

Lab Benchmark Testing Report

Joint Solution: Syncsort Backup Express (BEX) and NetApp Deduplication

Comparative Data Reduction Tests



www.scasicomp.com



www.syncsort.com



www.netapp.com

Lab Benchmark Testing Report

Joint Solution - Syncsort Backup Express (BEX) and NetApp Deduplication

Comparative Data Reduction Tests

CONTENTS

I.	EXECUTIVE SUMMARY	1
	A. Preview of Results	
II.	BACKGROUND	2
III.	SCASICOMP (INDEPENDENT OBSERVER)	4
IV.	LAB TEST: SCENARIOS AND CONFIGURATION	4
	A. Source Data Set Generation	
	B. Testing Scenarios	
	C. Configuration	
V.	LAB TEST: OBSERVATIONS	7
	A. Test Scenario 1 – Syncsort/NetApp	
	B. Test Scenario 2 – Client-Based (Source Side) Data Protection Product	
	C. Test Scenario 3 – File Level Backups with VTL Deduplication	
VI.	LAB TEST: CONCLUSIONS	11
VII.	THE FEATURED SOLUTION	12
	A. Syncsort / NetApp Joint Solution	
	B. NetApp	
	C. Syncsort	

Lab Benchmark Testing Report

Joint Solution - Syncsort Backup Express (BEX) and NetApp Deduplication

Comparative Data Reduction Tests

I. EXECUTIVE SUMMARY

According to a February 2009 User Survey Analysis by industry analyst Gartner, data deduplication implementations are picking up momentum. Gartner estimates that there are more than 32,000 target-side and client-side production deployments of deduplication. Twenty-six percent of organizations are considering implementing some form of data deduplication in the next twelve months.

The goal of this benchmark lab report is to objectively identify an optimal secondary storage deduplication or data reduction solution for a typical customer environment. The lab tests include comparisons across scenarios and products. They measure secondary storage requirements for the backup of a realistic data set that evolves over time in a Windows 2003 environment. Other values including backup speed, client and network resource utilization, and recovery efficiency are observed during the testing.

This study looks at three types of data reduction solutions. One uses a traditional backup application with a "target side" appliance. Another uses a newer type of deduplication technology that deduplicates the data at the source side (application server) before sending it to the target (destination server). The third is a hybrid solution that utilizes an efficient source side data reduction technique in combination with a target side deduplication product.

The lab tests indicate that this hybrid solution (Syncsort BEX + NetApp FAS) not only delivers a comparable deduplication ratio (comparable amount of storage required on the backup system) but also achieves the best results for most other key parameters.

