EMC²
where information lives®

CLARiiON® and FCIP – A Practical Inter-Continental DR and HA Solution

EMC Proven Professional Knowledge Sharing 2009

Jaison K. Jose
Sr. Storage Administrator
EMC Corporation
Jose_jaison@emc.com

EMC² PROVEN PROFESSIONAL

Table of Contents

Abstract

This document explains the functional design and implementation of a disaster recovery solution using CLARiiON and FCIP. This solution uses the advantages of various array features such MirrorView/A™, SnapView™, SnapShot and SnapView Clones along with the VSAN and FCIP fabric features.

Along with Navisphere® Manager, Cisco Fabric Manager and CLI are used for the configurations. PowerPath® and NaviAgent are used in the hosts for multipathing and host array communications.

I assume that anyone reading this article has basic knowledge of SAN infrastructures, CLARiiON and MDS SAN switches.

Introduction

EMC CLARiiON offers options to meet almost all today's business needs. We can easily achieve complex industry requirements when we join this small magic box to other technology. I would like to share an Intercontinental disaster recovery solution that I achieved with the help of features such as CLARiiON, MirrorView/A, SnapView Clones, SnapView SnapShots, FCIP and VSAN.

The business required us to implement a primary site in Europe and a disaster recovery (DR) site in the United States. This was for a customer facing EMC application, so it was very important to have a robust solution with a proper DR plan.

FCIP was our first choice to manage the data movement over the Atlantic. We could take advantage of internet connectivity and VPN to create a tunnel between the two sites. Data Replication traffic was segregated from production by separating replication ports to a special VSAN that used FCIP to extend it to the DR site.

CLARiiON was the obvious choice of midrange application. The data volume was huge but not dynamic. The primary concern was data availability. MirrorView/AS was the best option to connect two sites separated by thousands of miles. We used MirrorView/A to send the data through the FCIP tunnel to the DR CLARiiON.

A snap of the replicated LUNs was presented to the DR hosts. This avoided promoting the secondary image in order for the DR hosts to view the replicated data.

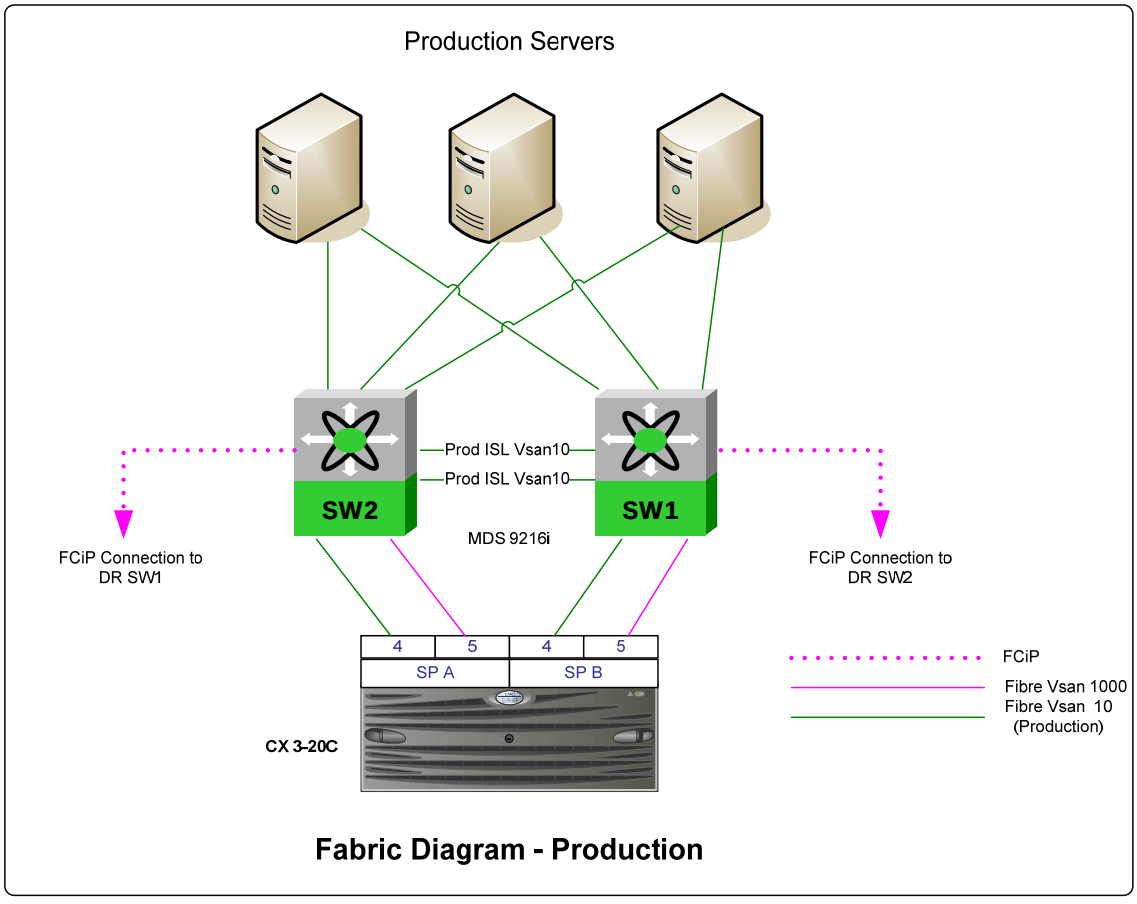
Now, we needed a backup solution. This environment was hosted in a third party data hosting facility. To have a tape based or external backup option can cost more money, so SnapView clones were our answer. A gold copy of Production and DR LUNs were set up to protect against data corruption or data loss caused by a human error.

