Extended Distance SAN with MC/ServiceGuard Opens New Disaster Recovery Opportunities

Joseph Algieri
Senior Consultant
Network Storage Solutions
Hewlett-Packard Company

HP WORLD 2002
Conference & Expo

Overview

- What is a SAN
- Extended distance SAN
- Deploying MC/ServiceGuard clusters on extended distance SAN

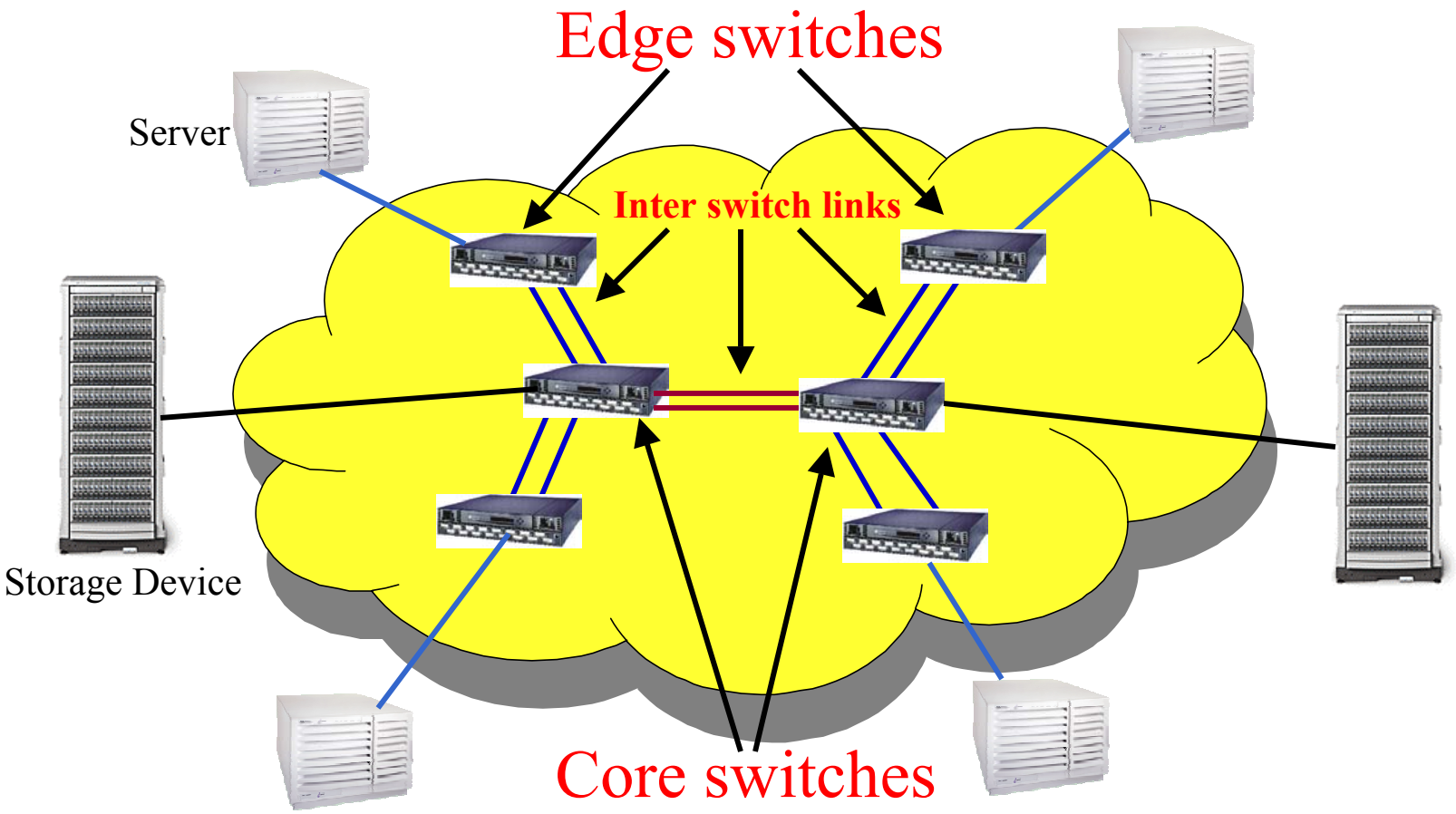
What is a SAN?

- SAN – Storage Area Network
- A specialized network used to connect servers and storage devices together
- Built using Fibre Channel switches

SAN Example

- Servers connect to “edge” switches
- Storage connects to “core” switches
- ISLs (Inter Switch Links) are used to connect switches together
- Generally, only ISLs between core switches will be > 500 meters

SAN Example



*Note: no hubs shown. Hubs would connect to edge switches

Extended Distance SAN

- Generally, long distance ISLs are used between “core” switches and short distance are used between “edge” switches
- ISL distances of up to 100km between switches

Extended Distance SAN

- Distance provided in two manners
  - Long distance GBICs (GigaBit Interface Controller)
  - DWDM (Dense Wavelength Division Multiplexing)

ISL Guidelines and Issues

- In general,the maximum number of ISL's between any pair of switches should equal one-half the number of ports on the switch (e.g. 1-8 for Brocade 2800; 1-4 for Brocade 2400)

- It is strongly recommended that a minimum of 2 ISL's exist between any pair of switches for both performance, redundancy, and to limit SAN fabric reconfiguration

- Not all switches support dynamic load balancing between ISL's. Port-port connections are assigned on a round-robin basis at switch startup and remain fixed regardless of loading. ISL re-assignment after a link failure is done automatically by the switch

SAN ISL Considerations

- Distance between sites
- SAN Bandwidth requirements
- Ensure a sufficient number of ISLs to support IO workload

SAN ISL Considerations

- Number of optical fibers available between sites
- Fiber Cost
- Consider DWDM if a large number of fibers are necessary (cost analysis)
- DWDM can support multiple connections over a single pair of fibers

ISL Distances

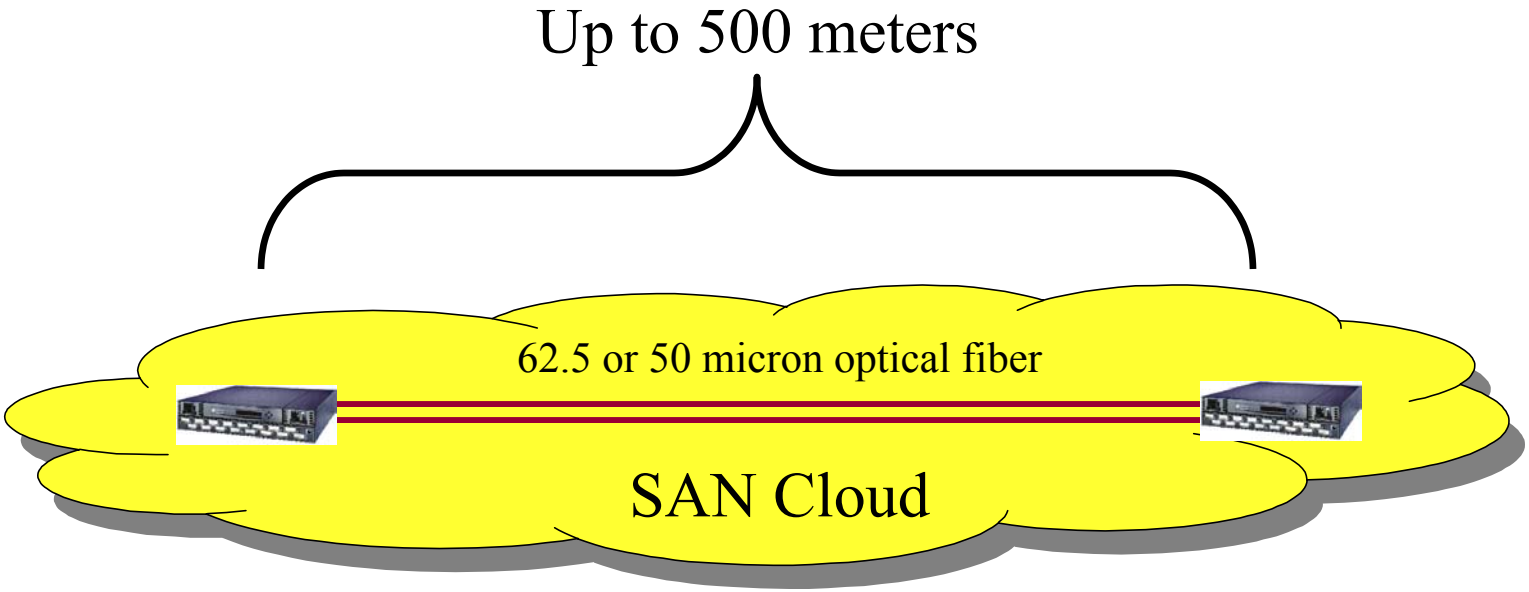
- Short-wave GBIC
  - Distances up to 500 meters
- Long-wave GBICs
  - Distances up to 10km
- Long-haul GBICs
  - Distances up to 80km
- DWDM (Dense Wavelength Division Multiplexing)
  - Distances up to 100km

