PURESTORAGE®
Uncomplicate Data Storage, Forever

TECHNICAL WHITE PAPER

Rapid Backup and Restore for Epic Systems Using Pure Storage FlashBlade//S, with Cohesity DataProtect

A guide on how to get the best performance.

Contents

Summary

Epic Systems helps manage a patient's entire journey, starting with scheduling an appointment, moving into the clinic or operating room as the doctor records allergies or X-rays and then to the back office for billing and follow-ups. Epic electronic medical record systems are used by more than 250 health care organizations nationwide, and roughly 45 percent of the US population have their medical records in an Epic system.

Epic Systems administrators are faced with ensuring the availability of mission-critical business applications daily, for any number of stakeholders within the organization. Doctors, nurses, surgeons, and technicians all need timely access to electronic medical records 24x7x365.

When an IT organization needs to protect Epic records, they need a solution that provides operational recovery with very low recovery time objectives (RTOs) and can address multiple recovery point objectives (RPOs).

While there is no shortage of tools and solutions that offer protection for Epic Systems, many don't focus on the most important aspect: the recovery. When business-critical records and data are unavailable, every second counts. So having a solution that can't meet recovery goals of the business creates a liability.

This white paper focuses on protecting Epic Systems environments with Pure Storage® FlashBlade//S™ with Cohesity DataProtect. Readers will be taken through the performance testing of FlashBlade in the backup and recovery of Epic records.

Introduction

Data continues to grow at an unprecedented rate, and the rise in the value and volume of data has coincided with the increased adoption of all-flash in enterprises worldwide. Yet many backup environments remain unchanged and are unable to meet demanding new recovery requirements, cloud services demand, and the data reusability needs of modern use cases such as DevOps and analytics. Therefore, the need to ensure flash-enabled fast, reliable, and simple backup and recovery of data—at scale—is greater than ever before. Increased occurrences of ransomware attacks, including the recent are a constant reminder that this is becoming an even more critical problem.

Technology leaders are looking for better answers to the challenges associated with the:

- Growing recovery demands of modern applications and data
- Move toward cloud services
- Heightened concerns over ransomware
- Need to address their infrastructure complexities.

Pure Storage and Cohesity forged a strategic partnership and brought to market a first-of-its-kind solution to address all these challenges. Pure Storage FlashBlade//S with Cohesity DataProtect combines all-flash storage with modern data protection solution for rapid recovery at scale. This solution is simple, fast, and reliable backup and recovery for enterprise data that also allows organizations to do more with their backup data, restoring petabytes of data in hours, not days or weeks.

Today's Business Environment

In the era of digital transformation, all businesses are being disrupted. Almost every organization today is producing more data than ever before. The value of data and its importance to businesses today is higher than at any point in the past. Most customers depend on their data as a competitive advantage.

As organizations continue to look for ways to generate value from their data, it has coincided with the growing adoption of all-flash storage. While the declining cost of flash has played a role, the performance, simplicity, and usability of flash have been a significant factor. And while flash had traditionally been used for only mission-critical workloads but with time, we continue all-flash storage being ideal for all use-cases.

IT disasters are far more common than people believe. The worst part about these events is downtime, which is expensive and disruptive for businesses. This situation has made business continuity and disaster recovery top of mind concerns for IT leaders.

While organizations have multiple approaches to data protection, backups are the last line of defense. While typical backup solutions may be optimized to back up the data efficiently, they are missing an essential element: recovering the data quickly and at scale to minimize the impact on business.

With the rise in cybercrimes, more and more bad actors are attempting to infiltrate organizations' IT environments or carry out ransomware attacks. No business is immune from those attacks, and all organizations need a strategy if (or, unfortunately, when) they are impacted. When a business can't recover quickly, the ramifications are severe—from lost revenue, financial liabilities, loss of goodwill, and more.

Organizations today need a strategy that includes the right storage and backup strategy that can provide them a competitive advantage to maximize the value from their data and protect their business for any unforeseen failure and attacks on their infrastructure.

Pure Storage FlashBlade//S with Cohesity DataProtect Overview

All-flash storage dominates storage spend in the enterprise, but the market wasn't always this way. A decade ago, Pure Storage disrupted the market by bringing wide-scale flash storage to the enterprise storage market, a disruption that forever changed the customer expectation from their primary storage.

Current business demands from our customers have called for another change, this time in the backup industry.

While backup vendors like Cohesity have disrupted the backup market, demanding recovery requirements have meant that the purpose-built backup appliance (PBBA) market is impacted to an extent like never before. Backup and storage markets are adjacent in many ways and share many similar customers' pain points.

Pure Storage FlashBlade//S with Cohesity DataProtect combines modern all-flash storage with a modern data protection application to create a solution for rapid recovery at scale. The jointly solution is simple, fast, reliable, and independent scaling of storage and compute for backup and recovery of enterprise data. With storage innovation from Pure Storage and modern backup from Cohesity, the solution delivers:

- **High performance**: Get up to three faster backup and restore throughput than disk-based alternatives.
- **Simplified management**: Gain the ease of management, auto-discovery, and configuration of Pure Storage FlashBlade®; non-disruptive upgrades; and native cloud integration for flash-to-flash-to-cloud backup, recovery, and archive.
- **Scalable and efficient data reuse**: Get disaggregated compute and storage for independent scaling of backup and recovery processes and reuse of backup data on FlashBlade for modern applications.
- **Recovery at scale**: Restore and recover petabytes of data in mere hours.

The Pure Storage FlashBlade//S with Cohesity DataProtect solution has three components:

- **Cohesity DataProtect software**: Cohesity DataProtect is a simple, comprehensive, enterprise-grade backup and recovery software for your traditional and modern data sources.
- **Cohesity-certified compute nodes**: Cohesity DataProtect runs on Cohesity-certified compute (PXG2) nodes for FlashBlade//S with Cohesity DataProtect with no local drives[1].
- **Pure FlashBlade //S**: The storage is provided by Pure Storage FlashBlade and the file systems are mounted to FlashBlade//S Cohesity DataProtect compute nodes via NFS v3.

Business Value

The solution introduces tremendous value to enterprises in three key areas:

Performance

There's value in finishing backups faster. Backups offer protection services to production applications and data. With data and application sets growing at a rapid rate, creating backups in a timely fashion has become more difficult. By being able to back up more often' and write quicker, a business can meet growing demand while meeting the most stringent recovery point objectives (RPOs).

While an "incremental forever" approach to backups can shrink backup windows, during recoveries there are no such optimizations. Restore performance is often dictated by how fast you can read from your backup target and how efficiently your backup target file system is designed. Flash eliminates defragmentation related concerns and inherently brings performance to restores. High-throughput FlashBlade systems offer massive recovery capabilities, allowing recovery of thousands of VMs or petabyte scale environments.

