LOGNOVATIONS

WHITE PAPER

# More data. Same pipelines. Lower cost.

Controlling transmission and data storage costs as data creation skyrockets.

LOGNOVATIONS

# Summary

## 1. Data volume is growing exponentially.

More data was created in the last two years than in all previously recorded time. With continually rising user and device numbers, digital information will grow at a rate of 29.8% CAGR from 12 ZB in 2015 to 163 ZB in 2025.[1]

## 2. Neither bandwidth nor storage cost reductions are keeping up.

Data is growing exponentially faster than bandwidth. And storage costs are projected to balloon from $139.45 per user in 2015 to $520.21 per user in 2025, even taking into account expected cost decreases. Why?

- Existing compression methods are based on a theorem of limited compressibility.
- Existing compression methods use multiple complex algorithms which require continual maintenance.
- Multiple points of failure in existing compression methods pose significant security risks along with their related financial consequences.

## 3. Until now, that is.

The most cost-effective solution is “low latency” compression. This allows more data through the same pipeline and/or the same data to use less bandwidth, depending on an organization’s needs. The SQZ’s low latency compression algorithm:

- Offers previously-unattainable lossless compression rates that actually improve as file size increases.
- Losslessly compresses all file formats in one single process, with no upper size limit.
- Encrypts as it compresses, exceeding current standards.
- Transmits more data faster through the same pipelines at lower cost.

The unprecedented efficiency of the SQZ Algorithm offers organizations distinct financial advantages now and in the future.

Is it magic? Nope. We just have a better way. Read on.

[1] IDC White Paper: Data Age 2025 Study, April 2017.

LOGNOVATIONS

WHITE PAPER

# More data. Same pipelines. Lower cost.

How to control data storage costs as data creation skyrockets.

## In the last hour, more data was created than in the previous hour.

The global data-sphere is getting bigger and faster every day. In 2025, the world will use over 80 billion connected devices generating 180 trillion GB of data.[2] Increases are projected in all sectors, including larger AI and embedded data file types. This upward trend shows no sign of letting up.

Cisco reports that:[3]

- The percentage of global internet users rises daily, moving from 44% of the world population in 2016 to a projected 58% in 2021.
- The number of sensors collecting information is growing exponentially, from 2.3 devices per capita in 2016 to 3.5 devices per capita in 2021.
- Average speed per device will nearly double in 5 years, from 27.5 Mbps in 2016 to 53.0 Mbps in 2021.
- Data per user will reach 34,680.9 GB per user by 2025, from 3669.7 GB in 2015 - a 25.18% CAGR increase.

As a digital workforce expands around the globe, organizations expect applications that work seamlessly and effortlessly in near real time.

## Existing data storage and transmission methods aren't keeping up with the flood.

Although advances in physical storage density have somewhat reduced storage costs, the rising flood of data creation and collection far exceeds these cost savings. Even with the cost

---

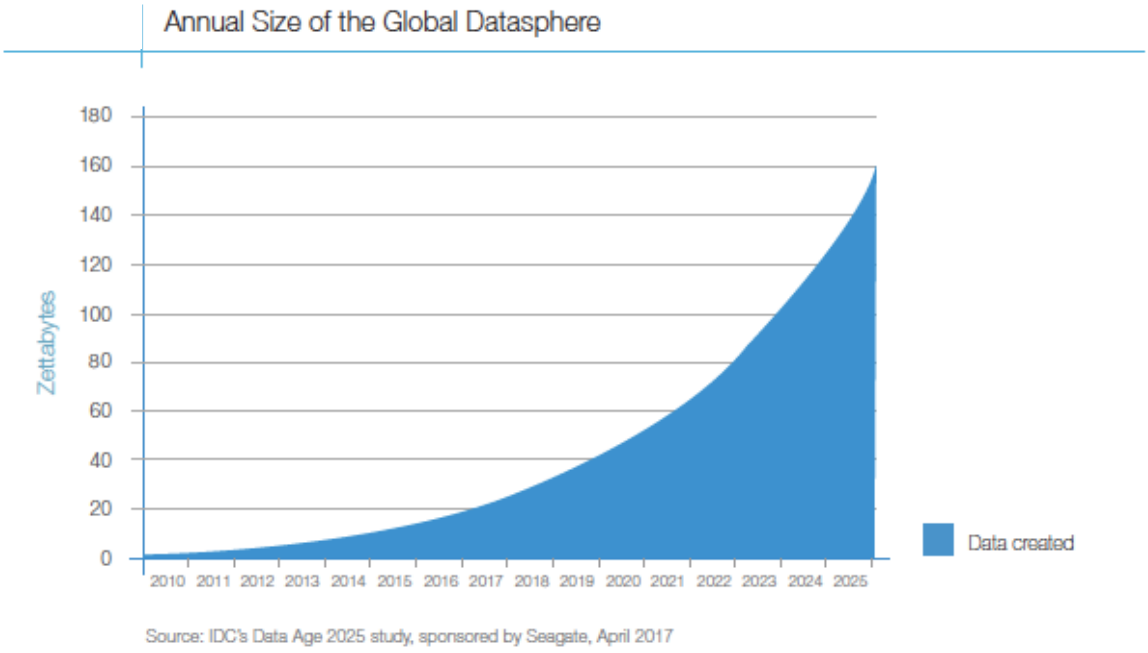
[2] IDC White Paper: Data Age 2025 Study, April 2017.

