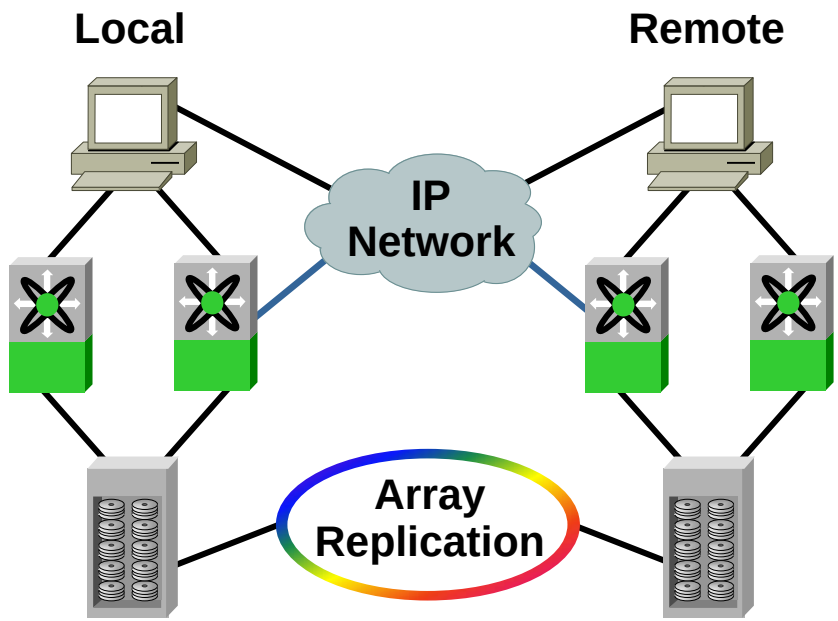
Disasters
What Components Provide Recovery?

How do the various levels help recover from a disaster without using a remote facility? Recovery can be done at each level

1. **Application performs recovery: Replay Logs, Restart Instance, Clustering done at App layer**
2. **OS performs recovery: VCS, MSCS, VxVM. Powerpath/DMP**
3. **Data: Journaling FS, Tape Restore, FS/VolMgr Snapshot**
4. **HW/Server: VMWare, Ghost Image, take faulty HW out of service (cpu/ram)**
5. **SAN: FSPF, PortChannel, Virtualization, SANTap**
6. **Array: RAID, SRDF, BCVs, Snapshots**



Site DR—Requires use of Remote Facility. Zero data loss at remote facility

1. **Application: Application transfers transactions to remote instance of application.**
2. **OS: VolMgr, "SANTAP type" driver replicates IO. Continuous backups**
3. **SAN: SANTap, Virtualization provides synchronous replication**
4. **Array: Synchronous Replication to remote array**

Replication: Modes of Operation

- **Synchronous**

  All data written to local and remote arrays before I/O is complete and acknowledged to host

  Speed of Light = 3 x 10^8m/s (Vacuum) ≈ 3.3µs/km

  Speed through Fiber ≈ ⅔ c ≈ 5µs/km

  2 RTT per write I/O = 20µs/km

- **Asynchronous**

  Write acknowledged and I/O is complete after write to local array. Changes (writes) are replicated to remote array asynchronously

Synchronous Vs. Asynchronous Trade-Off

Synchronous

- ⊖ Impact to Application Performance
- ⊖ Distance Limited (Are Both Sites within the Same Threat Radius)
- ⊕ No Data Loss

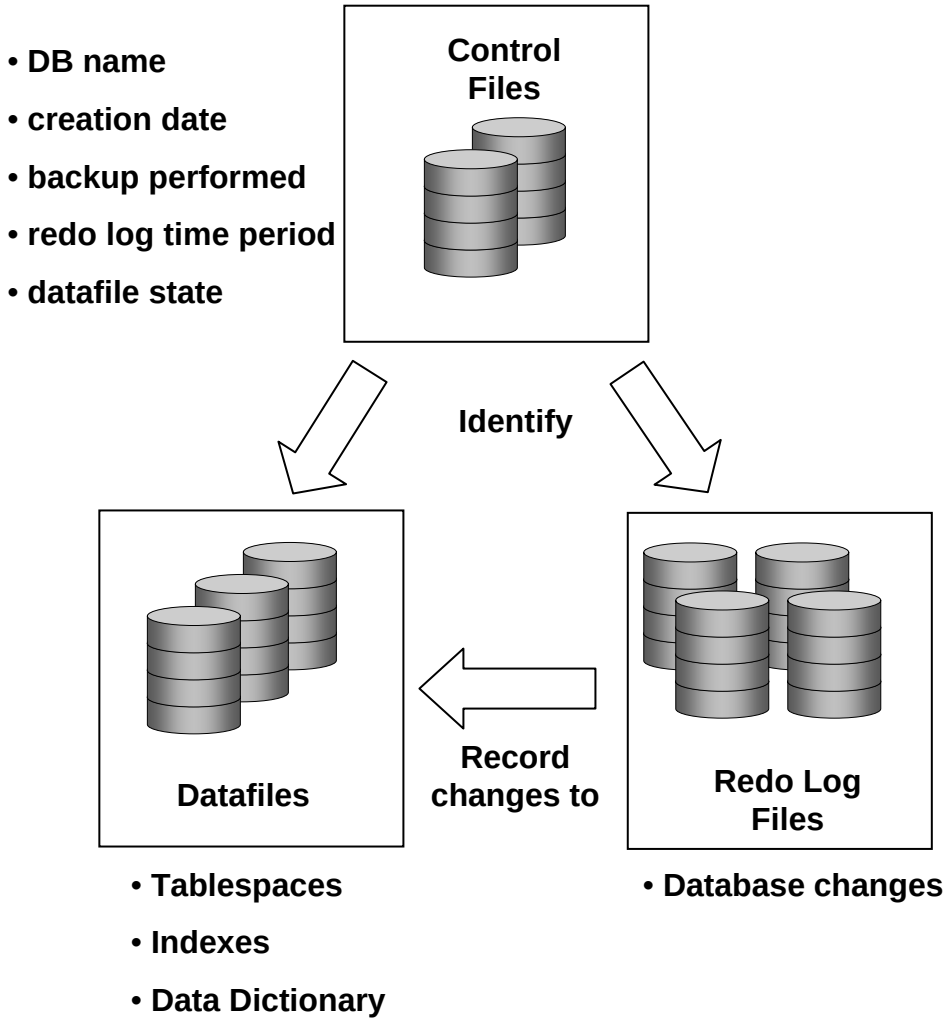
Asynchronous

- ⊕ No Application Performance Impact
- ⊕ Unlimited Distance (Second Site Outside Threat Radius)
- ⊖ Exposure to Possible Data Loss