Simple

Backups have always been considered complex solutions. Cohesity has introduced a variety of innovations including non-disruptive upgrades to backups. When combined with Flash Blade's non-disruptive upgrades, the solution becomes a simple,

[1] A compute node does not have any storage except a system disk to run the Cohesity software.

scalable one. Pure Storage FlashBlade//S with Cohesity DataProtect will auto-discover and configure the storage required for backups in a simple wizard-driven approach.

Customers can now store data on FlashBlade for rapid recoveries, ransomware protection, re-use of data and seamlessly integrate with the Cloud for longer term retention or for re-use of data in the cloud.

Scalable

Backups are predominantly storage intensive. By providing a disaggregated, hyperconverged model for backups, customers can scale based on their needs. If additional storage is required, they can just add additional blades to the FlashBlade and grow their environment. Independent scaling allows efficient use of resources.

Customers don't have to pre-plan their backup purchases, which the backup admins typically have no control of. They can grow their environment based on the changing business needs. By leveraging the unified fast file and object (UFFO) storage of FlashBlade, customers have the benefit of consolidating all their modern data needs.

Why Disaggregated Storage and Compute Is Important for Data Protection

One of the unique design aspects of the combination of FlashBlade and Cohesity's solution is its disaggregated storage and compute architecture. By having this disaggregated approach, it allows organizations to scale storage and compute separately. Why is this so important for data protection?

One of data protection's biggest challenges is the ability to scale out. Primary storage continues to grow at an unprecedented rate, and this inevitably impacts your data protection strategy, requiring it to keep pace with the needs of your primary storage. Traditional data protection solutions deliver this by adding a backup server and storage or another backup appliance. This will require not only deployment, but also calibrating backup and DR policies. The other factor to be considered is that backup deduplication only works at the backup server level, so careful policy management is needed to maintain levels of deduplication efficiency.

Contrast this to the disaggregated approach—simply add more storage. No backup policy maintenance is required, deduplication continues to be performed across all backed up data. The disaggregated approach also gives other advantages for example, if you want to increase backup or recovery performance, you can simply add additional compute nodes. In a more traditional data protection model that likely requires a significant redesign, whilst in the disaggregated model, it's simply a case of adding more compute nodes.

The disaggregated approach preserves a single name space, eliminates islands of deduplication, and greatly simplifies the process of scaling a data protection solution—both in respect to adding more storage and more performance.

Use Case: Protecting Epic Caché Databases Files Using a Proxy Server

Pure Storage FlashBlade with Cohesity DataProtect achieves rapid restore at scale, defends all your backup data wherever it resides, and gets the most value from your data. The integrated solution delivers high performance backup and restore, simplified management, and scalable and efficient data reuse.

To illustrate the simplicity and performance, we have outlined the setup of the solution and tested its performance in protecting and recovering the Caché databases files using a proxy server and backing it as physical server filesystem backup.

The setup and creation of policies has been simplified as illustrated below:

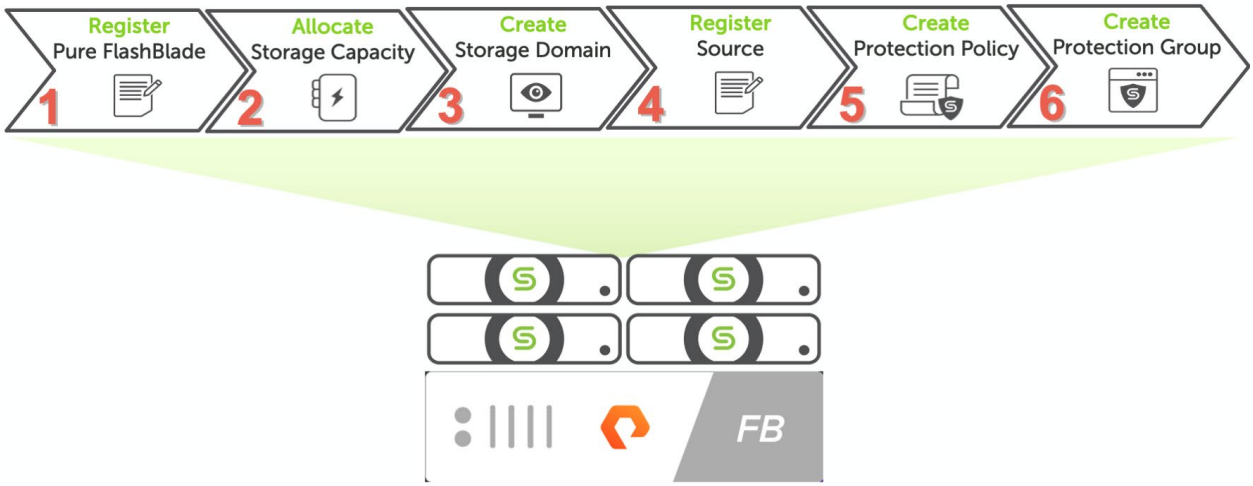

Figure 1: Pure Storage FlashBlade//S and Cohesity DataProtect, steps to protect Oracle databases

Epic Solutions Architecture

Epic is a software company that makes software for midsize and large medical groups, hospitals, and healthcare organizations. Epic software has been widely adopted by major medical institutes and hospitals. It has various modules that target behavioral health, genomics, radiology, rheumatology, clinical trials, oncology, and many other medical related requirements.

Epic is one of the main electronic health records (EHRs) in use at U.S. healthcare organizations today. Epic's production OLTP database server—also known as the operational database, or ODB—has well-defined storage response requirements for both reads and writes to ensure low response times for high performing interactive user workflows to meet the needs of providers and patients in the health system. The underlying InterSystem Caché database on which the Epic application runs has a write cycle limit which is in a few seconds. And to keep things humming, the flush of writes from the database must be completed before that limit before the next write cycle begins, even under the heaviest of workloads. And, to maintain great response times and user experience, the ongoing voluminous random application reads must be reliably fast. There are tests executed for the response time on FlashArray//X™ and that meets and exceeds Epic's requirement, hence FlashArray is the most compelling storage architecture for Epic deployments. It is beyond the scope of this document to cover the wide span of various FlashArray™ response time results and tests with Epic.

Epic deployment consists of a production server which comprises Epic server databases, a reporting server, a read only server, and an Epic disaster recovery server. Each of the Epic deployed components is hosted on Epic Caché database. The storage of Epic Caché databases is hosted on a high-performance primary storage such as FlashArray. The data protection for the Epic Caché database is conducted via exposing the volume of the associated databases from the disaster recovery site to a Linux proxy server or AIX server and later this filesystem is backed up. This document will cover the performance of backup and restore of the Caché datafiles by Pure Storage FlashBlade and Cohesity DataProtect. Figure 2 depicts the Epic deployment and the FlashBlade//S with Cohesity DataProtect data protection of Epic Caché data files.