MirrorView A/S, SnapView SnapShots, SnapView Clones, FCIP and VSAN – all these features were utilized in this unique DR solution. This is not just a concept; it has been implemented and is currently working in production.

Fabric Setup – Production

The majority of the storage production is for ESX Servers. The application uses multiple VMs to meet customer requirements. However, each VM has unique and critical data. The majority of this information is less dynamic, resulting in minimal incremental changes.

A highly available production setup was important for this application. However, the amount of data is much less compared to other major applications. A CX3-20C could meet the storage requirements of this production environment. Two fabric switches were deployed for hardware redundancy. We chose MDS9516i switches because they come with two Gigabit Ethernet ports that can be used with FCIP for remote replication.



Fabric Diagram - Production

Note that CX3-20C comes with 8 iSCSI and two fibre channel front end ports. 0, 1, 2, and 3 ports are iSCSI on both SPs. Ports 4 and 5 were fibre front end. SPA5 and SPB5 were connected to a different VSAN to segregate MirrorView ports from production traffic. Production HBAs and FA ports were connected to VSAN10, and VSAN1000 was used for remote replication.

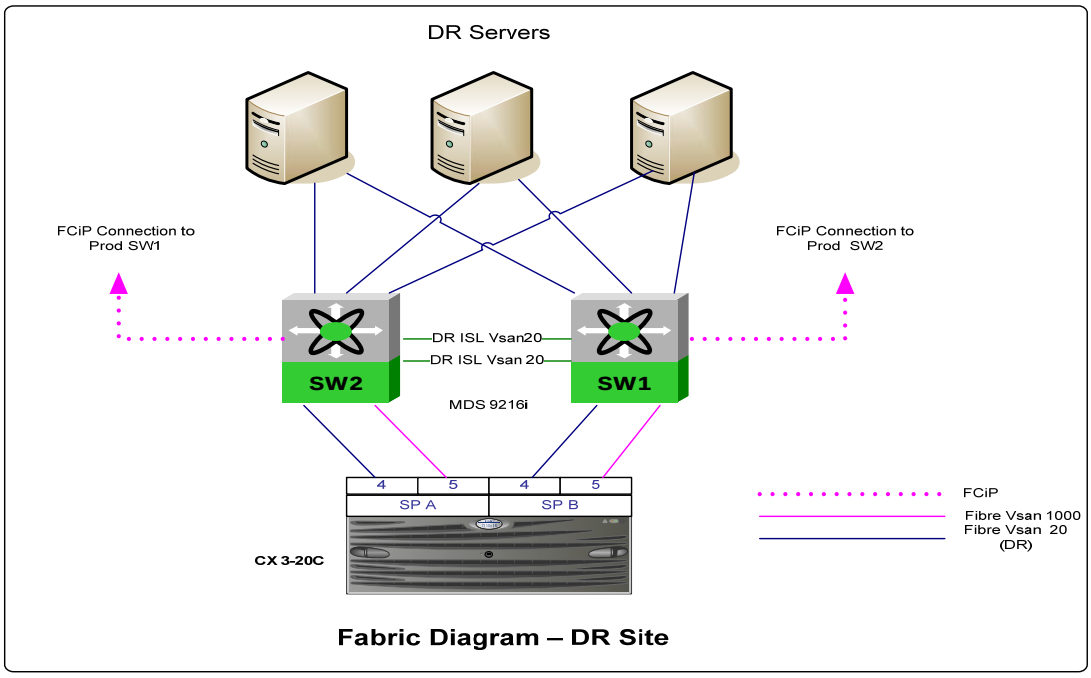
Each host was connected to both the switches using two different HBAs. Each HBA was zoned to both SPA4 and SPB4 with the help of a production ISL to reduce LUN trespasses.

