



Feature Brief: Asynchronous Replication

Lenovo ThinkSystem DS2200, DS4200, DS6200

June 2017

David R. Vestal
WW Product Marketing Manager



Lenovo.com/systems

What you'll learn:

- Overview of new AR Functionality
- Basic steps for setting up AR between

Introduction

The Lenovo ThinkSystem DS Series Arrays bring enterprise-class performance and features to small and medium businesses at an affordable price. Recently we introduced a new set of software features with firmware release 1.1. These features are an extension of Lenovo's previously released Lenovo SAN Manager software. This brief highlights one of the new features available in the 1.1 release - Asynchronous Replication (henceforth referred to as AR).

AR is a licensable feature that allows for creating remote copies of a virtual volume, virtual volume-group, or snapshot on a remote system.

The screenshot shows the Lenovo SAN Manager web interface. The top navigation bar includes 'Action' and 'REPLICATIONS'. The main content area is divided into two sections: 'Peer Connections' and 'Replication Sets'. Both sections have search bars, 'Clear Filters' buttons, and 'Show' dropdown menus. The 'Peer Connections' table has columns for Name, Status, Health, Type, Local Ports, and Remote Ports. The 'Replication Sets' table has columns for Name, Primary Volume, Secondary Volume, Status, Last Successful Run, and Estimated Completion. Both tables currently display 'No data available in the table'. The bottom status bar shows system metrics like IOPS and MB/s, and a timestamp: 09:50:23: Finished loading configuration data.

Figure 1: V3 Home Screen – Replication Topic

Licensing

AR is an optional feature and requires appropriate licensing be purchased from your storage provider. This is shown under the License Settings window – accessed by selecting the Home Icon, then Action > Install License. The License screen pop-up is shown in Figure 2 below.

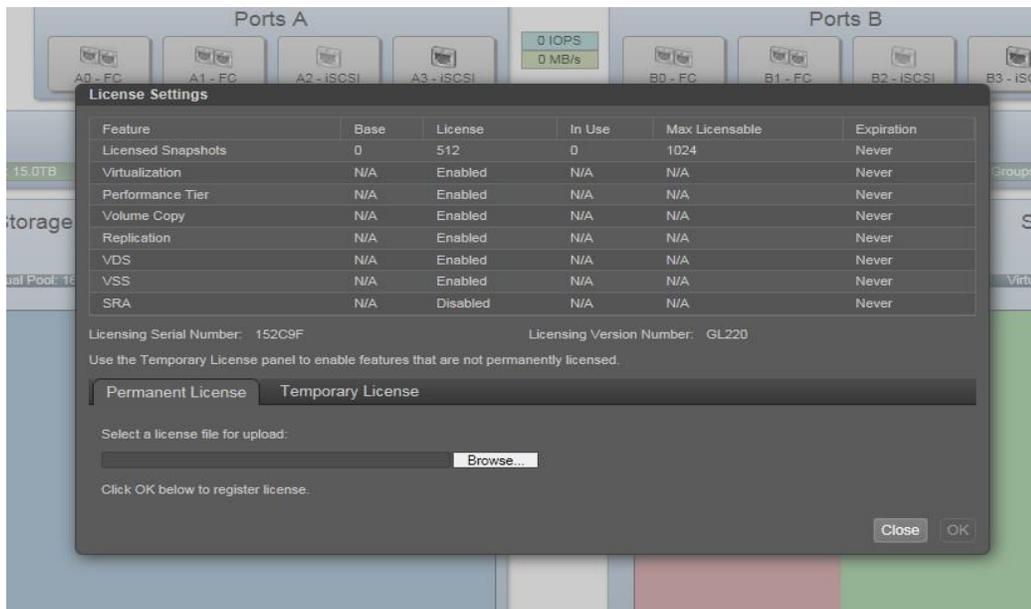


Figure 2: License Setting Screen

If a system has the Replication License installed, then it may perform replication with a remote peer if the peer has the Replication license installed as well. The license controls the number of volumes to be replicated – but AR currently supports a larger number.

