SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

SCHOOL OF COMPUTING

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

UNIT – I – Cloud Computing – SCSA7023

Unit – I

Understanding Cloud Computing

Topics:

Introduction -History of Cloud Computing-Cloud computing architecture-NIST architectural Framework-Types of Cloud-Service Based - Location Based-Merits and Demerits of Cloud Computing-Difference: Cloud and Web 2.0-Key challenges in cloud computing-Major cloud players-Cloud deployment model-Virtualization in cloud computing-Types of virtualization-Parallelization in cloud computing-Cloud resource management-Dynamic resource allocation-Optimal allocation of cloud models-Web services-Key Concepts

1.1. Introduction to Cloud Computing

Cloud computing is the on-demand availability of computer system resources, especially data storage (cloud storage) and computing power, without direct active management by the user. The term is generally used to describe data centers available to many users over the Internet. Large clouds, predominant today, often have functions distributed over multiple locations from central servers. If the connection to the user is relatively close, it may be designated an edge server.

1.2 History of Cloud computing

Cloud computing has as its antecedents both client/server computing and peer-to-peer distributed computing. It's all a matter of how centralized storage facilitates collaboration and how multiple computers work together to increase computing power.

During the 1960s, the initial concepts of time-sharing became popularized via RJE (Remote Job Entry) this terminology was mostly associated with large vendors such as IBM and DEC. Full-time-sharing solutions were available by the early 1970s on such platforms as Multics (on GE hardware), Cambridge CTSS, and the earliest UNIX ports (on DEC hardware). Yet, the "data center" model where users submitted jobs to operators to run on IBM mainframes was overwhelmingly predominant.

In the 1990s, telecommunications companies, who previously offered primarily dedicated point-to-point data circuits, began offering virtual private network (VPN) services with comparable quality of service, but at a lower cost. By switching traffic as they saw fit to balance server use, they could use overall network bandwidth more effectively. They began to use the cloud symbol to denote the demarcation point between what the provider was responsible for and what users were responsible for. Cloud computing extended this boundary to cover all servers as well as the network infrastructure. As computers became more diffused, scientists and technologists explored ways to make large-scale computing power available to more users through time-sharing. They experimented with algorithms to optimize the infrastructure, platform, and applications to prioritize CPUs and increase efficiency for end users.

In August 2006, Amazon created subsidiary Amazon Web Services and introduced its Elastic Compute Cloud (EC2). In April 2008, Google released the beta version of Google App Engine.

In early 2008, NASA's OpenNebula, enhanced in the RESERVOIR European Commission-funded project, became the first open-source software for deploying private and hybrid clouds, and for the federation of clouds.

By mid-2008, Gartner saw an opportunity for cloud computing "to shape the relationship among consumers of IT services, those who use IT services and those who sell them" and observed that "organizations are switching from company-owned hardware and software assets to per-use service-based models" so that the "projected shift to computing will result in dramatic growth in IT products in some areas and significant reductions in other areas."

In 2008, the U.S. National Science Foundation began the Cluster Exploratory program to fund academic research using Google-IBM cluster technology to analyze massive amounts of data.

In February 2010, Microsoft released Microsoft Azure, which was announced in October 2008.

In July 2010, Rackspace Hosting and NASA jointly launched an open-source cloud-software initiative known as OpenStack. The OpenStack project intended to help organizations offering cloud-computing services running on standard hardware. The early code came from NASA's Nebula platform as well as from Rackspace's Cloud Files platform. As an open source offering and along with other open-source solutions such as CloudStack, Ganeti and OpenNebula, it has attracted attention by several key communities. Several studies aim at comparing these open sources offerings based on a set of criteria.

On March 1, 2011, IBM announced the IBM SmartCloud framework to support Smarter Planet. Among the various components of the Smarter Computing foundation, cloud computing is a critical part. On June 7, 2012, Oracle announced the Oracle Cloud. This cloud offering is poised to be the first to provide users with access to an integrated set of IT solutions, including the Applications (SaaS), Platform (PaaS), and Infrastructure (IaaS) layers.

In May 2012, Google Compute Engine was released in preview, before being rolled out into General Availability in December 2013.

In 2019, it was revealed that Linux is most used on Microsoft Azure.

1.3 Cloud Computing Architecture

1.3.1 NIST Architectural Framework - The NIST model

The United States government is a major consumer of computer services and, therefore, one of the major users of cloud computing networks. The U.S. National Institute of Standards and technology.

The National Institute of Standards and Technology(NIST) model originally did not require a cloud to use virtualization to pool resources, nor did it absolutely require that a cloud support multi-tenancy in the earliest definitions of cloud computing. Multi-tenancy is the sharing of resources among two or more clients. The latest version of the NIST definition does require that cloud computing networks use virtualization and support multi-tenancy.

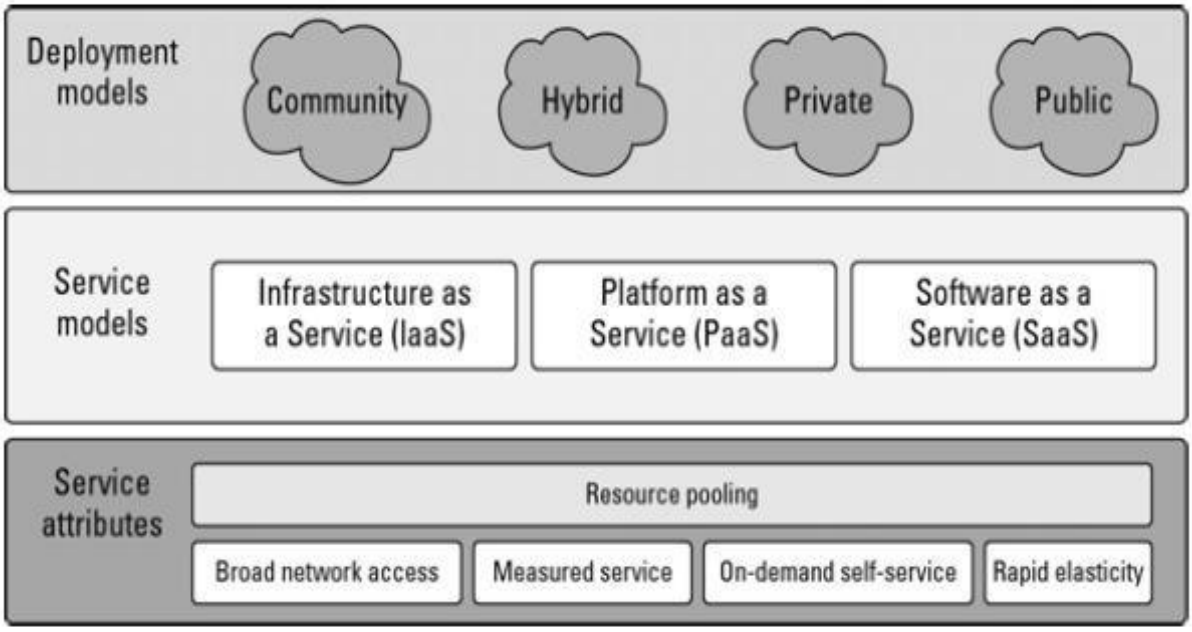


Fig 1.1 NIST Architectural Framework

The NIST cloud computing definitions

Because cloud computing is moving toward a set of modular interacting components based on standards such as the Service Oriented Architecture(SOA), we might expect that future versions of the NIST model may add those features as well. The NIST cloud model doesn't address a number of intermediary services such as transaction or service brokers, provisioning, integration, and interoperability services that form the basis for many cloud computing paradigms.

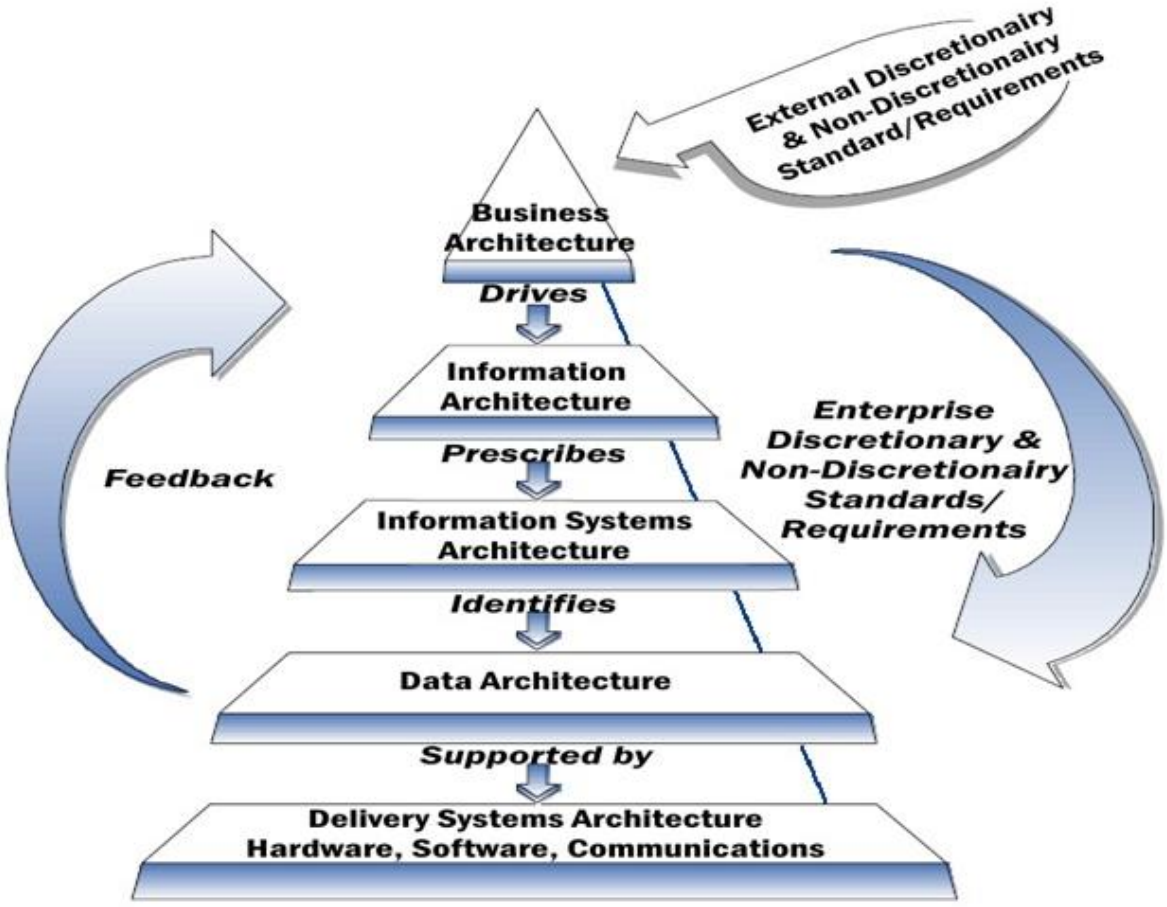


Fig 1.2 NIST Architecture Levels

NIST Architecture Levels

- Business Architecture level: This level can picture the total or a subunit of any corporation, which are in contact with external organizations.
- Information architecture level: This level specifies types of content, presentation forms, and format of the information required.
- Information systems architecture level: Specifications for automated and procedure-oriented information systems.
- Data Architecture level: Framework for maintenance, access and use of data, with data dictionary and other naming conventions.
- Data Delivery Systems level: Technical implementation level of software, hardware, and communications that support the data architecture.

1.4 Types of Cloud

Cloud models are classified into two major categories. They are Location based as Public, Private, Hybrid and Service Based as Software as a Service, Platform as a Service and Infrastructure as a Service.

1.4.1 Service Based Cloud Types

Cloud computing services are divided into three classes, according to the abstraction level of the capability provided and the service model of providers, namely: (1) Infrastructure as a Service, (2) Platform as a Service, and (3) Software as a Service

There are many different service models described in the literature, all of which take the following form:

'X'aaS, or "<Something> as a Service"

Three service types have been universally accepted:

Infrastructure as a Service:

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

The IaaS service provider manages all the infrastructure, while the client is responsible for all other aspects of the deployment. This can include the operating system, applications, and user interactions with the system.

Platform as a Service:

PaaS provides virtual machines, operating systems, applications, services, development frameworks, transactions, and control structures. The client can deploy its 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.

Software as a Service:

SaaS is a complete operating environment with applications, management, and the user interface. In the SaaS model, the application is provided to the client through a thin client interface (a

browser,usually), and the customer's responsibility begins and ends with entering and managing its data and user interaction. Everything from the application down to the infrastructure is the vendor's responsibility.

The three different service models taken together have come to be known as the SPI model of cloud computing. It is useful to think of cloud computing's service models in terms of a hardware/software stack. At the bottom of the stack is the hardware or infrastructure that comprises the network. As we move upward in the stack, each service model inherits the capabilities of the service model beneath it. IaaS has the least levels of integrated functionality and the lowest levels of integration, and SaaS has the most. Examples of IaaS service providers include:

- Amazon Elastic Compute Cloud (EC2)
- Eucalyptus
- GoGrid
- FlexiScale
- Linode
- RackSpace Cloud
- Terremark

All these vendors offer direct access to hardware resources. On Amazon EC2, considered the classic IaaS example, a client would provision a computer in the form of a virtual machine image, provision storage, and then go on to install the operating system and applications onto that virtual system Amazon has a number of operating systems and some enterprise applications that they offer on a rental basis to customers in the form of a number of canned images, but customers are free to install whatever software they want to run. Amazon's responsibilities as expressed in its Service Level Agreement, which is published on Amazon's Web site, contractually obligates Amazon to provide a level of performance commensurate with the type of resource chosen, as well as a certain level of reliability as measured.

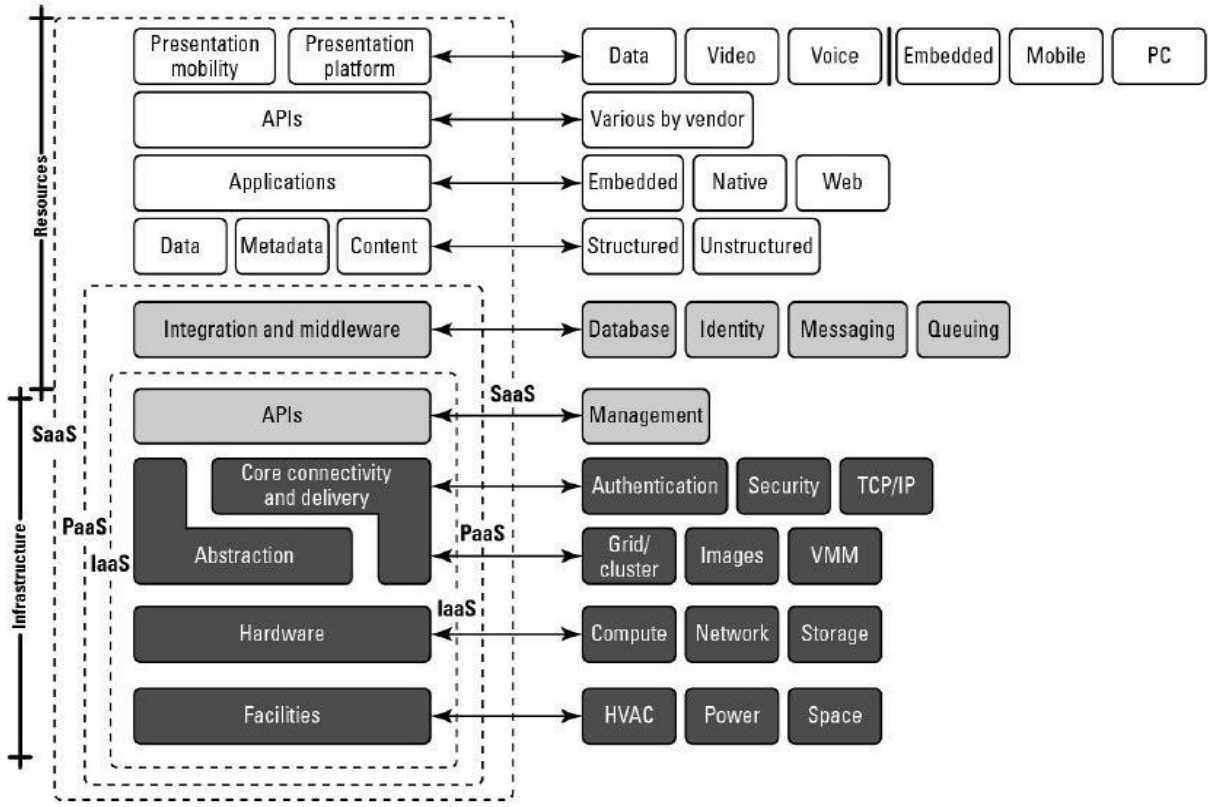


Fig 1.3 The Cloud Reference Model

A PaaS service adds integration features, middleware, and other orchestration and choreography services to the IaaS model. Examples of PaaS services are:

- Force.com
- GoGrid CloudCenter
- Google AppEngine
- Windows Azure Platform

When a cloud computing vendor offers software running in the cloud with use of the application on a pay-as-we-go model, it is referred to as SaaS. With SaaS, the customer uses the application as needed and is not responsible for the installation of the application, its maintenance, or its upkeep

1.4.2 Location Based cloud types

Cloud deployment describes the way a cloud platform is implemented, how it's hosted, and who has access to it. All cloud computing deployments operate on the same principle by virtualizing the computing power of servers into segmented, software-driven applications that provide processing and storage capabilities.

Public Cloud

Some public cloud examples include those offered by Amazon, Microsoft, or Google. These companies provide both services and infrastructure, which are shared by all customers. Public clouds typically have massive amounts of available space, which translates into easy scalability. A public cloud is often recommended for software development and collaborative projects. Companies can design their applications to be portable, so that a project that's tested in the public cloud can be moved to the private cloud for production. Most cloud providers package their computing resources as part of a service. Public cloud examples range from access to a completely virtualized infrastructure that provides little more than raw processing power and storage (Infrastructure as a Service, or IaaS) to specialized software programs that are easy to implement and use (Software as a Service, or SaaS).

The great advantage of a public cloud is its versatility and "pay as we go" structure that allows customers to provision more capacity on demand. On the downside, the essential infrastructure and operating system of the public cloud remain under full control of the cloud provider. Customers may continue to use the platform under the terms and conditions laid out by the provider, but they may have difficulty repatriating their assets if they want to change providers. Should the provider go out of business or make significant changes to the platform, customers could be forced to make significant infrastructure changes on short notice.

Private Cloud

Private clouds usually reside behind a firewall and are utilized by a single organization. A completely on-premises cloud may be the preferred solution for businesses with very tight regulatory requirements, though private clouds implemented through a colocation provider are gaining in popularity. Authorized users can access, utilize, and store data in the private cloud from anywhere, just like they could with a public cloud. The difference is that no one else can access or utilize those computing resources. Private cloud solutions offer both security and control, but

these benefits come at a cost. The company that owns the cloud is responsible for both software and infrastructure, making this a less economic model than the public cloud.

The additional control offered by a private cloud makes it easier to restrict access to valuable assets and ensures that a company will be able to move its data and applications where it wants, whenever it wants. Furthermore, since the private cloud isn't controlled by an outside vendor, there's no risk of sudden changes disrupting the company's entire infrastructure. A private cloud solution will also not be affected by a public cloud provider's system downtime. But private clouds also lack the versatility of public clouds. They can only be expanded by adding more physical compute and storage capacity, making it difficult to scale operations quickly should the business need arise.

Hybrid Cloud

Hybrid clouds combine public clouds with private clouds. They are designed to allow the two platforms to interact seamlessly, with data and applications moving seamlessly from one to the other.

The primary advantage of a hybrid cloud model is its ability to provide the scalable computing power of a public cloud with the security and control of a private cloud. Data can be stored safely behind the firewalls and encryption protocols of the private cloud, then moved securely into a public cloud environment when needed. This is especially helpful in the age of big data analytics, when industries like healthcare must adhere to strict data privacy regulations while also using sophisticated algorithms powered by artificial intelligence (AI) to derive actionable insights from huge masses of unstructured data.

There are two commonly used types of hybrid cloud architecture. Cloud bursting uses a private cloud as its primary cloud, storing data and housing proprietary applications in a secure environment. When service demands increase, however, the private cloud's infrastructure may not have the capacity to keep up. That's where the public cloud comes in. A cloud bursting model uses the public cloud's computing resources to supplement the private cloud, allowing the company to handle increased traffic without having to purchase new servers or other infrastructure.

The second type of hybrid cloud model also runs most applications and houses data in a private cloud environment, but outsources non-critical applications to a public cloud provider. This arrangement is common for organizations that need to access specialized development tools (like

Adobe Creative Cloud), basic productivity software (like Microsoft Office 365), or CRM platforms (like Salesforce). Multi-cloud architecture is often deployed here, incorporating multiple cloud service providers to meet a variety of unique organizational needs.

What is a Multi-Cloud Model?

In some cases, a single public cloud isn't enough to meet an organization's computing needs. They turn instead to multi-clouds, a more complex hybrid cloud example that combines a private cloud with multiple public cloud services. While a hybrid cloud always consists of a public and private cloud, a multi-cloud environment is a bit more varied on a case-to-case basis. In this arrangement, an organization's IT infrastructure consists of multiple public clouds from multiple providers, although it may access those clouds through a single software-defined network. A private cloud could certainly be part of a multi-cloud architecture, but it is usually more isolated from its public cloud counterparts.

The purpose of a multi-cloud model is versatility and specialization. In enterprise-level organizations, for example, not every department has the same cloud needs. A marketing department, for instance, needs different types of cloud computing tools than a research or human resources department. Rather than trying to create a one-size-fits-all solution, companies can pick and choose from existing public cloud providers to ensure that each department has a solution catered to their specific needs.

Multi-cloud models also offer reassurance because they don't leave organizations dependent upon a single cloud provider. This can decrease costs and increase flexibility in the long run while also avoiding the problem of vendor lock-in. When combined with private cloud assets, multi-cloud deployments allow organizations to accomplish multiple goals at one time without having to radically expand or rethink their existing infrastructure.

Hybrid Cloud Vs Multi-Cloud:

The key differentiator to keep in mind is that multi-cloud models involve using separate cloud environments to perform separate tasks. If an organization needs its IT infrastructure to be able to accommodate the conflicting demands of different departments, then it probably needs to pursue a multi-cloud deployment. The sales team may need the CRM features offered by a specific cloud provider, while software programmers may favour different types of cloud computing environments that offer superior storage and processing capacity.

Large organizations with divisions existing in separate “silos” will typically find that multi-cloud solutions address more of their business needs. At the executive level, CIOs will find the cost efficiencies and versatility of multi-cloud strategies appealing because it gives them the power to leverage providers against one another to drive down IT costs. It also helps them maintain a level of independence that protects them against any sudden changes a cloud vendor may spring on them once the organization is already locked into and dependent upon a single platform.

On the other hand, hybrid cloud models offer a lot of advantages. Since they only involve the interconnections between two environments, they are easier to set up and scale. By using the private cloud to locate sensitive data and running front-end applications in the public cloud, organizations can reduce their exposure to potential security threats and keep a closer watch over activity within their cloud ecosystem. Since public cloud computing services can be offered on a “pay for what we use” model, hybrid clouds can reduce overall IT spend while still allowing companies to scale up processing power when they need it.

Since a hybrid cloud model is more custom built to cater to the specific needs of an organization, it gives IT decision makers more control over their deployments. This level of customization can be tremendously valuable for smaller companies that have a very clear idea of what they need from their infrastructure and it should be optimized to deliver superior services.

Of course, for many organizations, the choice between a hybrid cloud model and a multi-cloud model is a false dichotomy. There’s no reason why a multi-cloud environment can’t incorporate the features of a hybrid cloud. While this is necessarily a more complex solution that requires careful implementation and security considerations, private cloud environments can be integrated into multiple public clouds to allow different users across an organization to access both the data and cloud services they need to do their work more effectively.

1.5 Merits and Demerits of Cloud Computing

Cloud Computing: Advantages Lower-Cost Computers for Users

Here’s a quantitative financial advantage: We don’t need a high-powered (and accordingly high-priced) computer to run cloud computing’s web-based applications. Because the application runs in the cloud, not on the desktop PC, that desktop PC doesn’t need the processing power or hard disk space demanded by traditional desktop software. Hence the client computers in cloud computing can be lower priced, with smaller hard disks, less memory, more efficient processors,

and the like. In fact, a client computer in this scenario wouldn't even need a CD or DVD drive, because no software programs have to be loaded and no document files need to be saved.

Improved Performance

Computers in a cloud computing system will boot up faster and run faster, because they'll have fewer programs and processes loaded into memory.

Lower IT Infrastructure Costs

In a larger organization, the IT department could also see lower costs from the adoption of the cloud computing paradigm. Instead of investing in larger numbers of more powerful servers, the IT staff can use the computing power of the cloud to supplement or replace internal computing resources. Those companies that have peak needs no longer have to purchase equipment to handle the peaks (and then lay fallow the rest of the time); peak computing needs are easily handled by computers and servers in the cloud.

Fewer Maintenance Issues

Speaking of maintenance costs, cloud computing greatly reduces both hardware and software maintenance for organizations of all sizes. First, the hardware. With less hardware necessary in the organization, maintenance costs are immediately lowered. As to software maintenance, remember that all cloud apps are based elsewhere, so there's no software on the organization's computers for the IT staff to maintain.

Lower Software Costs

There is the issue of software cost. Instead of purchasing separate software packages for each computer in the organization, only those employees actually using an application need access to that application in the cloud. As to the cost of that software, it's possible that some cloud computing companies will charge as much to "rent" their apps as traditional software companies charge for software purchases. However, early indications are that cloud services will be priced substantially lower than similar desktop software. In fact, many companies (such as Google) are offering their web-based applications for free, which to both individuals and large organizations is much more attractive than the high costs charged by Microsoft and similar desktop software suppliers.

Instant Software Updates

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

Increased Computing Power

We are no longer limited to what a single desktop PC can do, but can now perform super computing - like tasks utilizing the power of thousands of computers and servers. In other words, we can attempt greater tasks in the cloud than we can on our desktop.

Unlimited Storage Capacity

Similarly, the cloud offers virtually limitless storage capacity. Consider that when our desktop or laptop PC is running out of storage space. Our computer's 200GB hard drive is peanuts compared to the hundreds of petabytes (a million gigabytes) available in the cloud. Whatever we need to store, we can.

Increased Data Safety

The data in the cloud is automatically duplicated, so nothing is ever lost. That also means if our personal computer crashes, all our data is still out there in the cloud, still accessible. In a world where few individual desktop PC users back up their data on a regular basis, cloud computing can keep data safe.

Improved Compatibility between Operating Systems

We can connect our Windows computer to the cloud and share documents with computers running Apple's Mac OS, Linux, or UNIX. In the cloud, the data matters, not the operating system.

Improved Document Format Compatibility

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

Easier Group Collaboration

Sharing documents leads directly to collaborating on documents. To many users, this is one of the most important advantages of cloud computing: the ability for multiple users to easily collaborate on documents and projects. Imagine that we, all need to work together on an important project. Before cloud computing, we had to email the relevant documents from one user to another, and work on them sequentially. Now each of we can access the project's documents simultaneously; the edits one user makes are automatically reflected in what the other users see onscreen. It's all possible, of course, because the documents are hosted in the cloud, not on any of our individual computers. All we need is a computer with an Internet connection, and we're collaborating. Of course, easier group collaboration means faster completion of most group projects, with full participation from all involved. It also enables group projects across different geographic locations. No longer does the group have to reside in a single office for best effect. With cloud computing, anyone anywhere can collaborate in real time. It's an enabling technology.

Universal Access to Documents

Ever get home from work and realize we left an important document at the office? Or forget to take a file with us on the road? Or get to a conference and discover we forgot to bring along our presentation? Not a problem—not anymore, anyway. With cloud computing, we don't take our documents with us. Instead, they stay in the cloud, where we can access them from anywhere, we have a computer and an Internet connection. All our documents are instantly available from wherever we are. There's simply no need to take our documents with us as long as we have an Internet connection.

Latest Version Availability

And here's another document-related advantage of cloud computing. When we edit a document at home, that edited version is what we see when we access the document at work. The cloud always hosts the latest version of our documents; we're never in danger of having an outdated version on the computer we're working on.

Removes the Tether to Specific Devices

Finally, here's the ultimate cloud computing advantage: we're no longer tethered to a single computer or network. Change computers, and our existing applications and documents will follow us through the cloud. Move to a portable device, and our apps and docs are still available. There's

no need to buy a special version of a program for a particular device, or save our document in a device-specific format. Our documents and the programs that created them are the same no matter what computer we're using.

Cloud Computing: Disadvantages

That's not to say, of course, that cloud computing is without its disadvantages. There are a number of reasons why we might not want to adopt cloud computing for our particular needs. Let's examine a few of the risks related to cloud computing.

Requires a Constant Internet Connection

Cloud computing is, quite simply, impossible if we can't connect to the Internet. Because we use the Internet to connect to both our applications and documents, if we don't have an Internet connection, we can't access anything, even our own documents. A dead Internet connection means no work, period and in areas where Internet connections are few or inherently unreliable, this could be a deal breaker. When we're offline, cloud computing just doesn't work. This might be a more significant disadvantage than we might think. Sure, we're used to a relatively consistent Internet connection both at home and at work, but where else do we like to use our computer? If we're used to working on documents on our deck, or while we're at a restaurant for lunch, or in our car, we won't be able to access our cloud based documents and applications unless we have a strong Internet connection at all those locations, of course. A lot of what's nice about portable computing becomes problematic when we're depending on web-based applications.

Doesn't Work Well with Low-Speed Connections

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

Can Be Slow

Even on a fast connection, web-based applications can sometimes be slower than accessing a similar software program on our desktop PC. That's because everything about the program, from

the interface to the document we're working on, has to be sent back and forth from our computer to the computers in the cloud. If the cloud servers happen to be backed up at that moment, or if the Internet is having a slow day, we won't get the instantaneous access we're used to with desktop apps.

Features Might Be Limited

This particular disadvantage is bound to change, but today many web-based applications simply aren't as full-featured as their desktop-based brethren. Compare, for example, the feature set of Google Presentations with that of Microsoft PowerPoint; there's just a lot more we can do with PowerPoint than we can with Google's web-based offering. The basics are similar, but the cloud application lacks many of PowerPoint's advanced features. So if we're an advanced user, we might not want to leap into the cloud computing waters just yet. That said, many web-based apps add more advanced features over time. This has certainly been the case with Google Docs and Spreadsheets, both of which started out somewhat crippled but later added many of the more niche functions found on Microsoft Word and Excel. Still, we need to look at the features before we make the move. Make sure that the cloud-based application can do everything we need it to do before we give up on our traditional software.

Stored Data Might Not Be Secure

With cloud computing, all our data is stored on the cloud. That's all well and good, but how secure is the cloud? Can other, unauthorized users gain access to our confidential data? So security is a major issue in cloud environment.

Problem will arise if Data loss occurs

Theoretically, data stored in the cloud is unusually safe, replicated across multiple machines. But on the off chance that our data does go missing, we have no physical or local backup. Put simply, relying the cloud puts us at risk if the cloud lets us down. This is one of the major disadvantage of cloud.

1.6 Difference between Cloud Computing and Web 2.0

Cloud Computing	Web 2.0
It is more specific and definite	Programming and business models
It is a way of searching through data.	It is sharing entire pieces of data between different websites.
Cloud computing is about computers.	Web 2.0 is about people.
The internet as a computing platform	Attempt to explore and explain the business rules of that platform
Google apps are considered in Cloud computing.	A web-based application is considered in Web 2.0.
It is a business model for hosting these services.	It is a technology which allows webpages to act as more responsive applications

1.7 Key challenges in cloud computing

Bandwidth cost

Cloud computing saves the hardware acquisition costs but their expenditure on bandwidth rises considerably. Sufficient bandwidth is required to deliver intensive and complex data over the network.

Continuous monitoring and supervision

It is important to monitor the cloud service continuously as well as to supervise its performance, business dependency and robustness.

Security concerns

To prevent cloud infrastructure damages, some of the measures include tracking unusual behaviour across servers, buying security hardware and using security applications.

Data access and integration

In cloud where the data stored, how to access it, who is the owner and how to control it. Companies are often concerned about data ownership and loss of data control while moving to cloud.

The integration of existing applications in the cloud for smooth running is another challenge.

Proper usability

Enterprises need to have a good and clear view of how to use the technology to add value to their unique businesses

Migration issues

Migrating data from system to the cloud can pose major risks, if it not handled properly. It need to develop migration strategy that integrates well with the current IT infrastructure.

Cost assessment

Scalable and on-demand nature of cloud services makes the assessment of cost difficult.

Current Cloud Computing Challenges

Managing multi-cloud environments:

Most of the organizations are not using one cloud. From the survey of Right Scale findings, it is estimated that almost 81% of all the entrepreneurs are using multi-cloud strategy. The multi-cloud environment is increasing the complexity faced by IT companies. Doing practices like training employees, doing research, actively managing vendor relationships, tooling and rethinking process.

Migration:

Moving the existing application into the cloud is the most difficult task. A dimension research study found that mostly 62% of the cloud migration projects are most difficult than it is to be expected.55% exceeding their budget while 64% is taking longer than to be expected. All the migration projects are troubleshooting, time-consuming, slow migration of data and difficulty security configuration, difficulty syncing data before cutover, downtime during migration and

trouble during migration tools to work properly. To overcome all these major IT challenges hiring in-house experts, set a longer project timeline, increase budget.

Vendor Lock-In:

Many vendors like Microsoft Azure, IBM cloud, Google cloud platform, Amazon Web Services are dominating the public cloud market. That's because of the buyer's caution. Ensure that the services we use are unique and can be transported to other providers, and most importantly, understand the requirements. It could be that this service is not standard or there is no decent vendor replacement. Entering the cloud computing agreement is easier than leaving it.

Reliability of new technology:

Security threads increase because they do not know and where information is stored and processed. Usually, employers are hesitant to share organizational information with unknown service providers. This is a fact of human nature that we believe in things that are before our eyes. They think that the information stored in their offices is safer and more accessible. By using cloud computing, they fear losing control of data. They think that data is taken from them and submitted to unknown third parties. The fear of these unknown service providers must be peacefully dealt with and removed from their minds.

Operational Security:

A major fear of cloud service providers is cyber-attacks which is an operational security issue. Most of the data stored in the cloud are at a risk of cyber attack where an only limited amount of data is to be stored. Vulnerability assessment on the overall security measures of providers against external attacks is an effective way to ensure that data in the cloud is adequately protected. The high threat level is because the targets are often by malware, virtual machines and brute force attacks. Even as most cloud providers have strict security measures, cyber attacks are always looming. The high threat level is because the targets are often by malware, virtual machines and brute force attacks.

Cost Barrier:

To transfer complex and intensive data over the network, we must have sufficient bandwidth. This is a major obstacle in front of small organizations, which limits them from implementing cloud technology in their business. Businesses can reduce costs on hardware but they have to spend a large amount on bandwidth. For smaller application costs it's not a big problem but

for large and complex applications it's a major problem. For efficient cloud computing work, we have to bear high bandwidth costs.

Reducing the risk of threats:

Organizations must observe and examine threats very seriously. Every organization may not have enough mechanisms to reduce this type of threat. It's very complicated to say that cloud service providers meet security standards and threat risks. These security threats and risks examine the application of cloud solutions.

Hacking of the brand:

Cloud providers accommodate many clients; each can be influenced by actions taken against one of them. When there is a threat coming to the main server it also affects all other clients too. Hard computing brings several main risk factors such as hacking. Some professional hackers can hack applications by breaking down firewalls efficiently and stealing sensitive information from organizations. As in the rejection of requests for server attack services that flooded providers of widely distributed computers.

Password security:

Password security plays a crucial role in cloud security. Since many people are having access to cloudless security concern is there. If anyone who had knowledge about our password can access our account which we stored in the cloud. Multi-factor authentication should be employed to make sure password is protected properly particularly when the staff members leave. Access rights should be given to those who are related to username and password and it should be allocated for those who require them.

Management of cloud:

Cloud services can be easily changed and updated by business users. It does not involve the direct involvement of the IT department. It is the responsibility of the service provider to manage information and disseminate it throughout the organization. Many famous dramatic predictions about the impact of cloud computing. People think that traditional IT departments will be outdated and research supports the conclusion that cloud impacts tend to be more gradual and less linear. It consists of many technical challenges. So it's difficult to manage all the functions of complex cloud computing

1.8 Major Cloud Players

- Microsoft Azure
- Amazon Web Services (AWS)
- Google Cloud
- Alibaba Cloud
- IBM Cloud
- Oracle
- Salesforce
- SAP
- Rackspace Cloud
- VMWare

*Amazon Web Services offers many types of cloud computing services. See the website for details.

**VMware is a provider of the technology underlying cloud computing and does not provide hosting services.

The Top 5 Cloud-Computing Vendors: #1 Microsoft, #2 Amazon, #3 IBM, #4 Salesforce, #5 SAP

#1 Microsoft remains an absolute lock at the top due to four factors: its deep involvement at all three layers of the cloud (IaaS, PaaS and SaaS); its unmatched commitment to developing and helping customers deploy AI, ML and Blockchain in innovative production environments; its market-leading cloud revenue, which is estimated to be about $16.7 billion.

#2 Amazon might not have the end-to-end software chops of the others in the Top 5 but it was and continues to be the poster-child for the cloud-computing movement: the first-moving paradigm-buster and category creator. It is believed that Amazon will make some big moves to bolster its position in software, and no matter how we slice it, the $16 billion cloud revenue from AWS is awfully impressive.

#3 IBM has leapfrogged both Salesforce.com (formerly tied with Amazon for #2 and now in the #4 spot) and SAP (formerly #4) on the strength of its un-trendy but highly successful emphasis on transforming its vast array of software expertise and technology from the on-premises world to the cloud. In so doing, IBM has quietly created a $15.8-billion cloud business (again on trailing-12-month basis) that includes revenue of $7 billion from helping big global corporations convert legacy systems to cloud or cloud-enabled environments. And like #1 Microsoft, IBM plays in all

three layers of the cloud—IaaS, PaaS and SaaS—which is hugely important for the elite cloud vendors because it allows them to give customers more choices, more seamless integration, better cybersecurity, and more reasons for third-party developers to rally to the IBM Cloud. Plus, its relentless pairing of "cloud and cognitive" is an excellent approach toward weaving AI and ML deeply into customer-facing solutions.

#4 Salesforce.com falls a couple of spots from its long-time tie with Amazon at #2 but—and this will be the case as long as founder Marc Benioff is CEO—remains a powerful source of digital innovation and disruptive strategy. However, to remain in the rarified air near the top of the Cloud Wars Top 10, Benioff and Salesforce must find a way to extend their market impact beyond their enormously successful SaaS business and become more of a high-impact player in the platform or PaaS space. At this stage, it's simply not possible for Salesforce to become a player in IaaS, so Benioff needs to crank up the genius machine and hammer his way into top contention as a platform powerhouse.

#5 SAP has what all of the other cloud vendors would kill for: unmatched incumbency within all of the world's leading corporations as the supplier of mission-critical business applications that run those companies. It's also fashioned, under CEO Bill McDermott, powerful new partnerships with Amazon and Google to complement its long-standing relationships with IBM and Microsoft, all of which give customers a heightened sense of confidence that SAP will be willing and able to play nice in heterogeneous environments. Plus, SAP's HANA technology is now in full deployment across thousands of businesses, and as it takes root and SAP continues to rationalize its massive product portfolio around HANA in the cloud, SAP has a very bright future ahead of it in the cloud.

1.9 Cloud Deployment Models

The four types of Cloud Deployment Models identified by NIST.

- Private cloud
- Community cloud
- Public cloud
- Hybrid cloud

Private Cloud

- Cloud environments are defined based on hardware location and owner. Private clouds are accessible only to a respective customer residing either on-site or be outsourced by a third party.
- The cloud infrastructure is operated solely for an organization.
- Contrary to popular belief, private cloud may exist off premises and can be managed by a third party.
- Thus, two private cloud scenarios exist, as follows:

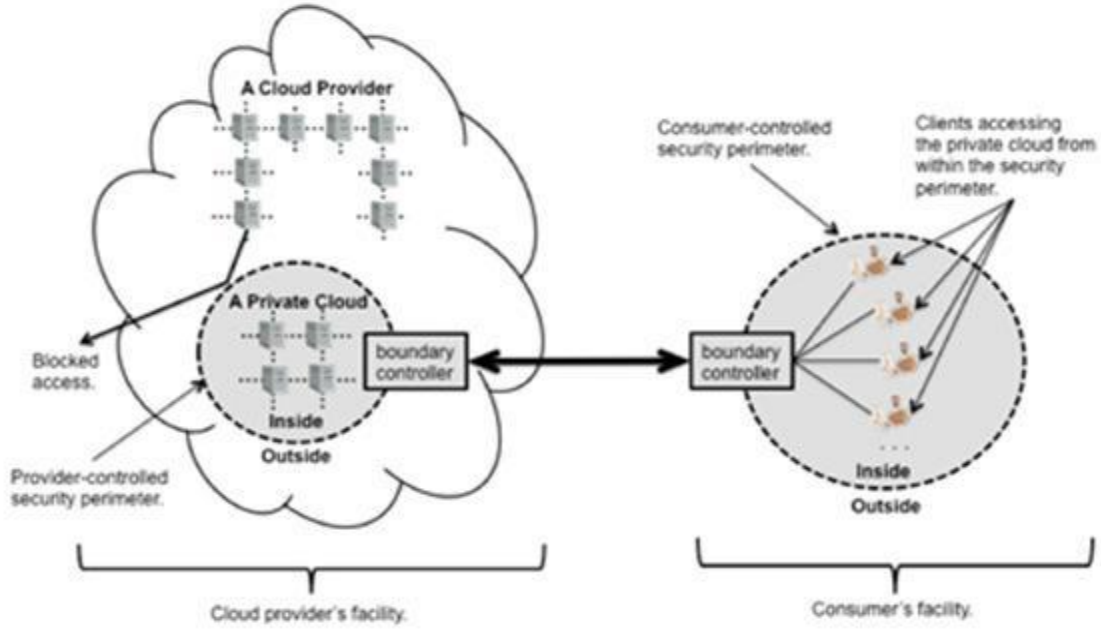
o On-site Private Cloud

- Applies to private clouds implemented at a customer's premises.

o Outsourced Private Cloud

- Applies to private clouds where the server side is outsourced to a hosting company.

- Examples of Private Cloud:
  - Eucalyptus
  - Ubuntu Enterprise Cloud - UEC (pooured by Eucalyptus)
  - Amazon VPC (Virtual Private Cloud)
  - VMware Cloud Infrastructure Suite
  - Microsoft ECI data center



.Fig 1.4: Private Cloud

Community Cloud

- The cloud infrastructure is shared by several organizations and supports a specific community that has shared concerns (e.g., mission, security requirements, policy, and compliance

considerations). Government departments, universities, central banks etc. often find this type of cloud useful.

- Community cloud also has two possible scenarios:
  - On-site Community Cloud Scenario
    - Applies to community clouds implemented on the premises of the customers composing a community cloud
  - Outsourced Community Cloud
    - Applies to community clouds where the server side is outsourced to a hosting company.
- Examples of Community Cloud:
  - Google Apps for Government
  - Microsoft Government Community Cloud

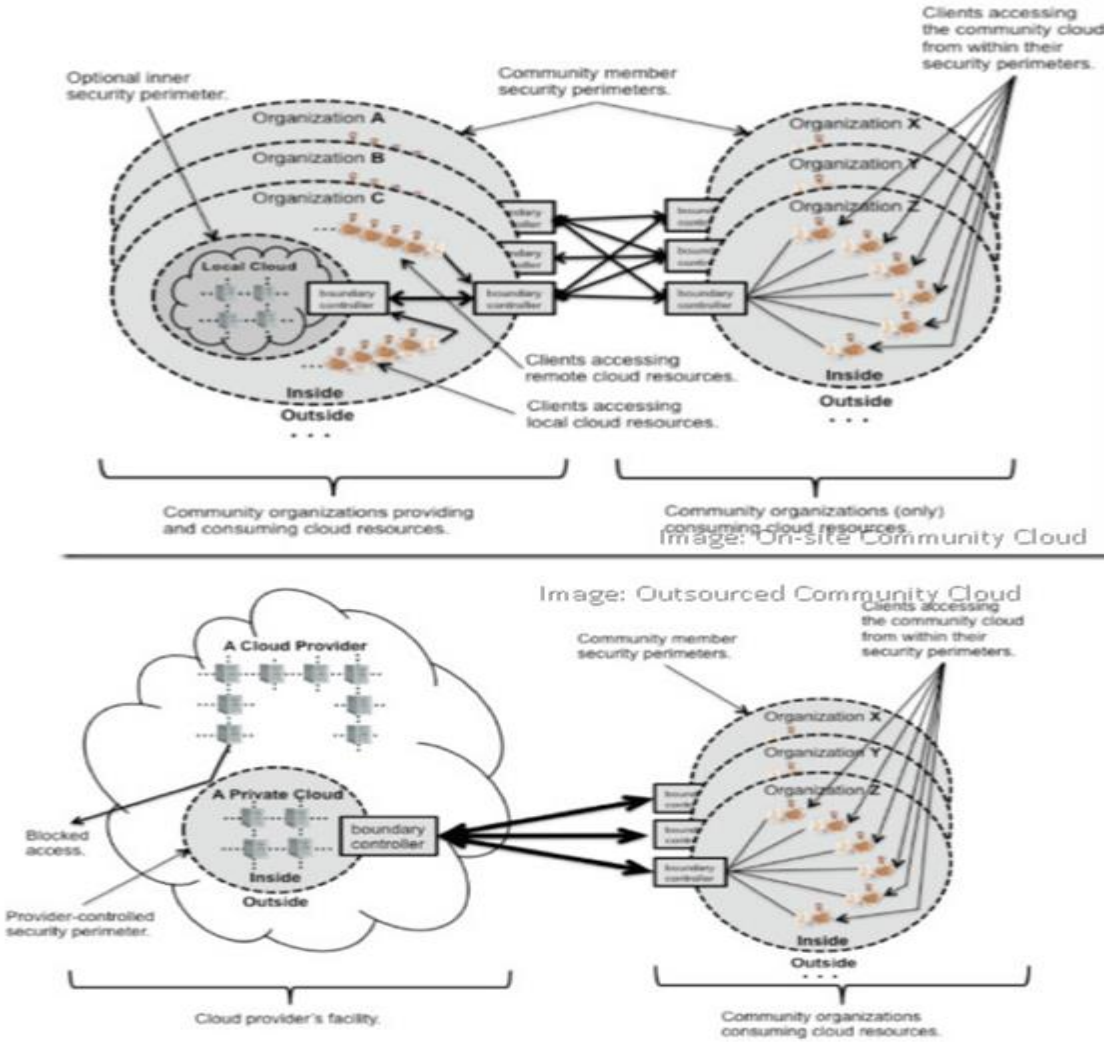


Fig 1.5 Community Cloud

Public Cloud

- The most ubiquitous, and almost a synonym for, cloud computing is 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.
- Examples of Public Cloud:
- Google App Engine
- Microsoft Windows Azure O IBM Smart Cloud
- Amazon EC2

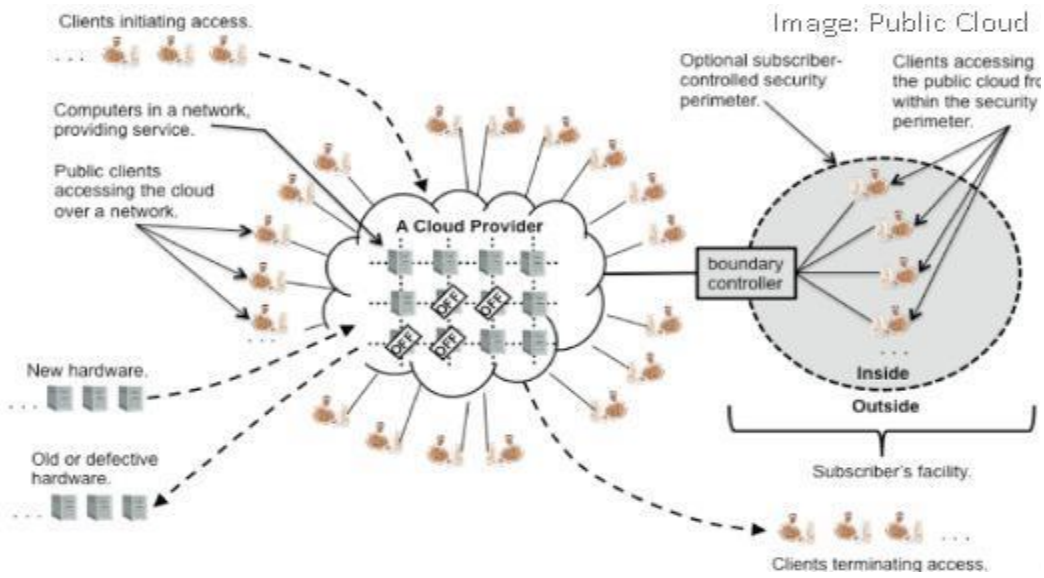


Fig 1.6 Public Cloud

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).
- Examples of Hybrid Cloud:
    - Windows Azure (capable of Hybrid cloud)
    - VMware vCloud (Hybrid Cloud Services)

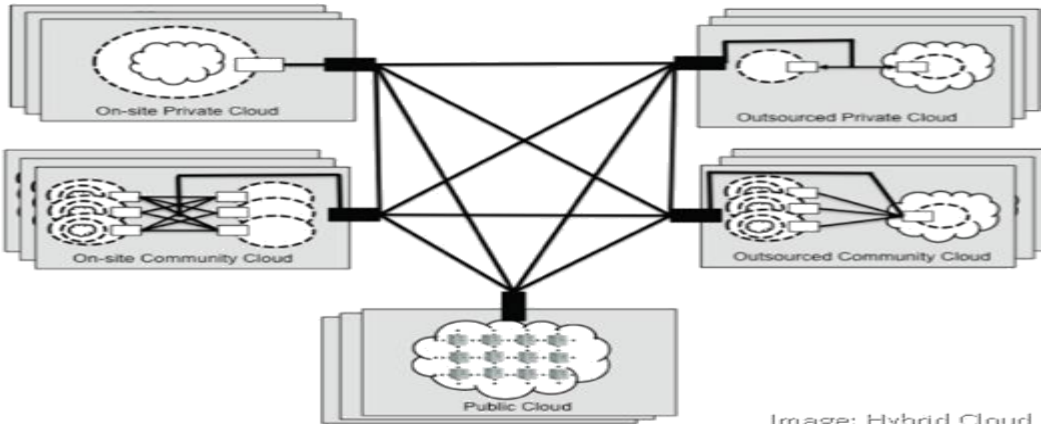


Fig 1.7 Hybrid Cloud

Cloud Computing Life Cycle

Fig 1.8 Cloud LifeCycle

Cloud Components

It has three components

- Client computers
- Distributed Servers
- Data centers

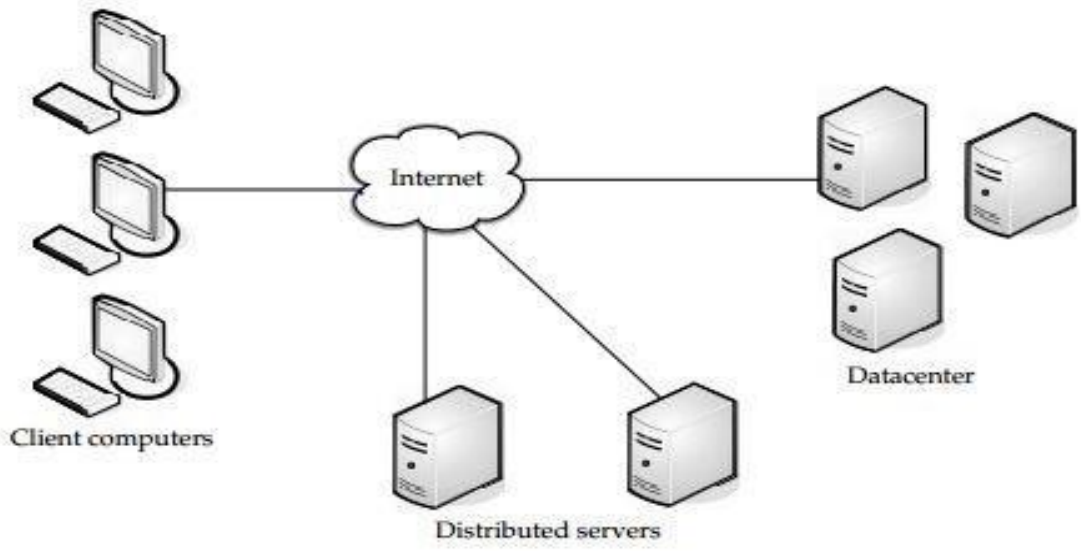


Fig 1.9 Cloud Components

- **Clients**

  Clients are the device that the end user interact with cloud. Three types of clients:

  - Mobile
  - Thick
  - Thin (Most Popular)

- **Datacenter**

  It is collection of servers where application is placed and is accessed via internet.