The four paths on each host are accomplished by zoning:

1. HOST_HBA1_CLARIIONPROD_SPA4
2. HOST_HBA1_ISL_CLARIIONPROD_SPB4
3. HOST_HBA2_CLARIIONPROD_SPB4
4. HOST_HBA2_ISL_CLARIIONPROD_SPA4

Fabric Setup – DR

The DR site is almost an exact copy of the production site with the small exception of a different VSAN id for the DR Host traffic.



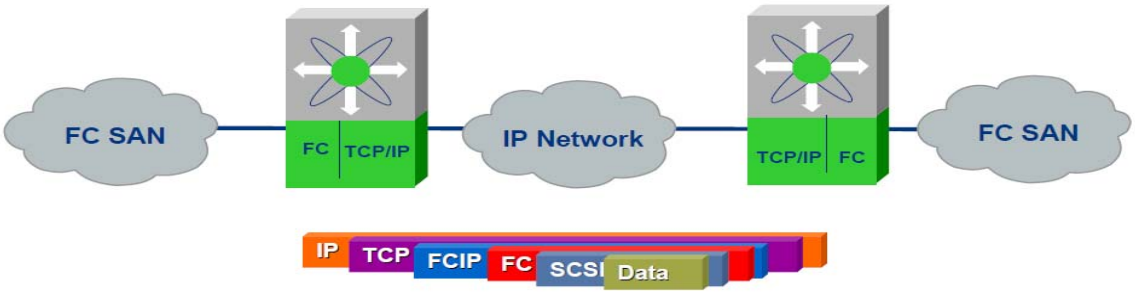
Fabric Diagram – DR Site

The VSAN 20 at DR site segregated Production and DR fabrics. Only VSAN1000 was allowed over FCIP to extend and merge with production. Other configurations remain the same as in the primary site.

FCIP – our carrier over the Atlantic!

Then, we can connect the production and DR sites. Fibre Channel over IP (FCIP) is a protocol that allows sending Fibre Channel (FC) data using IP Protocol. In this mechanism, FC blocks are encapsulated into FCIP packets and transferred over the network traffic. The FCIP mechanism uncovers the encapsulated data and sends it to the appropriate Fiber Channel Port at the secondary site.

A typical example of an FCIP SAN extension is illustrated below. We can see how the data is encapsulated through various protocols.