[3] Cisco. (2017). Cisco Visual Networking Index: Forecast and Methodology: 2016 2021 (Rep.). Cisco.

of storage decreasing 8.9% per year through 2021, the net effect of the higher data growth rate is a 14.1% CAGR increase, from $139.45 per user in 2015 to $520.21 per user in 2025.

The ever expanding data-sphere is forced into the same limited transmission pipelines resulting in deteriorating WAN performance and increased costs.  A corollary to Parkinson's First Law, it seems that the need for mass storage and transmission increases at least twice as fast as storage and transmission capacities improve.

Industry experts see this sea change coming, but they're not sure what to do about it. IDC's Data Age 2025 paper discusses "discardable data" and development of a "discerning data retention policy" rather than the possibilities of a better technology. This mindset presents a major obstacle for businesses and CTOs who will continue to rely on granular data analysis to gain a consistent competitive advantage - no matter how unwieldy the data-sphere gets.




The math is clear: Data is expanding at a much faster rate than storage and transmission costs are decreasing. Businesses are realizing it's time to strap up.

## The problems with multiple-algorithm compression

We now have more data than can affordably be collected, store, and analyze. The primary reason for this challenge comes with the myriad of data collecting devices out there. Different devices collect information in different formats, sizes, resolutions, and bit depths.

To compress these multiple data types optimally, prevailing theory has required the use of multiple different compression algorithms. The problem is that when faced with continually escalating data volumes, this method is proving financially inefficient and increasingly insecure.

### Complex and inefficient

The use of many different compression algorithms creates high storage complexity that's increasingly costly to maintain. Further, the cost of transmitting data is severely constrained by bandwidth. For example, the U.S. Department of Defense currently spends about $1B per year on Com-Sat-Com. If the DoD stays with its current compression method, this expense is expected to increase 11.33% CAGR to $5B per year by 2030.

### Insecure and risky

The numerous points of failure inherent in many algorithms expose organizations to data security breaches at a time when cyber attacks are increasing in frequency and scale. The expected average cost of a breach in 2020 is at least $150M, not including damage to a company's reputation.[4]

## Introducing SQZ

SQZ is a universal lossless compression software that enables seamless, effortless applications and fast data transmission. SQZ accomplishes this by using a content agnostic approach which identifies the location of the input file within an immutable dictionary without altering or encoding the contents of the file. This eliminates the need for multiple complex algorithms for numerous file types.

This proprietary method yields optimal compression rates of any file format and robust data security within a single process.

### SQZ works fast in any pipeline.

The SQZ algorithm compresses files at significantly higher rates in a low latency environment, thus moving more data through the same pipeline at lower costs. This allows far more data through the same pipelines and/or the same data to use less bandwidth, depending on an individual organization's needs.

### SQZ exceeds encryption security standards.

The SQZ algorithm encrypts files as they are being compressed by changing the sort order of the dictionary. This permits only the nodes with the same settings to talk to each other. An attacker attempting to decrypt a 1 MB PDF would need to search every 1 MB file that could exist to find the file.

