Nexsan 10 Minute

White Paper

Deduplication 2.0: Twice the Efficiency... Half the Cost

Quick Summary

As much as first generation deduplication storage was a step in the right direction, there are so many inefficiencies within Deduplication 1.0 that, in many cases, it is having the OPPOSITE effect on the very problem it was intended to solve – reduction of storage costs.

Enter Deduplication 2.0, the first highly efficient storage solution that addresses all the shortcomings of Deduplication 1.0 to achieve what storage efficiency is supposed to achieve – significantly lower total cost of ownership.

IT professionals are facing one of the most unique challenges in this new era of data expansion and economic discipline. Last year, digital information grew 57%[1] while IT spending grew a meager 2.6%[2].

The next 18 months will see digital information explode two-fold[1] while the economy struggles to make incremental gains. With that, rules have changed. IT professionals have had to move away from 'just adding more storage' to adding 'storage efficiency' to drive down costs.

Many IT professionals find themselves caught against a wall of budgetary constraints, space inhibitors and power/cooling limitations while their data grows exponentially. Recent years have seen storage technologies emerge with the intent to drive down storage inefficiencies and lower total cost of ownership to give businesses a better return on their storage investment.

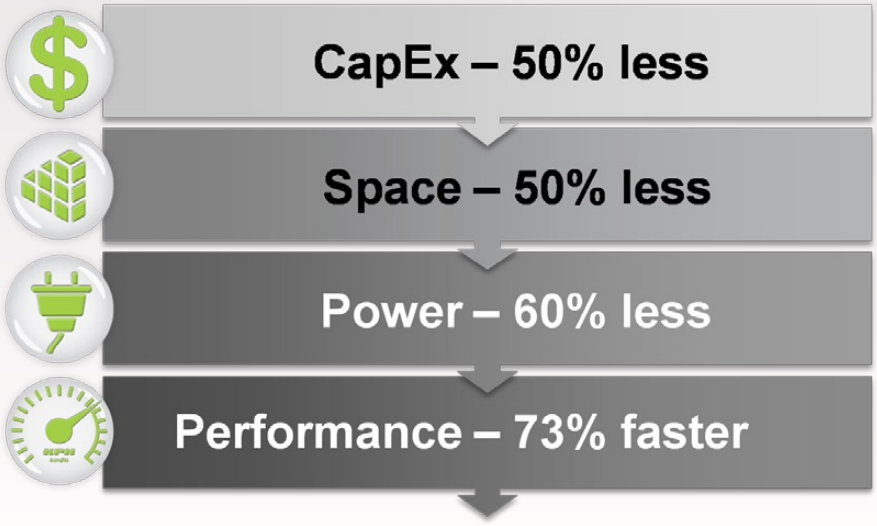
One of those emerging technologies is deduplication which eliminates redundant data that compounds in an organization's backup storage. As much as first generation deduplication storage was a step in the right direction, there are so many inefficiencies within Deduplication 1.0 that, in many cases, it is having the OPPOSITE effect on the very problem it was intended to solve – reduction of storage costs.

> "...the cost is not really being reduced, **it's simply being traded** - trading disk storage costs for server and memory costs."

The irony around Deduplication 1.0 is that rather than help organizations drive down storage costs, it has INCREASED storage costs for many. While organizations may be efficiently storing less data, it is costing them MORE to have less.

IT professionals who own Deduplication 1.0 are having a hard time justifying the return on investment after

Deduplication 2.0 Benefits over Deduplication 1.0



Deduplication 2.0 = Twice the Efficiency...Half the Cost

[1]Source: IDC Digital Universe

[2]Source: IDC Worldwide Black Book

exercising the capital expense to reduce the cost of storage only to find they have to add more costly servers, memory and storage to increase performance to meet the backup window.

The dirty little secret about first generation deduplication storage is that, in many cases, the cost is not really being reduced, **it's simply being traded** - trading disk storage costs for server and memory costs.

Moreover, first generation deduplication storage arrays struggle from the same old inefficiencies that have plagued storage arrays for years when it comes to power and space. They lack high density to save space and inexplicably consume FULL power even when idle. Organizations that need a more holistic solution to reduce storage costs and increase efficiencies have been left out in the cold.

