



WHAT, WHY, AND HOW OF “CLOUD STORAGE GATEWAY”



Balaji Panchanathan
Engineer
EMC – BRS
Balaji.panchanathan@emc.com

Table of Contents

Introduction	3
Cloud Storage Gateway	3
What is Cloud Storage gateway.....	3
Why the customers would need it	3
Avoids Vendor Lock-in.....	3
Provides increased security.....	4
Common deployment backup to the cloud.....	5
File vs Block vs Object level approach for the cloud storage gateway deployment	7
Opportunities for EMC.....	8
Avamar.....	11
Cloud Tiering Appliance	12
Data Domain	13
VNX:.....	13
How to decide on the Cloud Storage Gateway	13
1. Current environment and how the storage is organized	13
2. Future roadmap and growth needs	14
3. How majority of users are accessing the data.....	14
4. Whether the cloud is used only for backup.	14
Evaluation of the Cloud storage gateway.....	15
Business Environment and how the market looks	15
Defining the market segment.....	15
What is Storage-as-a-Service.....	16
Customers want more control and visibility through Reporting and Analytics.....	17
Major segments in this cloud storage gateway and the major companies.....	18
Whether Cloud storage gateway has the potential to disrupt the storage market.....	19
Conclusion	19
References	20

Disclaimer: The views, processes, or methodologies published in this article are those of the author. They do not necessarily reflect EMC Corporation's views, processes, or methodologies.

Introduction

This article will be most useful to IT decision makers and storage administrators in IT firms who are using cloud for their storage needs and professionals in Cloud Storage industry. Primarily, the four sections of this article address these questions.

What is cloud storage gateway? Why is it needed and how can it be used in the current environment?

What EMC products serve the functionality of cloud storage gateway? What functionalities are missing and how can they be added to the EMC portfolio?

Section 1 details what is cloud storage gateway and the deployment scenarios. Section 2 looks at EMC products that can be integrated with the cloud storage gateway functionality. Section 3 discusses how to evaluate cloud storage gateway vendors and which factors affect the decision. Section 4 examines the business environment and its potential to affect the current market.

Cloud Storage Gateway

What is Cloud Storage Gateway? A cloud storage gateway sits on the edge of the enterprise network and translates the requests on the enterprise to the cloud. Basically, they translate between File level/Block level requests to cloud storage service provider API's. One big advantage cloud storage gateway provides is the ability of the enterprise to switch between cloud service providers easily without having to rewrite numerous API's because those API's are built-in to the cloud storage gateway.

Why customers would need it: Security and vendor lock-in are the two main reasons why customers are unwilling to go to public cloud. The cloud storage gateway aids in removing both of these disadvantages.

Avoids Vendor Lock-in: Customers develop custom applications using API's provided by the cloud service provider. These API's differ from provider to provider. One example of Amazon API for Amazon S3 is:

POST / HTTP/1.1

Host: *destinationBucket*.s3.amazonaws.com

User-Agent: *browser_data*

Accept: file_types

Accept-Language: *Regions*
Accept-Encoding: *encoding*
Accept-Charset: *character_set*
Keep-Alive: 300
Connection: keep-alive
Content-Type: multipart/form-data; boundary=9431149156168
Content-Length: *length*

The format for Rackspace is:

```
PUT /<api version>/<account>/<container>/<object> HTTP/1.1
Host: storage.clouddrive.com
X-Auth-Token: eaaafd18-0fed-4b3a-81b4-663c99ec1cbb
ETag: 8a964ee2a5e88be344f36c22562a6486
Content-Length: 512000
X-Object-Meta-PIN: 1234
```

As you can see, each vendor has a different format and we need to write the code or get the code from a third party. However, this layer of complexity is handled by the cloud storage gateway, such as EMC Atmos®.

Figure 1 illustrates how the cloud storage gateway helps avoid vendor lock-in.

