



Inside Track Research Note

In association with



Everyone's talking about Dedupe

Where does dedupe and data
compression make sense?

March 2016

About this Inside Track

The insights presented in this document are derived from independent research conducted by Freeform Dynamics. Inputs into this include in-depth discussions with IT vendors and service providers on the latest technology developments, along with intelligence gathered from mainstream enterprises during broader market studies.

Is the use of data compression and deduplication technologies a good fit for all enterprise storage use cases?

From a practical perspective, data deduplication and data compression are IT solutions designed to minimise the total amount of physical data written to disk.

In a nutshell

Storage is a very hot topic in every data centre and computer room. Data volumes keep rising and users demand ever better performance and availability, all without pushing storage costs any higher. Over the past few years a new form of storage has enjoyed great success shaking up what was previously a very stable ecosystem. Flash storage and SSDs offer great performance compared to traditional spinning disks, but until recently they also came with a very high cost. In fact the costs were high enough that the use of flash storage could often only be justified for workloads that demanded the very best performance and lowest latency.

In order to expand the potential range of use cases, vendors began to offer flash storage solutions that employed special functionality to limit the total data that had to be stored. 'Data deduplication', usually truncated to dedupe, and 'data compression' are now positioned as simple ways to limit storage expenditure. In fact such functionality is now so widely marketed that data 'compression' and 'dedupe' have become standard fixtures on RFP's for many storage solutions, not just the new flash based systems but also those using traditional spinning disks.

This begs the question, is the use of data compression and deduplication technologies something that is a good fit for all enterprise storage use cases? More importantly, should you be applying these techniques to everything by default? This paper considers the advantages and challenges associated with data compression and deduplication solutions and discusses what to think about when looking at their use in the acquisition of new storage.

What happens during dedupe and data compression?

From a practical perspective, data deduplication and data compression are IT solutions designed to minimise the total amount of physical data written to disk, regardless of the 'natural' size of what's being stored. Some people even regard deduplication to be a specialised case of more general data compression technologies.

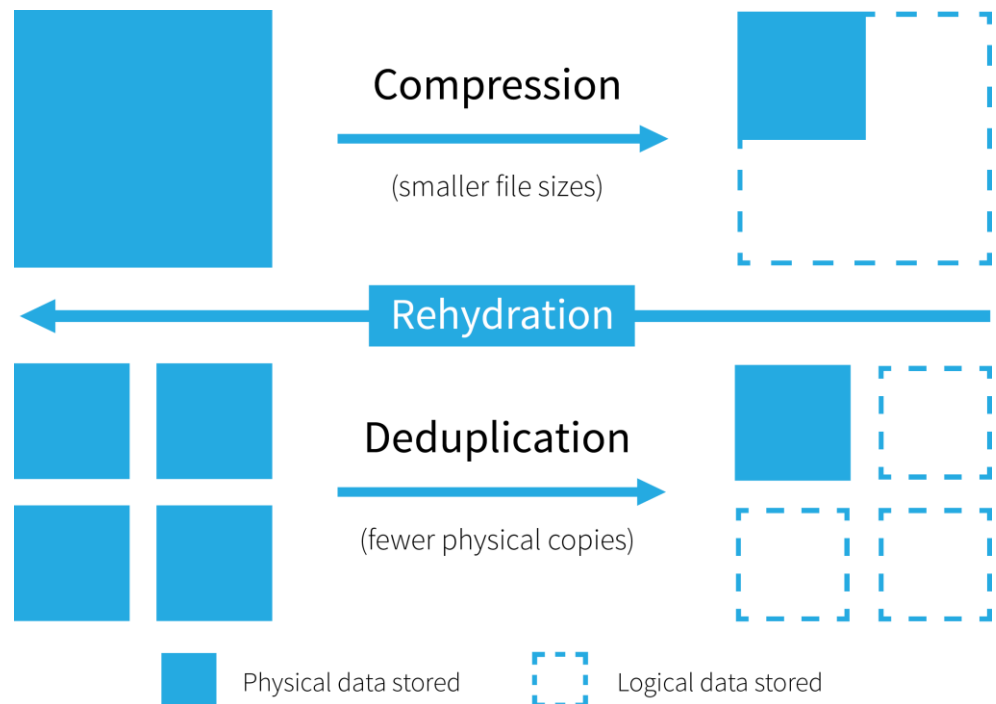
There are many different architectural approaches involved in making the 'magic' happen, the details of which are too complex to discuss here in detail, but let's consider some of the basic principles.

