



QSM Deduplication White Paper

XCubeNAS & XCubeNXT series



QSAN Technology, Inc.
www.QSAN.com



Copyright

© Copyright 2021 QSAN Technology, Inc. All rights reserved. No part of this document may be reproduced or transmitted without written permission from QSAN Technology, Inc.

July 2021

This edition applies to QSAN Unified Storage series. QSAN believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

Trademarks

All products and trade names used in this manual are trademarks or registered trademarks of their respective companies.

QSAN Technology, Inc

4F, No. 103, Ruei Hu, St. Neihu Dist, Taipei 114 Taiwan, ROC

Tel ; +886-2-7720-2118

Fax : +886-2-7720-0295

Email :

sales@qsantechology.com

Website :

www.qsan.com

Notices

This Unified Storage series white paper is applicable to all Unified Storage models and QSM version 3.3.4 or above.

Information contained in document has been reviewed for accuracy. But it could include typographical errors or technical inaccuracies. Changes are made to the document periodically. These changes will be incorporated in new editions of the publication. QSAN may make improvements or changes in the products. All features, functionality, and product specifications are subject to change without prior notice or obligation. All statements, information, and recommendations in this document do not constitute a warranty of any kind, express or implied.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products.

All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

Table of Contents

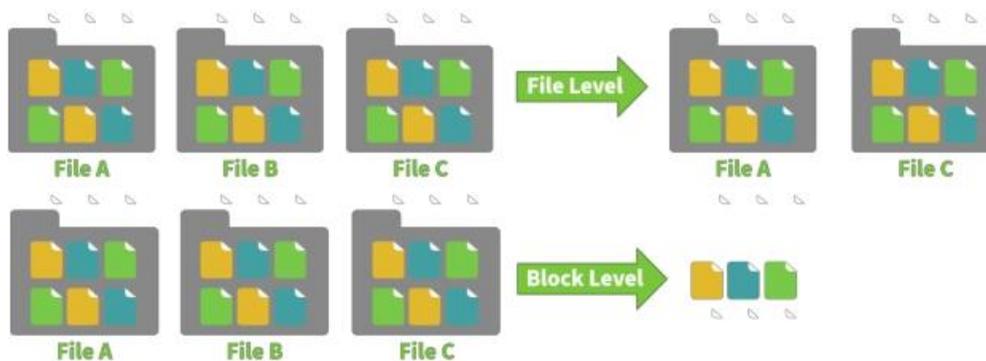
Notices	i
QSAN Deduplication Application Note	3
QSAN Deduplication technology	3
Prerequisites	4
How to set up your deduplication	5
Data Deduplication Ratio	6
Data Deduplication Efficiency	8
Deduplication Best practice.....	9
Deduplication Use case	9
Summary	11
Appendix	12
Related Documents	12
Technical Support.....	12

QSAN Deduplication Application Note

QSAN Deduplication technology

QSAN deduplication technology provides inline, block-level function that checks the block similarity of data as it enters the system. Deduplication will automatically remove the redundant data object to reduce the usage of storage capacity.

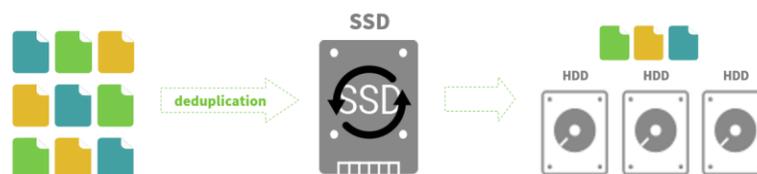
Block level deduplication is generally more efficient than file level one. Inline means that as the data is copying into the storage, the deduplication is processing the data in real time.