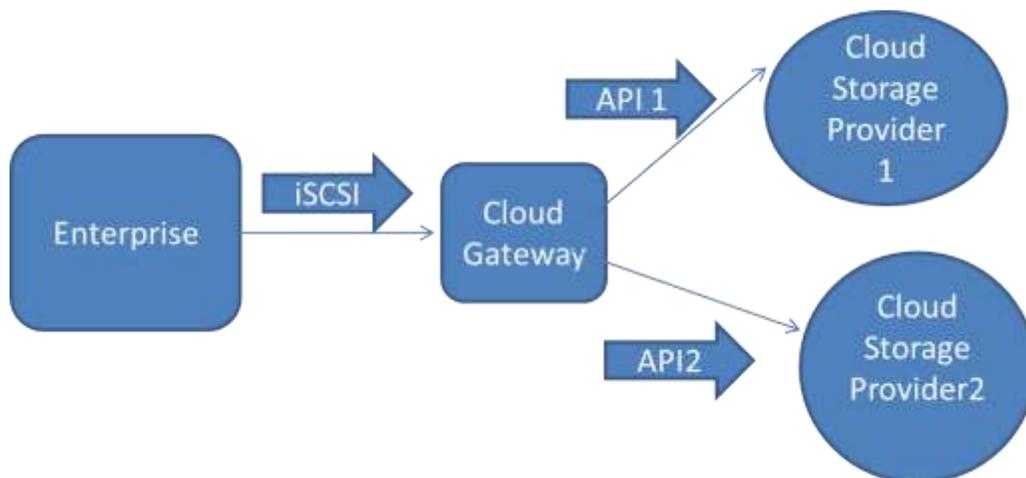


Figure1

Provides increased security: Companies in this space include CipherCloud and PerspecSys. These firms, which have cloud gateway products that encrypt the data real-time when it is posted to the cloud, are currently focused on securing specific

applications such as Salesforce.com, gmail, Oracle CRM, and so on. Figure 2 illustrates that.

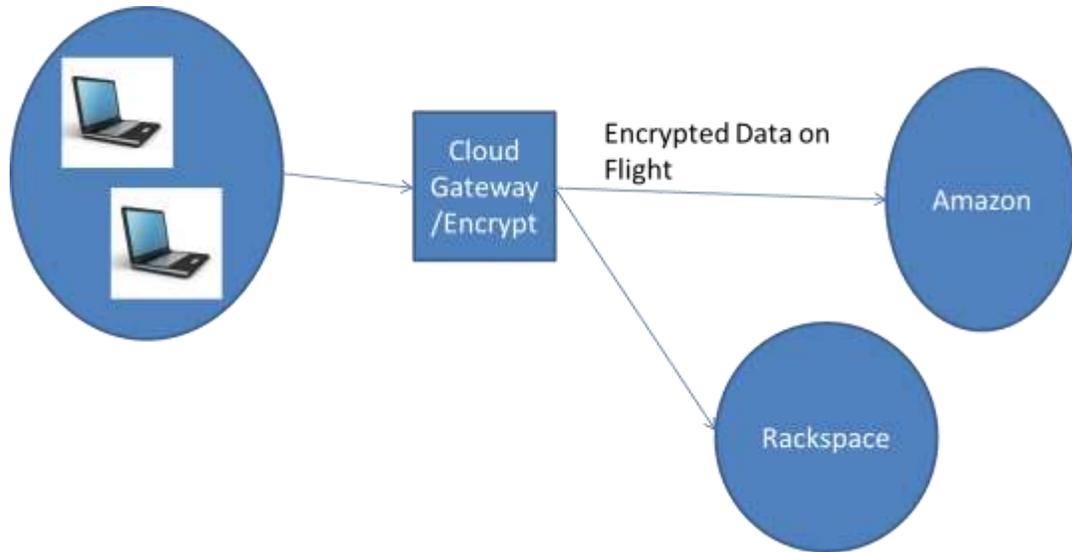


Figure 2

Common deployment - Backup to the cloud: Cloud Storage Gateway helps when backing up data to the cloud and aids in disaster recovery. Using a tiered approach, data can be backed up to the cloud based on the value of the data (Information Lifecycle Management). Figure 3 depicts this type of topology.

