

# The Business Benefits of Data Efficiency

Scott D. Lowe, vExpert February, 2016



## **Table of Contents**

Table of Contents	2
Introduction	3
The Data Problem	4
The Data Iceberg	4
Data Efficiency	5
Improving Line of Business Application Performance	5
Hybrid Storage	5
The Data Efficiency Amplifier	6
Backup and Recovery	6
Disaster Recovery and the Impact on the Network	8
WAN Acceleration	8
Summary	9
About the Author	9



ActualTech Media © 2016. All rights reserved.

Under no circumstances is this document be sold, copied, or reproduced in any way except with written permission.

The information contained with the document is given in good faith and is believed to be accurate, appropriate, and reliable at the time it is given, but is provided without any warranty of accuracy, appropriateness or reliability.

The author does not accept any liability or responsibility for any loss suffered from the reader's use of the advice, recommendation, information, assistance or service, to the extent available by law.

All trademarks are the property of their respective owners.



## Introduction

Business decision makers, such as the CIO and the CFO, are interested in improving the overall returns on technology investments while, at the same time, fixing what is often considered a fundamentally broken operating model for IT as a whole.

In recent years, new market segments have been invented to solve these challenging issues. Now that IT is a core part of the overall business, there is a growing need for technology products that make IT simpler and less expensive to manage while also providing the levels of performance that demanding business applications require.

Hyperconverged infrastructure (HCI) is a quickly growing segment of the data center technology market that is successful because it addresses the primary challenges facing IT: cost, performance, and complexity. However, this emerging data center technology, while powerful in its own right, is only part of the broader story.

As is the case with many popular technologies, there are new entrants into the market all the time. Although many vendors provide compelling products in this space, SimpliVity has taken their solution to a whole new level through their Data Virtualization Engine. A core feature of the Data Virtualization Engine is *data efficiency* — the focus of this paper.



When non-technical people think of the term *data efficiency*, they usually think about how the actual data works across the organization, and how it lends itself to improved business processes. However, in the context of the Data Virtualization Engine, *data efficiency* is used to describe how much physical and virtual capacity is necessary in order to house an organization's data, which includes far more than the data users see in their programs or e-mail.

Additionally, there are constructs that are stored as data the data center as well. Such constructs include the virtual servers (software-based versions of the computer on your desk) that business applications depend on to operate. Considering this, what you end up with is a very large data iceberg.

#### THE DATA ICEBERG

Think of your company's data as a very expensive iceberg. Users see only the pieces that stick above the water. Below it, but supporting everything that the business is doing, is a hidden foundation that is critically important and makes up a significant part of the company's data storage cost.

Storing all of this data has become an overwhelming business and technical challenge. Organizations need to turn to IT to meet complex needs. As such needs arise, IT builds out systems to support these needs, all of which require data storage capacity. They add on to the iceberg.

Now, take that data storage a step further. Most companies have robust backup processes that create copies of all of the data on a periodic basis. Others make copies of data that are then sent across the network to remote disaster recovery sites. Every time data is copied *en masse*, the data and its associated overhead (the hidden, underwater parts of the iceberg)must be copied as well to ensure business continuity in the face of a disaster.

The end result is a gigantic iceberg of bloated storage systems, expensive upgrades, business disruption, and additional complexity in the operation as IT works through all of the issues that arise.

To combat some of these problems, IT departments add additional services over the years, thus compounding the complexity. For example, to support newer workloads that require faster storage, separate storage devices custombuilt for performance are added. To make sure that copies of data can transfer efficiently across the network to a backup site, acceleration technologies are implemented. Again, the iceberg simply continues to grow.



The Business Benefits of Data Efficiency

Page 4

## **Data Efficiency**

Thanks to the SimpliVity Data Virtualization Engine's data efficiency features, a solution is available, and it has a number of benefits to help you avoid the data chaos. Due to the way that SimpliVity handles data reduction and due to the additional services built into the platform, there are opportunities for other farreaching benefits beyond savings in storage capacity. This is the rest of the data efficiency story.

#### IMPROVING LINE OF BUSINESS APPLICATION PERFORMANCE

In addition to needing raw capacity, modern business applications also need storage systems that can keep pace with the performance demands. This can be a challenge that's harder to solve than it would first appear. SimpliVity's OmniCube platform provides multiple solutions to this challenge.