- **Distributed servers**

  Often servers are in geographically different places, but server acts as if they are working next to each other

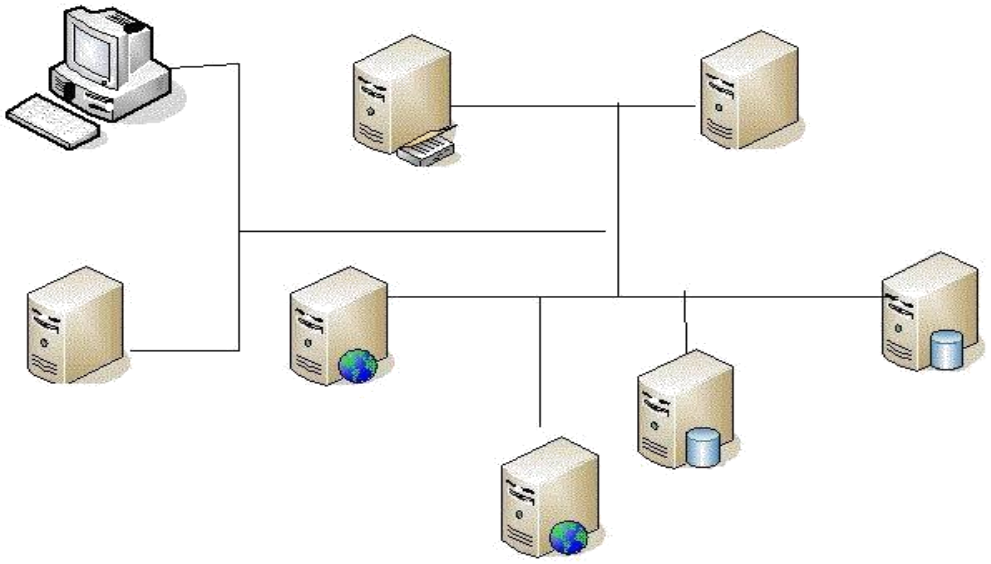


Fig 1.10 Distributed Servers

Who Benefits from Cloud Computing

- Collaborators
- Road Warriors
- Cost-Conscious Users
- Cost-Conscious IT Departments
- Users with Increasing Needs

Who Should not be using Cloud Computing

- The internet impaired
- Offline workers
- The security conscious
- Anyone married to existing applications

Dark Clouds: Barriers to use web-based applications

- Technical issues
- Business model issues
- Internet issues
- Security issues
- Compatibility issues
- Social issues

Cloud Computing Applications

- Clients would be able to access their applications and data from anywhere at any time.
- It could bring hardware costs down.
- The right software available in place to achieve goals.
- Servers and digital storage devices take up space.
- Corporations might save money on IT support.
- The client takes advantage of the entire network's processing power.

1.10 Virtualization in cloud computing

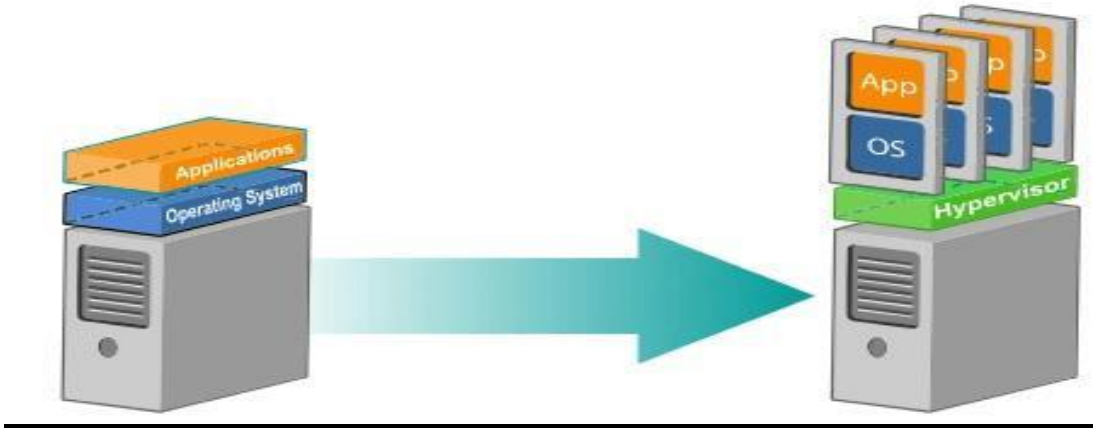


Fig 1.11Virtualization in cloud computing

- **Virtualization** is Nothing but Creating Virtual Version of Something, in Computer Terms it can be OS, Storage Device, Network Resources.

- **Virtualization** is the ability to run multiple operating systems on a single physical system and share the underlying hardware resources.
- It is the process by which one computer hosts the appearance of many computers.

- Virtualization is used to improve IT throughput and costs by using physical resources as a pool from which virtual resources can be allocated.
- With VMware virtualization solutions we can reduce IT costs while increasing the efficiency, utilization and flexibility of their existing computer hardware.

- **Virtualization Architecture**

A Virtual machine (VM) is an isolated runtime environment (guest OS and applications)

Multiple virtual systems (VMs) can run on a single physical system

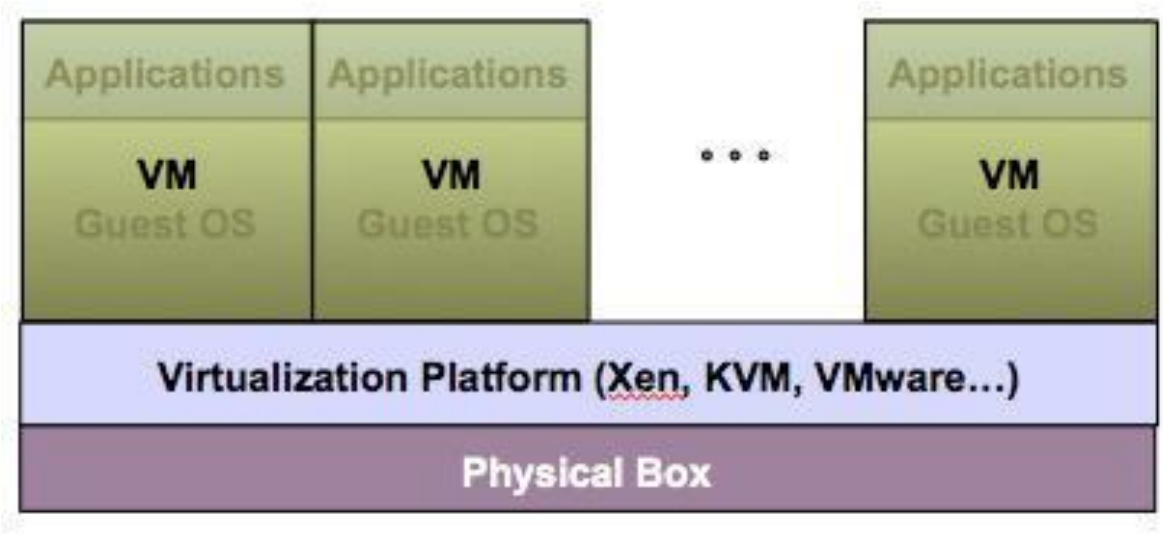


Fig 1.12 Vitualization Architecture

- **Before Virtualization**

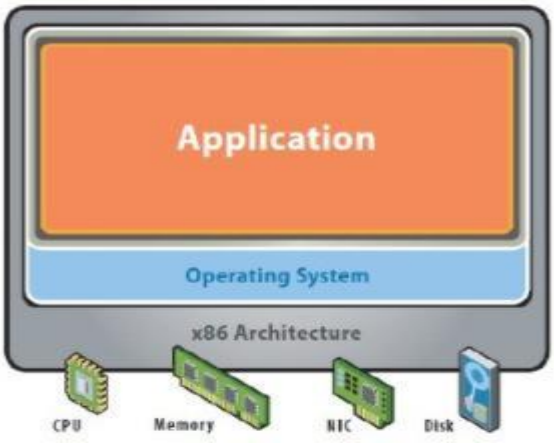


Fig 1.13 Before Virtualization

- Single OS image per machine
- Software and hardware tightly coupled
- Running multiple applications on same machine often creates conflict
- Inflexible and costly infrastructure

After virtualization

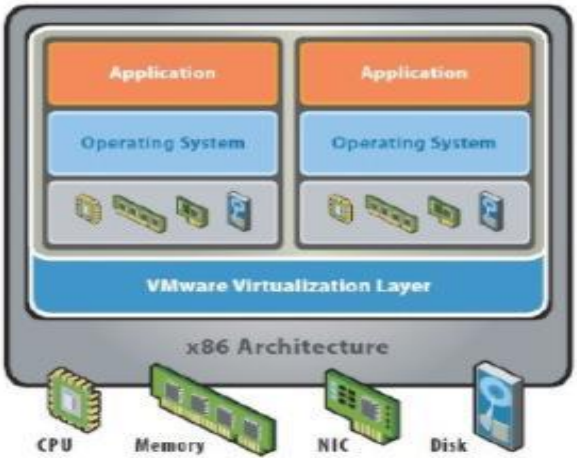


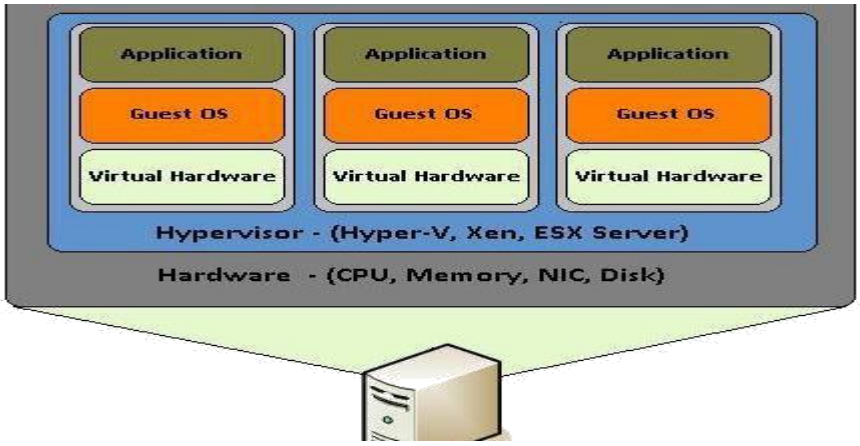
Fig 1.14 After Virtualization

- Hardware-independence of operating system and applications
- Virtual machines can be provisioned to any system
- Can manage OS and application as a single unit by encapsulating them into virtual Machines

- **Benefits of Virtualization**
  - Sharing of resources helps cost reduction.
  - Isolation: Virtual machines are isolated from each other as if they are physically separated.
  - Encapsulation: Virtual machines encapsulate a complete computing environment.
  - Hardware Independence: Virtual machines run independently of underlying hardware.
  - Portability: Virtual machines can be migrated between different hosts.

- **What makes virtualization possible?**
  - There is a software that makes virtualization possible. This software is known as a Hypervisor, also known as a virtualization manager.

- It sits between the hardware and the operating system and assigns the amount of access that the applications and operating systems have with the processor and other hardware resources.



- **Hypervisor**

The term hypervisor was first coined in 1956 by IBM

- Hypervisor acts as a link between the hardware and the virtual environment and distributes the hardware resources such as CPU usage, memory allotment between the different virtual environments.
- A hypervisor or virtual machine monitor (VMM) is computer software, firmware or hardware that creates and runs virtual machines. A computer on which a hypervisor runs one or more virtual machines is called a host machine, and each virtual machine is called a guest machine.
- A hypervisor is a hardware virtualization technique that allows multiple guest operating systems (OS) to run on a single host system at the same time. The guest OS shares the hardware of the host computer, such that each OS appears to have its own processor, memory and other hardware resources.
- A hypervisor is also known as a virtual machine manager (VMM).

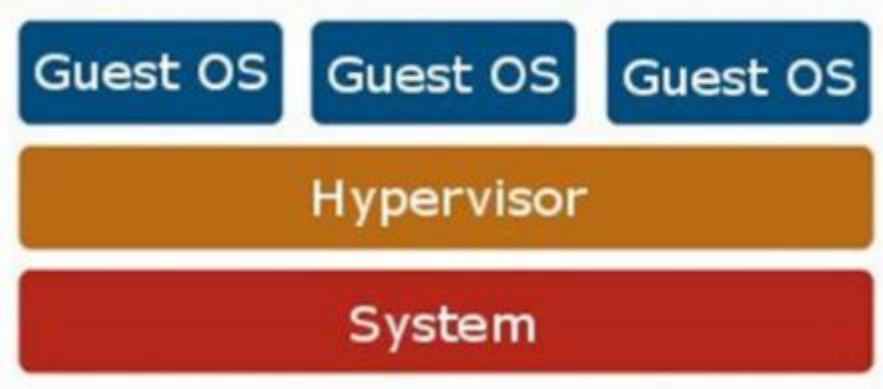


Fig 1.15 Hypervisor

1.11 Types of virtualization

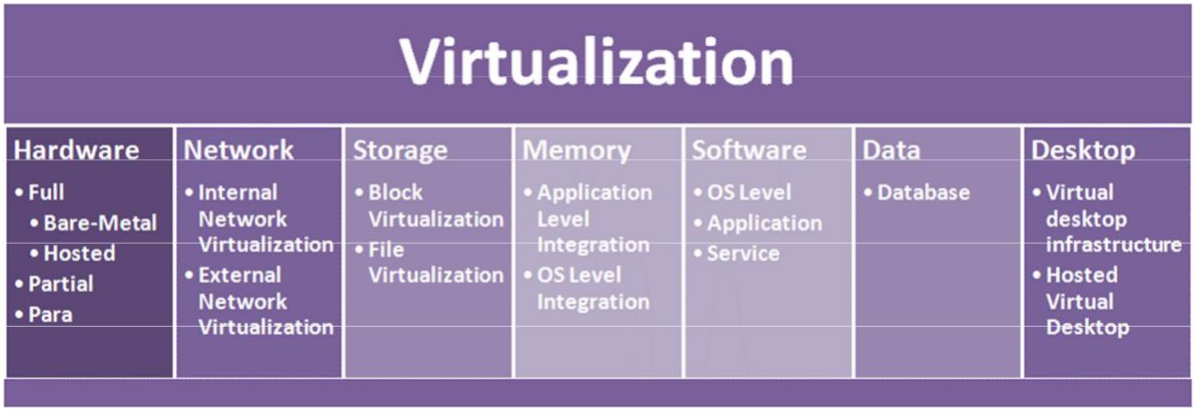


Fig 1.16 Types of Virtualization

- The 7 Types of Virtualization

Hardware Virtualization. Software Virtualization. Network Virtualization. Storage Virtualization Memory Virtualization. Data Virtualization.Desktop Virtualization.

Hardware Virtualization

- Hardware or platform virtualization means creation of virtual machine that act like **real computer**.
- Ex. Computer running Microsoft Windows 7 may host the virtual machine look like a Ubundu
- Hardware virtualization also knows as hardware-assisted virtualization or **server virtualization**.

- The basic idea of the technology is to combine many small physical servers into one large physical server, so that the processor can be used more effectively and efficiently.
- Each small server can host a virtual machine, but the entire **cluster of servers** is treated as a single device by any process requesting the hardware.
- The hardware resource allotment is done by the **hypervisor**.
- The advantages are increased processing poour as a result of **maximized hardware utilization and application uptime.**
- Hardware virtualization is further subdivided into the following types

Full Virtualization – Guest software does not require any modifications since the underlying hardware is fully simulated.

Para Virtualization – The hardware is not simulated and the guest software run their own isolated domains.

Partial Virtualization – The virtual machine simulates the hardware and becomes independent of it. The guest operating system may require modifications.

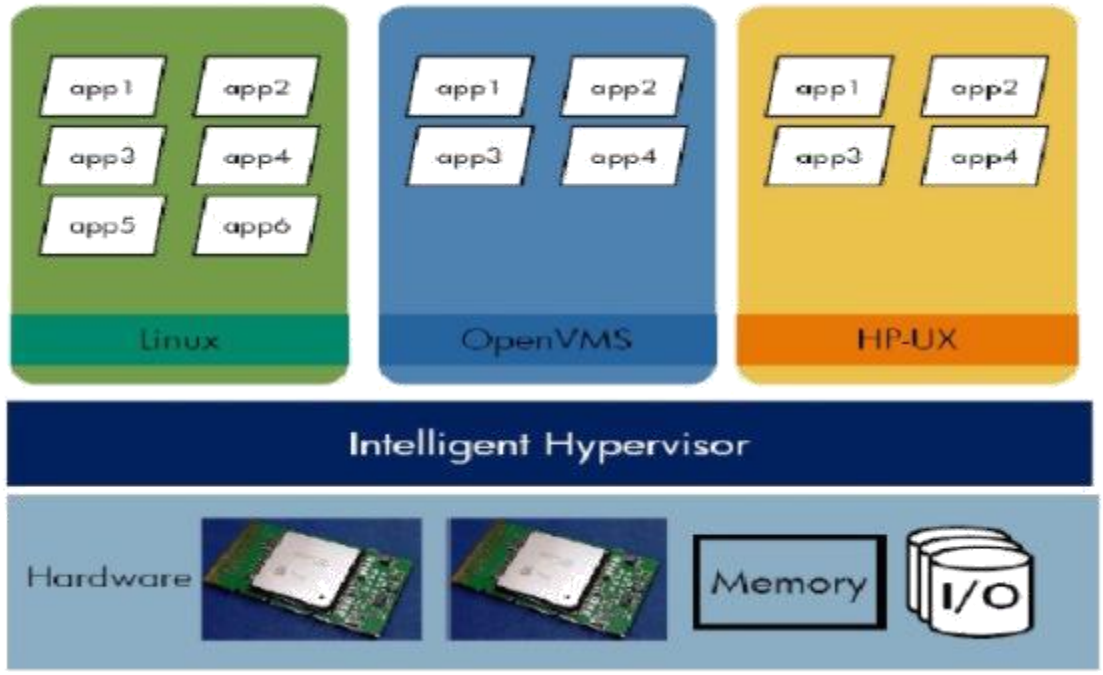


Fig 1.16. Hardware Virtualization

Software Virtualization

- The ability to computer to run and create **one or more virtual environments**.
- It is used to enable a **computer system** in order to allow a guest OS to run.
- Ex. Linux to run as a guest that is natively running a Microsoft Windows OS
- Subtypes:

Operating System Virtualization – Hosting multiple OS on the native

Application Virtualization – Hosting individual applications in a virtual environment separate from the native OS

Service Virtualization – Hosting specific processes and services related to a particular application

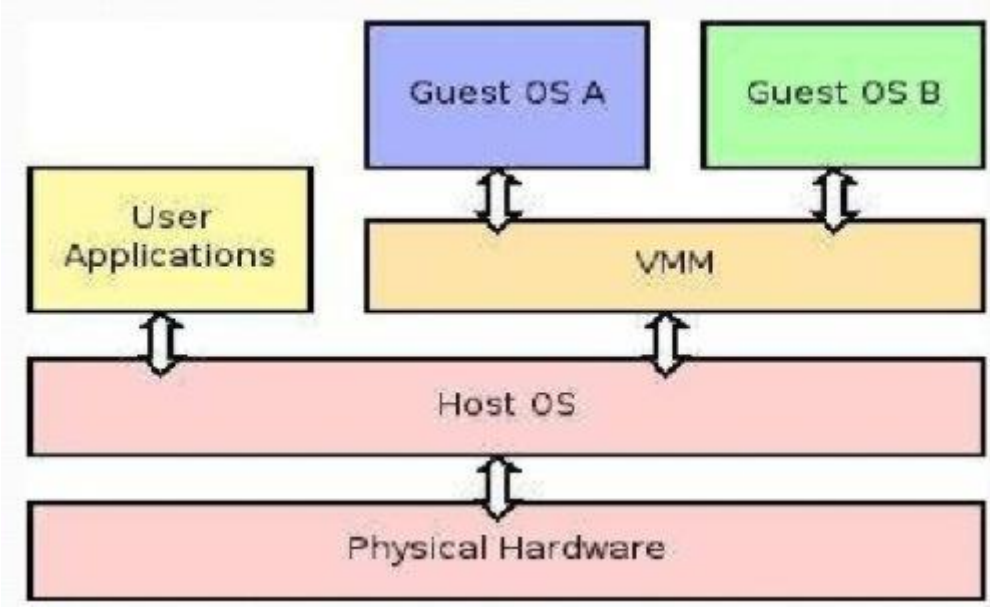


Fig 1.17 Software Virtualization

Network Virtualization

- It refers to the **management and monitoring of a computer network** as a single managerial entity from a single software-based administrator's console.
- **Multiple sub-networks** can be created on the same physical network, which may or may not is authorized to communicate with each other.
- It allows **network optimization** of data transfer rates, scalability, reliability, flexibility, and security
- **Subtypes:**

Internal network: Enables a single system to function like a network O **External network**: Combine many networks, or parts of networks into a virtual unit

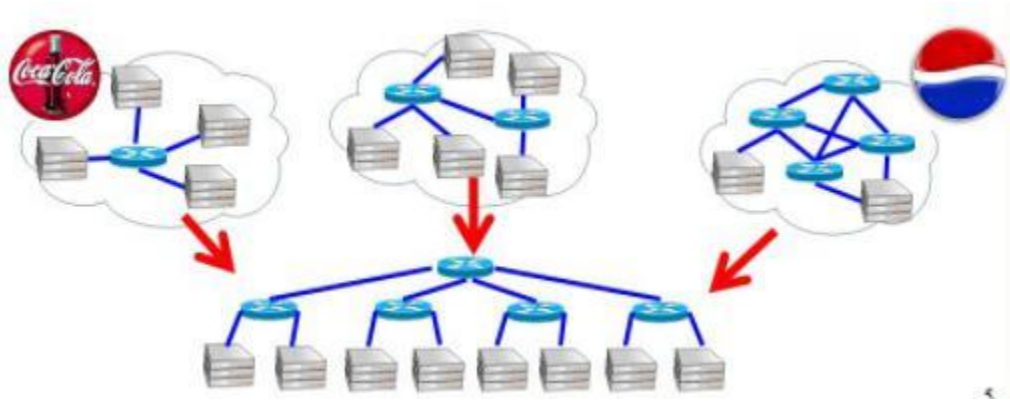


Fig 1.18 Network Virtualiztion

Storage Virtualization

- **Multiple physical storage devices** are grouped together, which look like a single storage device.
- Ex. **Partitioning our hard drive** into multiple partitions
- **Advantages**

  Improved storage management in a heterogeneous IT environment

  Easy updates, better availability

  Reduced downtime

  Better storage utilization

  Automated management

- Two types

  Block- Multiple storage devices are consolidated into one

  File- Storage system grants access to files that are stored over multiple hosts

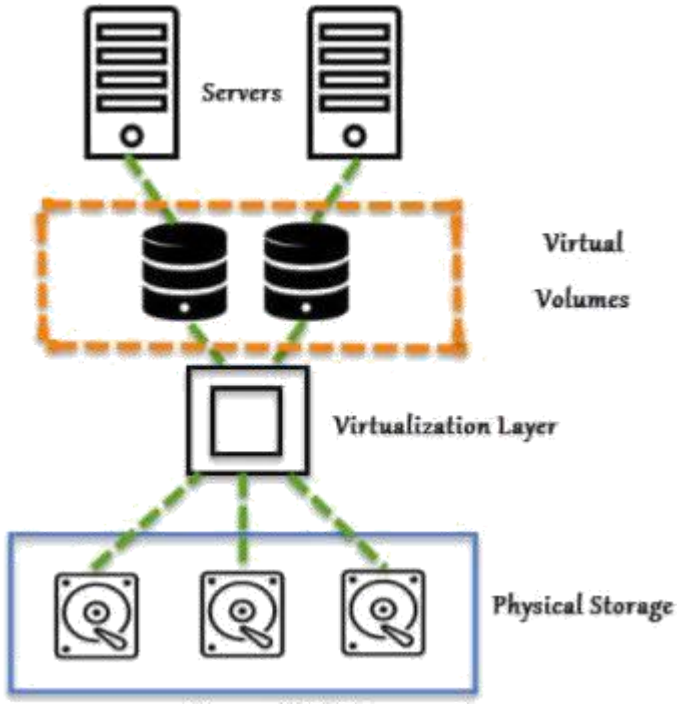


Fig 1.19 Storage Virtualizaion

Memory Virtualization

- The way to **decouple memory from the server** to provide a shared, distributed or networked function.
- It enhances performance by providing **greater memory capacity** without any addition to the main memory.

- **Implementations**

  Application-level integration – Applications access the memory pool directly

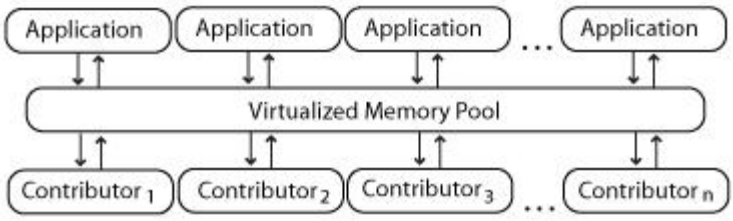


Fig 1.20 Application level integration

Operating System Level Integration – Access to the memory pool is provided through an operating system.

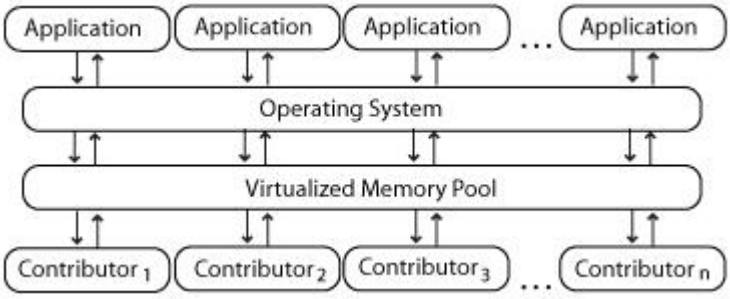


Fig 1.21 Operating System Level integration

Data Virtualization

- Without any technical details, we can **easily manipulate data** and know how it is formatted or where it is physically located.
- It decreases the data errors and workload
- The data is presented as an abstract layer **completely independent of data structure and database systems**
- The user's desktop is stored on a remote server, allowing **the user to access his/her desktop from any device or location**.
- It provides the **work convenience and security**
- It provides a lot of flexibility for employees to **work from home or on the go**
- Since the **data transfer takes place over secure protocols**, any risk of data theft is minimized

Fig 1.22 Data Virtualization

Which Technology to use?

- Virtualization is possible through a wide **range of Technologies** which are available to use and are also Open Source.
- They are,

  XEN O KVM

  OpenVZ

1.12 Parallelization in cloud computing

- Parallel computing is a type of computing architecture in which several processors execute or process an application or computation simultaneously.
- Parallel computing helps in performing large computations by dividing the workload between more than one processor, all of which work through the computation at the same time.
- Most supercomputers implemented parallel computing principles to operate.
- Parallel computing is also known as parallel processing.
- Parallel processing is generally implemented in operational environments/scenarios that require massive computation or processing poour.
- The primary objective of parallel computing is to increase the available computation poour for faster application processing.
- Typically, parallel computing infrastructure is housed within a single facility where many processors are installed in a server rack or separate servers are connected together.
- The application server sends a processing request that is distributed in small components, which are concurrently executed on each processor/server.
- Parallel computation can be classified as bit-level, instructional level, data and task parallelism.

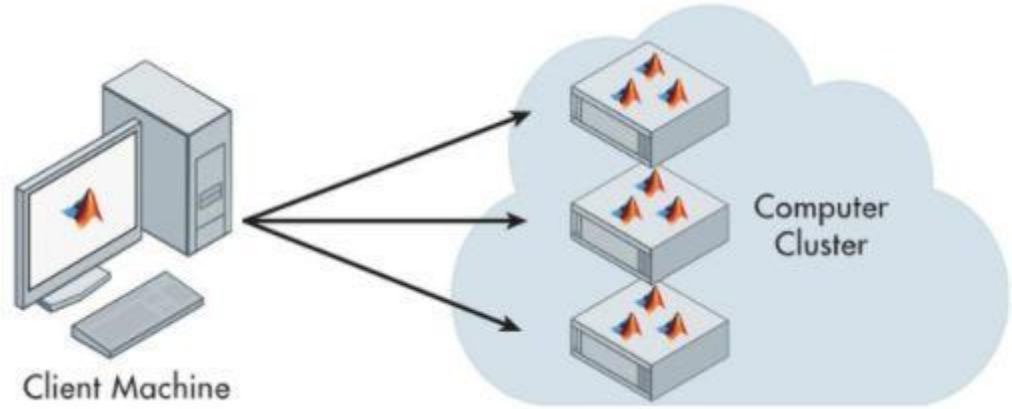


Fig 1.23 Parallelization in Cloud Computing

1.13 Cloud resource management

- **Critical function** of any man-made system.
- It affects the **three basic criteria** for the evaluation of a system: Functionality. Performance. Cost.
- **Scheduling** in a computing system deciding how to allocate resources of a system, such as CPU cycles, memory, secondary storage space, I/O and network bandwidth, between users and tasks.
- **Policies and mechanisms** for resource allocation.
  Policy: principles guiding decisions.

  Mechanisms: the means to implement policies
- **Cloud resources**
  Requires complex policies and decisions for multi-objective optimization.

  It is challenging - the complexity of the system makes it impossible to have accurate global state information.

  Affected by unpredictable interactions with the environment, e.g.,system failures, attacks.

  Cloud service providers are faced with large fluctuating loads which challenge the claim of cloud elasticity

  Strategies for resource management for IaaS, PaaS, and SaaS are different.
- **Cloud resource management (CRM) policies**

Admission control: prevent the system from accepting workload in violation of high-level system policies.

Capacity allocation: allocate resources for individual activations of a service.

Load balancing: distribute the workload evenly among the servers.

Energy optimization: minimization of energy consumption

Quality of service (QoS) guarantees: ability to satisfy timing or other conditions specified by a Service Level Agreement

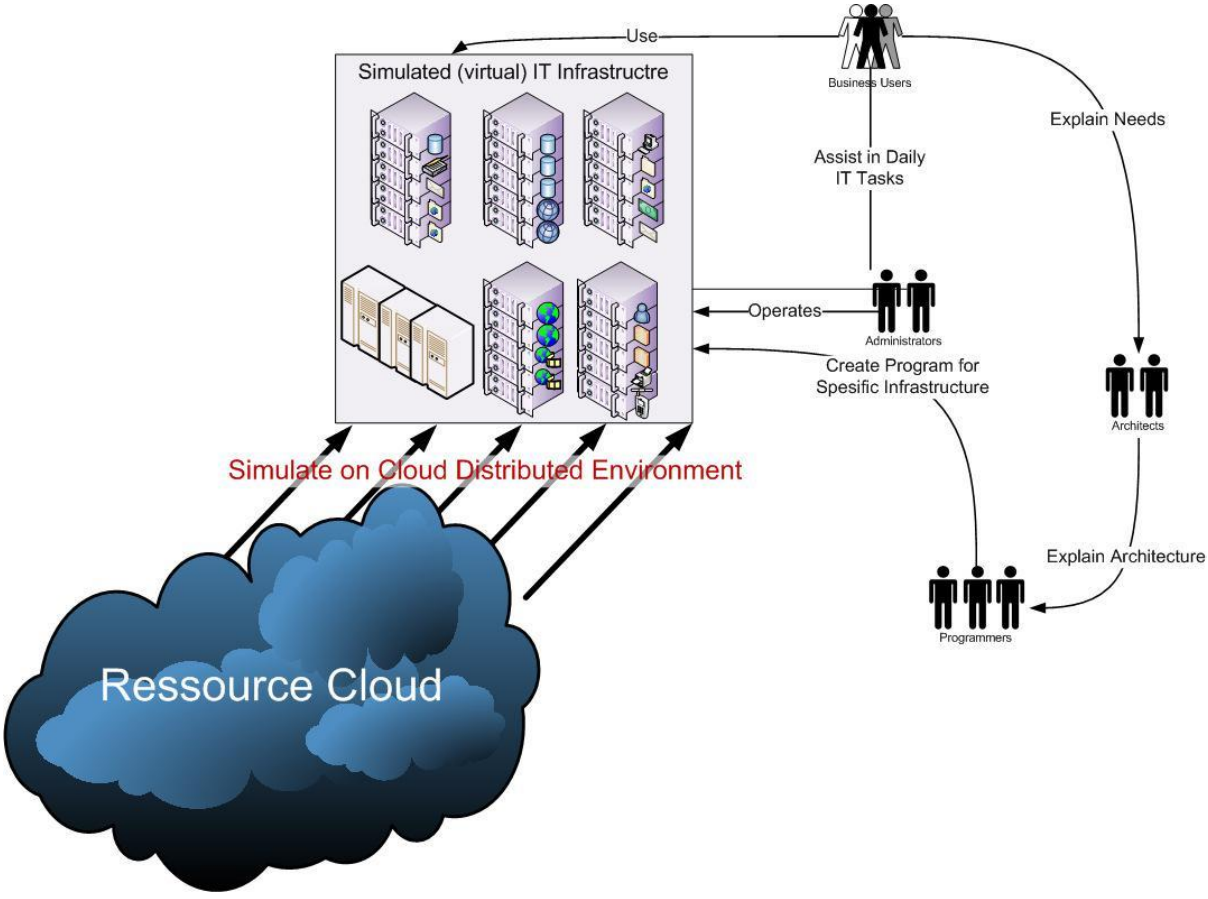


Fig 1.24 Cloud Resource Management

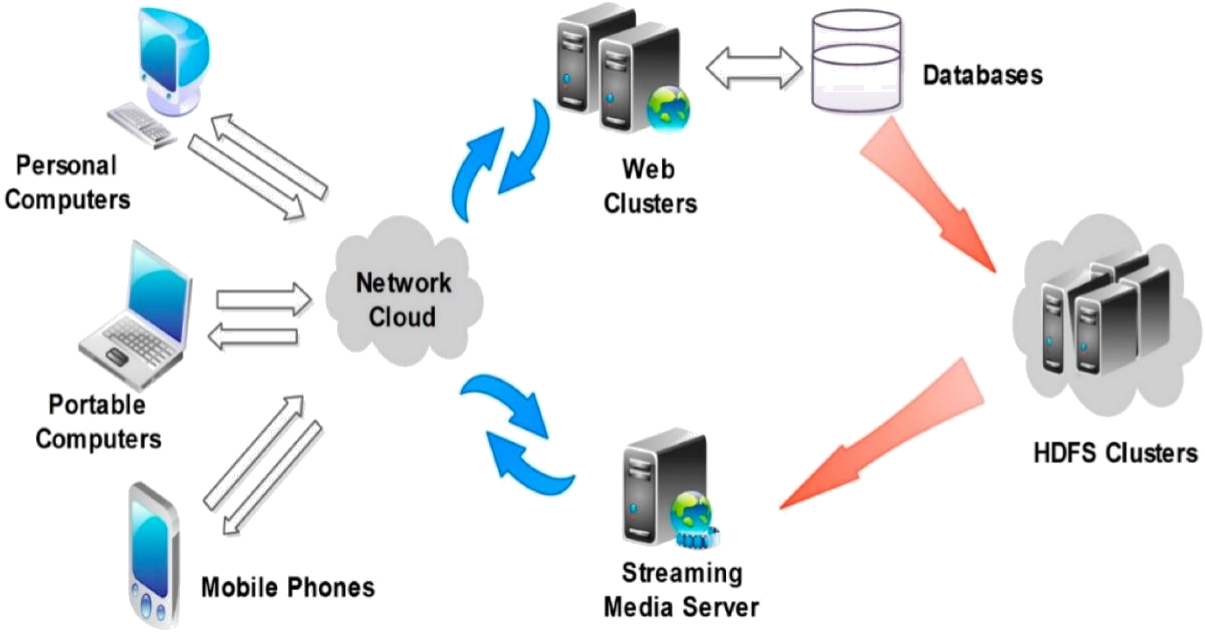


Fig 1.25 Devices connected across the cloud

1.14 Dynamic resource allocation

- Cloud Computing environment can supply of computing resources on the basis of demand and when needed
- Managing the customer demand creates the challenges of on-demand resource allocation.
- Effective and dynamic utilization of the resources in cloud can help to balance the load and avoid situations like slow run of systems.
- Cloud computing allows business outcomes to scale up and down their resources based on needs.
- Virtual Machines are allocated to the user based on their job in order to reduce the number of physical servers in the cloud environment
- If the VM is available then job is allowed to run on the VM.
- If the VM is not available then the algorithm finds a low priority job taking into account the job's lease type.
- The low priority job is paused its execution by pre-empting its resource.
- The high priority job is allowed to run on the resources pre-empted from the low priority.
- When any other job running on VMs are completed, the job which was paused early can be resumed if the lease type of the job is suspendable.

- If not, the suspended job has to wait for the completion of high priority job running in its resources, so that it can be resumed.
- There are three types

  Cancellable: These requests can be scheduled at any time after their arrival time

  Suspendable: Suspendable leases are flexible in start time and can be scheduled at any time after their ready time

  Non-Preemptable: The leases associated with such requests cannot be pre-empted at all.

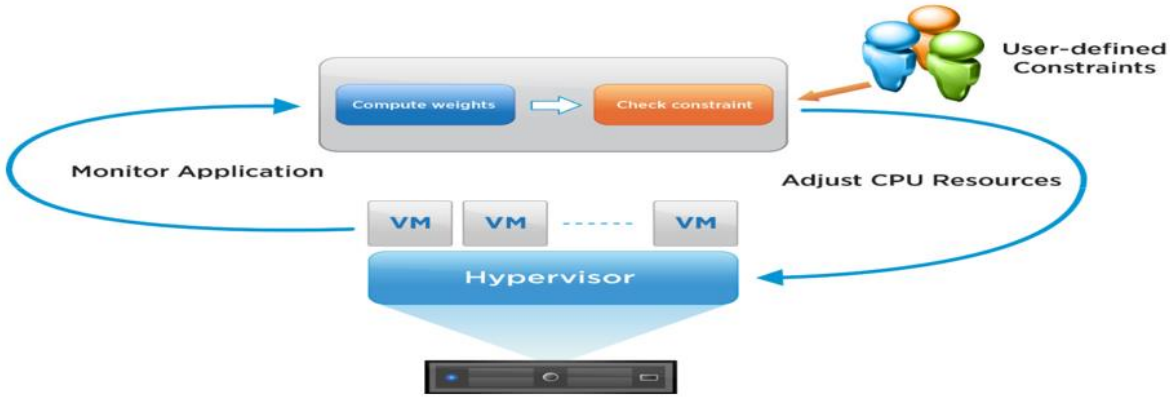


Fig 1.25 Dynamic Resource Allocation

1.15 Optimal allocation of cloud models

- The optimal allocation of computing resources is a core part for implementing cloud computing.
- High heterogeneity, high dynamism, and virtualization make the optimal allocation problem more complex than the traditional scheduling problems in grid system or cloud computing system

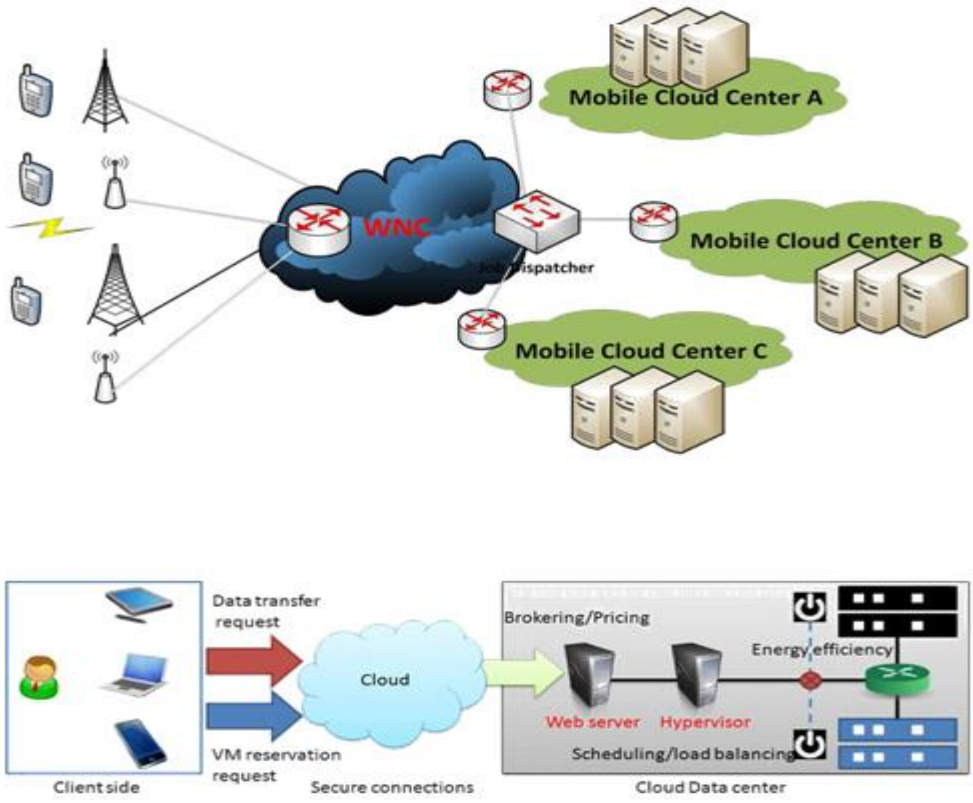


Fig 1.26 Optimal Allocation of Cloud

1.16 Web Services

Web services are XML-centered data exchange systems that use the internet for A2A (application-to-application) communication and interfacing. These processes involve programs, messages, documents, and/or objects.

Functions of Web Services:

- Available over the internet or intranet networks
- Standardized XML messaging system
- Independent of a single operating system or programming language
- Self-describing via standard XML language
- Discoverable through a simple location method

Types of Web Services:

XML-RPC (Remote Procedure Call) is the most basic XML protocol to exchange data between a wide variety of devices on a network. It uses HTTP to quickly and easily transfer data and communication other information from client to server.

UDDI (Universal Description, Discovery, and Integration) is an XML-based standard for detailing, publishing, and discovering web services. It's basically an internet registry for businesses around the world. The goal is to streamline digital transactions and e-commerce among company systems.

SOAP, is an XML-based Web service protocol to exchange data and documents over HTTP or SMTP (Simple Mail Transfer Protocol). It allows independent processes operating on disparate systems to communicate using XML.

RESTprovides communication and connectivity between devices and the internet for API-based tasks. Most RESTful services use HTTP as the supporting protocol.

Web services which are using markup languages:

- Web template
- JSON-RPC
- JSON-WSP
- Web Services Description Language (WSDL)
- Web Services Conversation Language (WSCL)
- Web Services Flow Language (WSFL)
- Web Services Metadata Exchange (WS-MetadataExchange)
- XML Interface for Network Services (XINS)

WSDL

- WSDL stands for Web Services Description Language
- WSDL is used to describe web services
- WSDL is written in XML
- WSDL is a W3C recommendation from 26. June 2007

WSDL Element Type	**Description**
<types>	Defines the (XML Schema) data types use by the web service

<message>	Defines the data elements for eac operation
<portType>	Describes the operations that can b performed and the messages involved.
<binding>	Defines the protocol and data format fo each port type

UDDI:

UDDI is an XML-based standard for describing, publishing, and finding web services.

- UDDI stands for **Universal Description, Discovery, and Integration.**
- UDDI is a specification for a distributed registry of web services.
- UDDI is a platform-independent, open framework.
- UDDI can communicate via SOAP, CORBA, Java RMI Protocol.
- UDDI uses Web Service Definition Language(WSDL) to describe interfaces to web services.
- UDDI is seen with SOAP and WSDL as one of the three foundation standards of web services.
- UDDI is an open industry initiative, enabling businesses to discover each other and define how they interact over the Internet.

  UDDI has two sections −

- A registry of all web service's metadata, including a pointer to the WSDL description of a service.
- A set of WSDL port type definitions for manipulating and searching that registry.

References:

1. Cloud concepts : https://en.wikipedia.org/
2. Web Services: https://www.cleo.com/blog/knowledge-base-web-services
3. WSDL: https://www.w3schools.com/xml/xml_wsdl.asp
4. UDDI: https://www.tutorialspoint.com/uddi/uddi_overview.htm
5. Cloud types: https://www.vxchnge.com/blog/different-types-of-cloud-computing

SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

SCHOOL OF COMPUTING

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

UNIT – II – Cloud computing – SCSA7023

UNIT 2

Cloud Service Models

Topics:

Software as a Service (SaaS) - Infrastructure as a Service (IaaS)- Platform as a Service (PaaS)- Service Oriented Architecture (SoA) - Elastic Computing - On Demand Computing - Parallelization in Cloud Computing – cloud resource management – dynamic resource allocation- Optimal allocation of cloud models.

2.0Cloud Services

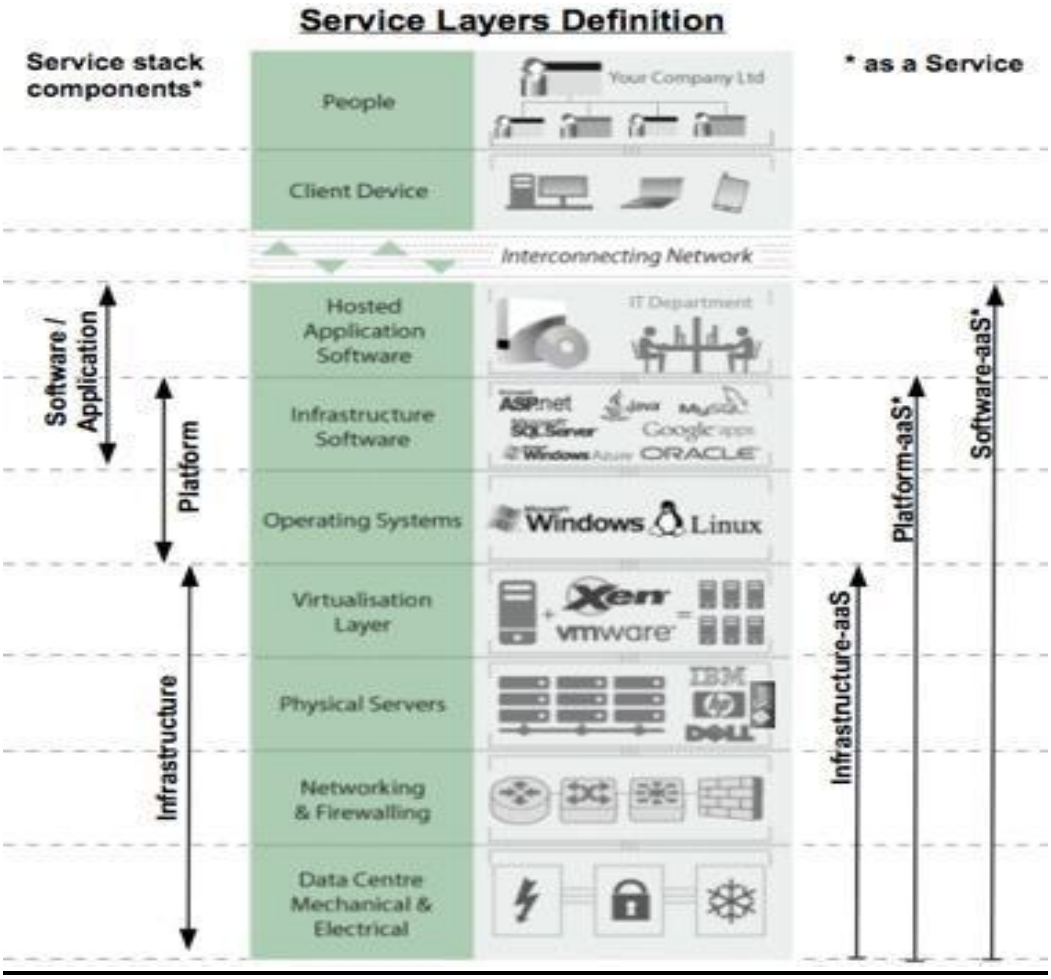
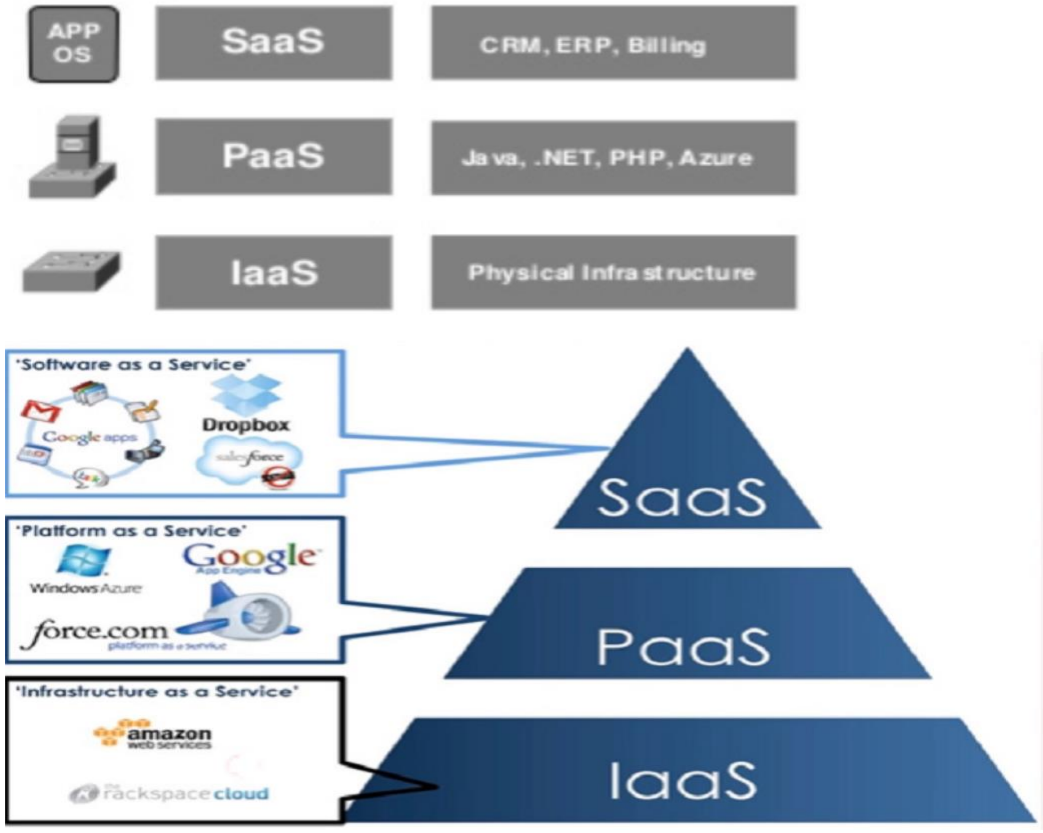


Fig 2.1 Cloud Services

Software as a Service

- SaaS (software-as-a-service). WAN-enabled application services (e.g., Google
- Software as a Service (SaaS) This is a public cloud service model where the application is 100% managed by the cloud provider.
- SaaS removes the need for organizations to install and run applications on their own computers or in their own data centers.
- This eliminates the expense of hardware acquisition, provisioning and maintenance, as well as software licensing, installation and support.
- Software-as-a-Service (SaaS) has evolved from limited on-line software delivery in 1990s to a fully matured “direct-sourcing” business model for enterprise applications.
- SaaS is one of the fastest growing concepts: more than 10 million companies will be using SaaS in the next 5 - 10 years; more than 50% of all Fortune 500 companies are already using SaaS.
- According to influential IT institutes, SaaS is the leading business model of choice for 2008/2009
- Virtually all big software/service vendors (IBM, Microsoft, Oracle, Cisco) are investing heavily in SaaS

- With the continuously increasing bandwidth and reliability of the internet, using web services over the (public) internet has become a viable option.
- Microsoft Office 365 is available with the Azure cloud platform.



