



EMEA Engineering and Integration

Frankfurt Data Centre SAN Infrastructure and Storage Design

Document Status Release

Design Paper

Revision 5.0

Version Control

Version and Date of Issue	Author	Peer Review	Comment
Draft 01/06/2007	Colin N Bradley	Storage ENG, PIT, DCIP	Draft RFC
Release 17/07/2007	Colin N Bradley	-	Updated after RFC
Version 3 14/08/2007	Ade Okusanya		Updated SAN storage proposed solution
Version 4 28/08/2007	Colin Bradley		Updated SWAN Design
Version 4	Pat Fogarty		Revised Storage Requirement
Version 4	Carl P. Williams		Cabling Requirements and Infrastructure diagrams
Version 5	Ade Okusanya	Carl Woolward, Steven Hopkins, Adrian Pengelly, Pat Fogarty, Parul Jain	Updated After RFC
Version 6	Ade Okusanya	Pat Fogarty	Additional details added to storage section



Legal Notice

Information contained herein is for Internal Use and may be used only for business purposes authorized by Citigroup.

CITI, CITIBANK, CITICORP, CITIGROUP and the Umbrella Device are trademarks and service marks of Citicorp and are used and registered throughout the world. Citibank, Citicorp, Citigroup and their subsidiaries also claim rights in certain other trademarks and service marks contained in these pages.

Contact

Please address any comments or questions regarding the information contained in this document via email to the “*EMEA Infrastructure Engineering” distribution list.

Latest Versions

It is recommended that the latest version of the document be used from the Infrastructure Engineering website <http://www.emea.citigroup.net/emeainfrastructureengineering>. Newer versions may include expanded scope, or comments on issues discovered after deployment.

Table of Contents

Version Control.....	2
Legal Notice.....	3
Contact.....	3
Latest Versions	3
Table of Contents.....	4
1 Introduction.....	5
1.1 Audience	5
2 Document Scope.....	6
2.1 In Scope Phase 1	6
2.2 Out of Scope Phase 1.....	6
2.3 In Scope Phase 2.....	6
2.4 In Scope Phase 3.....	Error! Bookmark not defined.
2.5 Requirements	Error! Bookmark not defined.
3 Priority Applications.....	8
4 Infrastructure.....	11
4.1 Switch Infrastructure Hardware	11
4.2 Edge Switch Architecture.....	12
4.2.1 Core & Backbone Switch Architecture.....	12
4.3 Port Bandwidth.....	13
4.3.1 Host to Switch Connectivity	13
4.3.2 Switch to Switch Bandwidth.....	14
4.3.3 Switch to Storage Bandwidth.....	14
5 Storage Infrastructure Hardware	15
5.1 Storage Protection Type	15
5.2 Asynchronous Remote Replication.....	15
5.2.1 SRDF/A Group.....	16
6 Network Infrastructure.....	17
6.1 Storage Wide Area Network	17
7 Data Centre Topology.....	18
7.1 Riverdale Data Centre (RDC)	18
7.2 Frankfurt Data Centre (FDC)	18
8 Proposed SAN Solutions	19
8.1 SAN Topology.....	19
8.1.1 Fabric Naming	20
8.2 SAN Configuration for RDC using MDS 9513.....	20
8.3 SAN Configuration for FDC using MDS 9513	23
8.4 Scalability of FDC SAN using MDS 9513 Switches	26
8.5 SAN Management	27
9 Storage Requirements.....	29
9.1 Timefinder BCVs or Timefinder Clones at the Remote Site.....	29
10 Replication SAN.....	33
10.1 Predicted Bandwidth Utilisation	35
10.1.1 Mainframe Estimated Bandwidth Utilisations.....	35
10.1.2 Distributed Measured Bandwidth Utilisations.....	36
10.2 Replication SAN Hardware.....	36
11 Cabling Requirements	37
12 References.....	39
13 Appendix A FDC Floor Layout.....	40
14 Appendix B using MDS 9509	41
15 Appendix C FDC room 1A SAN Topology.....	42
16 Appendix D Replication Fabric Topology.....	44
17 Appendix E RDC SAN Topology.....	45

1 Introduction

The proximity of the London data centres Citigroup Centre (CGC) and Riverdale Data Centre (RDC) has been identified as a business risk. As such an out of region solution is being developed where applications in London will be remotely replicated to a “new” COB data centre located in Frankfurt. This Frankfurt Data Centre will have a dual purpose of providing COB for production applications in London and for hosting German production applications for asynchronous replication to its COB in London.

