



## ExaGrid

*Cost-effective Disk-Based Backup with  
Data De-Duplication*



# Table of Contents

Changing Data Protection Requirements .....	3
A New Data Protection Approach .....	3
ExaGrid Disk-based Backup with Data De-duplication .....	4
ExaGrid Architecture .....	5
ExaGrid Functional Overview .....	5
Highest Backup Performance with Post Processing.....	6
Highest Restore and Tape Copy Performance with Byte-level Data De-duplication.....	6
Highest Performance for Backup, Restore and Tape Copy .....	6
Byte-level Data De-duplication .....	7
ExaGrid Optional Second Site System for Long Term Offsite Retention or Data Disaster Recovery.....	8
Cross-Site Data Protection.....	8
WAN Bandwidth Efficiency with Byte-level Data De-duplication .....	9
GRID Scalability.....	10
Data Protection Redundancy .....	11
Data Integrity Verification.....	11
Advanced Disk-based Backup and Security Management .....	11
ExaGrid Uses Your Backup Application Data Retention Management.....	11
ExaGrid Security .....	12
ExaGrid in Your Environment.....	13
The ExaGrid Advantage .....	14
Scalability Without Added Complexity .....	14
Performance Scaling with Data Growth.....	15
Right Sizing to the Data .....	15
Efficient Scaling of De-duplication.....	15
ExaGrid Cost-effective Disk-based Backup - Summary .....	16

## Changing Data Protection Requirements

New business imperatives, along with the continuous growth in the volume of data that IT organizations have to manage are creating the need for a more intelligent data protection solution. Organizations today face a range of challenges, including:

- Tape backups that take more time while IT organizations have less time
- Backup windows that are expanding outside of the allotted time
- Project deadlines that require faster and more reliable restores
- Increasing legal and audit requests for fast access to random collections of backup data
- Requests from auditors, lawyers, regulators, customers, partners, and suppliers demanding extended data retention periods and increased data security policies
- The need for faster recovery from site disasters

Tape-based backup is proving insufficient to meet these new requirements. Restores from tape are slow and unreliable because:

- During critical restore operations, tapes may be corrupted, damaged, or blank
- Tapes are often unlabeled or mislabeled
- Tapes may be lost or stolen
- Locating and retrieving data from tapes can take hours or days to complete
- Old tapes can't be read by new tape drives or new backup/archive software releases

For the past five decades, magnetic tape as a backup storage medium has been the only cost-effective solution. However, new business demands for faster and more reliable backups and restores have pushed the use of tape beyond many of its capabilities.

## A New Data Protection Approach

A new class of Disk-based backup solutions with compression and data de-duplication is now available to address these issues. These innovative systems replace onsite and, if required, offsite tapes with low cost, efficient, online disk-based storage. Since all data is protected online, these solutions ensure the highest levels of long-term data integrity, availability, and security. In addition, these systems dramatically reduce the administration required to perform backups and to restore files or recover from site disasters.

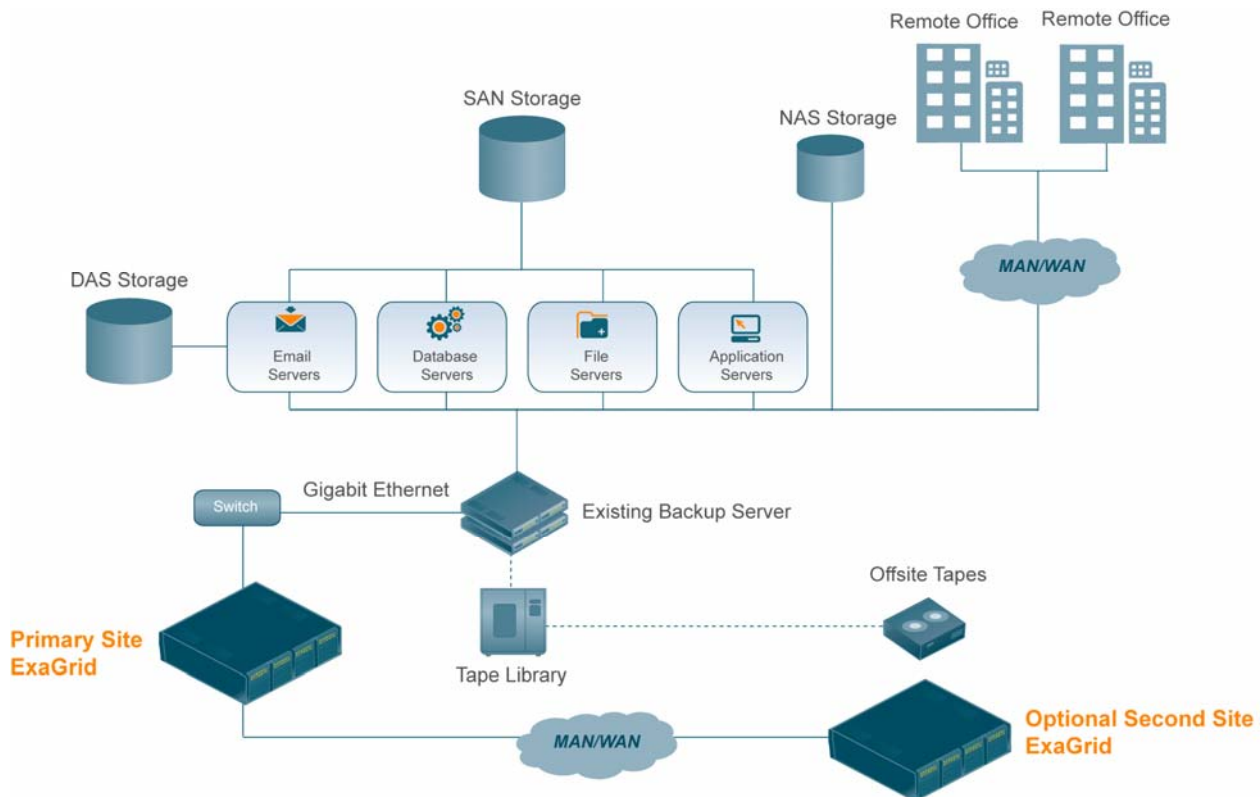
This paper explores the architecture and functionality of ExaGrid's Cost-effective Disk-based Backup solution.

