Data Center Backend Infrastructure — Solutions for Disaster Recovery

BRKDCT-2004

Gilles Chekroun

Cisco Networkers 2007

HOUSEKEEPING

- We value your feedback, don’t forget to complete your online session evaluations after each session and complete the Overall Conference Evaluation which will be available online from Friday.
- Visit the World of Solutions on Level -01!
- Please remember this is a ‘No Smoking’ venue!
- Please switch off your mobile phones!
- Please remember to wear your badge at all times including the Party!
- Do you have a question? Feel free to ask them during the Q&A section or write your question on the Question form given to you and hand it to the Room Monitor when you see them holding up the Q&A sign.

Agenda

- Introduction to “Backend” Infrastructure
- Local Recovery Methods
- Remote Recovery Methods
- MDS Building Blocks
- DR: The Game Plan
- DR: Building Your Plan

Introduction to Backend Infrastructure

Disaster Tolerance vs. Recovery

Disaster Tolerance: The Ability to Survive an Expected Failure with Zero Downtime

- Identify the issue
- Resist the attack on availability
- Repair the issue
- Prevent future occurrences

Recovery

Disaster Recovery:
The Ability to Minimize an Outage

- Identify the impact and scope of the damage
- Identify the best method to recover with
- Repair the damage
- Recover from the impact
- Prevent future occurrences

Where to Begin?

- Disaster prevention and recovery can be achieved using tools at all levels of the backend infrastructure

  Host: applications, logging filesystems, multipathing software, RAID and clustering

  SAN: VSANs, Zones, IVR, SANtap

  Storage: RAID, Replication, snapshots

- Leverage multiple levels of defense and multiple strategies as no one tool will solve all of your problems

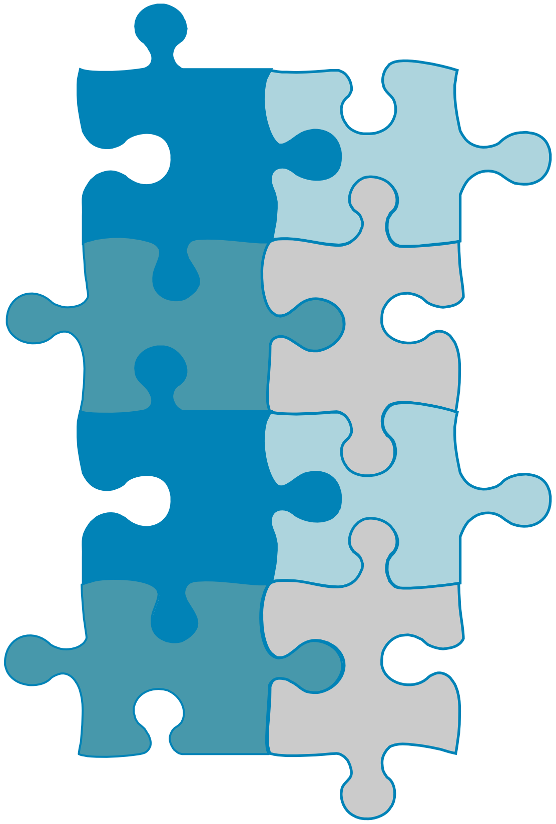
Using the Right Tool for the Job

There Is No “Cure All” for Any Disaster

- Implementing safeguards and preventative measures on all levels ensures choices are available on recovering from a disaster

  For example, depending on replication may prevent outages due to array failure, but may not protect against data corruption

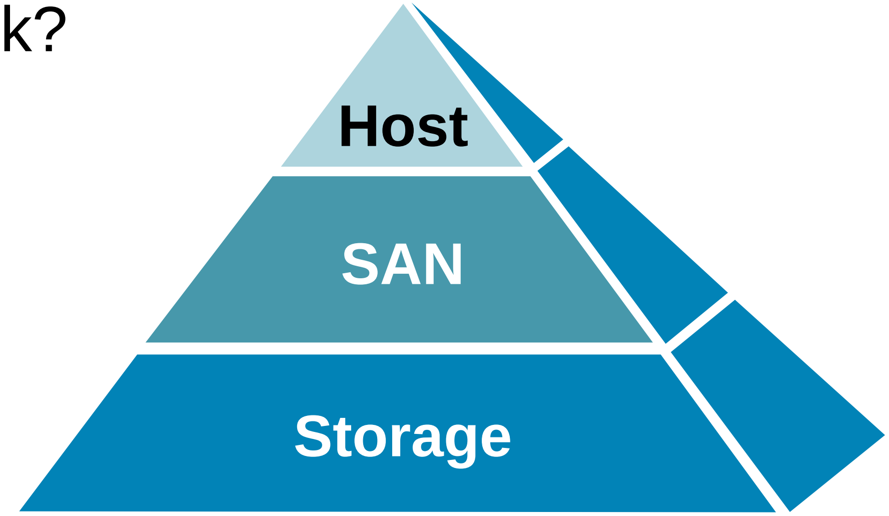
- Protect yourself against failures at any level of your architecture, including the one operating the equipment



Push Recovery "Down the Stack"

Keeping the Application Unaware of the Threat

- Application based recovery almost always requires an outage, and is not transparent to the clients
- Can multipathing software redirect the IO?
- Can the SAN route around or isolate the damage?
- Can the array mirror the disk?



“What If…” Planning?

- When assessing your infrastructure, point to a device or feature and say, “What if it failed?” and then determine if your application would survive that failure?

  No device guarantees 100% uptime

  The more often someone modifies it, the greater the chance of an ‘unintended’ issue

- Weigh the cost of protection vs. the cost of being offline

  It may not be feasible to mirror an entire disk array, but some applications may justify the cost of mirroring their data across arrays

Local Recovery Methods

Agenda

- Introduction to “Backend” Infrastructure
- Local Recovery Methods
- Remote Recovery Methods
- MDS Building Blocks
- DR: The Game Plan
- DR: Building Your Plan

Local Recovery

- The ability to tolerate or recover from a disaster without having to fail the application to a remote site
- Starting with the Host, what can provide tolerance and recovery?

  Application: Rolling forward redo logs, restarting an instance, application based clustering

  Filesystem: Journaling and FS based snapshots

  Volume Management: Mirroring/RAID, snapshots, replication

  Clustering: Operating System level or Logical Partitioning/Virtualization based

  Server Hardware: Hot swap components

Local Recovery on the SAN

Adding Intelligence in the SAN Enables It to Aid in Conflict Detection, Identification, Resolution and Prevention

- Detection: Online diagnostics, Callhome, SNMP traps, RMON
- Identification: Debug, fcanalyzer, SAN Extension Tuner, SPAN, FMServer, FCPing, FCTraceroute and AAA
- Resolution: FM, Scriptable CLI, SAN Health Check, NASB, SANTap
- Prevention: SANTap, RBAC, Port-Channels, VSANs, IVR, FSPF

Enabling the SAN to Work for You

Even Though Your SAN May Contain Intelligence, Don’t Paint Yourself into a Box Through Poor Designs

- Distribute storage across multiple switches
- Create hardened, diverse paths between switches
- Provide users with enough privileges to perform their task
- Restrict ports to specific modes (E/Fx)
- Provide isolation between devices that do not need to communicate (IVR, VSAN ACLs)

Lastly, the Storage Arrays

- There may be clustered hosts, multiple SAN fabrics—but often times—only one disk array
  - Physical Redundancy: RAID, cache protection, battery backup
  - Data Protection: Snapshots, replication
- In a local disaster, your goal is to make sure the data in question on the array is easily recoverable

Know Your Environment

- Continued education
- Understand and test the features to be deployed
- Test your ability to troubleshoot the features
- Bring in resources prior to an incident to review your environment
- The difference between a quick resolution and a long outage is the ability of the administrative team to leverage all of their technology

Remote Recovery Methods

Agenda

- Introduction to “Backend” Infrastructure
- Local Recovery Methods
- Remote Recovery Methods
- MDS Building Blocks
- DR: The Game Plan
- DR: Building Your Plan

Remote Recovery

When the Local Infrastructure Can No Longer Support the Application

- The remote site may be required to either host the application, or provide the recovery mechanism to resume local operations
- In either case, the data in the remote site must be easily accessible and valid as of a known point in time

Easily Accessible?

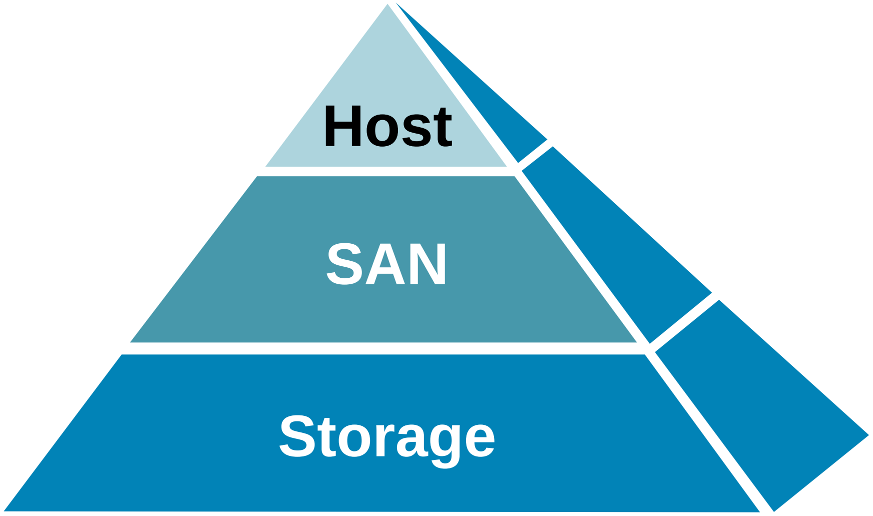
- If the data is to be used by a host in the remote site, a pre-built SAN should exist providing recovery hosts with access comparable to those in the primary site
- However, if the data is to be recalled from the remote site to the primary site, then the WAN connection needs to be of sufficient bandwidth and latency to transfer the data to meet RTO

Valid Data?