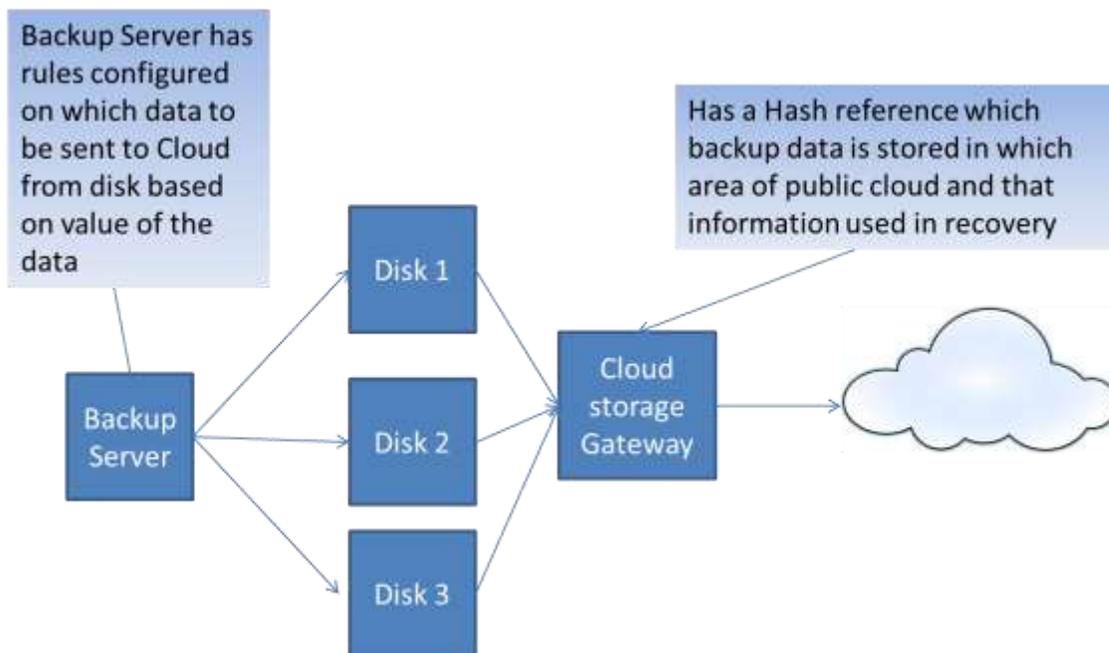


Figure 3

Another approach is to use deduplication technology which saves cost by storing only unique data to the cloud. Figure 4 shows a topology on how to implement it.

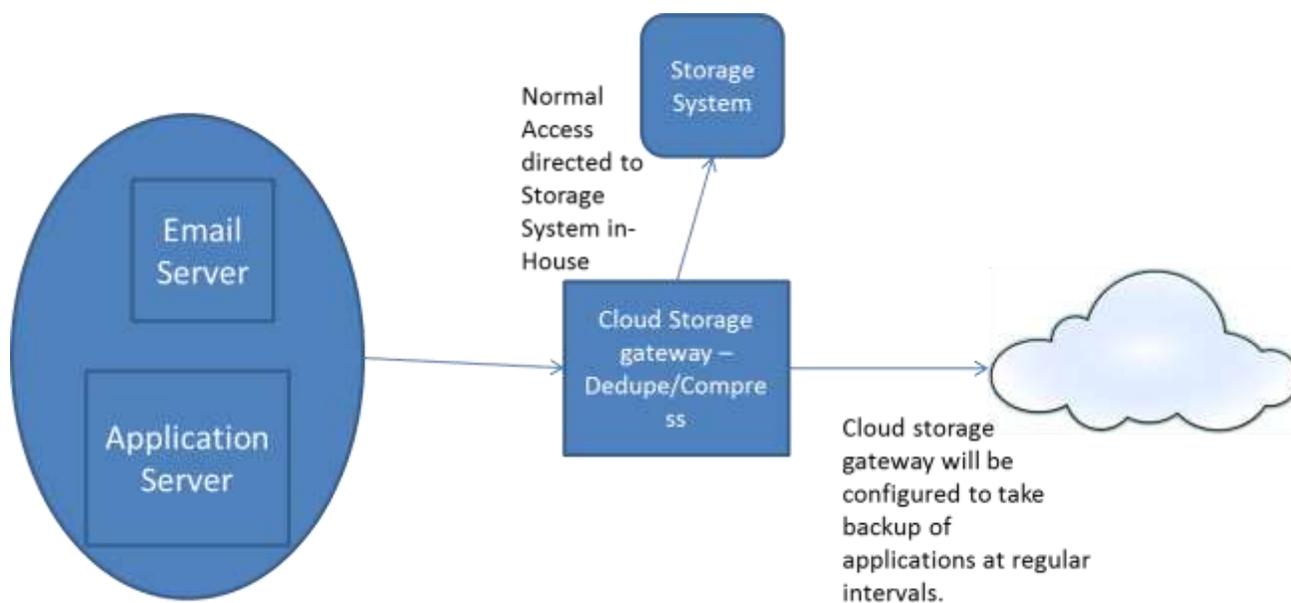


Figure 4

Cloud Storage Gateway will have a hash which will have two pointers. One will point to data in the internal storage system while the other pointer indicates how to retrieve the same data from cloud storage. Before storing it on the storage system and the public cloud, the cloud storage gateway will deduplicate and store only the unique data. One of the biggest bottlenecks is performance. When users update or write data to a file the cloud storage gateway has to deduplicate the data and only update the file. This might take time and the user may experience delay. More research is needed in this area on the type of storage system and algorithm on cloud storage gateway should be used to get a good response time.

Approach for cloud storage gateway deployment: File vs Block vs Object

The type of approach depends on the type of deployment used, as shown in Table 1.