This document details the Riverdale Data Centre (RDC) and Frankfurt Data Centre (FDC) SAN Infrastructure, Storage and SAN management design. The document will also contain the Storage Wide Area Network (SWAN) design. Due to the complexity and scale of the design there are a number of different options that need to be considered, therefore this document will present a number of solutions that will fulfil the requirements of the SAN infrastructure. The advantages and disadvantages of each will be listed and the recommended design put forward.

Due to the scale of the design the document has been broken down into a number of phases:

1. Phase 1 – will focus on the infrastructure design options, network connectivity between locations and cabling/connectivity infrastructure design.
2. Phase 2 – will focus on the management infrastructure needed to manage the SAN infrastructure that will be deployed.
3. Phase 3 - will focus on the more detailed aspects of the design i.e. mapping of applications to storage arrays, host-to-port mappings etc. Due for completion end of Q4.

1.1 Audience

This document has been written specifically for the personnel involved in the Frankfurt Data Centre project with a strong emphasis on those engaged in the design and implementation of the data centre. Knowledge of SAN technology is preferred.

2 Document Scope

Due to the phased approach that is being taken with the design the document scope has been broken down so as to make the scope of each initiative clear. The scope for each of the phases is presented in the following sections.

2.1 In Scope Storage Infrastructure

- All designs will adhere to Citigroup Engineering best Practices
- The SAN fabric design will be a scalable dual fabric Core-Edge Topology
- Detail the SAN Infrastructure design for Riverdale Data Centre (RDC).
 - SAN design will be based on MDS95XX architecture
 - SAN design will include recommendations for VSAN deployment
 - Inter-VSAN routing will form part of the design
- Detail the SAN fabric design for Frankfurt Data Centre (FDC).
 - SAN design will be based on MDS95xx architecture
 - SAN design will include recommendations for VSAN deployment
 - Inter-VSAN routing will form part of the design
- The design will include the Storage Wide Area Network (SWAN) sizing and fabric
 - SWAN design will include SRDF/A network between RDC and FDC
 - SWAN design will include connection between RDC and CGC
 - SWAN design will include connection between FDC and Welle
 - SWAN design will include connection between CGC and Welle
 - SWAN design will be scoped for all distributed applications
 - SWAN design will be scoped such that mainframe asynchronous replication traffic can be accommodated on the same fabric.
- The design will only provide sufficient capacity for phase 1 priority applications that will be remediated by year end
 - Sufficient port count
 - Sufficient storage capacity
 - Sufficient WAN bandwidth
- All designs will provide adequate scalability to accommodate all remaining applications to be migrated and reasonable future growth.
- Asynchronous block data replication will be provided between RDC and FDC; this will only be available on the new Cisco SAN fabric.

2.2 Out of Scope Storage Infrastructure

- Everything else is out of scope for phase 1

2.3 In Scope Management Infrastructure

- The SAN Management Infrastructure designs will be completed as part of this document, this will include;
 - Solutions Enabler
 - SANScreen
 - Fabric Manager
 - ECC
 - Storage Navigator
 - Vendor Remote Access
 These designs will be completed with input from each of the vendors.

2.4 Detail Design

- Zoning information
- Application to array mapping
- VSAN information
- Inter-VSAN routing information
- SRDF/A RA group to application Mapping

2.5 Deliverables

Through discussions with the EMEA storage manager and the GDCS program the following have been identified as high-level Deliverables

- Provide detailed design of RDC and FDC Cisco SAN Fabric
 - Ensure designs are scalable to meet Citigroup future growth requirements
 - Ensure designs are flexible to meet Citigroup business demands
 - Ensure designs are highly available to meet Citigroup business commitments
- Ensure designs are cost effective to meet Citigroup budgetary commitments

3 Priority Applications

Priority applications have been defined as those applications that are primarily involved in clearing and settlement. These applications are concentrated in two business and they are affected by the proximity risk of the London production and COB data centres. They are GTS and GCOT encompassing CMB.

