LTO PROGRAM TOTAL COST OF OWNERSHIP (TCO) TOOL USER GUIDE

ULTRIUM LTO

TAXONOMY

DATA SCHEME

This option establishes whether the solution is traditional data storage directly storing a copy of the data on tape or Erasure coded data across multiple tapes and possibly libraries. The fundamental difference in this choice is the resulting capacity of the solution will be 26% greater in consumed storage than the direct scheme.

OPERATIONS COST

Represents the support required from an information technology organization for security, planning, networking, servers, and tape infrastructure support.

COST OF MAINTENANCE

The services, hardware and other infrastructure required in the continuous operations of the infrastructure. This includes hardware replacement costs and service costs of personnel.

WATTS

A standard unit of energy.

kWh

Kilo (1000) watts per hour. A standardized unit of measurement for energy consumption.

BTU

British Thermal Units. A standardized measurement of thermal energy required to cool a given device at a specific power usage level.

TELCO

The calculated value for charges related to exchange of data over standard internet connectivity, formulated to be $1.49 US per MBps.

TERABYTE (TB)

Unit of digital storage measurement equivalent to 1000 GB.

PETABYTE (TB)

Unit of digital storage measurement equivalent to 1000 TB.

JBOD

Just a Bunch of Disks. Shorthand indicating a Controller/Expansion enclosure designed to house many HDDs to store data.

The controller has no special function or features in the design, making the costs of the enclosures relatively low when compared to specialty function HDD enclosures.

In this tool JBOD is calculated using Open Compute Project: Bryce Canyon enclosure specifications.

MANAGED SERVICE PROVIDER (MSP)

Often referred to as cloud providers, Managed Service Providers provide managed storage at multiple levels and charge based on agreed upon metrics for the storage and management of data. In this tool Amazon AWS Glacier is the comparative solution.

SUSTAINABILITY CHARGE

The CO_2e impact as calculated against raw values without offsets applied. This is the measurement of the individual product lifecycle without regional considerations for recycling and energy production impact reductions. Any offsets would apply artificially, resulting in comparative offset contribution to all solutions as a ratio against the original CO_2e impact of the products.

The LTO TCO tool is not a guarantee of pricing, service, SLA, availability of data, durability of data or functionality. When available publicly available links have been provided for cross referencing to HDD and AWS pricing comparisons. This tool is cross checked independently against available information and tools from vendors and organizations used in the compare.

TAPE

- Retail pricing and discount levels have been used from multiple tape vendors to determine pricing for hardware.

- Media pricing uses publicly available data from the internet collected at the time of the tool publication.

- Software solutions include hardware, support and software for tape data management. Support includes US averages for cost of employees in the software field.

- Commercial Software purchase verses open-source development of software incurs similar cost structures.

- Energy requirements assume 24x7x365 operations at 85% utilization.

- Media capacity is based on roadmap projections as of the date of the tool release. Refreshes and new hardware utilize the forecasted capacity.

- Market media pricing decreases are based on 10 analysis of market trends.

- When Erasure Coding is chosen the factor for EC is 1.26 and spread across tape media and automation choice.

- Supporting hardware of servers, connectivity switches, and cabling is calculated using publicly available pricing and configured to 4 tape drives per server and supporting infrastructure.

- Calculations of each level of storage are calculated on the yearly basis with most costs generated by hardware and media purchased in any given year. The variance of smaller units of measurement do not result in more accurate outcomes as deployment of capacity may vary during any given year.

- CO_2e is derived from the IBM tape sustainability report. Value of CO_2e in dollars is based on 2022 $185 per metric ton indicated in the referenced public report on CO_2e impacts.

CLOUD

- AWS chosen as lowest cost service provider with most flexibility in the lowest cost tier for Active Archive/Deep Archive data.

- All AWS Glacier pricing and Service Level Agreements (SLA) are gathered directly from AWS publicly referenced pricing.

- Pricing is up to date as of the published date of this tool.

- See AWS pricing link in the references section.

- Recall and data transfer out (egress) are different services in Glacier. Egress is the charge associated with moving data, back to premises of an organization or user, the terminology in this tool is retrieval. This is calculated per month, estimates of yearly should be divided by 12 and entered as the monthly whole percentage value.

- Telco are calculated charges directly associated with moving data over standard internet connections.

  - No consideration for direct connectivity to AWS services.
    - High degree of variability.
    - Simple added charge to cloud above and beyond the simple TCO compare.