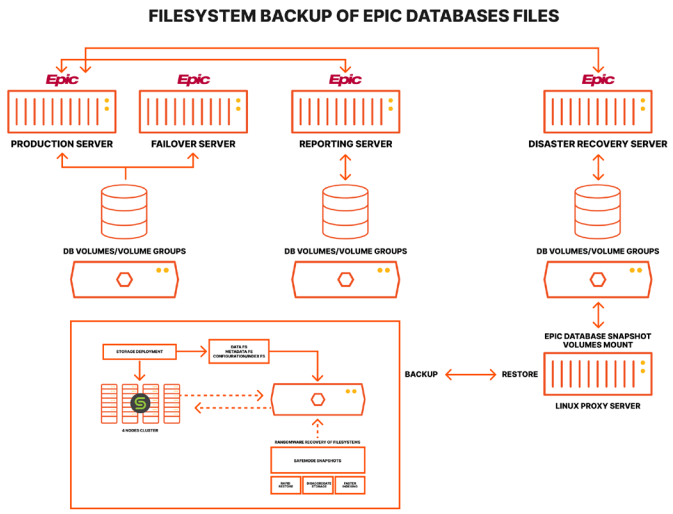

Figure 2: Epic deployment and data protection of Epic data files

The source applications Epic Caché databases can be hosted on any storage, but for fast backup and fast restore it is imperative that it is hosted on an all-flash array like Pure Storage FlashArray. The primary objective in using FlashArray is to ensure Pure Storage can meet the requirements of achieving low consistent latency of Epic workloads activity. Figure 3 shows the advantages of using FlashArray as the primary storage for Epic.

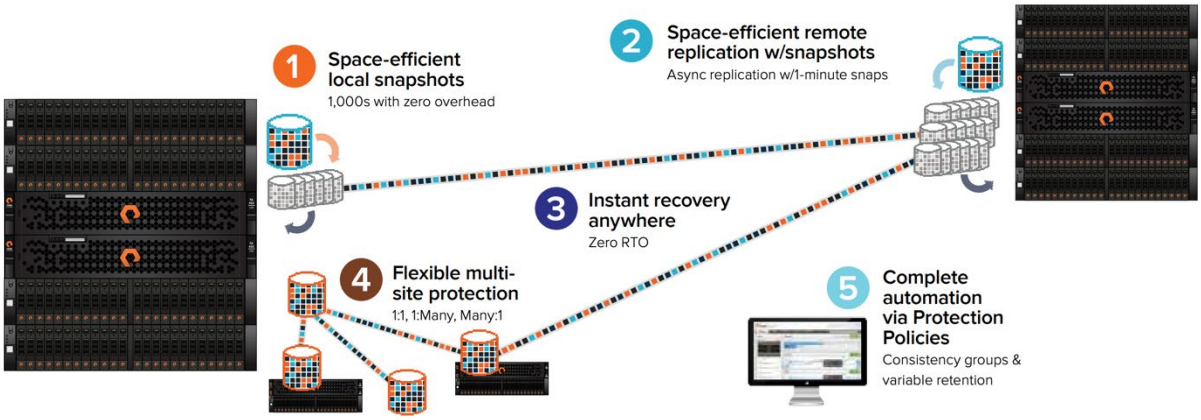

Figure 3: Advantages of using FlashArray

- Infinitely simple, enterprise-class storage array (simple to stand up, simple to use, and simple to maintain)
- Highly available hardware platform, without any single point of failure
- Highly available client connectivity, using native multi-path connectivity to hosts over the SAN
- Non-disruptive software and hardware upgrades with zero loss in performance
- A highly performant solution with predictable low latency (consistent <1ms average latency)
- Unmatched inline and post process data reduction (not one but all forms of reduction, i.e., zero/pattern removal, deduplication, compression inline as well as additional post process compression and deduplication)

- On-array data security using always on native AES-256-bit encryption that requires no external key management applications
- All software and features are included with the Array purchase at no additional cost (snaps, clones, replication, consistency group technology)

Test Configuration

To simulate the real-world Epic deployment eight different volumes of 3TB each and total size of 24TB are created on a FlashArray//X70R2 and a volume group is created with the eight volumes. The following screenshots depict the volumes and volume group for Epic simulated datasets.

