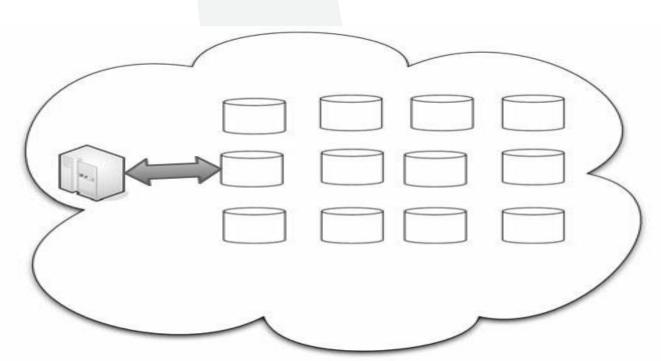


- Cloud storage has a number of advantages over traditional data storage. If you store your data on a cloud, you can get at it from any location that has Internet access.
- Workers don't need to use the same computer to access data nor do they have to carry around physical storage devices.
- There are hundreds of different cloud storage systems, and some are very specific in what they do. Some are nicheoriented and store just email or digital pictures, while others store any type of data.
- Some providers are small, while others are huge and fill an entire warehouse.
- A cloud storage system just needs one data server connected to the Internet. A subscriber copies files to the server over the Internet, which then records the data.
- When a client wants to retrieve the data, he or she accesses the data server with a web-based interface, and the server then either sends the files back to the client or allows the client to access and manipulate the data itself.



- Cloud storage systems utilize dozens or hundreds of data servers. Because servers require maintenance or repair, it is necessary to store the saved data on multiple machines, providing redundancy.
- Without that redundancy, cloud storage systems couldn't assure clients that they could access their information at any given time.
- Most systems store the same data on servers using different power supplies. That way, clients can still access their data even if a power supply fails.
- Many clients use cloud storage not because they've run out of room locally, but for safety. If something happens to their building, then they haven't lost all their data.





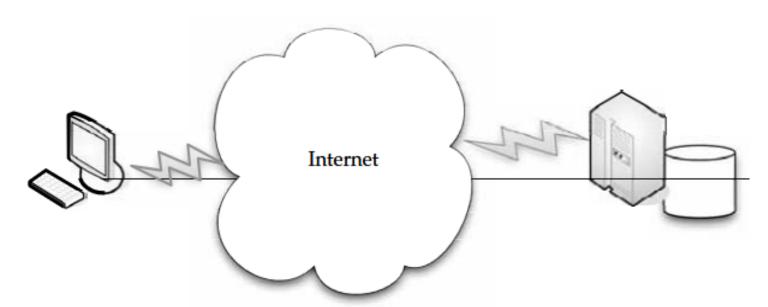
A cloud service provider can simply add more commodity hard drives to increase the organization's capacity.



Storage as a Service

- The term Storage as a Service (another Software as a Service, or SaaS, acronym) 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.
- It is also ideal when technical personnel are not available or have inadequate knowledge to implement and maintain that storage infrastructure.
- Storage service providers are nothing new, but given the complexity of current backup, replication, and disaster recovery needs, the service has become popular, especially among small and medium-sized businesses.
- The biggest advantage to SaaS is cost savings. Storage is rented from the provider using a cost-per-gigabyte-stored or cost-per-data-transferred model.
- The end user doesn't have pay for infrastructure; they simply pay for how much they transfer and save on the provider's servers.
- A customer uses client software to specify the backup set and then transfers data across a WAN. When data loss occurs, the customer can retrieve the lost data from the service provider.





Clients rent storage capacity from cloud storage vendors.



can then be edited using a Google application.