Distance, Wavelength, and Optical Fiber Specification

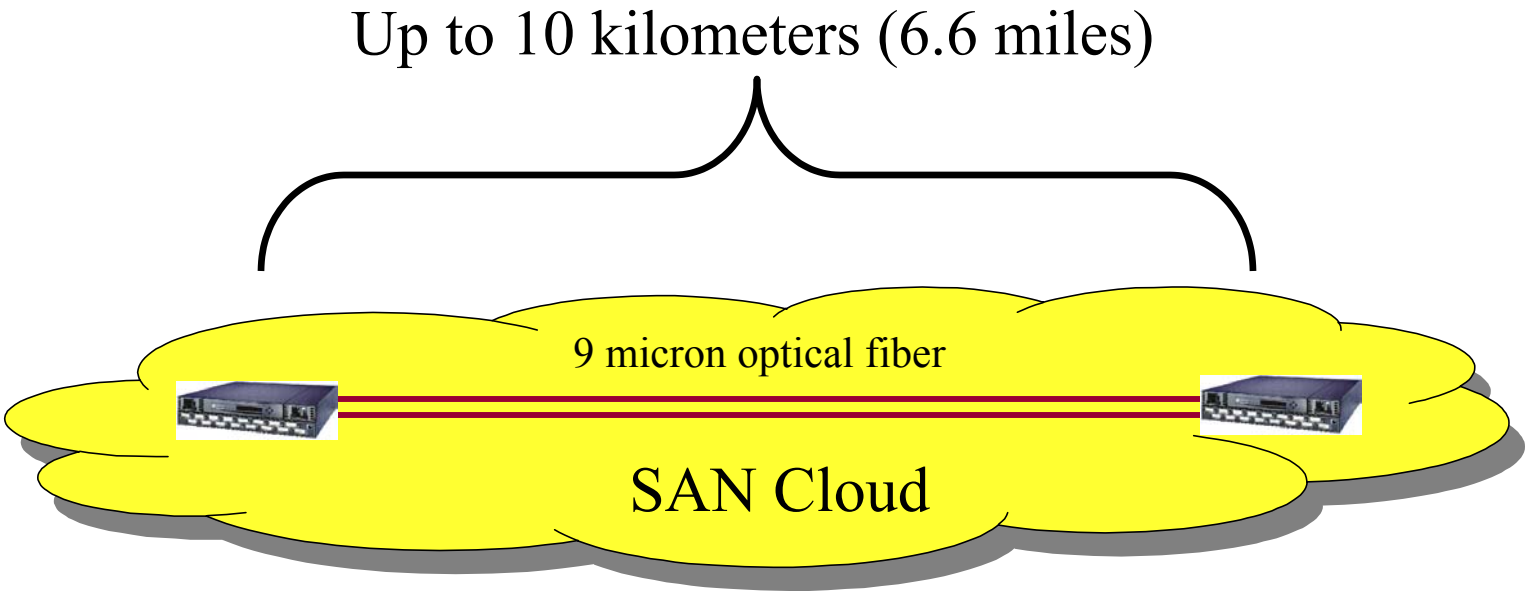
Optical Fiber Specification	62.5/125	50/125	9/125
Short Wave GBIC	175 meters	500 meters	NA
Long-wave GBIC	NA	NA	10 Km
Long-haul GBIC	NA	NA	80 Km
DWDM*	NA	NA	100KM

*can accept either long-wave or short-wave input. Consult your DWDM vendor for details