Scenario	Type of access	Reason
Public Cloud is their primary storage option	File system approach	Cloud storage gateway will have local cache. Managing the local cache will be better with file-based approach.
Public cloud is used for backup	Block-level approach	Deduplication will be done at cloud storage gateway and for deduplication. Block-level approach is best for providing savings.
Public cloud is their primary storage option and only used for unstructured data and least used data	Object level	Only benefit is that the cloud storage gateway enables seamless movement from one service provider to another.
Data does not change much	Block level	Deduplication will be done and as the data does not change, there is more likelihood of developing algorithms in such a way that result in having a single block in maximum number of files.

Table 1

Figure 5 illustrates the typical data structure used in the cloud storage gateway.

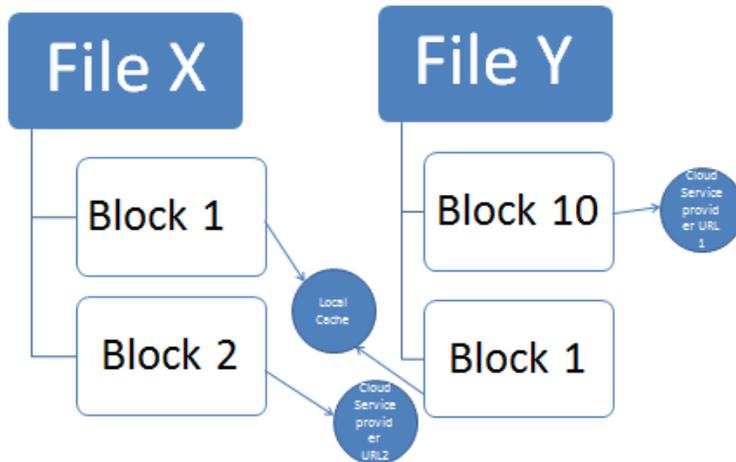


Figure 5

In this example, deduplications occur at the block level, but access is file level either by CIFS or NFS protocol. Thus, the Cloud storage gateway acts as a NAS device. Block 1 is part of both file 1 and file 2 that is stored in the local cache and Block 10 and Block 2 are stored in the Public cloud. The respective URLs to access that data is in there. There can be sophisticated algorithms which indicate that when reference to the block increases by a specified number, save that block in local cache. Reference count of a block increases by 1 every time a file has a link to it.

Opportunities for EMC

This section discusses types of cloud storage gateway functionalities available in EMC products, the missing elements of the cloud storage functionality, and how to fill the gaps. The major functionalities of cloud storage gateway are support for avoiding vendor lock-in (multi cloud service provider support), deduplication, compression, and security

Products with some of the functionalities of cloud storage gateway are shown in Table 3.

Product	Functionality
Cloud Tiering Appliance with built-in RSA Security	Encrypts the data before sending to the cloud.
Cloud Tiering Appliance	Support for multiple public clouds
Cloud Tiering Appliance	Move inactive and recently used data to cloud
EMC VNX [®] Platform	It is used to store both file level and block level data. File level deduplication is there and can be easily integrated with Cloud Tiering Appliance
Cloud Tiering Appliance	Compression capabilities
Avamar [®]	Source deduplication
Data Domain [®]	Target deduplication
Avamar and Data Domain	Backup and recovery

Table 3

Table 4 illustrates how the total portfolio of EMC products support the features of cloud storage gateway.

Functionality	Support Level	Comments
Multi-Vendor Support	Partial	No option available to automatically migrate from one vendor to another
Security	Full	
Deduplication	Partial	Even though Avamar and Data Domain supports deduplication, integration options are not clear
Compression	Full	
Backup and recovery	Partial	Even though Avamar and Data Domain support backup and recovery, integration options with Cloud Tiering Appliance is not clear

Table 4

The missing element seems to be the integration option of deduplication technology which is available in Avamar and Data Domain. An important element of cloud storage gateway is that both source and destination deduplication is done at the same place. Clients do not perform deduplication; rather, it is done by the Cloud storage gateway which then stores the deduplicated data in the cloud, like combining Avamar and Data Domain functionality in a single device.

The biggest challenge is performance. What is the best way to achieve an acceptable performance? With Avamar, the backup data or file is fetched only when there is a restore. However, in the case of cloud storage gateway, the file is fetched every time the file is read or written.

We will explore how each of the products can be modified to achieve this.

Avamar

Avamar can build the functionality of Translators and there are two ways to implement it. One is to have the Cloud storage gateway act as Avamar Server with the storage part in the public cloud. Figure 6 illustrates that scenario.



