Technical Review

Pure Storage ActiveDR: Near-zero RPO with Continuous Asynchronous Replication

Date: June 2020 **Author:** Tony Palmer, Senior Validation Analyst
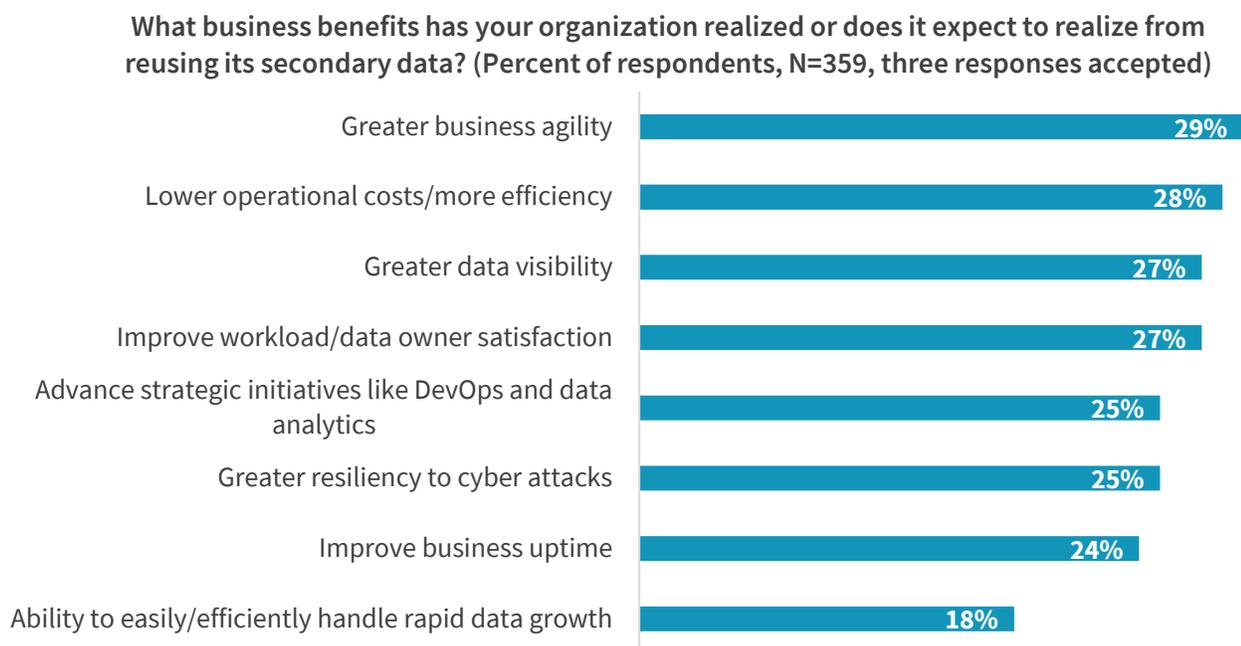
Abstract

In this Technical Review, ESG examines Pure Storage ActiveDR with a goal of validating the ability of the software to provide near-zero RPO protection with continuous replication and fast recovery. ESG also tested ActiveDR's zero-impact failover and ease of deployment and management.

The Challenges

High data availability is essential in today's business environment. Data drives most businesses today, and they must have data continuously available not only to remain productive and efficient to generate revenue, but also to ensure compliance with corporate governance, industry requirements, and government regulations. For that reason, organizations depend on data protection and disaster recovery (DR) strategies with fast recovery times and recent recovery points.

However, savvy organizations also realize that the ability to make additional use of copy data can provide significant benefits—rather than leaving secondary data at an offsite location, sitting idle and costing money while the organization waits for a failure to occur. Organizations are now seeking the opportunity to use these secondary copies to their advantage for tasks such as application development, failover testing, cybersecurity and compliance testing and reporting, data mining, and more. According to ESG research, organizations that are reusing their secondary data are gaining numerous benefits, including better business agility, lower costs, better data visibility, ability to advance DevOps and analytics, and cyber resiliency.[1]

Figure 1. Benefits of Reusing Secondary Data

What business benefits has your organization realized or does it expect to realize from reusing its secondary data? (Percent of respondents, N=359, three responses accepted)

Benefit	Percent
Greater business agility	29%
Lower operational costs/more efficiency	28%
Greater data visibility	27%
Improve workload/data owner satisfaction	27%
Advance strategic initiatives like DevOps and data analytics	25%
Greater resiliency to cyber attacks	25%
Improve business uptime	24%
Ability to easily/efficiently handle rapid data growth	18%

Source: Enterprise Strategy Group

---

[1] Source: ESG Master Survey Results, *The Evolution from Data Backup to Data Intelligence*, January 2020.