Cloud Storage

Go, change the world[®]

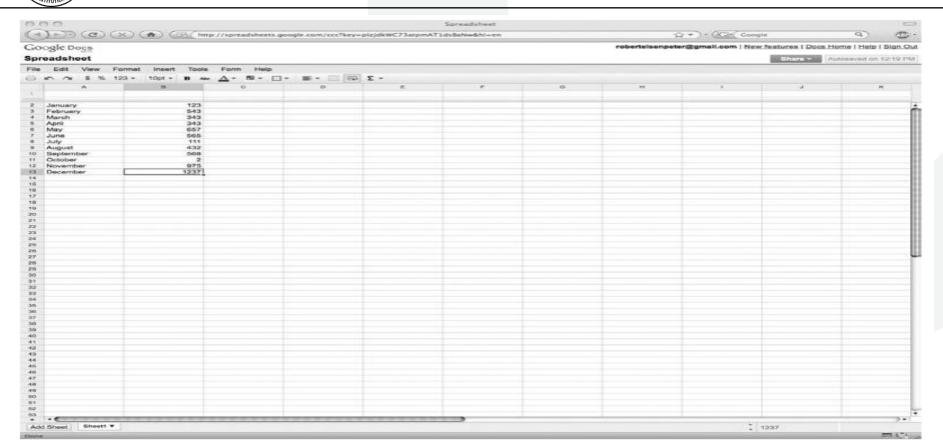
Providers:

- There are hundreds of cloud storage providers on the Web, and more seem to be added each day. Not only are there general-purpose storage providers, but there are some that are very specialized in what they store
- Google Docs allows users to upload documents, spreadsheets, and presentations to Google's data servers. Those files
- Web email providers like Gmail, Hotmail, and Yahoo! Mail store email messages on their own servers. Users can access their email from computers and other devices connected to the Internet.
- access their email from computers and other devices connected to the Internet.
 Flickr and Picasa host millions of digital photographs. Users can create their own online photo albums. YouTube hosts

millions of user-uploaded video files. Hostmonster and GoDaddy store files and data for many client web sites.

- Facebook and MySpace are social networking sites and allow members to post pictures and other content. That content is stored on the company's servers. MediaMax and Strongspace offer storage space for any kind of digital data.
- Many of these services are provided for free, but others charge you per stored gigabyte and by how much information
 is transferred to and from the cloud. More and more providers offer their services, prices have tended to drop, and
 some companies offer a certain amount for free.

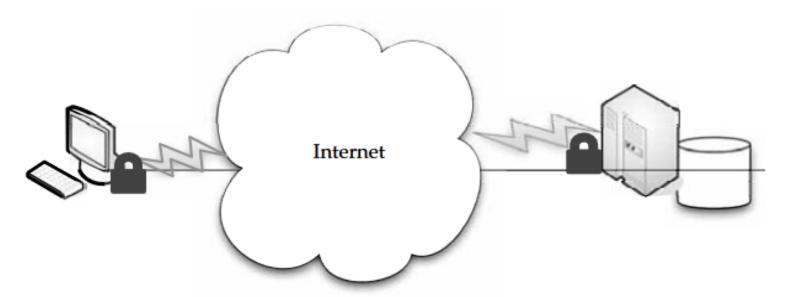
Go, change the world[®]



Security:

- Encryption A complex algorithm is used to encode information. To decode the encrypted files, a user needs the encryption key.
- While it's possible to crack encrypted information, it's very difficult and most hackers don't have access to the amount of computer processing power they would need to crack the code.
- Authentication processes This requires a user to create a name and password. Authorization practices The client lists the people who are authorized to access information stored on the cloud system.
- Many corporations have multiple levels of authorization. For example, a front-line employee might have limited access to data stored on the cloud and the head of the IT department might have complete and free access to everything.
- There are still concerns that data stored on a remote system is vulnerable. There is always the concern that a hacker will find a way into the secure system and access the data. Also, a disgruntled employee could alter or destroy the data using his or her own access credentials.





Encryption and authentication are two security measures you can use to keep your data safe on a cloud storage provider.