## ExaGrid Disk-based Backup with Data De-duplication

ExaGrid's turnkey system works in conjunction with your existing backup applications to replace onsite or onsite and offsite tape backups with a disk-based backup system. ExaGrid is easy to install and use, and fits seamlessly into your existing environment, enabling you to leverage your existing technology investments. You can still maintain your existing offsite tape strategy or with an optional second site system you can eliminate offsite tapes.

ExaGrid is extremely cost-effective, providing data compression for the most recent backup along with byte-level data de-duplication for all previous backups. ExaGrid's byte-level data de-duplication technology stores only the changes from backup to backup instead of storing full file copies. This unique approach reduces the amount of SATA disk space needed by a range of 10 to 1, to upwards of 50 to 1, or more, resulting in a solution that is a fraction the cost of standard SATA drives and about the cost of tape backup. A second ExaGrid system can replace offsite tape because the byte-level data de-duplication technology only moves changes, requiring minimal WAN bandwidth.

The following diagram shows how ExaGrid is deployed:



At the primary site, existing backup applications can write data to the ExaGrid in place of sending backup data to tape. This is called backup to disk or disk-based backup. ExaGrid also enables you to perform disk-to-disk-to-tape backup (D2D2T), where data can be written from the ExaGrid to a tape system for offsite tapes. Lastly, ExaGrid can replace tape with an onsite ExaGrid and an offsite ExaGrid system where all backups are kept both on and offsite, on an ExaGrid system. All backup data is maintained on high quality disks in the onsite system as well as at the optional second site system.

ExaGrid is designed to work with what you already have in your environment. It leverages:

- Your existing backup software applications
- Standard backup to disk features
- Standard Ethernet IP network connectivity within and between sites
- Standard VPN connections between sites for security
- Standard 19" data center racks

## **ExaGrid Architecture**

The ExaGrid System appears to your backup application as a Network Attached Storage (NAS) device. This system is created from industry-standard components including Intel's dual core Xeon processors, Seagate SATA drives with RAID 6 plus hot spare protection.

The ExaGrid system is mounted in a standard 19" rack and are connected to the backup server via a standard Ethernet switch.. Additional ExaGrid servers can be connected to the switch to create a scalable system.

The ExaGrid primary and optional second site systems communicate with each other using standard IP WAN connectivity. Security of data transferred between sites is achieved with standard VPN encryption.

## **ExaGrid Functional Overview**

### ***A Disk-based Target for Backup Applications***

The ExaGrid system acts as a target for your existing backup applications. Instead of sending nightly and weekend backups to a traditional tape system, files are written directly to one or more NAS shares using the standard NAS disk target feature in the existing backup application.

Through the use of your existing backup application, all data can be protected, regardless of whether primary storage is SAN, DAS or NAS and whether the file system is Windows, NetWare, Linux, UNIX, VMware and other popular operating system environments.

Many of the industry's most popular backup and data protection products are supported by

ExaGrid including: CA ARCserve, Symantec Backup Exec, Symantec NetBackup, CommVault Galaxy, EMC Networker, Microsoft SQL dump, VMware (VMDK) as well as many others. Consult ExaGrid for an expanded list of supported backup applications.

