Malla Reddy Engineering College
(Autonomous)
Department of Computer Science and Engineering

2017-18 Onwards (MR-17)	**MALLA REDDY ENGINEERING COLLEGE (Autonomous)**	**B.Tech. VI Semester**		
Code: 70533	**CLOUD COMPUTING [Professional Elective - II]**	**L**	**T**	**P**
Credits: 4		**3**	**2**	**-**

Prerequisites: Computer Networks

Course Objectives:

This course provides the students to gain knowledge in the cloud computing environment, security architecture and development of cloud services. Students will also examine the collaboration of real time cloud services and analyze the case studies from various cloud development tools.

MODULE I: Introduction [13 Periods]

Understanding Cloud Computing - Cloud Computing – History of Cloud Computing – Cloud Architecture – Cloud Storage –Why Cloud Computing Matters.

Issues in Cloud Services - Advantages of Cloud Computing – Disadvantages of Cloud Computing

–Companies in the Cloud Today – Cloud Services.

MODULE II: Development of Services [13 Periods]

Developing Cloud Services - Web-Based Application – Pros and Cons of Cloud Service Development – Types of Cloud Service Development – Software as a Service – Platform as a Service.

Web Services -Web Services – On-Demand Computing – Discovering Cloud Services, development Services and Tools – Amazon EC2 – Google App Engine – IBM Clouds.

MODULE III: Cloud ComputingSecurity Architecture [13 Periods]

A:Cloud Security - Cloud security fundamentals - Vulnerability assessment tool for cloud- Privacy and Security in cloud. Cloud computing security architecture: Architectural Considerations- General Issues-Trusted Cloud computing- Secure Execution Environments and Communications-Micro-architectures.

B:Identity Management - Identity Management and Access control, Identity management- Access control, Autonomic Security.

MODULE IV: Community Services [13 Periods]

Cloud Computing For Everyone - Centralizing Email Communications – Collaborating on Schedules – Collaborating on To-do Lists – Collaborating Contact Lists.

Community in Services - Cloud Computing for the Community – Collaborating on Group Projects and Events – Cloud Computing for the Corporation.

MODULE V: Case Studies and Applications [12 Periods]

Cloud Computing Case Studies - Cloud computing case studies: Google App Engine – IBM Clouds –Windows live.

Applications - Micro soft dynamic CRM- Salesforce.com CRM- App Exchange – Amazon S3 – Oracle OBIEE.

TEXTBOOKS

1.John W.Rittinghouse, James F.Ransome, "Cloud Computing: Implementation, Management and Security", CRC Press, 2012.

2.Anthony T.Velte, Toby J Velte Robert Elsenpeter, **"Cloud Computing a practical approach"**, TMH, 2010.

REFERENCES

1.Michael Miller, "Cloud Computing: Web-Based Applications That Change the Way you

Work and Collaborate Online", Que Publishing, 2008.

2.Haley Beard, "Cloud Computing Best Practices for Managing and Measuring Processes for On-demand Computing, Applications and Data Centers in the Cloud with

SLAs", Emereo Pty Limited, 2nd Edition, 2009.

3.Raj Kumar Buyya, **"Mastering Cloud computing"**, TMH, 2013.

E-RESOURCES

1 http://www.motc.gov.qa/sites/default/files/cloud_computing_ebook.pdf

2http://www.ishuchita.com/C.S.E/Cloud%20Computing/Cloud%20Computing%20Practical%2 0Approach.pdf

3 http://ndl.iitkgp.ac.in/document/zyMnqgZQXCJME6wgSqrU87VCGcelOw5mZ-5ybmrhKBj79VQPP0_ZQHLqcOopPDoaFWhZybCrPg_joTbBU8ZpGA

4 http://www.springer.com/computer/communication+networks/journal/13677

5http://nptel.ac.in/courses/106106129/28

Course Outcomes:

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

1.**Infer** the main concepts, strengths and limitations of Cloud computing.

2.**Explain** the development of cloud and web services.

3.**Identify** the core issues of cloud computing security architecture and their execution environments.

4.**Create** new ideas and innovations in cloud computing.

5.**Apply** the appropriate technologies, algorithms, and approaches for the related issues.

Module I: Understanding Cloud Computing

Cloud Computing – History of Cloud Computing – Cloud Architecture – Cloud Storage–Why Cloud Computing Matters – Advantages of Cloud Computing – Disadvantages of Cloud Computing – Companies in the Cloud Today – Cloud Services

Cloud Computing

Cloud Computing provides us means of accessing the applications as utilities over the Internet. It allows us to create, configure, and customize the applications online.

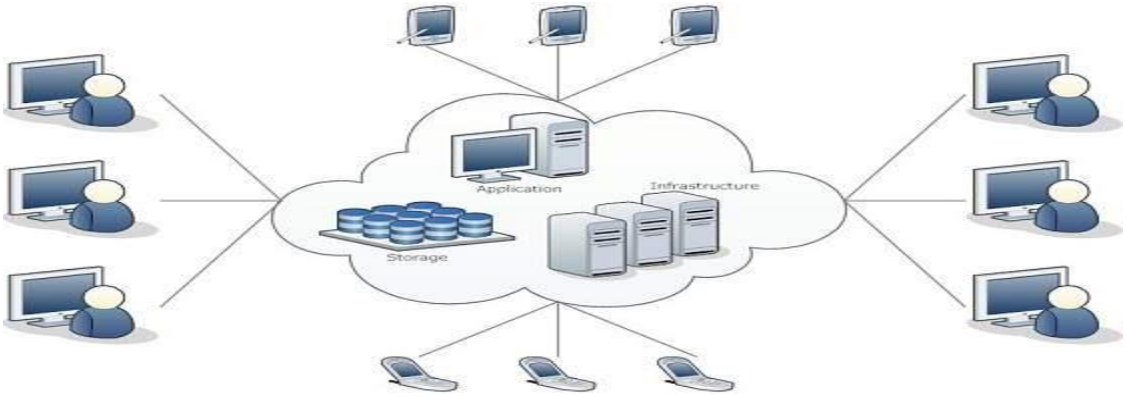
What is Cloud?

The term **Cloud** refers to a **Network** or **Internet.** In other words, we can say that Cloud is something, which is present at remote location. Cloud can provide services over public and private networks, i.e., WAN, LAN or VPN.

Applications such as e-mail, web conferencing, customer relationship management (CRM) execute on cloud.

What is Cloud Computing?

Cloud Computing refers to **manipulating, configuring,** and **accessing** the hardware and software resources remotely. It offers online data storage, infrastructure, and application.



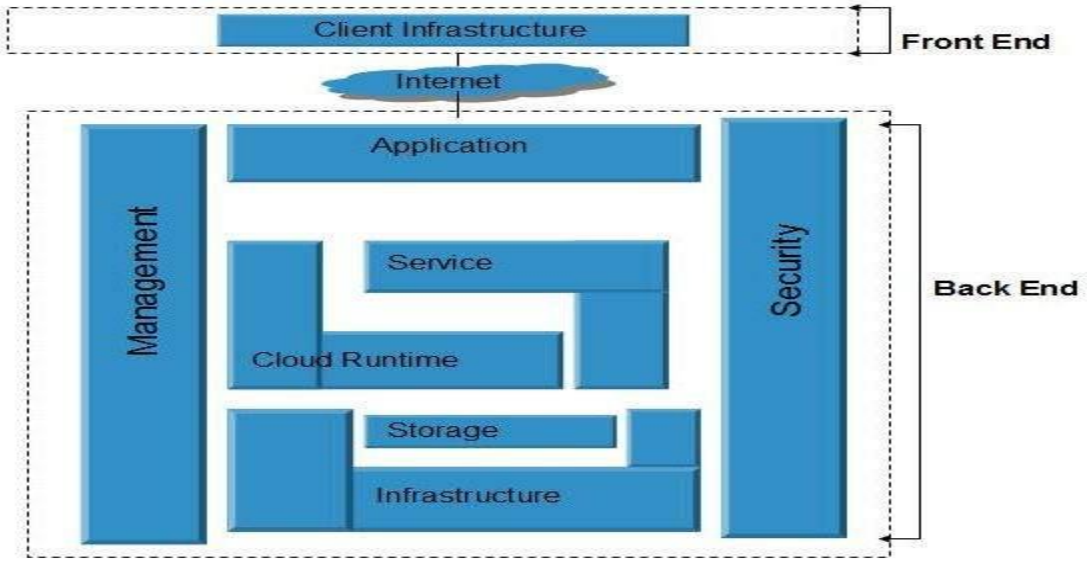
Cloud computing offers **platform independency,** as the software is not required to be installed locally on the PC. Hence, the Cloud Computing is making our business applications **mobile** and **collaborative.**

Cloud Computing Architecture

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

- Front End
- Back End

Each of the ends is connected through a network, usually Internet. The following diagram shows the graphical view of cloud computing architecture:



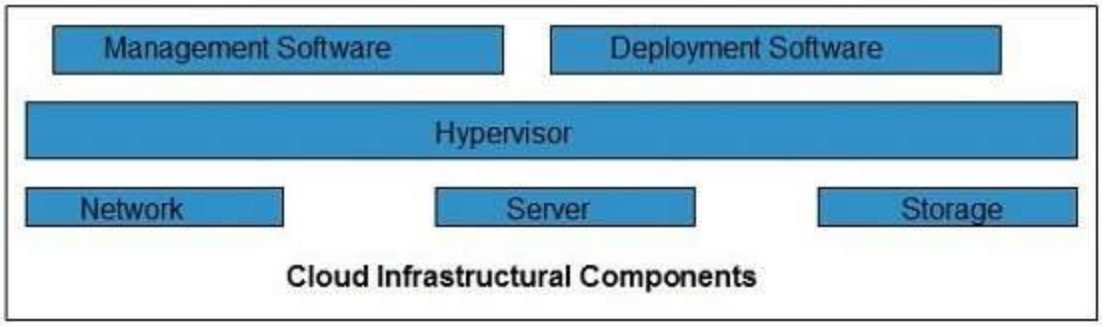
Front End

The **front end** refers to the client part of cloud computing system. It consists of interfaces and applications that are required to access the cloud computing platforms, Example - Web Browser.

Back End

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

Cloud infrastructure consists of servers, storage devices, network, cloud management software, deployment software, and platform virtualization.



Hypervisor

Hypervisor is a **firmware** or **low-level program** that acts as a Virtual Machine Manager. It allows to share the single physical instance of cloud resources between several tenants.

Management Software

It helps to maintain and configure the infrastructure.

Deployment Software

It helps to deploy and integrate the application on the cloud.

Network

It is the key component of cloud infrastructure. It allows to connect cloud services over the Internet. It is also possible to deliver network as a utility over the Internet, which means, the customer can customize the network route and protocol.

Server

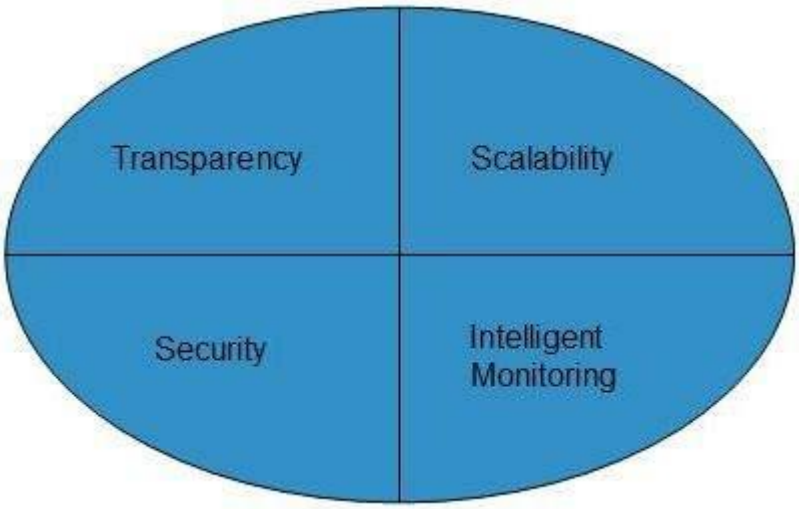
The **server** helps to compute the resource sharing and offers other services such as resource allocation and de-allocation, monitoring the resources, providing security etc.

