Snapshots

for dummies®
A Wiley Brand

Solve data gravity issues

Maneuver data nimbly

Protect data with confidence

Melody Zacharias

Janet Lafleur

Ted Coombs

Pure Storage Special Edition

About Pure Storage

Pure Storage (NYSE: PSTG) uncomplicates data storage, forever. Pure delivers a cloud experience that empowers every organization to get the most from its data while reducing the complexity and expense of managing the infrastructure behind it.

Pure Storage all-flash storage solutions simplify and accelerate the move to always-on, user-centric digital business operations — internal and customer facing — with an always-improving infrastructure that enhances how data is stored, mobilized, and protected.

Whether you choose to own or consume it as-a-service, Pure's unlimited scalability and Evergreen subscription-to-innovation model lets you take control of your applications while lowering your total cost of ownership (TCO). You can also rapidly deliver a stable and reliable data experience to customers and staff — no matter where they are.

Pure Storage's portfolio of solutions also helps organizations accelerate their digital ambitions without friction or compromise by future-proofing IT operations, enhancing existing workloads, and unlocking new ones like real-time analytics, artificial intelligence (AI)/machine learning (ML), the Internet of Things (IoT), and more. With a certified customer satisfaction score in the top one percent of B2B companies, Pure's ever-expanding list of customers are among the happiest in the world. For more information, visit **www.purestorage.com**.

About the Authors

Melody Zacharias is a senior Microsoft solution manager at Pure Storage. You can connect with her at sqlmelody.com or on Twitter @SQLMelody. **Janet Lafleur** brings 20 years of storage hardware, data management software, and cloud experience to her role as senior product marketing manager at Pure Storage. **Ted Coombs** is a multidisciplinary scientist with a long IT career who has authored many books over the last three decades.

Snapshots

Pure Storage Special Edition

by Melody Zacharias, Janet Lafleur, and Ted Coombs

for
dummies®
A Wiley Brand

Snapshots For Dummies, Pure Storage Special Edition

Publisher's Acknowledgments

Table of Contents

Introduction

Massive amounts of data are created every day, and that amount is continuing to grow at an exponential rate. Data is being collected now more than ever to meet the demands of data analytics, artificial intelligence, and data transformation for government and industry compliance regulations and considerably more. This over-abundance of data brings challenges of saving, moving, using, and protecting it. Volume snapshots help make managing this data much easier.

Snapshots make many data management tasks — disaster recovery, data migration, database cloning and archiving—more efficient. You can take a point-in-time copy of data, shortening development and testing (dev/test) or backup and restore cycles from hours and days to moments. You can even use them to selectively retrieve data out of database backups that needs restoring instead of performing a full restore.

Data migration to new platforms, new devices, or moving data to the cloud is made far simpler and more efficient by using snapshots. The ability to quickly and easily roll back to previous snapshots makes them the perfect solution in a dev/test environment. Avoid testing on the production database while having all the benefits of using production data. The ability to scale applications using snapshots makes them the ideal multipurpose solution for the growing data center.

Snapshots are space-efficient, fast, and secure. The security of snapshots even helps protect against ransomware attacks by creating snapshots that are immutable and destroyed only when your company decides they're ready to go, making it impossible for hackers to destroy these snapshots.

About This Book

This book introduces you to using snapshots created by using the Pure Storage FlashArray and Pure Cloud Block Store products. You discover how easy it is to scale applications, protect them from

disaster and malware such as ransomware, and improve your development team's ability to quickly roll back using snapshots.

Icons Used in This Book

Like most *For Dummies* books, you find icons in the margins that help you spot important information highlighted in this book.

The Tip icon points out helpful information.

Find information marked with the Remember icon that you may want to file away for later use or remind you of important details.

The Technical Stuff icon points out bits of information that are of a more technical nature.

The Warning icon alerts you to information that may save you from making decisions that may be harmful.

Beyond the Book

Entire volumes could be written about the topics covered briefly in this small book. For information beyond what I have given you in these pages, visit purestorage.com/accelerateyourdata.

Other helpful resources include

- » **Real-World Data Heroes e-book:** bit.ly/3B1uv18

- » **Using Databases with FlashArray Volume Snapshots:** bit.ly/3ebu37R

- » **Test Drive Modern Database Storage:** bit.ly/3Rpvs93

- » **Pure Code Developer Community:** bit.ly/3pX15v1

Chapter **1**

Introducing Snapshots

The impact of meeting massive data collection challenges can't be underestimated. This is particularly true when it comes to backing up data so it can be recovered if necessary either due to a disaster or through the normal course of application development and testing (dev/test).

In this chapter, you discover data gravity and its impact on data.

Discovering the Challenges with Data Gravity

Data gravity is a concept that says that data and applications are attracted to one another in a way similar to the law of gravity. It is a term first used by software engineer David McCrory and states that as the mass of data increases, so does the level at which the quantity and speed of services increases along with the number of applications and customers.

Seeing the consequences of data gravity

REMEMBER

Artificial intelligence and machine learning increasingly add to the massive data collection by enterprises. These are forming the basis for new insights, intelligent decision-making, and innovation. As the amount of data in storage increases, it becomes increasingly difficult and expensive to move.

WARNING

Data accumulation, both structured and unstructured, creates some harsh consequences:

» Innovation is hindered.

» Performance is limited.

» Productivity is reduced.

» Data storage costs increase.

Organizations going through data transformations and preparing for increases in data analytics require massive data sets to be truly successful. As enterprises grow these data sets, they become costly and risky to move among various storage systems. This becomes even more challenging when data sets reside in the cloud, on-premises, or both.

TECHNICAL STUFF

The data footprint of enterprises is growing at an incredible rate. According to Forbes Global 2000, enterprises will create data at a rate of 1.1 million gigabytes per second, requiring 15,635 exabytes of additional data storage per year.

Defying gravity

Defying gravity requires a bit of magic. Your first action should be to move the data as close to the applications and services that create and use the data. This step also means breaking down data silos within the enterprise and moving data to a single, consolidated storage platform, which reduces the distance and time in data set processing.

TIP

By using a data-centric architecture, enterprises can support both traditional and next-generation workloads and applications. When using modern flash storage, you can have the following:

- >> Built-in data reduction with high data reduction rates
- >> Reduced complexities of data silos using a high-performance, scalable data platform
- >> Increased data and application mobility across cloud and on-premises deployments
- >> The ability to scale and modernize without limits

Differentiating Backups and Snapshots

People are most familiar with computer backup and restore. This process involves periodically copying your data, sometimes to a different location or media type, so you can later restore your data in the event of catastrophic data loss. This type of data loss can occur during a power outage, ransomware attack, or other disaster. This practice is considered part of most business continuity and disaster recovery plans. Restoring data allows you to roll back your system to the time of your last backup.