- The data volumes in the remote facility, not only must be consistent, but should be regularly checked and tested against the application
- Determine how much data you may lose if you switch over to the remote volumes

Who Enables Recovery Using the Remote Site?

- Each layer can be used, as there is no single tool for every application
- All can send and retrieve data to/from the remote site



Remote Recovery Using the Host

- Application: Replicate transactions to a remote instance
- Volume Manager: Replicate volumes to remote storage
- Operating System: Inband Continuous Data Protection (CDP) agents, rsync, ftp or even tape backups

Remote Recovery Using the Host (Cont.)

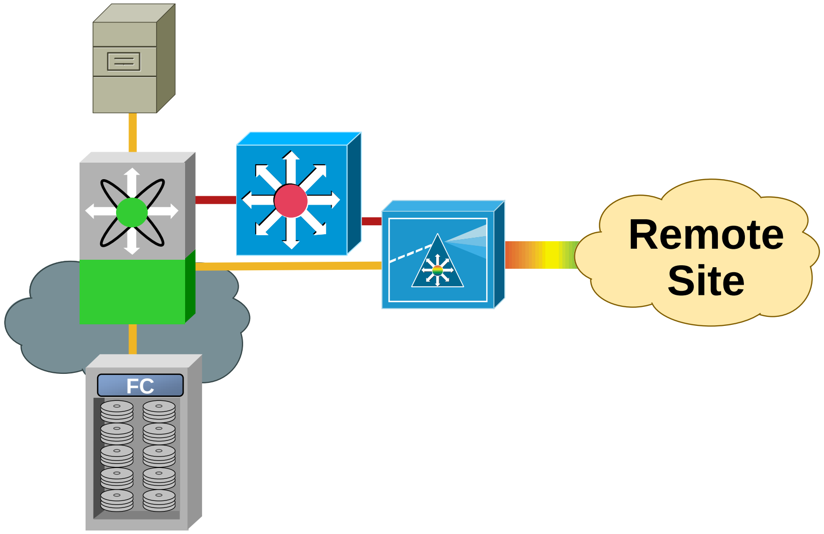
- Maximum flexibility when using the host to move the data as there are many choices
- Can lead to a lack of standardization of processes
- Heterogeneous operating systems and hardware platforms can lead to different solutions, thus adding complexity to the DR activity
- Proliferation of many different strategies based upon project, platform or department

How Can the SAN Leverage the Remote Site?

- Provides native FC or FCIP connectivity to the remote site for replication by either host or storage
- Can provide local hosts access to remotely replicated data
- Can host SANTap based CDP applications which are not tied to either host or storage
- Can host array virtualization applications which can access the remote site without having to modify the host or array

SAN Enhanced Infrastructure

- Extend Fiber Channel (FC) based applications over IP based WAN circuits
- ISL consolidation and aggregation to increase utilization
- Any SAN attached platform can leverage the provided services
- Operates at the at the FC and SCSI layers so is independent of host or array type



How Can the Array Provide Remote Recovery?

- Array based replication
- Offloaded and independent from the host
- Can be done synchronously, or asynchronously
- Currently the most established form of long distance replication. Done at the track level

Array Replication

So, If Array Replication Has Been the Standard, Why Not Continue Down This Path?

- Cannot use mixed array types
- Cannot replicate between different vendors
- Still dependent on a SAN to provide WAN connectivity
- Host operations done at the block level, such as defragging filesystems, can wreak havoc with replication. Although the files didn’t change, many blocks have, resulting in the array performing a full sync
- Data corruption is automatically replicated to the remote site, and cannot be undone, unless additional protection (snapshots) are used

The Human Factor

- Configure devices to protect themselves against human errors
- Enable access controls, accounting and remote logging on all devices
- Document everything
- Disasters aren’t only created by fires, floods or tornados, users ‘fat fingering’ keyboards cause them, too

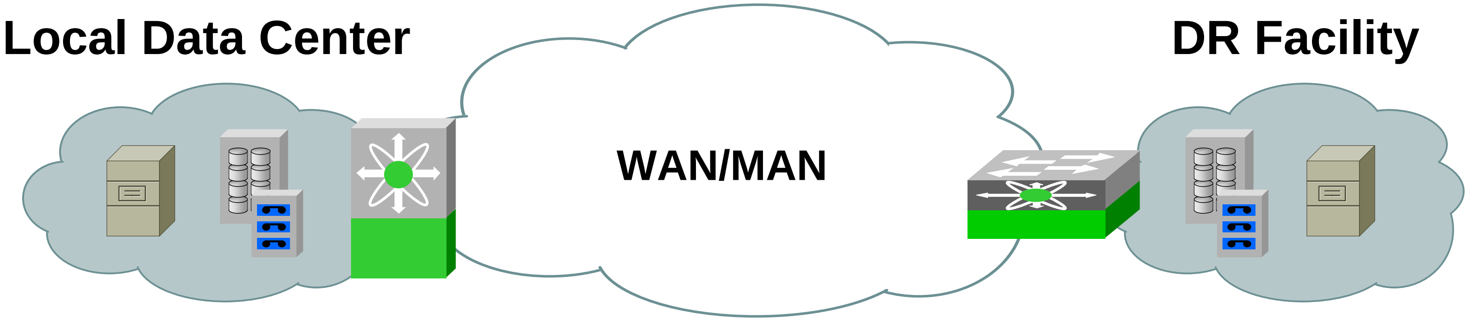
MDS Building Blocks

Agenda

- Introduction to “Backend” Infrastructure
- Local Recovery Methods
- Remote Recovery Methods
- MDS Building Blocks
- DR: The Game Plan
- DR: Building Your Plan

FCIP Primer

- Acts as a transport layer for both FCP (SCSI) and FICON traffic
- Independent of layers above (tape or disk IO), or below, requiring only IP
- Builds an ISL between two switches which provides the same functionality as a FC based ISL
  - Port Channeling for increased bandwidth and resiliency
  - TE port for trunking multiple VSANs to the remote facility



FCIP over FC?

- FCIP does not require dedicated FC-based infrastructure

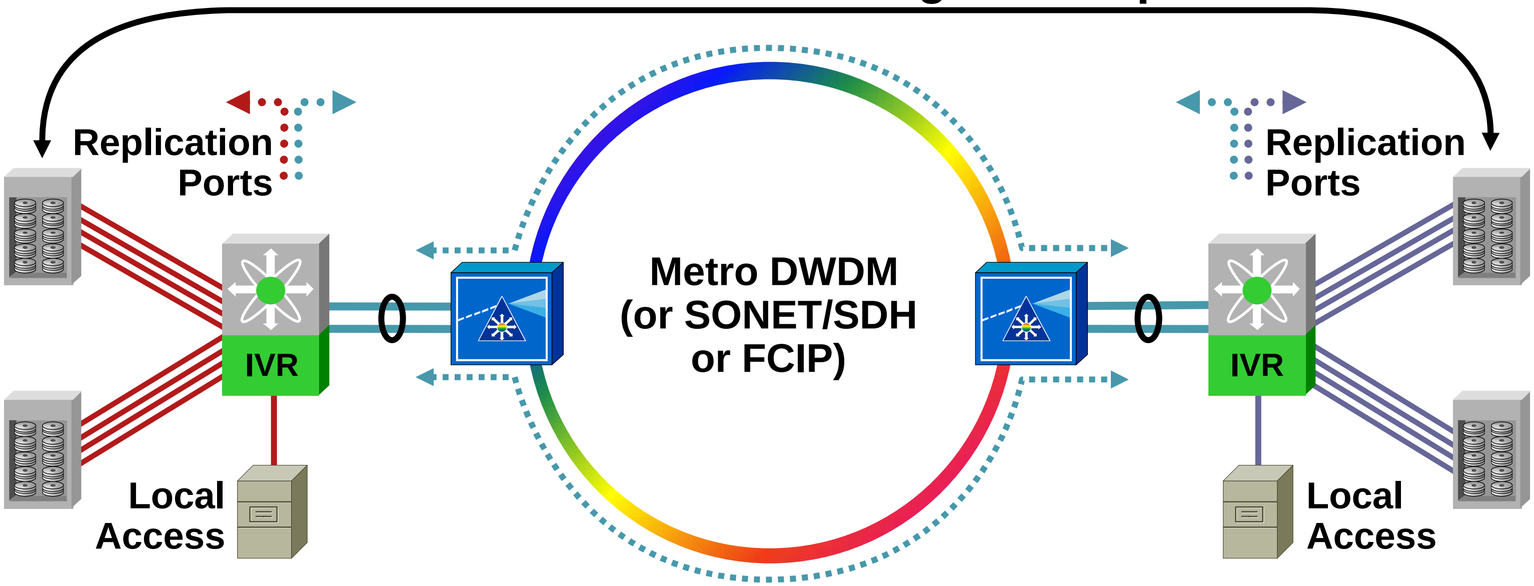
  A single 10Gigabit Ethernet circuit could be provisioned to carry both SAN traffic using FCIP and other IP based traffic such as telephony or web

- FCIP can optimize the higher level protocols (SCSI and FICON) to alleviate the latency of longer distances
- Using encryption, the data is protected while traversing foreign networks

InterVSAN Routing (IVR)

- Enables devices in different VSANs to communicate
- Enable replication while preventing local devices or VSANs from traversing WAN

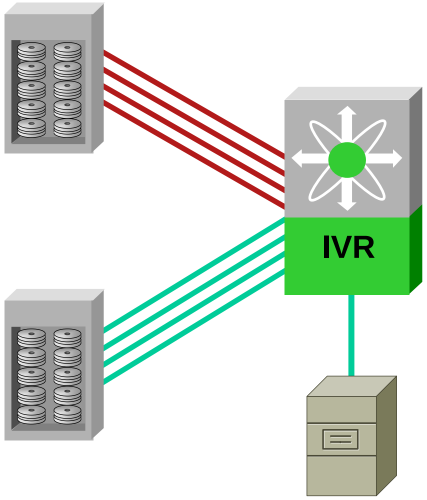
Inter-VSAN Connection Enabling Disk Replication



When Is IVR Used?