Storage

Cloud keeps multiple replicas of storage. If one of the storage resources fails, then it can be extracted from another one, which makes cloud computing more reliable.

Infrastructural Constraints

Fundamental constraints that cloud infrastructure should implement are shown in the following diagram:



Transparency

Virtualization is the key to share resources in cloud environment. But it is not possible to satisfy the demand with single resource or server. Therefore, there must be transparency in resources, load balancing and application, so that we can scale them on demand.

Scalability

Scaling up an application delivery solution is not that easy as scaling up an application because it involves configuration overhead or even re-architecting the network. So, application delivery solution is need to be scalable which will require the virtual infrastructure such that resource can be provisioned and de-provisioned easily.

Intelligent Monitoring

To achieve transparency and scalability, application solution delivery will need to be capable of intelligent monitoring.

Security

The mega data center in the cloud should be securely architected. Also the control node, an entry point in mega data center, also needs to be secure.

Cloud Storage

Cloud Storage is a service that allows to save data on offsite storage system managed by third-party and is made accessible by a **web services API.**

Storage Devices

Storage devices can be broadly classified into two categories:

- Block Storage Devices
- File Storage Devices

Block Storage Devices

The **block storage devices** offer raw storage to the clients. These raw storage are partitioned to create volumes.

File Storage Devices

The **file Storage Devices** offer storage to clients in the form of files, maintaining its own file system. This storage is in the form of Network Attached Storage (NAS).

Cloud Storage Classes

Cloud storage can be broadly classified into two categories:

- Unmanaged Cloud Storage
- Managed Cloud Storage

Unmanaged Cloud Storage

Unmanaged cloud storage means the storage is preconfigured for the customer. The customer can neither format, nor install his own file system or change drive properties.

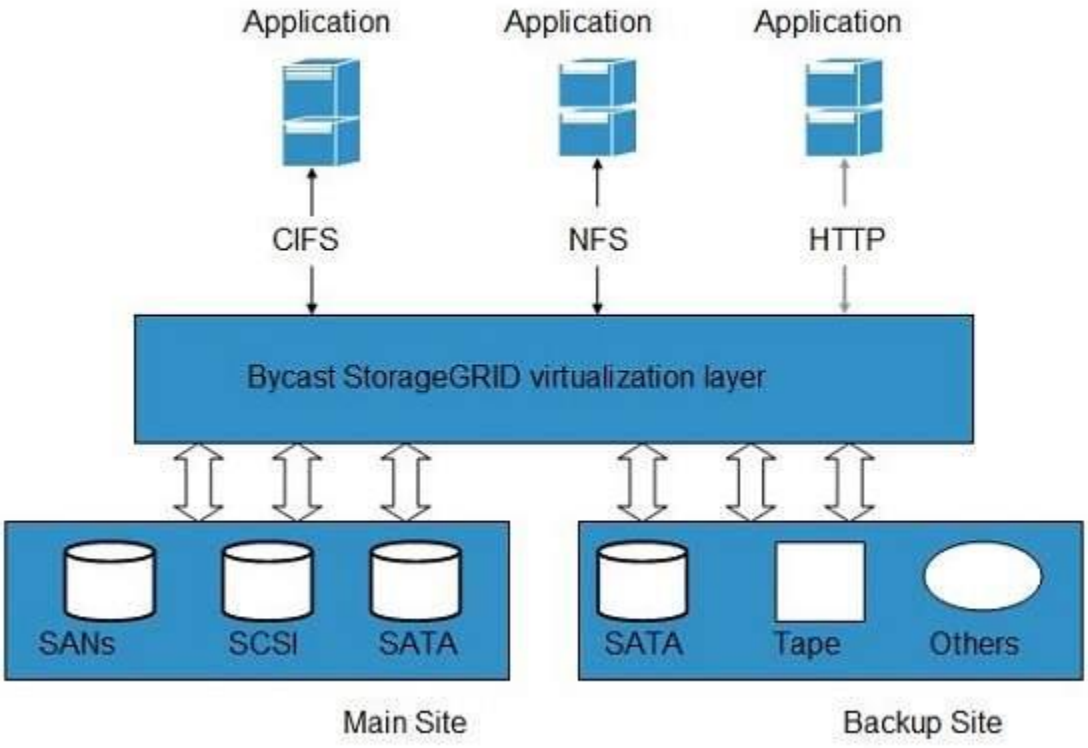
Managed Cloud Storage

Managed cloud storage offers online storage space on-demand. The managed cloud storage system appears to the user to be a raw disk that the user can partition and format.

Creating Cloud Storage System

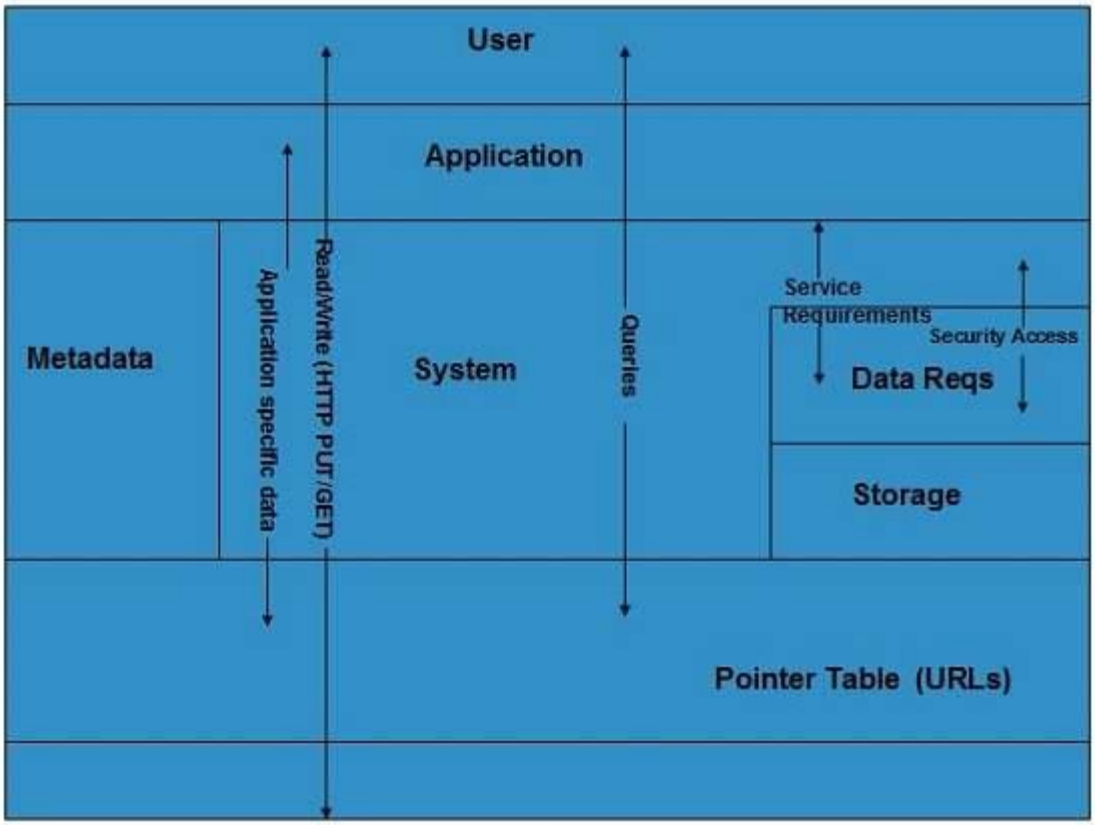
The cloud storage system stores multiple copies of data on multiple servers, at multiple locations. If one system fails, then it is required only to change the pointer to the location, where the object is stored.

To aggregate the storage assets into cloud storage systems, the cloud provider can use storage virtualization software known as **StorageGRID.** It creates a virtualization layer that fetches storage from different storage devices into a single management system. It can also manage data from **CIFS** and **NFS** file systems over the Internet. The following diagram shows how StorageGRID virtualizes the storage into storage clouds:



Virtual Storage Containers

The **virtual storage containers** offer high performance cloud storage systems. **Logical Unit Number (LUN)** of device, files and other objects are created in virtual storage containers. Following diagram shows a virtual storage container, defining a cloud storage domain:


User
Application
Metadata
Application specific data
Read/Write (HTTP PUT/GET)
System
Queries
Service Requirements
Security Access
Data Reqs
Storage
Pointer Table (URLs)

Why Cloud Computing Matters

Why is cloud computing important? There are many implications of cloud technology, for both developers and end users.

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

For end users, cloud computing offers all those benefits and more. A person using a web-based application isn't physically bound to a single PC, location, or network. His applications and documents can be accessed wherever he is, whenever he wants. Gone is the fear of losing data if a computer crashes. Documents hosted in the cloud always exist, no matter what happens to the user's machine. And then there's the benefit of group collaboration. Users from around the world can collaborate on the same documents, applications, and projects, in real time. It's a whole new world of collaborative computing, all enabled by the notion of cloud computing.

And cloud computing does all this at lower costs, because the cloud enables more efficient sharing of resources than does traditional network computing. With cloud computing, hardware doesn't have to be physically adjacent to a firm's office or data center. Cloud infrastructure can be located anywhere, including and especially areas with lower real estate and electricity costs. In addition, IT departments don't have to engineer for peak-load capacity, because the peak load can be spread out among the external assets in the cloud. And, because additional cloud resources are always at the ready, companies no longer have to purchase assets for infrequent intensive computing tasks. If you need more processing power, it's always there in the cloud—and accessible on a cost-efficient basis.

Cloud Computing Advantages

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

- **Cost efficiency –** The biggest reason behind shifting to cloud computing is that it takes considerably lesser cost than an on-premise technology. Now the companies need not store the data in disks anymore as the Cloud offers enormous space available saving money and resources of the companies.

- **High Speed –** Cloud computing lets you deploy the service quickly in fewer clicks. This quick deployment lets you get the resources required for your system within fewer minutes.

- **Excellent accessibility –** Storing the information in cloud allows you to access it anywhere, anytime regardless of the machine making it highly accessible and flexible technology of present times.

- **Back-up and restore data –** Once the data is stored in Cloud, it is easier to get the back-up and recovery of that, which is quite a time taking process on-premise.

Cloud Computing Disadvantages

Every technology has positive and negatives aspects that are highly important to discuss before implementing it. Aforementioned points highlight the benefits of using cloud technology and following discussion will outline the potential cons of Cloud Computing-

- **Security issues –** At time storing data in cloud may pose serious challenge of information theft in front of the company. Though advanced security measures are deployed on cloud, still storing a confidential data in cloud can be a risky affair.

- **Low bandwidth –** At times the bandwidth is low as many users are accessing cloud at the same time which causes its bandwidth to go down. With less speed the benefits of cloud computing cannot be realized.

- **Flexibility issues –** The cloud services run on remote servers which make it hard for the companies to have control over software and hardware. The services at times do not run the way it should.

- **Incompatibility –** Since entire infrastructure gets virtualized, incompatibility issues may arise at times that may pose serious challenges on the way of smooth running of services.

Companies in the cloud today

Microsoft Azure IaaS and PaaS computing for development, deployment, and managemen
"The whole of Microsoft has been re-tooled to cloud first. I can't say that for any other organization."

- Founder and CEO, Phidiax

Made for: enterprise clients familiar with Microsoft products, robust development and deployment.
Not for: managed cloud, or those unfamiliar with Microsoft products.

Amazon Web Services Highly scalable, complete cloud platform
"The whole of Microsoft has been re-tooled to cloud first. I can't say that for any other organization."

- Founder and CEO, Phidiax

Made for: enterprise clients familiar with Microsoft products, robust development and deployment.
Not for:managed cloud, or those unfamiliar with Microsoft products.

Google Cloud Platform Developer products and cloud technologies hosted by Google
"Two words: "power" and "simplicity", were the ultimate combination that has pushed our company towards Google's Clou..."
Made for: developers seeking a streamlined cloud ecosystem for development and deployment.
Not for: user seeking a managed cloud platform, simple cloud-based tasks.