Go, change the world[®]

Reliability:

- The other concern is reliability. If a cloud storage system is unreliable, it becomes a liability. No one wants to save data on an unstable system, nor would they trust a company that is financially unstable.
- Most cloud storage providers try to address the reliability concern through redundancy, but the possibility still exists that the system could crash and leave clients with no way to access their saved data.
- Reputation is important to cloud storage providers. If there is a perception that the provider is unreliable, they won't have many clients. And if they are unreliable, they won't be around long, as there are so many players in the market.

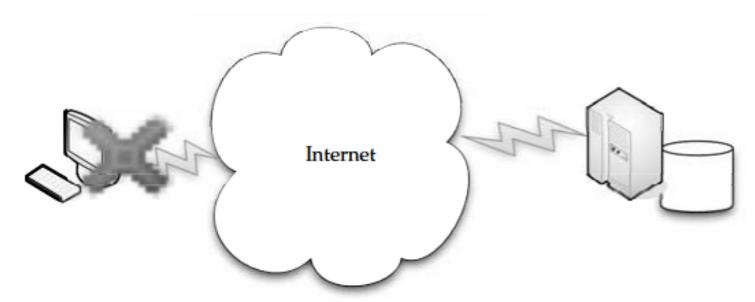
Advantages:

- Cloud storage is becoming an increasingly attractive solution for organizations. That's because with cloud storage, data resides on the Web, located across storage systems rather than at a designated corporate hosting site.
- Cloud storage providers balance server loads and move data among various datacenters, ensuring that information is stored close and thereby available quickly to where it is used.



- Storing data on the cloud is advantageous, because it allows you to protect your data in case there's a disaster. You may have backup files of your critical information, but if there is a fire or a hurricane wipes out your organization, having the backups stored locally doesn't help.
- Having your data stored off-site can be the difference between closing your door for good or being down for a few days or weeks.
- Storage vendor to go with can be a complex issue, and how your technology interacts with the cloud can be complex.
- For instance, some products are agent-based, and the application automatically transfers information to the cloud via FTP. But others employ a web front end, and the user has to select local files on their computer to transmit.
- Amazon S3 is the best-known storage solution, but other vendors might be better for large enterprises. For instance, those who offer service level agreements and direct access to customer support are critical for a business moving storage to a service provider.





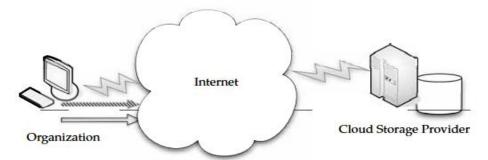
If there is a catastrophe at your organization, having your files backed up at a cloud storage provider means you won't have lost all your data.

Cautions:

- A mixed approach might be the best way to embrace the cloud, since cloud storage is still immature. That is, don't commit everything to the cloud, but use it for a few, noncritical purposes.
- Large enterprises might have difficulty with vendors like Google or Amazon, because they are forced to rewrite solutions for their applications and there is a lack of portability.
- A vendor like 3tera, however, supports applications developed in LAMP, Solaris, Java, or Windows.NET.
- The biggest deal-breakers when it comes to cloud storage seem to be price and reliability.
- This is where you have to vet your vendor to ensure you're getting a good deal with quality service. One mistake on your vendor's part could mean irretrievable data.
- A lot of companies take the "appetizer" approach, testing one or two services to see how well they mesh with their existing IT systems. It's important to make sure the services will provide what you need before you commit too much to the cloud.



- Legal issues are also important. For instance, if you have copyrighted materially like music or video that you want to maintain on the cloud, such an option might not be possible for licensing reasons.
- Vendors offer different assurances with the maintenance of data. They may offer the service, but make sure you know exactly what your vendor will or will not do in case of data loss or compromise.
- The best solution is to have multiple redundant systems: local and offsite backup; sync and archive.



Many companies test out a cloud storage vendor with one or two services before committing too much to them. This "appetizer" approach ensures the provider can give you what you want.