When a Single Device, Must Access Devices in Other VSANs

- A disk array port may be providing primary storage to a local host, while replicating volumes to a remote site
- A host, performing host based mirroring, is accessing local disk as well as remote disk using a single host bus adapter (HBA)



Green Host Accessing Both Red and Green VSAN-Based Storage

IVR Basics

- IVR Topology: The list of VSANs that are eligible to be routed between
- Transit VSAN: An intermediary VSAN between two IVR enabled switches used to carry interVSAN traffic. Can contain:

    End Devices

    3rd Party Switches

    MDSs not running IVR

IVR Zoning

- IVR Zone: A container or access control, containing two or more devices in different VSANs

  Standard zones are still used to provide intraVSAN access

- IVR Zoneset: A collection of IVR zones that must be activated to be operational

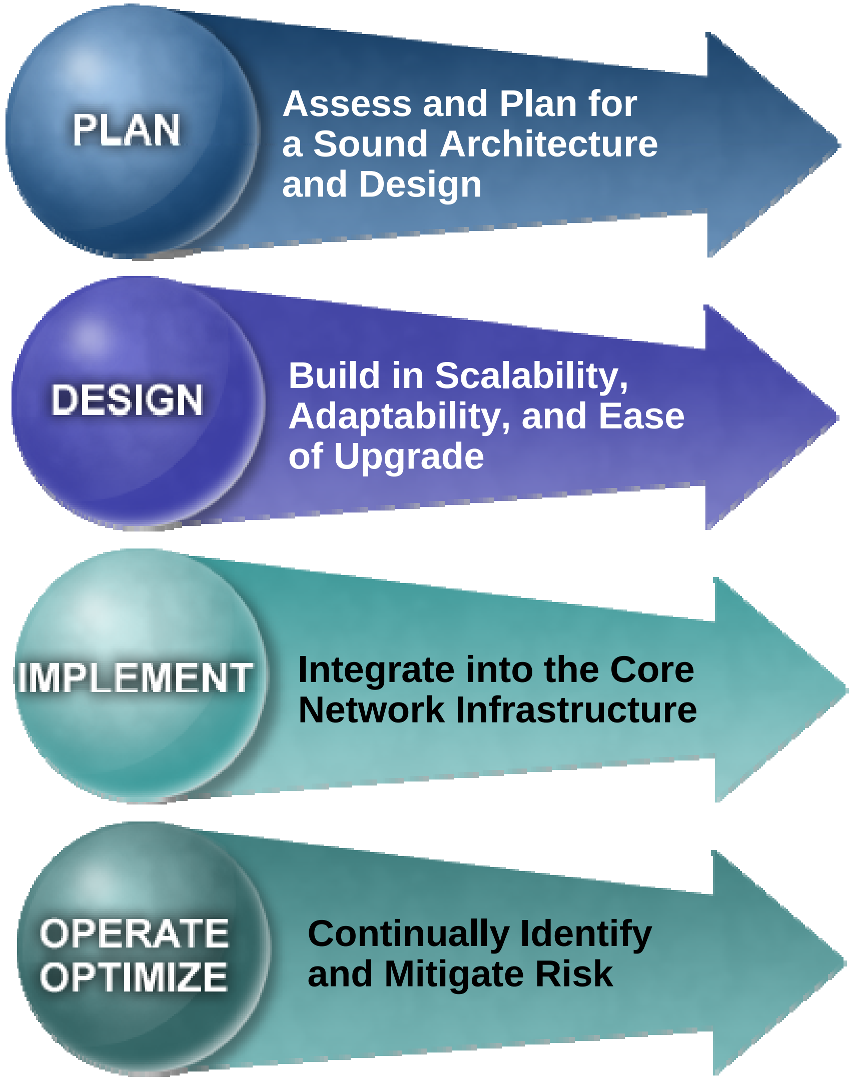
For More Information on IVR, See:

DCT-3008: Advanced SAN Fabric and Storage Virtualisation

Agenda

- Introduction to “Backend” Infrastructure
- Local Recovery Methods
- Remote Recovery Methods
- MDS Building Blocks
- DR: The Game Plan
- DR: Building Your Plan

Measure Twice, Cut Once



PLAN

Assess and Plan for a Sound Architecture and Design

- Determine both Business and Technology Requirements
- Evaluate RPO and RTO vs Cost

DESIGN

Build in Scalability, Adaptability, and Ease of Upgrade

- Design solution to meet requirements
- Determine actions for personnel to perform in case of disaster

IMPLEMENT

Integrate into the Core Network Infrastructure

- Build solution
- Test all layers of infrastructure as it is built

OPERATE OPTIMIZE

Continually Identify and Mitigate Risk

- Modify infrastructure to accommodate new projects and applications
- Tune processes and procedures as new functionality is added
- Practice “fire drills” to rehearse recovery

Phase 1
Plan

- Identify Business Requirements for Disaster Recovery:
  - Cost of downtime?
  - Federal Regulations
  - Different Departments all requiring recovery
- Technical Requirements:
  - Multiple Host Platforms
  - Scalable for capacity growth over a period of time
  - Easy to manage during a crisis
  - Identifying tools to provide different recovery methods

Phase 2
Design

- Determine components to be used and how they will fit into your overall DR strategy
- Creation of technical infrastructure and how the pieces will fit together
- Determine how existing infrastructure as well as not new will be integrated
- Creation of the processes and procedures that will guide personnel in how the infrastructure is to be used

Phase 3
Implement

- Based upon the Design Phase, with minimal impact and change management, build the infrastructure
- Update documentation
- Test infrastructure
  - An outage is not when a DR process or technology should be validated
  - Testing validates the design against both business and technical requirements
- Perform proper handoff to operational personnel

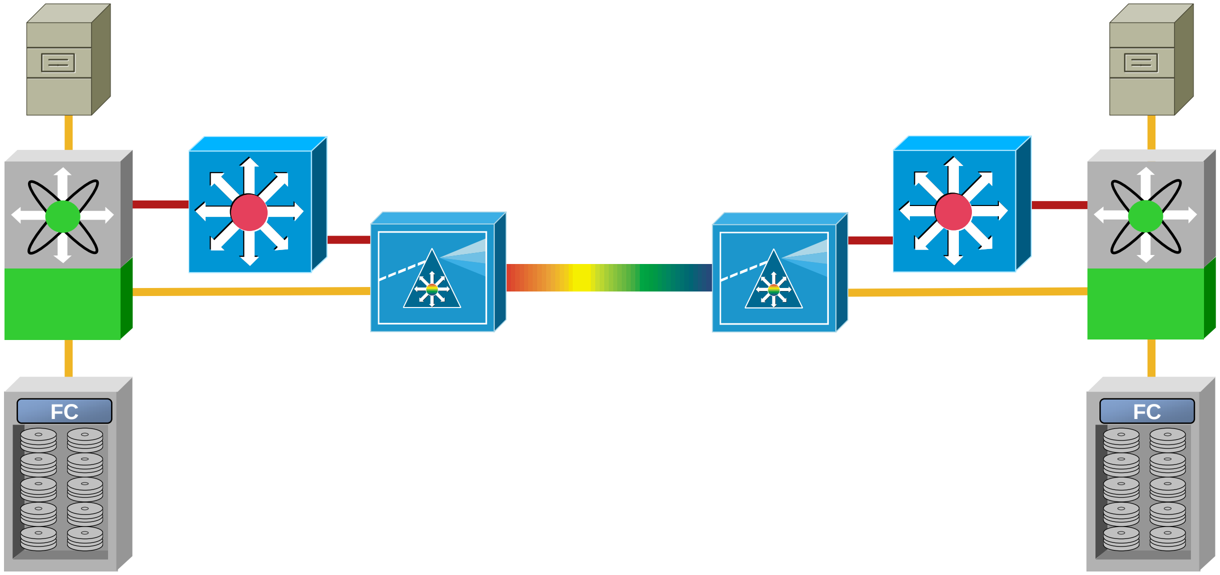
Phase 4
Operate and Optimize

- Monitor resources against predetermined metrics such as performance and utilization
- Perform recovery drills whereby replicated data is validated
- Verify not just one component, but the application, middleware and backend database need to be tested together
- Take lessons learned during this phase as input into future designs

