

Pure Storage, Catalogic Software and Copy Data Appliances

Introduction

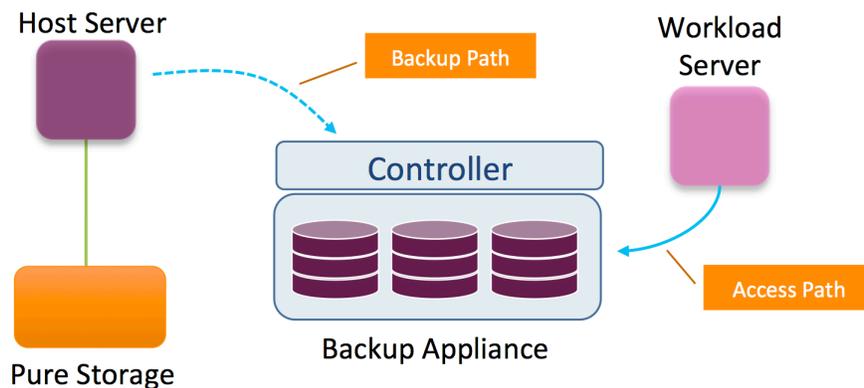
Pure Storage partners with many organizations that fall under the heading of “data protection” or “copy data management.” These include Catalogic Software as well as vendors such as Rubrik, Cohesity, Actifio, Commvault, Veeam, etc. Because they fall into the same categories, there is a misconception that Catalogic competes with these other vendors. This is not the case as Catalogic is a fundamentally different kind of solution. This document explains the differences and shows why Catalogic is not competitive to Rubrik, etc., but a complementary solution.

The Appliance Approach: Rubrik, Cohesity, Actifio, etc.

In the backup appliance model, data is moved from the primary Pure Storage source to a target storage device. The appliance is usually sold as a hardware implementation, but it may also be software-defined. The data is typically moved by a client process on the host server or virtual machine. The impact is not zero as the backup will affect host CPU utilization, memory and bandwidth. In terms of architecture, these solutions are identical to traditional backup, though they may offer additional capabilities.

When using the data copies (for reporting, dev-test, etc.), a workload server connects to the backup appliance. If changes need to go back to the production side, they would be copied back to production over the network. Multiple workload servers can connect to a single copy, which is how efficiency benefits are realized. But performance characteristics of the appliance of course will be different than the production storage.

This model can be simply diagrammed as follows:

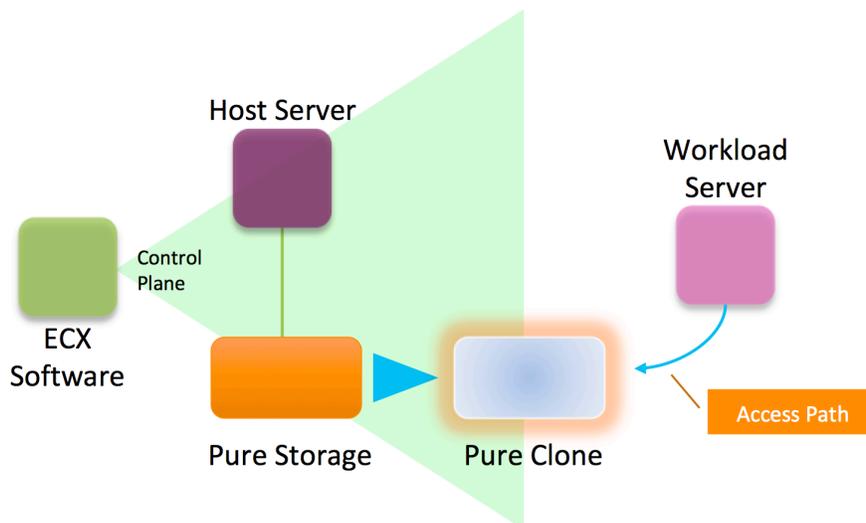


Catalogic In-Place Approach

Catalogic uses an In-Place Copy Data Management model. This In-Place approach uses the Pure Storage array copy processes – snapshots, replication, and clones – and leverages Pure deduplication and compression. Catalogic ECX doesn't replace any storage array features; it makes use of them, providing features like automation, self-service, application integration (for data consistency, log management, etc.), hypervisor integration, and so on. There is no separate storage device involved. Catalogic communicates to both the host server and the production storage, driving copy processes.