### **Highest Backup Performance with Post Processing**

The ExaGrid system uses post processing to perform its backup data compression and de-duplication. This means that the backup data is written directly from the backup server to disk without any processing to interfere. Once the backup job is complete and off the network, then the ExaGrid system compresses and de-duplicates the data. Since nothing is done to the data, it can be written at the highest possible rate, meaning the smallest possible backup window.

Each ExaGrid server is built around this architecture and has the right amount of processing power, memory and storage capacity to allow room for the data to land and be post-processed. Post-processing is also not protocol dependent, so the ExaGrid system performs just as well in Windows (CIFS) environments as well as UNIX (NFS) environments.

The alternative technology, in-line processing, means data is processed as it comes in from the backup server. Since each bit of data has to be de-duplicated on the fly, this can slow the backup down significantly. With ExaGrid's post process approach, backups run as fast as disk technology can write, resulting in the shortest backup window possible.

### **Highest Restore and Tape Copy Performance with Byte-level Data De-duplication**

The ExaGrid system stores the most recent version of the data in its complete form compressed two to one. Since 90% of restores typically come from the most recent data, and that data requires no processing to retrieve, the ExaGrid will provide the fastest possible restore.

Even for restores from older data, starting from the recent data in its complete form and working backwards to the version required is faster than having to start with the oldest data and work forwards through all the version, as block level de-duplication has to do.

A standard backup job can run to copy the most recent backup from ExaGrid to tape. The tape copy performance is fast since ExaGrid keeps the most recent backup in its complete form.

### **Highest Performance for Backup, Restore and Tape Copy**

ExaGrid's unique architectural and technological advantages optimize the system for backup and restore/tape copy performance. This means shorter backup windows, tapes ready for off-site sooner, and faster restores.

## ***Byte-level Data De-duplication***

Data backups make inefficient use of storage capacity. Even when files don't change from week to week, full backup file sets are still created. The same data is backed up over and over again.

ExaGrid stores the most recent backup, and for all previous backups only stores the bytes that change from backup to backup.

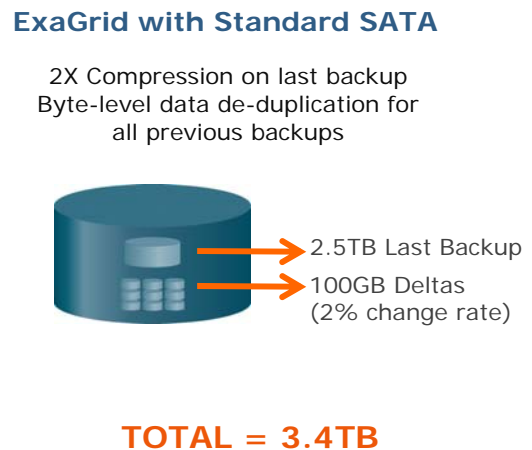
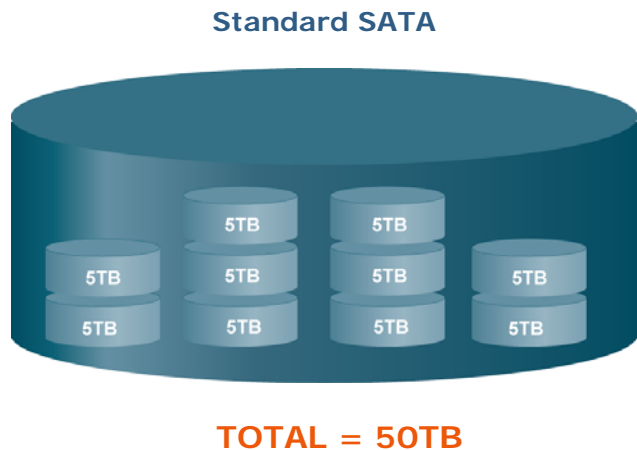
These efficiencies enable ExaGrid to be more cost effective than standard SATA disk with RAID. An ExaGrid system typically requires approximately 1/20<sup>th</sup> to 1/50<sup>th</sup> the SATA storage space of standard disk, resulting in a solution that costs a fraction the price of standard disk, and about the price of tape backup.

Since backup history is kept for years, potentially hundreds of copies of the same data are stored over and over again. When a backup occurs night after night, the data backed up is almost identical to the night before. ExaGrid minimizes the amount of data to be stored by providing standard data compression for the most recent backups, along with byte-level data de-duplication for all previous backups. ExaGrid's byte-level data de-duplication technology stores only the bytes that change from backup to backup instead of storing full file copies, significantly reducing the amount of disk space required.

