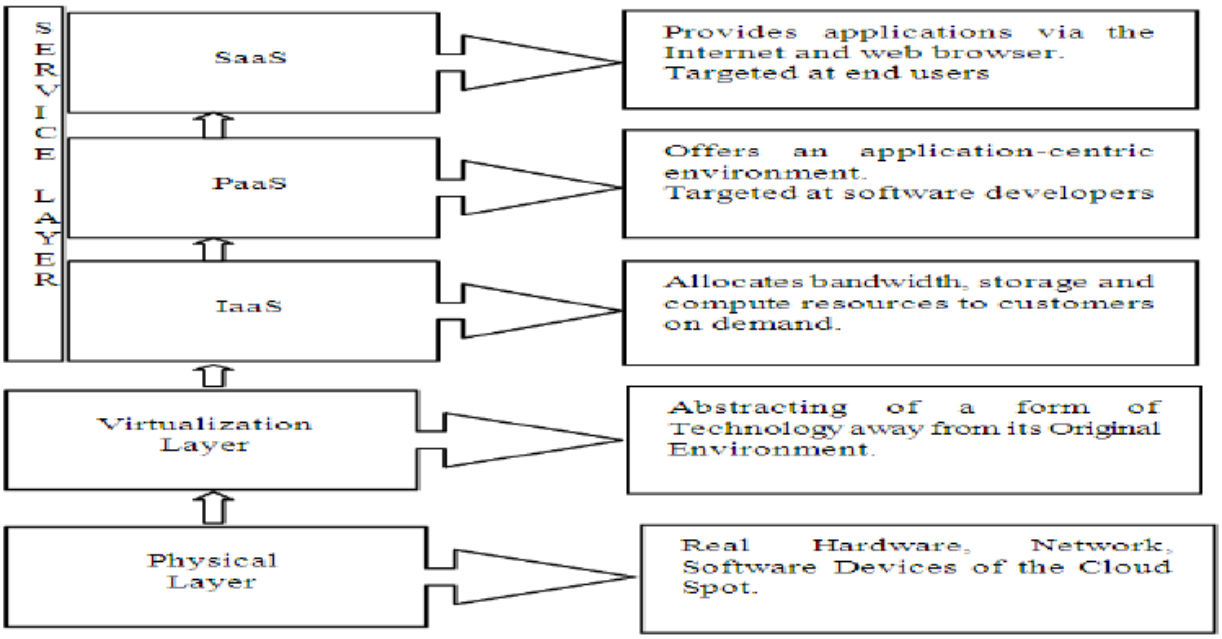
CLOUD COMPUTING – Unit 3

1)CLOUD ARCHITECTURE:

LAYERED CLOUD ARCHITECTURE:



Virtualization and dynamic provisioning of resources are the principles on which cloud computing works. In terms of architecture, the cloud hosting can be sliced into four different layers.

The Physical Layer:

This layer comprises of physical servers, network and other aspects that can be physically managed and controlled.

The Infrastructure Layer:

This includes storage facilities, virtualized servers, and networking. Infrastructure as a Service or IaaS points to delivery of services in hosted format. They include hardware, network and servers, delivered to end users. Consumers can enjoy access to scalable storage and compute power as and when needed.

Platform Layer:

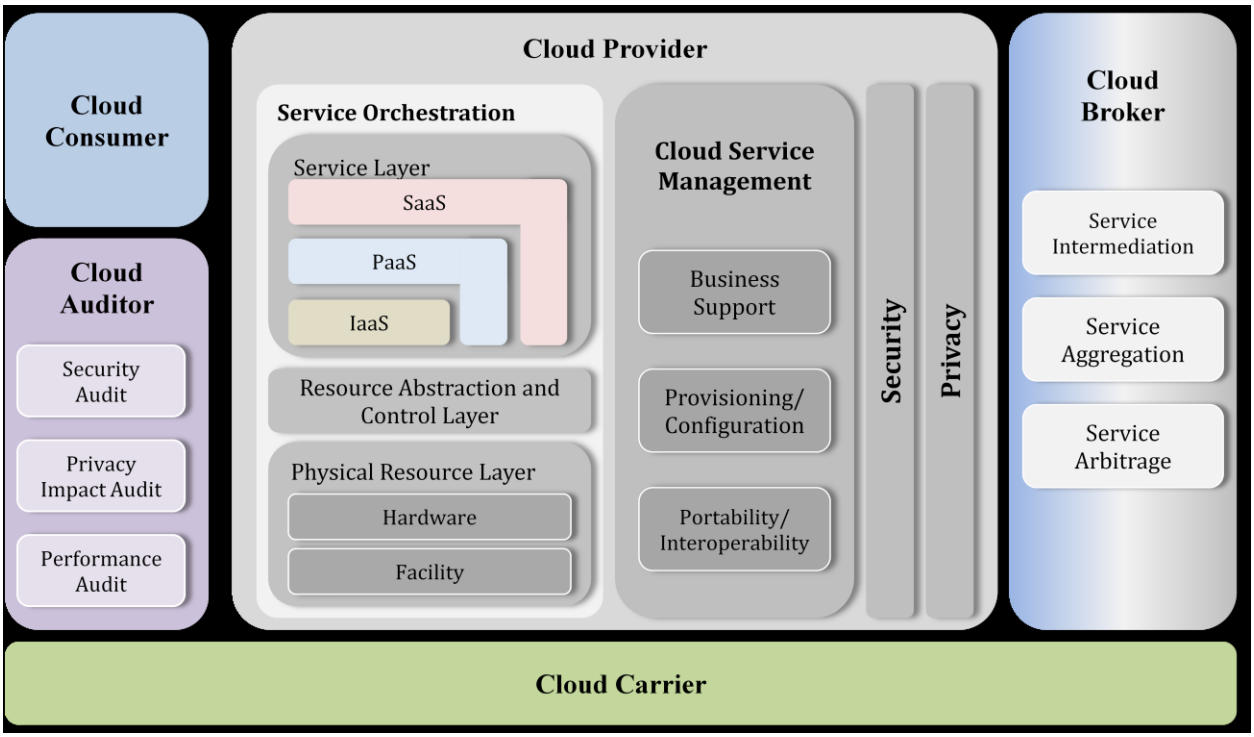
This layer includes services such as OS and Apps. It serves as a platform for development and deployment. The Platform layer provides the right platform for development and deployment of applications vital for the cloud to run smoothly.

Application Layer:

The Application Layer is the one that end users interact with in a direct manner. It mainly comprises of software systems delivered as service. Examples are Gmail and Dropbox. SaaS or Software as a Service ensures delivery of software in hosted form which can be accessed by users through the internet. Configurability and scalability are the two key features of this layer. Customers can easily customize their software system using Meta data.

These layers allow users to use cloud computing services optimally and achieve the kind of results they are looking for from the system.

NIST CLOUD REFERENCE ARCHITECTURE:



This Figure presents an overview of the NIST cloud computing reference architecture, which identifies the major actors, their activities and functions in cloud computing. The diagram depicts a generic high-level architecture and is intended to facilitate the understanding of the requirements, uses, characteristics and standards of cloud computing. NIST cloud computing reference architecture defines five major actors: *cloud consumer*, *cloud provider*, *cloud carrier*, *cloud auditor* and *cloud broker*

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

Cloud Consumer

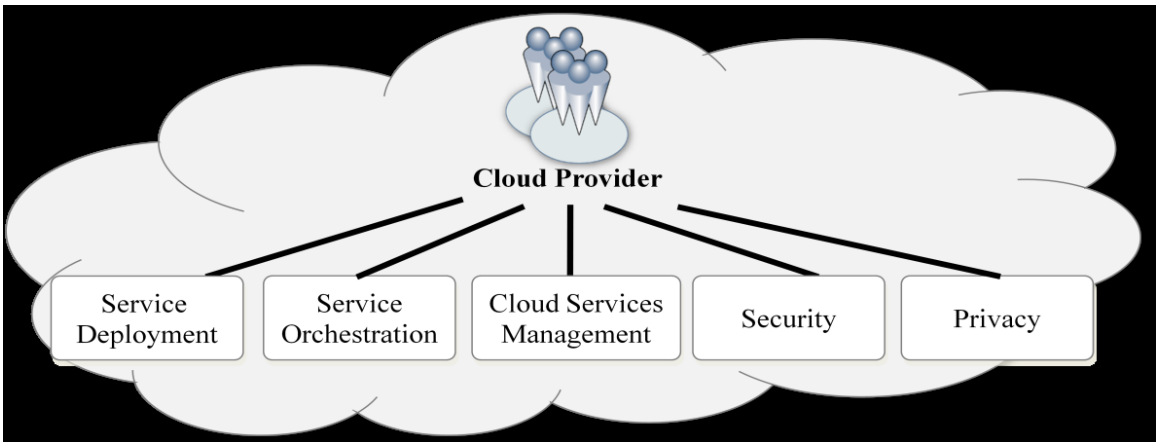
The cloud consumer is the principal stakeholder for the cloud computing service. A cloud consumer represents a person or organization that maintains a business relationship with, and uses the service from a cloud provider. A cloud consumer browses the service catalog from a cloud provider, requests the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

Cloud consumers need SLAs to specify the technical performance requirements fulfilled by a cloud provider. SLAs can cover terms regarding the quality of service, security, remedies for performance failures. A cloud provider may also list in the SLAs a set of promises explicitly not made to consumers, i.e. limitations, and obligations that cloud consumers must accept. A cloud consumer can freely choose a cloud provider with better pricing and more favorable terms. Typically a cloud provider's pricing policy and SLAs are non-negotiable, unless the customer expects heavy usage and might be able to negotiate for better contracts.