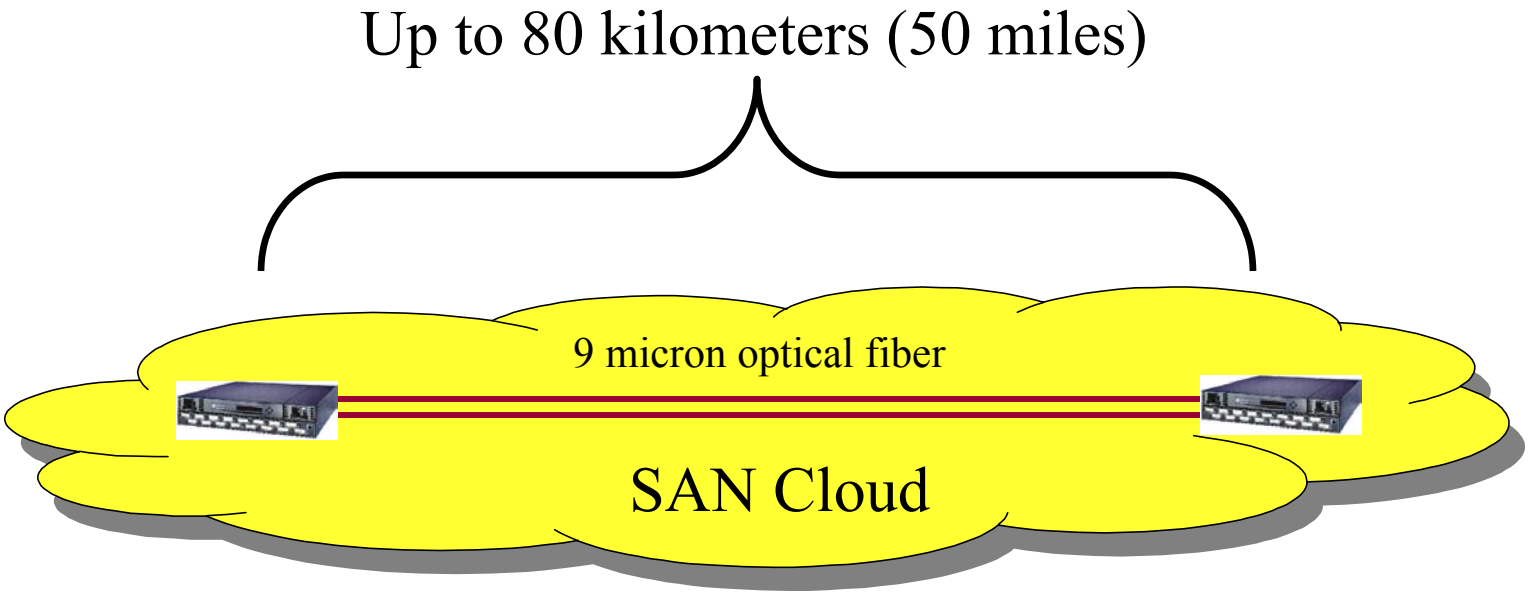
Short-Wave GBIC



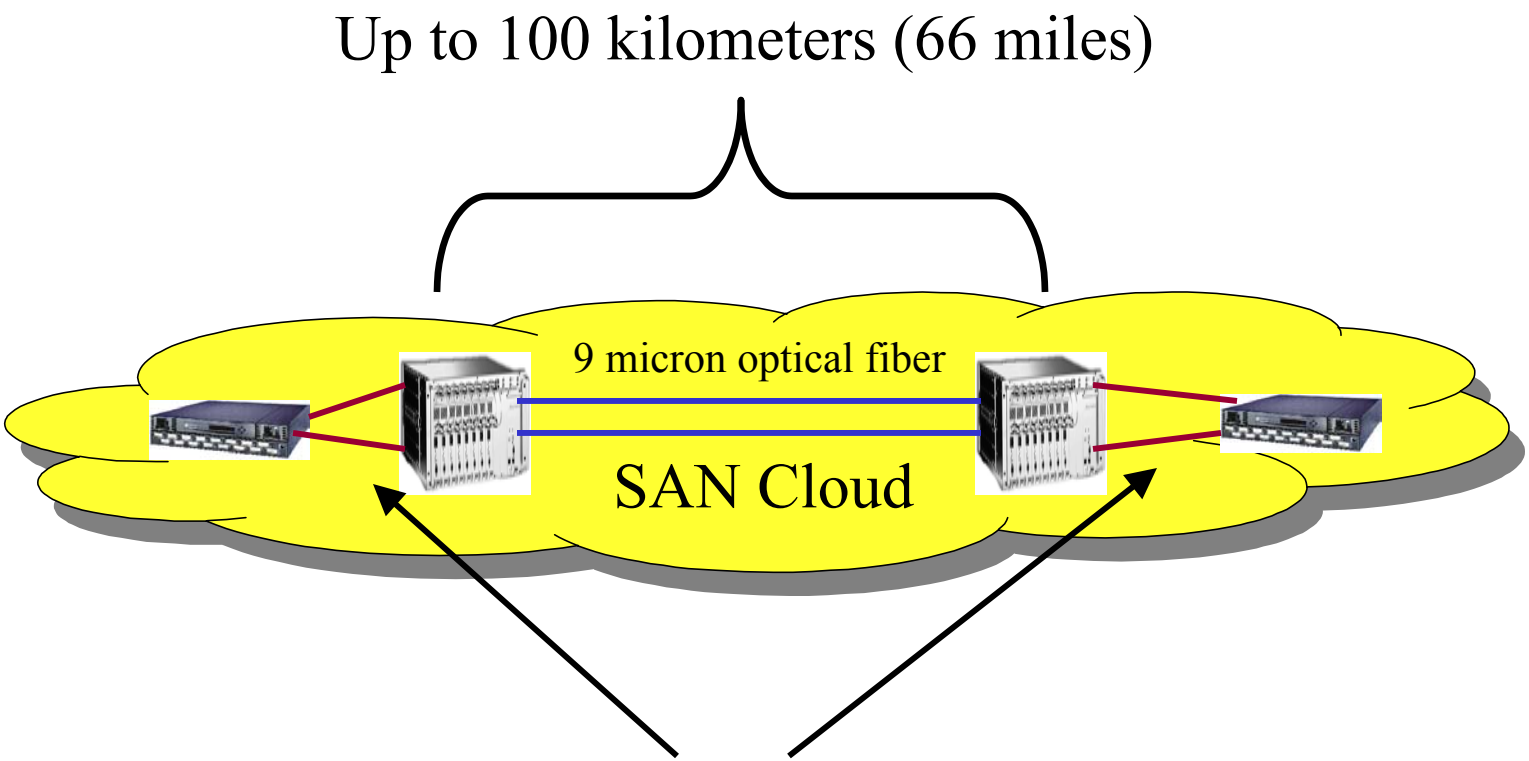
Long-Wave GBIC



Long-Haul GBIC

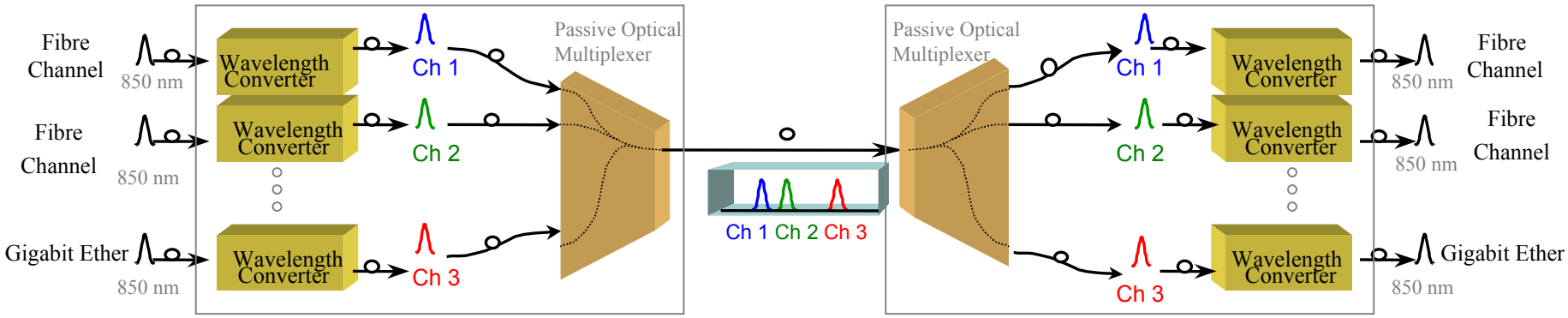


DWDM



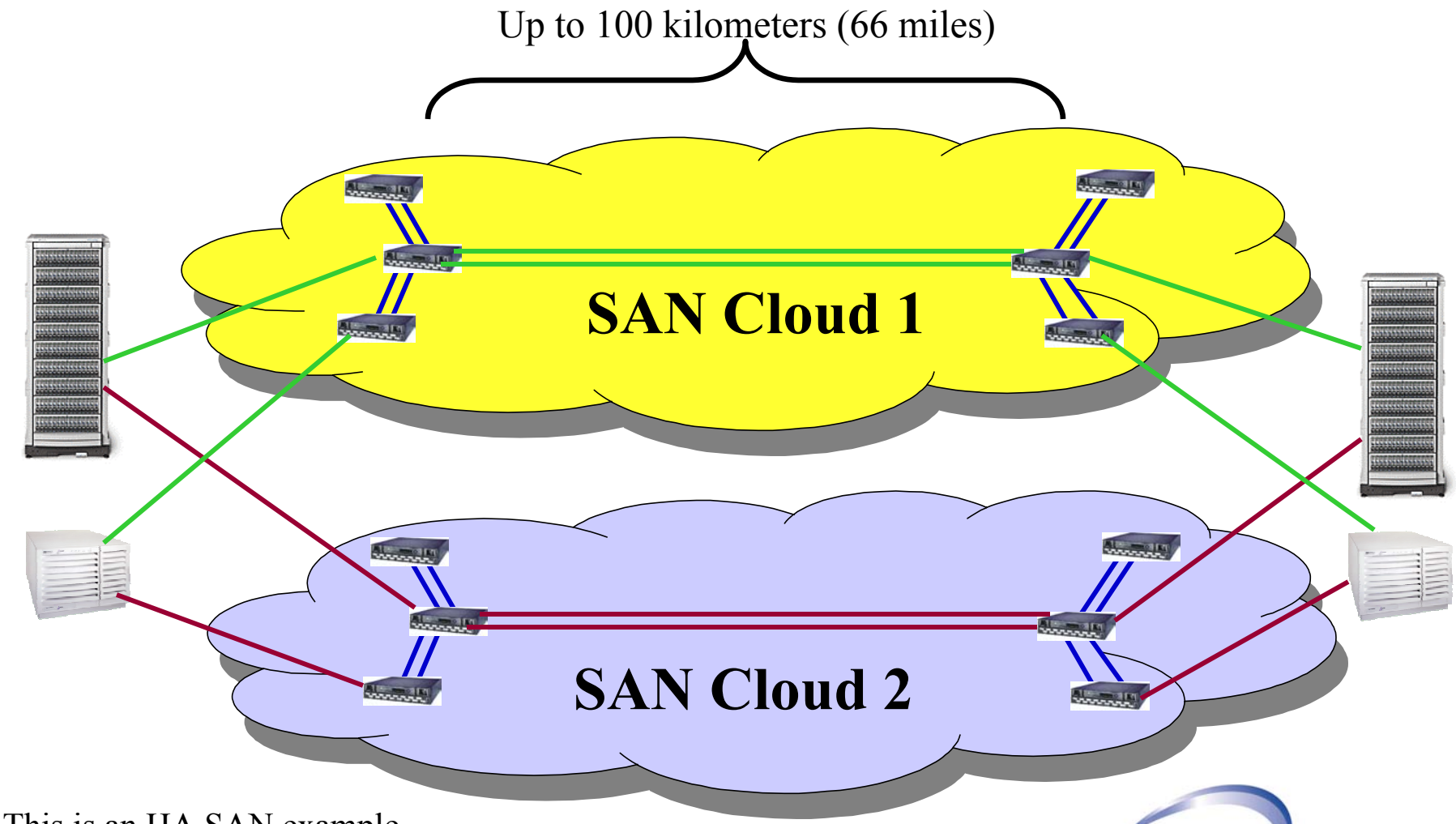
The FC switch to DWDM connection can be a short-wave or long-wave connection

DWDM Illustrated



- The number of channels available depends on the vendor and model of the DWDM equipment used

Example storage and server attach over extended distance SAN*



* This is an HA SAN example

Deploying ServiceGuard Clusters on Extended Distance SANs

- Why?
- SAN Design Considerations
- Solution Design Considerations
- Other issues

Why Extended ServiceGuard Clusters?

Clustering w/ MC/ServiceGuard +

Data Replication w/ MirrorDisk/UX +

Storage Infrastructure w/ Extended Distance SAN =

Low cost entry level DR solution

Why Extended ServiceGuard Clusters?

- Any storage supported by MC/ServiceGuard can be used in an extended distance cluster solution
- Use low cost modular storage in DR solutions
- Leverage existing storage

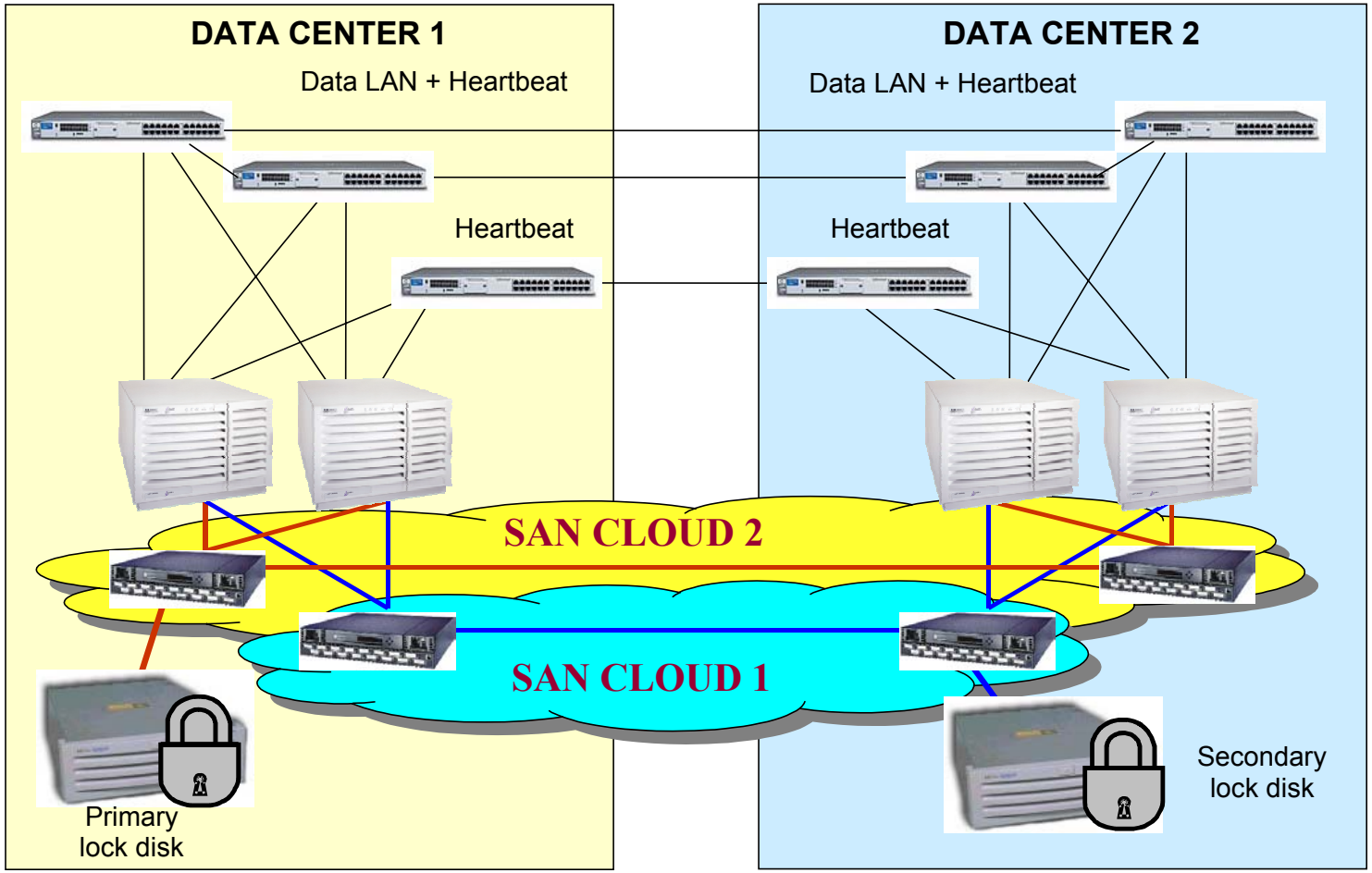
Extended Distance SAN Design Considerations

- Redundant SAN connectivity between servers and storage is highly recommended.
- SAN ISL cables between data centers must follow separate physical paths
- All legal and supported SAN configurations from HP are supported for extended distance clusters

