

## The Promise & Challenges of Cloud Storage

- The Promise of Cloud Storage
- The Challenges of Cloud Storage
- Types of Cloud Storage Solutions Available
- Delivering an Information-Centric Cloud Storage Infrastructure

### The Promise of Cloud Storage

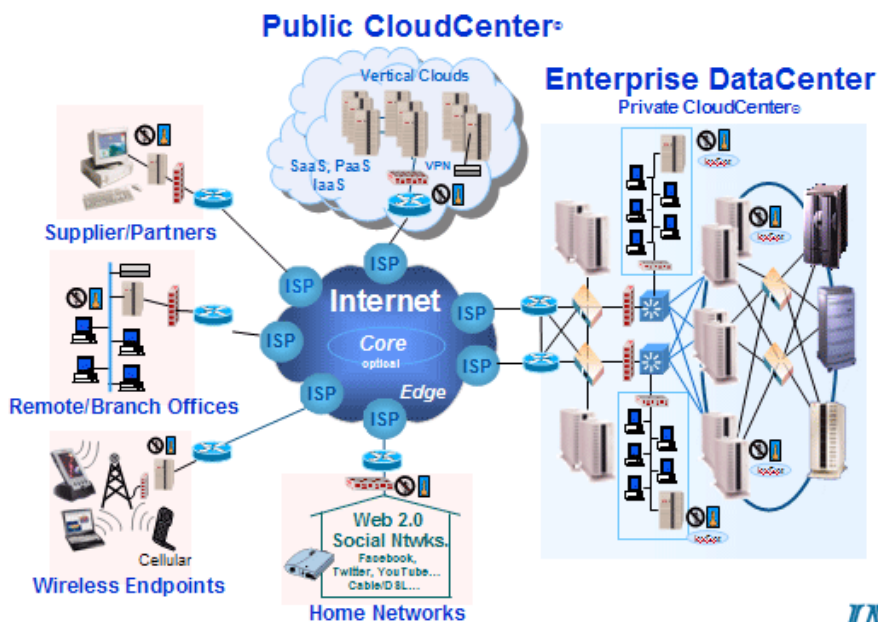
With the onset of cloud computing, and more specifically cloud storage, organizations see potential in being able to reduce cost and complexity for these key applications. The promise of capacity on-demand – being able to pay only for what is used and for what is transferred – is appealing,

### ► Enterprise and Cloud Infrastructure (Secure, HA, Interoperable)

#### Industry Report

Table of Contents

- 1 Executive Summary
- 2 Market Drivers/ Industry Dynamics
- 3 Market Segmentation & Product Requirements
- 4 Market Demand Forecast & Market Shares
- 5 Enabling Technologies - Trends & Standards
- 6 Competitive Product Positioning & Strategies
- 7 Suppliers: Portfolios & Strategies
- 8 Channels of Distribution
- 9 Recommendations
- 10 Methodology & Appendices



©2003-2010 IMEX Research All rights Reserved



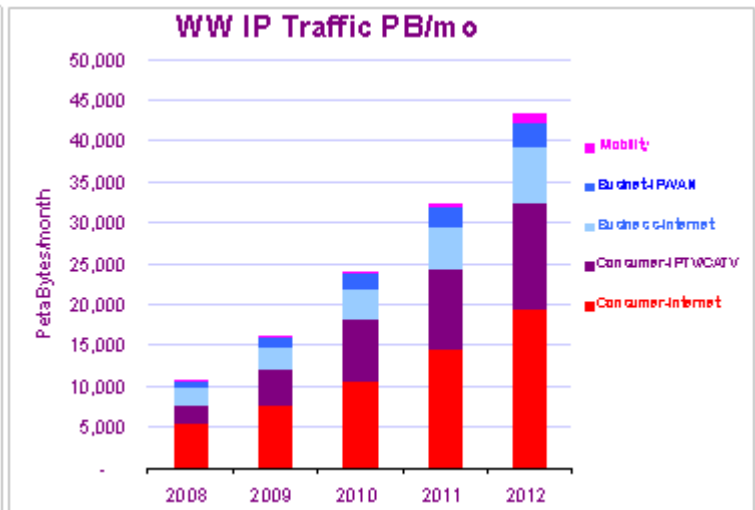
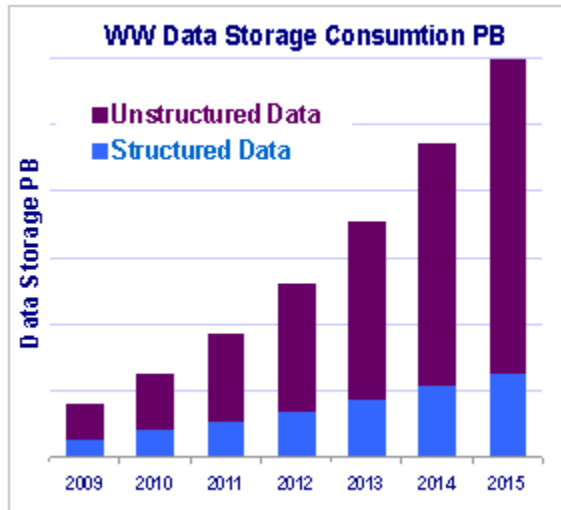
especially given that the price of (most) public cloud storage services are pennies on the dollar compared to traditional on-premises data center storage. The economies of scale and shared infrastructure allow service providers to deliver cloud-based storage at extremely low price points compared to that of a traditional infrastructure. But cloud-based computing and storage go far beyond this benefit by extending storage capabilities through the use of advanced API sets.

