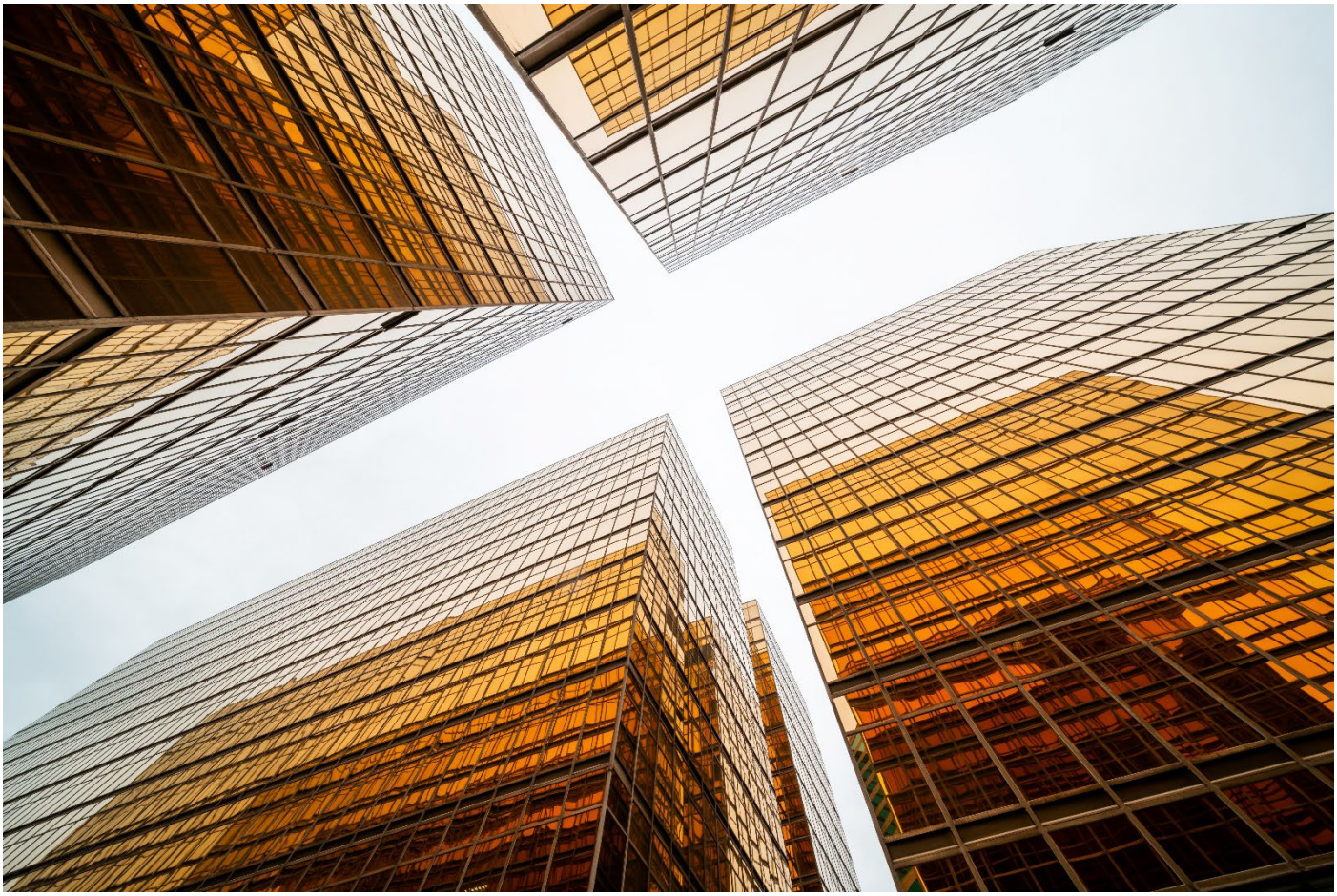
HPE GreenLake

Technical white paper

Ransomware data recovery architectures

The ability to recover your data if you are the victim of a ransomware attack



Contents

Introduction ... 3
Overview ... 3
Target audience ... 4
Terms and concepts ... 4
Perimeter defense: The foundation of a good ransomware protection solution is security ... 4
Detecting a ransomware attack ... 6
The 3–2–1–1 rule for data protection ... 6
Clean rooms ... 7
Immutable, read-only snapshots with retention time ... 8
Backup clean room ... 9
Array replication clean room vault ... 10
Checking clean room data for integrity ... 11
Backup validation process ... 12
Backup clean rooms are based on HPE StoreOnce and HPE StoreOnce Catalyst ... 13
Recovery after a ransomware attack ... 15
Ransomware data recovery architectures ... 15
Data recovery architecture #1 ... 15
Data recovery architecture #2 ... 16
Data recovery architecture #3 ... 17
Data recovery architecture #4 ... 19
Data recovery architecture #5 ... 20
Ransomware data recovery architectures for Zerto ... 22
Zerto architecture #1 ... 23
Zerto architecture #2 ... 23
Zerto architecture #3 ... 24
Zerto architecture #4 ... 26
Combining architectures ... 27
Conclusion ... 28

Introduction

Ransomware is malware that prevents an enterprise from accessing and using its data by encrypting it inconspicuously so that the victim does not know the attack is occurring. These attacks are extremely insidious in nature and can be very difficult to detect until long after significant damage has been done to the enterprise. After encrypting the data, the attacker suddenly denies access to it and demands a ransom to provide the encryption key that is needed to decrypt the data and make it accessible again—hence the term "ransomware." The ransom demanded might be in the thousands if not hundreds of thousands or even millions of dollars. These attacks are on the rise and are currently foremost in the minds of many CEOs and CIOs.

The ability to recover from a ransomware attack on the enterprise without having to pay a ransom is becoming a requirement that is higher in value than the ability to recover from a data center disaster such as a power outage, flood, fire, or other natural disaster. Ransomware protection is more important because a ransomware attack can infect and render unusable not only the primary production site data but also the disaster recovery (DR) site data, and even backup data, making recovery impossible unless special precautions are taken to protect the data. This means that a typical disaster-tolerant or DR solution is not sufficient to protect the data and enable recovery from a ransomware attack.

The means that hackers use to execute a ransomware attack and the holds they put on data to make it inaccessible are not only cunning but also constantly evolving. The attackers go to great lengths to hide the fact that an attack is occurring so that they can spend weeks or even months working their way deeper into the enterprise—compromising data and backups before announcing the attack and demanding a ransom. Because of this invisibility, the tools and processes used to detect a ransomware attack must be run regularly and consistently, must be constantly kept up to date, and must be run against data stored at all protection levels—from live production data all the way to backup data.

It is beyond the scope of this document to comprehensively address the tools, processes, and methods used to detect that a ransomware attack is occurring or has occurred in the enterprise. This paper focuses on solutions (that is, architectures) that the enterprise can deploy to provide a means of protecting an uninfected copy of data and restoring it after a ransomware attack has been detected. The paper also provides some rudimentary ideas and processes that can be implemented to try to detect a hidden ransomware attack (one that cannot be detected by current state-of-the-art virus scan software).

Each time a new tool, process, or set of virus definitions is deployed or updated, Hewlett Packard Enterprise recommends that the user run the new tool or the latest set of virus definitions on the last copy of data that was saved at each protected data layer in the enterprise. This precaution can detect an infection that is present but was undetectable by older versions of the tools and virus definitions.

Overview

Special measures must be taken to protect data to provide the potential for recovery following a ransomware attack. This paper describes various solution architectures that can be deployed to offer the potential for data recovery if the enterprise suffers an attack. For recovery to be achieved, a clean, uninfected copy of the data must have been saved somewhere before an attack occurs. That data in turn must be protected so that it cannot be attacked and compromised. Because the length of time an attack might go undetected is undetermined, the amount of time that data must be protected to provide a recoverable data store is also undetermined. Hewlett Packard Enterprise recommends saving at least a six-month repository of data for recovery, if needed. Therefore, Hewlett Packard Enterprise recommends using a very long-term archive storage layer such as backup to cloud and tape as the last line of defense.

The solution architectures described in this paper resemble standard DR solution architectures. (That is, they use many of the same tools and processes.) However, a solution that can provide recoverable data after a ransomware attack is different, and it must be deployed, managed, and treated completely differently from a normal DR solution. Unlike a traditional DR solution, the intent is not to provide a guaranteed recovery point objective (RPO), recovery time objective (RTO), or location to fail over to in the event of an attack. Instead, the intent is to try to ensure that a non-infected recovery point is available somewhere in the environment if an attack occurs. A corollary is that a solution designed to recover from a ransomware attack might be used quite successfully to recover from a DR scenario (such as a data center failure), although with more complicated recovery processes than a traditional DR solution requires and with a larger RTO than a traditional DR solution can generally provide. Think of it this way: A ransomware attack is a DR scenario in which the disaster occurred sometime in the distant past, but you are just now becoming aware of it, and you must recover data from before the point at which the disaster occurred.

Note
This paper uses HPE Alletra 9000 in its sample architectures, but everything it covers also applies to HPE Primera and HPE 3PAR arrays.

Target audience

The intended audience for this white paper is HPE Sales, HPE systems engineers, HPE partners, and customers who are interested in learning more about preventing ransomware and protecting against it.

Terms and concepts

- **Immutability**—Immutable data is data that cannot be modified or deleted by a client application/server that has access to it. HPE Primera and HPE Alletra 9000 arrays provide immutable volumes through read-only snapshots. Data in a read-only snapshot cannot be modified. HPE StoreOnce offers immutability for backups by applying a retention time that must expire before the backup application can delete them. Scality and public cloud offer S3 immutability to volumes as well.
- **Compliance**—A storage device that archives data marked as immutable should protect the data not only from client applications and hosts with storage access through the usual storage protocols, but also from storage administrators who might use the storage console to access the storage array for the purpose of deleting volumes or removing their immutability attributes.
  - HPE Primera and HPE Alletra 9000 offer compliance protection by applying a "virtual lock" to a read-only snapshot. After it is applied, a virtual lock makes it impossible for any storage administrator to delete the snapshot or its base virtual volume (BaseVV), or to change its RO attribute. A virtual lock can be applied to both base volumes (TPVVs or ADR volumes) and to snapshots on an HPE Alletra 9000 array. A virtual lock specifies the length of time, relative to the current time, that the volume must be retained before it can be deleted from the system. A volume with a virtual lock applied to it cannot be removed even by changing the system time to a time in the future.
  - HPE StoreOnce offers compliance protection by enabling the **dual authorization mode**. After this mode is set, every console operation that might lead to the deletion of a volume (also known as a Catalyst Store) requires a double authorization. An administrator starts the operation on the console, but it remains pending until a second special "security officer" user approves it. For added security, the credentials for this security officer account can be kept on paper in a closed envelope, and the account can be deleted after its use.
- **Perimeter defense**—Includes all sanitation and best practices that keep the enterprise safe from hacking and virus attacks, such as the following:
  - Keeping all software and firmware up to date on all equipment in the enterprise
  - Keeping all virus scan software up to date and running it on a regular basis
  - Making sure that all proper and recommended software settings are in place for the server OS and server application software
  - Training IT staff and enterprise employees to help them avoid becoming the vector that allows an intrusion into the enterprise through phishing or other social engineering gambits
  - Configuring networks properly and monitoring them to catch security holes that might allow intrusions into the enterprise, and if an intrusion happens, preventing it from migrating east-west between networks (see the term "air gap" in this list)
- **Clean room**—Contains an immutable copy of the data being protected. Any server using clean room data should be **air-gapped** (think firewalled) from the internet, from the production environment, and from other clean rooms. Each level of clean room has its own perimeter defense to prevent, to the maximum extent possible, the east-west migration of any infection that might get into the enterprise.
- **Air gap**—The concept of protecting data in a clean room by restricting access to it as much as possible. Think of an air gap as a firewall. It might include restricting network or administrator access, ensuring that the servers in a clean room that are used for virus scanning and backup are isolated from any production networks, and so forth. For example, an air-gapped tape is a tape cartridge that has been removed from a tape library.

Perimeter defense: The foundation of a good ransomware protection solution is security

Being able to recover data following a ransomware attack is necessary and a good practice, but at the heart of any solid ransomware protection solution is a secure environment in the enterprise that can prevent an attack from occurring in the first place. Having a solid line of defense against all malware (referred to as a **perimeter defense**) is necessary to prevent ransomware attacks.