On writing data to storage, fundamentally, data compression uses complex mathematics to reduce the actual number of bits required to hold a file or block of data. Data deduplication, on the other hand, essentially looks for data patterns that match others already held on the system, so it can store pointers to those rather than storing the same data over again. Both processes take place automatically and invisibly to the user or requesting application.

When reading back data these processes are reversed, and the data requested is recreated from either it's compressed or deduplicated form. The 'rehydration' process is again automatic and invisible.

Hopefully our description here has made sense so far, but to clarify further, take a look at our visual schematic of compression and deduplication in action, which will help to visualise one of the key principles we have been discussing (Figure 1).

Figure 1
Data deduplication and compression basics – it's all about reducing storage volumes



Practical considerations

There are a number of practical considerations to bear in mind when looking at the use of dedupe and compression in your environment, some positive, others you need to be wary of.

Dedupe and compression rates vary considerably depending on the type of data, but in many cases the amount of physical storage required to hold data can be substantially reduced.

Storage volume

Dedupe and compression rates vary considerably depending on the type of data, but in many cases the amount of physical storage required to hold data can be substantially reduced, thereby saving acquisition costs, as well as optimising certain consequential operational costs. The latter can include the power consumed by the storage system, but other softer benefits can be even more valuable. One commonly referenced is the reduction in time it takes to protect data during backup operations. Another is a reduction in the volume of traffic that needs to be transported over a WAN to replicate data.

Performance of writes and reads

The very processes associated with data compression and deduplication inevitably place overheads on both read and write operations. The magnitude of the overhead induced will vary depending upon many factors, including how the data is stored and the architecture of the storage platforms. In particular the latency induced in both read and write operations is normally too great to be considered practical for use with systems built using hard disks. Depending on the type of data being written, in systems where dedupe and compression are always on, performance impacts may be visible without providing significant storage reduction benefits.

Data compression and deduplication inevitably place overheads on both read and write operations.

Material changes to stored data

Data compression clearly changes the data stored as you are not writing it in its original, natural form. In many business use cases this may not be an issue, but some

While both compression and deduplication technology is generally very reliable, it is not totally fool proof.

data sets may have external regulatory or audit requirements that make it inappropriate to use the kind of techniques we have been discussing without creating a compliance risk.

Other concerns

While both compression and deduplication technology is generally very reliable, it is not totally fool proof. Applying advanced mathematics on the fly, which is essentially what's going on, can occasionally lead to anomalies and even loss of data. It is always worth checking with suppliers and potentially regulators and auditors to seek assurances that there are no significant risks given the types of data you store.

Use cases for and against dedupe and compression

As dedupe and compression offer both potential benefits and disadvantages, it is clear that using systems with these capabilities may make sense for some use cases, but not others.

It is clear that using dedupe and compression capabilities may make sense for some use cases, but not others.

The exact nature of the storage reduction benefits achieved depend entirely on the data sets concerned. In 'backup' environments where data is frequently stored multiple times, although it's difficult to generalise, compression ratios can be very dramatic. Gains in relation to 'live' data stored in production environments, however, are generally much less, with volume reduction ratios typically in the range of 3:1 to 5:1. There are also many important scenarios where data deduplication may have no significant impact on data volumes stored but can add to storage latency and hence to service performance degradation.

The computational overhead and increase in latency on both writes and reads of data usually makes compression and deduplication difficult on storage systems built using hard disks, unless such systems are designed explicitly to cater for these functions. The inherently lower latency found in flash based storage systems can make data dedupe and/or compression a lot more viable, provided the other concerns mentioned above have been addressed.

These factors, along with the naturally conservative nature of IT professionals looking after corporate data has meant that up until now, data dedupe and compression have only been utilised extensively in a few scenarios.

The naturally conservative nature of IT professionals looking after corporate data has meant that up until now, data dedupe and compression have only been utilised extensively in a few scenarios.

Based on some of these factors, let's take a closer look at where the opportunities for effective use of dedupe and compression technologies lie, and where they might be less applicable.