For IT customers, cloud-based storage is scalable, accessible, manageable and even more distributable, than a traditional storage infrastructure. Once the locality of data becomes irrelevant, users can integrate data from anywhere. And when the location of data is no longer important, it is easy to scale performance by distributing or moving data across any system according to demand. Exponential data growth particularly coming from unstructured data, is forcing companies to look at new, cost-effective ways to move their information from expensive primary disks to archive storage. Additionally an explosion of interest in large-scale, elastic and cloud-based data storage and processing using open-source Hadoop extract new value from both complex and structured data is being implemented both in-house and using cloud-based infrastructure.

Now cloud storage is being adopted as a reliable platform for long term archive needs given that new efficient tools are now available from vendors allowing users to quickly and securely move corporate information to the cloud.

1474 Camino Robles Ct.  
San Jose, CA - 95120  
Tel: 408.268.0800  
Fax: 408.268.2300

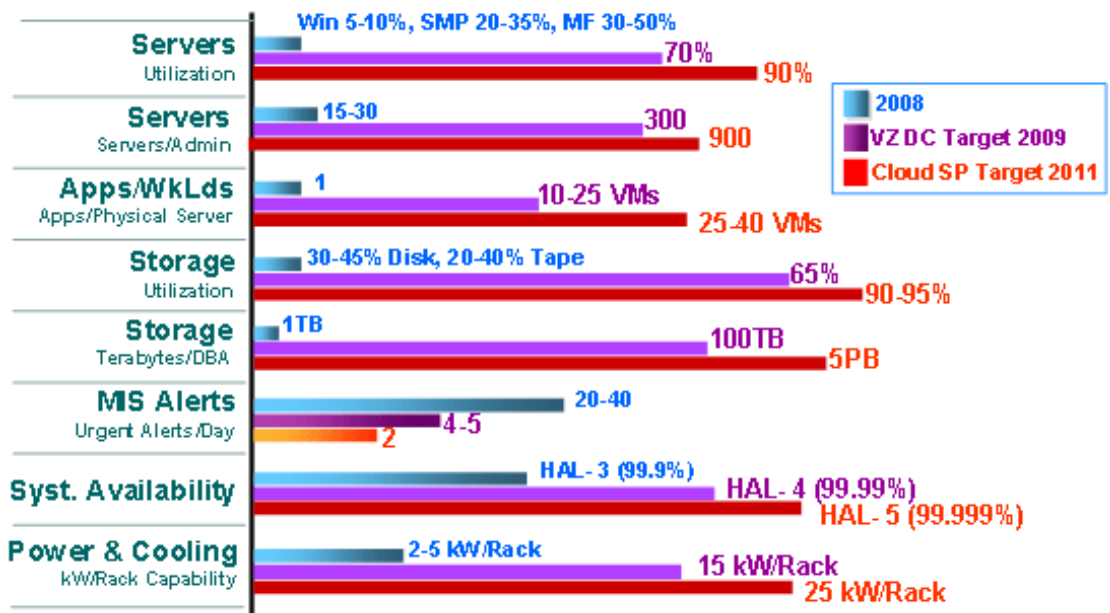
Click for email address  
[www.imexresearch.com](http://www.imexresearch.com)



A typical example of emerging VertiClouds® reveals tremendous benefits of using cloud storage services. The U.S. healthcare system spends \$10–15 billion each year on redundant radiological exams that stem from physicians' lack of access to patients' prior imaging exams. Medical Imaging brings with it a significant and never ending demand for storage. The cloud computing and storage infrastructures and

Vertical Cloud® platforms create an opportunity to address the significant need for providing a consistent set of networked healthcare services for radiology, cardiology and patient records transport and management to enable sharing diagnostic imaging information. By using a 10GbE network from end-users to Cloud, ultimately supports thousands of hospitals, imaging centers and hundreds of thousands of physicians and millions of patients. In turn, users benefit from a virtualized, secure, reliable and elastic nature of cloud services ensures a capability to scale the performance and operation of the network for peak usage periods. Also the pay-for-use cloud storage model provides an affordable model for health care facilities of different sizes. © 2008-10 IMEX Research

### ► New Metrics for Cloud-Aware VZ DC Infrastructure



© 2008-2010 IMEX Research. All rights reserved.



Cloud Storage changes the game for distributed enterprises by delivering LAN-like application performance over the WAN and enabling a wave of IT consolidation from the branch into the data center or private cloud through optimizing network access for cloud storage deployments with new solutions. Now, enterprises are beginning to migrate storage to the public cloud to take advantage of the dramatic operational and cost benefits.

Having implemented virtualization in their data centers, organizations are now establishing creating new metrics and goals metrics to make their data centers cloud-ready.

