activities:

- Maintaining the status of virtual processors
- Support of privileged instructions (trap and simulate privileged instructions)
- Support of paging within VM
- Console functions

Inefficiency and Degraded User Experience- Virtualization can sometime lead to an inefficient use of the host. In particular, some of the specific features of the host cannot be exposed by the abstraction layer and then become inaccessible. In the case of hardware virtualization, this could happen for device drivers: VM can sometime simply provide a default graphic card that maps only a subset of the features available in the host. In the case of programming-level VMs, some of the features of the underlying OSs may become inaccessible unless specific libraries are used. For example, in the first version of Java the support for graphic programming was very limited and the look and feel of applications was very poor compared to native applications. These issues have been resolved by providing a new framework called Swing for designing the user interface, and further improvements have been done by integrating support for the OpenGL libraries in the software development kit.

Security Holes and New Threats- Virtualization opens the door to a new and unexpected form of phishing. The capability of emulating a host in a completely transparent manner led the way to malicious programs that are designed to extract sensitive information from the guest. In the case of hardware virtualization, malicious programs can preload themselves before the operating system and act as a thin virtual machine manager toward it. The operating system is then controlled and can be manipulated to extract sensitive information of interest to third parties.

Unit 10: Virtualization Concepts

Software Licensing Considerations- This is becoming less of a problem as more software vendors adapt to the increased adoption of virtualization, but it is important to check with your vendors to clearly understand how they view software use in a virtualized environment.

Possible Learning Curve- Implementing and managing a virtualized environment will require IT staff with expertise in virtualization. On the user side a typical virtual environment will operate similarly to the non-virtual environment. There are some applications that do not adapt well to the virtualized environment – this is something that your IT staff will need to be aware of and address prior to converting.

Summary

- Virtualization opens the door to a new and unexpected form of phishing. The capability of emulating a host in a completely transparent manner led the way to malicious programs that are designed to extract sensitive information from the guest.
- Virtualization raises abstraction. Abstraction pertains to hiding of the inner details from a particular user. Virtualization helps in enhancing or increasing the capability of abstraction.
- Virtualization enables sharing of resources much easily, it helps in increasing the degree of hardware level parallelism, basically, there is sharing of the same hardware unit among different kinds of independent units.
- In protection ring architecture, the rings are arranged in hierarchical order from ring 0 to ring 3. The Ring 0 contains the programs that are most privileged, and ring 3 contains the programs that are least privileged.
- In a bare metal architecture, one hypervisor or VMM is actually installed on the bare metal hardware. There is no intermediate OS existing over here. The VMM communicates directly with the system hardware and there is no need for relying on any host OS.
- The para-virtualization works differently from the full virtualization. It doesn't need to simulate the hardware for the VMs. The hypervisor is installed on a physical server (host) and a guest OS is installed into the environment.
- The software-assisted full virtualization is also called as Binary Translation (BT) and it completely relies on binary translation to trap and virtualize the execution of sensitive, non-virtualizable instruction sets.
- Memory virtualization is an important resource virtualization technique. In the main memory virtualization, the physical main memory is mapped to the virtual main memory as in the virtual memory concepts in most of the OSs.

Keywords

- ***Virtualization:*** Virtualization is a broad concept that refers to the creation of a virtual version of something, whether hardware, a software environment, storage, or a network.
- ***Hardware-assisted full virtualization:*** Hardware-assisted full virtualization eliminates the binary translation and directly interacts with hardware using the virtualization technology which has been integrated on X86 processors since 2005.
- ***Data Virtualization:*** Data virtualization offers the ability to retrieve the data without knowing its type and the physical location where it is stored.
- ***Application Virtualization:*** Application virtualization is the enabling technology for SaaS of cloud computing that offers the ability to the user to use the application without the need to install any software or tools in the machine.
- ***Memory Virtualization:*** It involves the process of providing a virtual main memory to the VMs known as memory virtualization or main memory virtualization.

Cloud Computing

- **Network Virtualization:** It is a type of resource virtualization in which the physical network can be abstracted to create a virtual network.

Self Assessment

1. _____ technology allows to run multiple operating systems as virtual machines.
 - A. Operating System
 - B. Virtualization
 - C. Fostering
 - D. Embellishment
2. In a _____ virtualization, the VMM communicates directly with system hardware rather than relying on a host operating system.
 - A. Bare-metal
 - B. Hybrid
 - C. Storage
 - D. Hosted
3. Which of the following is an example of Hosted virtualization hypervisors?
 - A. VMware Workstation
 - B. Oracle Virtual Box
 - C. Microsoft Virtual PC
 - D. All the above
4. Cloud computing is an abstraction based on the notion of pooling physical resources and presenting them as a _____ resource.
 - A. limited
 - B. restricted
 - C. virtual
 - D. homogeneous
5. _____ virtualization requires an operating system installed on the computer with the virtualization layer installed as an application on the OS.
 - A. Bare-metal
 - B. Hybrid
 - C. Storage
 - D. Hosted
6. CapEx stands for
 - A. Capacity expenses
 - B. Capital expenditure
 - C. Capital explosion
 - D. Capacity explosion
7. Which of the following is/are benefits of virtualizing the resources?
 - A. To increase the resource utilization
 - B. To increase the returns on investment

Unit 10: Virtualization Concepts

-
- C. Ability to transform hardware into software
 - D. All of the above
8. What is most commonly used for managing the resources for every virtual system?
- A. Hypervisor
 - B. Router
 - C. Cloud
9. Which is not a benefit of virtualization?
- A. Run on single operating system
 - B. Flexible and efficient allocation of resources
 - C. Lowers the cost of IT infrastructure
 - D. Remote access and rapid scalability
10. Which of the following properties does not specify the virtualization requirements?
- A. Efficiency property
 - B. Resource control property
 - C. Cluster property
 - D. Equivalence property
11. _____ are used to isolate the OS from untrusted user applications.
- A. Usage rings
 - B. Protection rings
 - C. Privacy rings
 - D. Trustful rings
12. An operating system running on a Type _____ VM is full virtualization.
- A. 1
 - B. 2
 - C. 3
 - D. 4
13. In _____ the virtual machine simulates hardware, so it can be independent of the underlying system hardware.
- A. Paravirtualization
 - B. full virtualization
 - C. emulation
 - D. None of the above
14. The software that supports virtual machine is called?
- A. VMM
 - B. Hypervisor
 - C. Kernal
 - D. Both A and B
15. Which of the following type of virtualization is also characteristic of cloud computing?
- A. Storage
 - B. CPU

Cloud Computing

- C. Application
- D. All of the above

Answers for Self Assessment

1. B	2. A	3. D	4. C	5. D
6. B	7. D	8. A	9. A	10. C
11. B	12. A	13. C	14. D	15. D

Review Questions

1. What are the pros and cons of virtualization?
2. Explain the concept of virtualization?
3. What are the different types of virtualization?
4. Write a short note on:
 - (a) Para-virtualization
 - (b) OS virtualization
 - (c) Storage virtualization
5. Differentiate storage and data virtualization?
6. Discuss about virtualized environments?



Further Readings

- Mastering Cloud Computing by "Rajkumar Buyya, Christian Vecchiola, S ThamaraiSelvi, Tata McGraw-Hill Education, 2021.
- Cloud Computing: Concepts, Technology and Architecture by Thomas Erl, Pearson Education.
- Cloud Computing Black Book by Kailash Jayaswal,Jagannath Kallakurchi, Donald J. Houde, Deven Shah, Kogent Learning Solutions, DreamTech Press.
- Virtualization: A Manager's Guide by Dan Kusnetzky, O'Reilly, 2011.



Web Links

1. What is Virtualization? (tutorialspoint.com)
2. What is Virtualization? Definition from SearchServerVirtualization (techtarget.com)
3. Virtualization - Wikipedia
4. What is Virtualization? How does it Work? [Understanding Virtualization] (kuberty.io)
5. What is Virtualization? - Definition from Techopedia
6. What is virtualization? (redhat.com)
7. <https://youtu.be/iBI31dmqSX0>
8. <https://youtu.be/Pl45CQYN3zI>

Unit 11: Virtual Machine

CONTENTS

- Objectives
- Introduction
- 11.1 Virtualization
- 11.2 Virtual Machine Attributes
- 11.3 Interpretation and Binary Translation
- 11.4 Hypervisors
- Summary
- Keywords
- Self Assessment
- Answers for Self Assessment
- Review Questions
- Further Readings

Objectives

After this lecture, you will be able to,

- Discover about Virtual Machine (VM) and their properties.
- Explore the different types of Virtual Machines.
- Understand the concept of interpretation and binary translation.
- Learn about hypervisors and their types.
- Know about different hypervisors such as: HLL VM: Xen, KVM, VMware, Virtual Box, Hyper-V.

Introduction

A software that creates a virtualized environment between the computer platform and the end-user in which the end user can operate software. It provides an interface identical to the underlying bare hardware. The Operating System (OS) creates the illusion of multiple processes, each executing on its own processor with its own (virtual) memory. Virtual machines are "an efficient, isolated duplicate of a real machine"- Popek and Goldberg. Popek and Goldberg introduced conditions for computer architecture to efficiently support system virtualization.

Virtual machine is a software that creates a virtualized environment between the computer platform and the end user in which the end user can operate software. The concept of virtualization applied to the entire machine involves:

- mapping of virtual resources or state to real resources.
- use of real machine instructions to carry out actions specified by the virtual machine instructions.
- Implemented by adding a layer of software to a real machine to support the desired VMs architecture.

VMs are a number of discrete identical execution environments on a single computer, each of which runs an OS (Figure 1). These allow applications written for one OS to be executed on a machine which runs a different OS which provide a greater level of isolation between processes than is achieved when running multiple processes on the same instance of an OS.

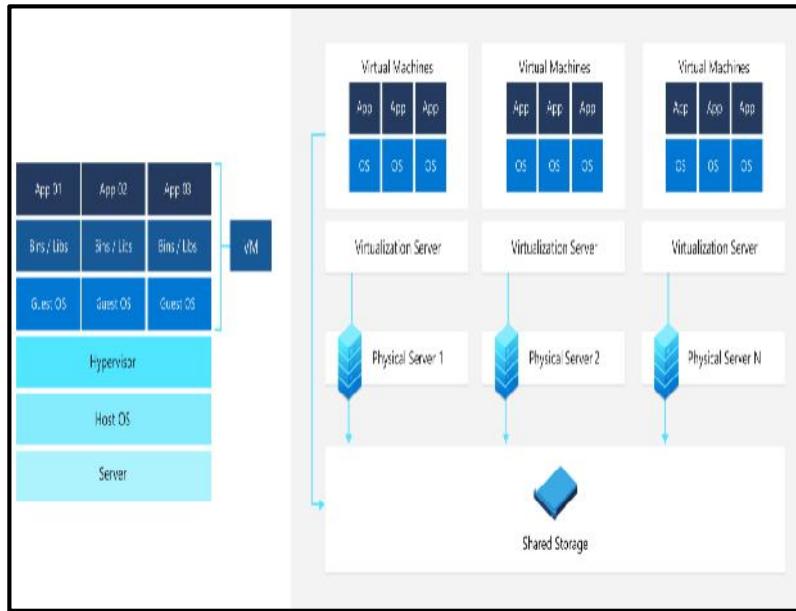


Figure 1: Virtual Machine Scenario

Virtual Machines: Virtual Computers Within Computers

VM is no different than any other physical computer like a laptop, smart phone or server. It has a CPU, memory, disks to store your files and can connect to the internet if needed. While the parts that make up your computer (called hardware) are physical and tangible, VMs are often thought of as virtual computers or software-defined computers within physical servers, existing only as code.

11.1 Virtualization

In computing, virtualization or virtualisation is the act of creating a virtual (rather than actual) version of something, including virtual computer hardware platforms, storage devices, and computer network resources. Virtualization began in the 1960s, as a method of logically dividing the system resources provided by mainframe computers between different applications. Since then, the meaning of the term has broadened. Virtualization technology has transformed hardware into software. It allows to run multiple Operating Systems (OSs) as virtual machines. Each copy of an operating system is installed in a virtual machine.



Figure 2: Virtualization Scenario

You can see a scenario (Figure 2), we have a VMware hypervisor that is also called as a Virtual Machine Manager. On a physical device, a VMware layer is installed out and, on that layer, we have six OSs that are running multiple applications over there, these can be the same kind of OSs or these can be the different kinds of OSs in it.

Why Virtualize

1. Share same hardware among independent users- Degrees of Hardware parallelism increases.

-
2. Reduced Hardware footprint through consolidation- Eases management and energy usage.
 3. Sandbox/migrate applications- Flexible allocation and utilization.
 4. Decouple applications from underlying Hardware- Allows Hardware upgrades without impacting an OS image.

Virtualization enables sharing of resources much easily, it helps in increasing the degree of hardware level parallelism, basically, there is sharing of the same hardware unit among different kinds of independent units, if we say that we have the same physical hardware and on that physical hardware, we have multiple OSs. There can be different users running on different kind of OSs. Therefore, we have a much more processing capability with us. This also helps in increasing the degree of hardware parallelism as well as there is a reduced hardware footprint throughout the VM consolidation. The hardware footprint that is overall hardware consumption also reduces out the amount of hardware that is wasted out that can also be reduced out. This consequently helps in easing out the management process and also to reduce the amount of energy that would have been otherwise consumed out by a particular hardware if we would have invested in large number of hardware machines would have been used otherwise. Virtualization helps in sandboxing capabilities or migrating different kinds of applications that in turn enables flexible allocations and utilization of the resources. Additionally, the decoupling of the applications from the underlying hardware is much easier and further aids in allowing more and more hardware upgrades without actually impacting any particular OS image.

Virtualization raises abstraction. Abstraction pertains to hiding of the inner details from a particular user. Virtualization helps in enhancing or increasing the capability of abstraction. It is very similar to how the virtual memory operates. It helps to access the larger address spaces physical memory mapping is actually hidden by an OS with the help of paging. It can be similar to hardware emulators where codes are allowed on one architecture to run on a different physical device such as virtual devices central processing unit, memory or network interface cards etc. No botheration is actually required out regarding the hardware details of a particular machine. The confinement to the excess of hardware details helps in raising out the abstraction capability through virtualization.

Basically, we have certain requirements for virtualization, first is the efficiency property. Efficiency means that all innocuous instructions are executed by the hardware independently. Then, the resource control property means that it is impossible for the programs to directly affect any kind of system resources. Furthermore, there is an equivalence property that indicates that we have a program which has a virtual machine manager or hypervisor that performs in a particular manner, indistinguishable from another program that is running on it.

11.2 Virtual Machine Attributes

Virtual machine is a software that creates a virtualized environment between the computer platform and the end user in which the end user can operate software. The section below discusses the different characteristics for the same.

History of Virtual Machine

Virtualization was first introduced in the 1960s to allow partitioning of large, mainframe hardware. In the 1990s, researchers began to see how virtualization could solve some of the problems associated with the proliferation of less expensive hardware, including under-utilization, escalating management costs and vulnerability. The concept of a VM was introduced around 1960. The evolution of time-sharing technique, where each program has full access to all computer resources but at a time, only one program will be executed. The system switch between programs in time slices while saving and restoring program states each time. With the use of the time-sharing method, multiple users can use the computer system concurrently. IBM research centers evolved the time-sharing method as VMs. CP-67 was the first available VM architecture. Evidently, the systems with multiple VMs on a single host and single VM on multiple hosts were developed.

Properties of a Virtual Machine

Virtual Hardware

- o Each VM has its own set of virtual hardware (e.g., RAM, CPU, NIC, etc.) upon which an operating system and applications are loaded.

Cloud Computing

- OS sees a consistent, normalized set of hardware regardless of the actual physical hardware components.

Partitioning

- Multiple applications and OSs can be supported within a single physical system.
- There is no overlap amongst memory as each Virtual Memory has its own memory space.

Isolation

- VMs are completely isolated from host machine and other VMs. If a VM crashes, all others are unaffected.
- Data does not leak across VMs.

Identical Environment

- VMs can have a number of discrete identical execution environments on a single computer, each of which runs an OS.

Other VM Features

- Each VM has its own set of virtual hardware (e.g., RAM, CPU, NIC, etc.) upon which an operating system and applications are loaded.
- OS sees a consistent, normalized set of hardware regardless of the actual physical hardware components.
- Host system resources are shared among the various VMs. For example, if a host system has 8GB memory where VMs are running, this amount will be shared by all the VMs, depending upon the size of the allocation.
- One of the best features of using Virtual machines is we can run multiple OSs/VMs in parallel on one host system.
- VMs are isolated from one another, thus secure from malware or threat from any other compromised VM running on the same host.
- Direct exchange of data and mutual influencing are prevented.
- Transfer of VMs to another system can be implemented by simply copying the VM data since the complete status of the system is saved in a few files.
- VMs can be operated on all physical host systems that support the virtualization environment used.

Virtual Machine Architecture

- Runtime software is the virtualization software that implements the Process VM. It is implemented at the API level of the computer architecture above the combined layer of OS and Hardware. This emulates the user-level instructions as well as OS or library calls.
- For the system VM, the virtualization software is called Virtual Machine Monitor(VMM).
- This software is present between the host hardware machine and the guest software.
- VMM emulates the hardware ISA allowing the guest software to execute a different ISA.

Virtual Machine Taxonomy

Figure 3 depicts the taxonomy for the virtual machines. Let us discuss each one of them below:

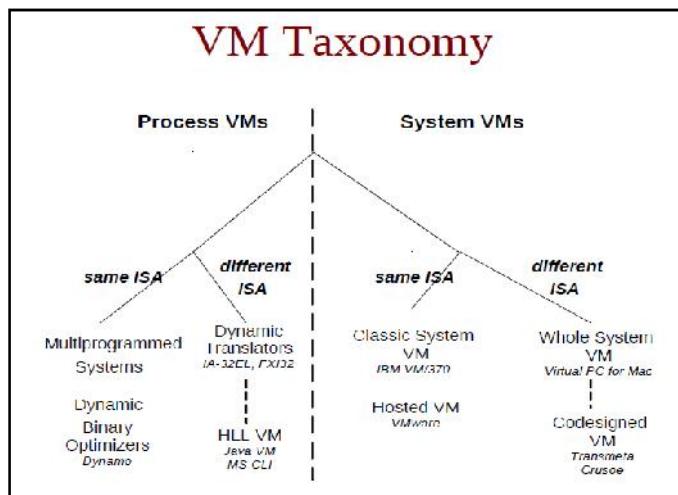


Figure 3: Taxonomy of Virtual Machines

Process Virtual Machines: These are also known as Application VM (Figure 4). The virtualization below the API or ABI, providing virtual resources to a single process executed on a machine is called as the process virtualization. It is created for the process alone, destroyed when process finishes.

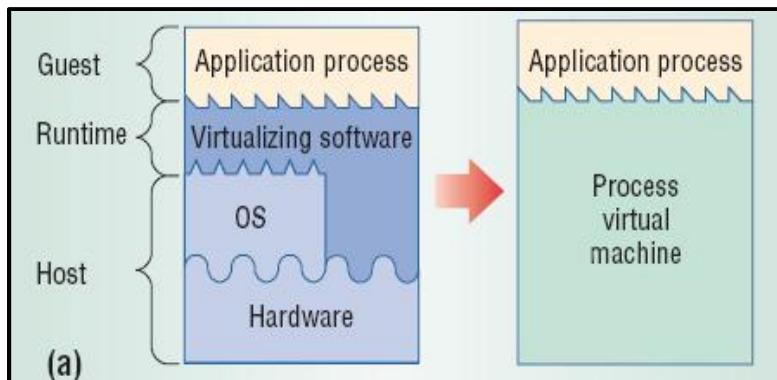


Figure 4: Process VM

Multiprogrammed Systems: Each application is given effectively separate access to resources, managed by the OS.

Emulators and Translators:

- Executes program binaries compiled for different instruction sets.
- Slower, requiring hardware interpretation.
- Optimization through storing blocks of converted code for repeated execution.

Optimizers, same ISA: Perform code optimization during translation and execution.

High-Level-Language VM:

- Cross-platform compatibility.
- Programs written for an abstract machine, which is mapped to real hardware through a VM.
 - Sun Micro systems Java VM
 - Microsoft Common Language Infrastructure, .NET framework.

System Virtual Machines: These correspond to the virtualized hardware below the ISA. The single host can run multiple isolated OSs (Figure 5). The servers running different OSs but in isolation between concurrent systems. The hardware managed by the Virtual Machine Manager (VMM).Classically, the

VMM runs on bare hardware, directly interacting with resources. It intercepts and interprets guest OS actions.

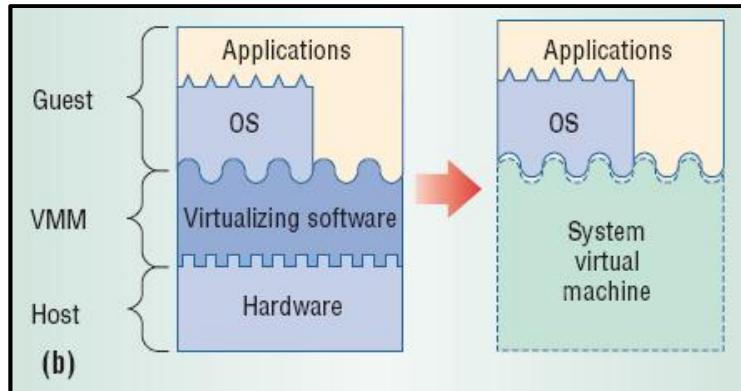


Figure 5: System Virtual Machines

Uses of Virtual Machines

- Building and deploying apps to the cloud.
- Trying out a new operating system (OS), including beta releases.
- Spinning up a new environment to make it simpler and quicker for developers to run dev-test scenarios.
- Backing up your existing OS.
- Accessing virus-infected data or running an old application by installing an older OS.
- Running software or apps on operating systems that they were not originally intended for.

Benefits of Virtual Machines

While VMs run like individual computers with individual operating systems and applications, they have the advantage of remaining completely independent of one another and the physical host machine. A piece of software called a hypervisor or virtual machine manager, lets you run different operating systems on different virtual machines at the same time. This makes it possible to run Linux VMs, for example, on a Windows OS or to run an earlier version of Windows on more current Windows OS. As the VMs are independent of each other, they are also extremely portable. You can move a VM on a hypervisor to another hypervisor on a completely different machine almost instantaneously. Because of their flexibility and portability, virtual machines provide many benefits, such as:

Cost Savings- Running multiple virtual environments from one piece of infrastructure means that you can drastically reduce your physical infrastructure footprint. This boosts your bottom line—decreasing the need to maintain nearly as many servers and saving on maintenance costs and electricity.

Agility and Speed- Spinning up a VM is relatively easy and quick and is much simpler than provisioning an entire new environment for your developers. Virtualisation makes the process of running dev-test scenarios a lot quicker.

Lowered Downtime- VMs are so portable and easy to move from one hypervisor to another on a different machine—this means that they are a great solution for backup, in the event the host goes down unexpectedly.

Scalability- VMs allow you to more easily scale your apps by adding more physical or virtual servers to distribute the workload across multiple VMs. As a result you can increase the availability and performance of your apps.

Security Benefits- Because VMs run in multiple OSs, using a guest operating system on a VM allows you to run apps of questionable security and protects your host OS. VMs also allow for better security forensics and are often used to safely study computer viruses, isolating the viruses to avoid risking their host computer.

Isolated environment provided by VMs- If you are a tester or security analyst then VMs will be a good idea to run multiple applications and services in an isolation using VMs because they do not affect each other.

Easy to Backup and Clone- All the VMs are stored on the physical hard drive of our host or physical machine in the file format. Thus, they can be easily back up, moved, or cloned in real-time is one of the popular benefits we get from running a virtual machine.

Faster Server Provisioning- VMs are easy to install, eliminating the cumbersome and time-consuming installation of applications on servers. For example, if you want a new server to run some application then it is very easy and fasts to deploy pre-configured VM templates instead of installing a new server OS on a physical machine. The same goes for cloning existing applications to try something new.

Beneficial in Disaster Recovery- As VM doesn't depend upon the underlying hardware, thus they are independent of the hardware or CPU model on which it is running. Hence, we can easily replicate VMs to cloud or offsite, so in some disaster situations, it would be easy to recover and get online in less span of time as we don't need to care for some particular server manufacturer or server model.

Use Older Applications for a Longer Time- Well, still many companies are using old applications but crucial to them and couldn't support modern hardware or operating system. In such situations, even the company wants, the IT would never prefer to touch them. However, we can pack such applications in a VM with the compatible old operating system and old virtual hardware. In this way, it will be possible to switch to modern hardware while keeping the old software stack intact.

Virtual Machine is Easily Portable- A single server running with some particular operating system software is not easy to move from one place to another, whereas if we have virtualized the same, then it becomes very easy to move data and OS from one physical server to another, situated somewhere else with the minimal workforce and without heavy transportation requirements.

Better Usage of Hardware Resources- Our modern computer or server hardware is quite powerful, using a single operating system and a couple of applications can't churn out the maximum juice of it. Thus, using VMs not only efficiently use the power of the CPU but allows the companies to save hundreds of bucks from spending on hardware.

Made Cloud Computing Possible- Yes, without VMs there will be no cloud computing because the whole idea behind it to provide an instant provision of machines running either Windows or Linux OS; it is only possible with the help of pre-build templates ready to deploy as VMs on some remote data center hardware. For example, Digital Ocean, AWS, and Google Cloud. So, next time whenever you heard "Cloud hosting" or "Virtual Private Server" hosting, remember it is a VM running on data center hardware.

11.3 Interpretation and Binary Translation

Emulation is required for implementing many VMs. It is the process of implementing the interface and functionality of one (sub)system on a (sub)system having a different interface and functionality. There are terminal emulators, such as for VT100, xterm, putty. There are various methods for enabling a (sub)system to present the same interface and characteristics as other ways of implementing emulation.

Interpretation: Interpretation involves relatively inefficient instruction-at-a-time.

Instruction Set Emulation- Binaries in source instruction set can be executed on machine implementing target instruction set. e.g., IA-32 execution layer.

Binary Translation: Binary translation involves block-at-a-time optimization for repeated. Example: Execution of programs compiled for instruction set A on a machine that executes instruction set B. The other features of binary translation include:

- Intercept OS code: Run-time translation of some OS instructions.
- User-level code is directly executed on the real hardware.
- No modifications to the OS are needed: the guest OS is not aware of virtualization.
- Specific device drivers are required.

Table 1 shows the difference between Interpretation vs Binary Translation.

Table 1: Interpretation vs Binary Translation

Interpretation	Binary Translation
Simple and easy to implement, portable	Complex implementation
Low performance	High initial translation cost, small execution cost
Threaded interpretation	Selective compilation

11.4 Hypervisors

VMs are widely used instead of physical machines in the IT industry today. The VMs support green IT solutions, and its usage increases resource utilization, making the management tasks easier. Since the VMs are mostly used, the technology that enables the virtual environment also gets attention in industries and academia. The virtual environment can be created with the help of a software tool called hypervisors.

Hypervisors are the software tool that sits in between VMs and physical infrastructure and provides the required virtual infrastructure for VMs. Hypervisors are also called as Virtual Machine Manager (VMM) (Figure 6). These are the key drivers in enabling virtualization in cloud data centers. Different hypervisors are being used in the IT industry. Some of the examples are VMware, Xen, Hyper-V, KVM, and OpenVZ.

The virtual infrastructure means virtual CPUs (vCPUs), virtual memory, virtual NICs (vNICs), virtual storage, and virtual I/O devices. The fundamental element of hardware virtualization is the hypervisor, or VMM that helps to recreate a hardware environment in which Guest Operating Systems (OSs) are installed.

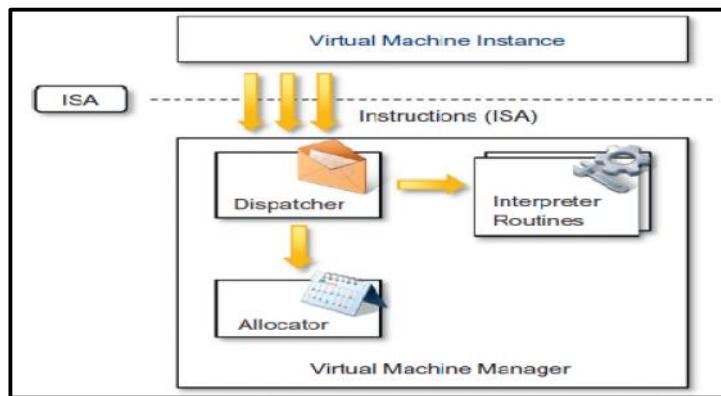


Figure 6: Internal Organization of a Virtual Machine Manager

There are 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 when ever a VM executes a privileged instruction: a trap is triggered and the corresponding routine is executed.

The design and architecture of a VMM, together with the underlying hardware design of the host machine, determine the full realization of hardware virtualization, where a guest OS can be transparently executed on top of a VMM as though it were run on the underlying hardware.

The criteria that need to be met by a VMM to efficiently support virtualization were established by Goldberg and Popekin 1974. The 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: VMM should be incomplete control of virtualized resources.
- Efficiency: A statistically dominant fraction of the machine instructions should be executed without intervention from the VMM.

Before the hypervisors are introduced, there was a one-to-one relationship between hardware and OSs. This type of computing results in underutilized resources.

After the hypervisors are introduced, it became a one-to-many relationship. With the help of hypervisors, many OSs can run and share a single hardware.