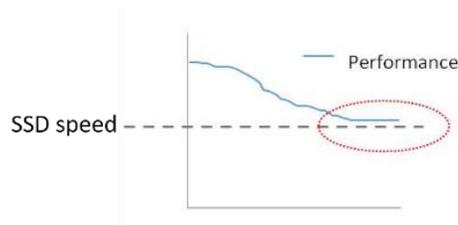
The space that is saved by deduplication will NOT be reflected in the share or iSCSI volume directly. This saved space will be accumulated in the free space of the storage pool. So you can use these “growing” space for other applications.

The greater the data redundancy is, the more the deduplication can help. This paper helps you estimate how much you can save from using Qsan deduplication technology.

The integration of QSAN deduplication and hybrid SSD cache transfers the computing tasks concerning deduplication to SSD.



In this way, the performance of deduplication relates itself to the drives. On QSAN storage, you may adopt NVMe SSD to speed up deduplication.



Deduplication is ideal for highly redundant operations like backup, which requires repeatedly copying and storing the same data set multiple times. As a result, enterprises of all sizes rely on deduplication to reach a cost-effective backup solution.

Prerequisites

To apply deduplication, we recommend to setup your system in the configuration below.

Software

This deduplication is applicable on QSM 3.3.4 or later.

Hardware

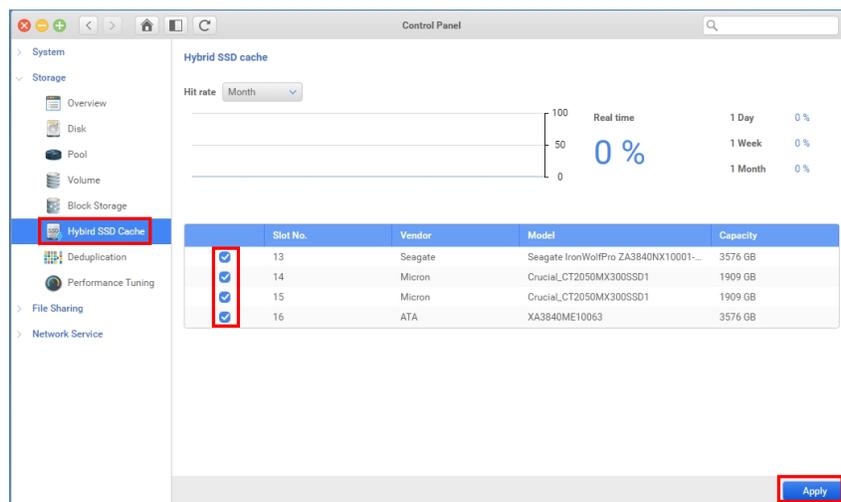
1. Deduplication must run with **2 SSD** or more, due to the deduplication data is very important and need to be mirrored.
2. The minimum SSD capacity to enable deduplication is **512Gb**, because the capacity need to leverage to both read/write cache as well.
3. When replacing SSD for deduplication, the new SSD must have same capacity or above with the original one. Due to the deduplication data is mirrored, so the capacity can not become lesser.

How to set up your deduplication

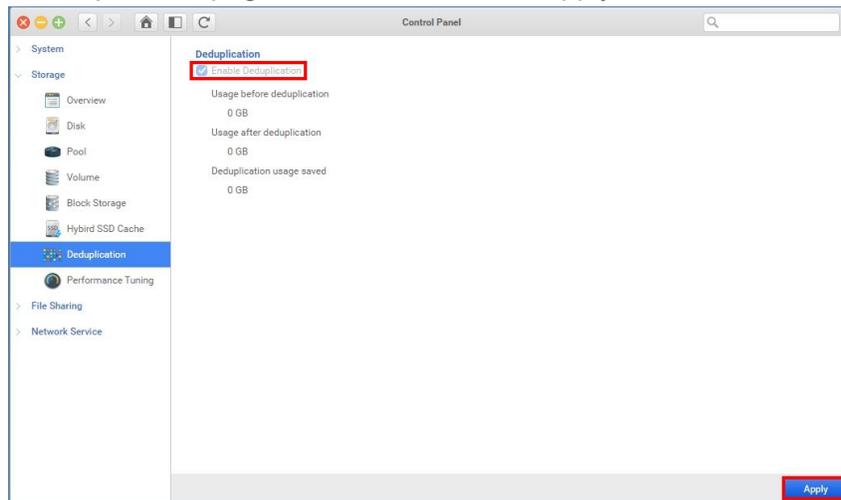
Qsan deduplication technology is based on hybrid SSD cache. It can improve the storage efficiency without impacting performance.