Replication Prerequisites

Before a user can start replicating a volume, a peer-connection and replication set must be defined. A peer connection establishes bi-directional communication between a local and remote system. In addition to the appropriate licensing, the systems must also have:

1. iSCSI Host connectivity (**Note:** Replication will not be available over FC ports in the initial release. Hybrid systems – FC and iSCSI – will support replication over iSCSI ports)
2. Minimum of one reachable iSCSI port address of remote system
3. At least one pool per system

To establish a Peer connection, the local system communicates to the desired peer (remote) system with a user provided Host port IP address (in-band) - along with a unique, user-defined name for the peer connection. The systems then use in-band communications to exchange information and establish the peer connection – without user intervention. All the interconnected ports between the two systems will be made available for peer-connection I/O traffic. The user can later block ports on either side from being used for AR I/O, or they may wish to add physical connections and modify the Peer Connection to increase the number of paths that exist between the peers. Refer to Figure 3 for the Peer connection information available with the web browser interface.

The screenshot displays the 'REPLICATIONS' management interface. It features two main sections: 'Peer Connections' and 'Replication Sets'. The 'Peer Connections' section shows a table with one entry: 'Primary to Remote', which is 'Online' and 'OK' in health, using 'iSCSI' type and ports 'A3,B3' on both local and remote systems. The 'Replication Sets' section is currently empty, displaying 'No data available in the table'. A 'Peer Connection Information' pop-up window is overlaid on the 'Replication Sets' section, providing detailed information for the 'Primary to Remote' connection, including its serial number, connection type (iSCSI), status (Online), and specific local and remote ports.

| Name | Status | Health | Type | Local Ports | Remote Ports |
|-------------------|--------|--------|-------|-------------|--------------|
| Primary to Remote | Online | OK | iSCSI | A3,B3 | A3,B3 |

| Name | Primary Volume | Secondary Volume | Status | Last Successful |
|--------------------------------|----------------|------------------|--------|-----------------|
| No data available in the table | | | | |

Peer Connection Information

- Name: Primary to Remote
- Serial Number: 00c0ff1d98fa000037b4415601020000
- Connection Type: iSCSI
- Connection Status: Online
- Local Port: A3 - 172.16.254.112
- Local Port: B3 - 172.16.254.116
- Remote Port: A3 - 172.16.254.104
- Remote Port: B3 - 172.16.254.108

OK

Figure 3: Peer connection Information Pop-up

Note: The ethernet management communication path between peers is not required for this information to be accessed.

Once a peer connection is defined, the user creates a replication-set by specifying a volume or volume group on the local system (referred to as a primary volume(s)) and identifying the peer connection upon which to execute AR operations.

The volume (or volume group) copies on the remote peer system (referred to as secondary volume(s)) are created automatically as part of creating the replication-set. The system also creates and maintains a pair of snapshots of both the primary and secondary volumes which are hidden to the user but are used for recording and transferring changes to the primary volume (or volume group). Figure 4 shows the Create Replication Set pop-up window.