Types of Hypervisors

Hypervisors are generally classified into two categories:

- Type 1 or bare metal hypervisors
- Type 2 or hosted hypervisors

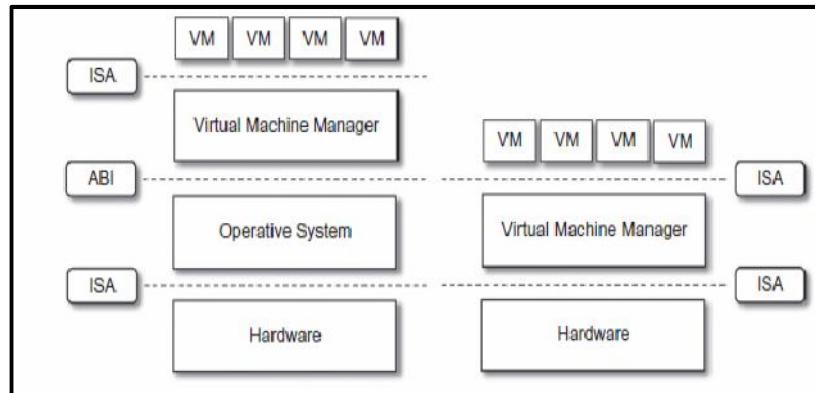


Figure 7: Hosted (left) and Native (right) VMs

Type I Hypervisors run directly on top of the hardware. Therefore, they take the place of the OSs 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 OSs. These are also called a native VM since it runs natively on the hardware. The other characteristics of Type I hypervisors include:

- Can run and access physical resources directly without the help of any host OS.
- Additional overhead of communicating with the host OS is reduced and offers better efficiency when compared to type 2 hypervisors.
- Used for servers that handle heavy load and require more security.
- Examples- Microsoft Hyper-V, Citrix XenServer, VMWare ESXi, and Oracle VM Server for SPARC.

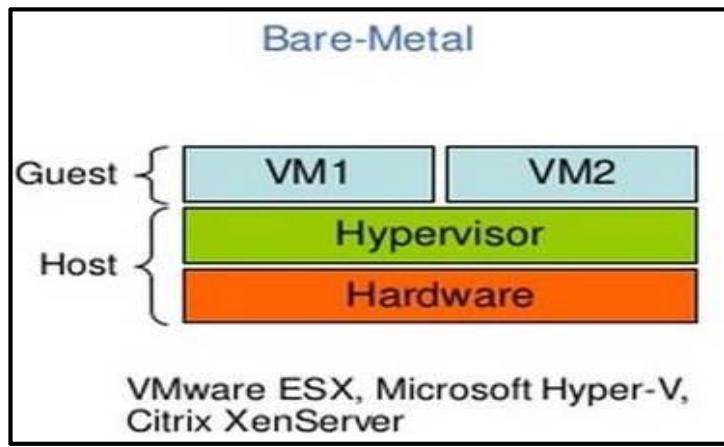


Figure 8: Bare-Metal Virtualization

Type II Hypervisors require the support of an operating system to provide virtualization services (Figure 9). This means that they are programs managed by the OS, which interact with it through the API and emulate the ISA of virtual hardware for guest OSs. This type of hypervisor is also called a hosted or embedded VM since it is hosted within an OS (Figure 10). Hosted virtualization requires the host OS and does not have direct access to the physical hardware. The host OS is also known as physical host, which has the direct access to the underlying hardware. However, the major disadvantage of this approach is if the host OS fails or crashes, it also results in crashing of VMs. So, it is recommended to use type 2 hypervisors only on client systems where efficiency is less critical. Examples- VMWare Workstation and Oracle Virtualbox.

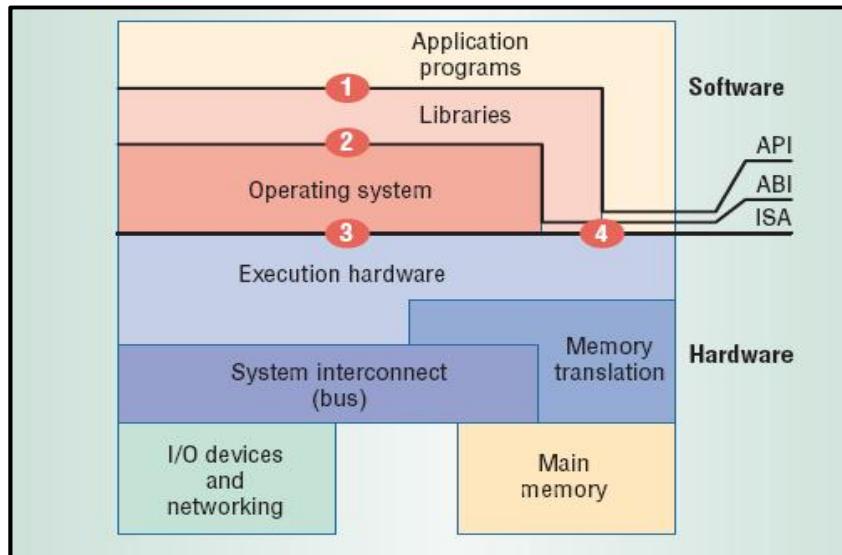


Figure 9: Type II Hypervisor

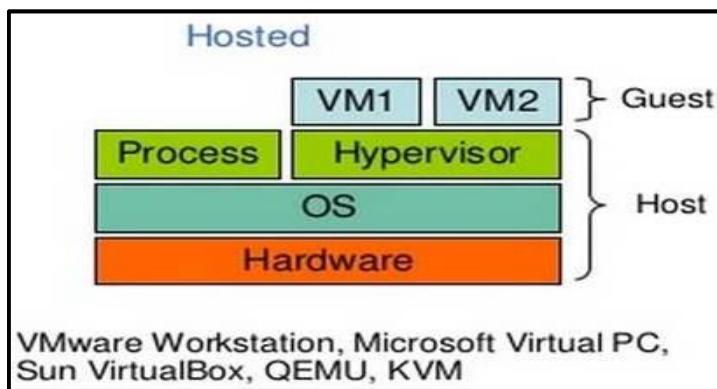


Figure 10: Hosted Virtualization

Summarized Implementation of Hypervisors

Vary greatly, with options including:

- Type 0 Hypervisors- Hardware-based solutions that provide support for virtual machine creation and management via firmware. Example: IBM LPARs and Oracle LDOMs are examples.
- Type 1 Hypervisors- Operating-system-like software built to provide virtualization. Example: Including VMware ESX, JoyentSmartOS, and Citrix XenServer.
- Type 1 Hypervisors- Also includes general-purpose operating systems that provide standard functions as well as VMM functions. Example: Microsoft Windows Server with HyperV and RedHat Linux with KVM.
- Type 2 Hypervisors- Applications that run on standard OSs but provide VMM features to guest OSs. Example: VMware Workstation and Fusion, Parallels Desktop, and Oracle VirtualBox.

Other Variations Include: Much variation exists due to breadth, depth and importance of virtualization in modern computing.

Para Virtualization- Technique in which the guest operating system is modified to work in cooperation with the VMM to optimize performance.

Programming-environment Virtualization- VMMs do not virtualize real hardware but instead create an optimized virtual system. It is used by Oracle Java and Microsoft .Net.

Emulators- Allow applications written for one hardware environment to run on a very different hardware environment, such as a different type of CPU.

Application Containment- Not virtualization at all but rather provides virtualization-like features by segregating applications from the operating system, making them more secure, manageable. It is included in Oracle Solaris Zones, BSD Jails, and IBM AIX WPARs.

Xen

An open-source initiative implementing a virtualization platform based on paravirtualization. Xen is a VMM for IA-32 (x86, x86-64), IA-64 and PowerPC 970 architectures. It allows several guest OSs to be executed on the same computer hardware concurrently. It was initially created by University of Cambridge, Computer Laboratory and is now developed and maintained by Xen community as free software, as well as Citrix XenServer Commercial version variant. It is the central part of Amazon.com's cloud computing platform, EC2 (Elastic Compute Cloud) that allows the users to rent virtual computers on which to run their own computer applications.

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). Recently Xen has been advanced to support full virtualization using hardware-assisted virtualization. The most popular implementation of paravirtualization, which, in contrast with full virtualization, allows high performance execution of guest OSs. This is made possible by eliminating the performance loss while executing instructions that require special management. This is done by

Cloud Computing

modifying portions of the guest OSs run by Xen with reference to the execution of such instructions. Therefore, it is not a transparent solution for implementing virtualization. This is particularly true for x86, which is the most popular architecture on commodity machines and servers.

Xen Architecture

Figure 11 depicts the Xen architecture consisting of three different layers as discussed below:

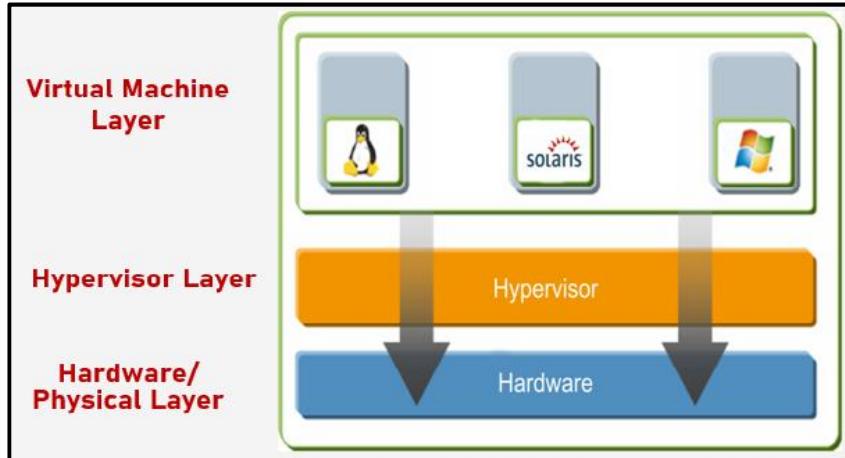


Figure 11: Xen Architecture

- **Hardware or Physical Layer:** Physical hardware components including memory, CPU, network cards, and disk drives.
- **Hypervisor Layer:** Thin layer of software that runs on top of the hardware. The Xen hypervisor gives each virtual machine a dedicated view of the hardware.
- **Virtual Machine Layer:** OS hosted on the hypervisor and appearing to the user as a separate physical computer. However, the machine shares physical resources with other virtual machines, and it is portable because the virtual machine is abstracted from the physical hardware.

A Xen-based system is managed by the Xen hypervisor, which runs in the highest privileged mode and controls the access of guest OS to the underlying hardware. Guest OSs are executed within domains, which represent VM instances. Moreover, specific control software, which has privileged access to the host and controls all the other guest OSs, is executed in a special domain called Domain 0. This is the first one that is loaded once the VMM has completely booted, and it hosts a Hyper Text Transfer Protocol (HTTP) server that serves requests for VM creation, configuration, and termination. This component constitutes the embryonic version of a distributed VMM, which is an essential component of cloud computing systems providing Infrastructure-as-a-Service (IaaS) solutions.

KVM (Kernel-based Virtual Machine (KVM))

An open-source virtualization technology built into Linux®. It is a full virtualization solution for Linux on x86 hardware containing virtualization extensions (Intel VT or AMD-V). It consists of a loadable kernel module, kvm.ko, that provides the core virtualization infrastructure and a processor specific module, kvm-intel.ko or kvm-amd.ko. KVM lets you turn Linux into a hypervisor that allows a host machine to run multiple, isolated virtual environments called guests or virtual machines (VMs).

KVM was first announced in 2006 and merged into the mainline Linux kernel version a year later. As KVM is part of existing Linux code, it immediately benefits from every new Linux feature, fix, and advancement without additional engineering. KVM converts Linux into a type-1 (bare-metal) hypervisor. All hypervisors need some operating system-level components—such as a memory manager, process scheduler, input/output (I/O) stack, device drivers, security manager, a network stack, and more—to run VMs.

How does KVM work?

- KVM has all these components because it's part of the Linux kernel.
- Every VM is implemented as a regular Linux process, scheduled by the standard Linux scheduler, with dedicated virtual hardware like a network card, graphics adapter, CPU(s), memory, and disks.

Implementing KVM

You have to run a version of Linux that was released after 2007 and it needs to be installed on X86 hardware that supports virtualization capabilities. If both of those boxes are checked, then all you have to do is load 2 existing modules (a host kernel module and a processor-specific module), an emulator, and any drivers that will help you run additional systems. But implementing KVM on a supported Linux distribution—like Red Hat Enterprise Linux—expands KVM's capabilities, letting you swap resources among guests, share common libraries, optimize system performance, and a lot more.

KVM Features

Security- KVM uses a combination of security-enhanced Linux (SELinux) and secure virtualization (sVirt) for enhanced VM security and isolation. SELinux establishes security boundaries around VMs. sVirt extends SELinux's capabilities, allowing Mandatory Access Control (MAC) security to be applied to guest VMs and preventing manual labeling errors.

Storage- KVM is able to use any storage supported by Linux, including some local disks and network-attached storage (NAS). Multipath I/O may be used to improve storage and provide redundancy. KVM also supports shared file systems so VM images may be shared by multiple hosts. Disk images support thin provisioning, allocating storage on demand rather than all up front.

Hardware Support- KVM can use a wide variety of certified Linux-supported hardware platforms. Because hardware vendors regularly contribute to kernel development, the latest hardware features are often rapidly adopted in the Linux kernel.

Memory Management- KVM inherits the memory management features of Linux, including non-uniform memory access and kernel same-page merging. The memory of a VM can be swapped, backed by large volumes for better performance, and shared or backed by a disk file.

Live Migration- KVM supports live migration, which is the ability to move a running VM between physical hosts with no service interruption. The VM remains powered on, network connections remain active, and applications continue to run while the VM is relocated. KVM also saves a VM's current state so it can be stored and resumed later.

Performance and Scalability- KVM inherits the performance of Linux, scaling to match demand load if the number of guest machines and requests increases. KVM allows the most demanding application workloads to be virtualized and is the basis for many enterprise virtualization setups, such as data centers and private clouds (via OpenStack®).

Scheduling and Resource Control- In the KVM model, a VM is a Linux process, scheduled and managed by the kernel. The Linux scheduler allows fine-grained control of the resources allocated to a Linux process and guarantees a quality of service for a particular process. In KVM, this includes the completely fair scheduler, control groups, network name spaces, and real-time extensions.

Lower Latency and Higher Prioritization- The Linux kernel features real-time extensions that allow VM-based apps to run at lower latency with better prioritization (compared to bare metal). The kernel also divides processes that require long computing times into smaller components, which are then scheduled and processed accordingly.

VMware

VMware Workstation is the most dependable, high-performing, feature-rich virtualization platform for your Windows or Linux PC (Figure 12 and Figure 13). It allows one physical PC to run multiple operating systems at the same time. Actually, no restarting or hard-drive partitioning is required. The software developers rely on workstation to develop and test client-server, Web and cloud applications in a replica of their production environments.

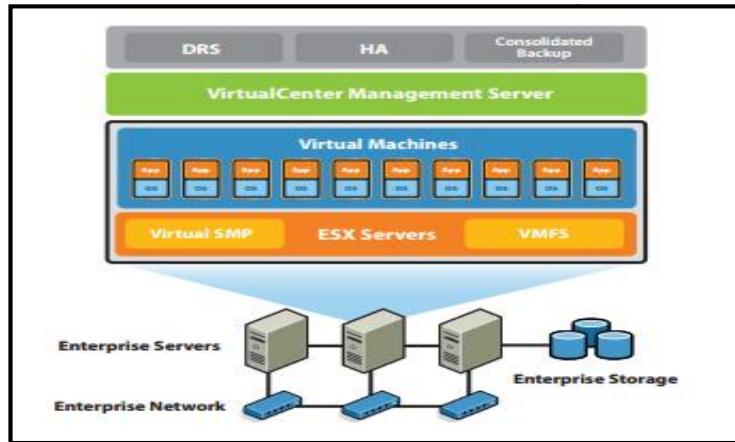


Figure 12: VMware Scenario

VMware's technology is based on the concept of full virtualization, where the underlying hardware is replicated and made available to the guest OS, 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.

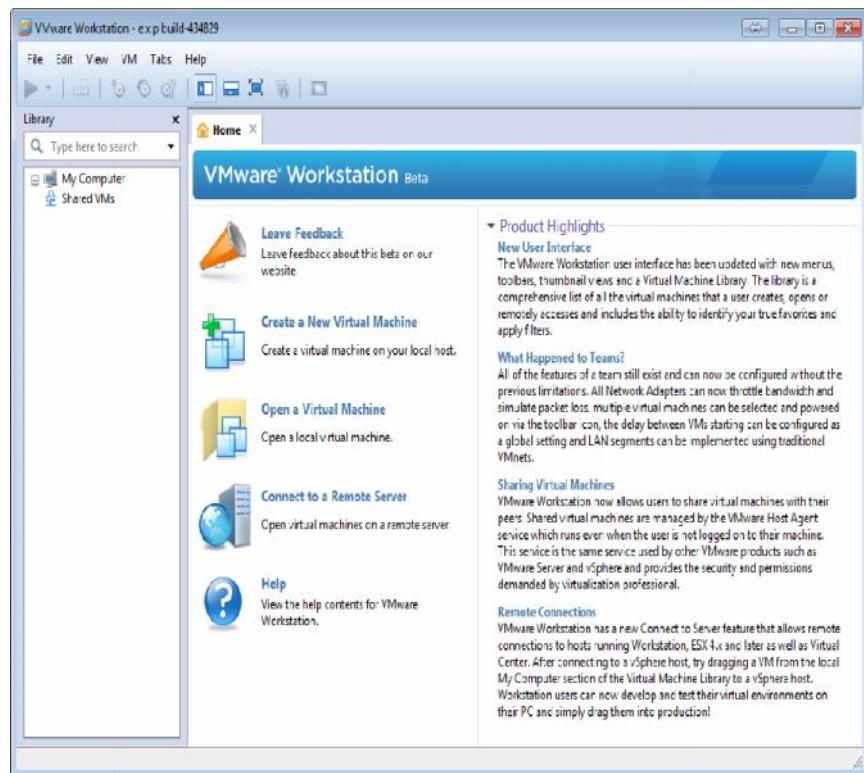


Figure 13: VMware Workstation Layout

Key Benefits of VMware

- Access anytime, anywhere
- Run applications in Windows, Linux and other systems at the same time without restarting.
- Remotely access VMs running on VMware.
- Run as a server to host applications for your team, department or anyone in your organization.

- Create VMs that are encrypted, block USB devices and have read-only settings.

Virtual Box

It is a powerful x86 and AMD64/Intel64 virtualization product for enterprise as well as home use. An extremely feature rich, high performance product for enterprise customers also, the only professional solution that is freely available as Open Source Software under the terms of the GNU General Public License (GPL) version 2. Presently, it runs on Windows, Linux, Macintosh, and Solaris hosts and supports a large number of guest OSs including but not limited to Windows (NT 4.0, 2000, XP, Server 2003, Vista, Windows 7, Windows 8, Windows 10), DOS/Windows 3.x, Linux (2.4, 2.6, 3.x and 4.x), Solaris and OpenSolaris, OS/2, and OpenBSD.

Oracle VM VirtualBox-

- Cross-platform virtualization application.
- Installs on your existing Intel or AMD-based computers, whether they are running Windows, Mac OS X, Linux, or Oracle Solaris operating systems (OSes).
- Extends the capabilities of your existing computer so that it can run multiple OSes, inside multiple virtual machines, at the same time.
- Oracle VM VirtualBox is deceptively simple yet also very powerful.
- Can run everywhere from small embedded systems or desktop class machines all the way up to data center deployments and even Cloud environments.

Microsoft Hyper-V

Hyper-V is a primary engine that drives Windows Server 2008 “beyond virtualization” initiative. The primary responsibility of Windows Server 2008 Hyper-V is to provide the tool kit that organizations will use to create a shared pool of compute, network, and storage resources where servers and applications can be virtualized for consolidation, scalability, and mobility purposes.

Microsoft Hyper-V Architecture

Hyper-V supportshypervisor-based virtualization platform and an enabling technology for one of Windows Server 2008 R2’s marquee features, Live Migration. The guest OSs running in a Hyper-V VM provide performance approaching the performance of an OS running on physical hardware if the necessary virtual server client (VSC) drivers and services are installed on the guest OS.

Hyper-V virtual server client (VSC) code, is also known as Hyper-V enlightened I/O, enables direct access to the Hyper-V “Virtual Machine Bus” and is available with the installation of Hyper-V integration services. Both Windows Server 2008 R2 and Windows 7 support Hyper-V enlightened I/O with Hyper-V integration services. Hyper-V supports isolation in terms of a partition. The Microsoft hypervisor must have at least one parent, or root, partition, running Windows Server 2008 R2. The root partition then creates the child partitions which host the guest OSs. The child partitions also do not have direct access to other hardware resources and are presented a virtual view of the resources, as virtual devices (VDevs).

Despite its straightforward installation as a component of the host OS, Hyper-V takes control of the hardware, and the host OS becomes a VM instance with special privileges, called the parent partition.

- 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 OSs, and creates child partitions through the hypervisor.
- Child partitions are used to host guest OSs 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.

Hypervisor is the component that directly manages the underlying hardware (processors and memory). It is logically defined by the following components:

Cloud Computing

Hyper calls Interface: This is the entry point for all the partitions for the execution of sensitive instructions. This is an implementation of the Para virtualization approach discussed with Xen. This interface is used by drivers in the partitioned OS to contact the Hypervisor using the standard Windows calling convention. The parent partition also uses this interface to create child partitions.

Memory Service Routines (MSRS): These are the set of functionalities that control the memory and its access from partitions. By leveraging hardware-assisted virtualization, the hypervisor uses the Input/Output Memory Management Unit (I/O MMU or IOMMU) to fast-track access to devices from partitions by translating virtual memory addresses.

Advanced Programmable Interrupt Controller (APIC): This component represents the interrupt controller, which manages signals coming from the underlying hardware when some event occurs (timer expired, I/O ready, exceptions and traps). Each virtual processor is equipped with a synthetic interrupt controller (SynIC), which constitutes an extension of the local APIC. The hypervisor is responsible of dispatching, when appropriate, the physical interrupts to the synthetic interrupt controllers.

Scheduler: This component schedules the virtual processors to run on available physical processors. The scheduling is controlled by policies that are set by the parent partition.

- Address manager. This component is used to manage the virtual network addresses that are allocated to each guest OS.

Partition Manager: This component is in charge of performing partition creation, finalization, destruction, enumeration, and configurations. Its services are available through the hypercalls interface API.

Microsoft Hyper-V Performance Characteristics

Improved Hardware Sharing Architecture: Hyper-V provides improved access and utilization of core resources, such as disk, networking, and video when running guest OSs with a hypervisor-aware kernel and which are equipped with requisite virtual server client (VSC) code (known as Hyper-V enlightened I/O).

Critical Disk Performance for I/O Intensive Applications: Disk performance is critical for disk I/O intensive enterprise applications such as Microsoft BizTalk Server and in addition to Hyper-V enlightened I/O; Hyper-V provides “Passthrough” disk support which provides disk performance on par with physical disk performance.

Processor Hardware-assisted Virtualization Support: Hyper-V takes full advantage of processor hardware-assisted virtualization support that is available with recent processor technology.

Multi-core (SMP) Guest OS Support: Hyper-V provides the ability to support up to four processors in a virtual machine environment, which allows applications to take full advantage of multi-threading functionality in a VM.

Both 32-bit and 64-bit Guest OS Support: Hyper-V provides broad support for simultaneously running different types of OSs, including 32-bit and 64-bit systems across different server platforms, such as Windows, Linux, and others.

Advantages of Microsoft Hyper-V Architecture

Consolidation of Hardware Resources: Multiple physical servers can be easily consolidated into comparatively fewer servers by implementing virtualization with Hyper-V. Consolidation accommodates full use of deployed hardware resources. Hyper-V in Windows Server 2008 R2 can now access up to 64 logical CPUs on host computers.

Ease of Administration:

- Consolidation and centralization of resources simplifies administration.
- Implementation of scale-up and scale out is accommodated with much greater ease.

Fault Tolerance Support through Hyper-V Clustering: Because Hyper-V is a cluster aware application, Windows Server 2008 SP2 provides native host clustering support for virtual machines created in a Hyper-V virtualized environment.

Ease of deployment and management:

- Consolidation of existing servers into fewer physical servers simplifies deployment.

- A comprehensive Hyper-V management solution is available with System Center VMM.

Proven Track Record- Key Microsoft web sites MSDN (<http://msdn.microsoft.com>) and TechNet (<http://technet.microsoft.com>) are hosted in Hyper-V environments.

Comprehensive Product Support- Because Microsoft enterprise applications (such as Exchange Server and SQL Server) are fully tested running in Hyper-V, Microsoft provides code fix support for these applications when deployed and run in a Hyper-V environment.

Scalability- Additional processing power, network bandwidth, and storage capacity can be accomplished quickly and easily by apportioning additional available resources from the host computer to the guest VM(s).

Summary

- Virtualization raises abstraction. Abstraction pertains to hiding of the inner details from a particular user. Virtualization helps in enhancing or increasing the capability of abstraction.
- Virtualization enables sharing of resources much easily, it helps in increasing the degree of hardware level parallelism, basically, there is sharing of the same hardware unit among different kinds of independent units.
- In a bare metal architecture, one hypervisor or VMM is actually installed on the bare metal hardware. There is no intermediate OS existing over here. The VMM communicates directly with the system hardware and there is no need for relying on any host OS.
- Type I Hypervisors run directly on top of the hardware. Therefore, they take the place of the OSs 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 OSs.
- Type II Hypervisors require the support of an operating system to provide virtualization services. This means that they are programs managed by the OS, which interact with it through the ABI and emulate the ISA of virtual hardware for guest OSs.
- Xen is an open-source initiative implementing a virtualization platform based on paravirtualization. Xen is a VMM for IA-32 (x86, x86-64), IA-64 and PowerPC 970 architectures.
- KVM is part of existing Linux code, it immediately benefits from every new Linux feature, fix, and advancement without additional engineering. KVM converts Linux into a type-1 (bare-metal) hypervisor.
- VMware Workstation is the most dependable, high-performing, feature-rich virtualization platform for your Windows or Linux PC.

Keywords

- **Virtualization:** Virtualization is a broad concept that refers to the creation of a virtual version of something, whether hardware, a software environment, storage, or a network.
- **Type 0 Hypervisors-** Hardware-based solutions that provide support for virtual machine creation and management via firmware. Example: IBM LPARs and Oracle LDOMs are examples.
- **Type 1 Hypervisors-** Operating-system-like software built to provide virtualization. Example: Including VMware ESX, JoyentSmartOS, and Citrix XenServer. It also includes general-purpose operating systems that provide standard functions as well as VMM functions. Example: Microsoft Windows Server with HyperV and RedHat Linux with KVM.
- **Type 2 Hypervisors-** Applications that run on standard OSs but provide VMM features to guest OSs. Example: VMware Workstation and Fusion, Parallels Desktop, and Oracle VirtualBox.
- **Interpretation:** Interpretation involves relatively inefficient instruction-at-a-time.
- **Binary Translation:** Binary translation involves block-at-a-time optimization for repeated.

- **Para Virtualization-** Technique in which the guest operating system is modified to work in cooperation with the VMM to optimize performance.
- **Programming-environment Virtualization-** VMMs do not virtualize real hardware but instead create an optimized virtual system. It is used by Oracle Java and Microsoft.Net.
- **Emulators-** Emulators allow the applications written for one hardware environment to run on a very different hardware environment, such as a different type of CPU.

Self Assessment

1. _____ is software that creates a virtualized environment between the computer platform and the end-user in which the end user can operate software.
 - A. Virtual machine
 - B. FogSim
 - C. Firmware
 - D. Freeware

2. Which of the following is characteristic of virtualization concept?
 - A. Mapping of virtual resources or state to real resources
 - B. Use of real machine instructions to carry out actions specified by the virtual machine instructions
 - C. Implemented by adding a layer of software to a real machine to support the desired VM's architecture
 - D. All of the above

3. Process virtual machines are also known as _____ virtual machines.
 - A. Primary
 - B. Hype
 - C. Application
 - D. Sigma

4. _____ is the process of implementing the interface and functionality of one (sub)system on a (sub)system having a different interface and functionality.
 - A. Inheritance
 - B. Emulation
 - C. Multitenancy
 - D. Interfacing

5. Which of the following is not true about Interpretation?
 - A. Simple and easy to implement
 - B. High performance
 - C. Threaded interpretation
 - D. Portability

6. _____ is a form of binary recompilation where sequences of instructions are translated from a source instruction set to the target instruction set.
 - A. Compilation
 - B. Blocking

- C. Binary translation
 - D. Coordination
7. Which of the following is/are significant feature(s) of binary translation?
- A. Run-time translation of some OS instructions
 - B. User-level code is directly executed on the real hardware
 - C. Guest OS is not aware of virtualization
 - D. All of the above
8. Virtualization allows multiple virtual machines, with heterogeneous OSs to run in isolation, side-by-side on the _____ physical machine(s).
- A. Same
 - B. Different
 - C. One
 - D. Two
9. In virtualization, the virtual environment can be created with the help of a software tool called _____.
- A. Hypervisors
 - B. Emulators
 - C. Compilers
 - D. Interpreters
10. A _____ is a fundamental element of hardware virtualization.
- A. Bare-metal device
 - B. Virtual machine manager
 - C. Storage
 - D. Software stick
11. In a hypervisor, the _____ constitutes the entry point of the monitor and reroutes the instructions issued by the virtual machine instance to one of the two other modules.
- A. Virtual layer
 - B. Interpreter
 - C. Dispatcher
 - D. Translator
12. Which of the following property(s) specify the virtualization requirements?
- A. Efficiency property
 - B. Resource control property
 - C. Equivalence property
 - D. All of the above
13. Type II hypervisors require the support of an operating system to provide virtualization services. Such hypervisors are also called as
- A. Hybrid
 - B. Hosted
 - C. Bare-metal

D. Native

14. Which of the following is an example of hosted virtualization hypervisors?

- A. VMware Workstation
- B. Oracle Virtual Box
- C. Microsoft Virtual PC
- D. All the above

15. In a _____ virtualization, the VMM communicates directly with system hardware rather than relying on a host operating system.