REMEMBER

You can back up to the same data storage device, but best practices suggest that backing up to another device gives you the best chance at recovery in the event the principal storage device is compromised.

Choosing a backup type

There are various types of backups. The four most common backup types are

- >> **Full-image:** This backup makes a complete copy of the data that allows a full restore. This process can be lengthy in large data sets.
- >> **Differential:** This type copies only the changes made since the last full-image backup. Restoring requires both the full-image backup file and the differential backup. This type of backup consumes fewer resources.
- >> **Incremental:** This backup type starts with an initial full backup and continues in smaller incremental backups, each starting from the last change. As long as these incremental parts are applied in the correct order without gaps, it constitutes a full backup. This has the benefit of allowing for specific point in time recovery.

CHAPTER 1 **Introducing Snapshots** 5

>> **Real-time:** Data is copied continuously to another device in a manner sometimes known as a *hot backup.* This method is the most granular and comprehensive backup but also the most resource intensive.

Using snapshots for the quick pic

Snapshots copy the state of a system as a virtual picture at a particular point in time. Think selfie. Unlike a backup, snapshots copy the settings and metadata necessary to restore your data in case of a disruption, not the data itself.

When systems are restored using a snapshot, they're recovered to the point in time the snapshot was taken. These are best used for short-term storage. For this reason, they're often used during dev/test when frequent roll back is required.

TIP Use snapshots for a quick save of your system before a major update. If you're unhappy with the update, roll back to the version saved in the snapshot. And remember, source files for snapshots need to be stored in a different location to retrieve them.

Defining Architecture Basics

To best understand how snapshots work, you need some basic concepts. Understanding these helps you see how truly compelling snapshots are.

Two kinds of snapshots

There are two kinds of snapshots: an application layer snapshot and a storage layer snapshot. They both are a reference to the contents of an object at a point in time. This can be either an application such as a database or virtual machine, or a snapshot of the actual storage volume. Combining an application layer snapshot with a storage layer snapshot provides additional consistency and guarantees for the state of an application during recovery.

While these snapshots are similar, they have some important differences. Applications layer snapshots place an object into a state

where that state can be guaranteed in the future. It retains all the settings and application state at a particular point in time. It's transparent to existing and future running operations while also tracking the changes between the existing snapshot and the current state of the application.

In an application snapshot, you have overhead in maintaining the state of applications. For this reason, it consumes considerably more storage space than a storage layer snapshot (also called a *volume snapshot*). It's recommended that application snapshots be retained for the shortest period of time.

Storage layer snapshots, particularly those created using the Purity operating environment on FlashArray or Pure Cloud Block Store, don't have the same overhead as application snapshots. You can save storage snapshots for longer periods of time, giving them all sorts of uses. Volume snapshots are thin-provisioned, deduplicated, and compressed (capacity reduction), and require no capacity reservations.

Virtual RAID

In the not-so-distant past, people striped data across multiple redundant hard drives in a way that allowed data on any one of them to be restored in the event one of the drives died. This group of hard drives, called Redundant Array of Independent Drives (RAID), also provided a fragment of performance enhancement as access to the drives was distributed among them.

Flash array storage is significantly faster, less costly, and more dependable than spinning magnetic hard drives.

In the past, RAID disks would be configured for different server types, such as database servers, email servers, and file servers. This process was inefficient because you couldn't predict how much space or performance each server needed. Storage largely remained unused until there were large demands — for example, nightly backups or chkdsk operations. Storage vendors then had pools of storage and striped your data across many drives. The storage devices simply kept track of where the data was stored. It was much faster, and it led to huge advances in storage.

Letting the SAN map data storage locations

The Storage Area Network (SAN) promotes blocks to faster storage and also keeps track of where your data is stored. You can update your data in a couple of ways:

» Update the data exactly where it lives on the drive, overwriting the existing data.

» Write data to a new location, leaving a copy of the old data.

The storage keeps track of the history of the content, allowing you to roll back to previous versions of the data.

Creating new versions and being able to roll back is the same way that software code repositories work.

Writing time-stamped new copies of data

Writing new copies of data as it changes allows for roll back to previous versions. This does require that you have enough space because changes to the data will all be written to new locations.

Snapshots versus SAN clones

Snapshots are copies of point-in-time maps (the pointers to the underlying raw data map), not copies of data. Copying or cloning the data takes a great deal of time while copying the map as a snapshot is fast and efficient. Efficiency increases because you don't need to copy the entire map; you only copy the parts where data has changed. You can then keep as many snapshots of your data as your storage allows.

After you have all these snapshots, restoring becomes as simple as switching maps. No data is rewritten. You're simply pointing to data that has already been written.

When restoring databases, it's most common to need only partial sets of data, such as a data table that was improperly dropped or an update gone wrong. Snapshots let you grab only the data you need instead of rewriting an entire database.

Getting to Know Snapshot Types

In this section, you discover more distinguishing characteristics of different types of snapshots. Although there are two basic types, these are further differentiated into categories by how they will be used. The next few sections delve deeper into how to work with snapshots and why each type has its own use case.

Working with crash-consistent snapshots

A *crash-consistent snapshot* captures the state of the data whenever a crash occurred, whether it was a hardware failure or the dreaded power failure. In the case of crashes, most of the data will have been written to disk, but some of the data may have been in transit or waiting in memory to be written to the disk. This data may be lost. This is usually fine for projects such as print queues where data can be resent.

Crash consistent snapshots are less draining on your resources but don't provide a perfect recovery. Therefore, they aren't suitable for databases that require point in time recovery. Recovery using crash consistent snapshots is usually possible, but not in every case. Recovery will possibly include integrating or replaying a log file into a database.

REMEMBER

When deciding to use crash consistent snapshots, you should weigh the pros of lower resource drain against the risk of possibly losing some of your data in the event of a crash. Check to see which of your applications may automate crash recovery and which ones are mission critical where data loss is unacceptable. When data loss can't be tolerated, crash consistent snapshots are unacceptable.

Applying volume shadow copy snapshots

The volume shadow copy service (VSS) specific to Microsoft Windows initiates the creation of a shadow copy on different volumes so restores can happen in the future. There are clones as well as differential or copy-on-write methods.