Where dedupe and compression make most sense

Several generic use cases have been shown to make effective use of data deduplication and compression with virtual environments? Examples include virtual servers and desktop virtualisation scenarios where it is likely common data sets will be utilised in virtual system after virtual system, e.g. operating system files etc. The same applies to backup systems where in many architectures the same file may be stored as data is protected repeatedly, time after time. Let's look at a few typical use cases in more depth:

It is very easy to clone virtual servers, thus the opportunity for deduplication to reduce data volumes stored is very clear.

Dedupe in particular has achieved good adoption in relation to long term data protection systems.

The inappropriate application of dedupe and / or compression technologies can lead to performance degradation with no benefits to offset this disadvantage.

Large scale 'web' services

Very large web services, especially those used in consumer markets, can achieve great benefits in terms of reduced data volumes, and hence cost avoidance, by employing compression and data deduplication.

Server and desktop virtualisation

It is very easy to clone virtual servers, thus in many organisations it is possible that several copies, in some scenarios potentially hundreds or even thousands, of certain virtual machines may need to be stored at any one time. The opportunity for deduplication to reduce data volumes stored is very clear. Equally, in situations where desktop virtualisation is deployed there may be scenarios where many desktops are essentially identical, once again providing opportunities for the effective use of data deduplication technologies.

Backup and recovery

An area where dedupe in particular can be very effective and in which it has achieved good adoption is in long term data protection systems. The reason for this is simple – depending on the information being protected and how often it changes many identical copies of the same data can be stored. As a consequence the application of data deduplication can result in very high data reduction rates. Indeed, it is possible for many data sets to be reduced in size by very high factors, with results of compression ratios of 30 to 1 or more not uncommon depending on how frequently backups are performed.

The reduction in data volumes can also have benefits in the time required to protect data, reducing the overall backup window. It is also a task where latency is not usually a sensitive matter hence making the compression and dedupe overhead far easier to entertain. However, it should be noted that while backup windows can be significantly reduced, the time to restore data from deduped sets can be extended due to the extra processing required to retrieve information, and this consideration must be borne in mind when looking at recovery time objectives.

Archiving

Similar arguments also apply to the use of compression and dedupe in long term archiving. The amount of information that needs to be archived continues to grow rapidly, and the cost benefits achieved by reducing data volumes via dedupe and compression can be substantial. Once again, the nature of archiving makes any latency associated with the deduplication and compression processes straightforward to absorb. However, dedupe ratios tend to be lower than for backup solutions as data is written only once.

Where data deduplication and compression may prove ineffective

There are certain areas where the application of data compression and deduplication technologies may deliver little in the way of storage volume reductions, but given the processing overhead inherent in such solutions they still add latency.

Compressed data

Some data sets, for example video and other media content, are natively stored in very compact formats making them unlikely to benefit significantly from using deduplication and data compression solutions unless multiple copies of the same content are stored time after time. Thus the inappropriate application of such

It is usually impossible to use data compression and deduplication on data sets that are encrypted.

capabilities can, once again, lead to performance degradation with no benefits to offset this disadvantage.

Undermined by encryption

Equally, it is usually impossible to use data compression and deduplication on data sets that are encrypted. After all, the very purpose of encrypting data is to make it unreadable and essentially this involves removing any 'patterns' and 'repetitions' in the data stream that could provide a clue to the contents. This, in turn, makes it impossible to compress and deduplicate.

Databases

The way that many database management systems store data can mean that data deduplication and compression solutions have little effect on data volumes. The precise benefit gain in data storage varies between different database systems, but it is always the case that the latency overhead inherent in such systems will reduce effective service performance on both data writes and reads. As a consequence, up to the time of writing data deduplication solutions have not been applied widely to such systems.

In situations where absolute performance is key the use of deduplication and compression technologies is unlikely to be considered.

High performance systems

Today any system which requires the best possible performance is almost certain to utilise all flash array storage. In situations where absolute performance is key, the use of deduplication and compression technologies is unlikely to be considered as these incur additional latency, however small. If the storage is planned to support very high performance systems along with other enterprise services, an all flash array should be capable of selectively utilising dedupe / compression only for workloads where it makes sense.

Use case evolution

To date, the overhead of data reduction technologies has limited their use in the storage of 'active' data, i.e. the data that is required by applications and users in everyday business processes. This is certainly the case in the context of hard disk based solutions, where the latency overhead induced can be difficult, certainly expensive, to overcome.