- A. Bare-metal
- B. Hybrid
- C. Storage
- D. Hosted

Answers for Self Assessment

- | | | | | |
|-------|-------|-------|-------|-------|
| 1. A | 2. D | 3. C | 4. B | 5. B |
| 6. C | 7. D | 8. A | 9. A | 10. B |
| 11. C | 12. D | 13. B | 14. D | 15. A |

Review Questions

1. Differentiate Native and Hosted virtualization?
2. Explain the concept of Hosted virtualization?
3. Write a short note on:
 - (a) Xen
 - (b) KVM
 - (c) Hyper-V
4. What are the benefits of using Microsoft Hyper-V?
5. Elaborate the working on VMware?
6. Discuss about Interpretation and Binary Translation?
7. What is Virtualization and why it is required?
8. How does Virtual Machines work? Discuss the characteristics.
9. Give an overview of Virtual Machine Taxonomy.
10. Compare the Process VM with System VM?



Further Readings

- Mastering Cloud Computing by "Rajkumar Buyya, Christian Vecchiola, S Thamarai Selvi, Tata McGraw-Hill Education, 2021.
- Cloud Computing: Concepts, Technology and Architecture by Thomas Erl, Pearson Education.
- Cloud Computing Black Book by Kailash Jayaswal, Jagannath Kallakurchi, Donald J. Houde, Deven Shah, Kogent Learning Solutions, DreamTech Press.

Unit 11: Virtual Machine

- Virtualization: A Manager's Guide by Dan Kusnetzky, O'Reilly, 2011.
- Virtual Machines, by James Edward Smith, Elsevier, 2021.

**Web Links**

- 1.What is Virtualization? (tutorialspoint.com)
- 2.What is Virtualization? How does it Work? [Understanding Virtualization] (kuberty.io)
- 3.What is a Virtual Machine? | VMware Glossary
- 4.What Is a Virtual Machine and How Does It Work | Microsoft Azure
5. Virtual Machine - Javatpoint
- 6.Types of Virtual Machines - GeeksforGeeks
- 7.Home - Xen Project
- 8.What Is Hyper-V & How Do You Use It? A Beginner's Guide (cloudwards.net)
<https://youtu.be/d7J9p2uHkEU>

Unit 12: Security and standards in Cloud

CONTENTS

- Objectives
- Introduction
- 12.1 Cloud Computing Security
- 12.2 Security Challenges in Cloud Computing
- 12.3 Software-as-a-Service Security
- 12.4 Standards in Cloud Computing
- 12.5 Standards for Application Developers
- 12.6 Standards for Messaging
- 12.7 Standards for Security
- Summary
- Keywords
- Self Assessment
- Answers for Self Assessment
- Review Questions
- Further Readings

Objectives

After this lecture, you will be able to,

- Know about security in clouds and Software as a Service Security.
- Explore different security challenges in clouds.
- Know about the Open Cloud Consortium and the Distributed Management Task Force.
- Understand the cloud standards with detailed exploration of the standards for application developers, Standards for messaging, and the Standards for security.

Introduction

The term computer security is used frequently, but the content of a computer is vulnerable to few risks unless the computer is connected to other computers on a network. As the use of computer networks, especially the Internet, has become pervasive, the concept of computer security has expanded to denote issues pertaining to the networked use of computers and their resources. Threat is defined as a computer program, a person, or an event that violates the security system. A threat causes loss of data and attacks the data privacy. Security is the protection of assets. The three main aspects are:

- prevention
- detection
- re-action

Security is actually related to protection of assets, assets pertaining to something that we have, like our computer, our laptop, a mobile, all these are assets. The main aspects related to security is prevention of how we can prevent the loss of our particular asset, then detection, if we have lost it, how we can detect it, then the reaction that is pertaining to how we will react to a particular security breach, if it takes place.

Traditional Security v/s Computer Security

Cloud Computing

In case of traditional security mechanisms, they were excellent chances that the information can be stolen, but you still have the confidential information, it may get copied or it may get be sold, but there can be situations with the theft does not get detected. So, criminals may be on the other side of the world that is one important aspect related to the computer security whereas if we talk about computers versus physical security computer security is more concerned with the loss of data and information. If we talk about the physical security of a particular computer system, even our mobile phones or our smartphones or any other laptop devices are physical things. The computer security it means to protect the information. It deals with the prevention, detection and offers any unauthorized actions by the users of a particular computer. The computer security has been extended to include privacy, confidentiality and integrity.

Computer security is the protection afforded to an automated information system in order to attain the applicable objectives of preserving the integrity, availability and confidentiality of information system resources (includes hardware, software, firmware, information/data, and telecommunications). The major focus of security lies on:

- **Application-level Security-** Determines who can access data and what tasks they are allowed to perform.
- **Transport-level Security-** Transport-level mostly determines the security of data as it is transmitted.

Security Triad

The major technical areas of computer security are usually represented by the initials CIA: confidentiality, integrity, and authentication or availability. Confidentiality means that information cannot be accessed by unauthorized parties. Confidentiality is also known as secrecy or privacy; breaches of confidentiality range from the embarrassing to the disastrous. Integrity means that information is protected against unauthorized changes that are not detectable to authorized users; many incidents of hacking compromise the integrity of databases and other resources. Authentication means that users are who they claim to be. Availability means that resources are accessible by authorized parties; “denial of service” attacks, which are sometimes the topic of national news, are attacks against availability.

The other important concerns of computer security professionals are access control and non-repudiation. Maintaining access control means not only that users can access only those resources and services to which they are entitled, but also that they are not denied resources that they legitimately can expect to access. Non-repudiation implies that a person who sends a message cannot deny that he sent it and, conversely, that a person who has received a message cannot deny that he received it.

Confidentiality: Confidentiality is further subdivided into two main aspects one is a data confidentiality and second is privacy of data, so data confidentiality, it shows that confidential information it is not disclosed to any unauthorized individual. If we talk about privacy, it shows that individual control or influence what information may be collected and what information can be stored. Confidentiality pertains to keeping the information secret or private. It is very important aspect. For example, suppose we have the student grade that is an important information thereby confidentiality is considered as very high. It says that the grades should only be available to the students or their parents. The student enrollment information has certain moderate confidentiality rating. If it gets disclosed, there is lesser damage and then that can be caused.

Integrity: Integrity is further divided into two main parts one is data integrity second is system integrity. Now if I talk about data integrity, it shows that information and the programs, they are changed only in a specified and authorized manner. System integrity means that the system performs its operations in an unimpaired manner. Also, there is no hindrance in performing its actions in case of system integrity. Integrity prevents any unauthorized writing modification or manipulation of information. It means there is an external consistency that is existing in a system, everything is as it is, as it is expected to be like, if I transfer data from one source to a destination then in between that transport-level security is being maintained that means that data does not get manipulated, it does not get modified in between that transition or transport. Data integrity means that data stored on a computer is same as that on the source document. Example: Suppose in healthcare sector, if we talk about a particular hospital it is maintaining the patient's record. The patient's allergy information is a very highly integrated data like doctors should be able to trust that the information is correct. Considering a real-world scenario like you go to a hospital, you're

Unit 12: Security and standards in Cloud

working in a hospital, or you have a friend who is a nurse in that hospital. By mistake the nurse has written some data related to a particular patient that she did not mention that this patient is diabetic into the patient's record, but being a diabetic, is considered as the most important aspect in a patient's treatment aspect.

A doctor comes up give certain medicines to the patient, unaware of the thing that the patient is diabetic, so that can cause certain, you know blunder in that case. A missed information or wrong information by the nurse impacts the doctor's treatment. The patient might get certain treatment that would actually exaggerate his diabetic levels. It is a blunder on the aspect of that particular nurse that she did not mention, or maybe one has mentioned but by default that page gets deleted out or that record gets deleted out so in that case the integrity related to a patient's record is extremely important.

What if a particular nurse deliberately falsifies the data, the database should be restored to a trusted basis and falsified information should be traced back to the person who actually did it. Even in online news group registration data, it is of moderate-level, it does not require very much high-level of integrity security.

Availability: Availability assures that the systems work promptly, and the service is not denied to any unauthorized users. The information should be accessible, as in when it is required, it should be reasonable, or usable upon appropriate demand by an app or by an authorized user. The availability pertains to prevention of any unauthorized from withholding of information. There are certain attacks that are related to availability like Denial of Service (DoS) attack. In such attacks, the service might get unavailable for the authorized users. Example: If a system provides authentication, it requires higher availability if the customers they can are not able to access their resources. The loss of services could actually result in certain financial loss. The access to a public website of a particular university is an example of moderate availability scenario. If I say a particular online telephone directory lookup is not available or is unavailable for certain reasons. This can cause certain amount of annoyance among the people but this is an example of low-availability scenario.

Other Concepts Related to Security

Authenticity: Authenticity is the property of being genuine and being able to be verified and trusted; confident in the validity of a transmission, or a message, or its originator. It may be obtained by the provision of a password or a scan of your retina.

Accountability: Accountability generates the requirement for actions of an entity to be traced uniquely to that individual to support non-repudiation, deference, fault isolation etc. The system managers are accountable to scrutiny. It can involve audit trails that must be selectively kept protected so that actions affecting security can be traced back to the responsible party. Logs are also used for keeping the account of security and events.

Need for Security

Administrators normally find that putting together a security policy that restricts both users and attacks is time consuming and costly. Users also become disgruntled at the heavy security policies making their work difficult for no discernable reason, causing bad politics within the company. Planning an audit policy on huge networks takes up both server resources and time, and often administrators take no note of the audited events. A common attitude among users is that if no secret work is being performed, why bother implementing security.

There is a price to pay when a half-hearted security plan is put into action. It can result in unexpected disaster. A password policy that allows users to use blank or weak passwords is a hacker's paradise. No firewall or proxy protection between the organization's private local area network (LAN) and the public Internet makes the company a target for cybercrime.

Organizations need to determine the price they are willing to pay in order to protect data and other assets. This cost must be weighed against the costs of losing information and hardware and disrupting services. The idea is to find the correct balance. If the data needs minimal protection and the loss of that data is not going to cost the company, then the cost of protecting that data will be less. If the data is sensitive and needs maximum protection, then the opposite is normally true.

- Cyber-Crime Rise Issues: No firewall or proxy protection between organization's private LAN & public Internet makes a company a target for cyber-crime.

Cloud Computing

- Security Loss Costs are High: Organizations will need to determine the price they are willing to pay in order to protect data and other assets. This cost must be weighed against the costs of losing information and hardware and disrupting services.
- Security Policy: Administrators normally find that putting together a security policy that restricts both users and attacks is time consuming and costly.
- Audit Policy: Planning an audit policy on huge networks takes up both server resources and time, and often administrators take no note of the audited events. The lack of security can cause unexpected disasters.
- Password policy: Allows users to use blank or weak passwords is a hacker's paradise.

12.1 Cloud Computing Security

Cloud computing security is the set of control-based technologies and policies designed to adhere to regulatory compliance rules and protect information, data applications and infrastructure associated with cloud computing use. It refers to a broad set of policies, technologies, applications, and controls utilized to protect virtualized IP, data, applications, services, and the associated infrastructure of cloud computing. It can be adjudged as a sub-domain of computer security, network security, and, more broadly, information security.

The security concerns associated with cloud computing are typically categorized in two ways: as security issues faced by cloud providers (organizations providing software-, platform-, or infrastructure-as-a-service via the cloud) and security issues faced by their customers (companies or organizations who host applications or store data on the cloud). However, the responsibility is shared, however, and is often detailed in a cloud provider's "shared security responsibility model" or "shared responsibility model". The provider must ensure that their infrastructure is secure and that their clients' data and applications are protected, while the user must take measures to fortify their application and use strong passwords and authentication measures.

When an organization elects to store data or host applications on the public cloud, it loses its ability to have physical access to the servers hosting its information. As a result, potentially sensitive data is at risk from insider attacks. According to a 2010 Cloud Security Alliance report, insider attacks are one of the top seven biggest threats in cloud computing. Therefore, cloud service providers must ensure that thorough background checks are conducted for employees who have physical access to the servers in the data center. Additionally, data centers are recommended to be frequently monitored for suspicious activity. In order to conserve resources, cut costs, and maintain efficiency, cloud service providers often store more than one customer's data on the same server. As a result, there is a chance that one user's private data can be viewed by other users (possibly even competitors). In order to handle such sensitive situations, cloud service providers should ensure proper data isolation and logical storage segregation.

The extensive use of virtualization in implementing cloud infrastructure brings unique security concerns for customers or tenants of a public cloud service. Virtualization alters the relationship between the OS and underlying hardware – be it computing, storage or even networking. This introduces an additional layer – virtualization – that itself must be properly configured, managed and secured.

Cloud Security Controls

Cloud security architecture is effective only if the correct defensive implementations are in place. Efficient cloud security architecture should recognize the issues that will arise with security management. The security management addresses these issues with security controls. These controls are put in place to safeguard any weaknesses in the system and reduce the effect of an attack. While there are many types of controls behind a cloud security architecture, they can usually be found in one of the following categories:

Deterrent controls- These controls are intended to reduce attacks on a cloud system. Much like a warning sign on a fence or a property, deterrent controls typically reduce the threat level by informing potential attackers that there will be adverse consequences for them if they proceed. (Some consider them a subset of preventive controls.)

Preventive controls- Preventive controls strengthen the system against incidents, generally by reducing if not actually eliminating vulnerabilities. Strong authentication of cloud users, for

Unit 12: Security and standards in Cloud

instance, makes it less likely that unauthorized users can access cloud systems, and more likely that cloud users are positively identified.

Detective controls- Detective controls are intended to detect and react appropriately to any incidents that occur. In the event of an attack, a detective control will signal the preventative or corrective controls to address the issue. System and network security monitoring, including intrusion detection and prevention arrangements, are typically employed to detect attacks on cloud systems and the supporting communications infrastructure.

Corrective controls- Corrective controls reduce the consequences of an incident, normally by limiting the damage. They come into effect during or after an incident.

 **Notes:** Restoring system backups in order to rebuild a compromised system is an example of a corrective control.

Security Boundaries in Cloud

A particular service model defines the boundary between the responsibilities of service provider and customer. In cloud, there is a Cloud Security Alliance (CSA) stack model. The CSA stack model defines the boundaries between each service model and shows how different functional units relate to each other.

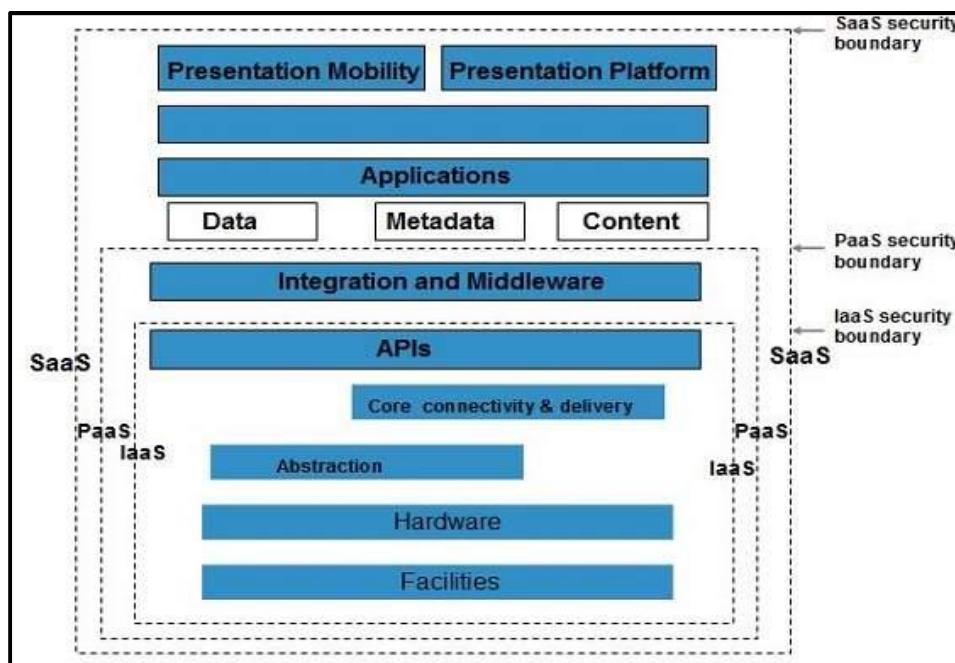


Figure 1: Cloud Security Alliance (CSA) Stack Model

Key Points to CSA Model

- IaaS is the most basic level of service with PaaS and SaaS next two above levels of services.
- Moving upwards, each of the service inherits capabilities and security concerns of the model beneath.
- IaaS provides the infrastructure, PaaS provides platform development environment, and SaaS provides operating environment.
- IaaS has the least level of integrated functionalities and integrated security while SaaS has the most.
- This model describes the security boundaries at which cloud service provider's responsibilities end and the customer's responsibilities begin.
- Any security mechanism below the security boundary must be built into the system and should be maintained by the customer.

*Cloud Computing***12.2 Security Challenges in Cloud Computing**

Number of security challenges associated with cloud computing that must be adequately addressed.

Managing Complex Environments: Today's modern enterprise IT infrastructure is usually a mix of multi-cloud or hybrid environments scattered around a few countries, especially in a global setting. Managing everything across the entire environment is not easy, especially when it comes to security. This requires tools that can handle all these different environments safely, even spanning multiple continents in some cases. Many things can go wrong in such a setting, so it's crucial that you use tools from a provider that allows you to do all this safely.

Just a single misconfigured setting somewhere can cause a malicious actor to breach your defences. And they only have to breach it once. Your defense needs to be up 24/7, as every single error has the potential to be exploited. No pressure!

Compliance with Rules and Regulations: In the US, there's HIPAA or The Health Insurance Portability and Accountability Act. In California, there's the CCPA or California Consumer Privacy Act. In the EU, there's the GDPR or General Data Protection Regulation. Depending on the business itself, where you do business and, in addition to that, on the data you process, these are laws you have to comply with. Laws like these often make it mandatory to disclose data breaches and hold you responsible for storing data safely. It's therefore good to pick tools that allow you to gain absolute control of your environment, whether it's a single public cloud, single private cloud, multi-cloud or hybrid cloud environment.

Lack of Visibility: We touched upon the immense task of keeping cloud environment secure earlier on. And it's not as simple as just creating a less complex environment. The larger and older the enterprise or organisation, the more you have to work with or around legacy software or a particular way of doing a specific process that doesn't translate 1:1 with the new cloud tools you're currently using for that process. The environment can quickly become so overwhelmingly complex that you lose sight of both the big picture and the detailed insights. That's where it all goes downhill because this lack of oversight and insight means you can't keep everything up-to-date and don't know what is and isn't misconfigured. To regain control, you need to have detailed insight into every asset within your cloud environment.

Data Center or Physical Security Issues: When service providers build data centers, they must have various aspects in mind; the geographical location and the environment thereof, political and governmental restrictions and rules, availability of power and other such requirements, possibilities of energy-savings and so on. All these would help the provider to keep a fault-tolerant cloud available all the time reaching a goal of 100% uptime. Data centers qualities are categorized in tiers with the lowest level being 1 and the highest level being 4. A highly reliable and available data center with special cooling techniques is assigned to tier 4 level. When so much goes in a data center, the physical security of such facilities also is a great concern and attracts attention. Data centers must have appropriate security mechanisms equipped with video cameras and other techniques to prevent any physical violations.

Other Security Challenges

Let us discuss the major security challenges in detail:

Security Challenge #1: Organizations are afraid of losing their data stored in the cloud

Business leaders confirmed that when moving to the cloud, their biggest security concern is the loss and leakage of their cloud data, especially if we are talking about business-critical information. The truth is that the data stored in the cloud can be lost or leaked not only by malicious tampering but also by accident or when an act of nature brings down a cloud service provider. Regardless of the situation, the impact will be disastrous for the business.

Security Challenge #2: Threats to data privacy put cloud computing at risk

Generally speaking, enterprise-grade cloud services are more secure than legacy architecture but with hackers getting more experienced in breaching security parameters within the cloud, the risk of a data breach is higher than ever. When moving to the cloud, businesses have to relinquish some degree of control to a third-party provider. This means that in the eventuality of a data breach, the business has expertise if critical business systems are offline or a breach is being managed.

Security Challenge #3: Breaches of confidentiality challenge the integrity of cloud computing

Unit 12: Security and standards in Cloud

Businesses are extremely concerned when storing sensitive data in the cloud, especially business-critical information. Usually, the data sent to the cloud is not stored in a single cloud storage server but in fact, they are replicated to different cloud data centers located in different geographic areas. This means that the data is at risk of being hacked from any of those cloud data centers and this is the reason why businesses are so much concerned about maintaining the confidentiality of their data.

Security Challenge #4: Changing Service Provider

Vendor lock In is also an important Security issue in Cloud Computing. Many organizations will face different problems while shifting from one vendor to another. For example, An Organization wants to shift from AWS Cloud to Google Cloud Services then they ace various problem's like shifting of all data, also both cloud services have different techniques and functions, so they also face problems regarding that. Also, it may be possible that the charges of AWS are different from Google Cloud, etc.

Security Challenge #5: Lack of Skill

While working, shifting o another service provider, need an extra feature, how to use a feature, etc. are the main problems caused in IT company who doesn't have skilled Employee. So it requires a skilled person to work with cloud computing.

Security Challenge #6: Data Leakage or Loss

Data Loss is one of the issues faced in Cloud Computing. This is also known as Data Leakage. As we know that our sensitive data is in the hands of Somebody else, and we don't have full control over our database. So, if the security of cloud service is to break by hackers then it may be possible that hackers will get access to our sensitive data or personal files.

Security Challenge #7: Lack of Control

One of the biggest benefits of using cloud-based technologies is that the customer doesn't have to manage the resources needed to keep it working (such as servers). However, handing off the responsibility for managing the day-to-day maintenance of a software, platform, or computing asset can result in having less visibility and control over that asset.



Notes: Why is this one of the more important cloud security challenges that organizations need to address?

Because, it affects the ability of the organization to:

- Verify the efficacy of their security controls (because there's no visibility into the tools and data on the cloud platform);
- Enact incident response plans (since they may not have complete control over cloud-based assets); and
- Analyze information about their data, services, and users (which is often necessary to recognize abnormal use patterns inherent to a security breach).

12.3 Software-as-a-Service Security

Cloud computing models of the future will likely combine the use of SaaS (and other XaaS's as appropriate), utility computing, and Web 2.0 collaboration technologies to leverage the Internet to satisfy their customers' needs. SaaS will likely remain the dominant cloud service model for the foreseeable future and the area where the most critical need for security practices and oversight will reside. With a managed service provider, corporations or end users will need to research vendors' policies on data security before using vendor services to avoid losing or not being able to access their data.

Security Issues

The Technology analyst & consulting firm Gartner lists seven security issues which need to be discussed with a cloud vendor:

- Privileged User Access: Inquire about who has specialized access to data, and about the hiring and management of such administrators.

Cloud Computing

- Regulatory Compliance: Make sure that the vendor is willing to undergo external audits and/or security certifications.
- Data Location: Does the provider allow for any control over the location of data?
- Data Segregation: Make sure that encryption is available at all stages, and that these encryption schemes were designed and tested by experienced professionals.
- Recovery: Find out what will happen to data in the case of a disaster. Do they offer complete restoration? If so, how long would that take?
- Investigative Support: Does the vendor have the ability to investigate any inappropriate or illegal activity?
- Long-term Viability: What will happen to data if the company goes out of business? How will data be returned, and in what format?

Some Security Practices

SaaS providers need to incorporate and enhance security practices used by the managed service providers and develop new ones as the cloud computing environment evolves.

Security Management and Governance: One of the most important actions for a security team is to develop a formal charter for the security organization and program. These foster a shared vision among the team of what security leadership is driving toward and expects, and will also foster “ownership” in the success of the collective team. The security steering committee should be developed whose objective is to focus on providing guidance about security initiatives and alignment with business and IT strategies.

Risk Management and Risk Assessment: The effective risk management entails identification of technology assets; identification of data and its links to business processes, applications, and data stores; and assignment of ownership and custodial responsibilities. A formal information security risk management process should proactively assess information security risks as well as plan and manage them on a periodic or as-needed basis.

Security risk assessment is critical to helping the information security organization make informed decisions when balancing the dueling priorities of business utility and protection of assets. The formal risk assessment process should be created that allocates security resources linked to business continuity.

Security Portfolio Management: Security portfolio management is a fundamental component of ensuring efficient and effective operation of any information security program and organization. The lack of portfolio and project management discipline can lead to projects never being completed; unsustainable and unrealistic workloads and expectations because projects are not prioritized according to strategy, goals, and resource capacity; and degradation of the system or processes due to the lack of supporting maintenance and sustaining organization planning. The portfolio and project management capabilities can be enhanced by developing methodology, tools, and processes to support the expected complexity of projects that include both traditional business practices and cloud computing practices.

Secure Software Development Life Cycle (SecSDLC): SecSDLC involves identifying specific threats and the risks, followed by design and implementation of specific controls to counter those threats and assist in managing the risks they pose to the organization and/or its customers. It consists of six phases, and there are steps unique to the SecSDLC in each of phases:

- Phase-1 Investigation: Define project processes and goals, and document them in the program security policy.
- Phase-2 Analysis: Analyse existing security policies and programs, to analyse current threats and controls, examine legal issues, and perform risk analysis.
- Phase-3 Logical design: Develop a security blueprint, plan incident response actions, plan business responses to disaster, and determine the feasibility of continuing and/or outsourcing the project.

Unit 12: Security and standards in Cloud

- Phase-4 Physical design: Select technologies to support the security blueprint, develop a definition of a successful solution, and design physical security measures to support technological solutions, and review and approve plans.
- Phase-5 Implementation: Buy or develop security solutions. At the end of this phase, present a tested package to management for approval.
- Phase-6 Maintenance: Constantly monitor, test, modify, update, and repair to respond to changing threats.

Physical Security of Data Centres: Data centres must deliver multilevel physical security because mission-critical Internet operations require the highest level of security. The key components of data centre physical security are the following:

- Physical access control and monitoring: The data centres should be protected from unauthorized access and for monitoring various security equipment need to be installed.
- Policies, processes, and procedures: As with information security, policies, processes, and procedures are critical elements of successful physical security that can protect the equipment and data housed in the hosting centre.

Environmental controls and backup power: Heat, temperature, air flow, and humidity should all be kept within optimum ranges for the computer equipment housed on-site. Alternate power sources should be available.

12.4 Standards in Cloud Computing

In Internet circles, everything eventually gets driven by a working group of one sort or another. The working group is an assembled, cooperative collaboration of researchers working on new research activities that would be difficult for any one member to develop alone. The working group can exist for anywhere between a few months and many years. The working groups generally strive to create an informational document a standard, or find some resolution for problems related to a system or network. These are sometimes also referred to as task groups or technical advisory groups.

The Open Cloud Consortium (OCC)

The Open Cloud Consortium supports the development of standards for cloud computing and frameworks for interoperating between clouds; develops benchmarks for cloud computing; and supports reference implementations for cloud computing, preferably open-source reference implementations (Figure 2). The OCC manages a testing platform and a test-bed for cloud computing called the Open Cloud Test-bed. The group also sponsors workshops and other events related to cloud computing.

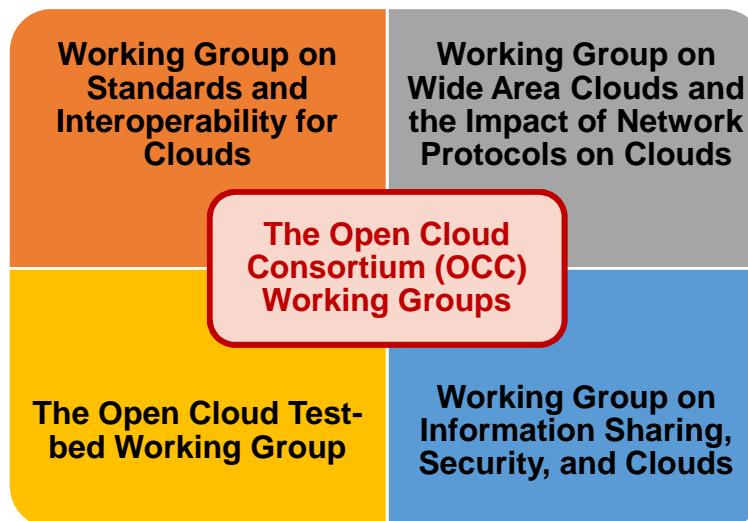


Figure 2: Open Cloud Consortium Working Groups

Cloud Computing

The OCC is organized into several different working groups-

Working Group on Standards and Interoperability for Clouds: Provide On-Demand Computing Capacity focuses on developing standards for interoperating clouds that provide on-demand computing capacity. One architecture for clouds that was popularized by a series of Google technical reports describes a storage cloud providing a distributed file system, a compute cloud supporting MapReduce, and a data cloud supporting table services. The open source Hadoop system follows this architecture. These types of cloud architectures support the concept of on demand computing capacity.

Working Group on Wide Area Clouds and the Impact of Network Protocols on Clouds: Focus of this working group is on developing technology for wide area clouds, including creation of methodologies and benchmarks to be used for evaluating wide area clouds. This working group is tasked to study the applicability of variants of TCP (Transmission Control Protocol) and the use of other network protocols for clouds.

The Open Cloud Test-bed Working Group: It uses Cisco C-Wave and the UIC Teraflow Network for its network connections. C-Wave makes network resources available to researchers to conduct networking and applications research. Experimental and production networks exist side by side but are physically and operationally separate. Production networks support cutting-edge applications by providing users guaranteed levels of reliability, availability, and performance. At the same time, experimental networks enable the deployment and testing of new networking technologies, providing researchers national-scale testbeds without the limitations typically associated with production networks.

Working Group on Information Sharing, Security, and Clouds: Focus on standards and standards-based architectures for sharing information between clouds. This is especially true for clouds belonging to different organizations and subject to possibly different authorities and policies. This group is also concerned with security architectures for clouds.

Distributed Management Task Force

With more than 3,500 active participants, the Distributed Management Task Force, Inc. (DMTF) is a not-for-profit, vendor-neutral, collaborative body that is leading the development, adoption, and unification of management standards and initiatives for desktop, enterprise, and Internet environments (Figure 3). Distributed Management Task Force (DMTF) is involved in the development, adoption, and interoperability of management standards and initiatives for enterprise and Internet environments. The aim is to handle the exchange of management information in a platform-independent and technology-neutral way, streamlining integration and reducing costs by enabling end-to-end multi-vendor interoperability in management systems.