TIP

At times, data from application workloads spans across multiple FlashArray volumes, and files can be open or in an inconsistent state, making the coordination done by the VSS service very important. When there's a requirement for the application to have a point-in-time copy, you need to use the VSS.

Employing application-based snapshots

Application-based snapshots are normally based on differencing disks, virtual hard disks that store changes made to a different virtual hard disk or the guest operating system. The reason to use this method is to track changes so you can roll them back if necessary.

TIP

The number one advantage of application-level snapshots is consistency. File-system data snapshots are almost always file-system consistent, not leaving you with partially written files.

Using storage-based snapshots

Storage-level snapshots occur at the block level. They don't incur the performance impact of software-based snapshots. The downside is that it's possible to take a storage-based snapshot while a file update is occurring, leaving you with a potentially corrupt file if a roll back restores the incomplete file.

Chapter **2**

Recovering Data in a New Way

Maintaining data comes with many risks, but the one thing that lets you sleep at night may be knowing that you can restore data in the event of a catastrophe — natural, accidental, or malicious. When it comes to restoring data, you don't want to spend forever doing it. Mission-critical systems could be completely down during a restore, costing time, money, and reputation.

This chapter discusses lowering the time it takes to restore data, making it simpler. I also cover some points on recovering from the dreaded ransomware attack.

Decreasing Backup and Restore Times with Snapshots

Traditional database backup-and-restore systems are powerful but time consuming. Traditional backup methods start copying at the start of a file or file system and continue until it reaches the end. This type of a backup quickly becomes obsolete, because if changes are made while the backup is taking place, those changes

aren't captured. Recovering the data loss from a typical backup system can also keep your business in limbo until the data has been restored. A better way to backup is by using a crash-consistent snapshot, which drastically speeds up recovery times.

Improving backups with crash-consistent snapshots

A crash-consistent backup captures all the data at the same point in time, a snapshot of the data. When you restore a crash-consistent snapshot, you restore all the data up to the point in time the snapshot was taken. Chapter 1 discusses how snapshots work to quickly capture references to data, making them a speedy way to save your data.

While crash-consistent backups are faster than traditional backups, they have some of the same drawbacks. Like an "inconsistent" backup, crash-consistent backups don't capture data in memory at the time of the crash. Snapshots are fast during recovery as well. Traditional backups require that files be copied from a storage location, which takes longer than simply pointing to the snapshot.

Reduce restore times by maintaining regular crash-consistent backups. While not absolutely perfect, this type of snapshot has a lower demand on your resources.

Using a better backup method

An application-consistent backup gives you the ultimate level of protection and consistency. Unlike crash-consistent backups, these also capture not only what is stored on disk but also what is in memory.

In a Microsoft Windows environment, application-consistent backups are done using the volume shadow copy service (VSS). This service freezes any I/O, flushes memory to disk, and then takes a block-level snapshot of the volume.

Simplifying Disaster Recovery

A much simpler recovery method than a traditional database restore is possible if your company uses an all-flash storage array and you've been taking FlashArray application snapshots

regularly. This type of snapshot maintains the entire state of the application at a point in time. This method of restoring involves two simple steps:

1. **Take the database offline.**
2. **Rollback to the last good snapshot.**

Before launching into a rollback, a more measured approach is to see whether you're able to fix the corrupt or missing data using a more surgical approach with a snapshot. You should consider this when most of the application has remained unaffected. Here are the steps:

1. **Limit access, when possible, to the data by setting it to read-only.**

   When this step isn't possible, shut down the application and stop any remove-database access.

2. **Create a clone of the database by using the snapshot.**
3. **After agreeing on the needed fixes with the application owner, apply the fixes to the production database by using data from the clone.**
4. **Restore full access to the production database.**

   Following this approach reduces the overall impact on your applications and any dependent applications.

Recovering from Ransomware Attacks with SafeMode

A ransomware attack is no longer the result of some hacker writing malicious code. Ransomware as a Service (RaaS) now exists that offers pay-for-use malware. When it comes to malware attacks and the potential impact to a business, ransomware attacks can be one of the most devastating. Having your data encrypted or otherwise made unavailable until a ransom has been paid is an ever-increasing way for criminals to monetize malware. You aren't even guaranteed to get your data back after paying the ransom. The only way to safeguard your business from this type of attack is having a good and current backup system such as using snapshots.

In the event your business is suffering a ransomware attack, you want to make certain that recovering from the attack is simple and fast. One of the features built into FlashArray is a data-protection system known as SafeMode from Pure Storage.

SafeMode is a compliance-driven policy for volumes and snapshots. When either a volume or snapshot is destroyed on your system, these objects are saved into a special "destroyed" area. These destroyed objects are then recoverable for at least 24 hours unless an array admin clicks the trash can icon in the GUI and permanently destroys the objects. As a safeguard, a count-down eradication timer slowly clicks down allowing the destruct sequence to be cancelled. Normally, a destroyed object remains recoverable, like an undo, for a full 24 hours.

TIP

Because 24 hours may not be enough time during a ransomware attack window, you should change the timer setting for a longer duration, possibly 14 days. You can set it for periods of up to 30 days. As a further protection, mechanism policy changes to the timer can only be done by contacting Pure Storage support as an authorized contact. See the next section, "Enabling SafeMode" for more information.

One of the steps taken by a malicious intruder is that after they've encrypted your data sets, they will then attempt to delete your production data and all backups. SafeMode makes it impossible to delete the snapshots you have of your data.

When protected with SafeMode, recovering your data is as simple as removing access to the malware-encrypted volumes and replacing them with snapshots. The results are minimal impact to your systems and your company doesn't have to pay a ransom to recover your data.

Enabling SafeMode

In Purity//FA 6.3.0 and beyond, SafeMode can be enabled by customers directly through the FlashArray GUI, CLI, and REST API. To shorten the retention time for SafeMode, you need to contact Pure Storage support. A conference call is then setup between at least two authorized members of your organization and your Pure Storage account team. This extra step ensures attackers can't remove your critical recovery data.

REMEMBER Five contacts from your company can be authorized to make changes to SafeMode, each given a six-digit PIN.