The first and foremost goal of the FDC project is to ensure that all priority applications have their proximity risk resolved by 15th of December 2008. Table 1 below is a list of the priority applications. The details in table 1 were compiled by the Application Analysis Stream. This list doesn't include the applications that have a dependency on one or more of those listed below. The eventual list should state the type of dependency and how it might be impacted during an application failover. Identifying where relevant the need for inter-application consistency as well as data consistency. This is a crucial piece of information in designing the replication solution and has been highlighted to the programme as a risk.

PRIORITY APPLICATION	PRIORITY APPLICATION	PRIORITY APPLICATION
Account Server- London	Customer Directory Management System	Global Transfer Agency System (Poland)
ATIS	Customer Information System	Global Transfer agency System (UK)
Business Activity Monitoring (EMEA)	Direct Debit Management System	GMI Interfaces
Cairo	EBA Step 2 Participant System	Hamper
CBII	EMEA Clear Portfolio (Confluence)	IBM Cosmos Frankfurt
Checks – 32	Enhanced Transmission Service Client Gateway	IBM Cosmos Milan
Citi9	Equity-e	IBM Liquidity Manager
CitiDirect for Funds Unit Matching	ETEBAC - FR	IBM Message Delivery
CitiDirect IPA	EUCLID PC	ILOC Customer Functions
CitiDocs (CEEMEA)	Eurocars	IOWA
CitiGateway	Euroclear High Speed Link	ISSAC-Intergrated Swiftnet SVCS access for Citig
CitiNetting	eXtensible Security Management System	ITA GARI
CitiSwitch EMEA	Faster Payments System	Liquidity Monitoring System
Client Money	Flexcube (CEEMEA)	Loans Agency Flexcube Product Processor – UK
Collateral Debt Obligation System	FT Engine	MessageWay
Communications Sub-System	FTI Front-End	Millennium
CPC	Generic Rec	Multifonds Fund accounting (Global)
CPSE (Cross Processing Services Europe)	Global Interdict (EMEA)	New corporate Actions Server

PRIORITY APPLICATION	PRIORITY APPLICATION	PRIORITY APPLICATION
Orbiflow	SecRouter	SIT++ - FR
Quantum TSG	Secure	SMAC2000
Recon Frontier	Securities Front End –Europe	SMI – Spain
ReconTrade	Securities Market interface – Greece	SST- Monitor
Regional Balance Control Systems	Securities Market Interface – Milan	SWIFTNet SAG (SWIFT Alliance Gateway)
Regional Comms Server	Securities Market Interface – Paris	Timestamp
Regional Communications Exchange	Securities Market Interface – Zurich	Tradebus
Relationship Management Application	Securities Market Interface- Amsterdam	Transaction Reporting
SAM – Satellite application for Messaging	Securities Market Interface- CREST	Western Europe Flexcube
Sanctions Check	Securities Market Interface- Frankfurt	Zero Balancing /X-BDR Sweep
Saturn	Securities Market Interface- Toronto	
Secore – EMEA	SFE – NikkoCitiTrust	

Table 1: GTS Application and Server Count

The table below shows the combined storage requirements for the GCOT & CMB applications based at CGC and RDC.

Business	Data	Contingency	Production	Grand Total
GCOT & CMB	Sum of Local Storage (GB)	1,921.88	1,921.88	3,843.76
	Sum of SRDF Storage (GB)	4108.85	5,596.96	9,777.81

Table 2: Combined GCOT Storage and Port Requirements

The table below shows the storage requirements for GTS applications based at CGC and RDC.

Business	Data	Contingency	Production	Grand Total
GTS	Sum of Local Storage (GB)	373.1	340.16	66,076.44
	Sum of SRDF Storage (GB)	50,490.77	76,924.73	127,415.50

Table 3: GTS Storage Requirements

The table below shows the port requirements for priority applications based at RDC and CGC.

Location	Data	RED Fabric	BLUE Fabric	Grand Total
RDC	Sum of Number of Ports	65	65	130
Location	Data	RED Fabric	BLUE Fabric	Grand Total
CGC	Sum of Number of Ports	84	84	168

Table 4: Port Usage for RDC and CGC

4 Infrastructure

4.1 Switch Infrastructure Hardware

Citigroup engineering has certified the Cisco MDS 95xx range of switches as the platform of choice for future Citigroup SAN implementations. As such this design document will focus solely on the MDS 95xx range of switches and will use these to accommodate the port requirements of all servers and storage arrays.

