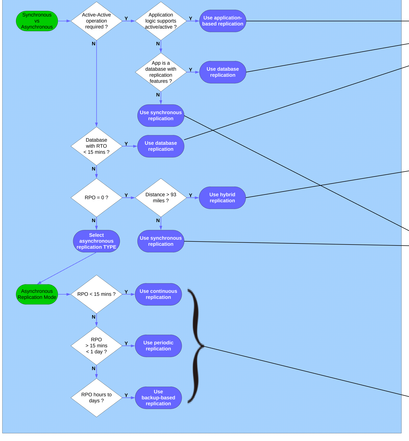
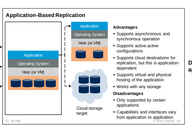
Virginia Information Technologies Agency

VITA

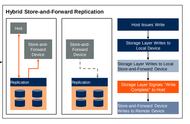
Enterprise Architecture
Data Replication Decision Point Process

Selecting the Right Replication Mode



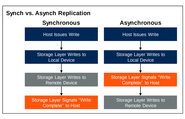


Database platforms are the most common implementation of application-based replication.



Synchronous replication should only be used for applications that must have zero data loss because of the impact on write performance.

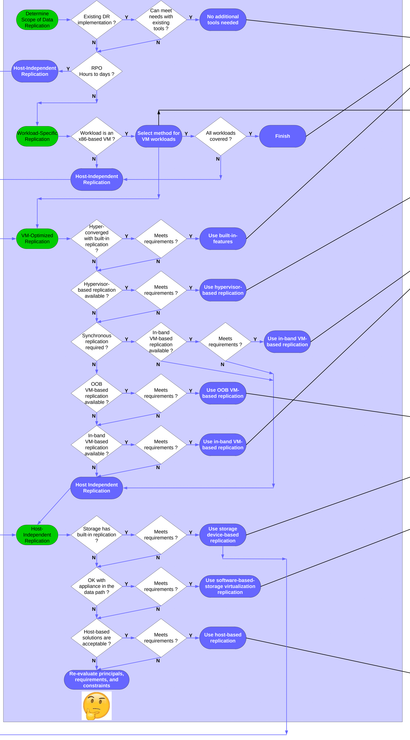
Synchronous replication: Input/output (I/O) transactions are replicated in real time to the remote site, and the I/O transaction isn't complete until it's stored on both locations. Synchronous replication is used for almost all mission-critical applications where zero data loss is required. Due to transmission latency, truly synchronous replication is limited to remote sites within approximately 150 km (~93 miles) of the primary site.



If none of the requirements for synchronous replication are met, then appropriate asynchronous replication should be chosen.

Asynchronous replication: Asynchronous replication encompasses several technologies including snapshots and CDP. With an asynchronous approach, I/O transactions are written immediately at the primary site, with delayed replication to the remote site occurring independently. Making the replication independent of the initial write removes the distance restrictions seen with synchronous approaches. Asynchronous replication is used for applications that need to be recovered in minutes or hours. WAN bandwidth and latency requirements: There must be enough bandwidth available to handle the volume of data created by the replication. Latency is also important because it impacts RPO for asynchronous replication.

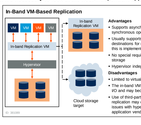
Selecting Replication Implementation

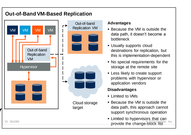


Virtualization-optimized replication mechanisms are increasingly popular because of their superiority when replicating virtual machines (VMs) and their support for cloudbased DR and DR as a service (DRaaS) providers.
VM-optimized replication: Many solutions are specifically designed for VMs. The implementation can be in the hypervisor, part of a hyperconverged infrastructure platform or a separate VM. Examples include Microsoft Hyper-V Replica, VMware vSphere Replication, Zerto IT Resilience Platform and Nutanix hyperconverged infrastructure.

If the hypervisor used to support the VMs offers replication as an option, then it makes sense to use this rather than VM-based methods because of superior integration with the hypervisor's provisioning and orchestration features. If the hypervisor-based replication doesn't meet the requirements, then continue the selection process.

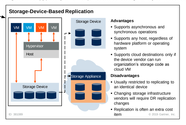
In-band VM-based replication uses a VM to manage all the storage assigned to the host and provides storage services to the rest of the VMs. In-band VM-based replication should be the last choice for the VM-optimized methods because of its potential performance impact on writes to storage. If the inband VM-based replication doesn't meet the requirements, then it's necessary to look at the host-independent approaches to replication typically used for non-VM workloads.