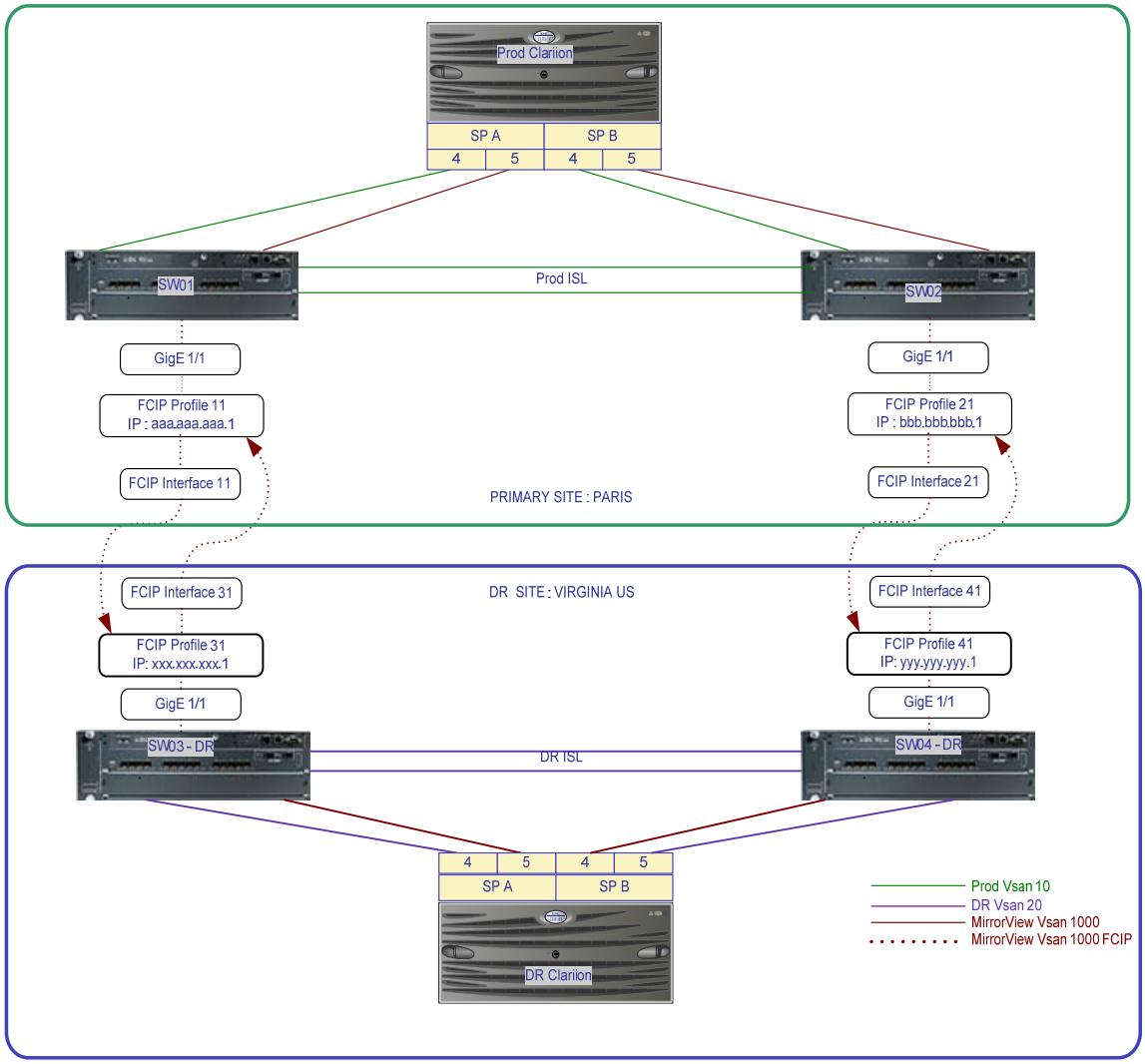
In our scenario, our network team offered a secure VPN tunnel between both sites. This tunnel was used for further FCIP data communication between the sites.

We chose MDS9216i for SAN because it integrates Fibre Channel and IP into a single form factor, providing fourteen 2-Gbps Fibre Channel interfaces and two Gigabit Ethernet ports that can be configured to support Fibre Channel over IP (FCIP) for long distance SAN extension. In addition, we did not need to purchase additional licenses to use FCIP with MDS9216i.

FCIP uses Profiles; interfaces must accomplish the link between two sites. You must create a profile n Cisco MDS Switches when configuring FCIP. The profile holds the TCP parameters such as IP, subnet mask etc. Create at least one FCIP profile for each participating GigE port.

The FCIP Tunnel (interfaces) holds the peer interfaces' IP information. We had to create this logical interface on each end of the link.

A more detailed view of our configurations is illustrated below. For security reasons, the actual IP addresses have been masked to alphabets. (example xxx.xxx.xxx.1)



FCIP Configuration Details

The output of the configuration of DR switch 3 follows. The GigE ports are set to use ip address xxx.xxx.xxx.1

```
sw03-mgt# show ip interface gigabitethernet 1/1
GigabitEthernet1/1 is up
  Internet address is xxx.xxx.xxx /24
  Broadcast address is 255.255.255.255
```

FCIP profile configuration

```
sw03-mgt# config
Enter configuration commands, one per line.  End with CNTL/Z.
sw03-mgt(config)# FCIP profile 31
sw03-mgt(config-profile)# ip address xxx.xxx.xxx.1
```

The characteristics of the FCIP profile are displayed below.

```
sw03-mgt(config-profile)# do show FCIP profile 31
 FCIP Profile 31
    Internet Address is xxx.xxx.xxx.1
    Tunnels Using this Profile: FCIP31
    Listen Port is 3225
    TCP parameters
        SACK is enabled
        PMTU discovery is enabled, reset timeout is 3600 sec
        Keep alive is 60 sec
        Minimum retransmission timeout is 200 ms
        Maximum number of re-transmissions is 4
        Send buffer size is 0 KB
        Maximum allowed bandwidth is 1000000 kbps
        Minimum available bandwidth is 500000 kbps
        Configured round trip time is 1000 usec
```

```
        Congestion window monitoring is enabled, burst size is 50
KB
        Auto jitter detection is enabled
sw03-mgt(config-profile)# end
```

At this point, we had created the FCIP profile on DR sw3, which was associated with the gigE port 1/1 and its IP address.