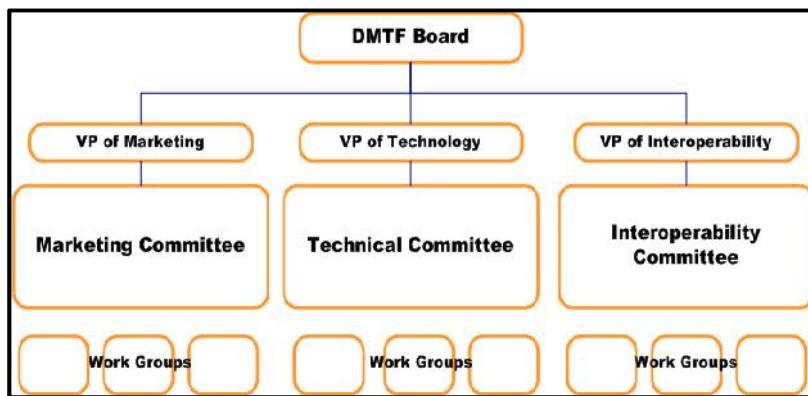


Figure 3: DMTF Organization

DMTF Board: The DMTF Board is responsible for the overall direction, strategy and activity of the DMTF, including managing DMTF finances, approving technical and marketing initiatives and leading DMTF committee work.

Technical Committee: The Technical Committee develops standards and initiatives for the DMTF and is responsible for coordinating all the technical activities of the DMTF including the Common Information Model (CIM), Web-Based Enterprise Management (WBEM), Management Profiles and

Unit 12: Security and standards in Cloud

Management Initiatives (i.e. SMASH, CDM). The Technical Committee oversees many working groups.

Interoperability Committee: The Interoperability Committee supplements the resources of the DMTF such that multi-vendor implementations of DMTF technologies can be compatible in the industry.

Marketing Committee: The Marketing Committee communicates with the industry, the public, and members about the activities of the organization.

DMTF Standards and Initiatives

DMTF is leading the development, adoption, and unification of management standards and initiatives for desktop, enterprise and Internet environments. Some of the approved specifications that the DMTF has made public.

- Common Information Model (CIM): CIM is a common data model of an implementation-neutral schema for describing overall management information in a network/enterprise environment.
- Web-Based Enterprise Management (WBEM): WBEM is a set of management and Internet standard technologies developed to unify the management of enterprise computing environments.
- Systems Management Architecture for Server Hardware (SMASH) Initiative: The DMTF SMASH Initiative is a suite of specifications that deliver architectural semantics, industry standard protocols and profiles to unify the systems management of the data centre.
- Alert Standard Format (ASF): The ASF specification defines remote control and alerting interfaces that best serve clients' OS-absent environments.
- System Management BIOS (SMBIOS): The SMBIOS specification addresses how motherboard and system vendors present management information about their products in a standard format by extending the BIOS interface on Intel architecture systems.
- Virtualization Management Initiative (VMAN): The VMAN unleashes the power of virtualization by delivering broadly supported interoperability and portability standards to virtual computing environments.
- Open Virtualization Format (OVF): It is a fairly new standard that has emerged within the VMAN Initiative. The OVF simplifies interoperability, security, and virtual machine life-cycle management by describing an open, secure, portable, efficient, and extensible format for the packaging and distribution of one or more virtual appliances.



Task: Conduct an in-depth exploration of DMTF initiatives taken since its establishment.

12.5 Standards for Application Developers

There are standards for application developers. The purpose of application development standards is to ensure uniform, consistent, high-quality software solutions.

Browsers(Ajax): Ajax (Asynchronous JavaScript and XML) is a group of interrelated web development techniques used to create interactive web applications or rich Internet applications. Using Ajax, web applications can retrieve data from the server asynchronously, without interfering with the display and behavior of the browser page currently being displayed to the user. The use of Ajax has led to an increase in interactive animation on web pages.

Data (XML, JSON)

XML- Extensible Markup Language (XML)

- Specification for creating custom markup languages.

Cloud Computing

- Its purpose is to enable sharing of structured data.
- An XML document has two correctness levels, well-formed and valid.
- A well-formed document conforms to the XML syntax rules.
- A valid document is well formed and additionally conforms to semantic rules which can be user-defined or exist in an XML schema.
- XML documents must conform to a variety of rules and naming conventions.
- XML provides a general, data model-oriented framework for the development of application-specific languages.

JavaScript Object Notation (JSON)

- JSON is a lightweight computer data interchange format. It is a text-based, human-readable format for representing simple data structures and associative arrays (called objects).
- JSON format is often used for transmitting structured data over a network connection in a process called serialization.
- Main application is in Ajax web application programming, where it serves as an alternative to the XML format.
- Based on a subset of the JavaScript programming language.
- Considered to be a language-independent data format.
- Code for parsing and generating JSON data is readily available for a large variety of programming languages.

Solution Stacks (LAMP and LAPP)

- LAMP
 - LAMP is a popular open source solution commonly used to run dynamic web sites and servers.
 - The acronym derives from the fact that it includes Linux, Apache, MySQL, and PHP (or Perl or Python).
 - It is considered by many to be the platform of choice for development and deployment of high-performance web applications which require a solid and reliable foundation.
 - When used in combination, they represent a solution stack of technologies that support application servers.
- Linux, Apache, PostgreSQL, and PHP(or Perl or Python)- LAPP:
 - LAPP stack is an open-source web platform that can be used to run dynamic web sites and servers.
 - LAPP offers SSL, PHP, Python, and Perl support for Apache2 and PostgreSQL.
 - There is an administration front-end for PostgreSQL as well as web-based administration modules for configuring Apache2 and PHP.
 - LAPP stack considered as a more secure out-of-the-box solution than the LAMP stack.

12.6 Standards for Messaging

A message is a unit of information that is moved from one place to another.

Simple Message Transfer Protocol (SMTP): SMTP is usually used for either sending a message from a workstation to a mail server or for communications between mail servers. It was designed so that sender and recipient information could be transmitted with the message. It is a two-way protocol that usually operates using TCP (Transmission Control Protocol) port 25.

Post Office Protocol (POP): POP is a lightweight protocol whose single purpose is to download messages from a server. This allows a server to store messages until a client connects and requests them. Once the client connects, POP servers begin to download the messages and subsequently

Unit 12: Security and standards in Cloud

delete them from the server (a default setting) in order to make room for more messages. The users respond to a message that was downloaded using SMTP. It is defined by RFC 1939 & usually functions on TCP port 110.

Internet Messaging Access Protocol (IMAP): IMAP allows messages to be kept on the server but viewed and manipulated (usually via a browser) as though they were stored locally. IMAP is a part of the RFC 2060 specification, and functions over TCP port 143.



Task: Explore in detail about SMTP, POP and IMAP services.

Syndication (Atom, Atom Publishing Protocol, and RSS)

The content syndication provides citizens convenient access to new content and headlines from government via RSS (Really Simple Syndication) and other online syndication standards. It has the following benefits:

- Ability to scan headlines from many sources, all in one place, through a newsreader.
- Time-saving awareness of new content from government, if the RSS feed or feeds are designed properly.
- Ability to monitor new content from across the council, as well as display feeds on their own web site.
- Awareness of new content position councilors as guides to government for citizens.
- Ability to aggregate new content or headlines from across multiple office locations and agencies.

Limitations:

- Dissemination via syndication is a new concept to governments just getting used to the idea of remote online public access to information.
- Governments need to accept that while they control the content of the feed, the actual display of the headlines and content will vary.
- Popular RSS feeds can use significant amounts of bandwidth.
- Automated syndication requires use of a content management system.
- Most viable content management systems have integrated RSS functions, but the sophistication, ease of use, and documentation of these tools vary.

RSS: RSS is a family of web feed formats used to publish frequently updated works- such as blog entries, news headlines, audio, and video in a standardized format. An RSS document includes full or summarized text, plus metadata such as publishing dates and authorship. RSS feeds can be read using software called a reader that can be web-based, desktop-based, a mobile device, or any computerized Internet-connected device. A standardized XML file format allows the information to be published once and viewed by many different programs. The user subscribes to a feed by entering the feed's URI into the reader or by clicking an RSS icon in a browser that initiates the subscription process. The RSS reader checks the user's subscribed feeds regularly for new work, downloads any updates that it finds, and provides a user interface to monitor and read the feeds.

Atom and Atom Publishing Protocol (APP): The name Atom applies to a pair of related standards. The Atom Syndication Format is an XML language used for web feeds, while the Atom Publishing Protocol (AtomPub or APP) is a simple HTTP-based protocol for creating and updating web resources, sometimes known as web feeds. The web feeds allow software programs to check for updates published on a web site. In order to provide a web feed, a site owner may use specialized software (such as a content management system) that publishes a list (or "feed") of recent articles or content in a standardized, machine-readable format. The feed can then be downloaded by web sites that syndicate content from the feed, or by feed reader programs that allow Internet users to subscribe to feeds and view their content. A feed contains entries, which

Cloud Computing

may be headlines, full-text articles, excerpts, summaries, and/or links to content on a web site, along with various metadata. The Atom format was developed as an alternative to RSS.

Web Services (REST)-Representational State Transfer: REST corresponds to the style of software architecture for distributed hypermedia systems such as the WWW. REST refers to a collection of network architecture principles which outline how resources are defined and addressed. An important concept in REST is the existence of resources, each of which is referenced with a global identifier (e.g., a URI in HTTP). REST provides improved response time and reduced server load due to its support for the caching of representations. REST improves server scalability by reducing the need to maintain session state. REST requires less client-side software to be written than other approaches, because a single browser can access any application and any resource.

SOAP (Simple Object Access Protocol): SOAP is a protocol specification for exchanging structured information in the implementation of Web Services in computer networks. SOAP can form the foundation layer of a web services protocol stack, providing a basic messaging framework on which web services can be built. The SOAP architecture consists of several layers of specifications for message format, message exchange patterns (MEPs), underlying transport protocol bindings, message processing models, and protocol extensibility. SOAP is platform-independent, language-independent, and it is simple and extensible.

Communications (HTTP, SIMPLE, and XMPP)

Hypertext Transfer Protocol (HTTP): HTTP is an application-level protocol for distributed, collaborative, hypermedia information systems. Its use for retrieving linked resources led to the establishment of the World Wide Web. HTTP is a request/response standard between a client and a server. A client is the end-user, the server is the web site.

SIMPLE (Session Initiation Protocol for Instant Messaging and Presence Leveraging Extensions): An instant messaging (IM) and presence protocol suite based on the Session Initiation Protocol (SIP). The core protocol methods provide SIP extensions for subscriptions, notifications, and publications. The methods used, subscribe and notify. The subscribe allows a user to subscribe to an event on a server. The notify is the method used whenever the event arises and the server responds back to the subscriber.

XMPP (Extensible Messaging and Presence Protocol): An XML-based protocol used for near-real-time, extensible instant messaging and presence information. XMPP remains the core protocol of the Jabber Instant Messaging and Presence technology. Jabber provides a carrier-grade, best-in-class presence and messaging platform.

12.7 Standards for Security

The standards for security define the processes, procedures, and practices necessary for implementing a security program. It also applies to cloud-related IT activities and include specific steps that should be taken to ensure a secure environment is maintained that provides privacy and security of confidential information in a cloud environment. These standards are based on a set of key principles intended to protect this type of trusted environment.

SAML: Security Assertion Markup Language (SAML): SAML is an XML-based standard for communicating authentication, authorization, and attribute information among online partners. It allows businesses to securely send assertions between partner organizations regarding the identity and entitlements of a principal. SAML protocol refers to what is transmitted, not how it is transmitted.

Open Authentication (OAuth): An open protocol to allow secure API authorization in a simple, standardized method for various types of web applications. The method for publishing and interacting with protected data. The core deals with fundamental aspects of the protocol, namely, to establish a mechanism for exchanging a user name and password for a token with defined rights and to provide tools to protect the token. With OAuth, sites use tokens coupled with shared secrets to access resources. Secrets, just like passwords, must be protected.

OpenID: OpenID is an open, decentralized standard for user authentication and access control that allows users to log onto many services using the same digital identity. It is a single-sign-on (SSO) method of access control. As such, it replaces the common log-in process (i.e., a log-in name and a

Unit 12: Security and standards in Cloud

password) by allowing users to log in once and gain access to resources across participating systems.

SSL/TLS: Transport Layer Security (TLS) and its predecessor, Secure Sockets Layer (SSL), are cryptographically secure protocols designed to provide security and data integrity for communications over TCP/IP. TLS and SSL encrypt the segments of network connections at the transport layer. The several versions of the protocols are in general use in web browsers, email, instant messaging, and voice-over-IP.

Summary

- Threat corresponds to any circumstance or event with the potential to harm an information system through unauthorized access, destruction, disclosure, modification of data, and/or denial of service. Threats arise from human actions and natural events.
- Computer security is the protection afforded to an automated information system in order to attain the applicable objectives of preserving the integrity, availability and confidentiality of information system resources (includes hardware, software, firmware, information/data, and telecommunications).
- The Open Cloud Consortium supports the development of standards for cloud computing and frameworks for interoperating between clouds; develops benchmarks for cloud computing; and supports reference implementations for cloud computing, preferably open-source reference implementations.
- DMTF is leading the development, adoption, and unification of management standards and initiatives for desktop, enterprise and Internet environments.
- The DMTF Board is responsible for the overall direction, strategy and activity of the DMTF, including managing DMTF finances, approving technical and marketing initiatives and leading DMTF committee work.
- SOAP is a protocol specification for exchanging structured information in the implementation of Web Services in computer networks.
- RSS is a family of web feed formats used to publish frequently updated works- such as blog entries, news headlines, audio, and video in a standardized format. An RSS document includes full or summarized text, plus metadata such as publishing dates and authorship.
- There are standards for application developers. The purpose of application development standards is to ensure uniform, consistent, high-quality software solutions.

Keywords

- **Authentication:** The process of verifying that users are who they claim to be when logging onto a system. Generally, the use of user names and passwords accomplishes this. More sophisticated is the use of smart cards and retina scanning. The process of authentication does not grant the user access rights to resources-this is achieved through the authorization process.
- **Availability:** The prevention of unauthorized withholding of information or resources. This does not apply just to personnel withholding information. Information should be as freely available as possible to authorized users.
- **Computer security:** Computer security means to protect information. It deals with the prevention and detection of unauthorized actions by users of a computer.

Cloud Computing

- **Confidentiality:** The prevention of unauthorized disclosure of information. This can be the result of poor security measures or information leaks by personnel. An example of poor security measures would be to allow anonymous access to sensitive information.
- **Integrity:** The prevention of erroneous modification of information. Authorized users are probably the biggest cause of errors and omissions and the alteration of data. Storing incorrect data within the system can be as bad as losing data. Malicious attackers also can modify, delete, or corrupt information that is vital to the correct operation of business functions.
- **Open Authentication (OAuth):** An open protocol to allow secure API authorization in a simple, standardized method for various types of web applications. The method for publishing and interacting with protected data.
- **OpenID:** OpenID is an open, decentralized standard for user authentication and access control that allows users to log onto many services using the same digital identity. It is a single-sign-on (SSO) method of access control.
- **Standards for Security:** The standards for security define the processes, procedures, and practices necessary for implementing a security program.

Self Assessment

1. Availability in the CIA security triad includes:
 - A. Assures that system works promptly and service is not denied to authorized users
 - B. Information should be accessible and useable upon appropriate demand by an authorized user.
 - C. Prevention of unauthorized withholding of information
 - D. All of the above
2. Which of the following property is not part of CIA security triad?
 - A. Locking
 - B. Integrity
 - C. Authentication
 - D. Confidentiality
3. _____ assures that confidential information is not disclosed to unauthorized individuals.
 - A. Authorizer
 - B. Data confidentiality
 - C. Integrity
 - D. Authenticity
4. _____ assures that a system performs its operations in unimpaired manner.
 - A. Data integrity
 - B. Data confidentiality
 - C. System integrity
 - D. Availability
5. The security concerns associated with cloud computing are typically categorized in certain ways that include(s):
 - A. Security issues faced by cloud providers
 - B. Security issues faced by their customers
 - C. Both cloud providers and customers

Unit 12: Security and standards in Cloud

- D. None of the above
6. The cloud security management involves certain security controls. Which of the following is a security control?
- Deterrent controls
 - Determinant controls
 - Creative control
 - Collective controls
7. _____ controls in cloud security strengthen the system against incidents, generally by reducing if not actually eliminating vulnerabilities.
- Preventive
 - Protective
 - Trust
 - Safety
8. Which of the following is/are a cloud security challenge?
- Managing complex environments
 - Compliance with rules and regulations
 - Data center or physical security issues
 - All of the above
9. Risk Management and Risk Assessment involve
- Formal risk assessment process
 - Identification of technology assets
 - Identification of data and its links to business processes, applications, and data stores
 - All of the above
10. SecSDLC stands for
- Secure Software Development Life Cycle
 - Secure Software Delivery Life Cycle
 - Secure Service Delivery Life Cycle
 - Secure Service Development Life Cycle
11. Working groups generally strive to create an informational document a standard, or find some resolution for problems related to a system or network. Working groups are also called as
- Tasklist groups
 - Technical advisory groups
 - Advising group
 - Resolution group
12. Open Cloud Consortium (OCC) offer(s) open-source reference implementations that
- manages a testing platform and a test-bed for cloud computing
 - sponsors workshops and other events related to cloud computing
 - develops benchmarks for cloud computing to support the development of standards for cloud computing and frameworks for interoperate between clouds
 - all of the above

13. Which of the following working group is tasked to study the applicability of variants of TCP and the use of other network protocols for clouds?
 - A. Working Group on Wide Area Clouds
 - B. Open Cloud Test-bed Working Group
 - C. Working Group on Standards and Interoperability
 - D. None of the above

14. DMTF stands for
 - A. Datacentres Maintenance Task Force
 - B. Datacentres Management Task Force
 - C. Distributed Management Task Force
 - D. Distributed Maintenance Task Force

15. The interoperability committee of DMTF performs the
 - A. Communication with the industry, the public, and members about the activities of the organization.
 - B. Supplementation of the resources of the DMTF such that multi-vendor implementations of DMTF technologies can be compatible in the industry.
 - C. Development of standards and initiatives for the DMTF
 - D. All of the above

Answers for Self Assessment

- | | | | | |
|-------|-------|-------|-------|-------|
| 1. D | 2. A | 3. B | 4. C | 5. C |
| 6. B | 7. A | 8. D | 9. D | 10. A |
| 11. B | 12. D | 13. A | 14. C | 15. D |

Review Questions

1. What are the different standards for messaging and communication?
2. Explore the standards for application developers.
3. The standards for security define the processes, procedures, and practices necessary for implementing a security program. Explain the various standards for security?
4. What is the functionality of Open Cloud Consortium?
5. Discuss the composition of Distributed Management Task Force?
6. Explain about security triad with respect to cloud security?
7. What are the different challenges in cloud security?
8. Describe Software-as-a-Security Service?
9. Explain:
 - (a) DMTF
 - (b) OCC



Further Readings

- Cloud Computing: Concepts, Technology and Architecture by Thomas Erl, Pearson Education.
- Cloud Computing Black Book by Kailash Jayaswal, Jagannath Kallakurchi, Donald J. Houde, Deven Shah, Kogent Learning Solutions, DreamTech Press.
- Cloud Computing Security: Foundations and Challenges by John. R. Vacca, CRC Press, 2016.
- Cloud Computing Security: Foundations and Challenges by Chris Dotson, O'Reilly, 2019.



Web Links

1. Cloud Computing Security (tutorialspoint.com)
2. Cloud Computing Security - GeeksforGeeks
3. Cloud Computing Security | How Cloud Computing Security Works? (educba.com)
4. What Is Cloud Computing Security? | McAfee
5. Home | DMTF
6. About DMTF | DMTF
7. Cloud Security: An Essential Guide | IBM
8. "The Open Cloud Consortium (OCC): supports the development of standards for cloud computing and..." - SmartData Collective
9. <https://youtu.be/jI8IKpjiCSM>
10. <https://youtu.be/uekNp9kEDY8>
11. <https://youtu.be/rD8kK1fz80U>

Unit 13: Application of Cloud Computing and Related Technologies**Dr. Tarandeep Kaur, Lovely Professional University****Unit 13: Application of Cloud Computing and Related Technologies****CONTENTS**

- Objectives
- Introduction
- 13.1 Why use EUC?
- 13.2 How Cloud Computing Can Empower End Users?
- 13.3 Reasons for Wide Adoption of Cloud Computing
- 13.4 Applications of Cloud Service in Various Areas of Life
- 13.5 Internet-of-Things (IoT)
- 13.6 Why Is Internet of Things (IoT) so Important?
- 13.7 History of IoT
- 13.8 Characteristics of IoT
- 13.9 Structure and Working of IoT
- 13.10 Working of IoT
- 13.11 Building Blocks of IoT
- 13.12 Applications of IoT
- 13.13 Cloud of Things (CoT)
- 13.14 Role of Cloud Computing in CoT
- 13.15 Architecture of Cloud of Things (CoT)
- 13.16 Mobile Internet of Things (Mobile IoT)
- 13.17 Edge Computing
- 13.18 Cloud Computing vs Edge Computing
- 13.19 Why Edge Computing?
- 13.20 Fog Computing
- 13.21 Blockchain Technology
- Summary
- Keywords
- Self Assessment
- Answers for Self Assessment
- Review Questions
- Further Readings

Objectives

After this lecture, you will be able to,

- know about end user access to cloud computing and End-user Computing.
- learn the different applications of cloud service in various areas of life.
- discover the computing technologies such as Internet of Things (IoT), Cloud of Things (CoT), and mobile IoT.

Cloud Computing

- explore the concepts of edge computing, fog computing, and blockchain technologies.
- analyze the role and applications of edge and fog computing.

Introduction

In today's world, many organizations have incorporated public cloud services into their workspace strategy. It is very common to see Salesforce used for CRM, Slack for collaboration, Google Drive for file sharing, and Okta for identity management. Example: Microsoft Office365 and Google Apps have reinvented the modern workspace. The organizations have transitioned from a design-build-support model to an as-a-service model. As-a-service has transformed everything from customer relationship management to user authentication. Cloud has impacted the way virtual desktops and applications are delivered to end-users.

Traditionally the end-user computing, which encompasses the management of these virtual desktops and apps, has been dominated by on-premise solutions like Virtual Desktop Infrastructure (VDI). The switch from CAPEX to OPEX procurement has allowed organizations to pay for usage over time, which lowers upfront costs. The security has dramatically improved, new features can be rolled out instantly and updates are made automatically. Also, by using a service which is operated by someone else, organizations are also able to reduce complexity and risk.

The End-User Computing (EUC) refers to computer systems and platforms that help non-programmers create applications. However, there is a lot more to EUC and its related technology, VDI, which essentially hosts desktop environments on a central server. It can be considered a form of desktop virtualization.

From Desktop as Software to Desktop-as-a-Service

The shift from desktop as software (VDI) to Desktop-as-a-Service (DaaS) is already underway. This is because more and more customers see the huge benefits of DaaS such as flexible usage model, unlimited scale, elastic design, and easy setup make the case for an expedited time to market.

With DaaS, the customers don't need to worry about running a virtual desktop platform on their own. They don't need to worry about people with the right skills or worry about the tight integration of EUC infrastructure components anymore. This is because the organization who developed the solution is responsible for running it.

The key to the success of an EUC/VDI initiative is bringing the components of the platform to users and managers without requiring systems expertise, allowing a more fluid introduction of new technologies to end users. Given the scope of roles and responsibilities, any platform that is used to support an initiative should be intuitive, agile, scalable, and manageable. These features reduce the level of expertise required to keep up with the changing digital workplace. A well-designed EUC/VDI plan can allow users to access the digital platforms they need to be productive, both on-premises and working remotely in the cloud.

13.1 Why use EUC?

EUC offers solutions to support mixed workload environments without the need for specialized knowledge of compute, network, and storage platforms with cloud computing support. Presently, the work environment has changed dramatically. Businesses require new ways of working to survive, remain successful and return to growth. Resiliency & business continuity are driving organizations to redesign & recalibrate operations. EUC/VDI is at forefront as application delivery to a remote workforce becomes critical for success.

Remote Offices- For organizations with a diverse physical footprint, it can be challenging to deliver consistent digital workplace resources. An EUC/VDI solution offers an opportunity to deliver improved consistency in the desktop and access to user applications in remote offices. It's easy to add new users and troubleshooting tasks can be minimized.

Remote Workers and BYOD Users- Supporting a remote workforce is now mandatory in almost every organization. This reality requires a more diverse set of digital platforms. An EUC/VDI environment is the first step toward a hybrid multi-cloud experience that can support this diversity without breaking organizational standards of operation. And an EUC/VDI solution can be designed that provides desktops and applications so that remote users can easily consume

Unit 13: Application of Cloud Computing and Related Technologies

them across multiple devices and can increase productivity while adhering to the organization's platform requirements.

Changing Requirements: At the same time, today's users have changed. As creators and consumers of content they demand access to increasingly rich data, images, video & sound in order to do their jobs, while some of more extreme users need to interact with images at high precision. Many organizations have all types of users, from task-based up to 3D graphical image user.

The ability to change and adapt to the new requirements defines an organization. To achieve the required level of desktop flexibility, agility and mobility often requires new cloud optimized infrastructures that can be scaled on demand with simple, automated, processes and tools.

Compliance and Licensing Control- For organizations that are concerned with regulatory compliance, or that want to improve management of their software licensing, an EUC/VDI environment can deliver a centralized platform for all user desktops and applications. With this centralized platform, security can be tightly managed and software licensing can be controlled. The EUC/VDI environment can be designed to provide everything that users need to perform their jobs while closely maintaining the range of external software installations by individual users.

Work smarter with End-user Computing Solutions- Bringing the digital workplace to life with innovative end user computing solutions that deliver the optimal user experience wherever your teams are located. Many end-user computing solutions and services are intuitive, agile and scalable, and consolidate desktops, applications and databases into a single platform. IT teams benefit from simplified management, the streamlined introduction of new technologies and lower costs.

The users can access the cloud computing using networked client devices, such as desktop computers, laptops, tablets and smartphones. Some of these devices - cloud clients- rely on cloud computing for all or a majority of their applications so as to be essentially useless without it. Examples are thin clients and the browser-based Chromebook.

Many cloud applications do not require specific software on the client and instead use a web browser to interact with the cloud application. With Ajax and HTML5 these web user interfaces can achieve a similar or even better look and feel as native applications. Some cloud applications, however, support specific client software dedicated to these applications (e.g., virtual desktop clients and most email clients). Some legacy applications (line of business applications that until now have been prevalent in thin client Windows computing) are delivered via a screen-sharing technology.

13.2 How Cloud Computing Can Empower End Users?

There are several applications that are successful SaaS offerings proving beneficial to the end users such as:

- YouTube, an online video repository, has an amazing hold on the global audience.
- Collaboration suites such as Zimbra both enhance mobility and allow you to maintain a virtual office in the cloud.
- Social networking with Facebook has become very popular, especially in academic settings.
- Zoho is a SaaS vendor to watch. Backed by Google, Zoho offers something for everyone.
- For presentations and online sharing, Dimdim is a good choice.

13.3 Reasons for Wide Adoption of Cloud Computing

Cloud computing is one of the strategic digital technologies considered important enablers for productivity and better services. The most substantial impact of cloud technology could come in the cost savings and increased competitiveness of IT services available to public and private organisations, as well as opportunities leading to new services, but cloud technology could positively increase other benefits for end users.

- Consumer-driven innovation
- Rise of power collaborators

Cloud Computing

- Changing economics
- Lowering of barriers to entry

Lower Computer Costs-A high-power computer is not needed to run cloud-based applications. Since applications run in the cloud, a desktop computer does not need the processing power or hard disk space demanded by traditional desktop software, or even a DVD drive.

Improved Performance-With fewer programs using the computer's memory at the same time, it will perform better. Desktop computers that use cloud-based services may boot and run faster because they have fewer programs and processes loaded into memory.

Reduced Software Costs-Instead of purchasing software applications, cloud computing applications can often be obtained for free. An example is the Google Docs suite for consumers.

Instant Software Updates-Updates to a cloud-based application generally occur automatically and are available on logging into the cloud. When accessing a web-based application, the latest version is usually instantly available, without need for an upgrade.

Improved Document Format Compatibility-All documents created by web-based applications can be read by any other user accessing that application. There are fewer format incompatibilities when all users are sharing documents and apps in the same cloud.

Unlimited Storage Capacity- Cloud computing can offer virtually limitless storage. A computer's current hard drive space is small compared to the storage space available in the cloud. Note however that large scale storage is generally not available for free, even in a cloud environment.

Increased Data Reliability- Unlike desktop computing, in which a hard disk crash can destroy personal data, a computer crashing in the cloud should not affect the storage of data, as typically cloud services provide multiple layers of security.

Universal document access- Documents stay in the cloud, and can be accessed from wherever with an Internet-capable device and an Internet connection. Documents are instantly available independent of location, removing the need to carry them when travelling.

Latest Version Availability- When editing a document from one location (example at home), that edited version is identical to the document accessed from another location (e.g. at work).

Easier Group Collaboration- Multiple users can collaborate easily on documents and projects. Because the documents are hosted in the cloud, not on individual computers, an Internet connection is all that is needed to collaborate.

Device Independence- When changing computers or moving to a portable device, existing applications are still available. There is no need for a special version of a program for a particular device, or to save a document in a device-specific format.

13.4 Applications of Cloud Service in Various Areas of Life

Cloud computing offers distributed; agile; transparent; seamless integration and delivery of services to the users and thus extends a wide horizon of opportunities for supporting highly pervasive technological progress and innovations. Moreover, by facilitating technological advancements that empower interactions across the globe, cloud computing generously proffers ample and awless delivery of large number of services and applications.

The capability to host large number of applications that can be delivered to the users speedily at minimal cost has augmented and attracted cloud computing as a favourable paradigm for the IT industry.