The bottom line behind storage efficiency is one thing – it's about saving. Highly efficient deduplication systems should save more than just storage. They should:

- Save Money
- Save Power
- Save Space
- Save Time

Enter Deduplication 2.0, the first highly efficient storage solution that addresses all the shortcomings of Deduplication 1.0 to achieve what storage efficiency is supposed to achieve – lower total cost of ownership.

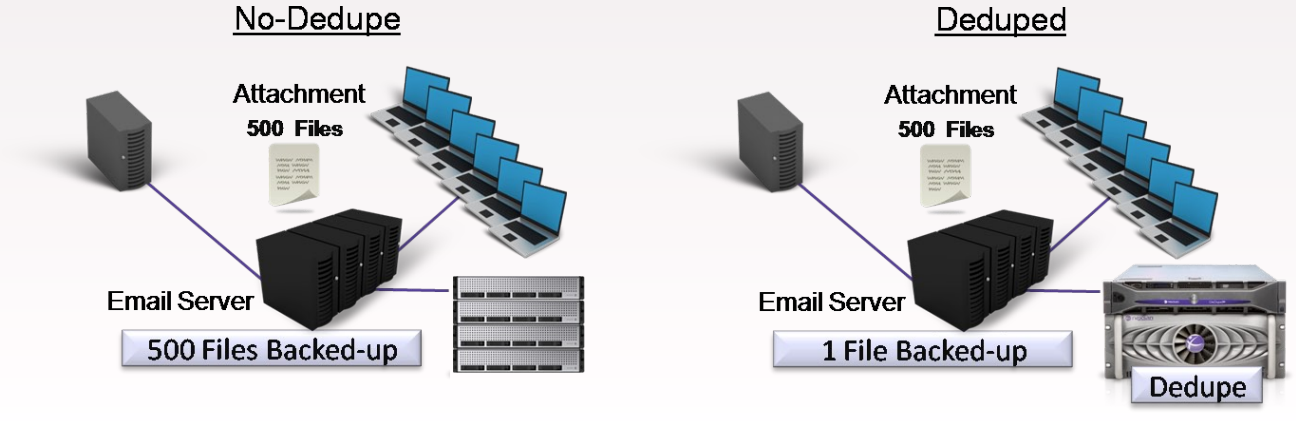
Whereas first generation deduplication was supposed to reduce expense by eliminating duplicate data, second-generation deduplication also eliminates duplicate data but doesn't stop there. Deduplication 2.0 uniquely and

What is Data Deduplication?

Data Deduplication is:

- A method for eliminating redundant data from the backup process
- It works by saving a single copy of data and replaces any further identical instances with pointers to the master record

Data deduplication backs up just one instance of the attachment's data and replaces the other 499 instances with pointers back to that copy.



significantly reduces operating expense, risk, power and space while delivering greater performance, reliability and simplicity.

GEN 1 vs. GEN 2 - POWER & SPACE

While Gen 1 does save energy from the reduction of storage devices necessary to store backups, the underlying storage architecture has no energy savings whatsoever. Even between backup windows when the system is idle, Gen 1 continues to spin at full power.

In addition, since first generation deduplication is dependent on certain performance levels to meet backup windows, multiple deduplication systems have to be added to match data growth. Inherent to this expensive model is the fact that every Gen 1 system means more power to operate and cool along with more space to house the additional servers and storage.

In sharp contrast, Gen 2 is highly efficient when it comes to power and space. When comparing Nexsan's DeDupe SG™ storage system against a typical competitor, the underlying storage infrastructure consumes 60% less power and cooling while delivering 50% less floor space over Gen 1 deduplication storage systems.