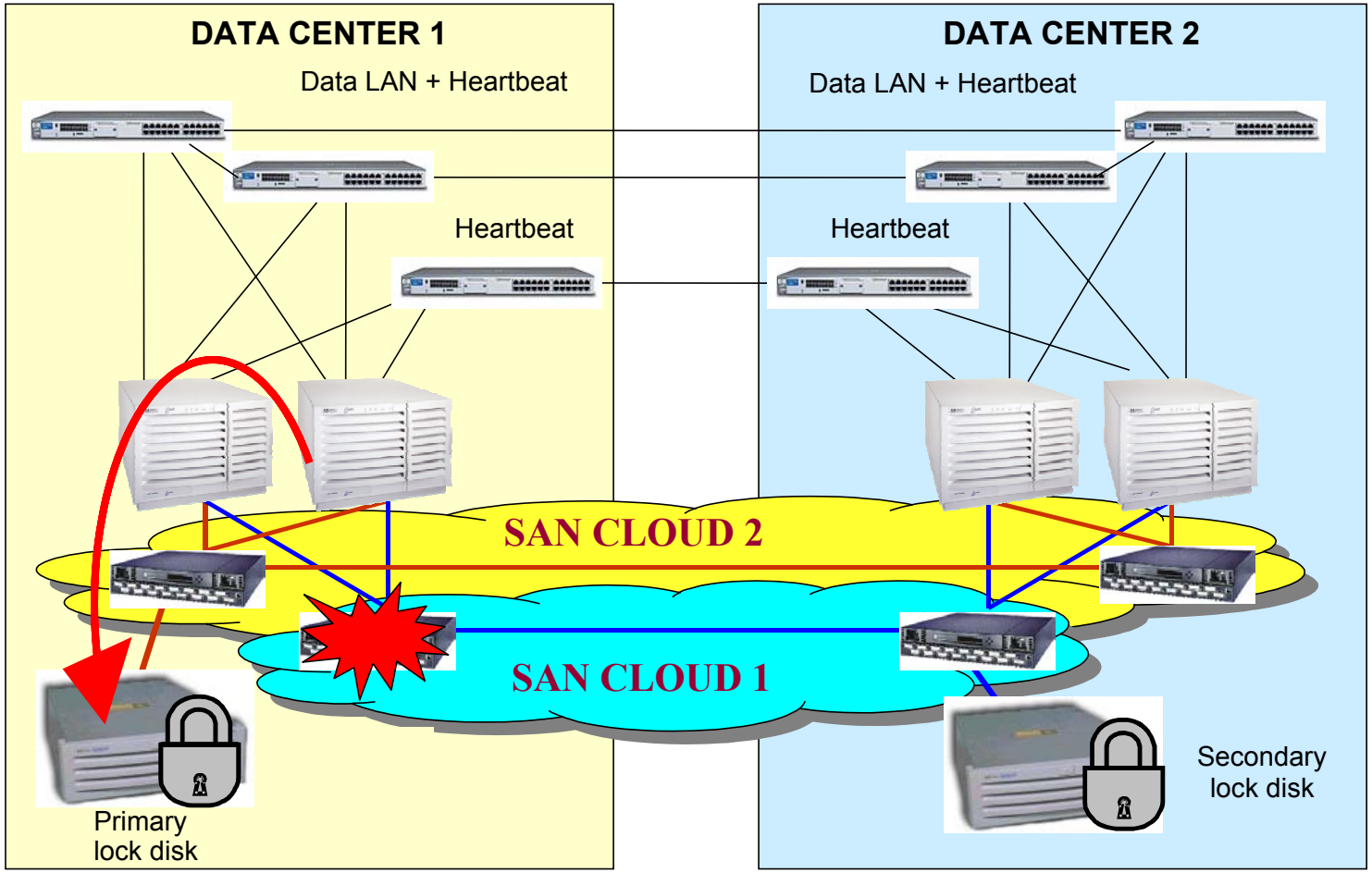
Extended Distance SAN Examples

- Dual SANs without PV-Link support
- Dual SANs with PV-Links support
- Dual SANs and networks sharing a DWDM site interconnect

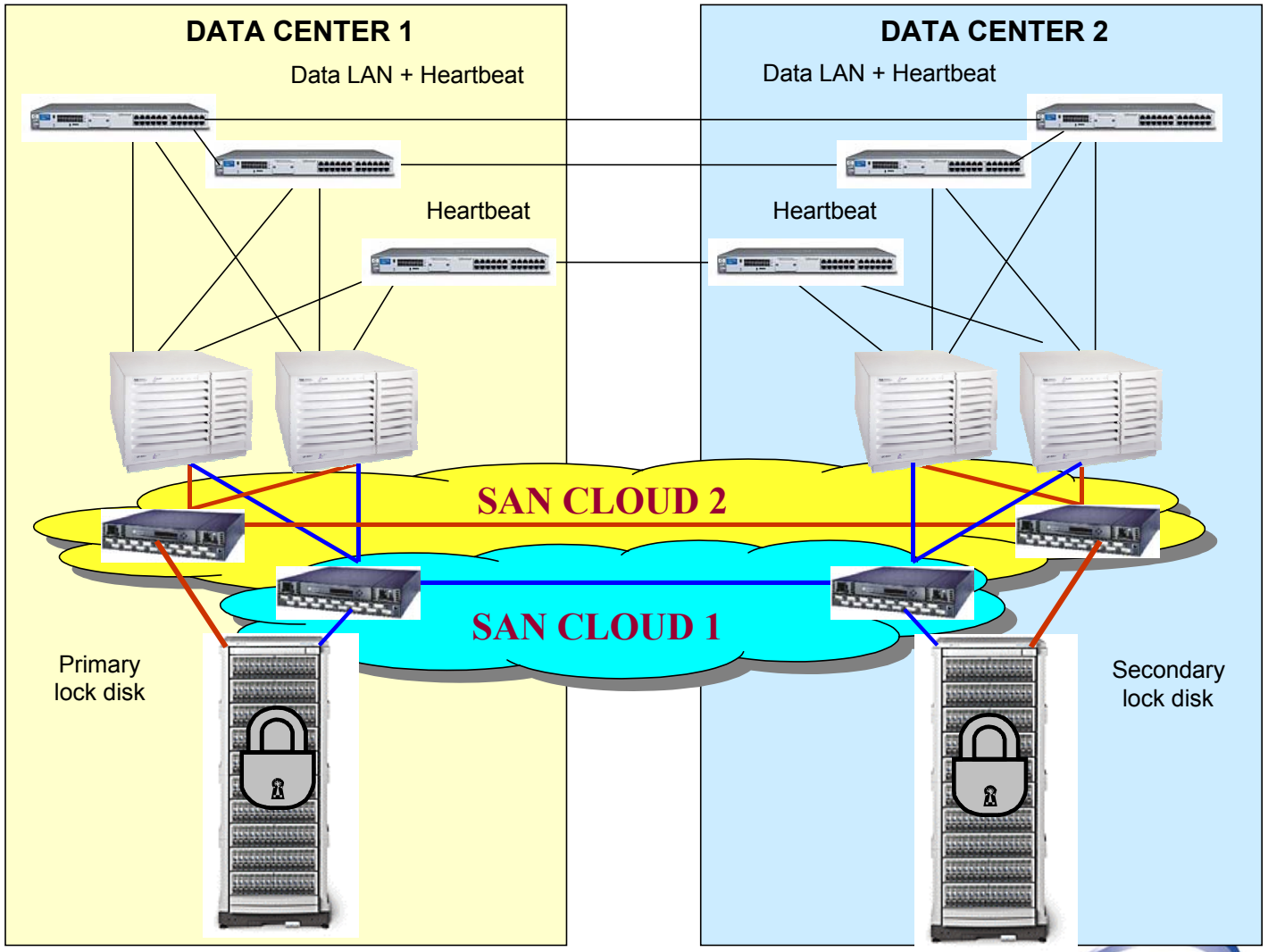
Dual SAN clouds without PV-Link support



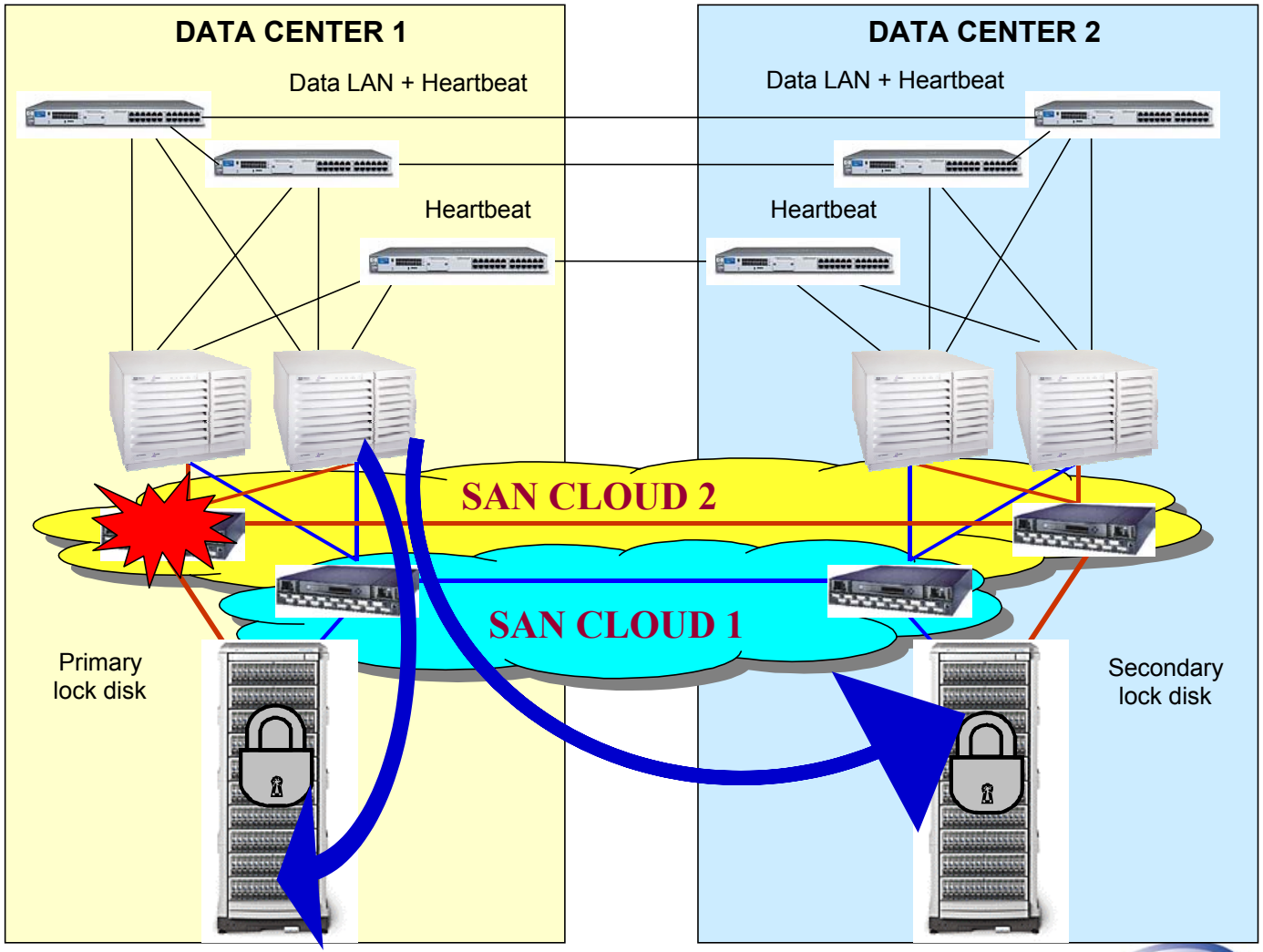
Failure without PV-Links - interrupted data replication