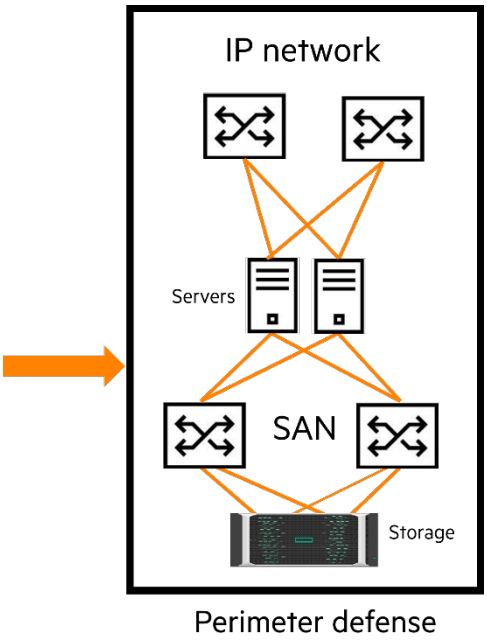


Figure 1. The best ransomware attack is the one that is stopped by the perimeter defense before it occurs

A strong perimeter defense protects the enterprise's servers, software, storage, networking, and SAN infrastructure from attack. It encompasses, but is not limited to, the hardware infrastructure in the enterprise. It ensures that essential precautions are taken:

- All appropriate software updates and necessary patches are applied promptly.
- All software is appropriately configured to close all known security holes.
- Secure passwords are in place.
- Users are properly trained not to inadvertently open the door to an attack vector through phishing or other social engineering schemes.
- The most current malware detection software is in place and is run on a regular basis.

It also includes customer security measures such as the following:

- Requiring personnel to be trained on protecting the enterprise through dual authentication or dual authorization (when available) for data-destructive administrative tasks
- Checking space usage on the storage arrays to help detect whether deduped and compressed data is being encrypted and thus is using more capacity than would normally be expected
- Investigating whether backups are consuming more space than expected, or whether unexpected data (historically static data) is suddenly being backed up

Protecting from a ransomware attack is not a one-time, fire-and-forget operation that you execute once and then ignore. It requires constant vigilance because the threats are constantly changing and evolving.

Note
For more information about securing the enterprise against malware attacks, refer to the HPE white paper Ransomware: Ensuring protection from an increasingly complex threat, and discover the latest additional documentation at hpe.com.

Detecting a ransomware attack

This paper does not provide detailed information about methods for detecting an active unnoticed ransomware attack. Hewlett Packard Enterprise recommends that you deploy available software tools (virus scan software, security monitoring software, data scanning software, and so forth) that might help detect that a ransomware attack is occurring or has occurred in the enterprise.

The most worrisome ransomware attack is one that is undetectable by current state-of-the-art virus scan software and tools. If an attack can be detected early by virus scan software (or by other means that detect a security breach of the enterprise), it can be recovered from quickly with minimal data loss. However, an attack that goes undetected for a long period of time can wreak irreparable harm on the enterprise, making recovery without paying the ransom next to impossible. The single best way to detect an attack is to inspect the data on the array to see if it has been attacked because this method can detect even a hidden ransomware attack that current state-of-the art detection software cannot detect. Checking the data on the storage array can be a cumbersome endeavor and difficult to do (and is not guaranteed 100% effective), but it should be seriously considered.

Note
For some proposed ideas about how to inspect clean room data for a ransomware attack, see the section Checking clean room data for integrity in this paper.

Note
For more information about securing the enterprise against malware attacks, refer to the HPE white paper Ransomware: Ensuring protection from an increasingly complex threat.

The 3–2–1–1 rule for data protection

The best way to ensure that you can recover from a ransomware attack is to have a solid data protection strategy in place, one that includes live snapshots, data replication, and a solid backup solution. When protecting from ransomware, skimping on your data protection strategy is not a good idea. A best practice data protection rule that can help effectively mitigate the threat of a malware attack is the **3-2-1-1 rule** for protecting data. The 3-2-1-1 rule specifies that you should:

- Have at least three copies of the data: the primary data and two copies.
  - **Minimum requirement example:** Primary, snapshot, and remote backup
  - **Usual configuration example:** Primary, snapshot, local backup, remote backup
- Store the copies on two different and independent storage devices: tape, disk, disk arrays, or cloud.

  For instance, you might use primary storage and backup storage. You must use two different storage devices. It is better if they are also different models with independent software so that if one fails because of a hardware or software issue, the other is not impacted by the same cause of failure. There is a consensus that replicated primary storage arrays are not considered independent and thus do not comply with the rule.
- Keep one backup copy offsite in the event of a site-wide issue or local hazards or infections within the network.
- Keep at least one backup copy air-gapped.

  This is a recent extension to the 3-2-1 rule aimed specifically at protecting against cyberattacks such as ransomware. This copy of data cannot be altered or deleted by anyone, even if they have administrative credentials and a remote connection to the backup console or the storage repository. The classic example of an air-gapped solution is tape media that has been extracted from the library and stored in a vault. A new approach is based on immutability and compliance; that is, the backup copy is marked as immutable on the storage device, and even an administrator with remote access to the storage cannot alter the immutability attribute or its expiration date.

Hewlett Packard Enterprise recommends that you seriously consider the air-gap requirement for any existing or new data protection solution. In a world where ransomware attacks are constantly evolving and can instantly take you offline, it is a vital precaution.

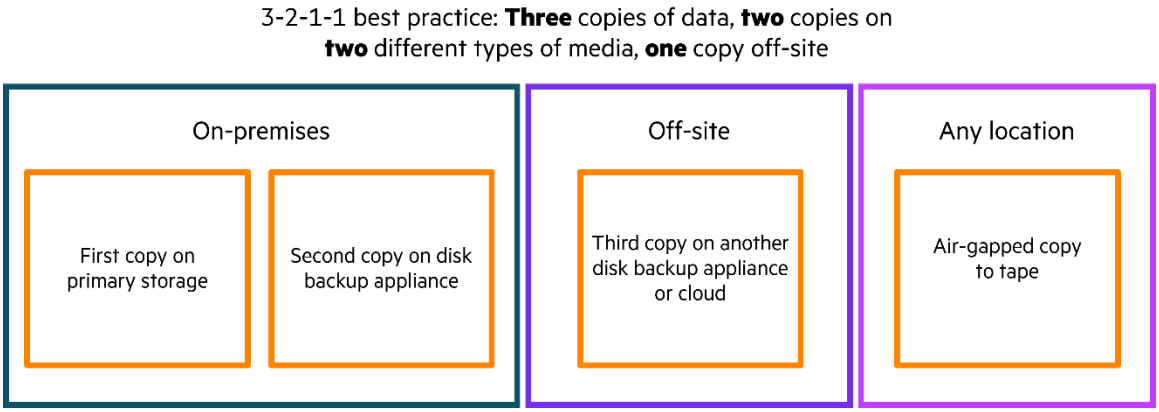


Figure 2. The 3-2-1-1 best practice rule for data protection

Clean rooms

The ransomware data recovery architectures presented in this paper use the concept of **clean rooms** where protected data resides. A clean room is a copy of data that is air-gapped from the production environment and from other clean rooms in the enterprise. It contains an immutable and locked copy of the data being protected so that the data cannot be modified and cannot be removed for a predetermined interval of time. Creating read-only snapshots of volumes creates a copy of data that cannot be modified. A virtual lock is applied to read-only snapshots on HPE Alletra 9000 arrays to prevent the snapshots from being removed.

After an immutable copy of data is put into a clean room, it can be inspected to determine whether it has been silently attacked. If it is found to be good, it is locked and designated as a potential recovery point. This data is then available to be used by the next level of clean room and can be used for recovery if necessary in the event of an attack. Each successive level of clean rooms should be air-gapped from the previous level (and from the production environment), making it more secure than the previous level of clean room and therefore less likely to be affected by a malware attack than the previous clean room.

Note
Each time a new ransomware detection tool, new process, or new ransomware anti-virus definition is deployed, Hewlett Packard Enterprise recommends that you run it against the last copy of data that was saved in each clean room to check for an intrusion that was not detected by the older version of the tools and processes.

Clean rooms should be kept as isolated and "sterile" as possible so that an attack on the enterprise cannot easily move from one clean room to the next, infecting data. The arrays, servers, network gear, software, and so forth in each clean room should be isolated as much as possible from production and from other clean rooms to prevent the spread of infection. Having each successive clean room air-gapped from other clean rooms helps prevent the spread of any attack. Care must be taken with the avenues used to move data from one clean room to the next so that they do not become vectors in the spread of the attack. For example, with regard to the replication clean room "vault" array discussed later in this paper, Hewlett Packard Enterprise recommends using either direct connect RCIP links or RCFC[1] for replication between the enterprise production array and the clean room vault array so that the replication network cannot become a vector for ransomware spread.

How frequently new copies of data are created and moved into a clean room and how long the data is maintained in a clean room before aging out might depend on a number of factors. Like a disaster-tolerant solution, the choice of how long to retain a copy of data in a clean room is most often driven by cost. The smaller the RTO desired from the solution, the more expensive the solution becomes. For a smaller RPO and RTO, more data must be stored more frequently and for a longer period of time in expensive clean rooms closer to the production environment (immutable snapshots on the production array) to help deliver the smaller RTO because ransomware attacks are insidious and might not be detected or announced until long after the attack has initially occurred. The farther a clean room is from the production data, the safer the clean room data is, but the longer it will take to restore data to a usable production state if its room data must be used, which ultimately increases the RTO.

[1] Direct connect RCFC is not supported, but switched FC cannot be used as a vector for ransomware spread.

Data gets into a clean room by one of three means:

- Immutable (read-only) snapshots that have a retention time assigned to them
- Immutable backups, including having an immutable backup catalog
- Array replication (after being replicated to a "vault" target array, an immutable copy of the data is created through read-only snapshots with retention times on the vault array)

As stated earlier, ransomware attacks are insidious, hidden, and can be ongoing for a long time before being detected. Data must be protected for a period longer than an attack has been occurring to ensure there is a copy of data to recover from. Figure 3 provides some perspective on the timeline of a ransomware attack.

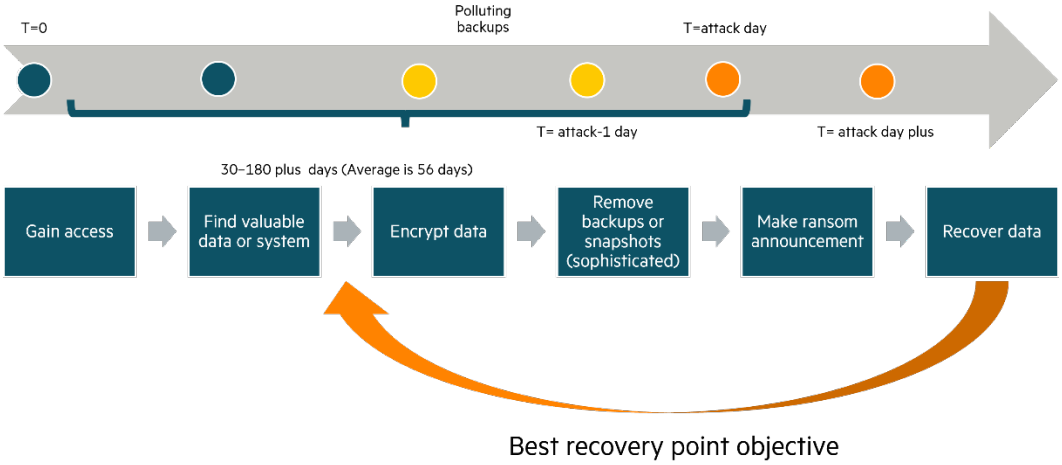


Figure 3. Ransomware attack timeline

The next sections take a deeper look at each of the different types of clean room.

Immutable, read-only snapshots with retention time

The first level of clean room is created on the production enterprise array and consists simply of read-only snapshots with a defined retention time (virtual lock) applied to them. Because these snapshots are created as read-only, they cannot be modified, and because they have a retention time applied to them (the virtual lock), they cannot be deleted until the retention time expires. After they are created, the data they represent can be inspected for data integrity; if it is found to be good, they are locked.

See the section Checking clean room data for integrity for a discussion about how to inspect clean room data to determine whether it has been attacked.

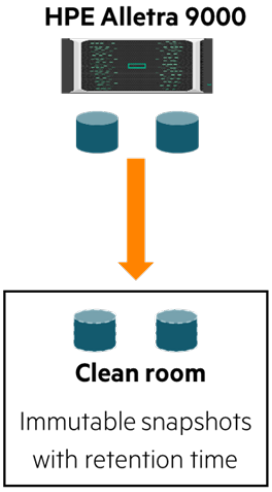


Figure 4. Basic snapshot clean room

Important
A volume (BaseVV or snapshot) with a retention time (virtual lock) applied to it cannot under any circumstances be removed from the array until the retention time expires. This includes any volume with a snapshot that has a retention time applied to it. Neither the volume nor the snapshot can be removed until the virtual lock expires. Hewlett Packard Enterprise recommends applying retention times of a reasonable to both BaseVVs and snapshots and extending those retention times when they expire rather than setting retention times of extended duration.

Note
If the array runs out of space, snapshots on the array go stale and become unavailable to use for recovery. Monitor space usage on the array to prevent it from running out of space. Unusually high space consumption on the array is a sign of a possible ransomware attack that is encrypting data on the array and that might have the potential to make the array run out of space.