For example, in a scenario using standard SATA disks, if 5TB of primary data (full backup) is backed up ten times, then 50TB of data is stored at the end of the period., without additional overhead for RAID

Typically, only 2% of data changes between backups. Byte level data de-duplication only stores those changes at the byte level. If 5TB is backed up and the next day backed up again, only 100GB will have changed. Instead of storing 5TB ten times, the first 5TB is stored and then the 100GB of changes for each previous backup. With compression and data de-duplication, the first 5TB is compressed 2 to 1 resulting in 2.5TB plus nine 100GB changes will be stored for a total of 3.4TB. In this example the data de-duplication is about 15 to 1.

Longer retention periods result in even greater storage savings. For 5TB of data with 20 full backups, 100TB of storage without RAID, compression and data de-duplication. With compression and data de-duplication only 4.4TB of storage is required. (2.5 last back compressed plus 19 x 100GB changes). This results in a data reduction ratio of 23 to 1.



### ***ExaGrid Optional Second Site System for Long Term Offsite Retention or Data Disaster Recovery***

The need for fast and simple restores from long term offsite data backup storage or for data recovery during a site disaster is becoming a business imperative for many companies. With today's tape-based protection, the longest delays in recovering from a system or site disaster are caused by the need to acquire, load, and restore data from offsite tapes. Many times, tapes are lost, stolen or misplaced. In addition, the time it takes to complete a recovery operation increases considerably if one or more of the backup or archive tapes is found to be corrupted, damaged, blank, or incomplete. When this occurs, older backup tapes must be used. This results in a longer overall recovery process and the data finally restored to users and applications may be weeks old.

With ExaGrid, an available second site system provides offsite tape replacement and faster and more reliable data recovery. The byte level changes from the primary site ExaGrid are replicated efficiently to a second site. Since only the bytes that change from backup to backup are replicated typically only about 1/50<sup>th</sup> of the backup traverses the WAN. When a primary site disaster occurs, it takes just a few seconds to initiate a data recovery process at the second site using the existing backup application. All data is available because protected data is continually replicated into the second site repository.

### ***Cross-Site Data Protection***

Since the ExaGrid can be located at both a primary and a second site, each site can act as backup target storage and can cross-site protect each other. This allows the second site to act as a data recovery site if the primary site is lost and for the primary site to act as a data recovery site if the second site is lost. The data repositories at each site retain all of the backup files.

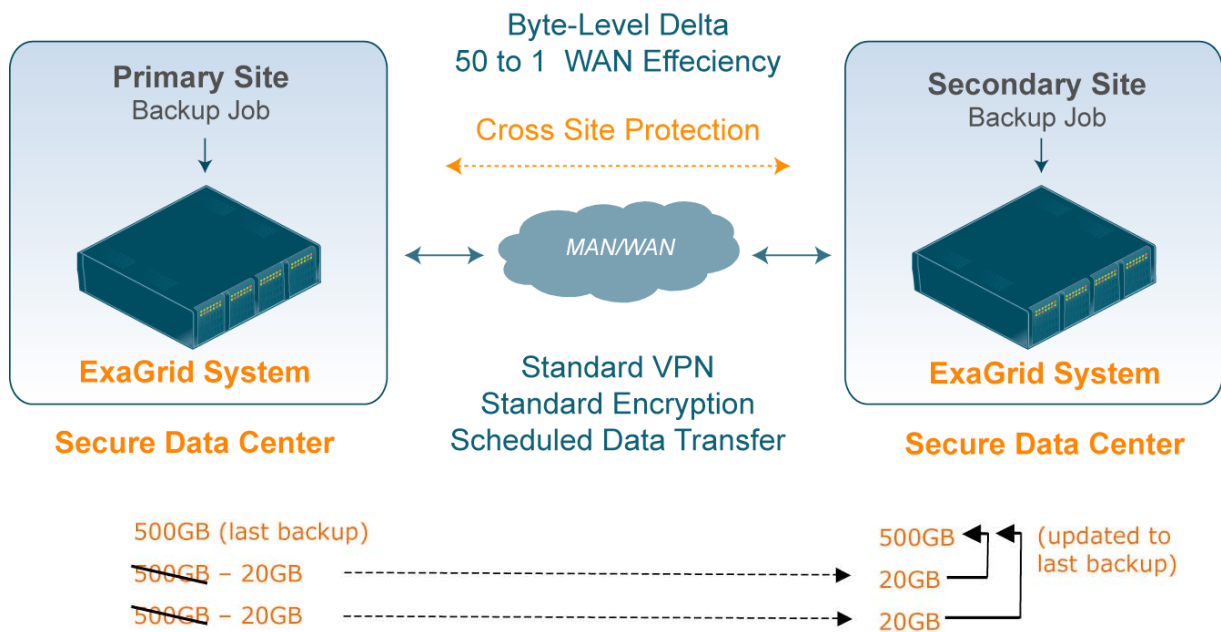


## WAN Bandwidth Efficiency with Byte-level Data De-duplication

Sending a large collection of files or a very large database or email backup file across a WAN is not feasible with the limited bandwidth of most WAN connections.