# The Challenges of Cloud Storage

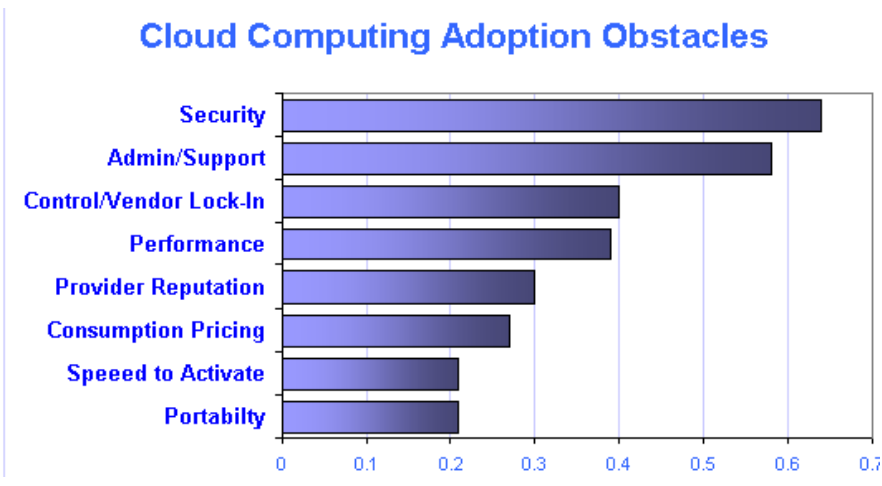
Given the obvious benefits of Cloud Storage, raises the question as to why haven't more organizations already adopted cloud storage for on-premises applications and alleviate present expensive storage.

The answer – Right now, there are traditional IT barriers which must be overcome when considering integration of cloud storage into today's production environments.

- Performance and data transfer rates become key issues as the distance between the data and the user increases - which is what happens in cloud computing.
- Even unlimited bandwidth without solving the latency problem will not improve the performance because it is the latency - or the chattiness - of the protocols, plus the speed of light limitations that cause the end user experience to be very poor.
- Not all data access patterns are well suited to the cloud, particularly if there are large distances to cover. In such cases, bandwidth becomes not only a challenge but a financial consideration.
- Cloud storage isn't about to replace the storage network in the data center any time soon, at least not for data-intensive, high-performance, low-response time, transactional applications and for mission-critical data. But we will see many use cases where companies and organizations of all sizes will augment their on-premise storage with cloud storage potentially from various vendors in a hybrid model deployment. However, hybrid models tend to bring interoperability issues and the need to deal with different tools, API's, management infrastructure etc.
- The popular storage use cases tend to be infrequently accessed data scenarios including archiving, backup, DR, and offsite data protection. Performance and data transfer rates become key issues as the distance between the data and the user increases - which is what happens in cloud computing.
- It will be a hybrid cloud storage world for many years to come with a lot of that storage living in private clouds. Many large enterprises deal with petabytes of data and their processing.

A survey conducted by IMEX Research reveals that both enterprises and SMBs today are concerned about:

- **Security & Data Availability** - Concerned about the security of their data, once it is in the hands of the cloud provider whose multi-tenancy infrastructure is shared by others. Numerous questions arise when considering cloud storage, including "is my data secure?", "am I in compliance?", "what happens if a provider loses a disk drive?", "is my data protected?", "how do I know if my data is truly unusable when I delete it?", and many others. StorSimple allows you to provide an encryption key – or we can generate one for you – and all data sent to the cloud is encrypted. They are rightfully worried about the availability of their data and how that impacts their day-to-day operations. What happens if a cloud storage service is offline for a period of time?
- **Performance** - Legitimately concerns about application performance if the application storage is in the cloud. Will the cloud storage service satisfy my workloads?
- **Bandwidth limitations** – Bandwidth is a limiting factor when accessing a public storage cloud, as they are accessed over the Internet. Primary storage deduplication and compression, minimizes bandwidth consumption dramatically while also improving performance.
- **Latency constraints** – Latency is the silent killer of application performance, both in terms of response time and throughput. StorSimple takes advantage of parallelization, persistent connections, and TCP optimizations to overcome latency and improve performance
- **Manageability** - Are concerned about being locked into their proprietary cloud storage infrastructure and applications services. They don't have vendor independent tools or industry standards to evaluate the applicability or measure the effectiveness of cloud storage for their environment.
- **Interoperability/Protocol translation** – A serious concern exists today is: Most of today's on-premises applications use block protocols such as FC, iSCSI etc. But Cloud storage protocols predominantly speak only in the language of file protocols (CIFS, NFS) and both public and private storage clouds are accessed via REST HTTP-based, or SOAP APIs. Since these applications expect block access to storage, introducing a cloud storage system to the application is like trying to have a conversation in Spanish when you only speak English. So how will current applications even be interoperable between existing storage infrastructure and cloud storage? Taking advantage of these conversion tools and seamlessly integrating with cloud storage eliminates the need for complex application re-programming and integration (see below - Importance of Cloud Storage API set).
- **Costs** - Many cloud storage services charge their customers based on the amount of storage capacity consumed and the number of IOPs performed or the amount of bandwidth consumed. Much as reducing cloud storage costs through deduplication and optimization and taking advantage of pay-as-you-grow schemes helps but choosing a cloud storage vendor which meets all other criteria remains a challenge.