Outages:

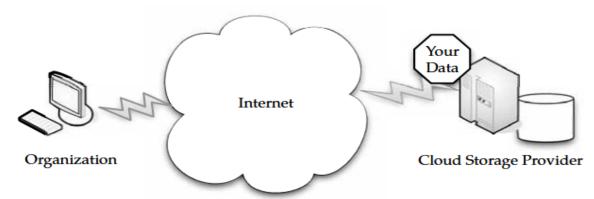
- Further, organizations have to be cognizant of the inherent danger of storing their data on the Internet. Amazon S3, for example, dealt with a massive outage in February 2008.
- The result was numerous client applications going offline. Amazon reports that they have responded to the problem, adding capacity to the authentication system blamed for the problem.
- They also note that no data was lost, because they store multiple copies of every object in several locations.
- The point remains, however, that clients were not able to access their data as they had intended, and so you need to use caution when deciding to pursue a cloud option.

Theft:

• You should also keep in mind that your data could be stolen or viewed by those who are not authorized to see it. Whenever your data is let out of your own datacenter, you risk trouble from a security point of view.



- Storage providers put everything into one pot. Your company's data could be stored next to a competitor's, and the risk of your competition seeing your proprietary information is real.
- If you do store your data on the cloud, make sure you're encrypting data and securing data transit with technologies like SSL.



Whenever you let your data out of your organization, you give up a measure of security.



Cloud Storage Providers:

- There are hundreds of them and new players every day. This is simply a listing of what some of the big players in the game have to offer, and you can use it as a starting guide to determine if their services match your needs.
- Amazon and Nirvanix are the current industry top dogs, but many others are in the field, including some well-known names. Google is ready to launch its own cloud storage solution called GDrive.
- EMC is readying a storage solution, and IBM already has a number of cloud storage options called Blue Cloud.

Amazon Simple Storage Service (S3):

- The best-known cloud storage service is Amazon's Simple Storage Service (S3), which launched in 2006. Amazon S3 is designed to make web-scale computing easier for developers.
- Amazon S3 provides a simple web services interface that can be used to store and retrieve any amount of data, at any time, from anywhere on the Web.
- It gives any developer access to the same highly scalable data storage infrastructure that Amazon uses to run its own global network of web sites. The service aims to maximize benefits of scale and to pass those benefits on to developers.



Amazon S3 is intentionally built with a minimal feature set that includes the following functionality:

- 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.
- Uses standards-based REST and SOAP interfaces designed to work with any Internet-development toolkit.

Design Requirements:

- Amazon built S3 to fulfill the following design requirements:
- Scalable Amazon S3 can scale in terms of storage, request rate, and users to support an unlimited number of webscale applications.



- Reliable Store data durably, with 99.99 percent availability. Amazon says it does not allow any downtime.
- Fast Amazon S3 was designed to be fast enough to support high-performance applications. Server-side latency must be insignificant relative to Internet latency. Any performance bottlenecks can be fixed by simply adding nodes to the system.
- Inexpensive Amazon S3 is built from inexpensive commodity hardware components. As a result, frequent node failure is the norm and must not affect the overall system. It must be hardware-agnostic, so that savings can be captured as Amazon continues to drive down infrastructure costs.
- Simple Building highly scalable, reliable, fast, and inexpensive storage is difficult. Doing so in a way that makes it easy to use for any application anywhere is more difficult. Amazon S3 must do both.
- A forcing function for the design was that a single Amazon S3 distributed system must support the needs of both internal Amazon applications and external developers of any application.
- This means that it must be fast and reliable enough to run Amazon.com's web sites, while flexible enough that any developer can use it for any data storage need.



Design Principles:

- Amazon used the following principles of distributed system design to meet Amazon S3 requirements:
- Decentralization It uses fully decentralized techniques to remove scaling bottlenecks and single points of failure.
- Autonomy The system is designed such that individual components can make decisions based on local information.
- Local responsibility Each individual component is responsible for achieving its consistency; this is never the burden of its peers.

Controlled concurrency Operations are designed such that no or limited concurrency control is required.

