Data Storage Technology, Tape, and Sustainability

By Fred Bonner and Peter Guglielmino

Executive Summary

Mostly hidden from view, the environmental costs of storing data are considerable. It can be estimated that computer hard disk drive (HDD) storage consumes roughly **2.5 percent of the electricity produced worldwide.** It is also true that a great amount of stored data will only be accessed infrequently, and **much of this data could easily be stored in ways that are far more energy efficient.**

Data Object Sizes

1 Bit = 1 or 0 8 Bits = 1 Byte
1 Kilobyte (KB) = 1000 Bytes
1 Megabyte (MB) = 1000 KB
1 Gigabyte (GB) = 1000 MB
1 Terabyte (TB) = 1000 GB
1 Petabyte (PB) = 1000 TB
1 Exabyte (EB) = 1000 PB
1 Zettabyte (ZB) = 1000 XB

Technologies exist to greatly reduce, or even *bring to virtually zero,* the energy consumption of stored data. With a goal of replacing over time data that can be appropriately moved to Digital Tape (essentially inert storage) a significant amount of data-storage carbon emissions can be reduced, without sacrificing the integrity of the stored information. In analyzing total cost of ownership, Digital Tape is also *the most cost effective way to archive data*, beating both cloud and HDD in cost per TB per month.

In looking to comply with COP21 carbon emissions reduction targets, the domain of data storage—often completely overlooked as a significant energy drain—offers sustainability leaders an easy pathway to help achieve their emission reduction targets while also reducing costs.

I. Exponential Data Growth is Disaggregated and Largely Invisible

The sheer volume of data that humans generate and choose to keep has expanded exponentially and continues to expand year after year. And yet, as obvious as the ubiquitous nature of data may be, there are two reasons it is equally as difficult to quantify.

First, humans are surrounded with devices that by historical standards can now hold great volumes of data. For example: Apollo 11 landed on the moon using a computer that had 1,300 times *less* processing power than an iPhone 5S.[1] A typical iPhone 6 may store 64GB to 128GB of data; Back in the 1980s desktop disk drives cost $500 and contained 1/1000 that much data-—40MB. Today, clearly, stored data exists in relatively high volumes everywhere.

Second, technologies like Cloud have removed a great preponderance of institutional data from our direct purview, putting it out of sight and out of mind. Yet, large volumes of data from many institutions are being archived for very long periods of time. The question remains, where is the data, and what impact does it have on our sustainability goals?

II. How Much Data Is Out There?

Estimates have been made of both data at rest, and data in motion. Wikibon.org estimated that in 2010 there were over 1.2 ZB of data in the known universe-—1,200,000,000,000 GB.[2] Using that statistic and a growth rate of 10 percent to 15 percent a year, the known amount of data at rest has probably surpassed 2 ZB (2 trillion GB) today.

In terms of data in motion, according to Cisco's Visual Networking Index initiative, internet traffic is now in the Zettabytes as well. Internet traffic includes data flowing from one destination to another. Cisco estimated that by the end of 2016, global Internet traffic was 1.1 ZB per year, and they further anticipated that by 2019, global traffic will hit 2 ZB per year.[3]

The energy challenge to data, however, occurs when it is at rest. There is a distinct *physicality* to the storage of data when it stops moving. It takes up space and requires infrastructure. Moving data to a storage medium and extracting it from a storage medium draws energy. A constant source of power is required to keep storage devices operational, powered up, and cooled. And with “retention windows”-—the amount of time that someone wants to keep data-—iincreasing, the problem of how much data humans create is compounded by the fact that most people want to keep their digital history forever. Big Data is aptly named.

---

[1]Computer Weekly, Apollo 11: The computers that put man on the moon, July 2009 https://goo.gl/5KfBz

[2] Wikibon, Information Explosion & Cloud Storage, https://goo.gl/eF1Ee1

[3] Live Science, How Big is the Internet, Really? March 2016 https://goo.gl/GJegSU

III. Storing the World's Data Draws Enormous Amounts of Energy

Stored data today has significant energy requirements because volumes are high and retention windows are long. And notwithstanding technologies like Flash, essentially solid state storage, and optical storage devices like CDs and DVDs, by a vast margin the two most prevalent ways of storing data "permanently" are HDD systems and robotic Digital Tape systems.

The easiest way to get a handle on the energy costs of storing data on HDDs is to calculate the number of disk drives spinning today. According to sales statistics, there are approximately **2.6 billion disk drives** currently in operation, burning on average 15 Watts of power each.[4]

Here are some factors that affect the analysis of HDD energy consumption:

- A single HDD will last on average about five years.
- Most HDDs spin all day.
- If the average HDD uses 15 watts of power, that means it draws 15 watts in any given instant, and it uses 15 "Watt Hours" per day.
- Some technologies have been developed to reduce the consumption of disk arrays, notably MAID, Massive Arrays of Idle Disk.