A. Preview of Results

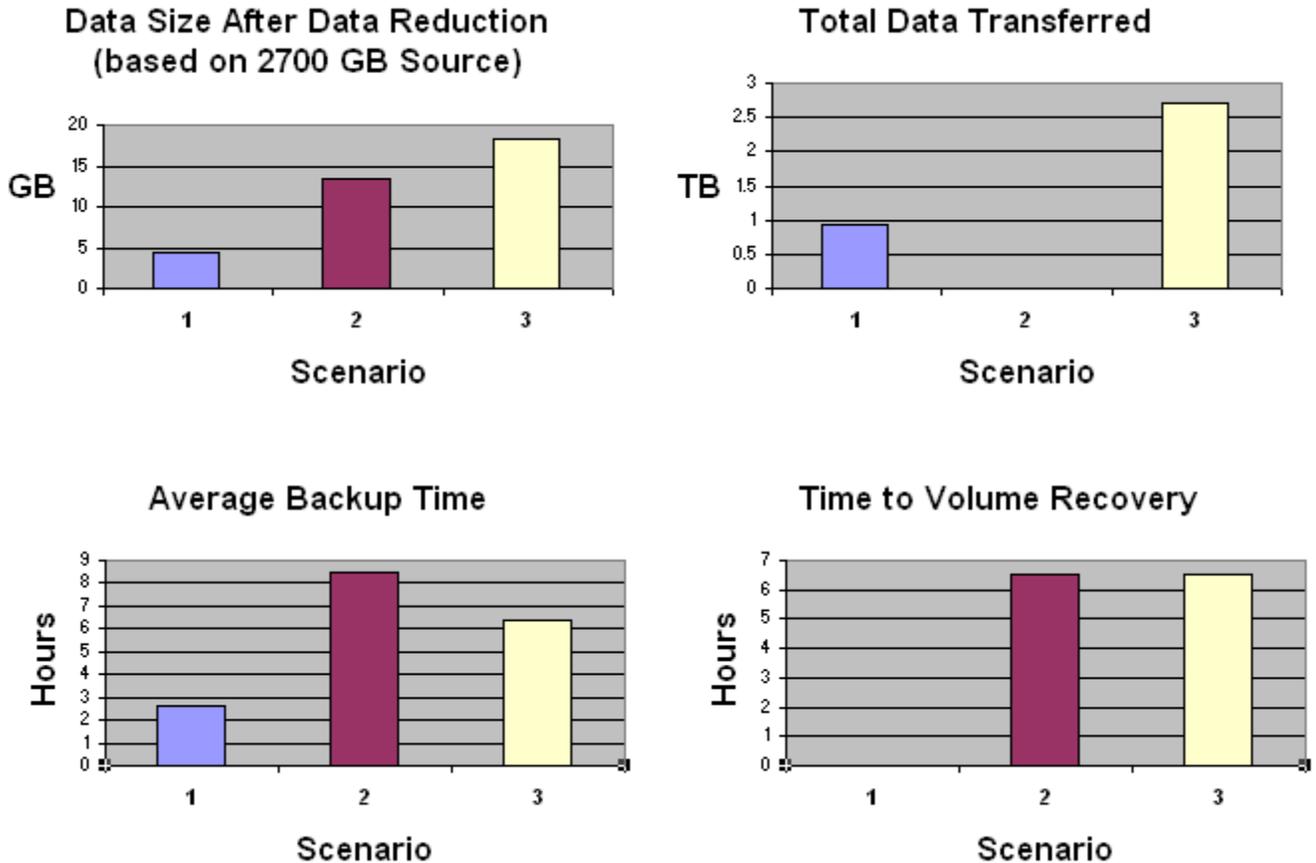
Three data reduction solutions were compared in the lab tests. Briefly, they were:

TEST SCENARIO 1. The first solution, the hybrid solution, used Syncsort's Backup Express (BEX) data protection software and NetApp's FAS technology including NetApp Deduplication. Syncsort Backup Express uses a data reduction technique that transfers only blocks that have changes since the last backup. NetApp Deduplication performs target-based deduplication.

TEST SCENARIO 2. The second solution used a top-tier data protection product that performs client-based (source side) deduplication.

TEST SCENARIO 3. The third solution used a top-tier backup product that performs file-based backups combined with a leading Virtual Tape Library (VTL) product that performs target-based deduplication.

The graphs below illustrate the results for four parameters measured:



For three of the four measurement parameters, the Syncsort/NetApp solution (Scenario 1) indicated the most favorable result.

II. BACKGROUND

Across all industries and sectors, the last two decades have witnessed tremendous escalation of data volumes and mushrooming demand for data and application availability. The need for **Data Reduction** arose when demand for rapid recovery of data exceeded that which could be achieved using backup-to-tape paradigms. **Backup-to-disk** paradigms became critical for meeting SLAs. The cost of disk, being significantly greater than tape, begged for solutions that would reduce the storage needs of the backups.

However, the goal of reducing disk storage needs could not be studied in isolation. Fast, successful backups and rapid recovery with multiple recovery points were also required to meet organizations' ever-increasing availability demands.

The ideal **Data Reduction** model has many goals. These include:

- Reducing the amount of disk space needed to store data,
- Decreasing the amount of data transferred,
- Speeding data transfer and enabling more frequent backups,
- Increasing backup success rates,
- Minimizing CPU usage and network contention,
- Optimizing recovery time.

Achievement of these goals ultimately results in the ability to meet SLAs, improved positioning for future growth, and cost savings. In the current data protection marketplace, both hardware and software vendors are attempting to satisfy some of these goals. Few have succeeded in satisfying all of these goals.

The first goal, reducing disk space needed to store secondary data, can be achieved through a process called **deduplication**. Deduplication achieves data reduction by eliminating redundant information, saving only one copy. In the context of disk storage, deduplication refers to any algorithm that searches for duplicate data objects and discards those duplicates. When duplicate data is detected, it is not retained, but instead a data "pointer" is modified so that the storage system references an exact copy of the data objects already stored on disk.