The architecture uses an application instance instead of server instances. There is no actual migration of company servers to the cloud. The SaaS model provides single-tenant and multiple tenant services.

The single-tenant dedicates the application instance to the assigned tenant. The multiple tenant application is shared by multiple tenants. The company can manage the security and storage with the single-tenant model. The SaaS application is well suited to internet connectivity. The employees along with their partners and customers can access the application with a variety of network access devices. The SaaS billing model is based on either per usage or monthly subscription. O The security compliance requirements for some applications prevented deployment to the SaaS cloud.

Some SaaS providers offer Virtual Private SaaS (VPS) services for additional application security. It is a hybrid deployment model that allows peering with an enterprise or VPC database server.

The peering is for storage purposes only and used for security compliance. Salesforce.com is a leading SaaS provider with a CRM application to customers.

Benefits of SaaS

- Flexible payments
- Scalable usage
- Automatic updates
- Accessibility and persistence
- On Demand Computing

Opportunities of SaaS

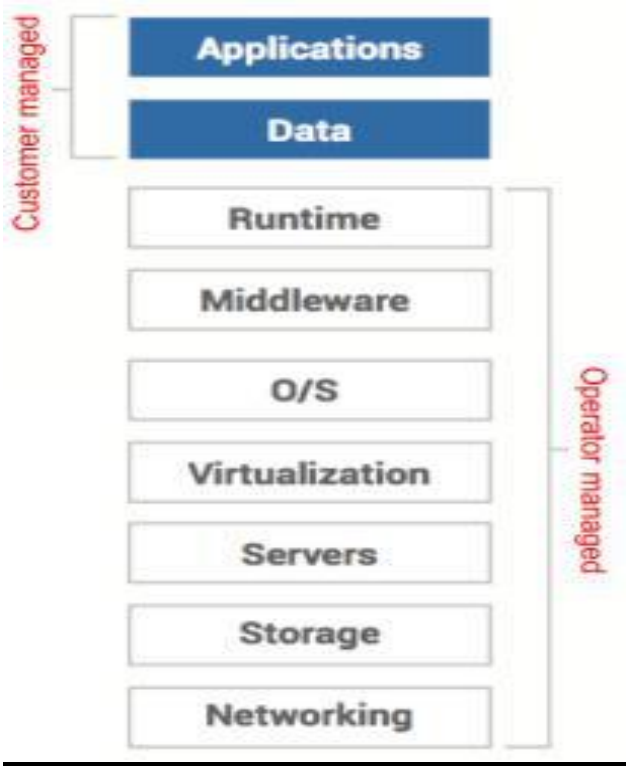
- Software provided as a service by a software vendor to multiple customers with the following main characteristics:
  - Standardization of software
  - Service including maintenance, support and upgrades
- Web based – usage over the (public) internet
- SaaS offers potential for loouring the Total Cost of Ownership
- Loour operational costs
- No large scale, costly, high risk implementations of applications
- Need few operational resources for application management
- No platform and hardware (maintenance) costs for application servers
- Reduced operational complexity: software delivered as a transparent service through the web
- Minimized software development costs – No lengthy software development and testing cycles
- Loour costs for software use
- No software license and annual maintenance fees
- No expensive software upgrades
- Loour application consultancy and support costs
- SaaS allows corporations to focus on core business activities and responsibilities

- Transparent overview and usage of electronic data and information
- Automation of iterative, manual tasks
- Faster Time to Market – easy to scale software
- More flexibility in changing and modifying application services for business needs – Full scale integration of business processes
- Control over IT
- Minimized IT Service Management efforts mainly focused on availability – Well-defined SLAs between the corporation and the IT vendor
- More predictable cash flow – easier licensing based on access/usage of software
- Increased productivity and improved user satisfaction O Automatic software upgrades with minimal outage

Limitations

Businesses must rely on outside vendors to provide the software, keep that software up and running, track and report accurate billing and facilitate a secure environment for the business' data.

Platform as a Service



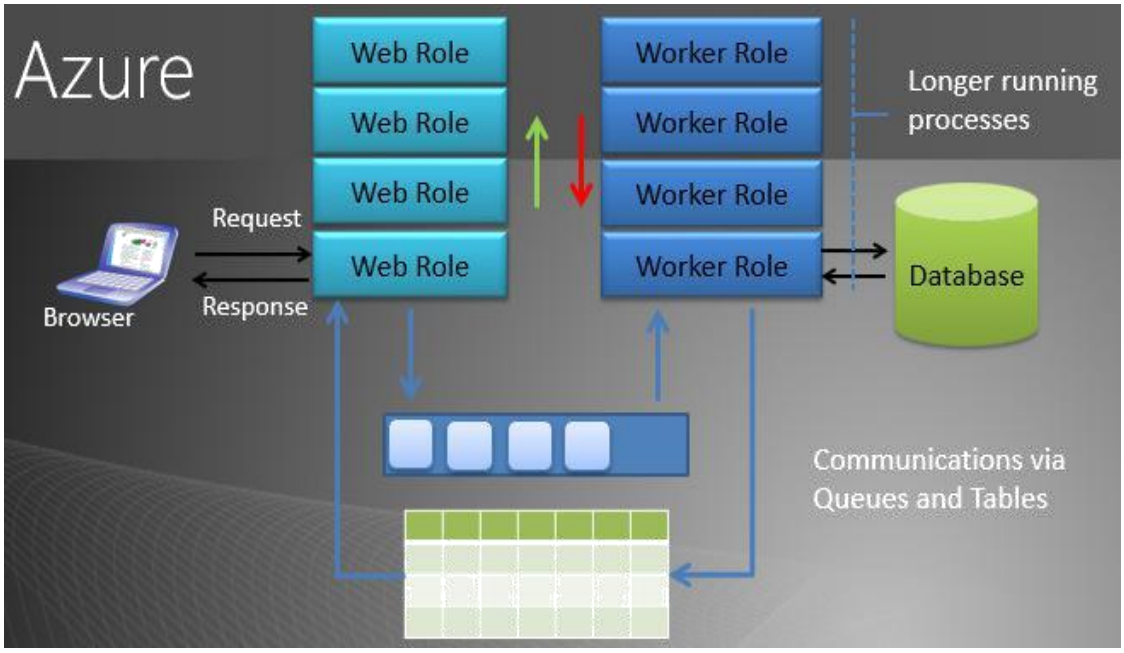
1. The Platform as a Service (PaaS) is a way to rent hardware, operating systems, storage and network capacity over the Internet.
2. PaaS services are,
    - Data services
    - Application runtime
    - Messaging & queueing
    - Application management.

- The PaaS is a computing platform that abstracts the infrastructure, OS, and middle-ware to drive developer productivity.
- The PaaS is foundational elements to develop new applications
- E.g., Google Application Engine, Microsoft Azure, Coghead.
-
- **Microsoft Azure**

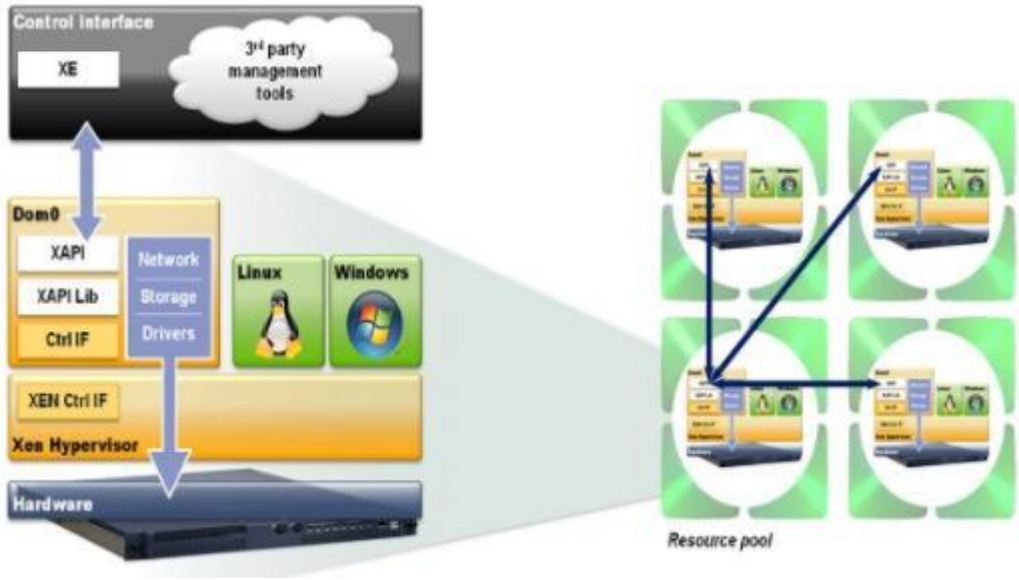
  Pay per role instance

  Add and remove instances based on demand

    - Elastic computing!
    - Load balancing is part of the Azure fabric and automatically allocated



- The PaaS is the delivery of a computing platform and solution stack as a service
- **The Solution stack** is integrated set of software that provides everything a developer needs to build an application for both software development and runtime.



-PaaS offers the following

Facilities for application design Application development

Application testing, deployment Application services are,

- Operating system
- Server-side scripting environment

- Database management system
- Server Software

- Support
- Storage
- Network access
- Tools for design and development
- Hosting

All these services may be provisioned as an integrated solution over the web

Properties and characteristics of PaaS

Scalability

Availability

Manageability

Performance

Accessibility

PaaS Features

It delivers the computing platform as service

The capacities to abstract and control all the underlying resources

It helps to providers any smallest unit of resources

To provide a reliable environment for running applications and services

Act as a bridge between consumer and hardware

Do not need to care about how to build, configure, manage and maintain the backend environment

It provides a development and testing platform for running developed applications

Reduce the responsibility of managing the development and runtime environment

Advantages of PaaS

It helps to provide deployment of application without the cost and complexity of buying and managing the hardware and software

It provides all the required to support the complete life cycle of building and delivering web applications and services entirely available from the internet

Disadvantages of PaaS

Less flexible than IaaS

Dependency on provider

Adoption of software / system architecture required

Evolving from different standards

Evolving “upwards” from IaaS

 Amazon (Mail, Notification, Events, Databases, Workflow, etc.) Evolving “downwards” from SaaS

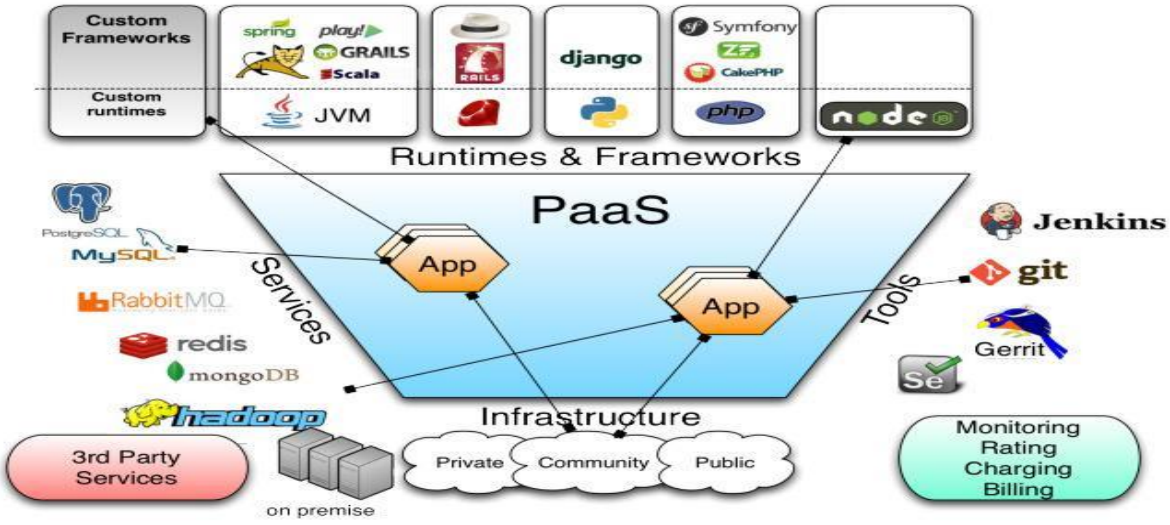
 Force.com – a place to host additional per-tenant logic.

 Google App Engine

Evolving “sideways” from middleware platforms

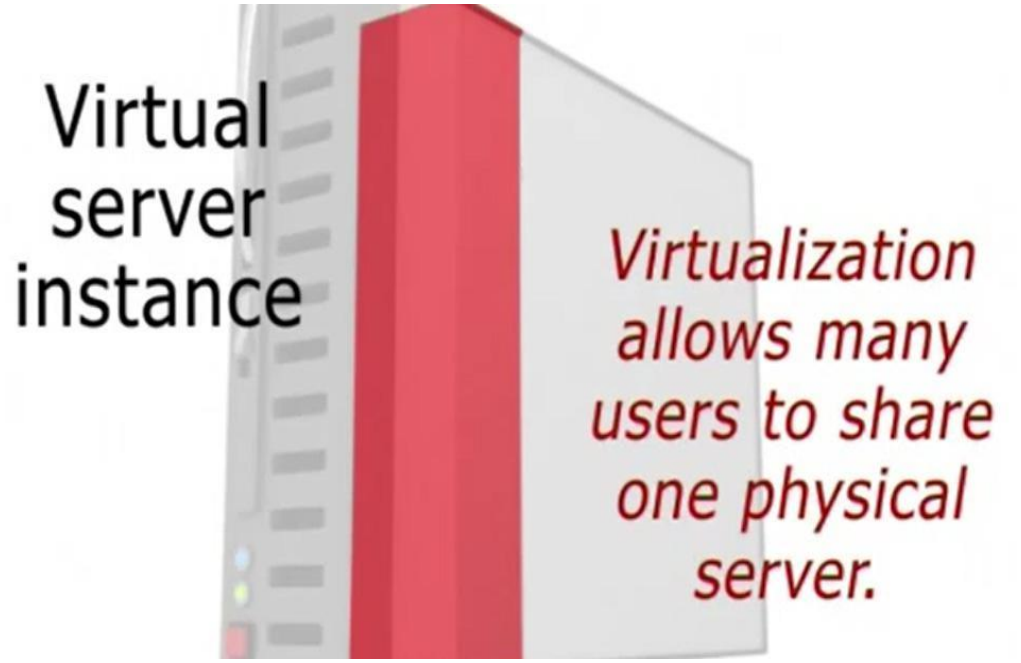
 WSO2, Tibco, vmWare, Oracle, IBM

Generic PaaS Model



Infrastructure as a Service

This service offers the computing architecture and infrastructure i.e. all types of computing resources. All resources are offered in a virtual environment, so that multiple users can access it.

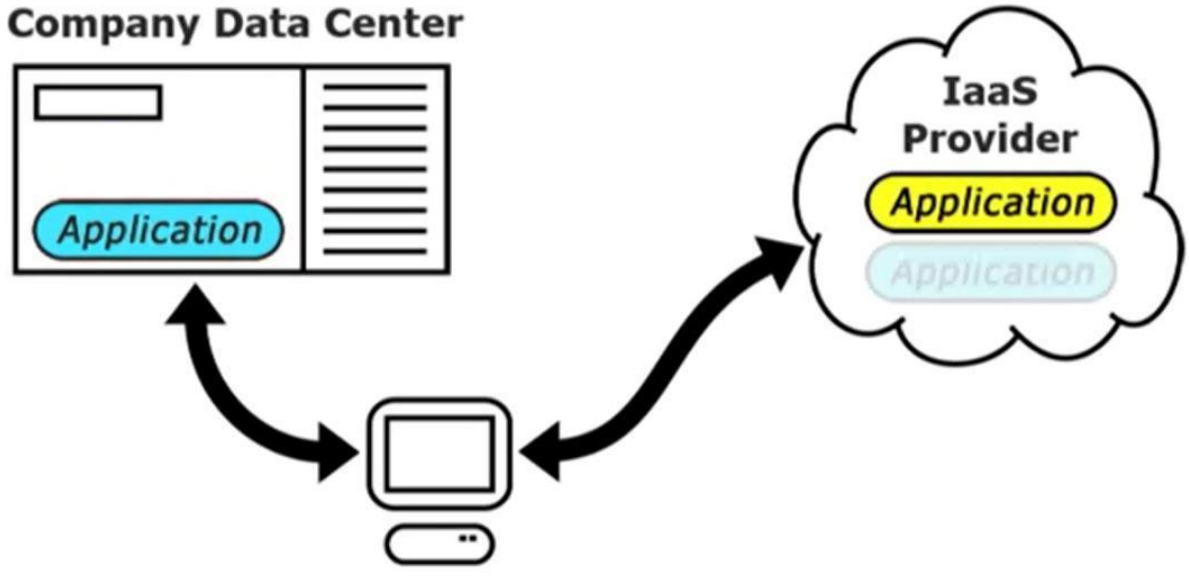


The resources are including, Data storage

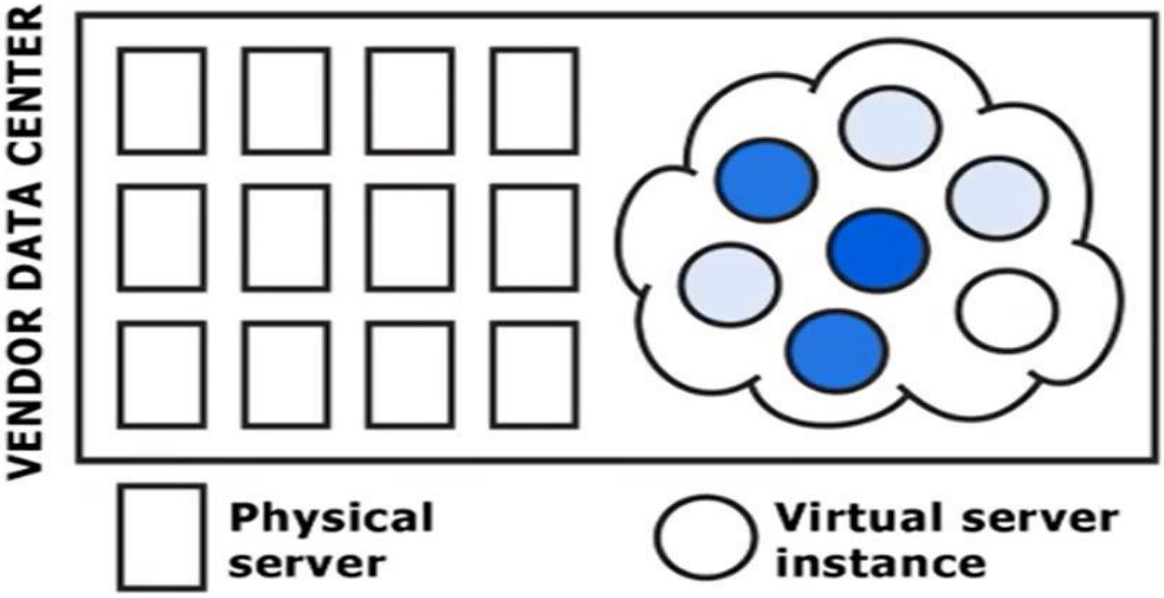
Virtualization Servers

Networking

- The vendors are responsible for managing all the computing resources which they provided.
- It allows existing applications to be run on a supplier's hardware.



User Task in IaaS Cloud



Multiple user can access Virtual instances

- The user responsible for handling other resources such as,
  - Applications
  - Data
  - Runtime
  - Middleware

- Example IaaS service providers

- AWS EC2 / S3 / RDS
- GoGrid
- RackSpace

- **Pros**

  The cloud provides the infrastructure

  Enhanced scalability i.e. dynamic workloads are supported

  It is flexible

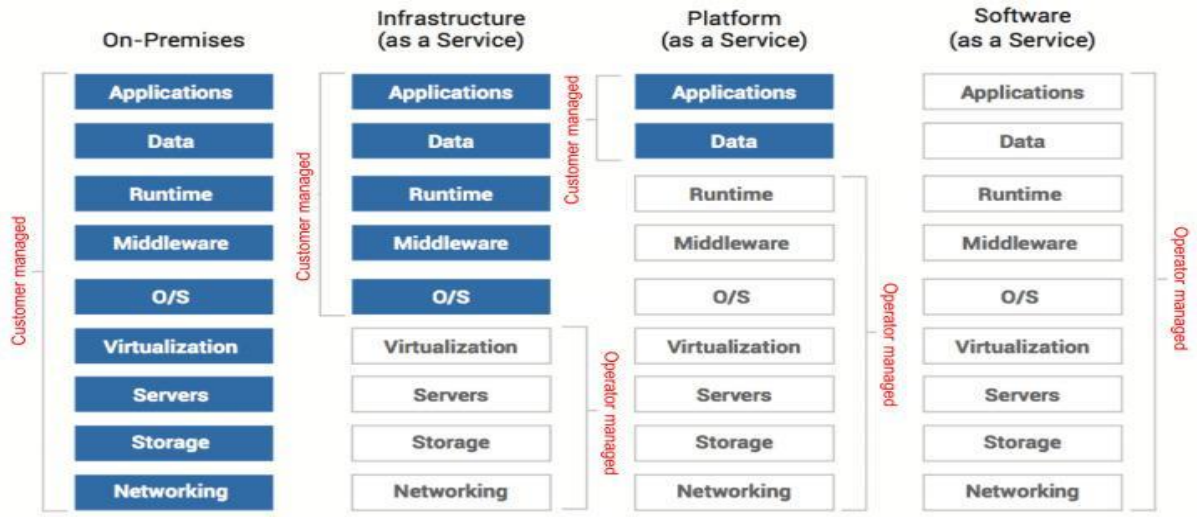
- **Cons**

Security issues

Network and service delay

Comparison of cloud services

Blue indicates the levels owned and operated by the organization / Customer White levels are run and operated by the service provider / Operator



Cloud Computing Services

Pros

1. Loour computer costs
2. Improved performance:
3. Reduced software costs
4. Instant software updates
5. Improved document format compatibility
6. Unlimited storage capacity
7. Increased data reliability
8. Universal document access
9. Latest version availability
10. Easier group collaboration
11. Device independence

Cons

- Requires a constant Internet connection
- Does not work well with low-speed connections
- Features might be limited
- Can be slow
- Stored data can be lost
- Stored data might not be secure

Service Oriented Architecture

- **Service**

  A service is a program we interact with via message exchanges

  A system is a set of deployed services cooperating in a given task

Architecture

It serves as the blueprint for the system

Team structure

Documentation organization

Work breakdown structure

Scheduling, planning, budgeting

Unit testing, integration

Architecture establishes the communication and coordination mechanisms among components

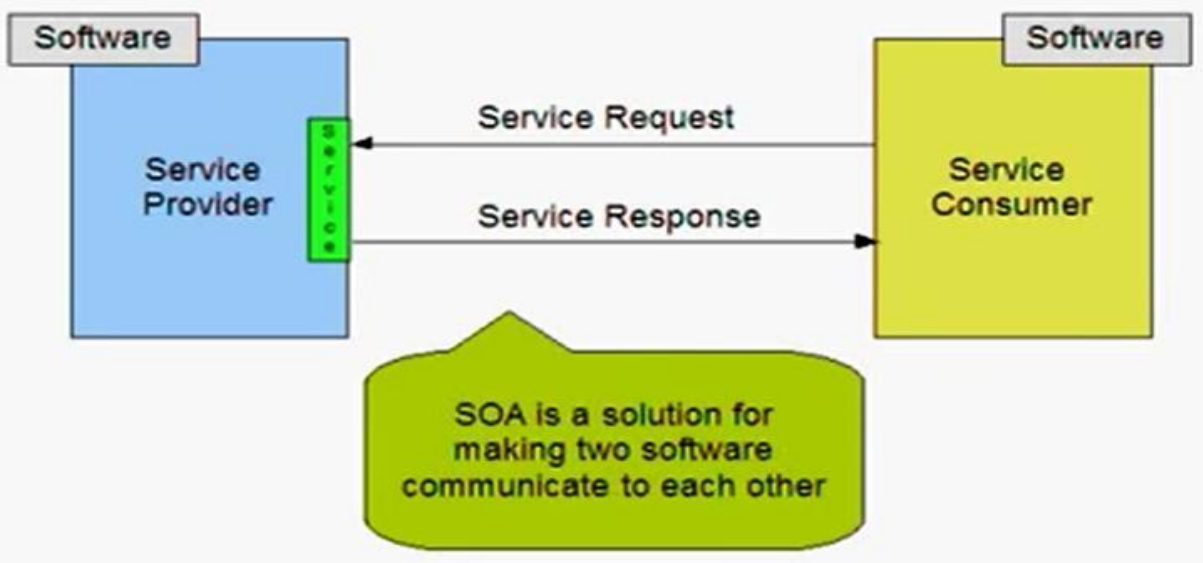
- **Software Architecture**

  It is collection of the fundamental decisions about a software product/solution designed to meet the project's quality attributes (i.e. requirements).

  The architecture includes the main components, their main attributes, and their collaboration (i.e. interactions and behavior) to meet the quality attributes.

  Architecture can and usually should be expressed in several levels of

abstraction (depending on the project's size). Architecture is communicated from multiple viewpoints



Why SOA?

- **SOA**

  SOA stands for Service Oriented Architecture

  It is a design pattern or software architecture which provides application functionality as a service to other applications.

  The basic principles of service-oriented architecture are independent of vendors, products and technologies.

The services are provided to the other components through a communication protocol over a network.

Every service has its own business logic

- **SOA Architecture**

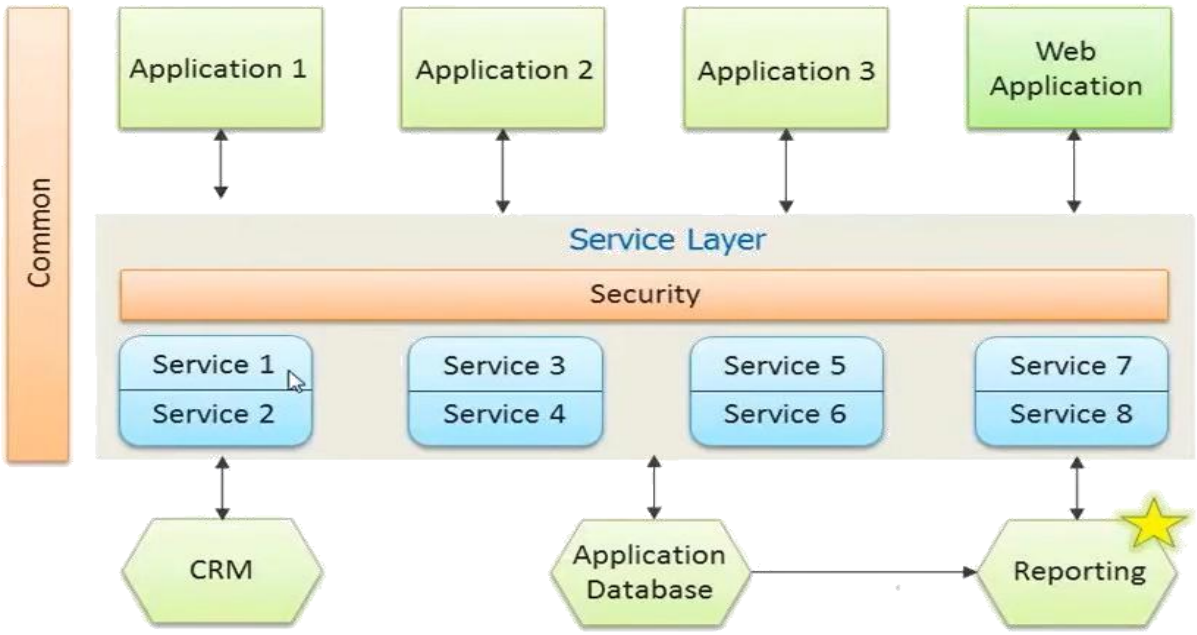
Consumer interface layer – this layer is used by the customer

Business process layer – it provides the business process flow

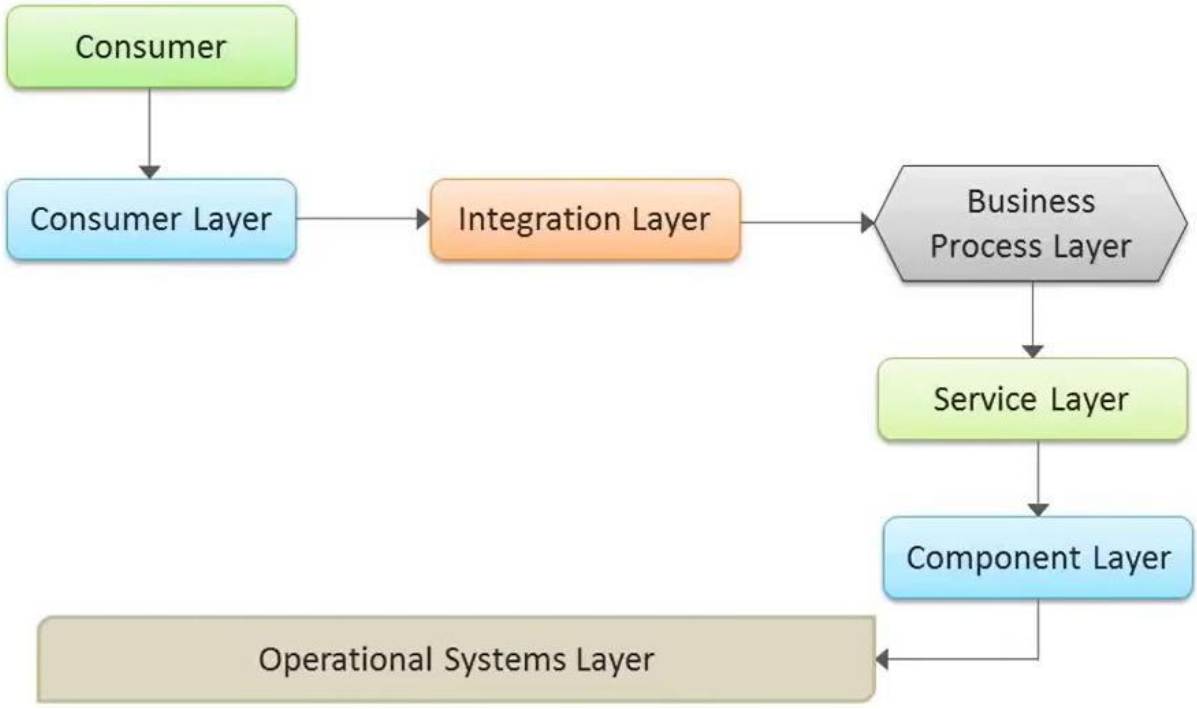
Service layer – this layer comprises of all the services in the enterprises

Component layer – this layer has the actual service to be provided

Operational system layer – this layer contains the data model



SOA Architecture



SOA – Architecture in details

Principles of SOA

Service loose coupling – service does not have high dependency implementation from outside world

Service reusability – services can be used again and again instead of rewriting them

Service statelessness – they usually do not maintain the state to reduce the resource consumption

Service discoverability – services are registered in registry, so that the client can discover them in the service registry.

Applications

Manufacturing – E.g. Inventory management

Insurance – Take up the insurance of the employees in companies

Companies using SOA

Banking Sector

- ICICI Bank
- HDFC Bank

- UTI Bank etc..

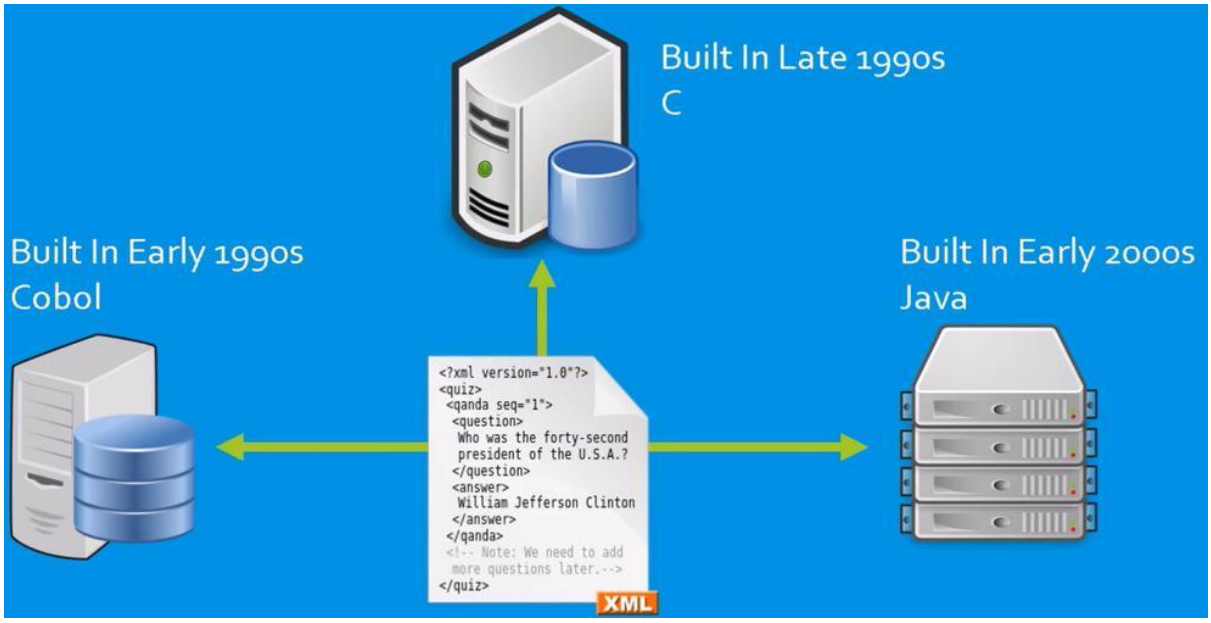
Manufacturing Sector

- Apollo Tyres
- Maruthi
- Hyundai

Advantages

Interoperability

- Programs to run different vendors / locations
- To interact with different networks
- Different operating systems
- Solution: XML

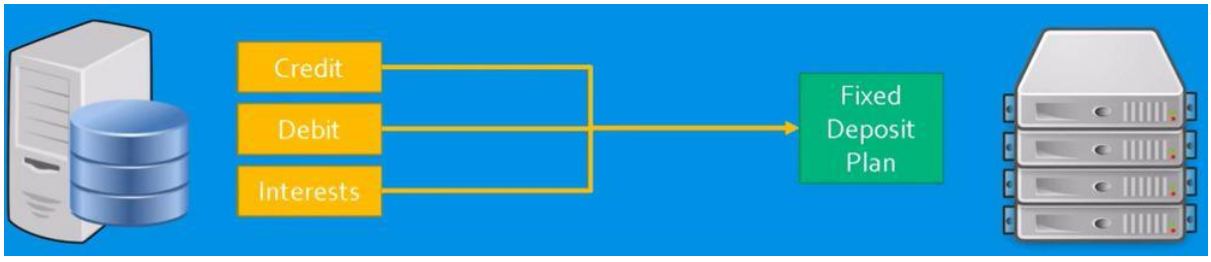


Scalability

- To extend the processing poour of the servers

Reusability

- If any new systems are introduced, no need to create a new service for every time.



- Parallel application development
- Modular approach
- Easy maintenance
- Greater Reliability
- Improved Software Quality
- Platform Independence
- Increased Productivity

Disadvantages

- Stand alone, non-distributed applications
- Homogenous application environments
- GUI based applications
- Short lived applications
- Real time applications
- One-way asynchronous communication applications

Elastic Compute Cloud (EC2)

To access the **Elastic Compute Cloud (EC2)** functionality, access the **Amazon EC2** table in the **AWS Console**.

Elastic Compute Cloud (EC2) is the engine room of AWS. This is where our servers will operate and run on a day-to-day basis. However, the 'elastic' in EC2 is there for a reason. EC2 is much more than just a bunch of servers! EC2 provides 'resizable compute capacity', or in other words can scale tremendously depending on our capacity requirements at a particular point in time.

EC2 provides the ability to start and stop multiple servers from a single server image, as well as modifying the number of these instances dynamically.

However, there have been some significant differences in the past on how this is implemented, which requires some up-front planning on how we will use the EC2 environment.

General roles of EC2 in the architecture

EC2 is the backbone of the architecture where our servers are implemented. EC2 will not only run our servers but will manage the capacity that they produce.

Using EC2

To start using EC2 we must start with an EC2 '**bundle**' or **Amazon Machine Image (AMI).** Both Amazon and third parties such as RightScale and IBM provide images. For this project, we will be using the default Windows Server Basic AMIs provided by AWS.

Each AMI is a starting point for our instance. Once we have started our Windows instance, we may need to wait up to 15 minutes for AWS to generate our password, so be patient, before we can log on using **Remote Desktop Protocol (RDP)**.

Once our instance has started and we have RDP'd to it, we now have access to install any software that we need onto this instance. But beware, if we fail to create another bundle from our running instance—which we can use to start it next time—then all of our changes will be lost

This is the major difference between standard instances in EC2 and the servers, which we have been familiar with up to now. When installing software onto a server that exists in our own server room, the software tends to remain installed. If we install software on an Amazon EC2 instance, our software (and data) will disappear when our instance is '**terminated**'.

However, recently Amazon has introduced the concept of persistent EC2 images. These are AMIs, which are created on **Elastic Block Store (EBS)** disk. In this specific instance, changes made to the image are persisted when we '**stop**' the image. However, if we terminate the image, the changes are lost.

On Demand Computing

On-demand computing packages computer resources (processing, storage, and so forth) as a metered service similar to that of a public utility. In this model, customers pay for as much or as little processing and storage as they need.Companies that have large demand peaks followed by much loour normal usage periods particularly benefit from utility computing. The company pays more for their peak usage, of course, but their bills rapidly decline when the peak ends and normal usage patterns resume.

Clients of on-demand computing services essentially use these services as offsite virtual servers. Instead of investing in their own physical infrastructure, a company operates on a pay-as-we-go plan with a cloud services provider. On-demand computing itself is not a new concept, but has acquired new life thanks to cloud computing. In previous years, on-demand computing was provided from a single server via some sort of time-sharing arrangement. Today, the service is based on large grids of computers operating as a single cloud.

Parallelization in Cloud Computing

Over the past decade, scientific and engineering research via computing has emerged as the third pillar of the scientific process, complementing theory and experiment. Several national studies have highlighted the importance of computational science as a critical enabler of scientific discovery and national competitiveness in the physical and biological sciences, medicine and healthcare, and design and manufacturing.

As the term suggests, computational science has historically focused on computation: the creation and execution of mathematical models of natural and artificial processes. Driven by opportunity and necessity, computational science is expanding to encompass both computing and data analysis. Today, a rising tsunami of data threatens to overwhelm us, consuming our attention by its very volume and diversity. Driven by inexpensive, seemingly ubiquitous sensors, broadband networks, and high-capacity storage systems, the tsunami encompasses data from sensors that monitor our planet from deep in the ocean, from land instruments, and

from space-based imaging systems. It also includes environmental measurements and healthcare data that quantify biological processes and the effects of surrounding conditions. Simply put, we are moving from data paucity to a data plethora, which is leading to a relative poverty of human attention to any individual datum and is necessitating machine-assisted winnowing.

This ready availability of diverse data is shifting scientific approaches from the traditional, hypothesis-driven scientific method to science based on exploration. Researchers no longer simply ask, "What experiment could I construct to test this hypothesis?" Increasingly, they ask, "What correlations can I glean from extant data?" More tellingly, one wishes to ask, "What insights could I glean if I could fuse data from multiple disciplines and domains?" The challenge is analyzing many petabytes of data on a time scale that is practical in human terms.

The ability to create rich, detailed models of natural and artificial phenomena and to process large volumes of experimental data created by a new generation of scientific instruments that are themselves pooured by computing makes computing a universal intellectual amplifier, advancing all of science and engineering and poouring the knowledge economy. Cloud computing is the latest technological evolution of computational science, allowing groups to host, process, and analyze large volumes of multidisciplinary data. Consolidating computing and storage in very large datacenters creates economies of scale in facility design and construction, equipment acquisition, and operations and maintenance that are not possible when these elements are distributed. Moreover, consolidation and hosting mitigate many of the sociological and technical barriers that have limited multidisciplinary data sharing and collaboration. Finally, cloud hosting facilitates long-term data preservation -- a task that is particularly challenging for universities and government agencies and is critical to our ability to conduct longitudinal experiments.

It is not unreasonable to say that modern datacenters and modern supercomputers are like twins separated at birth. Both are massively parallel in design, and both are organized as a network of communicating computational nodes. The individual nodes of each are based on commodity microprocessors that have multiple cores, large memories, and local disk storage. They both execute applications that are designed to exploit massive amounts of parallelism. Their differences lie in their evolution. Massively parallel supercomputers have been designed to support computation with occasional bursts of input/output and to complete a single massive

calculation as fast as possible, one job at a time. In contrast, datacenters direct their poour outward to the world and consume vast quantities of input data.

Parallelism can be exploited in cloud computing in two ways. The first is for human access. Cloud applications are designed to be accessed as Web services, so they are organized as two or more layers of processes. One layer provides the service interface to the user's browser or client application. This "Web role" layer accepts users' requests and manages the tasks assigned to the second layer. The second layer of processes, sometimes known as the "worker role" layer, executes the analytical tasks required to satisfy user requests. One Web role and one worker role may be sufficient for a few simultaneous users, but if a cloud application is to be widely used -- such as for search, customized maps, social networks, weather services, travel data, or online auctions -- it must support thousands of concurrent users.

The second way in which parallelism is exploited involves the nature of the data analysis tasks undertaken by the application. In many large data analysis scenarios, it is not practical to use a single processor or task to scan a massive dataset or data stream to look for a pattern -- the overhead and delay are too great. In these cases, one can partition the data across large numbers of processors, each of which can analyze a subset of the data. The results of each "sub-scan" are then combined and returned to the user.

This "map-reduce" pattern is frequently used in datacenter applications and is one in a broad family of parallel data analysis queries used in cloud computing. Web search is the canonical example of this two-phase model. It involves constructing a searchable keyword index of the Web's contents, which entails creating a copy of the Web and sorting the contents via a sequence of map-reduce steps. Three key technologies support this model of parallelism: Google has an internal version, Yahoo! has an open source version known as Hadoop, and Microsoft has a mapreduce tool known as DryadLINQ. Dryad is a mechanism to support the execution of distributed collections of tasks that can be configured into an arbitrary directed acyclic graph (DAG). The Language Integrated Query (LINQ) extension to C# allows SQL-like query expressions to be embedded directly in regular programs. The DryadLINQ system can automatically compile these queries into Dryad DAG, which can be executed automatically in the cloud.

Microsoft Windows Azure supports a combination of multi-user scaling and data analysis parallelism. In Azure, applications are designed as stateless "roles" that fetch tasks from queues, execute them, and place new tasks or data into other queues. Map-reduce computations

in Azure consist of two pools of worker roles: mappers, which take map tasks off a map queue and push data to the Azure storage, and reducers, which look for reduce tasks that point to data in the storage system that need reducing. Whereas DryadLINQ executes a static DAG, Azure can execute an implicit DAG in which nodes correspond to roles and links correspond to messages in queues. Azure computations can also represent the parallelism generated by very large numbers of concurrent users.

This same type of map-reduce data analysis appears repeatedly in large-scale scientific analyses. For example, consider the task of matching a DNA sample against the thousands of known DNA sequences. This kind of search is an "embarrassingly parallel" task that can easily be sped up if it is partitioned into many independent search tasks over subsets of the data. Similarly, consider the task of searching for patterns in medical data, such as to find anomalies in fMRI scans of brain images, or the task of searching for potential weather anomalies in streams of events from radars.

Finally, another place where parallelism can be exploited in the datacenter is at the hardware level of an individual node. Not only does each node have multiple processors, but each typically has multiple computer cores. For many data analysis tasks, one can exploit massive amounts of parallelism at the instruction level.

For example, filtering noise from sensor data may involve invoking a Fast Fourier Transform (FFT) or other spectral methods. These computations can be sped up by using general-purpose graphics processing units (GPGPUs) in each node. Depending on the rate at which a node can access data, this GPGPU-based processing may allow us to decrease the number of nodes required to meet an overall service rate.

The World Wide Web began as a loose federation of simple Web servers that each hosted scientific documents and data of interest to a relatively small community of researchers. As the number of servers grew exponentially and the global Internet matured, Web search transformed what was initially a scientific experiment into a new economic and social force. The effectiveness of search was achievable only because of the available parallelism in massive datacenters. As we enter the period in which all of science is being driven by a data explosion, cloud computing and its inherent ability to exploit parallelism at many levels has become a fundamental new enabling technology to advance human knowledge.

Cloud Resource Management

Service providers seek scalable and cost-effective cloud solutions for hosting their applications. Despite significant recent advances facilitating the deployment and management of services on cloud platforms, a number of challenges still remain. Service providers are confronted with time-varying requests for the provided applications, inter- dependencies between different components, performance variability of the procured virtual resources, and cost structures that differ from conventional data centers. Moreover, fulfilling service level agreements, such as the throughput and response time percentiles, becomes of paramount importance for ensuring business advantages.In this thesis, we explore service provisioning in clouds from multiple points of view. The aim is to best provide service replicas in the form of VMs to various service applications, such that their tail throughput and tail response times, as well as resource utilization, meet the service level agreements in the most cost effective manner. In particular, we develop models, algorithms and replication strategies that consider multi-tier composed services provisioned in clouds. We also investigate how a service provider can opportunistically take advantage of observed performance variability in the cloud. Finally, we provide means of guaranteeing tail throughput and response times in the face of performance variability of VMs, using Markov chain modeling and large deviation theory. We employ methods from analytical modeling, event-driven simulations and experiments. Overall, this thesis provides not only a multi-faceted approach to exploring several crucial aspects of hosting services in clouds, i.e., cost, tail throughput, and tail response times, but our proposed resource management strategies are also rigorously validated via trace-driven simulation and extensive experiments.

Dynamic Resource Allocation

Cloud computing allows business customers to scale up and down their resource usage based on needs. Many of thetouted gains in the cloud model come from resource multiplexing through virtualization technology. A system that uses virtualization technology to allocate data center resources dynamically based on application demands and support greencomputing by optimizing the number of servers in use . The concept of “skewness” is to measure the unevenness in t hemulti-dimensional resource utilization of a server. By minimizing skewness, we can combine different types of workloads nicely andimprove the overall utilization of server resources.A set of heuristics is developed that prevent overload in the system effectively

whilesaving energy used. Trace driven simulation and experiment results demonstrate that our algorithm achieves good performance.

Optimal allocation of cloud models

Cloud computing is a prominent technology that becomes popular in providing on demand resources to the end users on payment basis via internet. The major challenge of any cloud service provider is to allocate the suitable resource packages efficiently that meet the user's need. Because of fluctuation in user's requirements, it is very challenging task to allocate cost effective resources to the consumers or end users as per the need. Hence it is mandatory to have an efficient resource allocation techniques that fulfilled the user's requirements to achieve QoS (Quality of Service).

Cost Optimization Strategies:

- Resource Rightsizing
- Instance Scheduling
- Instance Reservation
- Reserved Instance (RI) Utilization
- Terminate Unassociated Resources
- Identify Under-utilized Resources

Cost Allocation Strategies:

- Visualize tags
- Define tagging nomenclatures
- Identify incorrect tags
- Auto-correct tags

References

1. Dennis Gannon and Dan Reed , Parallelism and the Cloud, October 16, 2009, https://www.drdobbs.com/parallel/parallelism-and-the-cloud/220601206

SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

SCHOOL OF COMPUTING

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

UNIT – III – Cloud Computing– SCSA7023

Unit 3

Cloud Deployment Models

Topics:

Deployment models: Public cloud – Private Cloud –Hybrid cloud – Community cloud - Deployment of applications on the cloud - Hypervisor -Case studies- Xen, Hyper V, Virtual Box, Eucalyptus, Amazon Cloud Computing - Amazon S3 – Amazon EC2-, Windows Azure – creating deploying cloud services, Working with Google App engine, Disaster Recovery.

Deployment models

A deployment model defines the purpose of the cloud and the nature of how the cloud is located. The NIST definition for the four deployment models is as follows:

• **Public cloud:** The public cloud infrastructure is available for public use alternatively for a large industry group and is owned by an organization selling cloud services.