The cloud services and applications are distributed across multiple domains and are accessible to the users without any location confinements. Consequently, big IT giants such as Amazon, Google, Microsoft, Salesforce, Facebook have significantly adopted cloud computing as principal technological platform for their operation. Owing to the extensive implementation of virtualized servers, consolidation and useful service provisioning, cloud computing has unexpectedly revolutionized and dominated the business and service delivery patterns of the IT applications. The prominent cloud-based ventures include (Figure 1):

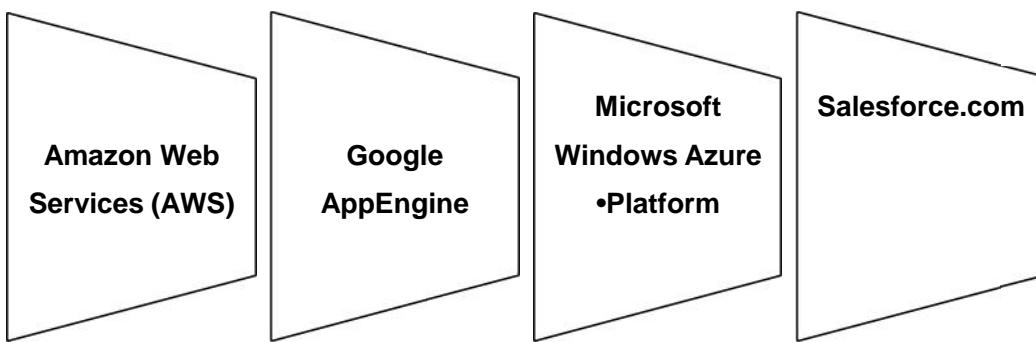
Unit 13: Application of Cloud Computing and Related Technologies

Figure 1: Major Cloud Services

• **Amazon Web Services (AWS):** AWS is a platform supporting design and development of highly flexible applications that can sustain scalability, storage; messaging across elastic infrastructure vested by Amazon. The prominent compute services include: Amazon Elastic Compute Cloud (EC2), Amazon Elastic MapReduce etc. Similarly, the storage services include: Amazon Simple Storage Service (S3), Amazon SimpleDB etc.

• **Google AppEngine:** IT bigwigs like Google largely offers PaaS through its venture, Google AppEngine. Google AppEngine provides services for development and hosting of web applications on the distributed, dynamic and scalable infrastructure of Google. The application developers can develop their applications in Java, Python and Go languages. It constantly monitors the application's usage of resources and services.

• **Microsoft Windows Azure Platform:** The Windows Azure platform is deployable as an appliance on third- party data centers and constitutes the infrastructure controlling the physical machines in the data center. The prime Windows Azure platform appliance products include Windows Azure, SQL Azure and Microsoft- assisted network; storage; and physical configurations.

• **Salesforce.com:** Salesforce is primarily a SaaS solution providing support for Customer Relationship Manager (CRM) applications. Certain cloud computing offerings and their prominent features are listed in Table 1 below:

Table 1: Cloud Services and their Descriptions

Product	Type Description
Amazon Web Services (AWS)	A collection of large number of web services facilitating computing, storage and other advanced services especially for IaaS services.
Google AppEngine	Supports development of simple and scalable web applications
Microsoft Azure	A kind of cloud operating system enabling development of scalable applications.
SalesForce.com	A SaaS solution for creating CRM applications vested on Force.com platform.
RightScale	A management platform for cloud providing scalable infrastructural support.

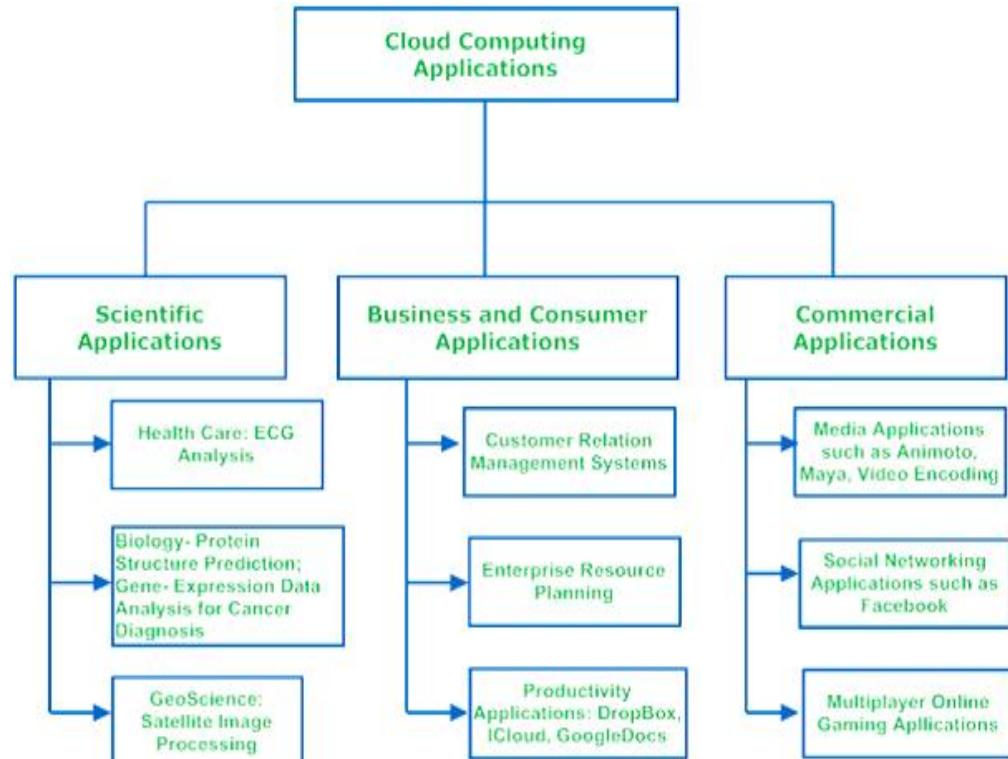


Figure 2: Cloud Computing Applications

Cloud computing can lucratively host pervasive applications ranging from scientific consumer and business domains driven by its capability to liberally map existing applications to it without much alterations. Accordingly, several cloud-specific applications ranging from scientific to business domains and consumer applications (Figure 2).

The scientific applications exploit the scalability and elasticity features of cloud environments that further facilitates higher degree of customization for deploying and executing scientific applications. Also, the economic benefits of cloud computing aid business and consumer applications in reducing or eliminating incurrence; operational; and maintenance costs for the infrastructure. The extended storage and easy accessibility offered by cloud makes it ideal for development of productive applications. Similarly, media applications and social networks benefit from continuous capacity extension and service reliability of cloud computing.

E-commerce Application: Cloud-based e-commerce allows responding quickly to the opportunities which are emerging. Users respond quickly to the market opportunities as well as the traditional e-commerce responds to the challenges quickly. Cloud-based e-commerce gives a new approach to doing business with the minimum amount as well as minimum time possible. Customer data, product data, and other operational systems are managed in cloud environments.

Cloud computing in Education: Cloud computing in the education sector brings an unbelievable change in learning by providing e-learning, online distance learning platforms, and student information portals to the students. It is a new trend in education that provides an attractive environment for learning, teaching, experimenting, etc to students, faculty members, and researchers. Everyone associated with the field can connect to the cloud of their organization and access data and information from there.

Technology-enhanced Learning, Education as a Service (EaaS): There are the following education applications offered by the cloud such as:

- **Google Apps for Education:** Google Apps for Education is the most widely used platform for free web-based email, calendar, documents, and collaborative study.
- **Chromebooks for Education:** Chromebook for Education is one of the most important Google's projects. It is designed for the purpose that it enhances education innovation.
- **Tablets with Google Play for Education:** It allows educators to quickly implement the latest technology solutions into the classroom and make it available to their students.

Unit 13: Application of Cloud Computing and Related Technologies

E-Governance Application: Cloud computing can provide its services to multiple activities conducted by the government. It can support the government to move from the traditional ways of management and service providers to an advanced way of everything by expanding the availability of the environment, making the environment more scalable and customized. It can help the government to reduce the unnecessary cost in managing, installing, and upgrading applications and doing all these with help of cloud computing and utilizing that money public service.

Cloud Computing in Healthcare: In the medical field also nowadays cloud computing is used for storing and accessing the data as it allows to store data and access it through the internet without worrying about any physical setup. It facilitates easier access and distribution of information among the various medical professional and the individual patients. Similarly, with help of cloud computing offsite buildings and treatment facilities like labs, doctors making emergency house calls and ambulances information, etc can be easily accessed and updated remotely instead of having to wait until they can access a hospital computer.



Example: Telehealth, Medi-Cloud, Technology-enhanced Healthcare

Entertainment Applications: Many people get entertainment from the internet, in that case, cloud computing is the perfect place for reaching to a varied consumer base. Therefore, different types of entertainment industries reach near the target audience by adopting a multi-cloud strategy.

The cloud-based entertainment provides various entertainment applications such as online music/video, online games and video conferencing, streaming services, etc and it can reach any device be it TV, mobile, set-top box, or any other form. It is a new form of entertainment called On-Demand Entertainment (ODE). The entertainment industries use a multi-cloud strategy to interact with the target audience. Cloud computing offers various entertainment applications such as online games and video conferencing.

Today, cloud gaming becomes one of the most important entertainment media. It offers various online games that run remotely from the cloud. The best cloud gaming services are Shaow, GeForce Now, Vortex, Project xCloud, and PlayStation Now.

The video Conferencing Apps provides a simple and instant connected experience that allows us to communicate with our business partners, friends, & relatives using a cloud-based video conferencing. The benefits of using video conferencing are that it reduces cost, increases efficiency, and removes interoperability. Example- GoToMeeting provides video conferencing and online meeting apps, which allows you to start a meeting with your business partners from anytime, anywhere using mobile phones or tablets. Using GoToMeeting app, you can perform the tasks related to the management such as join meetings in seconds, view presentations on the shared screen, get alerts for upcoming meetings, etc.

Art Applications: Cloud offers various art applications for quickly and easily design attractive cards, booklets, & images. Some most commonly used art applications are:

- Moo: One of the best cloud art applications. It is used for designing & printing business cards, postcards, & mini cards.
- Vistaprint: Vistaprint allows us to easily design various printed marketing products such as business cards, Postcards, Booklets, and wedding invitations cards.
- Adobe Creative Cloud: Adobe creative cloud is made for designers, artists, filmmakers, and other creative professionals. It is a suite of apps which includes PhotoShop image editing programming, Illustrator, InDesign, TypeKit, Dreamweaver.

Management Applications: Cloud computing offers various cloud management tools which help admins to manage all types of cloud activities, such as resource deployment, data integration, and disaster recovery. These management tools also provide administrative control over the platforms, applications, and infrastructure. Some important management applications are:

- Toggl: Toggl helps users to track allocated time period for a particular project.

Cloud Computing

- Evernote: Evernote allows you to sync and save your recorded notes, typed notes, and other notes in one convenient place. It is available for both free as well as a paid version. It uses platforms like Windows, macOS, Android, iOS, Browser, and Unix.
- Outright: Outright is used by management users for the purpose of accounts. It helps to track income, expenses, profits, and losses in real-time environment.

Social Applications: Social cloud applications allow a large number of users to connect with each other using social networking applications such as Facebook, Twitter, LinkedIn, etc. There are following cloud-based social applications:

- Facebook: Facebook is a social networking website which allows active users to share files, photos, videos, status, more to their friends, relatives, and business partners using the cloud storage system. On Facebook, we will always get notifications when our friends like and comment on the posts.
- Twitter: Twitter is a social networking site. It is a microblogging system. It allows users to follow high profile celebrities, friends, relatives, and receive news. It sends and receives short posts called tweets.
- LinkedIn: LinkedIn is a social network for students, freshers, and professionals.

Business Applications- Cloud computing has made businesses more collaborative and easier by incorporating various apps (Table 2).

Table 2: Business Applications of Cloud

MailChimp	It offers an e-mail publishing platform. It is widely employed by the businesses to design and send their e-mail campaigns.
Chatter	Chatter app helps the employee to share important information about organization in real time. One can get the instant feed regarding any issue.
Google Apps for Business	Google offers creating text documents, spreadsheets, presentations, etc., on Google Docs which allows the business users to share them in collaborating manner.
Quickbooks	It offers online accounting solutions for a business. It helps in monitoring cash flow, creating VAT returns and creating business reports.

Data Storage and Backup- Box.com, Mozy, Joukuu are the applications offering data storage and backup services in cloud (Table 3).

Table 3: Data Storage and Backup Cloud Applications

Box.com	Box.com offers drag and drop service for files. The users need to drop the files into Box and access from anywhere.
Mozy	Mozy offers online backup service for files to prevent data loss.
Joukuu	Joukuu is a web-based interface. It allows to display a single list of contents for files stored in Google Docs, Box.net and Dropbox.

Unit 13: Application of Cloud Computing and Related Technologies

With respect to this as a cloud, the market is growing rapidly and it is providing various services day by day. So, in the future cloud computing is going to touch many more sectors by providing more applications and services.

13.5 Internet-of-Things (IoT)

The Internet of things (IoT) describes the network of physical objects – a.k.a. "things" – that are embedded with sensors, software, and other technologies for the purpose of connecting and exchanging data with other devices and systems over the Internet.

The Internet of Things (IoT) describes the network of physical objects: "things" – that are embedded with sensors, software, and other technologies for the purpose of connecting and exchanging data with other devices and systems over the internet. These devices range from ordinary household objects to sophisticated industrial tools. With more than 7 billion connected IoT devices today, experts are expecting this number to grow to 10 billion by 2020 and 22 billion by 2025. Oracle has a network of device partners.

In the most general terms, IoT includes any object – or "thing" – that can be connected to an Internet network, from factory equipment and cars to mobile devices and smart watches. But today, the IoT has more specifically come to mean connected things that are equipped with sensors, software, and other technologies that allow them to transmit and receive data – to and from other things. Traditionally, connectivity was achieved mainly via Wi-Fi, whereas today 5G and other types of network platforms are increasingly able to handle large datasets with speed and reliability.

Of course, the whole purpose of gathering data is not merely to have it but to use it. Once IoT devices collect and transmit data, the ultimate point is to analyse it and create an informed action. Here is where AI technologies come into play: augmenting IoT networks with the power of advanced analytics and machine learning.

Things have evolved due to the convergence of multiple technologies, real-time analytics, machine learning, commodity sensors, and embedded systems. Traditional fields of embedded systems, wireless sensor networks, control systems, automation (including home and building automation), and others all contribute to enabling the Internet of things. In the consumer market, IoT technology is most synonymous with products pertaining to the concept of the "smart home", including devices and appliances (such as lighting fixtures, thermostats, home security systems and cameras, and other home appliances) that support one or more common ecosystems, and can be controlled via devices associated with that ecosystem, such as smartphones and smart speakers. The IoT can also be used in healthcare systems.

13.6 Why Is Internet of Things (IoT) so Important?

Over the past few years, IoT has become one of the most important technologies of the 21st century. Now that we can connect everyday objects – kitchen appliances, cars, thermostats, baby monitors – to the internet via embedded devices, seamless communication is possible between people, processes, and things.

By means of low-cost computing, the cloud, big data, analytics, and mobile technologies, physical things can share and collect data with minimal human intervention. In this hyperconnected world, digital systems can record, monitor, and adjust each interaction between connected things. The physical world meets the digital world – and they cooperate.

- Dynamic control of industry and daily life- IoT is used in our daily life to automatically make decisions and optimize power consumption. Google Home, Amazon echo, etc. are examples of some of the IoT based home automation devices where IoT and machine learning are used heavily.
- Improves the resource utilization ratio- improving resource efficiency is the lack of suitable capabilities to collect, exchange and share real-time data among various stakeholders. Having such capabilities would provide improved awareness and visibility of resource use and help make better decisions that drive overall productivity.

Cloud Computing

- Integrating human society and physical systems- Real-time mapping of human environment & physical systems.
- Flexible configuration- With Cloud IoT Core, you can control a device by modifying its configuration. A device configuration is an arbitrary, user-defined blob of data. After a configuration has been applied to a device, the device can report its state to Cloud IoT Core.
- The Internet of Things (IoT) is an ecosystem that consists of physical devices that can receive, store, process, and send digital information. Implementing these smart devices requires a variety of skills. In addition to the physical installation, connected devices must be configured and secured properly. That's why the company you choose to implement your IoT solutions must be well versed in cyber-security and the configuration of technology infrastructure.
- Acts as technology integrator that helps in integrating various technologies. Help you narrow down the best system integrator for your wireless IoT product or service.
- IoT is a global inter-connection for devices, objects, and things that contain embedded technologies to sense, communicate, and interact.

13.7 History of IoT

The main concept of a network of smart devices was discussed as early as 1982, with a modified Coca-Cola vending machine at Carnegie Mellon University becoming the first Internet-connected appliance, able to report its inventory and whether newly loaded drinks were cold or not. Mark Weiser's 1991 paper on ubiquitous computing, "The Computer of the 21st Century", as well as academic venues such as UbiComp and PerCom produced the contemporary vision of the IOT. In 1994, Reza Raji described the concept in IEEE Spectrum as "[moving] small packets of data to a large set of nodes, so as to integrate and automate everything from home appliances to entire factories". Between 1993 and 1997, several companies proposed solutions like Microsoft's at Work or Novell's NEST. The field gained momentum when Bill Joy envisioned device-to-device communication as a part of his "Six Webs" framework, presented at the World Economic Forum at Davos in 1999.

The term "Internet of things" was coined by Kevin Ashton of Procter and Gamble, later MIT's Auto-ID Center, in 1999, though he prefers the phrase "Internet for things" (Figure 1). At that point, he viewed radio-frequency identification (RFID) as essential to the Internet of things, which would allow computers to manage all individual things. The main theme of the Internet of Things is to embed short-range mobile transceivers in various gadgets and daily necessities to enable new forms of communication between people and things, and between things themselves.

Defining the Internet of things as "simply the point in time when more 'things or objects' were connected to the Internet than people", Cisco Systems estimated that the IoT was "born" between 2008 and 2009, with the things/people ratio growing from 0.08 in 2003 to 1.84 in 2010.

13.8 Characteristics of IoT

According to the definition of IoT, it is the way to interconnection with the help of the internet devices that can be embedded to implement the functionality in everyday objects by enabling them to send and receive data. Today data is everything and everywhere. Hence, IoT can also be defined as the analysis of the data generate a meaning action, triggered subsequently after the interchange of data.

Intelligence: IoT comes with the combination of algorithms and computation, software & hardware that makes it smart. Ambient intelligence in IoT enhances its capabilities which facilitate the things to respond in an intelligent way to a particular situation and supports them in carrying out specific tasks. In spite of all the popularity of smart technologies, intelligence in IoT is only concerned as means of interaction between devices, while user and device interaction is achieved by standard input methods and graphical user interface.

Unit 13: Application of Cloud Computing and Related Technologies

Connectivity: Connectivity empowers Internet of Things by bringing together everyday objects. Connectivity of these objects is pivotal because simple object level interactions contribute towards collective intelligence in IoT network. It enables network accessibility and compatibility in the things. With this connectivity, new market opportunities for Internet of things can be created by the networking of smart things and applications.

Dynamic Nature: The primary activity of Internet of Things is to collect data from its environment, this is achieved with the dynamic changes that take place around the devices. The state of these devices changes dynamically, example sleeping and waking up, connected and/or disconnected as well as the context of devices including temperature, location and speed. In addition to the state of the device, the number of devices also changes dynamically with a person, place and time.

Enormous Scale: The number of devices that need to be managed and that communicate with each other will be much larger than the devices connected to the current Internet. The management of data generated from these devices and their interpretation for application purposes becomes more critical. Gartner (2015) confirms the enormous scale of IoT in the estimated report where it stated that 5.5 million new things will get connected every day and 6.4 billion connected things will be in use worldwide in 2016, which is up by 30 percent from 2015. The report also forecasts that the number of connected devices will reach 20.8 billion by 2020.

Sensing: IoT wouldn't be possible without sensors which will detect or measure any changes in the environment to generate data that can report on their status or even interact with the environment. Sensing technologies provide the means to create capabilities that reflect a true awareness of the physical world and the people in it. The sensing information is simply the analogue input from the physical world, but it can provide the rich understanding of our complex world.

Heterogeneity: Heterogeneity in Internet of Things as one of the key characteristics. Devices in IoT are based on different hardware platforms and networks and can interact with other devices or service platforms through different networks. IoT architecture should support direct network connectivity between heterogeneous networks. The key design requirements for heterogeneous things and their environments in IoT are scalabilities, modularity, extensibility and interoperability.

Security: IoT devices are naturally vulnerable to security threats. As we gain efficiencies, novel experiences, and other benefits from the IoT, it would be a mistake to forget about security concerns associated with it. There is a high level of transparency and privacy issues with IoT. It is important to secure the endpoints, the networks, and the data that is transferred across all of it means creating a security paradigm.

Active Engagement: IoT makes the connected technology, product, or services to active engagement between each other.

Endpoint Management: It is important to be the endpoint management of all the IoT system otherwise, it makes the complete failure of the system. For example, if a coffee machine itself orders the coffee beans when it goes to end but what happens when it orders the beans from a retailer and we are not present at home for a few days, it leads to the failure of the IoT system. So, there must be a need for endpoint management.

There are a wide variety of technologies that are associated with IoT that facilitate in its successful functioning. IoT technologies possess the above-mentioned characteristics which create value and support human activities; they further enhance the capabilities of the IoT network by mutual cooperation and becoming the part of the total system.

13.9 Structure and Working of IoT

IoT can be viewed as a gigantic network consisting of networks of devices and computers connected through a series of intermediate technologies like RFIDs, WiFi. There are wireless connections that may act as enablers of this connectivity (Figure 3).

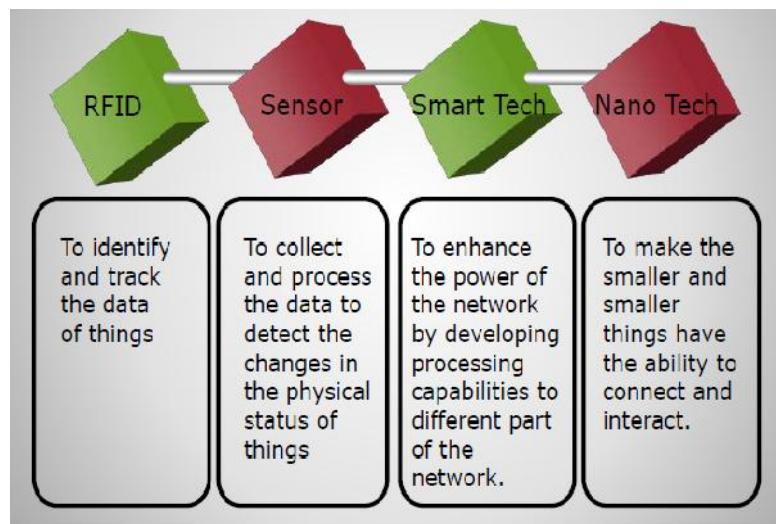


Figure 3: IoT Structural Components

Tagging Things: These include real-time item traceability and addressability done by Radio-Frequency Identification Device (RFID). RFIDs use electromagnetic fields to automatically identify and track tags attached to objects. An RFID system consists of a tiny radio transponder, a radio receiver and transmitter. When triggered by an electromagnetic interrogation pulse from a nearby RFID reader device, the tag transmits digital data, usually an identifying inventory number, back to the reader. These are widely used in Transport and Logistics and are easy to deploy including the RFID tags and RFID readers.

Feeling Things: Sensors act as primary devices to collect data from the environment. Today, almost everybody probably has a smartphone. The smartphone isn't just a mobile phone that has access to the Internet. The iPhone has a lot of different types of sensors.

Shrinking Things: Miniaturization & Nanotechnology has provoked the ability of smaller things to interact & connect within the "things" or "smart devices."

Thinking Things: Embedded intelligence in devices through sensors has formed network connection to the Internet. It can make the "things" realize intelligent control.

13.10 Working of IoT

Just like Internet has changed the way we work and communicate with each other, by connecting us through the World Wide Web (internet), IoT also aims to take this connectivity to another level by connecting multiple devices at a time to the internet thereby facilitating man to machine and machine to machine interactions. IoT is also referred to as Internet of Everything (IoE) or Machine-to-Machine (M2M), Skynet being an ecosystem of connected physical objects that are accessible through the Internet. IoT consists of all the web-enabled devices that collect, send and act on data they acquire from their surrounding environments using embedded sensors, processors and communication hardware. The devices are called "connected" or "smart" devices, can sometimes talk to other related devices, a process called machine-to-machine (M2M) communication, and act on the information they get from one another. The devices collect useful data with the help of various existing technologies and then autonomously flow the data between other devices.

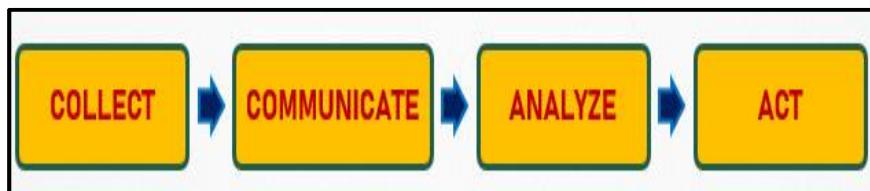


Figure 4: IoT Processing Stages

IoT devices are empowered to be our eyes and ears when we can't physically be there. Equipped with sensors, devices capture the data that we might see, hear, or sense. They then share that data

Unit 13: Application of Cloud Computing and Related Technologies

as directed, and we analyse it to help us inform and automate our subsequent actions or decisions. There are four key stages in this process (Figure 4):

- Capture or Collect the data: Through sensors, IoT devices capture data from their environments. This could be as simple as the temperature or as complex a real-time video feed.
- Communicate the data: Using available network connections, IoT devices make this data accessible through a public or private cloud, as directed
- Process/ Analyse the data: At this point, software is programmed to do something based on that data – such as turn on a fan or send a warning. It can include creating information from the data. It includes visualizing the data; building reports and filtering data
- Action on the data: Accumulated data from all devices within an IoT network is analysed. This delivers powerful insights to inform confident actions and business decisions. It includes:
 - Communicate with another machine (M2M)
 - Sending a notification (email, SMS, text)
 - Talk to another system

13.11 Building Blocks of IoT

Four things form basic building blocks of the IoT system- sensors, processors, gateways, applications. Each of these nodes has to have its own characteristics in order to formulate useful IoT system. The smart systems and IoT are driven by a combination of (Figure 5):

Sensors/Devices: Sensors are the front-end of the IoT devices and are the so-called “Things” of the IoT system. Their main purpose is to collect data from its surroundings (sensors) or give out data to its surrounding (actuators). Each sensor has to be uniquely identifiable devices with a unique IP address so that they can be easily identifiable over a large network. These are active in nature which means that they should be able to collect real-time data. All of this collected data can have various degrees of complexities ranging from a simple temperature monitoring sensor or a complex full video feed. A device can have multiple sensors that can bundle together to do more than just sense things. For example, our phone is a device that has multiple sensors such as GPS, accelerometer, camera but our phone does not simply sense things. The most rudimentary step will always remain to pick and collect data from the surrounding environment be it a standalone sensor or multiple devices.

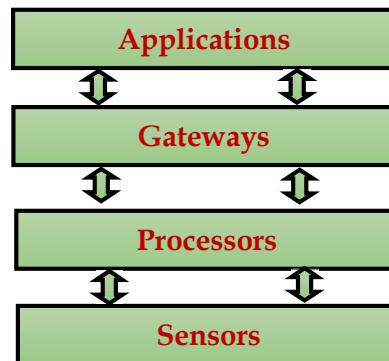


Figure 5: Simplified Block Diagram of the Basic Building Blocks of IoT

Connectivity: Next, that collected data is sent to a cloud infrastructure but it needs a medium for transport. The sensors can be connected to the cloud through various mediums of communication and transports such as cellular networks, satellite networks, Wi-Fi, Bluetooth, wide-area networks (WAN), low power wide area network and many more. Every option we choose has some specifications and trade-offs between power consumption, range, and bandwidth. So, choosing the best connectivity option in the IoT system is important. There can be multiple connectivity options such as:

Cloud Computing

Applications: The applications form another end of an IoT system. They are essential for proper utilization of all the data collected. These can be any cloud-based applications which are responsible for rendering effective meaning to the data collected. The applications are controlled by users and are delivery point of particular services. Examples: Home automation apps, security systems, industrial control hub, etc.

Gateways: Gateways are responsible for routing the processed data and send it to proper locations for its (data) proper utilization. They help in to and fro communication of the data and provide network connectivity to the data. The network connectivity is essential for any IoT system to communicate.

Processes and People: The processors formulate the brain of the IoT system with their main function of processing the data captured by the sensors and process them so as to extract the valuable data from the enormous amount of raw data collected. They give intelligence to the data and mostly work on real-time basis and can be easily controlled by the applications. Processors are also responsible for securing the data- that is performing encryption and decryption of data. Example: Embedded hardware devices, microcontroller, etc. are the ones that process the data

Once the data is collected and it gets to the cloud, the software performs processing on the acquired data. This can range from something very simple, such as checking that the temperature reading on devices such as AC or heaters is within an acceptable range. It can sometimes also be very complex, such as identifying objects (such as intruders in your house) using computer vision on video.

13.12 Applications of IoT

IoT can help the organizations in reducing cost through improved process efficiency, asset utilization and productivity. It aids in the growth and convergence of data, processes and things on the internet would make such connections more relevant and important. IoT supports creation of more opportunities for people, businesses and industries across several sectors.

Corporate Aspect of IoT

- IoT is a transformational force- Any physical object can be transformed into an IoT device if it can be connected to the internet to be controlled or communicate information. A lightbulb that can be switched on using a smartphone app is an IoT device, as is a motion sensor or a smart thermostat in your office or a connected streetlight.
- Companies can improve performance through IoT analytics and IoT Security to deliver better results.
- Businesses in the utilities, oil and gas, insurance, manufacturing, transportation, infrastructure and retail sectors can reap the benefits of IoT by making more informed decisions, aided by the torrent of interactional and transactional data at their disposal.