In this model, there is no data movement as storage snapshots are created on the array. Workload server access is provided by using the Pure Storage zero-footprint cloning features and connecting the snapshot clone to the server. With the In-Place approach, recoveries or data copying are far quicker because little or no data movement is involved, and nothing traverses the network.

This approach can be visualized as follows:



If array separation is desired, ECX can use the Pure Storage replication features to move data from one array to another. Because the data remains on Pure Storage, all the compression and deduplication savings on the primary are maintained on the secondary and during transfer. All restore functions can be performed on the replica copy.

Solutions Working Side by Side

Since each approach works differently, there is no reason they cannot be used together. In fact, Catalogic is typically deployed in environments that already have a backup solution running and ECX is not designed to be a broad-based backup solution.

Catalogic is primarily focused on operational recovery and data re-use for high performance database applications such as Oracle, SQL Server and InterSystems Cache. These applications require the kinds of performance and efficiency that can only be achieved by keeping the data in a Pure Storage environment.

Where ECX is the Right Fit

ECX is the preferred solution for the following scenarios.

Protection and Recovery of Large Databases

When databases grow too large for conventional appliance-based backups across the wire, ECX application-aware snapshot management provides the necessary functionality to make Pure snapshots and replication the principal method of protecting and recovering databases.

- ECX manages snapshot and replication scheduling, retention, and application awareness, similar to how backup software manages backups.
- Databases of any size can be protected in seconds.
- Compression and dedupe-aware replication from Pure to Pure greatly reduces copy time and bandwidth requirements.
- Log management and point-in-time recovery provide tight RPOs and pin-point recovery.

Data Re-use for High IO Workloads: Reporting, Analytics, Test-Dev, etc.

Appliance-based solutions can provide quick access to data sets, but their performance is limited. It may be sufficient for workloads that don't demand too many IOPs. But for high performance needs, it is far more effective to keep the workloads on Pure FlashArray and use ECX to manage the copy delivery process.

- ECX provides push-button delivery of data sets using Pure Storage clones.

- Clone creation can be automated and scheduled to refresh on a regular basis with no user intervention.
- End-user self-service allows data access by non-IT staff: developers, data analysts, etc.

Large Data Sets and High Copy Frequency

Organizations that require frequent creation of large data copies cannot wait for appliance solutions that can take many hours to transfer data sets. By keeping all data on Pure Storage, data compression and deduplication are maintained across arrays, significantly decreasing copy time, as does the ultra-high performance of Pure Storage. On average, Pure achieves nearly 5 to 1 data reduction. These benefits are lost using appliance solutions.

- Because ECX uses the native copy processes of the Pure array, all data reduction benefits are maintained during data transfer and on any array used as a copy target.

Dev-Test Data Format and Performance Consistency

In the appliance model, development work is done on a storage system that is fundamentally different from the production Pure storage array. It has different software, a different file system, different disks, etc. This introduces an element of uncertainty and risk in development work since when you move the application back to production you can't be certain it will work properly. When something goes wrong, it leads to finger-pointing as development blames IT for an infrastructure problem and IT blames development for problems with their code. Troubleshooting this kind of problem can be nightmarish, because the source of the problem could be anywhere on either side of the equation.

In the Catalogic approach, the same Pure array family and firmware are used in prod and non-prod, so all Dev work is done on a storage stack that is 100% identical to production. This is a no-risk infrastructure model for software development.

© Copyright Catalogic Software 2018
50 Tice Boulevard, Suite 110
Woodcliff Lake, NJ 07677
201.249.8980
U.S.A.
catalogicsoftware.com

United Kingdom: +44 (0) 207 712 1667
Germany: +49 (0) 2102 307614 0
Netherlands: + 31 (0) 20 347 23 88



Catalogic is a registered trademark of Catalogic Software Inc. ECX is a trademark of Catalogic Software Inc. All other company and product names used herein may be the trademarks of their respective companies.