The Solution: Pure Storage ActiveDR

The ActiveDR feature is available with version 6 of the Purity Operating Environment with no additional licenses or fees, and is included as part of Pure's non-disruptive upgrades and evergreen subscription. It provides continuous, asynchronous, bidirectional replication for FlashArrays (FA) to a secondary site, providing near-zero recovery point objectives (RPOs) while enabling use of the replica for multiple purposes. This is a significant improvement over Pure's asynchronous periodic replication, which offered 10-minute RPOs and replication every five minutes. With Active DR, organizations with high bandwidth networks can achieve near-zero RPO. Like ActiveCluster, ActiveDR makes use of storage containers called pods to reduce storage management overhead.

Once a primary site pod is linked to a secondary site pod, writes are streamed to the target and compressed to reduce bandwidth requirements. This results in minimal data loss in case of failover. The secondary site stores a replica for protection and recovery, while also enabling that replica to be used for QA, test/dev, etc., without interrupting the DR replication process. The target pod can be used to test failover—also without stopping DR replication—a task that is traditionally difficult and carries enough risk that many organizations choose not to perform it, leading to insecurity about their ability to survive a failure.

CLI, GUI, or REST operations performed in the source FA pod, such as volume creation, snapshots, resizing, and cloning, are automatically replicated to a secondary pod, including the snapshot history, protection groups, and scheduling. The volumes on the target pod are read-only but can be easily activated to make them writable. Replication performance can be monitored from either pod, and the replica link direction can be easily reversed for failback. DR hosts can be pre-connected to volumes on the target pod for fast recovery time, and there are no journal devices to manage. Organizations gain fast recovery, failover, and failback, and can test failover without compromising the recovery point.