Employing immutable snapshots

Pure Storage snapshots are immutable — they can't be changed. If you use SafeMode, snapshots can't be eradicated, making them an incredibly secure way of restoring data in the event of a ransomware attack. Because snapshots aren't traditional backups, they're incredibly fast to create — it only takes milliseconds to create a persistent data structure. They're placed in groupings of one or more volumes so policies can be applied to them; these groupings are called *Protection Groups.*

SafeMode includes the prevention of hackers to interfere with snapshots being sent to Protection Group targets. Additionally, intruders can't set a snapshot retention to zero, which eradicates the existing snapshots. Protection Groups offer robust and configurable snapshot polices. These policies include

» Frequency of snapshots

» Snapshot retention policy

» Sending snapshots to destinations such as FlashArray//C, FlashBlade, AWS, Azure, S3, and NFS shares

REMEMBER Snapshot retention can be increased, but it can only be decreased by two authorized admins using their PINs.

Chapter **3**

Moving Data with Snapshots

Migrating data isn't as common as backing up for data protection, but it's still part of the everyday operations of businesses as they move to new platforms, into the cloud, or a hybrid of both.

The process of migrating data from FlashArray to FlashArray, FlashArray to Cloud Block Storage and from Cloud Block Storage to another Cloud Block Storage instance is similar and easy to do. I cover those platforms in this chapter.

Upgrading to New Application Platforms

One of the scenarios you may find yourself faced with is upgrading your database to a new platform. The traditional and time-consuming way is to take a backup of your database and restore it onto the new platform. Consider using snapshots for this purpose instead.

Maintain application consistency

To maintain application consistency, the application needs to be placed in a state in which it can be ensured that recovery is possible until the point in time that it needs to be recovered.

Readying a database, for example, requires that you ensure that all changes are flushed to the physical files, and then you need to make sure to stop any new changes. You can also redirect changes to a different set of files.

After the database has been readied, an application consistent storage snapshot can be created. This snapshot can later be recovered to a new or different platform.

Using snapshots during migration

When migrating an Oracle database to new architecture, for example, snapshots can be used in more than one place. The first one is more of a safety measure. You take a snapshot of the protection group on the source, including volumes for the source database, to enable the ability to restore volumes in the event something goes wrong in subsequent steps of moving the database files. The snapshot then allows the process to be rolled back.

When ready to perform the actual switch of the database to the new environment, follow these steps:

1. **Capture the source metadata into an XML file.**
2. **Take a snapshot of the protection group covering the source volumes.**
3. **Copy the snapshot to the new volumes.**

   After the snapshots are copied, you can attach the volumes to the target host. Mount the new disks and transfer the XML file.

Migrating to the Cloud

Cloud services like AWS, GCP, and Azure have tools for migrating databases. After databases have been migrated, they're often missing some of the toolsets previously provided. If you use FlashArray or Pure Cloud Block Store, you can avoid this problem.

In this section, I give you an example of moving an on–premises SAP HANA database to the cloud using FlashArray and Pure Cloud Block Store for AWS. While this example is specific to an SAP HANA database and AWS, similar steps are used to move other database types to different cloud providers, such as Azure.

TIP

SAP HANA can be used as either a development and testing (dev/ test) database once migrated to the cloud or for running analytics due to its unique column table structure. These processes are fine– tuned for both online transaction processing (OLTP) workloads and online analytical processing (OLAP) analytics.

You want to begin migration by using the following steps (performed on–premises):

1. **Connect the Pure Cloud Block Store on AWS to the on-premises FlashArray as an asynchronous-replication target.**

   There is a protection group for the on-premises SAP HANA data volume. This protection group is configured to be replicated to the Pure Cloud Block Store.

2. **Freeze the data volume's file system to an application-consistent snapshot.**

3. **Configure your pgroup with a replication schedule or use replicate-now.**

4. **Take an application-consistent snapshot of the on-premises SAP HANA system by putting it in back up mode (snapshot mode).**

5. **Unfreeze the file system and exit the SAP HANA backup mode.**

   Data is asynchronously replicated to the Pure Cloud Block Store.

6. **After replication is complete, take the following steps to recover the SAP HANA instance on AWS:**

   a. Take the snapshot created from the on-premises FlashArray and apply it to the SAP HANA data volume.

   b. Start the recovery process for the SAP HANA instance on the AWS EC2.

   After the recovery is complete, use the SAP HANA system in the cloud as a dev/test instance or build star schemas on the column stores for analytics.

Moving a VMware Volume or File System to AWS

In this section, I give you a case where a VMware virtual machine stored on a vVol datastore and is moved to AWS using the AWS Application Migration Service (MGN) and the replication technology of FlashArray and Pure Cloud Block Store.

The MGN is used to migrate and convert the VM to AWS. This includes its boot volume. Array-based replication will then replicate the data volumes from FlashArray to Pure Cloud Block Store.

REMEMBER

Data volume replication requires that the VM's Virtual disk be backed by a vVol or RDM. The replicated data volume is then presented to the new AWS EC2 instance through in-guest iSCSi. iSCSi is a storage network standard for linking data storage facilities. It provides block-level access to storage devices by sending commands over TCP/IP.

Here is a high-level overview of the steps necessary to perform this migration:

1. **Verify that any prerequisites have been met. You can do that here: https://bit.ly/3EdjQm0**
2. **Install the AWS replication agent on the source machine.**
3. **Configure the agent to ignore desired data vVols.**
4. **Launch the replicated instance from the migration service and connect the Pure Cloud Block Store volume.**
5. **Confirm that the migration was successful.**

Migrating Data Externally

At times, you may want to move data to an external source. This could be for any number of reasons, such as moving data centers or moving data into an external source for data analysis. Moving

data out of the network is known as *data egress,* and when you move data using snapshots, you save a considerable amount of time and effort.

Moving data between clouds

Pure Cloud Block Store has an abstraction layer that allows applications to be agnostic to the private or public cloud on which it runs. This enables customers to seamlessly move their data between all the major public cloud vendors such as AWS and Azure using the native replication capabilities of Pure Cloud Block Store.

REMEMBER Pure Cloud Block Store data reduction is preserved when data is replicated, so, the required bandwidth, time, and any potential data egress costs are reduced.

Migrating with array-based replication

Array-based replication lets customer copy and migrate data between Pure Cloud Block Store instances and FlashArrays. Pure Cloud Block Store supports bi-directional asynchronous replication, continuous replication with a feature called ActiveDR, and synchronous replication by using a feature called ActiveCluster.

TIP Advantages of using this technology include

- » Reduction of underlying data footprint
- » Reduction of data bandwidth requirements
- » Deduplicated and compressed data prior to replication
- » Lower egress charges from the public cloud vendor due to deduplication

Metadata is shared between replicating parties that prevents data from being replicated from the blocks already existing on the target volume. Data can be replicated between FlashArrays and Pure Cloud Block Store instances only while there is a connection between them.