Enterprises Must Evaluate the Trade-Offs

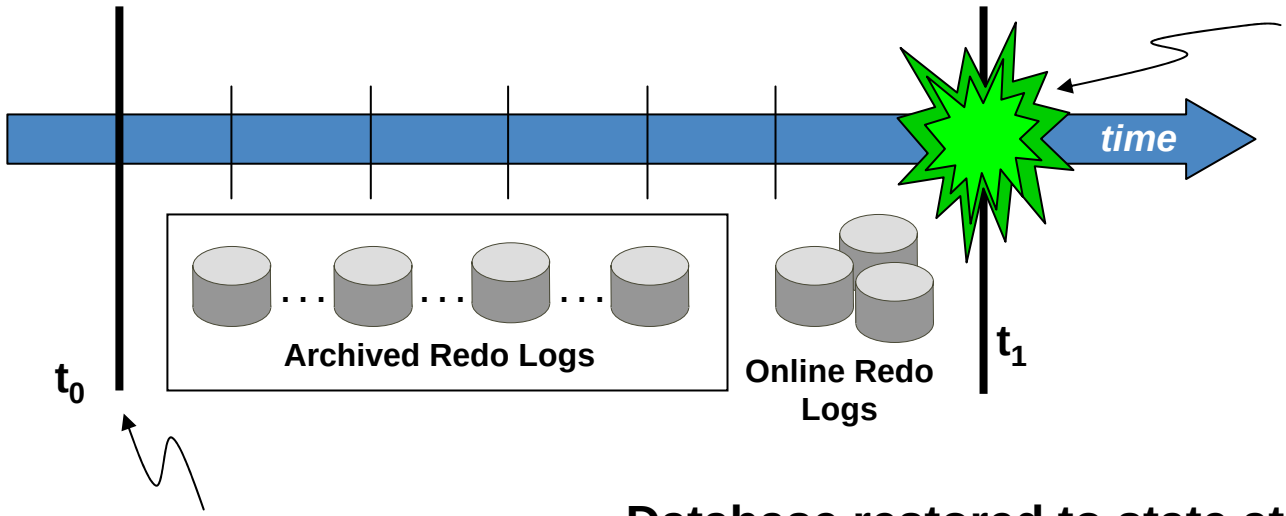
- Maximum tolerable distance ascertained by assessing each application
- Cost of data loss

Data Replication with DB Example



- **Control Files identify other files making up the database and records content and state of the db.**
- **Datafile is only updated periodically**
- **Redo logs record db changes resulting from transactions**
  - **Used to play back changes that may not have been written to datafile when failure occurred**
  - **Typically archived as they fill to local and DR site destinations**

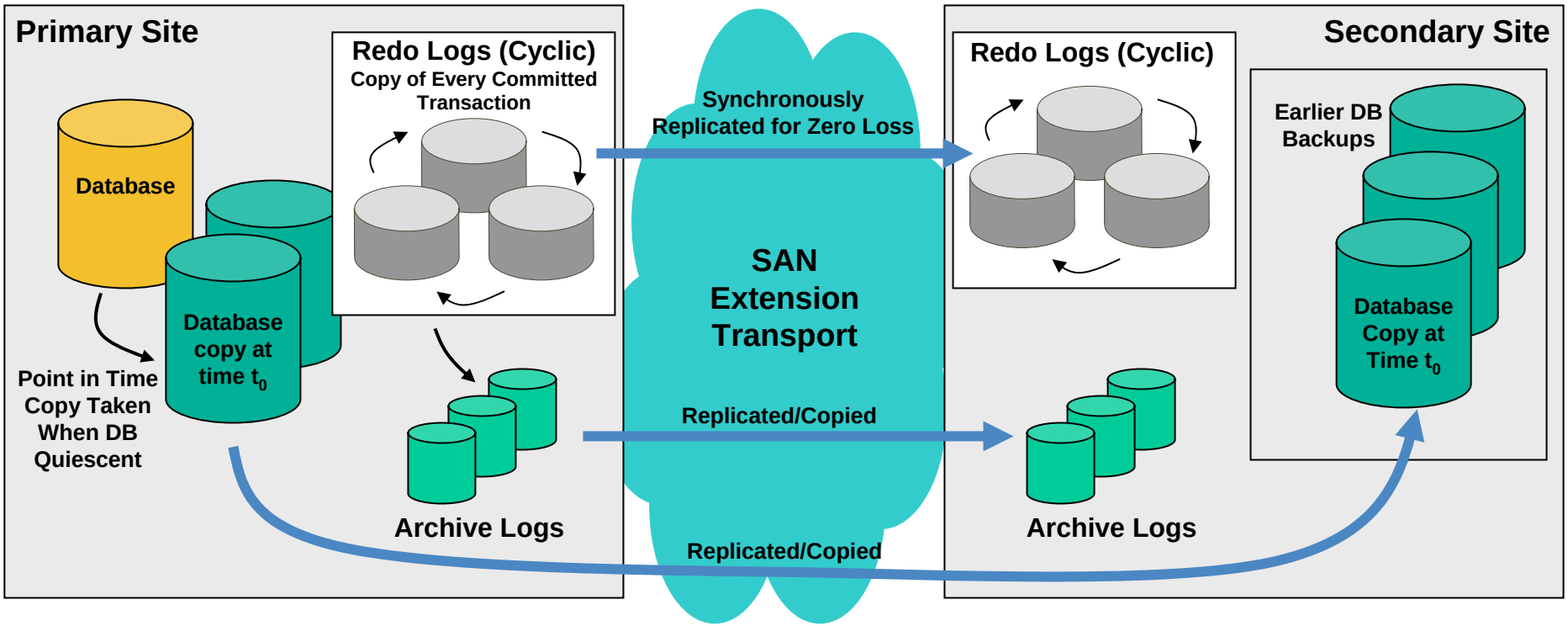
Data Replication with DB Example (Cont'd)



Failure or disaster occurs at time t_1
- **Media Failure (e.g. disk)**
- **Human Error (datafile deletion)**
- **Database Corruption**

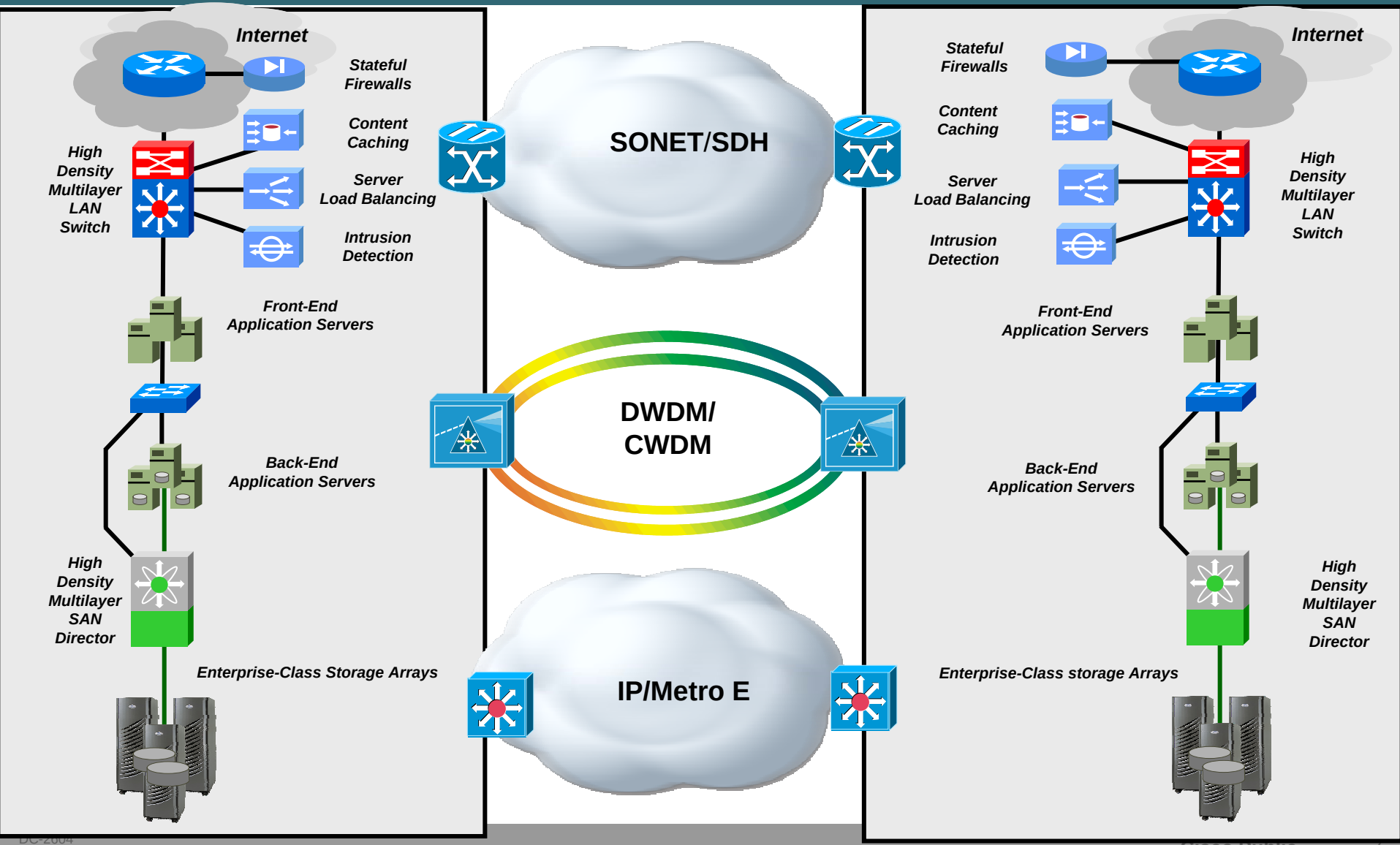
- **Database restored to state at time of failure (time t1) by:**
  1. **Restoring Control Files & Datafiles from last Hot Backup (time t0)**
  2. **Sequentially replaying changes from subsequent Redo Logs (archived and online) – changes made between time t0 and t1**