Backup clean room

The next level of clean room contains immutable and offline backups. These backups can land on a public or private cloud if desired, but Hewlett Packard Enterprise highly recommends that off-line tape also be used. Landing the data on a public or private cloud provides a copy of the data that can be used for relatively quick recovery if necessary. Even when immutable backup to cloud is used, Hewlett Packard Enterprise recommends also creating an offline backup to tape. An offline air-gapped tape copy of the data is very safe, and it is difficult for it to be compromised. That is why Hewlett Packard Enterprise highly recommends having an offline tape copy of the data for the highest level of protection.

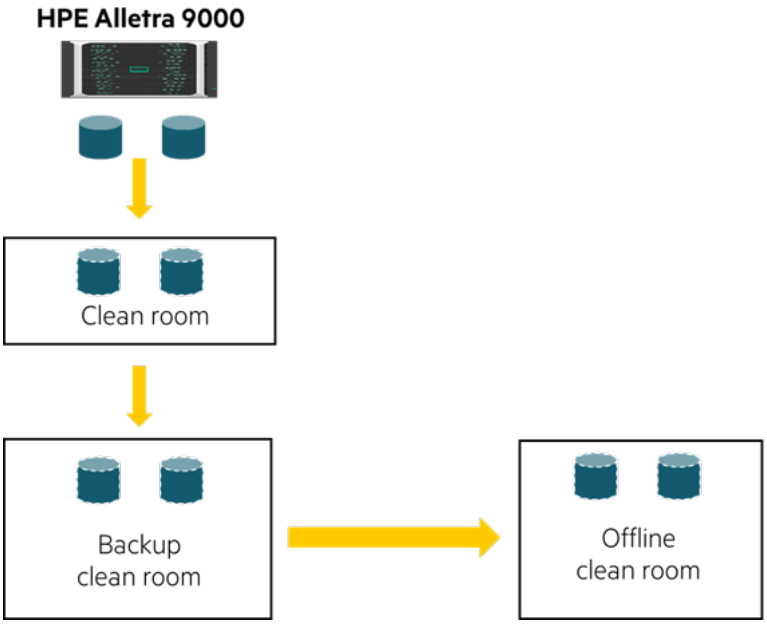


Figure 5. Basic snapshot clean room feeding a backup clean room

Array replication clean room vault

Array replication can be leveraged to provide an extra level of backup clean room protection by providing data to a backup clean room. The replication target array and associated servers (backup, virus scan, and so forth) should reside on an air-gapped or "segmented" network that is isolated from the enterprise's production network to the greatest extent possible. The vault array and associated servers are not used to run production, test, or development—it is a replication target clean room that should be used only to check data integrity and provide data for backups, and for recovery if necessary. You can think of it as a vault to which no production servers should have access. Its clean room should be hardened, isolated, and restricted as much as possible to protect it from a direct ransomware attack.

The only servers that should have access to the clean room vault array are the ones used to check data for integrity and the ones used for backup. Servers in the vault clean room should be given access only to snapshots of the clean room volumes on the array and not to the Remote Copy replication target volumes themselves. All manner of precautions should be taken to isolate any server connected to the clean room vault array from all networks and attack vectors that have the potential to result in infection from the outside. Hewlett Packard Enterprise recommends using direct connect RCIP or switched RCFC replication links rather than switched RCIP links. Replication target volumes on the clean room vault array should never under any circumstances be exported to any production servers. The clean room vault array is not intended as and should not be used as a DR failover target. Directly exporting the volumes on the clean room vault array to production servers exposes them to the very ransomware attack vectors from which you are trying to protect them.

Periodically, production data is replicated to the clean room vault array. Immutable snapshots of the volumes are created, inspected for integrity, and—if determined to be clean—are then locked.

Copies of the immutable snapshots in the replication vault clean room can be used to feed a backup clean room if desired. Hewlett Packard Enterprise recommends that backup clean rooms use a clean room vault array's data as opposed to using the production array directly as the backup source because a vault array provides an extra level of protection for the backup servers and data.

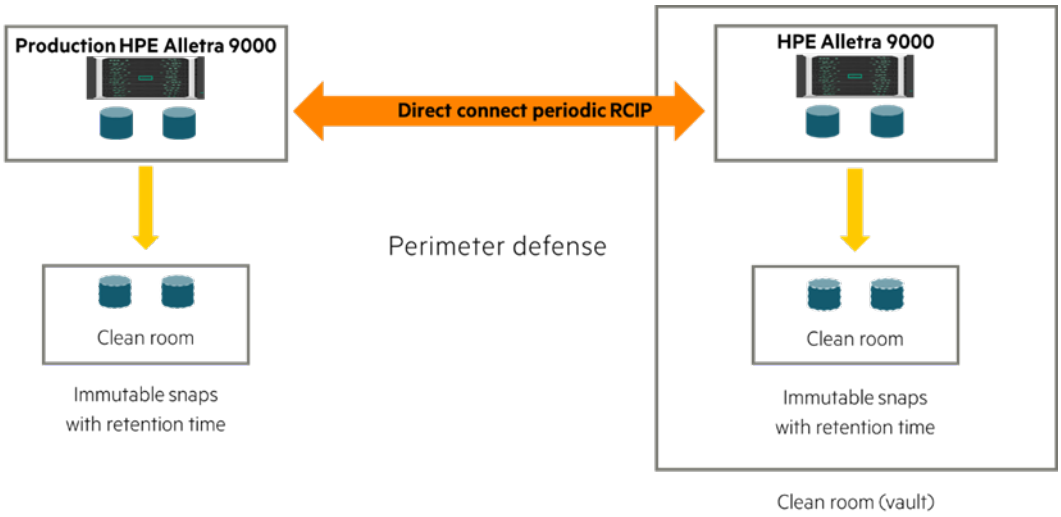


Figure 6. Array replication to an air-gapped clean room vault

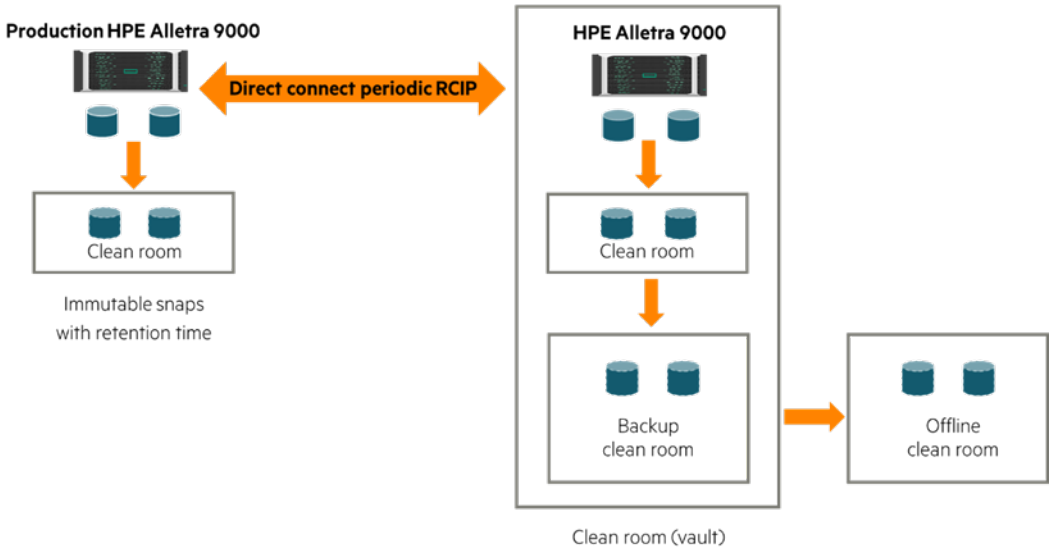


Figure 7. Backup clean room on the clean room from the vault array

Checking clean room data for integrity

A number of products advertised by third-party software companies claim to be able to inspect data on disk or backups and report if it has been attacked by ransomware. Also, many backup vendors claim to be able to detect ransomware in the enterprise or detect when data has been attacked by ransomware based on statistical analysis of data change rates. Although these offerings can be effective at finding some attacks against data, none of these solutions seems to be 100% effective at detecting all ransomware attacks. In addition, the bad actors have been known to purchase these software products, learn how they work, and then design their ransomware to avoid detection by them. It is beyond the scope of this paper to review and analyze ransomware data detection solutions offered by third-party vendors, but it is important to be aware that they exist.

Ransomware attackers go to great lengths to avoid being detected. They encrypt so little data in each file data that simple metrics such as increased space consumption are not usable in detecting the attack. You are in essence “looking for a needle in a haystack” with regard to detecting affected data on the array. And if the data written by a server is being legitimately encrypted by the application, then simply checking to see whether the data on the disk is encrypted will not be sufficient to detect an attack.

Directly inspecting the data on the array—looking to see if it has been encrypted surreptitiously by a ransomware attack—is the surest way to detect an attack. (Even data legitimately encrypted by an application can be re-encrypted by ransomware.) A new silent attack not known

to the latest-and-greatest virus scan software can be detected by inspecting the data on the array and looking for anomalies in that data. After all, it is not the servers or the storage array that are being attacked—it is the data that is being attacked (encrypted), and the servers are simply a vector the attackers use to get to the data. Inspecting data on the array, looking to see if it has been compromised, can be very effective, but it can be very cumbersome as well. The larger the data pool and the smaller the change rate, the more difficult it becomes to detect an attack by looking at the data. Understand, however, that if an attack has occurred, it is 100% guaranteed that data on the array has been encrypted, and if this encryption can be detected, the attack can be uncovered.

In addition to off-the-shelf ransomware detection software product offerings, Hewlett Packard Enterprise recommends implementing some simple processes that can be run against data on your enterprise array after it is put into a clean room. These checks help to determine whether the data has been affected by ransomware. For the following processes to be effective, it is absolutely imperative that all server, hypervisor, and virtual machine (VM) images used to inspect the data have fresh **golden images** that you are certain have not been infected by ransomware. To achieve this, Hewlett Packard Enterprise recommends booting from SAN using golden images that were installed on the storage array from a fresh air-gapped server. After the golden image boot device for a server is installed on the array, you should create an immutable read-only snapshot of it with a virtual lock. Then, every time a clean room server is to be used, the read-only snapshots of the golden image boot device for that server are promoted to the golden image boot device before the server or VM is used. This approach sanitizes the golden image boot device, wiping any infection that might have made its way onto the boot device the last time the server was used. The server, hypervisor, or VM is then rebooted, using this freshly "sanitized" golden image boot device. This process of creating golden boot devices and taking immutable snapshots that are promoted every time before a server is used should also be applied to any backup server, any server from Zerto, a Hewlett Packard Enterprise company, or any other server that is utilizing data in a clean room or moving data between clean rooms.

The following processes can be implemented to perform some checking of the data in a clean room. Note that these are simple and easy-to-implement checks and that none of these processes guarantees 100% success in detecting compromised data. The larger the percentage of data inspected, the better the chance of detecting a silent attack.

- The first option, if the data being checked is filesystem data, is to have a set of "test" servers in the snapshot clean room that are air-gapped from the production environment and that have known good, uninfected golden OS images on them. Read-write snapshots are created of the immutable read-only snapshots placed into the clean room. These read-write snapshots are used to bring up the filesystem on the clean room servers, and filesystem integrity checks can then be run against the data. In addition, random checks of the data in the files themselves can be conducted, looking for corrupt or altered data.
- For non-filesystem data (an Oracle® database or an SQL database, for instance), an instance of Oracle is spun up on an Oracle golden test server in the clean room, and the database is started using read-write snapshots of the immutable read-only snapshot data put into the clean room. If the database has been attacked, Oracle Database should complain of database corruption while starting, and if not, a database integrity check can be run against the data looking for database corruption. (A ransomware attack will appear as a database corruption event.) In addition to testing the database for integrity (which checks database metadata), some database test queries can be run against the data to check for infection. If the data is found to be clean, then the immutable read-only snapshots can be marked as potentially safe and usable for recovery if needed.
- A "honeypot" can be created in the production environment to which a software agent on the production server writes a known data pattern. The honeypot can be a completely standalone volume, or multiple volumes, with a filesystem if desired, but simply creating honeypot files in the production filesystem for use might be adequate. When snapshots of the clean room data are mounted on the clean room filesystem test server, an agent on that test server checks the contents of the honeypot file to ensure that only the proper data pattern has been written to it. If the data found in the honeypot is not what was expected, this is an indication that the data has been compromised.

The larger the volume of data inspected by using these techniques, the better the chance of detecting a silent ransomware attack. After the data in a clean room has been tested and is thought to be "clean," it can be marked as available to use for recovery if needed and can be used as the source to provide data to the next level of clean room.

Backup validation process

It is very important to have a backup validation process that inspects the backup data for integrity to the greatest extent possible. Having a corrupted backup is not only useless but also provides a false feeling of protection. The backup validation process is independent from any standard antivirus solution that is run in the enterprise to check production servers for attack. It can, in theory, detect a new attack that is not detectable by the latest ransomware detection software. The backup data must be inspected by a known certified, clean server that resides in a clean room. This is necessary to ensure that the server used to perform the validation is not infected with malware that would prevent it from detecting and reporting any data corruption.