Moving to non-Pure Storage targets

If you aren't moving data between FlashArrays or into the cloud using a Pure Cloud Block Store, there may be times when you can

still take advantage of the FlashArray's capabilities when moving a block storage volume to a separate network file system (NFS) or cloud storage such as Amazon S3 or Azure blob storage.

Offloading volume snapshots to NFS or cloud storage utilizes protection group policies. Create a protection group on one or more FlashArrays so the volumes in the group are protected to one offload target. When a snapshot is created and an offload policy is defined, the snapshot will be copied to the non-Pure Storage target. The protection group snapshot is a point-in-time consistency group. When moving to a non-Pure Storage target, begin by provisioning an NFS share or cloud storage remotely. The protection group will send snapshots and store them on the target storage. It will use the same policies as FlashArray-based protection groups. These include both the replication interval as well as the retention settings.

Snapshots offloaded to non-Pure Storage targets are self-defining and will be recoverable to any other FlashArray or Pure Cloud Block Store in the event of a disaster. Alternatively, Pure CloudSnap can be used to replicate snapshot data to Amazon S3 or Azure blob storage as a low-cost disaster recovery option.

Chapter **4**

Integrating Snapshots into DevOps

One of the great things about snapshots is how well they work for development and testing (dev/test). With fast roll back you can quickly develop application features, test them, and then instantly roll your database and environment back to where you started.

Creating a dev/test environment that's isolated from your production database is made easier by using the features of FlashArray and Pure Cloud Block Store. In this chapter, I cover how this works.

Developing Applications Using Snapshots

Developers have long understood the need for the ability to roll back during development. Regular testing and debugging requires point-in-time recovery. In fact, this feature for code roll back is built right into the source code control software. While reverting code to previous versions has been around for a long time, it hasn't always been as easy to create a development-friendly database.

This database needs to

>> Mimic what's happening in the actual production database.

>> Easily return to previous versions without massive data copying.

TIP While some developers use the production database for development and testing, it makes better sense and lowers the risk of damage to your mission-critical data if you develop a testbed database. Snapshots, simply metadata, make it simple to roll back to previous snapshots without the need to overwrite data. They also make it fast and simple to update data tables used in your development environment.

Developing a cloud-based test database

You have a choice, as a developer, whether you want your test database local or hosted in the cloud. Both scenarios work. More development is moving to the cloud to simplify global development team collaboration.

Cloud service providers have physically separate locations known as availability zones. These have their own power and network connectivity. They are also regionally located so network latency is reduced. Setting up your development environment in a new availability zone can be ideal. Check out Figure 4-1.
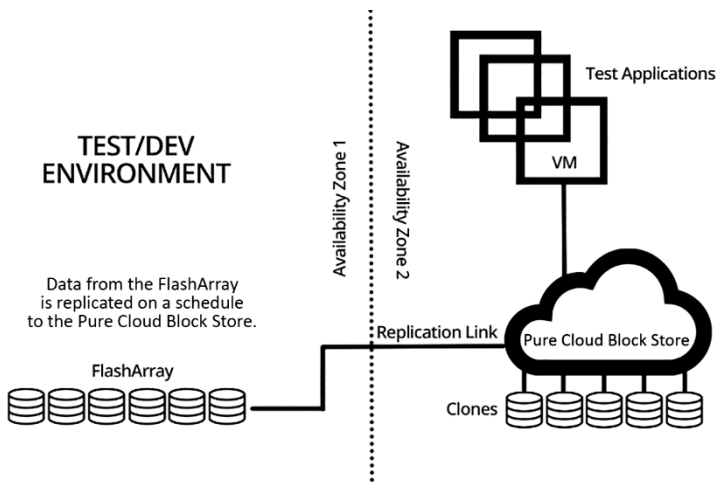
FIGURE 4-1: Data is moved from a FlashArray or Pure Cloud Block Store into the new availability zone used for development.

Setting up in a new availability zone

Choosing a different availability zone or data center for dev/test is done so your development workload, with its dynamic changes, isn't happening in the same environment as your production workload. Of course, it's possible to have them both running in the same location, but it's not recommended.

TIP Isolating your development and production environments creates a safe place to do development and lowers the impact on the production environment.

Within an availability zone, you can also create a virtual private cloud (VPC) that uses the resources of a cloud provider such as Azure to create a virtual network. This VPC will be similar to one you may have on-premises, but it has the advantage of being scalable.

Cloning for the Dev/Test Pipeline

To start cloning for your dev/test pipeline, you want to create a replication link between your primary data source — for example, your production database — and secondary data location, for example Pure Cloud Block Store. After that, this section gives you the remaining steps.

Setting up a replication link

You want to set up your replication link so new clones are sent to the development environment on a schedule that best meets your needs. This could be done during off-hours, nightly. The development database is then refreshed every 24 hours.

REMEMBER Data transfer between the two storage systems, the FlashArray and the Pure Cloud Block Store, is done using asynchronous replication. It is advantageous to have both the physical and the virtual cloud systems sharing the same features, as do all Pure storage products.

Replicated snapshots of the production database gives you copies at the granularity and the location you need for development. You can make as many copies as you want without increasing

the data storage requirement. You can also make multiple dev/ test replication copies from one production source to horizontally scale development and improve productivity and increase testing throughput.

Provisioning copies of your development project

Provision copies of your application or development environments running your application in the cloud. These can be provisioned onto VMs or into containers, shown previously in Figure 4-1 in section "Developing a cloud-based test database." Depending on your application, both work equally well.

Linking your test application to the replicated database instances

As development and testing are done against the clones in the Pure Cloud Block Store, the test databases can be restored instantly because the FlashArray and Pure Cloud Block Store snapshots are instantaneous. They're also space saving and easily overwritten in a single operation.

To link your source and development environments, follow these steps:

1. **Get a connection key from the local array and the Pure Cloud Block Store.**
2. **Create an asynchronous replication connection between the two environments using the key.**
3. **Create a protection group on the source array and add the source volumes.**
4. **Set up a replication schedule.**
5. **Connect the replicated volumes.**

When you use Pure Storage FlashArray and Pure Cloud Block Store to create your development environment, you reap the following benefits:

>> **Save space:** When you make copies no additional space is consumed on the array. The space used is equal to the original dataset.

>> **Rapid transfers:** Because the transfers are simply metadata operations, they take no time at all.

>> **Easy overwrites:** As new data arrives or if there's a new copy you want to source, you can easily overwrite the volume that's in place.