Taking all of this into consideration, the following table estimates the daily and annual energy draw of HDDs worldwide:

Power Consumption of Presently Operational HDDs July 2017		
Drives Sold Last 5 Years	2,600,000,000	
Watts Per Drive	15	
Total Watts All Drives	39,000,000,000	
Watt Hours All Drives	39,000,000,000	
Watt Hours / Day	936,000,000,000	
KWH Per Day	936,000,000	
MWH Per Day	936,000	
MWH Per Year	341,640,000	
Reductions		
% Time Powered On	273,312,000	80
% MAID	245,980,800	10
Total MWH / Year HDD	**245,980,800**	

[4] Sales data from Statistica.com. There are many different kinds of HDDs using varying amounts of power. High estimates push wattage usage to 25 watts per drives. Some disk systems might average 10 watts per drive or less. For this analysis our calculations are using a 15 watt per drive metric.

With the aforementioned estimate for energy consumption of worldwide HDDs, environmental control must also be factored into the analysis. Converting megawatts directly into heat (British Thermal Units / BTUs), the number of BTUs generated in any environment must be countered with cooling to keep a constant ambient temperature, which necessitates doubling the annual energy usage for HDDs.. Doing so generates a new total for the energy impact of worldwide HDDs—491,961,600 megawatt hours per year.

How does that compare to worldwide electrical usage? The Organization for Economic Co-operation and Development estimates world electricity consumption per year at 20,000 terawatt hours, or 20,000,000,000 (20 billion) megawatt hours.

This suggests that HDD storage draws approximately 2.5 percent of *all electrical consumption worldwide.*

IV. Digital Tape Can Often Replace Hard Disks, Using Much Less Energy

What is the alternative to HDD storage, and when is it appropriate? For an equivalent amount of data, HDDs use 70 times *more* energy than Digital Tape. This fact has been born out by numerous studies conducted by the Clipper Group in Boston.[5]

The most successful and prevalent Digital Tape format in use today is LTO (Linear Tape Open). According to the statistics of the industry consortium that sells LTO (LTO.com), there are approximately 385,000,000 Terabytes of data being stored on LTO Tape, approximately 200,000,000 (Two Hundred Million) cartridges worldwide. Since cartridge densities have continued to increase, the average cartridge in the field today holds 2 TB. And therefore, in terms of installed base, there are at least 10 times *more* HDDs than Digital Tape cartridges with the same average density. So while Digital Tape is vastly more energy efficient, HDDs are today vastly more prevalent.

Digital Tape cartridges, like HDDs, can be organized into banks, drawers, shelves, or slots, and work together as a unit in storage architectures. Industry measures the density of a storage system by the data volume of its fundamental storage unit -- the single tape or the single HDD. And densities have been doubling every 2.5 years for the last several decades for both tape and disk. The latest tape technology, LTO7, holds 6TB of data, the same amount as the most popular and cost effective HDDs.

[5] In Search of the Long-Term Archiving Solution: Tape Delivers Significant TCO Advantage Over Disk, December 2010, Clipper Group https://goo.gl/OxbW3

HDD performance is measured by the speed with which data can be placed on or removed from a disk – known as I/O (Input/Output). And here lies the major difference in general between HDDs and Digital Tape: HDDs, especially used in groups, can store and access data with a much faster instantaneous I/O compared to Digital Tape. In evaluating the use of disk versus tape, the difference in I/O speed is critical for some use cases, *but insignificant for others.* The challenge is to find the use cases that do not require fast disk I/O, and to switch those workloads to far more energy efficient tape.

In terms of pollutants, 3 PB of disk, running all year, can produce over 14 metric tons of CO_2. The tapes, being largely inert, produced almost zero CO_2 over the same time.

To understand the CO_2 impact of computer storage, convert the wattage used by storage systems into BTUs (British Thermal Units – 3.14 BTUs per watt hour.) The following table shows a comparison among three types of computer storage systems (two disk, one tape), with the BTU impact for each.