- Failure toleration The system considers the failure of components to be a normal mode of operation and continues operation with no or minimal interruption.
- Controlled parallelism Abstractions used in the system are of such granularity that parallelism can be used to improve performance and robustness of recovery or the introduction of new nodes.



Go, change the world[®]

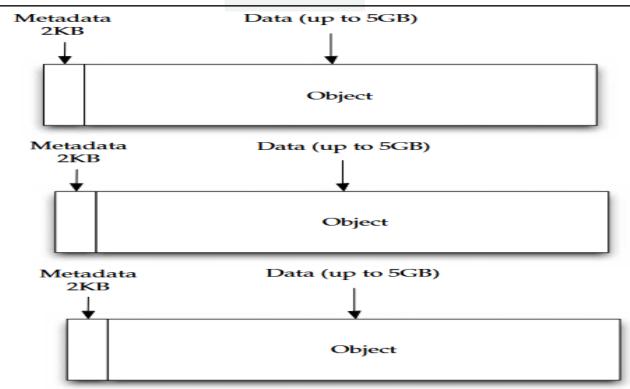
- Small, well-understood building blocks Do not try to provide a single service that does everything for everyone, but instead build small components that can be used as building blocks for other services.
- Symmetry Nodes in the system are identical in terms of functionality, and require no or minimal node-specific configuration to function. Simplicity The system should be made as simple as possible, but no simpler.

How S3 Works:

- Amazon keeps its lips pretty tight about how S3 works, but according to Amazon, S3's design aims to provide scalability, high availability, and low latency at commodity costs.
- 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.
- Buckets and objects are created, listed, and retrieved using either a REST-style or SOAP interface. Objects can also be retrieved using the HTTP GET interface or via BitTorrent. An access control list restricts who can access the data in each bucket.
- Bucket names and keys are formulated so that they can be accessed using HTTP. Requests are authorized using an access control list associated with each bucket and object, for instance:

Go, change the world $^{\circ}$

- http://s3.amazonaws.com/examplebucket/examplekey
- http://examplebucket.s3.amazonaws.com/examplekey
- The Amazon AWS Authentication tools allow the bucket owner to create an authenticated URL with a set amount of time that the URL will be valid.
- For instance, you could create a link to your data on the cloud, give that link to someone else, and they could access your data for an amount of time you predetermine, be it 10 minutes or 10 hours.
- Bucket items can also be accessed via a BitTorrent feed, enabling S3 to act as a seed for the client. Buckets can also be set up to save HTTP log information to another bucket.
- This information can be used for later data mining. "Amazon S3 is based on the idea that quality Internet-based storage should be taken for granted," said Andy Jassy, vice president of Amazon Web Services.
- "It helps free developers from worrying about where they are going to store data, whether it will be safe and secure, if it will be available when they need it, the costs associated with server maintenance, or whether they have enough storage available.
- Amazon S3 enables developers to focus on innovating with data, rather than figuring out how to store it."



Multiple objects are stored in buckets in Amazon S3.

Go, change the world

Early S3 Applications:

- 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.
- These images will be delivered to 100,000 volunteers around the world who scan the images looking for dust particles from comet Wild2.
- "We quickly ran into challenges when we started the project using our own infrastructure," said Andrew Westphal, project director of Stardust@Home. "Using Amazon S3 has allowed us to proceed without having to worry about building out the massive storage infrastructure
- The fact that Amazon S3 is an Internet-connected storage service is particularly useful to us as we expect the data examination phase of the project to take only a few months. We can quickly ramp up and back down again without a huge investment."

Nirvanix:

• Nirvanix uses custom-developed software and file system technologies running on Intel storage servers at six locations on both coasts of the United States. They continue to grow, and expect to add dozens more server location.