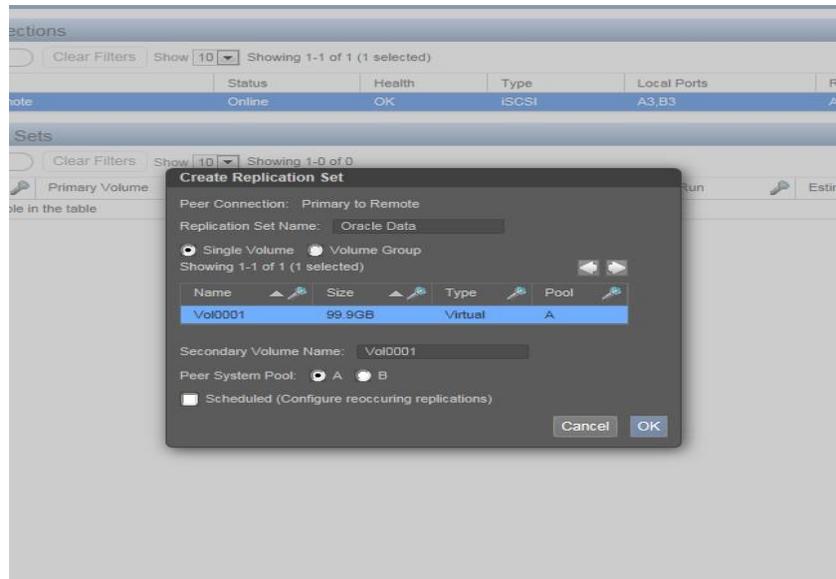


Figure 4: Create Replication Set pop-up

Replication Process

After you create a peer connection and replication set, you can then replicate volumes between the systems. There are two kinds of replications that will take place: an initial replication and subsequent replications. The initial replication creates a full copy of the primary volume to the secondary volume. Depending on how large your source volume is and the speed of the network connection, this initial replication may take some time. It is recommended to perform the initial copy locally and then move the peers system to a remote location to reduce the total time.

Subsequent replications are completed by taking a hidden snapshot of the primary volume and comparing the changes to the last time the hidden snapshot was taken of the volume. The system writes any changes it finds on the hidden primary snapshot to the hidden secondary snapshot, after which the hidden secondary snapshot is promoted to the secondary volume.

This activity is automatically managed by the Lenovo SAN Manager software – and is kept hidden in an effort to simplify the user experience.

Once the replication-set is created, there are two primary methods to initiate a replication. This is done by first selecting the Replication topic, then selecting the Peer connection and appropriate Replication Set – then click the Action button:



Figure 5: Two primary methods to initiate a replication

The first option is “Initiate Replication” – this is a manual method which will start a replication immediately (See figure 5). The second option is “Schedule Replication” – in which the user can define a schedule to initiate replications on a user-defined schedule (See figure 6).

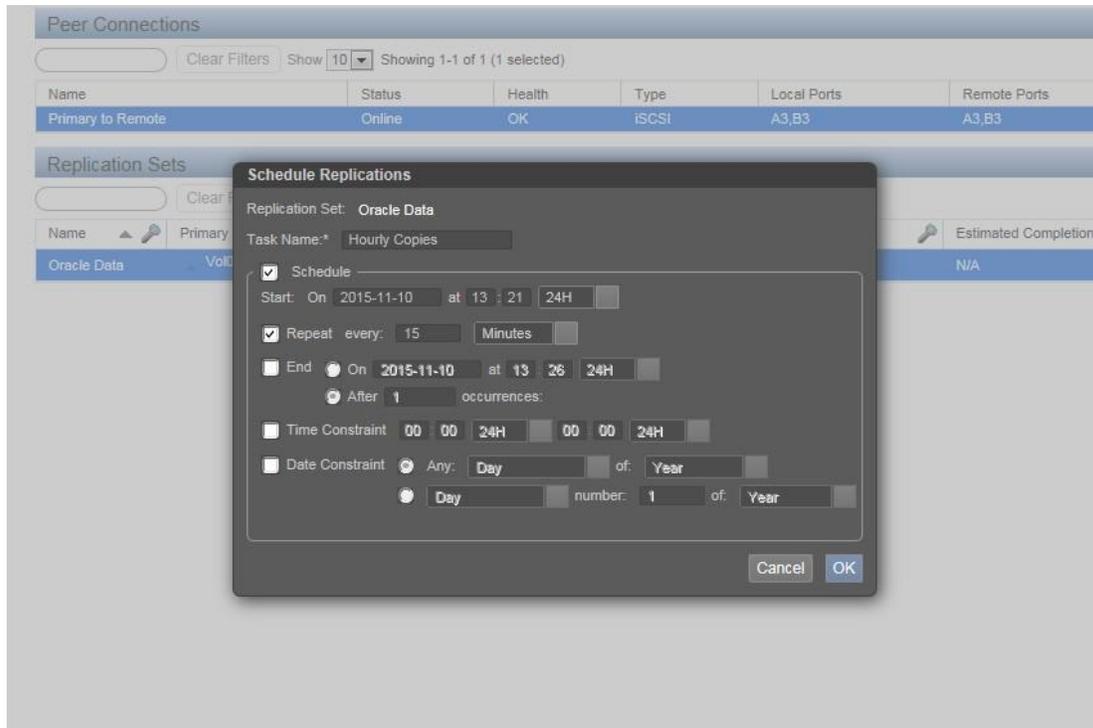


Figure 6: Schedule replication screen

Primary Volume Considerations

The following Volume operations cannot be performed on a Volume that is the Primary Volume of a Replication Set:

- The volume cannot be deleted. The replication set must be deleted before the volume may be deleted.
- The volume cannot be defined in more than one replication set.