- Operations and management charge considerations:

  - Billing
  - Information security
  - Development
  - Cloud monitoring
  - Access administration
  - Information management

- All charges are configured in Ohio-2 US as a default configuration. This is the least expensive zone in the AWS choices.

- Charges that result in increases in cost of AWS such as multi-zone replication, higher cost zones, etc are not included in this tool due to the complexity of reflecting a fair comparison in the solutions.

HDD JBOD

- Roadmaps are forward looking and may have variances from this tool.

- All pricing based on publicly available pricing and discounts communicated with vendors and sales professionals.

- Individual unit pricing is configurable for price and capacity.

  - Adjustments to the starting variables of HDDs, will be carried through the tool for future projections.

- HDD controller and JBOD are based on Open Compute Project Bryce Canyon specifications.

  - All related outputs are calculated from the open specification.

  - Bryce Canyon rack configuration is 4 expansions per controller, maximum 10 drawers per rack.

HDD JBOD (continued)

- Energy requirements assume 24x7x365 operations at 85% utilization.

- Media capacity is based on roadmap projections as of the date of the tool release.

  - Refreshes and new hardware utilize the forecasted capacity based on starting capacity and pricing.

- Market media pricing decreases are based on 10 analysis of market trends.

- Erasure Coding factor of 1.26 is included in the capacity requirements.

- TCO Tool assumes data archive usage of HDD.

  - No performance factoring.

  - The tool calculates infrastructure including number of HDDs based on capacity. Capacity is rounded up per drawer year.

  - Storage is calculated on the yearly basis with most costs generated by hardware and media purchased yearly.

  - Smaller units of measurement do not result in more accurate out comes as deployment of capacity may vary year to year.

- Maintenance of HDD is established based on the analysis of multiple industry comparisons for OCP solutions in Hyperscale environments and research infrastructures.

  - Actual maintenance cost can vary. This tool makes every effort to calculate an average cost of support.

- HDD unit failure rate is factored into the TCO.

  - The failure rate is calculated per year, per chosen refresh cycle and resets with refresh cycles.

  - Failure rate is based on publicly available Rackspace Inc. data (see resources)

- No attempt is made to compare to commercially available, feature rich performance solutions such as Netapp, Dell Data Domain®, IBM DS8000®, Puresystems® or others.

  - These systems are performance focused or offer features that are not likely to be fairly compared.
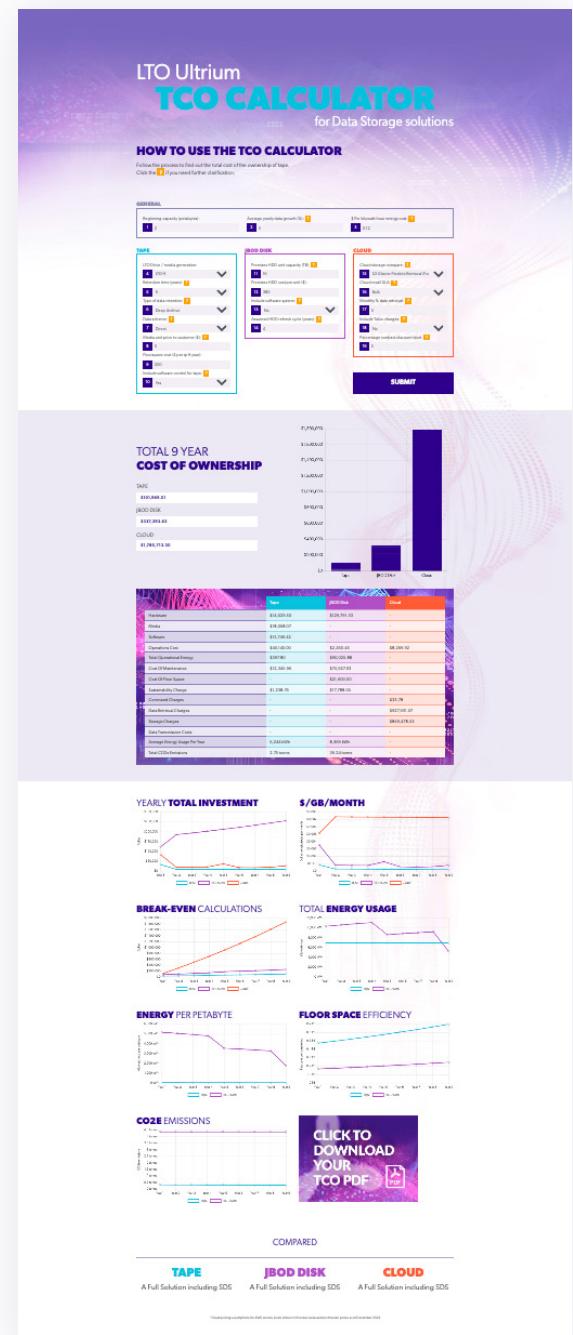
Required Information:

- Beginning capacity in petabytes.
- Type of data retention (higher retrieval should be Active Archive, not Deep Archive).
- Length of time to compare solutions.