REMEMBER

The DevOps pipeline ensures quality by heavily testing applications as they're being developed. You may also want to set up your automated testing software in the same environment to make the pipeline more efficient. Using data cloned from production ensures that your application will operate correctly when released and operating against the production database.

Chapter **5**
Scaling Out Applications

Horizontal scaling of database applications refers to adding more databases or dividing your database into smaller chunks across additional nodes to handle increased workloads. This differs from scaling up, which means adding additional computing power to a single node.

In this chapter, you discover some of the challenges of scaling out and how to overcome them. Getting data into the initial replicas is one of those challenges best handled by using snapshots to move data.

Growing Database Applications Horizontally

The decision to scale horizontally usually comes when adding additional compute resources to a database server is no longer an option. When databases are spread across additional nodes, those nodes can be scaled up or down as needed. This combination of scaling horizontally as well as vertically has become common.

Horizontal scaling provides high availability by ensuring that a loss of a database node won't disrupt access to the database. Members of the same replication topology exist in two categories:

» A primary database system

» The secondary database, also known as a *replica,* accepting only read operations

This type of horizontally scaled topology improves an application's responsiveness and read efficiency.

Seeding Database Replicas

Seeding is populating a database with initial data. One of the challenges with adding this initial information is that moving data is time-consuming. With snapshots for initial seeding, you save time.

When a database needs to have its data copied for the purpose of seeding, the copying process can be done using a volume snapshot to a protection group, or to a new volume and attaching it to the system where the replica will live. This process can be performed using crash-consistent snapshots with an entire availability group, in a single protection group, or using a volume snapshot.

TIP

When adding members to an existing replication topology, you need to perform an initial data synchronization. This process involves using a donor database to provide data to an intended replica.

Using MySQL or MariaDB as examples, several replication technologies are available to create replication topologies. Here are three of them:

» **Asynchronous and semi-synchronous replication:** This process allows secondary database servers to receive updates from a single primary database.

» **Group replication:** This is virtually synchronous replication between replica systems with automatic primary election. This is a core component of InnoDB cluster. This can be implemented in both a single or multiple primary database mode.

>> **Galera cluster:** This is virtually synchronous replication between three or more nodes with each node in the cluster serving as a primary. Virtually synchronous replication is the update of several data repositories simultaneously.

Replication technologies exist in most popular databases. The following gives you some databases and the associated replication technologies:

>> Microsoft SQL: Always On availability groups or Log Shipping

>> Oracle Database: Data Guard/Physical Standby

>> PostgreSQL: Log Shipping or Streaming Replication

>> MongoDB: Replica Set

As you add new members to a replication topology, you must first perform an initial data synchronization. This process involves using the primary or donor database, sending the data to the new replica (recipient or joiner) as shown in Figure 5-1.
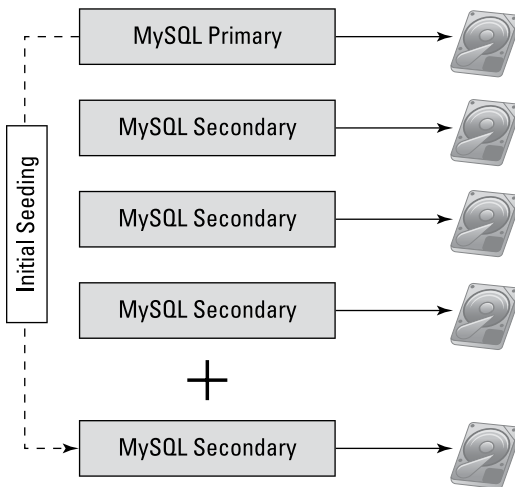
FIGURE 5-1: Performing a data synchronization from the primary.

WARNING

The initial seeding of a database can negatively impact an application's performance due to data transfer time. When data is written or read, a block is placed on the database to avoid data conflicts. When read and write blocking occurs during the seeding process, particularly on busy systems, the replicas being written to may struggle to keep up with changes to the source database.

Volume snapshots on FlashArray eliminates the need to perform initial seeding for replication, group replication, and galera cluster nodes for MySQL or MariaDB. You simply need to take a volume snapshot of the database (check out the next section).

Employing Volume Snapshots in Purity

Purity//FA is the heart of FlashArray's data services and does more than take snapshots; it intelligently manages your data, similar to an operating system. Volume snapshots in Purity are immutable, point-in-time images of your data in one or more volumes. They can be created individually or as part of a protection group.

Volume snapshots on FlashArray avoid the need for initial seeding for replication. Adding a replica is as simple as taking a volume snapshot of the volumes on the primary database instance. After you've taken the snapshot, you copy it to a new or existing volume to be used as the intended replica. You can perform these steps manually using the Purity Graphic User Interface (GUI) or command-line interface (CLI). You can automate it using the Representational State Transfer Application Programming Interface (REST API).

TIP

MySQL Group replication using volume snapshots also eliminates the limit on a donor transferring to only one replica at a time.

FlashArray volume snapshots are faster and more efficient than other initial seeding techniques. With FlashArray volume snapshots, there are a couple of differences:

» With group replication and galera cluster for MySQL or MariaDB, the use of the clone plugin or system suitability test (SST) is automated for new nodes/instances joining the topology.

>> The speed of using volume snapshots with FlashArray is limited to how quickly a user can perform the steps using the GUI or CLI. If the process is automated using the REST API, the limit is the speed of code execution.

REMEMBER All replicas on the same FlashArray will consume no additional storage space.

Cloning for Application Performance

Clones of databases are made for several reasons such as migrating copies for development and test, the creation of reporting databases and intra-instance Extract Transform Load (ETL) scenarios. FlashArray management can be used in place of native database tools for cloning and copy operations, particularly when point-in-time recovery for cloning is required. Volume snapshots are appropriate to use when point-in-time recovery for cloning isn't required.

TIP Microsoft SQL Always On availability groups can be created to provide high availability or read scale. Before you can improve availability or read scale, you must first copy the data to a new instance by copying a volume snapshot or protection group to a new volume and then attaching that volume to the system where the replica will live. Figure 5-2 shows you this process.
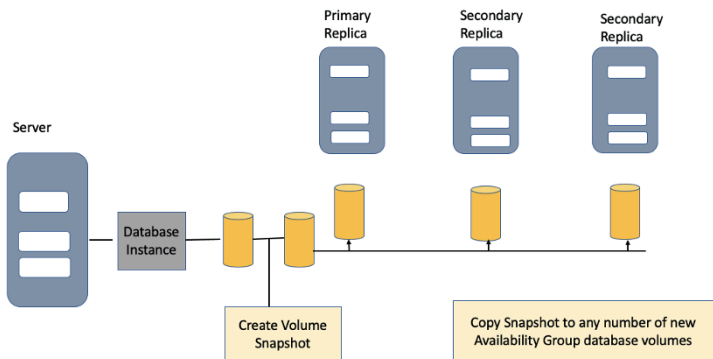
FIGURE 5-2: Create a snapshot and copy it to new availability group database volumes.