The MDS9513 switch is the largest of the MDS95xx range and has 11 slots in which line-cards (LC) can be inserted. However, a single slot in each switch will be reserved for future functionality i.e. the Storage Services Module (SSM).

The MDS range of switches has 3 Generation-2 Line cards that can be inserted into each of the slots. Generation two Line cards increase the port density to 528 compared with a 256 port restriction when using first generation line cards. Below is a table of the three generation 2 Line cards.

	1 Gbps FC	2 Gbps FC	4 Gbps FC
12 port	Full Rate	Full Rate	Full Rate
24 port	Full Rate	Full Rate	Oversubscribed 2:1
48 port	Full Rate	Oversubscribed 2:1	Oversubscribed 4:1

Table 5: Cisco MDS Generation 2 Line Cards

The MDS9513 switch provides maximum flexibility and scalability to meet the demands of the Citigroup environment and as such it is preferable to use this switch to meet the edge and core requirements of the Frankfurt data centre.

These switches will also be deployed for the Storage Wide Area Network (SWAN) that will be used for replication between RDC and FDC. FCIP will be the protocol used to transport replication data between these Sites. A section of this document will provide an overview of the infrastructure and the assumptions made on the design.

The introduction of these switches, enable us to take advantage of oversubscription. This is a concept that is presently unique to CISCO Fibre Channel Networks. Oversubscription is the ratio of potential bandwidth to available bandwidth. A line card has a total bandwidth of 51.8GB; this is shared by all ports on that line card. A 12 port line card can operate all its ports at 4Gbps full line rate, as the number of ports increase above 12 the bandwidth available to each port decreases resulting into oversubscription. An analysis of host port utilisation in our environment reveals that most hosts barely use 20% of their 2 GB bandwidth, opening up the opportunity to deploy 24 & 48 port cards. Therefore using oversubscription a fabric with more ports may be deployed for a similar price to a smaller fabric without the use of oversubscription. They will provide our customers with alternatives that suit their requirements based on price and performance. Even though these cards have an oversubscription ratio of 2:1 and 4:1 they can be used where 2Gbps and 1Gbps speeds are suitable. The majority of our hosts fall into this category.

4.2 Edge Switch Architecture

The installation of 24 & 48 port Line cards on the edge switch will adequately meet host requirements for the 88 priority applications and the phase of migrations. In line with GDSE recommendations 12 port line cards will be installed for inter switch links between the Edge and Core switches. This will ensure that ISL's operate at 4Gbps full line rate. Figure1 is a picture of a partially populated MDS9513.

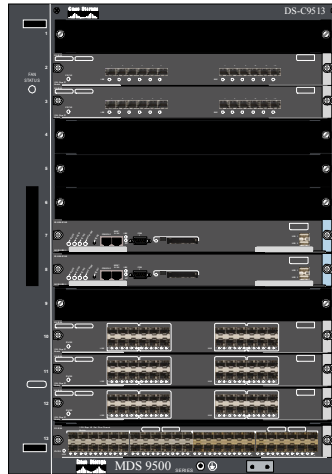


Figure 1: Cisco MDS9513 Switch

The decision to select 24 and 48 port line cards for host connectivity based on two assumptions;

- Any existing production/COB hosts using 2*2Gbps connectivity to the SAN Fabric will continue to use a 2*2Gbps connections.
- Development hosts and UAT hosts that are dedicated to that purpose only require 1Gbps links.

Both factors have been presented to the BAU and Integration team for input. A definitive response has not been obtained on using the 48 port LC but discussions are still ongoing. All edge switches will connect to both the Storage and Backup Core switches with VSAN's deployed for segmentation. This mixed environment is subject to final approval.

4.2.1 Core & Backbone Switch Architecture

As with the edge switch, the recommended hardware for the backbone and core is the MDS9513 Cisco director. However, unlike the edge switch, the choice of line-card for inclusion in the core and backbone switch will be limited to the 12-port line-card. There are a number of reasons driving this selection, and these are:

- 12-port line-cards provide 4Gbps FLR with no oversubscription.
- Each individual enhanced TE-port will run at 4Gbps.
- Connectivity to storage will run at 4Gbps.