Figure 6

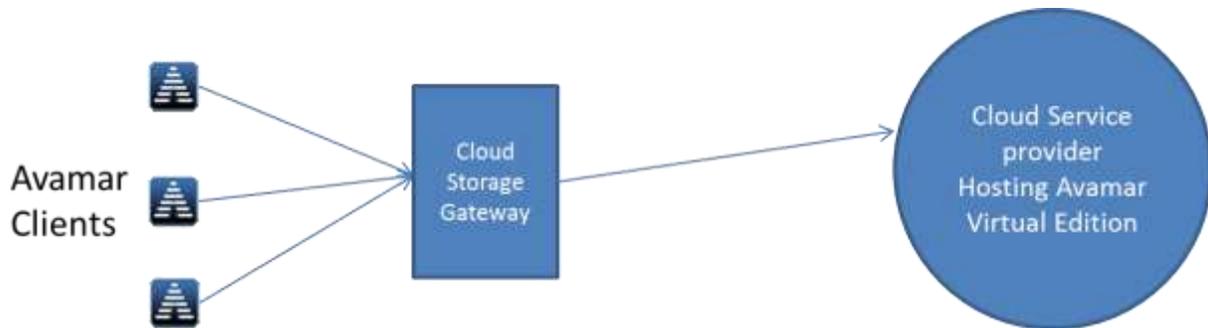
Avamar server will have a table with two columns, as shown in Table 5. The first column will have the block number and second column will have the corresponding URL where the block is stored. Avamar Server will have a table of which block can be accessed by accessing the appropriate URL that then uses that table to fetch the block and perform restore and backup.

File x – Block 1	url=http://doc.s3.amazonaws.com/2006-03-01 Bucket = avamar object=backupfile
File x – Block 2	url=http://doc.s3.amazonaws.com/2006-03-01 Bucket = avamar object=backupfile1
File x – Block 3	url=http://doc.s3.amazonaws.com/2006-03-01 Bucket = avamar object=backupfile2

Table 5

The second approach is to have a cloud storage gateway and form partnerships with different cloud service providers—such as Amazon, Rackspace, and so on—and host an Avamar Server on their environment.

This would be like buying an Avamar Virtual Edition and using that as server and the cloud storage gateway as client. Hence, the data can be backed up. In Cloud storage gateway, avtar needs to be installed and that needs to do the backup in addition to its translation functionalities.



Cloud Tiering Appliance

This appliance has both backup and archiving functionalities, support for multiple cloud service providers, and can provide the support for file level deduplication by integrating with EMC VNX. Thus, if it can provide block level deduplication with good performance then this can be easily deployed as cloud storage gateway. Another functionality which will make this very attractive is support for automatic and seamless migration of data from one cloud service provider to another Table 6 illustrates the probable steps that can be followed by the Cloud tiering appliance.

The appliance should automatically retrieve the objects from Amazon S3 and then create space on the new Rackspace storage and store the old contents and update the file location for each of the pointers. This option can be made available in the cloud tiering appliance after the customer submits payment, gets the list of URLs from Rackspace, and provides that to the cloud tiering appliance.

File	Amazon location – Premigration Pointer on the CTA	Post Migration – Rackspace, Create data	Post Migration – pointer on the CTA
File X -> 1-4096 Bytes	url=http://doc.s3.amazonaws.com/2006-03-01 Bucket = avamar object=backupfile	PUT /<api version>/<account>/<container>/<object> HTTP/1.1 Host: storage.cloudrive.com X-Auth-Token: eaaafd18-0fed-4b3a-81b4-663c99ec1cbb ETag: 8a964ee2a5e88be344f36c22562a6486 Content-Length: 4096 X-Object-Meta-PIN: 1234	url=<api version>/account/container/object X-auth-Token=" token"

Table 6

The steps are:

Retrieve data from old service provider -> create space on the new cloud service provider -> store the data -> update the pointers on the cloud tiering appliance.

Data Domain

Data Domain performs target level deduplication and probably can be used by customers to attach to cloud tiering appliance or VNX and deduplicate and store the least recent or archived data.

VNX

The integrated solution of VNX and cloud tiering appliance has the potential to act as cloud storage gateway. The product roadmap should include the following features:

- Block-level deduplication
- Provide features which will help the customer seamlessly migrate from one service provider to another

How to decide on the Cloud Storage Gateway

A customer should determine requirements after defining:

- The current environment and how the storage is organized.
- The future roadmap of the data center, the storage needs, and usage of different applications by the user.
- Future performance requirements.
- How the majority of users access the data.
- If you plan to use the cloud only for backup and archiving purposes.
- The regulatory compliance requirements in the respective industry

Each of the factors above change the requirements for cloud storage gateway.

1. Current environment and how the storage is organized: The factors to look for are:

- Data stored on multiple data centers spread across different locations. Clearly, if data is stored on multiple locations, the cloud storage vendor must support multi-site. Whether the vendor supports collaboration across multi-site also comes into play.
- The current throughput of the data center on the storage side. This will help in evaluating the data sheet of the cloud storage vendor, its performance, and whether it exceeds the current throughput.
- The type of SAN and storage arrays are used (CLARiiON[®], Symmetrix[®], Netapp Filers, etc.). Compatibility with existing storage arrays is a must for near-line storage because this will help preserve investment of existing infrastructure

- The type of backup used and whether tape is used (EMC Avamar®, Symantec , Commvault, etc.). Compatibility and migration capabilities of the vendor will come into play.

