7.5.7.4 DEDUPLICATION BEST PRACTICES

From *Acronis Backup Advanced 11.5 User Guide*
http://www.acronis.com/en-us/download/docs/aba11.5/userguide

Deduplication is a complex process that depends on many factors.
The most important factors that influence deduplication speed are:

- The speed of access to the deduplication database
- The RAM capacity of the storage node
- The number of deduplicating vaults created on the storage node.

To increase deduplication performance, follow the recommendations below.

Place the deduplication database and deduplicating vault on separate physical devices

To increase the speed of access to a deduplication database, the database and the vault must be located on separate physical devices.

It is best to allocate dedicated devices for the vault and the database. If this is not possible, at least do not place a vault or database on the same disk with the operating system. The reason is that the operating system performs a large number of hard disk read/write operations, which significantly slows down the deduplication.

Selecting a disk for a deduplication database

- The database must reside on a fixed drive. Please do not try to place the deduplication database on external detachable drives.

- Minimal disk access time is important. The best hard drive choice is an enterprise-grade Solid-State Drive (SSD), though using a fast IDE hard drive (7200 RPM or faster) or a SCSI drive is acceptable.

- The volume to store the deduplication database should have at least 10 GB of free space. When backing up a large number of machines, the required free space may exceed 10 GB.

- The disk space required for a deduplication database can be estimated by using the following formula:

$$S = U / 32 + 10$$

Where

S = disk size, in GB

U = planned amount of unique data in the deduplication data store, in GB.

 For example, if the planned amount of unique data in the deduplication data store is U=5 TB, the deduplication database will require the free disk space not less than

$S = U / 32 + 10$

Selecting a disk for a deduplicating vault

For the purpose of data loss prevention, we recommend using RAID 10, 5 or 6. RAID 0 is not recommended since it not fault tolerant. RAID 1 is not recommended because of relatively low speed. There is no preference to local disks or SAN, both are good.

3 GB of RAM per 1 TB of unique data

It is not necessary to follow this recommendation if you do not experience a deduplication performance problem. However, if the deduplication is too slow, adding more RAM to the storage node may significantly raise the deduplication speed. (In general, the more RAM you have, the greater the deduplication database size can be, provided that the deduplication speed is the same.)

Only one deduplicating vault on each storage node

It is highly recommended that you create only one deduplicating vault on a storage node. Otherwise, the whole available RAM volume may be distributed in proportion to the number of the vaults.

64-bit operating system

The storage node must be installed in a 64-bit operating system. The machine with the storage node should not run applications that require much system resources; for example, Database Management Systems (DBMS) or Enterprise Resource Planning (ERP) systems.

Multi-core processor with at least 2.5 GHz clock rate

Use a processor with the number of cores not less than 4 and the clock rate not less than 2.5 GHz.

Sufficient free space in the vault

Indexing of a backup requires as much free space as the backed-up data occupies immediately after saving it to the vault. Without a compression or deduplication at source, this value is equal to the size of the original data backed up during the given backup operation.

High-speed LAN

1-Gbit LAN is recommended. It will allow the software to perform 5-6 backups with deduplication in parallel, and the speed will not reduce considerably.

Back up a typical machine before backing up several machines with similar contents

When backing up several machines with similar contents, it is recommended that you back up one machine first and wait until the end of the backed-up data indexing. After that, the other machines will be backed up faster owing to the efficient deduplication. Because the first machine's backup has been indexed, most of the data is already in the deduplication data store.

Back up different machines at different times

If you back up a large number of machines, spread out the backup operations over time. To do this, create several backup plans with various schedules.

Use fast cataloging

Indexing of a backup starts after its cataloging has been completed. To reduce the overall time required for backup processing, switch automatic cataloging (p. 109) to the fast mode. You can start full cataloging manually outside of the backup window.

Configure alert notifications

It is recommended that you configure the "Vaults" alert notification in the management server options. So you can promptly react in out-of-order situations. For example, a timely reaction to a "There is a vault with low free space" alert can prevent an error when next backing up to the vault.