The next five goals listed above are equally important. These can be achieved with data protection software that handles data reduction by sending only changed blocks, rather than changed files, to secondary storage – resulting in extremely small backup footprints. In the textbook case, when capturing changed blocks on the physical or virtual client, the backup process bypasses the file system resulting in minimal impact on backup clients, servers and applications. Because of the small amount of backup data transferred and the fact that the backups have no impact on other processes on the network, high backup success rates are achieved. Further, the small backups are ideally reconstructed on the destination storage as full backup images – without using additional space – eliminating the need to aggregate backups for point-in-time restores.

III. SCASICOMP (INDEPENDENT OBSERVER)

Created in 1994, Scasicomp is one of the leading storage, backup and archiving solution integrators and developers in the French market. Scasicomp was selected to conduct and author this study and document the results. Their qualifications include their unique and specific competencies as independent consultants, integrators and architects of storage and backup solutions, as mutual channel partner, and as data reduction experts.

Scasicomp's systems engineers defined and configured the lab environment for this series of benchmark tests at the Scasicomp headquarters demo centre in Boulogne-Billancourt (Paris) and have run the tests for several scenarios. Scasicomp selected market leading backup products for their comparisons of data reduction solutions. By the end of the test, Scasicomp identified and demonstrated a solution that delivers the optimal method for achieving minimal amount of storage when protecting a customer environment, without sacrificing other critical criteria.

IV. LAB TEST: SCENARIOS AND CONFIGURATION

A. Source Data Set Generation

The benchmark tests described herein were designed to compare several data reduction solutions. To ensure an accurate comparison of all the solutions, data growth was simulated over a period of time, assuring that all solutions tested were presented with the same source data. A data generator was used to simulate five days of data growth. After the first (base) backup, 50% of the data set was modified. For each of the subsequent iterations, 15% of the data set was modified. In order to redo the tests with different backup software but with the same source data, for each iteration a snapshot was taken using NetApp's "Snapshot" and "Flexclone" capabilities.

Data was constructed to give maximum benefit to the traditional source or target deduplication technologies. To do so, almost all blocks were identical in the base backup, so these technologies would have the 'best' case for the comparison. In addition, in order to also maximize the benefits of the traditional source side deduplication technologies, some of the backups after the base were constructed to have data that contained blocks that were 100% duplicates, thus giving the traditional source side technology the advantage of transferring 0 bytes to the target storage.

B. Testing Scenarios

Three groups of tests were run representing three data reduction solutions.

TEST SCENARIO 1. The first solution tested used Syncsort's Backup Express data protection software and NetApp's FAS technology including FAS Deduplication. Syncsort Backup Express uses a data reduction technique that transfers only blocks that have changes since the

Lab Benchmark Testing Report

last backup. NetApp Deduplication performs target-based deduplication. For additional information about this solution, see Section VII of this paper.

TEST SCENARIO 2. The second solution tested used a top-tier data protection product that performs client-based (source side) deduplication.

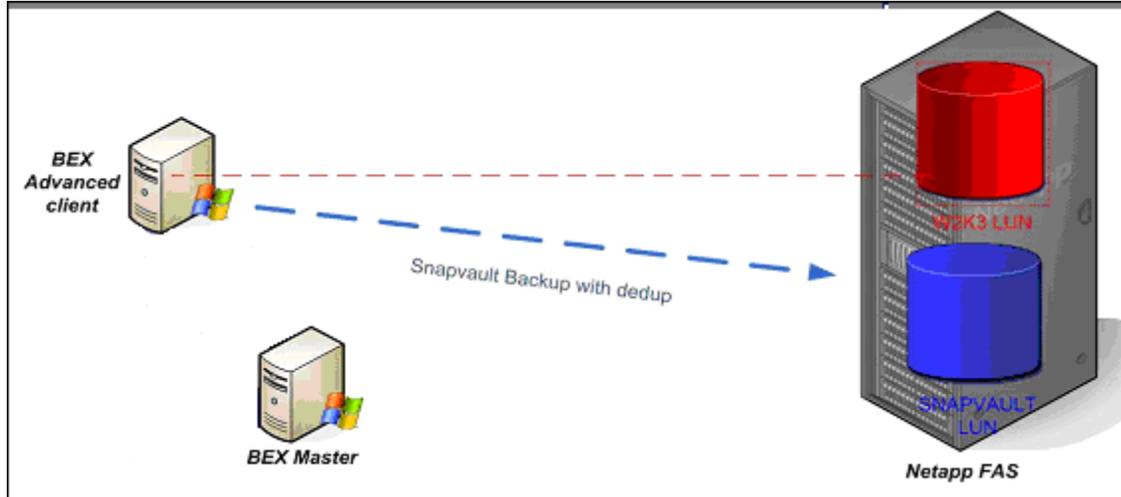
TEST SCENARIO 3. The third solution tested used a top-tier backup product that performs file-level backups combined with a leading Virtual Tape Library (VTL) product that performs target-based deduplication.