The obvious disadvantage of this line card is the low port to switch ratio when compared to the edge switches. The increased bandwidth per port makes up for the reduced port count. Port channelling (aggregation of ISL links into a trunk sharing bandwidth between virtual fabrics) will provide resilience and increased availability. Sufficient bandwidth will ensure the "front-end"¹ bandwidth is not a constraining factor in the performance of an application.

¹ Here the Front-End is being defined as the logical path from HBA to Storage Cache and everything in between.

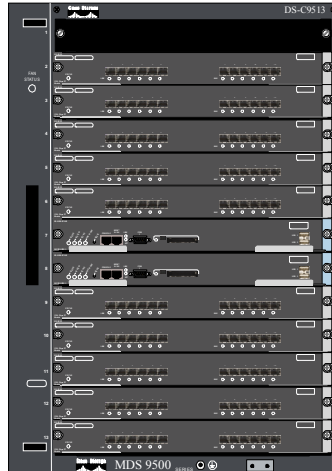
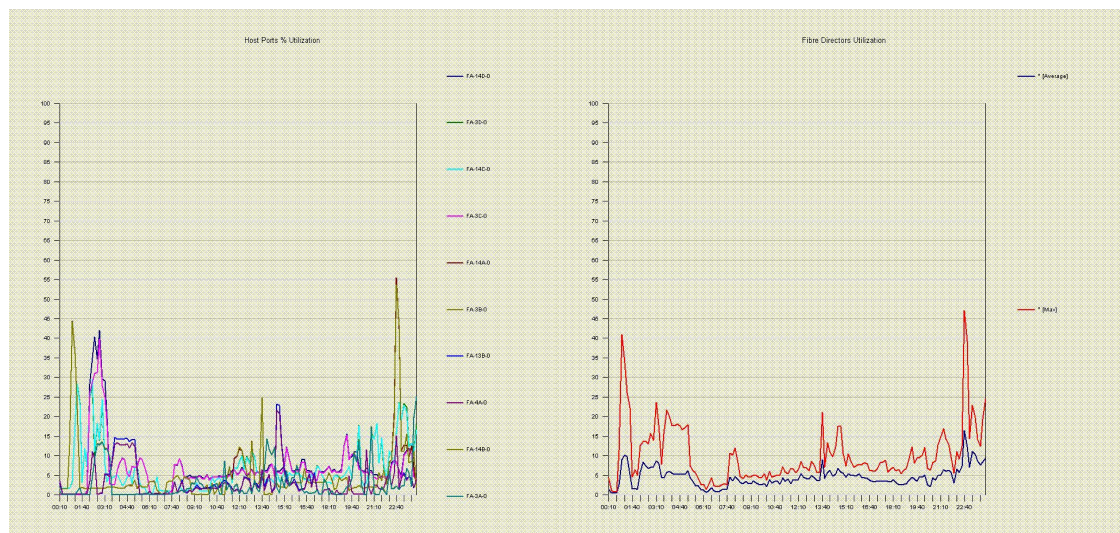


Figure 2: Cisco MDS9513 Switch

4.3 Port Bandwidth

4.3.1 Host to Switch Connectivity

The existing host to switch bandwidth that each host has been allocated is 2Gbps. The current plan is to provide equivalent bandwidth to each of the priority applications on the new infrastructure. However throughput analysis using ECC Performance Manager indicates that in general, each server uses only a fraction of this bandwidth. This analysis was done at the DMX port level, where typically averages of 10 servers share the port bandwidth. An example of the throughput that was observed is shown graphically below. Full page graphs are available in appendix F and G.



Graph 1: Graphs showing the Port Utilisation and Fibre Director Utilisation

This trend is consistent over all the DMX storage arrays in London. These statistics raise a question over the actual physical bandwidth requirements of each server. For planning purposes 2Gbps for each production host to switch connection will be used. This would result in the 24-port LC being used as the blade of choice for production servers with an oversubscription² ratio of 2:1.

² Oversubscription is the ratio of potential bandwidth to available bandwidth.

4.3.2 Switch to Switch Bandwidth

The switch-to-switch connectivity is provided by Inter-Switch Links (ISL's). ISL's provide a data path for data to flow between switches. Cisco uses different terminology when referring to ISL's; they are sometimes referred to as "enhanced E-ports". A minimum of two ISL's are required to form a Port Channel and a maximum of 16.