Out-of-band (OOB) VM-based replication relies on the hypervisor providing a change block list to a specialized VM that replicates to the remote site. If OOB VM-based replication meets the requirements, then it should be chosen ahead of in-band VM-based replication because it doesn't impact I/O performance significantly for VMs running on the host. In this approach, the VM responsible for replication is not in the I/O path between the VMs and physical storage. Instead, the VM relies on getting a list of changed blocks from the hypervisor (often referred to as change block tracking). The changed blocks are then replicated asynchronously. The diagram outlines the basic operation, along with some key advantages and disadvantages.

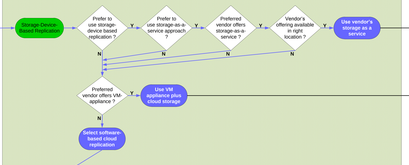
Storage-device-based data replication: This approach relies on features built into the storage array or NAS. Examples include Dell EMC Symmetrix Remote Data Facility (SRDF) and the NetApp SnapMirror functions in OnTap-based NAS devices. Storage-device-based replication formed the basis for most DR replication for many years and is still widely used in many organizations. In this approach, the replication functions are built in to the storage device (typically an enterprise-class storage array or NAS device). Storage-device-based replication requires centralized storage with built-in replication features.



Use software-based replication if there is no suitable device-based approach or the organization prefers pure software-based approaches, then the only option is to use either the native service offered by the cloud provider or third-party replication software. Software-based approaches provide flexibility in the choice of secondary storage infrastructure, failover location and public cloud support. In addition, changes to hardware infrastructure have less impact on existing DR plans when software is used for replication. Given the advances being made in integration of software-based replication technologies with DR orchestration, Gartner recommends that organizations be more open to replacing some of their existing approaches to replication. This is particularly true in virtualized environments and for situations where replication to the cloud is important. Storage-device-based and software-based storage virtualization replication solutions can offload all the work of replicating data to a storage array controller or storage virtualization appliance. Given the advances being made in integration of software-based replication technologies with DR orchestration, Gartner recommends that organizations be more open to replacing some of their existing approaches to replication. This is particularly true for virtualized environments and for situations where replication to the cloud is important. Storage-device-based and software-based storage virtualization replication solutions can offload all the work of replicating data to a storage array controller or storage virtualization appliance. Storage-device-based and software-based storage virtualization replication solutions require some additional host CPU, memory and I/O processing cycles to operate. For example, any in-band or out-of-band VM-based replication uses up resources on the VM host. Host-independent: In this class of solution, replication is completely independent from the hosts running the applications, so it can support a wide variety of hardware platforms, operating systems and applications. Two implementation approaches fall into this category: 1) Software-based storage virtualization replication; 2) Storage-device-based replication

Host-based replication requires replication capabilities in the host operating system. Implemented in the operating system it is a general-purpose approach that can support any application. Host-based replication leverages software running on each server or VM to copyreplicate data. It can be built in to an application or implemented in a file system or volume manager. Examples include Carbonite (Double-Take Software), Oracle Solaris ZFS file system, Veritas Volume Replicator, and database-specific solutions like Microsoft SQL Server Always On. Host-based replication relies on features of the application, operating system, or third-party volume managers and file systems to replicate data. Application-based replication is primarily used for active-active configuration or where the RTO is zero. Operating system or third-party volume managers should only be considered as a last resort. Host-based solutions are the least favored option because of the need to manage replication separately for each host and the potential need for different solutions for different operating systems and applications.

Replication for Cloud-based Applications



If the organization prefers an as-a-service approach and the vendor offers its service in right physical locations, then the organization can select this option.

If the organization doesn't like the as-a-service model or the vendor doesn't offer it in the right locations (or doesn't offer it all), then the VM-appliance plus cloud storage option is worth exploring. If the vendor supports it, then this approach will maintain the centralized management of the organization's existing replication without being tied to specific cloud locations.