---

[4] Cybersecurity, Cybint News. (2018, March 16). 12 Alarming Cyber Security Facts and Stats. Retrieved April 14, 2018, from https://www.cybintsolutions.com/cyber-security-facts-stats/

A centerpiece of the SQZ algorithm is security when data is at rest and when it is in transit and at rest. We are able to encrypt files amongst all other possible files of the same size. This means that the encryption is stronger the larger the input file. For example a 16 bit file is encrypted to 2^16, while a 8,000,000 bit (1 megabyte) file is encrypted to 2^8,000,000. This far exceeds the highest current standards available in the market.

## Provides fast, efficient compression that's way ahead of the pack

SQZ compresses and encrypts any file a computer can read or execute by working at the bitstring level.

- Achieves losslessness without storing massive amounts of unique variables.
- Eliminates the need for multiple algorithms to compress different file formats.
- Retain's the full integrity of the each of the original file's bits.

## SQZ's compression rate has no upper limit.

In this example, the SQZ algorithm compresses data losslessly to an output file 0.004% of the original file, as shown in Figure 1. This depicts that the upper bounds of the compression rate

**Figure 1**

| | |
|---|---|
| **Original Size** | 3706306 bytes |
| **Squeezed Size** | 169 bytes |
| **Squeeze Compression** | 21930.80 |
| **Squeeze Time** | 7.109375 seconds |
| **LZW Size** | 759336.00 bytes |
| **LZW Compression** | 4.88 |
| **LZW Time** | 8.72 seconds |

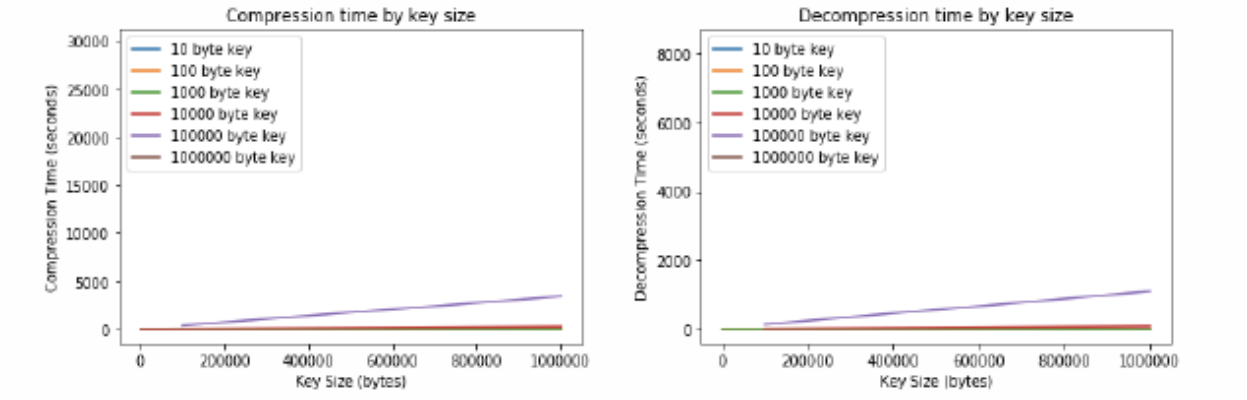
has no limit; SQZ compression improves as file size increases.

The SQZ algorithm, on average, compresses data losslessly to an output file 13% of the input file size compared with a typical Lempel-Ziv lossless file compression of 43%. Dictionary based compression algorithms are the most commonly used lossless file compression algorithms.

## Speed: The bigger the file, the faster SQZ compresses and decompresses.

Owing to the unique features of the SQZ algorithm, the algorithm's Big O notation is linear, as illustrated in Figure 2. The result? This lends itself to being predictably efficient on CPU time.