ExaGrid's delta technology can also deliver the equivalent of a 50:1 reduction in the amount of bandwidth consumed on a WAN as only byte level deltas are transmitted to the second site repository. The 50:1 reduction is an average based on the fact that in most environments about two percent of the overall data, at the byte level, changes from week to week. ExaGrid only moves the two percent that changed, at the byte level, since the full backup job already resides on the second site system. When the bytes transfer to the second site they are stored but also merged into the full backup on to bring it up to date with the most recent full backup.

In a typical ExaGrid dual-site implementation, two T1's, or about 3Mbps of bandwidth is required per 1TB of primary data backed up during a full backup. Two percent of 1TB is 20GB, and two T1's of bandwidth will allow that 20 GB of changes to be transmitted to the remote site in well under 24 hours. 2TB will require four T1's, 3TB will require six T1's and so on. Unusual environments including complete full backups every night, high data change rates, or a requirement to have the data off-site under specific time constraints could require more bandwidth.



**On-site and Off-site are Identical**

*For each 1TB full, two T1's are required (assuming a 2% byte change rate)*

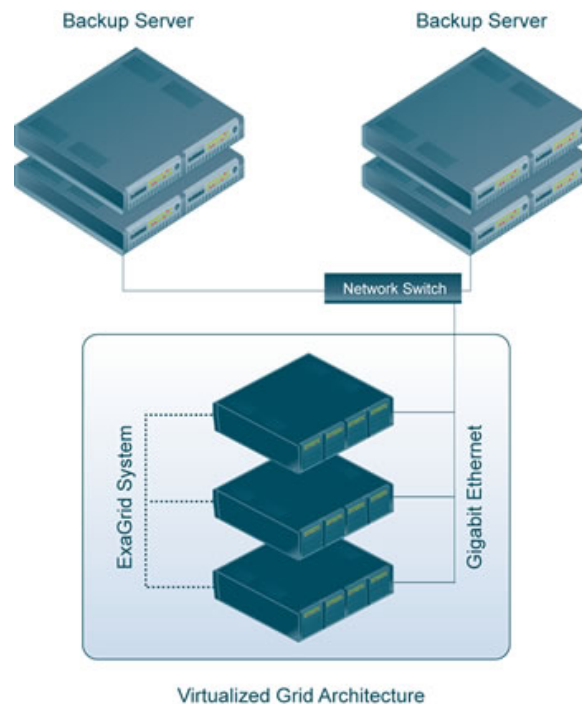
## GRID Scalability

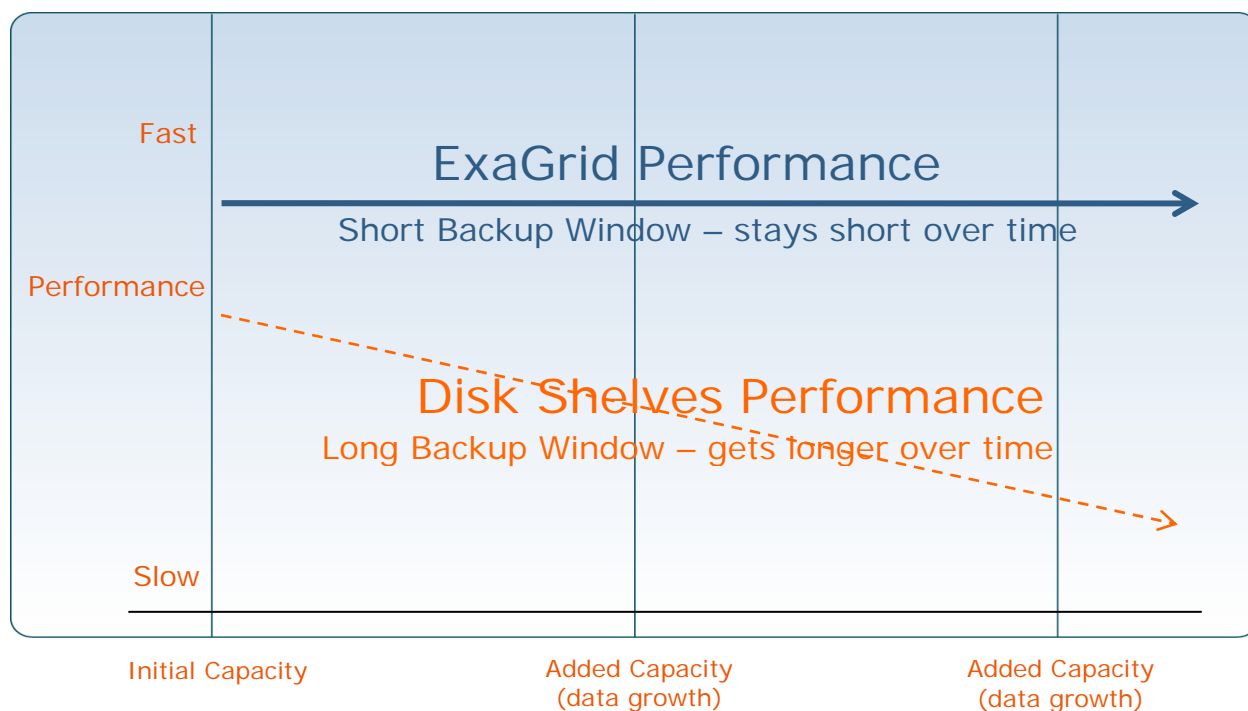
ExaGrid's byte-level data de-duplication does not depend on a large in-memory hash table, since it does not use blocks and hashes to find duplication, nor does it process in line. This allows data and processing to be automatically distributed across multiple systems using a peer GRID computing architecture.