Cloud Provider

A cloud provider is a person, an organization; it is the entity responsible for making a service available to interested parties. A Cloud Provider acquires and manages the computing infrastructure required for providing the services, runs the cloud software that provides the services, and makes arrangement to deliver the cloud services to the Cloud Consumers through network access.

For Software as a Service, the cloud provider deploys, configures, maintains and updates the operation of the software applications on a cloud infrastructure so that the services are provisioned at the expected service levels to cloud consumers. The provider of SaaS assumes most of the responsibilities in managing and controlling the applications and the infrastructure, while the cloud consumers have limited administrative control of the applications.



Cloud Auditor

A cloud auditor is a party that can perform an independent examination of cloud service controls with the intent to express an opinion thereon. Audits are performed to verify conformance to standards through review of objective evidence. A cloud auditor can evaluate the services provided by a cloud provider in terms of security controls, privacy impact, performance, etc.

A privacy impact audit can help Federal agencies comply with applicable privacy laws and regulations governing an individual‟s privacy, and to ensure confidentiality, integrity, and availability of an individual‟s personal information at every stage of development and operation.

Cloud Broker

As cloud computing evolves, the integration of cloud services can be too complex for cloud consumers to manage. A cloud consumer may request cloud services from a cloud broker, instead of contacting a cloud provider directly. A cloud broker is an entity that manages the use, performance and delivery of cloud services and negotiates relationships between cloud providers and cloud consumers.

In general, a cloud broker can provide services in three categories :

Service Intermediation: A cloud broker enhances a given service by improving some specific capability and providing value-added services to cloud consumers. The improvement can be managing access to cloud services, identity management, performance reporting, enhanced security, etc.

Service Aggregation: A cloud broker combines and integrates multiple services into one or more new services. The broker provides data integration and ensures the secure data movement between the cloud consumer and multiple cloud providers.

Service Arbitrage: Service arbitrage is similar to service aggregation except that the services being aggregated are not fixed. Service arbitrage means a broker has the flexibility to choose services from multiple agencies. The cloud broker, for example, can use a credit-scoring service to measure and select an agency with the best score.

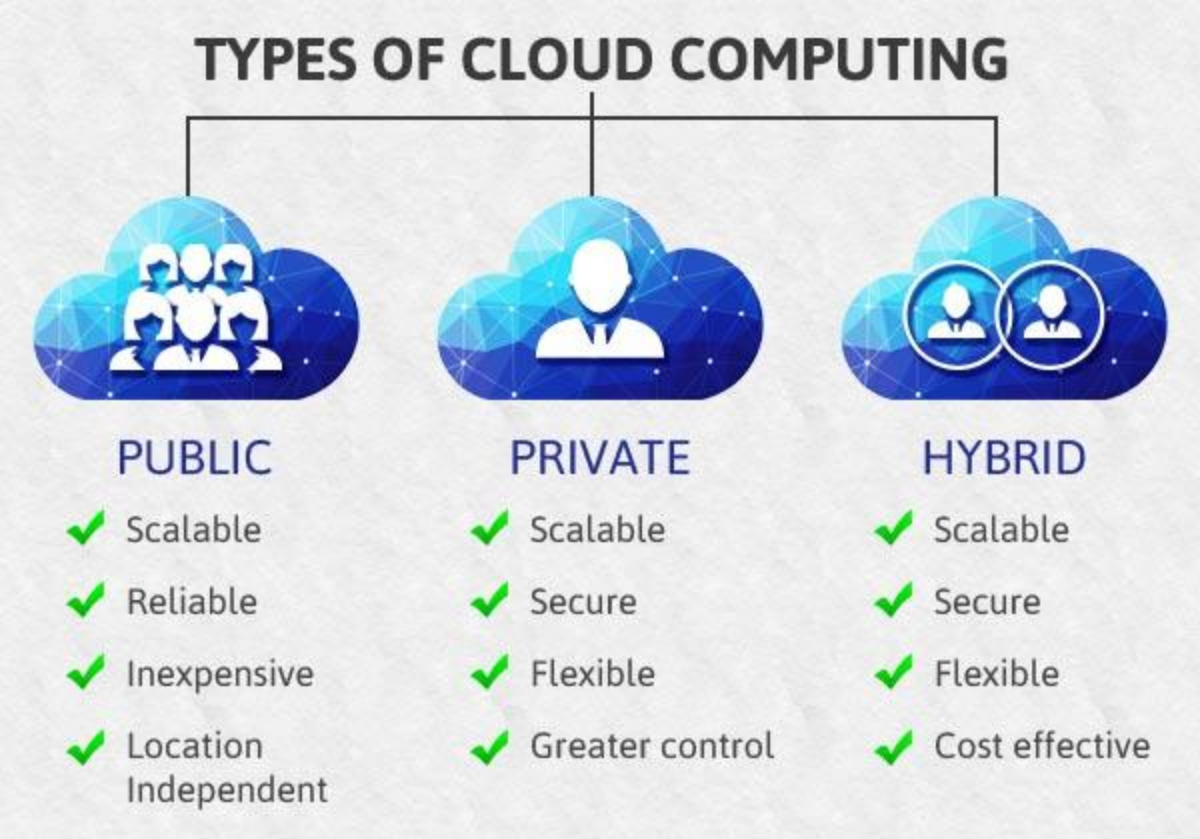
Cloud Carrier

A cloud carrier acts as an intermediary that provides connectivity and transport of cloud services between cloud consumers and cloud providers. Cloud carriers provide access to consumers through network, telecommunication and other access devices. For example, cloud consumers can obtain cloud services through network access devices, such as computers, laptops, mobile phones, mobile Internet devices (MIDs), etc [1]. The distribution of cloud services is normally provided by network and telecommunication carriers or a transport agent, where a transport agent refers to a business organization that provides physical transport of storage media such as high-capacity hard drives. Note that a cloud provider will set up SLAs with a cloud carrier to provide services consistent with the level of SLAs offered to cloud consumers, and may require

the cloud carrier to provide dedicated and secure connections between cloud consumers and cloud providers.

2) PUBLIC PRIVATE HYBRID CLOUD

Difference	Private	Public	Hybrid
Tenancy	Single tenancy: there's only the data of a single organization stored in the cloud.	Multi-tenancy: the data of multiple organizations in stored in a shared environment.	The data stored in the public cloud is usually multi-tenant, which means the data from multiple organizations is stored in a shared environment. The data stored in private cloud is kept private by the organization.
Exposed to the Public	No: only the organization itself can use the private cloud services.	Yes: anyone can use the public cloud services.	The services running on a private cloud can be accessed only the organization's users, while the services running on public cloud can be accessed by anyone.
Data Center Location	Inside the organization's network.	Anywhere on the Internet where the cloud service provider's services are located.	Inside the organization's network for private cloud services as well as anywhere on the Internet for public cloud services.
Cloud Service Management	The organization must have their own administrators managing their private cloud services.	The cloud service provider manages the services, where the organization merely uses them.	The organization itself must manage the private cloud, while the public cloud is managed by the CSP.
Hardware Components	Must be provided by the organization itself, which has to buy physical servers to build the private cloud on.	The CSP provides all the hardware and ensures it's working at all times.	The organization must provide hardware for the private cloud, while the hardware of CSP is used for public cloud services.
Expenses	Can be quite expensive, since the hardware, applications and network have to be provided and managed by the organization itself.	The CSP has to provide the hardware, set-up the application and provide the network accessibility according to the SLA.	The private cloud services must be provided by the organization, including the hardware, applications and network, while the CSP manages the public cloud services.



What is Public Cloud Computing?

A cloud platform that is based on standard cloud computing model in which service provider offers resources, applications storage to the customers over the internet is called as public cloud computing. The hardware resources in public cloud are shared among similar users and accessible over a public network such as the internet. Most of the applications that are offered over internet such as Software as a Service (SaaS) offerings such as cloud storage and online applications uses Public Cloud Computing platform. Budget conscious startups, SMEs not keen on high level of security features looking to save money can opt for Public Cloud Computing.

Advantage of Public Cloud Computing

1. It offers greater scalability
2. Its cost effectiveness helps you save money.
3. It offers reliability which means no single point of failure will interrupt your service.
4. Services like SaaS, (Paas), (Iaas) are easily available on Public Cloud platform as it can be accessed from anywhere through any Internet enabled devices.
5. It is location independent – the services are available wherever the client is located.

