Are you aware of the hidden costs of storing data in the cloud?

Every enterprise, regardless of the industry, struggles with the management of data. Average data growth is between 35% and 65%, compounding yearly. This growth is managed within IT budgets that are only growing at an average of 7% annually. In order to meet budget requirements, while maintaining the value of data to the business, many IT and Storage professionals have turned to cloud-based storage for storing second copies of data or for long-term retention of infrequently accessed data.

Cloud Storage makes a good first impression

At first glance, cloud storage solutions from leading providers of these types of services appear to be "too good to be true". Each vendor offers a fixed cost for storing the data on a per Gigabyte (GB) per month basis, with free uploads of the data to the cloud provider's site. For the IT Storage professional this is a fixed cost that can be moved into the Operating Expense (OPEX) category of their budgets. Moving long-term storage seems like a great idea, it reduces the impact to the overall IT capital budget and makes accounting for the data fairly simple. The issue with these services is that the motivation for the cloud vendors to offer low per GB per month pricing is often hidden in the fine print.

The two most prominent vendors in the long-term cloud storage space are Amazon Glacier and Google Cloud Storage Nearline. The lowest cost provider in the cloud offering space is Oracle Archive Storage. These three vendors understand the storage market very well. They realize that very little data is 100% untouched forever and users can't predict which data will be required and when. The cloud storage vendors have put pricing factors in place to ensure that they can make a profit from this periodic need to access archived data. Furthermore, if the stored data is not being used in the compute portion of these vendors' environments, the cloud vendors impose a cost to deliver the data back to the owners.

Consider the facts

Retrieval and export of data are based on a per GB basis, with the first GB being free to retrieve. The customer pays a charge to retrieve each subsequent GB using a tiered pricing model. Crucially, this tiered pricing is maintained over the specified period. This means that once a couple of GBs have been retrieved and/or exported, ***the rest of the period is at the higher tiered data retrieval pricing***. In many instances data retrieval and export charges dwarf the cost of the storage.

More data equals exponentially increased fees

All three of the vendors mentioned above charge additional fees for commands sent to retrieve data from long-term storage. The fee seems very small, a couple of pennies per 1,000 commands, but the number of commands to retrieve data can grow rapidly if not managed carefully. Commands for single files can rapidly mount up and before users know it, the cost of restoring just a tiny fraction of the data stored in the cloud, can cost several dollars or more.

Since long-term data can be viewed as: a source for historical analytics; a secondary copy (air gap protection); or even as primary low-cost storage for infrequently accessed data, there is a reasonable probability that a non-recurring percentage of data will be accessed on a monthly basis. With this in mind, whenever IT professionals are looking for a low-cost storage solution for infrequently accessed data, they should include a monthly cost to retrieve some percentage of their data in the business case. This is the only way to make a true and fair comparison to other long-term storage mediums. Based on a survey conducted by Solutions North Consulting, an average of 10-15% of archived, long term retention of secondary copy data is retrieved monthly by users or administrators.

For any enterprise, these hidden costs add up over the life cycle of data stored with apparently "low-cost" cloud storage providers. The costs can increase enormously in a very short time simply by changing the amount of data to be retrieved by just a couple of percentage points. By the time the data is pulled back onsite for usage by the end user, the cost of the data is higher than using primary storage in many cases.

LTO Technology as an alternative solution for long-term retention

It's crucial to state that LTO Technology has none of these costs associated with storing and retrieving data. In fact, LTO tape technology tends to get less expensive over time due to the price erosion of key elements such as tape media. Even *after* factoring in the costs of administration, power, floor space, energy consumption and capital expenses for hardware and software, LTO tape technology offers a much lower Total Cost of Ownership for rarely accessed data. LTO tape technology also offers the peace of mind that the data can be validated and stored for many years with no retrieval costs or retrieval command costs. Plus, users need not worry about the reliability or financial viability of a third party to keep their data safe.

Every enterprise has a different use case scenario. That is why the LTO TCO Tool allows you to input your own scenario, and receive instant feedback on the real costs associated with managing data.

Unless you are storing less than 300 TeraBytes (TB) of data, and regardless of OPEX or CAPEX, the cost of storing data in the cloud can change dramatically.