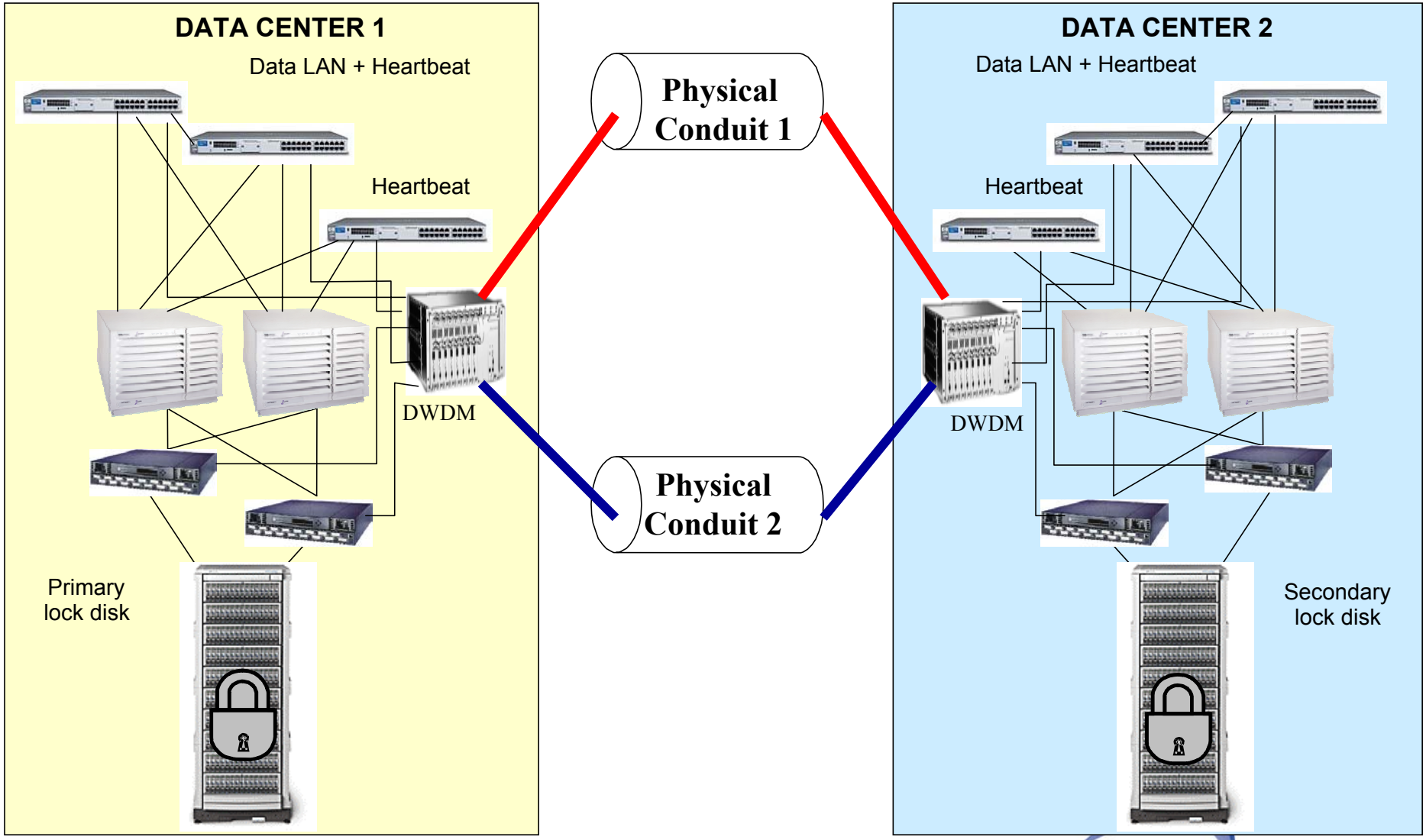
Dual SAN clouds with PV-links – preferred design



Uninterrupted data replication with PV-Links



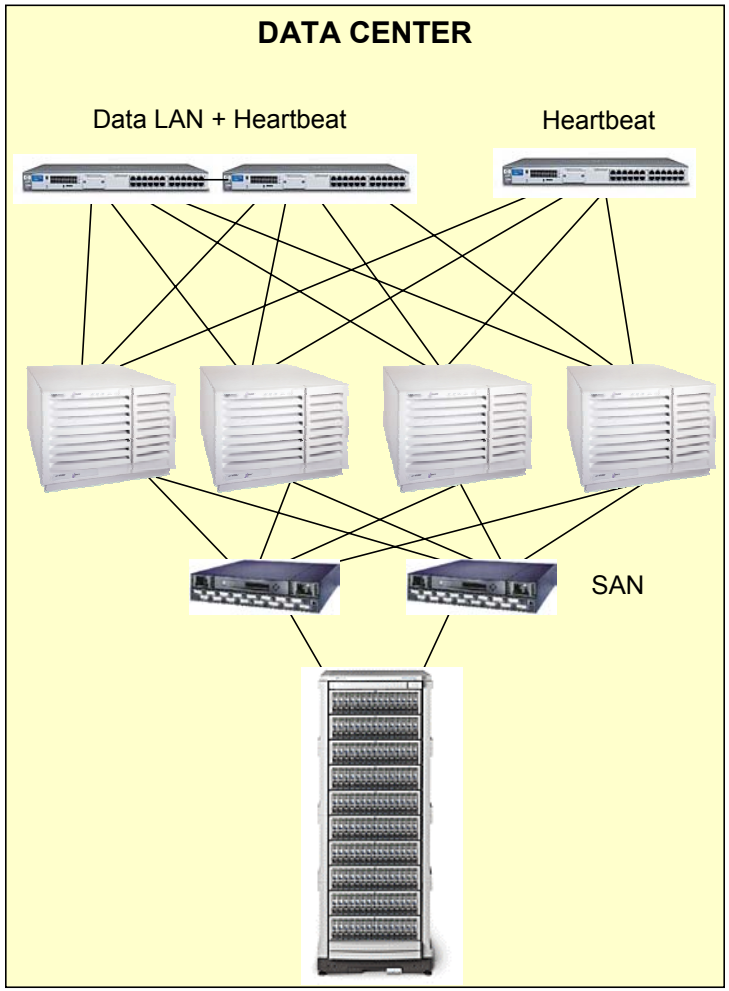
Extended Cluster with DWDM Site Interconnect



MC/ServiceGuard Cluster in a Single Data Center

- MC/ServiceGuard's intended use
- A properly designed solution protects from all single points of failure in the infrastructure.
- Protection limited to failures within the data center

MC/ServiceGuard Cluster - Single Data Center



Extended Campus Cluster Solution

- Extended Campus clusters moves MC/ServiceGuard into the DR space
- MC/ServiceGuard is the basis for all of HP's cluster DR solution product (MetroClusters & ContinentalClusters)

Extended Cluster Solution

- Takes an MC/ServiceGuard cluster and turns it into a low cost DR solution
- Uses an extended distance SAN to provide a medium for host based replication of the cluster data

Extended Cluster Solution Considerations

- Many special considerations
- There are no special MC/ServiceGuard configuration requirements
- All MC/ServiceGuard configuration rules and limitation must be adhered to for the solution to be supportable

Extended Cluster Solution Considerations

- There are special cluster architecture requirements
- There are special solution infrastructure requirements

Extended Cluster Solution Considerations

- To ensure a complete DR solution, solution monitoring must be addressed
- Protection from a rolling-disaster is not automatic and can be very tricky

Extended Cluster Solution Considerations

- A rolling disaster occurs when a data center failure occurs while recovery from an initial failure requiring data resynchronization is in progress
- Results in a total loss of data requiring a restore from backup

Extended Cluster Solution Considerations

- Protecting from a rolling disaster is difficult because MirrorDisk/UX is the data replication tool being used

- MirrorDisk/UX does not inherently contain functionality required to protect from a rolling disaster

Extended Cluster Solution Considerations

- Other tools must be integrated into the solution to help protect from a rolling disaster
- This portion of the solution will be custom and will be driven by the customer's availability requirements

Extended Cluster Architecture Requirements

- Primary data centers must always contain the same number of nodes
- No lock disk support for clusters with > 4 nodes (MC/ServiceGuard limitation)
- Single cluster lock disk solutions ARE NOT supported

Extended Cluster Architecture Requirements

- Solutions requiring more than 4 nodes must either use a three data center solution or a two data center solution with quorum server (lock disks cannot be used)
- Maximum number of nodes in the cluster is 16 (MC/ServiceGuard limitation)

Extended Cluster Architecture Requirements

- Data replication is host based via MirrorDisk/UX over extended distance SAN
  - Can put performance pressure on the servers
- Cluster must be deployed so that the failure of an entire data center can be recovered from by the MC/ServiceGuard cluster quorum protocols

Extended Cluster Infrastructure Requirements

- Two separate SANs clouds required between data centers
- Two separate IP networks required between data centers to carry cluster heartbeat
- Redundant cables between data centers must follow different physical routes to avoid the back-hoe problem

Combining Extended Distance SAN and MC/ServiceGuard

- No special or custom configurations or licenses are required
- Solution utilizes standard SAN support
- Solution utilizes standard MC/ServiceGuard support

Extended Cluster Monitoring Requirements

- Extended clusters require extensive solution monitoring
  - Event Management Service
  - OpenView
  - Clusterview
  - BMC
  - CA
  - Operator vigilance
  - ??