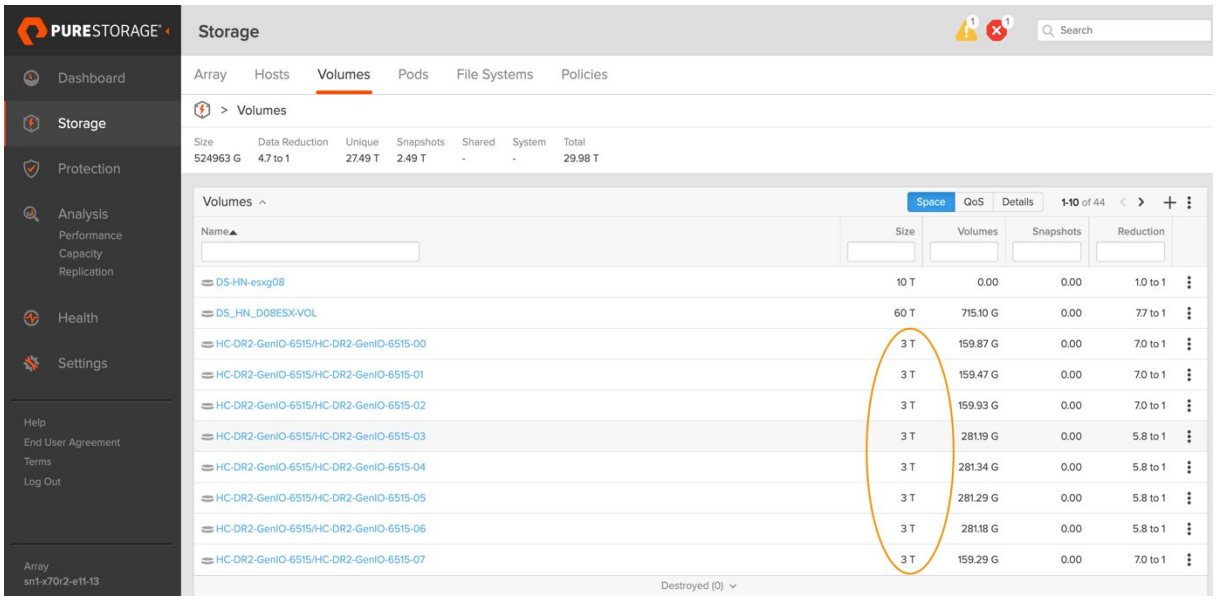

Figure 4: FlashArray Volumes for Epic

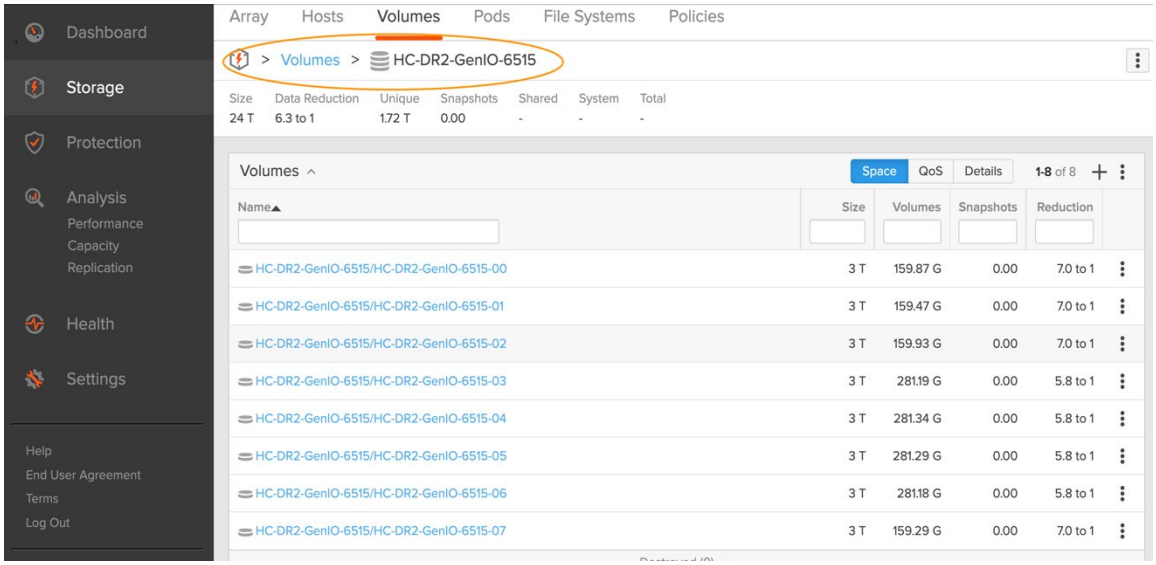

Figure 5: FlashArray Volume Group

Table 1 depicts the FlashArray health and configuration.

COMPONENT	DESCRIPTION
FlashArray	X70r2
Connectivity	8 x 16Gb/s Redundant scsi-fc 1Gb/s Redundant Ethernet (management port)
Raw Capacity	20 SSD x 7.93TB Total Raw Capacity 158TB
Software	Purity//FA 6.1.7

Table 1: FlashArray Configuration

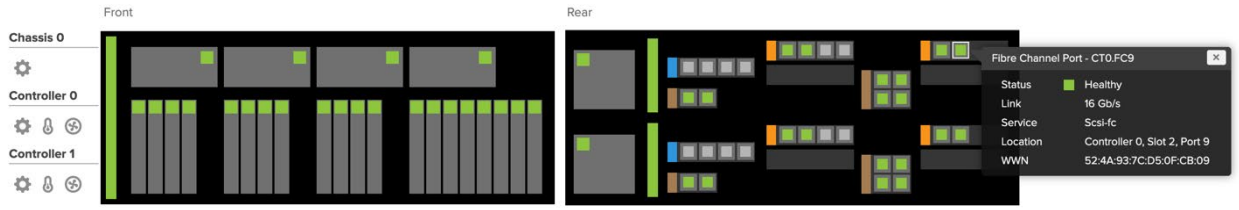

Figure 6: FlashArray health

The following Figure depicts the FlashArray Fibre Channel connection and replication network connectivity.

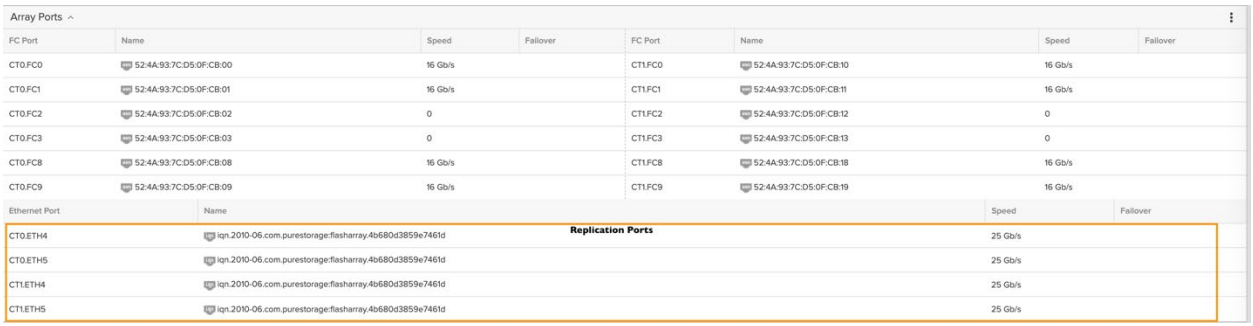

Figure 7: FlashArray network configuration

Epic GenerateIO

GenerateIO a proprietary dataset generation and performance tool which comprises Datagen.pl. The datagen.pl script generates the data files that simulate the Epic Caché database files. This tool is provided by Epic, a data gen tool that would generate data in a very close format compared to the real Caché database, it will create a number of files (like the cache.dat files) stored within a set of directories. To simulate the real-world Epic dataset implementation each file will be created with data reduction (compressibility factor). The files are usually located in the same directory which is either a mount point or a

level below the mount point in this case /epic is the mount point and /epic/prd01 is the directory where all the subdirectories (dir01 to dir80 in this case) are located each file are named with the capacity as suffix.

Most of our customers have a 3:1 data reduction, therefore a compressibility factor of 30 is recommended for the FlashArray , and the following command is used to generate the dataset for backup and recovery test with FlashBlade with Cohesity DataProtect.

```
/usr/bin/perl ./dgen.pl -d /epicdr2/prd0 -c 30 --jobs 10
```

For more information on GenerateIO, please refer to the appendix.

For this test we simulated the Epic load using the Epic GenIO tool. In our solutions lab, we tested the backup recovery workflow of Epic dataset which is hosted on the FlashArray array as a primary storage target. The FlashArray volumes and volume groups are created to host the Epic datasets, the volume group is later shared/mounted on a Linux proxy server which is connected to FlashArray through dual Emulex Corporation LPe35000/LPe36000 FC adapter. Eight volumes sized 3TB each are configured on FlashArray and a single volume group for 8 volumes is created and exposed to Linux proxy server, a volume group is configured on the Linux server and one logical volume is created with XFS filesystem on the logical volume. A datagen.pl script is executed to create an 18TB total size of dataset that represents the Epic Caché database files.