IBM Cloud Multifaceted cloud offerings for business operations
"IBM cloud provides more scalability and reliability with high-end security. It is used for data storage..."
Made for: business users seeking, bare metal servers, infrastructure, and analytics tools.
Not for: minimal cloud functions, simple storage and computing needs.

Rackspace Dedicated servers and infrastructure services
"Rackspace handles so much that it's hard to imagine our business running smoothly without them..."
Made for: *Made for:* ***powerful managed hosting with various managed services.***
Not for:complete cloud ecosystems, interconnected PaaS and IaaS.

VMware Virtualization public cloud with hybrid capabilities
"VMware has the best 3rd party applications and support, and is the gold standard in virtualization."
Made for:ongoing testing and deployment of applications, users familiar with vSphere.
Not for:simple, private computing platform, or those without a need for virtualization.

Red Hat Open cloud technology for enterprise clients
"Red hat was chosen for the cost and flexibility. It's used for security of data..."
Made for: users preferring open-source networks and Linux systems.
Not for: turnkey service with a complete set of cloud services and tools.

Oracle Cloud Integrated public cloud for business applications
"Oracle cloud has increased business agility, lowered costs and reduced complexity..."

Made for:enterprise-grade cloud computing capabilities, including SaaS, IaaS, and PaaS.
Not for: single-user or small business user client, simple turnkey cloud solution.

CloudSigma International IaaS provider with hosting and computing features
"I'm glad that CloudSigma exists because they protected my data for a long time, which also benefited my organization..."
Made for: developers seeking a robust cloud environment with PaaS capabilities
Not for: simple website services and hosting tools.

DigitalOcean Simplified cloud servers for development
"It has been great so far, we expect they will grow even bigger."
Made for: users seeking a simple, robust cloud server platform with high scalability

Cloud Computing Services

Cloud computing is based on Service model.

Categories of service model

The service models are categorized into three basic models:

1)Software-as-a-Service(SaaS)
2)Platform-as-a-Service(PaaS)
3)Infrastructure-as-a-Service(IaaS)

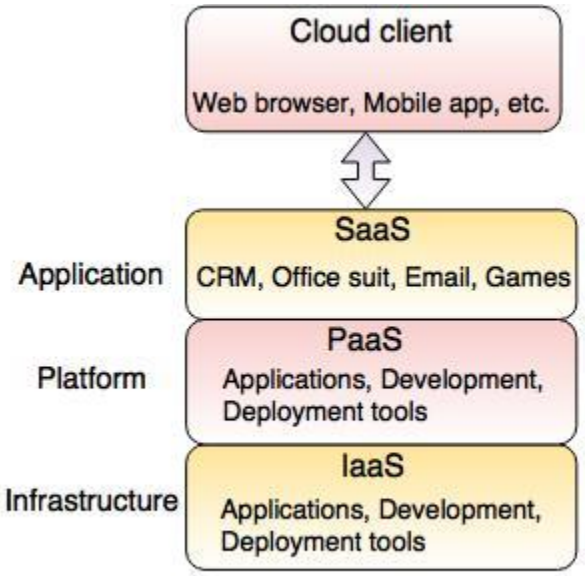


Fig. - Categories of Cloud Computing

1) Software-as-a-Service (SaaS)

- SaaS is known as **'On-Demand Software'.**
- It is a software distribution model. In this model, the applications are hosted by a cloud service provider and publicized to the customers over internet.
- In SaaS, associated data and software are hosted centrally on the cloud server.
- User can access SaaS by using a thin client through a web browser.
- CRM, Office Suite, Email, games, etc. are the software applications which are provided as a service through Internet.
- The companies like Google, Microsoft provide their applications as a service to the end users.

Advantages of SaaS

- SaaS is easy to buy because the pricing of SaaS is based on monthly or annual fee and it allows the organizations to access business functionalities at a small cost, which is less than licensed applications.
- SaaS needed less hardware, because the software is hosted remotely, hence organizations do not need to invest in additional hardware.
- Less maintenance cost is required for SaaS and do not require special software or hardware versions.

Disadvantages of SaaS

- SaaS applications are totally dependent on Internet connection. They are not usable without Internet connection.
- It is difficult to switch amongst the SaaS vendors.

2) Platform-as-a-Service (PaaS)

- PaaS is a programming platform for developers. This platform is generated for the programmers to create, test, run and manage the applications.
- A developer can easily write the application and deploy it directly into PaaS layer.
- PaaS gives the runtime environment for application development and deployment tools.
- Google Apps Engine(GAE), Windows Azure, SalesForce.com are the examples of PaaS.

Advantages of PaaS

- PaaS is easier to develop. Developer can concentrate on the development and innovation without worrying about the infrastructure.

- In PaaS, developer only requires a PC and an Internet connection to start building applications.

Disadvantages of PaaS

- One developer can write the applications as per the platform provided by PaaS vendor hence the moving the application to another PaaS vendor is a problem.

3) Infrastructure-as-a-Service (IaaS)

- IaaS is a way to deliver a cloud computing infrastructure like server, storage, network and operating system.
- The customers can access these resources over cloud computing platform i.e Internet as an on-demand service.
- In IaaS, you buy complete resources rather than purchasing server, software, datacenter space or network equipment.
- IaaS was earlier called as Hardware as a Service(HaaS). It is a Cloud computing platform based model.
- HaaS differs from IaaS in the way that users have the bare hardware on which they can deploy their own infrastructure using most appropriate software.

Advantages of IaaS

- In IaaS, user can dynamically choose a CPU, memory storage configuration according to need.
- Users can easily access the vast computing power available on IaaS Cloud platform.

Disadvantages of IaaS

- IaaS cloud computing platform model is dependent on availability of Internet and virtualization services.

Characteristics of Cloud Computing:

1. Resources Pooling

It means that the **Cloud provider** pulled the computing resources to provide services to multiple customers with the help of a multi-tenant model. There are different physical and virtual resources assigned and reassigned which depends on the demand of the customer. The customer generally has no control or information over the location of the provided resources but is able to specify location at a higher level of abstraction

2. On-Demand Self-Service

It is one of the important and valuable features of Cloud Computing as the user can continuously monitor the server uptime, capabilities, and allotted network storage. With this feature, the user can also monitor the computing capabilities.

3. Easy Maintenance

The servers are easily maintained and the downtime is very low and even in some cases, there is no downtime. Cloud Computing comes up with an update every time by gradually making it better. The updates are more compatible with the devices and perform faster than older ones along with the bugs which are fixed.

4. Large Network Access

The user can access the data of the cloud or upload the data to the cloud from anywhere just with the help of a device and an internet connection. These capabilities are available all over the network and accessed with the help of internet.

5. Availability

The capabilities of the Cloud can be modified as per the use and can be extended a lot. It analyzes the storage usage and allows the user to buy extra **Cloud storage** if needed for a very small amount.

6. Automatic System

Cloud computing automatically analyzes the data needed and supports a metering capability at some level of services. We can monitor, control, and report the usage. It will provide transparency for the host as well as the customer.

7. Security

Cloud Security, is one of the best features of cloud computing. It creates a snapshot of the data stored so that the data may not get lost even if one of the servers gets damaged. The data is stored within the storage devices, which cannot be hacked and utilized by any other person. The storage service is quick and reliable.

8. Pay as you go

In cloud computing, the user has to pay only for the service or the space they have utilized. There is no hidden or extra charge which is to be paid. The service is economical and most of the time some space is allotted for free.

9. Measured Service

Cloud Computing resources used to monitor and the company uses it for recording. This resource utilization is analyzed by supporting charge-per-use capabilities. This means that the resource usages which can be either virtual server instances that are running in the cloud are getting monitored measured and reported by the service provider. The model pay as you go is variable based on actual consumption of the manufacturing organization.

Module II: Developing Cloud Services

Web-Based Application – Pros and Cons of Cloud Service Development – Types of Cloud Service Development – Software as a Service – Platform as a Service – Web Services – On-Demand Computing – Discovering Cloud Services Development Services and Tools – Amazon Ec2 – Google App Engine – IBM Clouds.

Why Develop Web-Based Applications?

The needs of a typical IT department are daunting: They must deliver adequate computing power and data storage to all users within the company. This must be done, of course, within a set budget, and there is the rub; to meet peak needs or to add capacity for new users can often send an IT budget soaring.

For most companies, it is not financially prudent to add capacity that will be used only a small percentage of the time. What the IT department needs is a way to increase capacity or add capabilities without investing in new servers and networking gear, or licensing new software. It is to this need that cloud computing speaks.

Cloud services, in the form of centralized web-based applications, also appeal to the IT professional. One instance of an application hosted in the cloud is cheaper and easier to manage than individual copies of similar software installed on each user's desktop PC. Upgrading a cloud app only has to be done one time, where upgrading traditional software has to be done for each PC on which that software is installed. Then, of course, we have the promise of cloud-enabled collaboration, which just can't be done with traditional desktop apps.

The advantages of cloud services development are particularly notable to smaller businesses, who otherwise wouldn't have the budget or resources to develop large-scale applications. By hosting locally developed web applications within the cloud, the small business avoids the cost of purchasing expensive hardware to host similar software.

Let's face it, most small companies don't have the staff, resources, hardware, or budget to develop and maintain their own applications, or to deal with the rigors of maintaining secure environments. Although they could outsource their software development and hosting, moving those applications to the cloud, companies don't have to invest in locally hosted systems, freeing up their staff and resources to focus on the day-to-day running of their own businesses.

In short, there's a lot to be gained by investing in cloud services development. A company that develops its own web-based applications gains functionality while reducing expenses. The combined power of the cloud is accompanied by lower software purchase and management costs.

The Pros and Cons of Cloud Service Development

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

Advantages of Cloud Development

One of the underlying advantages of cloud development is that of economy of scale. By taking advantage of the infrastructure provided by a cloud computing vendor, a developer can offer better, cheaper, and more reliable applications than is possible within a single enterprise. The application can utilize the full resources of the cloud, if needed—without requiring a company to invest in similar physical resources.

Speaking of cost, because cloud services follow the one-to-many model, cost is significantly reduced over individual desktop program deployment. Instead of purchasing or licensing physical copies of software programs (one for each desktop), cloud applications are typically "rented," priced on a per-user basis. It's more of a subscription model than an asset purchase (and subsequent depreciation) model, which means there's less up-front investment and a more predictable monthly expense stream.

IT departments like cloud applications because all management activities are managed from a central location rather than from individual sites or workstations. This enables IT staff to access applications remotely via the web. There's also the advantage of quickly outfitting users with the software they need (known as "rapid provisioning), and adding more computing resources as more users tax the system (automatic scaling). When you need more storage space or bandwidth, companies can just add another virtual server from the cloud. It's a lot easier than purchasing, installing, and configuring a new server in their data center.

For developers, it's also easier to upgrade a cloud application than with traditional desktop software. Application features can be quickly and easily updated by upgrading the centralized application, instead of manually upgrading individual applications located on each and every desktop PC in the organization. With a cloud service, a single change affects every user running the application, which greatly reduces the developer's workload.

Disadvantages of Cloud Development

Perhaps the biggest perceived disadvantage of cloud development is the same one that plagues all web-based applications: Is it secure? Web-based applications have long been considered potential security risks. For this reason, many businesses prefer to keep their applications, data, and IT operations under their own control.

That said, there have been few instances of data loss with cloud-hosted applications and storage. It could even be argued that a large cloud hosting

Types of Cloud Service Development

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

Software as a Service

Software as a service, or SaaS, is probably the most common type of cloud service development. With SaaS, a single application is delivered to thousands of users from the vendor's servers. Customers don't pay for owning the software; rather, they pay for using it. Users access an application via an API accessible over the web.

Each organization served by the vendor is called a tenant, and this type of arrangement is called a multitenant architecture. The vendor's servers are *virtually partitioned* so that each organization works with a customized virtual application instance.

For customers, SaaS requires no upfront investment in servers or software licensing. For the application developer, there is only one application to maintain for multiple clients.

note An API (application development interface) is an interface that enables a remote program to communicate or use the resources of another program or service.