Note
At the time of this writing, HPE is not aware of any commercial data protection product that can automate the backup analysis from a clean room with a clean, sanitized server disconnected from the production backup server; therefore, some level of customization will be necessary.

The backup validation must be conducted directly on the data that the backup process has written to the backup target (for example, HPE StoreOnce). Often, backup is a two-step process: First, an I/O-consistent snapshot of the data to be backed up is generated, and second, the snapshot is used as a source for saving to a long-term backup repository such as HPE StoreOnce. (There might also be a step in which the backup server requests the application to quiesce itself to provide not only I/O-consistent data but also transactional consistency of the data before the snapshots are created.) In this case, the inspection should ideally be executed on the final copy of data after it has landed on the backup repository. This approach enables the detection of an attack that has affected the backup server itself.

Generally, backup repositories are tuned for capacity and sequential I/O performance and do not have enough performance headroom to run integrity checks after every backup cycle. To speed up the data integrity checking process, it is sufficient to inspect the snapshots that are used as the source for creating the backup instead of inspecting the backup data after it has landed on the backup store. If you choose to inspect the snapshots that are used to source the backup rather than check the actual backup data on the backup, it is imperative that the server performing the backup to the long-term retention repository be isolated in a clean room, booted from a certified clean OS image before backing the data up to the backup store. This is done to help prevent the backup operation itself from introducing corruption that would not be recognized until the data was restored and an attempt was made to use it. Hewlett Packard Enterprise recommends regularly restoring and inspecting the data saved to the long-term repository as the best way to ensure its integrity. To help reduce the impact of this inspection process, backups can be inspected at a lower frequency (for example, inspecting only one in every five backups rather than every backup) while still retaining the assurance that the data saved to the backup repository is good. Although this practice reduces the chance for fast detection, it still provides good protection but with a lower impact on the infrastructure resources and their cost and potentially a larger RPO if the data is attacked.

Many commercial backup applications have an automated process for inspecting and validating backups. For example, Veeam SureBackup can boot a server by using an image located on a storage array snapshot or even on HPE StoreOnce itself. After the server is booted, it is connected to an isolated network by Veeam SureBackup. The server's integrity is then checked. After its integrity is confirmed, the server can be used to check the integrity of the backup data.

Backup clean rooms are based on HPE StoreOnce and HPE StoreOnce Catalyst

The backup clean rooms referenced in these architectures are based on the HPE StoreOnce appliance and on HPE StoreOnce Catalyst. (HPE StoreOnce VSA is also supported.) The use of HPE StoreOnce is not required, but Hewlett Packard Enterprise highly recommends using it for the backup control and space savings it can provide.

Note
For more information about securing backups against malware attacks, refer to the HPE white paper Protecting Data from Ransomware with HPE StoreOnce Catalyst.

After you have a backup repository that has been inspected and validated for integrity, it is necessary to ensure that the repository's integrity is preserved for the entire length of the backup retention. This requires ensuring that the repository is immutable to both modification and removal because, during a ransomware attack, an attempt will be made to delete or corrupt backup data. This specific protection is described in the following sections.

Immutable backups

Backup processes save a copy of production data at regular intervals for the purpose of restoring the data in case production data is damaged, deleted, or erroneously altered. Clearly, the recovery process following a ransomware attack will fail if the backup data has been compromised or deleted by the same ransomware attack that damaged the production data. In the case of a ransomware attack, a proper data protection strategy must include protection for additional threats on top of the normal "legacy" threats such as human error and hardware failures. In addition, a ransomware attack might include two other types of threats:

- **Backup encryption:** When production data is infected by ransomware, even backup repositories can be attacked and encrypted as part of the attack. An example would be to have production services running on Windows Server instances and backing up that data to a file system on other Windows Server instances. In this situation, if malware infects the production environment, there is a high risk that the servers storing the backup data might be infected as well. When the backup data is used to perform a restore, it will be discovered that the backup data has been compromised, rendering the ability to restore good, usable data impossible.

- **Backup deletion:** A very common phase of many ransomware attacks includes a direct attempt to destroy backup data after the production data has been encrypted. If the attacker destroys backup data, the victims cannot use that data to recover from the attack. To destroy backup data, attackers might, for example, connect to the backup infrastructure by using stolen administrative credentials and delete every backup copy. An attacker might even bypass the backup application and delete data directly on the storage device—for instance, accessing the file system by using root/administrator privileges.

Backup data **immutability** and **compliance** are the most important aspects required to protect backups from attack. Not all immutability implementations offer the level of protection necessary to ensure that a solution has full protection of the backup data should an attack occur. Particular attention must be given to the backup target storage and the privileges assigned to its administrators. It is evident that if administrators can bypass the immutability enforcements and delete data, then the protection is weak and can be exploited by a bad actor, including a disgruntled employee.

The backup data (the backup store) and metadata (that is, the backup catalog) must both be immutable so that they cannot be encrypted, deleted, or overwritten during a ransomware attack. An attack that deletes or encrypts the backup catalog would render the backup data unusable. Some backup applications do not rely on a catalog to be able to restore backups. A backup to tape, for example, can be used to restore a backup even if the backup store or backup catalog has been compromised. A backup to tape for which the tape cartridge has been removed from the tape library (known as "air-gapped") is the highest level of backup protection; therefore, Hewlett Packard Enterprise highly recommends including backup to tape in addition to backup to online media such as disk and cloud.

Note
Regardless of where the backup store and catalog reside, Hewlett Packard Enterprise recommends also using tape as an offline/air-gapped backup target for further protection from a possible ransomware attack. A backup to tape can be used to recover, even if the backup catalog or backup store is compromised.

Data immutability with HPE StoreOnce

HPE StoreOnce provides a high level of immutability protection. When immutability is set on an HPE StoreOnce data object, that object remains immutable until the immutability time period set for it expires. HPE StoreOnce supports two methods of protection for defining the immutability period for objects written to a Catalyst Store (that is, to a volume inside HPE StoreOnce):

- **HPE StoreOnce managed immutability:** The HPE StoreOnce administrator defines a common immutability period attribute for a Catalyst Store. When an object is written to that Catalyst Store and the operation is marked as complete, HPE StoreOnce sets the immutability period specified in the attribute. Each object receives the same immutability duration.
- **Application-managed immutability:** The immutability period is defined by the backup application and is communicated to HPE StoreOnce for each data object. In this case, the backup application has the flexibility of assigning different immutability periods to different objects.

Immutability periods and backup retention periods are different concepts. To avoid generating errors when the backup application tries to delete an expired backup for which the immutability period has not yet expired, it is a good practice to set an immutability period that is shorter in duration than the intended backup retention period.

If an object is marked as immutable, neither the backup application's GUI, its CLI, nor its integration API can delete or alter that object—HPE StoreOnce will refuse to execute the operation. HPE StoreOnce also complies with the immutability enforcement when an administrator connects directly to the unit through a management interface such as the CLI or the GUI. By design, no HPE StoreOnce administrator has the privilege necessary to delete immutable data. Even if a bad actor were to obtain the HPE StoreOnce administrator credentials, he could not delete immutable data.

In rare situations, HPE StoreOnce administrators might legitimately need to delete immutable data before its immutability period has expired. This might be necessary, for example, to clean data after a non-production backup test, or to fix a configuration error in which the immutability period was accidentally set to 30 years instead of 30 days. In these situations, there is a way for the system owner to reclaim the capacity he paid for. This process is based on a dual authorization process. An administrator starts the "delete" operation at the HPE StoreOnce console, but the command remains pending until a special "security officer" approves it. The object is not actually removed until the security officer approves the removal. For added protection, the security officer account credentials can be kept printed on a piece of paper in a closed envelope, in a locked drawer, and the account can be deleted after its use. This way, a bad actor cannot reuse the credentials to create disruption. In addition, security officers cannot start any administrative operation—they can only approve or deny pending operations.

It is worth highlighting that the dual authorization model offers a higher level of protection than multi-factor authentication (MFA) because it protects against a different class of attack that MFA does not protect against, such as an internal attack by a disgruntled employee (not to mention that hackers are finding ways around dual authentication requirements). With MFA, an administrator can still create a disaster, and there are organizations (such as banks, financial institutions, and intelligence agencies) that do not want to assume that risk. With dual authorization, a disgruntled administrator cannot complete the disruptive operation on his own, but must have the security officer approve of any data destructive operation.

Recovery after a ransomware attack

If a ransomware attack is detected in the enterprise, not only must the data be recovered (restored) to a point in time before the attack occurred, but everything in the enterprise, including storage, servers, network switches, software, firmware, and so forth must all be sanitized to the point that no telltale sign of the ransomware is left in the enterprise before processing in the enterprise can resume. Recovering to a good point in the data only to have some latent ransomware attack the data again would be a catastrophe. It is beyond the scope of this paper to discuss how to sanitize the enterprise to remove all latent signs of any ransomware following an attack. However, this an extremely import aspect of any good ransomware protection and recovery plan. Remember that ransomware does not attack the storage devices themselves; it attacks the data stored on the storage devices, and this attack comes from the servers using those storage devices and is spread through the networks connecting those servers.

Ransomware data recovery architectures

This section and the next outline suggested architectures that can provide usable data for recovery in the event of a ransomware attack. Remember that a silent attack that goes on for a period of time before detection will result in a data recovery point (RPO) that can be very large. Implementing these architectures requires considering the operating environment in which they will be deployed. Factors such as array performance headroom and array free capacity must be taken into account when any of these solution architectures is deployed.

Data recovery architecture #1

The first data recovery architecture is very basic and is actually the foundation for all other architectures discussed in this paper. Hewlett Packard Enterprise does not recommend relying on this architecture alone because it does not adhere to the best practice 3-2-1-1 rule for data protection. However, it is a first layer of protection—and in many situations, snapshots are the fastest way to execute backups and restores. There are, however, limits to how often and how many snapshots of the production data can be maintained on the production array. Therefore, it is necessary to add other layers of protection that can be used to restore data in all the situations for which protection based on storage snapshots alone is not enough.

Note
Hewlett Packard Enterprise does not recommend using data recovery architecture #1 because it does not conform to the 3-2-1-1 best practice rules for data protection.

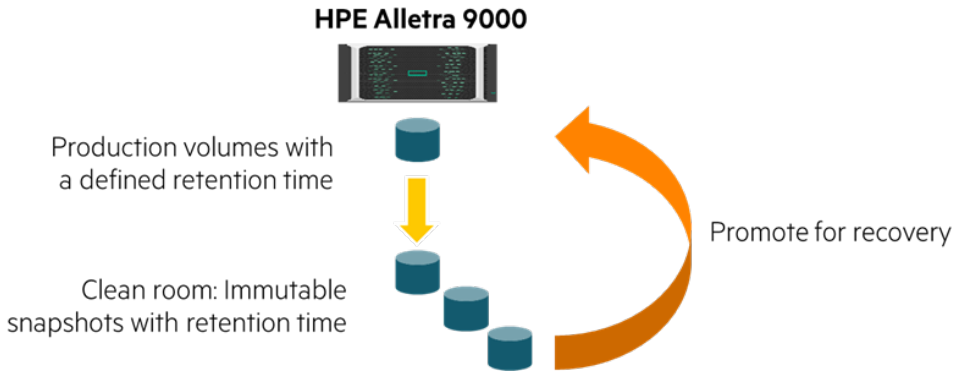


Figure 8. Data recovery architecture #1—Basic snapshot clean room

The clean room in this architecture contains simply a set of immutable (read-only) snapshots that are created from the production data volumes. After these snapshots are created, they are inspected for data integrity if desired. A retention time (virtual lock) is then applied to them so they cannot be deleted by a malware attack. Consideration must be given to the retention time placed on production data and the clean room snapshots. A volume with a retention time applied to it cannot under any circumstances be removed from the system until the retention time expires, unless the array is reinitialized—in which case, all data on the array is destroyed. Even a BaseVV with a snapshot

hanging off of it, for which the snapshot has a virtual lock applied to it, cannot be deleted until the virtual lock on the snapshot expires. Setting retention times of extremely long duration can result in a situation in which a volume that is desired to be removed or required to be removed (for example, after tuning to a different volume type) cannot be removed in a timely manner. The retention times chosen should be of short-to-moderate duration and can be extended if necessary. This time limitation provides the flexibility to remove volumes and snapshots if the need arises. If a ransomware attack occurs, the newest set of good clean room snapshots can be promoted to recover the production volumes back to a state from before the malware attack occurred.