Disadvantage of Public Cloud Computing

1. No control over privacy or security
2. Cannot be used for use of sensitive applications
3. Lacks complete flexibility as the platform depends on the platform provider
4. No stringent protocols regarding data management

What is Private Cloud Computing?

A cloud platform in which a secure cloud based environment with dedicated storage and hardware resources provided to a single organization is called Private Cloud Computing. The Private cloud can be either hosted within the company or outsourced to a trusted and reliable third-party vendor. It offers company a greater control over privacy and data security. The resources in case of private cloud are not shared with others and hence it offer better performance compared to public cloud. The additional layers of security allow company to process confidential data and sensitive work in the private cloud environment.

Advantage of Private Cloud Computing

1. Offers greater Security and Privacy
2. Offers more control over system configuration as per the company's need
3. Greater reliability when it comes to performance
4. Enhances the quality of service offered by the clients
5. Saves money

Disadvantage of Private Cloud

1. Expensive when compared to public cloud
2. Requires IT Expertise

What is Hybrid Cloud Computing?

Hybrid Cloud computing allows you to use combination of both public and private cloud. This helps companies to maximize their efficiency and deliver better performance to clients. In this model companies can use public cloud for transfer of non-confidential data and switch on to private cloud in case of sensitive data transfer or hosting of critical applications. This model is gaining prominence in many business as it gives benefits of both the model.

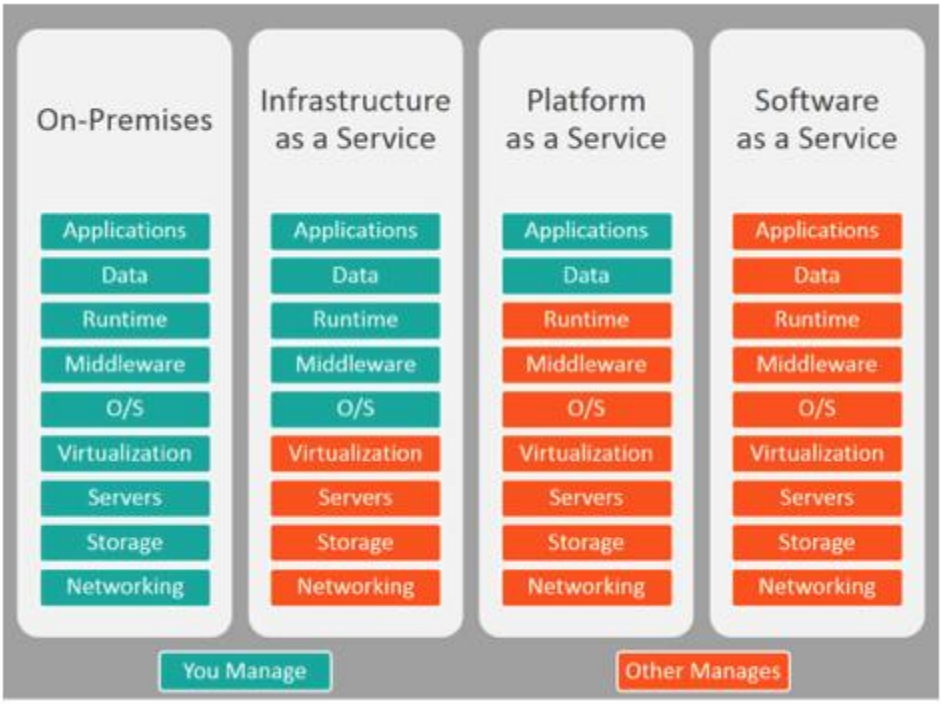
Advantage of Hybrid Cloud Computing

1. It is scalable
2. It is cost efficient
3. Offers better security
4. Offers greater flexibility

Disadvantage of Hybrid Cloud Computing

1. Infrastructure Dependency
2. Possibility of security breach through public cloud

3) IAAS, PAAS & SAAS:



IaaS: Infrastructure as a Service

This is a virtual equivalent of a traditional data center. Cloud infrastructure providers use virtualization technology to deliver scalable compute resources such as server s, network s and storage to their clients. This is beneficial for the clients, as they don't have to buy personal hardware and manage its component s. Instead, they can deploy their platforms and application s within the provider's virtual machines that offer the same technologies and capabilities as a physical data center.

An IaaS provider is responsible for the entire infrastructure, but users have total control over it. In turn, users are responsible for installing and maintaining apps and operating systems, as well as for security, runtime, middleware and data.

IaaS users can compare the cost and performance of different providers in order to choose the best option, as they can access them through a single API.

IaaS Key Features

- Highly scalable resources
- Enterprise-grade infrastructure
- Cost depends on consumption
- Multitenant architecture, i.e. a single piece of hardware serves many users
- The client gets complete control over the infrastructure

IaaS Advantages

- The most flexible and dynamic model
- Cost-effective due to pay-as-you-go pricing
- Easy to use due to the automate d deployment of hardware

- Management tasks are virtualized, so employees have more free time for other tasks

IaaS Disadvantages

- Data security issues due to multitenant architecture
- Vendor outages make customers unable to access their data for a while
- The need for team training to learn how to manage new infrastructure

When to Use IaaS

IaaS can be especially advantageous in some situations:

- If you are a small company or a startup that has no budget for creating your own infrastructure
- If you are a rapidly growing company and your demands are unstable and changeable
- If you are a large company that wants to have effective control over infrastructure but pay only for the resources you actually use

Examples of IaaS

The best-known IaaS solution s vendors are Microsoft Azure, Google Compute Engine (GCE), Amazon Web Services (AWS), Cisco Metapod, DigitalOcean, Linode and Rackspace.

PaaS: Platform as a Service

PaaS in cloud computing is a framework for software creation delivered over the internet. This is the offering of a platform with built-in software components and tools, using which developer s can create, customize, test and launch applications. PaaS vendors manage servers, operating system updates, security patches and backups. Clients focus on app development and data without worrying about infrastructure, middleware and OS maintenance.

The main difference between IaaS and PaaS lies in the degree of control given to users.

PaaS Key Features

- Allows for developing, testing and hosting apps in the same environment
- Resources can be scaled up and down depending on business needs
- Multiple users can access the same app in development
- The user doesn’t have complete control over the infrastructure
- Web services and databases are integrated
- Remote teams can collaborate easily

PaaS Advantages

- PaaS-built software is highly scalable, available and multi-tenant, as it is cloud-based
- The development process is quickened and simplified
- Reduced expenses for creating, testing and launching apps
- Automated company policy
- Reduced amount of coding required
- Allows for easy migrating to the hybrid cloud

PaaS Disadvantages

- Data security issues
- Compatibility of existing infrastructure (not every element can be cloud-enabled)
- Dependency on vendor's speed, reliability and support

When to Use PaaS

Such solutions are especially profitable to developers who want to spend more time coding, testing and deploying their applications. Utilizing PaaS is beneficial when:

- Multiple developers work on one project
- Other vendors must be included
- You want to create your own customized apps

Examples of PaaS

The best-known PaaS solutions vendors are Google App Engine, Amazon AWS, Windows Azure Cloud Services, Heroku, AWS Elastic Beanstalk, Apache Stratos and OpenShift.

SaaS: Software as a Service

With this offering, users get access to the vendor's cloud-based software. Users don't have to download and install SaaS applications on local devices, but sometimes they may need plugins. SaaS software resides on a remote cloud network and can be accessed through the web or APIs. Using such apps, customers can collaborate on projects, as well as store and analyze data.

SaaS is the most common category of cloud computing. The SaaS provider manages everything from hardware stability to app functioning. Clients are not responsible for anything in this model; they only use programs to complete their tasks. In this case, the client software experience is fully dependent on the provider.

SaaS Key Features

- The subscription model of utilizing
- No need to download, install or upgrade software
- Resources can be scaled depending on requirements
- Apps are accessible from any connected device
- The provider is responsible for everything

SaaS Advantages

- No hardware costs
- No initial setup costs
- Automated upgrades
- Cross-device compatibility
- Accessible from any location
- Pay-as-you-go model
- Scalability
- Easy customization

SaaS Disadvantages

- Loss of control
- Limited range of solutions
- Connectivity is a must

When to Use SaaS

Utilizing SaaS is most beneficial in the following situations:

- If your company needs to launch a ready-made software quickly
- For short-term projects that require collaboration
- If you use applications on a temporary basis
- For applications that need both web and mobile access

Examples of SaaS

The best-known SaaS solutions vendors are Google Apps, Dropbox, Gmail, Salesforce, Cisco WebEx, Concur, GoToMeeting, Office365.

ARCHITECTURAL DESIGN CHALLENGES:

Here are six common challenges you must consider before implementing cloud computing technology.

1. Cost

Cloud computing itself is affordable, but tuning the platform according to the company's needs can be expensive. Furthermore, the expense of transferring the data to public clouds can prove to be a problem for short-lived and small-scale projects.