Agenda

- Introduction to “Backend” Infrastructure
- Local Recovery Methods
- Remote Recovery Methods
- MDS Building Blocks
- DR: The Game Plan
- DR: Building Your Plan

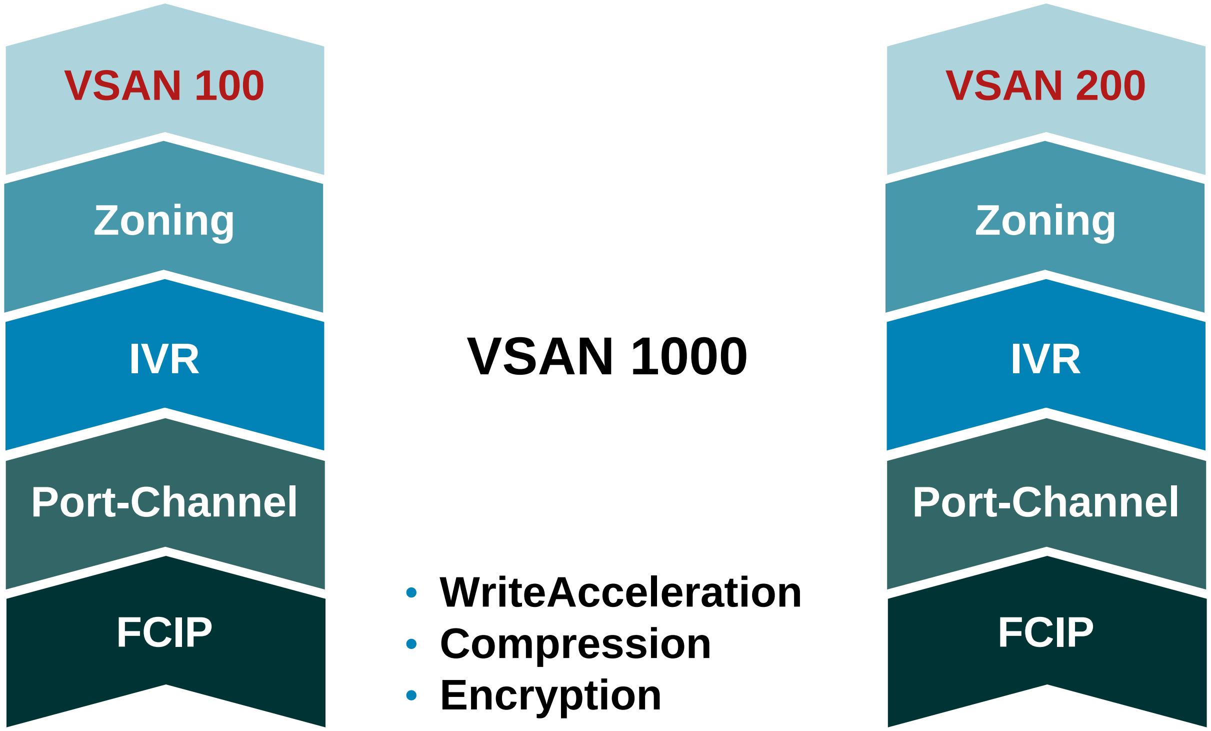
Physical Topology



- Only the first view of the infrastructure
- Doesn’t tell you how the devices are configured, just what equipment you have and connectivity
- Always based upon business and technical requirements

MDS Technology Hierarchy

Basic Infrastructure



- WriteAcceleration
- Compression
- Encryption

- Understand how the IO will flow from Primary to Remote site
- Helps you determine “why one device cannot communicate to another”
- Implement services from the “Bottom Up”

VSANs and Zones for DR?

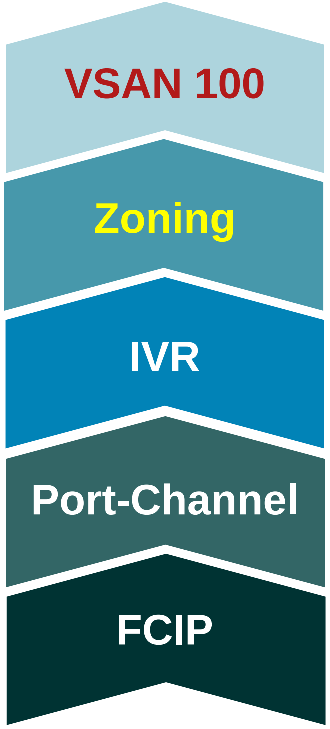
- VSAN: Provide isolation for devices and limit failure domains

  Provide ability to isolate primary from remote sites. Eliminates polluting a recovery method

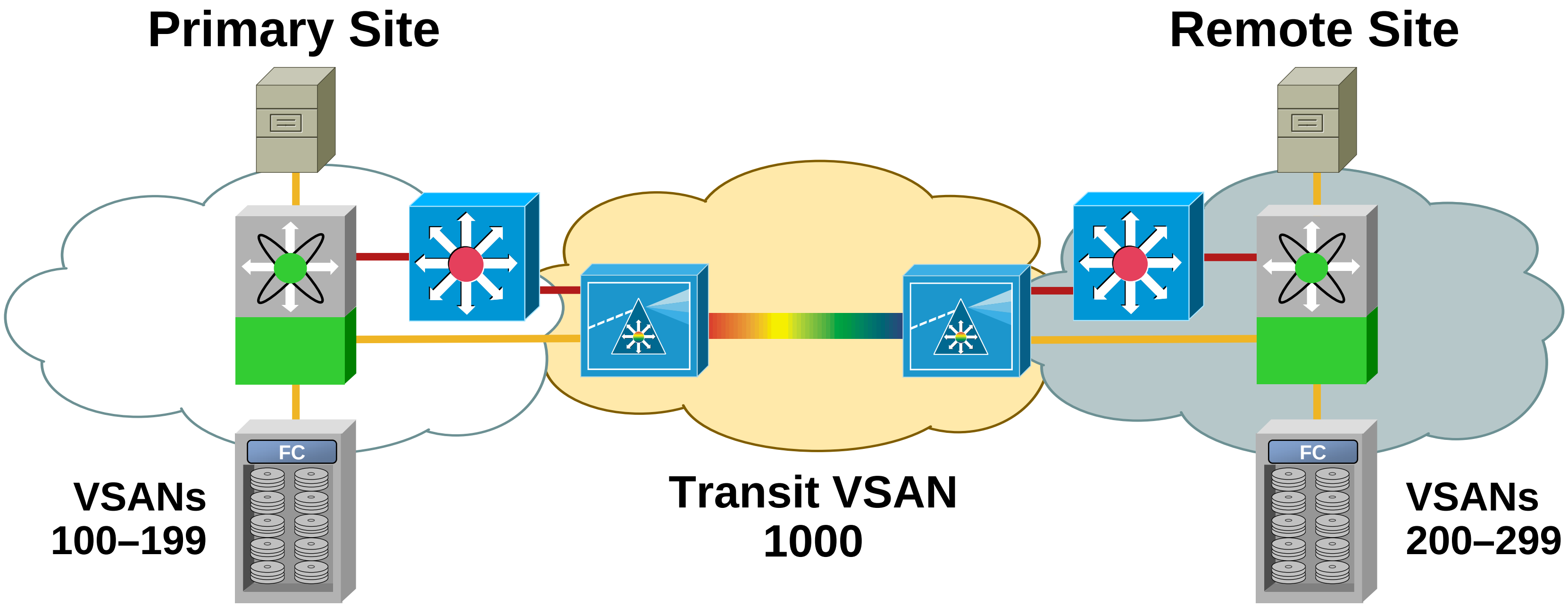
  Can represent different classes of recovery

  Can contain all the SAN devices representing an application stack (Web, Middleware, Database)