Data Replication with DB Example (Cont'd)

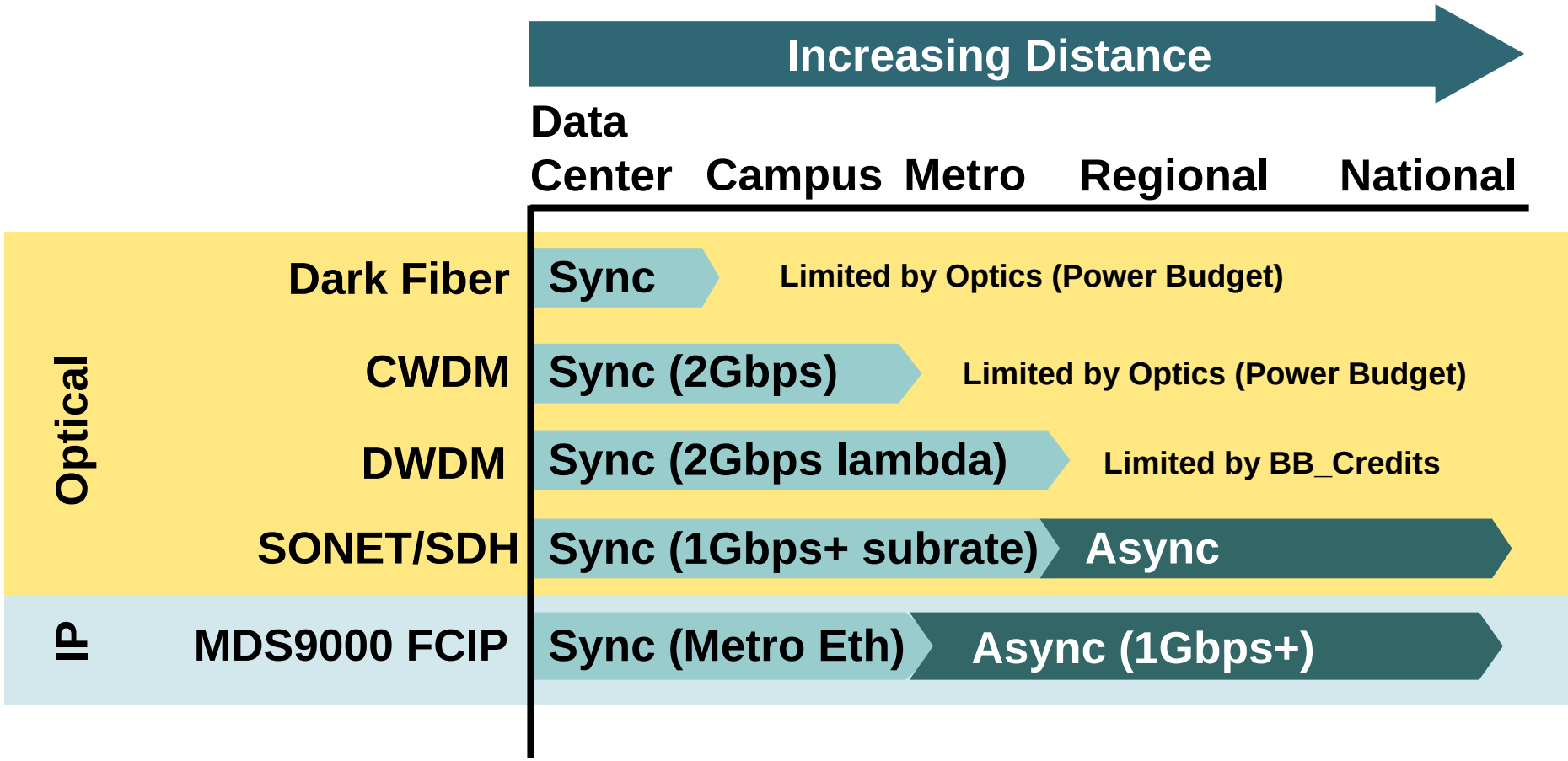


- **Mixture of sync and async replication technologies commonly used**
  - **Usually only redo logs sync replicated to remote site**
  - **Archive logs created from redo log and copied when redo log switches**
  - **Point in time (PiT) copies of datafiles and control files copied periodically (e.g. nightly)**