SDN Features:

- Nirvanix Storage Delivery Network (SDN) turns a standard 1U server into an infinite capacity network attached storage (NAS) file accessible by popular applications and immediately integrates into an organization's existing archive and backup processes.
- "Up until recently, cloud storage has primarily served as an on-tap back end for application developers," said Adam Couture, principal analyst at Gartner.
- "Today, we're starting to see enterprises begin to consider cloud storage as a low-cost storage tier for selective applications such as backup and archiving."
- Nirvanix has built a global cluster of storage nodes collectively referred to as the Storage Delivery Network (SDN), powered by the Nirvanix Internet Media File System (IMFS).
- The SDN intelligently stores, delivers, and processes storage requests in the best network location, providing the best user experience in the marketplace.



- With the ability to store multiple file copies in multiple geographic nodes, the SDN enables unparalleled data availability for developers, businesses, and enterprises.
- The Nirvanix CloudNAS for Linux mounts the Nirvanix Storage Delivery Network as a virtual drive that can be accessed via NFS, CIFS, or FTP.
- After installation, storage administrators can apply standard file, directory, or access permissions, and users on the network can then access the Nirvanix-mapped drive from their existing applications or storage processes.
- Additionally, storage administrators get access to the robust Nirvanix SDN functionality such as automated policy-based file replication, single global namespace that scales to petabytes, and storage of secure, encrypted data on one or more of Nirvanix's globally clustered storage nodes.



Benefits of CloudNAS:

- The benefits of cloud network attached storage (CloudNAS) include Cost savings of 80–90 percent over managing traditional storage solutions Elimination of large capital expenditures while enabling 100 percent storage utilization Encrypted offsite storage that integrates into existing archive and backup processes.
- Built-in data disaster recovery and automated data replication on up to three geographically dispersed storage nodes for a 100 percent SLA Immediate availability to data in seconds, versus hours or days on offline tape Nirvanix
- CloudNAS is aimed at companies that maintain repositories of archival, backup, or unstructured data that requires long-term, secure storage, or organizations that use automated processes to transfer files to mapped drives.
- Example use cases include long term archiving of data leveraging an established backup/archival solution; departments using a centralized, shared data repository; disk-to-disk-to-cloud replacing tape for archival of data; and simple backup of all computers within a department.



Google Bigtable Datastore:

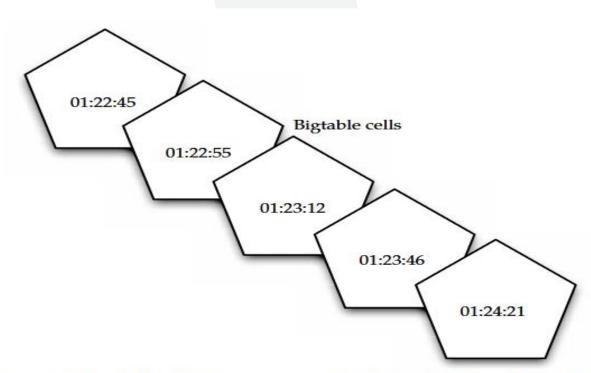
- In cloud computing, it's important to have a database that is capable of handling numerous users on an on-demand basis. To serve that market, Google introduced its Bigtable.
- Google started working on it in 2004 and finally went public with it in April 2008.
- Bigtable was developed with very high speed, flexibility, and extremely high scalability in mind. A Bigtable database can be petabytes in size and span thousands of distributed servers.
- Bigtable is available to developers as part of the Google App Engine, their cloud computing platform.

How Bigtable Works:

- Bigtable is a complex offering that is not easy to understand. If you have trouble sleeping, they offer a very technical explanation at http://labs.google.com/papers/bigtable-osdi06.pdf. Google describes Bigtable as a fast and extremely scalable DBMS. This allows Bigtable to scale across thousands of commodity servers that can collectively store petabytes of data.
- Each table in Bigtable is a multidimensional sparse map. That is, the table is made up of rows and columns, and each cell has a timestamp. Multiple versions of a cell can exist, each with a different timestamp. With this stamping, you can select certain versions of a web page, or delete cells that are older than a given date and time.



