# Detailed Product Description

## ExaGrid Backup Storage Appliances with Scale-out, Landing Zone, and Data Deduplication

# Table of Contents

# Executive Summary

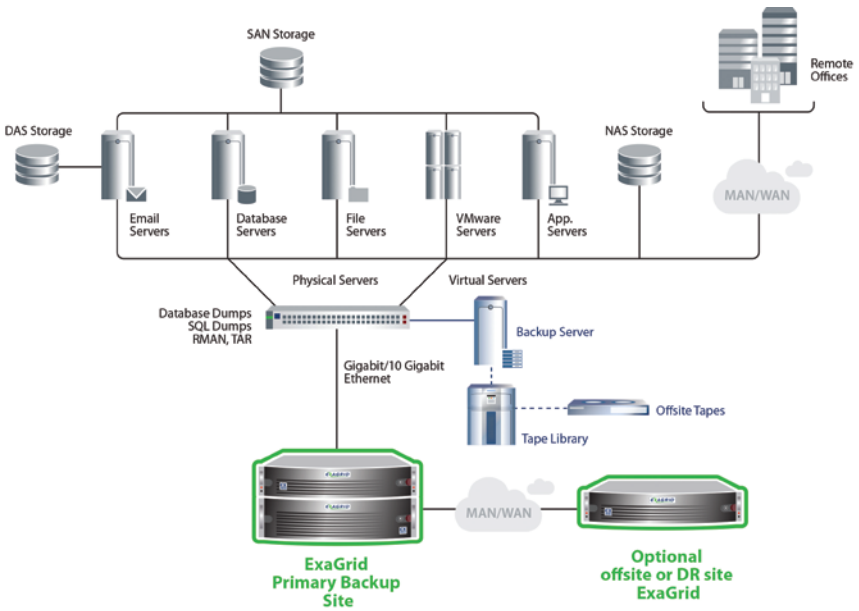
ExaGrid delivers better backups using the only disk-based backup appliance with data deduplication purpose-built for backup that leverages a unique architecture optimized for performance, scalability and price. ExaGrid customers achieve the fastest backup times as data is written at the speed of disk and the most recent backups are stored in their complete undeduplicated form for fast restores, fast VM instant recoveries, and fast offsite tape copies. ExaGrid delivers the fastest backup performance for the shortest backup window and maintains a fixed length backup window as data grows by adding full appliances into a scale-out GRID.

ExaGrid makes backup better with:

- Fast backup resulting in the shortest backup window
- A fixed length backup window as data grows, eliminating expensive future forklift upgrades
- Fastest restores
- Instant VM recoveries in seconds to minutes
- Fastest offsite tape copies
- Fast and reliable offsite disaster recovery
- Low cost up front and over time as data grows

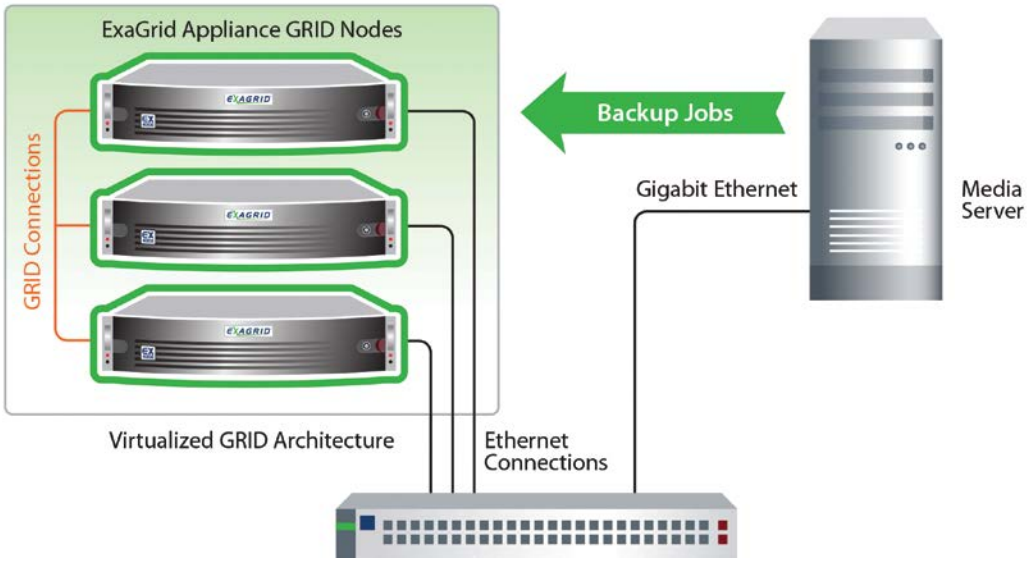
# ExaGrid Basic Concept

ExaGrid appliances work seamlessly with industry-leading backup applications and database utilities by presenting themselves as standard NAS shares (CIFS or NFS). Backup jobs are directed to the ExaGrid appliance. ExaGrid appliances are easily integrated into existing backup environments, as illustrated below.




**ExaGrid Sits Behind the Existing Backup Server and Replaces Tape Onsite or Offsite**

# ExaGrid Technology Overview




**ExaGrid Appliances Connect to Form a Scalable GRID**

ExaGrid appliances are comprised of Intel Quad Core XEON processors, RAID6 + Hot Spare storage using enterprise class SATA or SAS drives, and ExaGrid software. Each appliance plugs into a switch and is virtualized into a shared GRID.

The media server is connected to the same switch and sees the appliances as one or more NAS shares. Since each appliance includes the appropriate amount of processor, memory, landing zone disk, deduplicated repository disk, and bandwidth for the rated data size, performance increases as more appliances are added to the GRID.

## ExaGrid Appliance Models

The product line's multiple appliance models can be combined into a GRID configuration of up to 25 appliances in a single GRID for a total of 2.4PB raw capacity, 1.95PB of usable capacity and supports a full backup of up to 1PB, total backup logical data capacity of 16PB.

