

1 NetApp Data Reduction Technologies: Deduplication, Compression, and Compaction

Storage efficiency enables you to store the maximum amount of data in the smallest possible space and at the lowest cost. NetApp storage efficiency technologies are key to achieving data consolidation and managing future data growth while saving time and money.

Data storage is the largest and fastest growing IT expense incurred by businesses today. The NetApp® portfolio of storage efficiency technologies provides the ability to store the maximum amount of data for the lowest possible cost.

NetApp deduplication, data compression, and data compaction are the key components to delivering superior performance, scale, resource efficiency, and portability across multiple applications and storage tiers. No additional purchases are required to take advantage of the benefits that NetApp technologies deliver. They are:

- Free, no licence is required to enable and use deduplication
- Transparent to applications
- Lossless/zero data loss
- Supported both as inline operations and post-process operation

NetApp Deduplication

With NetApp deduplication, you can store just one copy of each unique data object, which reduces capacity requirements substantially. Deduplication automatically removes duplicate data blocks on a 4KB level across an entire volume and across volumes within an aggregate, reclaiming wasted storage to achieve significant space savings.

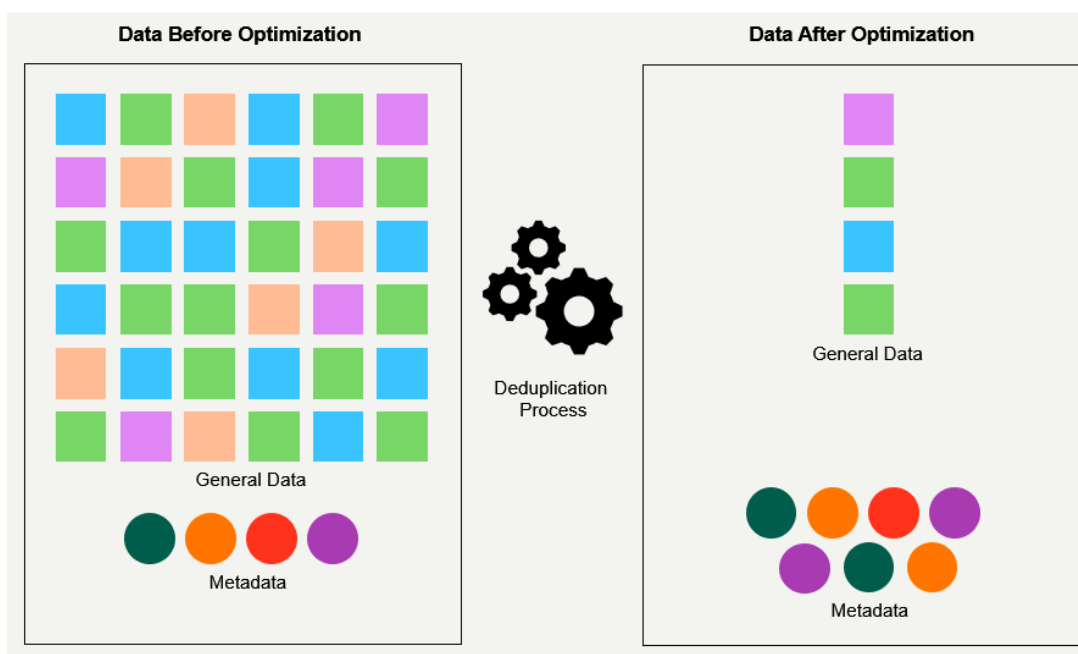


Figure 1: NetApp deduplication removes duplicate data blocks.

NetApp deduplication can be implemented across a wide variety of applications and file types, including:

- Primary data volumes
- Data backup
- Data archiving

NetApp deduplication can help you reclaim up to 95% of storage space, depending on application and file type.

There are two main deduplication methods: In-line and post-process deduplication. Inline deduplication provides immediate space savings; post-process deduplication first writes the blocks to disk and then dedupes the data at a scheduled time.

- **Inline deduplication:** For copy offload and VDI deployments. It's especially useful for VM patch apply use cases, VM provisioning, and database migrate operations that result in many duplicate blocks in memory that can be eliminated by inline dedupe.
- **Post-process deduplication:** When inline deduplication is not preferable; for example, for entry-level systems in which memory is limited to meeting client I/O SLAs. In general, NetApp does not recommend deduplication for use cases in which data is overwritten at a rapid rate.