Platform as a Service

In this variation of SaaS, the development environment is offered as a service. The developer uses the "building blocks" of the vendor's development environment to create his own custom application. It's kind of like creating an application using Legos; building the app is made easier by use of these predefined blocks of code, even if the resulting app is somewhat constrained by the types of code blocks available.

Web Services

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

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

A good example of web services are the "mashups" created by users of the Google Maps API. With these custom apps, the data that feeds the map is provided by the developer, where the engine that creates the map itself is provided by Google. The developer doesn't have to code or serve a map application; all he has to do is hook into Google's web API.

As you might suspect, the advantages of web services include faster (and lower-cost) application development, leaner applications, and reduced storage and bandwidth demands.

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

On-Demand Computing

As the name implies, on-demand computing packages computer resources (processing, storage, and so forth) as a metered service similar to that of a

Discovering Cloud Services Development Services and Tools

As you're aware, cloud computing is at an early stage of its development. This can be seen by observing the large number of small and start-up companies offering cloud development tools. In a more established industry, the smaller players eventually fall by the wayside as larger companies take center stage.

That said, cloud services development services and tools are offered by a variety of companies, both large and small. The most basic offerings provide cloud-based hosting for applications developed from scratch. The more fully

IBM

It's not surprising, given the company's strength in enterprise-level computer hardware, that IBM is offering a cloud computing solution. The company is targeting small- and medium-sized businesses with a suite of cloud-based on-demand services via its Blue Cloud initiative.

Blue Cloud is a series of cloud computing offerings that enables enterprises to distribute their computing needs across a globally accessible resource grid. One such offering is the Express Advantage suite, which includes data backup and recovery, email continuity and archiving, and data security functionality—some of the more data-intensive processes handled by a typical IT department.

To manage its cloud hardware, IBM provides open source workload-scheduling software called Hadoop, which is based on the MapReduce software used by

Module III: Cloud Computing Security Architecture

Cloud security fundamentals-Vulnerability assessment tool for cloud- Privacy and Security in cloud. Cloud computing security architecture: Architectural Considerations- General Issues-Trusted Cloud computing- Secure Execution Environments and Communications-Micro-architectures; Identity Management and Access control Identity management-Access control, Autonomic Security

Cloud security Fundamentals

Cloud security consists of the practices and technology that protect cloud computing environments from both external and internal cybersecurity threats. Cloud computing, which is the delivery of IT services over the internet, has become a mainstay for modern businesses and governments. To keep data and applications in the cloud secure from current and emerging threats, security solutions need to be put in place that prevent unauthorized access, along with best practices for managing those security resources.

Cloud security differs based on the category of cloud computing being used. There are three main categories of cloud computing:

- **Public cloud services, operated by a public cloud provider** — These include software-as-a-service (SaaS), infrastructure-as-a-service (IaaS), and platform-as-a-service (PaaS).
- **Private cloud services, operated by a public cloud provider** — These services provide a computing environment dedicated to one customer, operated by a third party.
- **Private cloud services, operated by internal staff** — These services are an evolution of the traditional data center, where internal staff operates a virtual environment they control, and provide access to employees over the internet. Many organizations are moving to a software-defined data center (SDDC) to accomplish this.

When using a cloud computing service operated by a public cloud provider, data is being stored with a third party, which marks a fundamental difference between cloud computing and traditional IT, where most data was held within a self-controlled network. Understanding the separation of data from owner to public cloud provider is the first step to building a cloud security strategy.

Vulnerability assessment tool for cloud

A vulnerability assessment is the process of defining, identifying, classifying and prioritizing vulnerabilities in computer systems, applications and network infrastructures and providing the organization doing the assessment with the necessary knowledge, awareness and risk background to understand the threats to its environment and react appropriately.

vulnerability assessment process that is intended to identify threats and the risks they pose typically involves the use of automated testing tools, such as network security scanners, whose results are listed in a vulnerability assessment report.

Organizations of any size, or even individuals who face an increased risk of cyberattacks, can benefit from some form of vulnerability assessment, but large enterprises and other types of organizations that are subject to ongoing attacks will benefit most from vulnerability analysis.

Because security vulnerabilities can enable hackers to access IT systems and applications, it is essential for enterprises to identify and remediate weaknesses before they can be exploited. A comprehensive vulnerability assessment along with a management program can help companies improve the security of their systems.

Importance of vulnerability assessments

A vulnerability assessment provides an organization with information on the security weaknesses in its environment and provides direction on how to assess the risks associated with those weaknesses and evolving threats. This process offers the organization a better understanding of its assets, security flaws and overall risk, reducing the likelihood that a cybercriminal will breach its systems and catch the business off guard.

Types of vulnerability assessments

Vulnerability assessments depend on discovering different types of system or network vulnerabilities, which means the assessment process includes using a variety of tools, scanners and methodologies to identify vulnerabilities, threats and risks.

Some of the different types of vulnerability assessment scans include the following:

- Network-based scans are used to identify possible network security attacks. This type of scan can also detect vulnerable systems on wired or wireless networks.
- Host-based scans are used to locate and identify vulnerabilities in servers, workstations or other network hosts. This type of scan usually examines ports and services that may also be

visible to network-based scans, but it offers greater visibility into the configuration settings and patch history of scanned systems.

- Wireless network scans of an organization's Wi-Fi networks usually focus on points of attack in the wireless network infrastructure. In addition to identifying rogue access points, a wireless network scan can also validate that a company's network is securely configured.
- Application scans can be used to test websites in order to detect known software vulnerabilities and erroneous configurations in network or web applications.
- Database scans can be used to identify the weak points in a database so as to prevent malicious attacks, such as SQL injection attacks.

Privacy and Security in cloud

Cloud technology has given opportunities to many businesses to showcase their potential in the business world. SMEs are not only getting an opportunity to grow, they are also taking their business operations to the next level. ***Cloud technology*** has opened a door for small & medium scale companies to acquire market share by entering the yard of bigger players. As the business requirements have become on-demand and need-based, it gave many companies a significant edge and allow them to complete in a much larger business space.

Cloud technology provides various advantages. Starting from data management, data storage, 0% downtime, CRM management, resource optimization to entire business automation. It also reduces a high amount of investment and saves a lot of time.

At the same time, cloud computing has raised multiple eyebrows with IT management, especially when it comes to ***data security in the cloud computing***. ***Data security*** and privacy protection are two major factors. These two factors are becoming more important for the future development of cloud computing technology in business, industry, and government. While addressing this fear, Google claimed that data stored in the cloud are much safer.

What are the Challenges?

Data Replication

Every business faces this challenge. Snapshots and data backups are taken on a daily basis. They automatically stored in the cloud. Are you aware where they have been stored and who can see and access them? Can you identify and control unauthorised copying of your data?

Data Loss

Data loss can be a disaster for any business. Virtual data can be easily lost or exposed as it moves between VMs or in the cloud. Are you sure that authorised users are accessing your data within

predefined policies? Do you have the authority to block any user who is violating data use policies?

New Class of Users

Cloud computing need cooperation between security, storage, application, and security admins. They all manage your sensitive business data. With more number of users, the risk also increases. If one admin went wrong, entire data in the system will be at risk.

Insecure APIs

Application Programming Interfaces (API) allow users to customize their cloud computing practices. APIs can be a threat to ***cloud security*** because of their nature. APIs give developers the tools to build solutions to integrate their applications with other software. The vulnerability of an API depends on the communication that takes place between applications. While this can help developers and businesses, they also issue serious security concerns.

Internal Threat

Never keep this point out of your mind. You may be thinking data is safe inside. But this is one of the biggest challenge company's face. Employees can use their access to an organisation's cloud-based services to misuse or access information related to finance, customer details etc.

How to Protect your Data?

You can protect your business data in the cloud from unauthorised access. All you need is a sharp eye and an extra effort. Here are few practical tips to keep your cloud data safe and secure.

Architectural Considerations for Cloud Security

As a CIO, one needs to understand the various factors that affect the implementation & performance of Cloud Security architecture. General issues involving regulatory requirements, standards compliance, security management, information classification and security awareness need to be considered along with more specific architectural related issues - trusted hardware & software, secure execution environment, secure communications and hardware augmentation using micro-architectures. These architectural issues are elaborated below.

General ISSUES

There are many factors which affect the performance and implementation of cloud security architecture. There are many general issues in cloud security architecture such as security management, security compliance, controls, security awareness and cloud administrative issues. In addition to general issues there are more specific issues such as trusted hardware and software which provide secure environment for establish a secure connection and application execution platform.

A. Compliance: The service provider in public cloud environment normally does not provide the information about the location of user's data stored to the user. The user of public cloud does not have knowledge of where its data is stored. So Cloud service provider must ensure the security of user data through some compliance certificate issued by cloud service provider.

B. Security Management: Security management is the part of security architecture. Cloud service provider builds trust through security management. Security management is the tool to securely manage the data of cloud user in best possible way. Cloud security management should have the ability to identify and address the issues related to access control, vulnerability analysis, change control, incident response, fault tolerance, and disaster recovery and business continuity planning. Trusted cloud computing- Trusted cloud computing is a type of computer security architecture. It is designed to protect cloud from attackers and hackers and ensure that cloud resources work properly when requested from customers. A trusted cloud computing system has the ability to protect data used by hypervisors and applications. It can also protect against unauthorized access to information and provide strong authentication, apply encryption to protect sensitive data that resides on stolen or lost devices. It also support compliance through hardware and software mechanisms.

C. Secure execution environment: In a cloud computing environment, there are many applications which run on different servers in a distributed mode. These applications interact with the outside world and other applications and may contain sensitive information. The inappropriate access of this sensitive information would be harmful to a client. In addition, cloud computing is increasingly being used to manage and store large amounts of data in database centres located

at different places. Therefore, it is extremely important for the cloud vendor to provide a secure execution environment and secure communications for client applications and storage. **D. Identity management and access control**: Identification and authentication are the most important access control systems. Identification means provision to identify a valid user usually with help of a username or user logon ID to the system. For identity management following methods can be applied a. Finger print scan b. Retina Scan c. Iris Scan d. Hand Geometry e. Voice f. Handwritten signature dynamics

E. Secure Communications: The application and data moves from from clout to outside in public cloud and within cloud in private cloud, therefore movement of data should be secured. Secure cloud communications involves the structures, transmission methods, transport formats, and security measures that provide confidentiality, integrity, availability, and authentication for transmissions over private and public communications networks.

F. API: Common vulnerabilities such as weak antivirus software, unattended computing platforms, poor passwords, weak authentication mechanisms, and inadequate intrusion detection that can impact communications must be more stringently analyzed, and proper APIs must be used

Trusted Computing:

Trusted cloud computing protects cloud systems from malicious intrusions, attacks, protects data in use by hypervisors & applications, provides for strong authentication, applies encryption to protect sensitive data & supports compliance through hardware & software. Protection domains are the execution & memory space assigned to each concurrent running process & they protect the programs from all unauthorized modification or executional interference. A trusted computing base is the total combination of protection mechanism including hardware, software & firmware that are trusted to enforce an organization's security policy.

Secure Execution Environment:

In the cloud, applications run on different servers in a distributed mode and may contain sensitive data. The cloud service provider should have a secure execution environment that enables protected data transfers via strong authentication mechanisms and clients must implement best practices to address privacy & confidentiality of information exchange.

Secure Communications:

Organizations should reevaluate their communications security policies once they move to the cloud as the cloud brings about newer challenges in this area. The communications referred here are both - data in motion & data at rest! Secure cloud communication revolves around structures, transmission methods, transport formats, security measures to provide confidentiality (network security protocols, authentication, data encryption), data integrity (firewall, communications security & intrusion detection), availability (logins, fault tolerance, backups, redundancy) & authentication for transmissions over public & private networks.

Micro-architectures: Micro-architectures can be designed as hardware accelerators for functions such as encryption, arithmetic functions & to secure web transactions to support cloud computing. Micro-architecture designs may include concepts related to Pipelining to increase performance by overlapping steps of different instructions, super-scalar processor to enable concurrent execution of multiple instructions and Very-Long Instruction Word Processing (VLIW) to specify a more than one concurrent operations in a single instruction.