#### Hybrid Storage

With traditional hard disk-based storage devices (which are the storage in use in many of today's data centers) storage performance can be hard to achieve.

Here's why: When a user decides to save a document, the mechanical components inside the hard drive on which that document will be stored jump into action. The *write head*, which is the device responsible for saving that data, positions itself a fraction of a hair's width above a magnetic platter that spins 7,200 to 15,000 revolutions per minute. Once positioned, the write head starts to write the file to the storage. However, there may not be enough continuous space on the platter to write the whole file at once, so the write head saves portions of the file all over the disk and then keeps a comprehensive index so that it knows where the pieces of every file are stored.

The more time it takes to write the file, the longer the user has to wait for the computer to save the file. The time between the command to write the file and the acknowledgement that the file has been written is known as *latency*.

Latency is one of the most serious problems in today's data centers, and it is exacerbated by newer business applications. Many newer applications introduce storage system performance requirements that increase latency. Eventually, the latency becomes so bad that it is a detriment to ongoing business operations. No one, from customers to company employees, likes to have to wait for computers or technology systems to perform basic operations.

To combat this growing problem, a new kind of storage has entered the market. Called *flash storage*, or *solid state storage*, this kind of storage doesn't suffer from the latency issues that plague traditional hard drives. However,



where modern hard drives can each hold 6 terabytes (6 TB) or more *per disk*, solid state disks can only storage a fraction of that amount (1 TB, or 1 trillion bytes, for example).

Further, solid state disks remain very expensive compared to hard disks, at least when it comes to capacity. You will pay far more per terabyte for a solid state disk than you will for a hard disk.

Because of this, many companies desire to leverage both kinds of storage. They use hard disks for capacity benefits and solid state disks for performance benefits. This mash-up is called a *hybrid storage solution*.

SimpliVity's OmniCube platform leverages hybrid storage right out of the box to provide customers with the best of both worlds when it comes to storage.

#### The Data Efficiency Amplifier

Understanding how SimpliVity's storage system works is important, but there's more here to leverage.

Data efficiency actually goes even further to help accelerate business workloads. As new data is saved, the Data Virtualization Engine continuously scans that data in real time. If the engine finds a duplicate piece of data — that is, if it finds that the particular data being saved already exists in the storage system — the engine just discards that save operation with no detrimental effect on the user or on the data. The end result is that the data-save operation never even needs to happen. This process is called *deduplication*.

Deduplication reduces the total amount of data stored, and is typically expressed as a ratio, like 5:1. The performance gains made possible by this feature are enormous; after all, how can things get any faster than not having to do any work at all?

The more data that can be reduced, the more that SimpliVity's Data Virtualization Engine can support.

#### BACKUP AND RECOVERY

In many companies, backing up critical business information remains an overnight event. In other words, it happens once a day. What does that mean in the event that data needs to be recovered? In essence, it means that the company has opted to operate with a 24-hour *recovery point objective* (RPO). In the world of IT, the RPO is a measure of the time and the amount of data that an organization is willing to sacrifice in the event that there is a problem that results in data loss. In this example, that means that many companies are willing to simply give up a full day's worth of work if there is a total loss of primary storage.

One of the biggest reasons why companies have accepted this abysmal situation is the fact that backing up all of a company's data takes a lot of time.



The Business Benefits of Data Efficiency

Page 6

Moreover, the entire environment slows down when massive amounts of data move to backup systems. Therefore, there is a reluctance to run these kinds of operations during business hours when IT performance levels need to be high.

Reducing the RPO so that there is the potential for less data loss has traditionally required separate infrastructure (separate storage devices and tape drives) and separate software dedicated to this task. Moreover, traditional operations have indicated that the shorter the RPO, the more expensive the solution.

Enter data efficiency.

Using the 5:1 reduction ratio example from earlier, there is already only onefifth of the stored data that needs to be protected. However, if a company were to try to use traditional backup tools to protect the data stored in the SimpliVity environment, they would find no benefit. This is because the data only stays reduced for as long as it stays within the SimpliVity environment. This is true of all technologies that include data reduction capabilities. So, as the data leaves the SimpliVity environment for the backup environment, it is rehydrated, or expanded back to its original size.

SimpliVity, however, has included backup capabilities in their OmniCube platform. These backup capabilities are designed to work on the *reduced* data and do not require that the data be expanded to be backed up. By using SimpliVity's built-in data backup capabilities, your company gets to work with a much smaller set of data, providing immediate advantages.