General Considerations

- Extended distance cluster solutions are very complex to design, monitor, and operate properly
- Properly designing, deploying, and managing an extended distance cluster is much more difficult to do than it looks
- Engage HP Consulting to ensure proper solution design, deployment, and management

Cluster Quorum

- Following the failure of a node or nodes in a cluster, a new cluster consisting of the surviving nodes must be able to form

- Care must be taken to ensure the cluster protocols can achieve quorum and a new cluster can form following a complete data center failure

Cluster Quorum

- Three ways to achieve quorum after a failure
  - New cluster contains > ½ of the nodes from the prior cluster
  - New cluster contains ½ of the nodes from the prior cluster and the quorum server votes for it
  - New cluster contains ½ of the nodes from the prior cluster and a cluster lock disk

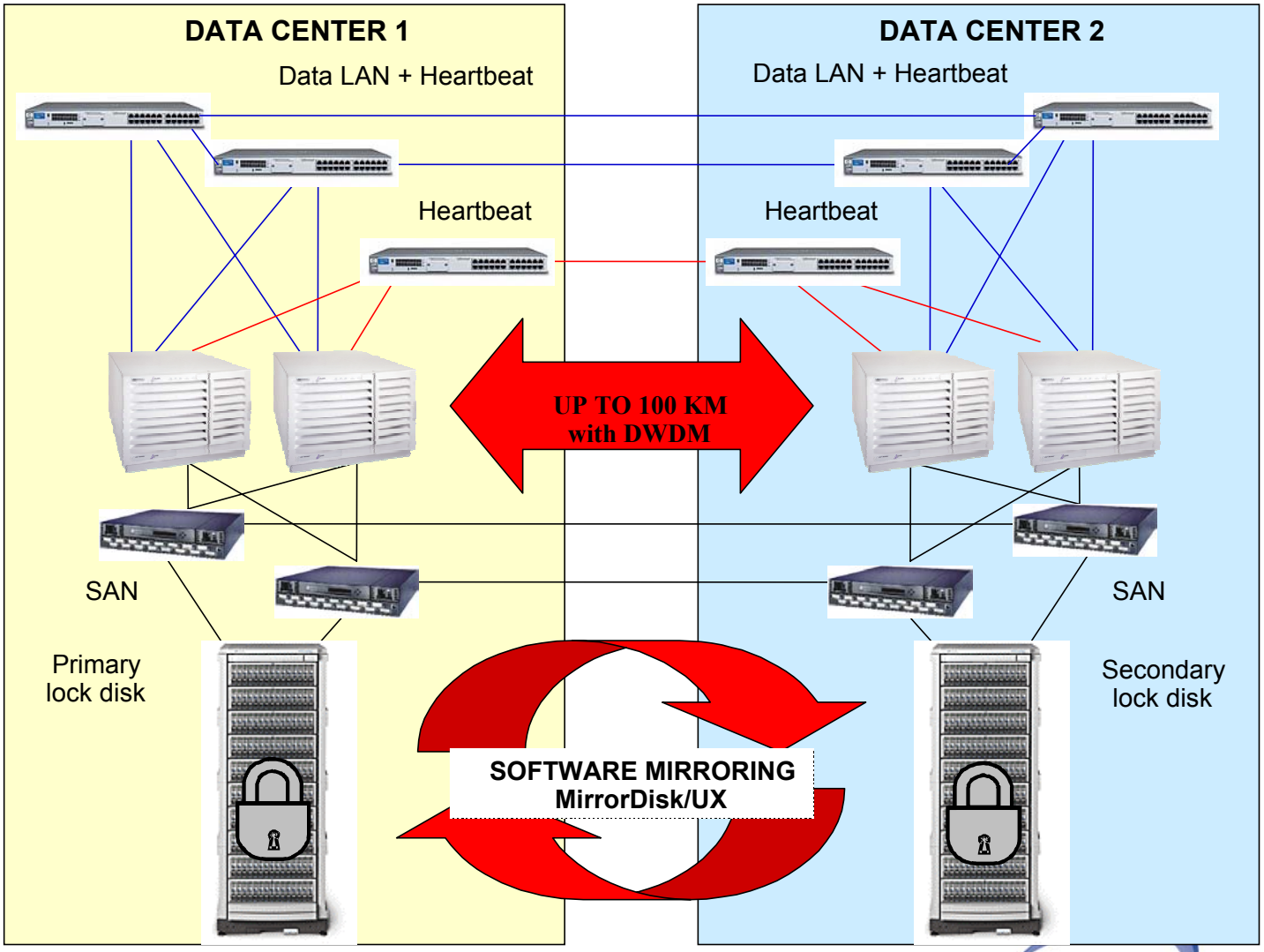
Cluster Quorum

- Both three-site clusters and two-site clusters with cluster quorum server will ensure quorum can be achieved after a site failure without the use of cluster lock disks
- A two site cluster with dual cluster lock disks will be able to achieve quorum after a complete data center failure

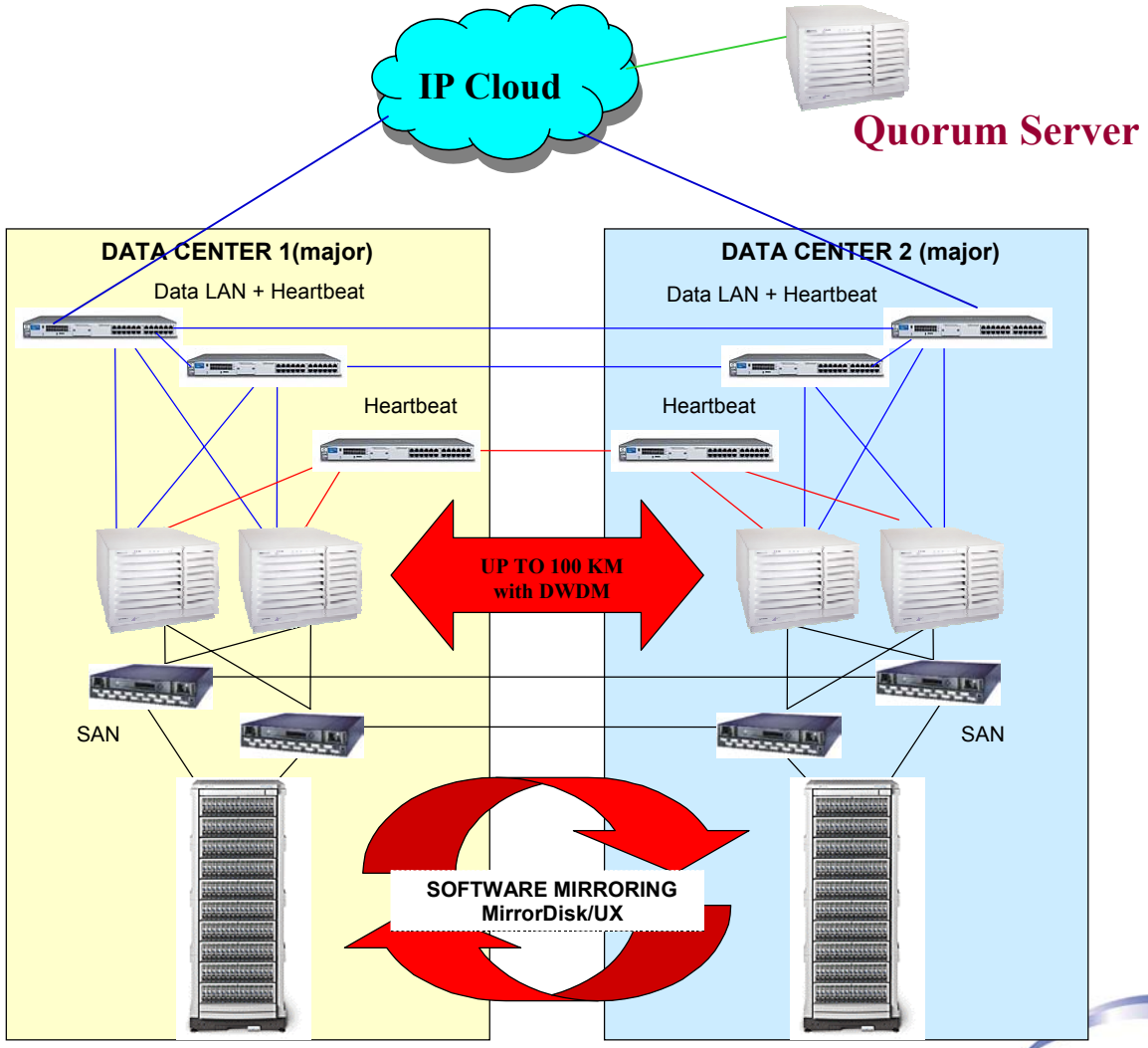
Extended Cluster Supported Topologies

- Two Data Center Design
  - Cluster Quorum Server
  - Dual Cluster Lock Disks
- Three Data Center Design