FCIP interface configuration

Creating the FCIP interface required two things: the profile it had to use and the IP address of the peer switch that it had to communicate with.

```
sw03-mgt# config
Enter configuration commands, one per line.  End with CNTL/Z.
sw03-mgt(config)# interface FCIP 31
sw03-mgt(config-if)# peer-info ipaddr aaa.aaa.aaa.1
sw03-mgt(config-if)# use-profile 31
sw03-mgt(config-if)# end
```

This will create the FCIP interface; we can verify the connections as below.

```
sw03-mgt# show int  FCIP 31
 FCIP31 is down (Tunnel port src interface unbound)
    Hardware is GigabitEthernet
    Port WWN is 00:00:00:00:00:00:00:00
    Admin port mode is auto, trunk mode is on
    snmp link state traps are enabled
    Port VSAN is 1000
    Using Profile id 31
    Peer Information
      Peer Internet address is aaa.aaa.aaa.1 and port is 3225
    Write acceleration mode is configured off
    Tape acceleration mode is configured off
```

```
    Tape Accelerator flow control buffer size is automatic
    Ficon Tape acceleration configured off for all VSANs
    IP Compression is disabled
    Special Frame is disabled
    Maximum number of TCP connections is 2
    Time Stamp is disabled
    QOS control code point is 0
    QOS data code point is 0
    B-port mode disabled
    TCP Connection Information
      0 Active TCP connections
      0 Attempts for active connections, 0 close of connections
    5 minutes input rate 0 bits/sec, 0 bytes/sec, 0 frames/sec
    5 minutes output rate 0 bits/sec, 0 bytes/sec, 0 frames/sec
      0 frames input, 0 bytes
         0 Class F frames input, 0 bytes
         0 Class 2/3 frames input, 0 bytes
         0 Reass frames
         0 Error frames timestamp error 0
      0 frames output, 0 bytes
         0 Class F frames output, 0 bytes
         0 Class 2/3 frames output, 0 bytes
         0 Error frames
```

A summary of profiles and interfaces on production and DR are as below:

Prod Sw01.

```
 FCIP Profile 11
    Internet Address is aaa.aaa.aaa.1  (interface
GigabitEthernet1/1)
    Tunnels Using this Profile: FCIP11
    Listen Port is 3225
```

```
    TCP parameters
        SACK is enabled
        PMTU discovery is enabled, reset timeout is 3600 sec
        Keep alive is 60 sec
```

FCIP Interface 11

```
Using Profile id 11  (interface GigabitEthernet1/1)
Peer Information
  Peer Internet address is xxx.xxx.xxx.1 and port is 3225
Trunk VSANs (admin allowed and active) (1,1000)
```

Please note that a trunk / tunnel and allowing only VSAN1 and vsaan1000 to communicate over the FCIP.

DR Sw03.

FCIP Profile 31

```
    Internet Address is xxx.xxx.xxx.1 (interface
GigabitEthernet1/1)
    Tunnels Using this Profile: FCIP31
    Listen Port is 3225
    TCP parameters
        SACK is enabled
        PMTU discovery is enabled, reset timeout is 3600 sec
        Keep alive is 60 sec
```

FCIP Interface 31

```
Using Profile id 31  (interface GigabitEthernet1/1)
Peer Information
  Peer Internet address is aaa.aaa.aaa.1 and port is 3225
Trunk VSANs (admin allowed and active) (1,1000)
```

Note that FCIP interface 11 on the production switch is pointing to the IP address of FCIP Profile 31 in the DR Switch. The FCIP interface on the DR switch points to the IP of the production switch. We are allowing only the VSAN1000 to communicate over the FCIP link.

Fine tuning the FCIP link

We fine tuned this connection parameter by observing and setting the optimal values. An extended ping was used to determine an average round trip time.

```
sw01-mgt(config)# do ping
Target IP address: xxx.xxx.xxx.1
Repeat count [5]:
Datagram size [100]: 1472
Timeout in seconds [1]:
Extended commands [n]: y
Source address or interface: aaa.aaa.aaa.1
Type of service [0]:
Set DF bit in IP header [n]: y
Data pattern [0xABCD]:
Sweep range of sizes [n]:
PATTERN: 0xabcd
PING   xxx.xxx.xxx.1   (xxx.xxx.xxx.1)   from   aaa.aaa.aaa.1   :
1472(1500) bytes of data.
1480 bytes from xxx.xxx.xxx.1: icmp_seq=1 ttl=46 time=150 ms
1480 bytes from xxx.xxx.xxx.1: icmp_seq=2 ttl=46 time=148 ms
1480 bytes from xxx.xxx.xxx.1: icmp_seq=4 ttl=46 time=148 ms
1480 bytes from xxx.xxx.xxx.1: icmp_seq=5 ttl=46 time=148 ms
```

