



Deduplication In Qsan Unified Storage

XCubeNAS& XCubeNXT Series

Copyright

Copyright©2004~2020, Qsan Technology, Inc. All rights reserved. No part of this document may be reproduced or transmitted without written permission from Qsan Technology, Inc.

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, Rue
i Hu, St. Neihu
Dist, Taipei
114 Taiwan,
R.O.C.

Tel: +886-2-7720-2118

Fax: +886-2-7720-0295

Email:

sales@QsanTechnology.com

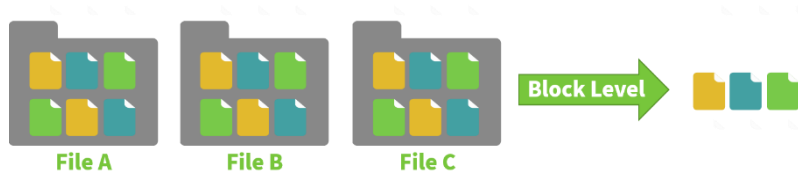
Website:

www.QsanTechnology.com

Qsan Deduplication technology

Qsan deduplication technology bases on ZFS and 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.





From market research data, different application environments have different degrees of data redundancy. The table below shows some well-known cases.

Data types	Data redundancy (space saving)
Backup data (data archive)	85~95%
VMware VMs	50~90%
Database backup	45~70%
Home directories	20~50%
Email archive	20~40%

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.

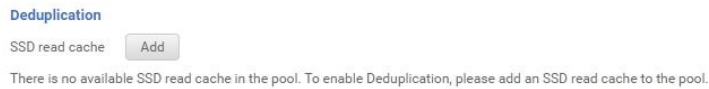
How to set up a deduplication pool?

Qsan deduplication technology is enabled on a per file-system and volume basis. It can be turned on and off on the fly depending on user’s needs. First, go to control panel > Storage > Deduplication

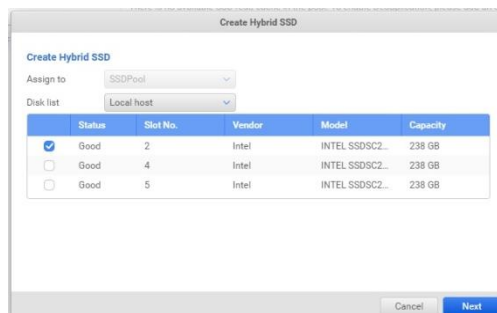
1. Select a SSD Pool, If do not have one, Click Add to create.

*note : Deduplication can only apply on a SSD pool.

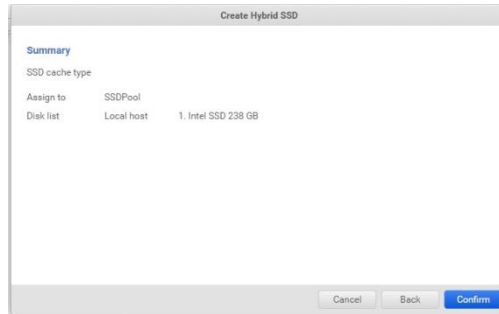
- a. Create SSD pool, Click Add.



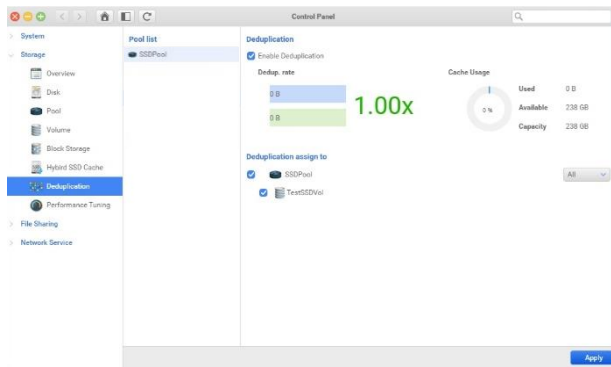
- b. Select SSD for your pool, then click next.



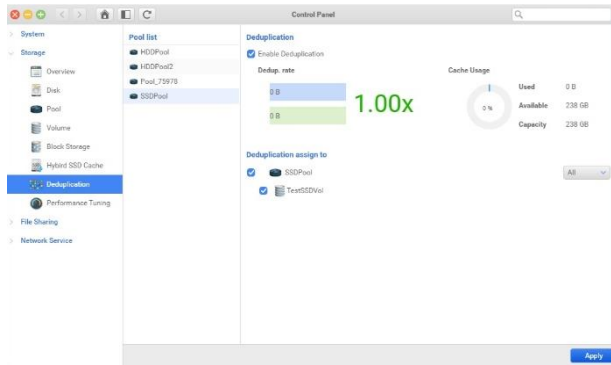
c. Confirm your pool, then click confirm.



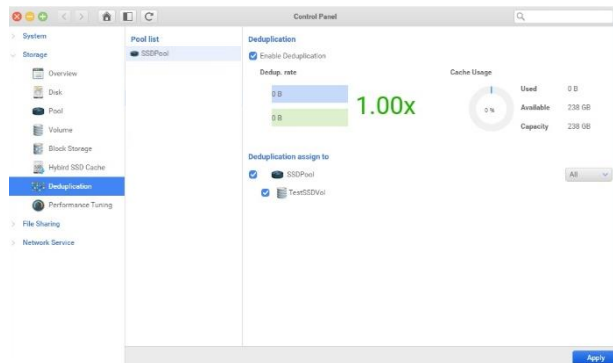
2. Click Enable deduplication checkbox to turn on/off deduplication.



3. Click Pool/Volume/LUN to assign deduplication.



4. When finish setting, click Apply to apply the change.



Data Deduplication Ratio

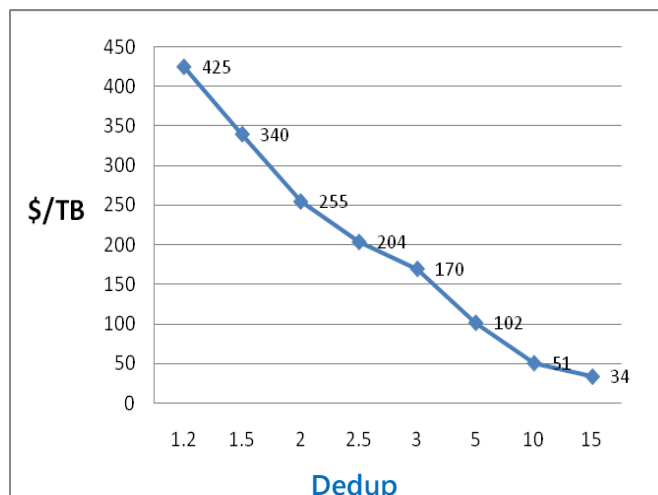
A data deduplication ratio is the measurement of data's original size versus the data's size after removing redundancy. Different user data or application data have different degrees of redundancy in them. The higher the redundancy, the more you can benefit by using deduplication function.

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.

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 ZFS's 256-bit 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.