2. Future roadmap and growth needs

Future roadmap may include building data centers in different locations or starting operations in new geographies. Thus, multi-site support will be a requirement.

Changes in the backup and disaster recovery requirements in the near future (Recovery Time Objective and Recovery Point Objective) may create a problem if backup requirements change and the vendor cannot back up in that backup window. Another changing requirement include instances where an application administrator wants control and visibility on where their data is stored. For example, this capability must be supported by the vendor to give visibility to the VMware administrator on where their VM's are backed up.

Company acquisitions will mean integrating a new data center, once again bringing multi-site support into play.

Performance requirements may change in the future. For example whether access time has to improve over a period of time from 10ms to 1ms or IOPS has to increase from 100 to 1000.

New applications may be added. For instance, performance requirements for database applications will be high. Consequently, check the vendor capabilities on handling database applications. Size and speed factors must be considered.

3. How the majority of users access the data.

Users may be accessing their data from mobile devices.

Accessing over VPN. Delay characteristics of the network and the application requirements must be considered.

4. Whether the cloud is used only for backup.

What are the RTO and RPO requirements for backup?

Is the backup done for servers or clients?

Evaluating these things will result in a list of requirements that can then be matched to vendor requirements to select the best match. In addition to matching the requirements, IT decision makers also have to evaluate on the set of categories shown below.

Evaluating Cloud storage gateway

- Ease-of-use: How easy is it to migrate from one cloud service provider to another?
- Analytics and Reporting: Does the Cloud storage gateway offer an integrated view of storage needs, predict problems and bottlenecks, and provide estimates on performance of each type of data stored in different places?
- Considering the future product roadmap.
- Compatibility with different storage vendors (Storage OEM).

Business Environment and how the market looks

Currently, cloud storage gateway vendors are dominated by start-ups and small firms such as Twin strata, StorSimple (acquired by Microsoft), Nasuni, Panzura, Riverbed, and others. Large companies such as EMC and NetApp have not ventured fully into it. They currently provide part of the solution, like EMC Atmos. According to Gartner, the market size was around \$10 million in 2010 and poised to grow to \$400 million by 2014. The main reason behind this expected growth is that public cloud storage adoption by enterprises is poised to grow at a faster rate and cloud storage gateway aids in public cloud storage adoption. The predicted CAGR from 2012-2016 is 26.5% for adoption of public cloud.

Defining the market segment

There are three major types of cloud offerings.

1. IaaS – Infrastructure as a Service. In this category the storage, network, and compute elements are offered as service
2. PaaS – Platform as a Service. Google apps fall under this category.
3. AaaS – Application as a Service. Salesforce.com comes under the category.

The target market for cloud storage gateway is those using IaaS and, more specifically, Storage-as-a-Service. Typical customers are those using cloud only as a storage option. Growth rate of these segments is shown in Figure 7. This report is taken from Gartner.