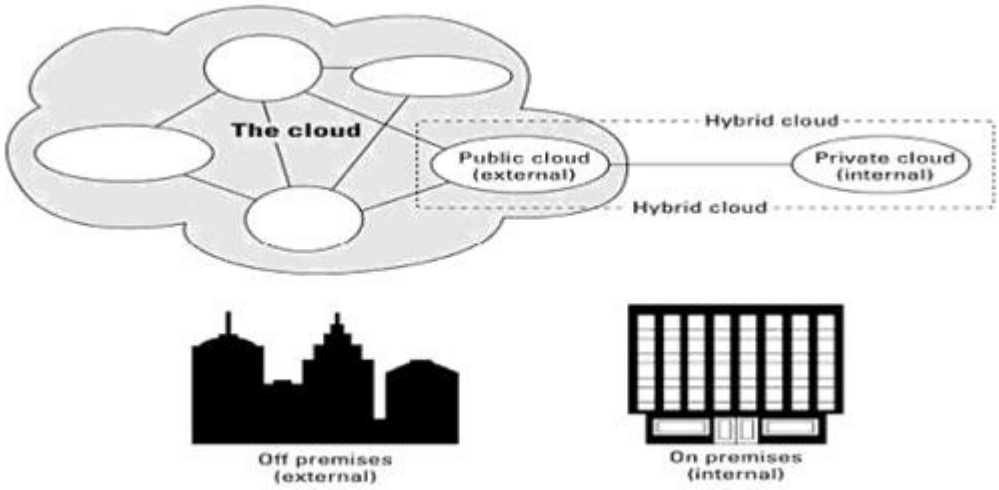
• **Private cloud:** The private cloud infrastructure is operated for the exclusive use of an organization. The cloud may be managed by that organization or a third party. Private clouds may be either on- or off-premises.

• **Hybrid cloud:** A hybrid cloud combines multiple clouds (private, community of public) where those clouds retain their unique identities, but are bound together as a unit. A hybrid cloud may offer standardized or proprietary access to data and applications, as well as application portability.

• **Community cloud:** A community cloud is one where the cloud has been organized to serve a common function or purpose.

It may be for one organization or for several organizations, but they share common concerns such as their mission, policies, security, regulatory compliance needs, and so on. A community cloud may be managed by the constituent organization(s) or by a third party.

The following figure shows the different locations that clouds can come in. In the sections that follow, these different cloud deployment models are described in more detail.



Deployment locations for different cloud types

Cloud Deployment of Applications on the cloud

Deployment to the Cloud

Cloud deployment refers to the enablement of SaaS (software as a service), PaaS (platform as a service) or IaaS (infrastructure as a service) solutions that may be accessed on demand by end users or consumers. A cloud deployment model refers to the type of cloud computing architecture

a cloud solution will be implemented on. Cloud deployment includes all of the required installation and configuration steps that must be implemented before user provisioning can occur.

SAAS Deployment & Cloud Deployment Models

Cloud deployment can be viewed from the angle of management responsibility for the deployment of the SaaS, PaaS and/or IaaS solutions in question. From this perspective, there are two possible approaches: the cloud solution(s) may be deployed by a third party (under a community cloud, public cloud or private cloud deployment model) or the cloud solution(s) may be deployed by a single entity (under a private cloud deployment model).

SaaS deployment is a type of cloud deployment that is typically initiated using a public cloud or a private cloud deployment model, however SaaS deployment may also be initiated using a hybrid cloud deployment model, when hybrid cloud resources are owned and/or managed by the same entity. Expanding on this theme is the existence of virtual private clouds that can be used for SaaS deployment as well. Virtual private clouds are technically public clouds that function the same as private clouds, since only trusted entities may gain access to the virtual private cloud resources.

Regardless of whether or not a SaaS solution is deployed in a public cloud, a private cloud , a virtual private cloud or a hybrid cloud; many SaaS solutions provide automatic deployment for the cloud services being delivered. SaaS deployment provides many additional benefits over the traditional model of software deployment, including scalability, where application users can be added or subtracted on demand without concerns over capital investments in additional hardware or software. SaaS deployment also provides above average up-time for enterprise applications as compared to on premise software deployment.

After cloud deployment has been completed for a SaaS, PaaS or IaaS solution, user provisioning can occur based on user permissions, where access is provided for cloud resources based on the consumer's classification as either a trusted or untrusted entity. Trusted entities may receive access permission to managed cloud, private cloud or hybrid cloud resources. Untrusted entities may receive access permission to public cloud, managed cloud or hybrid cloud resources. The key difference between trusted and untrusted entities is that untrusted entities never receive access permission to private cloud resources.

Hypervisor

Software that controls the layer between the hardware operating systems. It allows multiple operating systems to run on the same physical hardware. There are two types of hypervisors:

- Bare metal, which allows the hypervisor to run directly on the hardware
- Hosted architecture, in which the hypervisor runs on top of an existing operating system

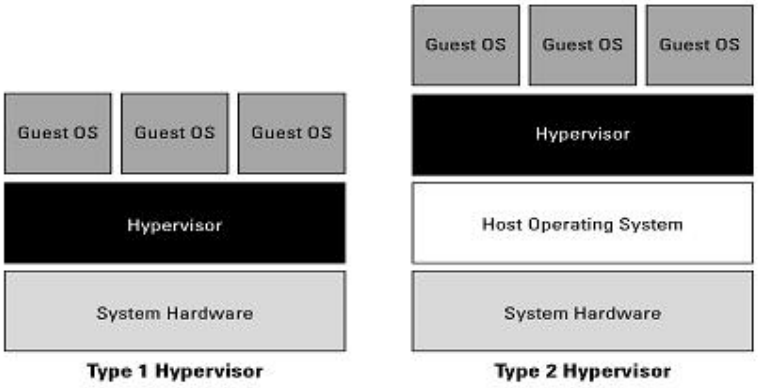
A low-level program is required to provide system resource access to virtual machines, and this program is referred to as the hypervisor or Virtual Machine Monitor (VMM). A hypervisor running on bare metal is a Type 1 VM or native VM. Examples of Type 1 Virtual Machine Monitors are LynxSecure, RTS Hypervisor, Oracle VM, Sun xVM Server, VirtualLogix VLX, VMware ESX and ESXi, and Wind River VxWorks, among others. The operating system loaded into a virtual machine is referred to as the guest operating system, and there is no constraint on running the same guest on multiple VMs on a physical system. Type 1 VMs have no host operating system because they are installed on a bare system.

An operating system running on a Type 1 VM is a full virtualization because it is a complete simulation of the hardware that it is running on. Not all CPUs support virtual machines, and many that do require that we enable this support in the BIOS. For example, AMD-V processors (code named Pacifica) and Intel VT-x (code named Vanderpool) oure the first of these vendor's 64-bit offerings that added this type of support.

Some hypervisors are installed over an operating system and are referred to as Type 2 or hosted VM. Examples of Type 2 Virtual Machine Monitors are Containers, KVM, Microsoft Hyper V, Parallels Desktop for Mac, Wind River Simics, VMWare Fusion, Virtual Server 2005 R2, Xen, Windows Virtual PC, and VMware Workstation 6.0 and Server, among others. This is a very rich product category. Type 2 virtual machines are installed over a host operating system; for Microsoft Hyper-V, that operating system would be Windows Server. In the section that follows, the Xen hypervisor (which runs on top of a Linux host OS) is more fully described. Xen is used by Amazon Web Services to provide Amazon Machine Instances (AMIs).

On a Type 2 VM, a software interface is created that emulates the devices with which a system would normally interact. This abstraction is meant to place many I/O operations outside the virtual environment, which makes it both programmatically easier and more efficient to execute device I/O than it would be inside a virtual environment. This type of virtualization is sometimes referred to as paravirtualization, and it is found in hypervisors such as Microsoft's Hyper-V and Xen. It is the host operating system that is performing the I/O through a para-API.

VMware's vSphere cloud computing infrastructure model



Type 1 and Type 2 hypervisors.

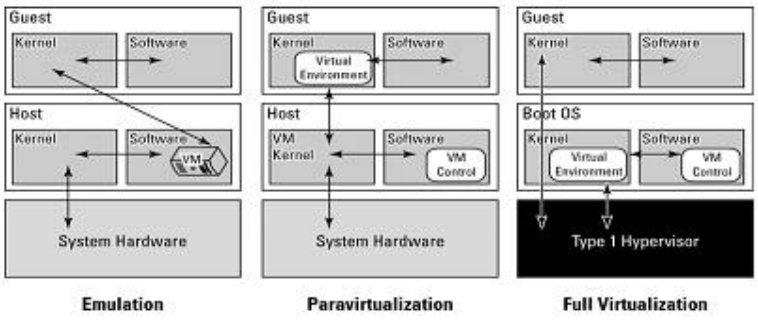
The above figure shows the difference between emulation, para virtualization, and full virtualization. In emulation, the virtual machine simulates hardware, so it can be independent of the underlying system hardware. A guest operating system using emulation does not need to be modified in any way. Para virtualization requires that the host operating system provide a virtual machine interface for the guest operating system and that the guest access hardware through that host VM. An operating system running as a guest on a paravirtualizationsystem must be ported to work with the host interface. Finally, in a full virtualization scheme, the VM is installed as a Type 1 Hypervisor directly onto the hardware. All operating systems in full virtualization communicate directly with the VM hypervisor, so guest operating systems do not require any modification.

Guest operating systems in full virtualization systems are generally faster than other virtualization schemes.

The Virtual Machine Interface (VMI) open standard (http://vmi ncsa.uiuc.edu/) that VMware haspropose‌d is an example of a paravirtualization API. The latest version of VMI is 2.1, and it ships as a default installation with many versions of the Linux operating system.

Wikipedia maintains a page called “Comparison of platform virtual machines” http://en.wikipedia.org/wiki/Comparison of platform virtual machines. The page contains a table offeatures of the most common Virtual Machine Managers.

We are probably familiar with process or application virtual machines. Most folks run the Java Virtual Machine or Microsoft's .NET Framework VM (called the Common Language Runtime or CLR) on their computers. A process virtual machine instantiates when a command begins a process, the VM is created by an interpreter, the VM then executes the process, and finally the VM exits the system and is destroyed. During the time the VM exists, it runs as a high-level abstraction.



Emulation, paravirtualization, and full virtualization types

Applications running inside an application virtual machine are generally slow, but these programs are very popular because they provide portability, offer rich programming languages, come with many advanced features, and allow platform independence for their programs. Although many cloud computing applications provide process virtual machine applications, this

type of abstraction isn't really suitable for building a large or high-performing cloud network, with one exception.

The exception is the process VMs that enable a class of parallel cluster computing applications. These applications are high-performance systems where the virtual machine is operating one process per cluster node, and the system maintains the necessary intra-application communications over the network interconnect.

Examples of this type of system are the Parallel Virtual Machine (PVM; see http://www.csm.ornl.gov/pvm/pvm home.html) and the Message Passing Interface (MPI; see http://www mpi-forum.org/).

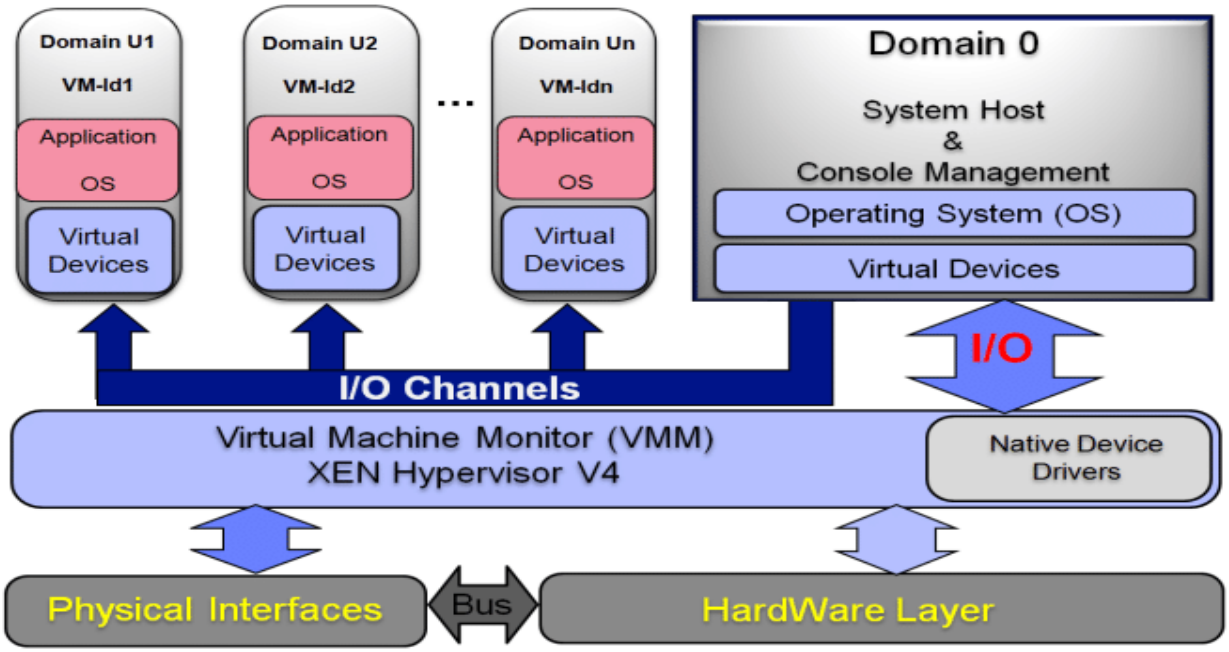
Some people do not consider these application VMs to be true virtual machines, noting that these applications can still access the host operating system services on the specific system on which they are running. The emphasis on using these process VMs is in creating a high-performance networked supercomputer often out of heterogeneous systems, rather than on creating a ubiquitous utility resource that characterizes a cloud network.

Some operating systems such as Sun Solaris and IBM AIX 6.1 support a feature known as operating system virtualization. This type of virtualization creates virtual servers at the operating system or kernel level. Each virtual server is running in its own virtual environment (VE) as a virtual private server (VPS). Different operating systems use different names to describe these machine instances, each of which can support its own guest OS. However, unlike true virtual machines, VPS must all be running the same OS and the same version of that OS. Sun Solaris 10 uses VPS to create what is called Solaris Zones. With IBM AIX, the VPS is called a System Workload Partition (WPAR). This type of virtualization allows for a dense collection of virtual machines with relatively low overhead. Operating system virtualization provides many of the benefits of virtualization previously noted in this section.

Xen Hypervisor

Xen Hypervisor Xen is a type-1 hypervisor, providing services that allow multiple computer operating systems to execute on the same computer hardware concurrently. It was developed by

the University of Cambridge. Now being developed by the Linux Foundation with support from Intel.



Xen is a hypervisor that enables the simultaneous creation, execution and management of multiple virtual machines on one physical computer.

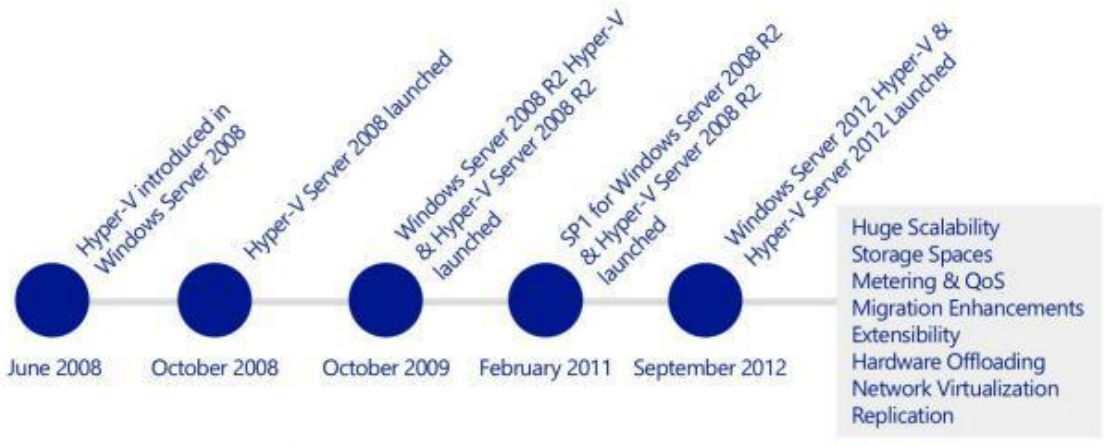
- Xen was developed by XenSource, which was purchased by Citrix Systems in 2007.
- Xen was first released in 2003.
- It is an open source hypervisor.
- It also comes in an enterprise version.
- Because it's a type-1 hypervisor, Xen controls, monitors and manages the hardware, peripheral and I/O resources directly.
- Guest virtual machines request Xen to provision any resource and must install Xen virtual device drivers to access hardware components.
- Xen supports multiple instances of the same or different operating systems with native support for most operating systems, including Windows and Linux.
- Moreover, Xen can be used on x86, IA-32 and ARM processor architecture.

Hyper V

Microsoft **Hyper-V**, formerly known as **Windows Server Virtualization**, is a native hypervisor; it can create virtual machines on x86-64 systems

running Windows. Starting with Windows 8, Hyper-V supersedes Windows Virtual PC as the hardware virtualization component of the client editions of Windows NT. A server computer running Hyper-V can be configured to expose individual virtual machines to one or more networks.

Hyper-V was first released alongside Windows Server 2008, and has been available without charge for all the Windows Server and some client operating systems since.



There are two manifestations of the Hyper-V technology:

Hyper-V is the hypervisor-based virtualization role of **Windows Server**.

Microsoft Hyper-V Server is the hypervisor-based server virtualization product that allows customers to consolidate workloads onto a single physical server. This is available as a free download.

With the launch of Windows Server 2008 R2 Hyper-V, in October 2009, Microsoft introduced a number of compelling capabilities to help organizations reduce costs, whilst increasing agility and flexibility. Key features introduced included:

Live Migration – Enabling the movement of virtual machines (VMs) with no interruption or downtime

Cluster Shared Volumes – Highly scalable and flexible use of shared storage (SAN) for VMs

Processor Compatibility – Increase the Flexibility for Live Migration across hosts with differing CPU architectures

Hot Add Storage – Flexibly add or remove storage to and from VMs

Improved Virtual Networking Performance – Support for Jumbo Frames and Virtual Machine Queue (VMq)

With the addition of Service Pack 1 (SP1) for Hyper-V, in October 2011, Microsoft introduced 2 new, key capabilities to help organizations realize even greater value from the platform:

Dynamic Memory – More efficient use of memory while maintaining consistent workload performance and scalability.

RemoteFX– Provides the richest virtualized Windows 7 experience for Virtual Desktop Infrastructure (VDI) deployments.

Windows Server 2012 Hyper V and Windows Server 2012 R2

Fast forward to September 2012, and the launch of Windows Server 2012. This brought an incredible number of new and an enhanced Hyper-V capabilities. These capabilities, many of which we'll discuss in this paper, ranged from enhancements around scalability, new storage and networking features, significant enhancements to the Live Migration capabilities, deeper integration with hardware, and an in-box VM replication capability, to name but a few. These improvements, new features and enhancements can be grouped into 4 key areas, and it's these key areas we'll focus on throughout this whitepaper, looking at both Windows

Server 2012 and R2, and how it compares and contrasts with vSphere 5.5. The 4 key areas are:

Scalability, Performance & Density – customers are looking to run bigger, more poourful virtual machines, to handle the demands of their biggest workloads. In addition, as hardware scale grows, customers wish to take advantage of the largest physical systems to drive the highest levels of density, and reduce overall costs.

Security &Multitenancy- Virtualized data centers are becoming more popular and practical every day. IT organizations and hosting providers have begun offering infrastructure as a service (IaaS), which provides more flexible, virtualized infrastructures to customers—"server instances on-demand." Because of this trend, IT organizations and hosting providers must offer customers enhanced security and isolation from one another, and in some cases, encrypted to meet compliance demands.

Flexible Infrastructure – In a modern datacenter, customers are looking to be agile, in order to respond to changing business demands quickly, and efficiently. Being able to move workloads flexibly around the infrastructure is of incredible importance, and in addition, customers want to be able to choose where best to deploy their workloads based on the needs of that workload specifically.

High Availability & Resiliency – As customers' confidence in virtualization grows, and they virtualize their more mission-critical workloads, the importance of keeping those workloads continuously available grows significantly. Having capabilities built into the platform that not only help keep those workloads highly available, but also, in the event of a disaster, quick to restore in another geographical location, is of immense importance when choosing a platform for today's modern datacenter.

Why Hyper-V?

Virtualization technologies help customers' loour costs and deliver greater agility and economies of scale. Either as a stand-alone product or an integrated part of Windows Server,

Hyper-V is a leading virtualization platform for today and the transformational opportunity with cloud computing.

With Hyper-V, it is now easier than ever for organizations to take advantage of the cost savings of virtualization, and make the optimum use of server hardware investments by consolidating multiple server roles as separate virtual machines that are running on a single physical machine. Customers can use Hyper-V to efficiently run multiple operating systems, Windows, Linux, and others, in parallel, on a single server. Windows Server 2012 R2 extends this with more features, greater scalability and further inbuilt reliability mechanisms.

In the data center, on the desktop, and now in the cloud, the Microsoft virtualization platform, which is led by Hyper-V and surrounding System Center management tools, simply makes more sense and offers better value for money when compared to the competition.

	Resource	Windows Server 2008 R2 Hyper-V	Windows Server 2012 R2 Hyper-V	Improvement Factor
Host	Logical Processors	64	**320**	**5×**
	Physical Memory	1TB	**4TB**	**4×**
	Virtual CPUs per Host	512	**2,048**	**4×**
VM	Virtual CPUs per VM	4	**64**	**16×**
	Memory per VM	64GB	**1TB**	**16×**
	Active VMs per Host	384	**1,024**	**2.7×**
	Guest NUMA	No	**Yes**	**-**
Cluster	Maximum Nodes	16	**64**	**4×**
	Maximum VMs	1,000	**8,000**	**8×**

Enhanced Storage Capabilities

Windows Server 2012 and subsequently, 2012 R2 Hyper-V also introduce a number of enhanced storage capabilities to support the most intensive, mission-critical of workloads. These capabilities include:

Virtual Fiber Channel – Enables virtual machines to integrate directly into Fiber Channel Storage Area Networks (SAN), unlocking scenarios such as fiber channel-based Hyper-V Guest Clusters.

Support for 4-KB Disk Sectors in Hyper-V Virtual Disks. Support for 4,000-byte (4-KB) disk sectors lets customers take advantage of the emerging innovation in storage hardware that provides increased capacity and reliability.

New in R2 - Storage Spaces with Tiering- Storage Spaces enables we to virtualize storage by grouping industry-standard disks into storage pools, and then create virtual disks called storage spaces from the available capacity in the storage pools. These pools now support a mix of HDD and SSD, providing a tiered pool, where hot data will reside on SSD and cold data on HDD. Fully supported as a repository for Hyper-V VMs.

Data Deduplication - Windows Server 2012 R2 also provides an inbox deduplication capabilities which utilizes sub-file variable-size chunking and compression to considerably reduce storage consumption for files and folders hosted on deduplicated Windows Server volumes. With Windows Server 2012 R2, support has been added for VDI deployments. Deduplication rates for VDI deployments can range as high as 95% savings and that includes VDI deployments that utilize differencing disks for rapid provisioning.

New Virtual Hard Disk Format. This new format, called VHDX, is designed to better handle current and future workloads and addresses the technological demands of an enterprise's evolving needs by increasing storage capacity, protecting data, improving quality performance on 4-KB disks, and providing additional operation-enhancing features. The maximum size of a VHDX file is 64TB.

Offloaded Data Transfer (ODX). With Offloaded Data Transfer support, the Hyper-V host CPUs can concentrate on the processing needs of the application and offload storage-related tasks to the SAN, increasing performance.

Online Checkpoint Merge. With the online checkpoint merge capability, customers who have taken checkpoints (snapshots), for a running virtual machine, no longer have to poour down the virtual machine in order to merge the checkpoint back into the original virtual disk file, ensuring virtual machine uptime is increased and the administrator gains increased flexibility.

New in R2 - Online Virtual Disk Resize. With the online virtual disk resize, administrators can grow and shrink virtual disks that are attached to a VM's virtual SCSI controller, providing an administrator with greater flexibility to respond to changing business needs.

Enhanced Networking Performance

Windows Server 2012 R2 Hyper-V also includes a number of performance enhancements within the networking stack to help customers virtualize their most intensive network workloads. These capabilities include:

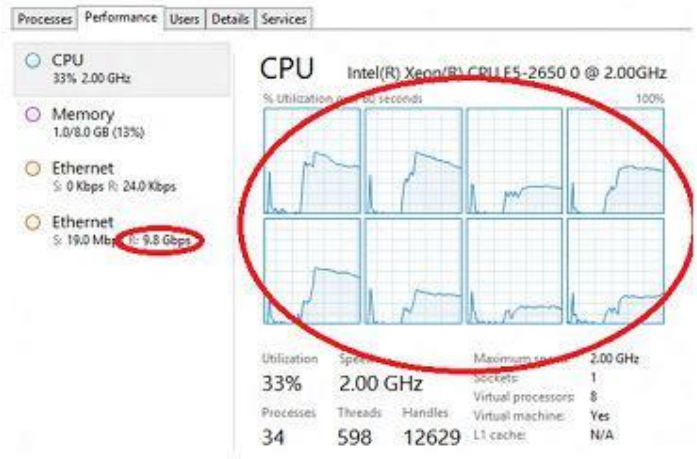
Dynamic Virtual Machine Queue – DVMQ dynamically distributes incoming VM network traffic processing to host processors (based on processor usage and network load). In times of heavy network load, Dynamic VMQ automatically recruits more processors. In times of light network load, Dynamic VMQ relinquishes those same processors

IPsec Task Offload - IPsec Task Offload in Windows Server 2012 R2 leverages the hardware capabilities of server NICs to offload IPsec processing. This reduces the CPU overhead of IPsec encryption and decryption significantly. In Windows Server 2012 R2, IPsec Task Offload is extended to Virtual Machines as well. Customers using VMs who want to protect their network traffic with IPsec can take advantage of the IPsec hardware offload capability available in server NICs, thus freeing up CPU cycles to perform more application level work and leaving the per packet encryption/decryption to hardware.

SR-IOV - When it comes to virtual networking, a primary goal is native I/O throughput. Windows Server 2012 R2 provides the ability to assign SR-IOV functionality from physical devices directly into virtual machines. This gives VMs the ability to bypass the software-based Hyper-V Virtual Network Switch, and more directly address the NIC. As a result, CPU overhead and latency is reduced, with a corresponding rise in throughput. This is all available, without sacrificing key Hyper-V features such as virtual machine Live Migration.

New in R2 – Virtual Receive Side Scaling - Prior to 10GbE networking, one modern processor was usually more than enough to handle the networking workload of a VM. With the introduction of 10GbE NICs, the amount of data being sent to and received from a VM exceeded what a single processor could effectively handle. In the physical host, this challenge had a solution, namely, Receive Side Scaling (RSS). RSS spreads traffic from the network

interface card (NIC), based on TCP flows, and to multiple processors for simultaneous processing of TCP flows. With Windows Server 2012 R2 however, similar to how RSS distributes networking traffic to multiple cores in physical machines, vRSS spreads networking traffic to multiple VPs in each VM by enabling RSS inside the VM. With vRSS enabled, a VM is able to process traffic on multiple VPs simultaneously and increase the amount of throughput it is able to handle.



Enhanced Resource Management

Windows Server 2012 R2 Hyper-V also includes a number of enhanced resource management capabilities that help customers to optimize the utilization of the virtualized infrastructure to drive higher levels of performance. These capabilities include:

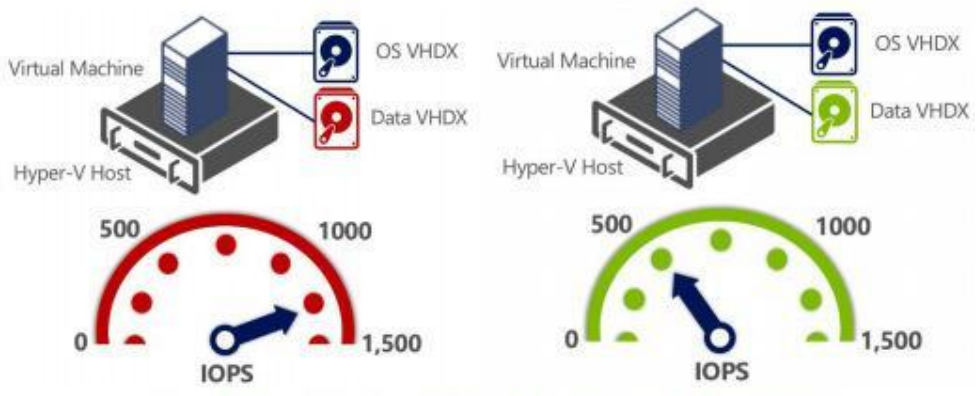
Dynamic Memory Improvements - These improvements dramatically increase virtual machine consolidation ratios and improve reliability for restart operations that can lead to loour costs, especially in environments, such as VDI, that have many idle or low-load virtual machines. Administrators can now more flexibly manage memory through the use of a

Startup, Minimum and Maximum configuration option, along with the ability to adjust the memory values whilst the VM is running, increasing flexibility for the administrator. Windows Server 2012 R2 Hyper-V also includes a capability known as Smart Paging, which provides a more reliable and robust solution for VM restarts when memory is under contention.

Resource Metering - In Windows Server 2012 R2 Hyper-V, Resource Metering, helps we track historical data on the use of virtual machines and gain insight into the resource use of specific servers. We can use this data to perform capacity planning, to monitor consumption by different business units or customers, or to capture data needed to help redistribute the costs of running a workload. Resource Metering captures metrics across CPU, Memory, Disk and Network.

Network Quality of Service - QoS provides the ability to programmatically adhere to a service level agreement (SLA) by specifying the minimum bandwidth that is available to a virtual machine or a port. It prevents latency issues by allocating maximum bandwidth use for a virtual machine or port.

New in R2 – Storage Quality of Service – Storage QoS provides storage performance isolation in a multitenant environment and mechanisms to notify we when the storage I/O performance does not meet the defined threshold to efficiently run our virtual machine workloads.



	Resource	Windows Server 2008 R2 Hyper-V	Windows Server 2012 R2 Hyper-V	Improvement Factor
Host	Logical Processors	64	**320**	**5×**
	Physical Memory	1TB	**4TB**	**4×**
	Virtual CPUs per Host	512	**2,048**	**4×**
VM	Virtual CPUs per VM	4	**64**	**16×**
	Memory per VM	64GB	**1TB**	**16×**
	Active VMs per Host	384	**1,024**	**2.7×**
	Guest NUMA	No	**Yes**	**-**
Cluster	Maximum Nodes	16	**64**	**4×**
	Maximum VMs	1,000	**8,000**	**8×**

Virtual Box

Oracle VM VirtualBox is cross platform virtualization software that allows we to extend our existing computer to run multiple operating systems at the same time. Designed for IT professionals and developers, Oracle VM VirtualBox runs on Windows, Mac OS X, Linux and Oracle Solaris systems and is ideal for testing, developing, demonstrating and deploying solutions across multiple platforms on one machine.

Key Benefits

- Run almost any type of application on our existing machine
- Quickly and easily try out new platforms
- Create a multiplatform test and development environment
- Build a multi1tier demonstration system on a single portable machine
- Extend the lifetime and usefulness of existing computers
- Run legacy platforms and applications on modern hardware
- Easily create isolated environments

Key Features

- Available for Windows, Mac OS X, Linux and Oracle Solaris host operating systems
- Supports a wide range of guest platforms
- Easy to use graphical user interface
- Poourful, scriptable command line interface
- Import and export virtual machines using OVF/OVA standards
- Shared folders between guest and host
- Seamless, resizable, and full screen window display modes
- Video and 3D (OpenGL, DirectX) acceleration
- Virtual webcam
- Multiple virtual screen support
- Poourful and flexible networking options
- USB 1.1/2.0/3.0 and serial ports
- SAS, SATA, SCSI and IDE storage controllers
- Built-in iSCSI initiator
- Built-in Remote Display Server
- Multi-generational branched snapshots
- Linked and full clones
- Controllable copy and paste
- Screen-recording facility
- Disk image encryption
- HiDPI support
- Drag and drop support

Oracle VM VirtualBox Manager Screen

Easy to Use, Fast and Poourful, Great Platform Coverage

Designed for use on systems ranging from ultrabooks to high end server class hardware, Oracle VM Virtual Box is lightweight and easy to install and use. Yet under the simple exterior lies an extremely fast and poourful virtualization engine. With a formidable reputation for speed and agility, Oracle VM Virtual Box contains innovative features to deliver tangible business benefits: significant performance improvements; a more poourful virtualization system and a wider range of supported guest operating system platforms.

Easy to Use

Improved Virtual Box Manager with further features – The Oracle VM Virtual Box Manager now supports hot-plug for SATA virtual disks and the option to customize status bar, menu bar and guest-content scaling for each virtual machine deployed;

New Introduced Headless and Detachable start options – The Oracle VM Virtual Box Manager now supports to start virtual machine in the background with a separate frontend process that can be closed while the virtual machine continues to work;

Easy to use Wizards – Wizards help with the creation of new virtual machines. Preconfigured settings are used based on the type of guest OS;

Easy import and export of appliances – Virtual machines can be created, configured and then shared by exporting and importing virtual appliances using industry-standard formats such as .ova;

Improved Huge Range of Guest Platforms – including the very latest Windows 10, Windows Server 2012 R2 and leading edge Linux platforms too.

Improved Virtual Box Guest Additions – Installed inside the guest virtual machine, the Guest Additions provide a more natural user experience. For example, guest windows can be easily resized to arbitrary resolutions, made full-screen or even operate in seamless mode. And data can be copy and pasted to and from, and between, concurrently running machines and the host platform. This functionality is now controllable as bi-directional, uni-directional, or disabled;

Shared Folders – Share our host platform's filesystem with the guest to facilitate real cross-platform computing;

Multi-touch support – Hosts supporting multi-touch interfaces can now also deliver this to their guests too;

Flexible Networking options – Oracle VM Virtual Box offers a rich range of networking models from easy-to-use NAT networking, to fully functional Bridged networking, and specialist Internal and Host-only networking too. The new "NAT Network" mode allows multiple guests to run on the same internal network, seeing each other, and also the outside world via a new NAT service;

IPv6 – IPv6 is now offered as an option in most networking modes alongside IPv4;

Virtual Media Manager – Oracle VM Virtual Box supports the widest range of virtual disk formats from its own native .vdi format to those offered by Microsoft (.vhd), VMware

(.vmdk), and Parallels (.vdd). The Virtual Media Manager tool now allows conversions between formats using an easy to use graphical user interface;

Video Capture – A built-in recording mechanism of the guest's screen contents. Easy to start and stop, recording one or more virtual screens to the standard webm format.

Performance and Poour

Improved Latest Intel and AMD hardware support – Harnessing the latest in chip-level support for virtualization, Oracle VM Virtual Box supports even the most recent AMD and Intel processors bringing faster execution times for everything from Windows to Linux and Oracle Solaris guests. But Virtual Box will also run on older hardware without VT support;

Improved Instruction Set extended – More instruction set extensions available to the guest when running with hardware-assisted virtualization; this include also AES-NI that improve the speed of applications performing encryption and decryption using Advanced Encryption Standard (AES);

New Para virtualization Support – Virtual Box allows exposing a para-virtualization interface to facilitate accurate and efficient execution of software by leveraging built-in virtualization support of modern Linux and Microsoft Windows;

New Disk Image Encryption – Virtual Box allows to encrypt data stored in hard disk images transparently for the guest. Virtual Box uses the AES algorithm and supports 128 or 256-bit data encryption keys;

Improved Bi-Directional Drag and Drop support – On all host platforms, Windows, Linux and Oracle Solaris guests now support "drag and drop" of content between the host and the guest. The drag and drop feature transparently allows copying or opening of files, directories, and more;

High-performance storage I/O subsystem – Oracle VM Virtual Box offers a wide range of virtual storage controllers including SAS, SATA, SCSI and IDE controllers. Virtual Box utilizes an asynchronous I/O virtual disk subsystem to achieve high-performance whilst maintaining high data integrity;

Built-in iSCSI Initiator – Oracle VM Virtual Box includes an iSCSI initiator that allows virtual disks to exist as iSCSI targets. The guest sees a standard storage controller but disk accesses are translated into iSCSI commands and sent across the network;

3D graphics and video acceleration – The Guest Additions feature new, improved display drivers that accelerate 3D graphics by intercepting OpenGL and Direct3D calls in the guest and leveraging the host's GPU to render the images and video onto the screen.

Remote Display Protocol – The unique built-in Virtual Box Remote Display Protocol (VRDP) enables poourful remote, graphical access to the console of the guest. Microsoft RDP capable clients can connect to one or more remote monitors, with USB device redirection when using rdesktop-based clients. VRDP is now also accessible over IPv6;

Improved Serial and USB connections – External devices can be connected to guests, with specific USB devices selected by a poourful filter mechanism; now Virtual Box supports up to USB 3.0 devices;

Virtual webcam – On hosts with cameras, Virtual Box now exposes a virtual webcam allowing guests running apps such as Skype or Google Hangouts to use the host camera;

High-Definition audio – Guests enjoy the rich audio capabilities of an Intel high definition audio card;

Full ACPI support – The host's poour status is fully available to the guest and ACPI button events can be sent to the guest to control the lifecycle of the virtual machine;

Linked and full clones – Oracle VM Virtual Box makes it easy to clone virtual machines. Clones can be full copies of configuration information and virtual disks, or may share a parent virtual disk for faster cloning and greater storage efficiency;

Multi-generational and branched snapshots – Snapshots allow a user to revert to previous known states. Take a snapshot before installing software, then revert to the snapshot to recover the pre-installation state;

Page Fusion – Traditional Page Sharing techniques have suffered from long and expensive cache construction as pages are scrutinized as candidates for de-duplication.

Taking a smarter approach, Virtual Box Page Fusion uses intelligence in the guest virtual machine to determine much more rapidly and accurately those pages which can be eliminated thereby increasing the capacity or VM density of the system;

Resource controls – Host resources such as CPU execution, disk and network I/O can be capped or throttled to protect against rogue guests consuming excessive amounts;

Guest automation – The guest automation APIs have been extended to allow host-based logic to drive operations in the guest including update of the Guest Additions;

Web services – A Web service API enables remote control of Virtual Box by authorized clients.

Platforms

Commercially supported platforms – Oracle VM Virtual Box enables we to install and run a huge range of host and guest platforms. Oracle offers commercial support for the most popular guest operating systems, assuring customers of expert help when they need it.

New Oracle Linux 7 – Support for the latest version of Oracle's flagship Linux platform;

New Ubuntu and Fedora – Support for both the desktop and server versions of the most popular Ubuntu Linux and Fedora distributions; New Mac OS X 10.10 “Yosemite” – The latest Mac OS X platform from Apple.

System Requirements			
Hardware Requirements:			
Processor	Any x86 compatible processor from Intel or AMD (with or without VT-x or AMD-V support)		
Memory	Minimum 1GB + RAM as required by running guests		
Host Platform Requirements (Commercially supported):			
Windows	**Mac OS X**	**Linux hosts (32-bit and 64-bit)**	**Oracle Solaris hosts (64-bit)**
• Windows Vista SP1 and later (32-bit and 64-bit) • Windows Server 2008 (32-bit and 64-bit) • Windows Server 2008 R2 (32-bit and 64-bit) • Windows 7 (32-bit and 64-bit) • Windows 8 (32-bit and 64-bit) • Windows 8.1 (32-bit and 64-bit) • Windows Server 2012 • Windows Server 2012 R2	• 10.8 (Mountain Lion, 32-bit and 64-bit) • 10.9 (Mavericks) • 10.10 (Yosemite)	• Oracle Linux 5, 6 and 7 • Ubuntu: 10.04 ("Lucid Lynx") to 15.04 ("Vivid Vervet") • Red Hat Enterprise Linux 5, 6 and 7 • SUSE Linux Enterprise Server 11, 12 • Fedora Core/Fedora 6 to 22	• Solaris 11, 11.1 • Solaris 10 (u10 and higher)

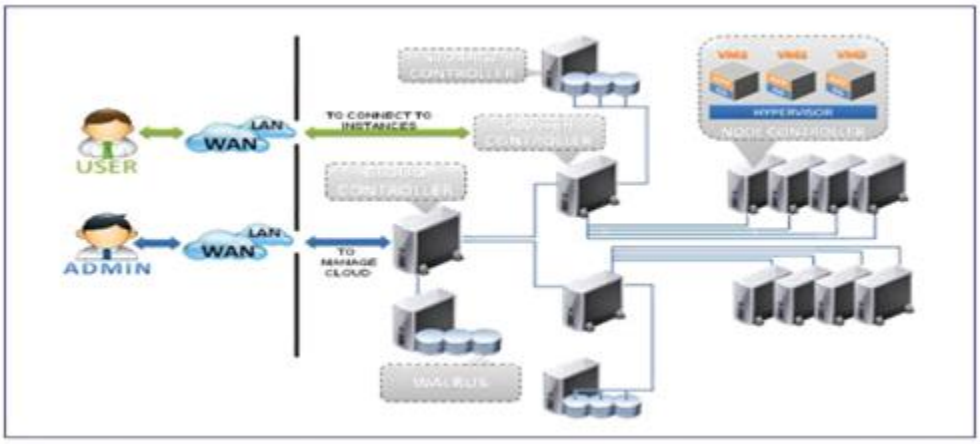
The Eucalyptus Open-Source Private Cloud

Eucalyptus is a Linux-based open-source software architecture that implements efficiency-enhancing private and hybrid clouds within an enterprise's existing IT infrastructure.

Eucalyptus is an acronym for "Elastic Utility Computing Architecture for Linking Our Programs to Useful Systems."

A Eucalyptus private cloud is deployed across an enterprise's "on premise" data center infrastructure and is accessed by users over enterprise intranet. Thus, sensitive data remains entirely secure from external intrusion behind the enterprise firewall.

Initially developed to support the high performance computing (HPC) research of Professor Rich Wolski's research group at the University of California, Santa Barbara, Eucalyptus is engineered according to design principles that ensure compatibility with existing Linux-based data center installations. Eucalyptus can be deployed without modification on all major Linux OS distributions, including Ubuntu, RHEL, Centos, and Debian. And Ubuntu distributions now include the Eucalyptus software core as the key component of the Ubuntu Enterprise Cloud.



Eucalyptus Components

Each Eucalyptus service component exposes a well-defined language agnostic API in the form of a WSDL document containing both the operations that the service can perform and the input/output data structures. Inter-service authentication is handled via standard WS-Security mechanisms. There are five high-level components, each with its own Web-service interface, that comprise a Eucalyptus installation (Fig a). A brief description of the components within the Eucalyptus system follows.

Cloud Controller

Cloud Controller (CLC) is the entry-point into the cloud for administrators, developers, project managers, and end-users. The CLC is responsible for querying the node managers for information about resources, making high level scheduling decisions, and implementing them by making requests to cluster controllers. The CLC, as shown in Figure 1, is also the interface to the management platform. In essence, the CLC is responsible for exposing and managing the underlying virtualized resources (servers, network, and storage) via a well-defined industry standard API (Amazon EC2) and a Web-based user interface.

Functions:

1. Monitor the availability of resources on various components of the cloud infrastructure, including hypervisor nodes that are used to actually provision the instances and the cluster controllers that manage the hypervisor nodes.
2. Resource arbitration – deciding which clusters will be used for provisioning the instances.
3. Monitoring the running instances.

In short, CLC has a comprehensive knowledge of the availability and usage of resources in the cloud and the state of the cloud.

Cluster Controller

Cluster Controller (CC) generally executes on a cluster front-end machine or any machine that has network connectivity to both the nodes running NCs and to the machine running the CLC. CCs gather information about a set of VMs and schedules

VM execution on specific NCs. The CC also manages the virtual instance network and participates in the enforcement of SLAs as directed by the CLC. All nodes served by a single CC must be in the same broadcast domain (Ethernet).

Functions:

1. To receive requests from CLC to deploy instances.
2. To decide which NCs to use for deploying the instances on.
3. To control the virtual network available to the instances.
4. To collect information about the NCs registered with it and report it to the CLC.

Node Controller

Node Controller (NC) is executed on every node that is designated for hosting VM instances. A UEC node is a VT-enabled server capable of running KVM as the hypervisor. UEC automatically installs KVM when the user chooses to install the UEC node. The VMs running on the hypervisor and controlled by UEC are called instances. Eucalyptus supports other hypervisors like Xen apart from KVM, but Canonical has chosen KVM as the preferred hypervisor for UEC.

The NC runs on each node and controls the life cycle of instances running on the node. The NC interacts with the OS and the hypervisor running on the node on one side and the CC on the other side.

NC queries the operating system running on the node to discover the node's physical resources – the number of cores, the size of memory, and the available disk space. It also learns about the state of VM instances running on the node and propagates this data up to the CC.

Functions:

1. Collection of data related to the resource availability and utilization on the node and reporting the data to CC.
2. Instance life cycle management.

Storage Controller

Storage Controller (SC) implements block-accessed network storage (e.g., Amazon Elastic Block Storage -- EBS) and is capable of interfacing with various storage systems (NFS, iSCSI, etc.). An elastic block store is a Linux block device that can be attached to a virtual machine but sends disk traffic across the locally attached network to a remote storage location. An EBS volume cannot be shared across instances but does allow a snapshot to be created and stored in a central storage system such as Walrus, the Eucalyptus storage service.

Functions:

1. Creation of persistent EBS devices.
2. Providing the block storage over AoE or iSCSI protocol to the instances.
3. Allowing creation of snapshots of volumes.

Walrus

Walrus (put/get storage) allows users to store persistent data, organized as eventually-consistent buckets and objects. It allows users to create, delete, list buckets, put, get, and delete objects, and set access control policies. Walrus is interface compatible with Amazon's

S3, and supports the Amazon Machine Image (AMI) image-management interface, thus providing a mechanism for storing and accessing both the virtual machine images and user data. Using Walrus, users can store persistent data, which is organized as buckets and objects. WS3 is a file-level storage system, as compared to the block-level storage system of Storage Controller.

For using Walrus to manage Eucalyptus VM images, we can use Amazon's tools to store/register/delete them from Walrus. Other third-party tools can also be used to interact with Walrus directly.

Third-Party Tools for Interacting with Walrus

1. S3curl: a command line tool that is a wrapper around curl.

http://open.eucalyptus.com/wiki/s3curl

2. S3cmd: a tool that allows command line access to storage that supports the S3 API.
http://open.eucalyptus.com/wiki/s3cmd

3. S3fs: a tool that allows users to access S3 buckets as local directories. http://open.eucalyptus.com/wiki/s3fs

Management Platform

Management Platform provides an interface to various Eucalyptus services and modules. These features can include VM management, storage management, user/group management, accounting, monitoring, SLA definition and enforcement, cloud-bursting, provisioning, etc.

Euca2ool

Euca2ools are command-line tools for interacting with Web services that export a REST/Query-based API compatible with Amazon EC2 and S3 services. The tools can be used with both Amazon's services and with installations of the Eucalyptus open-source cloud-computing infrastructure. The tools oure inspired by command-line tools distributed by Amazon (api-tools and ami-tools) and largely accept the same options and environment

variables. However, these tools oure implemented from scratch in Python, relying on the Boto library and M2Crypto toolkit.

Features:

1. Query of availability zones (i.e., clusters in Eucalyptus).
2. SSH key management (add, list, delete).
3. VM management (start, list, stop, reboot, get console output).
4. Security group management.
5. Volume and snapshot management (attach, list, detach, create, bundle, delete).
6. Image management (bundle, upload, register, list, deregister).
7. IP address management (allocate, associate, list, release).

Key Benefits

Build and manage self-service heterogeneous on-premise IaaS clouds using either existing infrastructure or dedicated compute, network and storage resources.

Support high-availability IaaS for the most demanding cloud deployments

Gain precise control of private cloud resources via enterprise-ready user and group identity management along with resource quotas.

Pool dynamic resources with built-in elasticity, allowing organizations to scale up and down virtual compute, network and storage resources.

Integrate robust storage, enabling IT to easily connect and manage existing storage systems from within Eucalyptus clouds.

Build hybrid clouds between on-premise Eucalyptus clouds and AWS and AWS-compatible public clouds.

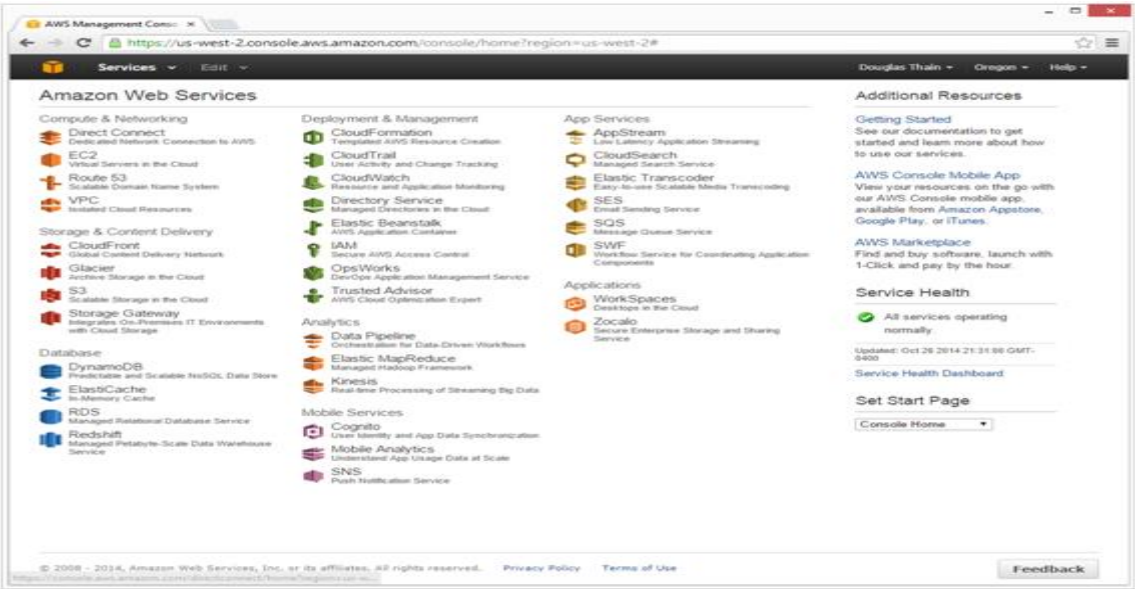
Run Eucalyptus or Amazon Machine Images as virtual cloud instances on Eucalyptus and AWS-compatible clouds.

Leverage vibrant AWS ecosystem and management tools to manage Eucalyptus IaaS clouds.

Amazon Cloud Computing

Amazon AWS

- Grew out of Amazon's need to rapidly provision and configure machines of standard configurations for its own business.
- Early 2000s – Both private and shared data centers began using virtualization to perform "server consolidation"
- 2003 – Internal memo by Chris Pinkham describing an "infrastructure service for the world."
- 2006 – S3 first deployed in the spring, EC2 in the fall
- 2008 – Elastic Block Store available.
- 2009 – Relational Database Service
- 2012 – DynamoDB



Terminology

- Instance = One running virtual machine.