The usable bandwidth of this connection can be obtained from network team

```
sw02-mgt(config-profile)# tcp max-bandwidth-mbps 4 min-available-
bandwidth-mbps 1 round-trip-time-ms 300

tcp max-retransmissions 6
```

Remember that this particular implementation was able to use very low bandwidth and roundtrip time because the incremental changes were minimal.

The steps above were repeated for the second pair of switches (Prod Sw2 and DR sw4).

Merged fabric

VSAN1000 on both the switches will merge once FCIP establishes the connection.

This allows us to zone Production SPA5 to DR SPA5 and Production SPB5 to DR SPB5. These connections were established over the SAN extension using the FCIP tunnel.

Zones on VSAN1000

1. CLARIIONPROD_SPA5_ FCIP_CLARIIONDR_SPA5
2. CLARIIONPROD_SPB5_ FCIP_CLARIIONDR_SPB5

MirrorView requires an FC connection between SP ports. We accomplished this using the FCIP tunnel.

CLARiiON Configurations

Navisphere Domain

Both the Production and DR CLARiiONs were configured to use the same Navisphere domain. This enabled an easy MirrorView configuration. The network switches were configured to provide a connection between sp ports. Ports 80/443 and 6389 are opened between CLARiiONs and the management host for the Navisphere and agent communications.

To avoid huge traffic over the WAN, both CLARiiONs were initially configured at the same location. Once the initial full synchronization of MirrorView LUNs was complete, both CLARiiONs were powered down and shipped to the appropriate locations overseas.

MirrorView/A

MirrorView/A is a storage-based application residing on the CLARiiON. It provides a host independent protection solution that duplicates changes in production site data (primary) to a secondary site (secondary) at regular intervals (after an initial full synchronization). Because MirrorView/A does not use a synchronous mechanism, it is distance-independent and allows replication over IP networks at extended distances. MirrorView/A ensures that there is a restartable, point-in-time copy of the data at the remote CLARiiON. MirrorView/A is storage-based software, thus uses no host CPU cycles. Host applications are unaffected by the latency of the network that connects the primary to the secondary. MirrorView/A operates in the background, transparent to any hosts or applications

The reserved LUN pools are created with the best practices guideline of 20% of the source data.

MirrorView Is configured with the following parameters.

Recovery Policy :Automatic

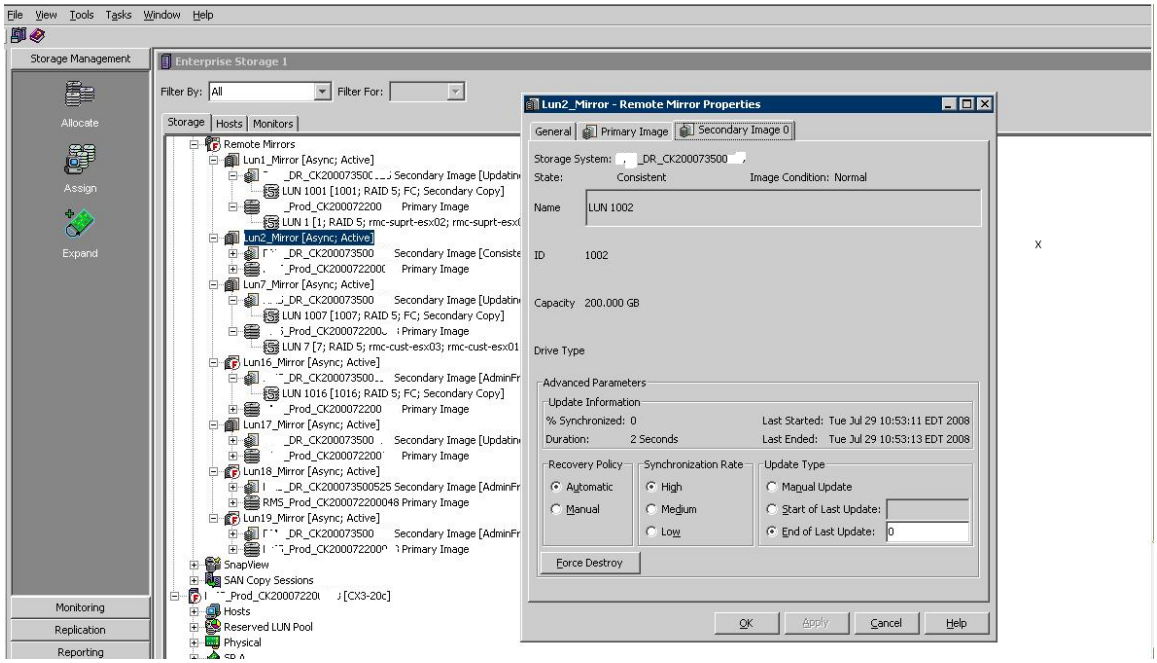
In case of disconnect because of a network glitch, MirrorView will reestablish the replication once the connection is recovered.