Until recently the additional cost associated with flash based storage systems has restricted their use to meet the requirements of applications and services where performance is an absolute. As a consequence, while some business cases have used projected data compression and deduplication factors to reduce the apparent cost per terabyte of flash storage solutions when compared with traditional spinning disks, the real world implementation of these technologies in environments where consistent, lowest latency performance is key has been sporadic.

To date, the overhead of data reduction technologies has limited their use in the storage of 'active' data.

As the price of flash falls and as the technologies mature flash systems are being used to support more mainstream enterprise applications and services. In these scenarios it is likely that the use of dedupe and compression will increase, especially if these capabilities can be selectively applied to specific data sets rather than as a binary on / off selection as is the case in some current offerings. However, on a pure cost per terabyte basis, the incentive to use deduplication and compression will equally be reduced for performance sensitive data sets.

The bottom line

Data dedupe and compression are frequently promoted as the holy grail for storage cost savings. But experience has shown that it is essential the organisation consider the specific usage scenarios.

Data dedupe and compression are frequently promoted as the holy grail for storage cost savings. But experience has shown that it is essential the organisation consider the specific usage scenarios they need to address to avoid deploying such technologies ineffectively, thus gaining little in the way of data reduction but still experiencing the performance penalty such systems entail. Dedupe and compression may be great for use in your backup environment, but only if you have also considered the potential impact on your recovery times. Equally, data dedupe may offer little, if any benefit to certain other systems, notably database systems and anything involving the storage of already encrypted data.

Thus, when considering whether to look for data compression and deduplication functionality in new storage solutions there are a few questions that need to be addressed. In essence these are no different from the standard questions that should be applied to any acquisition of a storage system.

The first item, as always, is to understand exactly what are the service capabilities required by the workloads to be supported on the system. Naturally enough these should include an accurate idea of the service levels to be delivered, especially in terms of application / service performance. Clearly the budget available for the projected acquisition will help shape what solutions can be considered.

When considering whether to look for data compression and deduplication functionality in new storage solutions there are a few questions that need to be addressed.

But it is then essential to consider whether there are any specific compliance or regulatory issues governing how the data can be stored, essentially identifying if data compression and / or deduplication is permitted by the demands of data governance. If data compression and dedupe is permitted for some data sets but not others or if different data sets need different performance qualities it is essential to determine which potential solutions provide enough management granularity.

Where dedupe and / or compression do make sense for a particular workload, it is important to understand there are several, significantly different, architectural approaches used in solutions. Each of these architectures have different advantages and constraints, so if dedupe and / or compression are required, additional research will be necessary to identify which approach best fits the needs of the services to be supported.

Despite all the marketing activity at the moment, do not assume that dedupe is broadly applicable, desirable or legal in your environment.

Despite all the marketing activity at the moment, do not assume that dedupe is broadly applicable, desirable or legal in your environment. By all means include it in your RFP, but make sure you can manage where and how it is applied within the system. This may cost more than systems having compression and dedupe on all the time (with no ability to control or manage), but in a mixed workload, shared storage environment, lack of control could be a problem.

As always, the key to making the right decision is to understand your requirements and constraints, and make sure the technology you are considering will meet your needs without creating any undesirable side-effects.

About Freeform Dynamics

Freeform Dynamics is an IT industry analyst firm. Through our research and insights, we aim to help busy IT and business professionals get up to speed on the latest technology developments, and make better informed investment decisions.

For more information, and access to our library of free research, please visit www.freeformdynamics.com.

About Fujitsu

Fujitsu is the leading Japanese information and communication technology (ICT) company offering a full range of technology products, solutions and services. Approximately 162,000 Fujitsu people support customers in more than 100 countries. We use our experience and the power of ICT to shape the future of society with our customers.

Terms of use

This document is Copyright 2016 Freeform Dynamics Ltd. It may be freely duplicated and distributed in its entirety on an individual one to one basis, either electronically or in hard copy form. It may not, however, be disassembled or modified in any way as part of the duplication process. Hosting of the entire report for download and/or mass distribution by any means is prohibited unless express permission is obtained from Freeform Dynamics Ltd or Fujitsu. The contents contained herein are provided for your general information and use only, and neither Freeform Dynamics Ltd nor any third party provide any warranty or guarantee as to its suitability for any particular purpose.