Tape vs Disk Carbon Footprint: 3 PB For One Year

System	Configuration	BTUs / Hour	BTUs / Day	Cooling BTUs	Total BTUs / Day	BTUs / Year	Metric Tons Carbon / Year
NAS Storage (Drives)	498 6TB Drives	17,607	422,568	422,568	845,136	308,474,640	18
Isilon (240 TB)	12 X 240 TB Isilon Archive Systems	43,200	1,036,800	1,036,800	2,073,600	756,864,000	44
EchoLeaf (Medium Config / IBM 3500)	IBM TS3500 Tape / 6 Drives / Lenovo RD540	510	12,240	12,240	24,480	8,935,200	0.5182

BTU/ Wattage Data from IBM Data Sheets for specific products:

- ➔ Watts = 3.41 BTUs / hour
- ➔ BTUs / Day = (BTUs / Hour) * 24
- ➔ Cooling Needed = Equivalent of BTUs produced.
- ➔ Assumption BTUs generated by cooling system are equivalent to BUTs removed in a given environment.
- ➔ Tape as % of Disk = (Tape /100) * Disk.
- ➔ BTUs/ Year = Daily BTUs * 365
- ➔ Metric Tons Carbon = [(BTUs/Year)*0.058]/1,000,000 (1 MM BTU= 0.58 metric tons carbon)
- ➔ Gas Consumption = Metric Tons * 2205 (Lbs in a ton) / 19.4 (19.4 Lbs CO2 in emissions of 1 gallon gas)

Digital Tape per TB stored uses less energy than HDDs by a factor of at least 70. So why are HDD's so much more prevalent in the market? Because HDDs and Digital Tape have different performance profiles. Yet while some applications certainly require the performance characteristics of HDDs, *an increasing number do not*.

Three primary factors explain why industries continue to favor disk storage over the low-energy Digital Tape option: Input/Output time, seek time, and data transfer speeds.

First, I/O, as previously discussed, offers almost instantaneous access to data and is essential for many applications, especially those requiring transactional information.

Second, Digital Tape requires "seek time" to find queried data. Once data is requested from a tape, it has to spin the tape to the section that has the needed data. This delay is known as "tape-to-data" time, and it is far longer on tape (sometimes minutes) than on disk (usually milliseconds).

Finally, while a *single* tape cartridge can move blocks of data faster than a *single* HDD, a highly tuned hard disk system—many HDDs working together—can move blocks of data from a storage *system* faster than tape.

However, there are many use cases where performance characteristics of disk against tape become irrelevant. When the volume of the data being moved is very large, disk I/O and tape seek-times are unimportant. For example, archiving a movie in an uncompressed digital format might take over a TB of storage. And that TB (1,000 Gigabytes) might take upwards of 30 or 40 minutes to move from place to place on an enterprise network. In that case, the disk system and tape will get the data to where it's going in about the same time.

While short seek times are expected for daily consumer data queries—text messages, emails, bank balances, studio schedules, food orders, etc.—they are often insignificant for pulling data from large archives that are rarely or never accessed, but which organizations wish to preserve. Examples might include raw movie footage, architectural plans, archived bank transactions, document archives, photo archives, etc. Archives like these can be much more effectively stored in inert archival formats, like Digital Tape.

Additionally, it is also clear that certain large data objects like video files or unstructured data streams are optimized with sequential storage access – for which Digital Tape is perfectly suited. If there is a true requirement for multiple concurrent user access, this can easily be supplied with hybrid architectures that use both tape and either disk or solid state devices.

With plentiful use cases suited to energy saving Digital Tape, simply moving just 33 percent of all stored data to Digital Tape could decrease worldwide electricity consumption by one percent per year. Designing archival protocols to optimize the use of inert data storage can be encouraged by corporate and government policy.

V. Digital Tape Has Evolved Dramatically

As late as the 1980s, Digital Tape suffered from several technical drawbacks that may today account for its lack of market penetration. During the 1980s and 1990s, tape was in many ways less reliable than disk, while the technologies that allowed many individual HDDs to perform in concert progressed rapidly. For example, even though individual HDDs could be in and of themselves somewhat unreliable, technologies such as RAID (Redundant Arrays of Independent Disks) masked the problem, and provided reliable performance. RAID allowed any single hard disk in a physical group to fail, while other disks in the group rebuilt the lost data. As a technology, HDD systems took on an aura of increased reliability. Simultaneously, Digital Tape suffered from several problems.

1. The areal density (bits per square inch of media) of tape media was low, and many tapes were needed in very large robotic systems to store quantities data.
2. Tape quality was less than perfect, sometimes causing tapes to break.
3. Tape transport mechanisms also caused wear and tear on tapes, increasing the likelihood that they might fail.
4. There was no true international standard for tape formats, so the industry was left with a series of proprietary formats, both in the actual size and construction of tapes, but also in how data on the tapes had to be formatted.
5. And finally tape storage was “block based” rather than file based, meaning an external database was needed to understand what was on any given tape. If the database was lost or corrupted, the tapes were rendered all but useless.