- Instance Type = hardware configuration: cores, memory, disk.
- Instance Store Volume = Temporary disk associated with instance.
- Image (AMI) = Stored bits which can be turned into instances.
- Key Pair = Credentials used to access VM from command line.
- Region = Geographic location, price, laws, network locality.
- Availability Zone = Subdivision of region the is fault-independent. **EC2**

Pricing Model

- Free Usage Tier
- On-Demand Instances
    - Start and stop instances whenever we like, costs are rounded up to the nearest hour. (Worst price)
- Reserved Instances

  Pay up front for one/three years in advance. (Best price)
    - Unused instances can be sold on a secondary market.
- Spot Instances
    - Specify the price we are willing to pay, and instances get started and stopped without any warning as the marked changes. (Kind of like Condor!)
- http://aws.amazon.com/ec2/pricing/

Free Usage Tier

- 750 hours of EC2 running Linux, RHEL, or SLES t2.micro instance usage
- 750 hours of EC2 running Microsoft Windows Server t2.micro instance usage
- 750 hours of Elastic Load Balancing plus 15 GB data processing
- 30 GB of Amazon Elastic Block Storage in any combination of General Purpose (SSD) or Magnetic, plus 2 million I/Os (with Magnetic) and 1 GB of snapshot storage
- 15 GB of bandwidth out aggregated across all AWS services
- 1 GB of Regional Data Transfer

Q: How many instances can I run in Amazon EC2?

You are limited to running up to 20 On-Demand Instances, purchasing 20 Reserved Instances, and requesting 5 Spot Instances per region. New AWS accounts may start with limits that are lower than the limits described here. Certain instance types are further limited per region as follows:

Instance Type	On-Demand Limit	Reserved Limit	Spot Limit
cg1.4xlarge	2	20	5
hi1.4xlarge	2	20	5
hs1.8xlarge	2	20	Not offered
cr1.8xlarge	2	20	5
g2.2xlarge	5	20	5
r3.4xlarge	10	20	5

Q: How many instances can I run in Amazon EC2?

You are limited to running up to 20 On-Demand Instances, purchasing 20 Reserved Instances, and requesting 5 Spot Instances per region. New AWS accounts may start with limits that are lower than the limits described here. Certain instance types are further limited per region as follows:

Instance Type	On-Demand Limit	Reserved Limit	Spot Limit
cg1.4xlarge	2	20	5
hi1.4xlarge	2	20	5
hs1.8xlarge	2	20	Not offered
cr1.8xlarge	2	20	5
g2.2xlarge	5	20	5
r3.4xlarge	10	20	5

Simple Storage Service (S3)

1. A **bucket** is a container for objects and describes location, logging, accounting, and access control. A bucket can hold any number of **objects**, which are files of up to 5TB. A bucket has a name that must be **globally unique**.
2. Fundamental operations corresponding to HTTP actions:

   http://**bucket**.s3.amazonaws.com/**object**

   POST a new object or update an existing object.

   GET an existing object from a bucket.

   DELETE an object from the bucket

   LIST keys present in a bucket, with a filter.
3. A bucket has a **flat directory structure** (despite the appearance given by the interactive web interface.)

Easily Integrated into Web Applications

```
<form action="http://examplebucket.s3.amazonaws.com/"
method="post" enctype="multipart/form-data">

<input type="input" name="key" value="user/user1/" /> <input
type="hidden" name="acl" value="public-read" /> <input
type="hidden" name="success_action_redirect"

value="http://examplebucket.s3.amazonaws.com/successful_uploa
d.html" />

. . .

<input type="text" name="X-Amz-Credential”

value="AKIAIOSFODNN7EXAMPLE/20130806/us-east-
1/s3/aws4_request" />

. . .

<input type="submit" name="submit" value="Upload to Amazon
S3"/></form>
```

Bucket Properties

- Versioning – If enabled, POST/DELETE result in the creation of new versions without destroying the old.
  - Lifecycle – Delete or archive objects in a bucket a certain time after creation or last access or number of versions.
  - Access Policy – Control **when and where** objects can be accessed.
  - Access Control – Control who **may** access objects in this bucket.
  - Logging – Keep track of how objects are accessed.
  - Notification – Be notified when failures occur.

S3 Weak Consistency Model

"Amazon S3 achieves high availability by replicating data across multiple servers within Amazon's data centers. If a PUT request is successful, our data is safely stored. However, information about the changes must replicate across Amazon S3, which can take some time, and so we might observe the following behaviors:

- A process writes a new object to Amazon S3 and immediately attempts to read it. Until the change is fully propagated, Amazon S3 might report "key does not exist."
- A process writes a new object to Amazon S3 and immediately lists keys within its bucket. Until the change is fully propagated, the object might not appear in the list.
- A process replaces an existing object and immediately attempts to read it. Until the change is fully propagated, Amazon S3 might return the prior data.
- A process deletes an existing object and immediately attempts to read it. Until the deletion is fully propagated, Amazon S3 might return the deleted data."

Amazon Elastic Compute Cloud (EC2)

- Amazon Machine Images (**AMIs**) are the basic building blocks of Amazon EC2

- An AMI is a template that contains a software configuration (operating system, application server and applications) that can run on Amazon's computing environment
- AMIs can be used to launch an INSTANCE, which is a copy of the AMI running as a virtual server in the cloud.
-

Getting Started with Amazon EC2

- Step 1: Sign up for Amazon EC2
- Step 2: Create a key pair
- Step 3: Launch an Amazon EC2 instance
- Step 4: Connect to the instance
- Step 5: Customize the instance
- Step 6: Terminate instance and delete the volume created

Creating a key pair

- AWS uses public-key cryptography to encrypt and decrypt login information.
- AWS only stores the public key, and the user stores the private key.
- There are two options for creating a key pair:
  - Have Amazon EC2 generate it for we
  - Generate it ourself using a third-party tool such as OpenSSH, then import the public key to Amazon EC2

Generating a key pair with Amazon EC2

- Open the Amazon EC2 console at http://console.aws.amazon.com/ec2/
- On the navigation bar select region for the key pair
- Click **Key Pairs** in the navigation pane to display the list of key pairs associated with the account.

P Click **Create Key Pair**

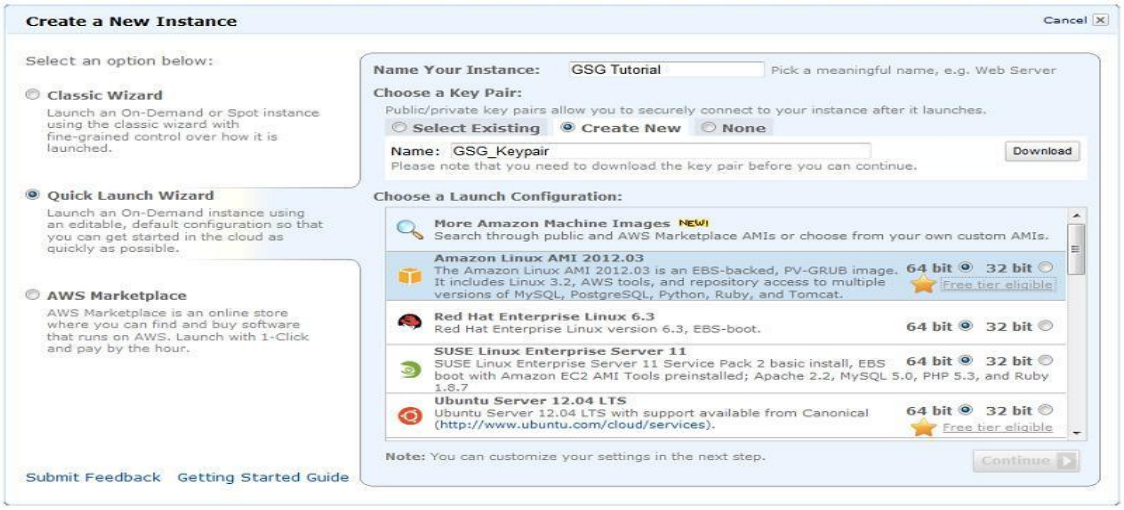
Q Enter a name for the key pair in the **Key Pair Name** field of the dialog box and click **Create**

R The private key file, with .pem extension, will automatically be downloaded by the browser.

Launching an Amazon EC2 instance

- Sign in to AWS Management Console and open the Amazon EC2 console at http://console.aws.amazon.com/ec2/
- From the navigation bar select the region for the instance

3. From the Amazon EC2 console dashboard, click **Launch Instance**



6. On the **Create a New Instance** page, click **Quick Launch Wizard**
7. In **Name Our Instance**, enter a name for the instance
8. In **Choose a Key Pair**, choose an existing key pair, or create a new one
9. In Choose a Launch Configuration, a list of basic machine configurations are displayed, from which an instance can be launched
10. Click continue to view and customize the settings for the instance
11. Select a security group for the instance. A **Security Group** defines the firewall rules specifying the incoming network traffic delivered to the instance. Security groups can be defined on the Amazon EC2 console, in **Security Groups** under **Network and Security**

12. Review settings and click **Launch** to launch the instance
13. Close the confirmation page to return to EC2 console

14. Click **Instances** in the navigation pane to view the status of the instance. The status is **pending** while the instance is launching

	Name	Instance	AMI ID	Root Device	Type	State	Public DNS
	GSG Tutorial	i-e1ab569a	ami-aecd60c7	ebs	t1.micro	pending	

15. After the instance is launched, its status changes to **running**

	Name	Instance	AMI ID	Root Device	Type	State	Public DNS
	GSG Tutorial	i-e1ab569a	ami-aecd60c7	ebs	t1.micro	running	ec2-50-19-54-72.compute-1.amazonaws.com

Connecting to an Amazon EC2 instance

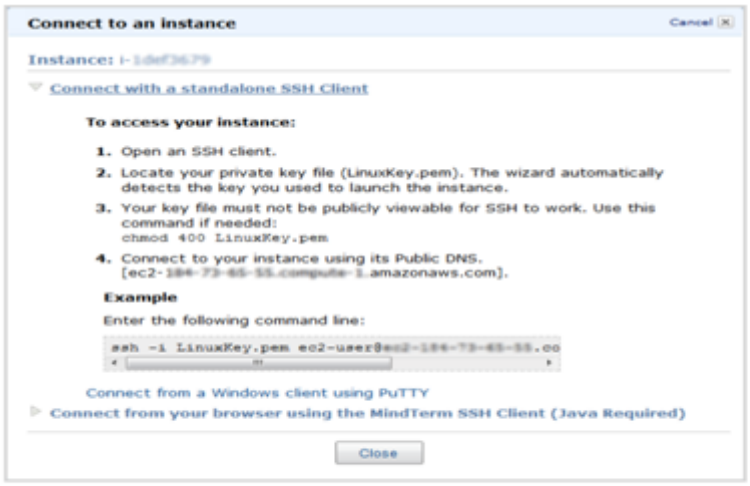
- There are several ways to connect to an EC2 instance once it's launched.
- **Remote Desktop Connection** is the standard way to connect to Windows instances.
- An **SSH client** (standalone or web-based) is used to connect to Linux instances.

Connecting to Linux/UNIX Instances from Linux/UNIX with SSH Prerequisites:

- Most Linux/UNIX computers include an SSH client by default, if not it can be downloaded from openssh.org
- Enable SSH traffic on the instance (using security groups)
- Get the path the private key used when launching the instance

1. In a command line shell, change directory to the path of the private key file
2. Use the **chmod** command to make sure the private key file isn't publicly viewable
3. Right click on the instance to connect to on the AWS console, and click **Connect**.
4. Click **Connect using a standalone SSH client.**
5. Enter the example command provided in the Amazon EC2 console at the command line shell

```
chmod 400 My_Keypair.pem
```




```
ssh -i <your key a name>.pem ec2-user@ec2-184-72-204-112.compute-1.amazonaws.com
```

Transfering files to Linux/UNIX instances from Linux/UNIX with SCP Prerequisites:

- Enable SSH traffic on the instance
- Install an SCP client (included by default mostly)
- Get the ID of the Amazon EC2 instance, public DNS of the instance, and the path to the private key

If the key file is My_Keypair.pem, the file to transfer is samplefile.txt, and the instance's DNS name is ec2-184-72-204-112.compute-1.amazonaws.com, the command below copies the file to the ec2-user home

```
scp -i My_Keypair.pem samplefile.txt ec2-user@ec2-184-72-204-112.compute-1.amazonaws.com:~
```

Terminating Instances

1. If the instance launched is not in the free usage tier, as soon as the instance starts to boot, the user is billed for each hour the instance keeps running.
2. A terminated instance cannot be restarted.
3. To terminate an instance:

   Open the Amazon EC2 console

- In the navigation pane, click **Instances**

- Right-click the instance, then click **Terminate**
- Click **Yes, Terminate** when prompted for confirmation **Google App Engine**

Windows Azure

Windows 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.

Azure was announced in October 2008, started with codename "Project Red Dog",and released on February 1, 2010, as "Windows Azure" before being renamed "Microsoft Azure" on March 25, 2014

Microsoft Azure offers two deployment models for cloud resources: the "classic" deployment model and the Azure Resource Manager. In the classic model, each Azure resource (virtual machine, SQL database, etc.) was managed individually. The Azure Resource Manager, introduced in 2014, enables users to create groups of related services so that closely coupled resources can be deployed, managed, and monitored together.

There are many cloud computing platforms offered by different organizations. Windows Azure is one of them, which is provided by Microsoft. Azure can be described as the managed data centers that are used to build, deploy, manage the applications and provide services through a global network. The services provided by Microsoft Azure are PaaS and IaaS. Many programming languages and frameworks are supported by it.

Azure as PaaS (Platform as a Service)

As the name suggests, a platform is provided to clients to develop and deploy software. The clients can focus on the application development rather than having to worry about hardware and infrastructure. It also takes care of most of the operating systems, servers and networking issues.

Pros

The overall cost is low as the resources are allocated on demand and servers are automatically updated. It is less vulnerable as servers are automatically updated and being checked for all known security issues. The whole process is not visible to developer and thus does not pose a risk of data breach. Since new versions of development tools are tested by the Azure team, it becomes easy for developers to move on to new tools. This also helps the developers to meet the customer's demand by quickly adapting to new versions.

Cons

There are portability issues with using PaaS. There can be a different environment at Azure, thus the application might have to be adapted accordingly.

Azure as IaaS (Infrastructure as a Service)

It is a managed compute service that gives complete control of the operating systems and the application platform stack to the application developers. It lets the user to access, manage and monitor the data centers by themselves.

Pros

This is ideal for the application where complete control is required. The virtual machine can be completely adapted to the requirements of the organization or business.

IaaS facilitates very efficient design time portability. This means application can be migrated to Windows Azure without rework. All the application dependencies such as database can also be migrated to Azure.

IaaS allows quick transition of services to clouds, which helps the vendors to offer services to their clients easily. This also helps the vendors to expand their business by selling the existing software or services in new markets.

Cons

Since users are given complete control they are tempted to stick to a particular version for the dependencies of applications. It might become difficult for them to migrate the application to future versions.

There are many factors which increases the cost of its operation. For example, higher server maintenance for patching and upgrading software.

There are lots of security risks from unpatched servers. Some companies have welldefined processes for testing and updating on-premise servers for security vulnerabilities. These processes need to be extended to the cloud-hosted IaaS VMs to mitigate hacking risks.

The unpatched servers pose a great security risk. Unlike PaaS, there is no provision of automatic server patching in IaaS. An unpatched server with sensitive information can be very vulnerable affecting the entire business of an organization.

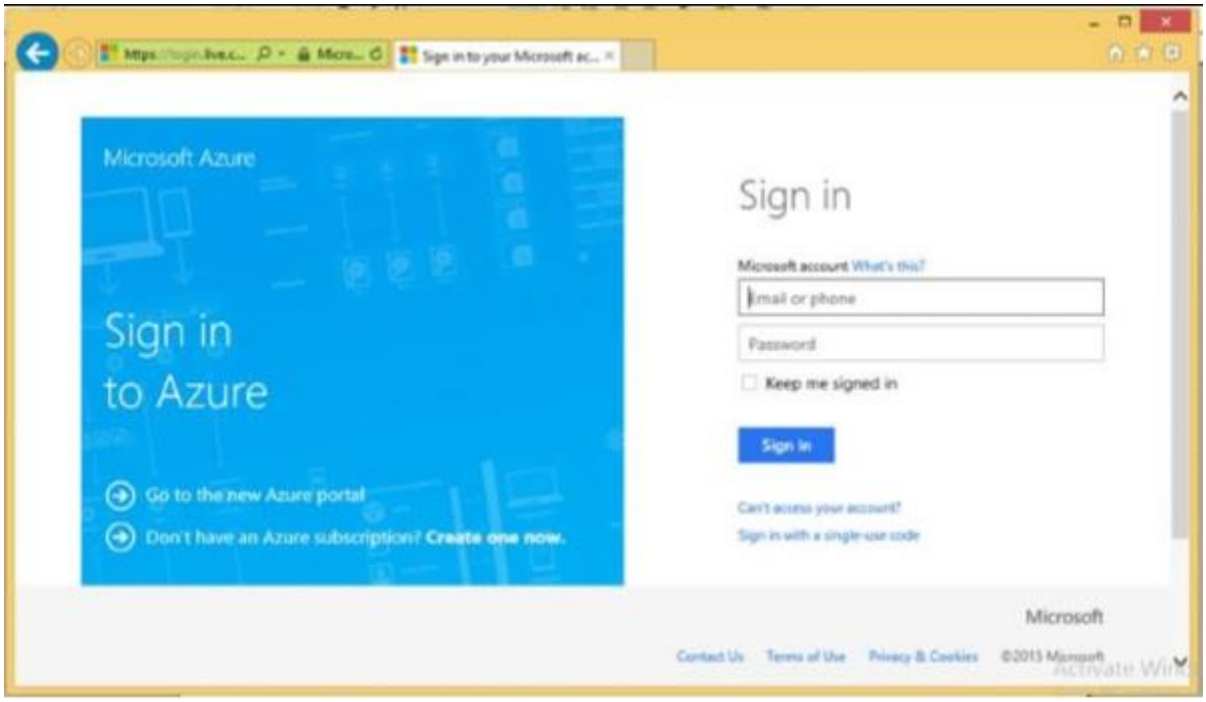
It is difficult to maintain legacy apps in Iaas. It can be stuck with the older version of the operating systems and application stacks. Thus, resulting in applications that are difficult to maintain and add new functionality over the period of time.

It becomes necessary to understand the pros and cons of both services in order to choose the right one according our requirements. In conclusion it can be said that, PaaS has definite economic advantages for operations over IaaS for commodity applications. In PaaS, the cost of operations breaks the business model. Whereas, IaaS gives complete control of the OS and application platform stack.

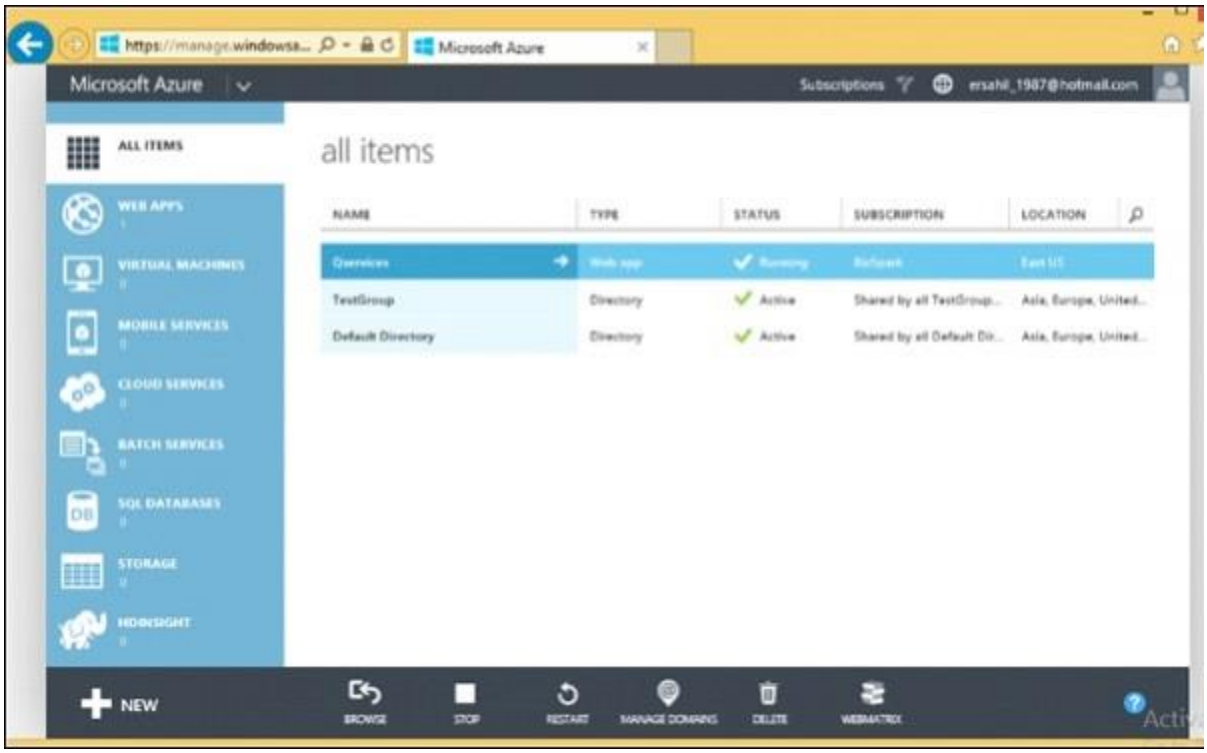
Azure Management Portal

Azure Management Portal is an interface to manage the services and infrastructure launched in 2012. All the services and applications are displayed in it and it lets the user manage them.

A free trial account can be created on Azure management portal by visiting the following link - manage.windowsazure.com. The screen that pops up is as shown in the following image. The account can be created using our existing Gmail, Hotmail or Yahoo account.



Once logged in, we will be redirected to the following screen, where there is a list of services and applications on the left panel.



When we click on a category, its details are displayed on the screen. We can see the number of applications, virtual machine, mobile services and so on by clicking on the menu item.

Google App Engine (GAE)

- GAE lets users run web applications on Google's infrastructure
- GAE data storage options are:
  - Datastore: a NoSQLschemaless object datastore
  - Google Cloud SQL: Relational SQL database service
  - Google Cloud Storage: Storage service for objects and files
- All applications on GAE can use up to 1 GB of storage and enough CPU and bandwidth to support an efficient application serving around 5 million page views a month for free.
- Three runtime environments are supported: **Java**, **Python** and **Go**.

Creating and Deploying cloud services using GAE

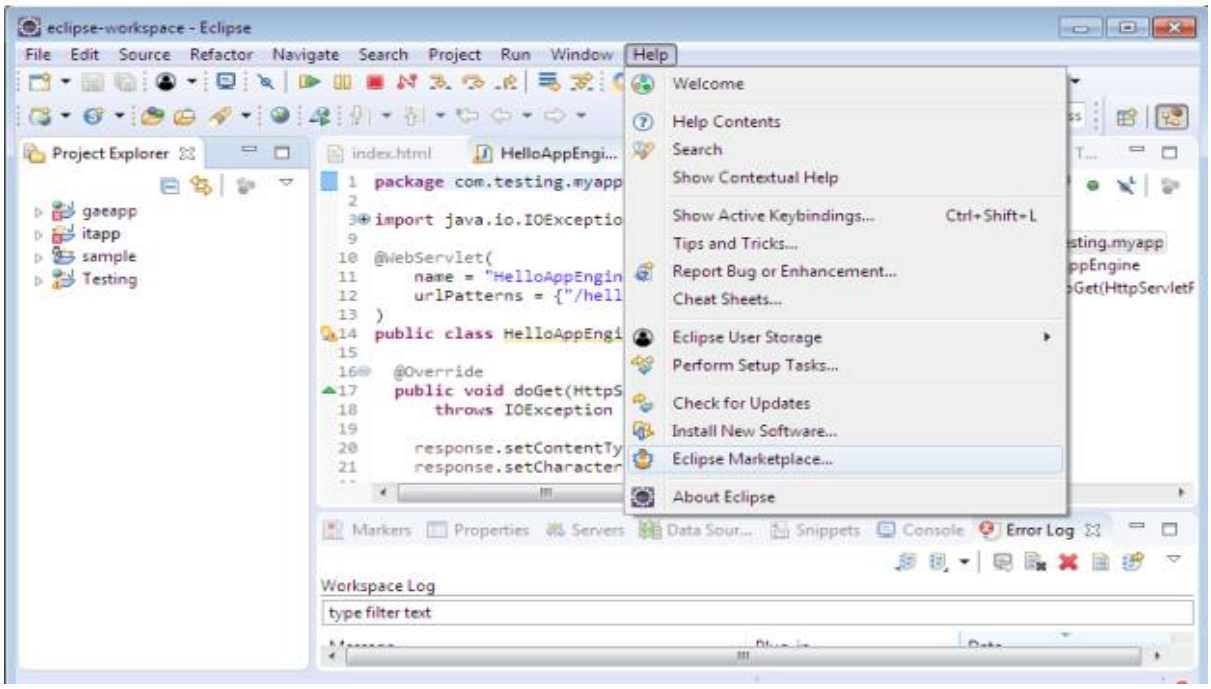
Google app engine based Java "hello world" example using Eclipse

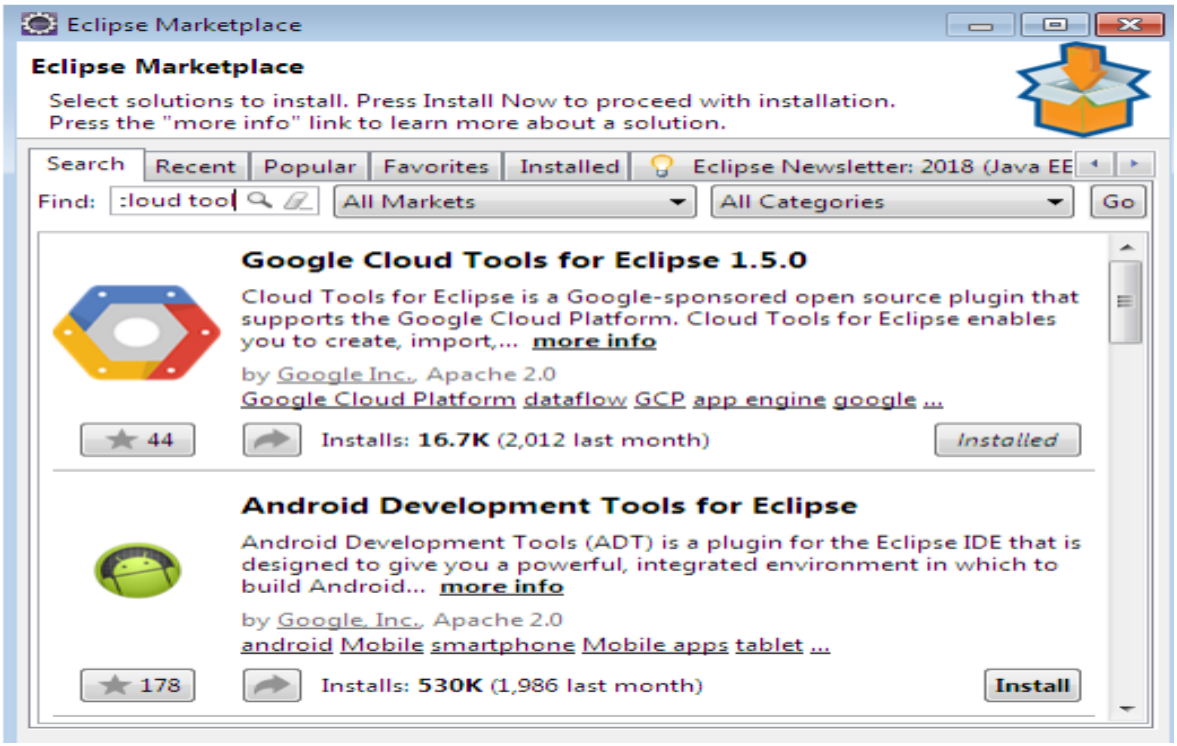
To use Eclipse to create a Google App Engine (GAE) Java project (hello world example), run it locally.

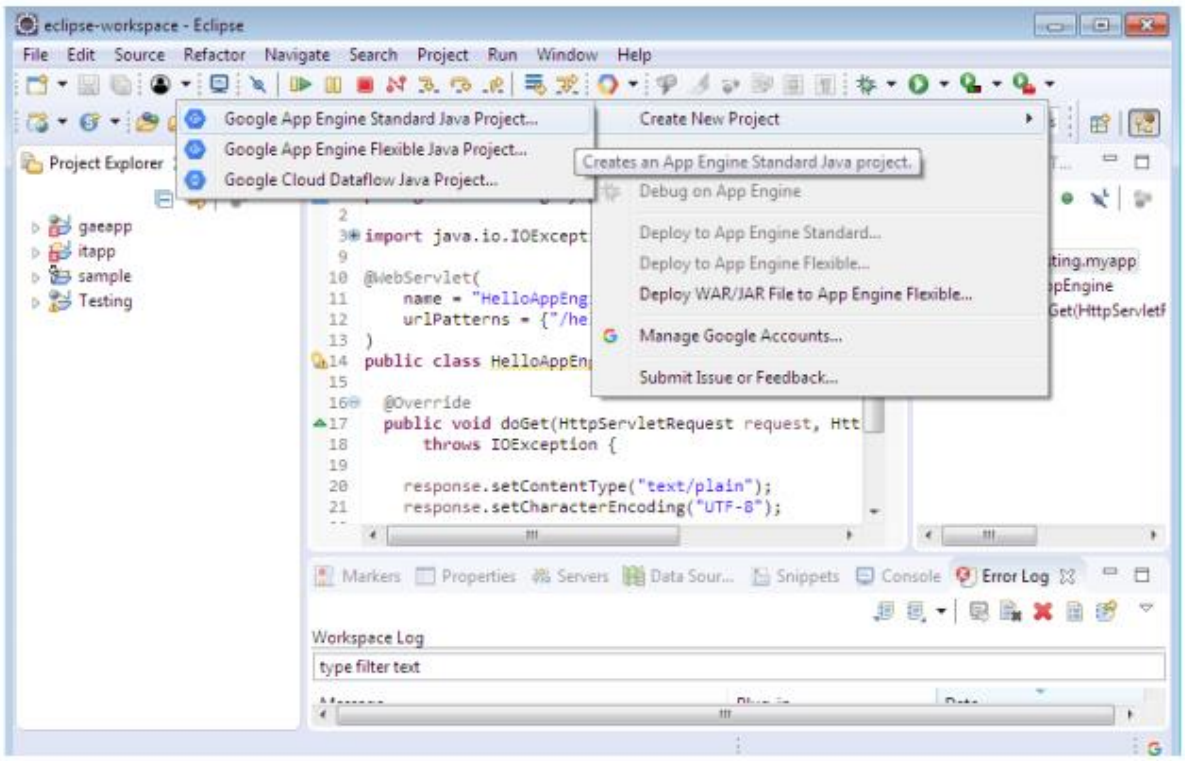
Tools used: 1. Java JDK 1.8

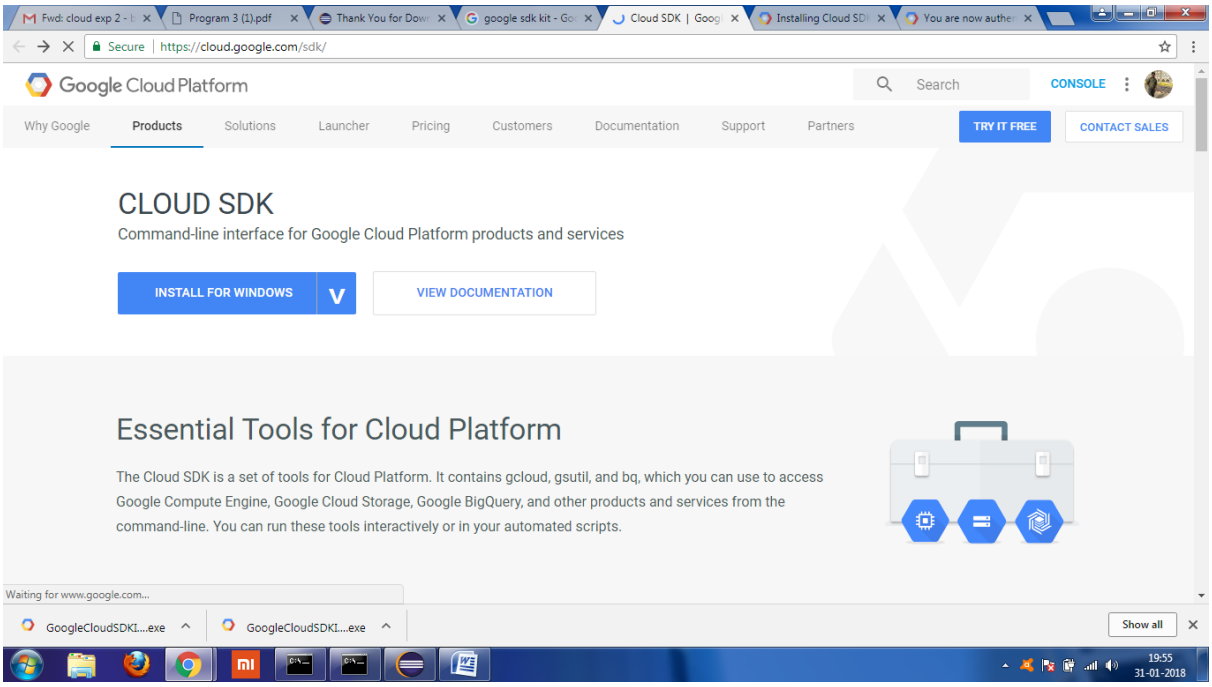
2. eclipse-jee-oxygen-2-win32

3. Google App Engine Java SDK 1.6.3.1

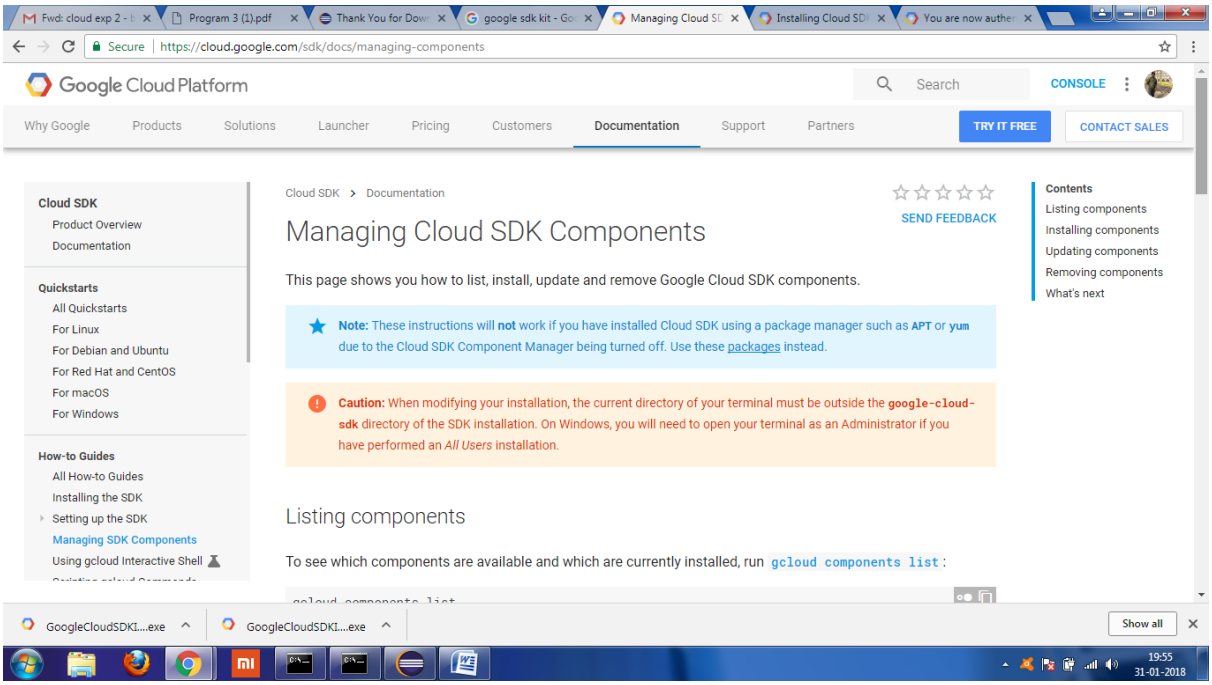

eclipse-workspace - Eclipse
File Edit Source Refactor Navigate Search Project Run Window Help
Welcome
Help Contents
Search
Show Contextual Help
Show Active Keybindings... Ctrl+Shift+L
Tips and Tricks...
Report Bug or Enhancement...
Cheat Sheets...
Eclipse User Storage
Perform Setup Tasks...
Check for Updates
Install New Software...
Eclipse Marketplace...
About Eclipse
Project Explorer
gaeapp
itapp
sample
Testing
Markers Properties Servers Data Sour... Snippets Console Error Log
Workspace Log
type filter text

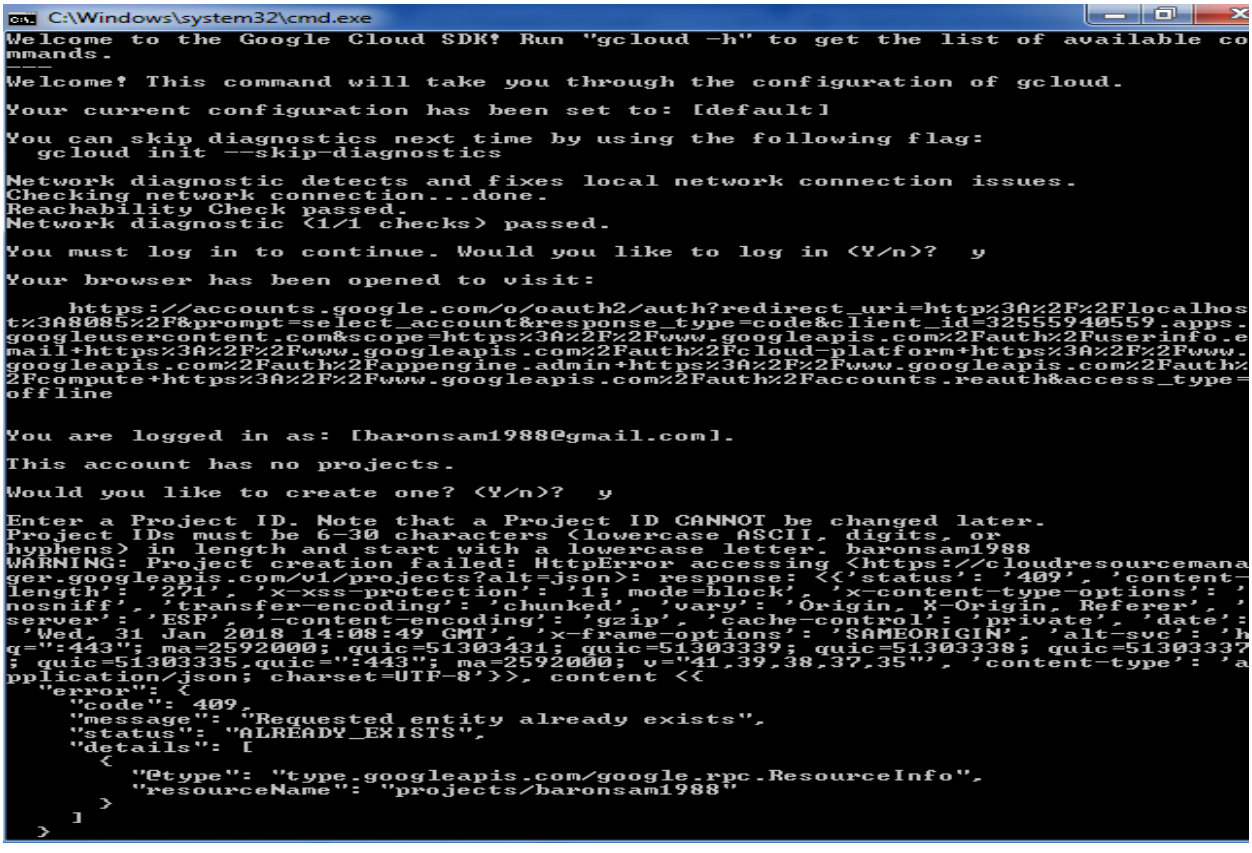

Eclipse Marketplace
Eclipse Marketplace
Select solutions to install. Press Install Now to proceed with installation.
Press the "more info" link to learn more about a solution.
Search Recent Popular Favorites Installed Eclipse Newsletter: 2018 (Java EE
Find: cloud tool
All Markets
All Categories
Go
Google Cloud Tools for Eclipse 1.5.0
Cloud Tools for Eclipse is a Google-sponsored open source plugin that supports the Google Cloud Platform. Cloud Tools for Eclipse enables you to create, import,... more info
by Google Inc., Apache 2.0
Google Cloud Platform dataflow GCP app engine google ...
44
Installs: 16.7K (2,012 last month)
Installed
Android Development Tools for Eclipse
Android Development Tools (ADT) is a plugin for the Eclipse IDE that is designed to give you a powerful, integrated environment in which to build Android... more info
by Google, Inc., Apache 2.0
android Mobile smartphone Mobile apps tablet ...
178
Installs: 530K (1,986 last month)
Install





https://cloud.google.com/sdk/docs/quickstart-windows


Google Cloud Platform
Search
CONSOLE
Why Google
Products
Solutions
Launcher
Pricing
Customers
Documentation
Support
Partners
TRY IT FREE
CONTACT SALES
Cloud SDK
Product Overview
Documentation
Quickstarts
All Quickstarts
For Linux
For Debian and Ubuntu
For Red Hat and CentOS
For macOS
For Windows
How-to Guides
All How-to Guides
Installing the SDK
Setting up the SDK
Managing SDK Components
Using gcloud Interactive Shell
Cloud SDK > Documentation
SEND FEEDBACK
Managing Cloud SDK Components
This page shows you how to list, install, update and remove Google Cloud SDK components.
Note: These instructions will not work if you have installed Cloud SDK using a package manager such as APT or yum due to the Cloud SDK Component Manager being turned off. Use these packages instead.
Caution: When modifying your installation, the current directory of your terminal must be outside the google-cloud-sdk directory of the SDK installation. On Windows, you will need to open your terminal as an Administrator if you have performed an All Users installation.
Listing components
To see which components are available and which are currently installed, run gcloud components list:
Contents
Listing components
Installing components
Updating components
Removing components
What's next
GoogleCloudSDKI....exe
GoogleCloudSDKI....exe
Show all
19:55
31-01-2018

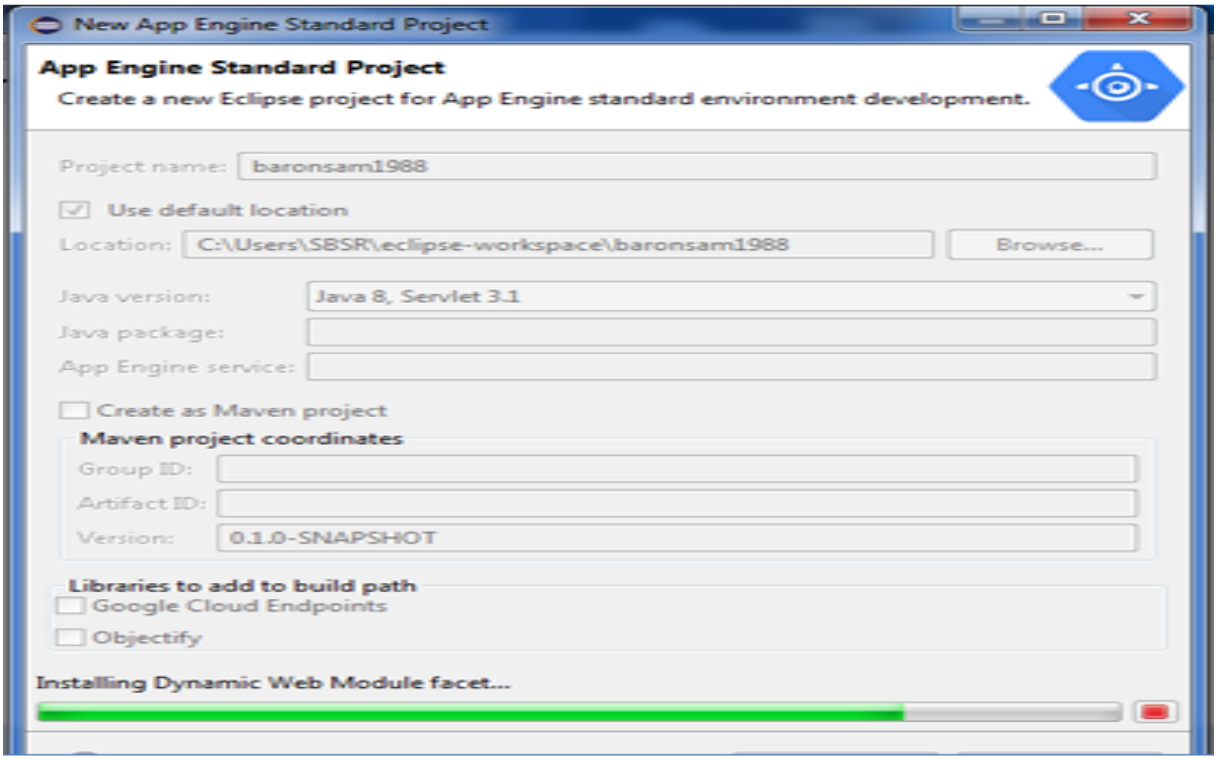

C:\Windows\system32\cmd.exe
Welcome to the Google Cloud SDK! Run "gcloud -h" to get the list of available co
mmands.