# Delivering an Information-Centric Cloud Storage Infrastructure

Organizations that presently use on-premise storage and wish to selectively federate data to public clouds, would need to create an integration that enables automated searching and tagging of on-premise data and creates a policy-based federation of data to public clouds to deliver a truly information-centric cloud infrastructure keeping in mind the following:

## Organization Goals for Primary Storage

In area of cloud storage, organization goals for primary storage include:

- **Improved storage economics** – Leverage appropriate tiers of storage according to the performance requirement of the data.
- **Performance consistency** – working set and hotspot data to be automatically tiered to the highest-performance tier of storage (SSD), whereas non working-set or non hotspot data is tiered to a lower performing tier dynamically
- **Transparent integration** – no complex policies to manage, no agents to install on servers, and the storage view from server to remain same in light of new integration with cloud storage
- **Deduplication** - of primary storage to eliminate the repeated storage of redundant segments of data. Deduplication to be completely transparent to the server, i.e. the server's view of the content of their storage volumes should completely remain unchanged.
- **Compression** - Achieve higher levels of compression even when encountering single-byte insertion scenarios.
- **Data Reduction** - Achieve 10X data reduction for specific storage workloads.
- **Databases** - Separate the storage of BLOBs (Binary Large Objects) from content databases in conjunction with appropriate application frameworks . BLOB externalization leads to smaller database sizes, reduced fragmentation and maintenance cycles, and consistent response times as the deployment scales
- **Performance consistency** –Hotspot elimination – Use technologies to automatically identify working sets, create integrated storage through deploying SSDs for hotspots removal and improve performance while leveraging , and low cost SATA for compressed, deduplicated data.
- **Version control support** – Version control can be enabled, thanks to some companies' implementation of primary storage deduplication, which minimizes storage capacity requirements and cost
- **Pay-as-you-grow** – use of cloud storage enables pay-as-you-grow capacity consumption, which minimizes cost of over-provisioned yet unutilized on-premises storage
- **Optimal Cost Structure** – Using deduplication, compression and encryption minimize cost for cloud storage capacity and cost for IO and data transfer,
- **Tape elimination** – Use Cloud Clones to enable consistent, point-in-time recovery snapshots and store independent copies in the cloud to eliminate need for tape
- **Advance API-enabled Capabilities** – Leverage advanced API sets to achieve following capabilities: Content search capabilities, Optimized Retrieval, Retention and Compliance Controls, Metadata Tagging, Security and others.

## Application Performance Optimization using Data Segmentation

One of the key requirements to improve performance of data access and computations is to identify application/workload characteristics (IOPs, latency and bandwidth), data usage patterns (including frequency of access, I/Os, age, priority and relationship with other data) as well as data management (data placements, storage device characteristics and locality of reference) and then store the data on appropriate tier of storage and eliminate access hotspots.

## Cloud Storage - Ideal for Certain Data Types

**Cloud Storage - a poor fit in following data types:**

### Transactional Data

- Frequent Read and Write Accessed Data (eg DB)
- Massive I/O requirements
- Database, Source Code, Active VMware images

### Active Corporate Data

- Advanced Data Protection Schemes
- Office Docs and spreadsheets
- Source Code

**Cloud Storage - good for the following data types:**

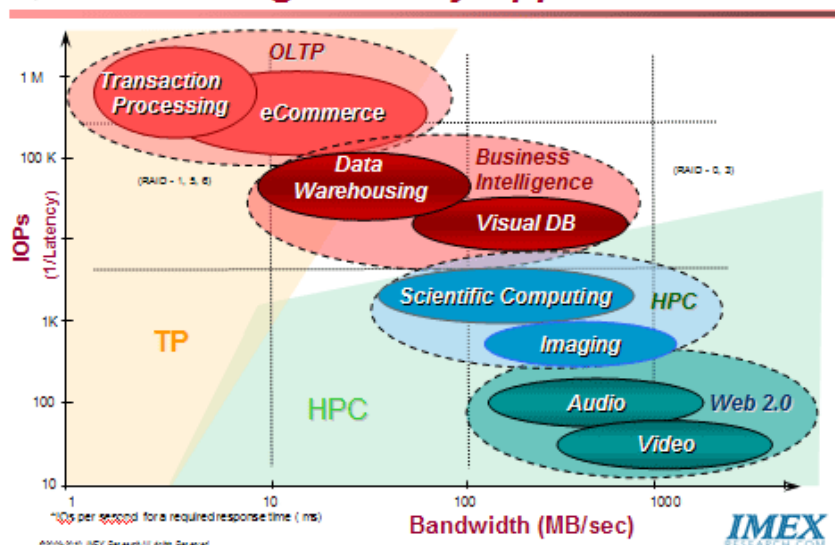
### Large files with mostly read accesses

- Digital Content, Streaming Media, Video, Music
- Parallel Streaming Writes
- Video Surveillance (Private Clouds)

### Long-Term Storage Files

- Backup and archival Files (Private Clouds)
- Medical Images, Energy Exploration, Genomics
- VMWare Backups

## Market Segments by Applications



## Geographically Shared Files

- Universal Access from anywhere, anytime (Public Clouds)
- Movie Trailers, Training Videos
- Corporate Marketing and Training Colaterals (docs/podcasts/videos)

## Achieving Optimized Cloud Storage using Tiered Data Sets

Key to optimizing I/O performance of storage is understanding the characteristics of Data set volumes and taking advantage of tiered storage on-premise plus cloud storage services to achieve CAPSIMS© goals of

- **Cost** (Achieve 10X data reduction for specific storage workloads. Deduplication of primary storage to eliminate the repeated storage of redundant segments of data, Achieve higher levels of compression, Enabling Version control to minimize storage capacity requirements and cost, use of cloud storage enables pay-as-you-grow capacity consumption, which minimizes cost of over-provisioned yet unutilized on-premises storage.)