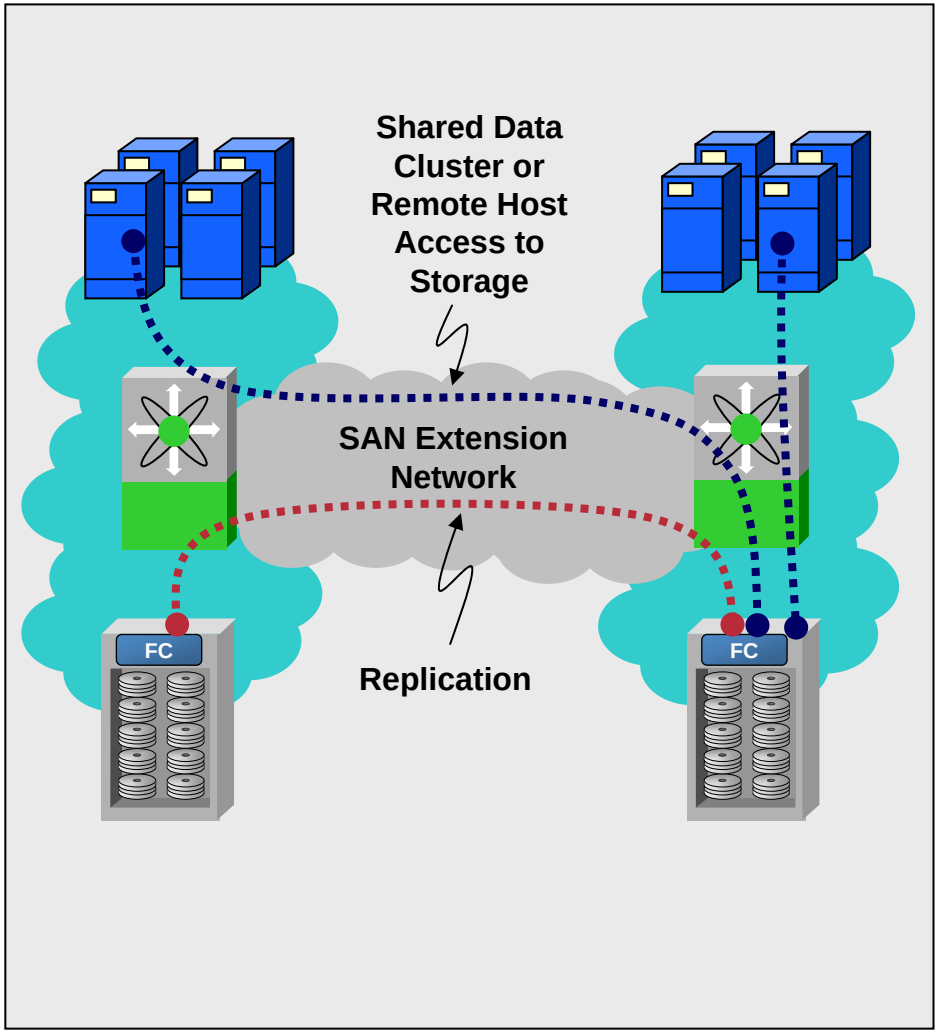
Data Center Interconnection Options



Data Center Transport Options

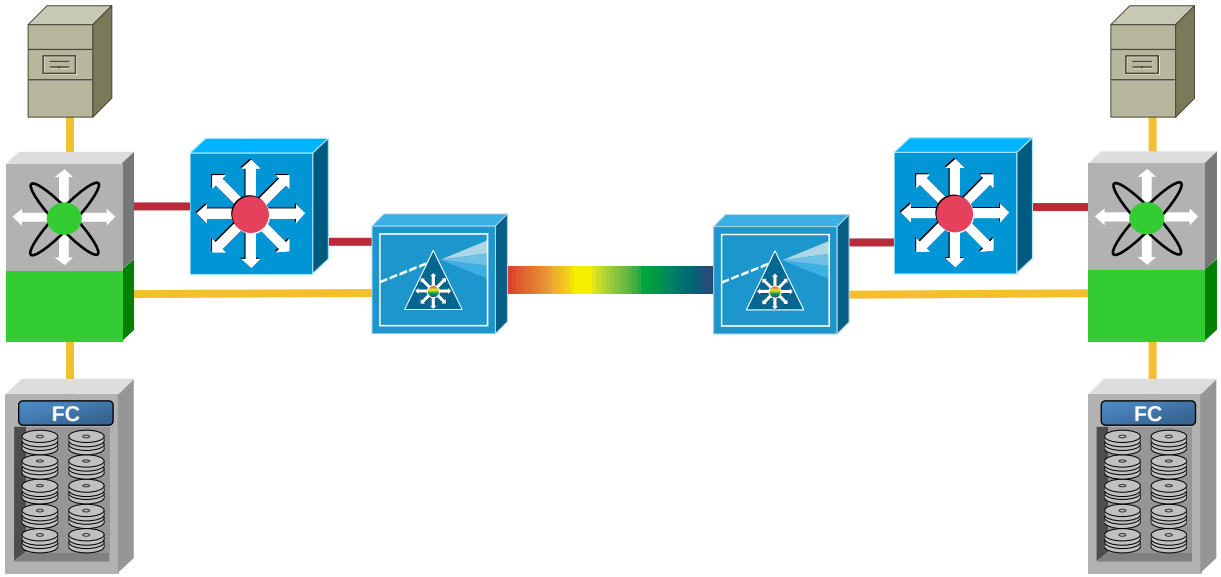


Data Center Replication with SAN Extension



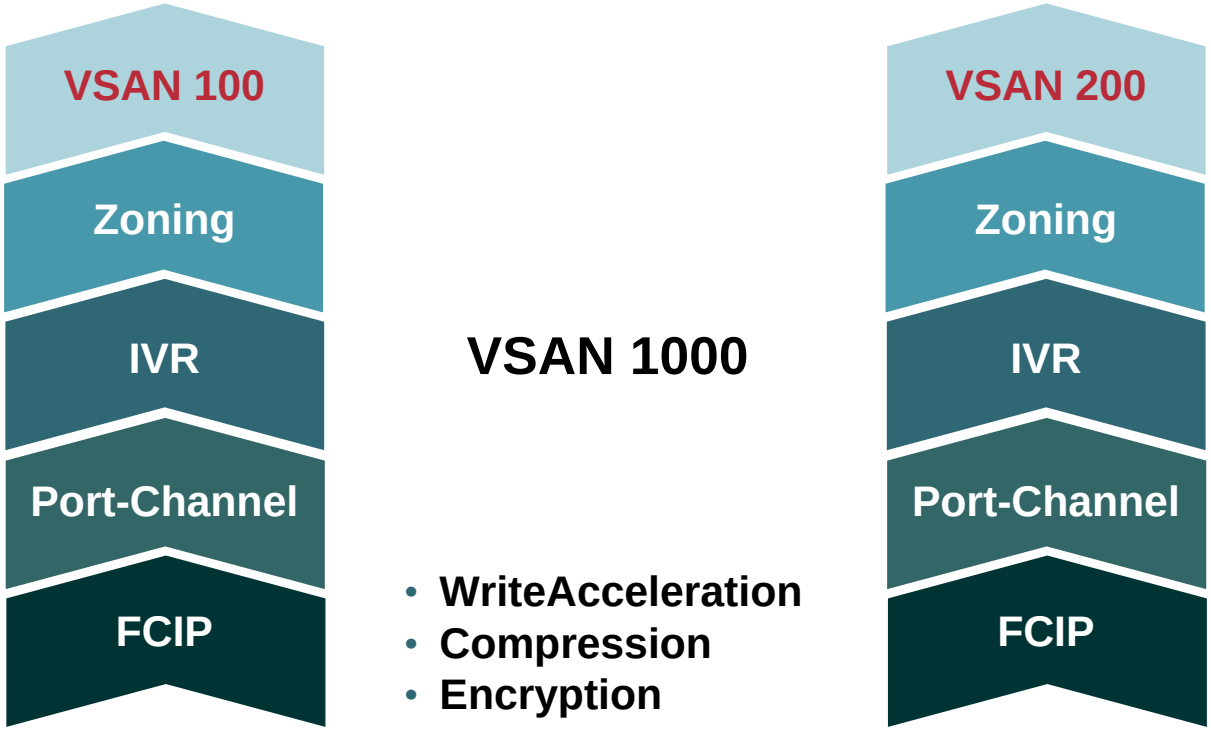
- **Extend the normal reach of a Fibre Channel fabric**
  - **Replication**
  - **Remote host to target array**
  - **Shared data clusters**

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**

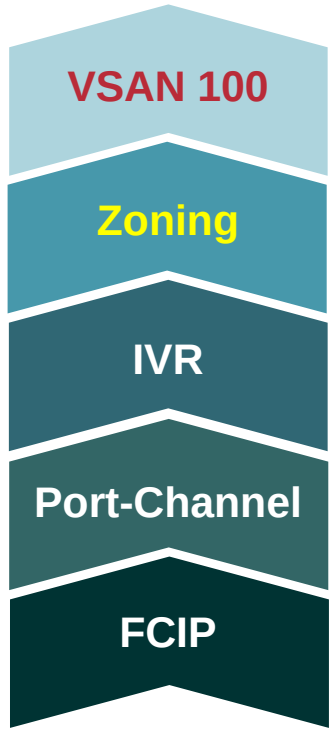
MDS Technology Hierarchy
Basic Infrastructure