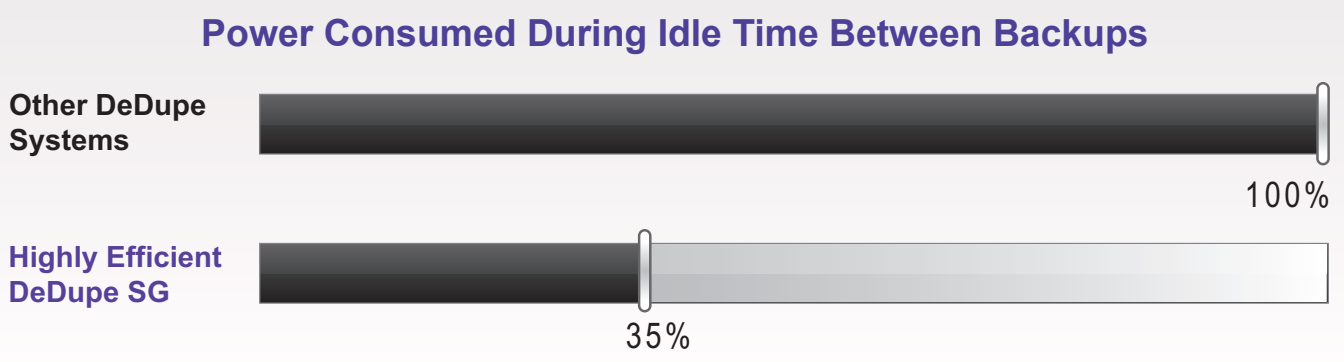
As a result, Nexsan's power and storage density ratios are unparalleled. At $.12 a kilowatt-hour, the energy cost for deduplicated storage on a DeDupe SG 52TB system (supporting 1PB of data) is 1/3 the cost of a typical competitor (see "The Power/Density Paradox" white paper for more information). The energy saved is the equivalent of powering 10 homes for a year and reduces the amount of CO_2 released into the atmosphere by 42 metric tons over 3 years.

GEN 1 vs. GEN 2 - BACKUP

Whereas Gen 1 and Gen 2 both remove duplicate data from backup storage, the Gen 1 approach to backup is far more expensive, slow and wasteful. Deduplication 1.0 requires a new server to process backup jobs, ingest backup data and then use very CPU-intensive methods to deduplicate data.

This "in-line" approach is typically 73% slower on average than Deduplication 2.0. Because it is so slow, more than one appliance is typically needed to get backups done inside the backup window. With ever-expanding data growth, there is an ever-expanding need with Gen 1 systems to add more costly and power hungry servers to achieve the necessary throughput to dedupe more data in the same backup window.

Power Consumption Comparison



Save Energy

Dedupe storage is primarily a backup target used only during the backup window. The typical backup window is 2-8 hours a day, yet all other dedupe solutions spin disks at full speed for the remaining 16-22 hours.

Would you leave your car running at full speed in the driveway when not using it? Probably not. So why would you keep your storage running at full speed between backup windows? Not only is it wasting power to operate but wasting power to cool.

By leveraging the green efficiencies inherent to Nexsan storage components and energy saving AutoMAID™ technology, DeDupe SG users not only get better performance during their backups, they save energy and reduce costs every hour the system is idle.

AutoMAID operates transparently in the background and self-optimizes energy efficiency by looking for opportunities to save energy. By uniquely delivering speed with green, DeDupe SG will gradually slow down and ultimately turn off its drives to conserve power while providing near-instantaneous access to data when needed.

What this means to users is that DeDupe SG can reduce energy costs as much as 95% when compared to normal disk storage and 60% when compared to other dedupe solutions.

Only Gen 2 deduplication storage systems use power saving technology

When it comes to backup, the problem with Deduplication 1.0 has to do with the conflict of deduping growing amounts of data within a static backup window. Since the size of the backup window remains static, the only way Deduplication 1.0 can get more data in that window is to add more servers for additional inline processing performance which ultimately translates to added expense to buy, support, power and cool.

By contrast, DeDupe SG requires no new computing capacity other than a single appliance used for deduplication, which is done after the business-as-usual backup completes. DeDupe SG is much faster and far more cost efficient than the inline server-based approach because it is not constricted by an in-line process. With Gen 2, the existing backup servers move data as usual at full backup performance. After a backup completes (or by using an "on-the-fly" technique while a backup is being captured), a separate, single server deduplicates backup data with up to 73% better performance than first generation deduplication storage systems.

With this next generation approach to deduplication, the conflict between deduping more data within a static backup window is resolved and the total cost of ownership is greatly reduced. The Gen 1 model requires SIGNIFICANTLY more servers and memory to keep pace with data growth.

GEN 1 vs. GEN 2 - RESTORE

With Gen 1, when restoring from deduplicated data, there are no performance accelerators. In fact, there are bottlenecks. When restoring, the restore volume must be "re-inflated" to restore all the duplicate copies. Gen 1 chugs away at this task blindly requesting data, re-inflating and restoring the original data.