- **Availability** (Cloud storage D2D data replication cloning to enable consistent, point-in-time recovery snapshots and store independent copies in the cloud to eliminate need for tape)

- **Performance** (Leverage appropriate tiers of storage according to the performance requirement of the data, working set and hotspot data to be automatically tiered to the highest-performance tier of storage (SSD), whereas non working-set or non hotspot data is tiered to a lower performing tier dynamically (SATA) Separate the storage of Binary Large Objects from content databases in conjunction with appropriate application frameworks. BLOB externalization leads to smaller database sizes, reduced fragmentation and maintenance cycles, and consistent response times as the deployment scales)

- **Security** (Leverage advanced API sets to achieve Security)

- **Interoperability** (Leverage advanced API sets to achieve following capabilities: Content search capabilities, Optimized Retrieval, Retention and Compliance Controls, Metadata Tagging, Security and others)

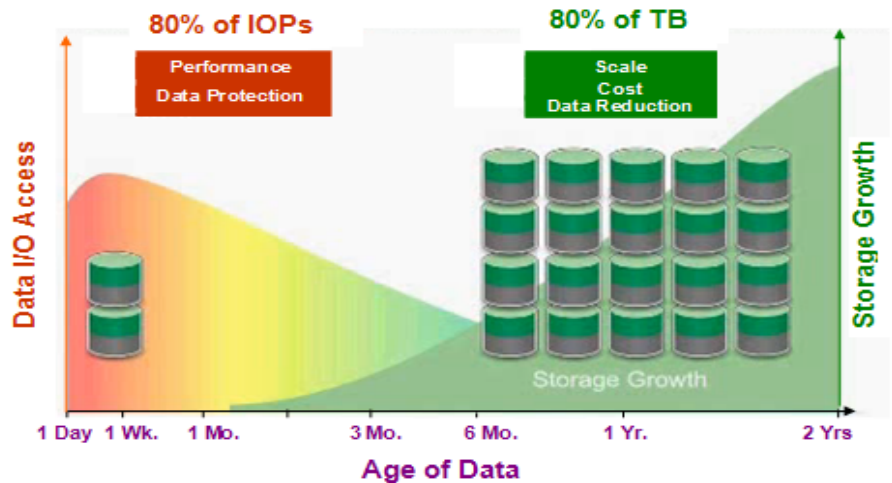
- **Manageability** (no complex policies to manage, no agents to install on servers, and the storage view from server to remain same in light of new integration with cloud storage, Leverage advanced API sets to achieve following capabilities: Content search capabilities, Optimized Retrieval, Retention and Compliance Controls, Metadata Tagging, Security and others.)

- **Solid State Drives (SSD)** – SSD provides the highest levels of throughput and the lowest response times. WSL helps ensure that the working set for a given volume resides on SSD to ensure consistent, high performance access

- **FC/FCoE/iSCSI/SAS** – Fibre Channel SANs are predominantly used in large enterprises while SMBs are more inclined towards using iSCSI SANs for primary storage and to achieve higher availability and performance from their storage systems.

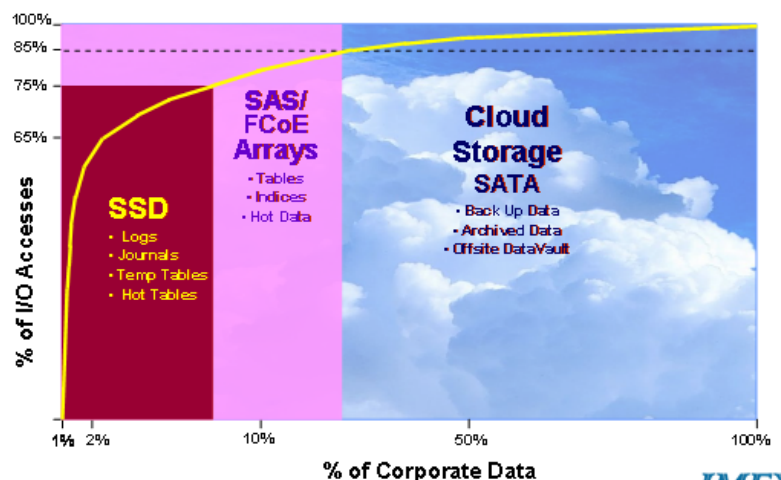
- **Serial ATA (SATA)** – SATA provides the secondary tier of internal storage, and is used for holding deduplicated working set and non working-set data. SATA provides high capacity and consistent levels of performance and response time

- **Cloud Storage** – Public or private cloud storage can be configured for a particular volume if so desired, and when enabled, is used as the lowest tier of storage for data. Cloud storage enables pay-as-you-grow capacity and the greatest levels of economic efficiency



## Corporate Cloud Storage Usage

### I/O Access Frequency vs. Percent of Corporate Data



IMEX  
RESEARCH.COM

# Types of Cloud Storage Solutions

Customers need to decide whether or not a cloud storage service, public or private, should be used for a given data set and which cloud storage service class should be used before deciding to migrate to cloud storage services.