FlashBlade//S with Cohesity DataProtect Architecture

The solution is made up of three main parts:

- **Cohesity DataProtect software:** Cohesity DataProtect is a simple, comprehensive, enterprise-grade backup and recovery software for traditional and modern data sources.
- **Cohesity-certified compute nodes**: Cohesity DataProtect runs on Cohesity-certified compute (PXG2) nodes for FlashBlade//S with Cohesity DataProtect with no local drives.
- Pure Storage FlashBlade//S: FlashBlade//S is the backend storage where the file systems are created and mounted to compute nodes via NFS v3 protocol.

The solution is built on the foundation that compute and storage are disaggregated, which enables enterprises to independently scale compute and capacity as needed. The solution is also designed with high availability in mind, with no single point of failure and no inherent bottlenecks.

This solution can be deployed in an automated manner, which means a user would only need to supply the IP address of the existing or new FlashBlade in the datacenter and token ID. Cohesity auto-deployment software will auto-detect the FlashBlade system and will validate if existing data VIPs are configured on FlashBlade to perform the deployment, as shown in Figure 8.

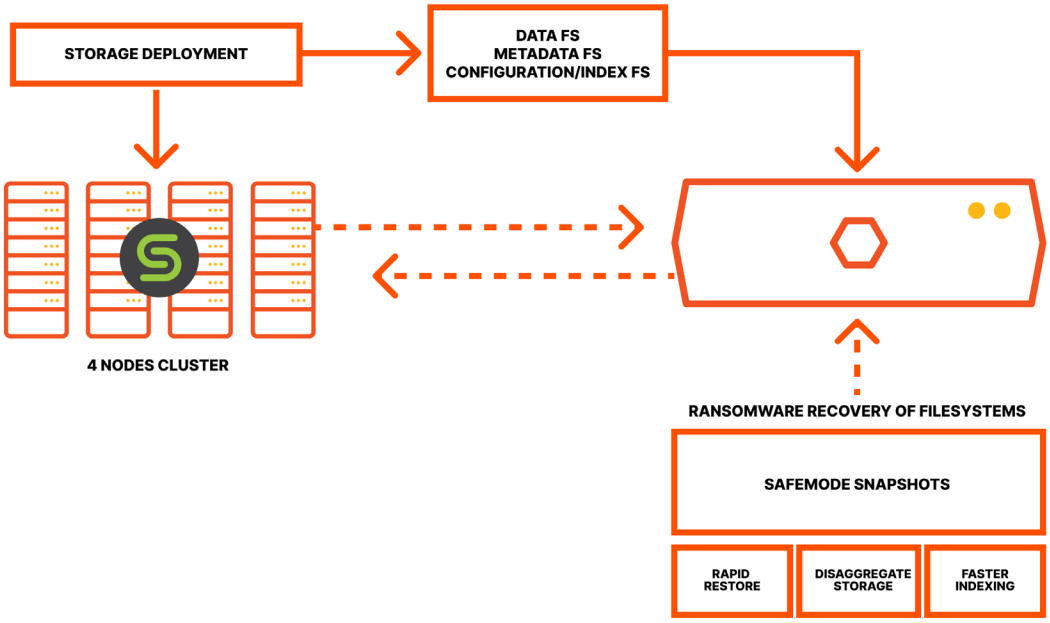

Figure 8: Pure Storage FlashBlade//S with Cohesity DataProtect™ architecture

The deployment creates the data, metadata, configuration, and indexing file systems on the FlashBlade system, as shown in the screenshot below. There are four main filesystem types that get created on the FlashBlade . These four types of filesystems are:

- **Data filesystems**: For backup data store
- **Metadata filesystems**: Uses a distributed key:value store for cluster metadata storage
- **Configuration filesystems**: Stores the cluster configuration and cluster node data
- **Indexing filesystems**: For indexing stores

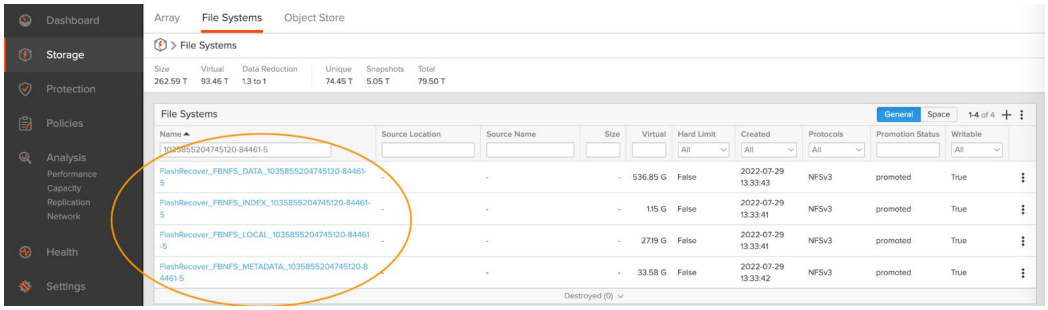

Figure 9: FlashBlade data, metadata, configuration, and indexing file systems

These four types of FlashBlade//S filesystems (data, metadata, configuration, and indexing file systems) are mounted over the data VIPs and balanced across all the available nodes. In the case of a node failure, the logical NFS mount points are redistributed on the remaining surviving compute nodes.

When the failed compute node is back online, the mount points are rebalanced across the FlashBlade//S cluster. Any new node addition to the configuration will create a new set of NFS filesystems.

FlashBlade//S Tuning

The Epic/Filesystem backup and restore optimization in performance will be effective for Cohesity DataProtect software version 6.8.1.u2 and above with some tuning in place. There are some Cohesity Gflags settings required on both the server side and the agent to attain better performance, and this tunable needs to be set by reaching out to Cohesity Support team.

Backup and Restore Tests

Objective

This test case illustrates full backups of Epic simulated data files through the Cohesity UI to demonstrate the backup rate.

Test Procedure

A Linux proxy host is registered as a source to Cohesity. A full backup and restore of Epic data are performed through the Cohesity DataProtect UI by default settings in the data protection policy. The backup and restore speed are captured from Cohesity job run and hence this shows the end-to-end backup timing.

Test Results

Based on our environment set up with four nodes, Cohesity cluster a single host backup performance achieved on an average of 2GB/S, and it took 2 hours and 27 minutes to finish 18.3TB which is almost four times more performance compared to any purpose-based disk backup appliance as shown in the below screenshots.