This document will follow as closely as possible the Citigroup best practices for the ISL sizing and design between switches. The main points of this best practice are:

- Assume fan-in ratio on the storage ports to be 10:1. This is a rule of thumb and can be modified depending on performance requirements if known.
- Switches should be interconnected by Port Channels. Each Port Channel should have at least two member links (ISL's), and each Port Channel member link should come from a different line card. This way the Port Channel can survive a line card failure.
- Each Port Channel can be scaled to 16 member links, and that should provide enough bandwidth for most topologies. A second Port Channel should only be considered when the first is fully populated with 16 links. Having a bigger Port Channel is desirable over having multiple Port Channels because failure of a Channel causes disruptive FSPF convergence to be triggered unlike the failure of a member link within a Channel.
- On the core and backbone switch, ISL's should always be run from full line rate 4 Gbps ports found on the 12 port line card. If 9506 or 9509 switches are being used on the edge, ISL's can terminate on 24 port or 48 port line cards after dedicating 4 Gbps bandwidth to the selected ports. Because smaller switches have a limited number of slots, dedicating two slots to 12 port line cards to run ISL's would be inefficient. Please note that dedicating bandwidth to one or more ports in a port group increases the oversubscription ratio for the rest of the ports in that port group. Oversubscription ratio should be managed by either taking ports out of service from that port group, or by distributing ISL ports between port groups, or both.
- The following can be used to calculate number of members in the Port Channel:

If number of host ports that will use this Port Channel is H, and each host connects with S Gbps speed, and the storage port fan-in ratio is F

Total port channel bandwidth required = $(H * S)/F$

Number of port channel members = $(H * S)/F/4$ Example: If there are 100 host ports operating at 2 Gbps and fan-in ratio is 10, number of members in the port channel = 5.

- The entire Port Channel should be provisioned upfront; although MDS allows for non-disruptive addition of members to the Port Channel, Operations may run out of ports by over-allocating ports to hosts, or may even forget to address Port Channel bandwidth as hosts are being added to the fabric.

4.3.3 Switch to Storage Bandwidth

Current switch to storage connectivity is provided by 2Gbps links. The introduction of storage arrays with 4Gbps front end ports from vendors, as well as Line Cards that are capable of 4Gbps at full line rate, requires the design to address these developments. On that basis 12 port line cards will be installed in the MDS95xx Core switches for storage connectivity. This is more than adequate to meet bandwidth requirements as the present work load on existing arrays are hardly pushing utilisation on the 2Gbps front end ports despite a fan in ratio of 14:1 in some instances.

5 Storage Infrastructure Hardware

5.1 Storage Protection Type

BAU and integration in EMEA have agreed four storage offerings with their customers. Each offer provides a different level of protection and price. The new infrastructure will be deployed in line with these recommendations. Since it is not a departure from what already obtains in the environment, save for an extra option, the following decisions have been made.

- Any priority application that is using RAID-1 storage will use RAID-1 storage on the new infrastructure.
- Any non-replicated storage being used for anything other than Production or Contingency will be migrated to a Lower Cost storage tier, as determined by the Storage Optimization initiative.

The above decisions might be changed as the project progresses due to a desire to move many of the applications to RAID-5 storage. The move from RAID-1 to RAID-5 will require proper analysis and understanding of potential impacts. Identifying appropriate applications for moves to RAID-5 storage is being handled by the Storage Optimization initiative.

In anticipation of these developments the storage requirements will be sized with both Raid types, to accommodate a move of applications from RAID-1 to RAID-5. Clarification of application requirements should be obtained from the **Application Analysis project stream**.

5.2 Asynchronous Remote Replication

The Current storage based replication in use between CGC and RDC is an EMC SRDF/S implementation. Due to the distance between RDC and FDC and latency involved when sending data such distance it is not technically feasible to implement storage based synchronous replication between the two data centres. As such only asynchronous storage replication will be offered between RDC and FDC.

Asynchronous replication is a remote mirroring solution for the symmetrix DMX. Unlike synchronous replication there is seconds of data exposure and has a longer distance restriction. The current Citigroup certified storage asynchronous replication solution is based on EMCs SRDF/A technology and is discussed in this section.