C. Configuration

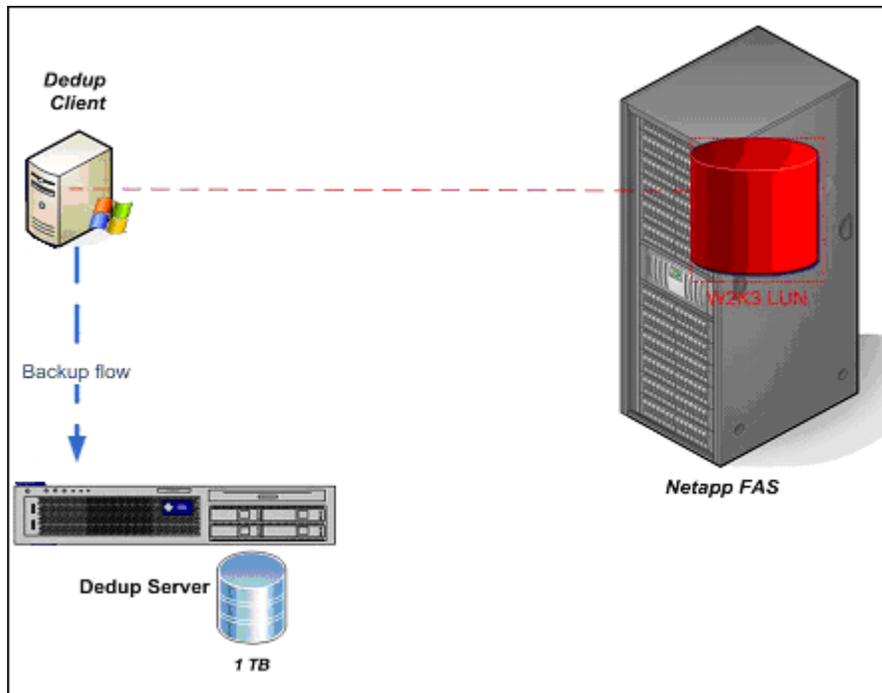
Lab Test Configuration	Scenario 1	Scenario 2	Scenario 3
Source Server	Windows 2003 Server connected by iSCSI to NetApp FAS 3020 Volume 1.75 TB		
Destination Server	NetApp FAS 3020 Volume with NetApp Deduplication 1.32 TB	Sun Fire x2270 Server (2.66 GHz)	Top-tier backup appliance with VTL interface and embedded deduplication technology, attached to 3 tape drives (HP LTO4) with 20 tape cartridges (400GB)
Backup/Recovery Software	Syncsort Backup Express	Top-tier backup & recovery software with source-based deduplication, which identifies redundant data at the backup source	Top-tier file level backup/restore product

The following illustrate the configurations of the three scenarios:

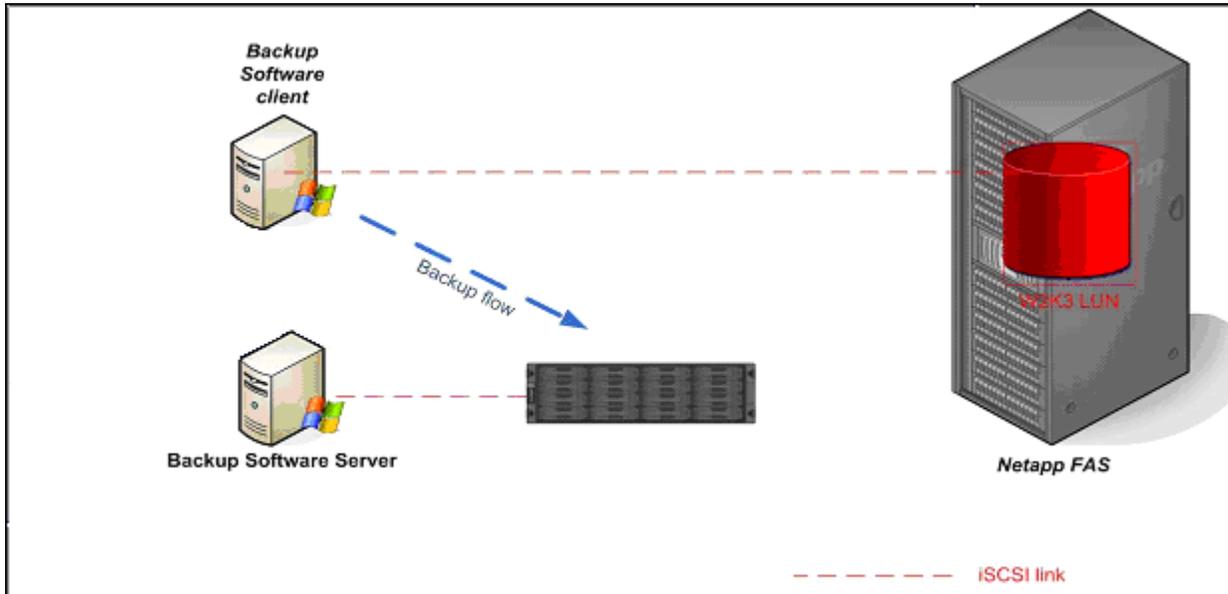
SCENARIO 1:



SCENARIO 2:



SCENARIO 3:



V. LAB TEST: OBSERVATIONS

The tests were performed, observed, and documented by Scasicomp, with the following fundamental axioms:

- For each different test scenario, the same data set was used. This was stored on a LUN attached to the backup client.
- The base data set and the five subsequent iterations were generated with the same tool.
- The base data set was about 417 GB. Each subsequent iteration has a size of approximately 457 GB.
- Effective Data Reduction Ratios are obtained by using the formula:

$$1 - \frac{\text{Data Size After Data Reduction}}{\text{Data on Source}}$$

A. Test Scenario 1 – Syncsort/NetApp

Scenario 1 Backup Instance	Data on Source (GB)	Data Trans- ferred (GB)	Elapsed Time HH:MM	Data Size After Data Reduction (GB)	Effective Data Reduction Ratio
Base	417	419	5:15	0.044	99.99%
Backup 2	456	223	4:43	3.226	99.29%
Backup 3	456	70	1:35	0.657	99.86%
Backup 4	456	70	1:37	0.152	99.97%
Backup 5	457	70	0:32	0.152	99.97%
Backup 6	457	70	2:05	0.151	99.97%
Totals		922	15:47:00	4.381	99.84%
Time to Volume Recovery			0:02		

Observations for Scenario 1:

- The first backup is a base backup. The whole volume is backed up to the NetApp FAS.
- Deduplication provided by the NetApp FAS allowed this solution to write only 0.04 GB against the 419 GB backed up for the base backup. Apart from the second backup which backed up 223 GB, all additional backups transferred 70 GB each against a source volume of 457 GB. This is because BEX moves only the actual block changes made to the data, not the files containing the changes. Each backup was further reduced by NetApp Deduplication.
- After six simulated backups including the base, the effective data reduction ratio was > 99.8%.
- Test backups took about 5 hours each for the base backup and the second backup, and then about 2 hours for the subsequent backups.
- Recovery took two minutes using the BEX Instant Availability feature. BEX Instant Availability provides rapid access to data and applications stored on SnapVault secondary systems. Instant Availability eliminates the need for data transfer and provides immediate access to any backup data on the destination host. Once Instant Availability is activated, destination storage volumes appear as local, fully accessible (read/write) drives. Files on the secondary storage disks are transparently available to users, essentially turning secondary storage (i.e. the backed up data) into primary storage. Note that volume recovery provided by BEX Instant Availability is temporary. Restore of the entire server will take longer and can be accomplished in the background with NetApp’s FlexClone capabilities.