How frequently a new set of clean room snaps are created on the production array and how long they are maintained depend on several factors:

- **RPO desired by the customer.** RPO is a definition of the amount of data that might be lost if a malware attack occurs and is defined as an amount of time. The smaller the desired RPO, the more frequently a new set of snapshots must be placed in the clean room. There is no guarantee that the desired RPO will be achieved; the malware attack might have started long in the past and gone unnoticed for an extended period, rendering the desired RPO unachievable.
- **System configuration.** There are system limits on the total number of volumes supported and the number of snapshots allowed per VV. Also, snapshots consume capacity on the array as the volumes they are created from are written to. These limits and the length of time a set of snapshots is retained might limit how frequently a new set of snapshots can be placed into a clean room and how long they can be retained.
- **RTO desired.** RTO is a measure of how long it takes to recover from an attack and to get back up and running. Before a set of snapshots in the clean room can be used for recovery, they must be declared free from attack. Although frequent clean room snapshots help provide a small RPO, if the clean room is being populated with snapshots too frequently, it might not be possible for the processes that inspect and certify their data integrity to execute quickly enough to keep up with a desired small RTO.
- **Duration.** A hidden ransomware attack can be months in the making, so it is not realistic to assume that snapshot clean room data can be retained long enough on the production array to guarantee detection of a ransomware attack before data starts aging out of the clean room. Because the duration from when a ransomware attack occurs until it is ultimately detected is undetermined and might be hours to days to weeks to months, it is entirely possible that recovery might have to be addressed by a second or even third level of clean room seeded with data from the production array clean room snapshots. These situations are covered in data recovery architectures 3 through 6.

Data recovery architecture #2

Architecture #2 extends the base architecture #1 to include a backup clean room that can provide onsite, cloud, and offline tape backup. This solution does adhere to the best practice 3-2-1 rule for data protection and to the HPE 3-2-1-1 rule if an air-gapped tape backup is included.

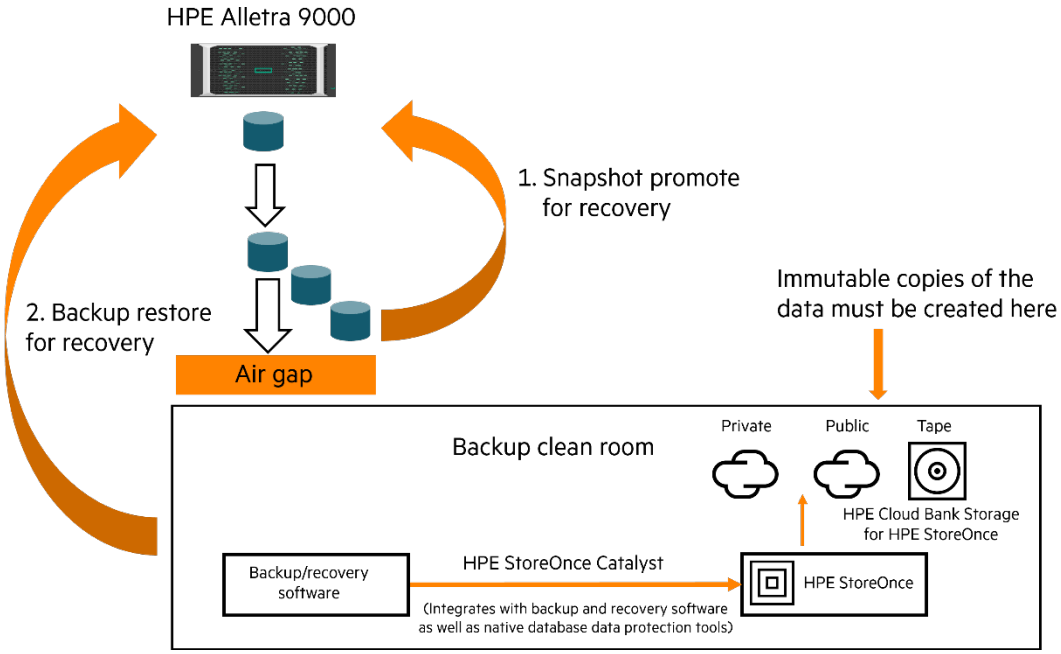


Figure 9. Data recovery architecture #2—Backup clean room with offline tape

In this solution, data from the production array snapshot clean room is used to feed a backup clean room. The backup solution should not use the production VVs, but should instead use the production array clean room snapshots that have been inspected and declared free from infection. If the backup software in use requires read/write copies of the data, then read/write snapshots of the production array clean room snapshots should be created for it to use. The backup can land wherever the customer wants, including in a private or a public cloud. A best practice would be to ensure that at least one copy lands on air-gapped tape.

With this solution, if a ransomware attack occurs, the process to recover data involves the following steps:

1. Recovery is first attempted by using snapshots from the snapshot clean room on the production array. If good data exists in the clean room on the production array, those snapshots are promoted to provide recovery.
2. If recovery cannot be achieved by using the production array snapshot clean room data, then successive backups are restored and checked until a clean copy of the data is found. The necessity to restore data from the backup clean room can have a significant effect on the RPO of the solution.

The backup servers and any servers used to check the snapshot clean room data for infection should be isolated as much as possible (air-gapped) from the networks that the production servers are using to make it more difficult for a single attack vector to affect both the production and the backup servers. In essence, they should be given a physically separate perimeter defense network to the extent possible, even though they share a common array.

Data recovery architecture #3

This data recovery architecture uses the periodic asynchronous Remote Copy feature to place a point-in-time copy of data from the production array onto a separate air-gapped clean room "vault" array. The vault array should not be used as a DR failover site to which production workloads are failed over and run on the replication target array following a ransomware attack or other disaster. Doing so would expose the data on the clean room vault array to attack.

The intent of the architecture is to make the clean room vault more secure than the production environment by limiting outside access to it. The vault array and all servers in the vault clean room are air-gapped away from the production environment. No production or failover servers should be connected to the clean room vault array. The only servers that should be connected to the vault array are the ones used to check the clean room data for integrity (and in the case of architecture #4, the backup servers).

Data moves from the production array to the vault array through asynchronous periodic Remote Copy. For the transport method between the two arrays, Hewlett Packard Enterprise recommends using direct-connect Remote Copy over IP (RCIP) or Remote Copy Fibre Channel (RCFC) through SAN switches if possible. (Direct connect RCIP is preferred.) If RCIP is used but a direct connect configuration cannot be used, then any switches used to connect the production and vault arrays together to provide the RCIP transport should be isolated as much as possible from the corporate network and the clean room servers. To prevent the network from being used as an attack vector, clean room

servers should not use the replication network for communication. After the volumes on the clean room vault array are updated through async periodic RC, immutable snapshots of the volumes are created. The snapshot data can then be checked for integrity. If they are found to be clean, retention times are then applied to them.

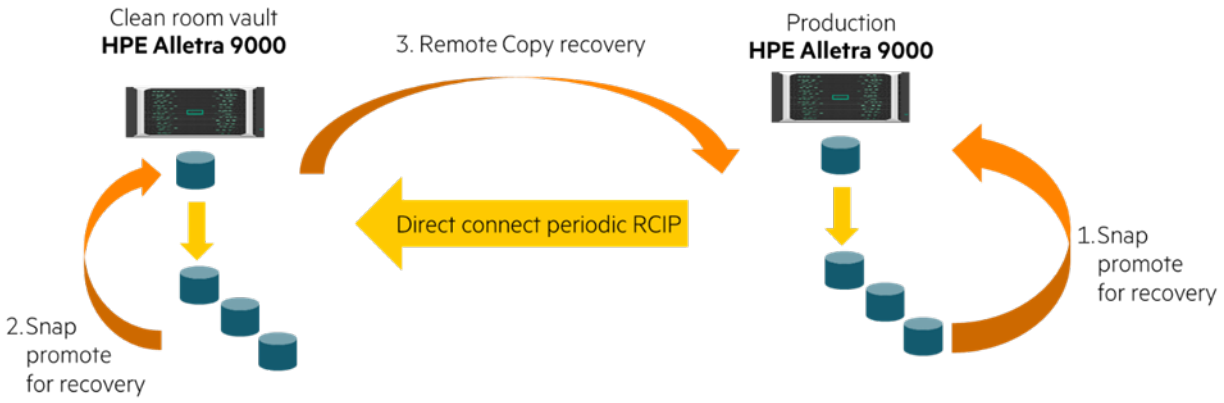


Figure 10. Data recovery architecture #3—Replicated clean room vault array

Note
A best practice is to use direct connect RCIP as the Remote Copy transport between the arrays in the Remote Copy configuration.

Hewlett Packard Enterprise recommends that asynchronous periodic Remote Copy not be configured for regular delta updates of the vault array. Instead, updates to the vault array should be controlled by setting the Remote Copy group delta resync interval to zero and updating the vault array through the `syncrcopy` CLI command on a planned schedule or by requesting coordinated snapshots of the volumes in the RC group through the `creates -rcopy` command (see Figure 11). After a Remote Copy resync to the vault array has completed, the data is inspected for intrusion. If it is declared clean, a set of immutable snaps of the replication target volumes are created and a retention period is applied to the clean room snaps. When this process is complete, another resync can occur immediately afterward if desired. The advantage of using this process over using a defined delta resync interval for the Remote Copy group is that all clean room data on the vault array is always checked for infection and is ready for use if necessary before the next update to the vault array occurs.

Note
As a best practice, Hewlett Packard Enterprise recommends that updates to the vault array be controlled through the `syncrcopy` CLI command executed on a controlled schedule that includes inspecting the data for intrusion between resyncs rather than allowing Remote Copy to replicate changes based on a standard delta resync interval.

Using coordinated snapshots to update the vault array

It is possible to create coordinated snapshots between the production array and the vault array. Coordinated snapshots are snapshots taken on the Remote Copy primary and secondary arrays that are created at the exact same point in time. These coordinated snapshots, created on the two arrays, can then be used as the clean room snapshots. (More frequent clean room snapshots can be created on the production array if desired). This process is shown in Figure 11.

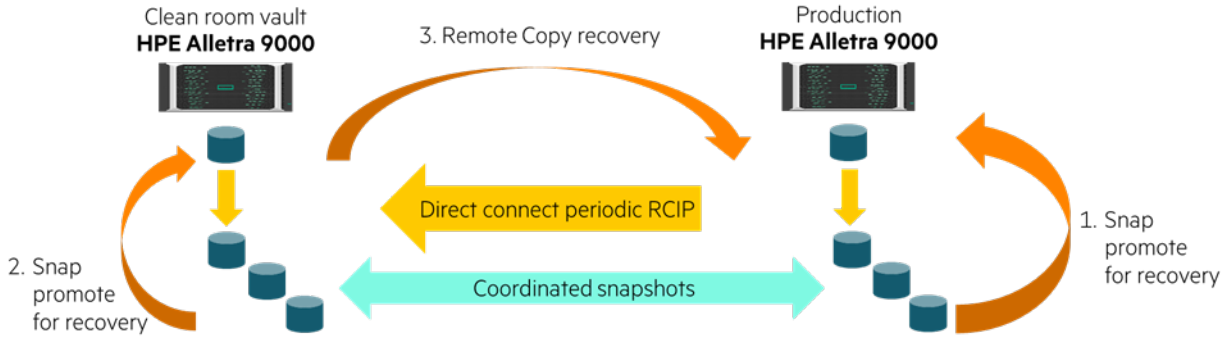


Figure 11. Coordinated snapshots to clean room vault array

If the delta resync interval on the Remote Copy group is set to zero, requesting a coordinated snapshot for the Remote Copy group results in a delta resync of the vault array before the coordinate snapshot is taken there. After the vault array is updated through the Remote Copy resync and the coordinated snapshots have been taken, the data on the vault array can be inspected for infection.

Note
A best practice is to set a delta resync interval of zero on the Remote Copy group and perform controlled resyncs of the Remote Copy group. After the resync is complete, immutable snapshots on the vault array should be checked for integrity. If the volumes are found to be clean, a retention time is then applied to the immutable snapshots.

With this solution, if a ransomware attack occurs, the process to recover data involves the following steps:

1. The environment must be sanitized to ensure that all ransomware infection has been purged from the enterprise before restoring the data to a clean point in time.
2. Data recovery is first attempted by using snaps from the snapshot clean room on the production array. If good data exists in the clean room on the production array, those snapshots are promoted to provide recovery.
3. If data recovery cannot be achieved by using the snaps in the production array clean room, a Remote Copy failover to the vault array is executed. (Production is not run on the vault array itself.) After the Remote Copy failover, successive clean room snapshots on the vault array clean room are then checked for infection (if this has not been done prior to the attack). When a clean copy of data is identified in the vault array clean room, those snapshots are promoted to the volumes on the vault array, and then a resync back to the production array occurs to restore the production array data back to this clean point-in-time copy of the data from the vault array.
4. After the vault array snapshots are promoted, the data on the vault array is resynced back to the production array. After the resync from the vault array to the production array has completed, a failover back to the production array is executed to bring production back online.

Note
Production should never be run on the clean room vault array.

Data recovery architecture #4

Hewlett Packard Enterprise recommends using data recovery architecture #4, which it considers the best practice architecture for maximum protection against ransomware. This architecture can be deployed for Zerto environments. For more information, see the Ransomware data recovery architectures for Zerto section.

This data recovery architecture moves the backup clean room off the production array to the clean room vault array discussed in data recovery architecture #2. This architecture makes the backup clean room more secure than having it on the production array and allows the data on the vault array to be inspected closely before it is used for backup.

Note
Hewlett Packard Enterprise recommends using this data recovery architecture for maximum data protection against a ransomware attack.