The ExaGrid system can be easily expanded by adding additional ExaGrid servers. Once installed, multiple ExaGrid servers act as a single virtual pool of storage capacity for your backups. This eliminates the need for you to migrate your data or start a new retention rotation in a new isolated system. The ExaGrid software ensures that there is always adequate room for your ongoing backups across the ExaGrid systems.

ExaGrid's scalable architecture has several advantages, including:

- In place, plug and play scalability without added complexity
- The ability to add systems without disruption
- The ability to mix and match systems of different disk capacities and processing power to create a larger, more powerful pool of capacity
- New higher capacity and higher performance systems can be added to an existing ExaGrid system over time. This eliminates the technology obsolescence problems associated with tape technologies today
- Right-sizing of the system to the data
- Performance is maintained as each ExaGrid is a server with disk, processor, memory and bandwidth. Performance related system resources accompany additional disk capacity.





### ***Data Protection Redundancy***

RAID 6 plus a hot spare is employed to protect against two simultaneous disk failures in order to ensure that your backup data is always available for restore. This method of data protection redundancy provides:

- System operation, even with a simultaneous disk drive failure
- The ability to automatically rebuild the RAID set, to the included spare drive, if a drive failure occurs
- How swappable drives; if a drive fails the system still runs; replace the drive at any time

### ***Data Integrity Verification***

All copies and versions of data in the ExaGrid system are continually checked using MD5 checksum technology. When data is first protected, a checksum is computed and stored along with the protected data.

## **Advanced Disk-based Backup and Security Management**

New business imperatives are driving advanced data retention requirements, new random data restore requirements, and increased security requirements.

### ***ExaGrid Uses Your Backup Application Data Retention Management***

Based on regulatory rules, business goals, customer agreements, partnership agreements and legal requirements, each collection of data that is created by an organization may need to be

retained for different periods of time. The ExaGrid system acts as a live disk-based data repository for your existing backup application. All retention is controlled by the backup application. If the backup application reaches its retention period all replicas and versions of that data is purged from the system. It's impossible to manage the retention and purging of data with magnetic tapes since every tape contains a mix of data in an interlaced format.

### ***ExaGrid Security***

Data theft from lost or stolen tapes is being reported by major corporations at an alarming rate. By its very nature, traditional tape is not secure, and there is an increased likelihood that vital company data will fall into the wrong hands:

- Tapes are handled by many individuals, including employees and third-party personnel
- The data on tape is rarely encrypted and if lost or stolen can be easily read