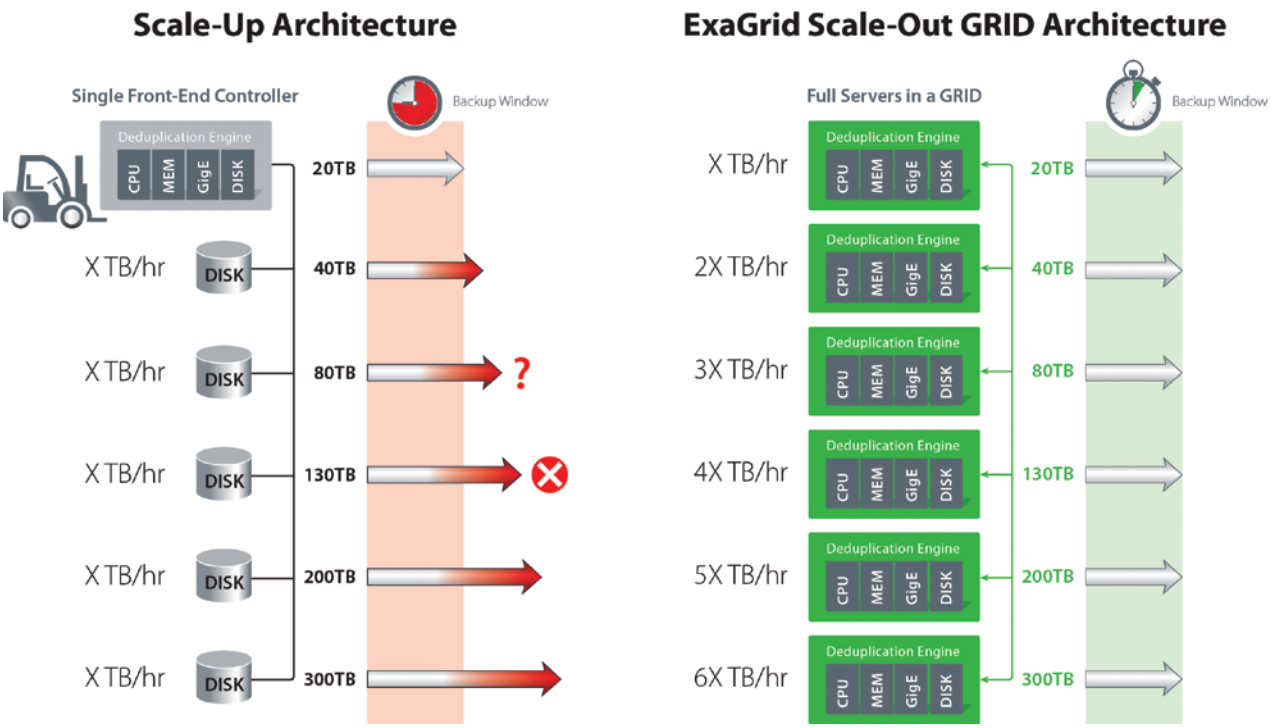
| ExaGrid Model | Raw Capacity (TB) | Usable Capacity (TB) | Capacity for Weekly Fulls (TB) | Capacity for 16 Weekly Fulls (TB) | Max Backup Thruput (TB/hr) |
|---|---|---|---|---|---|
| **ExaGrid Appliance Nodes** | | | | | |
| EX2000 | 7 | 4 | 2 | 32 | 0.36 |
| EX3000 | 9 | 6 | 3 | 48 | 0.72 |
| EX5000 | 16 | 10 | 5 | 80 | 1.08 |
| EX7000 | 20 | 14 | 7 | 112 | 2.40 |
| EX10000E | 26 | 20 | 10 | 160 | 2.40 |
| EX13000E | 32 | 26 | 13 | 208 | 2.40 |
| EX21000E | 56 | 44 | 21 | 336 | 4.32 |
| EX32000E | 72 | 63 | 32 | 512 | 7.56 |
| EX40000E | 96 | 78 | 40 | 640 | 8.00 |
| **ExaGrid Appliance Nodes with Encryption** | | | | | |
| EX7000-SEC | 20 | 14 | 7 | 112 | 2.40 |
| EX10000E-SEC | 26 | 20 | 10 | 160 | 2.40 |
| EX13000E-SEC | 32 | 26 | 13 | 208 | 2.40 |
| EX21000E-SEC | 56 | 44 | 21 | 336 | 4.32 |
| EX32000E-SEC | 72 | 63 | 32 | 512 | 7.56 |
| EX40000E-SEC | 96 | 78 | 40 | 640 | 8.00 |
| **Example GRID Configurations** | | | | | |
| EX61-G | 144 | 120 | 61 | 976 | 12.3 |
| EX80-G | 192 | 156 | 80 | 1280 | 16.0 |
| EX120-G | 288 | 234 | 120 | 1920 | 24.0 |
| EX160-G | 384 | 312 | 160 | 2560 | 32.0 |
| EX200-G | 480 | 390 | 200 | 3200 | 40.0 |
| EX240-G | 576 | 468 | 240 | 3840 | 48.0 |
| EX280-G | 672 | 546 | 280 | 4480 | 56.0 |
| EX320-G | 768 | 624 | 320 | 5120 | 64.0 |
| EX360-G | 864 | 702 | 360 | 5760 | 72.0 |
| EX400-G | 960 | 780 | 400 | 6400 | 80.0 |
| EX440-G | 1056 | 858 | 440 | 7040 | 88.0 |
| EX480-G | 1152 | 936 | 480 | 7680 | 96.0 |

Scales to 25 appliances in a scale-out GRID (1PB Full Backup / 200TB/hr.)

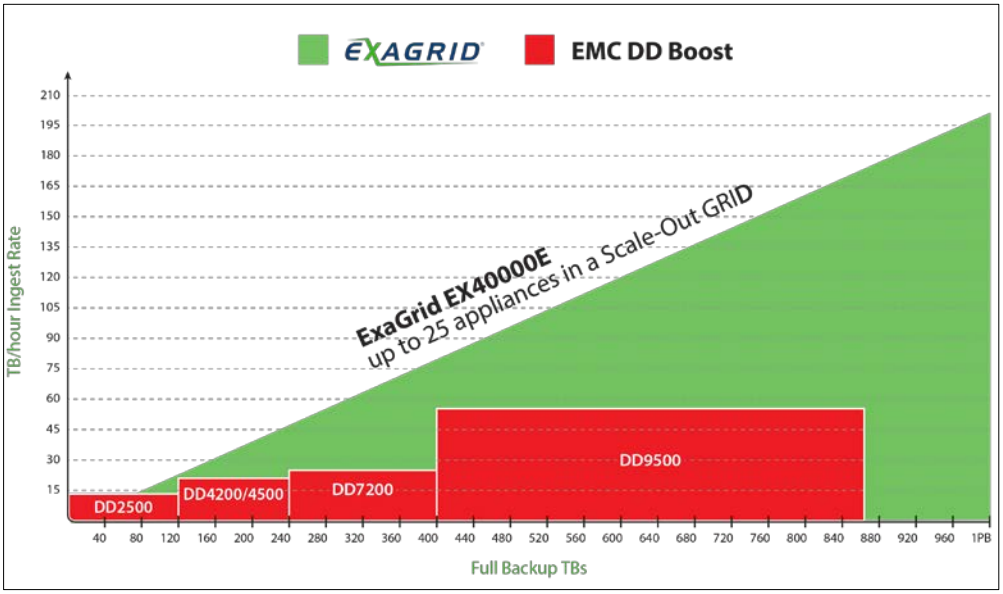
## Scalable GRID Architecture – Scale-Out

The typical business or organization is seeing data growth of 30% a year which doubles data every 2.5 years. This data growth presents challenges to IT to ensure their backup system can scale easily to support that growth. With scale-up disk backup solutions that have a front-end server/controller architecture and add just disk shelves as data grows, you start with a short backup window, but as data grows, the backup window expands because only disk is added but no additional deduplication processing resources. Eventually, the backup window expands to a point where you must replace the front-end server with a more powerful server via a costly "forklift upgrade."