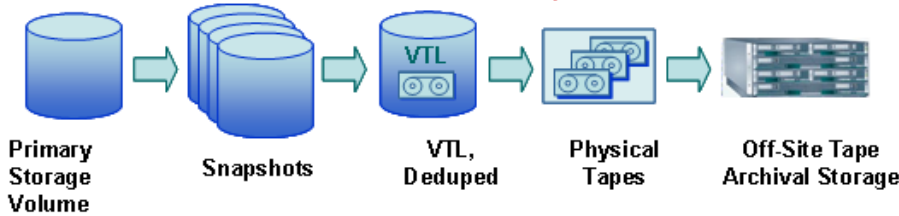
- **Classic DataCenter Storage** – Data set volumes configured as on-premise only do not take advantage of a cloud storage service but rather only take advantage of tiered storage system.
- **Private Cloud Storage** – Data set volumes configured in this mode take advantage of tiered storage on-premise plus a private cloud storage service on-premise that is used for data protection using a D2D Cloud Clone feature format compatible with public clouds. This helps minimize data protection infrastructure and cost
- **Federated Cloud Storage** – Data set volumes configured in this mode take advantage of both on-premises tiered storage and capacity made available through a cloud storage provider that is used for data protection

Cloud Clones are similar to snapshots in that they are a point-in-time, consistent copy of a series of application-related volumes. Whereas snapshots are typically mounted by a backup server and written to virtual or physical tape, which is then shipped off-site for protection, Cloud Clones on the other hand are stored persistently in the cloud. They are a replacement for VTL and tape libraries and allow one to mount and restore data at any point in time without having the hassle of retrieving tapes. Cloud Clones are de-duplicated and compressed, minimizing cloud capacity and data transfer costs, which makes them more cost compelling than tape as a form of archival storage. Cloud Clones are encrypted using a key the user supplies (or optionally the system can generate for the user) to ensure confidentiality while data sets are stored in the cloud using cloud clones.

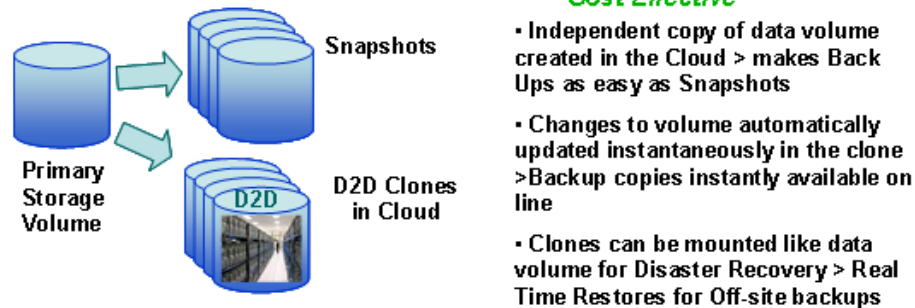
## Features of a good Cloud Storage Solution

- Fully integrates into traditional on-site storage as well as the cloud so users who want to store large amounts of content can connect to either in minutes

### Traditional Backup & Restore Today - *Complex, Laborious, Risky, Expensive*



### Backup & Restore in D2D Cloud Clones - *Simple, Automated, Reliable, Cost Effective*



- A hybrid storage solution that makes cloud storage appear like local data center storage that integrates into customers' existing storage and data management tools. It identifies and stores all the hotspot and bottleneck data on a tier of high-performance Solid State Drives (SSD), enabling the use of lower-cost SATA storage and/or cloud storage as primary storage. It also performs real-time data de-duplication to minimize the footprint of the data stored and provides the WAN optimization functions for cloud storage. All data that is stored in the cloud is encrypted.
- It intelligently identifies the bottleneck areas and stores the data on a tier of integrated high-performance Solid State Drives (SSD), thereby, allowing them to use low cost SATA storage and/or cloud storage as primary storage for the data needs. Their technology transparently examines all data, breaking it into smaller segments, and stores the data segments on the appropriate tier of storage according to its frequency of use, age, relationship with other segments, and

priority.

- To transparently protect and manage file, email and other applications' data copies across on premises and public cloud storage to improve service levels, while simplifying and greatly reducing data management costs.
- Native integration with the REST based API also delivers increased performance over alternate protocols. It also enables access to Cloud Storage from native CIFS or NFS clients.
- Turns Windows Explorer into a Cloud Storage Portal using REST API on Cloud Desktops, enabling native access and providing an ability to access content as if it were stored on the desktop with drag and drop migration.
- SMBs can now enhance their file servers to store content in the Cloud while implementing a smart cache that maintains LAN access speeds.
- Identifies and archives inactive data from applications and data warehouses—including master, reference and transactional data—to Cloud Storage Service that is readily accessible when needed, streamlining information lifecycle management while reducing management, software, and hardware costs.
- A simplified virtual file server that delivers cost-effective and completely secure cloud storage offsite for businesses yet retaining the local performance and functionality of a traditional NAS

- Provide a specialized data retention repository in which to preserve large amounts of historical data for regulatory or business purposes
- Key is to create an archive file system with the unified goals of providing unfettered access to Cloud Storage as an enterprise archiving storage tier on an Exabyte level scale by presenting the cloud as a standard network share while removing the hurdles associated with REST or SOAP integration.