To summarize, cloud computing security architecture is a critical element in establishing trust amongst users of the Cloud Services!

Identity Management?

There’s been a movement in the past few years to innovate in the identity management space. Most of that innovation has been around cloud identity management – shifting the process of authentication and authorization to the cloud. Unfortunately, the initial take on this was effectively single sign-on to web applications or what many have called IDaaS

The main goal of identity management is to ensure that only authenticated users are granted access to the specific applications, systems or IT environments for which they are authorized. This includes control over user provisioning and the process of onboarding new users such as employees, partners, clients and other stakeholders. Identity management also includes control over the process of authorizing system or network permissions for existing users and the offboarding of users who are no longer authorized to access organization systems.

Importance of identity management

Identity management is an important part of the enterprise security plan, as it is linked to both the security and productivity of the organization.

Business benefits of identity management

In addition to managing employees, the use of identity management along with access management enables a business to manage customer, partner, supplier and device access to its systems while ensuring security is the top priority.

Access Control

Access control is generally a policy or procedure that allows, denies or restricts access to a system. It may, as well, monitor and record all attempts made to access a system. Access Control may also identify users attempting to access a system unauthorized. It is a mechanism which is very much important for protection in computer security. Various access control models are in use, including the most common Mandatory Access Control (MAC), Discretionary Access Control (DAC) and Role Based Access Control (RBAC). All these models are known as identity based access control models. In all these access control models, user (subjects) and resources (objects) are identified by unique names. Identification may be done directly or through roles assigned to the subjects. These access control methods are effective in unchangeable distributed system, where there are only a set of Users with a known set of services.

ACCESS CONTROL METHODS: The first way a system provides security to its resources and data, is by controlling access to the resources and the system itself. However, access control is more than just controlling which users (subjects) can access which computing and network resources. In addition, access control manages users, files and other resources. It controls user's privileges to files or resources (objects). In access control systems various steps like, identification, authentication, authorization and accountability are taken before actually accessing the resources or the object in general.

What Is Access Control in Cloud Security?

Access Control in cloud security is a system with which a company can regulate and monitor permissions, or access to their business data by formulating various policies suited chosen by the company. Access control in cloud security helps companies gain macro-level visibility into their data and user behavior, which a cloud app may not be able to offer, given their on-demand services and mobility.

Today, data is the most valuable asset of a company, safeguarding it is the next thing to do! Access Control in cloud computing gives companies the control to restrict unauthorized user access and, at the same time, give enough access for smooth functioning at work.

Cloud Codes Access Control in cloud security lets companies formulate policies to restrict access through specific IP addresses, browsers, devices, and during specified time shifts. Here's an in-depth view of our Access Control in cloud computing solution.

Autonomic Security

Introduction

Resource monitoring in cloud computing environment is an important part of resource management of cloud computing platform. It provides the basis for resource allocation, task scheduling, and load balancing. With the extensive use of cloud computing services, users have made increasing demands on the security of cloud computing. Since the cloud computing environment has the characteristics of transparent virtualization and resource flexibility, it is infeasible for a traditional security program to protect the data security in the cloud platform, which hinders further development and application of cloud computing [1]. Therefore, it is of critical importance to develop new tools suitable for monitoring cloud platform data. However, the collection, transmission, storage, and analysis of a large number of monitored data will bring huge resource overhead, directly affecting system performance, timely detection of anomalies, and pinpoint accuracy of problem. In addition, because cloud computing is essentially developed on the basis of current technology, the existing security vulnerabilities will be inherited directly to the cloud computing platform, which may even bring greater security threat. It can be seen that, in the cloud computing environment, users basically lost the control of private information

and data, which triggered a series of security challenges, such as cloud data storage location, data encryption mechanism, data recovery mechanism, integrity protection, third-party supervision and auditing, virtual machine security, and memory security. At present, there is not enough research on cloud computing resource monitoring, but there are a lot of researches on distributed computing and grid computing, for instance, DRMonitor [2], Ganglia [3], and MDS (Monitoring and Discovery System) [4]. They play important roles in distributed systems or grid systems. However, if the above methods are applied directly in the cloud computing environment, there will be some shortcomings. On the one hand, the resource in the cloud computing environment is highly virtualized and flexible. Moreover, cloud computing provides services such as IaaS, PaaS, and SaaS, in addition to monitoring the resources of the physical server [5]. Users need to monitor the virtual machine running on it. On the other hand, cloud computing is a business model, and the cloud service provider will charge the user for usage accordingly. Monitoring information in existing resource monitoring system is not fine granularity, so it is unable to get to the process level of information and track consumption of CPU, memory, storage and other resources in real time during the user task execution process. Cloud computing environment is dynamic, random, complex, and open. Cloud providers need to collect user-related fees based on resource usage; as a result, original resource monitoring methods cannot fully meet the requirements of the cloud computing environment. Therefore, according to the characteristics of cloud computing itself, some resource monitoring methods for current distributed computing, and grid computing, cannot fully adapt to the cloud computing environment.

In order to adapt to the cloud computing environment, combining with abnormal data mining algorithm, we propose a data monitoring method under cloud environment based on autonomic computing model. In order to address the security challenges for data on the cloud platform, the model uses autonomic computing mechanism and the abnormal data mining idea to transmit the monitoring information to each other. The model is mainly composed of five modules: network monitoring module, data analysis module, response strategy module, system implementation module, and knowledge base. In the network monitoring module, the system gathers the data by collecting the data stream and generates the original data. In addition, through the data preprocessing mechanism, the original data are formatted. The data analysis module evaluates these processed data, extracts useful data from it to determine whether they are abnormal, and then feeds the analysis result back to the response strategy module to adjust the monitoring period. The data collection and analysis of storage are the core parts of this model, which provide users with essential data monitoring information. In the local computer deployment monitoring framework, the cloud is connected to data monitoring. Our contributions are as follows:(i)We propose a safe and effective model that enables the data on the cloud platform to be monitored in time, and the system adjusts the monitoring cycle to autonomously protect the data.(ii)We design a data mining algorithm, in which, based on an improved chaotic algorithm, data mining method was proposed for the frequently appearing abnormal data in the cloud computing environment. We also design and implement abnormal behavior detection based on the Poisson process to obtain accurate test results.(iii)We formally analyze the capability of abnormal behavior monitoring and implement all of these data security monitoring models based on autonomic computing. A large number of experiments are carried out in the simulation environment using prepared dataset, and the results show that our system achieves the desired goals.

This paper is organized as follows. Section 2 states the origin of autonomic computing theory, and its related work. Section 3 analyzes how to establish a security monitoring model based on autonomic computing for the cloud platform. Then we analyze the existing security model and safety monitoring method of autonomic computing to the cloud platform oriented metrics and the calculation method. Section 4 analyzes method of simulation and experiment and presents a security monitoring model based on autonomic computing analysis for the cloud platform. Section 5 gives the summary and points out the future research directions.

Module IV: Cloud Computing For Everyone

Centralizing Email Communications – Collaborating on Schedules – Collaborating on To-Do Lists – Collaborating Contact Lists – Cloud Computing for the Community – Collaborating on Group Projects and Events – Cloud Computing for the Corporation.

Centralizing Email Communications

Cloud computing for families by examining how a typical family an use cloud-based tools to help improve communications between family members. The key here is to enable anywhere/anytime access to email. Precloud computing, your email access was via a single computer, which also stored all your email messages. For this purpose, you probably used a program like Microsoft Outlook or Outlook Express, installed on your home computer. If you wanted to check your home email from work, it took a bit of juggling and perhaps the use of your ISP's email access web page. That web page was never in sync with the messages on your home PC, of course, which is just the start of the problems with trying to communicate in this fashion. A better approach is to use a web-based email service, such as Google's Gmail (mail.google.com), Microsoft's Windows Live Hotmail (mail.live.com), or Yahoo! Mail (mail.yahoo.com). These services place your email inbox in the cloud; you can access it from any computer connected to the Internet.

You can check your web based email whether you're in the office or on the road. Just make sure you're connected

to the Internet, and then open your web browser and log in to the Gmail or Windows Live Hotmail or Yahoo! Mail website. Go to your inbox and you'll find your spouse's message; reply as necessary and await your spouse's response. Even if you change

locations or computers, your spouse's message remains in your inbox, and your reply remains in your sent messages folder.

Collaborating on Schedules

It comes to coordinating multiple individuals or families in a community activity; you have your work cut out for you. Whether it's a youth sports team, community organization, school event, or some community event, trying to line up who's free and who's not on a given evening takes a lot of effort—unless, that is, you're using web-based scheduling tools.

One of the most common community activities: youth sports. whatever the sport, there's a lot of activities that need to be scheduled—practices, home games, away games, team meetings, you name it. Multiply the number of players (and coaches) on each team times the number of events, and you see the complexity; it gets even worse if you're trying to manage events for an entire league.

Collaborating on To-Do Lists

Let's start with simple task management, in the form of the old-fashioned todo list. These are web-based lists that multiple group members can access from any web browser. Tasks are entered (complete with due date) and checked off when completed.

(www.blablalist.com), Remember the Milk (www.rememberthemilk.com),

Ta-do List (www.tadalist.com), Tudu List (www.tudulist.com)

Collaborating Contact Lists

Salespeople have to deal with lots and lots of contacts. Not only is their address book full, they need to know when to contact certain clients, when follow-up calls are necessary, what the boss needs them to do today, and the like.

The solution, of course, is a web-based contact management or customer resource management (CRM) application. These programs are tailored to the needs of a busy salesperson and come complete with features such as activity scheduling, appointment

reminders, email templates, and the like. Among the most popular of these applications are Big Contacts

(www.bigcontacts.com), High rise (www.highrisehq.com), and the market leading Salesforce.com (www.salesforce.com).

Cloud computing for community

It has tremendous benefits for the entire community, from neighborhood groups to sports teams to school organizations. Any time any groups of people in the community need to communicate and collaborate; web-based applications are the way to go.

Communicating Across the Community

One of the key components of any community collaboration is communication. Many community activities are undertaken by people in their spare time—outside of normal work and home activities. Therefore, they might be communicating during office hours on their work computer, after hours on their home computer, or during any spare moment. Programs can be accessed from any computer connected to the Internet. You use your web browser to send and view email messages hosted on the web. You can send and receive messages at work, at home, or from wherever you happen to be. Everything you send and receive is stored in the cloud, accessible from anywhere at any time.

Collaborating on Group Projects and Events

Community groups often have a lot on their plates. Someone has to schedule the next fundraiser, someone else needs to print up flyers, someone else is in charge of recruiting new members…there's just a lot of stuff to do! How does your community group manage all these activities? In the new world of cloud computing, the best way is with a web-based application— which anyone in the group can access.

Collaborating on Task Management: For managing more complex tasks, a simple to-do list application might not cut the mustard. Instead, consider using a web-based task management application that lets you manage the multiple pieces and parts of large projects.

Collaborating on Event Management: When you're putting on a big event such as a concert or conference, you have a whole new set of challenges to face. Not only do you have to manage the tasks involved with putting together the event, you also have to handle
Attendee registration, event marketing, ticket sales, and the like. It's a massive effort— made somewhat easier by web-based event management tools.

Collaborating on Event Marketing: When it comes to promoting your community events, you want to go beyond the basics to more creative forms of marketing. For example, you may want to create a brochure or flyer to announce your event. Fine and dandy, but everybody in the group (includng all the community bigwigs) wants input on the final piece. This may have been difficult in precloud days, but now you can use a web-based application such as Google Docs (docs.google.com) to create your piece and make it available online for everyone to see and comment on.

Cloud Computing for the Corporation

Companies large and small recognize the cost savings and productivity enhancements of using web-based tools to manage projects, collaborate on documents and presentations, manage enterprise wide contacts and schedules, and the like.

Managing Schedules

Better way is web-based scheduling. Everyone places his or her schedule in the cloud, which then enables the meeting's organizer to easily see who's available when. The cloud-based app finds the best time for all involved and schedules the meeting. No more emails, no more phone calls; it all happens automatically, in the cloud.