NetApp Data Compression

NetApp data compression is a software-based solution that provides transparent data compression. It gives the ability to store more data in less space. You can use data compression to reduce the time and bandwidth required to replicate data during volume SnapMirror® transfers. Data compression saves space on regular files or LUNs.

Data compression works by compressing a small group of consecutive blocks known as a compression group. It can be run inline or postprocess and includes the capability to compress existing data. No application changes are required to use NetApp data compression. This process is enabled and managed by using a simple CLI or GUI such as System Manager or NetApp Active IQ Unified Manager.

As with our deduplication process, there are two main data compression methods: In-line and post-process compression. Inline compression provides immediate space savings; post-process compression first writes the blocks to disk as uncompressed and then at a scheduled time compresses the data.

- **Inline compression:** For customers who are not as performance sensitive and can handle some impact on new write performance as well as on CPU during peak hours. This type of compression is also useful for customers who want to minimise \$/gigabyte and the PE cycle of SSD drives to prolong the life of flash media. Common use cases under this category are primary and secondary workloads on AFF and Flash Pool configurations and secondary workloads on all HDD configurations.
- **Post-process compression:** For environments in which you want compression savings but don't want to incur a performance penalty associated with new writes. Common use cases under this category are primary workloads on all HDD configurations.

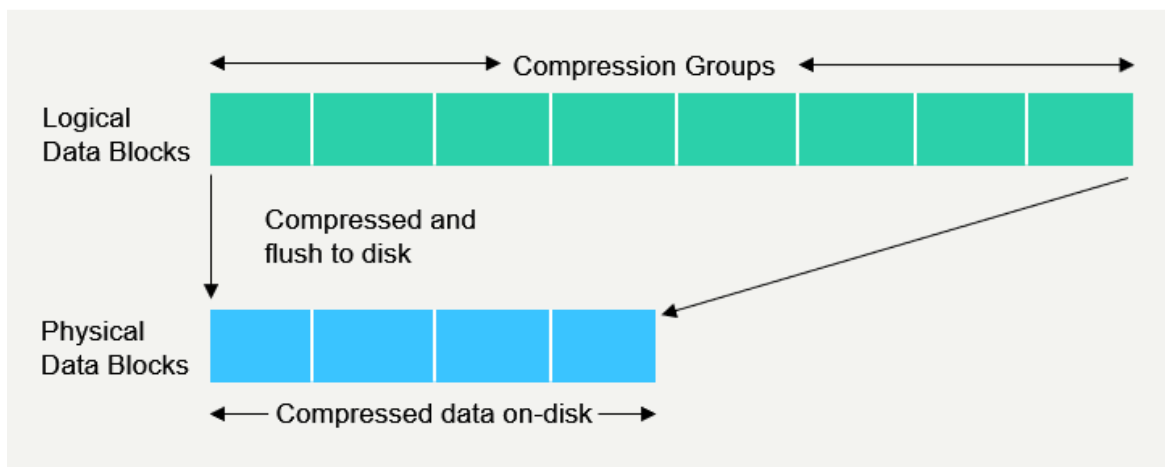


Figure 2: Compression groups are tested for compressibility before any compression takes place – They are then flushed to disk, compressed or uncompressed, depending on the results of the test.

NetApp Data Compaction

Compaction gives you the ability to further reduce the physical used space needed to store data. It's an inline operation and occurs after inline compression and inline deduplication. Data compaction happens on logical blocks as they are organised before being written to storage. It takes I/Os that normally consume a 4K block each on physical storage, and packs multiple I/Os into one physical 4K block.