Figure 2. Pure Storage ActiveDR—Active Data and DR Copies with Continuous Replication
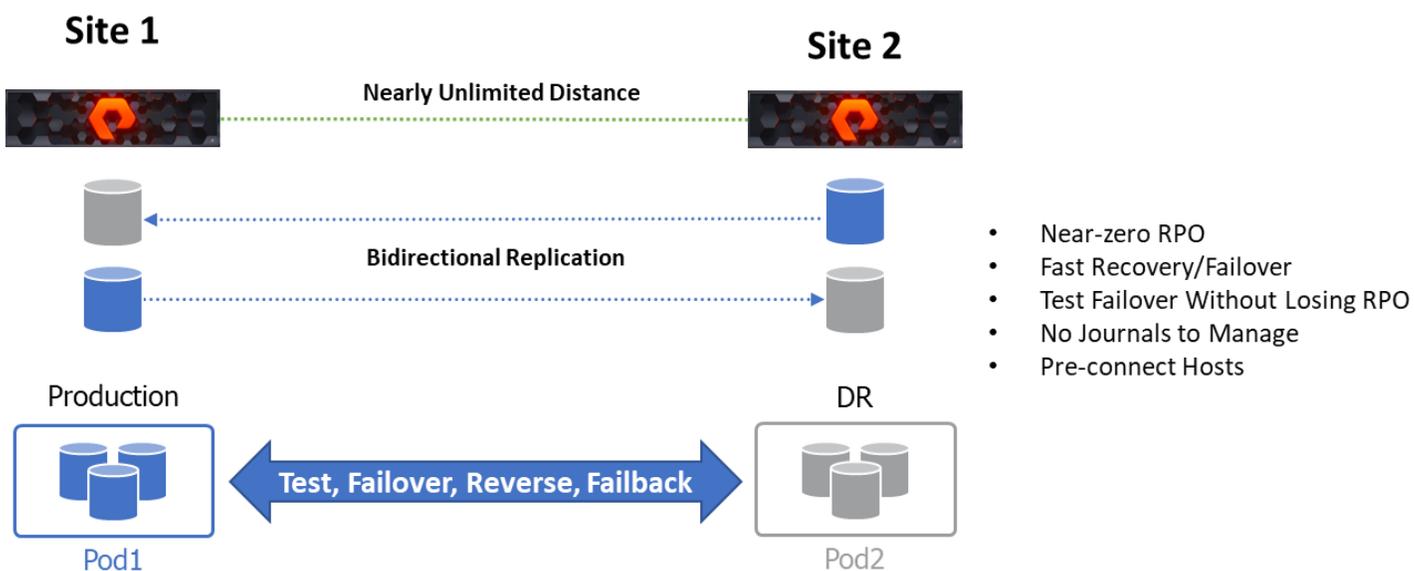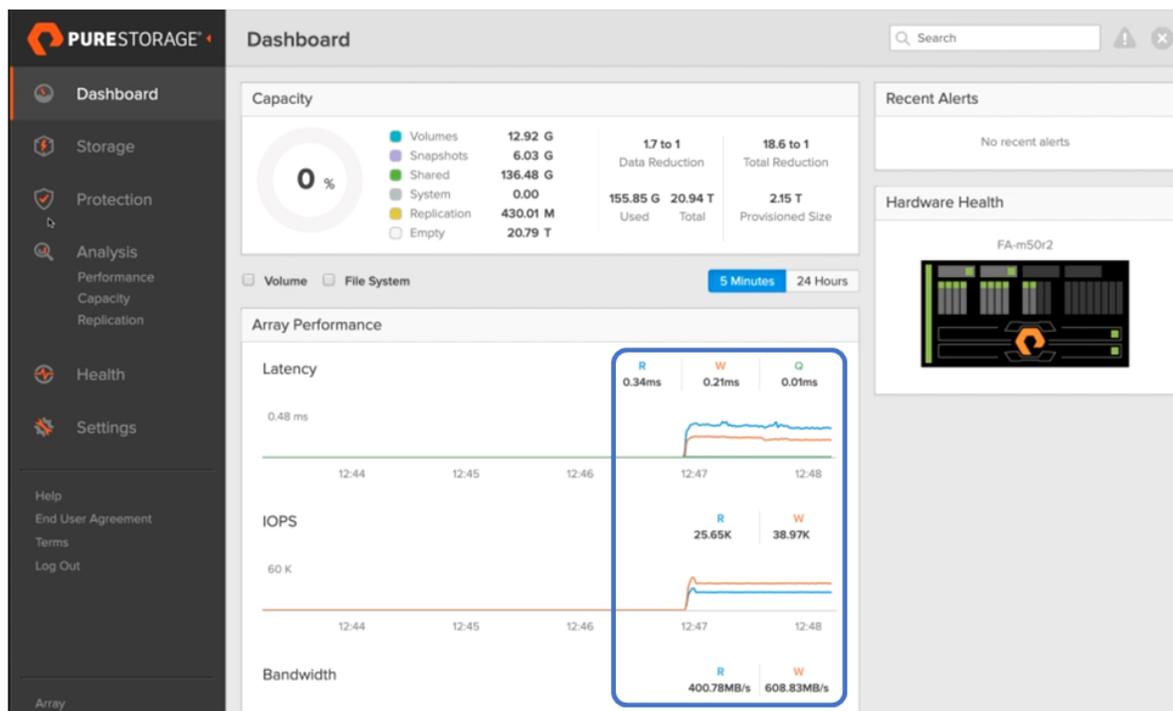

Source: Enterprise Strategy Group

ESG Tested

Testing was conducted with two Pure FlashArray//M50 systems, which we designated as production and DR, in a campus environment. The production system was configured with a pod named ProdPod1 containing two volumes. Vdbench was used to generate I/O against one volume, while a script was configured to write date and time stamps to a file—importantdata.txt—in the other volume. The workload profile was 100% random 16KB I/O, 40% reads, 60 % writes. When

we began the test, the production FlashArray was servicing 64,620 IOPS at sub-millisecond response times. The workload was driving more than 1GB/sec of throughput, as seen in Figure 3.