Web-based scheduling programs let you schedule both in-person meetings and teleconferences with attendees from multiple locations. You're not limited to just those people located in your office; you can work with the schedules of people around the country and even in different firms.

Managing Projects

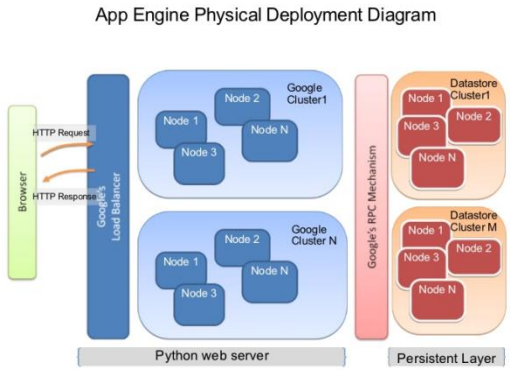
Most companies at one point or another have at least one big project going on—the type of project that involves multiple employees from multiple departments and perhaps multiple locations. Projects of this type have tons of individual pieces and parts, each of which dependent on the completion of a previous task.

Many project management applications include additional functions useful in the management of group projects. These features may include group to-do lists, web-based file sharing, message boards, time and cost tracking, and so on. And the most robust of these apps le ts you manage multiple projects simultaneously; users can schedule their time across multiple projects and make sure they're not doing two things at once.

Module V: Cloud Computing Case Studies

Cloud computing case studies: Google App Engine – IBM Clouds –Windows live – Micro soft dynamic CRM- Salesforce.com CRM- App Exchange – Amazon S3 – Oracle OBIEE.

Case Study – Google App Engine



Google App Engine (often referred to as GAE or simply App Engine, and also used by the acronym GAE/J) is a platform as a service (PaaS) cloud computing platform for developing and hosting web applications in Google-managed data centers. Applications are sandboxed and run across multiple servers. App Engine offers automatic scaling for web applications—as the number of requests increases for an application, App Engine automatically allocates more resources for the web application to handle the additional demand.

Google App Engine is free up to a certain level of consumed resources. Fees are charged for additional storage, bandwidth, or instance hours required by the application. It was first released as a preview version in April 2008, and came out of preview in September 2011.

Runtimes and frameworks

Currently, the supported programming languages are Python, Java (and, by extension, other JVM languages such as Groovy, JRuby, Scala, Clojure, Jython and PHP via a special version ofQuercus), and Go. Google has said that it plans to support more languages in the future, and that the Google App Engine has been written to be language independent.

Reliability and Support

All billed High-Replication Datastore App Engine applications have a 99.95% uptime SLA

Portability Concerns

Developers worry that the applications will not be portable from App Engine and fear being locked into the technology. In response, there are a number of projects to create open-source back-ends for the various proprietary/closed APIs of app engine, especially the datastore. Although these projects are at various levels of maturity, none of them is at the point where installing and running an App Engine app is as simple as it is on Google's service. AppScale and TyphoonAE are two of the open source efforts.
AppScale can run Python, Java, and Go GAE applications on EC2 and other cloud vendors. TyphoonAE can run python App Engine applications on any cloud that support linux machines.
Web2py web framework offers migration between SQL Databases and Google App Engine, however it doesn't support several App Engine-specific features such as transactions and namespaces.

Differences with other application hosting

Compared to other scalable hosting services such as Amazon EC2, App Engine provides more infrastructure to make it easy to write scalable applications, but can only run a limited range of applications designed for that infrastructure.
App Engine's infrastructure removes many of the system administration and development challenges of building applications to scale to hundreds of requests per second and beyond. Google handles deploying code to a cluster, monitoring, failover, and launching application instances as necessary.
While other services let users install and configure nearly any *NIX compatible software, App Engine requires developers to use only its supported languages, APIs, and frameworks. Current APIs allow storing and retrieving data from a BigTable non-relational database; making HTTP requests; sending e-mail; manipulating images; and caching. Existing web applications that require a relational database will not run on App Engine without

modification.
Per-day and per-minute quotas restrict bandwidth and CPU use, number of requests served, number of concurrent requests, and calls to the various APIs, and individual requests are terminated if they take more than 60 seconds or return more than 32MB of data.

Differences between SQL and GQL

Google App Engine's datastore has a SQL-like syntax called "GQL". GQL intentionally does not support the Join statement, because it seems to be inefficient when queries span more than one machine. Instead, one-to-many and many-to-many relationships can be accomplished using ReferenceProperty(). This shared-nothing approach allows disks to fail without the system failing. Switching from a relational database to the Datastore requires a paradigm shift for developers when modelling their data.
Unlike a relational database the Datastore API is not relational in the SQL sense.
The Java version supports asynchronous non-blocking queries using the Twig Object Datastore interface. This offers an alternative to using threads for parallel data processing.

IBM Cloud

AT A GLANCE The Company IBM Cloud operates a scalable, modular global cloud infrastructure platform that delivers exceptional data center performance and control to leading-edge customers around the world. Platform features include: • Hybrid, public and private cloud deployments • Bare metal and virtualized servers • Automated platform management • Network storage, CDN and backup services Objectives • Expand the global footprint of the IBM Hybrid Cloud offering • Offer a wider range of hybrid cloud scenarios • Decrease latency and speed up throughput by collocating servers and application hosting • Lower power costs and eliminate need for expensive PoP deployments • Deploy and scale faster • Guarantee private cloud security and platform uptime Results By choosing to partner with Digital Realty, IBM Cloud is able to offer its customers: • Direct Connect, no public internet or private networks required • No 3rd-party electronics • Higher network performance, consistency, and predictability • Streamlined and accelerated workload and data migration • Improved data and operational security • 1 GigE or 10 GigE speed • Diverse and redundant deployment • State-of-the-art, pay-as-you-go connectivity through Service Exchange The Challenge When on-demand cloud data center and hosting services provider IBM Cloud needed to expand its global data center footprint to meet burgeoning demand for its Infrastructure-as-a-Service offering, it needed the right data center partner to achieve the best available combination of power, space, and price.

IBM Cloud was shopping for a partner that could free them to do what they do best—run servers and manage services for their customers, while delivering low latency, high throughput and fast deployment of new installations. Delivery systems that put extra steps between server deployments, cloud and customers were too slow and indirect to deliver the low latency and high throughput IBM Cloud promises its customers. The right solution would put IBM Cloud's Infrastructure-as-a-Service business just milliseconds away from customers' secure private data storage.

Also on the must-have list were experience in provisioning and maintaining world-class data centers, enough distributed real estate to allow for nearly unlimited growth with a single vendor, and an impeccable record for uptime, resiliency and disaster recovery.

The Solution Proximity, collaboration and great design

Choosing to collocate at Digital Realty's hub locations in APAC and the U.S. puts IBM Cloud Data Centers right next door to Digital Realty's world-class data center deployments, with direct dark fiber connection between them.

This solution allows IBM Cloud to secure private cloud deployments while delivering industry-leading connectivity and throughput speeds. Modular cabinets make provisioning and expansion of colo facilities fast and promise repeatable success.

Leverage over power cost: The size of Digital Realty's global footprint, with more than 25 million sq. ft. of data center space worldwide and more than 2,000 customers, gives it the clout to secure a very favorable cost of power for its customers. This cost structure constitutes a measurable competitive advantage, which IBM Cloud is able to pass on to its customers

The importance of trust, the room to grow

For IBM Cloud, the choice of the right global data center partner was a business-critical decision. Digital Realty finished first because of its state-of-the-art data center management methodology, the depth and breadth of their global portfolio, and its durable commitment to the data center as its core business.

Innovations in connectivity

Through its new Service Exchange platform, Digital Realty offers IBM Cloud and its customers a way to virtualize cross connects as quickly and easily as the IBM Cloud platform creates virtual infrastructure. Delivered through MegaPort, Service Exchange enables businesses of all sizes to future-proof their private and hybrid cloud deployments with fast, secure, pay-as-yougo access to partners, customers and service providers throughout North America.

Windows Live

Introduction

Commercial search engines like Google, Yahoo! and Windows Live constantly crawl the Web and maintain huge searchable databases of the pages that they have found. Search engine results are now widely used for measurement purposes, not only by information researchers in Webometrics (Almind & Ingwersen, 1997; Bar-Ilan, 2004b), and related fields (Foot, Schneider, Dougherty, Xenos, & Larsen, 2003; Park, 2003; Pennock, Flake, Lawrence, Glover, & Giles, 2002) but also by commercial activities such as Web analytics and search engine optimisation. Hence, there is a need for research into the reliability of the results that search engines deliver and two relevant issues are discussed here.

First, search engine hit count estimates (e.g., 119,000 in "Results 1–10 of about 119,000") are often used in Webometrics research, for example to determine how many pages in one country link to another (Ingwersen, 1998). These hit count estimates are normally reported on each results page and can vary between results pages (e.g., the second results page might state: "Results 11–20 of about 116,000"). Hence, it is logical to question which estimate is the most reliable: that on the first page of result or that on a subsequent or the last page of results? Nevertheless, despite the continued use of search engines in Webometrics research, there has been no systematic study of how hit count estimates vary between results pages. Such a study could shed light on reasons for differences and any systematic biases as well as providing simple best practice advice.

Second, instead of hit count estimates, some Webometrics research requires lists of URLs matching a query, for example if the individual URLs need to be visited or their country of origin determined (Thelwall, Vann, & Fairclough, 2006). This is often problematic because search engines normally stop at about the 1000th result, with all other matching URLs remaining hidden from the user (Jepsen, Seiden, Ingwersen, Björneborn, & Borlund, 2004). It

is currently not known whether it is possible to use other methods to extract all of the remaining URLs in such cases. Moreover, search engines employ unreported methods to select which URLs they return, such as their page ranking algorithms (Chakrabarti, 2003), and so it is unclear whether their results are representative of their databases. Of course, because search engines do not index the whole Web, it is not possible to get a complete list of all pages matching a query.

In order to address the two issues above, this article introduces new methods to obtain extended lists of URLs for a search, including the initially hidden URLs, and to evaluate the hit counts reported by a search engine for queries with multiple pages of results. Previous research (reviewed below) has already developed several methods to assess various aspects of search engine performance but, surprisingly, none has fully investigated whether the hit count estimates are reliable reflections of the number of matching URLs in a search engine's database. We apply these methods to a case study of Windows Live, via its search service, and also present similar results for Google and Yahoo!. No previous study has evaluated Windows Live for Webometrics, and this is an important omission because it is currently the best for some types of investigation, as described below. Note that this article is concerned with extracting results from a single search engine and is not concerned with methods to obtain more complete URL lists or more comprehensive hit count estimates, such as through the use of multiple search engines (cf., Lawrence & Giles, 1998).

Webometric Methods for Search Engine Evaluation

This section briefly reviews research evaluating search engines to set the background for the current study. An important issue in the early years of the Web was to discover the percentage of the Web in the databases of major commercial search engines. A method has been developed to assess this: submitting a set of queries to search engines and comparing the results (lists of URLs from each search engine) to discover their degree of overlap. This method was also used to make inferences about the percentage of the whole Web (however defined) that each one indexed (Lawrence & Giles, 1998, 1999). The research showed that the search engines of the day covered up to 16% of the "indexable Web:" i.e., the pages that search engines could retrieve, in theory, by finding all Web site home pages and following their links recursively. From the Lawrence and Giles research we can be confident that no search engine today indexes the whole Web, and it also seems that any two unrelated search engines are likely to overlap by less than 50%.

Although there is no perfect method to evaluate search engine coverage, related research has continued. For example, a Web site sampling method has been used to show that search engine coverage has an almost inevitable international bias against Web newcomers, caused by the link structure of the Web (Vaughan & Thelwall, 2004). Others have focussed on the

ranking of search engine results in an attempt to propose an alternative ranking system that is not too biased towards popularity and against page quality (Cho & Roy, 2004; Cho, Roy, & Adams, 2005).