Further, rather than having to copy even that full set of data, the SimpliVity backup feature uses a *snapshot*, which is a quick point-in-time "picture" of what data looks like, in order to accomplish its data protection goals. As such, there is almost no impact on overall performance. These two facts mean that backups can be performed throughout the day, reducing that RPO to minimal levels and significantly reducing the risk of lost data. Even better, you can store hundreds or even thousands of these point-in-time backups without using a lot of storage capacity, thanks to the fact that deduplication works even in the backup process.

There is another metric that's important here: *recovery time objective* (RTO). RTO is the amount of time that it takes to bring a service back into operation if a problem arises that requires recovery.

With traditional systems, recovery time can be lengthy. In many cases, extended downtime for recovery translates directly into lost revenue.

With SimpliVity's data efficiency technology, recovery time is cut into a fraction of what would normally be required. This translates to immediate savings in the form of reduced impact on revenue if (and when) a recovery operation needs to take place.



#### DISASTER RECOVERY AND THE IMPACT ON THE NETWORK

Backup tools are usually used to back data up to local storage and are the simplest option in the data protection spectrum. Further along that spectrum lies disaster recovery. *Disaster recovery* is a set of services that enables an organization to quickly resume operations in the event that a disaster strikes the primary data center. Although disaster recovery is a critically important component of the data center arsenal, many organizations do not adopt disaster recovery processes due to the costs and complexity that can be involved. Disaster recovery is another area in which data efficiency can both reduce costs and make it possible for companies that could otherwise not afford disaster recovery to bolster their data protection capabilities.

#### WAN Acceleration

One of the most significant challenges in implementing disaster recovery is getting the data from the primary data center to a secondary one. This may sound simple, but think about that data storage iceberg from earlier. *All* of your organization's data, what the user sees as well as everything that supports it, has to be present in the secondary data center in order for the business application to operate.

Traditionally, IT departments have had to resort to devices known as *WAN accelerators* to even think about moving this amount of data. For disaster recovery purposes, WAN accelerators exist in both the primary data center as well as the secondary one. Workloads that are included in the disaster recovery plan are then replicated across the company's connection to the secondary data center.

The WAN accelerator in the primary data center is responsible for compressing the data being copied. By compressing the data, much less information needs to transfer across the connection. Bear in mind that, for most companies, the connection between sites is much slower than the connections that reside within the data center. As such, the speed of data transfer is a real concern when it comes to disaster recovery.

In the secondary data center, a receiving WAN accelerator accepts the compressed and deduplicated data and rehydrates (expands) it back to its normal form before storing it on that data center's storage devices.

Now, think about a scenario in which SimpliVity OmniCubes exist in both of those data centers. In the primary data center, the company's data is already stored in an efficient state so the organization doesn't need to buy WAN accelerators to improve the performance of the data transfer. As long as the data resides within the SimpliVity domain, it stays efficient. As such, organizations can implement disaster recovery systems that don't then require the addition of extra hardware and software, leading to a less expensive, less complex disaster recovery operation – without the need for WAN accelerators.



The Business Benefits of Data Efficiency

Page 8

### **Summary**

Very often, business decision makers have to make choices about technical infrastructure without fully understanding the potential benefits or consequences that may result from a particular acquisition.

When it comes to hyperconverged infrastructure, there are enormous benefits that can be achieved in the areas of cost and complexity. The simpler that the data center is to manage, the more time that IT staff can spend on business-critical tasks rather than technical ones.

Further, thanks to SimpliVity's data efficiency technologies, you can enjoy high-end data protection capabilities and can support more and bigger workloads in the data center without having to buy all costly and complex ancillary hardware to support those efforts.

#### **ABOUT THE AUTHOR**

**Scott Lowe** is co-founder of ActualTech Media and the Senior Editor of <u>EnterpriseStorageGuide.com</u>. Scott has been in the IT field for close to twenty years and spent ten of those years in filling the CIO role for various organizations. Scott is also a micro-analyst for Wikibon and an InformationWeek Analytics contributor. In addition, Scott has also written thousands of articles and blog postings and regularly contributes to such



sites as TechRepublic, Wikibon, and virtualizationadmin.com. Because of his unique blend of skills (CIO/strategic & Engineer/tactical) Scott is also a soughtafter resource for marketing collateral, including white papers and e-books, for a variety of technology firms.