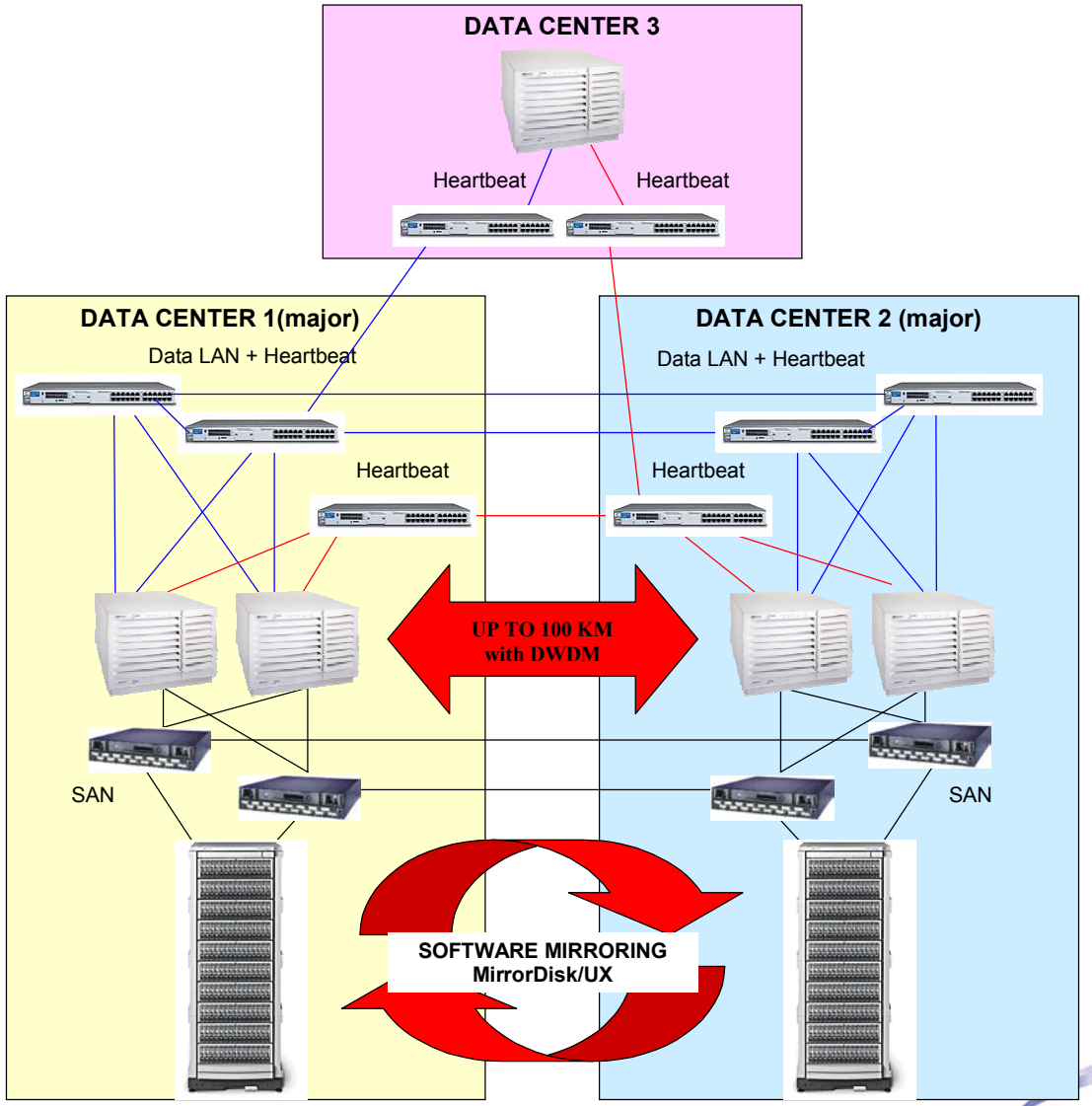
Two Data Center Design with Lock Disks



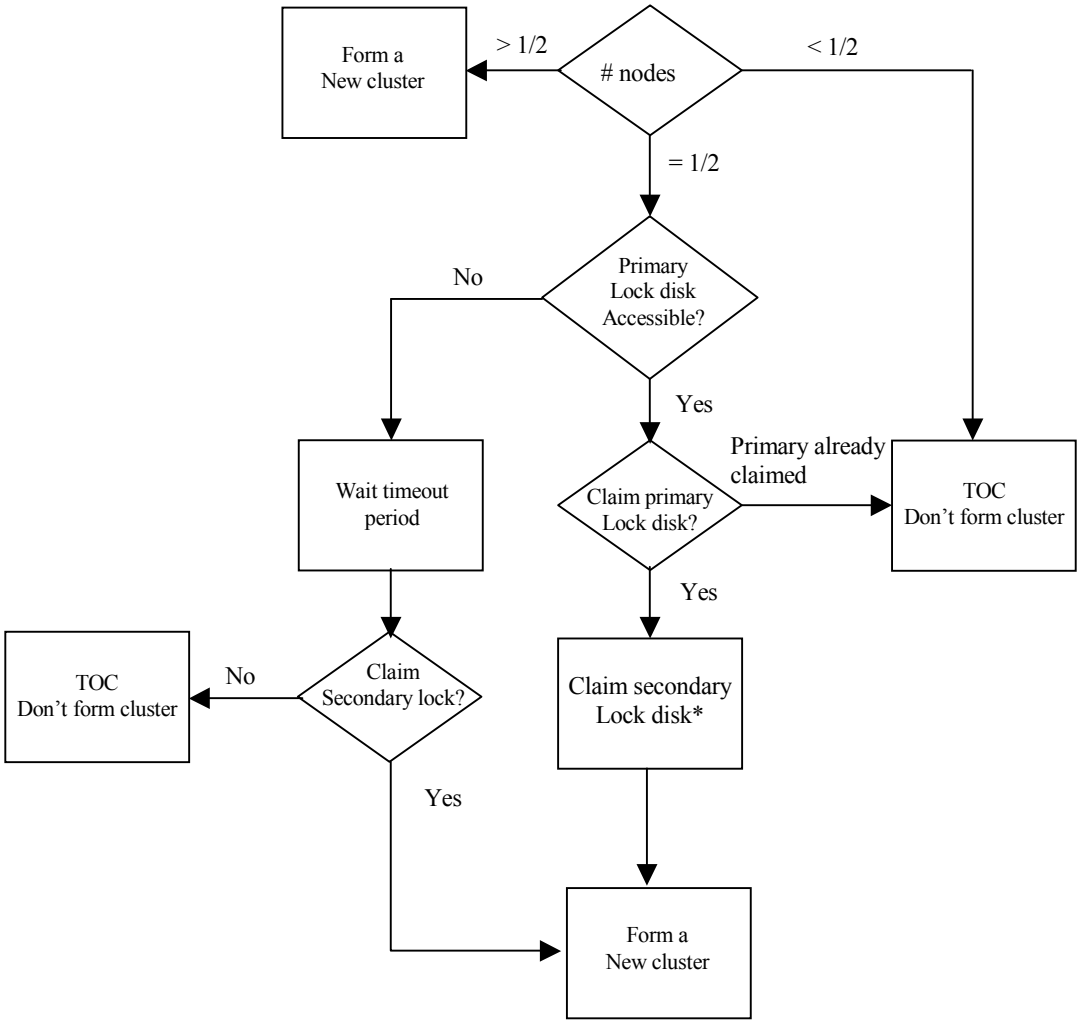
Dual Data Center Design with Quorum Server



Three Data Center Design



Dual Cluster Lock Disk Algorithm Flowchart



*The secondary cluster lock will not be claimed here if it is not accessible

Dual Data Center Extended Cluster Issues

- Dual data center solutions require the use of dual cluster lock disks or a cluster quorum server
- Use of dual cluster lock disks opens the door to "Split-Brain" syndrome
- Single cluster lock disk configurations are NOT supported as a single cluster lock is a SPOF

Split-Brain Syndrome

- Split-brain syndrome is when two separate viable clusters form from a single cluster
- Split-Brain can occur after a failure that breaks all network and SAN links between the data centers
  - The “back-hoe” problem
  - Failure of non redundant DWDM equipment