A separate research strand has focussed on the consistency of the results reported by search engines. Even though search engines do not cover the whole Web, the numbers that they report as hit count estimates for any query are interesting for at least two reasons. First, Webometric research has used these hit counts as the raw data for many studies of Web information (e.g., Aguillo, Granadino, Ortega, & Prieto, 2006; Ingwersen, 1998). Second, from an information retrieval perspective, it is useful to know how reliable the estimates reported by search engines are. In response, several researchers set out to systematically analyse variations in the results reported by commercial search engines. First, a comparison of results for the same query over short periods of time showed that fluctuations of several orders of magnitude could occur and also that sets of related queries could give inconsistent results (Snyder & Rosenbaum, 1999). Second, Rousseau (1999) tracked the variation over time of specific queries in NorthernLight and AltaVista, showing that the results tended to be quite stable but were subject to large fluctuations, presumably due to software or hardware upgrades. Bar-Ilan (1999) investigated the results of six search engines in more detail, discovering that they forgot information, in the sense that URLs were occasionally not reported in results, and that these URLs pointed to information that was not available elsewhere in the search engine results returned. Subsequent research encompassed Google and tracked the coverage of a large set of Web sites, finding a pattern of stability but with occasional sudden changes (Thelwall, 2001). The research of Mettrop and Nieuwenhuysen (2001) also used a time series approach but used a set of controlled seed URLs in order to get more detailed information on search engine performance. They confirmed that search engines sometimes did not report a page even when it matched a query and was in their index (Mettrop & Nieuwenhuysen, 2001). Bar-Ilan describes search engines as "concealing" pages when they do not report them, despite matching a query and the page being in their database (Bar-Ilan, 2002).

In conclusion, search engines should be viewed as engineering products, designed to produce fit-for-purpose results but not as mathematical "black boxes" that deliver logically correct results. Search engines may take shortcuts when estimating and returning results in order to improve their speed or efficiency of operation. For example, they may only search a fraction of their index for a query, stopping when they run out of time or have found enough results. See also Arasu, Cho, Garcia-Molina, Paepcke, & Raghavan, 2001; Bar-Ilan, 2004a2004b; Brin & Page, 1998 for technical issues that may impact on search engine results.

New Methods for Extracting URLs from Search Engines

This section contains details of two new methods for automatically extracting URLs from search engines for queries with more results than the search engine would normally return. First, however, the software environment and facilities used to implement the methods are described and a case is made for the value of Windows Live in Webometric research.

Data Collection—The Microsoft Search Web Service

The three major families of search engines: Google, Yahoo!, and Windows Live (Microsoft) all maintain at least one service to allow computer programs to automatically submit search queries and retrieve results. These are variously called Application Programming Interfaces (APIs) or Web services, and they operate in different ways but perform the same function. Each restricts the number of searches that can be made per day and has some conditions of use, although all of these conditions broadly allow them to be used for research purposes. Of the three, the Microsoft Search Web Service is the most useful for investigations and large scale deployment in applications for two reasons. First, its results are the same as those of the online search engine, which is not the case for the Google API (Mayr & Tosques, 2005) and the Yahoo! Web Search API (e.g., through our own testing and as claimed at http://bsd119.ib.hu-berlin.de/~ft/index_e.html on 8.2.2006). Second, the number of queries that can be submitted per day is 10,000, with each giving up to 50 results (http://search.msn.com/developer/). This is much more than the Google API (1,000 queries of 10 results, http://code.google.com/apis/soap-search/, support partially discontinued on Dec 6, 2006, in favour of a less powerful Ajax Search API) and the same as the Yahoo! Web Search API (5,000 queries of 100 results, http://developer.yahoo.com/search/Web/).

For this research, we extracted all data automatically using the Microsoft Search Web Service, submitting queries through the free LexiURL Searcher software, which was upgraded for this article with the two new methods described below (http://lexiurl.wlv.ac.uk). For each query submitted, a list of up to 50 URLs is returned, together with the title of the page and a description of the page (a snippet, as displayed in Windows Live). With LexiURL searcher a complete list of up to 1,000 URLs matching a query can also be obtained (the maximum allowed by Windows Live). This is achieved by submitting up to 20 separate queries, each giving a different starting point for the URL list (i.e., 1, 51, 101, … 951) and then merging all of the lists. Because the search service sometimes returns duplicate URLs and often returns identical pages of URLs without warning for the last pages of results, LexiURL Searcher automatically checks for duplicate URLs in all the results and discards any found.

Method 1: Automated Query Splitting

Query splitting is a technique previously used to increase the number of URLs returned by a search engine as matching a query (Thelwall, Vann, & Fairclough, 2006; as previously used by Bar-Ilan and Peritz, 2004 and elsewhere on an ad-hoc basis, e.g., http://www.Webmasterworld.com/forum7/1379.htm from November, 2005). For example, if a search engine reported 2,000 matches for the query "Webometrics," then it would return the just the first 1,000 results (e.g., in 20 pages of 50) and then stop. In order to get the full set of 2,000 results, it might be possible to split the query "Webometrics" into two separate logically disjoint queries by adding and subtracting the same word. For example, if the word was Cybermetrics, then this would give "Webometrics Cybermetrics" (i.e., all pages containing both the words Webometrics and Cybermetrics) and "Webometrics Cybermetrics" (i.e., all pages containing the word Webometrics but not the word Cybermetrics). If both of these queries gave 1,000 results, then the complete list of URLs matching "Webometrics" could theoretically be obtained by submitting the two subqueries and obtaining all 1,000 matching URLs for each one. In theory, any query, no matter how many URLs it matched, could be split into a large set of logically separate queries by recursively adding and subtracting extra words, and the results combined to create a complete list of matching URLs (e.g., see Appendix, Table 5). In practice, however, this process has a limit because of query length restrictions (see below).

The following example illustrates query splitting using the query "Ingwersen", for which Windows Live reported 33,779 results. Figure 1 shows the results of the first two query splitting levels. First, the query was split by adding and subtracting the word "peter". Then each new subquery was split differently, the first by adding and subtracting "j" and the second by adding and subtracting "und." In all cases, the total number of results was still above 1,000, so the splitting had to continue (not shown). This process eventually generated 39 separate queries (each with up to 20 results pages), terminating when each query returned under 1,000 results. For example, one of the leaf queries was "ingwersen -peter- und-information -de -geranium -1" with 627 results. Combining the results of all the leaf queries (i.e., those with less than 1,000 results) gave a new estimated result count (10,702) and combining the URLs extracted from the leaf queries gave a list of 10,702 URLs matching the query "Ingwersen."

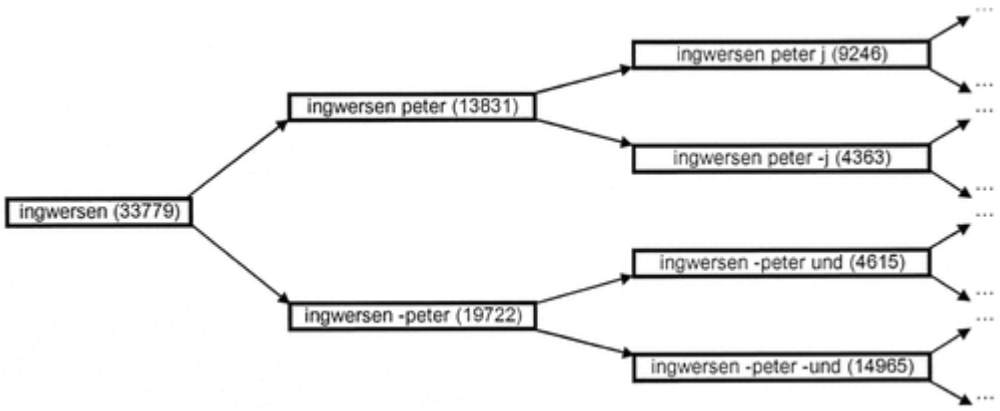


Figure 1

Open in figure viewerPowerPoint

An illustration of the start of query splitting for the query "Ingwersen." The number of matches for each query is in brackets.

Although query splitting has been previously implemented manually with human intervention to decide which words to add and subtract (Thelwall et al., 2006), here we propose a simple automatic method. This method is to create a list of all words used in the title or description of the matching URLs for a query and then to choose a "splitting" word that occurs in 15% of them. The value, 15%, was determined through testing as being the value that was most likely to produce an approximately even split of results, even though it often gives the most common word in the list that is not already in the query. In addition, very common words should not be used (we produced a list of 54) because they are ignored by Windows Live when conducting a search. Of course, words already in the query should clearly be excluded. This method normally works well but is occasionally poor because it selects a very common word that is rare in titles and descriptions (e.g., pronouns, common verbs) or a rare word that is common in titles and descriptions (e.g., "untitled").

Query splitting is fast because it only uses the data provided by the Microsoft Search Web Service. A more effective method would be to download the complete text of each page and choose a word that was in 50% of these pages, but this would take considerably longer and consume much more computing power and network bandwidth. A practical limitation is that it does not work for very large queries because search engines employ query length restrictions (e.g., on January 15, 2007, the Windows Live limit was 150 characters http://help.live.com/help.aspx?project=wl_searchv1&market=en-gb). In practice we apply it recursively up to nine times so it can, in theory, generate complete lists of up to $1000 \times 2^9 = 512{,}000$ URLs, but because the method is imperfect, 100,000 URLs is a more practical upper limit. Note that the recursive application of the method means that a new word list and query splitting word must be generated for each split query.

Query splitting can also be used for a second purpose: to gain revised *estimates* of the number of URLs matching a search by splitting the search into a set of multiple searches that combine to be equivalent to the original search.

Pseudocode for query splitting is given below in the form of a function designed to return an estimate of the total number of results available for a query based upon query splitting to produce constituent searches with a maximum of 950 results each (as required for investigation 2, below). The minimum number of results to trigger query splitting used was 950 because in some search engines the 1,000 maximum is not quite reached, even for queries with hundreds of thousands of results. Calling the function also produces a list of URLs matching the query, including URLs returned by recursive function calls, and this is the extended URL list (as required for investigations 3 and 4, below).

An alternative method to gain additional results from a search engine is to use the site: command to refine a search in the hope of gaining additional results from specific domains. This command restricts the results to a single domain name or domain name ending. For example, to gain additional Danish results matching "Ingwersen" the query "ingwersen site:dk" could be submitted, which would match all URLs of pages containing the word "Ingwersen" with domain names ending in .dk. Site searching can be automated in two ways.

- *Automated domain searching*: Conduct a normal search for a query *q* (with or without query splitting). For each URL in the results, extract its domain name *d* and submit the query "*q* site:*d*" (with or without query splitting). If multiple URLs have the same domain name d then of course the query "*q* site:*d*" only needs to be submitted once.

- *Automated TLD searching*: This is the same as automated domain searching except that the Top Level Domain (TLD) *t* of each URL (e.g., com, net, edu, uk, dk) is extracted and the query "*q* site:*t*" is submitted.

Automated domain searching is intuitively promising since search engines often report a maximum of one or two results per page from the same Web site. Automated TLD searching cannot be given a similar justification but is potentially useful as an additional method of identifying URLs with unknown domain names.

Microsoft Dynamics CRM

Customer Relationship Management (CRM) is a system for managing a company's interactions with current and future customers. It often involves using technology to organize, automate, and synchronize sales, marketing, customer service, and technical support. CRM can help reduce costs and increase profitability by organizing and automating business processes that nurture customer satisfaction and loyalty.

Microsoft Dynamics CRM is a customer relationship management software package developed by Microsoft focused on enhancing the customer relationship for any organization. Out of the box, the product focuses mainly on Sales, Marketing, and Customer Service sectors, though Microsoft has been marketing Dynamics CRM as an XRM platform and has been encouraging partners to use its proprietary (.NET based) framework to customize it. In recent years, it has also grown as an Analytics platform driven by CRM.

The CRM Solution can be used to drive the sales productivity and marketing effectiveness for an organization, handle the complete customer support chain, and provide social insights, business intelligence, and a lot of other out-of-the-box functionalities and features. As a product, Microsoft Dynamics CRM also offers full mobile support for using CRM apps on mobiles and tablets.