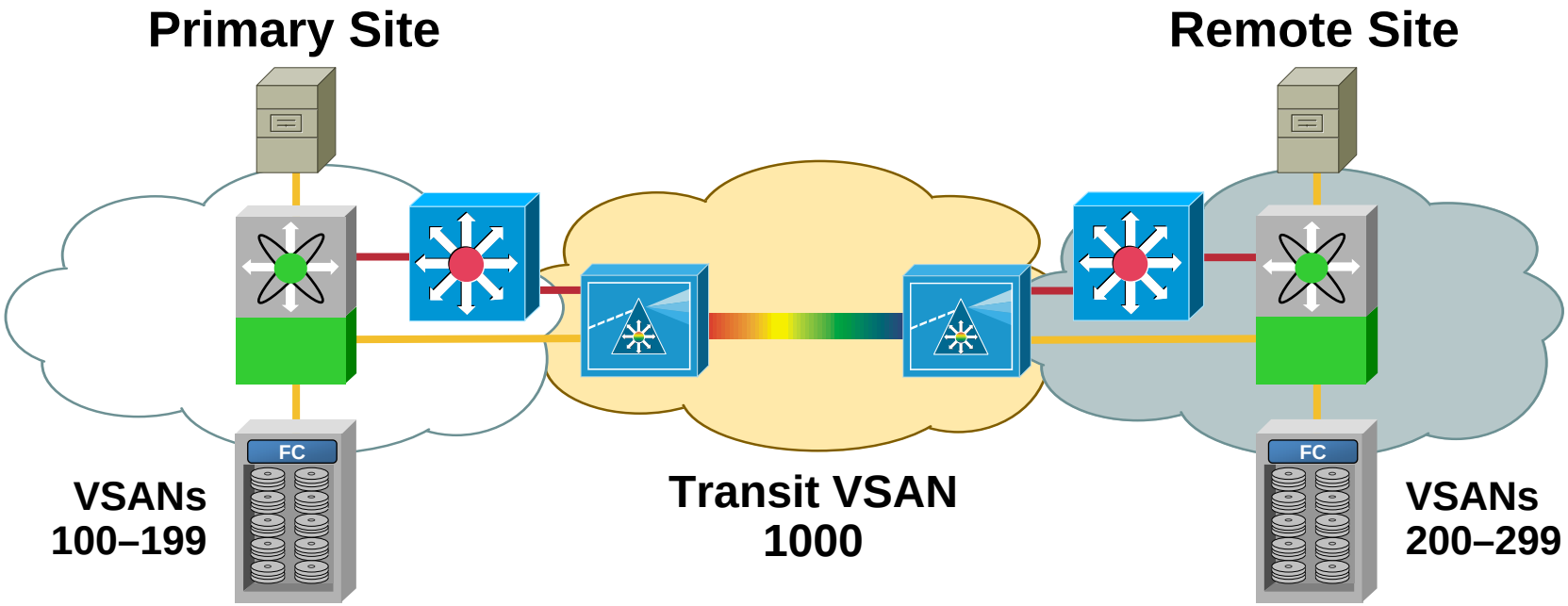
- Understand how the IO will flow from Primary to Remote site
- Helps you determine “why one device cannot communicate to another”
- Implement services from the “Bottom Up”

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**



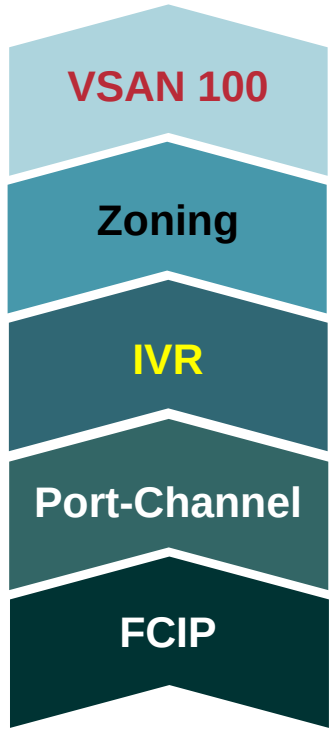
VSAN Topology



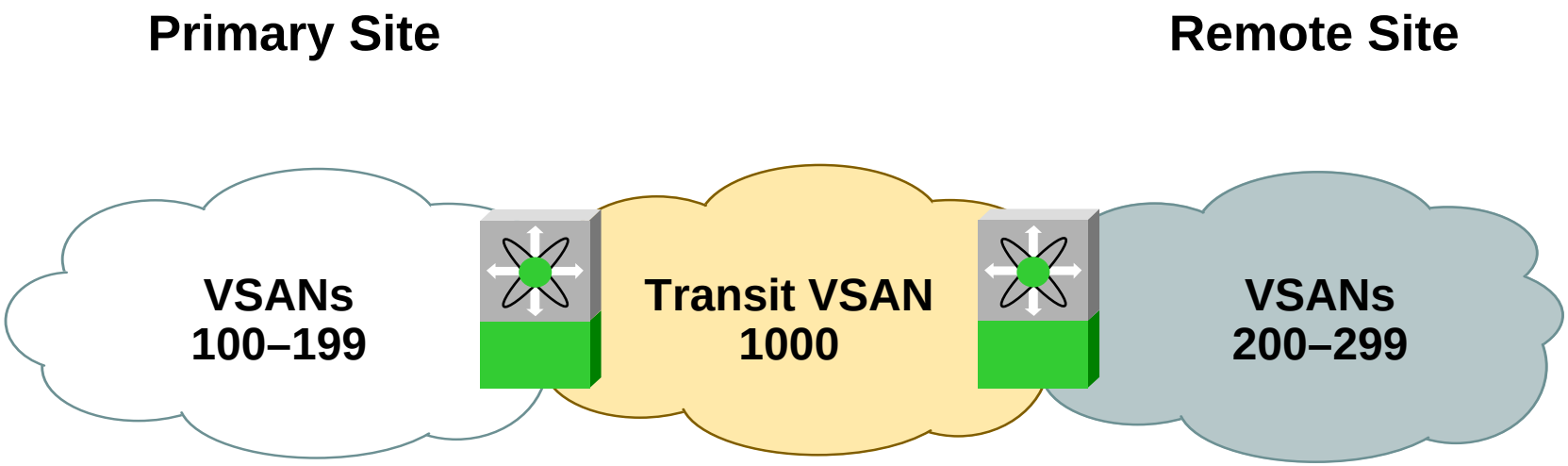
- Assign ranges of VSANs for future growth
- Provide ample room to prevent overlap
- Transit VSAN isolates Primary from Remote site