SRDF/A, at its simplest description, replicates data in discrete chunks called “delta sets” from the production array disk devices (R1), to the remote COB array disk devices (R2). SRDF/A provides a re-startable copy of the data on the R2 device. The nature of the asynchronous replication is such that the R2 device lags behind the R1 device, so in the event of a failure at the R1 end it is likely that data at the R2 end is a time interval behind the R1 device. It is therefore necessary to properly identify the Recovery Point Objective of each application before it can be considered suitable for SRDF/A replication. However, it has been necessary to make certain assumptions regarding SRDF/A for the purposes of this design document.

- “SRDF/A” functionality will only be made available on a new Cisco SAN infrastructure between RDC and FDC. The existing McData SAN will not offer “SRDF/A” functionality.
- HDS asynchronous replication HUR will not be certified by GDSE in the timescales required to deliver an asynchronous replication solution for priority applications.
- Any Phase 1 application that is now using SRDF/S for COB/DR will use SRDF/A or equivalent on the new infrastructure for replication.

The solution will be based on recommendations outlined in the GDSE Solutions Guide. The URL <http://engineering.citigroup.net/gdse/products/platforms/symmetrix/stdsdocs/05002-sg.pdf> points to the location of that document. SRDF/A is a complex solution and it is not the purpose of this document to cover the technical aspects of the technology.

In line with the guidelines stipulated by the solutions guide, EMC professional services will be engaged to produce the technical design and sizing of this environment. It must be noted that certain aspects of the design can only be achieved by the vendor. These tasks make use of proprietary software that is not released to customers. Feeding into this solution will be information gathered by the “Application Analysis” stream of the program.

The complete design will include the following information:

- Application to storage array mapping
- SRDF group to SRDF port mapping
- SRDF device to SRDF group mapping
- Cache requirements for each storage array
- Network bandwidth³ per storage array

5.2.1 SRDF/A Group

As has been mentioned previously the planning of the SRDF/A solution needs to be worked on closely with the vendor. One of the critical points in the planning process will be the mapping of applications to DMX arrays.

A primary difference between the implementation of SRDF/S and SRDF/A is that the fail-over is performed at the SRDF group level rather than at the SRDF device level. This has some serious implications and needs to be planned carefully to ensure that applications can be failed over individually.

For performance reasons the maximum number of dynamic SRDF groups that should be configured per RF has been advised by EMC performance engineering to be 6 not including an underlying static group. Excluding the underlying static group a pair of RFs can support 6 SRDF groups, 2 pairs of RFs can support 12 SRDF groups⁴ etc. This limitation has an impact on the number of host front end connections that can be supported. As the number of applications on a DMX increases beyond 12, then the additional RFs configured reduce the number of ports available for host connectivity⁵. EMC expect to announce an increase in the number of RA groups per port with the release of their latest microcode. If this is certified before the detailed design phase then the design will take advantage of this development.

³ Work has already been carried out as part of the SWAN analysis to estimate the network bandwidth to accommodate all current applications.

⁴ This is set to increase in 5772 code.

⁵ Assuming that port 1 of each FA is not used as per best practice recommendations.

6 Network Infrastructure

6.1 Storage Wide Area Network

To support asynchronous array based replication between RDC and FDC a dedicated Storage Wide Area Network (SWAN) is required. The SWAN design is specific to providing a redundant FCIP network fully dedicated to storage between RDC and FDC. The need for a dedicated storage network is driven by:

- Guaranteed bandwidth for asynchronous replication to ensure replication technologies will work as designed.
- A dedicated network will avoid any impact to production services that already exist on the corporate network.

The FCIP network will follow diverse routes and each FCIP “link” will have 10Gbps bandwidth.

To ensure that there is no short term need to upgrade the network once the migration of remaining applications begins; the 10Gbps link provided will accommodate the replication bandwidth requirements of all applications that currently use the synchronous array based replication SRDF/S that is in place today.