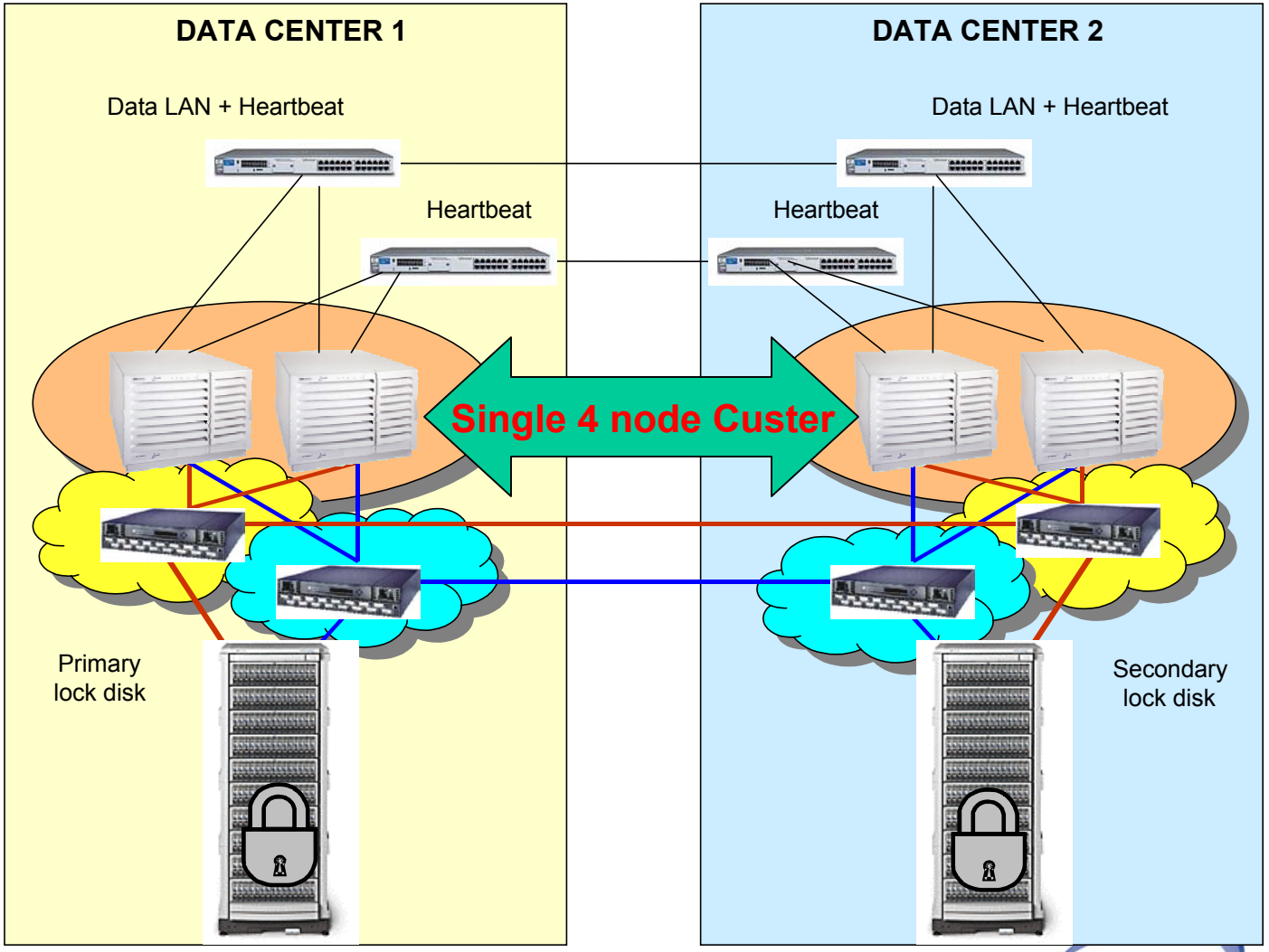
Split-Brain Syndrome

- Only a cluster configured with dual cluster lock disks can suffer split-brain
- Requires a failure of all inter-data center network and SAN links

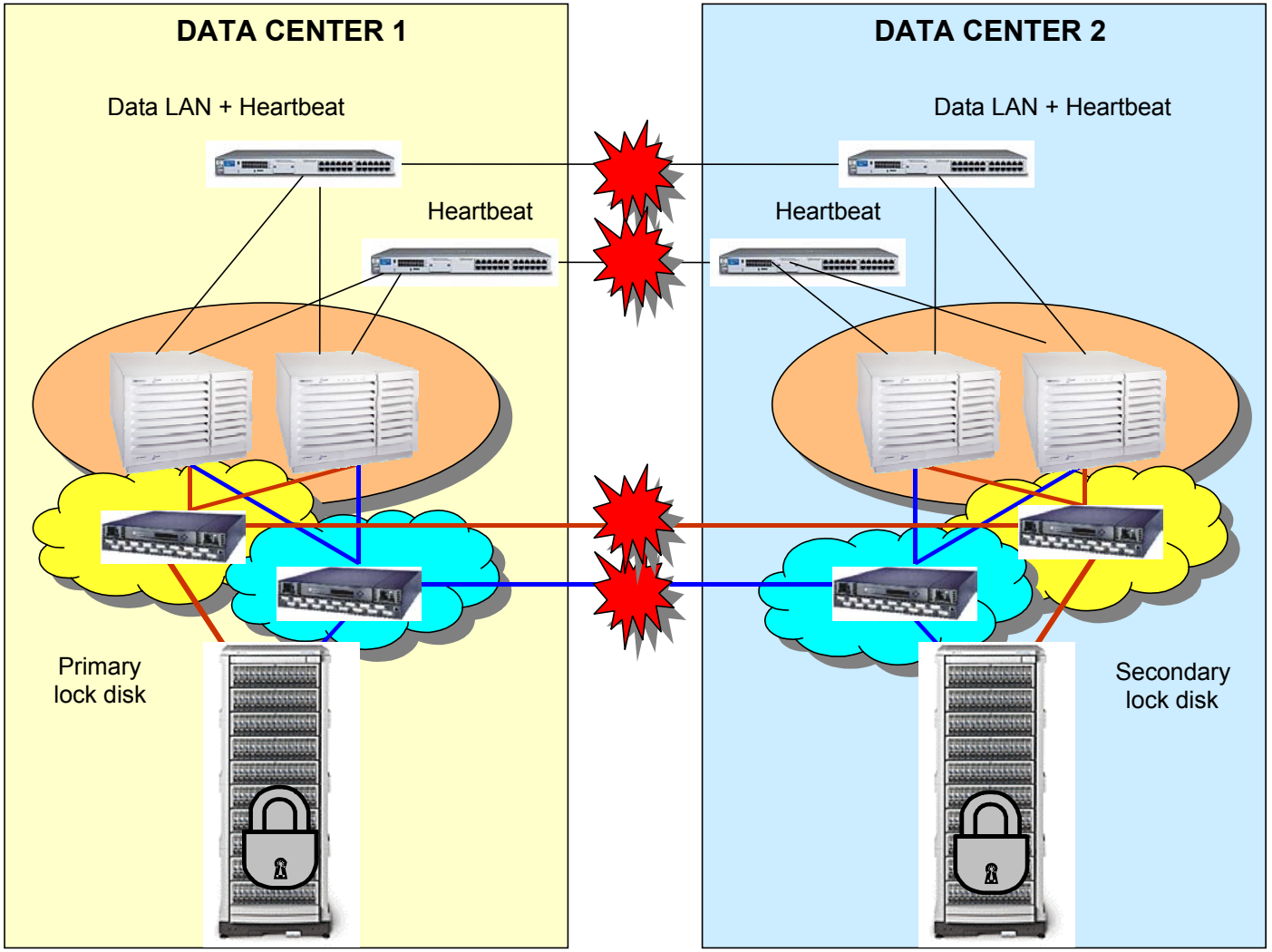
Split-Brain Syndrome

- Three-data center clusters and two-data center clusters using cluster quorum server cannot suffer split-brain syndrome
- Only a slight chance of split-brain occurring in a properly designed and deployed solution

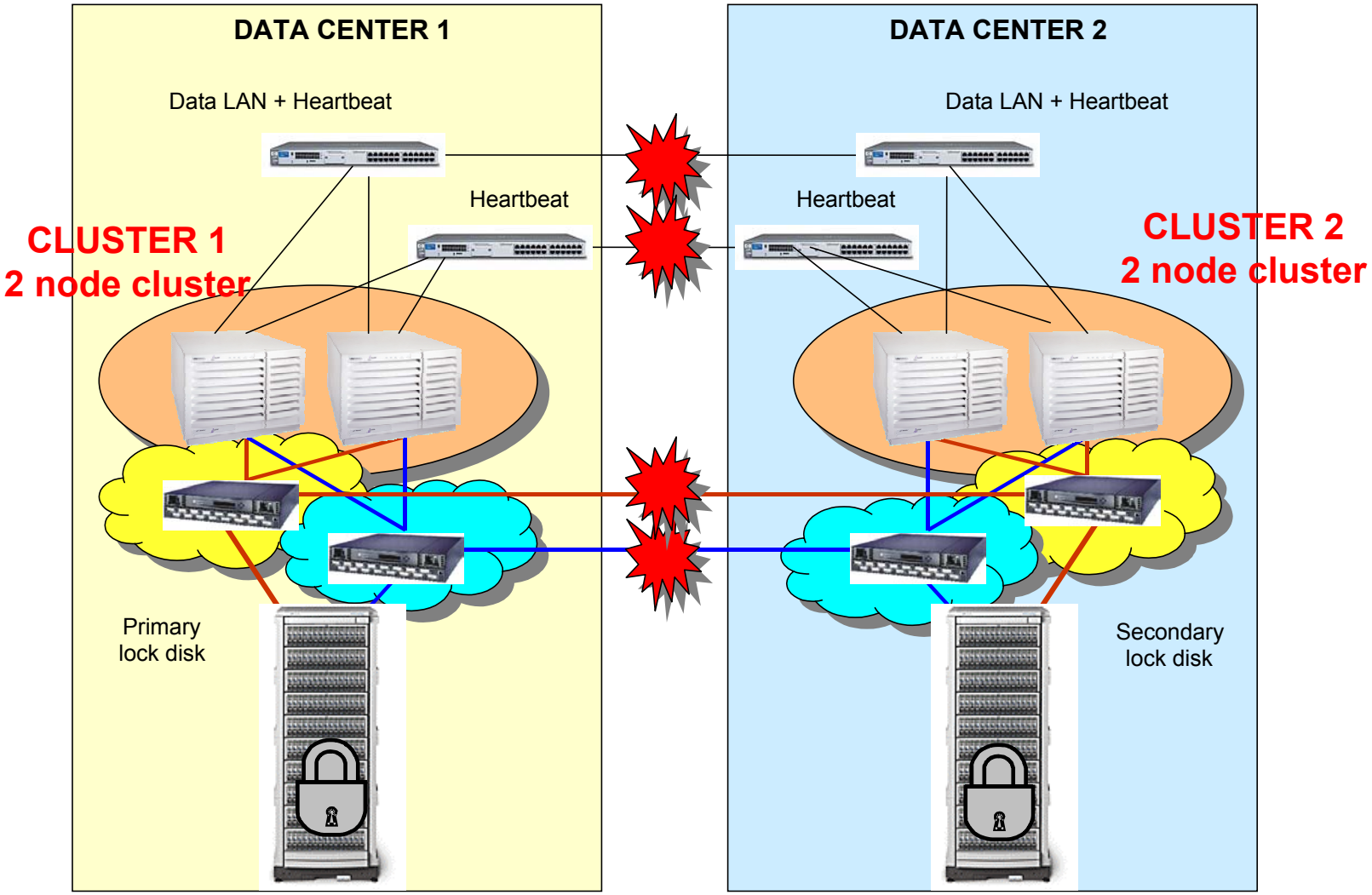
Split-Brain example: Single cluster before failure



Split-Brain example: Multiple connection failure



Split-Brain example: Two clusters form



Other Issues

- No Oracle Ops support
- Not supported with Veritas Volume Manager mirroring