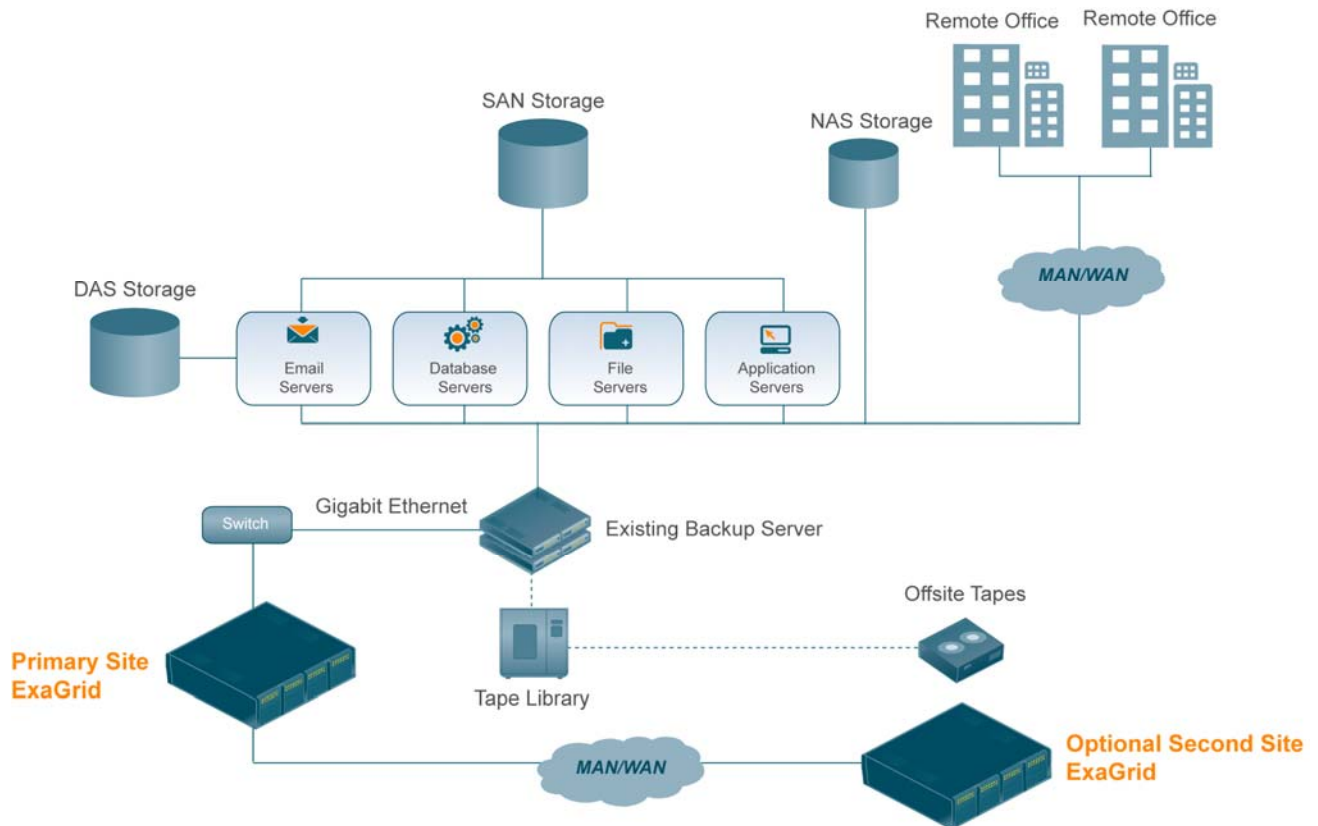
ExaGrid eliminates all of these data security issues. Standard building, data center and network security apply, as the primary ExaGrid system and optional second site system reside in standard data centers making ExaGrid as secure as the rest of your IT infrastructure.

In addition, as data is transmitted between repositories across the WAN, the data is encrypted using industry-standard VPN technology.

## ExaGrid in Your Environment

ExaGrid provides single site or two-site data protection for your IT environment. It leverages:

- Your existing backup applications
- Industry-standard Ethernet IP networking technology
- Standard Intel-based systems with high quality Seagate SATA drives
- Standard 19" data center rack mounted configurations
- A single site system maintaining standard offsite tapes or an end-to-end two site system



## The ExaGrid Advantage

ExaGrid is the most advanced disk-based backup system with compression and data deduplication on the market today. The ExaGrid solution has key advantages which should be considered when choosing a disk-based backup system.

These key advantages are:

- Highest performance for backups for short backup windows
- Fast restores as the last backup is kept in its complete form
- Fast tape copy for creation of offsite tapes
- Right sizing as ExaGrid can scale in 1TB, 2TB, 3TB, 4TB and 5TB increments
- A WAN efficient offsite / second site solution for offsite tape replacement
- GRID Scalability to maintain performance as data grows

### ExaGrid Server Configurations

**1TB – EX1000**  
**2TB – EX2000**  
**3TB – EX3000**  
**4TB – EX4000**  
**5TB – EX5000**



**Each server can store the rated full backup, plus 16 weeks retention**

- EX5000 stores up to a 5TB full backup, plus up to 16 weeks of retention

**Mix-and-Match in 1TB to 5TB increments to build larger systems**

- Example: 7TB ExaGrid System = 1 EX4000 + 1 EX3000
- Example: 10TB ExaGrid System = 2 EX5000's
- Example: 20TB ExaGrid System = 4 EX5000's

### Scalability Without Added Complexity

The GRID technology in the ExaGrid systems means that when the system needs to grow, it is a simple plug and play process. To add an ExaGrid server into an existing system, it is simply plugged into the rack, connected to the network, and a back-end gigabit Ethernet network between the systems is connected. The systems automatically recognize each other and virtualize into a single system. They then load and capacity balance automatically without user intervention. Add a 1TB system to a 2TB system and you end up with one 3TB system with a

single management view and single virtualized storage system with additional computing power, memory and network capacity.