Secondary Volume Considerations

The following operations apply to the Secondary Volume (the target of the replication):

- It is a base volume and for first release will be created by the system as part of creating the replication set.
- It cannot be mapped as write-able to a host. It is only updated by the replication set.
- It cannot be mapped as readable to a Host.
- It cannot be specified in an additional replication set. Note: A user is free to configure a system in the following way to provide replicating the contents of secondary volume to a third Peer:
 - Create a Snapshot of the secondary volume.
 - Define a "Snapshot Schedule" for the Snapshot that "Resets" it thereby picking up the changing contents of the Secondary Volume.
 - Define a Replication of the Snapshot to a third Peer.
 - It cannot be moved, expanded or deleted while the replication set exists.
 - The volume may be renamed.
 - Deleting the replication set leaves the volume intact with the last successfully replicated data image and it will be a normal base volume with no restrictions.
 - A secondary volume can be snapped at any time to create new volumes that can be mapped to a host and read or modified.
 - "Reset Snapshot" may be performed on a secondary volume's snapshots. ● A "Rollback to a snapshot" cannot be performed on a secondary volume.
 - A secondary volume will be shown as a new type of volume in "Show Volumes" from both CLI and WBI display.

Collision Handling

"Collision" refers to a replication being initiated while an existing replication is in progress. Asynchronous Replication does not queue replications on Collisions but instead simply discards the scheduled or user/application initiated replication. If a replication has failed and is "stuck" or otherwise long running, the user can use "Abort Replication" to clean up the existing and then manually force a desired replication.

Future releases will introduce a "Collision Policy" to allow specifying different behavior on replication collision.

Use Case: Disaster Recovery

Asynchronous Replication is intended to provide a cost-effective means to maintain a copy of a user's data at a remote location. The initial implementation is not a site-failover solution, but provides capabilities for manual disaster recovery support.

- A snapshot copy of Secondary Volume may be done at any time and then mapped and written as any other volume. This provides for the recovery and use of the replicated data without modifying or affecting the latest valid copy of the data from before the disaster.
- Forced Delete of the Replication while the primary site is down or cannot be communicated with is also provided. This allows for manual cleanup of the secondary site should the primary be permanently removed.

For the initial release, if a primary site goes down and the user wants to perform operations at the secondary site – there are two options:

- Use the secondary volume directly by deleting the replication set to make it a base volume which can be mapped to hosts. Deleting the replication set means the secondary volume becomes a base volume and is no longer the target of a replication.
- Take a snapshot of the volume which can be mapped to hosts.

Should the primary volume become available and you want to use it as is in preparation for another, a new replication set with a new secondary volume must be created in order to transfer the changes on the secondary site back to the original primary site.

The following support for additional Disaster Recovery support is planned for future releases:

- Site Failover: Allow unlocking of the Secondary volumes and make them directly writeable while providing for the efficient copying back of any updates made to the original Primary site while in "Site Failover" mode.
- Restore Replication: To come out of "Site Failover" and specify which Peers data images should be used as the current data.
- Reverse Direction: Ability to reverse the direction of Replication.

Replicating Volume Groups

The Volume Groups feature allows creation of a named group of virtual volumes that all belong to the same Pool. Volume Groups can be used to manage mappings for the volumes as a single group – convenient for environments where large amounts of Volumes are to be mapped in a similar fashion. The first release of AR provides the additional capability to define a replication set on these Volume Groups. The volumes in the group will then be replicated with consistency semantics. A replication set defined on a Volume Group is managed in the same manner as one defined on a single volume.

If the contents of the volumes in the group are consistent from the application view at the time the replication is initiated, the semantics of volume group replication maintain the consistency in the remote copies. The replication of the group occurs automatically and at each step of replication the operations are performed for all volumes on the group before proceeding to the next step. The data images to be copied will be the contents that all volumes in the group have at the time the replication is initiated. On power loss or failover of a single controller or loss of an entire storage system and its restoration, either all volume replications will be suspended and left at the same source time or all volume replications are successfully continued and the copies advanced to be images of the sources at the next later mutual time point.