All data in the entire data volume, including duplicate data, must be written back as part of the restore. It is important to note that Gen 1 is in the data path. For that reason, it can slow down more than backups or restorations; it can also slow down other applications that are trying to use the same SAN ports.

Aside from being slow to restore, Gen 1 is hindered by a very

serious business risk during restore operations. If a restore is ever required while a backup is running, the problem is that the server is tied up with very CPU intensive backup operations.

Tasking that server to restore deduplicated data requires equally intensive CPU operations and adds substantial processing load. In this case, something has to give. The net result is a backup job and restore job that can slow to the point of panic. Hunting for the correct backup job to kill in the middle of a cascading corruption failure can have significant impact to the business and is cause for unacceptable risk.

The same unparalleled performance that Deduplication 2.0 delivers to backup, it also delivers to restoring data with a see-through reference to data and advanced performance accelerators to enhance restore performance. From the restore application's point of view, it requests a block of data to restore and receives it without the need for re-inflation.

DeDupe SG simply delivers the correct blocks. Since data restore is sequential in nature, Gen 2 optimizes restore operations by anticipating and pre-fetching the next block instead of waiting for the request, which otherwise adds latency and slows down overall performance. Over the course of a restore, this next generation, pre-fetch intelligence saves significant amounts of time. Since Gen 2 deduplication storage is not in the data path, unlike Gen 1, and because Gen 2 anticipates and pre-fetches data, restore performance is greatly improved over Gen 1. Additionally, restore operations do not impact other applications.

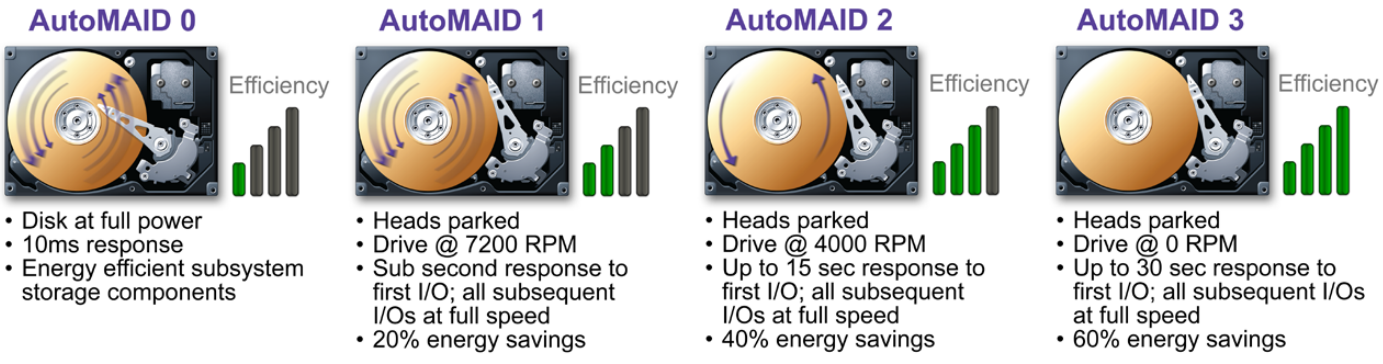
It is also important to note that Gen 2 avoids the risk exposure in Gen 1 by storing a backup in its disk cache before deduplicating. Therefore, if there is a need to restore data during a backup, it's read out of a disk cache to deliver a restore process without performance penalty.

GEN 1 vs. GEN 2 - SIMPLICITY

Based on performance limitations resulting in a greater number of appliances necessary to deliver the performance required to meet backup windows, Gen 1 management is more complex. Not only does Gen 2 have few components for easier management, the rate of expansion with Gen 2 is about half as much in response to data growth.

A single Nexsan DeDupe SG system can scale to protect a petabyte of storage capacity, thereby greatly reducing the likelihood of needing two systems. With DeDupe SG, users have the benefit of managing fewer components combined with a simple interface to save administrative resources.

AutoMAID™ - Speed with Green



Gen 1 creates even MORE cost and infrastructure complexities without a common or global deduplication scheme which ultimately duplicates cost with additional capital expense for more servers, memory and storage.

Two or more Gen 1 systems within the same IT infrastructure will share the same duplicate data between them because they can't coordinate deduplication across system boundaries. Since the repository for Gen 2 is larger on a system-to-system comparison, there is more space to have a single deduplication repository. This further increases the efficiency of Deduplication 2.0 by eliminating additional administrative overhead.