**Figure 2**




## How SQZ compression compares

The universal nature of the SQZ algorithm is shown in Figure 3 across various file formats. The upper bounds of the compression rates exceeds the nearest alternative algorithms. The closer SQZ gets to a one to one ratio of input file to chunk size the better the compression factor and the quicker the file processes. Calculations provided by The Canterbury Corpus, a

**SQZ File Compression Comparison**

| File Name | Raw Size (Bits) | Algorithm | SQZ Size (Bits) | SQZ Compression Factor | Nearest Alternative | Alternative's Compressed Size (Bits) | Alternative's Compression Factor |
|---|---|---|---|---|---|---|---|
| alice29.txt | 1216712 | SQZ | 1688 | **720.8009** | zpaq | 298888 | **4.0708** |
| asyoulik.txt | 1001432 | SQZ | 1496 | **669.4064** | zpaq | 277968 | **3.6027** |
| cp.html | 196824 | SQZ | 1280 | **153.7688** | zpaq | 53688 | **3.6661** |
| fields.c | 89200 | SQZ | 1272 | **70.1258** | zpaq | 20888 | **4.2704** |
| grammar.lsp | 29768 | SQZ | 896 | **33.2232** | brotli | 8968 | **3.3194** |
| kennedy.xls | 823792 | SQZ | 352 | **2340.3182** | zpaq | 126496 | **6.5124** |
| lcet10.txt | 3414032 | SQZ | 1656 | **2061.6135** | zpaq | 726112 | **4.7018** |
| plrabn12.txt | 3854888 | SQZ | 1800 | **2141.6044** | zpaq | 1021632 | **3.7733** |
| ptt5 | 4105728 | SQZ | 304 | **13505.6842** | zpaq | 220536 | **18.6170** |
| sum | 305920 | SQZ | 288 | **1062.2222** | lzma | 75144 | **4.0711** |
| xargs.1 | 33816 | SQZ | 1024 | **33.0234** | brotli | 11536 | **2.9313** |

Canterbury Corpus SQZ (Intel® Core™ i7|x86_64|2.3 GHz|16 GiB|Apple MacBookPro|OSX 10.13.3|GRPC) vs. Closest Alternative (hoplite|Intel® Core™ i7-2630QM|x86_64|2 GHz|6 GiB|Toshiba Satellite|A660-X|Fedora 22|4.1.4|gcc-5.1.1)

standardized file set used to benchmark compression algorithms.

## Significant cost savings

Data Storage Cost Per GB Benefit

| Year | 2015 | 2020 | 2025 |
|---|---|---|---|
| Information Created Worldwide (ZB) | 12.0 | 47.0 | 163.0 |
| Internet Users Worldwide | 3,270,000,000 | 4,100,000,000 | 4,700,000,000 |
| GB of Data / User | 3,669.7 | 11,463.4 | 34,680.9 |
| Storage Cost / GB | $0.04 | $0.02 | $0.015 |
| GB Storage Cost / User | $139.45 | $229.27 | $520.21 |
| Worldwide Info LogN Compressed (ZB) | 1.56 | 6.11 | 21.19 |
| Data Storage Reduction (ZB) | 10.44 | 40.89 | 141.81 |
| Storage Cost / GB | $0.04 | $0.02 | $0.015 |
| Gross Data Storage Cost Savings | $417.6B | $817.8B | $2,127.2B |
| **Cost Savings / User** | **$127.71** | **$199.46** | **$452.59** |

The Data Cost Per GB Benefit in Figure 4 illustrates the significant financial benefits to organizations leveraging the SQZ algorithm's powerful compression. If the entire data-sphere was to be compressed this way, total cost savings would exceed $2 trillion in 2025. Organizations may extrapolate their savings based on their own data storage projections.

# Conclusion

The expanding global data-sphere and its impending impact on organizations' bottom lines are cause for re-evaluation of current data storage and transmission methods. Existing compression methodology poses financial and security risks along with a conceivable limit to the quantity an organization can ultimately store. The SQZ Algorithm offers significant savings and outstanding efficiency on data storage. Secure, lossless, and simple, SQZ enables you to store more data and transmit it faster through the same pipeline at lower cost.

Contact Chris McElveen or Mike McElveen at 813-999-0869 for more information.

## About LogN

LogN is a SAAS company providing data compression and encryption at the enterprise level for maximally efficient data storage and transmission. We will disrupt the data archival, IoT, and transmission markets through the use of our patent pending compression and encryption algorithm, SQZ

We are in the pre-launch phase with Intellectual Property rights for data compression with significant benefits to file encryption and transmission. LogN has an operational version of the proprietary algorithm that proves the validity of across the board higher compression rates and the robustness of the encryption benefit.