



“The fact that Veeam’s data mover is integrated with the ExaGrid appliance is a massive benefit. It enables us to achieve our Gold-level SLA without any complexities or need to constantly move back and forth, thanks to ExaGrid’s Landing Zone and approach to deduplication. Veeam and ExaGrid working together in this way really inspires confidence.”

Joep Piscaer
Chief Technology Officer
Operator Groep Delft
The Netherlands



“We take advantage of Veeam’s source-side deduplication as well as deduplication on the ExaGrid. Veeam first deduplicates its backups to minimize the amount of data written across the network to the ExaGrid. Once the deduplicated Veeam data lands on the ExaGrid system, the ExaGrid deduplicates it further.”

Michael DiLalla
Sr. Network Technician
Penfield Central School District
Rochester, NY

ExaGrid and Veeam Deduplication for Virtualized Environments

Background

Deduplication technology is helping organizations cope with the extraordinary growth in data protection requirements and the steady march of virtualization. For backup, deduplication only stores unique blocks or bytes and, as a result, stores a lot less data than the primary data itself. The approach of only storing unique blocks or bytes greatly reduces the amount of disk storage required.

Deduplication can be accomplished in the backup application, at a target-side appliance, or both. Backup environments could see deduplication applied by the multiple components that make up today’s modern backup solutions. Backup applications can do an initial level of deduplication and the target-side disk-based backup appliance will do further deduplication.

When managing and assessing these backup solutions, you can no longer ask, “What is my deduplication ratio?” Instead, you need to ask, “What is the overall deduplication ratio achieved across the backup environment?”

How it Works

Hypervisors like VMware and Microsoft Hyper-V are able to track what changes are made to the datastores underlying the VM virtual disk; backup applications harness this facility to put only those changed blocks into the backup data, called “changed block tracking.” This is an excellent source of deduplication, since the overall footprint of the backup data is reduced by not having to include unchanged blocks.

Veeam Backup & Replication™ uses the information from VMWare and Hyper-V and provides deduplication on a “per-job” basis, finding the matching areas of all the virtual disks within a backup job and using metadata to reduce the overall footprint of the backup data. Veeam also has a “dedupe friendly” compression setting which further reduces the size of the Veeam backups in a way that allows the target-side disk-based appliance to achieve further deduplication. This approach typically sees a 2:1 deduplication ratio.

ExaGrid is architected from the ground up to protect virtualized environments and provide deduplication as backups are taken. ExaGrid will achieve a 3:1 up to 5:1 additional deduplication rate. The net result is a combined Veeam and ExaGrid deduplication rate of 6:1 upwards to 10:1, which greatly reduces the amount of disk storage required.

Veeam Example

Veeam Backup & Replication is protecting 80TB of virtual disks spread out across 40 virtual machines, divided into eight Veeam jobs. Veeam is configured to retain 14 recovery points, covering two weeks of daily backups.

The Veeam backup jobs have an “Advanced Storage” configuration of:

- Enable inline data deduplication (recommended)
- Compression level set to “dedupe-friendly”
- Optimize for “LAN target” (block size best for NAS targets like ExaGrid)

Each recovery point is fully recoverable by Veeam, effectively a full backup. This is equivalent to having 14 full backups, which would consume as much as 1.1PB of storage if no deduplication of any kind was used.

The hypervisor's changed block tracking reduces the 7 daily backups to approximately 8TB (10:1), and Veeam deduplication reduces this further by 2:1. The footprint of the Veeam backups is:
 $14 \times 8\text{TB (daily incrementals)} + 2 \times 80\text{ TB (synthetic) fulls} = 272\text{ TB @ }2:1 = 136\text{ TB.}$

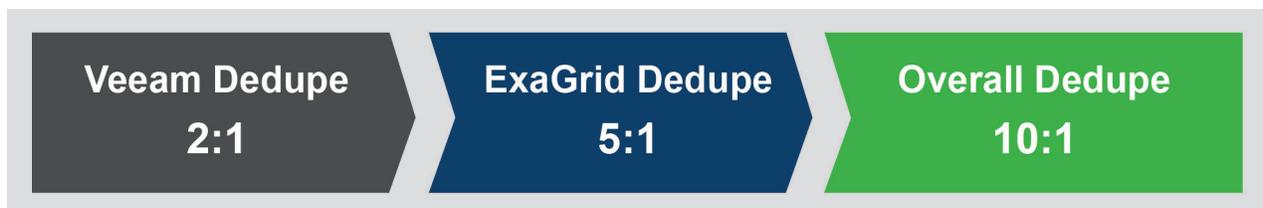
You can expect the deduplication achieved by Veeam to be about 2:1.

You can expect the deduplication achieved by the ExaGrid appliance to range from 3:1 to 5:1 so, for example, the ExaGrid consumption is: $136\text{ TB @ }5:1 = \text{only } 27.2\text{TB}$ of disk consumption on the ExaGrid appliance to fully protect this virtualized environment.

At the 10,000 foot view, 1.1PB of full VM backups requires only 27.2TB, which is a deduplication ratio of 41:1.

At the 5,000 foot view, Veeam and ExaGrid together take 272TB coming from the hypervisor down to only 27.2TB disk consumption on the ExaGrid, which is a deduplication ratio of 10:1.

The 2:1 dedupe achieved by Veeam is multiplied by as much as 5:1 achieved by the ExaGrid to produce a net overall dedupe of 10:1.



Summary

When sizing ExaGrid disk-based backup appliances for virtualized environments using Veeam Backup & Recovery, the deduplication achieved by both products is multiplied together and will generally be in the 6:1 to 10:1 range.