## Hybrid Storage – A new class of Storage Solution

Now a new class of storage solutions – namely hybrid storage solutions – are emerging that help address these exact challenges. A hybrid storage solution is:

- Deployed in the customer's data center
- Provides servers with access to storage using protocols that they understand
- Speaks cloud storage service protocols, and virtualizes cloud storage into usable data center capacity
- Overcomes performance limitations associated with cloud storage
- Addresses security concerns associated with cloud storage

These solutions typically have the following characteristics:

- They are application-aware and tune the storage system's behavior according to the application's needs.
- They have multi-class integrated storage with complicated caching algorithms or simple automated data tiering.
- They tend to operate like any one of the storage arrays that already exist in the data center.
- They sit between the servers and the cloud, and allow users to control where the data is stored.

### Reduced Storage Costs

These solutions help address the immediate requirement for adding capacity inexpensively. Aside from the cost of the solution hardware, the rest of expense on storage is based on how much capacity the user utilizes in the cloud and how much data they transfer (These rates vary from cloud services providers such as Amazon Web Services, Rackspace etc.)

### Application Aware Tiering

They have application-awareness that helps to address specific storage-related issues that impact the ability of an organization to scale or manage an application. This can be manifest in application-aware tiering, volume data location policies, and even plug-ins that integrate with frameworks provided by the application vendor.

### Storage Efficiency

They provide primary storage deduplication, which not only helps improve performance when using the cloud, but also minimize the cloud cost (since only deduplicated data would ever be read from or written to the cloud).

They also help address some of the concerns with Exchange and SharePoint regarding storage efficiency which is especially important, as Microsoft removed the Single Instance Store for attachments in Exchange 2010 and also for using versioning and extended recycle bins in SharePoint. Deduplication ensures that redundant email attachments are stored in a more space-efficient manner, meaning lower storage capacity consumption

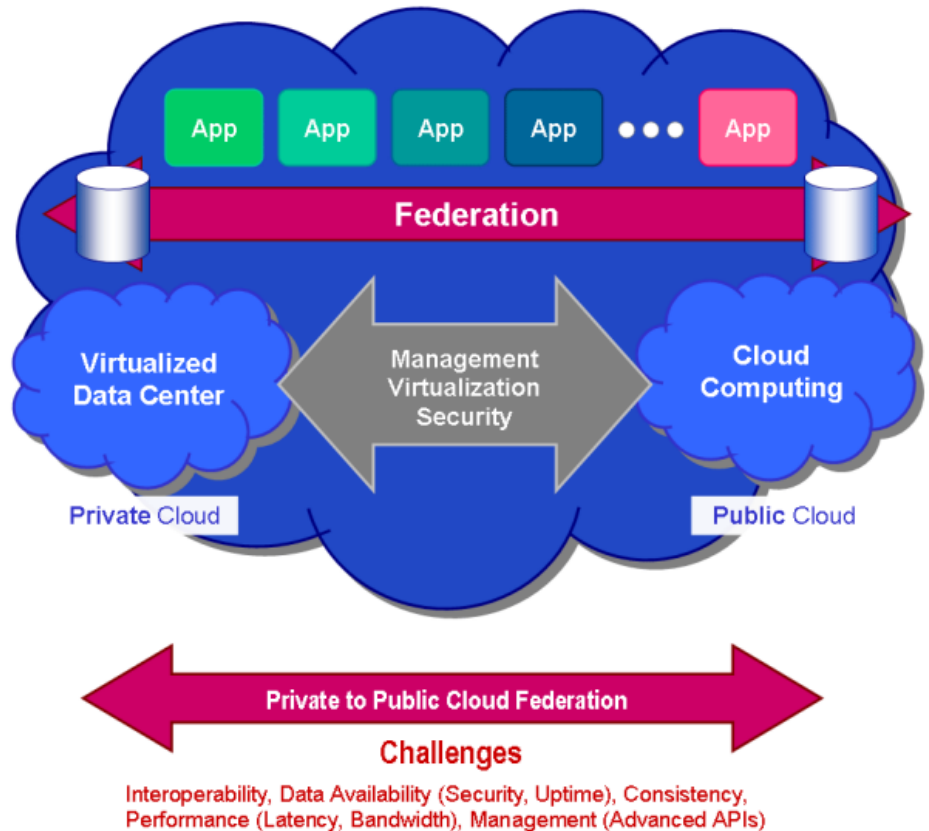
### Security

They allow users to control the encryption key, and it is never shared with the cloud provider. This increases confidence, as organizations don't need to worry about what happens if someone gets a hold of their data from the cloud, or if the cloud provider is requested to release the data – or hardware where data may reside – in support of investigation or litigation.

Coupled with other security services provided by the cloud provider, such as Virtual Private Networks, roles-based access control, multi-factor authentication etc., cloud storage can be as secure as traditional on-premises storage for certain applications. For these extra security features additional dedicated performance-boosting power is generally needed.