Host-based replication requires replication capabilities in the host operating system. Host-based replication implemented in the operating system is a general-purpose approach that can support any application. Replication implementation approaches fall into three categories: 1) Host-based; 2) VM-optimized; 3) Host-independent. Host-based solutions run on the same physical hardware as the application and fall into two categories: 1) Application-based replication; 2) Operating-system-based replication. Host-based replication relies on features of the application, operating system, or third-party volume managers and file systems to replicate data. Application-based replication is primarily used for active-active configuration or where the RTO is zero. Operating system or third-party volume managers should only be considered as a last resort.

Host-based replication requires replication capabilities in the host operating system. Host-based replication implemented in the operating system is a general-purpose approach that can support any application. Replication implementation approaches fall into three categories: 1) Host-based; 2) VM-optimized; 3) Host-independent. Host-based solutions run on the same physical hardware as the application and fall into two categories: 1) Application-based replication; 2) Operating-system-based replication. Host-based replication relies on features of the application, operating system, or third-party volume managers and file systems to replicate data. Operating system or third-party volume managers should only be considered as a last resort.

If the cloud vendor's replication suite meets the requirements, and the cost for additional storage is acceptable, then this should be the preferred option because of the integration with the cloud vendor's provisioning and management interfaces. If the cloud vendor's offerings fall short of the requirements or are cost-prohibitive, then the fallback option is to use third-party VM-optimized solutions. Experienced DR staff demand high salaries, so replication suites that simplify and automate such things as failover and failback are well worth the investment. This is particularly true for smaller organizations that don't have the expertise and can't afford to hire it.

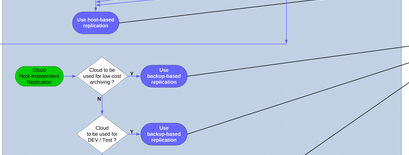
Replication for Container-based Applications

If DR orchestration tooling is already in use, then evaluation should focus on how the new replication solution will integrate with existing tooling.

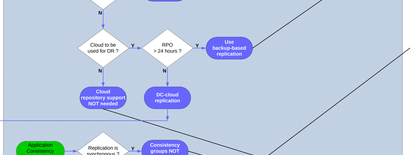
If the organization is adding a new application or replication capability to an existing DR implementation and is willing to use in-house expertise in creating the necessary orchestration of failover / failback, then it can continue to use its existing approach.

Given the advances being made in integration of software-based replication technologies with DR orchestration, Gartner recommends organizations be more open to replacing some of their existing approaches to replication. This is particularly true for virtualized environments and for situations where replication to the cloud is important. Blending of Backup and DR Solutions: Vendors in the replication and backup spaces are hungrily eyeing each other's markets and expanding their products to offer complete solutions for backup and DR. Leading vendors including Commvault, Dell EMC, Veeam and Veritas Technologies are adding DR orchestration to augment existing backup image replication. Newer vendors have built scale-out architectures with replication and instant VM boot from physical or virtual appliances that can be kept in sync across locations. Integrated backup appliances have made these solutions easier to deploy and support.DR configurations. Look for a continuation of blending the backup and DR use cases with vendors that are focused on solving both problems with a single integrated suite of products.

Selecting Additional Features



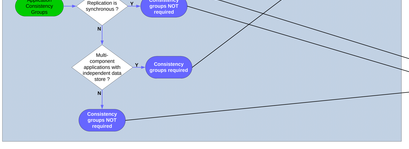
If the organization is using hyperconverged infrastructure that supports the necessary replication modes, then use the native replication services of the hyperconverged infrastructure.



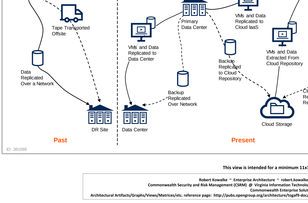
Host-based replication requires replication capabilities in the host operating system. Host-based replication implemented in the operating system is a general-purpose approach that can support any application. Host-based solutions run on the same physical hardware as the application and fall into two categories: 1) Application-based replication; 2) Operating-system-based replication. Replication implementation approaches fall into three categories: 1) Host-based; 2) VM-optimized; 3) Host-independent. Host-based replication relies on features of the application, operating system, or third-party volume managers and file systems to replicate data. Application-based replication is primarily used for active-active configuration or where the RTO is zero. Operating system or third-party volume managers should only be considered as a last resort.

