

Data Reduction-Deduplication and Compression

Data growth is the biggest data center hardware infrastructure challenge for large enterprises; Capacity-optimization technologies play a critical role in today's environment where companies need to increase storage efficiency and reduce costs – To do more with less .

Data Reduction is the process of minimizing the amount of data that needs to be stored in a data storage environment which can be achieved using several different types of technologies. The best-known data reduction technique is Deduplication and Compression.

Deduplication is the process of identifying duplicate data contained within a set of block storage objects and consolidating it such that only one actual copy of the data is used by many sources. This feature can result in significant space savings depending on the nature of the data. It can be done at source, inline or post-process. For example, suppose the same 10 MB PowerPoint presentation is stored in 10 folders for each sales associate or department. That's 100 MB of disk space consumed to maintain the same 10 MB file. File deduplication ensures that only one complete copy is saved to disk. Subsequent iterations of the file are only saved as references that point to the saved copy, so end-users still see their own files in place. Similarly, a storage system may retain 200 e-mails, each with a 1 MB attachment. With deduplication, the 200 MB needed to store each 1 MB attachment is reduced to just 1 MB for one iteration of the file .

Compression is the process of reducing data to use less capacity than the original format. Compression basically attempts to reduce the size of a file by removing redundant data within the file. By making files smaller, less disk space is consumed, and more files can be stored on disk. For example, a 100 KB text file might be compressed to 52 KB by removing extra spaces or replacing long character strings with short representations. An algorithm recreates the original data when the file is read. For example, a 2:1 compression ratio can ideally allow 400 GB worth of files on a 200 GB disk (or 200 GB worth of files would only take 100 GB on the disk). It's very difficult to determine exactly how much a file can be compressed until a compression algorithm is applied.