- Zoning: Limits host/storage access within a VSAN



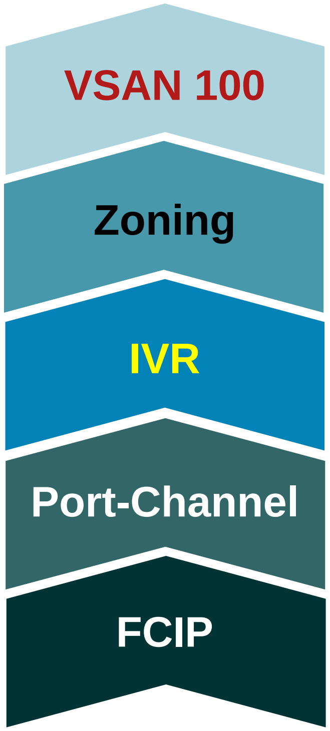
VSAN Topology



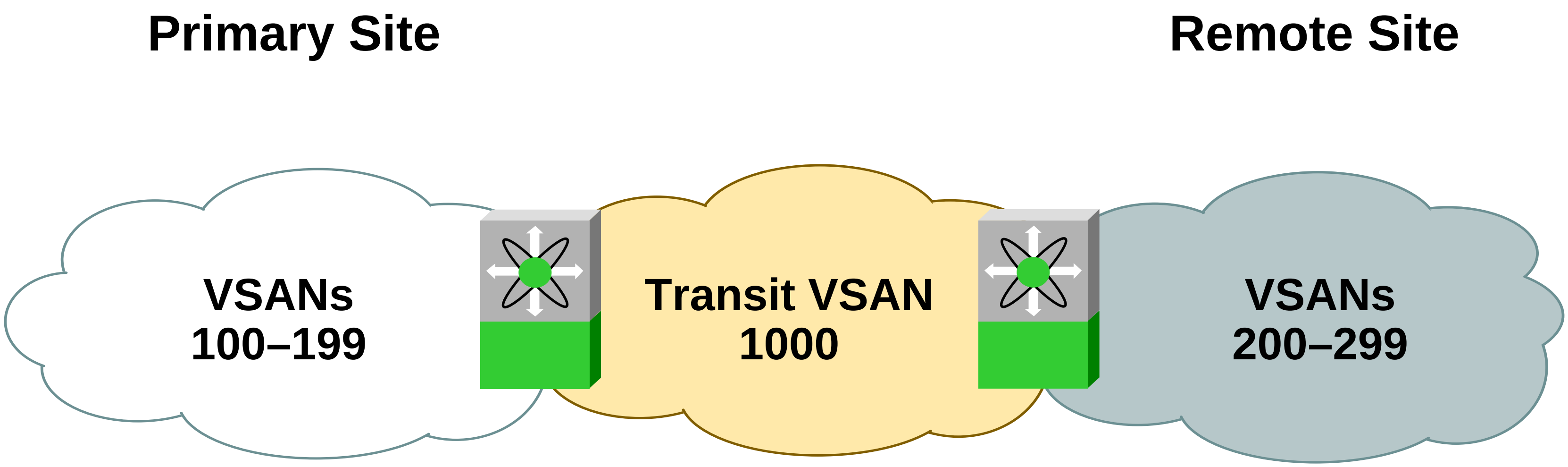
- Assign ranges of VSANs for future growth
- Provide ample room to prevent overlap
- Transit VSAN isolates Primary from Remote site

IVR
Adding Resiliency

- IVR: Enable hosts or storage arrays to access their replication peer
- Using IVR Network Address Translation (NAT) increases the scalability of the solution
- Transit VSANs ensure local and remote VSANs do not share resources, including switches
- Service Groups provide further isolation and enable different VSANs to use different transit VSANs

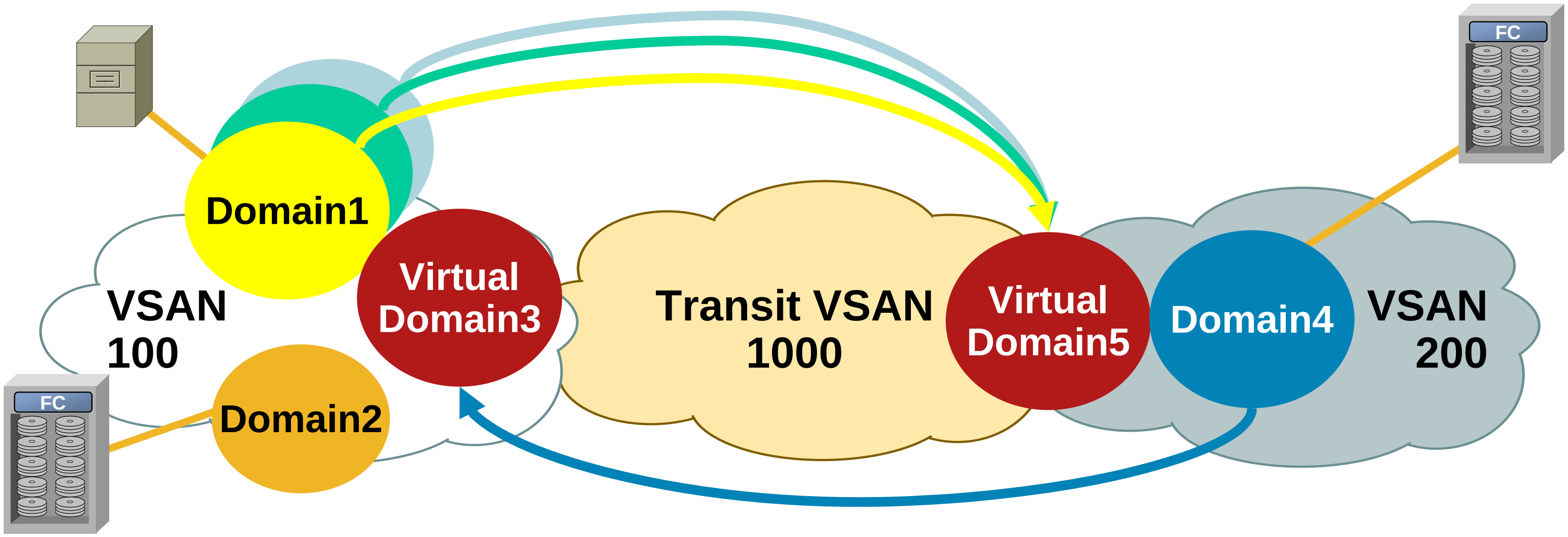


IVR View



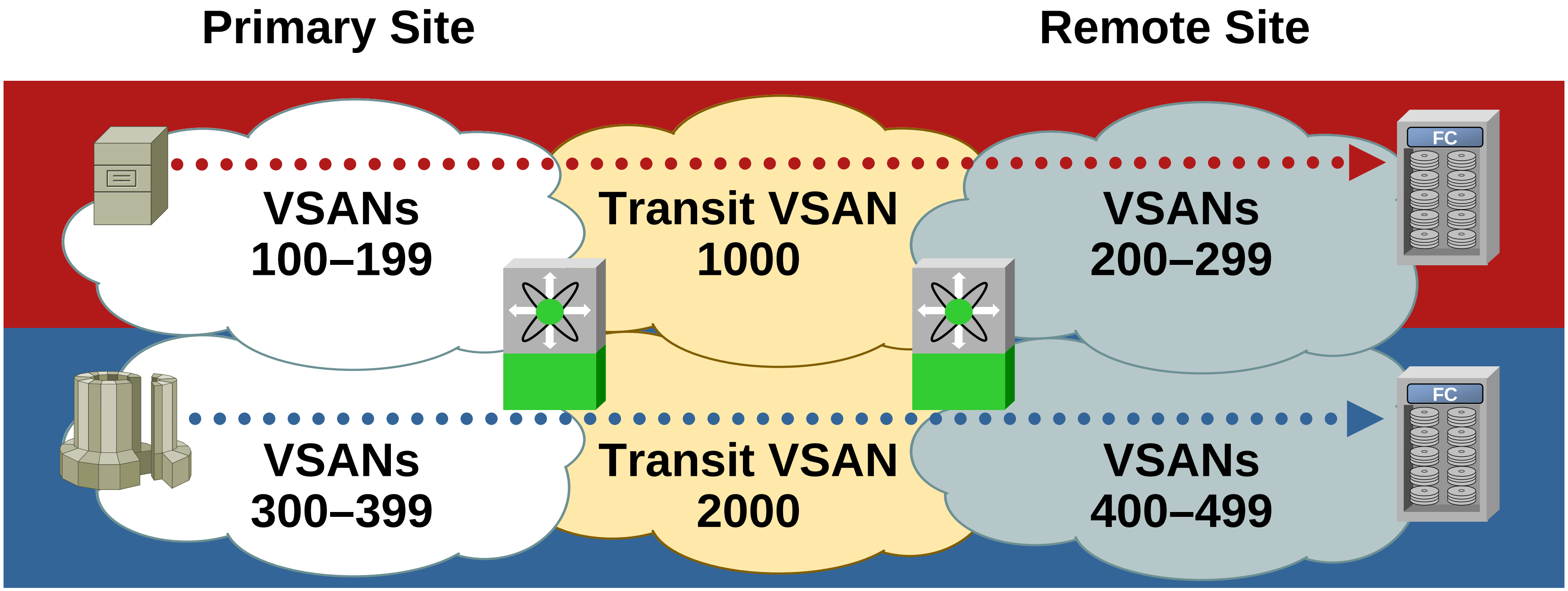
- Assign ranges of VSANs for future growth
- Provide ample room to prevent overlap
- Transit VSAN isolates Primary from Remote site

IVR with Network Address Translation



- NAT enables one virtual domain (5) to represent an entire VSAN and all of its domains
- Enables duplicate domainIDs within a fabric
- Can be used to provide connectivity for legacy fabrics to the remote site
- Transit VSAN isolates Primary from Remote site

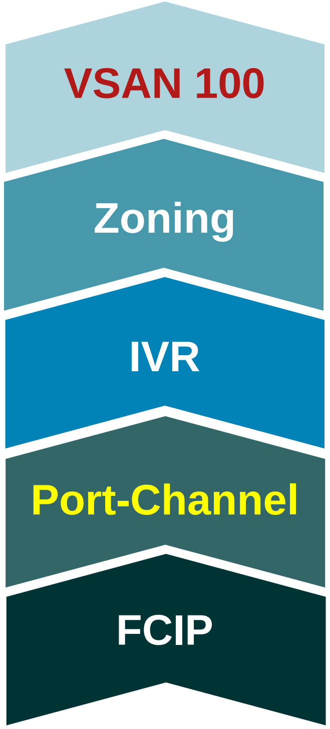
IVR Service Groups



- Divides up the IVR Topology into “sub-topologies”
- Limits IVR events to a single service group
- Enables the use of different transit VSANs per service group
- “Gold,” “Silver,” and “Bronze” transit VSANs