Companies can save some money on system maintenance, management, and acquisitions. But they also have to invest in additional bandwidth, and the absence of routine control in an infinitely scalable computing platform can increase costs.

2. Service Provider Reliability

The capacity and capability of a technical service provider are as important as price. The service provider must be available when you need them. The main concern should be the service provider's sustainability and reputation. Make sure you comprehend the techniques via which a provider observes its services and defends dependability claims.

3. Downtime

Downtime is a significant shortcoming of cloud technology. No seller can promise a platform that is free of possible downtime. Cloud technology makes small companies reliant on their connectivity, so companies with an untrustworthy internet connection probably want to think twice before adopting cloud computing.

4. Password Security

Industrious password supervision plays a vital role in cloud security. However, the more people you have accessing your cloud account, the less secure it is. Anybody aware of your passwords will be able to access the information you store there.

Businesses should employ multi-factor authentication and make sure that passwords are protected and altered regularly, particularly when staff members leave. Access rights related to passwords and usernames should only be allocated to those who require them.

5. Data privacy

Sensitive and personal information that is kept in the cloud should be defined as being for internal use only, not to be shared with third parties. Businesses must have a plan to securely and efficiently manage the data they gather.

6. Vendor lock-in

Entering a cloud computing agreement is easier than leaving it. “Vendor lock-in” happens when altering providers is either excessively expensive or just not possible. It could be that the service is nonstandard or that there is no viable vendor substitute.

It comes down to buyer carefulness. Guarantee the services you involve are typical and transportable to other providers, and above all, understand the requirements.

Cloud computing is a good solution for many businesses, but it’s important to know what you’re getting into. Having plans to address these six prominent challenges first will help ensure a successful experience.

**********END**********

Cloud Storage

Contents

What is Cloud Storage?

- Refers to saving data to an off-site storage system maintained by a third party.
- Instead of storing information to your computer's hard drive or other local storage device, you save it to a remote database.
- The Internet provides the connection between your computer and the database.

Clients rent storage capacity from cloud storage vendors.

In above figure shows that through internet end user can fetch the data anytime, anywhere.

Advantages of Cloud Storage

- **Universal Access:** Data Stored at cloud can be accessed universal. There is no bar.

- **Collaboration**: Any team member can work on same data without their physical existence, means they can fetch data through internet and share their data.

- **Scalability**: In cloud storage data can be increased or decreased according to requirements.
- **Economical:** For large organization data storage on cloud will be economical.
- **Reliability:** Due to redundancy of data, data is more reliable in case of cloud storage.

Storage as a Service

- The term Storage as a Service means that a third-party provider rents space on their storage to end users who lack the budget or capital budget to pay for it on their own.
- Also ideal when technical personnel are not available or have inadequate knowledge to implement and maintain that storage infrastructure.

Cloud Service Providers:

- Google Docs allows users to upload documents, spreadsheets and presentations.
- Web e-mail providers like Gmail, Hotmail and Yahoo! Mail store e-mail messages on their own servers.
- Sites like Flickr and Picasa host millions of digital photographs.
- YouTube hosts millions of user-uploaded video files.
- Web site hosting companies like StartLogic, Hostmonster and GoDaddy store the files and data for client Web sites.
- Social networking sites like Facebook allow members to post pictures and other content.
- Services like Xdrive, MediaMax and Strongspace offer storage space for any kind of digital data.

Challenges in Cloud Storage:

- Security (Data Leakage): Stored as well as data in transit
- Reliability
- Performance

Security can be provided through using a combination of techniques:

- Encryption
- Authentication
- Authorization

- **Reliability: Provided through Redundancy**

Disadvantages of Cloud Storage:

- Outsourcing increases attack surface area.
- Due to more locations, the risk of unauthorized physical access to the data increases.
- Number of people with access to data who could be compromised increases dramatically.
- Increases the networks over which the data travels.
- Access to other customers is also possible.

Type of storage:

DAS: Directly Attached Storage

Understanding DAS Analogy: A simple external USB drive.

Conceptually, the most basic form of directly attached storage is an external hard drive connected via a USB cable. That's as simple as it sounds. DAS. A directly attached storage device. However when we talk about storage we tend to mean multiple drives, an array of disks acting together in some way. The DAS concept is the same whether it's one or 24 drives. Similarly, it's the same concept regardless if we use a different cable. Actually, USB is much too slow for large DAS units.

You have probably already seen some external drives that have multiple disks. Western Digital has a Dual-drive external RAID device. Fancy that, it uses Thunderbolt instead of USB, which is a bit (many bits, actually!) faster, allowing it to saturate the read and write capacity of multiple disks! It might not carry the DAS label, but its conceptually the same.

Since DAS isn't an intuitive or marketing friendly acronym, we aren't going to see the term being used in the lower levels of hardware. "Dual-drive" explains the concept much better to consumers, and is more specific in regards to the Western Digital product offering above. When we move into larger arrays of storage, we begin to see the directly attached storage acronym: DAS. Again simply meaning, we are directly attaching the storage device to the computer without using a network.

A terrible idea, but super cheap and sometimes effective.

In theory, we could take our 6TB Super Desktop and pile on external drives. We so elegantly stuffed disks in there, I'm sure we could carefully cocoon the computer in external USB storage devices too! Using all those external USB drives would be more of an art project though, clearly not a serious business solution. There are a lot of problems with using USB this way, particularly speed and reliability. But let's move on, pass on the USB cocoon idea, and take a look at more serious DAS hardware to get a sense of what's available. What's a more legitimate DAS option?

A much more serious DAS solution.

Check out the Dell PowerVault MD1220 DAS units. These are pretty awesome and have advanced features not found in external drives. They are clearly more power intensive having to feed up to 24 hard disks, versus one or two disks in an external. You definitely can't power these machines over USB! But with equipment like this you're going to want redundant power supplies anyways. Not to mention that when connecting DAS devices of this caliber, you will need specialized hardware — not simple USB ports!

Connecting to DAS devices.

A variety of cables can connect DAS units, including fiber optics, SATA, SAS, etc. Sticking with the Dell MD1220 example above, the computer you're connecting to will need a specialized piece of hardware that provides SAS ports. Dell recommends the PERC H800 for this DAS unit. Just like Apple's Thunderbolt (and the Western Digital product above), we can daisy-chain the MD1220 DAS units for an even larger array. Now we're talking! We could stack 24 drives in one MD1220 with 24 drives from a second MD1220 and so on! I've done things like this, and let me just say that it's very cool to hear so many disks spin and whir up. Imagine writing your files across dozens of disks and watching all those disk lights blink!

DAS units might be a little fancier than external USB drives, but the external hard drive analogy is a good way to visualize the same concept. Your computer will treat the Dell MD1220 similarly to an internal drive, or an external hard drive that you just plugged in. For clarity, directly attached storage units like the Dell MD1220 will require some configuration and setup. They aren't as plug and play as a thumb drive, but they are pretty easy to manage nonetheless.

Things to remember about DAS.

There are some noteworthy things to mention about DAS and how it differs. Generally, you are not able to dedicate the hard disks to multiple computers. SAN and NAS are specifically designed to be shared resources by multiple computers, whereas DAS is designed to be used by (usually) a single computer. The Dell MD1220 actually allows you to split the 24 drives in half, dedicating 12 disks to one computer and the other 12 disks to second computer (if you wanted to do so).

So why would anyone choose DAS for their storage strategy over SAN and NAS? Simple. When it comes to speed, networks are much slower than directly attached storage. Especially in terms of latency, DAS is several times faster than SAN and NAS. Even entry-level direct attached storage is faster than enterprise-level network storage. You can get great performance out of SAN devices, but it just doesn't compare for certain applications.

The main downside to directly attached storage is that it's a dedicated resource for (usually) a single computer, and it can't be managed over a network. While DAS is cheaper in the strict sense, DAS is not necessarily cheaper in all scenarios. SAN units may be much more expensive, but they gain some economies of scale with many networked computers. It really depends on your situation if you're considering it from a cost-perspective. The upside to DAS is performance.

Advantages of DAS: Direct Attached Storage.

- Simpler to setup and configure over NAS / SAN
- Cheaper than NAS / SAN in terms of raw storage
- Networks not necessary, doesn't use IP addresses
- Faster, more performant and better latency over SAN / NAS
- Easier to deal with overall considering all things
- Disadvantages of DAS: Direct Attached Storage

Dedicated resource to a single computer

- No economies of scale in sharing the storage
- Can't manage DAS via a network
- Requires a special hardware connection

The Super Desktop 2.0

So now that we know a little bit more about DAS, we've gone ahead and added a RAID card to our Super Desktop. We had the option to expand the beautifully designed 6TB Super Desktop with a bunch of external hard drives over USB, but instead we spent a little bit more money on Dell's MD1220.