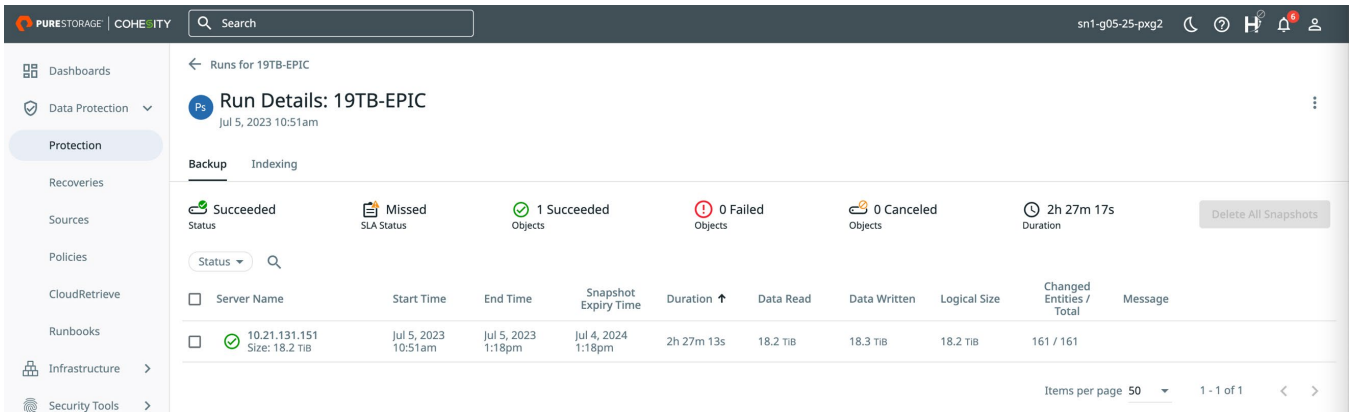

Figure 9 Total backup duration

Figure 10: Write performance from Cohesity dashboard for backup test

A full restore is triggered for the above backup data and captures the performance from the Cohesity UI, the restore performance achieved on an average of 2.4GBps on four nodes cluster as shown in the below screenshot. It took 2 hours and 20 minutes to finish the restore of the 18.3 TB Epic data set which is ~2.2GB/s end to end restore performance rate.

> **Note:** Since, Epic backup and restore is performed through a single proxy server which can leads to some limitations with the backup and restore performance, which may not show a linear scale in performance when adding more nodes at present release, and it may change in near future with more software optimization from Cohesity.

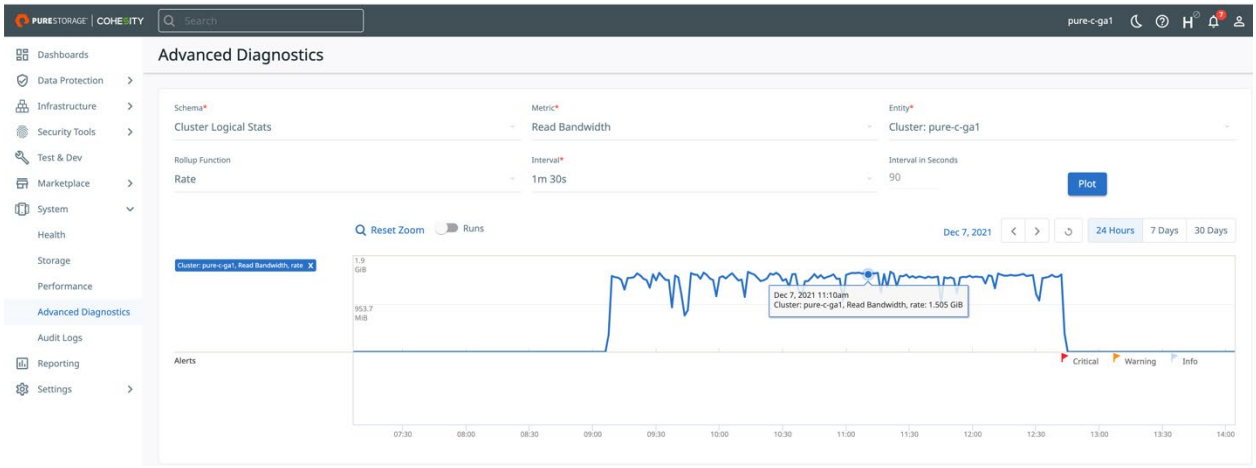

Figure 11: Restore performance on four nodes FlashBlade//S with Cohesity DataProtect cluster

Conclusion

Pure Storage FlashBlade//S with Cohesity DataProtect delivers a modern data protection experience, combining the best in a flash-based unified fast file and object storage with an industry-leading data protection application from Cohesity, giving mid-sized and enterprise customers alike a highly performant solution designed for fast recoveries at scale.

FlashBlade as an NFS storage target for deduplicated data offers an option for organizations seeking a storage-efficient solution that delivers “rapid restore” at scale of Virtual Machines in their environment.

By leveraging the key strengths of the solution’s data platform to consolidate secondary storage workflows, this architecture offers the industry’s top compute platform and disaggregated unified flash storage with integrated backup and recovery software.

FlashBlade//S with Cohesity DataProtect can deliver:

- Up to three times faster restore throughput than disk-based alternatives
- Improved storage efficiency when utilizing FlashBlade with Cohesity’s inline deduplication and FlashBlade compression
- Disaggregated compute and storage for independent scaling of backup and recovery processes
- Reuse of backup data on FlashBlade for modern apps
- Most simplified form of Epic data protection

Additional Resources

- Learn about how to how to combat ransomware in healthcare.
- Explore FlashBlade//S.
- Read about the Cohesity and Pure Storage partnership.

Appendix

FlashArray Storage Configuration

Configuring the Test Storage for Epic GENIO Linux

To simulate the real-world Epic deployment, eight different volumes of 3TB each and total size of 24TB are created on FlashArray//X70R2 and a volume group is created with the eight volumes. The following screenshots depict the volumes and volume group for Epic simulated datasets.