B. Test Scenario 2 – Client-Based (Source Side) Data Protection Product

Scenario 2 Backup Instance	Data on Source (GB)	Data Trans- ferred (GB)	Elapsed Time HH:MM	Data Size After Data Reduction (GB)	Effective Data Reduction Ratio
Base	417	0.7	8:15	0.732	99.82%
Backup 2	456	10.7	8:30	10.668	97.66%
Backup 3	456	1.8	9:00	1.8	99.61%
Backup 4	456	0.2	8:22	0.2	99.96%
Backup 5	457	0.0	8:24	0	100.00%
Backup 6	457	0.0	8:31	0	100.00%
Totals		13.4	51:02:00	13.4	99.50%
Time to Volume Recovery			6:30		

Observations for Scenario 2:

- On the full base backup of the 417 GB volume, 0.7 GB was transferred and 0.7 GB was stored. A high data reduction ratio was obtained (> 99.8%).
- The largest backup occurred after Backup 2, with 10GB being transferred. For backups 5 and 6, Scenario 2 software did not transfer any data as it considered all data to be duplicates.
- Each test backup took longer than 8h20m.
- Scenario 2 restored the original volume in 6h30m.

C. Test Scenario 3 – File Level Backups with VTL Deduplication

Scenario 3 Backup Instance	Data on Source (GB)	Data Trans- ferred (GB)	Elapsed Time HH:MM	Data Size After Data Reduction (GB)	Effective Data Reduction Ratio
Base	417	417	5:57	1.3	99.69%
Backup 2	456	456	6:34	11.6	97.46%
Backup 3	456	456	6:30	1.49	99.67%
Backup 4	456	456	6:30	1.31	99.71%
Backup 5	457	457	6:29	1.31	99.71%
Backup 6	457	457	6:31	1.3	99.72%
Totals		2699	38:31:00	18.31	99.32%
Time to Volume Recovery			6:30		

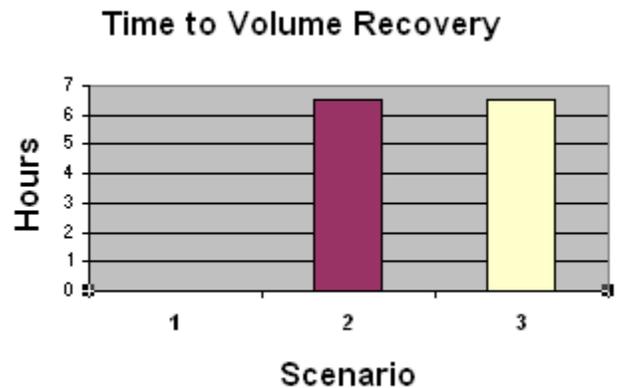
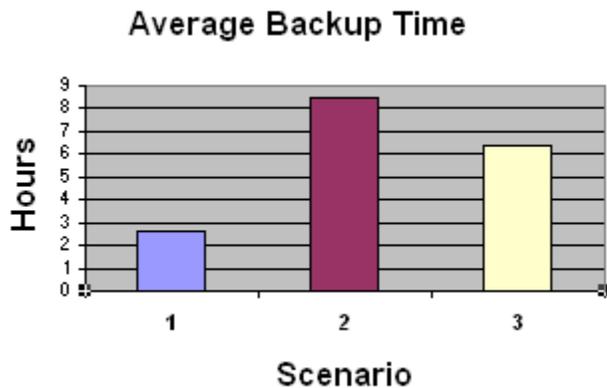
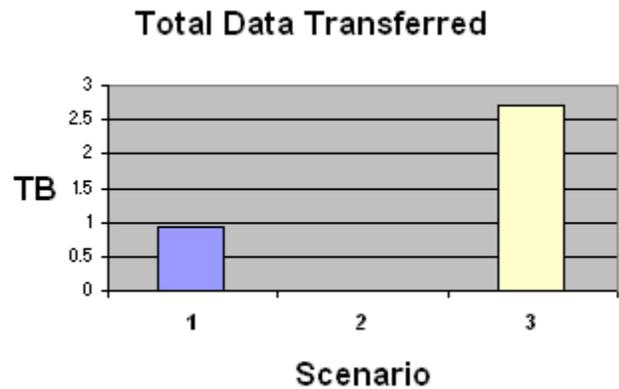
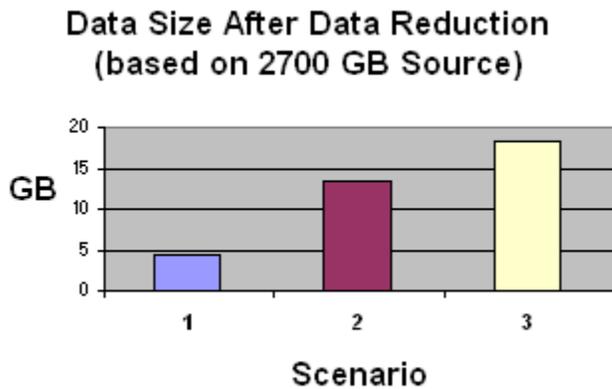
Observations for Scenario 3:

- While performing the test, Scenario 3’s data protection software, which acts at the file level, observed that every file changed. Therefore, a full backup was performed not only for the base but for each additional backup.
- The VTL replaces files with deduplicated files of a smaller size. For instance, the base backup of 417 GB is only 1.3 GB on the VTL.
- The effective data reduction ratio after all backups was > 99.3%.
- The total transferred data is equal to the sum of the base plus all additional backups, 2.7 TB.
- Against a logical size of 2.7 TB, only 18 GB were written on the VTL.
- Scenario 3 backup tests took an average of 6 hours each.
- Restoration of the whole volume took 6h30m.

VI. LAB TEST: CONCLUSIONS

- All three scenarios achieved effective data reduction ratios of > 99%.
- After 6 backups, Scenario 1 (Syncsort/NetApp) had less data written on the disk destination.
- Scenario 1 (Syncsort/NetApp) had the fastest backup time.
- Recovery time was significantly faster using Scenario 1 (Syncsort/NetApp).

The graphs below illustrate the results for four parameters measured:



VII. THE FEATURED SOLUTION

A. Syncsort / NetApp Joint Solution

Syncsort Incorporated began its partnership with NetApp in 1999. The combined Syncsort/NetApp solution provides powerful technologies for assuring the fastest, most efficient, and most complete data reduction available.

The coupling of Syncsort's Backup Express with NetApp's SnapVault, FAS and Deduplication technologies produces a seamlessly integrated data protection and recovery solution for physical or virtual environments with optimized data reduction.

Data centers deploying this solution achieve its benefits by efficiently protecting many servers – Windows, UNIX Solaris, Linux, or combinations – on a frequent basis to a single FAS volume. This is possible because BEX Zero-Impact Backups are extremely small, fast, and efficient, and highly successful. Because all the small backup images are sent to a single FAS Secondary volume, NetApp Deduplication is easily applied, assuring that any two identical blocks – even if they came from different source servers – are optimally deduplicated. Yet recovery times do not suffer; even Instant Availability remains on hand as a recovery option.

The advantage NetApp provides is the ability to deduplicate, at the block level, across multiple source servers. NetApp's SnapShot and NetApp Deduplication, which was introduced with Data OnTap 7G, are critical components to this solution. The advantages provided by Syncsort include the efficiencies garnered by minimizing time, bandwidth, CPU impact, and backup footprint, which are the explicit result of BEX Zero-Impact Backups.

The combined solution provides powerful technologies for assuring the fastest, most efficient, and most complete data reduction available.

B. NetApp

Data deduplication is an important new technology that is quickly being embraced by users as they struggle with issues of data proliferation. By eliminating redundant data objects, an immediate benefit is obtained through space efficiencies. When choosing a deduplication product, however, it is important to consider all aspects of design, including space savings efficiency, performance overhead, and resiliency against failure.

NetApp delivers the following deduplication benefits:

- NetApp deduplication operates with a high degree of granularity. Newly stored data is divided into small blocks. Each block of data has a digital "signature," which is compared to all other signatures in the volume. If an exact block match exists on the disk volume, the duplicate block is discarded and its disk space is reclaimed.

- NetApp deduplication is tightly integrated with Data ONTAP® software and the WAFL® file system. Because of this, deduplication is performed with extreme efficiency. Complex hashing algorithms and look-up tables are not required. Instead, NetApp deduplication is able to leverage existing Data ONTAP internal characteristics to create and search digital fingerprints, redirect data pointers, and free up redundant data areas—all with a minimal amount of user performance impact.
- Another key advantage of NetApp deduplication’s integration with Data ONTAP is the ability to utilize the error checking and recovery procedures that are inherent to Data ONTAP. This includes recovery from power failures, file inconsistencies, and file-system corruption.