Restrictions and notes on Volume Groups:

- All volumes in a Volume Group must belong to the same Pool. This is a current restriction of Volume Groups in release 1.0.
- Creating a replication on a group will cause secondary volumes to be created on the specified Peer and those volumes will also have a group created for them.
- A Volume Group cannot be deleted while a replication set is defined for it.
- For first release, a Volume Group that has a replication set defined for it cannot have volumes added to or removed from the group. The replication set must first be deleted to change the group membership and then recreated.
- Volumes within a volume group may be replicated separately from the group if the group does not have a replication defined for it. In a future release the restriction of only allowing a volume to participate in one replication will be lifted. The group will then be

allowed to have one replication set defined for it and the member volumes may have others for them individually.

The initial release containing Asynchronous Replication only provides the capability to replicate Volume Groups and does not provide capability for the user to snap or reset a Volume Group as an operation on a Volume Group (they can always operate on the individual volumes directly).

Maintaining Point-in-Time Copies of Replicated Images

The initial release of AR maintains only the most recent copy of the primary and secondary volume, leveraging snapshots hidden from the user to manage changed data between the two sites. The first release does not provide integrated snapshot creation triggered directly by the AR replication either at its start (for the Primary side) or completion (for the Secondary side).

Users may maintain a "loosely coupled" history by manually creating a separate snapshot schedule on the Secondary (or Primary) volume that is offset from the Replication schedule at a time when the replication is likely to have completed. This provides some history of the replicated images but may lag or miss some images or keep some duplicate images.

A future release will provide "Integrated Snapshot Creation" where the Replication creates a snapshot of the image being transferred:

- This can be specified for either Primary or Secondary volumes separately. It would be configured by defining a snapshot schedule on either a Primary or Secondary volume but instead of specifying time controls for the schedule the user would select the Replication as the trigger.
- If specified on a Primary volume it will create a snapshot of data image being replicated and create it when it starts the replication.
- If specified on a Secondary volume it will create a snapshot of the data image just transferred to the secondary volume when a replication successfully completes.
- The snapshots created would be immediately available to the user to map for read or write.

- Since these are separate snapshot schedules this provides the capability to store differing depths of history between the Primary and Secondary sides. Example: a replication may be set up to happen once a day and then separate schedules to save the last 7 images replicated on the Primary side and save the last 30 on the Secondary side.

Additional Replication Functions

Other functions available for Replication include:

- Aborting a Replication – User may abort a replication from the Primary system only.
- Suspending a Replication – User may Suspend a replication from the Primary system. When a replication is suspended, all replications in progress are paused and no new replications are allowed to occur (including schedule replications). Replications attempted during this time will fail.
- Resuming a Replication – User may resume a replication from the Primary system only. All paused Replications are resumed and new replications are allowed to occur.

Summary

With appropriate licensing, customers can leverage the built in capabilities the arrays provide as part of a base DR strategy.

Table 1 summarizes AR features. More detailed information will be available in the User Documentation and Online Help that comes with the array management interface.

Table 1: AR Feature Summary

| Feature | AR |
|--|-----------------------------------|
| Topology | iSCSI |
| # Peers | 1 |
| Max volumes in a volume group | 16 |
| Max number of replication volumes | 32 |
| Scheduling Capability | GL221 and higher |
| # of licenses needed | One for target and one for source |
| Volume expansion | Yes |
| Basic statistics monitoring replication status | Yes |
| Devices that can be replicated to/from | S3200/S2200 |
| Challenge-Handshake Authentication protocol (CHAP) | Yes |
| Point-in-Time History | Manual |
| VMWare SRM integration with SRA | No |
| Remote System Detection | In-band |
| Collision Handling | Not in first release |

LENOVO PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. This information could include technical inaccuracies or typographical errors. Changes may be made to the information herein; these changes will be incorporated in new editions of the publication. Lenovo may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any performance data contained herein was determined in a controlled environment; therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems, and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Any references in this publication to non-Lenovo Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this Lenovo product, and use of those Web sites is at your own risk.

Lenovo, the Lenovo logo, System x, ThinkServer, and ThinkSystem, are trademarks of Lenovo in the United States, other countries, or both.