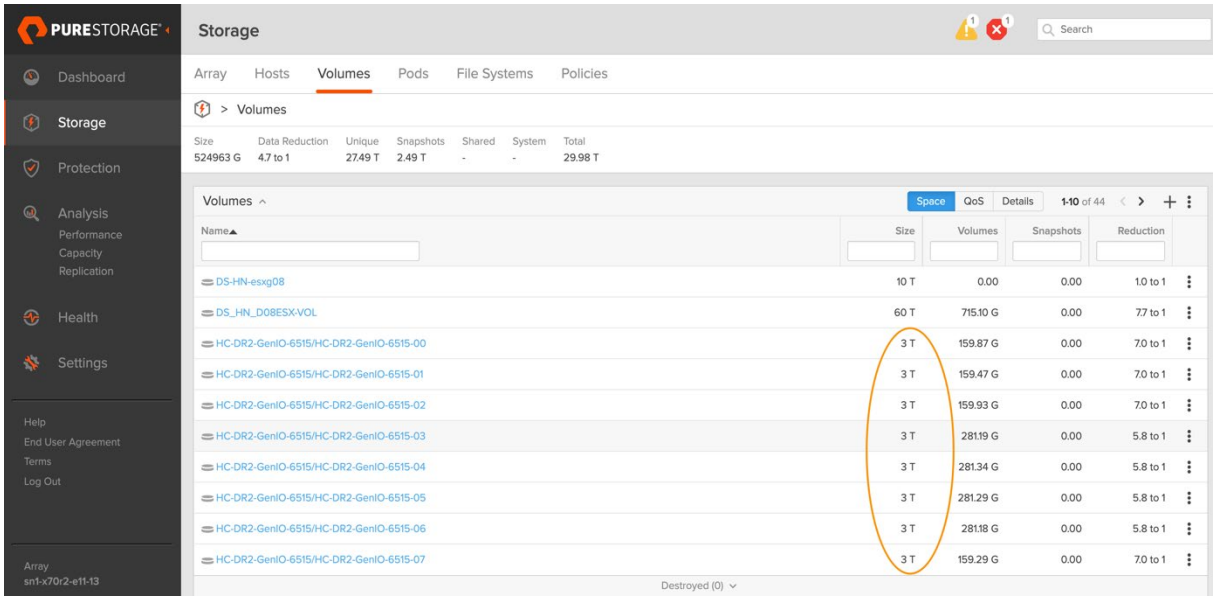

Figure 12: FlashArray test volumes creation

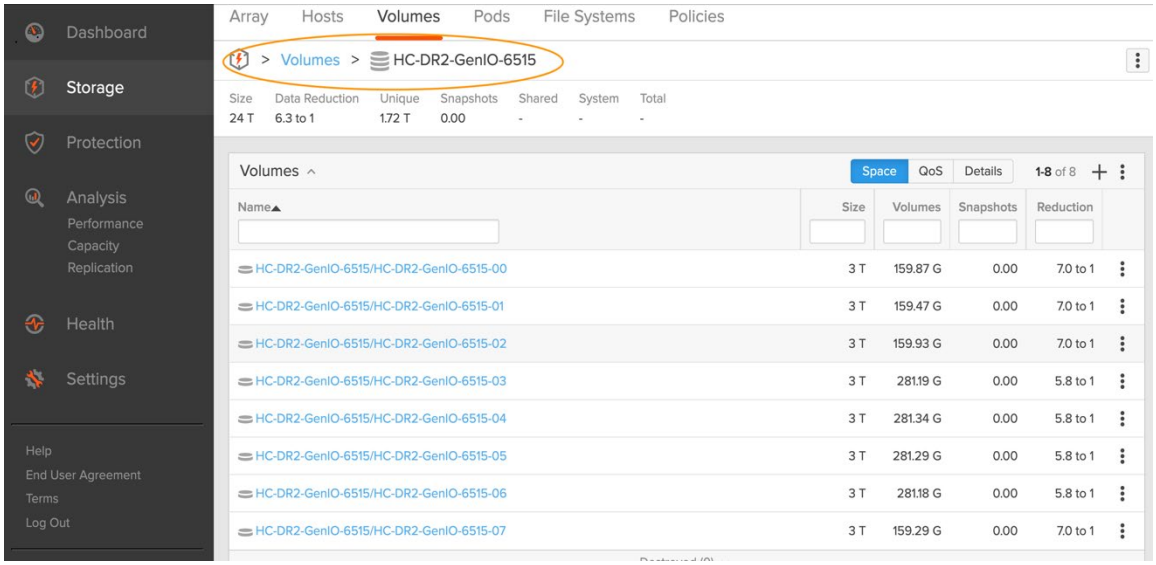

Figure 13: Volume group creation

Configure Physical LUNs

For the GenerateIO test we recommend using two to eight LUNs for “PRD” (four usually works fine) and one LUN for the WIJ (this doesn’t need to be copied or snap for backup this contains only a transient file).

Configure Physical LUNs

With a Pure Storage FlashArray, we recommend presenting four to eight LUNs, which will be managed within a single volume group. We recommend checking the "Queue_depth" for each adapter and have it at a high number: 64 to 121.

Configuration of Virtual Machine as a Proxy Server for Epic Backup

This section is based on Epic "Storage Configuration Quick Reference" guide.

The Linux machine used for Epic Caché usually run as a VMware Virtual Machine, and each volume presented as a RDM (raw device mapping) volume with a fixed IO path policy, LUN queue depth configuration with VMware is possible, however the defaults for most adapters are either 32 or 64 per path so we do not expect any adjustment to be necessary as long as there are a sufficient number of LUNs.

Create Volume Groups

Identify your devices (typically by size):

```
# lsblk -do KNAME,TYPE,SIZE,MODEL
KNAME TYPE SIZE MODEL
sda disk 16G    Virtual disk
sdg disk 800G
sdd disk 800G
```

If the size of the devices is not enough to identify them, the SCSI controller location can be viewed with lsscsi. If using physical RDMs or physical servers on RHEL7, the wwn can be displayed with lsscsi -sw or lsblk -d -o NAME,WWN. The default volume group physical extent size is 4MB. We typically use 16MB, but 4MB will also work. Keep in mind that the commands below assume 16MB. Create the volume groups (warnings about missing PV descriptors can be ignored):

```
# vgcreate -s 16M [Volume Group Name] [list of all disks for Epic]
```

Using the example above for prdvg:

```
# vgcreate -s 16M prdvg /dev/sd[h-k]
Physical volume "/dev/sdh" successfully created
Physical volume "/dev/sdi" successfully created
Physical volume "/dev/sdj" successfully created
Physical volume "/dev/sdk" successfully created
Volume group "prdvg" successfully created
```

Create Logical Volumes

We need to use LV striping to spread IO evenly across the devices to avoid creating contention of the OS level queues for the devices. We use 16MB as the VG physical extent size.

If you used the default (which is 4MB), make sure to use that instead. Make logical volumes according to what is specified in the file system layout guide:

lvcreate -n [Logical Volume Name] -L [Size of Logical Volume] -I [Number of LUNs in the volume group] -I 16M [Volume Group Name]

Using the prdvg example again:

```
# lvcreate -n prd01 -L 7.5T -i 4 -I 16M prdvgvi
Logical volume“”prd0”” created.
#
```

On top of the “prd01” logical Volume Add a “wij” logical volume of about 10% of the capacity of “prd01.”