In contrast, ExaGrid's scalable GRID-based approach adds full servers—including memory, processor, and bandwidth as well as disk. The figure below shows the differences between how the two different architectures cope with data growth over time.



### Scale-out vs Scale-up Backup Storage

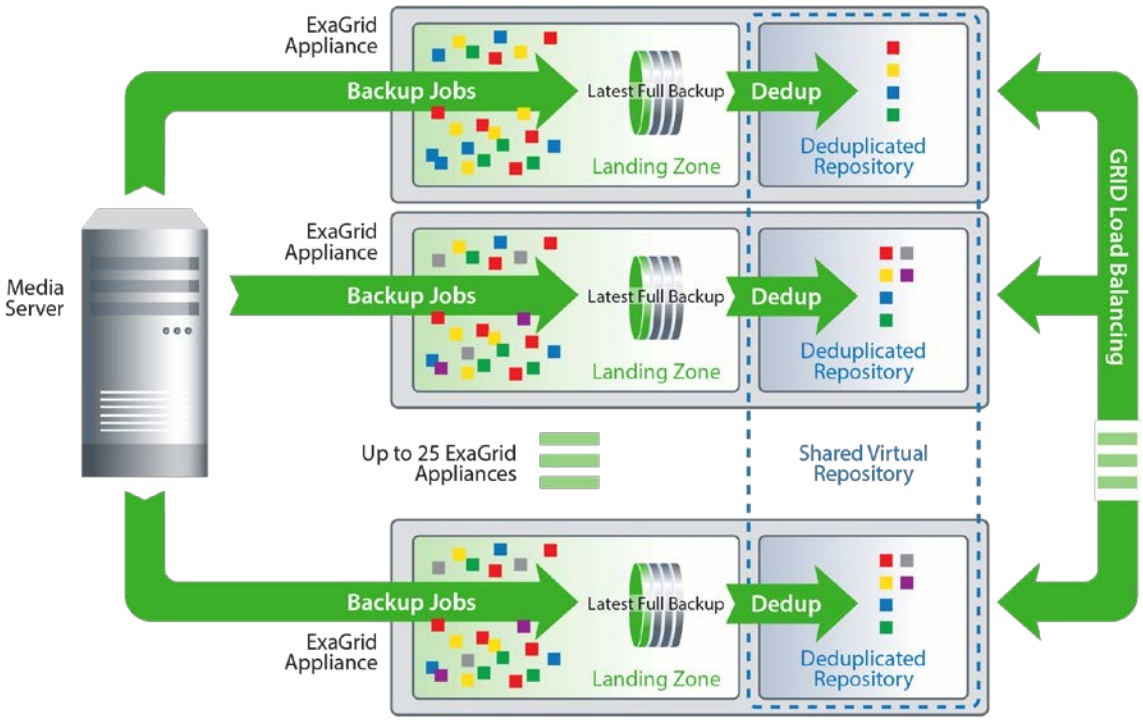



This unique approach provides the following benefits:

- **No Expansion of Backup Windows as Data Grows** – By adding full servers, ExaGrid maintains consistently fast backup performance and a fixed length backup window as data increases.

- **Seamless Scalability with No Forklift Upgrades** – ExaGrid's scalable GRID-based approach avoids forklift upgrades by adding modular capacity with full servers to the existing system in full backup increments of 2, 3, 5, 7, 10, 13, 21, 32, or 40TB with up to 25 appliances in a single GRID.

- **No Obsolescence of Previous Model Systems** – All ExaGrid appliances can be seamlessly added to any existing ExaGrid deployment and can be mixed and matched with older appliances of any size, thereby eliminating obsolescence. Adding a new ExaGrid appliance is as simple as plugging it in and pointing backup jobs at a NAS share.




ExaGrid disk backup appliances include GRID computing software which allows them to virtualize and share data storage capacity with one another. This GRID system shown above can expand as your data grows by adding appliances to the GRID, providing up to 2.4PB raw capacity, 1.95PB usable capacity, and allowing full backups of up to 1PB in a single GRID. Multiple GRIDs can be used at a single location.

# Designed for Performance

## Data Deduplication

ExaGrid's innovative approach to data deduplication minimizes the amount of data to be stored by using zone-level data deduplication across all received backups. ExaGrid's zone-level delta technology stores only the changed bytes from backup to backup instead of storing full copies.

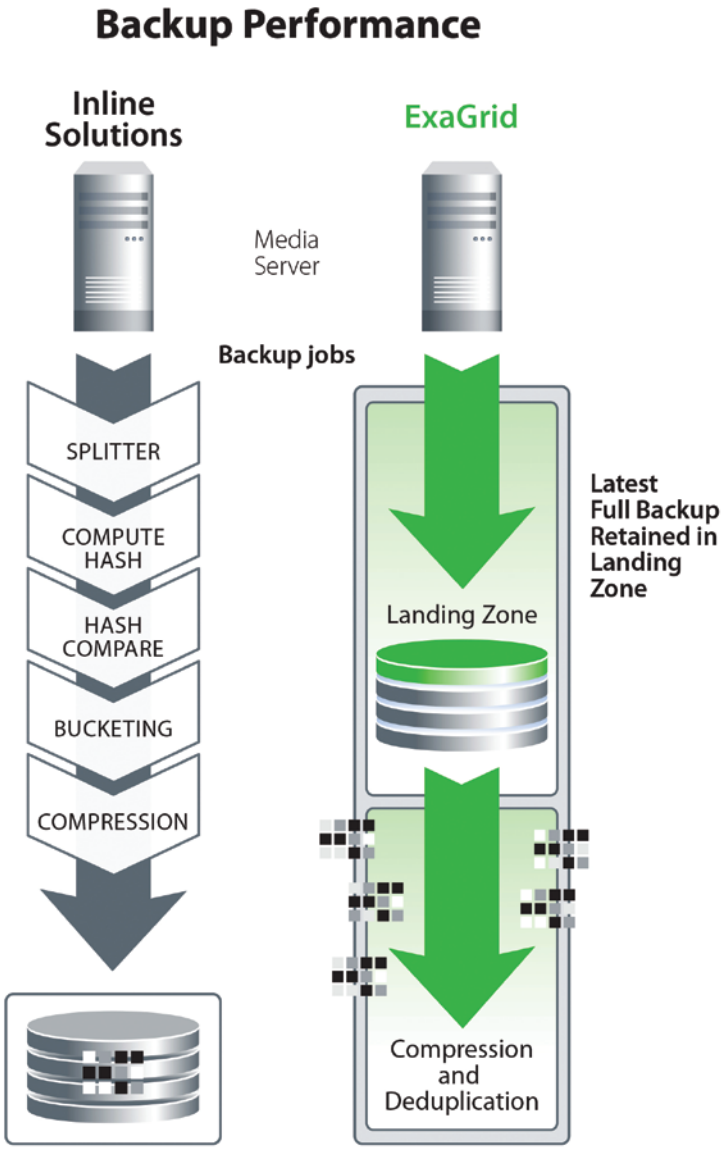
This unique approach reduces the disk space required by an average of 20:1 and from 10:1 up to 50:1 depending on data type, delivering unparalleled performance for the fastest backups and restores.