While reliability and standardization were the primary barriers to market penetration, several decades ago energy costs in general were lower, data volumes were still manageable, and disk technology was reliable enough, reducing market demand for Digital Tape. Consequently, competitors in the data storage business that did not sell tape systems promulgated a marketing message that tape was unreliable and high-risk. Sales number for HDDs soared. Digital Tape sales suffered.

In 1998 three industry giants -- IBM, Hewlett Packard, and Quantum Computer Corporation -- took it upon themselves to support a new industry standard called LTO – Linear Tape Open. In the 19 years since the creation of LTO, it has become the world’s leading Digital Tape format. In the same period, the combined technical efforts of the LTO consortium companies corrected all of the technical issues that haunted early tape systems.

- Today, tape density is very high, and continues to double almost every two years. LTO8, due out in early 2018, will store 12 TB of uncompressed data on a single cartridge.

- Tape film and substrates are reliable and continue to improve. They now perform with great tensile strength and minimum friction. Tape failures now are are highly uncommon.
- Tape transport mechanisms, that once created excess wear on tape, are now smooth and frictionless, adding to tape life.
- The LTO consortium all but guarantees that no single vendor can sunset the technology. And consequently the LTO form factor now is the most popular Digital Tape format by far.
- A new tape file system called LTFS – the Linear Tape File System – now allows tape to be file-based rather than block-based, making tape management even easier than HDD management. In June 2016 LTFS was adopted as an industry standard by the ISO (International Standards Organization.) The result is that data can now be held in a non-proprietary format, essential for long-term archives. But the greatest advantage of LTFS is that it places a file system *directly on each tape* -- similar to a USB or Flash drive. This critical innovation makes LTFS formatted tapes "self describing" and less reliant on external databases.

Because of its history of technical problems, Digital Tape is often ignored by corporations and governments. Today both HDD systems and Cloud storage dominate data archiving. Cloud is popular today for many reasons, but it should be noted that Cloud storage does not change the details about energy efficiency. HDDs burn energy whether they are spinning in mid-town Manhattan or in a desert Cloud data center. It should also be noted that some Cloud storage providers do use tape for certain tiers of their storage products.

For industry and government, *an awareness of how much energy is being consumed by data storage is the place to start in evaluating use cases for Digital Tape*. Clearly, converting some percentage of archive data storage to Digital Tape can be an easy target for sustainability officers to help achieve CO_2 emission reductions. In a world of so much "write-once, read-never" data, with many organizations choosing to save some data virtually forever, the energy profile of Digital Tape is essential to achieving sustainability goals.

VI. Conclusion: Maximizing Data Tape Usage Can Help Achieve CO_2 Reduction Targets

Here are the facts:

- HDDs make up 10 times the installed data storage market compared to tape
- HDD systems use 70 times more energy than tape

- Digital Tape is now a reliable and effective medium on which to store archival materials
- Simply transferring just 33 percent of stored data to tape could reduce global electricity consumption by 1 percent per year -- with no additional change in business practices or operations, saving millions of metric tons of CO_2.

If the volume of data in the world were not expanding exponentially with no end in sight, companies might be content with the growing energy efficiency of conventional storage. But the unabated growth of data and the relentless increase of retention windows recommends that even slightly shifting the ratio between HDDs and Digital Tape, on a global basis, could save many hundreds of millions of metric tons of CO_2 emissions year over year. Energy costs money, and the per-TB cost of tape storage is at this moment considerably less than all other types of computer storage, such that this shift could both save substantial carbon emissions *and* save those implementing the change substantial money.

Marketing forces often determine what technology is used for any particular application. Unfortunately, tapes shaky performance at a point of massive storage expansion has had considerable long-term costs. Creating a greater awareness of tape technology as a viable, and sometimes superior alternative to HDDs could have a substantial impact on personal, corporate, and global sustainability goals.

If only 33 percent of data now residing on hard disks could over time be shifted to tape technology, the world would use one percent less electricity.

It is really a matter of choosing the right tool for the job. Unfortunately, market forces have blurred the value of a fundamentally sustainable technology, Digital Tape. The intent should not be to warp natural market forces. But rather, with information about advances in Digital Tape performance and reliability, there should be a proper balance between the utility of HDD systems for their technical features, against the alternate features of Digital Tape, where lower cost and greatly reduced power consumption are legitimate factors in the decision matrix..

As multinational companies seek to achieve the COP21 carbon emissions reduction targets, many are undertaking expensive adjustments within their supply chains to do so. Data storage, a part of every organization's strategic infrastructure, may provide an easy path that will contribute to emission reduction initiatives -- while saving money in the bargain.

Fred Bonner
Chief Technology Officer
EchoLeaf Systems
fbonner@echoleafsystems.com

Peter Guglielmino
Digital Media / Distinguished Engineer
IBM Corporation
peteg@us.ibm.com