Create File Systems

XFS is the preferred file system. Use the -K option when creating the file system on flash-based storage (unless LUNS were previously used.) This avoids the mkfs command forcing an erase of the underlying storage for the luns which would cause the mkfs to take much longer.

Note that XFS automatically aligns itself with the underlying LV stride size (4k * 16k = 64MB in our example.)

mkfs.xfs -K [Logical Volume Device]

Example:

```
# mkfs.xfs -K /dev/prdvg/prd01
log stripe unit (16777216 bytes) is too large (maximum is 256KiB)
log stripe unit adjusted to 32KiB
meta-data=/dev/prdvg/prd01 isize=256 agcount=9, agsize=110592 blks
         = sectsz=512 attr=2, projid32bit=1
         = crc=0 finobt=0
data     = bsize=4096 blocks=917504, imaxpct=25
         = sunit=4096 swidth=16384 blks
naming   =version 2 bsize=4096 ascii-ci=0 ftype=0
log      =internal log bsize=4096 blocks=2560, version=2
         = sectsz=512 sunit=8 blks, lazy-count=1
realtime =none extsz=4096 blocks=0, rtextents=0
```

Create the “wij” filesystem, as well.

Mount File Systems

Update /etc/fstab with mount entries.

If using XFS, fstab fields 5 and 6 are ignored:

- /dev/mapper/prdvg-prd01 /epic/prd01 xfs defaults 0 0
- /dev/mapper/prdvg-wij /epic/wij xfs defaults 0 0

Create the mount points.

```
# mkdir /epic/prd01 /epic/wij
# chmod 755 /epic/prd01 /epic/wij
# chown root.root /epic/prd01 /epic/wij
```

If all entries were added to fstab, you can simply run 'mount -a'. Otherwise, mount each file system manually:

```
# mount /dev/mapper/prdvg-prd01 /epic/prd01
# mount /dev/mapper/prdvg-wij /epic/wij
```

Epic IO Simulator (GenerateIO)

In the Package provided by Epic, there are two main software components: GenerateIO which can be run using the command ramprun.pl and a “Data Generator” which is called dgen.pl.

Generate the data for the GenerateIO test : DataGen - dgen.pl

After installing the GenerateIO package provided by Epic, you can start to create the “simulated” Caché database.

The most reliable test would be to utilize existing data from a real Caché database, but if it is impractical or restoring a backup is not possible then dgen.pl could be used.

dgen.pl would generate data in a very close format compared to the real Caché database; it will create a number of files (like the cache.dat files) stored within a set of directories. Each file will be created with data reduction if needed (default is 2:1). The files are usually located in the same directory, which is either a mount point or a level below the mount point, in this case /epic is the mount point and /epic/prd01 is the Directory where all the subdirectories (dir01 to dir80 in this case) are located each file are named with the capacity as suffix. The default for dgen.pl is to create 80 directories and use 85 % of the filesystem capacity.

Here the information in the README file regarding dgen:

“DataGen is a program that is used to generate the input data files to GenerateIO. Writing out random blocks generates the files. To speed this process, a number of blocks are generated, and then written out in random sequence. By default, each block is just a single character. The creation of all the files is done in parallel.”

There is a perl script provided “dgen.pl” which will automatize the entire data generation process, in the latest version it creates the files with a 45 “--compressibility” factor which is a 2.5:1 data compression.

Most of our customer have a 3:1 data reduction, therefore a compressibility factor of 30 is recommended for the FlashArray array.

On AIX only, before running dgen.pl we recommend unmounting the /epic/prd01 directory and remount it without the “cio” option. Don’t forget to remount the filesystem with “cio” it will impact significantly the results.

Usually, the default command to run dgen.pl is:

#./dgen.pl -d /epic/prd0 -c 30 --jobs 10 –yestoall

The “/epic/prd0” is not a typo dgen will assume there is multiple directory prd01,prd02..prd08 if you create only one it will still need to use “/epic/prd0”:

```
# ls -laR prd01/dir0*
prd01/dir01:
total 723882208
drwxr-xr-x.  2 root root           37 Jan 17 17:29 .
drwxr-xr-x. 82 root root         4096 Jan 22 11:12 ..
-rw-r--r--.  1 root root 741255151616 Jan 22 16:03 random_testdata_706916M

prd01/dir02:
total 397900808
drwxr-xr-x.  2 root root           37 Jan 17 17:29 .
drwxr-xr-x. 82 root root         4096 Jan 22 11:12 ..
-rw-r--r--.  1 root root 407450419200 Jan 22 16:03 random_testdata_388575M
```

The second part is the GenerateIO program, which can be started using the RampRun.pl script. It is a Synthetic test, which simulates the behavior of a Caché database as utilized by Epic.

As such, it keeps the Write IO in memory until Written in phase two in the Write Immediate Journal and the “database”. This happens every 80 seconds and the type of data generated depends on the test. The data simulator has several options which include using “compressible” data.

The latest version of GenerateIO could use a journal mechanism and therefore can validate with High Availability Cluster solutions like Active Cluster.

Epic software simulation is a synthetic test; therefore, it can run on any type of data as long as they are presented in a file system format. Usually, Caché presents the Data file located in several Filesystems mounted under 8 main Filesystems (/Epic/prd01 … /Epic/prd08) and within that File System creates several files called “cache.dat”.

The most efficient way to test GenerateIO is to use a copy of the customer existing database (it could be an old copy from a month-old backup or a copy of the reporting Database). GenerateIO doesn't manage the data; it just reads at random places on the filesystems and writes random data on the filesystem as well.

Using a test copy of the customer database would give the appropriate "data reduction" factor for the test read and help to understand the benefit of using Pure Storage deduplication and compression. This will also supply a much more accurate reading of the performance that can be achieved with GenerateIO.

About the Author

Mandeep Arora is the Pure Storage data protection solutions architect responsible for defining data protection solutions partnered with various backup applications. He is a lead solutions architect for FlashBlade data protection solutions and responsible for defining other data protection solutions and reference architecture for primary workloads such as Oracle, SQL, VMware based on the company's products, and performance benchmarks.

Mandeep has over 14 years of his career spent with the data protection industry, he has the flavor working with various data protection products meant for small medium business and large enterprise business. He started the data protection career with IBM Tivoli storage manager in core software development and test team, a career followed by Isilon Systems where he was responsible for delivering the NAS backup solution to enterprise class customers. He was also a part of Veritas storage solutions team and was responsible for technical relationship and adviser with partners in data protection for VMware.

Pure Storage, Inc.
650 Castro Street, #400
Mountain View, CA 94041

PS2495-01-en 09/23