## Data Backup Performance

The ExaGrid appliance performs deduplication in parallel with the backup using idle system cycles, termed “Adaptive Deduplication.” This approach keeps the deduplication out of the backup path, resulting in faster backups and a shorter backup window. The Adaptive Deduplication approach gives full system resources to the backups for the shortest backup window but uses idle cycles to perform deduplication and offsite replication, allowing data to be replicated offsite for disaster recovery.

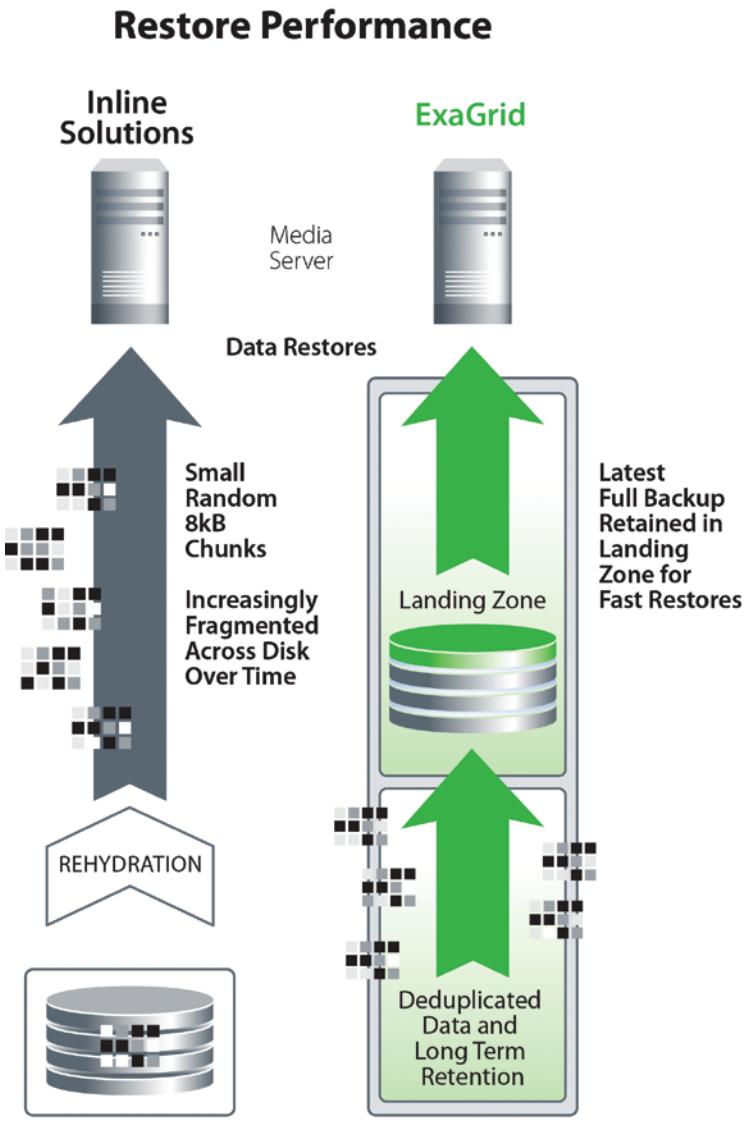
ExaGrid supports data deduplication globally across all NAS shares and all appliances in a scale-out GRID. Global deduplication ensures that all data is deduplicated regardless of the number of NAS shares targeted and/or the number of appliances in a GRID. In addition, global deduplication allows organizations the flexibility to redirect backup jobs to any NAS share, on any appliance, at any time while maintaining data deduplication globally across the GRID.

Once the backup job is complete and off the network, the data is protected and immediately available to be restored. When deduplication and replication are complete, the data is ready for restore on the ExaGrid appliances at the disaster recovery site. As shown in the figure below, the backup data is written directly from the backup server to ExaGrid's disk landing zone at the highest possible rate with no inline processing to interfere, resulting in the shortest possible backup window.



### Data Restore Performance

Because ExaGrid's appliance allows each full backup to first land on the landing zone, it maintains the most recent backup in a full undeduplicated form for fast restores, Instant VM recoveries in seconds to minutes, and fast offsite tape copies. Since over 90% of restores and 100% of instant VM recoveries and tape copies are done from the most recent backup, this approach avoids overhead incurred from "rehydrating" data during critical restores. As a result, restore, recovery, and copy times from an ExaGrid are an order of magnitude faster than solutions that only store deduplicated data.