Figure 3. Active DR—Production Workloads Running at Primary Site

Source: Enterprise Strategy Group

We created a Replica Link between the arrays. A Replica Link is a managed object that defines a replication relationship between two pods. While a Replica Link has a source, a target, and a direction, all we needed to define were the source pod, the remote array, and the remote pod name, which we selected from pull-down lists.

Figure 4. Creating a Replica Link
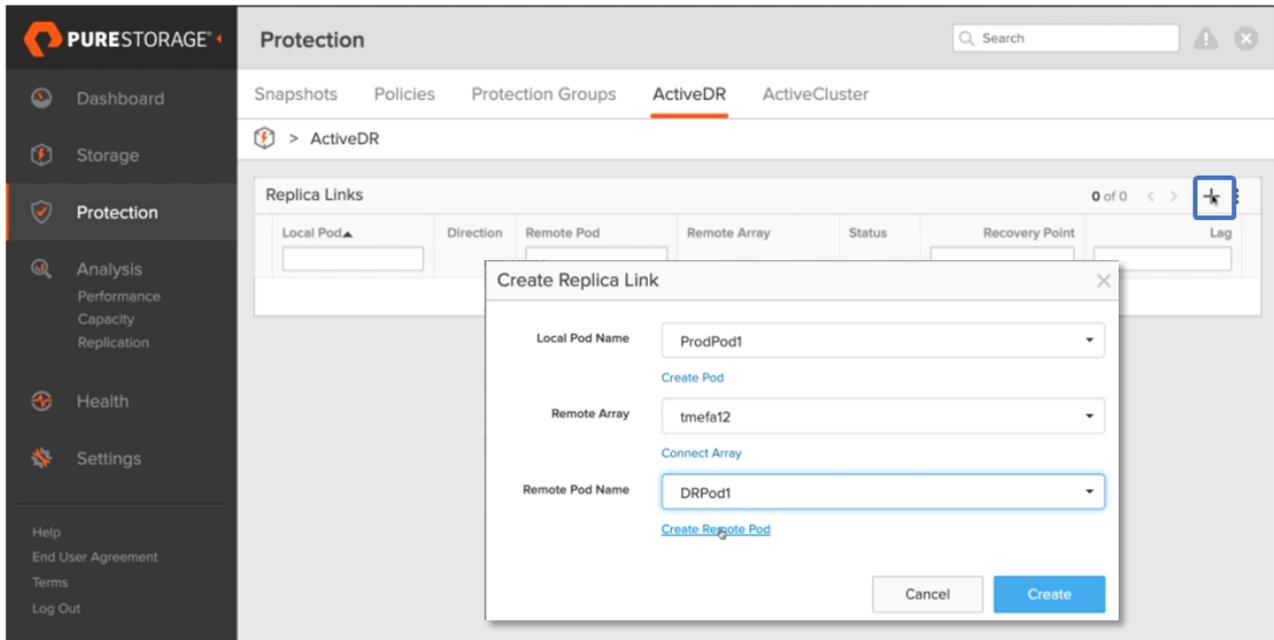
Once the link was created, the arrays automatically began baselining. This process is like the baselining process used in ActiveCluster and Async. This process populates the target with the initial copy of data and uses Pure's existing asynchronous replication engine that can preserve both compression and deduplication. Because ActiveDR does not present active volumes from both sites, no mediation system is required. Also, ActiveDR utilizes a mechanism to track writes with no journaling devices required.