IoT in Food Industry

IoT is making every aspect of the food industry smarter through a combination of IoT smart connected products gathering data throughout the supply chain and intelligent algorithms, converting them into smart insights.

IoT in Education

IoT can help us make education more accessible in terms of geography, status, and ability. There are boundless opportunities to integrate IoT solutions into school environments. It serves as a solid foundation on which to build a broader understanding of IoT applications in education such as:

- Foreign Language Instruction
- Connected / Smart Classrooms
- Task-Based Learning

Other IoT Applications in Education include:

- Special education
- Physical education
- School security
- Classroom monitoring using Video-as-a-Sensor technology
- Attendance monitoring automation

Unit 13: Application of Cloud Computing and Related Technologies

- Student physical and mental health
- Learning from home
- Personalized learning

IoT in Pharmaceuticals

Pharma IoT can bring in transparency into drug production and storage environment by enabling multiple sensors to monitor in real time multiple environmental indicators, such as: temperature, humidity, light, radiation, CO₂ level.

IoT in Retail

IoT is changing numerous aspects of the retail sector—from customer experience to supply chain management. With these transformations come new challenges. It allows these devices to communicate, analyze, and share data regarding the physical surrounding world via cloud-based software platforms and other networks. Retail sector has had a complete makeover over the last decade driven by technologies such as Machine learning, Artificial Intelligence, and IoT. The major benefits of IoT in retail sector include:

- **Enhanced Supply Chain Management-** IoT solutions like RFID tags & GPS sensors can be used by retailers to get a comprehensive picture regarding movement of goods from manufacturing to when its placed in a store to when a customer buys it.
- **Better Customer Service-** IoT also helps brick and mortar retailers by generating insights into customer data while opening opportunities for leveraging that data. For example, retailer IoT applications can synthesize data from video surveillance cameras, mobile devices, & social media websites, allowing merchants better to predict customer behavior.
- **Smarter Inventory Management** helps in automating inventory visibility otherwise inventory management can be a headache.
- Lack of accurate tracking for inventory can lead to stock-outs and overstock, costing retailers around the world billions annually.
- Smart inventory management solutions based on store shelf sensors, RFID tags, beacons, video monitoring, & digital price tags, retail businesses can enhance procurement planning.
- When product inventory is low, the system offers to re-order adequate amount based on the analytics acquired from IoT data.

IoT in Logistics

Transport and logistics is one of the first business sectors interested in IoT technologies. There are currently two systems that are already available and deployed: ConLock and ContainerSafe. There has been integration of light sensors, GPS and GSM. In IoT, there are things that could be an automobile with built-in sensors, that is, objects that have been assigned an IP address or a person with a heart monitor.

Other Applications Using IoT

There are several applications that are making use of IoT capabilities such as:

- **Google Traffic**
- **Jawbone UP:** One can link it to an iPhone application. It is not just a passive bracelet but is highly recommended to change life-style.
- **AutoBot:** These offer diagnostics service for cars and generate alerts relatives in case of an accident. Autobots can be used for discovery service of car position and can be easily integrated with several web services.
- **Daily Life and Domotics:** A framework has been developed for Home Automation applications: FreeDom.

IoT in Management

IoT is used in different types of management activities.

13.13 Cloud of Things (CoT)

Cloud of Things (CoT) involves the integration of Internet of Things (IoT) with Cloud Computing (CC). It is a high-performance cloud-based IoT application platform which allows to remotely monitor, manage and control the IoT enabled devices. It can be used to connect devices and machines and monitor and manage it. Using the cloud computing (CC), after integration with the Internet of Things (IoT), a new technological power/new paradigm is created known as Cloud of Things which provides a new business model with increased efficiency.

Why CoT is required?

Nowadays, the number of IoT enabled devices are so high, and also the volume of IoT generated data is increasing with respect to that storing data locally and temporarily is not possible. The virtual resource utilization and storage capacity requirement is also so high. That is why IoT is integrated with cloud computing resulting into Cloud of Things (CoT) or CloudIoT. It has helped the Internet of Things (IoT) a lot in processing/analysing the data and creating more usefulness from the data generated by IoT by allowing to develop more advanced smart applications.

13.14 Role of Cloud Computing in CoT

Cloud computing provides the ability to utilize the computing and storage powers of a computer without having it physically present in your study room or office. Consider you have a set of 500 sensors established in a private organization in a city which is used to measure all types of data like Pollution, Earthquake, etc. Now, consider that such data is generated by 100 cities and hence a lot of computing power is needed to process this information. So here, the role of cloud computing comes into the picture. So, you can say that IoT devices are used to collect the data and that data is processed in the cloud. Hence, the data collected through IoT devices is stored and processed on cloud. Based on the processing and learning of that data, a decision is detected (Figure 6).

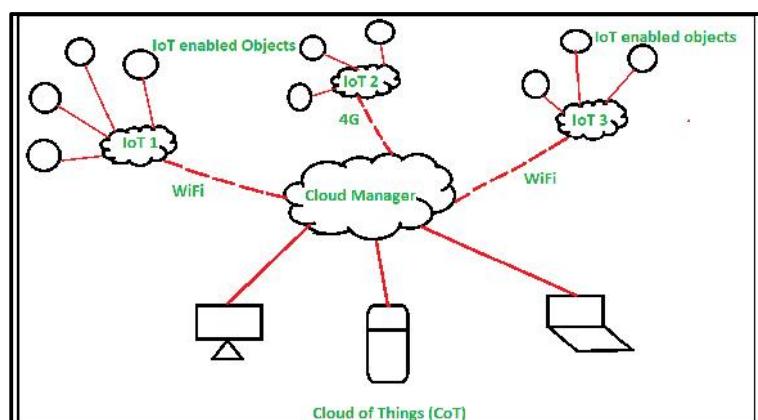


Figure 6: CoT Environment Scenario

13.15 Architecture of Cloud of Things (CoT)

Software-as-a-Service (SaaS), Platform-as-a-Service (PaaS), Infrastructure-as-a-Service (IaaS) are the different models of cloud computing for building, storing, creating and using data/application over the cloud. Similarly, Sensing-as-a-Service, Big-Data-Analytics-as-a-Service, Database-as-a-Service, Identity-and-Policy-Management-as-a-Service, Video-Monitoring-as-a-Service, Data-as-a-Service, Sensor-as-a-Service, etc are the different models/forms of Cloud of Things (CoT). These objects are interconnected through various different networked environments (Figure 7).

Unit 13: Application of Cloud Computing and Related Technologies

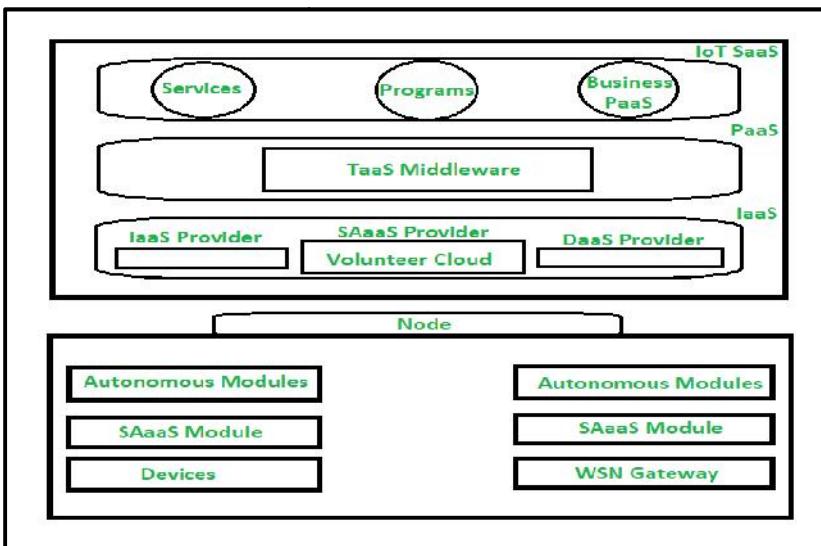


Figure 7: Architecture of Cloud of Things (CoT)

Application Areas of Cloud of Things (CoT)

There are several applications of CoT:

- o Healthcare
- o Smart home
- o Smart city
- o Smart energy
- o Smart mobility
- o Smart surveillance
- o Smart logistic
- o Environmental monitoring



Lab Exercise: Explore the role of Cloud of Things (CoT) in the Healthcare sector.

13.16 Mobile Internet of Things (Mobile IoT)

Mobile IoT refers to the low power wide area (LPWA) 3GPP standardised secure operator managed IoT networks in licensed spectrum. In particular, LPWA networks designed for IoT applications that are low cost, use low data rates, require long battery lives and often operate in remote and hard to reach locations. The existing cellular networks have evolved to deliver service to billions of new devices providing complete IoT connectivity. The key features related with Mobile IoT are:

- o Mobile IoT= Trusted IoT
- o Mobile IoT technologies are also inherently more secure than alternative solutions. This is because Mobile IoT networks use dedicated spectrum bands to avoid interference and because SIMs contain highly secure integrated circuits, but also because mobile operators are supplementing these inherent capabilities with additional security features.
- o GSMA Mobile IoT Initiative is designed to accelerate the commercial availability of low power wide area solutions in licensed spectrum.

Why do we need low power wide area networks?

Cloud Computing

- To support the further growth and development of the Internet of Things (IoT), the mobile industry together with 3GPP has standardised a new class of GSM technologies in record time. LPWA networks will support devices requiring, low power consumption, long range, low cost and security.
- LPWA networks will play an important role in connecting up the billions of new devices making up the IoT. LPWA technologies are expected to serve a diverse range of vertical industries and support a range of applications and deployment scenarios, which existing mobile technologies may not currently be best placed to connect.

Example of Mobile IoT Provider- BeWhere Inc.

- Backed by over 90 of the leading global mobile operators and vendors, and uniting over 2600 members from more than 1700 companies, the GSMA Mobile IoT Innovators are the voice of a vibrant and fast-growing ecosystem.
- There are great success stories from engaged members, such as BeWhere Inc., a member since 2017 when the company began working on Mobile IoT solutions.
- BeWhere have actively participated in the community and technology showcases, building valuable relationships around the world and leveraging the platform for increased visibility in the wider market. The company is now working with over 10 mobile operators globally, and in less than two years since the launch of their Mobile IoT solutions, has secured tens of thousands Mobile IoT product orders and a \$4 million capital raise that included a major operator, seeing their revenues more than double year over year during the first year of Mobile IoT sales, from \$1.7M to \$3.6M (2018).
- Mobile IoT Innovators membership is open to all developers and companies interested in or already working with Mobile IoT technologies.

13.17 Edge Computing

Edge computing is a distributed computing framework that brings enterprise applications closer to data sources such as IoT devices or local edge servers. This proximity to data at its source can deliver strong business benefits, including faster insights, improved response times and better bandwidth availability. The “edge” is where the device communicates with the Internet (Figure 8).

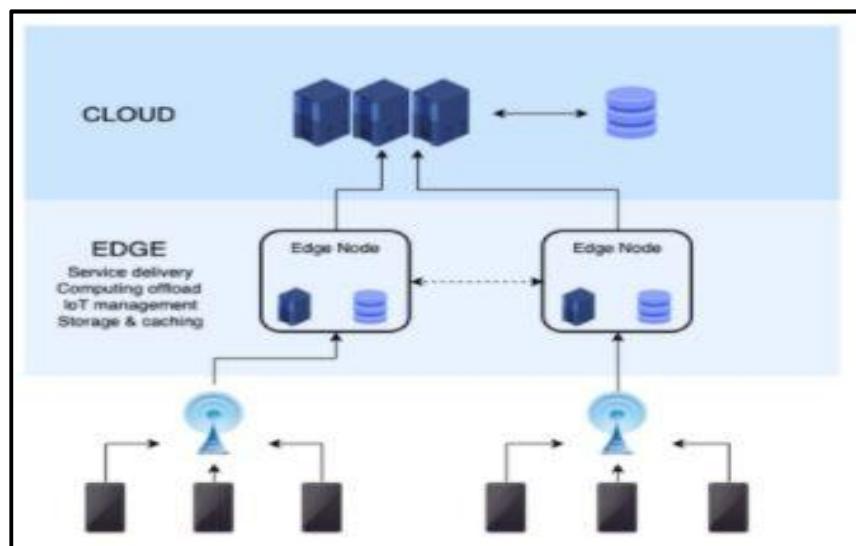


Figure 8: Edge Computing Scenario

Unit 13: Application of Cloud Computing and Related Technologies

“Edge” in Edge Computing-

- Edge can be the router, ISP, routing switches, integrated access devices (IADs), multiplexers, etc. The most significant thing about this network edge is that it should be geographically close to the device.
- ‘Edge’ is the geographic location on the planet to deliver services in a distributed manner.
- Gartner estimates that by 2025, 75% of data will be processed outside the traditional data center or cloud.

Edge vs Traditional Computing Concept

Edge Computing brings the decentralization of networks (Table 4):

- In a traditional setting, data is produced on a user's computer or any other client application. It is then moved to the server through channels like the internet, intranet, LAN, etc., where the data is stored and worked upon. This remains a classic and proven approach to client-server computing.
- However, the exponential growth in the volume of data produced and the number of devices connected to the internet has made it difficult for traditional data center infrastructures to accommodate them.
- The concept of edge computing is simple- instead of getting the data close to the data center, the data center is brought close to the data. The storage and computing resources from the data center are deployed as close as possible (ideally in the same location) to where the data is generated.

Table 4: Edge Computing vs Traditional Computing

Early computing	Applications run only on one isolated computer
Personal computing	Applications run locally either on the user's device or in a data center
Cloud computing	Applications run in data centers and processed via the cloud
Edge computing	Applications run close to the user; either on the user's device or on the network edge

13.18 Cloud Computing vs Edge Computing

The explosive growth and increasing computing power of IoT devices has resulted in unprecedented volumes of data. And data volumes continue to grow as 5G networks increase the number of connected mobile devices. Although cloud has automated and innovated actionable insights from data, but the unprecedented scale and complexity of data that's created by connected devices has outpaced network and infrastructure capabilities. Sending all that device-generated data to a centralized data center or to the cloud causes bandwidth and latency issues. Edge computing offers a more efficient alternative; data is processed and analysed closer to the point

Cloud Computing

where it's created. Because data does not traverse over a network to a cloud or data center to be processed, latency is significantly reduced. Table 5 lists the major differences between edge computing and cloud computing.

Table 5: Edge Computing vs Cloud Computing

Edge Computing	Cloud Computing
Good to be used for those organizations that have a limited budget to invest in financial resources. So, mid-level organizations can use edge computing.	Generally recommended for processing and managing a high volume of data that is complex and massive enough. Thus, such organizations that deal with huge data storage use cloud computing.
Can use different programming languages on different platforms, each having different runtime.	Works for one target platform using one programming language only.
Security in edge computing needs tight and robust plans such as advanced authentication methods, network security, etc.	No high and advanced security methods needed.
Processes time-sensitive data.	Process that data that is not driven by time, i.e., not time-driven.
Processes data at remote locations and uses the Decentralization approach.	Processes and deals with data at centralized locations by using a centralized approach.

13.19 Why Edge Computing?

Edge computing harnesses growing in-device computing capability to provide deep insights and predictive analysis in near-real time. This increased analytics capability in edge devices can power innovation to improve quality and enhance value. It also raises important strategic questions: How do you manage the deployment of workloads that perform these types of actions in the presence of increased compute capacity? How can you use the embedded intelligence in devices to influence operational processes for your employees, your customers and your business more responsively? In order to extract the most value from all those devices, significant volumes of computation must move to the edge.

Edge Computing is a new type of technology that not only saves time but also save the cost of servicing and other charges too. It allows smart applications and devices for responding to data very quickly as soon as it is created and thus removing the lag time. It enables the data stream acceleration that includes real-time processing of data without latency use. Data Stream acceleration is, however, critical for self-driving cars type of technologies and provides equal and vital benefits for the businesses. The efficient processing of data at large scale by allowing processing close to the source, and it saves the use of internet bandwidth also. Thus, it reduces the cost and enables effective accessibility to the applications in remote locations. The ability of edge computing to provide services and processing data at the furthest distance makes a secured layer for the sensitive data without keeping it into a public cloud.

Benefits of Edge Computing

Edge computing has emerged as one of the most effective solutions to network problems associated with moving huge volumes of data generated in today's world. Some of the most important benefits of edge computing:

Eliminates Latency- Latency refers to the time required to transfer data between two points on a network. Large physical distances between these two points coupled with network congestion can cause delays. As edge computing brings the points closer to each other, latency issues are virtually non-existent.

Unit 13: Application of Cloud Computing and Related Technologies

Saves Bandwidth- Bandwidth refers to the rate at which data is transferred on a network. As all networks have a limited bandwidth, the volume of data that can be transferred and the number of devices that can process this is limited as well. By deploying the data servers at the points where data is generated, edge computing allows many devices to operate over a much smaller and more efficient bandwidth.

Reduces Congestion- Although the Internet has evolved over the years, the volume of data being produced everyday across billions of devices can cause high levels of congestion. In edge computing, there is a local storage and local servers can perform essential edge analytics in the event of a network outage.

Applications of Edge Computing

Transportation: It is one of the most potential sectors where edge computing plays a vital role and particularly in autonomous vehicles. It is because autonomous vehicles are full of different sensors types from the camera to the radar system of the car. Such autonomous devices can essentially utilize edge computing for processing data too close to the vehicle through these sensors, and consequently, a good amount of time will be saved. The Automotive Edge Computing Consortium (AECC) mentioned in the year 2018 that it would be launching the operations which are focused on the connected car solutions. But not only the self-driving cars, edge computing is focusing upon trains, aeroplanes, and other transportation forms also.

HealthCare: People rely on fitness trackers, smartwatches, stamina measurement watches, etc., and find these health-monitoring wearables comfortable. However, the real-time analysis is essential for capturing the actual benefits of the collected data because many health wearables devices are directly connected to the cloud while others can be operated in an offline mode only. Certain health devices analyse the pulse rate and sleep pattern in an offline mode only, and doctors use to check and evaluate the patients on the spot using the analysis result. Such smart devices are useful for collecting and processing data in treating patients in any pandemic (such as COVID-19). Through edge computing, doctors are able to more fastly collect and process the data and may provide better plus fast care to the patients, including a security layer to PDHD (patient-generated health data). Through edge computing, hospitals and doctors are able to use and access more cloud applications more quickly, but the security and privacy of data is still a confusion.

Manufacturing: Edge computing in the field of manufacturing will reduce the data that goes to the cloud for applications like predictive maintenance, and it will move the operational technology to the edge computing platforms for running process similarly as processed in the cloud but with more speed and result. However, the maintenance of an on-premise deployment is still reliable on the cloud.

Grid Edge Control and Analytics: These smart grid controls work by creating two-way communication channels via WAN protocols between the power distribution infrastructure, consumers, and the utility head-end. However, edge grid computing is capable of providing advanced real-time monitoring and analytics and also producing actionable insights on distributed energy that generates resources like renewables. Such capability is available in edge computing technologies only. Edge Grid Computing can, however, reduce the overall cost, wastage of energy, avoid outages and overcompensation if electric vehicles, wind farms and hydroelectric dams that generate a massive amount of useful data can assist the utilities with analytics on requirements, peak usage predictions, availability, and energy production.

Remote monitoring of Oil and Gas: At present, IoT devices are providing modern safety monitoring, sensory devices for controlling, viewing, and sensing the temperature, pressure, moisture, humidity, sound, and radiation of oil and gas. With edge computing, real-time safety monitoring can be possible for safeguarding critical machinery infrastructure and oil and gas systems from disasters. Also, several edge IoT monitoring devices are being developed, keeping safety and reliability as the main focus points. Edge Computing allows analyzing, processing, and delivery of data to the end-users in real-time. Consequently, it will enable the control centers to access the data instantly as it occurs, preventing the malfunctions in the most optimized manner. Thus, oil and gas services are the critical infrastructures that can be catastrophic in nature if not maintained with safety and precautions.

Traffic Management: Traffic is the worst waste of time and needs to be optimized. The best way to optimize traffic is by maintaining and improving real-time data. For the traffic management process, the smart transportation systems such as self-driving cars, and other sensory systems make

Cloud Computing

extensive use of edge computing devices. Through edge computing, a massive amount of sensory and other data is analysed, then filtered, and finally compressed before transmitting on the IoT edge gateways to other systems for use. As a result, edge computing reduces network expenses, operating processing, and storage cost for undertaking traffic management solutions.

13.20 Fog Computing

Fog computing involves a system-level horizontal architecture that distributes resources and services of computing, storage, control and networking anywhere along the cloud-to-things continuum. It supports industry verticals and application domains, enables services and applications to be distributed closer to the data-producing sources, and extends from the things, over the network edges, through the cloud and across multiple protocol layers.

Fog Computing is the term coined by Cisco that refers to extending cloud computing to an edge of the enterprise's network. It is also known as Edge Computing or Fogging. It facilitates the operation of computing, storage, and networking services between end devices and computing data centers. It focuses on shifting certain cloud computing resources away from the cloud and closer to the individual devices.

Architecture of Fog Computing

With fog computing, an additional computer is placed at each plant, typically at the LAN level, to concentrate (or aggregate) data locally. It regulates which information should be processed locally and which can be offloaded to the cloud, acting as a mediator between local devices and remote servers.

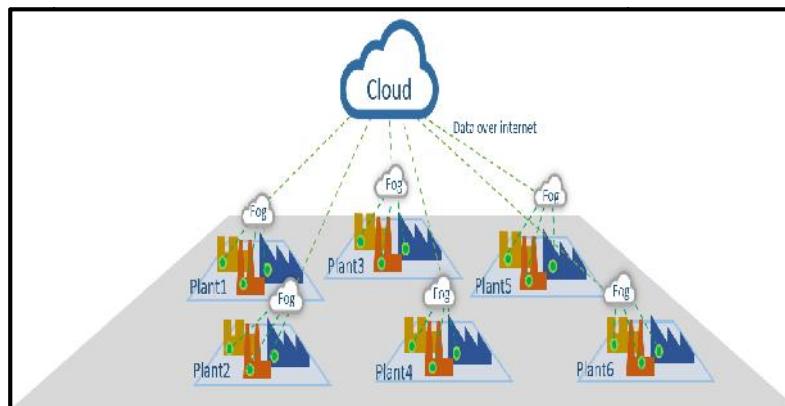


Figure 9: Architecture of Fog Computing

Benefits of Fog Computing

Fog helps in many ways such as:

- Reduction in the number of cloud connections necessary.
- Faster local processing – fog computer sends out consolidated and meaningful data to the cloud, where each plant's information is analysed in combination with other plants
- Reducing network latency and data traffic.
- Increasing the effective network bandwidth.
- Vertically isolating the network.
- Scaling up the deployment easily and efficiently.
- Controlling privacy and data security.



Task: Analyse the concept of fog computing and its applications in the various sectors.

13.21 Blockchain Technology

Blockchain can be defined as a shared ledger, allowing thousands of connected computers or servers to maintain a single, secured, and immutable ledger. Blockchain can perform user transactions without involving any third-party intermediaries. In order to perform transactions, all one needs is to have its wallet. A blockchain wallet is nothing but a program that allows one to spend cryptocurrencies like BTC, ETH, etc. Such wallets are secured by cryptographic methods (public and private keys) so that one can manage and have full control over his transactions.

Blockchain is a shared, immutable ledger for recording transactions, tracking assets and building trust (Figure 10). Blockchains carry with them philosophical, cultural, and ideological underpinnings that must also be understood. In the same way that billions of people around the world are currently connected to the Web, millions, and then billions of people, will be connected to blockchains. Blockchain technology permits transactions to be gathered into blocks and recorded; allows the resulting ledger to be accessed by different servers. Blockchain is a peer-to-peer decentralized distributed ledger technology that makes the records of any digital asset transparent and unchangeable and works without involving any third-party intermediary. It is an emerging and revolutionary technology that is attracting a lot of public attention due to its capability to reduce risks and frauds in a scalable manner.

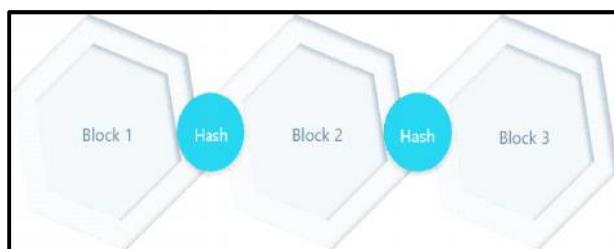


Figure 10: Blockchain Scenario

There are three complementary definitions of blockchain. Technically, blockchain corresponds to the backend database that maintains a distributed ledger that can be inspected openly. Business-wise, blockchain is an exchange network for moving transactions, value, assets between peers, without the assistance of intermediaries. As far as legal definition of blockchain is concerned, blockchain validates transactions, replacing previously trusted entities.

Characteristics of Blockchain Technology

Blockchain eliminates the need for any banks, government, intermediaries of any kind. Blockchain is a potentially revolutionary technology that promises to dramatically change the world. Blockchain is a meta-technology because it affects other technologies, and it is made up of several technologies itself. It is comprised of several pieces: a database, a software application, a number of computers connected to each other, clients to access it, a software environment to develop on it, tools to monitor it, and other pieces.

- **Blockchain is a data structure:** Blockchain is a growing list of data blocks. The data blocks are linked together, such that old blocks cannot be removed or altered.
- **Blockchain is a decentralized ledger:** Blockchain acts as a decentralized ledger of all transactions across a peer-to-peer network. Using this technology, participants can perform transactions without the need for a central certifying authority. The potential applications include fund transfers, settling trades, voting and many others. The business networks that benefit from connectivity - Connected customers, suppliers, banks, partners. Wealth is generated by the flow of goods and services across business network. Anything that is capable of being owned or controlled to produce value, is an asset. There are two fundamental types of asset - Tangible, e.g. a house - Intangible e.g. a mortgage.
- **Immutable records:** No participant can change or tamper with a transaction after it's been recorded to the shared ledger. If a transaction record includes an error, a new transaction must be added to reverse the error, and both transactions are then visible.

- **Smart contracts:** To speed transactions, a set of rules— called a smart contract— is stored on the blockchain and executed automatically. A smart contract can define conditions for corporate bond transfers, include terms for travel insurance to be paid and much more.

Benefits of Using Blockchain Technology

Business runs on information. The faster it's received and the more accurate it is, the better. Blockchain is ideal for delivering that information because it provides immediate, shared and completely transparent information stored on an immutable ledger that can be accessed only by permissioned network members. A blockchain network can track orders, payments, accounts, production and much more. And because members share a single view of the truth, you can see all details of a transaction end-to-end, giving you greater confidence, as well as new efficiencies and opportunities. The different benefits with blockchain technology are:

Immutability

In a traditional database, you have to trust a system administrator that he is not going to change the data. But with blockchain, there is no possibility of changing the data or altering the data; the data present inside the blockchain is permanent; one cannot delete or undo it.

Transparency

Centralized systems are not transparent, whereas blockchain (a decentralized system) offers complete transparency. By utilizing blockchain technology, organizations and enterprises can go for a complete decentralized network where there is no need for any centralized authority, thus improving the transparency of the entire system.

High Availability

Unlike centralized systems, blockchain is a decentralized system of P2P network which is highly available due to its decentralized nature. Since in the blockchain network, everyone is on a P2P network, and everyone has a computer running, therefore, even if one peer goes down, the other peers still work.

High Security

This is another major benefit that blockchain offers. Technology is assumed to offer high security as all the transactions of blockchain are cryptographically secure and provide integrity. Thus, instead of relying on third-party, you need to put your trust in cryptographic algorithms.

Distributed and Centralized Ledger

A distributed ledger is a database that exists across several locations or among multiple participants (Figure 11). By contrast, most companies currently use a centralized database that lives in a fixed location. A centralized database essentially has a single point of failure. There is one ledger. All nodes have some level of access to that ledger. All nodes agree to a protocol that determines “true state” of ledger at any point in time. The application of this protocol is called as “achieving consensus”.

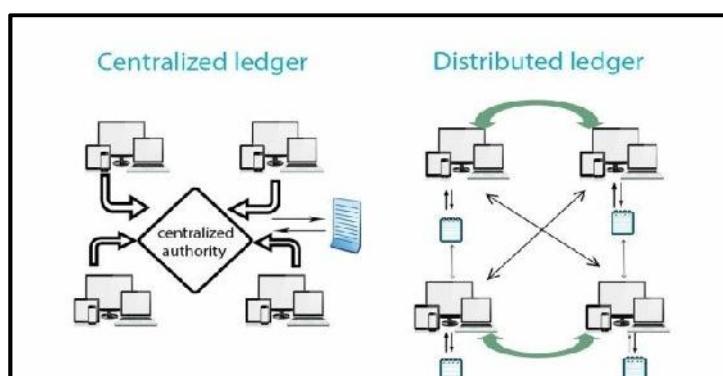


Figure 11: Distributed vs Centralized Ledger

Unit 13: Application of Cloud Computing and Related Technologies

A distributed ledger is decentralized to eliminate the need for a central authority or intermediary to process, validate or authenticate transactions. Enterprises use distributed ledger technology to process, validate or authenticate transactions or other types of data exchanges. Typically, these records are only ever stored in the ledger when the consensus has been reached by the parties involved. All the files in the distributed ledger are then timestamped and given a unique cryptographic signature. All of the participants on the distributed ledger can view all of the records in question. The technology provides a verifiable and auditable history of all information stored on that particular dataset.

In a centralized ledger, there are multiple ledgers, but Bank holds “golden record”. Client B must reconcile its own ledger against that of Bank, and must convince Bank of the “true state” of the Bank ledger if discrepancies arise.

How Distributed Ledger Works?

As illustrated in Figure 12, all the nodes in the distributed ledger have the certain level of access to that particular ledger that is a single ledger over there, then all the nodes they agree to a particular protocol that determines the true state of any ledger at any point of time at any particular point of time. The application of this protocol is sometimes called as achieving consensus. Then, we have the other type of ledger that is a centralized ledger used in the banks. The centralized ledger actually holds out the golden records there are multiple ledger's that are stored over there.