Unfortunately, for the purpose of this example, we decided to fill the DAS unit with a bunch of 73GB, 15K RPM drives in RAID 5. That added a total usable space of a bout 1.5TB. It wasn't the best decision, because we were actually trying to solve our storage capacity issue. So now we are running a very expensive 7.5TB Super Desktop with a big DAS unit connected. For clarity, we still have our 6TB partition and a 1.5TB partition.

We could have done this cheaper with a simple external hard drive, actually that probably would have given us more usable space. But as we learned, professional DAS devices like the MD1220 have the advantage of speed. In fact, they're really quite good for building servers that require low-latency — that is, DAS solutions like the MD1220 are AMAZING for demanding, hard-hitting, speed-absolutely-required applications. Unless we're considering distributed computing, implementing DAS is going to be your best choice for overpowered super servers and mission critical apps, that simply MUST GO ON. Which brings us to a special point.

Speed vs Storage Capacity are two ends of the storage spectrum

On one side of, we have speed. Increasing speed is usually at the cost of Capacity. Likewise, increase Capacity usually comes at the cost of Speed. We could probably add a third dimension to this for Cost. Because with enough money, you can achieve a satisfying mix of both.

Our 1.5TB partition on the Super Desktop 2.0 would be particularly great at running a decent sized database. Easily able to handle thousands of queries per second, no problem. Optimizing specific configurations towards a particular purpose is beyond the scope of this post, but let me mention that RAID 5 might not be your best choice for a heavily used database. Anyways, let's keep moving.

You decided to download Wikipedia and consumed the remaining terabytes of space pretty quickly. The Super Desktop 2.0 is starting to become a real pain in the neck. You could keep expanding with DAS devices, but you really need something more robust, and particularly you want something accessible by all your networked computers. What we need is some sort of network attached storage: NAS!

In theory, you could be using our Super Desktop as a NAS device. It's entirely possible to setup a computer like our Super desktop with a DAS (directly attached storage) array and then share that partition over the network. In fact, that's all a NAS device really is!

NAS: Network Attached Storage

Little did you know the Super Desktop was already a fancy Network Attached Storage device. A network storage device is conceptually simple. Any computer, regardless of number of disks and the size of storage space available, can be considered a NAS if it acts as a file server on the network. Said another way, a network attached storage (NAS) device is just a computer that shares files over the network. That's it!

I think where people become confused about network attached storage devices is when they use a special interface or operating system other than Windows. Or when that NAS has a fancy array of hard drives, perhaps even a DAS-like array built-in similar to the MD1220.

Let's go back to the beginning. Remember that NAS devices are a computer with any amount of disks or storage space available. Western Digital has a starter NAS device. Even though this only has a single hard drive, it's connected to the network, and is therefore a NAS. Conceptually this is almost identical to the external USB hard drive, except instead of a USB cable connection, a NAS will (usually) be using an Ethernet connection — or at least some networking cable.

Connected a NAS to the Network means it has an IP Address.

Computer networking goes beyond this post, but a NAS is essentially assigned an IP address and is then available via that IP over the network. It's worth pointing out that since both SAN and NAS are network storage mediums, they both have some sort of network interface and IP addresses.

You might be wondering to yourself: is this little Western Digital network hard drive really a computer? The answer is yes! It actually has its own processor, RAM, and all that! It's a true computer that probably runs some variant of Linux. You could, in theory, install software on it just like a regular computer, although it might not have the hardware inputs you require (like a mouse, keyboard, and graphics connection for your monitor). But trust me, it's a computer all the same!

Which brings up another important difference. Directly attached storage devices are generally not considered computers. They are more akin to a peripheral like your mouse or keyboard. In other words, DAS devices are simply connected to the computer. Not a computer themselves.

This is in contrast to SAN and NAS devices, which generally can be considered computers. They require a processor, operating system, RAM, and at least one network card. SAN and NAS devices are simply specialized computers that allow OTHER computers to access their storage over the network. There may be technical nuances here, regarding this and the previous statement about DAS devices not being computers, but I'm trying to illustrate overall concepts and not the exact technicalities. Nowadays, just about everything (even SIM cards) have some sort of microprocessor or ALU and could therefore be considered "a computer".

Why does this matter? Who cares if DAS, NAS, or SAN is a computer?

Your computer cares! It treats each differently. Your processor (and potentially RAID card) are going to be responsible for managing the directly attached storage unit. From the perspective of the computer, the DAS is physically a part of the same computer. In the Super Desktop example we've been using, the Desktop thinks the big MD1220 box off to the side is physically apart of the same computer — even though it doesn't physically appear to be enclosed as "one computer".

The above is in contrast with a network attached storage device. Your computer doesn't believe a NAS is physically part of the "same computer". The computer sees a NAS as storage provided by some other computer, accessible over the network. This is a big distinction!

How a DAS appear to computers:

Pretend we are now using a Macbook. If we connect an External USB drive (DAS analogy) to the notebook, we are going to see an additional drive become available. Mac is smart enough to categorize this as a "device" in the Finder sidebar. Your computer doesn't actually believe a USB device is physically a part of the same computer, but again the USB analogy is only meant to help conceptualize true DAS enclosures.

A true DAS enclosure like the Dell MD1220 would appear to be "within" the same computer. It would function as an "internal" drive (depending on how it is partitioned), even though it is clearly not internal in the literal, physical sense. The DAS analogy I provided breaks down partially because USB devices are designed to be Hot Swappable whereas DAS enclosures are not. On Windows the external USB analogy works much better.

On Windows, the USB drive appears and is made available as an unique partition. The USB drive is assigned its own unique letter, behaving much like an internal hard drive would on Windows — appearing as (C:\), (D:\), (E:\), etc. While Windows does also recognizes it is actually an external USB device, the lettered partitions make the DAS concept a easier to grasp. Connecting a Dell MD1220 would simply provide additional drive letters (depending on how you configured and partitioned it) similar to the external USB.

How NAS appears to computers:

Pretend again that we are on our Macbook. If we connected to the 6TB partition on the Super Desktop, it would appear distinctly different than an external USB drive. Remember that a DAS functions as if it were "within" the computer it is connected to. A NAS remains and functions as a separate computer that you will have to connect to with credentials (user and password).

Once connected to a NAS, you will see folders, or more aptly named "Shares". Shares are just special folders that are accessible over the network. NAS solutions are built around the concept of providing users access to "Shares". This is an area beyond the scope of this post, but to offer an example, we could say "My Share" on "Super Desktop" is now shared over the network.

We could even create multiple user logins and provide different levels of access here. You can probably start to piece together why NAS devices are useful. Network attached storage devices

are particularly good are sharing information and allowing for collaboration. Many businesses get a central file server (NAS) setup fairly quickly, so they can collaborate on files more easily across multiple computers. This might be changing a bit with Cloud Storage solutions, but I digress.

On Windows, you can find Shares provided by NAS devices under the “Network” icon. On Mac, you will see a “Shared” listing on the sidebar of Finder which is automatically populated. If the topic of file servers interests you, go ahead and research SMB and Samba.

Since we already mentioned the Dell MD1220 above, we can look at its equivalent NAS form factor. Remember, the major difference between a NAS and DAS is that one is a computer (NAS) and one needs to connect to a computer (DAS). Check out the Dell PowerVault NX3200 NAS. If you click on the specifications, they will actually list the processors within. Aha! Proof it is a computer!

Continuing with the Super Desktop, it would probably be wise to move Wikipedia off our low-capacity, high-speed MD1220 DAS unit, and over onto a new NX3200 NAS. We can repurpose that MD1220 for something more appropriate (like a really demanding application!) Woo!

Things to remember about NAS.

A network attached storage device is a computer that provides Shares to other computers and users over a network. A computer can be considered a NAS device regardless of number of disks and disk size, so long as it shares storage over a network. Usually though, when referring to NAS devices, it’s commonly understood to mean a specialized computer with multiple disk array(s).

A NAS device is a shareable resource. Multiple users and computers can benefit. This contrasts in comparison to a DAS (directly attached storage) device which is dedicated to the computer it directly connects to. The disk arrays in both NAS and DAS are similar in function and operation, meaning you can create similar RAIDS and partition styles on both.

While network attached storage is economical and ideal for collaboration and sharing, it really isn’t the best choice for performance critical applications. The downside to network attached

storage is complexity (introduced by relying on a network) and latency speed. You generally would not want to operate a heavily-used database from a NAS.

Advantages of NAS: Network Attached Storage.

- Economical way to provide large storage to many persons or computers
- Several times easier to setup and configure versus SAN
- Easy way to provide RAID redundancy to mass amount of users
- Allows users permissions, folder privileges, restricted access to documents, etc
- Higher utilization of storage resources

Disadvantages of NAS: Network Attached Storage

- Requires IP Address(es) and takes up network space
- Slower latency and potentially maximum data-transfer issues
- Performance can be affected by network status

Recall our triumph of human engineering from earlier, the Super Desktop 2.0! We have a MD1220 (DAS) connected to it. We are sharing our 6TB over the network (NAS). Now it's time to push our art project into the enterprise realm and start dealing with storage area networks, or SAN!