Backup-based replication: This may be a function of the backup software or backup storage target. Backup-based replication solutions are available from a wide variety of vendors including Dell EMC, Commvault, Rubrik, Cohesity, IBM, Actifio and many others. Backup-based replication completely decouples I/O at the primary site and relies on the replication of completed backup jobs to provide data redundancy at the remote site. Backup-based replication is used for applications that don't require synchronous or asynchronous replication.
Backup-based Replication: The final approach is the replication of backup repositories to a remote site. Backup-based replication is limited in terms of the RTO/RPO values it can support by the backup schedule. For example, a 24-hour backup schedule cannot support an RTO or RPO of less than 24 hours. In theory, backups can be run more frequently than a daily schedule, but in general, other replication methods are better fit for RPOs/RTOs of less than 24 hours. The critical advantage of backup-based replication is that technologies such as compression and deduplication, combined with incremental backup that only replicates changes since the last backup, can radically reduce replication bandwidth requirements.
Backup-based replication: Typically, backups are daily task, making them well-suited to meeting a 24-hour RPO. It is possible to meet more demanding RPOs by running backup more frequently; However, this places additional load on systems that must be scanned for backups. If there are applications that use more than one data store, or business processes that include multiple applications with separate data stores, then consistency groups are required for asynchronous or backup-based replication. RTO above 24 hours: Gartner recommends backup-based replication.

Application consistency groups: This capability allows the organization to define groups of storage and applications that need to be kept in sync during replication.
Recovery Granularity Requirements: Some customers have complex applications that require replication consistency groups. Such replication consistency groups ensure that all the data for an application or set of related applications that is spread across multiple VMs, devices, databases, and systems is kept in sync. Were such systems or applications at the target site to be out of sync,
applications could then be prone to corrupting the data. Only some replication solutions (typically, storage-array-based ones) enable such replication consistency groups.
Replicating data to meet RPO and RTO requirements is only part of the problem of creating a workable DR solution, and organizations should also consider making other capabilities as part of any evaluation of replication solutions. The capabilities of concern include Application consistency groups, Applications, VMs, systems, or databases that are replicated are often split across multiple logical unit numbers (LUNs) or other data entities (such as servers, VMs or file systems). In such cases, some data replication solution vendors have enabled consistency groups to avoid data corruption if one such LUN or entity is replicated and up-to-date while another is not. These consistency groups seek to ensure that the data replication of defined group elements is stopped and started in unison. Storage-device-based data replication and storage-virtualization-appliance-based replication solutions offer the most choices for those requiring such multi-LUN, multi-VM, or other multi-entity consistency.
Few other replication products offer consistency group capabilities at present (one example is Zerto's IT Resilience Platform (https://www.zerto.com/) , which offers consistency among VMs, but not physical servers).



Data Replication: Past and Present



VITA Draft Discussion Document // May-31-2019

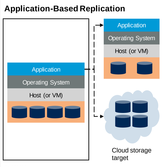
This view is intended for a minimum 11x17 printer paper

Robert Knowlke ~ Enterprise Architecture ~ robert.knowlke@vita.virginia.gov
Commonwealth Security and Risk Management (CSRM) ~ Enterprise Architecture ~ Richard.Mackie@vita.virginia.gov
Commonwealth Enterprise Solutions Center (CESC)

Architectural Artifacts/Graphs/Views/Matrices/etc. reference page: http://pubs.opengroup.org/architecture/togaf9-doc/arch/chap35.html

PURPOSE: To depict Gartner's recommended process for determining Data Replication decision points in a manner more easily digested by the VITA enterprise.

Application-Based Replication

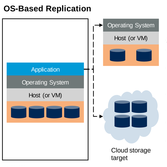


Advantages
- Supports asynchronous and synchronous operation
- Supports active-active configurations
- Supports cloud destinations for replication, but this is application-dependent
- Supports virtual and physical hosting of the application
- Works with any storage

Disadvantages
- Only supported by certain applications
- Capabilities and interfaces vary from application to application

ID: 381089 © 2019 Gartner, Inc.

OS-Based Replication

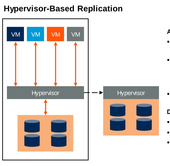


Advantages
- Supports asynchronous and synchronous operations
- Usually supports cloud destinations for replication, but this is implementation-dependent
- Works with local or SAN-attached storage

Disadvantages
- Capabilities and usability vary from one system to the next
- May require third-party products such as volume managers

ID: 381089 © 2019 Gartner, Inc.