REMEMBER

Crash consistent volume snapshots can be used when point-in-time clone recovery isn't required.

For Always On availability groups, keep in mind the following about snapshots:

» Not until SQL Server 2022 could you seed an Always On availability group with a crash consistent snapshot. New functionality was added in the 2022 version.

» You can snapshot and clone a database in an Always On availability group if it's on the primary. You can't on secondaries.

Chapter **6**
Improving Archives

Archiving data simple and efficient. Whether you use Pure Storage or simple file storage from any of the cloud service providers, saving point-in-time snapshots makes perfect sense.

Snapshot Parking with Pure Storage FlashBlade

Pure Storage FlashBlade is an all-flash storage designed for fast file and object access for critical applications such as high-performance computing applications and artificial intelligence (AI) workloads. But necessity is the mother of invention, and use of FlashBlade has grown beyond its original intended use. Its ability to restore as many as 1,000 virtual machines in hours, compared to weeks using traditional backup devices, has made it the go to virtual machine (VM) backup solution.

FlashBlade supports from 7 to 150 blades in a chassis making it super-fast and elastic. The Purity operating system also provides a low-cost file storage solution.

The Purity operating system for file and object access in FlashBlade arrays strengthens snapshot protection against ransomware

attacks and brings new remote site replication functionality. SafeMode, in Purity//FB 3.0 and beyond, write-protects snapshots until their expiration dates. Expired snapshots are then erased, leaving space for new snapshots. This protection was designed to protect them from ransomware (see Chapter 2 for more information on ransomware protection). Protecting the snapshots makes them available in the event a roll back becomes necessary.

REMEMBER

Snapshots are prime targets for ransomware attacks. Getting rid of the backups and ability to restore data often happens before data is encrypted or destroyed. Deleting a snapshot on purpose takes a great deal of effort and the involvement of Pure Storage tech support. This level of protection is what you should be looking for in archival storage of snapshots.

Retaining Data with Snapshots

There are a few reasons why a company wants or needs to retain and archive data. Many companies have a data retention policy that covers the desires of a company to retain data for long-term data analysis or the needs of a company to comply with regulations.

TIP

If you don't already have a data retention policy, you should work on putting one in place.

Recovering from disaster

As I wrote this book, I tried to emphasize how important retaining data is in order to recover from disaster, whether it's the flooding of your data center, a ransomware attack, or some other unforeseen circumstance. While striping your data across several nodes works well for small disasters, like losing a node, major disasters require more retention.

Consider how long you should keep your data around in preparation for disaster recovery, backups, snapshots, and any other data necessary to recover to a point-in-time. This should be part of any disaster recovery plan.

Retaining data for analysis

Data analysis is one of those things where you wish you'd saved more data. Whether it's discovering long-term trends or feeding

the learning data of an AI machine learning algorithm, the more data you have the better the results.

Because you're never sure what data you may need in the future, it's better to fall on the safe side of saving too much. Snapshots are an excellent way to archive data you may never need but will be glad you have just in case.

WARNING

Many folks are still trying to bring data back from magnetic disks and tape. These processes and mechanisms for data storage are fraught with risk and are time-consuming. Instead, FlashArray archival storage is a cost-effective way to maintain data so it's easily retrieved.

Complying with regulations

Most large companies are required to retain data. This includes financial data records, Human Resources (HR) records and regulatory data retention compliance. Safely storing those records means that they're unlikely to be lost and that they're safe from prying eyes.

Most companies are aware of the financial privacy requirements of credit card processing. These companies are ones that process medical data, such as insurance companies, medical record transcription companies, and, of course, hospitals and doctor's offices. For example, Health Insurance Portability and Accountability Act of 1996 (HIPAA) documents must be stored for a period of six years. This doesn't include medical records (which have no requirements to be saved) but does include HIPAA-related documents such as privacy practice notices, medical records disclosure requests, complaint and resolution documents, and more.

Global businesses must also comply with regulations such as those governing doing business with the European Union (EU). The General Data Protection Regulation (GDPR) also spells out security and data retention requirements.

REMEMBER

Data in FlashArrays remains available, even when major array components fail. FlashArray immutable snapshots provide unalterable records of data sets as key points in time.

Different regulations will spell out the data retention requirements, whether it's months, quarters, years, or even indefinitely. The regulation also indicates how often the data needs to be refreshed.

Destroying data

So what do you do when the data retention period has expired? Your data retention plan should include this consideration. Most of this data is sensitive and needs to be destroyed in a way that it can't be retrieved.

Destroying data provides both compliance with regulations on the destruction of data, and it frees up storage. Many regulations spell out the data destruction policies. For more information on these regulations, visit `www.ncsl.org/research/telecommunications-and-information-technology/data-disposal-laws.aspx`.

Each of the regulations lays out any specific requirements as to how data is destroyed once retention terms have been reached. To learn more about the best practices in data destruction, head to `www.ironmountain.com/blogs/2019/7-best-practices-for-data-destruction`.

Chapter **7**

Ten Things You Can Do with Snapshots

T he uses of snapshots are limited only by your imagination. Saving point-in-time snapshots of your data makes them perfect for everything from data migration to disaster recovery. In this chapter, you discover ten things you can do with snapshots.

Protecting Data

Backup and restore operations on large data sets are so cumbersome that companies are backing up less often because of the drag on their systems. It's also true that companies often underestimate how vulnerable their data truly is. Some of the potential ways to lose data are

» Hardware failures

» User errors

» Ransomware attacks

» Natural disasters

These scenarios and more can all result in your data being deleted or corrupted.

REMEMBER

The main difference between a traditional backup, which is copying all your data to a backup location, and using snapshots is that a snapshot is a picture of the state of your system at any one time. These are the settings and metadata needed to quickly restore your system.

Ransomware Recovery

Immutable snapshots are key to avoiding the cost and business interruption associated with ransomware attacks. Take snapshots as part of your cybersecurity plan.