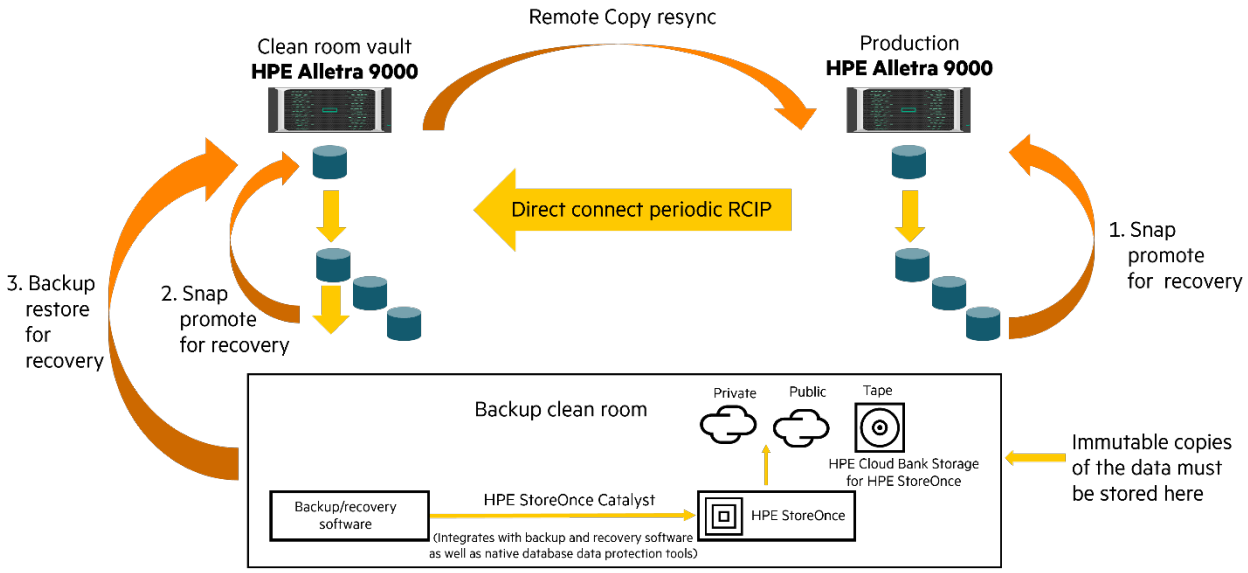


Figure 12. Data recovery architecture #4—Backup clean room off the vault array

With this solution, if a ransomware attack occurs, the process to recover data involves the following steps:

1. The environment must be sanitized to ensure that all ransomware infection has been purged from the enterprise before restoring the data to a clean point in time.
2. Recovery is first attempted by using snaps from the snapshot clean room on the production array. If good data exists in the clean room on the production array, those snapshots are promoted to provide recovery.
3. If recovery cannot be achieved by using the snaps in the production array clean room, a Remote Copy failover to the vault array is executed. (Production workloads should not be run on the vault array itself.) After the Remote Copy failover, successive clean room snapshots on the vault array clean room are checked for infection (if that has not already been done). If a clean copy of data is identified in the vault array clean room, those snapshots are promoted to the volumes on the vault array with a resync back to the production array followed by a failback to the production array.
4. If recovery cannot be achieved by using the vault array clean room snapshots, then successive backups are restored on the vault array and checked until a clean copy of the data is found. After a clean copy of data is identified from the backup clean room, a resync from the vault array to the production array occurs. After the resync from the vault array to the production array has completed, a failover back to the production array is executed to bring production back online. The necessity to restore data from the backup clean room can have a significant effect on the RPO of the solution.

Data recovery architecture #5

Data recovery architecture #5 extends data recovery architecture #4 to include an HA/DR option. With this architecture, the production workload is protected from disaster by two production arrays set up in a Peer Persistence configuration or a synchronous long-distance configuration. The data being replicated to the vault array is in what is known as a three-data center Peer Persistence (3DC-PP) solution or in a synchronous long distance (SLD) solution. Active Peer Persistence is not supported in a 3DC-Peer Persistence configuration at this time, so it cannot be deployed in this architecture if Peer Persistence is used. Unlike architecture #4, coordinated snapshots cannot be created between the production Peer Persistence arrays or synchronous SLD arrays and the clean room vault array.

Note

Active Peer Persistence is not currently supported in this architecture. Check with your HPE representative about support for Active Peer Persistence in a 3DC-PP configuration.

Note
The use of coordinated snapshots between the production arrays and the vault array is not currently supported in this architecture. Check with your HPE representative about support of coordinated snapshots to the asynchronous target array in a 3DC-PP or SLD configuration.

Note
Hewlett Packard Enterprise recommends using coordinated snapshots to create snapshots in the clean room of the two production sync arrays in the Peer Persistence (or SLD) configuration. If coordinate snapshots are used and created, it does not matter if an RC group has been failed over from one of the arrays to the other. Regardless of where the RC group is primary, the snapshots on that array, if determined to be clean, can be used to restore to good point-in-time data if necessary.

Hewlett Packard Enterprise recommends that periodic asynchronous Remote Copy not be configured for regular delta updates of the vault array. Instead, updates to the vault array should be controlled by setting the Remote Copy group delta resync interval to zero and updating the vault array by using the `syncrcopy` CLI command on a planned schedule for the group. After the Remote Copy resync to the vault array has completed, the volumes on the vault array should be checked for infection. If no infection is found, a set of immutable snaps with retention times is then created on the immutable snapshots. After this process is complete, another resync to the vault array can occur immediately afterward, if desired. The advantage of this approach is that all clean room data on the vault array is always checked for infection and is ready for use if necessary, resulting in a better RTO.

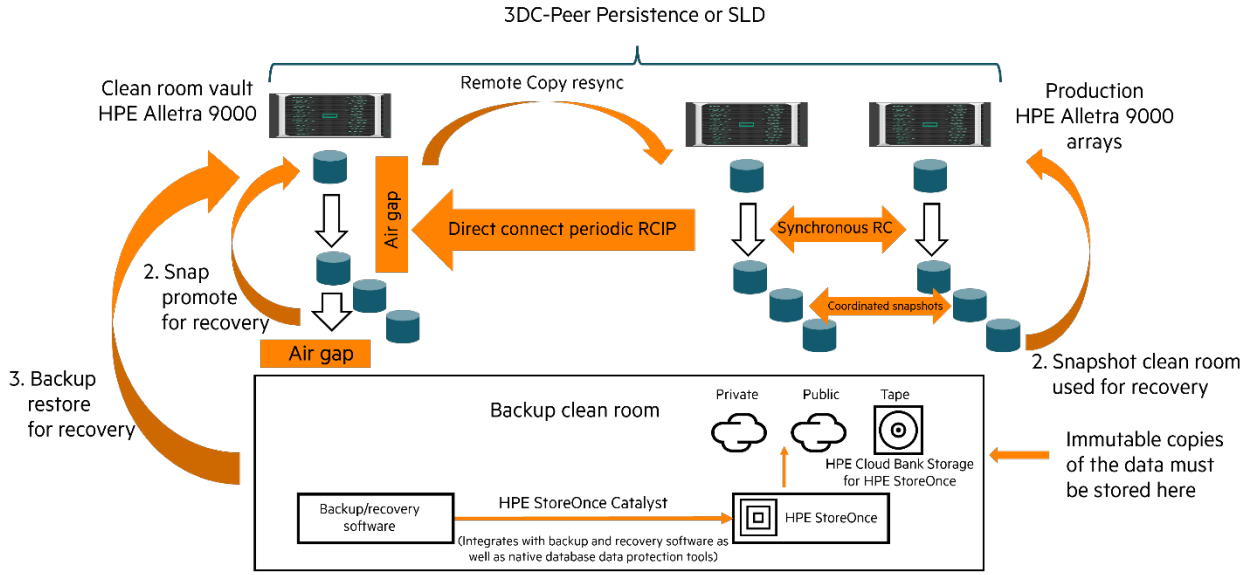


Figure 13. Data recovery architecture #5—Disaster recovery solution

If a ransomware attack occurs, snapshots in the clean rooms on the production arrays are checked for infection. If a clean copy of the data is present, then that copy is promoted to the production volumes. If no clean copy of the data exists in the clean rooms on the production arrays, a Remote Copy failover to the vault array must occur. Following the failover, clean room snapshots in the clean room on the vault array are checked for infection (if that has not already been done). If a clean copy of data is identified in the clean room on the vault array, those snapshots are promoted to the Remote Copy volumes on the vault array, and then the volumes must be resynced to the production 3DC-PP or SLD arrays. After the data is resynced, a failover back to the production Peer Persistence arrays occurs to bring production back online. If no clean copy of the data is found in the vault array clean room, successive backups must be restored to the vault array and checked until a good copy of the data is found. After a good backup copy is restored to the vault array, a Remote Copy resync back to the production arrays must occur, followed by a failover back to the production arrays to bring production back online.

With this solution, if a ransomware attack occurs, the process to recover data involves the following steps:

1. The environment must be sanitized to ensure all ransomware infection has been purged from the enterprise before restoring the data to a clean point in time.
2. Recovery is first attempted by using snaps from the snapshot clean room on the production array that is primary for the infected VVs. If good data exists in the clean room on the production array, those snapshots are promoted to provide recovery.

3. If recovery cannot be achieved by using the snaps in the production array clean room, a Remote Copy failover to the vault array is executed. (Production workloads should not be run using volumes on the vault array itself). After the Remote Copy failover, successive clean room snapshots on the vault array clean room are checked for infection. If a clean copy of data is identified in the vault array clean room, those snapshots are promoted to the volumes on the vault array with a resync back to the production array, followed by a failback to the production array
4. If recovery cannot be achieved by using the vault array snapshot clean room data, then successive backups are restored on the vault array and checked until a clean copy of the data is found. After a clean copy of data is identified from the backup clean room, a resync from the vault array to the production array occurs. After the resync from the vault array to the production array has completed, a failover back to the production array is executed to get production back online. The necessity to restore data from the backup clean room can have a significant effect on the RPO of the solution.

Ransomware data recovery architectures for Zerto

Using Zerto software to protect against a ransomware attack provides the potential for data recovery with a very small RPO and very fast RTO through the Zerto continuous data protection (CDP) capability. Zerto CDP enables recovery in RPO increments as small as five seconds, and Zerto-based solutions include recovery that spans a variety of ransomware infection scenarios, including but not limited to the following:

- **File infection:** If the ransomware infection is limited to files and folders on a VM, these can be restored back to their source location from a timestamp only 5–10 seconds before the infection occurred.
- **VM infection:** Similar to file infection, if one or more VMs are affected by ransomware, Zerto can restore those VMs back to production with no intermediate steps (for example, VMware vSphere® Storage vMotion®). Zerto can group VMs per application in virtual protection groups (VPGs); however, you can also test, restore, or perform any other operation on individual VMs within that VPG. If a multi-VM application is infected at different points in time, you can still recover individual VMs to different points in time.
- **Full workload contamination:** If the entire production environment has been infected, a live failover to the secondary site (Zerto DR) can bring operations back up and running in minutes. Recovery from the Zerto journal is preferable because of the granularity of recovery that is available. If recovery cannot be handled through the Zerto journal, then recovery through a snapshot clean room in the environment would be attempted. If that is not available, then recovery from a Zerto long-term repository (LTR) or a backup clean room would occur.

Although experience has shown that many ransomware attacks are detected within a relatively short time, attacks often continue for an extended length of time before being discovered (for weeks to months). When using Zerto CDP for ransomware recovery, it is important to remember that recovery can occur back in time for only the length of time the CDP journal has been configured to protect (from 1 hour up to 30 days). For example, if the Zerto CDP journal is configured to hold a maximum of seven days' worth of data, it is useful for recovering only from an attack that has occurred in the last seven days—not from one that occurred more than seven days in the past. For an attack that lasts more than seven days, other measures must be in place to enable recovery. Integrating the previously described data recovery architectures with Zerto provides the ability to recover from an attack that has been ongoing for longer than the period of time the Zerto journal was configured to cover.

Note
If you use the Zerto journal to recover from a ransomware attack, it is important to remember that you are protected only for the length of time for which the journal is configured to hold data.

Note
Zerto CDP can be configured for a maximum of 30 days' worth of data recovery.

Zerto provides recovery for the data used by the VMs in the environment. Because Zerto runs as a VM in the production environment with the production applications, it is potentially subject to any ransomware attack that occurs against the environment. The following architectures provide protection from an attack that has affected any Zerto VMs, enabling them to be restored so they can then be used to restore data for affected applications. In these architectures, the Zerto environment and the application environment combined are referred to as the "ransomware attack domain."