As of writing this tutorial, the latest version of CRM is CRM 2016. However, in this tutorial we will be using CRM 2015 Online version as it is the latest stable version as well as frequently used in many organizations. Nevertheless, even if you are using any other versions of CRM, all the concepts in the tutorial will still hold true.

Product Offerings

Microsoft Dynamics CRM is offered in two categories −

CRM Online

CRM Online is a cloud-based offering of Microsoft Dynamics CRM where all the backend processes (such as application servers, setups, deployments, databases, licensing, etc.) are managed on Microsoft servers. CRM Online is a subscription-based offering which is preferred for organizations who may not want to manage all the technicalities involved in a CRM implementation. You can get started with setting up your system in a few days (not weeks, months or years) and access it on web via your browser.

CRM On-Premise

CRM on-premise is a more customized and robust offering of Microsoft Dynamics CRM, where the CRM application and databases will be deployed on your servers. This offering allows you to control all your databases, customizations, deployments, backups, licensing and other network and hardware setups. Generally, organizations who want to go for a customized CRM solution prefer on-premise deployment as it offers better integration and customization capabilities.

From the functional standpoint, both the offerings offer similar functionalities; however, they differ significantly in terms of implementation. The differences are summarized in the following table.

CRM Online	CRM On-Premise
This is a cloud-based solution provided by Microsoft in which all the servers and databases are managed by Microsoft.	This is an on-premise solution provided by Microsoft in

	which the servers and databases are managed by the customer.
You can get started with an online offering in a matter of few days. You pay for the users and used space on-the-go.	Setting up an on-premise offering needs technical skills as well as sufficient time to setup the CRM instance and get it running.
It supports relatively less customizations and extensions.	It supports relatively more customization and extensions.
CRM Online does not give the ability to perform manual data backup and restore options, since the database is hosted on Microsoft servers. However, Microsoft performs daily backups of the database.	CRM on-premise gives complete ability to manage your database.
CRM Online has various plans based on the data storage limits such as 5GB, 20 GB, etc.	CRM on-premise does not have any such limits on storage size, since the data exists on your own servers.

CRM Online provides inbuilt capabilities of features such as insights, social listening, analytics, etc.	CRM on-premise has extra costs for these features.
CRM Online supports automatic updates to future version.	CRM on-premise updates need to be installed by the administrator.

Salesforce.com CRM

About Salesforce.com

- The Salesforce cloud is an on-demand customer relationship management (CRM) suite offering applications for small, midsize and enterprise organizations, with a focus on sales and support.
- The Salesforce app has capabilities that include sales management, marketing automation, partner relationship management and customer service. These applications help organizations manage customer accounts, track sales leads, conduct and monitor marketing campaigns and provide service post sale.
- Salesforce for Outlook allows users to synchronize contacts, calendars, emails and tasks in both applications. Salesforce for small business allows users to manage contacts, track sales deals, manage tasks and events, harvest leads and track performance. The Salesforce enterprise edition layers in call scripts, team-selling functionality, business workflow, setup approval and automation, custom applications, API integrations and more.
- CRM Salesforce solutions are available only for software-as-a-service (SaaS) deployment. Force.com provides a technology stack that covers database and security as well as workflow and user interface.

Salesforce is a company based out of San Francisco, California. They are the leaders in cloud technology.

Salesforce is also the name of their flagship product, which is a highly customizable CRM with attractive out of the box features like Web to Lead, Weeding out duplicate leads, Opportunity Forecasting, Email and Campaign management, Google Apps Integration etc… All this along with the standard CRM features like Lead, Contact, Account, and Opportunity Management.

What is salesforce CRM, you might ask. It's a top-notch CRM application built on the Force.com platform. It can manage all the customer interactions of an organization through different media, like phone calls, site email enquiries, communities, as well as social media. Salesforce handles all the customer relationships, by focusing on the sales, marketing and support processes. This is done by working with the standard objects (Shown below), and facilitating the relationships between them.

Force.com platform
Force.com is a platform for creating applications in the cloud with absolutely no software or hardware investment required. The apps thus created are data-centric and collaborative. In fact, the data is never lost here, because there are auto back-ups.

App exachange

AppExchange is the Salesforce store, empowering businesses to extend the functionality of Salesforce across every department and industry. It's an ecosystem of over 5,000 ready-to-install solutions, 80,000 peer reviews, and 6 million customer installs to help solve any business challenge. AppExchange includes all types of solutions too: You'll find apps, components, Lightning Data, Bolt Solutions, and Flows.

Just follow these seven steps.

1. Define success: Make a roadmap of your goals for both long- and short-term success.
2. Determine your budget: Do you have money for paid apps or do you need to start

with free apps?

3. Find your solution: Once you know your goals and budget, use AppExchange's personalized recommendations to narrow your search.
4. Take guided learning paths: To know more about AppExchange solutions, use resources such as Getting Started Trailhead guides, community content, and App Guides (solutions recommended by customers).
5. Read reviews: Inform your decision with over 80,000 peer reviews.
6. Try before you buy: Use a Salesforce sandbox to simulate your production environment and see how the app will perform.
7. Join the Salesforce community: With over 14 million members, including dedicated MVPs, the Salesforce community is willing to share its knowledge with you.

VISIT THE RECOMMENDED FOR YOU PAGE

How do I install an AppExchange solution?

Before installing one of our solutions into your production instance of Salesforce, make sure you:

1. Check edition compatibility: Verify that the solution you want to download is compatible with your Salesforce edition.
2. Check for an external service: Make sure your AppExchange solution has access to that service.
3. Test drive before installing: As an admin or read-only user, test drive your solution to review the core functionality.
4. Install into Developer Edition or Sandbox: Further testing your solution in either environment lets you test it beyond the constraints of a read-only user. Now you'll be able to create, edit, and delete records within your newly installed solution. Sign up for a free Development environment here.
5. If you're satisfied with your solution's functionality, install it into your production organization.

JOIN THE TRAILBLAZER COMMUNITY

Who will have access to my installed solution?

During the installation process, you'll be given three options. You can make your solution available to:

1. All users
2. Some users

3. Admins only. (Please note: If you choose admins only, you may make it available to other users later.)

Is there a limit to the AppExchange solutions I can download?

No. You have limits on the number of custom tabs and objects, but not the number of solutions.

Are AppExchange apps mobile?

AppExchange has hundreds of Salesforce mobile-ready solutions. In addition, some partners offer custom mobile solutions.

How do I list on AppExchange?

AppExchange partners have the ability to build, sell, and grow their business faster with Salesforce. If you're interested in getting up and running with AppExchange, or just want to learn the benefits of becoming an ISV or SI partner, visit our AppExchange Partner Program to find out more.

What are the best apps for my department or industry?

AppExchange can extend Salesforce across any department and industry you can imagine. That means there are lots of resources to help you find the best solutions.

1. A good place to start are the AppExchange collections. These are groups of listings curated by AppExchange experts, and they're a great way to zero in on the top solutions in a particular area, such as by industry or product.
2. You can also explore our series of App Guides — apps recommended by users themselves — in specific categories like sales, service, manufacturing, marketing, back office, and Salesforce Labs.
3. Finally, AppExchange includes Recommended for You solutions, enabling you to get a specific recommendation for a solution based on your profile data, install history, location, and more.

Amazon S3

Amazon Simple Storage Service (Amazon S3) is a scalable, high-speed, web-based cloud storage service designed for online backup and archiving of data and applications on Amazon Web Services. Amazon S3 was designed with a minimal feature set and created to make web-scale computing easier for developers.

How Amazon S3 works

Amazon S3 is an object storage service, which differs from block and file cloud storage. Each object is stored as a file with its metadata included and is given an ID number. Applications use this ID number to access an object. Unlike file and block cloud storage, a developer can access an object via a REST API.

The S3 cloud storage service gives a subscriber access to the same systems that Amazon uses to run its own websites. S3 enables customers to upload, store and download practically any file or object that is up to five terabytes (TB) in size, with the largest single upload capped at five gigabytes (GB).

Amazon S3 features

S3 provides 99.999999999% durability for objects stored in the service and supports multiple security and compliance certifications. An administrator can also link S3 to other AWS security and monitoring services, including CloudTrail, CloudWatch and Macie. There's also an extensive partner network of vendors that link their services directly to S3.

Data can be transferred to S3 over the public internet via access to S3 APIs. There's also Amazon S3 Transfer Acceleration for faster movement over long distances, as well as AWS Direct Connect for a private, consistent connection between S3 and an enterprise's own data center. An administrator can also use AWS Snowball, a physical transfer device, to ship large amounts of data from an enterprise data center directly to AWS, which will then upload it to S3.

S3 basics

In addition, users can integrate other AWS services with S3. For example, an analyst can query data directly on S3 either with Amazon Athena for ad hoc queries or with Amazon Redshift Spectrum for more complex analyses.

Amazon S3 storage classes

Amazon S3 comes in three storage classes: S3 Standard, S3 Infrequent Access and Amazon Glacier. S3 Standard is suitable for frequently accessed data that needs to be delivered with low latency and high throughput. S3 Standard targets applications, dynamic websites, content distribution and big data workloads.

S3 Infrequent Access offers a lower storage price for data that's needed less often, but that must be quickly accessible. This tier can be used for backups, disaster recovery and long-term data storage.

Amazon Glacier is the least expensive storage option in S3, but it is strictly designed for archival storage because it takes longer to access the data. Glacier offers variable retrieval rates that range from minutes to hours.

A user can also implement lifecycle management policies to curate data and move it to the most appropriate tier over time.

Working with buckets

Amazon does not impose a limit on the number of items that a subscriber can store; however, there are Amazon S3 bucket limitations. An Amazon S3 bucket exists within a particular region of the cloud. An AWS customer can use an Amazon S3 API to upload objects to a particular bucket. Customers can configure and manage S3 buckets.

Protecting your data

User data is stored on redundant servers in multiple data centers. S3 uses a simple web-based interface -- the Amazon S3 console -- and encryption for user authentication.

S3 buckets are kept private by default, but an admin can choose to make them publicly accessible. A user can also encrypt data prior to storage. Rights may be specified for individual users, who will then need approved AWS credentials to download or access a file in S3.

When a user stores data in S3, Amazon tracks the usage for billing purposes, but it does not otherwise access the data unless required to do so by law.

Oracle OBIEE server

Oracle BI Server is a query, reporting and analysis server and provides services to the other components of the Business Intelligence suite such as Data mining, Reporting, and Analytic Applications. The Server uses ODBC 2.0 which provides a standard software API method for using database management systems and JDBC (Java Database Connectivity) which is an

API for the Java programming language. The BI server compiles incoming query requests into an executable code and then execute the code. Clients of the BI Server work with a logical data independent of the data sources, and they submit them to the BI server. The server translates the data to some combination of physical SQL and sends them to the back-end databases.

Server administration functions include facilities such as monitoring, statistics logging, cancellation, session and query management. The administration of the server is managed internally with using standard protocols e.g. Active Directory is an implementation of LDAP services(Lightweight Directory Access Protocol) which is an application protocol for querying and modifying directory services running over TCP and IP.

Query compilation in BI server has five phases:

- Parsing
- Logical Request Generation
- Navigation
- Rewrite
- Code Generation

Parsing[edit]

The multi-functional parser accepts the full ANSI SQL92 syntax as its input and its main function is to generate a parse tree which is a tree data structure that represents the syntactic structure of a string as its output.

Logical Request Generation[edit]

A component of the logical request generation takes the parse tree and makes of it a logical request tree that describes the precise semantics of the requested data in simplifieds SQL.

Navigation[edit]

The navigator processes the navigation.

Rewrite[edit]

The rewrite's output is the execution plan. This phase distributes relational query optimization and suitable native SQL generation. The phase of rewriting covers:

- Multi-database join plan generation
- Function shipping
- Functional compensation analysis
- Optimized SQL generation

Code Generation[edit]

Code generation in the server produces native query processing language directives to communicate with remote databases. It also produces a code for any remaining query processing that has not been function shipped to remote databases. Parallel execution directives are inserted in this code and it is used for the analytics execution engine.