Synchronization rate: High
MirrorView is allowed to use all available bandwidth for replication.

Update type: Zero seconds after the end of last update.

With this setting, MirrorView will initialize another update process as soon as the previous update cycle is complete.

Following is a screen shot of the MirrorView configuration of a 200GB LUN. Site names and CLARiiON serial numbers are masked for security reasons.

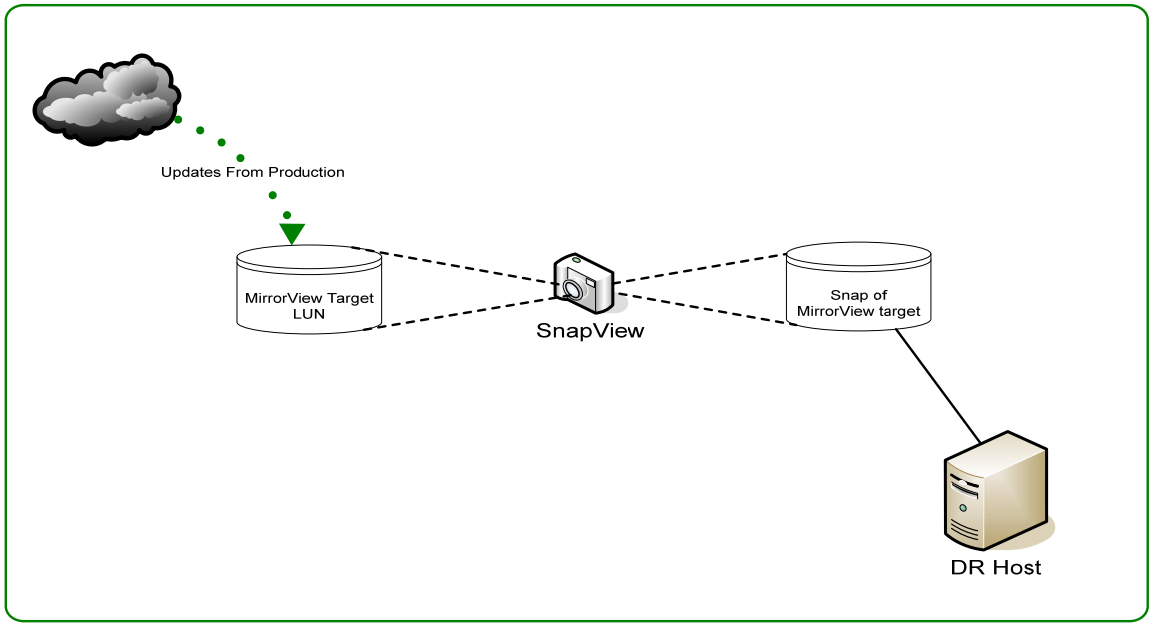


For the initial setup, we had both CLARiiONs in the same location and we were able to establish the initial synchronization locally and then send the CLARiiONs to separate locations. The incremental changes were updated though the FCIP connections. However, the secondary CLARiiON LUNs were not available for host applications unless we promoted them.

'Promote' is the operation by which the administrator changes an image's role from secondary to primary. As part of this operation, the previous primary image becomes a secondary image. The remote CLARiiON starts sending data back to primary. As our particular application included more static data, it was acceptable to discard the changes in the DR site. We set up a SnapView / SnapShot in the DR site to avoid the promoting function and overloading WAN traffic.

SnapView / SnapShot

CLARiiON's SnapView / Snap Shot provides a point in time copy of any given LUN. In our case, we created snapshots of MirrorView secondary LUNs and made them available to the DR hosts. The DR hosts see these LUNs as normal LUNs and are able to bring up the data and applications just as in production. Again, all the changes made by DR hosts are over written by new snaps whenever an updated snap is activated.