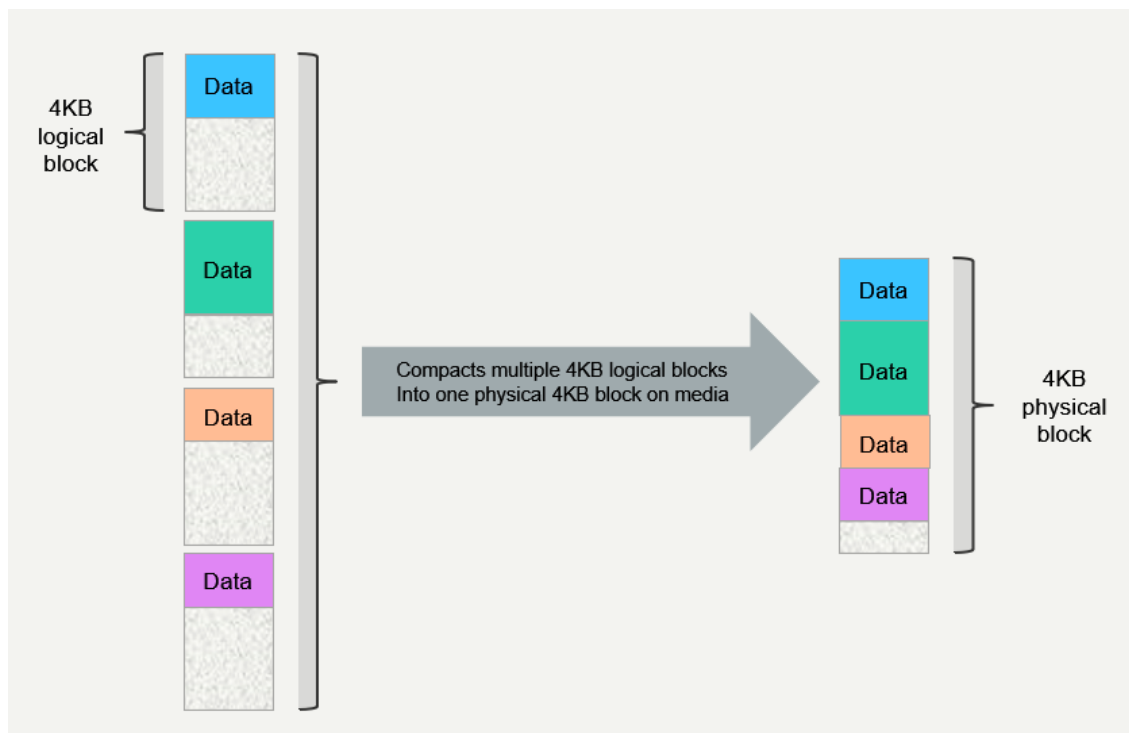


Figure 3: How data compaction works.

Compaction is enabled by default for NetApp All Flash FAS systems. It is an optional feature that can be turned on for FAS systems, with either HDD-only aggregates or NetApp Flash Pool™ aggregates. Compaction is a significant addition to our storage efficiency portfolio and complements NetApp's superior deduplication and compression technologies.

Space Savings

The table below lists the storage efficiency data reduction ratio ranges for different applications. A combination of synthetic datasets and real-world datasets has been used to determine the typical savings ratio range. The savings ratio range mentioned are only indicative.

Table 1: Typical savings ratios with ONTAP 9—Sample savings achieved with internal and customer testing¹.

Typical Savings Ratios with ONTAP 9	
Workload [with deduplication, data compaction, adaptive compression and FlexClone volumes (where applicable) technologies]	Ratio Range
Home directories	1.5:1.-2:1
Software development	2:1 - 10:1
VDI VMware Horizon full clone desktops (persistent) – NetApp Clones	6:1 - 10:1
VDI VMware Horizon linked clone desktops (nonpersistent)	5:1 - 7:1
VDI Citrix XenDesktop full clone desktops (persistent) – NetApp Clones	6:1 - 10:1
VDI Citrix XenDesktop MCS desktops (nonpersistent)	5:1 - 7:1
VDI Citrix Provisioning services desktops (nonpersistent)	3.3:1 - 5:1
Virtual Servers (OS and Applications)	2:1.-4:1
Oracle databases (with no database compression)	2.1 - 4:1
SQL 2014 databases (with no database compression)	2.1 - 4:1
Microsoft Exchange	1.6:1
Mongo DB	1.3:1 - 1.5:1
Precompressed data (such as video and image files, audio files, pdfs, etc.)	No Savings

Talk to your account representative on how to estimate savings for the data residing on non-NetApp systems.

¹ Actual customer savings will depend on data type and data layout.