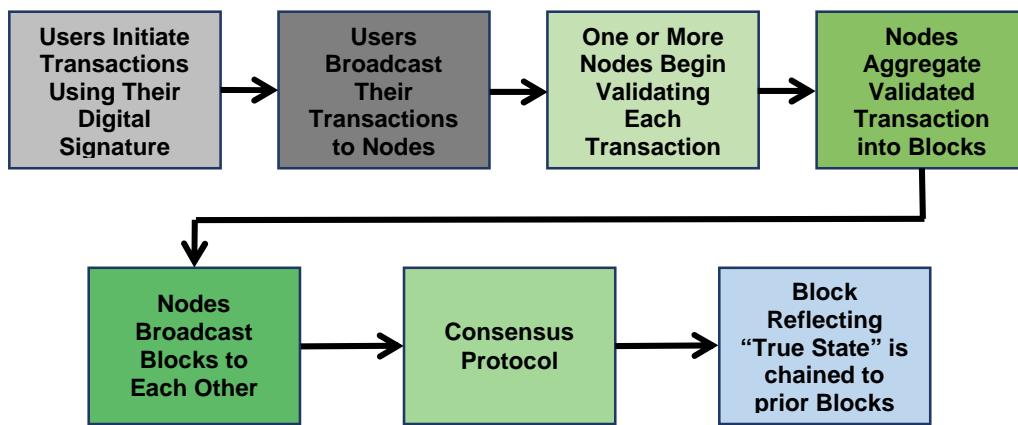


Figure 12: Scenario of Distributed Ledger

Private and Public Blockchain

A public blockchain is the permission-less distributed ledger technology where anyone can join and do transactions. It is a non-restrictive version where each peer has a copy of the ledger. This also means that anyone can access the public blockchain if they have an internet connection. One of the first public blockchains that were released to the public was the bitcoin public blockchain. It enabled anyone connected to the internet to do transactions in a decentralized manner.

The verification of the transactions is done through consensus methods such as Proof-of-Work (PoW), Proof-of-Stake (PoS), and so on. At the cores, the participating nodes require to do the heavy-lifting, including validating transactions to make the public blockchain work. If a public blockchain doesn't have the required peers participating in solving transactions, then it will become non-functional. There are also different types of blockchain platforms that use these various types of blockchain as the base of their project. However, each platform introduces more features in its platform aside from the usual ones. Examples of public blockchain: Bitcoin, Ethereum, Litecoin, NEO

Advantages of Public Blockchain:

- Anyone can join the public blockchain.
- It brings trust among the whole community of users.
- Everyone feels incentivized to work towards the betterment of the public network.
- Public blockchain requires no intermediaries to work.
- Public blockchains are also secure depending on the number of participating nodes.

Cloud Computing

- It brings transparency to the whole network as the available data is available for verification purposes.

A private blockchain is one of the different types of blockchain technology. A private blockchain can be best defined as the blockchain that works in a restrictive environment, i.e., a closed network. It is also a permissioned blockchain that is under the control of an entity. Private blockchains are amazing for using at a privately-held company or organization that wants to use it for internal use-cases. By doing so, you can use the blockchain effectively and allow only selected participants to access the blockchain network. The organization can also set different parameters to the network, including accessibility, authorization, and so on!

Advantages of Private Blockchain:

- Easy to Manage: The consortium or company running a private blockchain can easily change the rules of a blockchain, revert transactions, modify balances, etc.
- Cheap Transaction: Transactions are cheaper, since they only need to be verified by a few nodes that can be trusted to have very high processing power, and do not need to be verified by ten thousand laptops.
- Trustful and Fault Handling is Worth: Nodes can be trusted to be very well-connected, and faults can quickly be fixed by manual intervention, allowing the use of consensus algorithms which offer finality after much shorter block times.
- Privacy: If read permissions are restricted, private blockchains can provide a greater level of privacy.

Public vs Private Blockchain

The primary difference between public and private blockchain is the level of access participants are granted. In order to pursue decentralization to the fullest extent, public blockchains are completely open. Anyone can participate by adding or verifying data. The most common examples of public blockchain are Bitcoin (BTC) and Ethereum (ETH). Both of these cryptocurrencies are created with open source computing codes, which can be viewed and used by anyone. Public blockchain is about accessibility, and this is evident in how it is used.

Conversely, private blockchain (also known as permissioned blockchain) only allows certain entities to participate in a closed network. Participants are granted specific rights and restrictions in the network, so someone could have full access or limited access at the discretion of the network. As a result, private blockchain is more centralized in nature as only a small group controls the network. The most common examples of private blockchains are Ripple (XRP) and Hyperledger.

Additionally, some public blockchains also allow anonymity, while private blockchains do not. For example, anyone can buy and sell Bitcoin without having their identity revealed. It allows everyone to be treated equally. Whereas with private blockchains, the identities of the participants are known. This is typically because private blockchain is used in the corporate and business to business sphere, where it is important to know who is involved, but we'll discuss that more later.



Did you know?

Blockchain is extensively used in Business sector nowadays. Blockchain for business is valuable for entities transacting with one another. With distributed ledger technology, permissioned participants can access same information at same time to improve efficiency, build trust and remove friction. Blockchain also allows a solution to rapidly size and scale, & many solutions can be adapted to perform multiple tasks across industries. Blockchain for business delivers these benefits based on four attributes unique to the technology:

- **Consensus:** Shared ledgers are updated only after the transaction is validated by all relevant participants involved.
- **Replication:** Once a block— record of an event— is approved, it is automatically created across ledgers for all participants in that channel. Every network partner sees & shares a single “trusted reality” of the transactions.

Unit 13: Application of Cloud Computing and Related Technologies

- **Immutability:** More blocks can be added, but not removed, so there is a permanent record of every transaction, which increases trust among the stakeholders.
- **Security:** Only authorized entities are allowed to create blocks & access them. Only trusted partners are given access permission.

Applications of Blockchain

Blockchain Service: Blockchain testing and research require an ecosystem with multiple systems included in it which helps in the development and testing of the same. Many players have seen the benefits of offering blockchain services which include Amazon, Microsoft, IBM, and they have started providing blockchain services to their customers. The advantage of using the services that the user will not have to face the problem of configuring and setting up a blockchain that is working in nature along with no requirement of hardware investments. Examples include Accenture, ConsenSys, Ernst and young etc.

Blockchain First: This allows the users to directly work with blockchain tools and stack. share assembly is required for users who are not familiar with the technology and technicalities will not be able to do so. Change working directory with blockchain assures us an excellent degree of innovation example in building the decentralized applications.

Vertical Solutions: In case of financial services that is being the tremendous growth in the past few years in the segment. They provide industry-specific solutions and are mostly based on private blockchain. Examples include Axon, Clematis, R3 etc.

Smart Contracts: Smart contracts are like regular contracts except the rules of the contract are enforced in real-time on a blockchain, which eliminates the middleman and adds levels of accountability for all parties involved in a way not possible with traditional agreements. This saves businesses time and money, while also ensuring compliance from everyone involved.

Blockchain-based contracts are becoming more and more popular as sectors like government, healthcare and the real estate industry discover the benefits. Below are a few examples of how companies are using blockchain to make contracts smarter.

Cloud Storage: Simply using excess hard drive space, users could store the traditional cloud 300 times over.

Supply-Chain Communications and Proof-of- Provenance: Most of the things we buy aren't made by a *single* entity, but by a chain of suppliers who sell their components (e.g., graphite for pencils) to a company that assembles and markets the final product.

Electronic Voting: Blockchain can be possibly used in the next election or in voting because of its revolutionary unchanging nature. Voting will become more secure and fail proof by the help of blockchain. Delegated Proof of Stake (DPOS) is the fastest, most efficient, most decentralized, and most flexible consensus model available.

IoT(Internet of Things): The blockchain is also now used by IoT. This ensures that data which is going to transfer over or between the devices will be secure and encrypted without any interference.

Healthcare: Healthcare is also a domain where the uses of Blockchain technology has been used for storing the details of the patients. This technology ensures that anyone who has access to this particular blockchain can have access of patient data. This database will be highly secure and for checking the data related with the patient-doctor has to log in there with public key and details and he can check the data of the patients.

Banking: Nowadays blockchain is also replacing the existing or we can say overtaking the current Banking system. By the help of blockchain, we can transfer the fund from one to another person in a second because the validation of the transaction will take place by uses of blockchain and cryptography. It's a possibility that blockchain is going to cut down of 19.8 Billion Dollar which is going for middleman cost/year. Because of the blockchain, the hacking of account will become impossible.



Task: What are the different applications of Blockchain technology?

Summary

- End-User Computing (EUC) refers to computer systems and platforms that help non-programmers create applications.
- Internet of things (IoT) describes the network of physical objects—*a.k.a.* "things"—that are embedded with sensors, software, and other technologies for the purpose of connecting and exchanging data with other devices and systems over the Internet.
- The term "Internet of things" was coined by Kevin Ashton of Procter and Gamble, later MIT's Auto-ID Center, in 1999, though he prefers the phrase "Internet for things"
- RFIDs use electromagnetic fields to automatically identify and track tags attached to objects. An RFID system consists of a tiny radio transponder, a radio receiver and transmitter.
- Edge computing is a distributed computing framework that brings enterprise applications closer to data sources such as IoT devices or local edge servers.
- Mobile IoT refers to the low power wide area (LPWA) 3GPP standardised secure operator managed IoT networks in licensed spectrum. In particular, LPWA networks designed for IoT applications that are low cost, use low data rates, require long battery lives and often operate in remote and hard to reach locations.
- Blockchain is a peer-to-peer decentralized distributed ledger technology that makes the records of any digital asset transparent and unchangeable and works without involving any third-party intermediary.
- The decentralized blockchains are immutable, which means that the data entered is irreversible. For Bitcoin, this means that transactions are permanently recorded and viewable to anyone.
- Fog computing supports industry verticals and application domains, enables services and applications to be distributed closer to the data-producing sources, and extends from the things, over the network edges, through the cloud and across multiple protocol layers.
- Fog computing is the term coined by Cisco that refers to extending cloud computing to an edge of the enterprise's network. It is also known as Edge Computing or Fogging.
- Fog computing facilitates the operation of computing, storage, and networking services between end devices and computing data centers. It focuses on shifting certain cloud computing resources away from the cloud and closer to the individual devices.

Keywords

- **Sensors:** A sensor is a device, module, machine, or subsystem whose purpose is to detect events or changes in its environment and send the information to other electronics, frequently a computer processor. A sensor is always used with other electronics.
- **Cloud of Things (CoT):** Cloud of Things (CoT) involves the integration of Internet of Things (IoT) with Cloud Computing (CC). It is a high-performance cloud-based IoT application platform which allows to remotely monitor, manage and control the IoT enabled devices.
- **Mobile IoT:** Mobile IoT refers to the low power wide area (LPWA) 3GPP standardised secure operator managed IoT networks in licensed spectrum.
- **Radio Frequency Identification (RFID):** RFID uses electromagnetic fields to automatically identify and track tags attached to objects. An RFID system consists of a tiny radio transponder, a radio receiver and transmitter.
- **Infrared Sensor (IR Sensor):** IR Sensors or Infrared Sensor are light based sensor that are used in various applications like Proximity and Object Detection. IR Sensors are used as proximity sensors in almost all mobile phones.

Unit 13: Application of Cloud Computing and Related Technologies

- **Blockchain:** A blockchain is, in the simplest of terms, a time-stamped series of immutable records of data that is managed by a cluster of computers not owned by any single entity. Each of these blocks of data (i.e. block) is secured and bound to each other using cryptographic principles (i.e. chain).
- **Decentralized Ledger:** A distributed ledger is a database that exists across several locations or among multiple participants. By contrast, most companies currently use a centralized database that lives in a fixed location. A centralized database essentially has a single point of failure.
- **Fog computing:** Fog computing involves a system-level horizontal architecture that distributes resources and services of computing, storage, control and networking anywhere along the cloud-to-things continuum.

Self Assessment

1. The full form of VDI is
 - A. Virtual Display Information
 - B. Visual Display Interface
 - C. Visual Desktop Interface
 - D. Virtual Desktop Infrastructure

2. _____ refers to computer systems and platforms that help non-programmers create applications.
 - A. End-user computing
 - B. Interface computing
 - C. Process computing
 - D. Program computing

3. End-user computing can also be referred as
 - A. Storage virtualization
 - B. Desktop virtualization
 - C. Distributed virtualization
 - D. Network virtualization

4. Which of the following cloud benefit corresponds to availability of existing applications even when changing computers or moving to a portable device?
 - A. Device independence
 - B. Version availability
 - C. Universal document access
 - D. Improved document format compatibility

5. Google AppEngine provides services for development and hosting of web applications on the distributed, dynamic and scalable infrastructure of Google. This corresponds to which of the following cloud service model?
 - A. SaaS
 - B. FaaS
 - C. PaaS
 - D. IaaS

Cloud Computing

6. Cloud offers various art applications for quickly and easily design attractive cards, booklets, & images. Which of the following is an example of cloud-based art applications?
 - A. Moo
 - B. Vistaprint
 - C. Adobe Creative Cloud
 - D. All of the above

7. A business application that offers an e-mail publishing platform is
 - A. Chatter
 - B. Quickbooks
 - C. MailChimp
 - D. None of the above

8. The provision of various applications such as online music/video, online games and video conferencing, streaming services, etc that can reach any device be it TV, mobile, set-top box, or any other form is now possible. This is termed as
 - A. On-demand infrastructure
 - B. On-demand surface
 - C. On-demand entertainment
 - D. Distributed surfacing

9. _____ corresponds to the network of physical objects or "things" embedded with electronics, software, sensors, and network connectivity, which enables these objects to collect and exchange data.
 - A. Internet-of-Transport
 - B. Internet-of-Things
 - C. Distributed network
 - D. Sensor network

10. In _____ communication, the devices are called "connected" or "smart" devices, if they can talk to other related devices.
 - A. machine-to-device (M2D)
 - B. device-to-machine (D2M)
 - C. machine-to-machine (M2M)
 - D. network-to-network (N2N)

11. Internet-of-Things is also called as
 - A. Internet of Everything (IoE)
 - B. Machine-to-Machine (M2M)
 - C. Skynet
 - D. All of the above

12. RFID in IoT stands for
 - A. Remote-frequency Identification
 - B. Radio-frequency Identification
 - C. Radio-frequency Introspection

Unit 13: Application of Cloud Computing and Related Technologies

D. Remote-frequency Introspection

13. _____ are uniquely identifiable devices with a unique IP address in IoT network.

- A. Applications
- B. Gateways
- C. Procedures
- D. Sensors

14. Which of the following is not a benefit of Edge computing?

- A. Saves bandwidth
- B. Increases congestion
- C. Reduces congestion
- D. Eliminates latency

15. _____ technology permits transactions to be gathered into blocks and recorded; allows the resulting ledger to be accessed by different servers.

- A. Vistra
- B. Bitcore
- C. Pesa
- D. Blockchain

Answers for Self Assessment

- | | | | | |
|-------|-------|-------|-------|-------|
| 1. D | 2. A | 3. B | 4. A | 5. C |
| 6. D | 7. C | 8. C | 9. B | 10. C |
| 11. D | 12. B | 13. D | 14. B | 15. D |

Review Questions

1. What is blockchain and its features?
2. Discuss the characteristics and benefits of blockchain technology?
3. What does Internet-of-Things (IoT) means?
4. Explain the structure and working of IoT?
5. What are the building blocks of IoT?
6. Indicate the different applications of IoT in various sectors?
7. What is IoT? Discuss its characteristics.
8. Differentiate distributed and centralized ledger?
9. What are the different types of blockchain?
10. Differentiate public and private blockchain?
11. What are public blockchain? List the advantages and disadvantages?
12. Differentiate edge computing and cloud computing.
13. Compare and contrast edge computing with the traditional computing concept.
14. Discuss the role of Cloud of Things and its applications?
15. What is fog computing and its benefits?



Further Readings

Internet of Things: Principles and Paradigms by Rajkumar Buyya, Amir Vahid Dastjerdi

Internet of Things by Raj Kamal, McGraw-Hill Education

Blockchain Revolution: How the Technology Behind Bitcoin and Other Cryptocurrencies is Changing the World by Don Tapscott, Alex Tapscott.

Fog Computing: Theory and Practice by Assad Abbas; Samee U. Khan; Albert Y. Zomaya, Wiley Telecom 2020.

Fog and Edge Computing: Principles and Paradigms by Satish Narayana Srirama, Rajkumar Buyya, Wiley 2019.



Web Links

[What is the Internet of Things? | IoT Technology \(sap.com\)](#)

[What is the Internet of Things, and how does it work? \(ibm.com\)](#)

[What Is Edge Computing | IBM](#)

[What Is Edge Computing: Definition, Characteristics, and Use Cases \(simplilearn.com\)](#)

[Edge Computing vs Cloud Computing | Difference between Edge Computing and Cloud Computing - javatpoint](#)

[Edge Computing vs Cloud Computing: Key Differences \(phoenixnap.com\)](#)

[GSMA | Introduction | Internet of Things](#)

[GSMA | GSMA Mobile IoT Initiatives | Licensed Low Power Wide Area Technology](#)

[The Benefits of Mobile IoT Networks \(koombea.com\)](#)

[What is Fog Computing? - Definition from IoTAgenza \(techtarget.com\)](#)

[What Is Fog Computing and How Does It Work? - Fingent Technology](#)

[Fog Computing - GeeksforGeeks](#)

[Fog Computing | An Overview Guide to Fog Computing \(educba.com\)](#)

[What is Blockchain Technology? A Step-by-Step Guide For Beginners \(blockgeeks.com\)](#)

[Applications of Blockchain | 10 Most Popular Application of Blockchain \(educba.com\)](#)

Unit 14: Additional Cloud Concepts

CONTENTS

Objectives

Introduction

14.1 Traditional Server Concept

14.2 Concept of Virtual Servers

14.3 Cloud Storage

14.4 Cloud Storage Vs Traditional Storage

14.5 Benefits of Cloud Storage

14.6 Considerations for Storing Data to Cloud

14.7 Cloud Database Storage

14.8 Benefits of Using a Cloud Database

14.9 Resource Abstraction in Cloud Computing

14.10 Resource Management in Cloud Computing

14.11 Service Level Agreements (SLAs) in Cloud Computing

14.12 Case Study on Cloud Services Offered by Amazon Web Services (AWS)

14.13 Amazon Elastic Compute Cloud (EC2)

Summary

Keywords

Self Assessment

Answers for Self Assessment

Review Questions

Further Readings

Objectives

After this lecture, you will,

- learn about cloud server and cloud storage.
- Know about database storage offered through cloud computing.
- reviewthe concepts of resource abstraction, resource scheduling and management in cloud computing.
- discover what is service level agreements in cloud computing.
- investigate different services offered by Amazon Web Services (AWS).

Introduction

A powerful physical or virtual infrastructure that performs application- and information-processing storage. It is created using virtualization software to divide a physical (bare metal) server into multiple virtual servers. The organizations use an infrastructure-as-a-service (IaaS) model to process workloads and store information. It can access virtual server functions remotely through an online interface. The key features of the cloud server include:

Cloud Computing

- Computing infrastructure that can be physical (bare metal), virtual, or a mix of the two depending on use case.
- Has all the capabilities of an on-premises server.
- Enables users to process intensive workloads and store large volumes of information.
- Automated services are accessed on demand through an API.
- Gives users the choice of monthly or as-you-go payment.
- Users can opt for a shared hosting plan that scales depending on needs.

Why Cloud Server

- Cost Effectiveness- With cloud servers, organizations only pay for what they need and reduce the expense that comes with maintaining server hardware.
- Scalability- Users can scale computing and storage resources to meet changing needs. This is particularly helpful for organizations with fluctuating needs.
- Integration- An organization's cloud servers are networked to ensure uninterrupted communication and fast deployment. A "single pane" enables complete control.

14.1 Traditional Server Concept

The system administrators often talk about servers as a whole unit that includes the hardware, the OS, the storage, and the applications. The servers are often referred to by their function, that is, the Exchange server, the SQL server, the File server, etc. (Figure 1). If the file server fills up, or the Exchange server becomes overtaxed, then the system administrators must add in a new server.

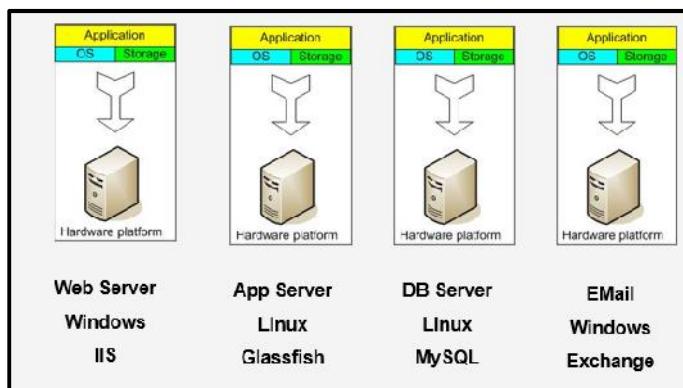


Figure 1: Server Concept

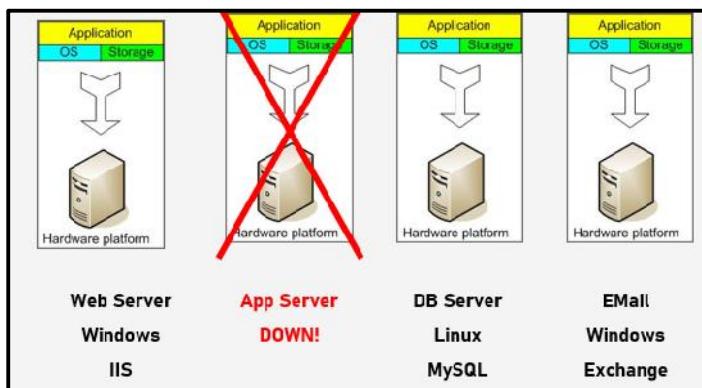


Figure 2: Scenario in which Server is Down

Unless there are multiple servers, if a service experiences a hardware failure, then the service is down (Figure 2). The system admins can implement clusters of servers to make them more fault

Unit 14: Additional Cloud Concepts

tolerant. However, even clusters have limits on their scalability, and not all applications work in a clustered environment.

Pros of Traditional Server Concept

- Easy to conceptualize.
- Fairly easy to deploy.
- Easy to backup.
- Virtually any application/service can be run from this type of setup.

Cons of Traditional Server Concept

- Expensive to acquire and maintain hardware
- Not very scalable
- Difficult to replicate
- Redundancy is difficult to implement
- Vulnerable to hardware outages
- In many cases, processor is under-utilized

14.2 Concept of Virtual Servers

The virtual servers seek to encapsulate the server software away from the hardware. It includes the OS, applications, and storage for that server (Figure 3). The servers end up as mere files stored on a physical box, or in enterprise storage. The virtual server can be serviced by one or more hosts, and one host may house more than one virtual server (Figure 4). The virtual servers can still be referred to by their function, that is, email server, database server, etc. If the environment is built correctly, virtual servers will not be affected by the loss of a host. The hosts may be removed and introduced almost at will to accommodate maintenance.

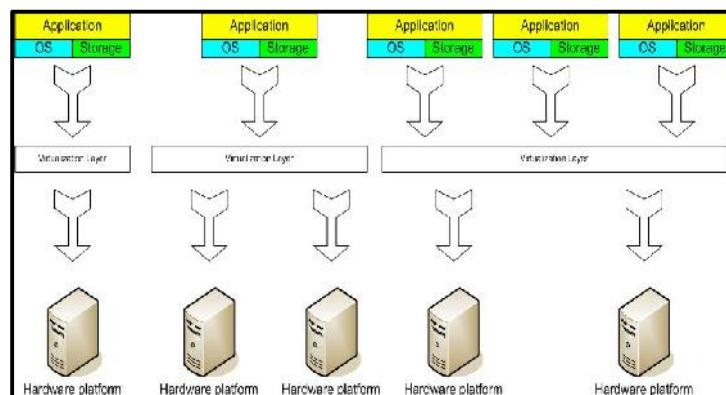


Figure 3: Virtual Servers Concept

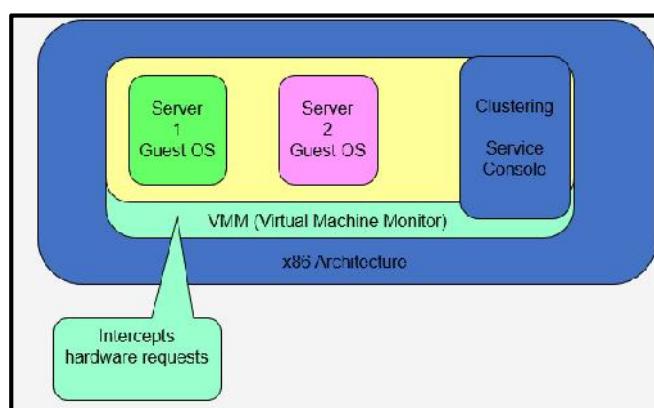


Figure 4: Within a Virtual Server

Cloud Computing

The virtual servers can be scaled out easily. If the administrators find that the resources supporting a virtual server are being taxed too much, they can adjust the number of resources allocated to that virtual server. The server templates can be created in a virtual environment to be used to create multiple, identical virtual servers. The virtual servers themselves can be migrated from host to host almost at will.

Pros of Using Virtual Server:

- o Resource pooling
- o Highly redundant
- o Highly available
- o Rapidly deploy new servers
- o Easy to deploy
- o Reconfigurable while services are running
- o Optimizes physical resources by doing more with less

Cons of Using Virtual Server:

- o Slightly harder to conceptualize
- o Slightly more costly (must buy hardware, OS, Apps, and now the abstraction layer)



Notes: Virtual Servers vs Physical Servers:
 1. Physical (bare metal) servers are best for data-intensive workloads.
 2. Virtual servers are better for highly variable workloads.

14.3 Cloud Storage

Cloud storage is the most innovative technology to store, access, and collaborate through scalable cloud technology. The cloud computing model that stores valuable data through the web and operates it through a storage as a device service.

Cloud storage is a model of computer data storage in which the digital data is stored in logical pools, said to be on "the cloud". The physical storage spans multiple servers (sometimes in multiple locations), and the physical environment is typically owned and managed by a hosting company. These cloud storage providers are responsible for keeping the data available and accessible, and the physical environment protected and running. People and organizations buy or lease storage capacity from the providers to store user, organization, or application data.

Cloud storage services may be accessed through a co-located cloud computing service, a web service application programming interface (API) or by applications that utilize the API, such as cloud desktop storage, a cloud storage gateway or Web-based content management systems. The cloud storage is based on highly virtualized infrastructure and is like broader cloud computing in terms of interfaces, near-instant elasticity and scalability, multi-tenancy, and metered resources. Cloud storage services can be utilized from an off-premises service (Amazon S3) or deployed on-premises (ViON Capacity Services).

Cloud storage typically refers to a hosted object storage service, but the term has broadened to include other types of data storage that are now available as a service, like block storage. The object storage services like Amazon S3, Oracle Cloud Storage and Microsoft Azure Storage, object storage software like Openstack Swift, object storage systems like EMC Atmos, EMC ECS and Hitachi Content Platform, and distributed storage research projects like OceanStore and VISION Cloud are all examples of storage that can be hosted and deployed with cloud storage characteristics ().

Figure 5: Different Cloud Storage Examples

Unit 14: Additional Cloud Concepts

Overall, the cloud storage is made up of many distributed resources, but still acts as one, either in a federated or a cooperative storage cloud architecture. It is highly fault tolerant through redundancy and distribution of data; highly durable through the creation of versioned copies and it eventually consistent with regard to the data replicas.

Comparatively, the traditional storage was done using the local physical drives to store the data at the primary location of the client and the user generally used the disk-based hardware to store data as well as for copying, managing, and integrating the data to software.

14.4 Cloud Storage Vs Traditional Storage

Storage Location

- Data (or files) are saved on a remote server, which is easily accessible from anywhere with internet access as well as from any device connected to the internet, including computers, tablets and smartphones.

Payment

- With traditional alternatives, what you pay for is what you get. If one has invested in a lot of storage to support a new project, then one is still left paying the same amount once it's over – even if one doesn't need the storage anymore.

Accessibility

- Businesses are bound to the traditional office-based 9-5 unlike cloud in which, "teams can access everything from wherever they are".
- Traditional storage solutions involve physical devices in which the monitoring, maintaining and patching the devices is user-dependent and thereby quite overwhelming.

Backup Options

- The better the backup plan, the faster the business can recover from disaster and be up and running as normal.
- Depending on the cause of the problem, it takes longer for traditional storage solutions to recover.

Flexibility

- Cloud storage is more flexible than traditional on-premise alternatives.
- Easy to create a tailored solution that suits a user's specific requirements.
- Freedom to choose a course of action based on a user's current setup and what servers are chosen.

14.5 Benefits of Cloud Storage

- All platforms can easily be accessed via a web browser.
- Offer apps for ease of access from a smartphone or tablet.
- Feature a directory structure similar to that of a computer drive; this facilitates navigation and organisation.
- Ease of Access
 - Access to the personal folders is perceived to be more cumbersome (involves 'more clicks').
 - Many people are always logged into Google (and hence Google Drive) in the background, both at home and at school
- Online Editing
 - OneDrive and Google Drive offer the possibility of editing documents inside a web browser.
 - No additional software is needed.
 - Folders or specific files can be shared with others; this facilitates collaboration.
- Online Collaboration
 - Documents and folders can be shared with colleagues.
 - Editing is possible without downloading the document, eliminating the need to email and save multiple versions of the same documents.

14.6 Considerations for Storing Data to Cloud

Various parameters that need to be taken into account whilst saving data on the cloud include:

- **Security-** As data is managed by third-party vendors, it is heavily encrypted when data is stored and moving through various points. These security systems work well with permissions and access controls that are pre-defined on the web.
- **Availability-** The data can be accessed whenever required. It mandatorily requires internet connectivity.