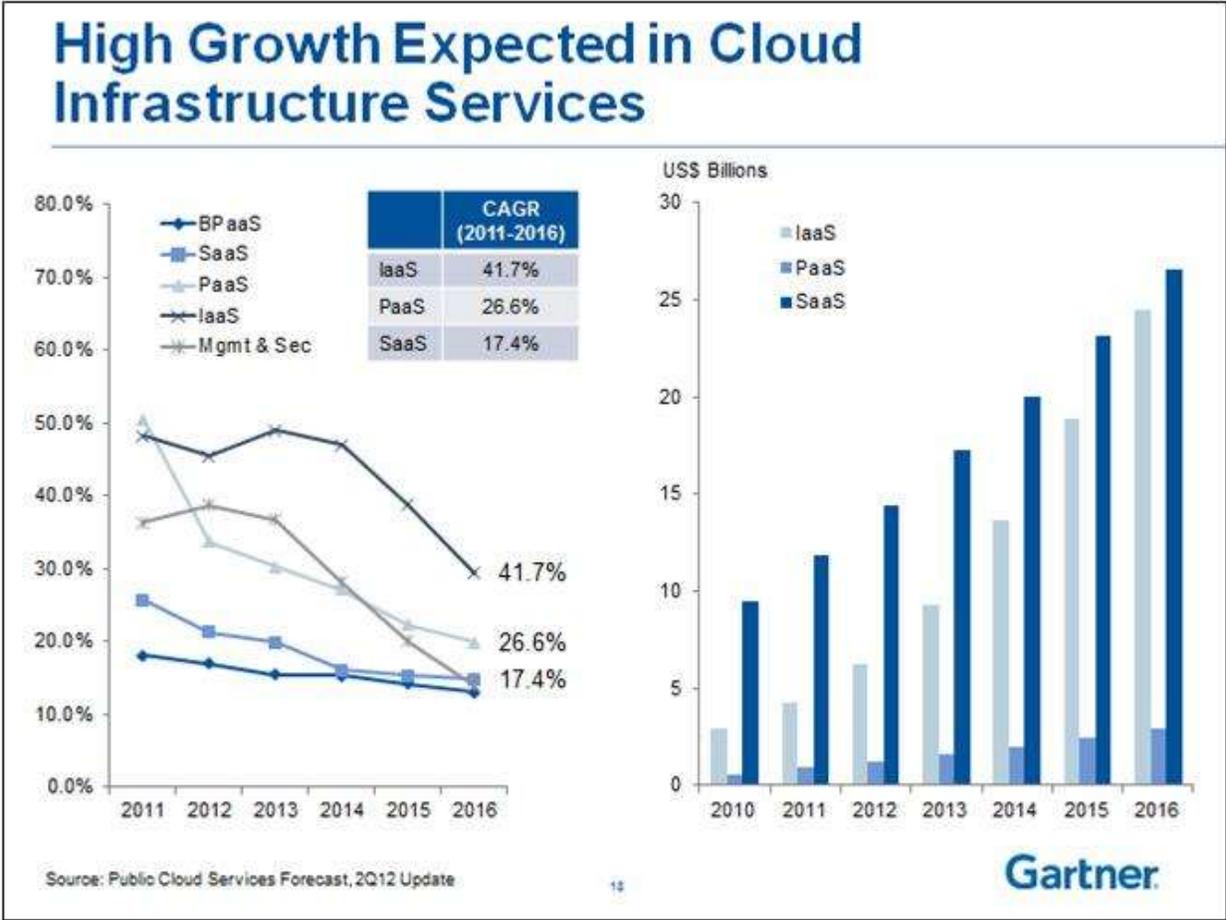


Figure 7

Storage-as-a-Service

Applications are hosted on the client premises and only the storage is on the cloud. Thus, when selecting cloud service providers, performance requirements of the applications are important.

Two things are particularly critical:

1. IOPS of the cloud service provider
2. Throughput

There are cloud service providers who provide storage on SSD for better performance. One such service provider—Rackspace— have published performance numbers. Some applications will have strict performance requirements and customers want to have it in-house or in private cloud, so most likely the deployment of cloud storage gateway will be in hybrid cloud. In addition to providing gateway functionalities to different public cloud service providers it should also have the capability to fetch data from private or public cloud based on the application. The customer can have private cloud using EMC Atmos platform and cloud storage gateway should integrate with that platform.

We will look at what types of applications are more suitable to cloud storage.

1. Email storage and archiving
2. Videos of your hosted website
3. Backup and disaster recovery
4. Database applications where the application is on the client and retrieves data from the cloud and to do queries

Customers want more control and visibility through Reporting and Analytics

Customers moving to cloud are often willing to pay a little more to gain more control and visibility. The cloud storage gateway should be able to display which data resides where and IOPS and throughput for that type of data. This type of analysis will help storage and application administrators determine whether automated tiering is working properly. In addition to that, the tool should also offer flexibility to move the data manually thereby meeting the performance requirements of the application. Greater control improves the confidence level of the administrators and they start investing their time in building more applications which will benefit the business.

Though not an exhaustive list, the set of reports below will be useful.

Type of Report	Why it is required
Where is particular data stored? Classifying the data based on customer need is the key to getting a useful report.	This will provide more visibility and also help in predicting the performance of each application.
Usage of each of the resources in each service provider category (service provider 1, service provider 2, private cloud) etc.	This will help to effectively use the resources and in planning ahead.
Cost and Performance in each of the storage options. Cost also should incorporate the cost of failure.	It will help in determining if we are over paying and in deciding which option is better.

Below is a detailed report from the list above.

Application/File	IOPS	Throughput READ Speed	Throughput Write Speed	Where it is stored
Testing.wmv	25	1Mbps	500Kbps	Amazon S3 – Geo location should be provided
Image.jpg for intranet	30	2Mbps	500Kbps	Rackspace - SSD

Major segments in cloud storage gateway and the major companies

Segment	Companies
Security	CipherSecurity, Perpesys
Cloud Storage gateway	Ctera, storsimple , Twinstrata
Visibility into cloud usage	JReport, NewVem, Cloudyn

Players in this space include JReport, NewVem, and Cloudyn.

NewVem helps customers show how they are using their Amazon web services and help them use them effectively to lower their bills. Cloud storage gateway should have this type of reporting capability also to give more control to the administrators.