Figure 5. Baselining
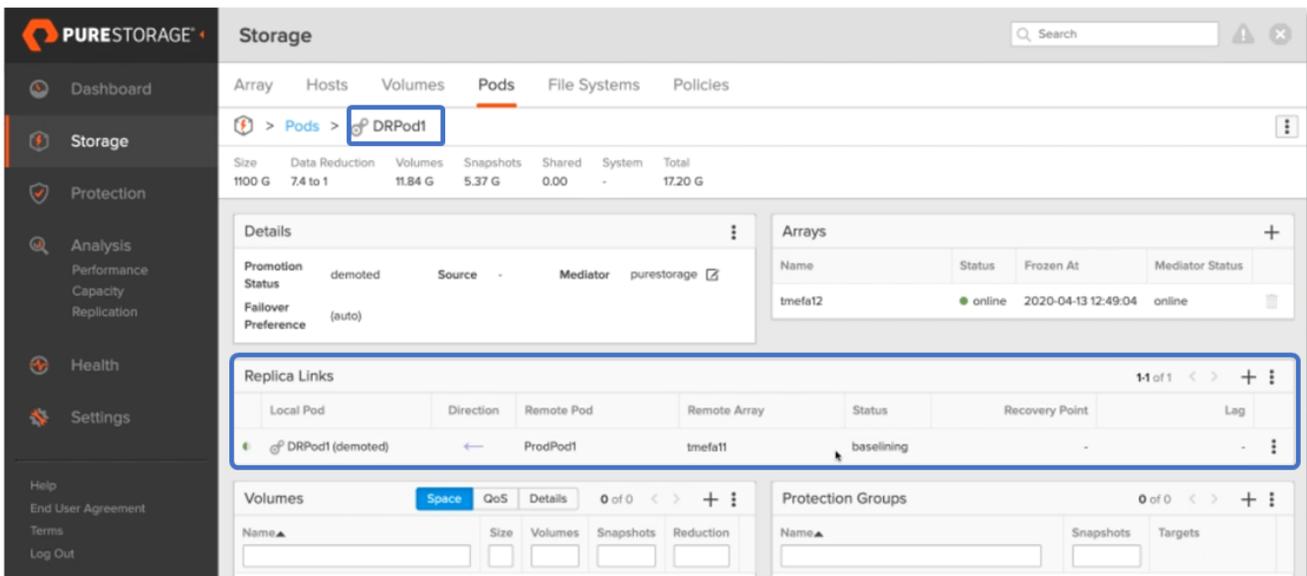
Once baselining was complete, the systems automatically switched over to continuous replication. When ActiveDR is replicating, writes are streamed continuously to the target rather than the periodic batch and forward method used by

Async to send changed data at a configured interval of five minutes or greater. This is one of the features that enables near-zero RPO.

Figure 6. Connecting Volumes to the DR Host
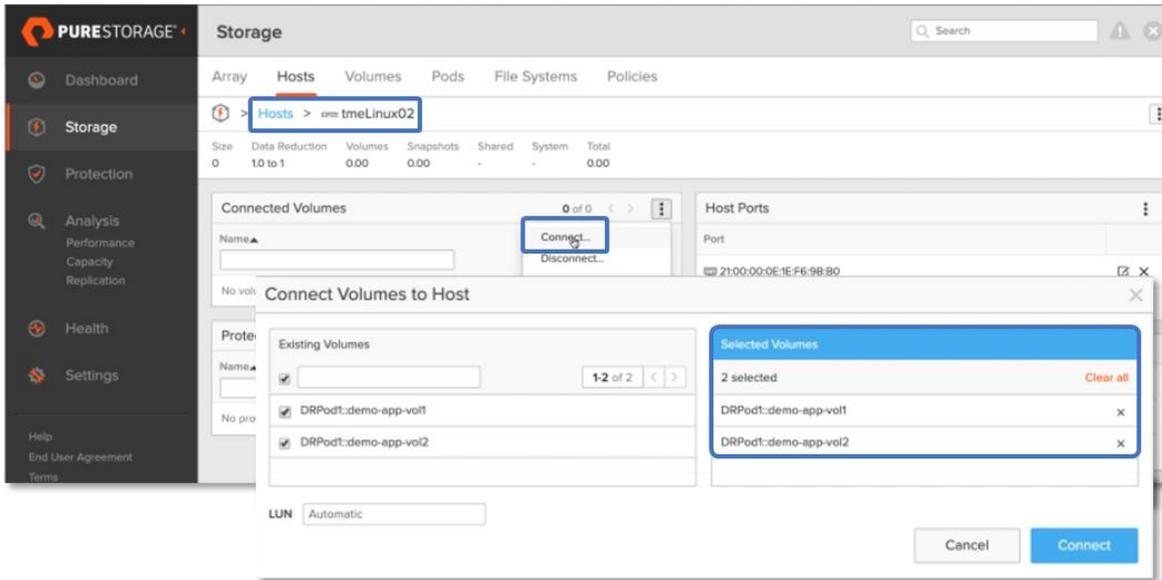
Source: Enterprise Strategy Group