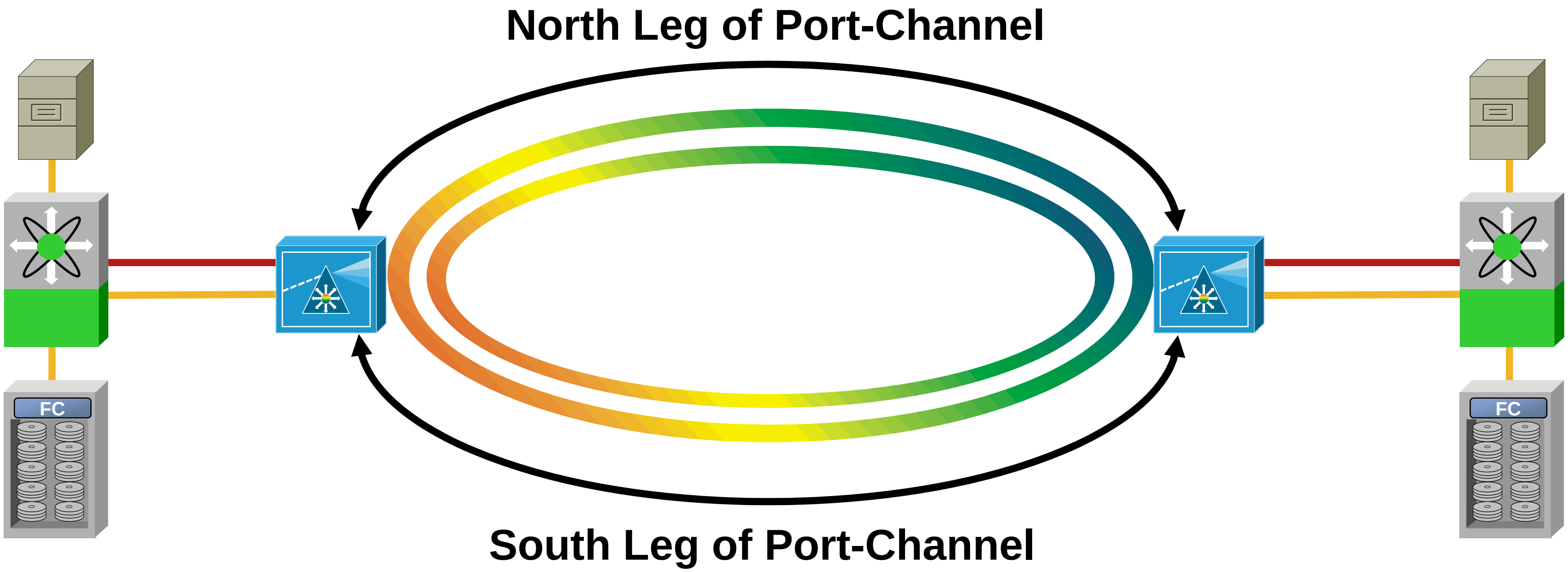
Port-Channels
Maintaining Connectivity

- Ability to load-balance traffic across multiple WAN circuits
- Consolidates multiple ISLs, into a single management object
- Independent of transport layer (FCIP, FibreChannel over Optical)
- Can trunk one or more VSANs to the remote facility carrying both FCP and FICON



Port-Channels

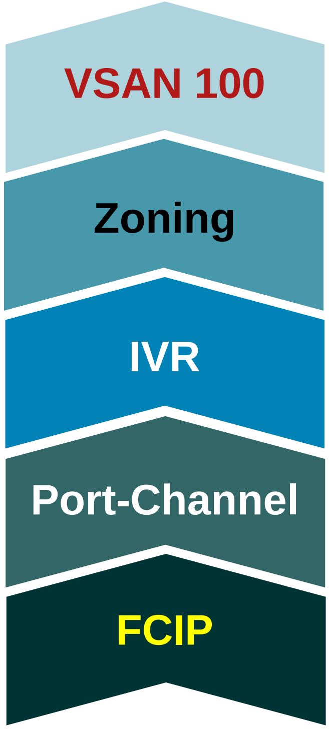
ISL Resiliency



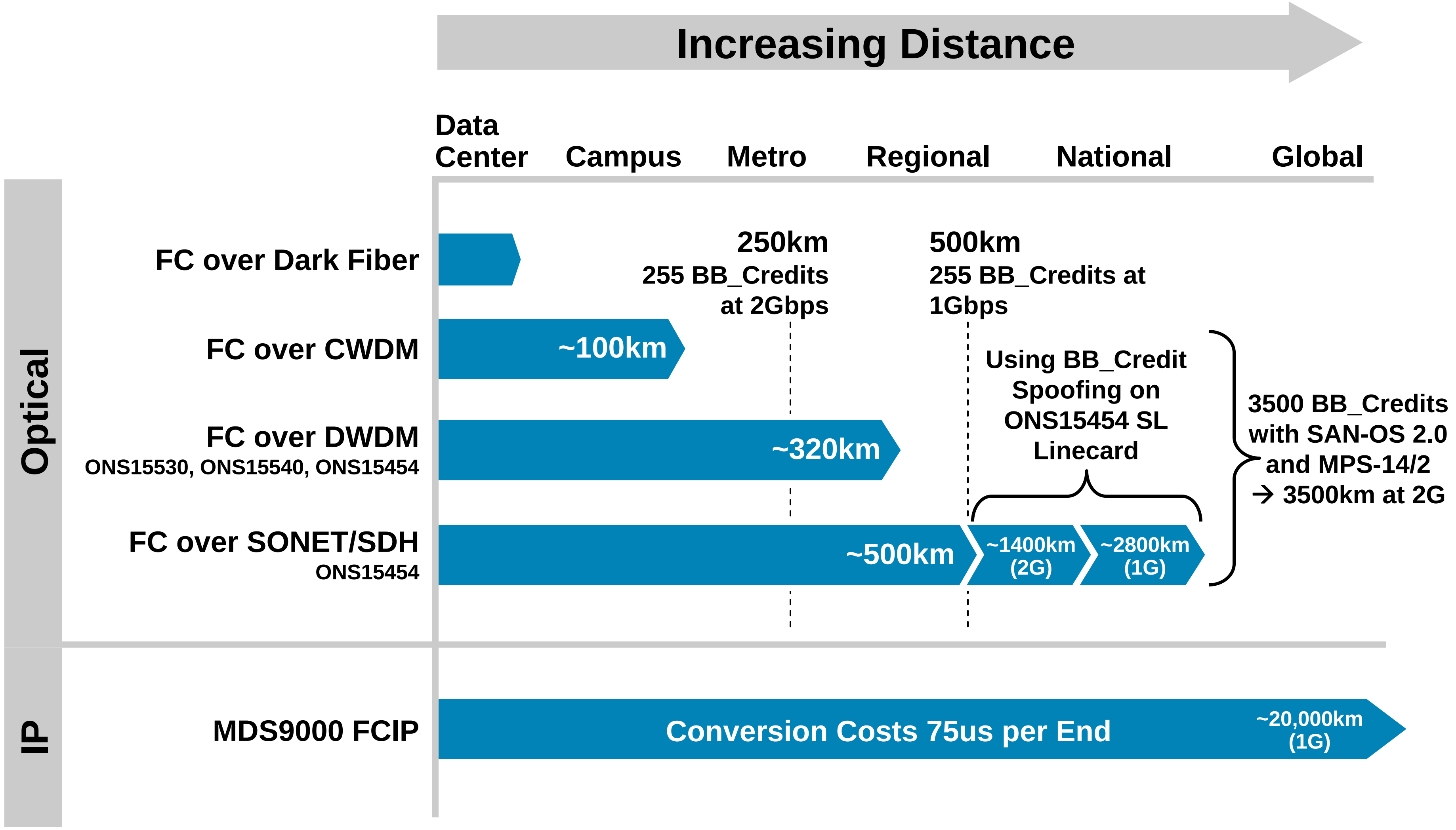
- Maintains switch connect even when members go down
- Can non-disruptively increase membership as bandwidth requirements scale to accommodate new DR projects
- FCIP and optical based port-channels are managed exactly the same

FCIP

- Cost effective long distance connectivity
- Common IP infrastructure
- Adaptive Compression, leverage smaller circuits between sites
- Write and tape acceleration, enable DR site to be located farther away. Synchronous replication over longer distances
- Encryption, protect data in flight



Determine Transport for Site Connectivity



Increasing Distance

Data Center | Campus | Metro | Regional | National | Global

Optical

FC over Dark Fiber

FC over CWDM
~100km

FC over DWDM
ONS15530, ONS15540, ONS15454
~320km

FC over SONET/SDH
ONS15454
~500km
~1400km (2G)
~2800km (1G)