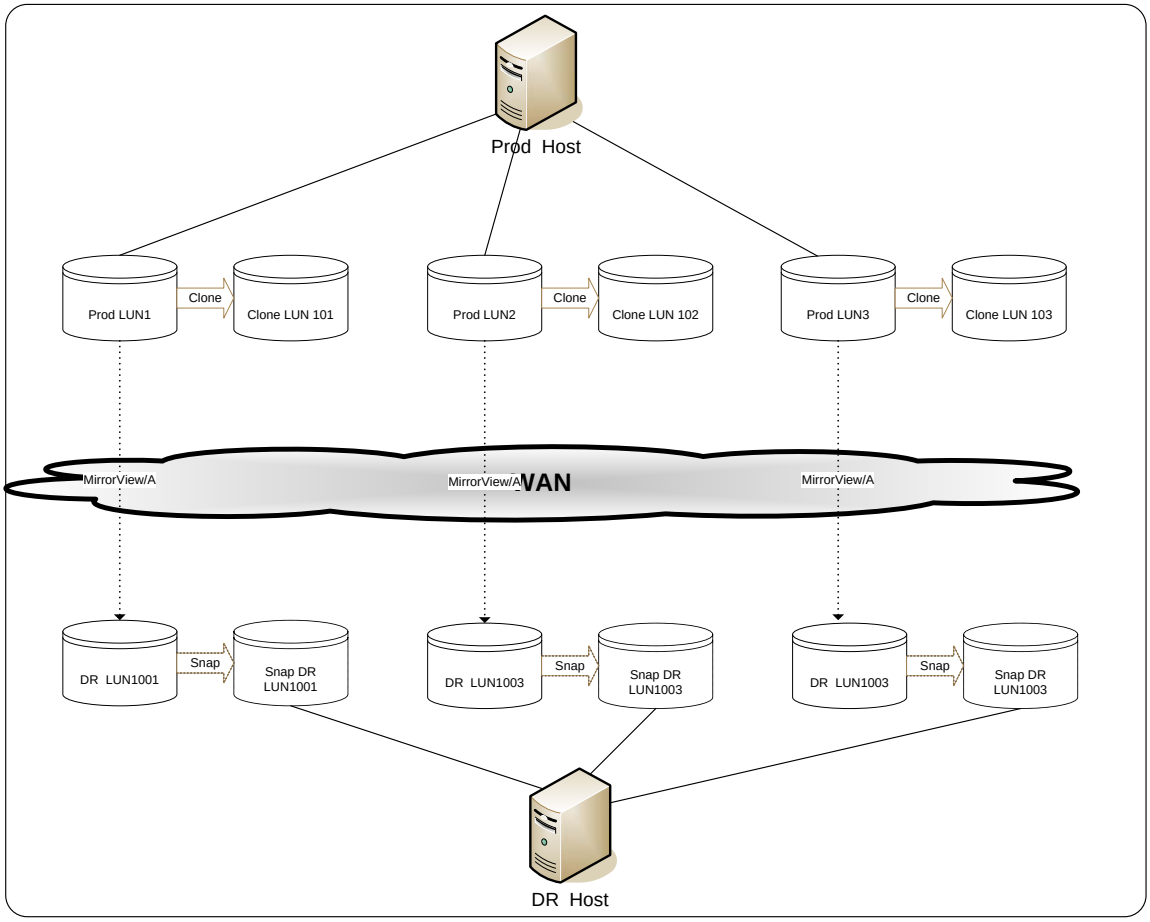


SnapView Clones

This setup was deployed at a third party secure datacenter. Implementing a dedicated data backup setup was inconvenient. We could make use of SnapView Clones for local information protection.

SnapView Clones offer a full replica of any given LUN on an array. This can be used for business continuity, decision making support, or as data backup. In production, SnapView clones are setup to take replicas of production LUNs. This increases recovery of production data in case of data corruption or hardware failure.

MirrorView, Clones and Snapshot are illustrated below.



The clones are refreshed daily, just like having a data backup run every night.

Summary

It seems like a very difficult task to accomplish DR between continents with limited resources. However, the step by step progress in this project gave us confidence that we can achieve complex results with appropriate technology strategies. It is amazing to see how the various technologies work together to accomplish the complex demands of today's business.

References

FCiP Definition and protocol diagram: Connectrix MDS Series Advanced Configuration and troubleshooting Guide.

MirrorView/A definition: MirrorView and SAN Copy foundations