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**

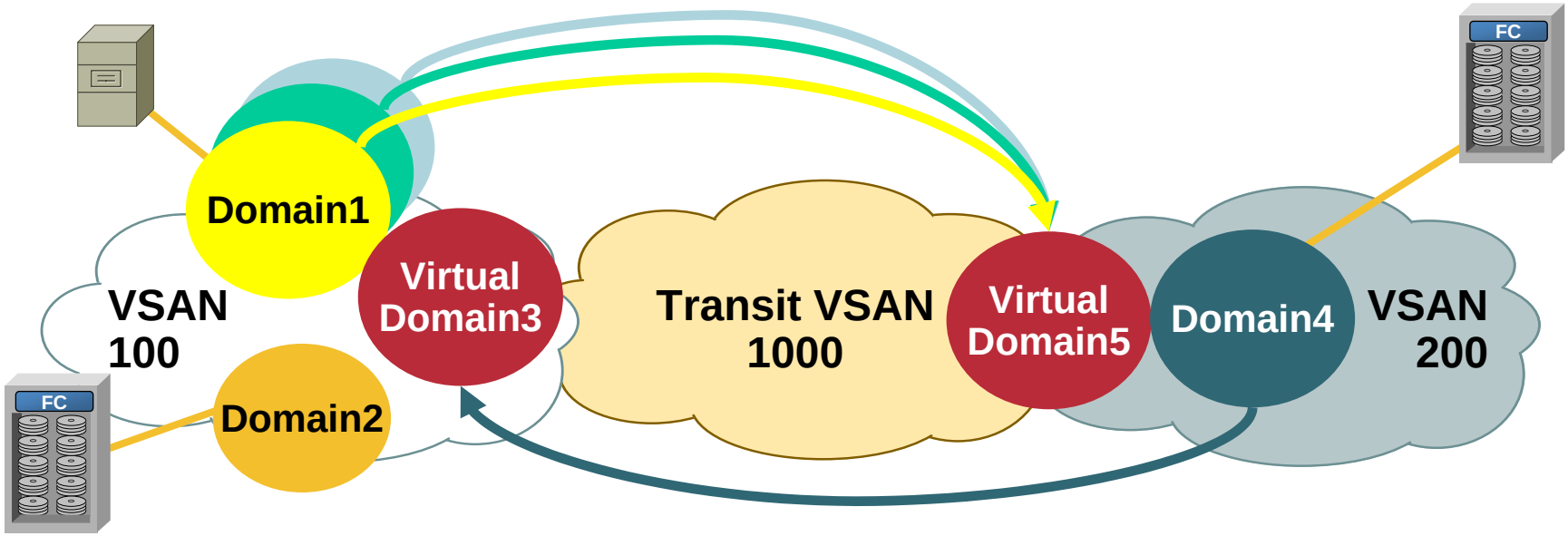


IVR View



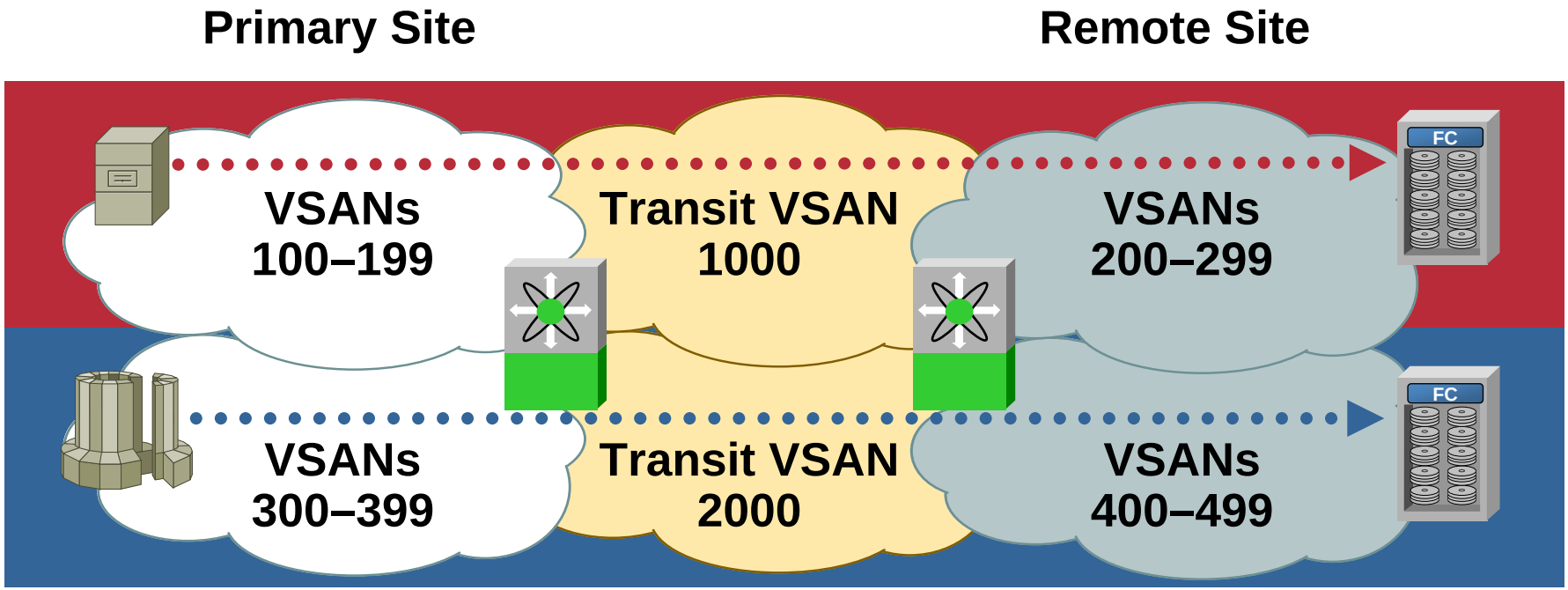
- Assign ranges of VSANs for future growth
- Provide ample room to prevent overlap
- Transit VSAN isolates Primary from Remote site

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

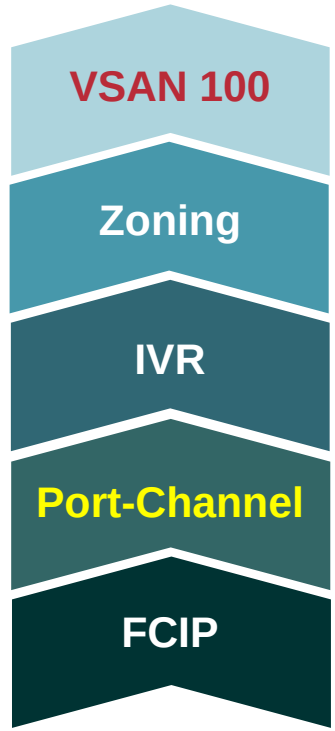
IVR Service Groups



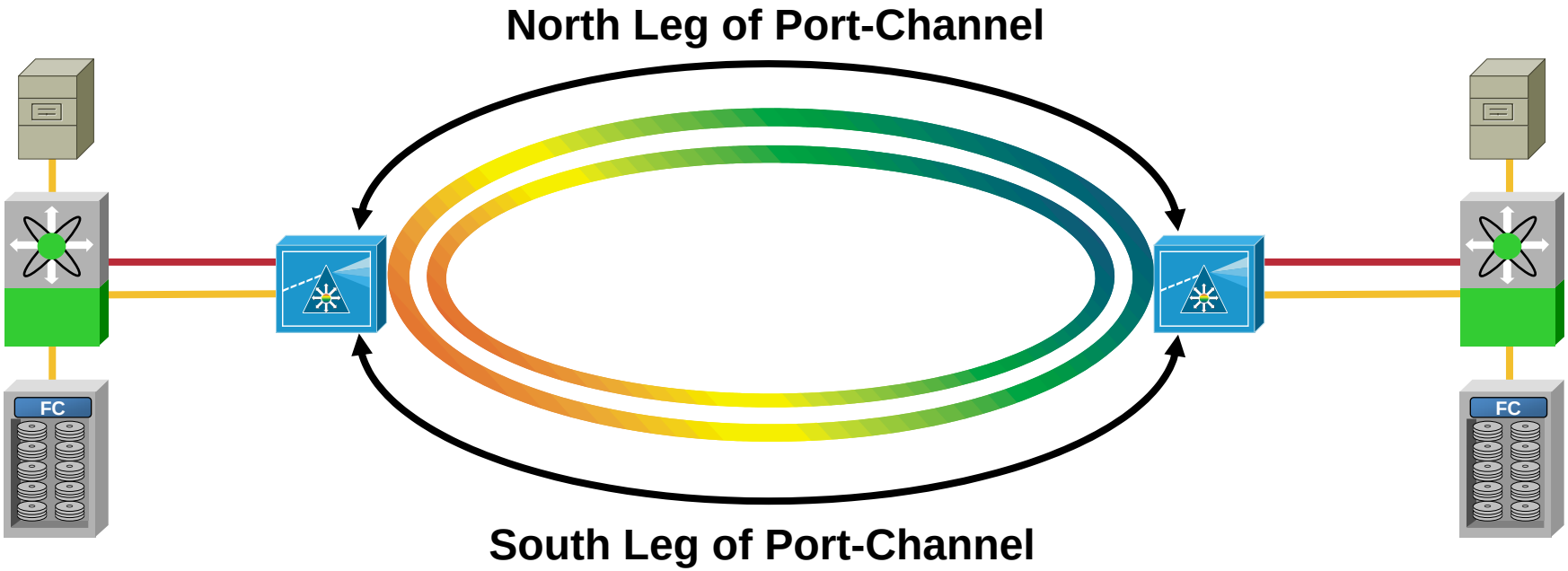
- Divides up the IVR Topology into “sub-topologies”
- Limits IVR events to a single service group
- Enables the use of different transit VSANs per service group
- “Gold,” “Silver,” and “Bronze” transit VSANs