Hosts at the target site can be pre-connected to the volumes in the DR pod to simplify failover. As seen in Figure 6, we connected the volumes to the failover host. At this point, the systems are completely ready for DR failover.

Figure 7. Executing DR—Promoting the Pod
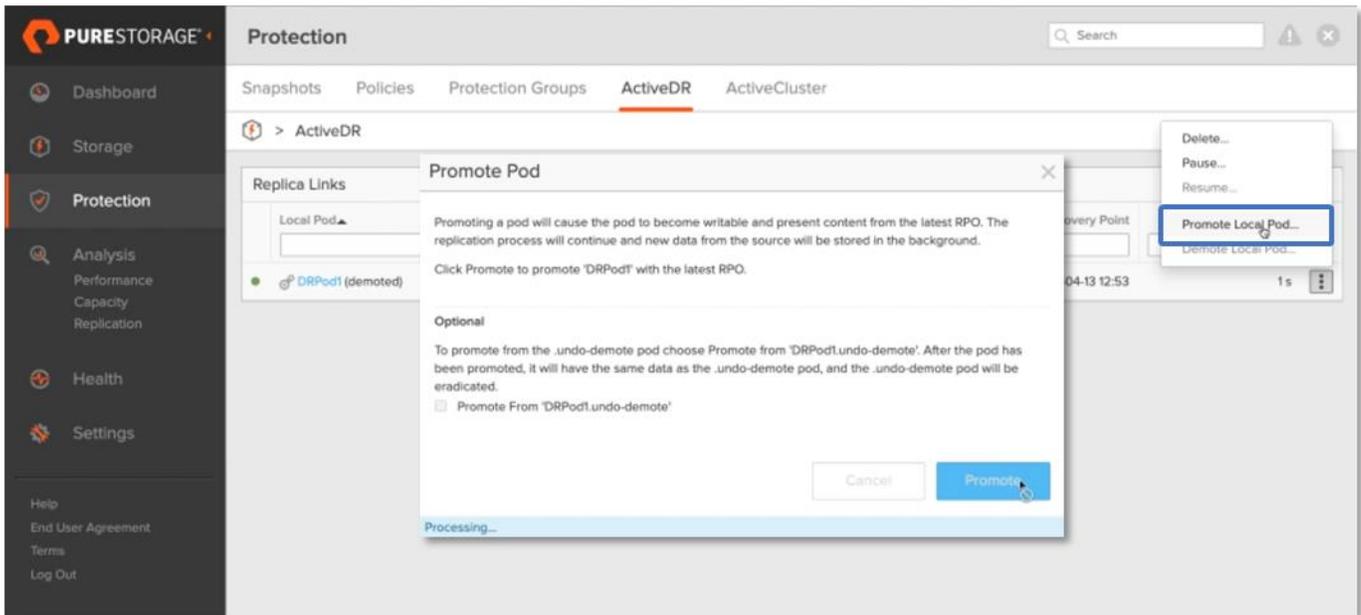
Source: Enterprise Strategy Group

To simulate an outage at the production site, we stopped the script writing date stamps, then clicked on ActiveDR at the target array, and selected *Promote Local Pod* from the pull-down menu. It's important to note that this is exactly the process one would use to test DR, since this brings the target volume online but continues replicating from the source system.