Follow the below steps to apply deduplication for your QSAN storage.

1. Go to Hybrid SSD cache page, select at least 2 SSD and apply.



2. Go to Deduplication page, clicc checkbox and apply.



You can check the below video to see full steps instruction of deduplication on QSM.

[Deduplication operation process](#)

Data Deduplication Ratio

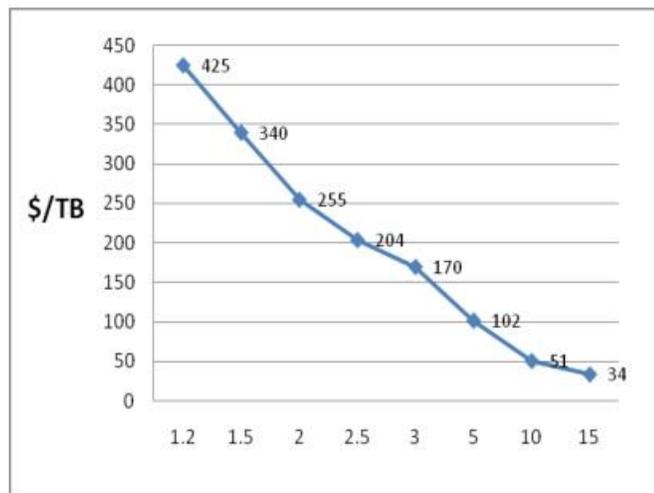
A data deduplication ratio is the measurement of data's original size versus the data's size after removing redundancy. Different user data or application data have different degrees of redundancy in them. The higher the redundancy, the more you can benefit by using deduplication function.

Dedup ratio is defined as total data over unique data (unique data means data being deduplicated.). As a result, 2 copies of the same data will give you a dedup ratio of 2 (2:1). 10 copies of the same data will yield a dedup ratio of 10 (10:1).

Think of this the other way around. Without deduplication function, higher dedup ratio means extra storage capacity you need to set aside to store these redundant data. This equals to extra money you need to spend on purchasing new drives to keep these redundant data. Therefore, the price per TB relation with and without deduplication is pretty simple.

$$\$/TB(\text{with deduplication}) = \frac{\$/TB(\text{without deduplication})}{\text{dedup ratio}}$$

Below is a rough estimation chart using market price of enterprise SATA and SSD drive.



As the deduplication ratio goes up, the dedup will generate less benefit. For example, 100:1 ratio remove 99% of the data. When it increases to 500:1, which only remove 99.8% of data, it's because most of the redundancy has already been removed.

In another word, the cost you can save by using deduplication becomes more and more apparent. If there is high redundancy in your application data, enabling deduplication with Qsan unified storage is a great solution to lower your Total Cost of Ownership and manage the data redundancy more efficiently.

In virtual machine environment where you need to store multiple VMs with similar settings, deduplication can remove those redundancy and save plenty of storage capacity for you. Another case would be the scheduled full backup instead of snapshot backup on a regular basis. In between the backup period, the data may not change much, which means a higher rate of redundancy. These are good examples where deduplication can help and make a difference.

Data Deduplication Efficiency

QSM deduplication can greatly improve data storage efficiency. However, the actual usage saved is based on the data itself. As mentioned above, deduplication can check possible usage savings and provide on a given data set.

Virtual machine files usually contain duplicate data and rarely modified. Deduplication on virtual machines of similar OS Images files can greatly reduce the storage usage. However, the possibility of performance degradation due to the calculation of data duplicate should be carefully considered first.