The great thing about snapshots is that you can set their expiration time (when they go away) and it's difficult (impossible for a hacker) to make them expire before their time. SafeMode snapshots involve contacting Pure Storage support that determines the request's legitimacy. It provides an approved code to authorized agents of your company that can then allow a snapshot to expire (disappear) before its planned expiration.

TIP

Set the expiration on snapshots you expect to serve as emergency backups in the event of a ransomware attack to a longer period, keeping them in storage for safekeeping. Hackers will want to try and get rid of data backups before encrypting your company data to enable their ransom scheme. Immutable SafeMode snapshots nip that in the bud.

Check out Chapter 2 for more about ransomware.

Moving Data

Data migration isn't always easy but with the support of Pure Storage products, there is a path to success. The built-in features of the Purity operating system used by FlashArray and Pure Cloud Block Store products make creating snapshots simple and efficient.

One example is migrating on-premises databases to a public cloud. Public clouds provide several tools to discover your data, replicate it, and move it to the cloud, but the toolsets previously provided by SAN storage are lost. These include tools like thin provisioning, data reduction, and instant snapshots.

By using the replication features of the Purity operating system, data sets can be directly replicated from on-premises FlashArray devices to cloud-deployed Pure Cloud Block Store. This means that features aren't lost during migration.

REMEMBER Pure Cloud Block Store offers the same data services as FlashArray.

The AWS and Azure Application Migration Services take care of migrating and converting the boot volume. At the same time, the data volume will be migrated using async replication between FlashArray and the Pure Cloud Block Store. Working together, these two applications offer a simple and efficient migration path.

Scaling Applications

Scaling data applications is an important use of snapshots. Scaling horizontally using volume snapshots on FlashArray or Pure Cloud Block Store allows you to scale your applications without suffering the drawbacks associated with other methods.

The beauty of volume snapshots is that they are thin-provisioned, deduplicated, and compressed (data storage reduction) and require no capacity reservations. Volume snapshots can be created individually or as part of a protection group that ensures that the snapshots created of volumes within the group are consistent with one another.

Making Snapshots Part of Your Development Pipeline

The modern DevOps pipeline consists of a set of automated processes and tools that allows for the collaborative development and deployment of software code to a production environment.

In short, many people will be pounding away on code stored in a versioning source code control system and then using automated testing tools to debug problems, fixing them, and starting the iterative process over again until the code is released into production.

Snapshots used to create test data in a development sandbox allow for point-in-time recovery resetting sample databases back to prior versions. If a bug in development accidentally wipes out a table, or erases the data for the current year, it's simple and fast to restore it. Snapshots help you avoid making new copies of the production data.

TIP

Quickly build sandboxes using infrastructure snapshots and clones to spin up a dev/test sandbox quickly and efficiently without impacting the production environment. Consider putting the sandbox in the cloud.

With the right software-defined storage solution, you can instantly provision production workload clones in the cloud to accelerate development cycles and expand your testing capabilities. Pure Cloud Block Store enables you to make snapshots and clones in the cloud, meaning you don't have to purchase and deploy new on-premises hardware. Also, Pure Cloud Block Store snapshots and clones consume very little cloud storage.

Recovering from Oops

To err is human, or divine; I forget which. But, having snapshots to fall back on when mistakes are made can save a great deal of heartache and possible business disasters. This is similar to having snapshots be part of your development pipeline because testing and debugging are all about finding mistakes in coding. But, it's not the only time mistakes are made.

In the case a database administrator accidentally leaves the WHERE clause from a DELETE command or misconfigures an UPDATE, you want the ability to rollback quickly and efficiently. Take application snapshots frequently so little data is lost.

Creating Data Analytics Resources

Like AI, data analytics thrives on Big Data. In fact, data analytics has led to storage systems such as data warehouses and data lakes. The trend toward long-term data usage has also powered the trend away from allowing data silos within companies. Data reuse is so popular now that companies are building data stores where data is marketed to departments for further use.

Pure Storage products support massive storage efforts, whether it is FlashArray and FlashBlade storage systems used on-premises or Pure Cloud Block Store. Creating data replicas and snapshots makes for some of the most efficient storage you can hope for. Use snapshots to migrate production data to archival storage in Pure Cloud Block Store.

Complying with Regulations

There are many compliance regulations for people doing business, whether it's complying with the rules of doing business in the EU or saving documents related to medical records. One part of many of the compliance regulations usually has to do with how long you need to retain data. Data retention shouldn't have to be a burden. Use the Pure Storage products to make sure they're created easily and stored safely.

A snapshot is a collection of pointers. The underlying data is compressed, deduplicated, and encrypted at rest.

REMEMBER

Archiving that Data

Whether it's normal business data retention or a compliance regulation, you can easily create snapshots of your data. Snapshots are immutable.

Volume snapshots require very little overhead and are compressed and deduplicated. Application snapshots aren't recommended for long-term storage because the general rule is that they should have a short life.

TIP

Use Pure Cloud Block Store to maintain your company archives safely in the cloud. Company execs can sleep better at night knowing their data is safely and efficiently stored using the most advanced storage hardware and software.

Testing Backups

Backups are a critical part of a database system but because of space and time they take as they grow, they are often left untested. Restores, recoveries and tasks like dbcc checkdb are easy to do with snapshots. It is important to remember the story of "Schrödinger's Backup" — the condition of any backup is unknown until a restore is attempted.

PURESTORAGE®

Uncomplicate Data Storage, Forever

Data gravity is increasing—so are performance issues and complexity. Only solutions that face these issues directly, as the foundation for how they work, will enable success. Pure Storage was built to make modern data storage simple so you can leave complexity, and performance issues, behind.

www.purestorage.com/accelerateyourdata

Snapshots are more than a throwaway backup

Your data is gaining in gravity and value all the time, making it more unwieldly and complex. Many copies of that data are often needed — in dev/test, in the cloud, and more. An immutable solution with deduplication and data reduction built in designed for enterprise workloads solves many use cases. With SafeMode added for ransomware mitigation, snapshots are really the overlooked hero of the toolbox.

PURESTORAGE®

Inside…

- Discover the difference between a backup and a snapshot
- Learn when to use which type of snapshot
- Find easy ways to migrate data
- Discover ransomware mitigation ideas
- Employ cloud egress resolutions
- Remedy data archival challenges

Go to **Dummies.com**™
for videos, step-by-step photos,
how-to articles, or to shop!

for
dummies®
A Wiley Brand

WILEY END USER LICENSE AGREEMENT

Go to www.wiley.com/go/eula to access Wiley's ebook EULA.