Solution Brief

Optimize Data Reduction Technology with Granular Control

Data reduction technologies include deduplication and compression. Data reduction can be powerful for the right applications, but it is not useful for all applications. Granular control is the best way to use these tools.

Highlights

Highlights

- Appropriate applications can benefit from significant capacity savings with data reduction tools.
- Applications that are poor fits for deduplication need to be treated differently
- Always on deduplication may actually hurt overall data center operations

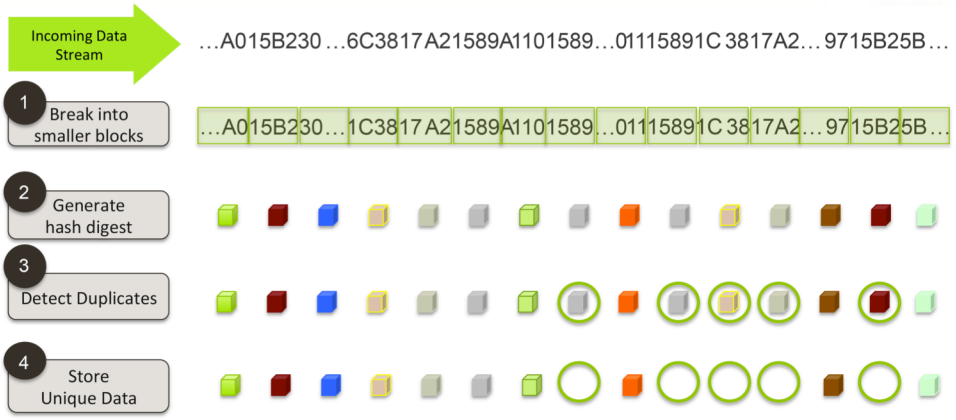
Granular Control

- Concerto 2200 provides inline deduplication and compression that is switchable by share or file
- Windows Flash Arrays provide post process deduplication on a file by file basis

When data reduction technology is appropriate for the workload, the results can be stunning. Imagine reducing the amount of storage required by 5x, 10x or more? The effective cost of storage plummets when data reduction is a good fit. Although data reduction technology has been available for a long time, the arrival of all-flash arrays has brought it to the forefront of storage discussions.

What is Deduplication?

Simply said, deduplication is the ability to avoid writing redundant blocks of data to reduce the amount of storage required. Instead of writing redundant data, systems use pointers to the original unique blocks of data. This can result in significant savings in the amount of storage used by reducing writes. There are two kinds of deduplication: inline which is processed before writing data, or post process: where deduplication reduces data after the initial write, to consolidate space needed.



Think of deduplication as a four step process:

- Incoming data is broken into smaller blocks
- Data blocks are then summarized as a mathematical expression (hash)
- Duplicate hash blocks are identified (easier to do with hash than actual blocks)
- Unique data is written to storage, and duplicate blocks are identified with a pointer to the original unique version

Violin Memory, Inc. | 4555 Great America Parkway, Santa Clara, CA 95054 USA | Tel: 1-650-396-1500 • Fax: 1-650-396-1543 | www.violin-memory.com

04-0417rev 1.0r 08/14

What is Compression?

Compression can also bring significant benefits to the right workload. The way it works is different, and the benefits are independent of deduplication. Where deduplication looks at blocks of data, compression looks at much smaller data, the bit stream. When the compression tool sees 1000 zeros in the bit stream, it will just note there are 1000 zeros, instead of actually writing zero a thousand times.

How Does Granular Control Work?

Data reduction works very well in the right environment, and can provide big savings for your storage budget.

Examples of a great fit for deduplication include virtual desktop infrastructure (VDI) and virtual server infrastructure (VSI). A great fit for deduplication is found when the files are mostly the same, so eliminating redundant blocks will have a big impact. When the files are all the same, there is no need for deduplication, a clone is created and replicated as needed. The best fit for identical files is really fast storage, like the Violin All Flash Arrays. Adding deduplication where it is not needed just adds latency, and makes things less productive. For instance, when dealing with VDI environments, persistent desktops are mostly the same with some unique elements, and are great fit for deduplication. For VDI environments with non-persistent desktops, the desktops are all the same, and a poor fit for deduplication. The lesson learned is to use deduplication where it adds value, and don't use it where it just adds latency.

A survey by IDC looking at user workloads identifies VDI and VSI as a small portion of the typical data center workload. Workloads that are typically not a good fit for deduplication would include databases, transaction-oriented applications, and analytics. In particular, consider deduplication's impact on databases tables. Tables are usually eliminated with deduplication, since the data exists elsewhere in the database. Tables exist to speed performance of the database, by reducing lookup time. By eliminating the tables in a database, performance could suffer. The right solution is to use the technology where it is a benefit, and not where it doesn't add a benefit.

There is a reason that some all-flash arrays have "always-on" deduplication. All-flash arrays come in two varieties: SSD-based and backplane based. The need for deduplication is different for the two architectures. SSD-based designs have two levels of management: one in the SSD itself, and an array-level controller. Deduplication can help SSD-based arrays manage their flash resiliency by limiting write traffic, since it cannot coordinate writes between its SSD and array-level controllers very well.

One of the few examples of a backplane design is Violin's Flash Fabric Architecture™ which does not use SSDs. Instead Violin puts flash chips on specially designed cards that plug into an optimized backplane and unites the management into a coherent unit. Hence there is no need for continual deduplication to manage writes, and the overhead of deduplication processing does not need to be incurred in a Violin array resulting in better performance and resiliency.

The Violin Memory Concerto 2200 provides inline deduplication and compression that allows you to decide on a share by share, or file by file basis what needs to be deduplicated, so no unnecessary latency is incurred by an unneeded, or even detrimental deduplication process. The Violin Windows Flash Array provides post process deduplication that also allows you to choose which files are to be deduplicated, and which to leave untouched. You have the option to use the technology, or not, as the application demands.

Results

Because Violin Memory's approach allows you to turn deduplication on and off, you can assess the value of data reduction technology for your applications. In the Concerto 2200 we have included a dashboard that allows you to see the actual deduplication rates for the data. If the deduplication rate doesn't justify the increase in latency, you can remove it from the 2200, and if it has a benefit, look for similar workloads that you want to add.

Violin Memory puts you in control of critical deduplication capability, allowing you to use it where appropriate, and to avoid the performance penalty where it is not needed.

Violin Memory, Inc. | 4555 Great America Parkway, Santa Clara, CA 95054 USA | Tel: 1-650-396-1500 • Fax: 1-650-396-1543 | www.violin-memory.com