Without the GRID technology, separate systems have to be deployed. This means that data management between the systems is manual, and has to be manually rebalanced if data sets grow at differing rates.

### **Performance Scaling with Data Growth**

Since the GRID technology grows by adding units of processing, memory, network connections and storage together, the total capability of the system grows with the data. This means that as the data grows, the backup window will remain short.

In alternative systems where system growth is accomplished by adding storage only, performance remains the same even if capacity increases. This means that double the data means double the backup window. It also means that when the capacity limit of the system is reached, no further growth is possible and the only option is separate systems.

### **Right Sizing to the Data**

ExaGrid's flexible combination of GRID technology, post-processing and byte-level data de-duplication allows ExaGrid to provide a solution that scales up and down in granular units to the backup data set. ExaGrid offers building blocks in 1, 2, 3, 4 and 5TB units. These not 1-5 physical TB, but rather systems designed for backing up their rated size of data AND include room for all the de-duplicated backup retention and history. Up to 4 ExaGrid servers can be plugged into a single GRID for up to a 20TB backup, plus 16 weeks of retention in a single GRID.

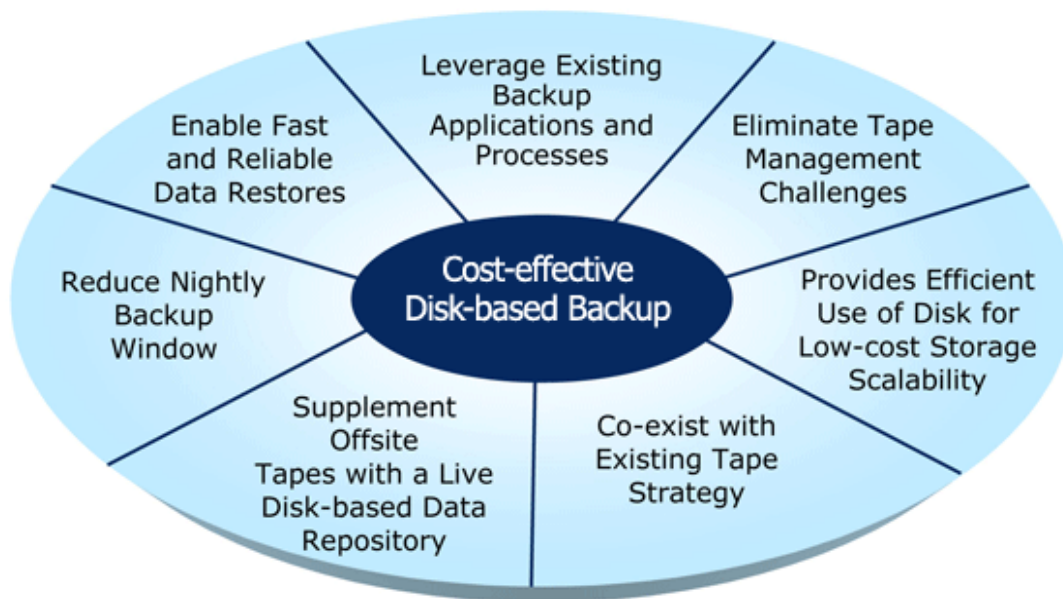
This allows for the most cost effective solution, since a system can be sized for the current data, or desired growth, and the increased later using the GRID technology. There is no need to overbuy now to account for growth with disk and processor prices continually falling. This approach also allows for starting slowly and adding later.

### **Efficient Scaling of De-duplication**

With a single-system GRID, all data history is stored together, and thus all data is de-duplicated. Without this architecture, separate systems do not talk to each other, and do not share de-duplication history. This means duplicate copies of the data will be kept, over-utilizing system capacity.

## ExaGrid Cost-effective Disk-based Backup - Summary

ExaGrid offers complete primary site and second site Cost-effective Disk-based Backup.



***ExaGrid is the only solution that satisfies all seven requirements of Cost-effective Disk-based Backup:***

1. Leverage existing backup applications and processes
2. Reduce nightly backup window
3. Enable fast and reliable data restores utilizing a disk-based verified data approach
4. Eliminate tape management challenges through the use of disk and server technology
5. Provide efficient and low cost scalability through compression and byte-level data de-duplication technology
6. Co-exist with existing tape strategy
7. Supplement offsite tapes with a live disk-based data repository





## ExaGrid

*Cost-effective Disk-Based Backup with  
Data De-Duplication*

<http://www.exagrid.com>

1-800-868-6985