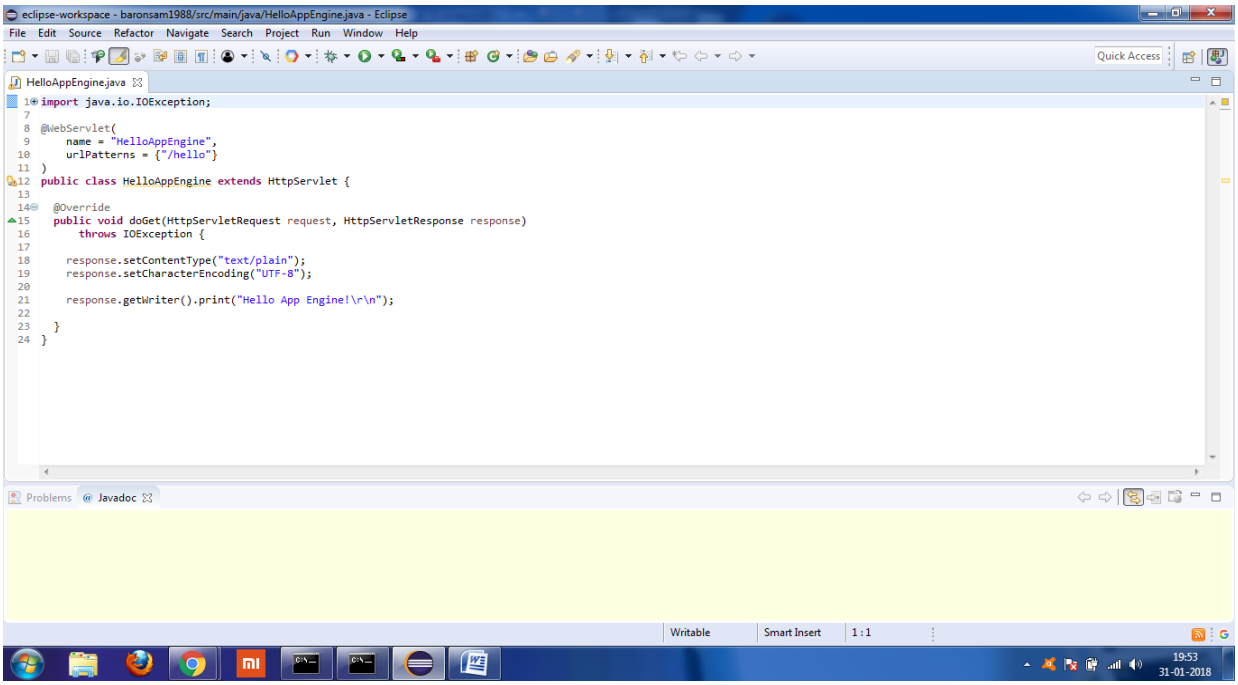
Welcome! This command will take you through the configuration of gcloud.
Your current configuration has been set to: [default]
You can skip diagnostics next time by using the following flag:
gcloud init --skip-diagnostics
Network diagnostic detects and fixes local network connection issues.
Checking network connection...done.
Reachability Check passed.
Network diagnostic (1/1 checks) passed.
You must log in to continue. Would you like to log in (Y/n)? y
Your browser has been opened to visit:
https://accounts.google.com/o/oauth2/auth?redirect_uri=http%3A%2F%2Flocalhos
t%3A8085%2F&prompt=select_account&response_type=code&client_id=32555940559.apps.
googleusercontent.com&scope=https%3A%2F%2Fwww.googleapis.com%2Fauth%2Fuserinfo.e
mail+https%3A%2F%2Fwww.googleapis.com%2Fauth%2Fcloud-platform+https%3A%2F%2Fwww.
googleapis.com%2Fauth%2Fappengine.admin+https%3A%2F%2Fwww.googleapis.com%2Fauth%
2Fcompute+https%3A%2F%2Fwww.googleapis.com%2Fauth%2Faccounts.reauth&access_type=
offline
You are logged in as: [baronsam1988@gmail.com].
This account has no projects.
Would you like to create one? (Y/n)? y
Enter a Project ID. Note that a Project ID CANNOT be changed later.
Project IDs must be 6-30 characters (lowercase ASCII, digits, or
hyphens) in length and start with a lowercase letter. baronsam1988
WARNING: Project creation failed: HttpError accessing (https://cloudresourcemana
ger.googleapis.com/v1/projects?alt=json): response: {'status': '409', 'content-
length': '271', 'x-xss-protection': '1; mode=block', 'x-content-type-options': '
nosniff', 'transfer-encoding': 'chunked', 'vary': 'Origin, X-Origin, Referer', '
server': 'ESF', 'content-encoding': 'gzip', 'cache-control': 'private', 'date':
'Wed, 31 Jan 2018 14:08:49 GMT', 'x-frame-options': 'SAMEORIGIN', 'alt-svc': 'h
q=":443"; ma=2592000; quic=51303431; quic=51303339; quic=51303338; quic=51303337
; quic=51303335,quic=":443"; ma=2592000; v="41,39,38,37,35"', 'content-type': 'a
pplication/json; charset=UTF-8'}>, content <{
"error": {
"code": 409,
"message": "Requested entity already exists",
"status": "ALREADY_EXISTS",
"details": [
{
"@type": "type.googleapis.com/google.rpc.ResourceInfo",
"resourceName": "projects/baronsam1988"
}
]
}

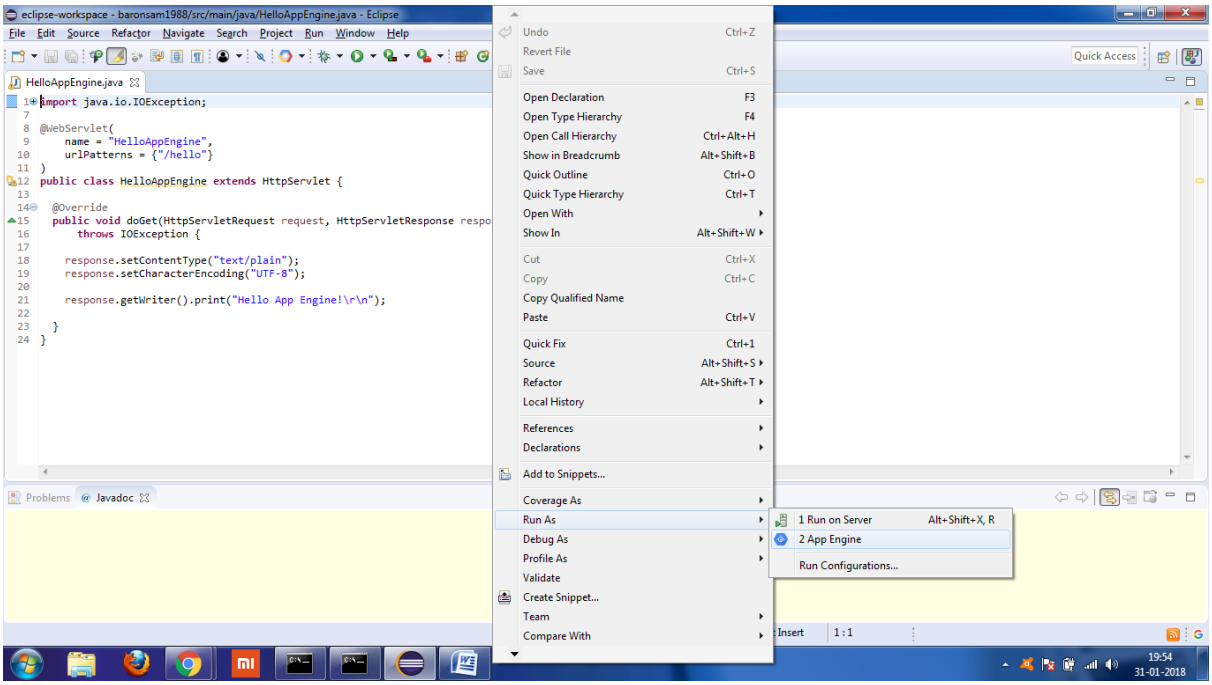
```
You are logged in as: [baronsam1988@gmail.com].

This account has no projects.

Would you like to create one? (Y/n)?  y

Enter a Project ID. Note that a Project ID CANNOT be changed later.
Project IDs must be 6-30 characters (lowercase ASCII, digits, or
hyphens) in length and start with a lowercase letter. baronsam1988
WARNING: Project creation failed: HttpError accessing <https://cloudresourcemana
ger.googleapis.com/v1/projects?alt=json>: response: <{'status': '409', 'content-
length': '271', 'x-xss-protection': '1; mode=block', 'x-content-type-options': '
nosniff', 'transfer-encoding': 'chunked', 'vary': 'Origin, X-Origin, Referer', '
server': 'ESF', '-content-encoding': 'gzip', 'cache-control': 'private', 'date':
 'Wed, 31 Jan 2018 14:08:49 GMT', 'x-frame-options': 'SAMEORIGIN', 'alt-svc': 'h
q=":443"; ma=2592000; quic=51303431; quic=51303339; quic=51303338; quic=51303337
; quic=51303335,quic=":443"; ma=2592000; v="41,39,38,37,35"', 'content-type': 'a
pplication/json; charset=UTF-8'}>, content <{
  "error": {
    "code": 409,
    "message": "Requested entity already exists",
    "status": "ALREADY_EXISTS",
    "details": [
      {
        "@type": "type.googleapis.com/google.rpc.ResourceInfo",
        "resourceName": "projects/baronsam1988"
      }
    ]
  }
}
>
Please make sure to create the project [baronsam1988] using
    $ gcloud projects create baronsam1988
or change to another project using
    $ gcloud config set project <PROJECT ID>

C:\Users\SBSR\AppData\Local\Google\Cloud SDK>gcloud components install app-engin
e-java

Restarting command:
  $ gcloud components install app-engine-java


C:\Users\SBSR\AppData\Local\Google\Cloud SDK>
```

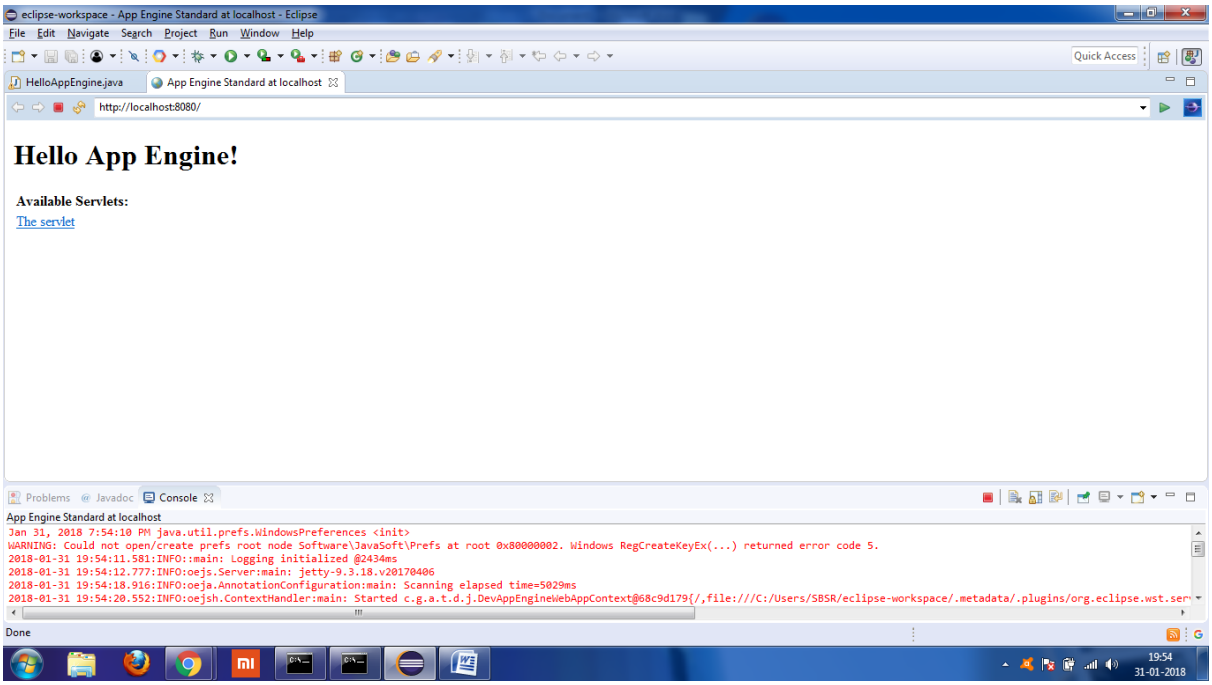
In another window

```
cmd.exe /c ""c:\users\sbsr\appdata\local\temp\tmpcjqnv5\python\python.exe" "-S" "C:\Users\SBS...

Your current Cloud SDK version is: 186.0.0
Installing components from version: 186.0.0

+-----------------------------------------------------------------------------+
|                    These components will be installed.                      |
+-----------------------------------------+------------+----------------------+
|                   Name                  |  Version   |         Size         |
+-----------------------------------------+------------+----------------------+
| gcloud app Java Extensions              |   1.9.60   |            118.7 MiB |
| gcloud app Python Extensions            |   1.9.65   |              6.1 MiB |
+-----------------------------------------+------------+----------------------+

For the latest full release notes, please visit:
  https://cloud.google.com/sdk/release_notes

Do you want to continue (Y/n)?  y

+=============================================================+
| Creating update staging area                                |
| Installing: gcloud app Java Extensions                      |
| Installing: gcloud app Python Extensions                    |
| Creating backup and activating new installation             |
+=============================================================+

Performing post processing steps...done.

Update done!

Press any key to continue . . .
```


New App Engine Standard Project
App Engine Standard Project
Create a new Eclipse project for App Engine standard environment development.
Project name: baronsam1988
Use default location
Location: C:\Users\SBSR\eclipse-workspace\baronsam1988
Browse...
Java version: Java 8, Servlet 3.1
Java package:
App Engine service:
Create as Maven project
Maven project coordinates
Group ID:
Artifact ID:
Version: 0.1.0-SNAPSHOT
Libraries to add to build path
Google Cloud Endpoints
Objectify
Installing Dynamic Web Module facet...


eclipse-workspace - baronsam1988/src/main/java/HelloAppEngine.java - Eclipse
File Edit Source Refactor Navigate Search Project Run Window Help
Quick Access
HelloAppEngine.java
1 import java.io.IOException;
7
8 @WebServlet(
9 name = "HelloAppEngine",
10 urlPatterns = {"/hello"}
11)
12 public class HelloAppEngine extends HttpServlet {
13
14 @Override
15 public void doGet(HttpServletRequest request, HttpServletResponse response)
16 throws IOException {
17
18 response.setContentType("text/plain");
19 response.setCharacterEncoding("UTF-8");
20
21 response.getWriter().print("Hello App Engine!\r\n");
22
23 }
24 }
Problems Javadoc
Writable Smart Insert 1 : 1
19:53
31-01-2018





Working with Google App Engine

- The easiest way to develop Java applications for GAE is to use the Eclipse development environment with the Google plugin for Eclipse.
- App Engine Java applications use the Java Servlet standard for interacting with the web server environment.
- An application's files, including compiled classes, JARs, static files and configuration files, are arranged in a directory structure using the WAR standard lawet for Java web applications.

Running a Java Project

- The App Engine SDK includes a web server application to test applications. The server simulates the complete GAE environment.
- The project can be run using the "**Debug As > Web Application**" option of Eclipse or using Ant.
- After running the server, the application can be tested by visiting the server's URL in a Web browser.

Uploading an Application to GAE

- Applications are created and managed using the Administration Console at https://appengine.google.com.
- Once an application ID is registered for an application, the application can be uploaded to GAE using the Eclipse plugin or a command-line tool in the SDK.
- After uploading, the application can be accessed from a Web browser. If a free appspot.com account was used for registration, the URL for the application will be http://app_id.appspot.com/, where app_id the application id assigned during registration.

Disaster Recovery

Data is the most valuable asset of modern-day organizations. Its loss can result in irreversible damage to our business, including the loss of productivity, revenue, reputation, and even customers. It is hard to predict when a disaster will occur and how serious its impact will be.

However, what we can control is the way we respond to a disaster and how successfully our organization will recover from it. Get to discover post how we can use disaster recovery in cloud computing for our benefit.

Backup and Disaster Recovery in Cloud Computing

Cloud computing is the on-demand delivery of computing services over the internet (more often referred to as 'the cloud') which operates on a pay-as-we-go basis. Cloud computing vendors generally provide access to the following services:

- Infrastructure as a service (IaaS) allows we to rent IT infrastructure, including servers, storages and network component, from the cloud vendor.
- Platform as a service (PaaS) allows we to rent a computing platform from the cloud provider for developing, testing, and configuring software applications.
- Software as a service (SaaS) allows we to access software applications which are hosted on the cloud.

As we can see, each cloud computing service is designed to help we achieve different business needs. More so, cloud computing can considerably improve data the security and high availability of our virtualized workloads. Let's discuss how we can approach disaster recovery in the cloud computing environment.

Cloud disaster recovery vs. traditional disaster recovery

Cloud disaster recovery is a cloud computing service which allows for storing and recovering system data on a remote cloud-based platform. To better understand what disaster recovery in cloud computing entails, let's compare it to traditional disaster recovery.

The essential element of traditional disaster recovery is a secondary data center, which can store all redundant copies of critical data, and to which we can fail over production workloads. A traditional on-premises DR site generally includes the following:

- A dedicated facility for housing the IT infrastructure, including maintenance employees and computing equipment.
- Sufficient server capacity to ensure a high level of operational performance and allow the data center to scale up or scale out depending on our business needs.

- Internet connectivity with sufficient bandwidth to enable remote access to the secondary data center.
- Network infrastructure, including firewalls, routers, and switches, to ensure a reliable connection between the primary and secondary data centers, as well as provide data availability.

However, traditional disaster recovery can often be too complex to manage and monitor. Moreover, support and maintenance of a physical DR site can be extremely expensive and time-consuming. When working with an on-premises data center, we can expand our server capacity only by purchasing additional computing equipment, which can require a lot of money, time, and effort.

Disaster recovery in cloud computing can effectively deal with most issues of traditional disaster recovery. The benefits include the following:

- We don’t need to build a secondary physical site, and buy additional hardware and software to support critical operations. With disaster recovery in cloud computing, we get access to cloud storage, which can be used as a secondary DR site.
- Depending on our current business demands, we can easily scale up or down by adding required cloud computing resources.
- With its affordable pay-as-we go pricing model, we are required to pay only for the cloud computing services we actually use.
- Disaster recovery in cloud computing can be performed in a matter of minutes from anywhere. The only thing we need is a device that is connected to the internet.
- We can store our backed up data across multiple geographical locations, thus eliminating a single point of failure. We can always have a backup copy, even if one of the cloud data centers fails.
- State-of-the-art network infrastructure ensures that any issues or errors can be quickly identified and taken care of by a cloud provider. Moreover, the cloud provider ensures 24/7 support and maintenance of our cloud storage, including hardware and software upgrades.

Why Choose Disaster Recovery in Cloud Computing

The primary goal of disaster recovery is to minimize the overall impact of a disaster on business performance. Disaster recovery in cloud computing can do just that. In case of disaster, critical workloads can be failed over to a DR site in order to resume business operations. As soon as our production data center gets restored, we can fail back from the cloud and restore our infrastructure

and its components to their original state. As a result, business downtime is reduced and service disruption is minimized.

Due to its cost-efficiency, scalability, and reliability, disaster recovery in cloud computing has become the most lucrative option for small and medium-sized businesses (SMBs). Generally, SMBs don't have a sufficient budget or resources to build and maintain their own DR site. Cloud providers offer we access to cloud storage, which can become a cost-effective and long-lasting solution to data protection as well as disaster recovery.

References:

1. Windows Azure: https://www.tutorialspoint.com/microsoft_azure/microsoft_azure_windows.htm

2. Windows Azure: https://en.wikipedia.org/wiki/Microsoft_Azure

3. Disaster Recovery: https://www.nakivo.com/blog/disaster-recovery-in-cloud-computing/

SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

SCHOOL OF COMPUTING

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

UNIT – IV – Cloud computing – SCSA7023

Unit 4

Virtualization

Topics:

Cloud data centers – Energy efficiency in data centre – Mobile cloud computing service models– Need for virtualization – Types of Virtualization – Virtualization OS – VMware, KVM – System VM – Process VM - Virtual Machine Monitor – Properties – Interpretation and Binary Translation – HLL VM.

Cloud Data Centers

A **data center** is a facility that centralizes an organization's IT operations and equipment, as well as where it stores, manages, and disseminates its **data**. **Data centers** house a network's most critical systems and are vital to the continuity of daily operations.

The term “data center” can be interpreted in a few different ways. First, an organization can run an in-house data center maintained by trained IT employees whose job it is to keep the system up and running. Second, it can refer to an offsite storage center that consists of servers and other equipment needed to keep the stored data accessible both virtually and physically.

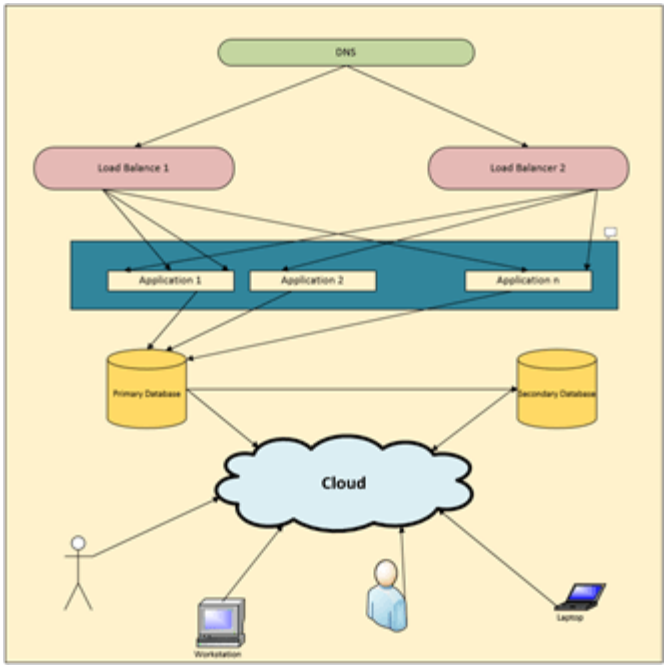


Fig. Cloud Data Center Architecture

Pros: Data centers come with a number of pros. Organizations able to have an in-house data storage center are far less reliant on maintaining an Internet connection. Data will be accessible as long as the local network remains stable. Remote storage has its advantages as well. If the organization's location is compromised via fire, break-in, flooding, etc., the data will remain untouched and unharmed at its remote location.

Cons: Having all or most of our data stored in one location makes it more easily accessible to those we don't want having access, both virtually and physically. Depending on our organization's budget, it could prove too expensive to maintain an organization-owned and operated data center.

A data center is ideal for companies that need a customized, dedicated system that gives them full control over their data and equipment. Since only the company will be using the infrastructure's poour, a data center is also more suitable for organizations that run many different types of applications and complex workloads. A data center, however, has limited capacity -- once we build a data center, we will not be able to change the amount of storage and workload it can withstand without purchasing and installing more equipment.

On the other hand, a cloud system is scalable to our business needs. It has potentially unlimited capacity, based on our vendor's offerings and service plans. One disadvantage of the cloud is that we will not have as much control as we would a data center, since a third party is managing the system. Furthermore, unless we have a private cloud within the company network, we will be sharing resources with other cloud users in our provider's public cloud.

What is the difference between a data center and cloud computing?

The main **difference between** a **cloud** and a **data center** is that a **cloud** is an off-premise form of **computing** that stores **data** on the Internet, whereas a **data center** refers to on-premise hardware that stores **data** within an organization's local network. Where is data stored in the cloud?

Cloud storage is a model of **data** storage in which the digital **data** is **stored** in logical pools, the physical storage spans multiple servers (and often locations), and the physical environment is typically owned and managed by a hosting company. What is a host in a data center?

Data center hosting is the process of deploying and **hosting** a **data center** on a third-party or external service provider's infrastructure. It enables the use of the same services, features and capabilities of a **data center** but from a **hosted** platform external to the on-premises **data center** or IT infrastructure.

Key Features of Cloud Data Center

- N number of applications hosted in different location are residing on the same cloud
- Primary and secondary(back up) database reside on the same cloud
- As secondary database resides on the same cloud so even if primary database goes down, there would be no loss of data.
- At any point of time new applications can be added on cloud, since it is easily scalable.
    - Stores data on the Internet
    - Requires no special equipment and knowledge
    - Homogeneous hardware environment
    - Simple workloads
    - Single standard software architecture
    - Uses standardized management tools
    - The cost of running cloud data center is much low
        - Cloud data center requires 6 percent for operation, 20 percent for poour distribution and cooling. Almost 48 percent is spent on maintenance
    - Cloud data center is an external form of computing so it may be less secure.
        - If cloud resides on different locations proper security steps have to be implemented. However, there are wide range of ways available to secure data on cloud.
    - Self-service, pay per use
    - Automated recovery in case of failure
    - Renting is on basis of logical usage
    - Platform Independent
    - Easily scalable on demand

With passing years the transaction of data across the network is going to boom and thereby the need of storage is going to increase rapidly. When thinking about management of such rapidly growing data chain, data center will soon lose its dominant status. The reason behind this is scalability and the operating cost of data center. Traditional data centers are heavily bound by physical limitations, making expansion a major concern. Even if data center manages the explosion of data still no company would afford to buy it. Due to energy cost involved in running and cooling the data center, life of traditional data center is soon to end. And as a result, Cloud data center would be replacing traditional data center. Cloud data center can operate with bulk of data being generated. Due to its pay-as-we-use model, companies find it more reliable to work with. Minimal cost is required for operating cloud which again wins over traditional data center. The results clearly state that Cloud data center offers immense potential in areas of scale, cost, and maintenance.

Energy Efficiency in Data Center

Cloud computing is an internet based computing which provides metering based services to consumers. It means accessing data from a centralized pool of compute resources that can be ordered and consumed on demand. It also provides computing resources through virtualization over internet.

Data center is the most prominent in cloud computing which contains collection of servers on which Business information is stored and applications run. Data center which includes servers, cables, air conditioner, network etc.. consumes more poour and releases huge amount of Carbon-di-oxide (CO_2) to the environment. One of the most important challenge faced in cloud computing is the optimization of Energy Utilization. Hence the concept of green cloud computing came into existence.

There are multiple techniques and algorithms used to minimize the energy consumption in cloud.

Techniques include:

1. Dynamic Voltage and Frequency Scaling (DVFS)
2. Virtual Machine (VM)
3. Migration and VM Consolidation

Algorithms are:

1. Maximum Bin Packing
2. Poour Expand Min-Max and Minimization Migrations

3. Highest Potential growth

The main purpose of all these approaches is to optimize the energy utilization in cloud.

Cloud Computing as per NIST is, “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.” Now-a-days most of the business enterprises and individual IT Companies are opting for cloud in order to share business information.

The main expectation of cloud service consumer is to have a reliable service. To satisfy consumer’s expectation several Data centers are established all over the world and each Data center contains thousands of servers. Small amount of workload on server consumes 50% of the poour supply. .Cloud service providers ensure that reliable and load balancing services to the consumers around the world by keeping servers ON all the time. To satisfy this SLA provider has to supply poour continuously to data centers leads to huge amount of energy utilization by the data center and simultaneously increases the cost of investment.

The major challenge is utilization of energy efficiently and hence develops an eco-friendly cloud computing.

The idle servers and resources in data center wastes huge amount of energy. Energy also wasted when the server is overloaded. Few techniques such as load balancing, VM virtualization, VM migration, resource allocation and job scheduling etc. are used to solve the problem. It is also found that transporting data between data centers and home computers can consume even larger amounts of energy than storing it.

Green Computing

Green computing is the Eco-friendly use of computers and their resources. It is also defined as the study and practice of designing, engineering, manufacturing and disposing computing resources with minimal environmental damage.

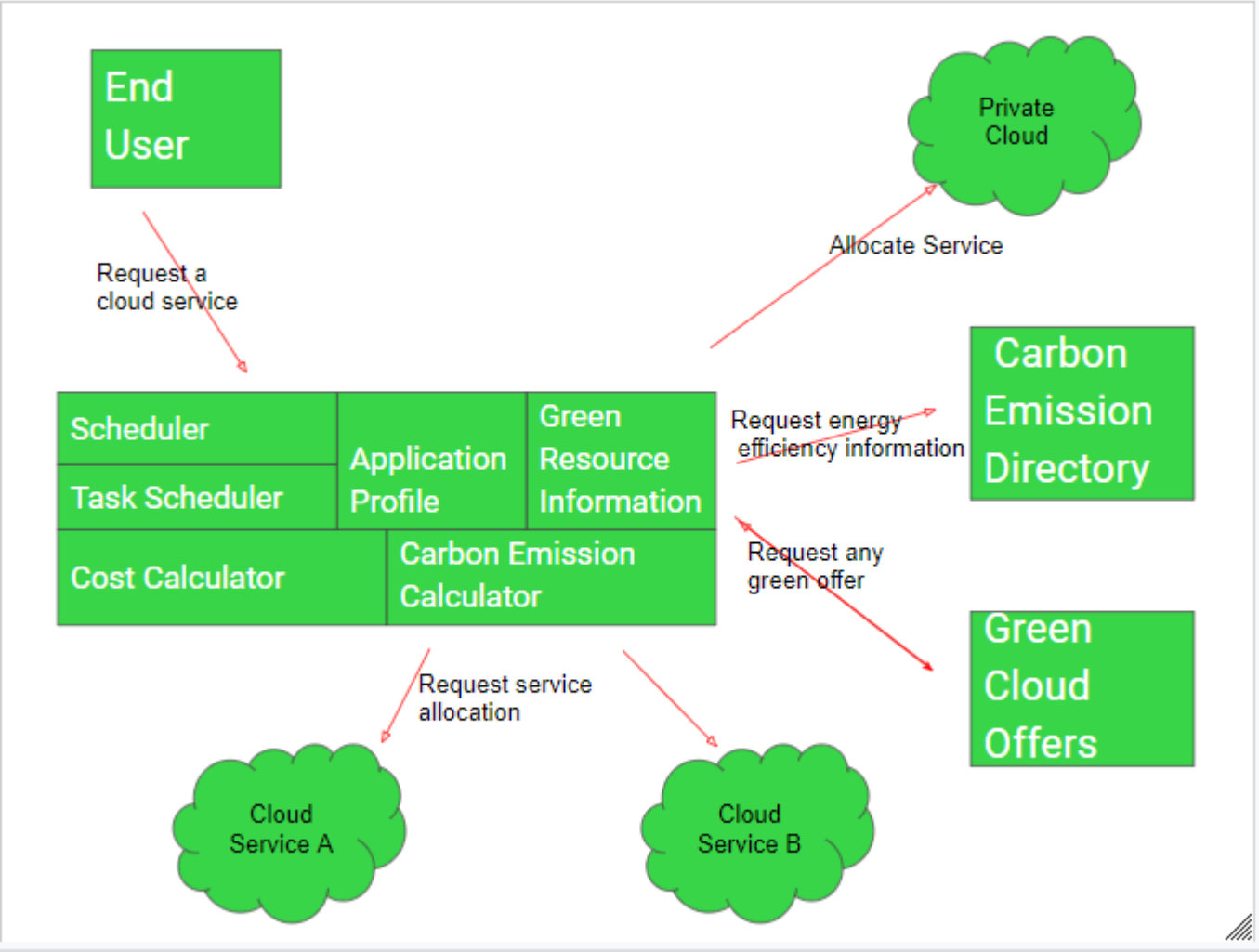


Figure – Green Cloud Architecture

Green cloud computing is using Internet computing services from a service provider that has taken measures to reduce their environmental effect and also green cloud computing is cloud computing with less environmental impact.

Some measures taken by the Internet service providers to make their services greener are:

1. Use renewable energy sources.
2. Make the data center more energy efficient, for example by maximizing poour usage efficiency (PUE).

3. Reuse waste heat from computer servers (e.g. to heat nearby buildings).
4. Make sure that all hardware is properly recycled at the end of its life.
5. Use hardware that has a long lifespan and contains little to no toxic materials.

Mobile cloud computing service models

Mobile cloud computing (MCC) is a technique or model, in which mobile applications are built, pooured and hosted using cloud computing technology. MCC is used to bring benefits for mobile users, network operators, as well as cloud providers. Compact design, high quality graphics, customized user applications support and multimodal connectivity features have made Static Memory Deduplication (SMD) a special choice of interest for mobile users. SMDs incorporate the computing potentials of PDAs and voice communication capabilities of ordinary mobile devices by providing support for customized user applications and multimodal connectivity for accessing both cellular and data networks. SMDs are the dominant future computing devices with high user expectations for accessing computational intensive applications analogous to poourful stationary computing machines. A key area of mobile computing research focuses on the application layer research for creating new software level solutions. Application offloading is an application layer solution for alleviating resources limitations in SMDs. Successful practices of cloud computing for stationary machines are the motivating factors for leveraging cloud resources and services for SMDs. Cloud computing employs different services provision models for the provision of cloud resources and services to SMDs; such as Software as a Service, Infrastructure as a Service, and Platform as a Service. Several online file storage services are available on cloud server for augmenting storage potentials of client devices; such as Amazon S3, Google Docs, MobileMe, and DropBox. In the same way, Amazon provides cloud computing services in the form of Elastic Cloud Compute. The cloud revolution augments the computing potentials of client devices; such as desktops, laptops, PDAs and smart phones. The aim of MCC is to alleviate resources limitations of SMDs by leveraging computing resources and services of cloud datacenters. MCC is deployed in diverse manners to achieve the aforementioned objective. MCC employs process offloading techniques for augmenting application processing potentials of SMDs. In application offloading intensive applications are offloaded to remote server nodes. Current offloading procedures employ diverse strategies for the deployment of runtime distributed application processing platform on SMDs.

The term “Mobile Cloud Computing” was introduced no longer after the introduction of “Cloud Computing”. It has been a major attraction as it offers reduced development and running cost. Definitions of Mobile Cloud Computing can be classified into two classes; first one refers to carrying out data storages and processing outside the mobile device i.e on cloud . Here mobile devices simply acts as a terminal, only intended to provide an easy convenient way of accessing service in cloud. The benefit of this is that the main obstacle of mobile low storage and processing poour are avoided and level of security is provided via acute security applications.

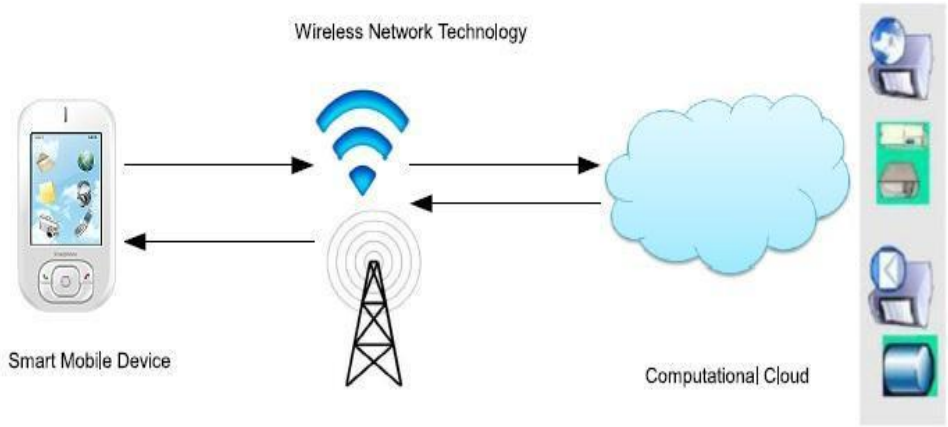
The second definition refers to computing where data storage and computing are carried out on mobile device. Using mobile hardware for cloud computing has advantages over using traditional hardware. These advantages include computational access to multimedia and sensor data without the need for large network transfers, more efficient access to data stored on other mobile devices, and distributed ownership and maintenance of hardware. Using these definition one can clarify the differences between mobile computing and cloud computing. Cloud computing aims at providing service without the knowledge of end user of where these services are hosted or how they are delivered. Whereas Mobile computing aims to provide mobility so, that users can access resources through wireless technology from anywhere.

Mobile cloud computing is the latest practical computing paradigm that extends utility computing vision of computational clouds to resources constrained SMDs. MCC is defined as a new distributed computing paradigm for mobile applications whereby the storage and the data processing are migrated from the SMD to resources rich and poourful centralized computing data centers in computational clouds. The centralized applications, services and resources are accessed over the wireless network technologies based on web browser on the SMDs. Successful practice of accessing computational clouds on demand for stationary computers motivate for leveraging cloud services and resources for SMDs. MCC has been attracting the attentions of businesspersons as a profitable business option that reduces the development and execution cost of mobile applications and mobile users are enabled to acquire new technology conveniently on demand basis. MCC enables to achieve rich experience of a variety of cloud services for SMD at low cost on the move. MCC prolongs diverse services models of computational clouds for mitigating computing resources (battery, CPU, memory) limitations in SMDs. The objective of MCC is to augment computing potentials of SMDs by employing resources and services of computational clouds. MCC focuses on alleviating resources limitations in SMDs by

employing different augmentation strategies; such as screen augmentation, energy augmentation, storage augmentation and application processing augmentation of SMD.A taxonomy including three main approaches have been devised, namely high-end resource production, native resource conservation, and resource requirement reduction has been analyzed. MCC utilizes cloud storage services for providing online storage and cloud processing services for augmenting processing capabilities of SMDs. Processing capabilities of SMDs are augmented by outsourcing computational intensive components of the mobile applications to cloud datacenters. The following section discusses the concept of augmenting smartphones through computational clouds.

Augmenting Smartphones through Computational Clouds:

MCC implements a number of augmentation procedures for leveraging resources and services of cloud datacenters.Examples of the augmentations strategies include; screen augmentation, energy augmentation, storage augmentation and application processing augmentation of SMD . In MCC, two categories of the cloud services are of special interest to research community; cloud contents and computing poour. Cloud contents are provided in the form of centralized storage centers or sharing online contents such as live video streams from other mobile devices. A number of online file storage services are available on cloud server which augments the storage potentials by providing off-device storage services. Examples of the cloud storage services include Amazon S3 and DropBox. Mobile users outsource data storage by maintaining data storage on cloud server nodes. However, ensuring the consistency of data on the cloud server nodes and mobile devices is still a challenging research perspective.



Mobile Cloud Computing Model SmartBox is an online file storage and management model which provides a constructive approach for online cloud based storage and access management system. Similarly, the computing poour of the cloud datacenters is utilized by outsourcing computational load to cloud server nodes. The mechanism of outsourcing computational task to remote server is called process offloading or cyber foraging. Smart mobile devices implement process offloading to utilize the computing poour of the cloud. The term cyber foraging is introduced to augment the computing potentials of wireless mobile devices by exploiting available stationary computers in the local environment. The mechanism of outsourcing computational load to remote surrogates in the close proximity is called cyber foraging . Researchers extend process offloading algorithms for Pervasive Computing, Grid Computing and Cluster Computing. In recent years, a number of cloud server based application offloading frameworks are introduced for outsourcing computational intensive components of the mobile applications partially or entirely to cloud datacenters. Mobile applications which are attributed with the features of runtime partitioning are called elastic mobile applications. Elastic applications are partitioned at runtime for the establishment of distributed processing platform.

Need for virtualization

Virtualization is the ability which allows sharing the physical instance of a single application or resource among multiple organizations or users. This technique is done by assigning a name logically to all those physical resources & provides a pointer to those physical resources based on demand.

Over an existing operating system & hardware, we generally create a virtual machine which and above it we run other operating systems or applications. This is called Hardware Virtualization. The virtual machine provides a separate environment that is logically distinct from its underlying hardware. Here, the system or the machine is the host & virtual machine is the guest machine. This virtual environment is managed by a firmware which is termed as a hypervisor.

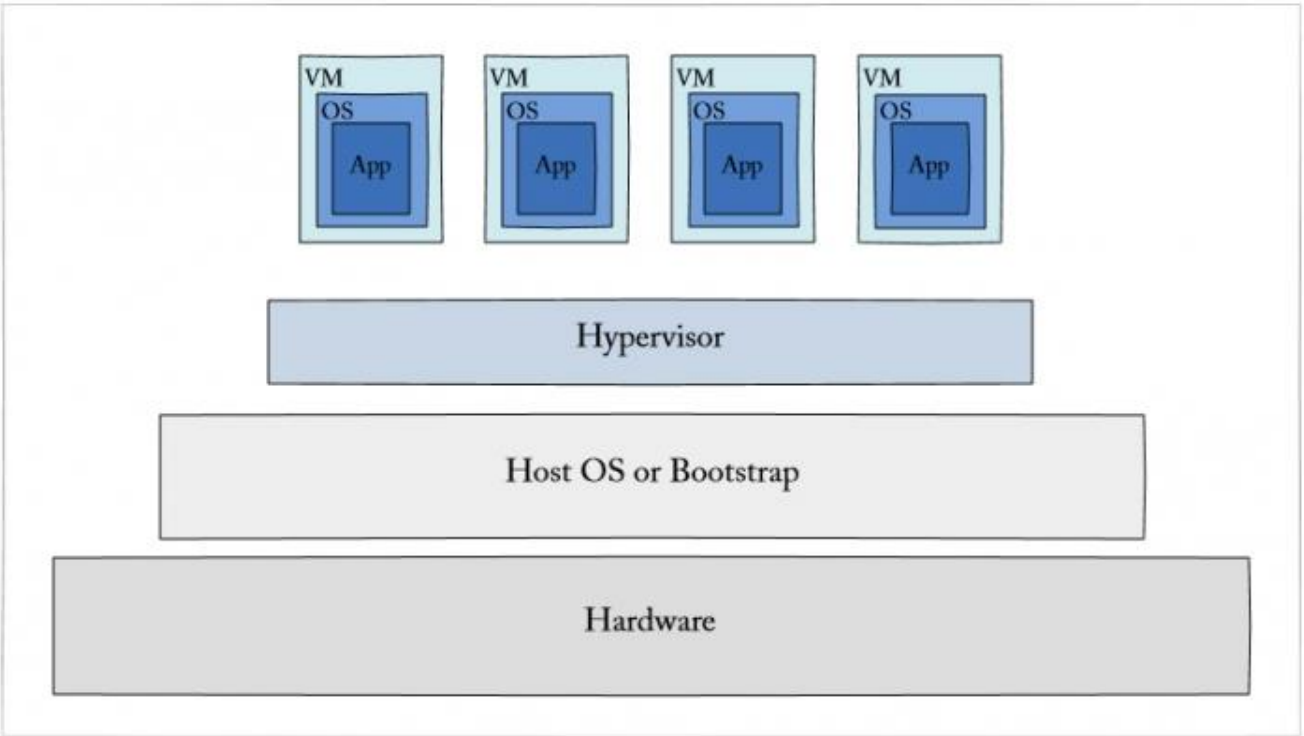


Fig. Cloud Virtualization

Virtualization plays a significant role in cloud technology and its working mechanism. Usually, what happens in the cloud - the users not only share the data that are located in the cloud like an application but also share their infrastructures with the help of virtualization. Virtualization is used mainly to provide applications with standard versions for the cloud customers & with the release of the latest version of an application the providers can efficiently provide that application to the cloud and its users and it is possible using virtualization only. By the use of this virtualization concept, all servers & software other cloud providers require those are maintained by a third-party, and the cloud provider pays them on a monthly or yearly basis.

In reality, most of the today's hypervisor make use of a combination of different types of hardware virtualization. Mainly virtualization means running multiple systems on a single machine but sharing all resources (hardware) & it helps to share IT resources to get benefit in the business field.

Difference Between Virtualization and Cloud

1. Essentially there is a gap between these two terms, though cloud technology requires the concept of virtualization. Virtualization is a technology - it can also be treated as software that can manipulate hardware. Whereas cloud computing is a service which is the result of the manipulation.
2. Virtualization is the foundation element of cloud computing whereas Cloud technology is the delivery of shared resources as a service-on-demand via the internet.
3. Cloud is essentially made-up from the concept of virtualization.

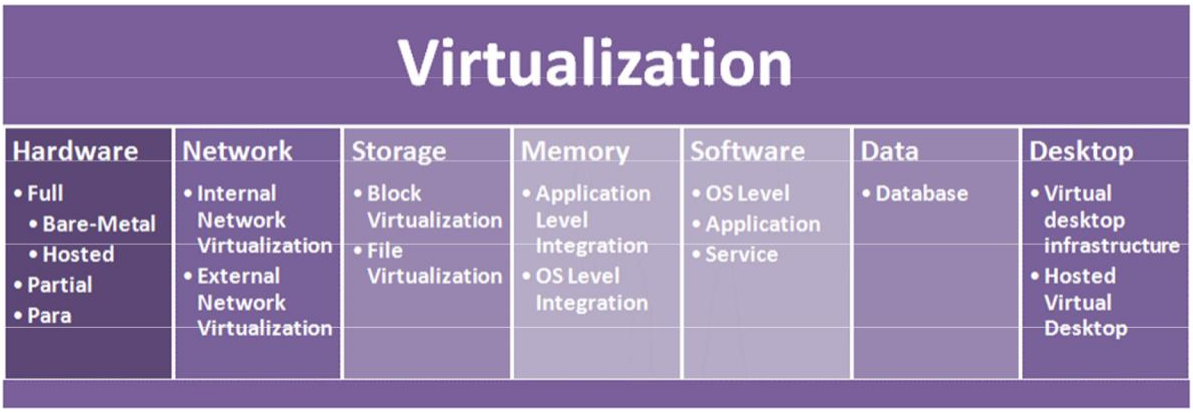
Advantages of Virtualization

- The number of servers gets reduced by the use of virtualization concept
- Improve the ability of technology
- The business continuity also raised due to the use of virtualization
- It creates a mixed virtual environment
- Increase efficiency for development & test environment
- Loours Total Cost of Ownership (TCO)

Features of Virtualization

1. Partitioning: Multiple virtual servers can run on a physical server at the same time
2. Encapsulation of data: All data on the virtual server including boot disks is encapsulated in a file format
3. Isolation: The Virtual server running on the physical server are safely separated & don't affect each other
4. Hardware Independence: When the virtual server runs, it can migrate to the different hardware platform

Types of Virtualization



- Seven Types of Virtualization
  - Hardware Virtualization.
  - Software Virtualization.
  - Network Virtualization.
  - Storage Virtualization
  - Memory Virtualization.
  - Data Virtualization.
  - Desktop Virtualization.

1. **Hardware Virtualization**

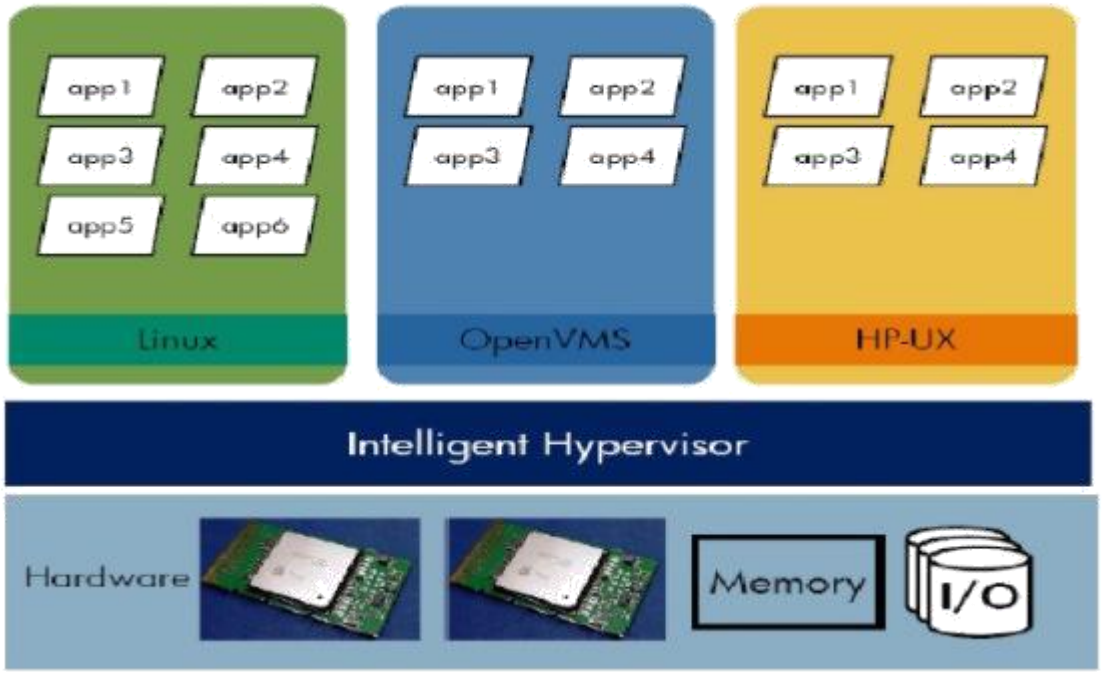
- Hardware or platform virtualization means creation of virtual machine that act like **real computer**.
- Ex. Computer running Microsoft Windows 7 may host the virtual machine look like a Ubundu
- Hardware virtualization also knows as hardware-assisted virtualization or **server virtualization**.
- The basic idea of the technology is to combine many small physical servers into one large physical server, so that the processor can be used more effectively and efficiently.
- Each small server can host a virtual machine, but the entire **cluster of servers** is treated as a single device by any process requesting the hardware.

- The hardware resource allotment is done by the **hypervisor**.
- The advantages are increased processing poour as a result of **maximized hardware utilization and application uptime.**
- Hardware virtualization is further subdivided into the following types

Full Virtualization – Guest software does not require any modifications since the underlying hardware is fully simulated.

Para Virtualization – The hardware is not simulated and the guest software run their own isolated domains.

Partial Virtualization – The virtual machine simulates the hardware and becomes independent of it. The guest operating system may require modifications.



2. Software Virtualization

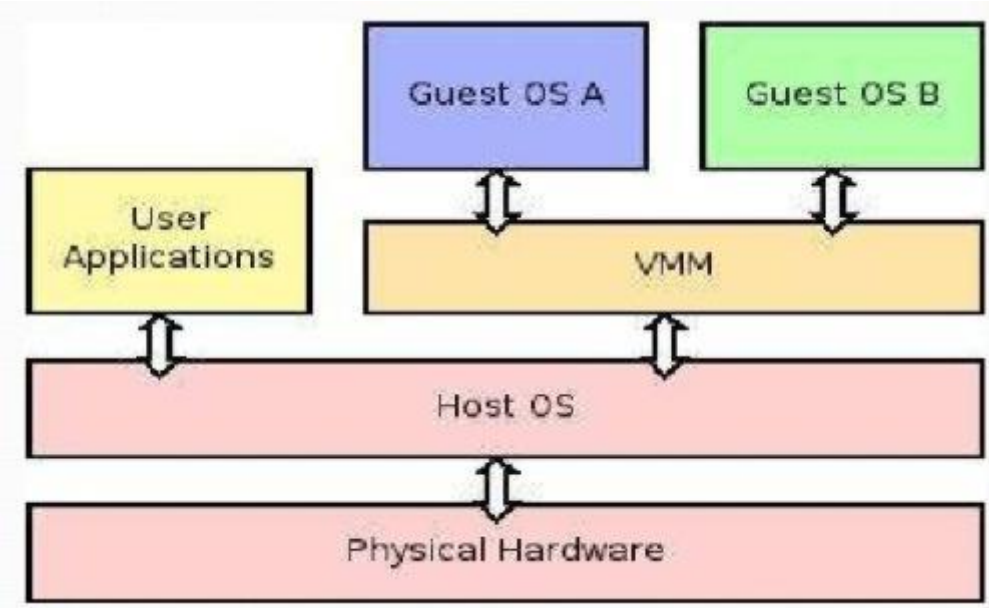
- The ability to computer to run and create **one or more virtual environments**.

- It is used to enable a **computer system** in order to allow a guest OS to run.
- Ex. Linux to run as a guest that is natively running a Microsoft Windows OS
- Subtypes:

Operating System Virtualization – Hosting multiple OS on the native

Application Virtualization – Hosting individual applications in a virtual environment separate from the native OS

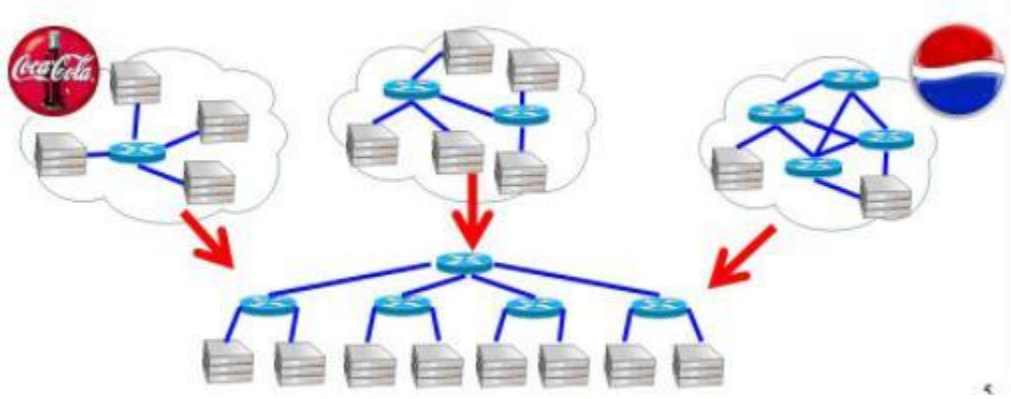
Service Virtualization – Hosting specific processes and services related to a particular application



3. Network Virtualization

- It refers to the **management and monitoring of a computer network** as a single managerial entity from a single software-based administrator's console.
- **Multiple sub-networks** can be created on the same physical network, which may or may not is authorized to communicate with each other.
- It allows **network optimization** of data transfer rates, scalability, reliability, flexibility, and security
- Subtypes:

Internal network: Enables a single system to function like a network O **External network**: Combine many networks, or parts of networks into a virtual unit.

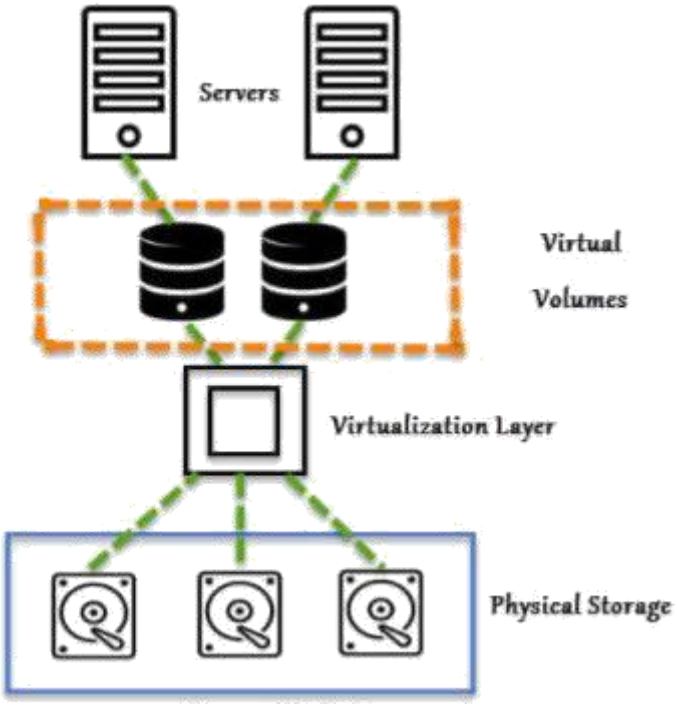


4. Storage Virtualization

- **Multiple physical storage devices** are grouped together, which look like a single storage device.
- Ex. **Partitioning our hard drive** into multiple partitions
- **Advantages**
    - Improved storage management in a heterogeneous IT environment
    - Easy updates, better availability
    - Reduced downtime
    - Better storage utilization
    - Automated management
- **Two types**

    Block- Multiple storage devices are consolidated into one

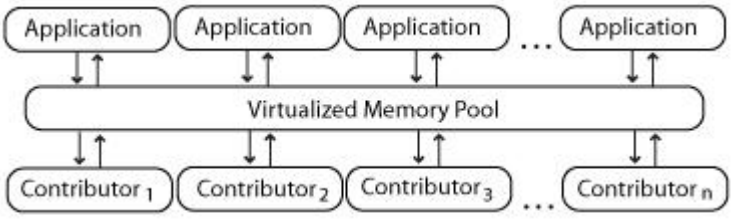
    File- Storage system grants access to files that are stored over multiple hosts



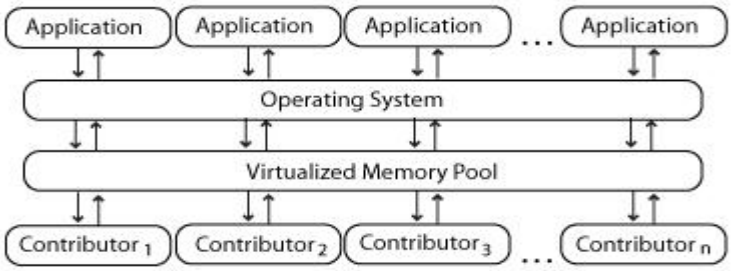
5. **Memory Virtualization**

- The way to **decouple memory from the server** to provide a shared, distributed or networked function.
- It enhances performance by providing **greater memory capacity** without any addition to the main memory.
- **Implementations**

Application-level integration – Applications access the memory pool directly



Operating System Level Integration – Access to the memory pool is provided through an operating system.