### Data Protection

Helping simplify data protection is one of the major benefits of Hybrid Storage Solutions. Having a virtually unlimited pool of storage sitting behind an on-premises hybrid storage appliance that provides data deduplication and encryption means that with the right application hooks and validation with data protection software vendors, users can take snapshots - that are not only crash-consistent



but also application-consistent - and store them in a space-efficient, secure format on the cloud as an always-on backup and available in case needed for disaster-recovery. So instead of having to continually fetch tape from an off-site vault, users can simply use a cloud backup copy to perform a mailbox restore, object restore or even recover the entire application.

## Seamless Integration of Data Storage between Data Centers & Cloud

Usually, when organizations decide to leverage cloud based storage solutions, they will be forced to rewrite their applications because most of the cloud storage providers give access to their storage using http or some REST based APIs. Traditional applications are not capable of accessing the storage that uses REST based API and it will require a rewrite of the applications or deployment of some sort of a connector. A start up company mitigate this using hybrid storage solutions that makes cloud storage from multiple vendors such as EMC, IronMoutain, Microsoft, and Amazon, appear like local data center storage that integrates into customers' existing storage and data management tools.

Much as these solutions are relatively nascent, many customers are achieving lower total cost of ownership, simplified backup and restore, better disaster recovery, and the ability to confidently scale many applications without concern over performance.

## Advanced Cloud Storage API enabled Capabilities

A cloud NAS gateway that many storage providers offer, provides an important translation between data center network storage protocols (e.g. CIFS/NFS) and a more internet-optimized protocol (e.g.WebDAV) that is less chatty. The cloud NAS gateway provides a viable quick fix to get data moving to the cloud. The advantage of a cloud NAS is that it enables a customer to begin archiving data to the cloud almost as easily as they could send data to a local network mounted disk. This allows for rapid adoption of cloud storage as an archive destination. But Interacting and performing efficient communication to the cloud is just the bare minimum of what an API should do. Advanced cloud storage API sets allow following improved capabilities:

**Content Search Capability** – An API set from within the programming language can provide the ability to deliver context searching during ingestion on the data that is being stored on it and create the index - all done on the back end, by the cloud storage system. Subsequent commands allow the application to query the index and retrieve results. These queries can be against the core content itself, or they can be metadata specified during ingestion and even display the results. This avoids having to issue subsequent retrieval requests to paint the search results. The query API set can return a set or a “page” at a time results, allowing the user browse through very large search results sets.

**Optimized Retrieval** - Advanced API sets can also provide commands to the cloud storage to store the data in a parsed format allowing an organization to retrieve just the components of the data that is needed. Having these items individually addressable provides an important workaround to mitigate the latency of the cloud. A typical example would be to command the API to parse the metadata (header, body and attachments) only. On search only the relevant metadata (header and or body) is pulled across the internet. With relevant email's header found, associated large attachments can then be brought across the internet.

**Retention and Compliance Controls** - The API set should be enable to set retention and WORM levels within the application itself optionally set by an administrator or user or automatically be activated tied to an internal analysis of the data. Retention commands should include the ability to set length of time to be stored, number of copies to be kept and if the data can be modified and allow for the assured destruction of data, confirming that the data is no longer recoverable.

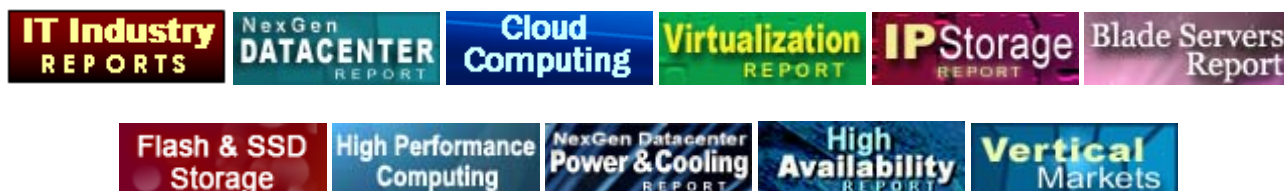
**Metadata Tagging** - Metadata tagging by an API set allows for keywords or reference points to be set on files to help with search criteria, retention criteria or compliance criteria. Being able to build search optimization, pre-parsing data for retrieval, setting retention, compliance and metadata upfront at the point of archive goes a long way toward those feature sets actually being used.

A feature-rich API set allows building classification into the process, enabling an interactive archive to become an upfront, quick task performed consistently will likely increase the accuracy of the classification as the application usually has the most context at that point compared to traditionally classification projects that are very large and involve processing of data long after it was written with specifics of that data forgotten.

**Security** – The API data should be encrypted during transmission and remain encrypted when stored at the cloud storage provider's facility. Not only does this prevent the infiltration of data in flight, it also insures that only that customer can read their data. The API set should have the ability to store multiple copies of that data not only locally but in geographically disperse locations as well, providing a comfort level in recovery from a failure or data corruption locally as well as globally.

---

Click on the following for additional information or go to <http://www.imexresearch.com>



IMEX Research, 1474 Camino Robles San Jose, CA 95120 (408) 268-0800 <http://www.imexresearch.com>  
If you wish not to receive any IT information/news from IMEX, please [click here](#).