Don't get caught off guard with the fine print and hidden costs of current cloud storage offerings.

About the deeper assumptions of the TCO tool

The calculations in the TCO tool are based on industry supplied information. The assumptions are listed for ease of review based on the technology in the TCO tool.

Cloud Storage Pricing and Assumptions

The costs of cloud storage are based on information supplied directly from each providers Website for lowest cost data storage. The calculations do take in to account the following pricing:

http://aws.amazon.com/glacier/pricing/
https://cloud.google.com/storage/pricing
https://cloud.oracle.com/en_US/storage/archive-storage/pricing

- Tiered data storage pricing based on storage amount
- Tiered data upload pricing based on upload amount (where applicable)
- Tiered data transfer command pricing for upload and /or retrieval based on the transferred data command count of 4GB of data per command.
- Tiered data retrieval pricing based on percentage of data retrieved
    - Tiered pricing during the applicable period of 30 days assumed to be per calendar month as a standard
- No on premise management costs are assumed in the calculations.
- Network transfer charges are not calculated in the tool
    - These charges are highly variable and would be required with any off-premise solution.
- All pricing as of 1Q2017
- A time based reduction of cost is assumed in the TCO tool, based on a sampling of pricing for Amazon Glacier over a 4 year period ending in 1Q of 2017

Disk based Object Storage

- Calculations assume lowest cost commodity disk storage as of April 2017.
- Assumes usage of distributed RAID reducing the requirement for multiple copies of the data
- Power is assumed to be 600W per controller and disk drawer.
    - Power consumption for controllers and disk drawers can vary
    - Model assumes constancy in the requirement for redundancy drives 60W per 4U device.
- No power down time is assumed in this model.
- No data compression is assumed in the model

- Disk pricing is assumed to continue to decrease on a per GB basis at the same constant rate that disk pricing has decreased over the 3 years preceding 1Q2017
- Disk capacity growth is assumed at a constant rate based on sampling over the 3 years preceding 1Q2017
- Assumes the usage of Industry software that comes with published support
    - The pricing for software was provided by 2 industry leading vendors.
    - Standard industry discounts of 50% are assumed
    - The value of the software is significant, but not in the overall pricing that may result from higher discounts on software.
    - Floor Space consumption is accounted for based on 60 Disk trays, in a 4U form factor installed in a 72U rack 2 disk subsystems per controller.

File System Based Tape Based Storage

- Tape Cartridge pricing:
    - Current pricing based on Sample of market pricing as of 1Q2017
    - Future pricing based on pricing per GB based on the 3 years prior to 1Q2017 and assumes a linear model
- Tape Capacity based on:
    - Current available capacity cartridges
    - Published roadmaps for capacity increases
    - No compression of the data is assumed in the model
- No power down time is assumed in this model.
- Hardware pricing
    - 50% industry published standard discount is assumed
    - Hardware pricing provided by 3 industry providers
        - Not all hardware pricing was supplied by LTO Consortium members.
    - Higher discounts of tape hardware only change the model by an average of $135,000 per solution.
        - This will only impact the smallest of the solutions and only in a positive manner giving tape an unfair advantage
        - Discounting above 65% is not a sustainable model for hardware suppliers that work in a discount sales model.
- Assumes the usage of Industry software that comes with published support
    - The pricing for software was provided by 2 industry leading vendors.
    - Standard industry discounts of 50% are assumed
    - The value of the software is significant, but not in the overall pricing that may result from higher discounts on software.
    - Floor Space consumption is accounted for based on the average floor space required to support the 2 most dense (in terms of capacity and tape drives) industry tape automation devices.

Energy Consumption

- The TCO tool assumes that 95% of IT organizations are not directly charged for power consumption. This assumption is based on a sample of data from 25 mid-large sized IT organizations from August 2016 thru December 2017.
- Although some organizations may be charge indirectly through charges for floor space, the calculations the TCO tool, help to determine the real costs of powering equipment and the impact benefit of those solutions.
- Power is calculated using United States National average for energy consumption per KwH as of November 2016.
- Energy cost is assumed to be flat over the 12 year maximum term of the tool
- Energy cost is significant, however in those solutions including Energy consumption making prediction on energy pricing would most likely create a variance of less than 10% in comparative pricing.