6. Data Virtualization

- Without any technical details, we can **easily manipulate data** and know how it is formatted or where it is physically located.
- It decreases the data errors and workload
- The data is presented as an abstract layer **completely independent of data structure and database systems**
- The user's desktop is stored on a remote server, allowing **the user to access his/her desktop from any device or location**.
- It provides the **work convenience and security**
- It provides a lot of flexibility for employees to **work from home or on the go**
- Since the **data transfer takes place over secure protocols**, any risk of data theft is minimized

Operating System Virtualization:

Operating system virtualization refers to the use of software to allow system hardware to run multiple instances of different operating systems concurrently, allowing we to run different applications requiring different operating systems on one computer system. The operating

systems do not interfere with each other or the various applications. Not to be confused with operating system-level virtualization, which is a type of server virtualization.

VMWare

VMware is a virtualization and cloud computing software provider based in Palo Alto, Calif. Founded in 1998, VMware is a subsidiary of Dell Technologies. EMC Corporation originally acquired VMware in 2004; EMC was later acquired by Dell Technologies in 2016. VMware bases its virtualization technologies on its bare-metal hypervisor ESX/ESXi in x86 architecture.With VMware server virtualization, a hypervisor is installed on the physical server to allow for multiple virtual machines (VMs) to run on the same physical server. Each VM can run its own operating system (OS), which means multiple OSes can run on one physical server. All the VMs on the same physical server share resources, such as networking and RAM. In 2019, VMware added support to its hypervisor to run containerized workloads in a Kubernetes cluster in a similar way. These types of workloads can be managed by the infrastructure team in the same way as virtual machines and the DevOps teams can deploy containers as they oure used to.

Diane Greene, Scott Devine, Mendel Rosenblum, Edward Wang and Edouard Bugnion founded VMware, which launched its first product -- VMware Workstation -- in 1999. The company released its second product, VMware ESX in 2001.VMware products include virtualization, networking and security management tools, software-defined data center software and storage software.VMware vSphere is VMware's suite of virtualization products. VMware vSphere, known as VMware Infrastructure prior to 2009, includes the following:

- ESXi
- vCenter Server
- vSphere Client
- vMotion

As of April 2018, the most current version is vSphere 6.7, which is available in three editions: Standard, Enterprise Plus and Platinum. There are also two three-server kits targeted toward small and medium-sized businesses named vSphere Essentials and Essentials Plus.With **VMware Cloud on AWS**, customers can run a cluster of vSphere hosts with vSAN and NSX in an Amazon data center and run

their workloads there while in the meantime manage them with their well-known VMware tools and skills.

Networking and security

VMware NSX is a virtual networking and security software offering created when VMware acquired Nicera in 2012. NSX allows an admin to virtualize network components, enabling them to develop, deploy and configure virtual networks and switches through software rather than hardware. A software layer sits on top of the hypervisor to allow an administrator to divide a physical network into multiple virtual networks.With the latest release of the product, NSX-T Data Center, network virtualization can be added to both ESXi and KVM as hypervisors, as well as to bare-metal servers. Also containerized workloads in a Kubernetes cluster can be virtualized and protected. NSX-T Data Center also offers Network Function Virtualization, with which functions such as a firewall, load balancer and VPN, can be run in the virtualization software stack.

VMware vRealize Network Insight is a network operations management tool that enables an admin to plan microsegmentation and check on the health of VMware NSX. VRealize Network Insight relies on technology from VMware's acquisition of Arkin in 2016. VRealize Network Insight collects information from the NSX Manager. It also displays errors in its user interface, which helps troubleshoot an NSX environment.

Software Defined Data Center (SDDC) platform:

VMware Cloud Foundation is an integrated software stack that bundles vSphere, VMware vSAN and VMware NSX into a single platform through the SDDC Manager. An admin can deploy the bundle on premises as a private cloud or run it as a service within a public cloud. An administrator can provision an application immediately without having to wait for network or storage.

Storage and availability

VMware vSAN is a software-based storage feature that is built into the ESXi hypervisor and integrated with vSphere; it pools disk space from multiple ESXi hosts and provisions it via smart policies, such

as protection limits, thin provisioning and erasure coding. It integrates with vSphere High Availability to offer increased compute and storage availability.

VMware Site Recovery Manager (SRM) is a disaster recovery management product that allows an administrator to create recovery plans that are automatically executed in case of a failure. Site Recovery Manager allows admins to automatically orchestrate the failover and failback of VMs. SRM also integrates with NSX to preserve network and security policies on migrated VMs.

VMware vCloud NFV is a network functions virtualization platform that enables a service provider to run network functions as virtualized applications from different vendors. NFV provides the same benefits of virtualization and cloud to a communications service provider that previously relied on hardware.

Cloud management platform

The **vRealize Suite** is a group of software that allows a user to create and manage hybrid clouds. The vRealize Suite includes vRealize Operations for monitoring, vRealize Log Insight for centralized logging, vRealize Automation for data center automation and vRealize Business for Cloud for cost management.

With this bundle, an administrator can deploy and manage VMs on multiple hypervisors or cloud platforms from a single management console. Released in 2019, VMware Tanzu allows customers to build containerized apps, run enterprise Kubernetes and manage Kubernetes for developers and IT.

Virtual desktop infrastructure

VMware Horizon allows organizations to run Windows desktops in the data center or in VMware Cloud on AWS. This removes the need to place and manage full desktops on the workplace and centralizes management and security for the user's environment. It integrates with the VMware products App Volumes and Dynamic Environment Manager for application delivery and Windows desktop management.

Digital workspace and enterprise mobility management

Workspace ONE allows an administrator to control mobile devices and cloud-hosted virtual desktops and applications from a single management platform deployed either in the cloud or on premises. The Workspace ONE suite includes VMware AirWatch, Horizon Air and Identity Manager. Identity Manager is an identity-as-a-service product that offers single sign-on (SSO) capabilities for web, cloud and mobile applications. Identity Manager gives SSO access to any application from any device, based on the policies created.VMware AirWatch is an enterprise mobility management (EMM) software platform that enables an administrator to deploy and manage mobile devices, applications and data.

Personal desktop

VMware Workstation is the first product ever released by the software company. It enables users to create and run VMs directly on a single Windows or Linux desktop or laptop. Those VMs run simultaneously with the physical machine. Each VM runs its own OS such as Windows or Linux. This enables users to run Windows on a Linux machine or vice versa simultaneously with the natively installed OS.VMware Fusion is software like VMware Workstation that virtualizes a Windows or Linux OS on Mac computers.

Benefits of VMware

- Security based on a zero-trust model, along with better security than container systems like Kubernetes;
- Better provisioning of applications and resources;
- Simplified Data Center Management
- Increased efficiency and agility of data center systems.

Drawbacks of VMware

- High licensing fees;
- Better Hyper-V and Xen hypervisor alternatives, according to some;
- Lack of support and several bugs when used alongside oracle products; and

- Hardware compatibility issues as not everything works well with VMware.

KVM

Kernel-based Virtual Machine (KVM) is an open source virtualization technology built into Linux®. Specifically, KVM lets we turn Linux into a hypervisor that allows a host machine to run multiple, isolated virtual environments called guests or virtual machines (VMs).

KVM is part of Linux. If we've got Linux 2.6.20 or neour, we've got KVM. KVM was first announced in 2006 and merged into the mainline Linux kernel version a year later. Because KVM is part of existing Linux code, it immediately benefits from every new Linux feature, fix, and advancement without additional engineering.

How does KVM work?

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. 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

We need to 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 we 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 we run additional systems.

But implementing KVM on a supported Linux distribution—like Red Hat Enterprise Linux—expands KVM's capabilities, letting we swap resources among guests, share common libraries, optimize system performance, and a lot more.

Migrating to a KVM-based virtual infrastructure

Building a virtual infrastructure on a platform we're contractually tied to may limit our access to the source code. That means our IT developments are probably going to be more workarounds than innovations, and the next contract could keep we from investing in clouds, containers, and automation. Migrating to a KVM-based virtualization platform means being able to inspect, modify, and enhance the source code behind our hypervisor. And there's no enterprise-license agreement because there's no source code to protect.

KVM features

KVM is part of Linux. Linux is part of KVM. Everything Linux has, KVM has too. But there are specific features that make KVM an enterprise's preferred hypervisor.

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 pooured 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 datacenters and private clouds

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.

Loour Latency and higher prioritization

The Linux kernel features real-time extensions that allow VM-based apps to run at loour 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.

Managing KVM

It’s possible to manually manage a handful of VM fired up on a single workstation without a management tool. Large enterprises use virtualization management software that interfaces with virtual environments and the underlying physical hardware to simplify resource administration, enhance data analyses, and streamline operations. Red Hat created Red Hat Virtualization for exactly this purpose.

KVM and Red Hat

We believe in KVM so much that it’s the sole hypervisor for all of our virtualization products, and we’re continually improving the kernel code with contributions to the KVM community. But since

KVM is part of Linux, it's already included in Red Hat Enterprise Linux—so why would we want Red Hat Virtualization?

Well, Red Hat has 2 versions of KVM. The KVM that ships with Red Hat Enterprise Linux has all of the hypervisor functionality with basic management capabilities, allowing customers to run up to 4 isolated virtual machines on a single host. Red Hat Virtualization contains an advanced version of KVM that enables enterprise management of unlimited guest machines. It's ideal for use in datacenter virtualization, technical workstations, private clouds, and in development or production.

System VM and Process VM

Two categories of virtual machines

Virtual machines are separated in two major categories, based on their use and degree of correspondence to any real machine. A system virtual machine provides a complete system platform which supports the execution of a complete operating system (OS). In contrast, a process virtual machine is designed to run a single program, which means that it supports a single process. An essential characteristic of a virtual machine is that the software running inside is limited to the resources and abstractions provided by the virtual machine — it cannot break out of its virtual world.

System Virtual Machines

System virtual machines (sometimes called hardware virtual machines) allow the sharing of the underlying physical machine resources between different virtual machines, each running its own operating system. The software layer providing the virtualization is called a virtual machine monitor or hypervisor. A hypervisor can run on bare hardware (Type 1 or native VM) or on top of an operating system (Type 2 or hosted VM).

Main advantages of system VMs

- Multiple OS environments can co-exist on the same computer, in strong isolation from each other;

- The virtual machine can provide an instruction set architecture (ISA) that is somewhat different from that of the real machine.

Main disadvantages of system VMs

- There's still an overhead of the virtualization solution which is used to run and manage a VM, so performance of a VM will be somewhat sloour compared to a physical system with comparable configuration
- Virtualization means decoupling from physical hardware available to the host PC, this usually means access to devices needs to go through the virtualization solution and this may not always be possible

Multiple VMs each running their own operating system (called guest operating system) are frequently used in server consolidation, where different services that used to run on individual machines in order to avoid interference are instead run in separate VMs on the same physical machine. This use is frequently called quality-of-service isolation (QoS isolation).

Process Virtual Machines

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

A process VM provides a high-level abstraction — that of a high-level programming language (compared to the low-level ISA abstraction of the system VM). Process VMs are implemented using an interpreter; performance comparable to compiled programming languages is achieved by the use of just-in-time compilation. This type of VM has become popular with the Java programming language, which is implemented using the Java virtual machine. Another example is the .NET Framework, which runs on a VM called the Common Language Runtime.

Virtual Machine Monitor

A Virtual Machine Monitor (VMM) is a software program that enables the creation, management and governance of virtual machines (VM) and manages the operation of a virtualized environment on top of a physical host machine. VMM is also known as Virtual Machine Manager and Hypervisor.

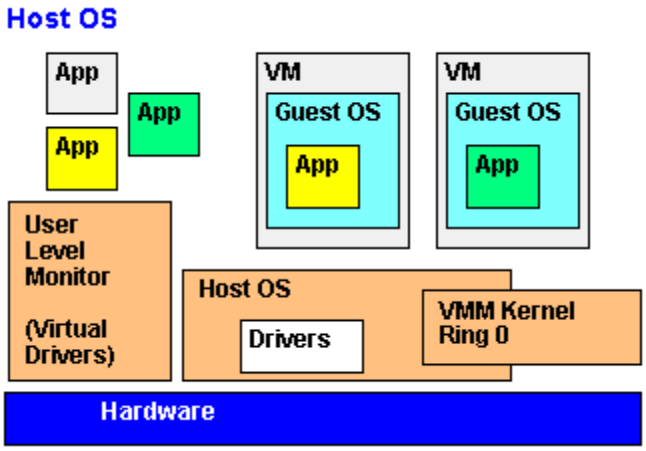
VMM is the primary software behind virtualization environments and implementations. When installed over a host machine, VMM facilitates the creation of VMs, each with separate operating systems (OS) and applications. VMM manages the backend operation of these VMs by allocating the necessary computing, memory, storage and other input/output (I/O) resources.

VMM also provides a centralized interface for managing the entire operation, status and availability of VMs that are installed over a single host or spread across different and interconnected hosts.

The software that creates a virtual machine (VM) environment in a computer In a regular, non-virtual computer, the operating system is the master control program, which manages the execution of all applications and acts as an interface between the apps and the hardware. The OS has the highest privilege level in the machine, known as "ring 0"

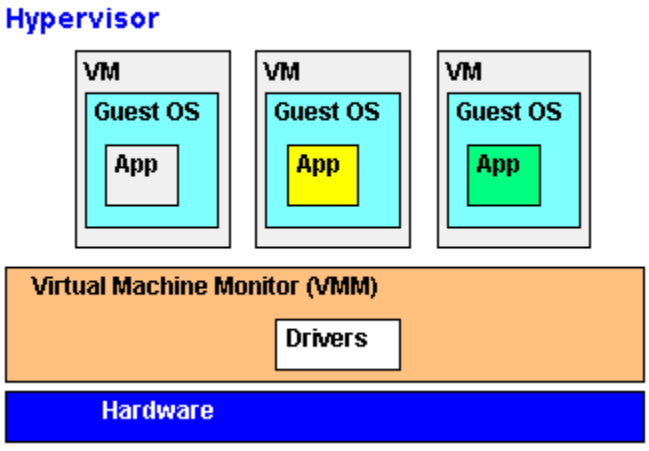
In a VM environment, the VM monitor (VMM) becomes the master control program with the highest privilege level, and the VMM manages one or more "guest operating systems." Each guest OS manages its own applications in a separate "virtual machine" (VM) in the computer, sometimes called a "guest OS stack."

The VM monitor (VMM) is an interface between the guest OS and the hardware. It intercepts calls to the peripheral devices and memory tables from each guest OS and intercedes on its behalf. In reverse, when a disk or SSD write creates an interrupt, the VM monitor injects that interrupt into the appropriate guest OS. Following are the major monitor types.



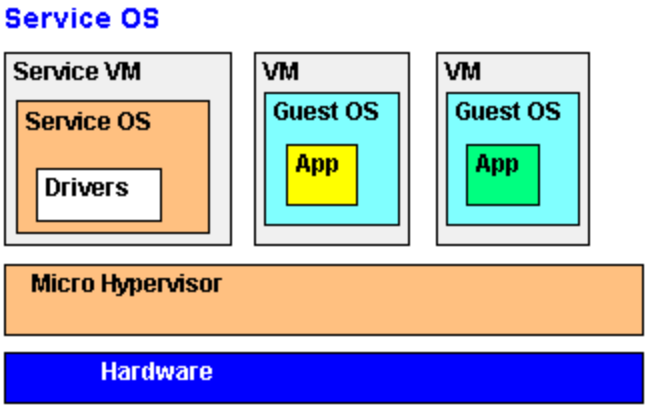
Host OS

This VM monitor (VMM) is installed in an existing, running computer. The VMM kernel runs alongside the host OS, and calls for I/O are redirected to virtual drivers that call the native API of the host OS. Examples of OS-hosted VMMs are VMware Workstation, VMware Server, Parallels Workstation and Parallels Desktop for Mac.



Hypervisor

The hypervisor monitor provides the most control, flexibility and performance, because it is not subject to limitations of a host OS. The hypervisor relies on its own software drivers for the hardware; however, they may limit portability to another platform. Examples of this method are VMware ESX and IBM's mainframe z/VM.



Service OS

This method combines the robustness of the hypervisor with the flexibility of the host model. In order to take advantage of the drivers in a popular OS, the Service OS runs as a component of the hypervisor in a separate VM. Xen, XenServer and Hyper-V are examples of the service VM approach. The VMM is in charge of running the virtual machines.

There are two main types of VMM:

Type 1: Native
Type 2: Hosted

Type 1: Native Hypervisors run directly on the host machine, and share out resources (such as memory and devices) between guest machines.

e.g. XEN, Oracle VM Server

Type 2: Hosted Hypervisors run as an application inside an operating system, and support virtual machines running as individual processes.

e.g. VirtualBox, Parallels Desktop, QEMU

Properties of a Virtual Machine

1.Efficiency: The majority of guest instructions are executed directly on the host machine.

2.Resource Control: The virtual machine monitor must remain in control of all machine resources.

3.Equivalence: The virtual machine must behave in a way that is indistinguishable from if it was running as a physical machine.

Efficiency

"All innocuous instructions are executed by the hardware directly, with no intervention at all on the part of the control program."

Normal guest machine instructions should be executed directly on the processor. System instructions need to be emulated by the VMM.

Resource Control

"It must be impossible for that arbitrary program to affect the system resources, i.e. memory, available to it; the allocator of the control program is to be invoked upon any attempt." The virtual machine should not be able to affect the host machine in any adverse way. The host machine should remain in control of all physical resources, sharing them out to guest machines.

Equivalence

"Any program K executing with a control program resident, with two possible exceptions, performs in a manner indistinguishable from the case when the control program did not exist and K had whatever freedom of access to privileged instructions that the programmer had intended." A formal way of saying that the operating system running on a virtual machine should believe it is running on a physical machine, i.e. the behaviour of the virtual machine (from the guest OS' point of view) is identical to that of the corresponding physical machine.

The two exceptions mentioned are: temporal latency (some instruction sequences will take longer to run) and resource availability (physical machine resources are shared between virtual machines)

Interpretation and Binary Translation:

•Interpretation

– simple and easy to implement, portable

– low performance

– threaded interpretation

•Binary translation

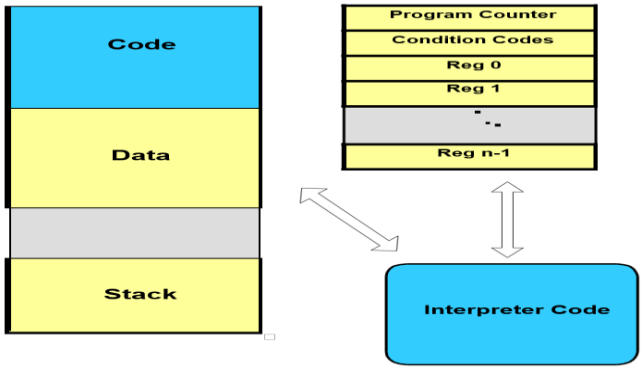
– complex implementation

– high initial translation cost, small execution cost

– selective compilation

Interpreter state:

• An interpreter needs to maintain the complete, architected state of the machine implementing the source ISA registers& memory(code, data, stack)



Decode and dispatch interpreter

– step through the source program one instruction at a time

– decode the current instruction

– dispatch to corresponding interpreter routine

– vry high interpretation cost

```
while (!halt && !interrupt) {
inst = code[PC];
opcode = extract(inst,31,6);
switch(opcode) {
case LoadWordAndZero: LoadWordAndZero(inst);
case ALU: ALU(inst);
case Branch: Branch(inst);
...}
}
```

Load Function:

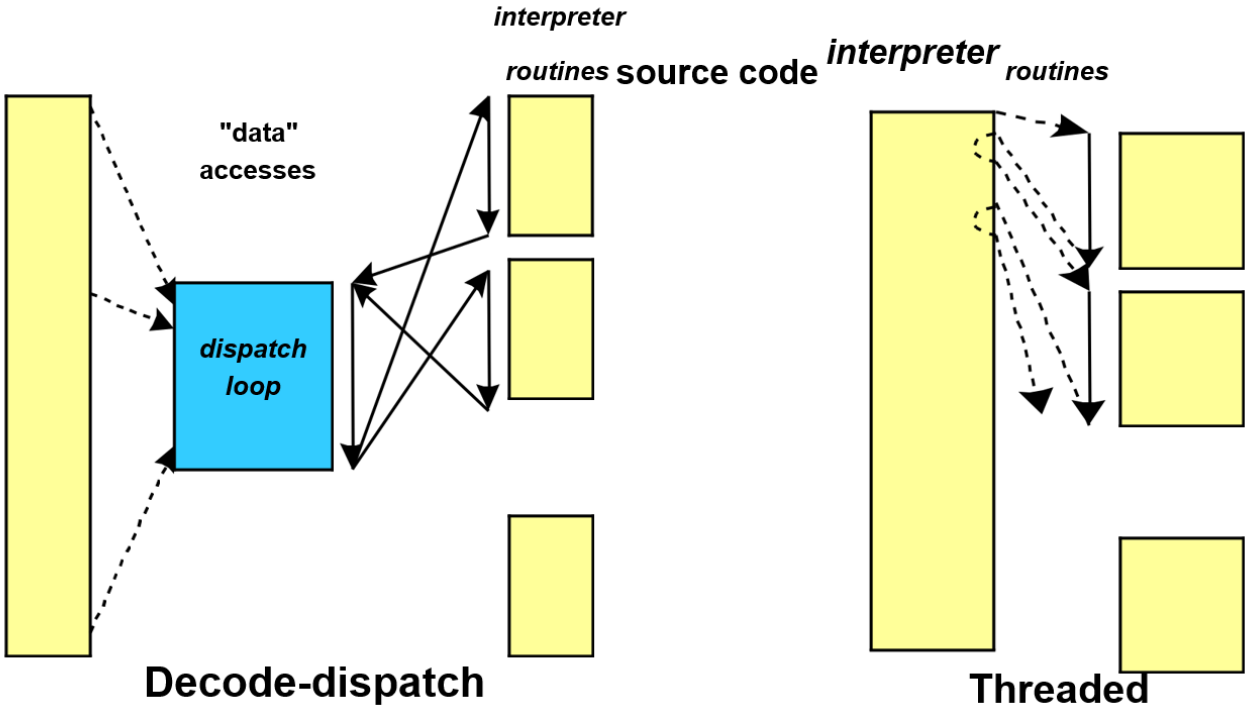
```
LoadWordAndZero(inst){
RT = extract(inst,25,5); RA = extract(inst,20,5); displacement =
extract(inst,15,16); if (RA == 0) source = 0; else source = regs[RA];


address = source + displacement; regs[RT] = (data[address]<< 32)>>
32; PC=PC+4;
```

```
}
```

ALU Function:

```
ALU(inst){

RT = extract(inst,25,5);

RA = extract(inst,20,5);

RB = extract(inst, 15,5);

source1 = regs[RA];

source2 = regs[RB];

extended_opcode = extract(inst,10,10); switch(extended_opcode) {

case Add: Add(inst);

case AddCarrying: AddCarrying(inst);

case AddExtended: AddExtended(inst);

.    . .} PC=PC+4;

}
```

Indirect Thread Interpretation:



Pre Coding:

- Parse each instruction into a pre-defined structure to facilitate interpretation

– separate opcode, operands, etc.

– reduces shifts / masks significantly

– more useful for CICS ISAs

```
lwz     r1, 8(r2)
add     r3, r3,r1
stw     r3, 0(r4)
```

07		
1	2	08

(load word and zero)

08		
3	1	03

(add)

37		
3	4	00

(store word)

```
struct instruction {
```

```
unsigned long op;

unsigned char dest, src1, src2; } code [CODE_SIZE];
```

Load Word and Zero:

```
RT = code[TPC].dest;

RA = code[TPC].src1;

displacement = code[TPC].src2;

if (RA == 0) source = 0;

else source = regs[RA];

address = source + displacement;

regs[RT] = (data[address]<< 32) >> 32;

SPC = SPC + 4; TPC = TPC + 1;

If (halt || interrupt) goto exit; opcode = code[TPC].op


routine = dispatch[opcode]; goto *routine;
```

Direct Threaded Interpretation

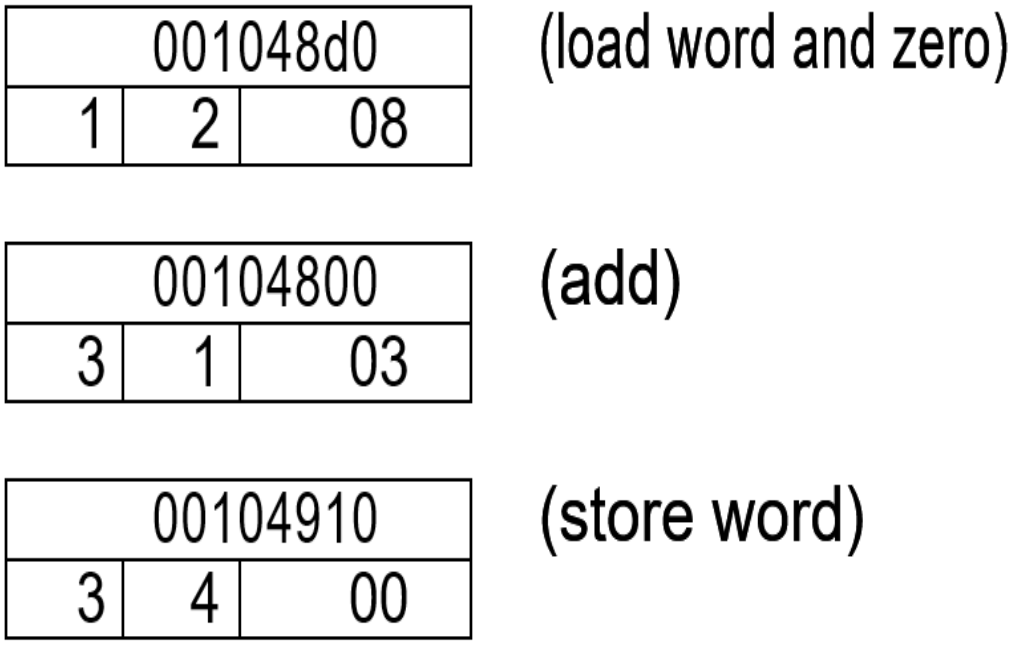
•Allow even higher efficiency by:

– Removing the memory access to the centralized table

– Requires predecoding

– Dependent on locations of interpreter routines

•Loses Portability



Load Word and Zero:

```
RT = code[TPC].dest;

RA = code[TPC].src1;

displacement = code[TPC].src2;

if (RA == 0) source = 0;

else source = regs[RA];

address = source + displacement;

regs[RT] = (data[address]<< 32) >> 32;

SPC = SPC + 4;

TPC = TPC + 1;

If (halt || interrupt) goto exit; routine = code[TPC].op; goto
*routine;
```

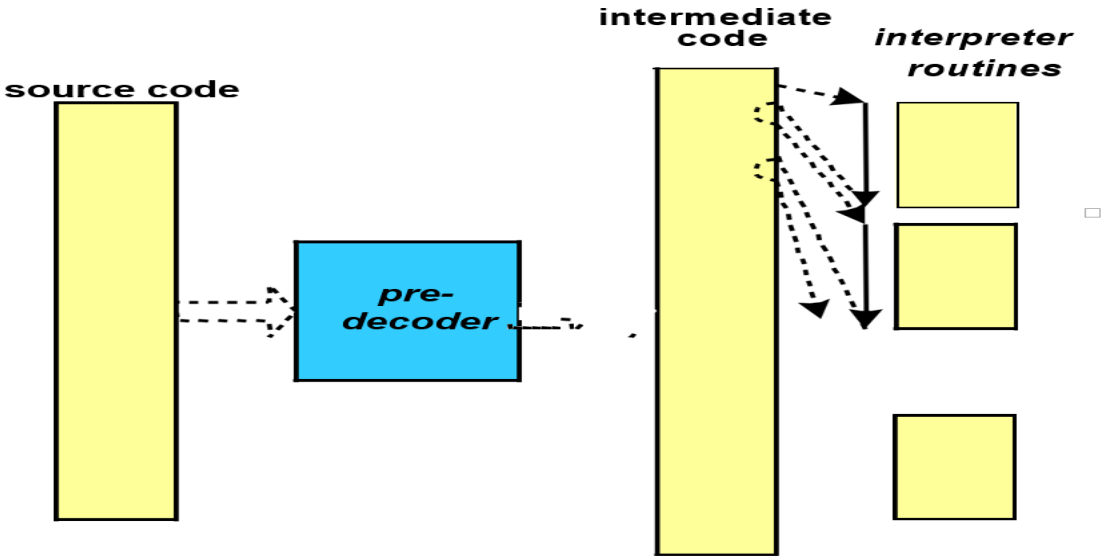
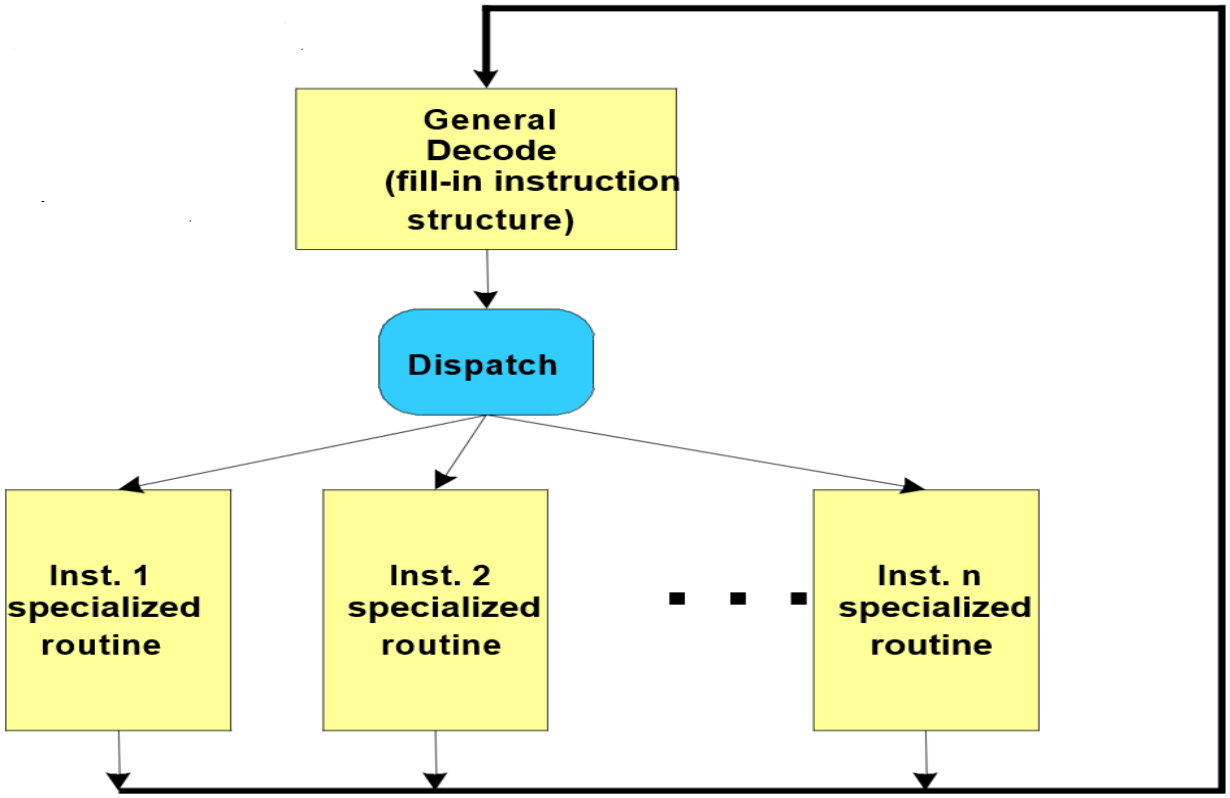


Fig. Direct Threaded interpretation

Interpretor control flow:

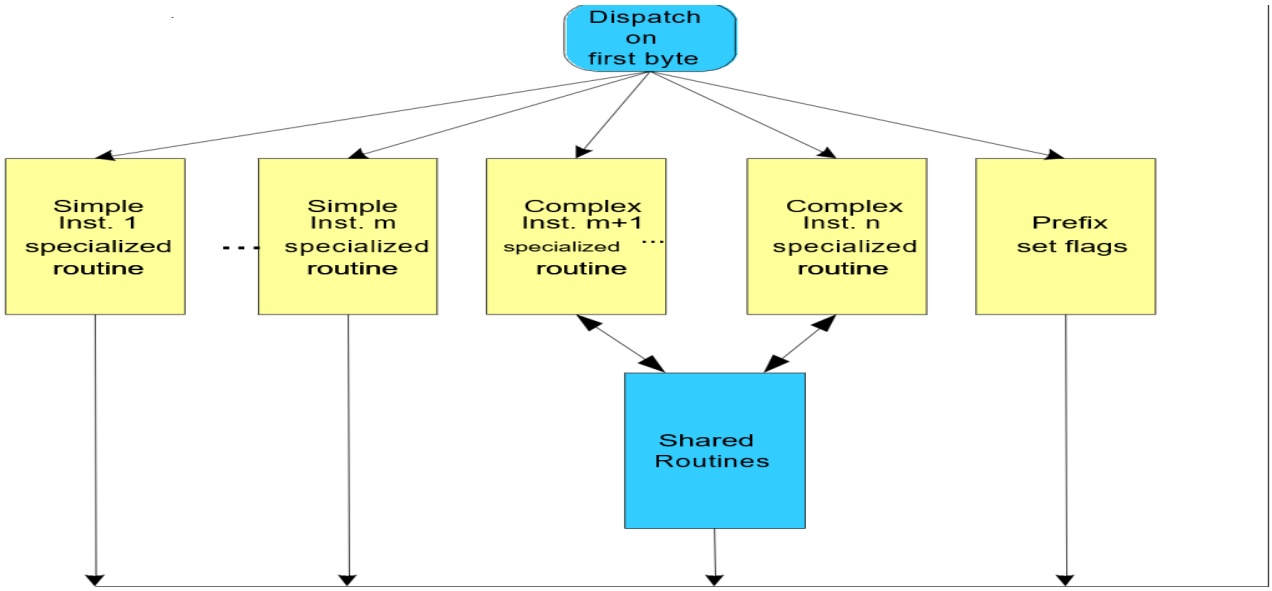
- Decode for CISC ISA
- Individual routines for each instruction



For CISC ISAs

– multiple byte opcode

– make common cases fast



Binary Translation

•Translate source binary program to target binary before execution

– is the logical conclusion of predecoding

– get rid of parsing and jumps altogether

– allows optimizations on the native code

– achieves higher performance than interpretation

– needs mapping of source state onto the host state (state mapping)

x86 Source Binary

```
addl %edx,4(%eax)

movl 4(%eax),%edx

add  %eax,4
```

Translate to PoourPC Target

```
r1 points to x86 register context block

r2 points to x86 memory image

r3 contains x86 ISA PC value

lwz  r4,0(r1)  ;load %eax from register block

addi r5,r4,4   ;add 4 to %eax

lwzx r5,r2,r5  ;load operand from memory

lwz  r4,12(r1) ;load %edx from register block

add  r5,r4,r5  ;perform add

stw  r5,12(r1) ;put result into %edx

addi r3,r3,3   ;update PC (3 bytes)
```

```
lwz  r4,0(r1)  ;load %eax from register block

addi r5,r4,4   ;add 4 to %eax

lwz  r4,12(r1) ;load %edx from register block

stwx r4,r2,r5  ;store %edx value into memory

addi r3,r3,3   ;update PC (3 bytes)

lwz  r4,0(r1)  ;load %eax from register block

addi r4,r4,4   ;add immediate

stw  r4,0(r1)  ;place result back into %eax

addi r3,r3,3   ;update PC (3 bytes)
```

State Mapping

•Maintaining the state of the source machine on the host (target) machine.

– state includes source registers and memory contents

– source registers can be held in host registers or in host memory

– reduces loads/stores significantly

– easier if target registers > source registers

Register Mapping

Map source registers to – spill registers if needed

•if target registers < source registers

– map some to memory

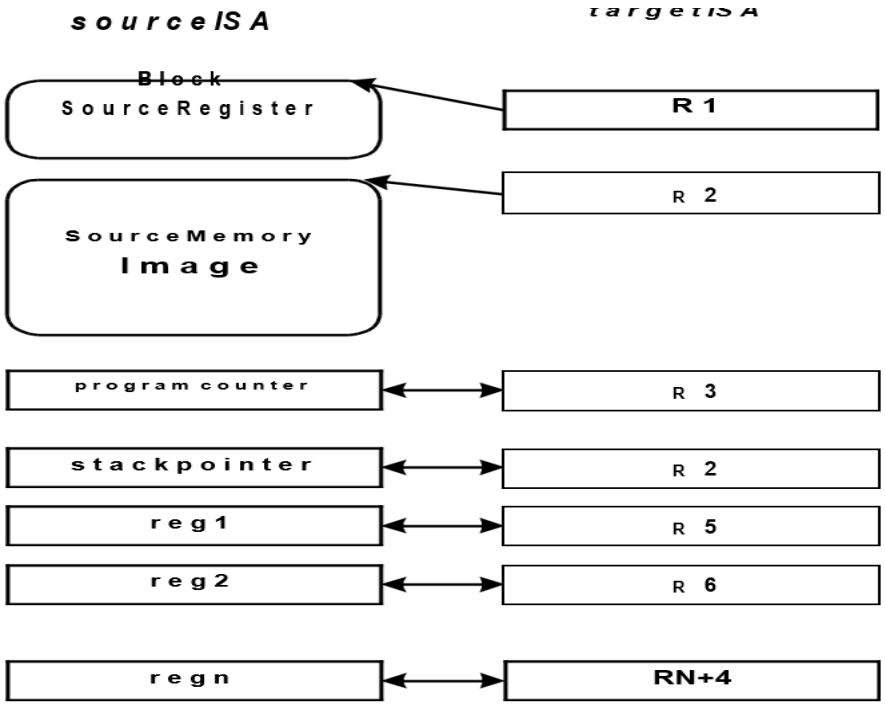
– map on per-block basis

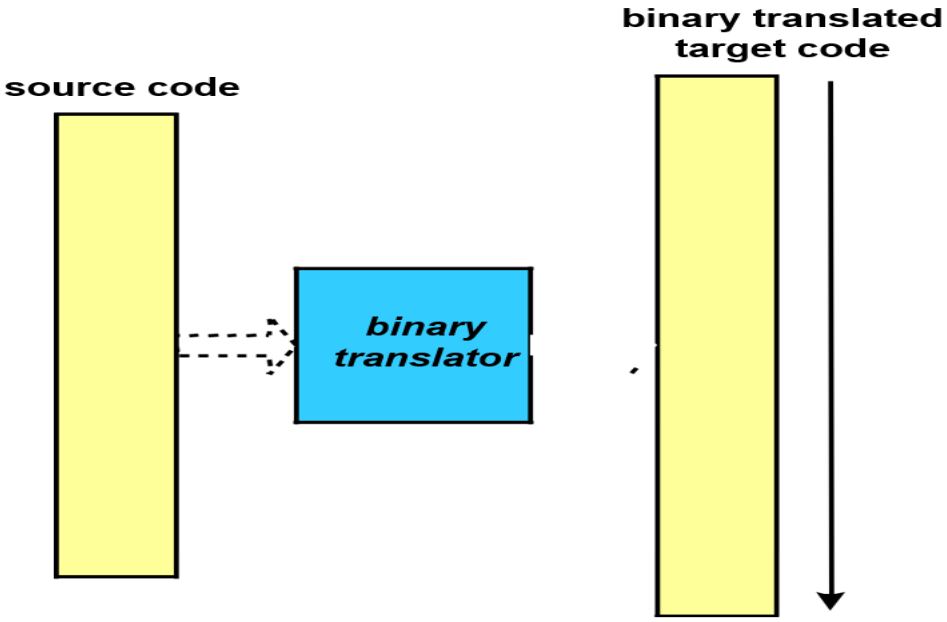
•Reduces load/store significantly

– improves performance

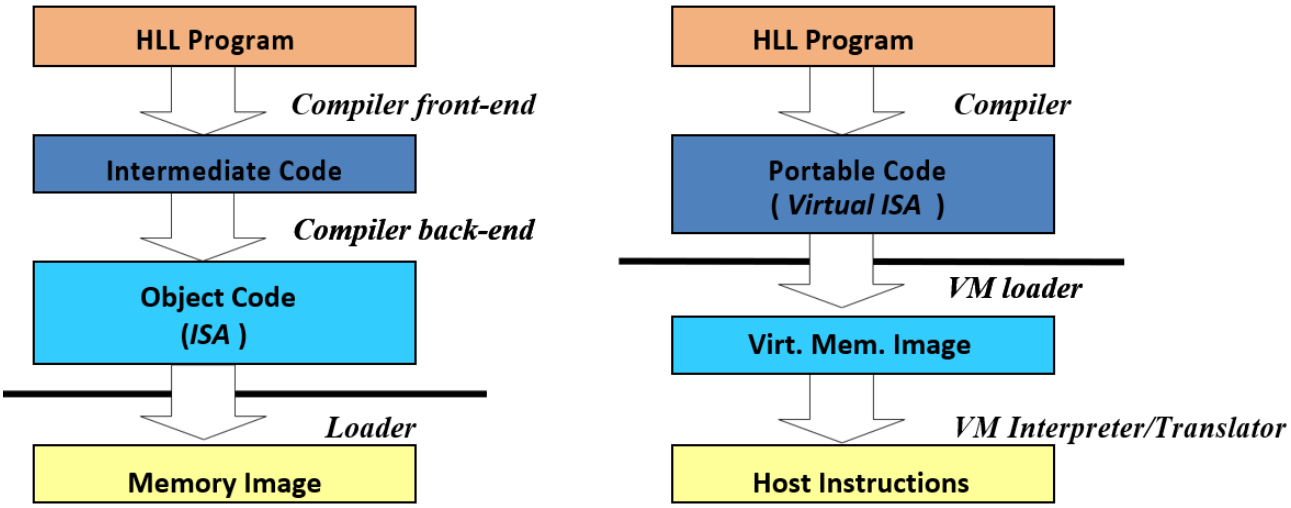
```
r1 points to x86 register context block
```

```
r2 points to x86 memory image

r3 contains x86 ISA PC value

r4 holds x86 register %eax

r7 holds x86 register %edx etc.

addi r16,r4,4  ;add 4 to %eax

lwzx r17,r2,r16     ;load operand from memory

add  r7,r17,r7 ;perform add of %edx

addi r16,r4,4  ;add 4 to %eax

stwx r7,r2,r16 ;store %edx value into memory

addi r4,r4,4   ;increment %eax

addi r3,r3,9   ;update PC (9 bytes)
```





High Level Language Virtual Machines(HLL VM)



Traditional **HLL VM**

Two major examples–

Java VM

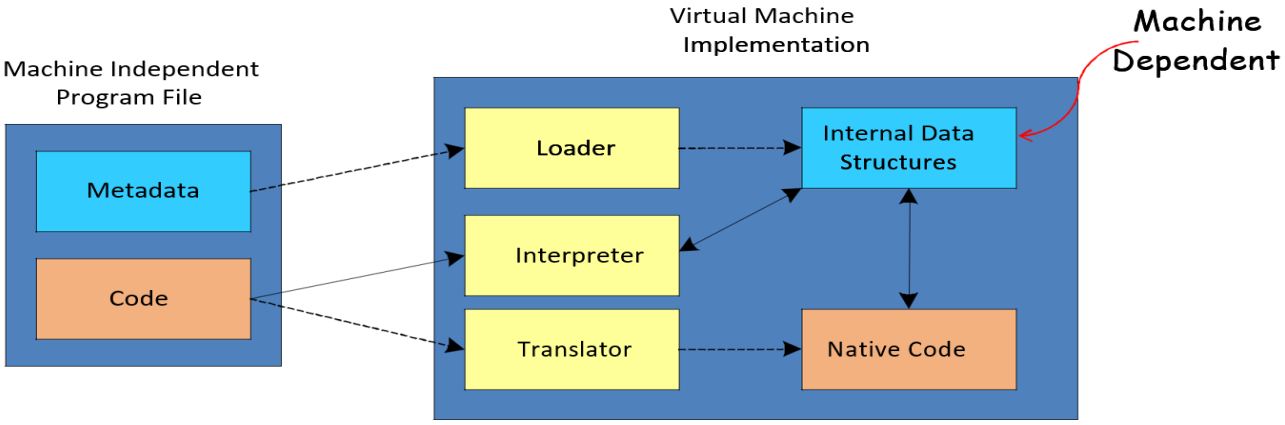
– Microsoft Common Language Infrastructure (CLI)

HLL VMS:

Compiler forms program files (e.g. class files)

–Standard format

Program files contain both code and metadata–



Java Virtual Machine Architecture & CLI	– Analogous to an ISA
Java Virtual Machine Implementation & CLR (Common Language Runtime)	– Analogous to a computer implementation
Java bytecodes & Microsoft Intermediate Language (MSIL), CIL, IL	-The instruction part of the ISA
Java Platform & .NET framework	– ISA + Libraries; a higher level ABI

Characteristics of HLL VMs

•Security

•Robustness

•Networking

•Performance

Security

•A key aspect of modern network-oriented VMs

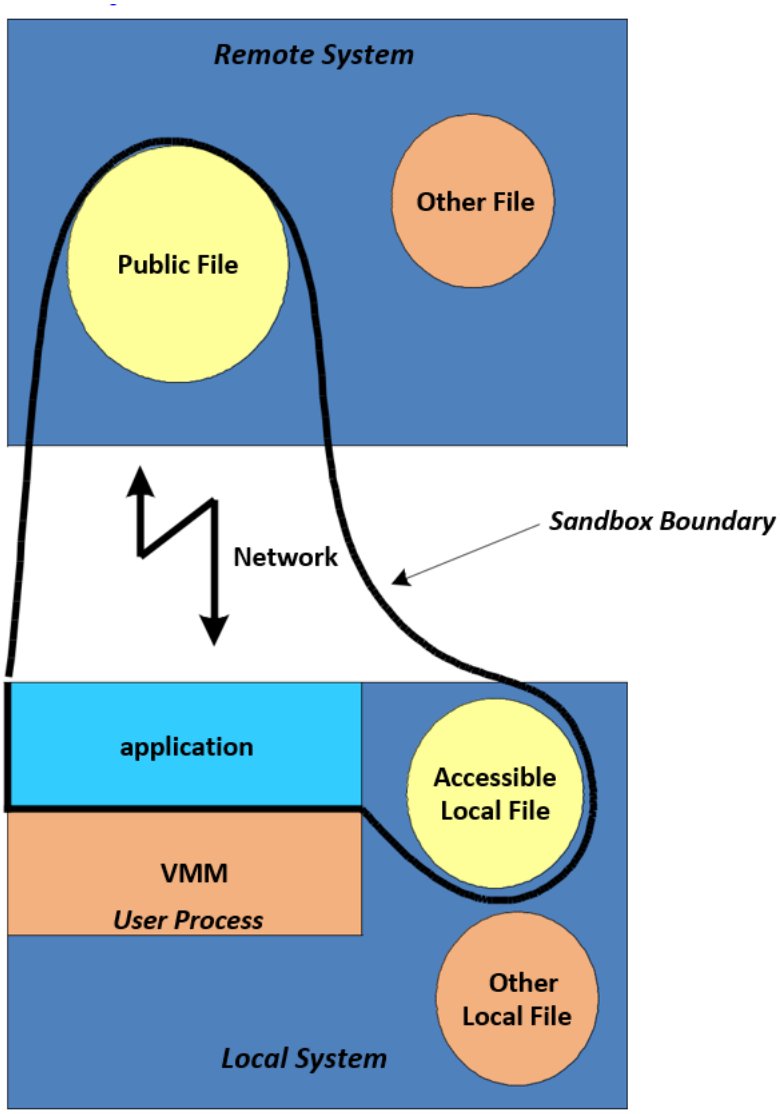
•Must protect:

– Local files and resources

– Runtime from user process

•The program runs in a sandbox at the host machine. It is managed by the VM runtime.

•The ability to load an untrusted application and run it in a managed secure fashion is a very big challenge!



Robustness:Object Orientation

•Objects

– Data carrying entities

– Dynamically allocated

– Must be accessed via pointers or references

•**Methods**

– Procedures that operate on objects

•**Class**

– A type of object and its associated methods

– Object created at runtime is an instance of the class

– Data associated with a class may be dynamic or static

OO programming paradigm has become the model of choice for modern HLL VMs.

Both Java and CLI are designed to support OO software.