Hypervisor-Based Replication

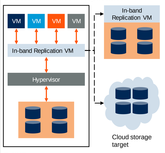


Advantages
- Supports asynchronous and synchronous operations
- Unlikely to create support problems with application vendors that support the underlying hypervisor platform
- Works with any storage managed by the hypervisor

Disadvantages
- Limited to VMs
- No direct-to-cloud replication
- Specific to each hypervisor brand in terms of exact capabilities and user interface

ID: 381089 © 2019 Gartner, Inc.

In-Band VM-Based Replication

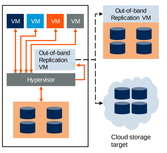


Advantages
- Supports asynchronous and synchronous operations
- Usually supports cloud destinations for replication, but this is implementation-dependent
- No special requirements for the storage
- Hypervisor independent

Disadvantages
- Limited to virtual machines
- The in-band VM has to handle all I/O and may become a bottleneck
- Use of third-party VM-based replication may create support issues with hypervisor and application vendors.

ID: 381089 © 2019 Gartner, Inc.

Out-of-Band VM-Based Replication

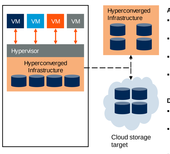


Advantages
- Because the VM is outside the data path, it doesn't become a bottleneck
- Usually supports cloud destinations for replication, but this is implementation-dependent
- No special requirements for the storage at the remote site
- Less likely to create support problems with hypervisor or application vendors

Disadvantages
- Limited to VMs
- Because the VM is outside the data path, this approach cannot support synchronous operation
- Limited to hypervisors that can provide the change-block list

ID: 381089 © 2019 Gartner, Inc.

HCI-Based Replication

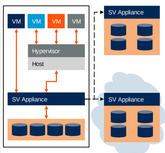


Advantages
- Supports asynchronous and synchronous operations
- Integrated with the hardware provisioning process
- Independent of the choice of hypervisor
- Usually supports cloud destinations for replication, but this is implementation-dependent

Disadvantages
- Requires matching hyperconverged infrastructure at the target
- Changing hyperconverged infrastructure vendors will require DR replication changes
- May require additional software in the cloud for cloud replication

ID: 381089 © 2019 Gartner, Inc.

Storage Virtualization Appliance Replication

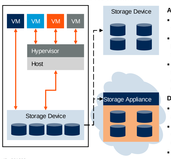


Advantages
- Supports asynchronous and synchronous operations
- Works with any underlying storage and supports any host, regardless of hardware platform or operating system
- Usually supports cloud destinations for replication by deploying SV appliance in the cloud
- Works equally well for physical and virtual workloads

Disadvantages
- Careful implementation required to avoid bottlenecks and single points of failure
- Use of appliance-based replication may create support issues with application vendors

ID: 381089 © 2019 Gartner, Inc.

Storage-Device-Based Replication

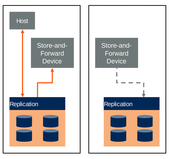


Advantages
- Supports asynchronous and synchronous operations
- Supports any host, regardless of hardware platform or operating system
- Supports cloud destinations only if the device vendor can run organization's storage code as cloud VM

Disadvantages
- Usually restricted to replicating to an identical device
- Changing storage infrastructure vendors will require DR replication changes
- Replication is often an extra cost item

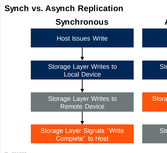
ID: 381089 © 2019 Gartner, Inc.

Hybrid Store-and-Forward Replication



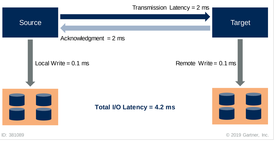
ID: 381089 © 2019 Gartner, Inc.

Synch vs. Asynch Replication



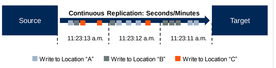
ID: 381089 © 2019 Gartner, Inc.

Synchronous Replication Latency



ID: 381089 © 2019 Gartner, Inc.

Continuous Data Protection



ID: 381089 © 2019 Gartner, Inc.

Snapshot-Based Replication

ID: 381089 © 2019 Gartner, Inc.

Backup-Based Replication

ID: 381089 © 2019 Gartner, Inc.