- The tables are so large, Bigtable splits them at row boundaries and saves them as tablets. Each tablet is about 200MB, and each server houses 100 tablets.
- Given this, data from a database is likely to be stored in many different servers—maybe not even in the same geographic location.
- This architecture also allows for load balancing. If one table is getting a lot of queries, it can remove other tablets or move the busy table to another machine that is not as busy.
- Also, if a machine fails, since the tablet is spread to different machines, users may not even notice the outage.
- When a machine fills up, it compresses some tablets using a Google-proprietary technique. On a minor scale, only a few tablets are compressed. On a large scale, entire tablets are compressed, freeing more drive space.
- Bigtable tablet locations are stored in cells, and looking them up is a three-tiered system. Clients point to the META0 table. META0 then keeps track of many tables on META1 that contain the locations of the tablets. Both META0 and META1 make use of prefetching and caching to minimize system bottlenecks.



In Google Bigtable, multiple copies of a cell exist, each with a different timestamp.



Issues:

- While Bigtable is a robust tool, developers have been cautious about using it. Because it is a proprietary system, they get locked into Google. That is also the case with Amazon's Web Services and other cloud providers.
- On the other hand, Google App Engine and Bigtable are affordable, costing about the same as Amazon's S3.

Costs are as follows:

- \$0.10–\$0.12 per CPU core-hour
- • \$0.15–\$0.18 per GB-month of storage
- \$0.11–\$0.13 per GB of outgoing bandwidth
- • \$0.09–\$0.11 per GB of incoming bandwidth



MobileMe:

- MobileMe is Apple's solution that delivers push email, push contacts, and push calendars from the MobileMe service in the cloud to native applications on iPhone, iPod touch, Macs, and PCs. MobileMe also provides a suite of ad-free web applications that deliver a desktop like experience through any modern browser.
- MobileMe applications (<u>www.me.com</u>) include Mail, Contacts, and Calendar, as well as Gallery for viewing and sharing photos and iDisk for storing and exchanging documents online.

MobileMe Features:

- With a MobileMe email account, all folders, messages, and status indicators look identical whether checking email on iPhone, iPod touch, a Mac, or a PC. New email messages are pushed instantly to iPhone over the cellular network or Wi-Fi, removing the need to manually check email and wait for downloads.
- Push also keeps contacts and calendars continuously up to date so changes made on one device are automatically pushed up to the cloud and down to other devices.

Go, change the world^{*}

- Push works with the native applications on iPhone and iPod touch, Microsoft Outlook for the PC, and Mac OS X applications, Mail, Address Book, and iCal, as well as the MobileMe web application suite.
- MobileMe web applications provide a desktop-like experience that allows users to drag and drop, click and drag, and even use keyboard shortcuts.
- MobileMe provides anywhere access to Mail, Contacts, and Calendar, with a unified interface that allows users to switch between applications with a single click, and Gallery makes it easy to share photos on the Web in stunning quality.
- Gallery users can upload, rearrange, rotate, and title photos from any browser; post photos directly from an iPhone; allow visitors to download print-quality images; and contribute photos to an album.
- MobileMe iDisk lets users store and manage files online with drag-and-drop filing and makes it easy to share documents too large to email by automatically sending an email with a link for downloading the file.
- MobileMe includes 20GB of online storage that can be used for email, contacts, calendar, photos, movies and documents.

Pricing and Requirements:

- MobileMe is a subscription-based service with 20GB of storage for US\$99 per year for individuals and US\$149 for a Family Pack, which includes one master account with 20GB of storage and four Family Member accounts with 5GB of storage each.
- A free, 60-day MobileMe trial at www.apple.com/mobileme and current Mac members will be automatically upgraded to MobileMe accounts.
- MobileMe subscribers can purchase an additional 20GB of storage for US\$49 or 40GB of storage for US\$99 annually.
- Using an iPhone or iPod touch with MobileMe requires iPhone 2.0 software and iTunes 7.7 or later. For use with a Mac, MobileMe requires Mac OS X Tiger 10.4.11 or the latest version of Mac OS X Leopard.
- For a PC, MobileMe requires Windows Vista or Windows XP Home or Professional (SP2), and Microsoft Outlook 2003 or later is recommended. MobileMe is accessible on the Web via Safari 3, Internet Explorer 7, and Firefox 2 or later.