250km
255 BB_Credits at 2Gbps

500km
255 BB_Credits at 1Gbps

Using BB_Credit Spoofing on ONS15454 SL Linecard

3500 BB_Credits with SAN-OS 2.0 and MPS-14/2
→ 3500km at 2G

IP

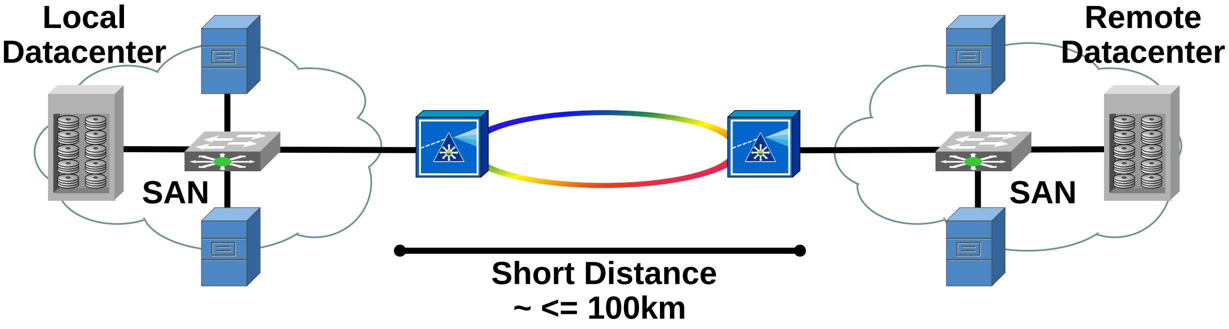
MDS9000 FCIP
Conversion Costs 75us per End
~20,000km (1G)

Multiprotocol Support

SAN Extension, IP SAN Extension

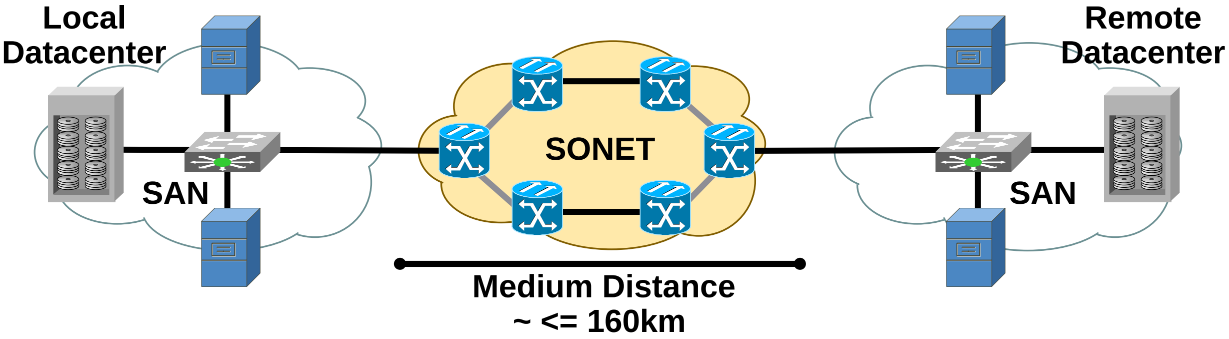
FC over DWDM/CWDM

- Short distance
- Dark fiber available
- Dedicated links
- Lowest latency—suitable for sync apps



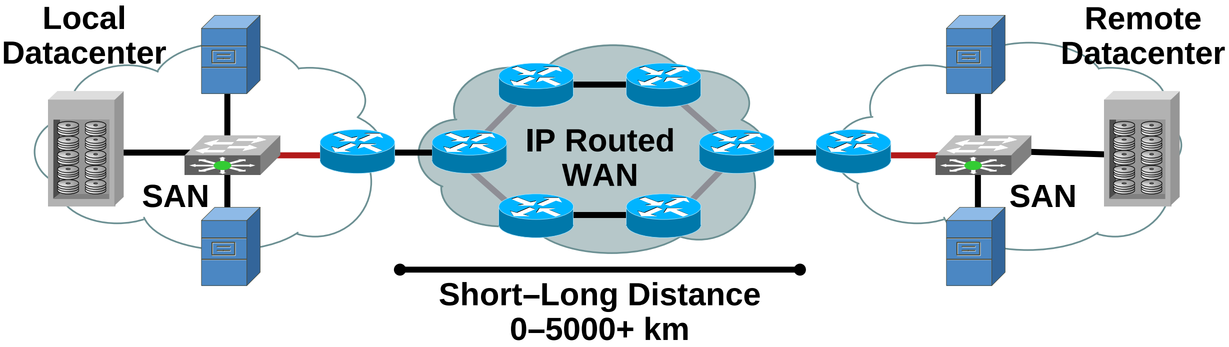
FC over SONET/SDH

- Short–intermediate distance
- Dark fiber not available—distance, cost, exhaust
- Links may be shared
- Suitable for most synchronous apps



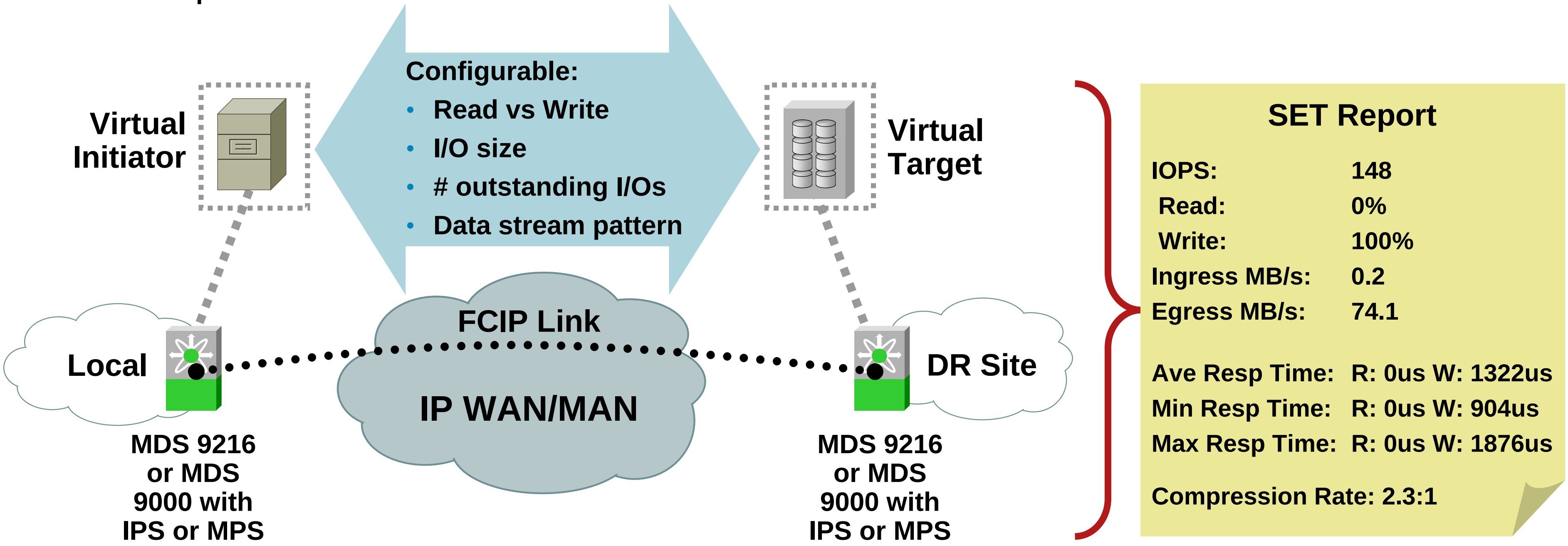
FC and FICON over IP

- Short–long distance
- Dark fiber not available
- Links may be shared
- Suitable for sync apps across metro Ethernet
- Suitable for async applications across WAN



Stop
Test Your FCIP

- Before adding additional services or features to the environment, use SAN Extension Tuner to validate WAN performance
- Baseline the configuration prior to running actual loads across
- Provide instant feedback for FCIP tuning, by simulating IO patterns of replication methods



Configurable:
- Read vs Write
- I/O size
- # outstanding I/Os
- Data stream pattern

Virtual Initiator
Virtual Target
Local
FCIP Link
IP WAN/MAN
DR Site
MDS 9216 or MDS 9000 with IPS or MPS
MDS 9216 or MDS 9000 with IPS or MPS

SET Report

IOPS:	148
Read:	0%
Write:	100%
Ingress MB/s:	0.2
Egress MB/s:	74.1
Ave Resp Time:	R: 0us W: 1322us
Min Resp Time:	R: 0us W: 904us
Max Resp Time:	R: 0us W: 1876us

Compression Rate: 2.3:1

Adding Advanced Features

- DPVM (Dynamic Port VSAN Manager), enables “WWN based VSANs” whereby a wwn logs into the same VSAN no matter which interface it is plugged into
- Device Aliases: Configure any feature using a user defined name, rather than a wwn
- Interop VSANs: Provide access to remote facility for legacy, third party SANs
- SANTap: Let the switch replicate the data, independent of host or array source or destination

Dynamic Port VSAN Membership

Adding Resiliency

- Decrease recovery time in case of switch hardware failure. The hba/storage port can be moved to a new port without reconfiguration. VSAN is assigned to the port based upon the pwwn logging in
- Reduce escalation time. Operation personnel just move the cable to an available port. No need to modify the switch’s configuration

Device Aliases

Keep It Simple

Reduce Recovery Time

Switch displays plain text name of hba/storage port instead of just the cryptic pwwn

```
VSAN 1000:
-----------------------------------------------------------------------
---
FCID    TYPE PWWN          (VENDOR)    FC4-TYPE:FEATURE
-----------------------------------------------------------------------
---
0x7f0004  N   10:00:00:00:c9:34:a5:be (Emulex)
         [ca-aix2_fcs0]
0x7f0006  N   10:00:00:00:c9:34:a5:94 (Emulex)
         [ca-aix2_fcs1]
0x7f0009  N   10:00:00:00:c9:34:a8:2a (Emulex)
         [ca-aix3_fcs1]
0xec0003  N   10:00:00:00:c9:34:a8:4e (Emulex)
         [ca-aix3_fcs0]
```