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**



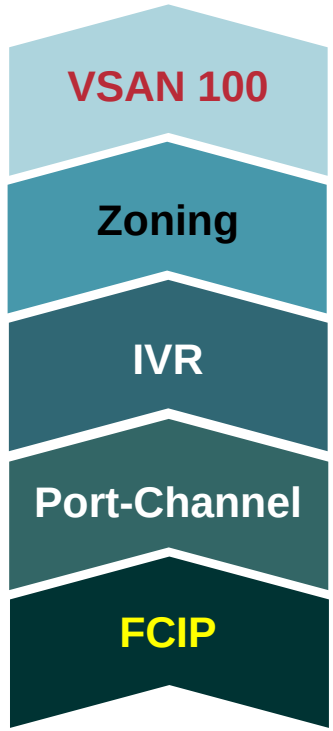
Port-Channels
ISL Resiliency



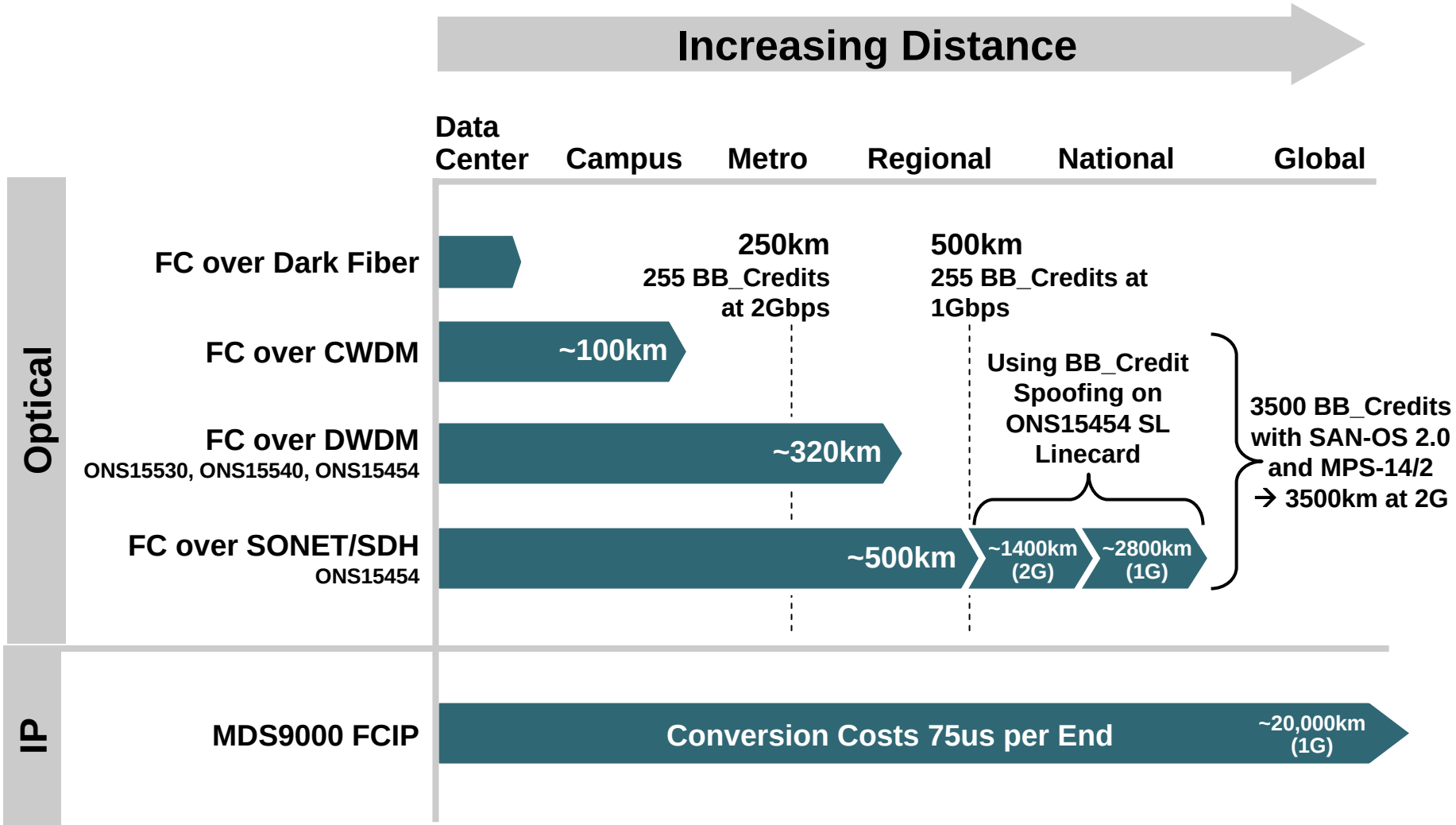
- Maintains switch connect even when members go down
- Can non-disruptively increase membership as bandwidth requirements scale to accommodate new DR projects
- FCIP and optical based port-channels are managed exactly the same