Because the Zerto Virtual Manager (ZVM) and Virtual Replication Appliance (VRA) are integrated as part of the production environment, it is not possible to populate a clean room directly by using Zerto alone. HPE solves this problem by integrating capabilities from the previously discussed data recovery architectures combined with Zerto to provide a clean room environment that can be used for data recovery and that can also be used to inspect data for hidden ransomware infection. The overall solution provides the ability to recover from

a short duration attack through Zerto CDP while also providing the ability to recover if the attack runs longer than the length of the Zerto journal or if the attack affects the Zerto VMs. Another advantage of these architectures is that they leverage Zerto DR and can be used as a standard DR failover solution through Zerto DR (providing failover in the event of a disaster such as a power failure or a datacenter failure) in addition to providing protection from a ransomware attack.

Zerto architecture #1

Hewlett Packard Enterprise does not recommend using the first Zerto architecture solution because it can support only a limited amount of clean room data. However, for a customer with a limited budget, deploying this architecture is better than doing nothing at all. Zerto can be used to protect virtualized servers while all other data is protected by a simple snapshot clean room. Note that while Zerto is protecting VM application data, the Zerto environment itself is protected by the snapshot clean room in the event that the Zerto data itself is attacked. This architecture can be integrated into the five recovery architectures described previously for environments that have a combination of virtualized and nonvirtualized servers with Zerto protecting the virtualized servers.

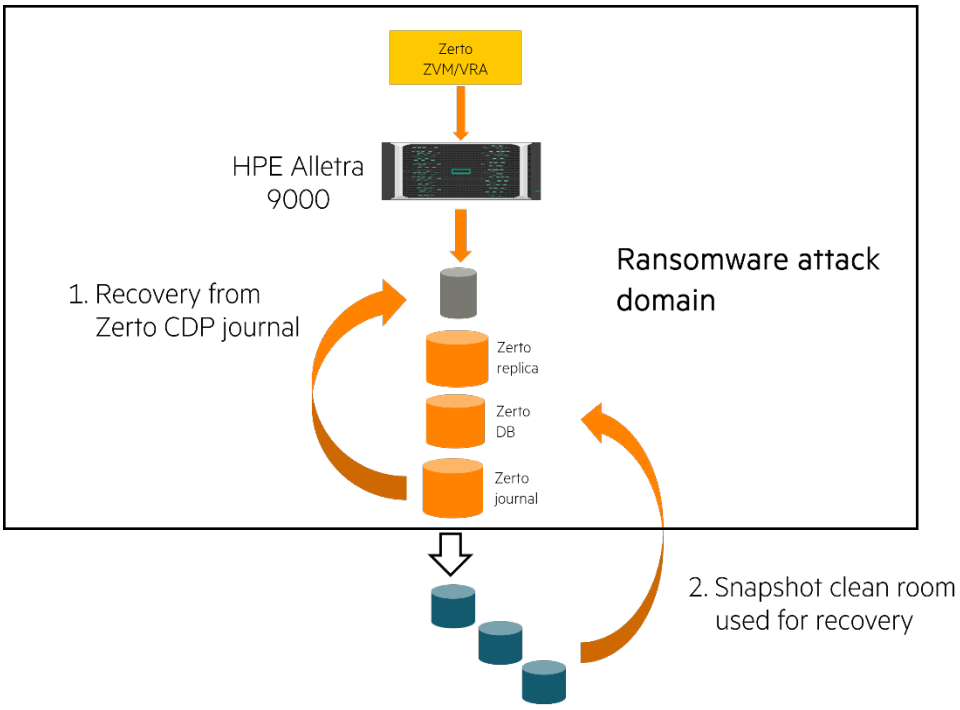


Figure 14. Zerto architecture #1

With this solution, if a ransomware attack occurs, the process to recover data involves the following steps:

1. Recovery is first attempted by using the Zerto CDP journal.
2. If recovery cannot be achieved by using the Zerto CDP journal, then clean room snapshots are promoted to provide a usable Zerto CDP journal, and then recovery occurs from the CDP journal.
3. Recovery for nonvirtualized servers would occur in the same manner as outlined in data recovery architecture #1.

Note

Hewlett Packard Enterprise does not recommend using the architecture shown in Figure 14.

Zerto architecture #2

The second Zerto architecture extends Zerto architecture #1 to enable the use of the Zerto remote DR capability. In this architecture, data is replicated from site A to site B through Zerto DR. Any storage can be used in site A, but the storage at the Zerto DR site must be HPE Alletra 9000 to provide a snapshot clean room for the Zerto data.

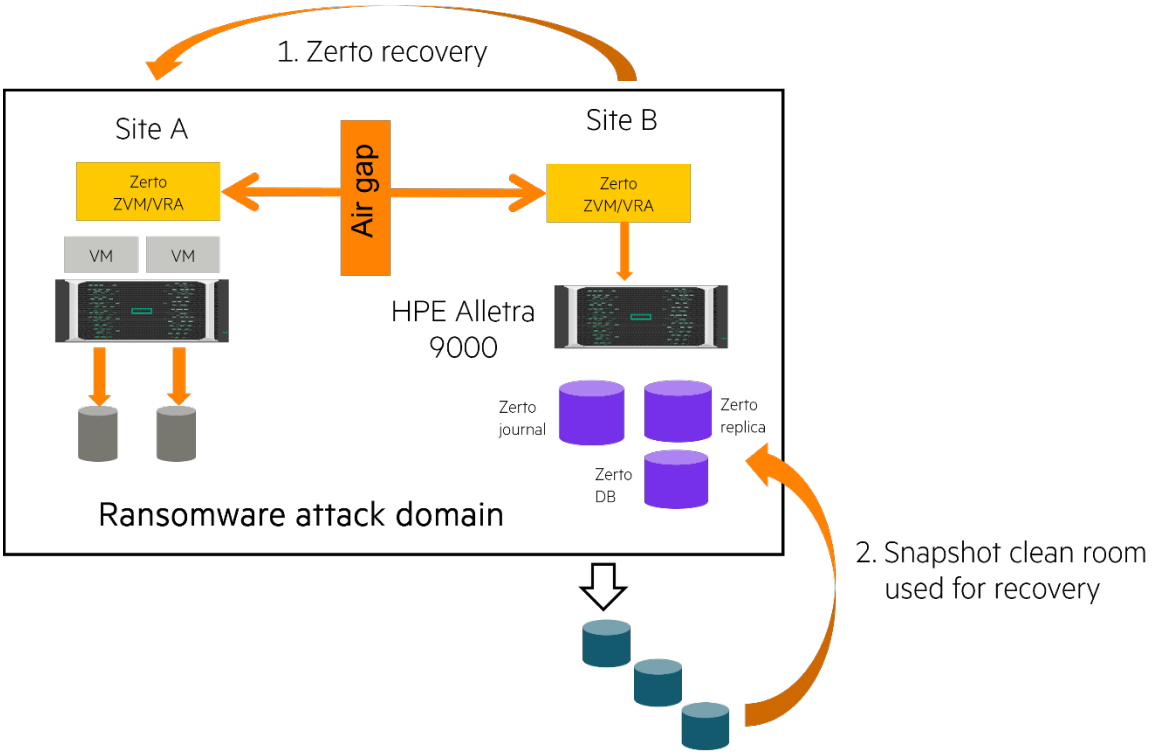


Figure 15. Zerto architecture #2

With this solution, if a ransomware attack occurs, the process to recover data involves the following steps:

1. Recovery is first attempted by using the Zerto journal at the Zerto DR site (site B). If good data exists in the CDP journal, it is used to restore the production volumes. Production can be resumed at site B while Zerto recovers good data back to the production environment in site A.
2. If recovery cannot be achieved by using the Zerto CDP journal at the DR site (site B), then HPE Alletra 9000 array clean room snapshots are used. Good data from the clean room snapshots is promoted, and then recovery is performed through the Zerto journal in site B. Production can be resumed at site B while Zerto recovers good data back to the production environment in site A. After site A is recovered, a planned failover event back to site A can be conducted.

Zerto architecture #3

Zerto architecture #3 adds both a Zerto clean room that is used to check data for integrity and a backup clean room that is used for long-term storage of the data. Data can be saved long-term to the backup clean room either through a traditional backup solution that uses the snapshot clean room data to source the backup or through the Zerto LTR to Scality S3 immutable storage.

The Zerto LTR allows daily checkpoints to be pulled from the Zerto journal and stored long term to use for recovery from an attack that has been occurring undetected for longer than the length of time for which the journal was configured. Hewlett Packard Enterprise recommends that the LTR be stored immutably on Scality by using the Scality S3 capability. After creation, an LTR checkpoint is recovered to the Zerto clean room and used to check data integrity. These checks of the data provide known good LTR checkpoints that can be used for recovery in the event of a full-site contamination. The LTR copies are also portable, and they can be reattached to a secondary site running Zerto even if it was not the original replication target. (The Zerto clean room in this architecture is an example of this situation.) Using clean or rebuilt infrastructure, you can reattach the LTR and restore from there, which is particularly helpful because these are restores of entire applications from a single consistent point in time rather than on a VM-by-VM basis.

The Zerto LTR must be saved to an immutable store to protect it from deletion or attack. Zerto can save an LTR to any S3-compatible interface. If desired, you can have Zerto can store the LTR directly to an S3 public cloud; however, Hewlett Packard Enterprise recommends storing the LTR on Scality through Scality's S3 interface rather than directly to an S3 public cloud. The reason for this recommendation is simple: After being saved, the LTR must then be restored to the Zerto clean room environment so that the data can be checked for integrity. Storing the LTR on Scality avoids the data egress charges that would be incurred if the LTR were stored in an S3 public cloud. Note that if Zerto LTR is used rather than a traditional backup of the Zerto environment, the Zerto configuration in the Zerto clean room can be a

minimal Zerto configuration if desired.[2] If a traditional backup solution is used, the backup must be restored to an environment that matches the production configuration at site B of the Zerto solution.

Whether you choose to back up the environment to HPE StoreOnce or to store Zerto LTRs to Scality, the site B array clean room snapshots can be used to check data integrity through the Zerto clean room. The array clean room snapshots can be used for data integrity checking by mounting a copy of them onto the configuration in the Zerto clean room. Note that if the array snapshot clean room data is used for checking data integrity, the Zerto configuration in the clean room must match the Zerto production environment in site B because the ZVM database contains pointers to managed VMware vCenter® resources such as hosts and vNets.

Note
Hewlett Packard Enterprise recommends saving a daily LTR checkpoint of the Zerto journal.

Note
Hewlett Packard Enterprise recommends saving the Zerto LTR in this architecture to a Scality appliance. The LTR can be restored from Scality to the Zerto clean room and used for data integrity checking without having to incur public cloud egress charges to populate the clean room.

Note
If the LTR is used for data integrity checking, the Zerto clean room configuration does not have to match the target Zerto configuration in site B; it can be a minimal Zerto configuration if desired. However, if the snapshot clean room data is to be used for data integrity checking, then the Zerto configuration in the clean room must match the Zerto configuration found in site B.

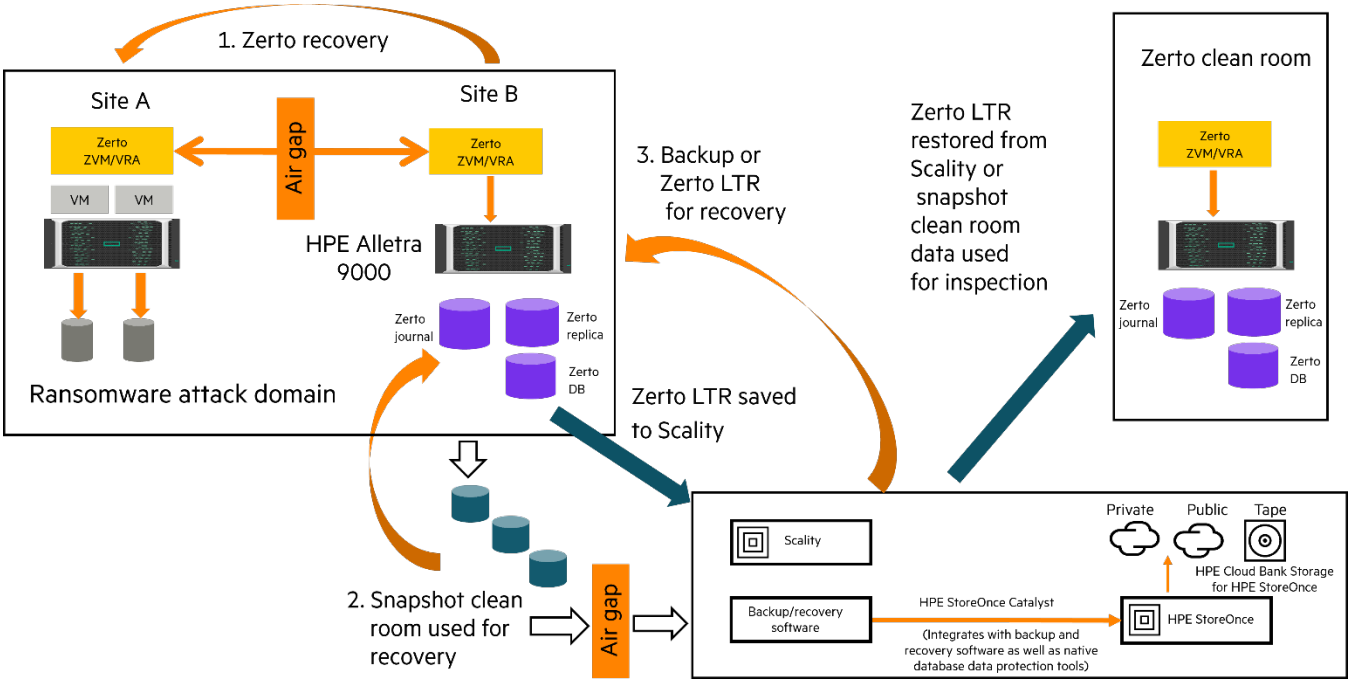


Figure 16. Zerto architecture #3, Zerto DR combined with immutable LTR on Scality or backup cleanroom on HPE StoreOnce

Using this solution to recover from a ransomware attack depends on whether data integrity was verified by using the array clean room snapshots or the Zerto LTR store:

If array clean room snapshot data was used to check data integrity, recovery involves the following steps:

[2] Contact your HPE representative for information about proper configuration of the Zerto clean room environment.