Networking:

•The application must use the available bandwidth (scarce) efficiently

– Application loaded incrementally dynamic linking

– Improves program startup-time

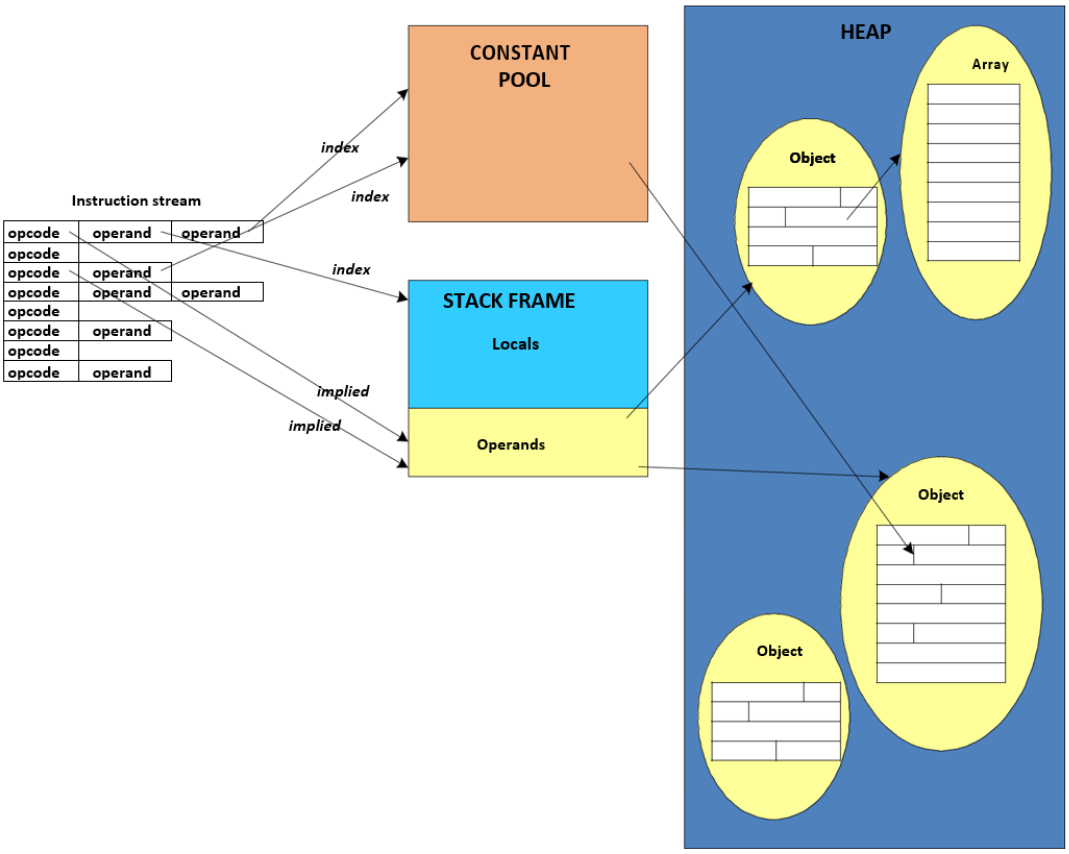


Fig: Memory Hierarchy in JVM

JVM: Bytcode Emulation

•Interpretation

– Simple, fast startup, but slow

•Just-In-Time (JIT) Compilation

– Compile each method when first touched

– Simple, static optimizations

•Hot-Spot Compilation

– Find frequently executed code

– Apply more aggressive optimizations on that code

– Typically phased with interpretation or JIT

•Dynamic Compilation

– Based on Hot-Spot compilation

– Use runtime information to optimize

References:

1. Cloud Data Center: https://www.emoneyindeed.com/traditional-data-center-vs-cloud-data-center/
2. Energy Efficiency: https://www.geeksforgeeks.org/energy-efficiency-in-cloud-computing/
3. Hitesh A. Bheda Jignesh Lakhani, Application Processing Approach for Smart Mobile Devices in Mobile Cloud Computing, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 8, August 2013, pp.1046-1054
4. VMWare : https://www.vmware.com/in/solutions/virtualization.html
5. VMWare:https://searchservervirtualization.techtarget.com/definition/server-virtualization
6. VMWare:https://searchvmware.techtarget.com/definition/VMware
7. Kvm: https://www.redhat.com/en/topics/virtualization/what-is-KVM
8. System VM and process VM https://www.desktop-virtualization.com/glossary/virtual-machine/
9. VMM: https://www.pcmag.com/encyclopedia/term/virtual-machine-monitor
10. Interpretation: http://www.ittc.ku.edu/~kulkarni/teaching/EECS768/slides/chapter2.pdf
11. HLLVM: https://cs.nyu.edu/courses/spring14/CSCI-GA.3033-015/lecture6.pdf

SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

SCHOOL OF COMPUTING

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

UNIT – V – Cloud computing – SCSA7023

Cloud Security

Topics:

Cloud security – Security threats and solutions in clouds – Auditing protocols – dynamic auditing – storage security –Privacy preserving — Fully Homomorphic Encryption – Big data security- Cloud availability- DoS attacks – Fault tolerance management in cloud computing- Cloud computing in India.

Cloud Security

Cloud security, also known as cloud computing security, consists of a set of policies, controls, procedures and technologies that work together to protect cloud-based systems, data, and infrastructure. These security measures are configured to protect cloud data, support regulatory compliance and protect customers' privacy as well as setting authentication rules for individual users and devices. From authenticating access to filtering traffic, cloud security can be configured to the exact needs of the business. And because these rules can be configured and managed in one place, administration overheads are reduced and IT teams empooured to focus on other areas of the business.

The way cloud security is delivered will depend on the individual cloud provider or the cloud security solutions in place. However, implementation of cloud security processes should be a joint responsibility between the business owner and solution provider.

For businesses making the transition to the cloud, robust cloud security is imperative. Security threats are constantly evolving and becoming more sophisticated, and cloud computing is no less at risk than an on-premise environment. For this reason, it is essential to work with a cloud provider that offers best-in-class security that has been customized for our infrastructure.

Cloud security offers many benefits, including:

Centralized security: Just as cloud computing centralizes applications and data, cloud security centralizes protection. Cloud-based business networks consist of numerous devices and endpoints that can be difficult to manage when dealing with shadow IT or BYOD. Managing these entities centrally enhances traffic analysis and web filtering, streamlines the monitoring of network events and results in feour software and policy updates. Disaster recovery plans can also be implemented and actioned easily when they are managed in one place.

Reduced costs: One of the benefits of utilizing cloud storage and security is that it eliminates the need to invest in dedicated hardware. Not only does this reduce capital expenditure, but it also reduces administrative overheads. Where once IT teams oure firefighting security issues reactively, cloud security delivers proactive security features that offer protection 24/7 with little or no human intervention.

Reduced Administration: When we choose a reputable cloud services provider or cloud security platform, we can kiss goodbye to manual security configurations and almost constant security updates. These tasks can have a massive drain on resources, but when we move them to the cloud, all security administration happens in one place and is fully managed on our behalf.

Reliability: Cloud computing services offer the ultimate in dependability. With the right cloud security measures in place, users can safely access data and applications within the cloud no matter where they are or what device they are using.

More and more organizations are realizing the many business benefits of moving their systems to the cloud. Cloud computing allows organizations to operate at scale, reduce technology costs and use agile systems that give them the competitive edge. However, it is essential that organizations have complete confidence in their cloud computing security and that all data, systems and applications are protected from data theft, leakage, corruption and deletion.

All cloud models are susceptible to threats. IT departments are naturally cautious about moving mission-critical systems to the cloud and it is essential the right security provisions are in place, whether we are running a native cloud, hybrid or on-premise environment. Cloud security offers all the functionality of traditional IT security, and allows businesses to harness the many advantages of cloud computing while remaining secure and also ensure that data privacy and compliance requirements are met.

Secure Data in the Cloud

Cloud data security becomes increasingly important as we move our devices, data centers, business processes, and more to the cloud. Ensuring quality cloud data security is acheived through comprehensive security policies, an orgnaizational culture of security, and cloud security solutions.

Selecting the right cloud security solution for our business is imperative if we want to get the best from the cloud and ensure our organization is protected from unauthorized access, data breaches and other threats. Forcepoint Cloud Access Security Broker (CASB) is a complete

cloud security solution that protects cloud apps and cloud data, prevents compromised accounts and allows we to set security policies on a per-device basis.

Security threats and solutions in clouds

There are five layers in cloud computing. They are Client Layer, Application Layer, Platform layer, Infrastructure layer and server layer. In order to address the security problems, every level should have security implementation.

Client Layer

In the cloud computing model, the cloud client consists of the computer hardware and the computer software that is totally based on the applications of the cloud services and basically designed in such way that it provides application delivery to the multiple servers at the same time, as some computers making use of the various devices which includes computers, phones, operating systems, browsers and other devices.

Application layer

The Cloud application services deliver software as a service over the internet for eliminating the need to install and run the application on the customer own computers using the simplified maintenance and support for the people which will use the cloud interchangeably for the network based access and management of the network software by controlling the activities which is managed in the central locations by enabling customers to access the applications remotely with respect to Web and application software are also delivered to many model instances that includes the various standards that is price, partnership and management characteristics which provides the updates for the centralize features.

Platform layer

In the cloud computing, the cloud platform services provides the common computing platform and the stack solution which is often referred as the cloud infrastructure and maintaining the cloud applications that deploys the applications without any cost and complexity of the buying and managing the hardware and software layers.

Infrastructure layer The Cloud Infrastructure services delivers the platform virtualization which shows only the desired features and hides the other ones using the environment in

which servers, software or network equipment are fully outsourced as the utility computing which will based on the proper utilization of the resources by using the principle of reusability that includes the virtual private server offerings for the tier 3 data center and many tie 4 attributes which is finally assembled up to form the hundreds of the virtual machines.

Server layer

The server layer also consist of the computation hardware and software support for the cloud service which is based on the multi-core processors and cloud specific operating systems and coined offerings.

Cloud Computing Attacks

As more companies move to cloud computing, look for hackers to follow. Some of the potential attack vectors criminals may attempt include:

a. **Denial of Service (DoS)** attacks - Some security professionals have argued that the cloud is more vulnerable to DoS attacks, because it is shared by many users, which makes DoS attacks much more damaging.
b. **Side Channel attacks** – An attacker could attempt to compromise the cloud by placing a malicious virtual machine in close proximity to a target cloud server and then launching a side channel attack.
c. **Authentication attacks** – Authentication is a weak point in hosted and virtual services and is frequently targeted. There are many different ways to authenticate users; for example, based on what a person knows, has, or is. The mechanisms used to secure the authentication process and the methods used are a frequent target of attackers.

d. **Man-in-the-middle cryptographic attacks** – This attack is carried out when an attacker places himself between two users. Anytime attackers can place themselves in the communication's path, there is the possibility that they can intercept and modify communications.

e. **Inside-job** – This kind of attack is when the person, employee or staffs who is knowledgeable of how the system runs, from client to server then he can implant malicious

codes to destroy everything in the cloud system.

Novel elements in the cloud threat model

The cloud computing threat model includes several novel elements. First, data and software are not the only assets worth protecting. Activity patterns also need to be protected. Sharing of resources means that the activity of one cloud user might appear visible to other cloud users using the same resources, potentially leading to the construction of covert and side channels. Activity patterns may also themselves constitute confidential business information, if divulging them could lead to reverse- engineering of customer base and Revenue size. Business reputation also merit protection. When using shared resources to do business-critical computations, it becomes harder to attribute malicious or unethical activity. Even if there are ways to clearly identify the culprits and attribute blame, bad publicity still creates uncertainty that can tarnish a long-established reputation. In addition, one must often accommodate a longer trust chain. For example, the application end-user could potentially use an application built by an SaaS provider, with the application running on a platform offered by a PaaS provider, which in turn runs on the infrastructure of an IaaS provider.

Auditing Protocols

Cloud Trust Protocol (CTP)

The Cloud Trust Protocol (CTP) is the mechanism by which cloud service consumers (also known as “cloud users” or “cloud service owners”) ask for and receive information about the elements of transparency as applied to cloud service providers. The primary purpose of the CTP and the elements of transparency is to generate evidence-based confidence that everything that is claimed to be happening in the cloud is indeed happening as described, …, and nothing else. This is a classic application of the definition of digital trust. And, assured of such evidence, cloud consumers become liberated to bring more sensitive and valuable business functions to the cloud, and reap even larger payoffs. With the CTP cloud consumers are provided a way to find out important pieces of information concerning the compliance, security, privacy, integrity, and operational security history of service elements being performed “in the cloud”.

These important pieces of information are known as the “elements of transparency”, and they deliver testimony about essential security configuration and operational characteristics for

systems deployed in the cloud. The elements of transparency empoour the cloud consumer with the right information to make the right choices about what processing and data to put in the cloud or leave in the cloud, and to decide which cloud is best suited to satisfy processing needs. This is the nature of digital trust, and reinforces again why such reclaimed transparency is so essential to new enterprise value creation. Transparency of certain important elements of information is at the root of digital trust, and thus the source of value capture and payoff.

Dynamic auditing Auditing as a Service

Auditing as a Service

1. A service to check the cloud data integrity
2. Conducted by a Third Party Auditor

Neither Data Owner nor Cloud Service Providers can provide unbiased auditing results, because they both have intention to cheat. The same as auditor to auditing a company.

Need for Third Party Auditing A third party auditor can provide unbiased auditing results. Benefit for both data owners and service providers Data Owners
– be ensured data integrity Service Providers
– Build good reputation Able to do a good job efficiently Professional Expertise Computing Capabilities

Research Issues

Privacy Preservation:
Keep the data confidential against the auditor Dynami Auditing
Allow dynamic updates of data in the cloud Batch Auditing
Combine multiple auditing tasks together to improve efficiency

Architecture of 3rd Party Auditing

Initialization: Data owner sends 1) encrypted data & verification tags to server, and 2) data index to auditor

Challenge: Auditor sends Challenge to cloud server

Proof: Server responses with Proof

Verification: Auditor verifies correctness of the Proof

Cloud Security Auditing

IT security audits determine whether an information system and its maintainers meet both the legal expectations of customer data protection and the company's standards of achieving financial success against various security threats. These goals are still relevant in the emerging cloud computing model of business, but they require customization.

Cloud computing, as defined by the National Institute of Standards and Technology (NIST), 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." In essence,cloud computing could be described as the use of computing resources—both hardware and soft ware— provided over a network, requiring minimal interaction between users and providers.

Three service models are commonly implemented in the cloud: software as a service (SaaS), platform as a service (PaaS), and infrastructure as a service (IaaS). In each of these service types, security is a signifi cant challenge. Security audits provide a clear and recognizable trail of resource access for various organizations.

Traditional IT audits typically fall into two main categories: internal and external. Internal audits refer to work done by an organization's own employees, concern very specific organizational processes, and focus primarily on optimization and risk management. External audits give an outside perspective on an organization's ability to meet the requirements of various laws and regulations. Organizations have used traditional IT audits to evaluate issues such as availability to authorized users and integrity and confidentiality in data storage and transmission.

But what happens when an organization's IT resources are moved to the cloud? Because cloud computing allows for multiple users across a large domain, it exposes novel security issues such as cloud-specifi c confi dentiality concerns. These threats pose new challenges for security auditing, but cloud advocates are responding to them. For instance, groups such as Cloud Security Alliance (CSA) are urging standardization of cloud confi dentiality, integrity, and availability auditing.

In this article, we highlight the challenges that separate cloud security auditing from traditional IT security auditing practices. These challenges illustrate the importance of special provisions for cloud security auditing in existing or newly emerging security auditing standards. We conducted a series of interviews with experienced cloud security

auditors and incorporated their insights and advice into our discussions.

Challenges

A traditional IT security audit is an examination of an IT group's checks, balances, and controls. Auditors enumerate, evaluate, and test an organization's systems, practices, and operations to determine whether the systems safeguard the information assets, maintain data integrity, and operate eff ectively to achieve the organization's business goals or objectives. 2 To support these objectives, IT security auditors need data from both internal and external sources.

In addition, cloud computing comes with its own set of security challenges. A cloud infrastructure is the result of a constant three-way negotiation among service organizations, cloud service providers (CSPs), and end users to ensure productivity while maintaining a reasonable degree of security. A CSP should keep data safe from security threats and yet give clients access anywhere with Internet service. In addition, the client organization must verify that the cloud computing enterprise contributes to its business goals, objectives, and future needs.

Although both conventional IT security auditing and cloud security auditing share many concerns, a cloud security audit must address unique problems typically not handled in traditional IT security audits. According to our interviews, the most immediate and obvious challenge lies in auditors acquiring sufficient knowledge of cloud computing. Effective cloud security auditors must be familiar with cloud computing terminology and have a working knowledge of a cloud system's constitution and delivery method. This knowledge ensures auditors pay attention to security factors that might be more important in cloud security auditing processes, including transparency; encryption; colocation; and scale, scope, and complexity .

Cloud Security:

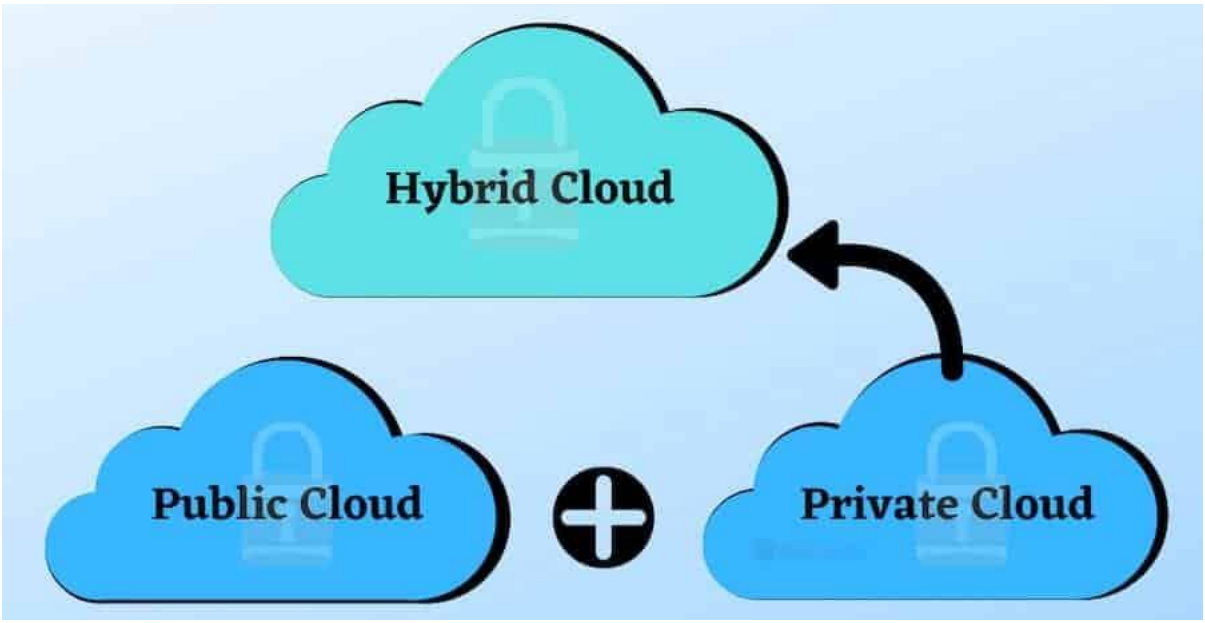
A Definition of Cloud Storage Security

Cloud-based internet security is an outsourced solution for storing data. Instead of saving data onto local hard drives, users store data on Internet-connected servers. Data Centers manage these servers to keep the data safe and secure to access.

Enterprises turn to cloud storage solutions to solve a variety of problems. Small businesses use the cloud to cut costs. IT specialists turn to the cloud as the best way to store sensitive data.

Any time we access files stored remotely, we are accessing a cloud.

Email is a prime example. Most email users don't bother saving emails to their devices because those devices are connected to the Internet.



How Secure is Cloud Storage?

All files stored on secure cloud servers benefit from an enhanced level of security. The security credential most users are familiar with is the password. Cloud storage security vendors secure data using other means as well.

Some of these include:

Advanced Firewalls: Firewalls inspect traveling data packets. Simple ones only examine the source and destination data. Advanced ones verify packet content integrity. These programs then map packet contents to known security threats.

Intrusion Detection: Online secure storage can serve many users at the same time. Successful cloud security systems rely on identifying when someone tries to break into the system. Multiple levels of detection ensure cloud vendors can even stop intruders who break past the network's initial defenses.

Event Logging: Event logs help security analysts understand threats. These logs record network actions. Analysts use this data to build a narrative concerning network events. This helps them predict and prevent security breaches.

Internal Firewalls: Not all accounts should have complete access to data stored in the cloud. Limiting secure cloud access through internal firewalls boosts security. This ensures that even a compromised account cannot gain full access.

Encryption: Encryption keeps data safe from unauthorized users. If an attacker steals an encrypted file, access is denied without finding a secret key. The data is worthless to anyone who does not have the key.

Physical Security: Cloud data centers are highly secure. Certified data centers have 24-hour monitoring, fingerprint locks, and armed guards. These places are more secure than almost all on-site data centers. Different cloud vendors use different approaches for each of these factors. For instance, some cloud storage systems keep user encryption keys from their users. Others give the encryption keys to their users.

Best-in-class cloud infrastructure relies on giving users the ideal balance between access and security. If we trust users with their own keys, users may accidentally give the keys to an unauthorized person.

There are many different ways to structure a cloud security framework. The user must follow security guidelines when using the cloud.

For a security system to be complete, users must adhere to a security awareness training program. Even the most advanced security system cannot compensate for negligent users.

Cloud Data Security Risks

Security breaches are rarely caused by poor cloud data protection. More than 40% of data security breaches occur due to employee error. Improve user security to make cloud storage more secure.

Many factors contribute to user security in the cloud storage system.

Many of these focus on employee training:

Authentication: Weak passwords are the most common enterprise security vulnerability. Many employees write their passwords down on paper. This defeats the purpose. Multi-factor authentication can solve this problem.

Awareness: In the modern office, every job is a cybersecurity job. Employees must know why security is so important and be trained in security awareness. Users must know how criminals break into enterprise systems. Users must prepare responses to the most common attack vectors.

Phishing Protection: Phishing scams remain the most common cyber attack vector. These attacks attempt to compromise user emails and passwords. Then, attackers can move through business systems to obtain access to more sensitive files.

Breach Drills: Simulating data breaches can help employees identify and prevent phishing attacks. Users can also improve response times when real breaches occur. This establishes protocols for handling suspicious activity and gives feedback to users.

Measurement: The results of data breach drills must inform future performance. Practice only makes perfect if analysts measure the results and find ways to improve upon them. Quantify the results of simulation drills and employee training to maximize the security of cloud storage.

Cloud Storage Security Best Practices

Cloud storage providers store files redundantly. This means copying files to different physical servers.

Cloud vendors place these servers far away from one another. A natural disaster could destroy one data center without affecting another one hundreds of miles away.

Consider a fire is breaking out in an office building. If the structure contains paper files, those files will be the first to burn. If the office's electronic equipment melts, then the file backups will be gone, too.

If the office saves its documents in the cloud, this is not a problem. Copies of every file exist in multiple data centers located throughout the region. The office can move into a building with Internet access and continue working.

Redundancy makes cloud storage security platforms failure-proof. On-site data storage is far riskier. Large cloud vendors use economies of scale to guarantee user data is intact. These vendors measure hard drive failure and compensate for them through redundancy.

Even without redundant files, only a small percentage of cloud vendor hard drives fail. These companies rely on storage for their entire income. These vendors take every precaution to ensure users' data remains safe.

Cloud vendors invest in new technology. Advances improve security measures in cloud computing. New equipment improves results.

This makes cloud storage an excellent option for securing data against cybercrime. With a properly configured cloud solution in place, even ransomware poses no real threat. We can

wipe the affected computers and start fresh. Disaster recovery planning is a critical aspect of cloud storage security.

Privacy Preserving in Cloud

To enable privacy-preserving public auditing for cloud data storage under the aforementioned model, our protocol design should achieve the following security and performance guarantee:

1) Public auditability: to allow TPA to verify the correctness of the cloud data on demand without retrieving a copy of the whole data or introducing additional on-line burden to the cloud users.

2) Storage correctness: to ensure that there exists no cheating cloud server that can pass the audit from TPA without indeed storing users' data intact.

3) Privacy-preserving: to ensure that there exists no way for TPA to derive users' data content from the information collected during the
auditing process.

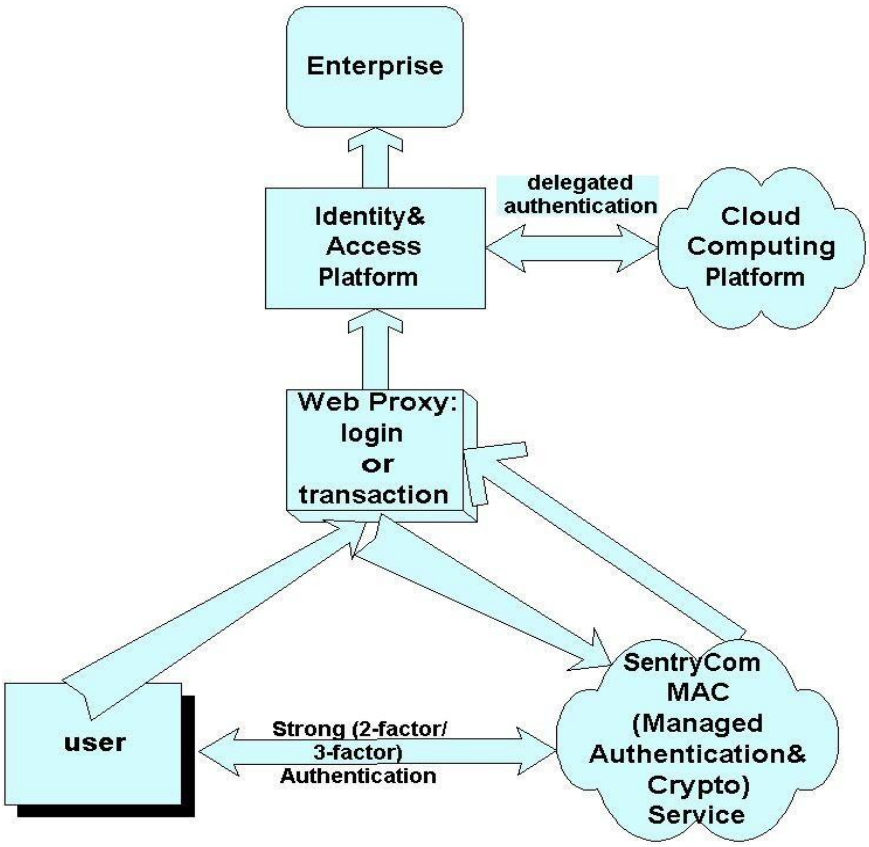
4) Batch auditing: to enable TPA with secure and efficient auditing capability to cope with multiple auditing delegations from possibly large number of different users simultaneously.

5) Lightweight: to allow TPA to perform auditing with minimum communication and computation overhead.

Algorithm:

A public auditing scheme consists of four algorithms (KeyGen, SigGen, GenProof, VerifyProof).

- KeyGen: key generation algorithm that is run by the user to setup the scheme
- SigGen: used by the user to generate verification metadata, which may consist of MAC, signatures or other information used for auditing
- GenProof: run by the cloud server to generate a proof of data storage correctness
- VerifyProof: run by the TPA to audit the proof from the cloud server Flowchart



Fully Homomorphic Encryption

Fully Homomorphic Encryption FHE allows for arbitrary computations on encrypted data. Computing on encrypted data means that if a user has a function f and want to obtain f(m1, . . . , mn) for some inputs m1, . . . , mn, it is possible to instead compute on encryptions of these inputs, c1, . . . , cn, obtaining a result which decrypts to f(m1, . . . , mn). In some cryptosystems the input messages (plaintexts) lie within some algebraic structure, often a group or a ring. In such cases the ciphertexts will often also lie within some related structure, which could be the same as that of the plaintexts. The function f in older homomorphic encryption schemes is typically restricted to be an algebraic operation associated with the structure of the plaintexts. For instance, consider ElGamal. If the plaintext space is a group G, then the ciphertext space is the product G×G, and f is restricted to the group operation on G. Indeed most schemes fit such a structure. The aim of fully homomorphic encryption to be to extend the function f to be any function. This aim can be achieved if the scheme is homomorphic with respect to a functionally complete set of operations and it is possible to iterate operations from that set. While it is always a requirement that encryption schemes are efficient in a theoretical sense, namely running in polynomial time in the security parameter, practical efficiency was not the first priority in obtaining the first FHE schemes. One reason for the lack of efficiency of these schemes is that they use a plaintext space consisting of a

single bit and are homomorphic with respect to addition and multiplication modulo 2. While any function of any complexity can be built up from such basic operations, that may require a large number of such operations.

Big data security

Security and privacy issues are magnified by velocity, volume and variety of big data, such as large scale cloud infrastructures, diversity of data sources and formats, streaming nature of data acquisition ,and high volume inter-cloud migration. The use of large scale cloud infrastructure with diversity of software platforms, spread across large networks of computers, also increases the attack surface of entire system. therefore traditional security mechanisms, which are tailored to securing small scale static(as opposed to streaming)data, are inadequate. Ex. analytics for anomaly detection would generate too many outliers.

Privacy and Security with a variety of personal data such as buying preference healthcare records, and location-based information being collected by big data applications and transferred over networks, the public‟s concerns about data privacy and security naturally arise. While there have been significant studies on protecting data centers from being attacked, the privacy and security loopholes when moving crow sourced data to data centers remain to be addressed. There is an urgent demand on technologies that endeavor to enforce privacy and security in data transmission. Given the huge data volume and number of sources, this requires a new generation of encryption solutions (e.g., homomorphic encryption). On the other hand, big data techniques can also be used to address the security challenges in networked systems. Network attacks and intrusions usually generate data traffic of specific patterns in networks. By analyzing the big data gathered by a network monitoring system, those misbehaviors can be identified proactively, thus greatly reducing the potential loss.

Security a Big Question of Big Data Big data implies performing computation and database operations for massive amounts of data, remotely from the data owner‟s enterprise. Since a key value proposition of big data is access to data from multiple and diverse domains, security and privacy will play a very important role in big data research and technology. The limitations of standard IT security practices are well-known, making the ability of attackers to use software subversion to insert malicious software into applications and operating systems a serious and growing threat whose adverse impact is intensified by big data. So, a big question is what security and privacy technology is adequate for controlled assured sharing for efficient direct

access to big data. Making effective use of big data requires access from any domain to data in that domain, or any other domain it is authorized to access. Several decades of trusted systems developments have produced a rich set of proven concepts for verifiable protection to substantially cope with determined adversaries, but this technology has largely been marginalized as “overkill” and vendors do not widely offer it.

With great poour of data comes great responsibility! A big data initiative should not only focus on the volume, velocity or variety of the data, but also on the best way to protect it. Security is usually an afterthought, but Elemental provides the right technology framework to get we the deep visibility and multilayer security any big data project requires. Multilevel protection of our data processing nodes means implementing security controls at the application, operating system and network level while keeping a bird's eye on the entire system using actionable intelligence to deter any malicious activity, emerging threats and vulnerabilities.

Big Data Security and Privacy Challenges

Secure Computations in Distributed Programming Framework Distributed programming framework utilize parallelism in computations and storage to process massive amounts of the data .A popular example is map reduce framework, which splits an input file into multiple chunks in the first phase of map reduce, a mapper for each chunk reads the data, perform some computation, and outputs a list of key/value pairs. In the next phase, a reducer combines the values belonging to each distinct key and outputs the result. There are two major attack prevention measure: securing the manners and securing the data in the presence of an untrusted manner.

Security Best Practices for Non Relational Data Stores Non relational data stores popularized by NoSQL databases are still evolving with respect to security infrastructure. For instance, robust solutions to NoSQL injection are still not mature each NoSQL DBs oure built to tackle different challenges posed by the analytics world and hence security was never part of the model at any point of its design stage. Developers using NoSQL databases usually embed security in the middleware .NoSQL databases do not provide any Support for Enforcing it explicitly in the database. However, clustering aspect of NoSQL databases poses additional challenge s to the robustness of such security practices.

Secure Data Storage and Transaction Logs Data and transaction logs are stored in multi-tiered

storage media manually moving data between tiers gives the it manager direct control over exactly what data is moved and when. However as the size of data set has been and continues to be, growing exponentially, scalability and availability necessited auto tiering for big data storage management. Auto tiering solutions do not keep track of where the data is stored ,which poses new challenges to secure data storage. new mechanisms are imperative to thwartun authorised access and maintain 24/7 availability.

End Point Input Validation/Filtering Many big data use cases in Enterprise settings require data collection from many sources, such as end point devices for example, a security information and event management system (SIEM) may collect event logs from millions of hardware devices and software application in an enterprise network . A key challenge in the data collection process is input validation :how can we trust the data? how can we validate that a source of input data is not malicious and how can we filter malicious input from our collection? input validation and filtering is a daunting challenge posed by untrusted input sources, especially with the bring our own device (BYOD) model.

Real –Time Security/Compliance Monitoring Real time security monitoring has always been a challenge ,given the number of alerts generated by (security)devices. These alerts (correlated or not)lead to many false positive , which are mostly ignored or simply “clicked away”, as humans cannot cope with the shear amount. This problem might even increase with the bid data given the volume and velocity of data streams however, big data technologies might also provide an opportunity, in the sense that these technologies do allow for fast processing and analytics of different types of data .Which in its turn can be used to provide, for instance, real time anomaly detection based on scalable security analytics.

Scalable and Compos able Privacy-Preserving Data Mining And Analytics Big data can be seen as a troubling manifestation of big brother by potentially enabling invasions of privacy ,invasive marketing, decreased civil freedoms ,and increase state and corporate control. A recent analysis of how companies are leveraging data analytics for marketing purpose identified an example of how a retailer was able to identify that teenager was pregnant before her father knew. Similarlly anonym zing data for analytics is not enough to maintain user privacy. For example AOL released anonymized search logs for academic purposes ,but users oure easily identified by their searchers .Netflix faced a similar problem when users of their anonymized data set oure identified by correlating their Netflix movie scores with IMDB scores. Therefore ,it is important to establish guidelines‟ and recommendations for

preventating inadvertent privacy disclosures.

Cryptographically Enforced Access Control And Secure Communication To ensure that the most sensitive private data is end to end secure and only accessible to the authorized entities, data has to be encrypted based on access control policies. Specific research in this area such as attribute-based encryption (ABE)has to be made richer, more efficient, and scalable. To ensure authentication, agreement and fairness among the distributed entities, a cryptographically secure communication framework has to be implemented.

Granular Access Control The security Property that matters from the perspective of access control is secrecy-preventing access to data by people that should not have access .The problem with course- grained access mechanisms is that data that could otherwise be shared is often swept into a more restrictive category to guarantee sound security granular access control gives data managers a scalpel instead of a sword to share data as much as possible without compromising secrecy.

Granular Audits With real time security monitoring ,we try to be notified at the moment an attack takes place. in reality, this will not always be the case(e.g, new attacks, missed true positives). In order to get to the bottom of the missed attack ,we need audit information. This is not only relevant because we want to understand what happened and what went wrong ,but also because compliance, regulation and forensics reasons .in that regard ,auditing is not something new, but the scope and granularity might be different. For example, we have to deal with more data objects, which probably are distributed.

Data Provenance metadata will grow in complexity due to large provenance graphs generated from provenance- enabled programming environments in big data applications. Analysis of such large provenance graphs to detect metadata dependencies for security/confidentiality applications is computationally intensive.

Cloud availability

High-availability is, ultimately, the idea of anywhere and anytime access to services, tools and data and is the enabler of visions of a future with companies with no physical offices or of global companies with completely integrated and unified IT systems. Availability is also related to reliability: a service that is on 24x7 but goes constantly offline is useless. For a

service to have true high-availability, it needs not only to be always-on, but also to have several "nines" (99.999[...]) of reliability.

It has long been the case that to build systems with this kind of reliability and availability means large costs for companies. For something like this, it's not enough to simply have a failover cluster of servers in a data center: we must also have multiple redundant energy sources for the data center and even to have replication between multiple geographical locations in case of disasters. With the exception of very large, multinational companies, almost no-one could afford such a setup.

With the advent of infrastructure-as-a-service and platform-as-a-service providers, however, the costs of building such a service have decreased dramatically. It is now possible for most cloud-based service providers, especially for software-based services, to offer very aggressive service level agreements.

DoS attacks

A Denial of Service Attack, is an action (or set of actions) executed by a malicious entity to make a resource unavailable to its intended users.
Another definition “a group of otherwise-authorized users of a specified service is said to deny service to another group of otherwise-authorized users if the former group makes the specified service unavailable to the latter group for a period of time that exceeds the intended (and advertised) waiting time.”

There are three basic types of attacks

1) Consumption of scarce, limited, or non-renewable resources
2) Destruction or alteration of configuration information
3) Physical destruction or alteration of network components.
4) The targeted resources can be network bandwidth, CPU, memory, I/O bandwidth, disk space, or any combination of them.

Detection Mechanisms

Detection is an important step before directing further actions to counter a DoS attack. DoS

response mechanisms depend on the attack related information discovered by detection mechanisms for countering the attack. Some response mechanisms rely on identification of the malicious actions, while others require identifying the entity that is performing the malicious actions. There are also few mitigation mechanisms that depend on discovering the fact that an attack is ongoing, in order to initiate the mitigation process.

Fault Tolerance management in cloud computing

In cloud computing the major problem area is fault tolerance. Fault tolerance is a major concern to guarantee availability and reliability of critical services as well as application execution. In order to minimize failure impact on the system and application execution, failures should be anticipated and handle. Fault tolerance techniques are used to predict these failures and take an appropriate action before or after failures occur.

The main benefits of implementing fault tolerance in cloud computing include failure recovery, loour cost, improved performance metrics. Robustness leads to the property to providing of a correct service in an adverse situation arising due to an uncertain system environment. Dependability is related to some QoS (Quality of Services) aspects provided by the system, it includes the attributes like reliability and availability.

Types of Faults

The faults can be classified on several factors such as:

Network fault: A Fault occur in a network due to network partition, Packet Loss, Packet corruption, destination failure, link failure, etc.

Physical faults: This Fault can occur in hardware like fault in CPUs, Fault in memory, Fault in storage, etc. Media faults: Fault occurs due to media head crashes

Processor faults: fault occurs in processor due to operating system crashes, etc. Process faults: A fault which occurs due to shortage of resource, software bugs, etc.

Service expiry fault: The service time of a resource may expire while application is using it.

A fault can be categorized on the basis of computing resources and time.

A failure occurs during computation on system resources can be classified as: omission failure, timing failure, response failure, and crash failure..

Permanent: These failures occur by accidentally cutting a wire, poour breakdowns and so on. It is easy to reproduce these failures. These failures can cause major disruptions and some part of the system may not be functioning as desired.

Intermittent: These are the failures appears occasionally. Mostly these failures are ignored while testing the system and only appear when the system goes into operation. Therefore, it is hard to predict the extent of damage these failures can bring to the system.
Transient: These failures are caused by some inherent fault in the system. However, these failures are corrected by retrying roll back the system to previous state such as restarting software or resending a message. These failures are very common in computer systems.

Existing fault tolerance techniques in cloud computing

Various fault tolerance techniques are currently prevalent in clouds:-
Check pointing - It is an efficient task level fault tolerance technique for long running and big applications .In this scenario after doing every change in system a check pointing is done. When a task fails, rather than from the beginning it is allowed to be restarted that job from the recently checked pointed state.

Job Migration - Some time it happened that due to some reason a job can- not be completely executed on a particular machine. At the time of failure of any task, task can be migrated to another machine. Using HA-Proxy job migration can be implemented.

Replication - Replication means copy. Various tasks are replicated and they are run on different resources, for the successful execution and for getting the desired result. Using tools like HA-Proxy, Hadoop and AmazonEc2 replication can be implemented.

Self- Healing - A big task can divided into parts .This Multiplication is done for better performance. When various instances of an application are running on various virtual machines, it automatically handles failure of application instances.

Safety-bag checks: In this case the blocking of commands is done which are not meeting the safety properties.

S-Guard- It is less turbulent to normal stream processing. S-Guard is based on rollback recovery. SGuard can be implemented in HADOOP, Amazon EC2.

Retry - In this case we implement a task again and gain. It is the simplest technique that

retries the failed task on the same resource.

Task Resubmission- A job may fail now whenever a failed task is detected, In this case at runtime the task is resubmitted either to the same or to a different resource for execution.**iming check:** This is done by watch dog. This is a supervision technique with time of critical function.

Rescue workflow- This technique allows the workflow to persist until it becomes unimaginable to move forward without catering the failed task.

Software Rejuvenation-It is a technique that designs the system for periodic reboots. It restarts the system with clean state and helps to fresh start.

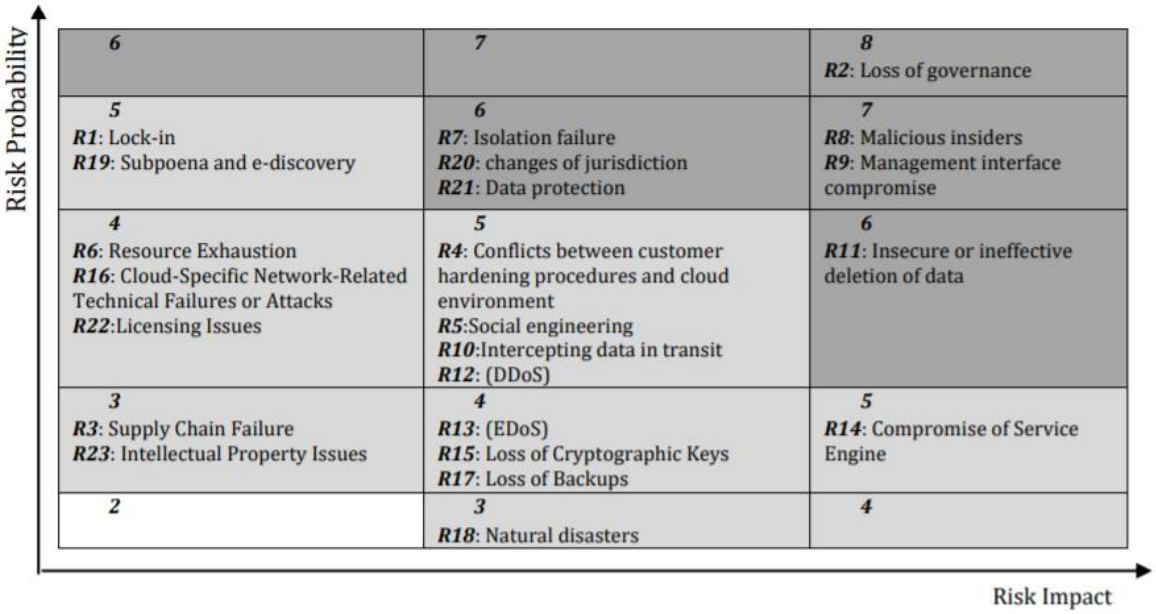
Pre-emptive Migration- Pre-emptive Migration count on a feedback-loop control mechanism. The application is constantly monitored and analysed.

Masking: After employment of error recovery the new state needs to be identified as a transformed state. Now if this process applied systematically even in the absence of effective error provide the user error masking.

Reconfiguration: In this procedure we eliminate the faulty component from the system.

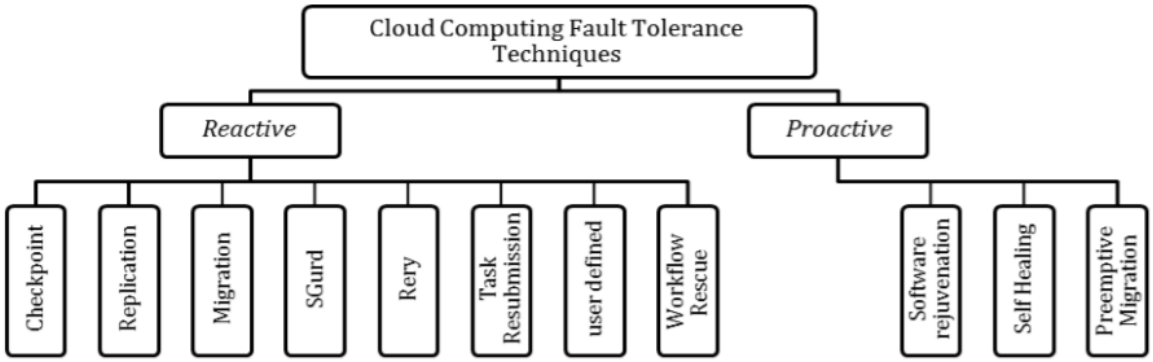
Reliability and Failures prepared a report carried out by European Network and Information Security Agency (ENISA) presented around twenty-three risks. They categorized the risks into three classes: policy and organizational, technical, and legal. A table for each risk contains the probability estimate, impact estimate, level of the risk. They presented risk assessment to assess

the impact of the risk on business and measure the likelihood of the risk.



The distribution of the risks probability and their impacts.

There are several types of failures that influence the reliability of cloud services. According to the failures ,they are classified as request stage failures and execution stage failures. Request stage failures affected the request before it accesses the required resource such as overflow and timeout. However, execution stage failures affected the request during its execution time such as data resource misses, computing resource missing, software failure, database failure, hardware failure, and network failure. For a cloud system to be reliable, it should be fault tolerant i.e. it should continue its operation despite any failures . Availability and reliability guarantees are mentioned by some cloud providers in the Service Level Agreement (SLA). For example, in Amazon EC2 SLA, the company provides its customers a service availability of at least 99.95% which is defined as a monthly uptime percentage, and in case of service unavailability the company assigns service credits to the users with a specific percentage as a penalty.



Fault tolerance techniques in cloud computing.

Proactive fault tolerance techniques take some preventative measures such as to avoid any failures in the application in future. Some of the techniques used are as follows:

- Software Rejuvenation: Restart the system with a clean state of the software.

- Self-Healing: It tolerates failure of application instances running on different Virtual Machines (VM)

. • Preemptive Migration: The application is monitored, analyzed and then preventive measures are taken. FTCC services are still under development and studies. Moreover, new techniques emerge in the cloud from time to time. A cloud service Failure-as-a-Service (FaaS) is proposed in here failure drills are run from time to time rather than waiting for unexpected failures to happen; when a drill finds a problem, recovery mechanism are initiated, thereby preventing an outage. The most popular FTCC techniques used are Checkpointing, Replication, and Job Migration. The details about the three techniques are described below:

Checkpointing: Checkpointing is a FT mechanism that takes snapshots of the system state and save it in a permanent storage (checkpoint). The system rollbacks to that state if a fault is detected by the system, instead of restarting from the beginning. There are several advantages of using checkpointing technique due to its low cost and high performance. However, overhead is the main disadvantage of using checkpointing.

Replication: Replication is copying all files to another storage device; the storage capacity is not an obstacle in order to improve system availability. The ideal number of replica is derived and the load balance is achieved using replication. The advantages of replication technique include increased parallelism (i.e. faster query execution), higher performance, and increased speed in processing. The drawbacks of replication technique include increased overhead and cost of replication.

Job Migration: Migration is the replacement of the running VM in another VM across distinct physical hosts which separates hardware and software to make management easier. Several papers attempted to determine the time to perform VM migration. A new method was proposed to predict the migration performance and energy cost. Advantages of job migration include easy management, load balancing, and service availability during migration. The disadvantages of job migration include cost and high overhead. Moreover, there is a possibility that the whole VM is malicious

Fault Tolerance Models in Cloud Computing (FTMCC) FTMCC is categorized into three groups Replication based fault tolerance: Cloud computing environments mostly have predefined in handling faults. For example, a passive replication model is capable of only tolerating crash faults while hardware redundancy based FT, as the response time is a significant parameter Replication based FTCC An efficient replication scheme is proposed in It transparently tolerates crash failures and offers high availability, high performance, generality, transparency, and seamless failure recovery. One of the main drawbacks of this model is the latency, as the network buffering causes performance overhead it requires additional hardware. It is an efficient replication based model but significantly introduces network delay. Thus, it is not suited for applications that are sensitive to network delay or latency. This drawback is overcome by largely reducing the external network buffering that caused the network latency. A middleware called Niagara is proposed that offers high availability and low latency. Shadow Replication is proposed to ensure successful job completion. Byzantine Fault Tolerance (BFT) models are replication based models that are used to tolerate byzantine failures. A byzantine fault tolerance framework for voluntary-resource cloud computing that tolerate faults such as: crash, arbitrary, and behaviours. The framework follows the same message algorithm that used in Practical Byzantine Fault Tolerance (PBFT) but they are different in the number of replicas. High scalability and less overhead are the advantages of using A Virtualization and Fault Tolerance (VFT) model is represented into reduce the service time. New replication is created based on selecting the finishing time of applications and dynamically generating the number of replicas. Adaptive Fault Tolerance model in Real time Cloud Computing (AFTRC) is proposed. AFTRC offers both backward and forward recovery.

Checkpoint based fault tolerance: Checkpointing consists of three main types: coordinated checkpointing, uncoordinated checkpointing, and Communication Induced Checkpointing (CIC). CIC is an equal cost checkpointing scheme with varying checkpoint interval .

Checkpoint based fault tolerance for cloud computing model uses Another Union File System (AUFS) in order to distinguish read-only properties from read and write parts in VM image. Another model with a union file system is presented in that uses time storing VM checkpoints.

Cloud Computing in India

Cloud computing is a tremendous innovation in the digital landscape that has changed the way IT solutions are delivered and how end-users put them to use. The cloud computing space is growing and will continue to do so.

According to a recent report by Gartner, cloud computing will constitute the bulk of IT spending by 2016. In India alone, it is predicted that the cloud market will reach over $3 billion by next year—an almost five-fold increase from 2012.

Considering the Indian context, cloud computing is set to transform how we do business and how we move up in the digital value chain. For India, it has a direct beneficial effect for small to medium sized businesses (SMBs), dotted across the country. This sector employs 40% of the workforce and is growing at a rate of 8% per year; it will also dictate the future course of Indian development. Latest figures indicate that 68% of SMBs having over 100 employees have resorted to cloud computing. According to Zinnov, a leading consulting company, the cloud computing market in India is expected to reach $4.5 billion with most users being SMBs. According to another report, if all SMBs in India oure to adopt cloud computing, the market could reach $56 billion, creating additional 1.1 million jobs in the near future. The cloud market continues to expand in India, so will the number of job prospects. It is estimated this year alone cloud computing will create over 2 million jobs. It is believed that players like TCS, Infosys, HCL and Tech Mahindra will bid for cloud computing service providers rather than developing solutions through their in-house research, as it requires huge funding. The competition among Indian service providers will have a telling effect on the pricing of solutions. In short, the scope for cloud computing and its successful applications in Indian companies are bound to go up.

The advantages of the cloud are well known and documented. But one aspect that has to be understood is how it leads to career opportunities. Cloud computing brings down the cost and manpoour requirements since the costly software packages and hardware systems needed to install them are not longer required. The software needed can be downloaded from the internet as per the requirement of the unit, by paying the service provider. Evidence

indicates it would be economical to download the software, rather than storing them in the hardware of the system.

Another important aspect is the manpoour requirements of service providers, who have to store and retrieve customised solutions from the internet. This will require huge investments in R&D professionals, besides employing people for day-to-day technological operations and other related work. The indications are there will be a proliferation of service providers, including global players like Google, Microsoft and Amazon who will roll out their Indian plans.

There are some key challenges. One is the accessibility of broadband internet, which is limited in India. Hopefully, with the rollout of Digital India, it will give a critical push to enterprises to make the switch to cloud computing.

India Inc has pledged R4.5 lakh crore for Digital India, which can create employment for some 18 lakh people. A good number of them will be in cloud computing. With the launch of 100 Smart Cities, 500 rejuvenated cities and numerous projects to create industrial hubs, a strong virtual backbone is a critical necessity to take the development process to the next level.

Skill formation for handling complex tasks of cloud computing is important. To evolve courses of various durations to cater to this need and popularise them from the school level itself to get the desired results and to give an option to the teeming millions to choose their field with the confidence that they can make a good career out of cloud computing.

The future of cloud computing as well as career opportunities in the field shines bright for India.

References:

1. Cloud Security: https://phoenixnap.com/blog/cloud-storage-security
2. Cloud data Security: https://www.forcepoint.com/cyber-edu/cloud-security
3. Ravi Jhawar Vincenzo Piuri, and Marco Santambrogio, Fault Tolerance Management in Cloud Computing: A System-Level Perspective, Article in IEEE Systems Journal · June 2013

4. Mohammad H. Alshayeji* , Mohammad Al-Rousan, Eman Yossef, Hanem Ellethy, A Study on Fault Tolerance Mechanisms in Cloud Computing, International Journal of Computer Electrical Engineering,pp.62-71