First, let's remember that NAS and DAS devices appear differently when they are connected to a computer. Understanding how a Shared Folder (from a NAS device) and an external USB device (DAS) differ in appearance is going to help conceptualize our final storage friend. Storage area networks (SAN) have some relationship to "how" these storages appear to computers.

SAN: Storage area networks

The essential difference between SAN and NAS is a clarification in how the network storage is being accessed. Well, actually that's not true. SAN and NAS are both accessed over a network by TCP / IP protocols. What I mean to say, is that other computers access SAN and NAS devices differently. Wait, I just said that? See, it is easy to confuse SAN and NAS. Let's start over.

Remember what DAS is good at? Speed. Speed. SPEED.

Remember what NAS is good at? Sharing. High Utilization. Flexibility.

Combining the best of DAS and NAS.

What if we wanted the performance of DAS with the benefits of NAS? We would find ourselves envisioning some sort of directly attached network storage (kinda) or SAN! And when it comes to enterprise-level, datacenter-level, or virtual computing environments, storage area networks are really the thing to do. The attempt to solve unify the best of both worlds.

I should mention that if you don't have a virtual environment or serious computing needs, a SAN isn't a realistic option, especially if you have a smaller budget. Even if a SAN is within your budget, they really are designed for higher-end computing, virtual computing, complex configurations, and serious economies of scale. This post is intended to help newcomers grasp the concepts of different storage solutions. If you found this post helpful, you will likely need to research networking and many other topics before considering a SAN.

We do have a NAS out on the network, which is really great for storing Wikipedia, but not so great for… say, running an Operating System from. While it's entirely possible to Map a network Drive on Windows and make a NAS Shared folder appear to as a local lettered drive, it's just not the same thing. Although, mapping a network drive is conceptually similar to the function that SANs provide.

I alluded to this earlier, but a SAN is a more sophisticated NAS device that allows others computers to treat its partitions like a DAS. Wow! In other words, a SAN is a network attached storage device, that other computer connect to over a network, and can treat as an "internal" or "directly attached" or "dedicated" hard drive. That does sound intriguing, don't you agree?

How does SAN magic work, you wonder? Didn't I say NAS was slow? (Yes.) How are we going to "directly attach" a network device? And what makes SAN more complex than NAS devices? These are good questions. Let me start by saying that SAN solutions vary quite a bit from implementation to implementation, require (at least some) network planning, and usually require some additional software to deal with. Storage area networks are not for faint of heart.

Firstly, SANs are complex because computer networks are (usually) complex. This is especially true in environments where SANs are commonly needed. While a NAS can be auto-assigned an IP address, SANS will definitely need dedicated IP addresses. Multiple addresses, probably going to different switches and fun stuff like that. This isn't a big deal, but SAN implementations sometimes involve virtual networks, switching rules, and planning for the physical wiring. That stuff is difficult to roll into an already very long post. So for the rest of this post I'll instead be focusing on more surface layer material.

Explaining the storage area network

To help explain the magical SAN, let's revisit our modern marvel, our trusty Super Desktop! If you remember, we have that low-capacity, but super fast 1.5TB partition provided by the MD1220 DAS unit (we put twenty-four 73GB, 15K RPM drives in it.) This DAS deserves some praise, and is quite awesome — you'll see why!

This MD1220 is fast. Really fast actually. It's difficult to imagine an array of 24 disks working together NOT being fast. Especially when we're talking about several thousand dollars worth of computing equipment. You may not know though, there will definitely be diminishing returns for each drive you add to a RAID. After 6 or so disks in RAID 5, the speed doesn't improve much and might get worse at SOME point. If we needed even higher performance, we would probably opt for RAID 10, solid state disks, or maybe a combination of both. Don't shy away from optimizing the application either!

I mentioned this briefly before, but we didn't choose the optimal configuration in our Super Desktop example, but I'm trying to illustrate the point that you should REALLY plan ahead and think designs through! Storage is a long-term decision, it's complicated, nuanced, and mildly annoying. But it is certainly worthy of your undivided attention.

Our Super Desktop wasn't designed to be a SAN, but it wasn't really designed to be a NAS either and we accomplished that. So let's keep going, and do A LOT of pretending along the way! We are going to take our MD1220 and repurpose it for our SAN. But instead of creating Shares (folders) on the network, we are going to be creating and sharing logical partitions. In

other words, our fancy SAN provides entire virtual drives over the network — which is a big contrast with a NAS that simply provides virtual folders over the network.

If you understand the difference between a folder and an entire hard drive, you can start to see why storage area networks are pretty powerful tools. You can't install an Operating System on a NAS — that doesn't even make sense! But you can install Windows, OS X, or Linux on a SAN partition. With SANs, you can actually let other computers boot from the network partitions. Again, we're really just treating a NAS as a directly attached, "internal" drive.

Our Super Desktop design may be the pinnacle of human achievement, but we're actually going to need to get a true SAN device in place for the remainder of this. I mean, there's only so much you can do with Windows Home Premium (haha, this example is amazing.) So let's get serious and put this art project behind us. Say hello to the Dell PowerVault MD3 SAN.

At first glance, the DAS, NAS, and SAN devices all look similar. It should not come as a surprise that many people get them confused, their physical appearance is dominated by a large array of disks.

Addressing confusion about DAS, NAS and SAN.

Actually, when I was first learning about business-class storage, I got hung up in the abstraction details. So let me address some minor points here. The array of disks in each of the DAS, NAS, and SAN function the same on ALL of the enclosures. However, the way in which these enclosures are connected, and put to action, is what separates them into their respective categories.

For clarity, we are able to create RAID 0, 1, 5, 10, 50, 60 logical disks on ALL of these enclosures types (DAS, NAS and SAN). Remember, physical disks are abstracted into logical RAID disks that are created by these enclosures and subsequently "seen" by the computer!

Individual, physical disks make up logical RAID disks. Logical disks are then made available to the computer (DAS: directly attached) or shared over the network (SAN). The operating system is installed on the logical disk, and doesn't care much whether it is provided via DAS or SAN. Once you have an OS running, you can create Share folders on the network (NAS).

A SAN could therefore have several DAS enclosures connected to it. The SAN+DAS enclosures store the virtual (or logical) disks for the operating system. This computer could in turn go on to be configured into a network file server (providing NAS services).

Similarly, a NAS could have several DAS enclosures connected to it. The NAS+DAS enclosures could provide massive amounts of storage to users and computers on the network, but the NAS+DAS solution could not provide "logical disks" to those computers. Special software is generally required to create a SAN.

A DAS needs to be directly connected to a computer, as the name implies!

Moving onwards with SAN from the Super Desktop

We have finally retired our brilliant Super Desktop design. We have moved onto a more legitimate network of devices; the MD3 SAN, the NX3200 NAS, and the MD1220 DAS. But wait!

What will we attach our MD1220 DAS to? Remember that a DAS is akin to a peripheral (like a mouse and keyboard) and is not a computer itself like a NAS or SAN. Was our DAS investment useless!? Au contraire, dear reader! I'm so glad we're talking about this! Because what makes these MD1220 DAS enclosures so very cool, is that we can directly attach them to our NX3200 NAS or our MD3 SAN! Wow!

Dell gets my HIGH FIVE for designing awesome enclosures.

There are hundreds of different form factors and DAS solutions that work similar and can help mitigate the risk of poor planning. Synology provides a few NAS+DAS hybrids that are pretty cool if you want to see other examples of the above. But let's move onto wrapping this post up. We're going to attach our MD1220 DAS to our MD3 and talk about SAN magic.

SAN Magic!

Without getting too advanced, the magic of SAN is combining the benefits of network attached storage and treating it as directly attached storage. Your typical SAN will not only have an array

of disks, but an array of ethernet or network ports to offload data. Multiple network ports is an effort to give SANs a serious speed boost. While SANs may or may not ever perform as well as DAS enclosures, it is possible to get pretty good results.

Consider that a 1Gbps network port will push about 125MB / second of data. Even with 8 x 1Gbps ports, we can only reach 1GB / second. That's probably enough to saturate 8 disks in RAID 10, but the MD1220 has a full array 24 disks! With three RAID 10's (8 physical disk each) we would likely run into some performance issues if we had, say… 3 different servers connected to our SAN, each with very intensive applications running. For reasons like this, DAS can be a smart option.

Although, some SANs do provide multiple 10Gbps network ports. But we can go back and forth between different configurations. The MD3 San actually supports multiple MD1220's daisy chained together. So it's a bit of a nuance (important to keep in mind), that your available network bandwidth and your disk read and write speeds should be comparable and sufficient for one another.

A lot of things can affect SAN performance though. With SANs, every connected computer's disk activity is can (and probably will) affect other computers. This performance hit is exacerbated in virtual environment where logical RAID disks are further divvied up into smaller logical disks. SANs and virtual computing is a sort of double-edge sword in that way — not having to worry about the underlying physical hardware, but not truly guaranteed physical resources. (Some will shoot back that virtual environments do, in fact, guarantee resources… But I'm not interested in discussing the specifics within this post.)