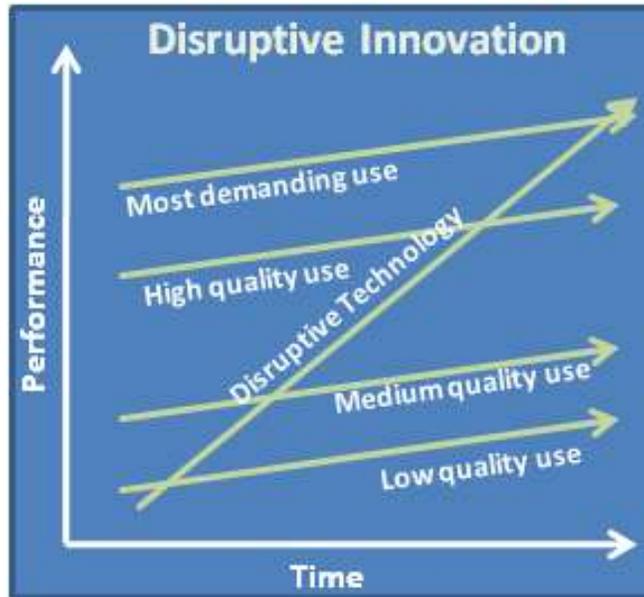
Backup to the cloud market is also poised to grow at a faster rate, as shown in Table 7.

	2011	2014
Cloud Storage	4	10
Backup in cloud	0.44	1.09
Total Backup	2.40	4.09
Percentage of revenue from cloud backup	18.18%	26.70%
		in \$Billion

Table 7

Does Cloud storage gateway have the potential to disrupt the storage market?

Disruptive things are new things which at first are not of great quality, targeted at niche markets, and provide a very different value proposition. However, as time passes, the quality improves and the new value proposition of that innovation becomes attractive and, as a result, is adopted by the mass market. Table 8 illustrates how cloud and mobile phone fall into that category.



Mobile Phone vs. Landline	Less noise quality but mobility
Cloud	Lesser cost but lesser security and performance

Table 8

Cloud storage gateway can be seen as a device which aids in adoption of cloud storage. Since Cloud Storage is disruptive, Cloud storage gateway also aids in disruption.

Conclusion

This article will be most useful to IT decision makers who are already using or planning to use Cloud Storage. It discusses the EMC products which fit in the Cloud storage gateway market and goes on to explore the missing options and how to fill those gaps from the EMC point of view.

References

<http://venturebeat.com/2012/12/28/5-enterprise-cloud-predictions-for-2013/>

<http://www.eweek.com/c/a/Cloud-Computing/Public-Private-Cloud-Markets-Set-to-Soar-as-Enterprise-Adoption-Grows-586161/>

<http://www.technavio.com/content/global-cloud-storage-gateway-market-2012-2016>

<http://searchcloudstorage.techtarget.com/feature/New-hybrid-cloud-storage-appliances-attract-attention-for-primary-data>

http://www.nasuni.com/blog/32-blocks_vs_files_which_approach_is_better_for_cloud

http://www.theregister.co.uk/2012/11/06/twinstrata_cloudarray_v4/

<http://www.rackspace.com/blog/cloud-block-storage/>

<http://www.forbes.com/sites/siliconangle/2011/12/07/20b-cloud-storage-market-heats-up-hp-bundles-box-net-for-consumers-nirvanix-takes-hybrid-clouds-to-the-enterprise-2/>

<http://www.enterprisestorageforum.com/storage-management/seven-solid-cloud-storage-apps.html>

<http://www.enterprisestorageforum.com/technology/features/article.php/3843151/Why-Cloud-Storage-Use-Could-Be-Limited-in-Enterprises.htm>

<http://venturebeat.com/2012/04/09/how-israeli-cloud-management-startup-newvem-is-cutting-its-clients-amazon-web-service-bills-by-25/>

<http://gigaom.com/2012/09/27/newvem-adds-amazon-s3-storage-analytics-to-its-roster/>

<http://searchcloudstorage.techtarget.com/feature/Buying-a-cloud-storage-gateway-appliance>

http://wikibon.org/wiki/v/Cloud_Storage_Gateways:_Opportunistic_features_or_the_next_generation_storage_systems

<http://www.businesscloud9.com/content/gartner-iaas-takes-over-traditional-data-centre-outsourcing/11472>

EMC believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

THE INFORMATION IN THIS PUBLICATION IS PROVIDED "AS IS." EMC CORPORATION MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND WITH RESPECT TO THE INFORMATION IN THIS PUBLICATION, AND SPECIFICALLY DISCLAIMS IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Use, copying, and distribution of any EMC software described in this publication requires an applicable software license.