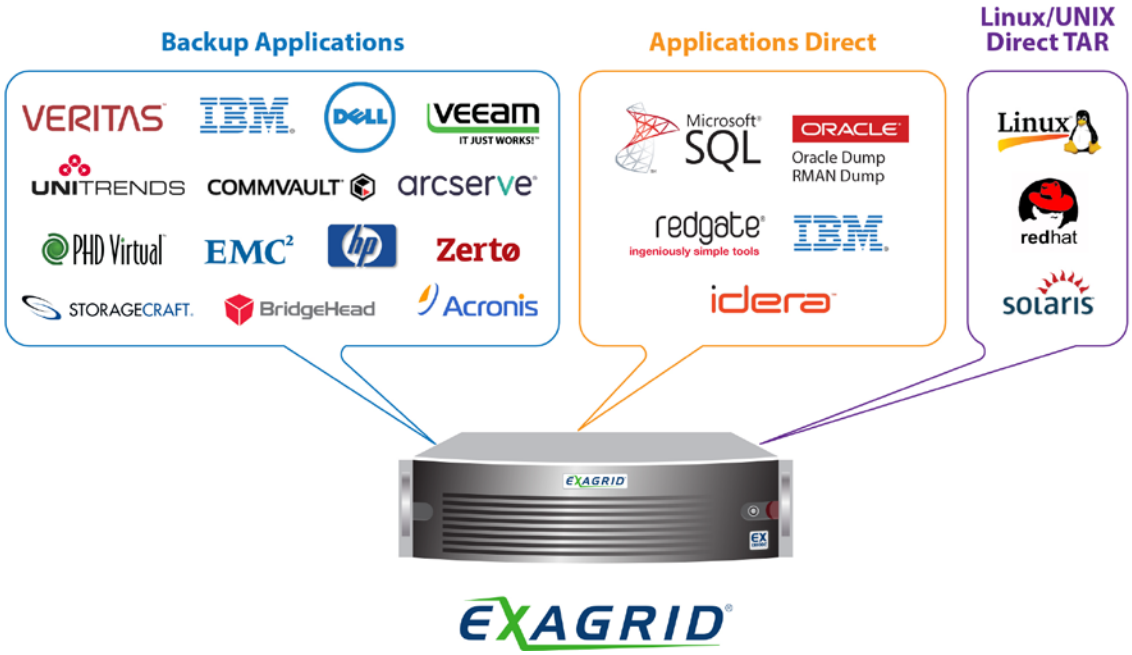


## Support for Heterogeneous Data and Utilities

Customer environments are made up of many backup approaches, backup applications, and utilities and different disk-based backup approaches support these in different ways. Customers may have any number of backups occurring in their environment, including traditional backup applications, specialized VMware and Hyper-V backup utilities, direct-to-disk SQL dumps, Oracle RMAN backups, and specific UNIX utilities such as UNIX TAR.

Disk-based backup appliances with data deduplication such as ExaGrid's are able to support backup data from multiple sources, including a variety of backup applications and database dump utilities. Performing deduplication in the backup software limits the ability to have all data from all sources stored and deduplicated in a single target device. Unless 100% of your backup data passes through that particular backup application, a purpose-built disk-based backup appliance such as ExaGrid's is the best choice to meet the requirements of your entire environment

In contrast, backup application software solutions that have incorporated deduplication by definition only support their own backup application, with its own backup server software and its own backup client agents. These solutions are not able to support backup data from other backup applications or utilities. For example, if you have a physical environment, for backup applications that employ data dededuplication but want to use a separate utility for VMWare and also do direct database dumps, only the data running through the physical system's backup application will be deduplicated. Also, deduplication in backup software ranges from 2:1 to 10:1 and therefore uses a lot more disk and bandwidth as retention grows versus target-side deduplication appliances that employ far more aggressive deduplication algorythms, since with appliances, resources are dedicated to the task. Dedicated appliances achieve ratios on average of 20:1 using 1/3 to 1/2 the disk and bandwidth.