Examples of Cloud Storage

The examples of cloud storage and their key characteristics are listed below (as depicted in Figure 5):

1. Google Drive

- 'Pure' cloud computing service, with all the apps and storage found online.
- Can be used via desktop top computers, tablets like iPad or on smartphones.
- All of Google's services can be considered cloud -based: Gmail, Google Calendar, Google Voice etc.
- Microsoft's OneDrive: Similar to Google Drive.



Figure 5: Different Cloud Storage Examples

2. Dropbox

- Commonly used to store documents and images.
- One can set his/her phone to automatically send all pictures taken with it into their Dropbox account, so that even if one loses their phone, the pictures will still be available to him/her up in space.
- One can use it to access documents at home, and then save changes to it.
- Sugarsync is another example.

3. Apple iCloud

- Apple's cloud service is primarily used by Apple users for online storage and synchronization of user's mail, contacts, calendar, and more.
- All the data needed is available to a user on whichever device he/she seeks to access it from, iOS, Mac OS, or Windows device.
- If a user makes a change to a document, say, on one of their devices, it will automatically update it so that when next access is made to the account, the amended version will be available on whatever device you use.
- If a user has loads of data up there (perhaps pictures or films have made) then one will need to pay for extra storage.

14.7 Cloud Database Storage

The cloud database is a collection of informational content, either structured or unstructured that resides on a private, public or hybrid cloud computing infrastructure platform. From a structural and design perspective, a cloud database is no different than one that operates on a business's own on-premises servers. However, the critical difference lies in where database resides. "Cloud database" can be one of two distinct things: a traditional or NoSQL database installed and running on a cloud virtual machine, or a cloud provider's fully managed database-as-a-service (DBaaS) offering. It can be said to be running own self-managed database in a cloud environment where an on-premises database is connected to local users through a corporation's internal local area network (LAN), a cloud database resides on servers and storage furnished by a cloud or database as a service (DBaaS) provider and it is accessed solely through the internet. The cloud DBaaS is natural database equivalent of software-as-a-service (SaaS). Cloud database access is based on the pay as you go model.

How Cloud Database Works?

There are two prominent cloud databases: Relational and Non-Relational databases. The relational database is written in Structured Query Language (SQL) and is a set of interrelated tables organized into rows & columns. The relationship between tables and columns (fields) is specified in a schema. Such databases are used in banking transactions or a telephone directory. The popular cloud platforms and cloud providers include MySQL, Oracle, IBM DB2 and Microsoft SQL Server. The non-relational database is sometimes also called as NoSQL as it does not employ a table model. It stores content, regardless of its structure, as a single document and is well-suited for unstructured data, such as social media content, photos and videos.

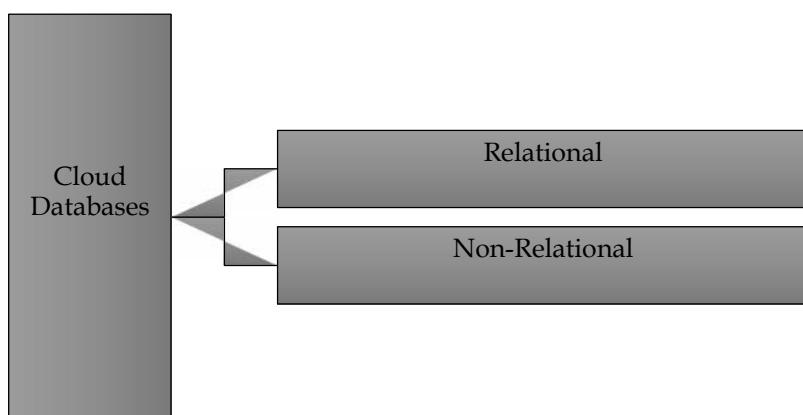


Figure 6: Categorization of Cloud Database

14.8 Benefits of Using a Cloud Database

- Ease of Access: Users can access cloud databases from virtually anywhere, using a vendor's API or web interface.
- Scalability: Cloud databases can expand storage capacities on run-time to accommodate changing needs. Organizations only pay for what they use.
- Disaster Recovery: In the event of a natural disaster, equipment failure or power outage, data is kept secure through backups on remote servers.
- Overall cost: The cost of adopting cloud databases can be considerably less than expanding your existing on-site server capability. There are reduced maintenance costs can also cut administrative costs substantially. Cloud's pay-as-you-go costs only increase if one expands or requires additional services.
- Flexible Solutions: Moving database solutions to the cloud can release a business from the demands and costs of managing own services. Cloud databases are massively efficient, as they have no inherent restrictions on their ability to expand.

Cloud Computing

- Mobile Access: Cloud databases add easiness in business expansion due to the ability of cloud platforms to be accessed and used from a range of remote devices. The applications can be built with geographically dispersed teams at no loss to efficiency or security.
- Disaster Recovery: The applications require reliable connections to the databases that power them. There is built-in redundancy and 24/7 uptime the norm. Cloud offers reliable platform for application development. Moreover, the robust cloud infrastructures are supported.
- Safe and Secure: Moving sensitive data to a cloud platform outside of a business's firewall could be risky. Cloud offers comprehensive security often more robust than that of on-site servers. The adoption of DBaaS infrastructures delivers world-class security.
- Scaling and Managing Database is Easier: Cloud service providers keep on evolving their services, any business can take instant advantage of these improvements, making both scaling and managing a database easier.

Considerations for Moving Database to Cloud

What is current database hosting infrastructure?

- Whether your business runs traditional, relational databases or has embraced NoSQL databases, it's essential to match your hosting environment to the right cloud infrastructure.
- Using DBaaS (Database-as-a-Service).
- Hosting services should be efficient and secure to migrate to and offer a flexible dynamic environment for your applications.
- Hosting environment should have the capacity and features to match these needs.

Are you expanding the capabilities of the applications you are creating?

- As your applications grow in size and complexity, a cloud database can grow along with it.
- Cloud databases have ability to handle rapid expansion and organization of complex data.
- Hosting these databases in the cloud lifts the limits on your application's growth.

Is 24/7 availability essential for your business's applications?

- Cloud databases have built-in back-up and recovery, ensuring database is 'always on.'
- Built-in back-ups help eliminate the risk of data-loss.
- Moving to a cloud database removes the risk of downtime issues.
- Offers steady and reliable connectivity.

Does your business build applications require large datasets to operate efficiently?

- Using a cloud database removes the issues of dealing with large datasets by giving access to data storage that expands to meet user needs.

Example of Cloud Database Service- MongoDB Atlas Cloud Database

MongoDB Atlas, part of MongoDB's broader Data-as-a-Service (DaaS) development platform. It is a powerful and compelling alternative to managing own NoSQL, or traditional database, or to using a cloud provider-specific managed offering. It offers fully managed database services through user's choice of cloud provider, including AWS, Azure, and GCP. The fully-managed database services handle the complexities of maintaining a consistently available, high-performance data cluster that developers can access as a single, globally available resource.

14.9 Resource Abstraction in Cloud Computing

Resource abstraction can be considered as a key feature where the resources are defined independently from the target OS. In other words, while writing any manifest file the user need not worry about the target machine or the OS, which is present on that particular machine.

- An abstraction layer simplifies the creation and use of Virtual Machines (VM) in cloud, and also make interoperability between providers from the view of users.
- Abstraction also enables opportunities for creating optimization layer (substituting, brokering)between the abstraction layer and cloud middleware.

Unit 14: Additional Cloud Concepts

The abstraction of cloud resources involves: Resources in the clouds (VMs, images, storages) are abstracted as objects and users can manipulate with them via methods provided by the objects. The resource abstraction allows changes in the backend without affecting functionalities and modification of developed applications in the abstraction layer.

The abstraction of complex systems involves: Done via mechanisms like inheritance, composition and polymorphisms, developers can make abstraction of more complex systems with several components easily, and deploy them with single click.

14.10 Resource Management in Cloud Computing

Resource management in the cloud computing is the process of allocating computing, storage, networking and indirectly energy resources to a set of applications. It attributes to jointly meet the performance objectives of the infrastructure providers, users of the cloud resources and applications as the major objective of the cloud users tend to focus on application performance. The resource management is a conceptual framework that provides a high-level view of the functional component of cloud resource management systems and all their interactions and is highly critical function that requires complex policies and decisions for multi-objective optimization. It is challenging as the complexity of the system makes it impossible to have accurate global state information.

The management of resources in cloud is highly affected by unpredictable interactions with the environment, e.g., system failures, attacks and is faced with large fluctuating loads which challenge the claim of cloud elasticity. There is extensive use of resource scheduling that decides how to allocate resources of a system, such as CPU cycles, memory, secondary storage space, I/O and network bandwidth, between users and tasks. Certain policies and mechanisms are used for resource allocation where the policies pertain to the principles guiding decisions whilst the mechanisms are the means to implement policies.

Resource Management Policies

- Admission control: To prevent system from accepting workload in violation of high-level system policies.
- Capacity allocation: Allocate resources for individual activations of a service.
- Load balancing: Distribute workload evenly among servers.
- Energy optimization: Minimization of energy consumption.
- Quality of service (QoS) guarantees: Ability to satisfy timing or other conditions specified by a SLA.

Resource Management Mechanisms

- Control theory: Uses the feedback to guarantee system stability and predict transient behavior.
- Machine learning: Does not need a performance model of the system.
- Utility-based: Require a performance model and a mechanism to correlate user-level performance with cost.
- Market-oriented/economic: Do not require a model of the system, e.g., combinatorial auctions for bundles of resources.

14.11 Service Level Agreements (SLAs) in Cloud Computing

Service Level Agreement (SLA) is the bond for performance negotiated between the cloud services provider and the client. Earlier, in cloud computing all Service Level Agreements were negotiated between a client and the service consumer. Nowadays, with the initiation of large utility-like cloud computing providers, most Service Level Agreements are standardized until a client becomes a large consumer of cloud services. A properly drafted and well-thought SLA should:

- State business objectives to be achieved in the provision of the services.
- Describe in detail the service deliverables.
- Define performance standards the customer expects in the provision of services by the service provider.

Cloud Computing

- Provide an on-going reporting mechanism for measuring the expected performance standards.

What should be included in an SLA?

- Provide a remedial mechanism and compensation regime where performance standards are not achieved, whilst incentivizing the service provider to maintain a high level of performance.
- Provide a mechanism for review and change to the service levels over the course of the contract.
- Give the customer the right to terminate the contract where performance standards fall consistently below an acceptable level.

Elements of Good Service Level Agreement (SLA)

- **Description of the Services:** The SLA should include a detailed description of the services. Each individual service should be defined, that is, there should be a description of what the service is, where it is to be provided, to whom it is to be provided and when it is required.
- **Overall objectives:** The SLA should set out the overall objectives for the services to be provided. For example, if the purpose of having an external provider is to improve performance, save costs or provide access to skills and/or technologies which cannot be provided internally, then the SLA should say so. This will help the customer craft the service levels in order to meet these objectives and should leave the service provider in no doubt as to what is required and why.
- **Critical Failure:** Service credits are useful in getting the service provider to improve its performance, but what happens when service performance falls well below the expected level? If the SLA only included a service credit regime then, unless the service provided was so bad as to constitute a material breach of the contract as a whole, the customer could find itself in the position of having to pay (albeit at a reduced rate) for an unsatisfactory overall performance.
- **Performance Standards:** Then, taking each individual service in turn, the customer should state the expected standards of performance. This will vary depending on the service. Using the “reporting” example referred to above, a possible service level could be 99.5%. However, this has to be considered carefully. Often a customer will want performance standards at the highest level. Whilst understandable, in practice this might prove to be impossible, unnecessary or very expensive to achieve.
- **Compensation/Service Credits:** In order for the SLA to have any “bite”, failure to achieve the service levels needs to have a financial consequence for the service provider. This is most often achieved through the inclusion of a service credit regime. In essence, where the service provider fails to achieve the agreed performance standards, the service provider will pay or credit the customer an agreed amount which should act as an incentive for improved performance.

Other Provisions in SLA

Changes to Pricing: Pricing may need to vary depending on a number of factors. SLA can include a pricing review mechanism or provisions dealing with the sharing of cost savings.

Contract Management: In longer term contracts, the parties will need to keep performance of the services under review. The provisions dealing with reporting, meetings, information provision and escalation procedures for disputes. It should be noted that it is vital if contract management procedures are agreed and are actually followed.

Change Control Procedures: The control procedures include the mechanisms for agreeing and recording changes to the agreement or services to be provided. In an agreement of any length or complexity, it is inevitable that changes will be made to the services. However, the properly implemented change control procedure is vital. Also, service credits and the right to terminate are main provisions in an SLA.

Sample Service Level Agreements (SLAs)

Windows Azure SLA: Window Azure has different SLA's for compute and storage. For compute, there is a guarantee that when a client deploys two or more role instances in separate fault and upgrade domains, client's internet facing roles will have external connectivity minimum 99.95% of the time. Moreover, all of the role instances of the client are monitored and there is guarantee of detection 99.9% of the time when a role instance's process is not runs and initiates properly.

SQL Azure SLA-SQL Azure clients will have connectivity between the database and internet gateway of SQL Azure. SQL Azure will handle a "Monthly Availability" of 99.9% within a month. Monthly Availability Proportion for a particular tenant database is the ratio of the time the database was available to customers to the total time in a month. Time is measured in some intervals of minutes in a 30-day monthly cycle. Availability is always remunerated for a complete month. A portion of time is marked as unavailable if the customer's attempts to connect to a database are denied by the SQL Azure gateway.

14.12 Case Study on Cloud Services Offered by Amazon Web Services (AWS)

Infrastructure as a Service (IaaS) includes all the cloud services with the capability to provision processing, storage, intra-cloud network connectivity services, and other fundamental computing resources of the cloud infrastructure. The Amazon Web Services (AWS) provides more of infrastructure as a service. IaaS, particularly involves providing infrastructure components as service like the virtual machines, servers, computational resources, networking components, load balancers. AWS provides the capability to provision, and to process storage intra cloud network connectivity services and other fundamental computing resources of the cloud infrastructure as involved. Platform-as-a-Service involves an Integrated Development Environment (IDE) that are provided to the users or the clients. Basically, IaaS facilitates on demand computing resources, eliminates the need for any head planning to be done, no upfront commitments, the user can actually start small and grow as in when they require it, pay for what is used. It involves Pay as you Go Model where a user pays for what is used such as CPU cycles. Moreover, there is no maintenance cost that is involved. A user is not concerned with maintaining the cloud services that are being offered. That is, these actually the headache of the service provider, the service vendor will be providing measured service as it is almost similar data what you are consuming, how much you are consuming everything is measured in that scenario, then it is quite scalable, and it is quite reliable in nature. Amazon Elastic Compute Cloud (EC2) is an example of Amazon's IaaS services.

14.13 Amazon Elastic Compute Cloud (EC2)

Amazon Elastic Compute Cloud (EC2) is a web service that provides resizable computing capacity that one uses to build and host different software systems. It has been designed to make web-scale computing easier for developers. A user can create, launch, and terminate server instances as needed, paying by the hour for active servers, hence the term "elastic". It provides scalable, pay as-you-go compute capacity. The Elastic here pertains to scaling in both directions.

Amazon Machine Images (AMI): An immutable representation of a set of disks that contain an operating system, user applications and/or data. From an AMI, one can launch multiple instances, which are running copies of the AMI (Figure 7).

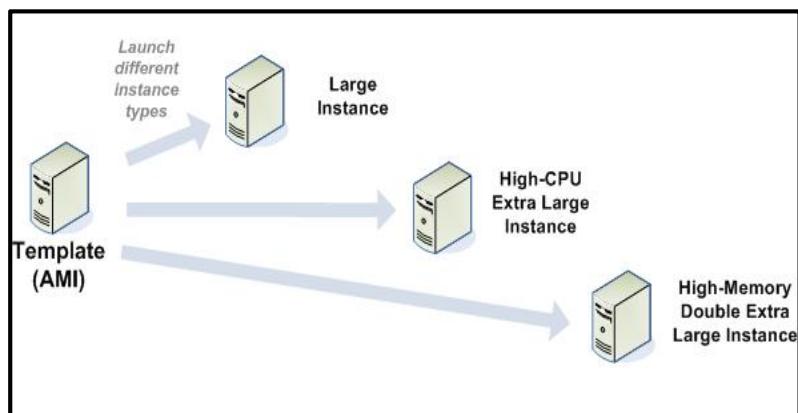
Cloud Computing

Figure 7: Amazon Machine Image

Amazon Machine Image (AMI) is a template for software configuration (Operating System, Application Server, and Applications) (depicted in Figure 8). An instance is a AMI running on virtual servers in the cloud. Each instance type offers different compute and memory facilities.

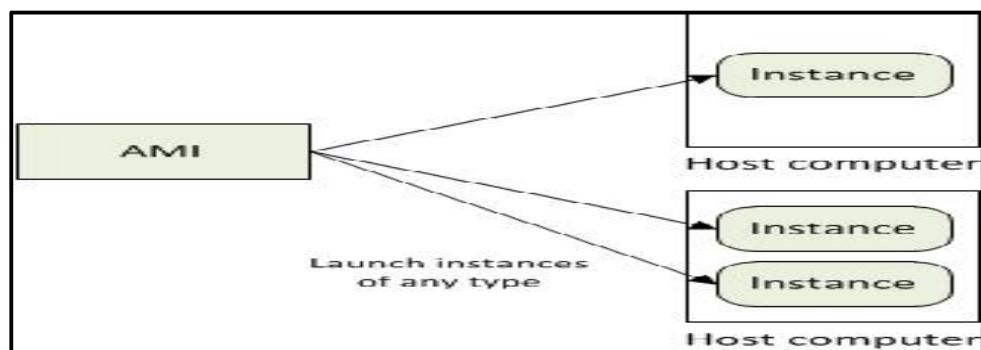


Figure 8: AMI Scenario

Figure 9 depicts the AWS EC2 console and different services (Figure 10) offered by Amazon.

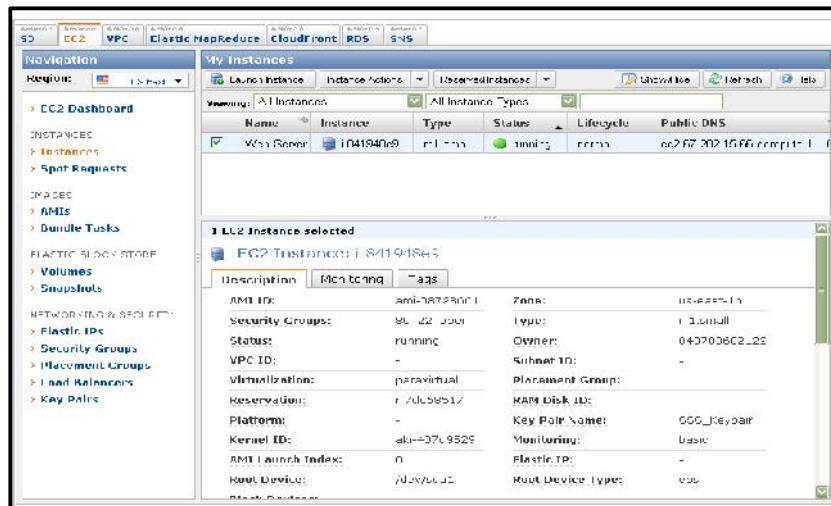


Figure 9: EC2 Console

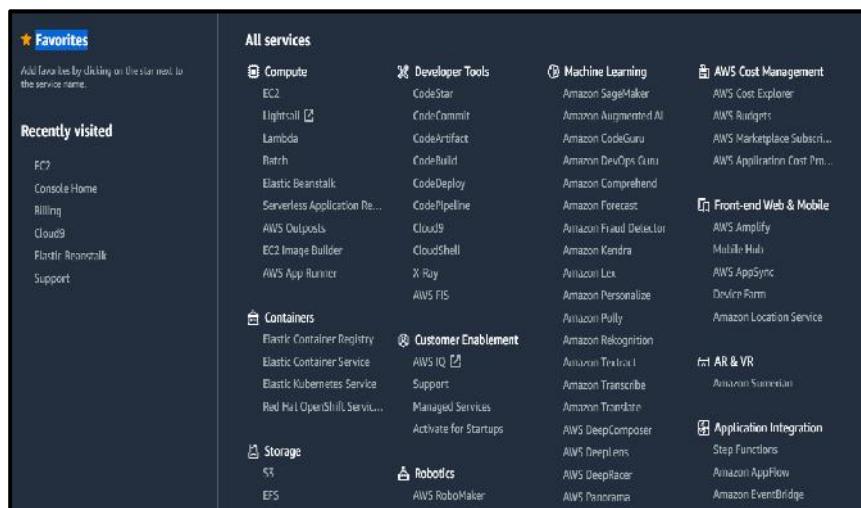


Figure 10: AWS Services

Region and Zones: Amazon have data centers in different regions across the globe. An instance can be launched in different regions depending on the need. Each region has set of zones where the zones are isolated from failure in other zones. It supports inexpensiveness, low latency connectivity between zones in the same region.

Storage:Amazon EC2 provides three type of storage options:Amazon EBS, Amazon S3, and Instance Storage.

Networking and Security: Instances can be launched on one of the two platforms.

- EC2-Classic
 - EC2-VPC

Each instance launched is assigned two addresses a private address and a public IP address.A replacement instance has a different public IP address.

- Instance IP address is dynamic: New IP address is assigned every time instance is launched
 - Amazon EC2 offers Elastic IP addresses (static IP addresses) for dynamic cloud computing

Remap the Elastic IP to new instance to mask failure

Separate pool for EC2-Classic and VPC

- Security Groups to access control to instances

Monitoring, Auto Scaling and Load Balancing: It involves monitoring the statistics of instances and EBS. Tools such as CloudWatch (monitor, manage and publish various metrics; configure alarms based on metrics). It automatically scales amazon EC2 capacity up and down based on rules such as add and remove compute resource based on demand; suitable for businesses experiencing variability in usage. It distributes incoming traffic across multiple instances supporting the elastic load balancing.

Amazon S3

In Amazon S3, the S3 stands for = Simple Storage Service. It is a Service Oriented Architecture (SOA) which provides online storage using web services. It allows read, write and delete permissions on the objects. It uses REST and SOAP protocols for messaging.

Amazon SimpleDB

Amazon SimpleDB is a highly available, flexible, and scalable non-relational data store that offloads the work of database administration. It creates and manages multiple geographically distributed replicas of your data automatically to enable high availability and data durability. The service

Cloud Computing

charges you only for the resources actually consumed in storing your data and serving your requests.

AWS Lambda

AWS Lambda is a Truly Serverless environment. You no longer have the overhead of administration. There is no need to manage any servers. The focus is entirely on your business logic.

AWS DynamoDB

- Database that delivers single-digit millisecond performance at any scale.
- Fully managed, cost effective, durable, multi-region NoSQL database.

AWS Amazon API Gateway

- Fully managed service.
- Simple way to create, publish, maintain, monitor, and secure APIs at any scale. RESTful APIs and WebSocket APIs are types of APIs you can create using API Gateway.

Summary

- Cloud computing signifies a major change in the way we run various applications and store our information. Everything is hosted in the “cloud”, a vague assemblage of computers and servers accessed via the Internet, instead of the method of running programs and data on a single desktop computer.
- The “cloud” itself is the key to the definition of cloud computing. The cloud is usually defined as a large group of interconnected computers. These computers include network servers or personal computers.
- With cloud storage, data is stored on multiple third-party servers, rather than on the dedicated servers used in traditional networked data storage.
- Resource management in the cloud computing is the process of allocating computing, storage, networking and indirectly energy resources to a set of applications.
- A Service Level Agreement (SLA) is the bond for performance negotiated between the cloud services provider and the client.
- The control procedures in SLAs include the mechanisms for agreeing and recording changes to the agreement or services to be provided. In an agreement of any length or complexity, it is inevitable that changes will be made to the services.
- The non-relational database is sometimes also called as NoSQL as it does not employ a table model.

Keywords

- **Cloud:** The cloud is usually defined as a large group of interconnected computers. These computers include network servers or personal computers.
- **Cloud Database:** Cloud database is a collection of informational content, either structured or unstructured that resides on a private, public or hybrid cloud computing infrastructure platform.

Unit 14: Additional Cloud Concepts

- **Relational cloud database:** The relational database is written in Structured Query Language (SQL) and is a set of interrelated tables organized into rows & columns.
- **Cloud Storage:** Cloud storage is a model of computer data storage in which the digital data is stored in logical pools, said to be on "the cloud".
- **Cloud Server:** Cloud server is a powerful physical or virtual infrastructure that performs application- and information-processing storage. It is created using virtualization software to divide a physical (bare metal) server into multiple virtual servers.
- **Infrastructure-as-a-Service (IaaS):** Infrastructure-as-a-Service includes all the cloud services with the capability to provision processing, storage, intra-cloud network connectivity services, and other fundamental computing resources of the cloud infrastructure.
- **Amazon Machine Image (AMI):** Amazon Machine Image (AMI) is a template for software configuration (Operating System, Application Server, and Applications).
- **Instance:** An instance is an AMI running on virtual servers in the cloud. Each instance type offers different compute and memory facilities.

Self Assessment

1. Which of the following offer online editing and storage service using cloud capabilities?
 - A. Google Drive
 - B. Zoho
 - C. Zoom
 - D. None of the above

2. What level of basic amount of free storage is available through Google Drive?
 - A. 5 GB
 - B. 10 GB
 - C. 15 GB
 - D. 1 TB

3. NoSQL databases are also called as
 - A. Relational Databases
 - B. Sequel Databases
 - C. Non-relational Databases
 - D. Table Databases

4. Which of the following is/are feature(s) of cloud server?
 - A. Computing infrastructure that can be physical (bare metal), virtual, or a mix of the two depending on use case.
 - B. Automated services are accessed on demand through an API.
 - C. Enables users to process intensive workloads and store large volumes of information.
 - D. All of the above

5. _____ servers seek to encapsulate the server software away from the hardware.
 - A. Cluster
 - B. Virtual
 - C. Consolidate
 - D. Dispersion

Cloud Computing

6. Nonrelational database are well-suited for _____.
 - A. structured data
 - B. distributed data
 - C. unstructured data
 - D. none of the above

7. _____ decides how to allocate resources of a system, such as CPU cycles, memory, secondary storage space, I/O and network bandwidth, between users and tasks.
 - A. Resource Utilization
 - B. Resource Scheduling
 - C. Resource Bandwidth
 - D. Resource Latency

8. Which of the following resource management policies specifies prevention of the system from accepting workload in violation of high-level system policies?
 - A. Load balancing
 - B. Capacity Allocation
 - C. Quality of Service
 - D. Admission Control

9. _____ enables opportunities for creating optimization layer between the abstraction layer and cloud middleware.
 - A. Brokering
 - B. Sectioning
 - C. Abstraction
 - D. Middle setting

10. Which of the following is not true for Service Level Agreements (SLA's)?
 - A. SLA measures the service provider's performance and quality.
 - B. SLA defines what the customer will not receive.
 - C. SLA is a contract between a service provider (either internal or external) and the end user.
 - D. SLA is a bond for performance negotiated.

11. In SLA, the customer has the right to terminate the contract where performance standards are consistently _____ an acceptable level.
 - A. below
 - B. above
 - C. close
 - D. none of the above

12. EC2 stands for?
 - A. Elastic Compute Cloud
 - B. Elastic Configuration Cloud
 - C. Elastic Cloud Configuration
 - D. Elastic Cloud Compute

Unit 14: Additional Cloud Concepts

13. Which of the following simply means to hardware or software load over web servers, that improver's the efficiency of the server as well as the application?
- Amazon Cloud-front
 - Elastic Load Balancer
 - Load Balancing
 - Security Management
14. What is a virtual server platform that allows users to create and run virtual machines on Amazon's server farm?
- Azure
 - EC2
 - EC3
 - EC5
15. _____ is an immutable representation of a set of disks that contain an operating system, user applications and/or data.
- Amazon Cloud9
 - Amazon CloudFront
 - Amazon Disk
 - Amazon Machine Image

Answers for Self Assessment

- | | | | | |
|-------|-------|-------|-------|-------|
| 1. A | 2. C | 3. C | 4. D | 5. B |
| 6. C | 7. B | 8. D | 9. C | 10. B |
| 11. A | 12. A | 13. C | 14. B | 15. D |

Review Questions

- Explain the difference between cloud and traditional storage?
- What are SLAs? What are the elements of good SLA?
- What is resource management in cloud computing?
- Write a short note on:
 - Resource abstraction in cloud computing
 - Amazon EC2
 - Amazon S3
- Differentiate Relational and Non-relation cloud database?
- How cloud storage works? What are different examples of cloud storage currently?
- What are the different services offered by Amazon?
- Discuss in detail atleast five Amazon Web Services.



Further Readings

Cloud Computing: Concepts, Technology and Architecture by Erl, Pearson Education.

Cloud Computing Black Book by Kailash Jayaswal,Jagannath Kallakurchi, Donald J. Houde, Deven Shah, Kogent Learning Solutions, DreamTech Press.

The Complete Guide to I.T. Service Level Agreements: Aligning It Services to Business Needs (Service Level Management) by Andrew N. Hiles, Rothstein Publishing 2003.

Resource Management in Utility and Cloud Computing by Han Zhao,Xiaolin Li, Springer 2013.



Web Links

[Cloud Computing Overview - Tutorialspoint](#)

[Cloud Computing \(w3schools.in\)](#)

[Cloud Management in Cloud Computing - GeeksforGeeks](#)

[Cloud Computing Management \(tutorialspoint.com\)](#)

[Service level agreements in Cloud computing - GeeksforGeeks](#)

[Service Level Agreement in Cloud Computing \(wright.edu\)](#)

[Fundamentals of Drafting Service Level Agreements \(SLAs\) \(lawsenate.com\)](#)

[Free Cloud Computing Services - AWS Free Tier \(amazon.com\)](#)

[Amazon Web Services \(AWS\) \(w3schools.in\)](#)

LOVELY PROFESSIONAL UNIVERSITY

Jalandhar-Delhi G.T. Road (NH-1)

Phagwara, Punjab (India)-144411

For Enquiry: +91-1824-521360

Fax.: +91-1824-506111

Email: odl@lpu.co.in

ISBN 978-93-94068-33-9



9 789394 068339