DeDupe SG goes beyond deduplicating data and delivers holistic efficiencies to cut costs, slash power consumption, save time and free space.

CONCLUSIONS

Nexsan and FalconStor recognized the glaring inefficiencies of first generation deduplication and teamed up to deliver the kind of highly efficient deduplication storage system needed by IT professionals in a new era of 21st century challenges.

To deliver a total cost of ownership that revolutionizes deduplication storage, Nexsan and FalconStor have taken a more holistic approach to storage efficiency to redefine what users expect from a deduplication storage system.

As industry trends for data growth and budget reductions continue, IT needs to do more with less. Whereas Deduplication 1.0 was a solution designed to reduce expense by eliminating duplicate data, DeDupe SG goes beyond deduplicating data and delivers holistic efficiencies to cut costs, slash power consumption, save time and free space.

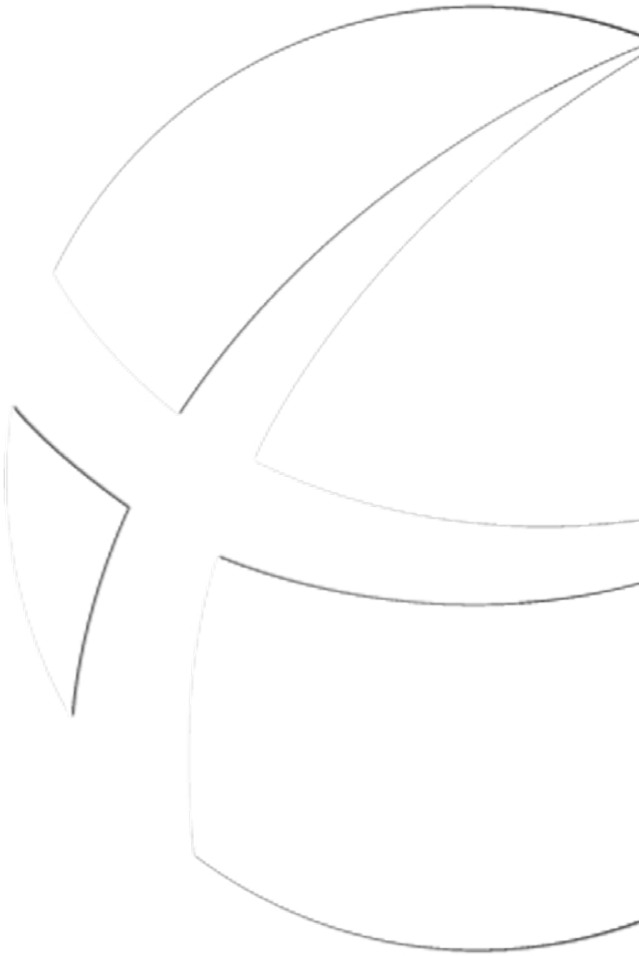
In doing so, DeDupe SG surpasses wasteful, slow and expensive approaches from the past to deliver unparalleled value by every measure. To learn more about how DeDupe SG can benefit your organization, please call 866.4.NEXSAN or email sales@nexsan.com. ❑

Presented by Nexsan Technologies

About Nexsan

Nexsan Corporation is a leading provider of energy-efficient, long-term storage systems. Nexsan delivers secure storage appliances and modular, capacity-optimized disk-storage systems for a broad range of applications including fixed content storage and archiving, email, medical imaging, compliance and litigation support, disk-based backup, digital video security, and rich media.

Nexsan's solutions are the choice of small and medium-sized companies as well as large global enterprises and major governmental agencies around the world who are seeking cost-correct, high density storage solutions. Founded in 1999 and based in Thousand, Oaks, Calif., Nexsan sells its products exclusively through a select global network of VARs, OEMs and system integrators. For more information, please see the company's website at www.nexsan.com.



nexsan TECHNOLOGIES

555 St. Charles Drive • Suite 202 • Thousand Oaks, CA 91360
Toll-free: (866) 4NEXSAN • Main Line: (805) 418-2700 • Fax: (805) 418-2799
European Office: +44 (0) 1332 291 600 • www.nexsan.com

 N10WP0904