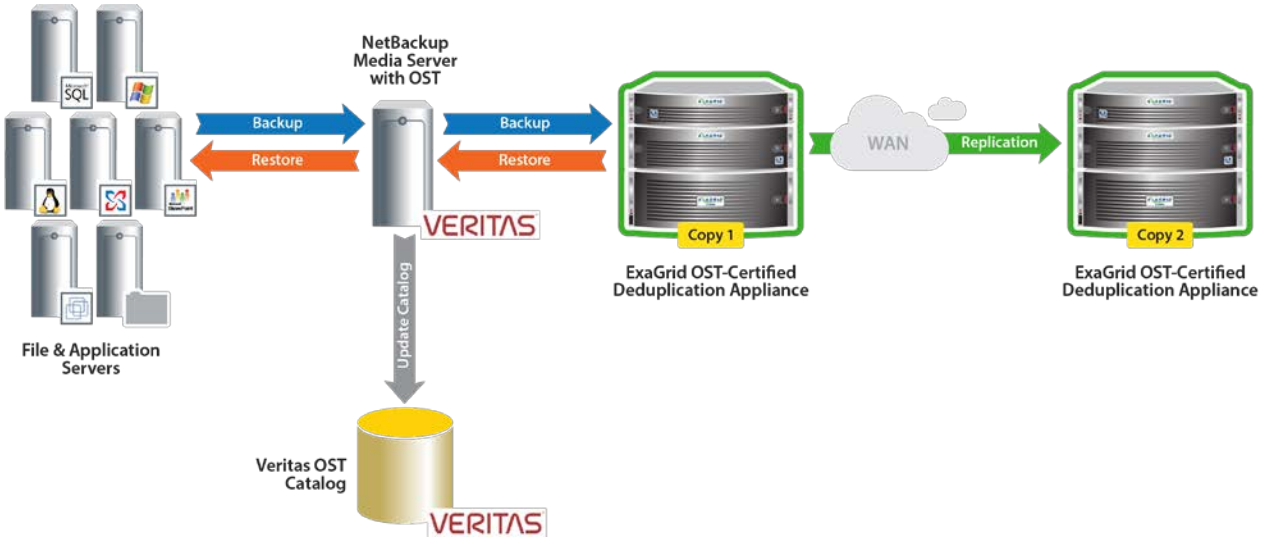


# Advanced Backup Features

## Veritas Open Storage Manager (OST)

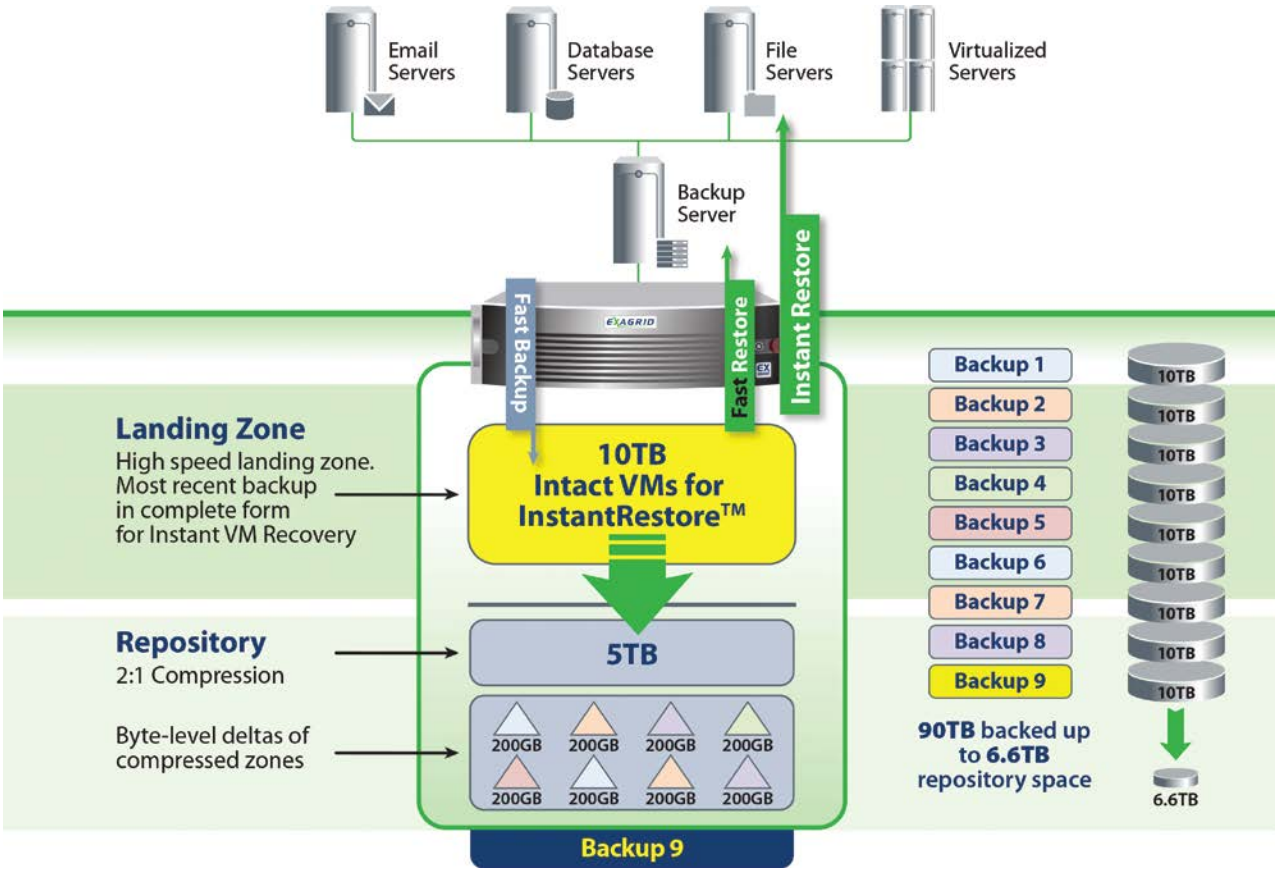
An important consideration when looking at disk-based backup solutions is how well a particular solution supports advanced backup application features such as Backup Exec GRT (Granular Restore) and Veritas' OST (Open Storage Technology) for Backup Exec and NetBackup. Some solutions do not integrate well with these features; poorly implemented GRT solutions, for example, may take hours to restore an individual e-mail or may not work at all.

Veritas' Open Storage is another popular feature that allows for more integrated offsite data protection, and it is important to check whether these features are supported if you are using Veritas NetBackup or Backup Exec.



### Veeam Accelerated Data Mover

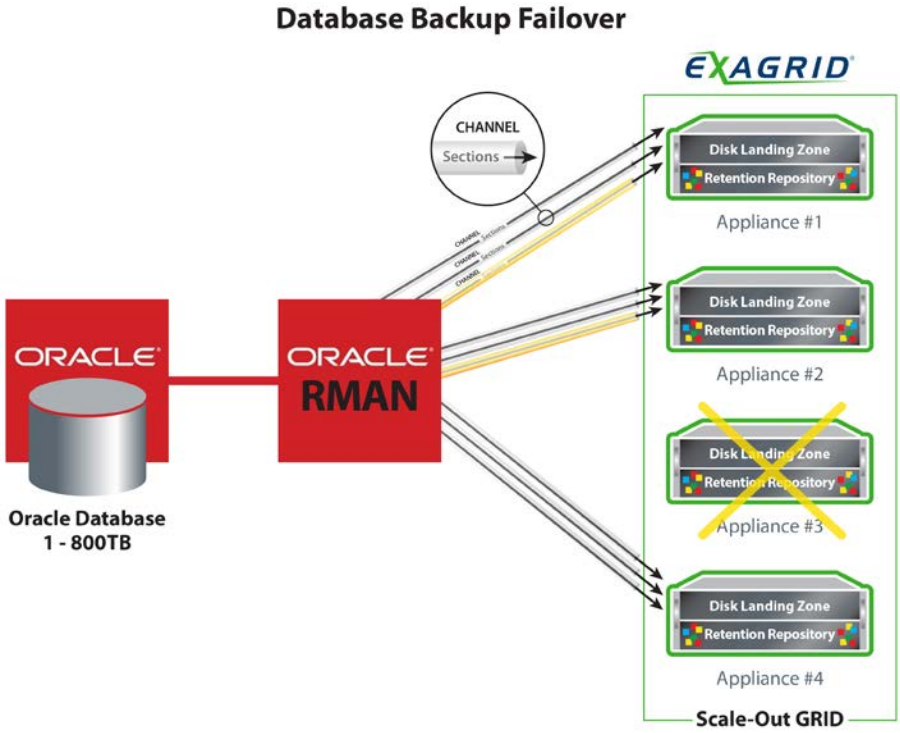
Most of Veeam's unique features such as Sure Backup, Virtual Lab, Instant VM Recovery, Copy and Replicate and other advanced features require an undeduplicated copy on disk. Only ExaGrid provides this with its unique landing zone. All other solutions only store deduplicated data. In addition, ExaGrid includes an integrated Veeam data mover with each appliance called the ExaGrid-Veeam Accelerated Data Mover. This improves all backup and restore processes and also allows a synthetic full to be created directly on the ExaGrid for increased performance.



## Oracle RMAN Channels

ExaGrid supports Oracle RMAN Channels targeted at multiple NAS shares across multiple appliances in a scale-out GRID. RMAN Channels automatically writes "sections" in parallel to all NAS shares and automatically redirects the next "section" based on available targets. RMAN Channels with ExaGrid has six major advantages.

1. Oracle database can be hundreds of terabytes in size and can be backed up in parallel to a single ExaGrid scale-out GRID.
2. The database backup performance is accelerated as the sections are backed up in parallel across multiple appliances in a scale-out GRID.
3. The database backup performance is maximized as each new section is automatically sent to the highest performance availability NAS shares and/or appliance, resulting in the best possible performance based on NAS share and appliance ingest availability.
4. If any appliance in a GRID fails, the segments are automatically redirected to the active appliance, providing for automatic failover.
5. The most recent database is stored in an undeduplicated form in the ExaGrid landing zone, allowing for fast restores while still allowing for storage efficiency as all long-term retention data is stored in deduplicated form. This avoids the lengthy data rehydration process of inline scale-up appliances that only store deduplicated data.
6. As the database data grows, the backup window stays fixed in length as full appliances are added into a scale-out GRID bringing compute with capacity. This eliminates the forklift upgrades associated with inline scale-up deduplication appliances.