At the very least, SAN performance requires a bit more work to "guarantee" that it is performing well. It's easy to mess up SANs, if the network connections aren't correct, virtual environments aren't correct, if the apps on those virtual computers aren't correct… A lot of variable and plenty of "uncertainty" there. Or, plenty of room for errors. Again, this is another reason why DAS remains an option (we stated earlier we wouldn't bring crazy distributed computing setups into this.). DAS is a great option for databases, file intensive operations, or other disk intensive ops. And it's just a heck of a lot easier to deal with than SAN.

Advantages of SAN: Storage Area Network

- Economies of scale similar to that of of NAS
- Higher hardware utilization, similar to that of NAS
- Speed similar or comparable to DAS
- Allows virtual environments, cloud computing, etc.

Disadvantages of SAN: Storage Area Network

- Performance affected by other SAN users
- Performance limited by network if configured incorrectly
- Better performance will still be found using DAS hardware
- Requires multiple static IP Addresses
- Generally consumes more IP addresses than NAS devices
- Complex networking planning is necessary
- May want to implement virtual networks / different subnets
- Physical network wiring may affect performance
- Physical network wiring should be thoroughly planned
- Generally more expensive than NAS or DAS

Cloud Service Providers

Contents

What is Cloud Service providers?

- Cloud service providers (CSP) are companies that offer network services, infrastructure, or business applications in the cloud. The cloud services are hosted in a data center that can be accessed by companies or individuals using network connectivity.

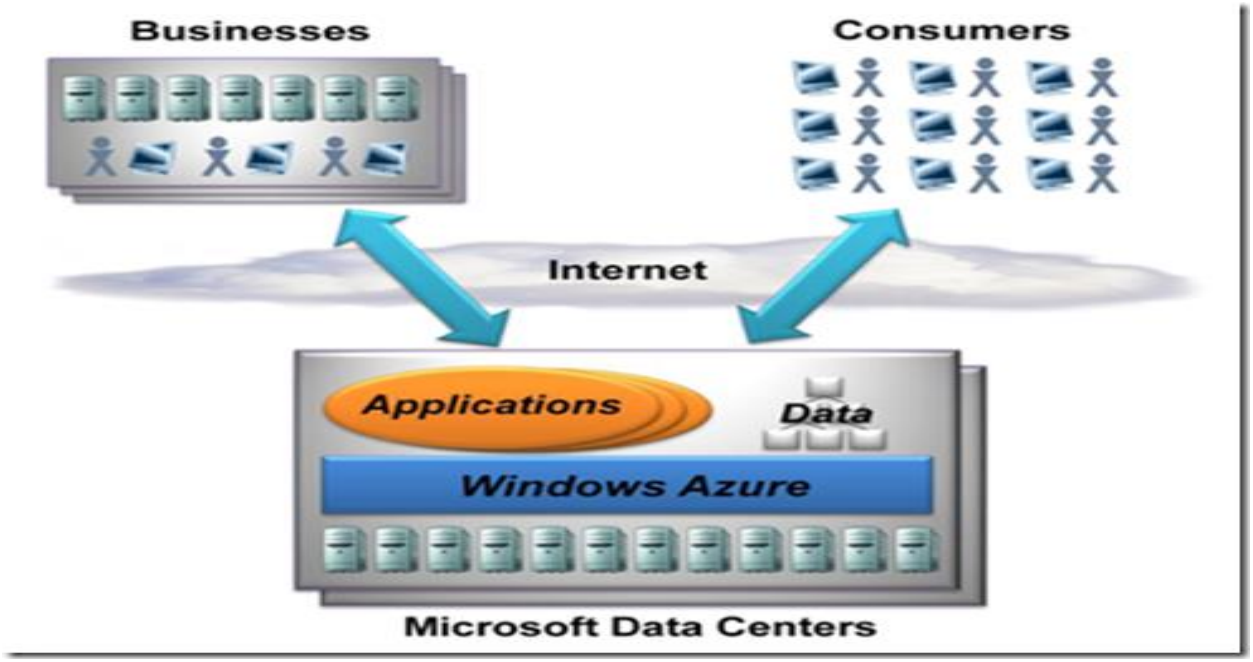
Examples:

- Microsoft
  - Windows Azure (PaaS)
- Amazon
  - EC2 (IaaS)
- Google
  - Google App Engine (PaaS)
- Salesforce.com (SaaS)

Window Azure:

- Microsoft cloud computing platform.

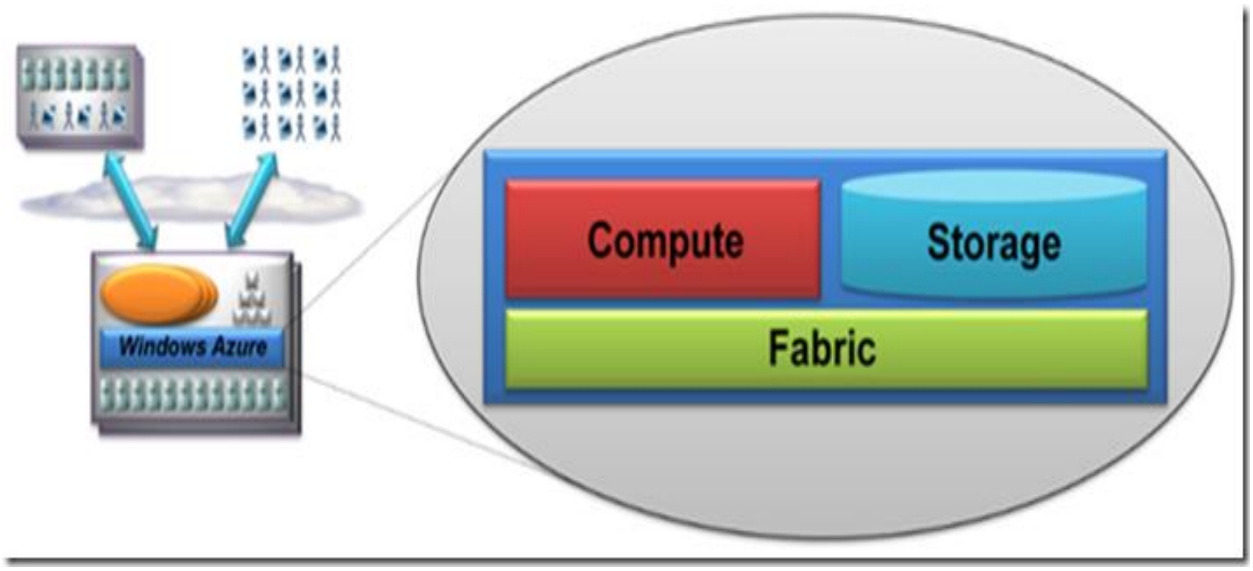
- Used to build, deploy and manage applications through a global network of Microsoft-managed datacenters.
- Allows for applications to be built using many different languages, tools or frameworks(PHP, Java, .NET)



Above figure shows the Microsoft Data Centers in which Window Azure is worked as PaaS, different application can run on this platform and used by end user .

Windows Azure has three core components:

1. **The Compute service**
2. **Storage service**
3. **Fabric.**



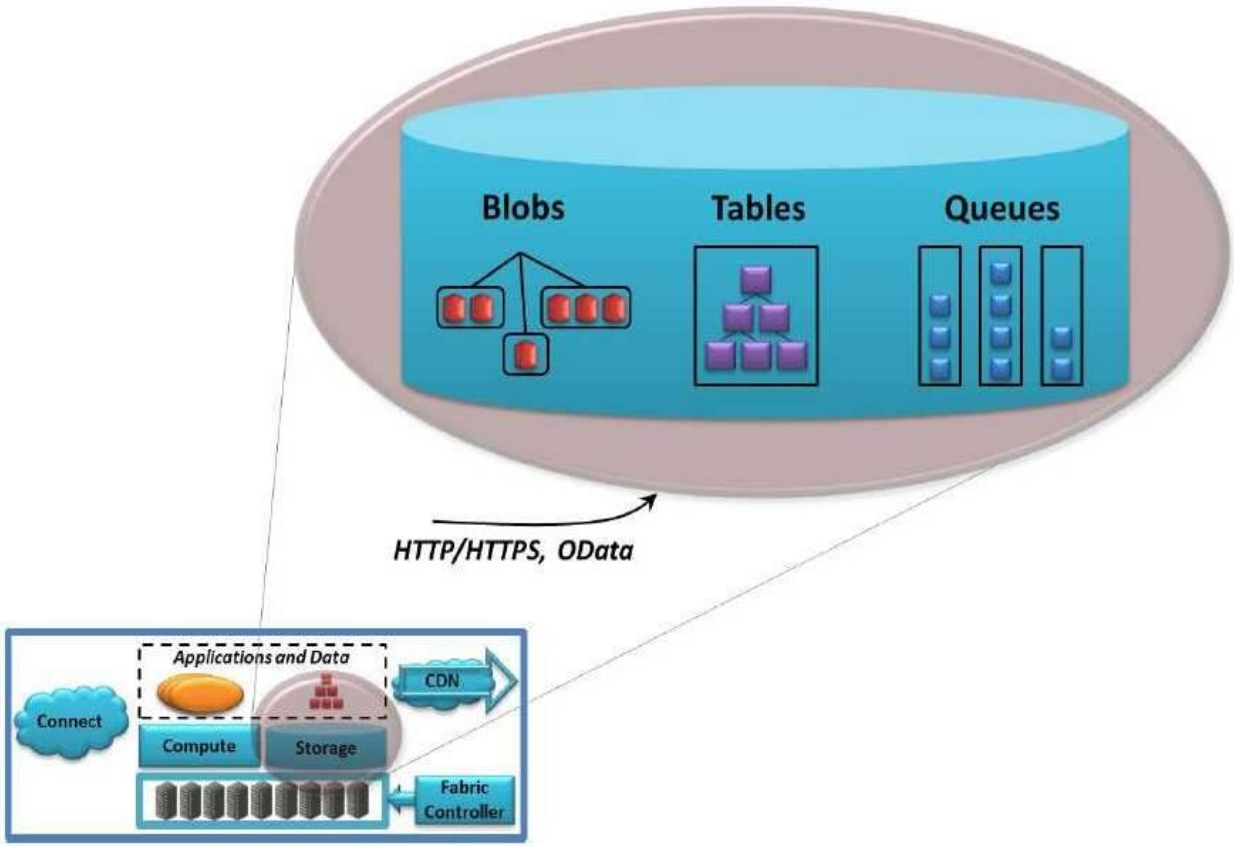
Above figure shows the core components of Windows Azure
Compute : Runs applications
Storage : Stores the data
Fabric : Provides a common way to manage and monitor applications that use this cloud platform.
Compute Service:

- The Windows Azure Compute service can run many different kinds of applications.
- A primary goal of this platform, however, is to support applications that have a very large number of simultaneous users.
- To allow this, a Windows Azure application can have multiple instances, each executing in its own virtual machine (VM).
- These VMs run 64-bit Windows Server 2008, and they're provided by a hypervisor (based on Hyper-V) that's been modified for use in Microsoft's cloud.
- To run an application, a developer accesses the Windows Azure portal through her Web browser, signing in with a Windows Live ID.
- She then chooses whether to create a hosting account for running applications, a storage account for storing data, or both.
- Once the developer has a hosting account, she can upload her application, specifying how many instances the application needs.

Windows Azure then creates the necessary VMs and runs the application.

- It's important to note that a developer can't supply her own VM image for Windows Azure to run. Instead, the platform itself provides and maintains its own copy of Windows.

Storage Service:



- Binary Large Objects – Fundamental storage in Azure
- Store binary data
- Can be as big as 50 GB each
- May have associated metadata

- TABLES
  - To let applications work with data in a more fine-grained way, Windows Azure storage provides tables.

    - These aren't relational tables. In fact, even though they're called "tables", the data they hold is actually stored in a simple hierarchy of *entities that contain properties.*

- Blobs and tables are both focused on storing and accessing data.
- The third option in Windows Azure storage, queues, has a quite different purpose.
- A primary function of queues is to provide a way for Web role instances to communicate with Worker role instances.
- All information held in Windows Azure storage is replicated three times.

FABRIC

- The Windows Azure Fabric consists of a (large) group of machines, all of which are managed by software called the fabric controller.
- The fabric controller is replicated across a group of five to seven machines, and it owns all of the resources in the fabric: computers, switches, load balancers, and more.
- Because it can communicate with a fabric agent on every computer, it's also aware of every Windows Azure application in this fabric.
- It monitors all running applications, for example, giving it an up-to-the-minute picture of what's happening in the fabric.
- It manages operating systems, taking care of things like patching the version of Windows Server 2008 that runs in Windows Azure VMs.

It also decides where new applications should run, choosing physical servers to optimize hardware utilization

AMAZON ELASTIC COMPUTE CLOUD (EC2)

- Amazon Elastic Compute Cloud (Amazon EC2) is a web service that provides resizable compute capacity in the cloud.

- An Amazon Machine Image (AMI) is a special type of pre-configured operating system and virtual application software which is used to create a virtual machine within the Amazon Elastic Compute Cloud (EC2).
- It serves as the basic unit of deployment for services delivered using EC2.

  Service Highlights
- Elastic
    - Capacity can be increased or decreased within minutes, not hours or days.
    - You can commission one, hundreds or even thousands of server instances simultaneously.
- Completely Controlled
- Flexible
    - You have the choice of multiple instance types, operating systems, and software packages.
- Designed for use with other Amazon Web Services.
- Reliable
- The Amazon EC2 Service Level Agreement commitment is 99.95% availability for each Amazon EC2 Region.
- Secure
- When launching Amazon EC2 resources within Amazon Virtual Private Cloud (Amazon VPC), you can isolate your compute instances by specifying the IP range you wish to use, and connect to your existing IT infrastructure using industry-standard encrypted IPsec VPN.

Google App Engine:

Service: PaaS

Tools you know and love: Python, Java, PHP, Go and Cloud SQL.

Used by millions: Snapchat, Rovio, and Khan Academy run on App Engine.

Build and deploy: No worrying about DBAs, servers and load balancers.

Scale: Auto scale to 7 billion requests per day.

Type: Web development

Google App Engine (often referred to as **GAE** or simply **App Engine**) is a Platform as a Service and 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 primarily supports Go, PHP, Java, Python, Node.js, .NET, and Ruby applications, although it can also support other languages via "custom runtimes". The service is free up to a certain level of consumed resources and only in standard environment but not in flexible environment. 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.

Amazon S3

- Online Storage Web Service offered by Amazon Web Services.
- Launched in 2006.
- S3's design aims to provide scalability, high availability, and low latency at commodity costs.
- S3 is designed to provide 99.999999999% durability and 99.99% availability of objects over a given year, though there is no SLA for durability.

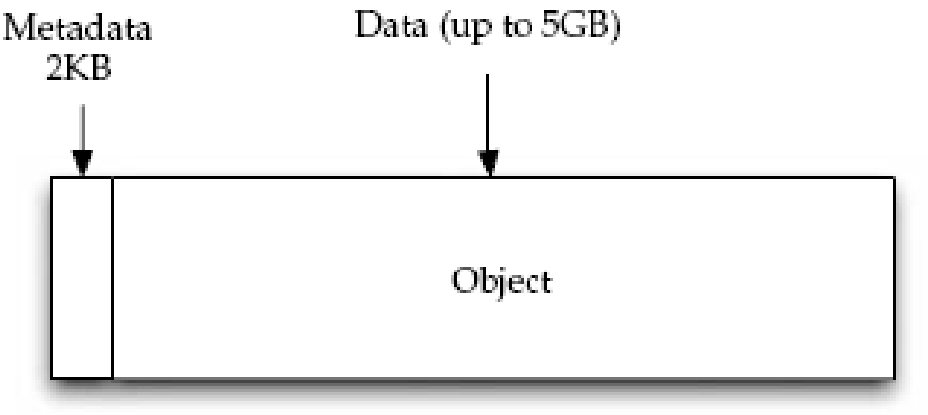
Feature:

- Write, read, and delete objects containing from 1 byte to 5 gigabytes of data each.
- The number of objects that can be stored is unlimited.
- Each object is stored and retrieved via a unique developer-assigned key.
- Objects can be made private or public, and rights can be assigned to specific users.

How S3 works?

- S3 stores arbitrary objects at up to 5GB in size, and each is accompanied by up to 2KB of metadata.

- Objects are organized by *buckets. Each bucket is owned by an AWS account and* the buckets are identified by a unique, user-assigned key.



- Amazon S3 Standard for general-purpose storage of frequently accessed data
-
- Amazon S3 Standard - Infrequent Access (Standard - IA) for long-lived, but less frequently accessed data
- Amazon Glacier for long-term archive.
- Amazon S3 also offers configurable lifecycle policies for managing your data throughout its lifecycle.
- Once a policy is set, your data will automatically migrate to the most appropriate storage class without any changes to your applications.

- As part of the AWS Free Usage Tier, you can get started with Amazon S3 for free.

Upon sign-up, new AWS customers receive 5 GB of Amazon S3 standard storage, 20,000 Get Requests, 2,000 Put Requests, and 15GB of data transfer out each month for one year

- Photo hosting service SmugMug has used S3 since April 2006.

- DropBox, StoreGrid and SyncBlaze are some of the many online backup and synchronization services that use S3 as their storage and transfer facility.
- MineCraft hosts game updates and player skins on the S3 servers.
- Tumblr, Formspring and Posterous images are hosted on the S3 servers.
- The science team at the University of California Berkeley responsible for NASA's

"Stardust@Home" project (http://stardustathome.ssl.berkeley.edu/) is using Amazon S3 to store and deliver the 60 million images that represent the data collected from their dust particle aerogel experiment.