For more accurate compares:

- Average yearly data growth percentage.
- Cloud monthly percentage data retrieval.
- Client HDD refresh cycle in years, this could also be the depreciation timeframe.

When entering data capacity, it should be entered in single copy, native data capacity. When the compression ratio for the data is known, enter the starting capacity as the compressed capacity. Here is an example: My data is a native capacity of 10PB, but the data compression ratio is 2:1, I will enter 5PB as the starting data in the tool. The tool assumes compression is available in the HDD software to reduce total data usage.

INPUTS

Entry	Options	Description
Beginning capacity (petabytes)	<customer information>	This is the day-1 capacity indicated by the client or user of the tool. Inputs in this field should align to a single copy, uncompressed capacity. Lowest capacity supported by the tool is 2 PB.
Average yearly data growth (%)	15%	CAGR% Enter the average over the expected time period. Global average is 13%, some clients may be growing much more rapidly. 15% provides clients with a reference of the growth in the different solution costs.
$ per kilowatt-hour energy cost	0.12	Baseline in US dollars, derived from the 12/2021 US national average. Use this field to account for renewable energy offsets or higher costs where applicable.
LTO Drive / media generation	LTO-9 LTO-8	The tape drive and media are assumed to be congruent generations when comparing the solution. If a plan to use LTO-8 media is expected, use LTO-8 as the option in this field.
Retention time (years)	1-10	The number of years the data will stay in the solutions. 3 years is the minimum recommended. Planning longer than 10 years is speculative and not considered realistic for comparison.
Type of data retention	Deep Archive Active Archive	This will change drive and media ratios for tape. LTO tape is an archive medium, as such the tape drive to tape media ratio is high for all comparisons and vary by automation choice. Deep Archive has the highest media to drive ratio. No performance tape option is considered.
Data scheme	Direct Erasure Coding	Direct indicates a single copy of data being saved to tape, two copies should just double the output config, while cutting support cost for second copy in half. Erasure utilizes 1.26 EC ratio to save data on tape.
Media unit price to customer ($)	<blank>	A blank or a 0 (zero) in this location will use the media default pricing. Blank uses the latest analyzed market pricing. Enter a desired pricing only when quoting the pricing is eminent.
Floorspace cost ($ per sq ft year)	200	Keep at $200 per year unless the end user desires something different. This is an average and does not significantly impact the output.
Include software control for tape	No	Choose Yes if the desired SDS is known or new to the client and they ask for a total compare of the solution. Feature enablement adds servers, software and the associated cost of development or commercial acquisition equivalent to the infrastructure.
Premises HDD unit capacity (TB)	16	Starting Capacity Default to 16TB. As of 2022 this is the capacity most likely to be implemented in JBOD solutions. The tool uses current HDD roadmaps to determine transition of new capacity and refresh and adjusts for the starting capacity and pricing.
Premises HDD Cost per unit ($)	385	Current average market price (as of the date of the tool).
Include software system	Yes No	This is a feature addition that includes the cost of software, servers and administration using colocation pricing. This would be software developed in house, as open source, or as commercial software, all incur the same rate of support or expenses. These charges tend to vary by customer but incur at the same rate regardless of the method of deployment.
Assumed HDD refresh cycle (years)	4	This is the lifecycle of data storage in years. Minimum of 3 years, maximum of 10 years. This is the number of years the client will retain the disk, also influences the failure schedule. Longer periods will not reduce costs as HDD fall out increases dramatically every year after 5. The refresh year always results in the refresh cost impact in the following year.
Cloud storage compare	S3 Glacier Instant Retrieval S3 Glacier Flexible Retrieval S3 Glacier Deep Archive	AWS is used in this TCO as it represents the combined market competitive position by offering. For AWS Deep Glacier, the standard recall time is 24 hours but the pricing per GB month is extremely low. This should be left as S3 Glacier Deep Archive unless specific knowledge is available otherwise.
Cloud recall SLA	Standard Bulk Expedited	No provisioned capacity unit charges, assumes data duration is always exceeded. This field varies by the choice made in the cloud storage compare field. Bulk has the lowest cost for recall, but also the longest SLA. Amazon does not guarantee SLA in the standard cloud contract.
Monthly % data retrieval	10%	These are egress and not recall to S3 for compute or transferring zones, the tool does not calculate zone transfer charges. 10% is the industry average for recall of archive data. Only the cloud solutions have a charge for recall and retrieval (egress as termed by AWS).
Include Telco charges?	Yes No	This will compute upload and egress charges based on the answers above against the reference standard of $1.49 per MBpS. Cost of internet data transmissions for cloud included or not. The TCO tool does not calculate dedicated connectivity charges due to the extreme variability.
Percentage contract discount level	5%	5% is known to be a reasonable discount for clients in the 43PB to 50PB range of data. Change this to the appropriate figure if you believe you receive a different discount level.