## Data Security

The data security capabilities in the ExaGrid product line, including optional enterprise-class Self-Encrypting Drive (SED) technology, provide a high level of security for data at rest and can help reduce IT drive retirement costs in the data center. All data on the disk drive is encrypted automatically without any action required by users. Encryption and authentication keys are never accessible to outside systems where they can be stolen. Unlike software-based encryption methods, SEDs typically have a better throughput rate, particularly during extensive read operations. Optional data encryption at rest is available for the EX7000 models and above.

Data can be encrypted during replication between ExaGrid systems. Encryption occurs on the sending ExaGrid system, is encrypted as it traverses the WAN, and is decrypted at the target ExaGrid system. This eliminates the need for a VPN to perform encryption across the WAN.

### Reliability and Redundancy

Organizations using a disk-based backup appliance to hold their invaluable backup data should consider carefully how the appliance is architected for reliability and redundancy. Compromises in a product's architecture or implementation may reduce product cost, but those savings are quickly dwarfed by the risk and real cost to an organization of a loss of some or all backup data.

ExaGrid's architecture and implementation have multiple facets of reliability and redundancy, which allows organizations that are considering disk-based backup appliances to make informed vendor selections:

- RAID6 internal storage with consistency checking
- Flash-backed RAID cache
- Backup data checksums with automatic repair
- Deduplicated metadata transactional consistency
- Internal database backups and self-describing metadata
- Logging filesystem

## Offsite Data Protection for Disaster Recovery

While keeping offsite copies of backups has traditionally meant maintaining a set of tapes at an offsite location, companies using ExaGrid appliances can easily maintain offsite backups through the use of an offsite ExaGrid appliance in conjunction with a primary site ExaGrid appliance.

Backing up your data to an ExaGrid appliance at your primary site dramatically reduces the amount of disk space required to store all of that data due to its high-performance data deduplication capability. In a multi-site ExaGrid environment, the onsite ExaGrid system is only sending deduplicated data—the backup data bytes that change between each backup—over the wide area network (WAN) to the offsite ExaGrid appliance. The offsite ExaGrid appliance is ready for data restore and fast recovery in the event of a disaster or other primary site outage.

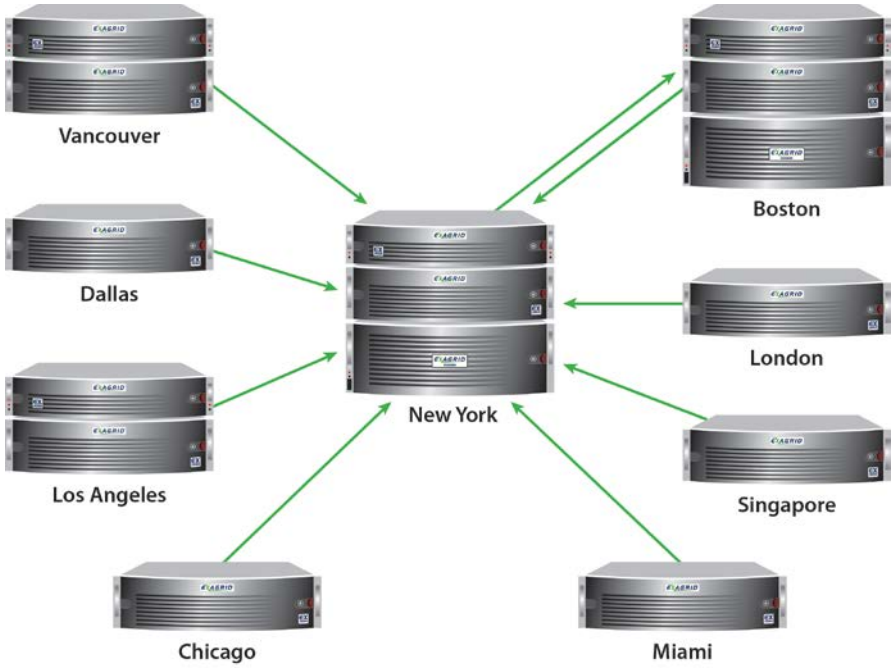
Replication between ExaGrid systems across a WAN can be scheduled for the day of the week and multiple times throughout each day. Each scheduled period allows for bandwidth throttling which limits replication to only use the assigned bandwidth. The combination of scheduling flexibility and bandwidth throttling allows for the maximum efficiency of WAN bandwidth used for replication.

Replicated data can be encrypted over the WAN using a cutomer's VPN or by utilizing the ExaGrid built-in replication encryption.


- Optimizes most recent backups for "Instant DR"
- Avoids delays from restoring from tape
- Restores faster from full copy vs. "rehydrating" data

ExaGrid supports three models for disaster recovery sites:

- *Unidirectional replication to offsite for disaster recovery*. In this use case, the entire offsite system can be configured for repository, allowing for a half-size system to be used offsite. ExaGrid is asymmetrical in this use case where all other solutions are symmetrical.
- *Cross protection*. In this use case, data is backed up at both the offsite and onsite systems and cross replicated such that each site becomes the disaster recovery site for the other.
- *Multiple data center sites*. ExaGrid can support up to 16 sites in a single star topology with 15 spokes to a hub. Full systems or individual shares can be cross replicated such that data center sites can serve as disaster recovery sites for each other.



## Total Cost of Ownership

Backup and disaster recovery is one area of IT spending which—though critically necessary—is typically viewed purely as cost. While backup is an extremely important area, organizations look to achieve appropriate protection so data is not lost while, at the same time, accomplishing this at the lowest possible cost. ExaGrid is the only vendor that has responded by creating a product that truly meets this different economic model warranted for backup spending. Backup spending has followed the same spending pattern as other IT infrastructure spending, which includes:

- Expensive forklift upgrades when a system is outgrown due to data growth
- Repurchasing of entire systems when an existing system simply "wears out"
- Complete rip-and-replace when a backup system becomes obsolete due to product end of life

ExaGrid redefines the economics of backup by helping you contain costs at every point in the life cycle—up front and as data grows over time.

### Cost Effectiveness Up Front

With ExaGrid disk-based backup appliances, backups are written directly to a disk landing zone, avoiding inline processing and ensuring the highest possible backup performance resulting in the shortest backup window. Adaptive Deduplication performs deduplication and replication in parallel with backups while providing full system resources to the backups for the shortest backup window. Available system cycles are utilized to perform deduplication and offsite replication for an optimal recovery point at the disaster recovery site. Once complete, the onsite data is protected and immediately available in its full undeduplicated form for fast restores, VM Instant Boot and Recoveries, and tape copies while the offsite data is ready for disaster recovery. This allows ExaGrid to be more cost effective up front than an inline/block deduplication system.