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**

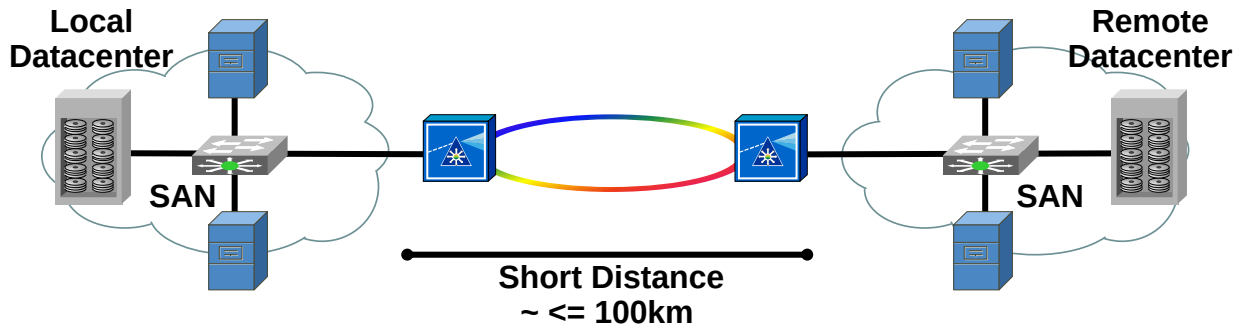


Determine Transport for Site Connectivity

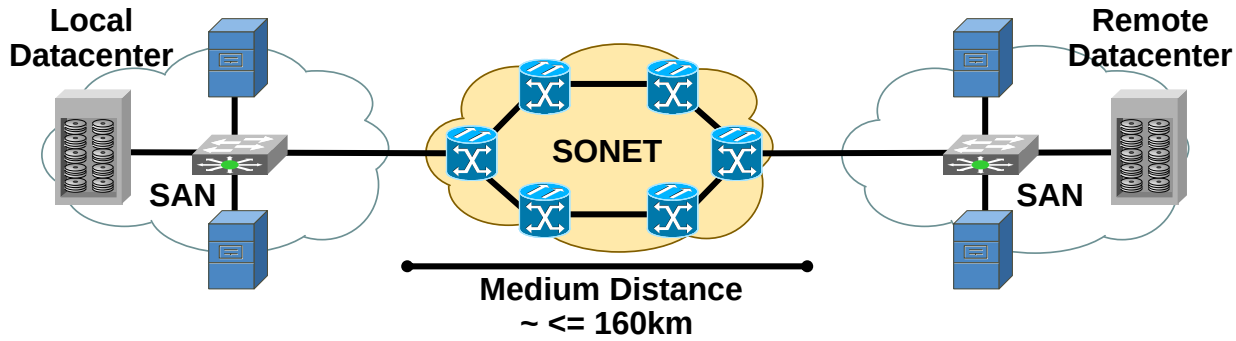


Multiprotocol Support
SAN Extension, IP SAN Extension

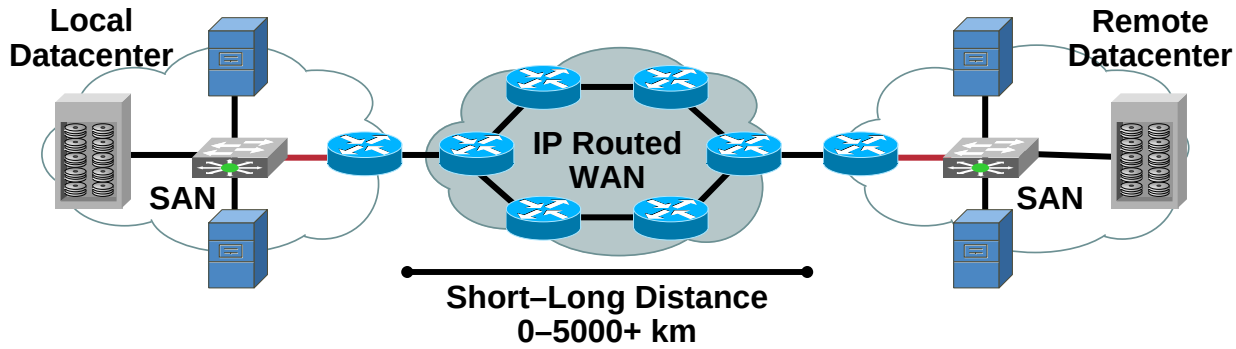
FC over DWDM/CWDM
- Short distance
- Dark fiber available
- Dedicated links
- Lowest latency—suitable for sync apps



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

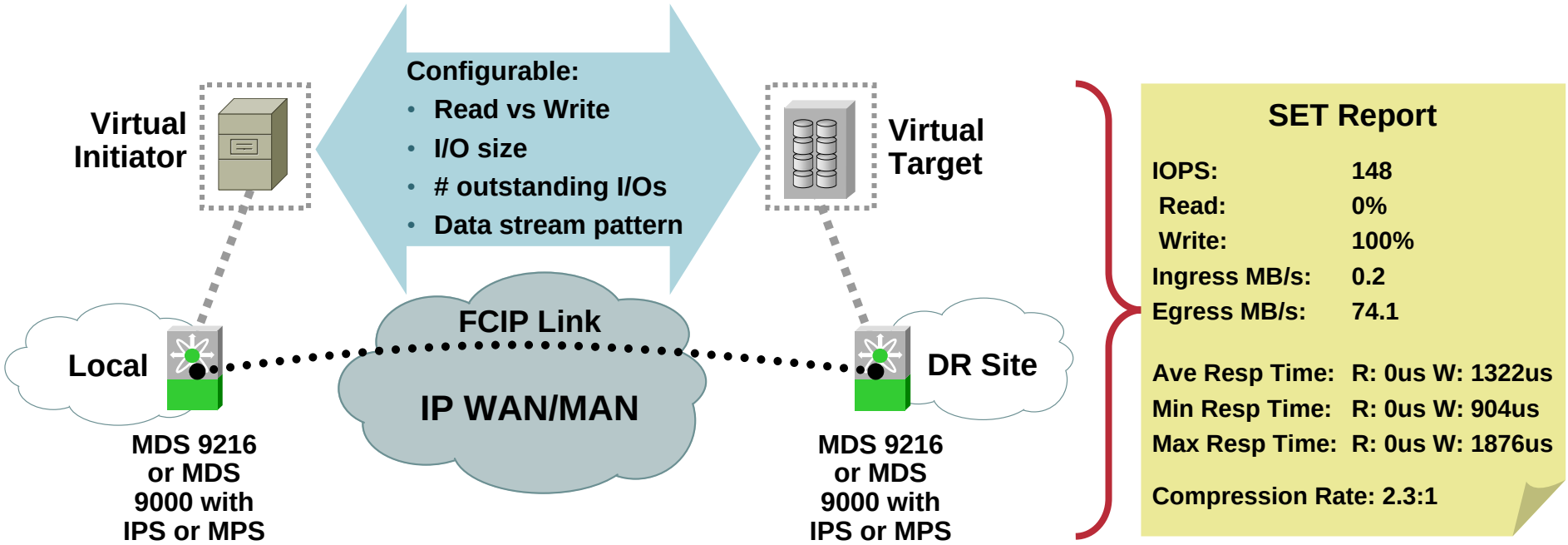


FC and FICON over IP
- Short–long distance
- Dark fiber not available
- Links may be shared
- Suitable for sync apps across metro Ethernet
- Suitable for async applications across WAN

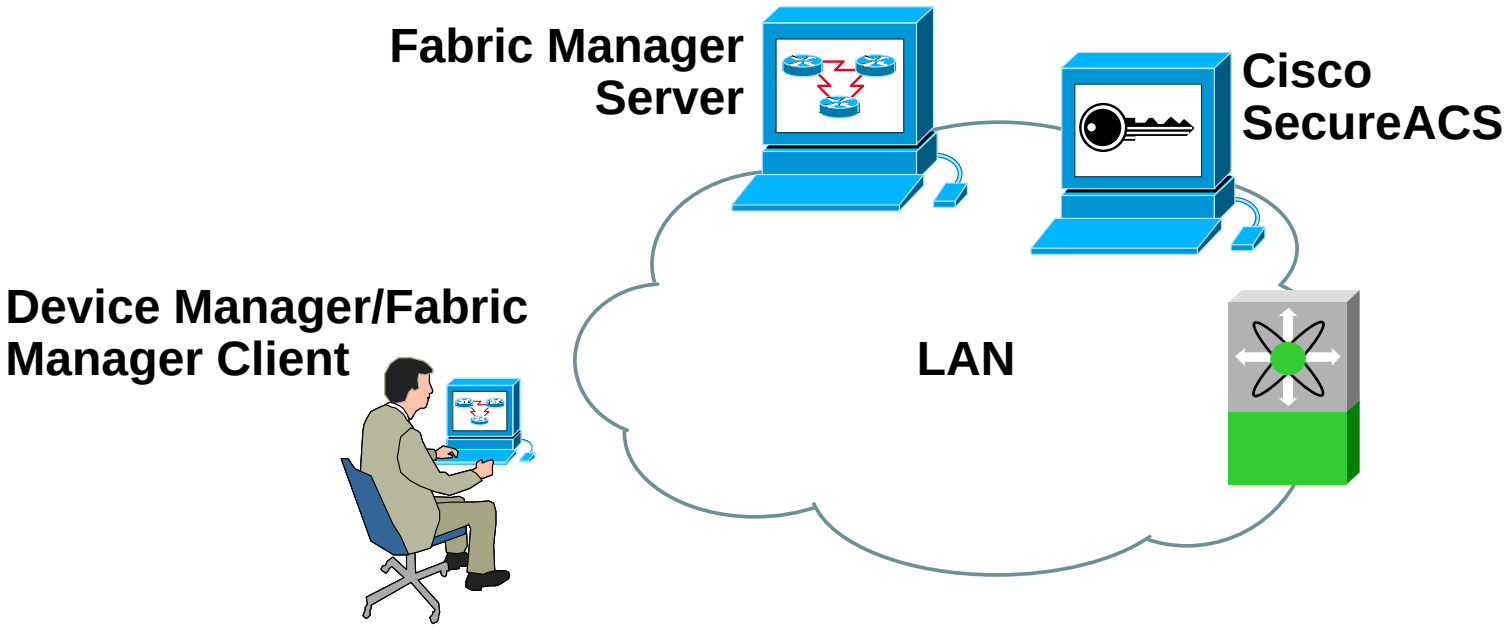


Stop
Test Your FCIP

- 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



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

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**

Q and A