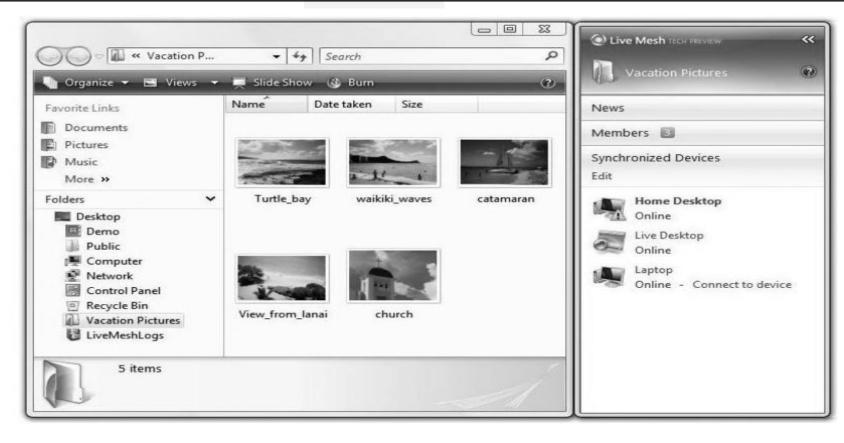
Live Mesh:

• Live Mesh is Microsoft's "software-plus-services" platform and experience that enables PCs and other devices to be aware of each other through the Internet, enabling individuals and organizations to manage, access, and share their files and applications seamlessly on the Web and across their world of devices.

Live Mesh has the following components:

- A platform that defines and models a user's digital relationships among devices, data, applications, and people made available to developers through an open data model and protocols.
- A cloud service providing an implementation of the platform hosted in Microsoft datacenters.
- Software, a client implementation of the platform that enables local applications to run offline and interact seamlessly with the cloud.
- A platform experience that exposes the key benefits of the platform for bringing together a user's devices, files and applications, and social graph, with news feeds across all of these.
- Microsoft promises an open data model, and developers will be able to help Live Mesh grow through the development of additional applications and services.







- The Live Mesh software, called Mesh Operating Environment (MOE), is available for
- Windows XP
- Windows Vista
- Windows Mobile
- Mac OS X
- The software is used to create and manage the synchronization relationships between devices and data. Live Mesh also incorporates a cloud component, called Live Desktop.
- This is an online storage service that allows synchronized folders to be accessible via a web site.
- It also includes remote desktop software called Live Mesh Remote Desktop, which can be used to remotely connect and manage any of the devices in the synchronization relationship.
- Live Mesh Remote Desktop allows you to control your devices from the Live Mesh application, as well as from any other PC connected to the Internet.

Go, change the world

Live Framework:

- For developers, there is a development component consisting of a protocol and APIs known as Live Framework. Live Framework is a REST-based API for accessing the Live Mesh services over HTTP.
- Live Framework differs from MOE in that MOE simply lets folders be shared. The Live Framework APIs can be used to share any data item between devices that recognize the data.
- The API encapsulates the data into a Mesh Object, which is the synchronization unit of Live Mesh. It is then tracked for changes and synchronization.
- A Mesh Object consists of data feeds, which can be represented in Atom, RSS, JSON, or XML. The MOE software also creates Mesh Objects for each Live Mesh folder so they can be synchronized.
- Like cloud computing itself, cloud storage takes its fair share of knocks for being used as a trendy term. If the term is used too often, it could wind up referring to any type of Internet-accessible storage.
- Organizations should think of cloud computing as scalable IT capabilities that are delivered to external customers using the Web.