In contrast, other appliances that use inline, block-level deduplication rely on a top-tier processor with large amounts of very fast memory and disk just to keep up with backup data. The premium cost of these components means higher cost compared to ExaGrid appliances. In addition, since the inline scale-up appliances have a fixed and limited ingest, in order to increase ingest, they need to use compute from elsewhere in the environment. To this end, they deploy software on media servers and certain application servers in order to do some of the deduplication work away from the inline appliance. Not only is deploying agents frowned upon by most IT organizations, but even with this approach, the ExaGrid will still be three times the ingest performance, and with ExaGrid there are no agents to deploy or manage.

When comparing ExaGrid appliances to non-appliance-based deduplication systems, such as deduplication via the backup application software, it is important to keep in mind that using deduplication in the backup application software typically requires greater resources on the backup server—more processing power, more memory, and more disk. Software deduplication merely shifts the backup performance bottleneck to the media server. Using data deduplication in the backup software uses more disk and bandwidth over time and does not allow for backup environment flexibility such as using a separate utility for virtualized backup, direct TAR backups, and direct database dumps such as SQL dumps or Oracle RMAN dumps.

### Cost Effectiveness as Data Grows

ExaGrid's landing zone with unique architecture—full appliances in a GRID architecture is the most cost-effective to scale as data grows. Each ExaGrid appliance added to the GRID includes a full server with additional processor, memory, bandwidth, and disk resources. Total backup capacity keeps pace with continued data growth over time by simply adding ExaGrid appliances to the GRID. There are no forklift upgrades and no additional future costs to consider.

Other appliances that use inline, block-level deduplication do not support a GRID architecture and are therefore more costly to scale. Instead of adding capacity by adding full servers, only disk shelves are added over time as data grows. But, at some point, the single front-end controller becomes a bottleneck due to its fixed processor, memory and bandwidth resources and can no longer handle the backup load. Eventually, the entire front-end server must be replaced with the next higher capacity unit in a "forklift upgrade." In fact, you may have to spend as much for the front-end controller upgrade as you originally spent on the original system, including disk shelves. In addition, all data is always deduplicated. For each restore, recovery, and copy request, the data has to be put back together, or "rehydrated," which can take hours to days.

In addition, unlike other appliances that go "end-of-life" in as little as 18 months and are incompatible with newer models from the same vendor, ExaGrid's GRID archictecture allows you to "mix and match" different capacities and generations of appliances within a GRID. Only ExaGrid protects your backup investment from obsolescence.

# Summary

When organizations evaluate the backup solution and company that can best meet their backup needs and address their challenges, more and more IT organizations are finding that ExaGrid offers not only the fastest backup and restore performance and best scalability but also a total cost of ownership that's 25-30% lower than competitive offerings.

Only with ExaGrid's full-servers-in-a-GRID architecture and adaptive deduplication do you get:

**1 Shortest Backup Window**

Back up straight to disk for high performance.

- All other solutions perform inline deduplication, slowing the backups down.
- All other solutions slow further if you additionally turn on replication and/or encryption as it takes resources from the processor performing the deduplication.
- Each appliance has additional bandwidth so that as data grows, so does the throughput.
- All other solutions have a front-end controller with fixed bandwidth up front and over time.

**2 Fixed Length Backup Window**

Each appliance has capacity and also compute to handle the increased deduplication load.

- All other solutions only add capacity and have fixed compute up front and over time.
- Each appliance has additional bandwidth so that as data grows, so does the required network bandwidth.
- All other solutions have a front-end controller with fixed bandwidth up front and over time.

**3 No Forklift Upgrades**

Capacity is always accompanied by compute to bring additional resources as data grows.

- All other solutions have a front-end controller with fixed compute and only add capacity.

**4 Pay as You Grow**

Up to 25 appliances can be mixed and matched in a single GRID system – buy what you need as you need it.

- All other solutions have a front-end controller – either overbuy up front or require a forklift upgrade later.

**5 No Obsolescence**

Appliances of all sizes and all ages can be used in a single GRID. Newer appliances work in the same GRID as older appliances.

- All other solutions have countless models over the years, leaving customers hanging.

**6 Fast Onsite Restores**

Unique landing zone keeps the most recent backups in their full form for fast and immediate restores.

- All other solutions only store deduplicated/dehydrated data that needs to be reassembled for every restore.

**7 Fast Offsite Tape Copy**

Unique landing zone keeps the most recent backups in their full form for fast and immediate tape copy.

- All other solutions only store deduplicated/dehydrated data that needs to be reassembled for every tape copy.

**8 Fastest Onsite Instant Recovery**

Unique landing zone keeps the most recent backups in their full form in order to boot a VM directly off the ExaGrid landing zone for Instant Recovery.

- All other solutions require a lengthier process to rehydrate the offsite disaster recovery (DR) data.

**9 Fastest Disaster Recovery**

Offsite data is stored in a unique format optimized for fast DR performance.

- All other solutions require a lengthier process to rehydrate the offsite DR data.

**10 Asymmetrical Onsite and Offsite Systems**

If the replication is one way to a DR site and there are no backups at the DR site, the ExaGrid system can be tuned to 100% repository, allowing for a half-size system at the DR site which saves valuable budget dollars.

- All other solutions require the exact same size system at both sites.

**11 Distributed Architecture**

There are multiple appliances in a GRID each with processor, memory, bandwidth, and disk. If an appliance fails, backup jobs targeted at the remaining appliances in the GRID will continue to run.

- All other solutions will be 100% inoperable as there is a single head-end controller, and if the head-end controller goes down, the entire system is unavailable.

## 12 Supports Heterogeneous Environments

Multiple backup applications, utilities, and dumps can be sent to a single ExaGrid. A customer can have one backup application for their physical environment, a separate backup application for their VM environment, and also do direct database dumps such as SQL dump or Oracle RMAN.

# About ExaGrid

Organizations come to us because we are the only company that implemented deduplication in a way that fixed all the challenges of backup storage. ExaGrid's unique landing zone and scale-out architecture provides the fastest backup — resulting in the shortest fixed backup window, the fastest local restores, fastest offsite tape copies and instant VM recoveries while permanently fixing the backup window length, all with reduced cost up front and over time. Learn how to take the stress out of backup at www.exagrid.com or connect with us on LinkedIn. Read how ExaGrid customers fixed their backup forever.

United States: 2000 West Park Drive | Westborough, MA 01581 | (800) 868-6985
United Kingdom: 200 Brook Drive | Green Park, Reading, Berkshire RG2 6UB | +44 (0) 1189 497 051
Singapore: 1 Raffles Place, #20-61 | One Raffles Place Tower 2 | 048616 | +65 6285 0302