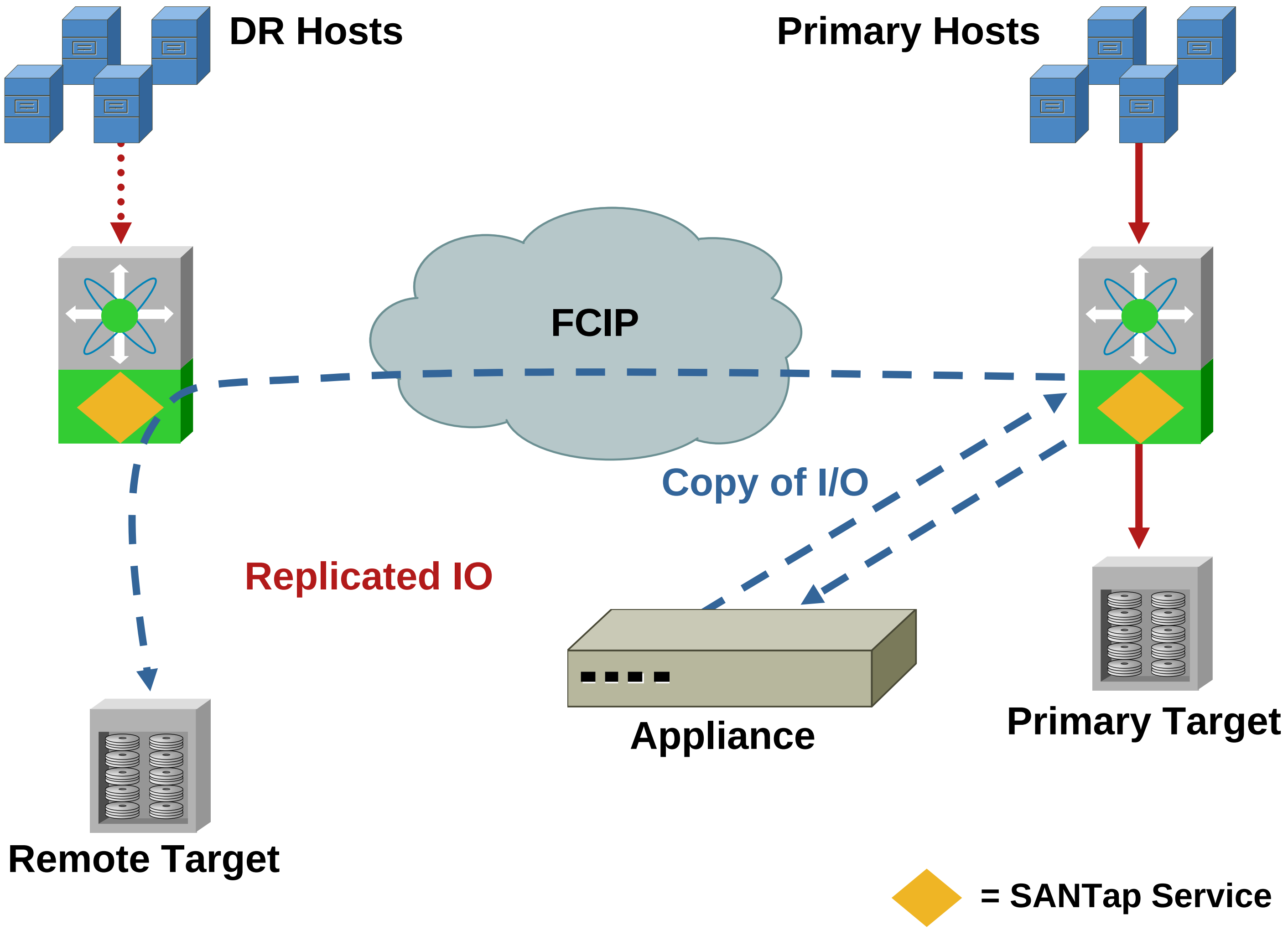
Interop
Provide DR for Legacy Islands

- Connect to legacy environments without having to migrate them to an all MDS environment
- Older switches can be uninstalled from the primary datacenter and reused in the remote facility
- Using IVR connect MDS, Brocade and McData fabrics via FCIP to the DR facility

SANTap

- Synchronous or Asynchronously replicate data
- Replicated data does not impact primary IO
- Can rollback LUNs to the IO before corruption took place
- Reduce standby capacity in the DR facility

SANTap

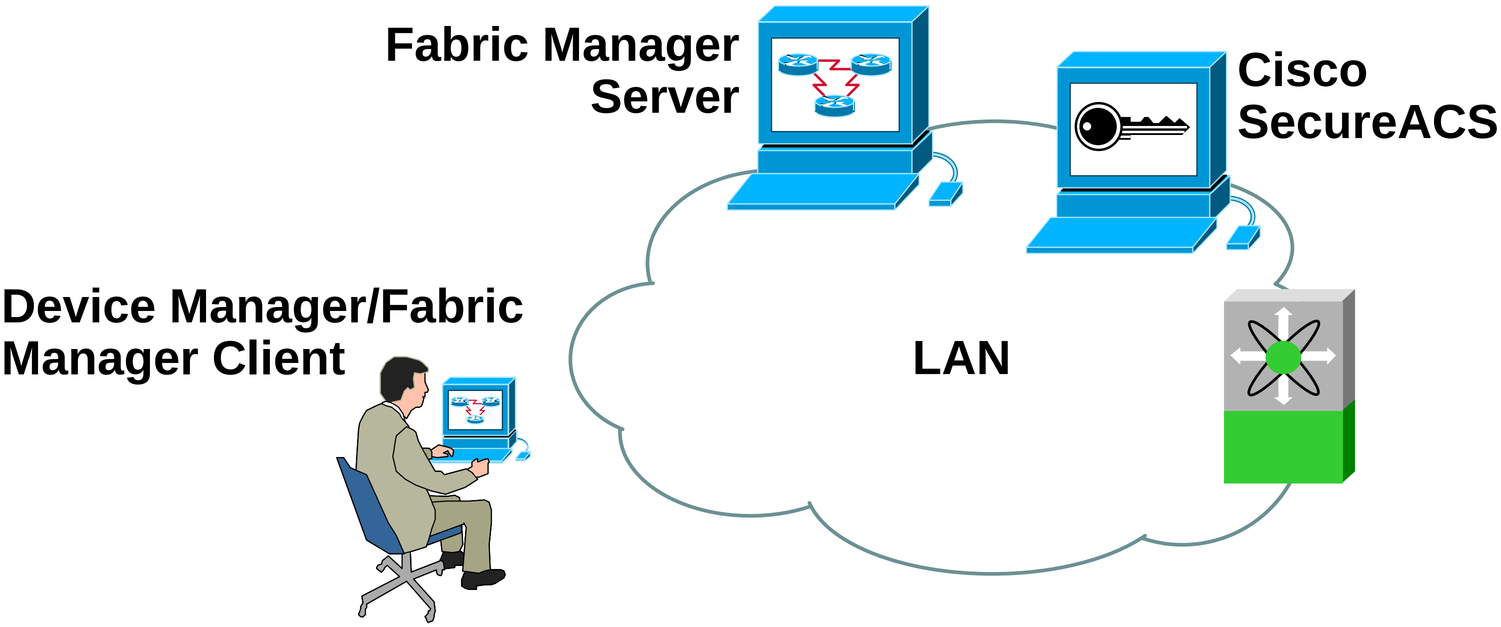

DR Hosts
Primary Hosts
FCIP
Copy of I/O
Replicated IO
Appliance
Primary Target
Remote Target
= SANTap Service

For More Information on SANTap, See:

DCT-3008: Advanced SAN Fabric and Storage Virtualisation

Management

Controlling the Disaster



- Fabric Manager Server, validate WAN usage and efficiency
- ACS for centralized user account management and accounting
- Role Based Access Controls to protect the SAN from users making accidental changes they should not be
- IP ACLs on the MDS to enhance security

Earning Your MBA in MDS

- Integrate the MDS with your NMS (Network Management System) and NOC for immediate detection of an issue via callhome and SNMP
- Automate daily backups of the switch configuration with the MDS scheduler
- Track accounting logs to find out if issues were user triggered
- Study Performance Manager reports, for over utilization trends

Key Takeaways

- Know your environment, not just the technology, but the interdependencies between applications within the datacenter
- Recovery is handled at all layers, host, switch and storage; One size does not fit all problems
- Disaster Tolerance and Recovery are not solved with just technology, but with proper processes, procedures and training
- Implement a complete, end to end solution, not a point solution

Meet the Experts
Data Centre

- Victor Moreno
  Technical Leader

Recommended Reading

BRKDCT -2004

- Continue your Networkers learning experience with further reading from Cisco Press.
- Visit the on-site Cisco company store, where the full range of Cisco Press books is available for you to browse.

CISCO SYSTEMS

Cisco Storage Networking Architectures Poster

Cisco Press

Q and A

For Easy-to-Follow Procedures and MDS Best Practices, See:

The MDS 9000 Cookbook for SAN-OS

Available on CCO