VTL BACKUP APPLIANCE

Post-Process Deduplication vs In-line Deduplication

What are the differences between a VTL backup appliance that achieves 50:1 deduplication on average and another device that promises up to 30:1 but only delivers between 17:1 and 3:1? The answer is simple, it depends on the backup deduplication process. All VTL Backup Appliance manufactures use different deduplication technologies and algorithms. However, how a VTL device deduplicates your backups affects both the speed (backup window) and its ability to reduce disk space.

In-Line Deduplication

An in-line process deduplicates the VTL backup in real-time, as it's written to disk. The in-line duplication process provides an immediate benefit of utilizing less disk space. However, in-line deduplication requires considerably more processing power to prevent slowing down the backup speed (increasing the backup window). As a result of increasing the processing power, backup appliances using in-line deduplication cost considerably more. An in-line deduplication VTL has two battles to win, backup speed vs real-time deduplication. The deduplication algorithm used will vary between VTL devices and vendors, but all in-line deduplication VTLs have to use a combination that will affect both the **backup window** and the **data reduction** of the deduplication process. If performance of the backup speed is favored in the algorithm, the data ingest rate will be reduced at the expense of deduplication process. In the same regards, if the deduplication process is favored, the performance of the backup speed will suffer.

Post-Process Deduplication

The alternate method to in-line deduplication is the post-process. A post-process deduplication VTL device first allows the backup to complete, which allows the maximum throughput performance to be achieved (smallest backup window possible), before deduplication occurs. Initially, a post-process deduplication process requires more disk space, but at significantly lower costs due to the low cost of storage. In addition to a smaller backup window and a lower cost per device, a VTL using post-process deduplication will also have higher deduplication ratios making more efficient use of your storage space.

Midland Information Systems, Inc.

Speed Matters.

Because the amount of time it takes to perform backups and restores is critical, a post-process VTL backup appliance is the right choice. A VTL using post-process deduplication sends backup data to the VTL at full speed, providing the smallest possible backup window. The necessary time is then taken to utilize the most efficient deduplication algorithms needed to provide the maximum data reduction, and without impacting your backup window.

The Honeymoon Ends.

Over-time, in-line deduplication algorithms on VTL devices become less effective due to excessive dictionary growth. Dictoinary growth is caused by new data that has been added and compared to existing content. Data centers with faster growth rates and a VTL using in-line deduplication will attest to these claims. Furthermore, backup restores are more likely to slow down since the data has to be reconstituted against a larger dictionary. As a result, the VTL device using in-line deduplication will also slow down as the amount of data stored grows.

Putting Deduplication Ratios Into Perspective.

In addition to the primary purpose of data reduction, deduplication also plays a key role in replication. The better your VTL's deduplication process is, the less bandwidth your data will consume while shortening the amount of time needed to complete replicating your data. For some data centers, achieving a 10:1 deduplication ratio vs a significantly higher 50:1 ratio will likely result in failure. Although vendor claims about deduplication ratios may confuse some customers, the fact is, ratios will vary by environment and are based on many variables. There are no guarantees for deduplication ratios, nor is there a means to provide an estimate for your environment. However, an honest vendor should provide you with real to life estimates based on existing customer experiences.

Planning For Unadvertised Costs.

Since in-line VTL devices run out of space prematurely, knowing the costs to expand the storage of your VTL is important. All too often companies underestimate the amount of capacity needed to store data when using in-line deduplication. The deduplication ratios expected aren't achieved, so the virtual tape library becomes prematurely full. Depending on the manufacturer, the cost of expanding the storage space can be greater than the cost of the original unit. However these issues can be avoided by finding VTL devices that perform deduplication more effiectively and can acheive a higher reduction rate.

A Note To The Customer.

Cybernetics' virtual tape libraries use post-process deduplication and are one of the only VTL vendors to do so. If you desire a VTL backup solution with the lowest initial costs, faster performance, more consistent backup and restore speeds, and the best-in-class deduplication ratios, turn to Cybernetics technology. Cybernetics offers a complete line of VTL devices for businesses of all sizes.

Midland
Information Systems, Inc.