Figure 8. Running Production at the DR Site

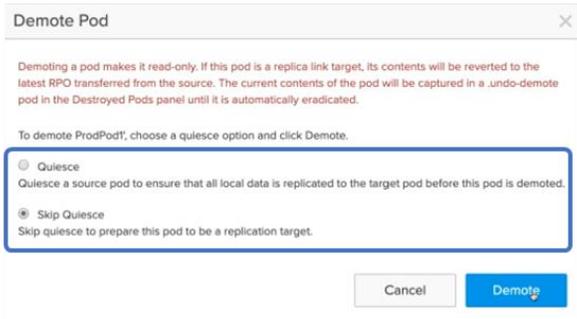Production Site   DR Site

While the target site was promoted, we verified that the volumes had been replicated to the DR site, then opened the importantdata.txt file and added the text: "Now Running Production At DR Site." It's important to note here that Pure is actively working on integrations, including an adapter for VMware Site Recovery Manager (SRM) that automates the entire failover process in VMware environments. ActiveDR is completely scriptable and can be easily integrated into existing manual or automated failover processes.

At this point, we were ready to failback to the original production site. First, we demoted the pod on the original production system and selected *Skip Quiesce* to enable replication from the DR site. We then opened the importantdata.txt file and verified that the data we inserted at the DR site had been replicated back to the production site.

Figure 9. Running at the DR Site

Production Site

Finally, we demoted the pod on the DR site and restarted our script. The entire process—creation of the replica link, failover, testing, and failback—took less than four minutes.

It's important to note here that the process to test failover is extremely similar, and just as easy. While Pure does have a command to pause replication, there is no need to use it during DR testing. ActiveDR is designed to keep replication running to allow organizations to test DR without impacting RPO.

ℹ️ Why This Matters

Data growth and the rapid proliferation of virtualized applications are increasing the cost and complexity of storing, securing, and protecting business-critical information assets. A storage solution with near-zero RPO and simple user tools that make it easy to deploy and centrally manage a complex, multi-site storage deployment can reduce the time and cost required to deliver business productivity and disaster recovery.

ESG was impressed with the ease of configuring ActiveDR. We validated that volumes, protection groups, and snapshots can be continuously replicated between local and remote arrays with a simple setup and a one-click failover process. We also validated that DR testing could be performed completely non-disruptively, with applications still running at the production site and replicating to the DR site while testing is occurring.

ActiveDR can be used for disaster recovery, test/dev, and data migration. It is included with Pure's Evergreen storage subscription, accessible with a simple operating system update. The entire process of connecting the arrays, testing failover, and failback took minutes with just a few clicks—no downtime, no professional services, and no interruption to productivity. Administrators could easily set it up with no specialized storage skills.

The Bigger Truth

Maintaining productivity is a key goal for any organization to remain competitive, and for that reason, organizations implement data protection and disaster recovery strategies. Data must be available for business use and compliance with industry and corporate requirements. However, data copies at a secondary site are a costly resource that sits idle, waiting to be needed. Storing this data consumes storage, power, cooling, and management resources; wouldn't it be helpful to use that data for business benefit?

Pure Storage has a proven record of providing the simple, cost-effective, enterprise data protection and data management features organizations need. Purity ActiveCluster already provides synchronous, bidirectional replication with stretched pods that enable active/active configurations. Now, with the addition of ActiveDR, Purity provides near-zero RPO with continuous *asynchronous* replication, enabling disaster recovery, failover, and failback, while making the replica available to be used for QA, test/dev, cybersecurity and compliance testing and reporting, data mining, and other activities.

The results that are presented in this technical review are based on testing in a controlled environment. Due to the many variables in each production data center, it is important to perform planning and testing in your own environment to validate the viability and efficacy of any solution. ESG believes that Pure's development of integrations extending Pure ActiveDR to work with third-party applications and systems will be a welcome enhancement.

ESG validated that Purity ActiveDR simplifies disaster recovery, making continuous replication, failover, and failback transparent and effortless. We validated the ease of setting up intelligent, efficient DR replication, eliminating the traditional complexity of manual setup. ActiveDR is configured without downtime, expensive professional services, or weeks of administrative effort.

Pure Storage strives for simplicity, and in ESG's view, the company has delivered yet again. ActiveDR makes minimizing RPO simple and cost-effective for any application and organization, across any distance. Given the competitive advantage that maximizing uptime delivers, ActiveDR is well worth a look.