OUTPUTS

Result	Description
Hardware	Total cost of hardware related to the solution.
Media	For tape this is the cost of tape media, For JBOD this is HDDs.
Software	Cost of implementing software regardless of open-source, in-house development or commercial acquisition.
Operations Cost	Costs associated with floor personnel, administrators, information security, physical security to operate the storage.
Total Operational Energy	Energy to operate the solution and thermal energy to cool the infrastructure. Lower cost impact such as utilization of renewable energy, scale proportionally by solution and do not effect the outcome in terms of scale comparisons.
Cost of Maintenance	Relative cost of continuous operation support service, regardless of a purchased agreement or internally self-service.
Cost of Floor Space	Calculated against the dimensions of each solution and scale, including isle-ways where required. This can be used to estimate the cost of real estate, depreciation of facilities or cost of co-location facility charges.
Sustainability Charge	Calculated self-imposed charge based on the total estimated CO2e emissions in metric tons/tonnes charged against $186 per metric ton/tonne global impact assessment. This charge is without offsets and should be considered the marked comparison value for the solutions.
Command Charges	Total cost of charge metric from AWS.
Data Retrieval Charges	Total charge by AWS for all transfers of data back on premises.
Storage Charges	Base charge for storing data for the time period designated in the lifecycle.
Data Transmission Costs	Cost of moving data to and from AWS Glacier over standard internet connections. Customized connection varies greatly, placing them outside the scope of this computation. It should be noted that customized direct connections incur both higher AWS charges, but also dedicated security and hardware considerations.
Average Energy Usage Per Year	This is the calculated per year energy usage in kilowatt-hours averages over the indicated solution lifecycle.
Total CO2e Emissions	This is the total calculated CO2e emissions from operations of the storage solution.

REFERENCES

https://www.opencompute.org/documents/facebook-bryce-canyon-storage-system-specification
https://www.opencompute.org/products/213/wiwynn-bryce-canyon-sas12g-storage-jbod-up-to-72-hot-pluggable-hdds
https://aws.amazon.com/s3/pricing/
https://www.ibm.com/tape-storage
https://www.backblaze.com/blog/backblaze-drive-stats-for-2022/
https://cloud.google.com/vpc/network-pricing
https://stealth.net/services/internet/dedicated
https://www.seagate.com/global-citizenship/product-sustainability/exos-x16-sustainability-report/
https://www.seagate.com/www-content/datasheets/pdfs/nytro-3000-sas-ssdDS1950-2-1711US-en_US.pdf
https://www.backblaze.com/blog/hard-drive-life-expectancy/
https://www.storagenewsletter.com/2022/01/25/how-long-do-hdds-last/
https://www.linkedin.com/pulse/operational-overhead-hybrid-charles-bystock/
https://www.itprotoday.com/business-resources/casb-and-saas-management-should-you-have-both
https://venturebeat.com/2022/02/18/security-service-edge-splits-off-from-sase-in-new-gartner-magic-quadrant/
Rennert, K., Errickson, F., Prest, B.C. et al. Comprehensive Evidence Implies a Higher Social Cost of CO2. Nature (2022).
https://doi.org/10.1038/s41586-022-05224-9
Brume, S.O., IBM Tape Sustainability Report, February 2022.

Other documentation may have been used to form directions related to governance, expectations and planning of archive storage that are not available for public reference.