1. Recovery is first attempted by using the Zerto journal in site B. If good data exists in the journal, it is used to restore the production volumes.
2. If recovery cannot be achieved by using the Zerto journal, a set of snapshots from the array snapshot clean room, known to contain verified data, is promoted to the Zerto production volumes in site B to achieve recovery.
3. If recovery cannot be achieved by using the production array clean room snapshots, it is achieved by using the backup clean room data. A good, certified backup is restored to the Zerto production volumes in site B to achieve recovery.

If the Zerto LTR was used to check data integrity, recovery involves the following steps:

1. Recovery is first attempted by using the Zerto journal in site B. If good data exists in the journal, it is used to restore the production volumes.
2. If recovery cannot be achieved by using the Zerto journal, a set of snapshots from the array snapshot clean room, known to contain verified data, is promoted to the Zerto production volumes in site B to achieve recovery.
3. If recovery cannot be achieved by using the production array clean room snapshots, it is achieved through the Zerto LTR. A clean Zerto LTR store is used to recover the Zerto environment. The Zerto LTR recovery can be run either to site B or to any site with a Zerto configuration.

Note
Hewlett Packard Enterprise recommends using the array snapshot clean room to source the backup clean room.

Zerto architecture #4

This final Zerto architecture combines a Zerto DR solution and a replication clean room vault array sourcing a backup clean room. It supports a Zerto clean room for checking data integrity but does not show it.

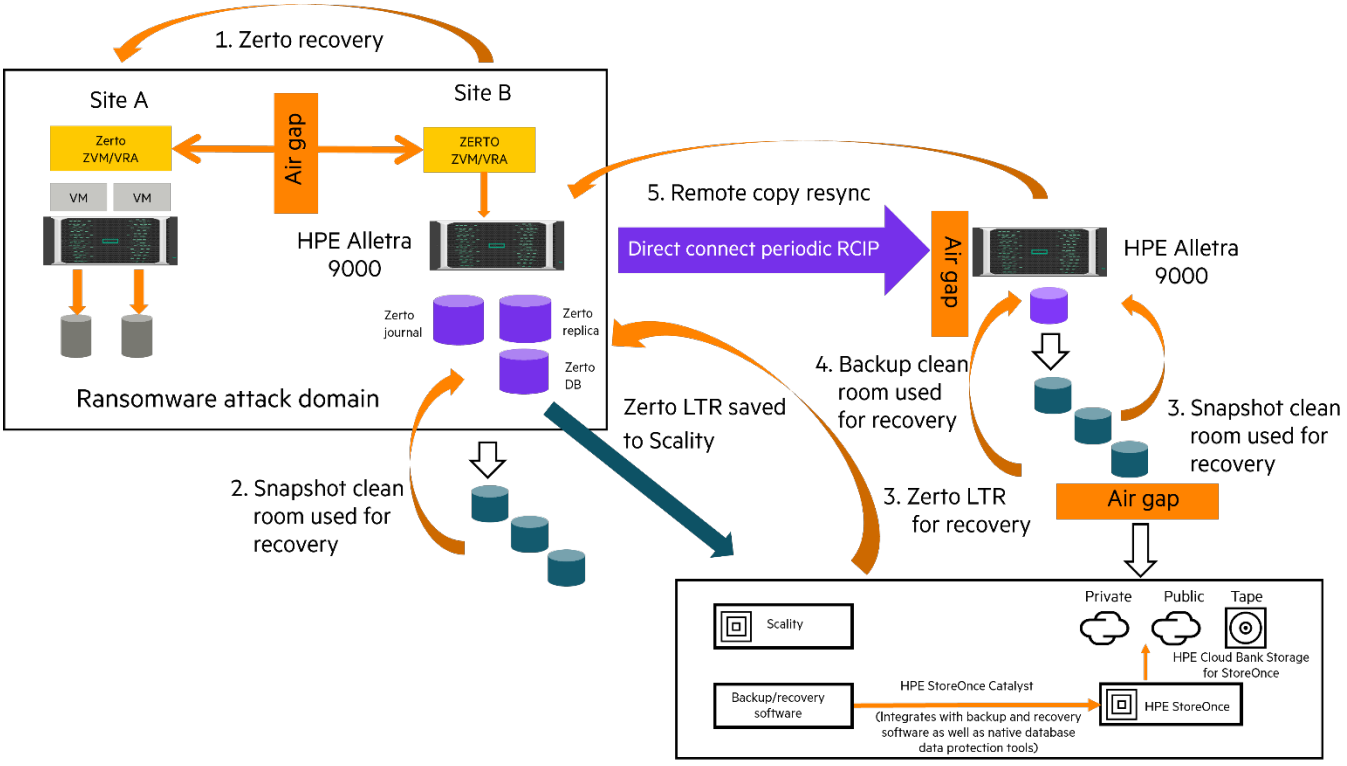


Figure 17. Zerto Architecture #4, Zero DR integrated with a vault array clean room and a backup cleanroom or LTR store (Zerto clean room not shown)

Similar to Zerto architecture #3, this architecture can support a Zerto clean room (not show in Figure 17) on the clean room vault array with a minimal Zerto deployment where the LTR can be restored and the data checked for infection. Alternatively, the Zerto clean room can contain a Zerto configuration matching the Zerto configuration at site B, and the vault array snapshot clean room data can be used to verify the data before being backed up.

Using this solution to recover from a ransomware attack depends on whether data integrity was verified by using the array clean room snapshots or the Zerto LTR:

If the vault array clean room data was used to check data integrity, recovery involves the following steps:

1. Recovery is first attempted by using the Zerto journal at site B. If good data exists in the journal, it is used to restore the production volumes.
2. If recovery cannot be achieved by using the site B journal, it is achieved by performing a failover from the site B HPE Alletra 9000 to the clean room vault HPE Alletra 9000. Good snapshot data in the vault array snapshot clean room is then promoted to the vault array base volumes. Next, a Remote Copy delta resync back to the site B HPE Alletra 9000 array is performed, followed by a failover to the site B HPE Alletra 9000 array.
3. If recovery cannot be achieved by using the vault array clean room snapshots, it is achieved by restoring backup clean room data on the vault array. A delta resync back to the site B HPE Alletra 9000 array is then performed, followed by a failover to the site B HPE Alletra 9000 array.

If the Zerto LTR was used to check data integrity, recovery involves the following steps:

1. Recovery is first attempted by using the Zerto journal at site B. If good data exists in the journal, it is used to restore the production volumes.
2. If recovery cannot be achieved by using the Zerto journal in site B, it is achieved through the Zerto LTR. A clean Zerto LTR store is used to recover the environment. The LTR is populated with daily checkpoints by Zerto on site B. Hewlett Packard Enterprise recommends using an HPE StoreOnce appliance as the local repository. The Zerto LTR recovery can be run either to site B or to any site with a bare-bones Zerto configuration.

Combining architectures

Many environments contain combined bare-metal server operations and virtualized operations. The architectures discussed in this paper can be combined to support these environments. For example, data recovery architecture #4 and Zerto architecture #4 can be integrated to share the HPE Alletra 9000 array in Zerto site B so that immutable snapshots can be created on that array and data can be replicated to a vault array where a backup clean room could be sourced from the vault array snapshots.

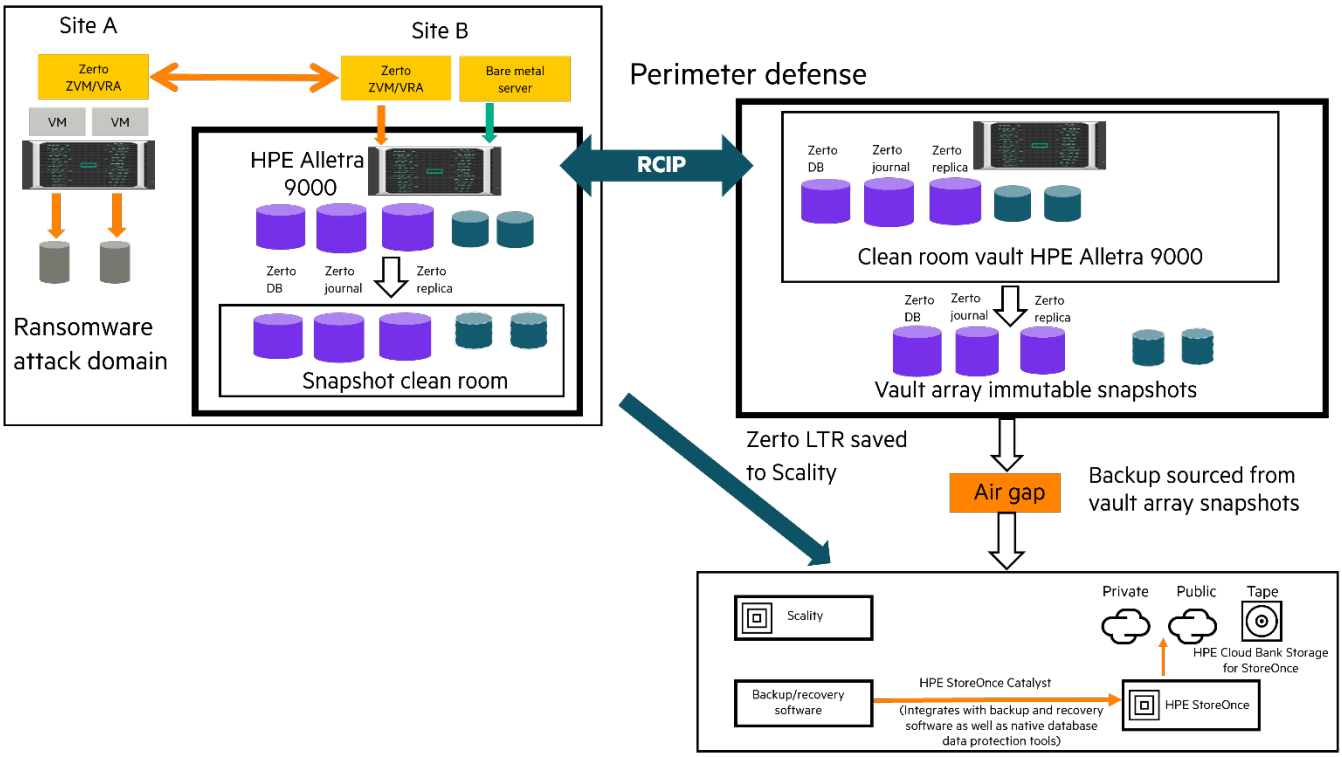


Figure 18. Data recovery architecture #4 integrated with Zerto architecture #4: The dark blue volumes represent the non-Zerto data being protected

Conclusion

Having a solid data protection scheme in place for recovering from a ransomware attack on the enterprise has become a requirement in today's world that is just as important as—if not more important than—a DR strategy for recovering from a physical disaster in the enterprise. As this paper has discussed, not only are ransomware attacks insidious, but the number of attacks on enterprises continues to grow. The attackers are constantly adapting and evolving their methods of attack, drilling deeper into the enterprise to corrupt data in an attempt to make it impossible to recover successfully without paying a ransom. Early detection of an attack enables recovery with minimal data loss, but protecting from an undetected attack that has been ongoing for an extended period of time is paramount. Having a solid, proven data protection solution architecture in place that provides multiple levels of protection against attacks on your data is insurance that a clean, usable copy of data will be available if a ransomware attack occurs. Hewlett Packard Enterprise has the products, tools, processes, software, and consulting resources to help you architect and design a ransomware data protection strategy that will meet your data protection requirements at a cost you can afford.

Resources

Ransomware: Ensuring protection from an increasingly complex threat

Protecting Data from Ransomware with HPE StoreOnce Catalyst

HPE StoreOnce Systems QuickSpecs

hpe.com

Zerto product documentation

Preventing Ransomware Isn't Always Possible—But Mitigating the Threat Is

HPE SPOCK

Learn more at

hpe.com/storage

a00119614ENW, Rev. 2

Hewlett Packard Enterprise