



PRESENTS

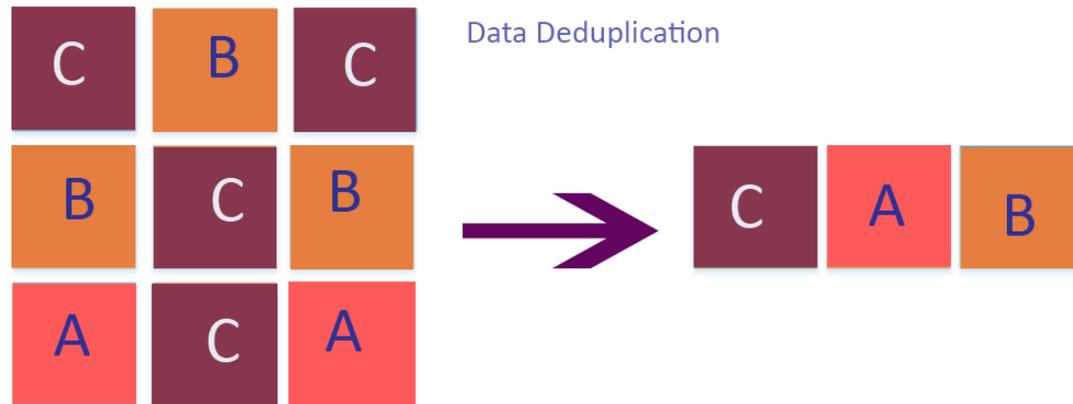
Guide to deduplication

Deduplication is the most affordable and complete data storage technique

What is data deduplication?

Data deduplication is synonym with “intelligent compression” or “single-instance storage”. In essence, it is a simple data compression process that eliminates all redundant copies of data and stores a single dataset.

Copies of data blocks are replaced with pointers to the unique copy.



How does it work?

The main data deduplication process consist of storing each unique data sequence only once. By discarding duplicate data elements, the deduplication engine sends only the changed data bits to the storage array. This way, the amount of transmitted data is significantly reduced and the speed of data backups increased. Data deduplication is also used for replication purposes in order to send only the unique blocks from the source to the target and if a block is not unique, it will send only the pointer to the right unique blocks to reduce bandwidth utilization.

Benefits of data deduplication

- The deduplication is the solution to data growth and data integrity; the finite amount of storage the organizations usually have access to limits the backup options and the reliability of the performed backup.
- Deduplication provides the best combination to address the challenges related to data storage and data integrity, enabling you to keep storage costs low and systems securely backed up.

More benefits of data deduplication

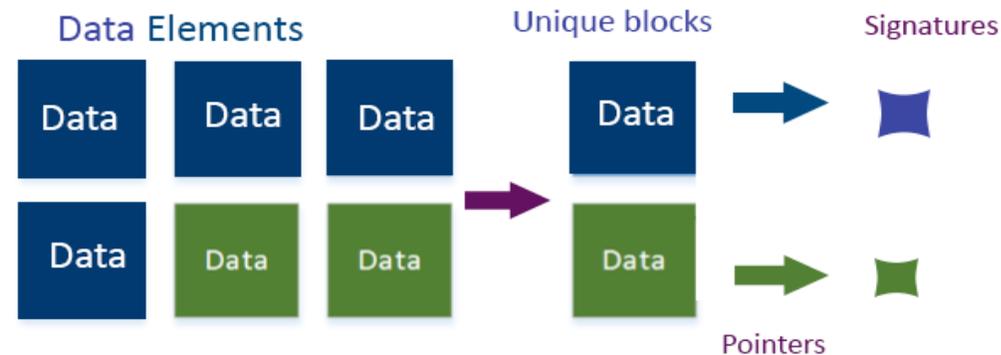
With data deduplication you will:

- Save cost on storage: the data is stored only once even if it's been subsequently updated and changed
- Improve backup replication speed by copying only the original data with its pointers
- Optimize storage requirements ensuring an efficient use of the infrastructure
- Keep the system securely backed up with data classification and storage automation

File-level vs. block-level deduplication

File deduplication compares a file that is about to be backed up or archived with any stored copy of this file. After checking the attributes of the file against an index, the deduplication engine removes any duplicate files and keeps only one copy. If the file is unique, but updated, only the stub that points to the original file is stored.

Block-level deduplication breaks the file into blocks with same fixed size. Each block is processed using hash algorithms; once the block is analyzed, its signature is compared against the deduplication table. Only if the block's signature is not found in the table, the data is backed up and stored on the disk.



Fixed vs. Variable deduplication

Fixed deduplication algorithm stores data into fixed size blocks. Based on the block size, the smallest identifiable difference in the data is stored as variant instance.

Variable compression groups data into chunks following the patterns in data itself. If the data pattern shifts during a backup, the chunking shifts along with it. Variable chunking recognizes duplicate data and resynchronizes the information according to the identified duplicates. Variable chunking is more efficient because it increases the deduplication ratio by matching smaller data more easily.



Multi-streaming deduplication

When data is sent to a backup appliance using multiple streams, the deduplication is performed in parallel through different channels, thus increasing I/O bandwidth and avoiding bottlenecks.

For environments with large amounts of data, the multi-streaming deduplication ensures a higher dedupe engine performance and a higher dedupe ratio.

Target vs. source deduplication

Deduplication at the source removes redundant blocks before sending any data to the backup target (client or server). Source-based deduplication reduces bandwidth and storage use. Moreover, it doesn't require any additional hardware.

Deduplication at the target is network-based: backups are sent through the network to a remote location. Even if it requires more network bandwidth, the target deduplication is more efficient than source deduplication, particularly for big data sets.

Data deduplication vs. compression

Although the two techniques serve the same purpose, they work differently. While deduplication seeks out redundant data, compression reduces the number of bits that represents the data.

Both techniques are data reducing techniques designed to optimize the storage capacity. Taken together, these techniques bring new benefits to deduplication:

- Reduced data print
- Lower bandwidth consumption
- Longer retention time
- Reduced tape backups
- Faster recovery time



5 things you should know about SPHiNX

1 – No agents or software to be installed on your POWER server !

SPHiNX emulate a tape device/library for which IBM i or AIX detects as a tape device or Virtual Tape Library !

2 – Backup to SPHiNX from every platform !

SPHiNX is a software-agnostic solution which gives you the option of using any backup management application for your open systems.

3 – Use any storage behind the appliance

SPHiNX is a vendor independent platform for which you can use any storage disks to backup to.

4 – No hassle maintenance

SPHiNX appliance is covered by default by a full five (5) years maintenance plan !

5 – Send to LTO tape device directly

SPHiNX can

To dedup or not dedup ?

SPHiNX offers many options to fit your needs ! Compression-only appliance, optimization engine which uses data deduplication technology, enhanced replication from your legacy/compression-only SPHiNX to your Managed Service Provider !



Choose the right technology, choose deduplication

Deduplication saves storage, time and ensures a better transfer rate across the network.

For more information about SPHiNX visit

www.sphinxbackup.com/success

or talk to our SPHiNX specialists

1-844-7SPHiNX



Quebec (Canada) – Florida (US), Oregon (US), Texas (US), Manchester (UK), Schwäbisch Gmünd (Germany)