C. Syncsort

Syncsort Backup Express (BEX) is a highly scalable data protection and recovery solution for SMBs, SMEs, and large enterprises. The solution reduces storage, bandwidth and CPU requirements, while enabling server and site recovery in just minutes.

The unique capabilities of the BEX Advanced Recovery disk-to-disk solution include bare metal recovery in minutes; archiving to tape; near instant access to Exchange, Oracle, SQL, and SharePoint images; and easy recovery to multiple points in time.

With a single console, BEX protects Windows, UNIX, and Linux networks, and behaves the same in physical and virtual environments. By providing a broad range of data protection needs, BEX reduces pain points associated with multiple products. For scalability, BEX provides LAN, SAN, WAN, and NDMP connectivity support, a flexible multi-tiered architecture, and a single catalog that tracks all components and backup/restore jobs.

BEX achieves data reduction by sending only changed blocks, rather than changed files, to secondary storage – resulting in extremely small backup footprints. These are referred to by Syncsort as “Zero-Impact Backups” because, when capturing changed blocks on the physical or virtual client, BEX bypasses the file system resulting in minimal impact on clients, servers and applications. Further, the miniscule BEX backups are reconstructed on the destination storage as full backup images – without using additional space – eliminating the need to aggregate backups for point-in-time restores. Because of the small amount of backup data transferred and the fact that BEX backups have no impact on other processes on the network, BEX achieves a backup success rate of over 99%.

Other tangible benefits from BEX Zero-Impact Backups include acceleration of the backup process, reduction in network bandwidth consumption, and the ability to run server backups very frequently – allowing users to meet recovery point objectives. From a single snapshot, granular or large scale recoveries can be performed, from file-level to site-level and everything in between. There is no need to perform or store multiple backups due to differing recovery objectives.

BEX Advanced Recovery also enables immediate recovery of critical applications and data via an iSCSI mount to a backup snapshot. When this technique, called Instant Availability, is applied, the result from the user's perspective is a real browsable and writable file system, yet no backup images are destroyed through this process. Users have live immediate access to file and application servers, and resume business operations in minutes, without a data transfer.

About NetApp

NetApp creates innovative storage and data management solutions that help accelerate business breakthroughs and deliver outstanding cost efficiency. Their dedication to principles of simplicity, innovation, and customer success has made them one of the fastest-growing storage and data management providers today.

Customers around the world choose NetApp for their "go beyond" approach and broad portfolio of solutions for business applications, storage for virtual servers, disk-to-disk backup, and more. Their solutions provide nonstop availability of critical business data and simplify business processes so customers can deploy new capabilities with confidence and get to revenue faster than ever before. Customers rely on NetApp's industry-leading solutions to lower the cost of protecting their data, business, and reputation.

About Syncsort

Syncsort is a global software company that speeds data processing, data integration, and data protection & recovery. Supported by a professional services team built on 40 years of expertise, Syncsort helps customers drive ROI with software solutions that combine high performance, optimal hardware utilization, and ease of use. With over 12,000 deployments worldwide, Syncsort customers include 90 of the Fortune 100 companies. Syncsort is headquartered in Woodcliff Lake, NJ, and has international subsidiaries in the United Kingdom, France, and Germany, with support centers in the United States and Netherlands. An international network of partners, resellers and distributors positions Syncsort products as an integral component of IT solutions.

About Scasicomp

Created in 1994, Scasicomp helps companies design, integrate and deploy solutions for information storage and data backup in multi-platform environments and on all types of

Lab Benchmark Testing Report

networks, while at the same time ensuring continuity of service and performance. Scasicomp proposes, recommends and integrates a vast range of hardware and software solutions via the products offered by its manufacturing and editing partners, thus meeting the current needs of IT managers.

Scasicomp's teams are qualified to recommend, install, deploy and implement solutions and integrate them into existing set-ups thus meeting your requirements in storage, archiving, backup, de-duplication, virtualization, Disaster Recovery Plan or intelligent automation of storage and archiving.