Below are some examples of data redundancy that can be achieved with QSM deduplication.

Application types	Data redundancy
Data archive	85~95%
Database backup	45~70%
Email archive	20~40%

Note: These capacity savings values are provided solely as rough guidance. Since no two data sets are alike (unless they're replicated), actual results can vary considerably from these examples.

Deduplication Best practice

QSAN recommends the following QSM deduplication best practices. Please note that some of this information may be covered elsewhere in this paper.

- Deduplication is most effective when applied to data sets with a low rate of change – for example, archived data.
- If your system is very far away and is hard to replace the hardware when any failure happen, we recommend using 3 SSD to have more availability.
- If your system is easy to do the hardware replacement, 2 SSD is recommended.
- To fit in most condition and for future expansion, QSAN recommend to use **960Gb** to enable deduplication to fullfill 90% enviroment.
- To choose the SSD size is according to the size of your system. QSAN recommends that if there is 100TB capacity, it should be setup with a 1TB SSD. If the space is larger, please use this base to caculate.

Deduplication Use case

Enterprise's data usually contains substantial quantities of redundant information. And home directories, file shares and data archives are consistently saved deduplicate data. Each time a document or email attachment is saved by multiple users, the same file is stored in multiple times, taking up valuable disk capacity. QSM Deduplication can be good used in the following ways:

Example A: File Sharing

By designing and configuring home directories and file-sharing repositories under a unified top-level directory, organizations can easily and effectively configure and run deduplication for these data sets.

From a performance perspective, file sharing are usually mid-tier workloads, usually involving concurrent access and a reasonable balance between read and write and data and metadata operations. Therefore, they are very suitable for QSM deduplication.

Example B: Storage Efficient Archiving

QSM deduplication is an ideal solution for large content repositories that are not frequently accessed. Examples include digital asset management workloads, seismic data archiving for energy exploration, document management repositories for legal discoveries, and compliance archiving for financial or medical records.

These are excellent use cases for data deduplication, because performance requirements are generally low, metadata operations are biased, and there is often a lot of duplicate data. In this way, in order to improve data efficiency, trading system resources will bring significant and substantial benefits to the bottom line. QSM deduplication is also ideal for immutable archives and offers a great storage efficiency.

Example C: Disaster Recovery Target

For performance-oriented environments that do not want to run deduplication on the primary data set, the typical approach is to deduplicate read-only data copies on the target cluster or disaster recovery (DR) cluster.

Summary

XCubeNXT and XCubeNAS unified storage systems provides block-level deduplication because this is the finest granularity that makes sense for a general- purpose storage system. Block-level dedup also maps naturally to block checksums, which provide unique block signatures for all blocks in a storage pool.

Also, SSD cache maximum deduplication limit and restore performance drop that causes by enabling deduplication. You can enjoy both cost TCO reduction and improved performance at the same time.

When running a full scale project, both deduplication and compression can save substantial space depending on the applications and the data types. Applications such as full backup, server virtualization, or Exchange server database will be suitable to use deduplication.

Appendix

Related Documents

There are related documents which can be downloaded from the website.

- [All Unified Storage Documents](#)
- [Unified Storage QIG \(Quick Installation Guide\)](#)
- [Unified Storage Hardware Manual](#)
- [Unified Storage Configuration Worksheet](#)
- [QSM Software Manual](#)
- [Compatibility Matrix](#)
- [White Papers](#)
- [Application Notes](#)

Technical Support

Do you have any questions or need help trouble-shooting a problem? Please contact QSAN Support, we will reply to you as soon as possible.

- Via the Web: <https://qsan.com/support>
- Via Telephone: +886-2-7720-2118 extension 136
(Service hours: 09:30 - 18:00, Monday - Friday, UTC+8)
- Via Skype Chat, Skype ID: qsan.support
(Service hours: 09:30 - 02:00, Monday - Friday, UTC+8, Summer time: 09:30 - 01:00)
- Via Email: support@qsan.com