Cisco do not currently have a generation 2⁶ line card that supports Gigabit Ethernet (Gig E), therefore the only option that remains is to utilise generation 1 line cards, specifically the MPS 14/2 card Figure 1, which provides:

- 14x 1-2 Gbps FC ports
- FRM
- 2X Gig E ports
- FCIP and iSCSI support

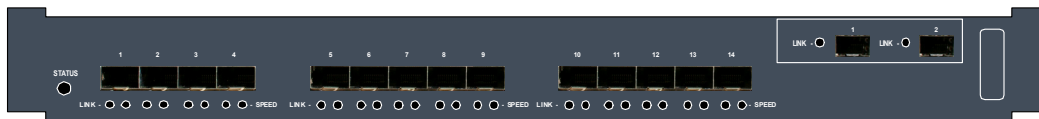


Figure 3: MPS-14/2 Blade

As asynchronous replication is latency sensitive it is imperative that replicating storage arrays do not negatively impact each other. Hopefully, the compression feature of the MPS line card and the 10Gbps bandwidth will ensure that there will be no contention on the link. More details on this subject can be found in the Replication SAN section of this document.

⁶ Generation 2 line-cards have a number of benefits over their Generation-1 counter parts e.g. the number of addresses that a Gen-2 line card can support is 1024, a Gen-1 line card is substantially lower at 256.

7 Data Centre Topology

This phase project consists of building two SAN environments in two different locations:

- Riverdale Data Centre (RDC)
- Frankfurt Data Centre (FDC)

This presents its own challenges as the physical infrastructure topology of each location is different. At each location there will be multiple fabrics, fabrics in this case defined as the physical connection of switches together, and in each fabric there will potentially be a number of Virtual SANs (VSAN's).

The next sections will outline the physical topology in each of the locations defined above.

7.1 Riverdale Data Centre (RDC)

RDC is an existing Data Centre in London. It contains a mix of production and COB services. A separate project will ensure that all production servers run out of RDC. CGC the target of the flip from RDC will take on the role as the COB datacentre, in anticipation of the migration of London COB to the FDC. For the purpose of this project a new SAN environment will be commissioned and will run in parallel with the existing SAN infrastructure. As RDC is a fully operational data centre the challenges that are presented here will be different from those in FDC. They are:

- This is a production data centre and as such is subject to stringent change control procedures.
- The data centre is split over a number of floors.
- The priority applications are distributed across these floors and this will impact the number of edge-switches that will be deployed.

There are 3 assumptions that are being made here:

1. Edge switches will need to be placed on each floor to accommodate the priority applications and eventually all other applications.
2. RDC level 4 will be used for the core switches and the storage arrays.
3. Any dual purpose hosts i.e. providing Contingency and another function such as UAT will be moved to FDC.

These assumptions have been tested and have been found to be correct. Therefore there will be no relocation of priority applications to RDC level 4 and any host connectivity will need to be provided by Cisco edge switches placed on every floor.

7.2 Frankfurt Data Centre (FDC)

FDC is located in Kallbach Frankfurt Germany. The FDC is a green field site. Construction of this datacentre at the time of writing has commenced. Early TI access to this data centre will be the 29th of February 2008.

Data Centre Site Management (DCIP) has taken the decision to modularise the data centre layout. There will be SAN attached modules and a non-SAN attached modules. Some of the SAN attached modules will be used for virtualised server farms and supplied with the necessary network infrastructure required for that technology.

In each SAN attached module there will be 56 SAN attached cabinets with cabling to support a maximum of 12 servers per cabinet. This equates to a potential of 672 servers per module. DCIP have settled on a figure of 482 hosts per module.

8 Proposed SAN Solutions

As the application analysis has not yet been completed the proposed solutions in this document are based on:

- Current knowledge of the existing environment
- Feedback from the Storage Optimisation initiative
- Assumptions outlined in this document and in the scope document
- Consultations with Storage Integration and EMEA Storage Manager
- GDSE certifications on Cisco MDS switches, Storage array Certifications and SRDF/A solutions guide

As per the GDSE Cisco MDS solutions guide, the solution proposed will follow a Core-Edge SAN topology design. It will also include a slight variation in one of the datacentres. This is covered in more detail in the next section.

****NOTE**

Data Centre Infrastructure Planning (DCIP) had concerns centred on the physical racking of the MDS 9513 switch (which is only supported in a 4 post rack). Calculations for a MDS 9509 switch reveal its lack of suitability and will not be in scope for the Storage to Host SAN. DCIP have agreed to install the MDS9513 and provide infrastructure to support its installation.

8.1 SAN Topology

The Core-Edge topology is a highly scalable design and allows the separation of host connectivity to the fabric restricted to the high port density edge switches and storage connectivity centralised in the core switches. A view of a simple core edge design is shown in Figure 4 below.

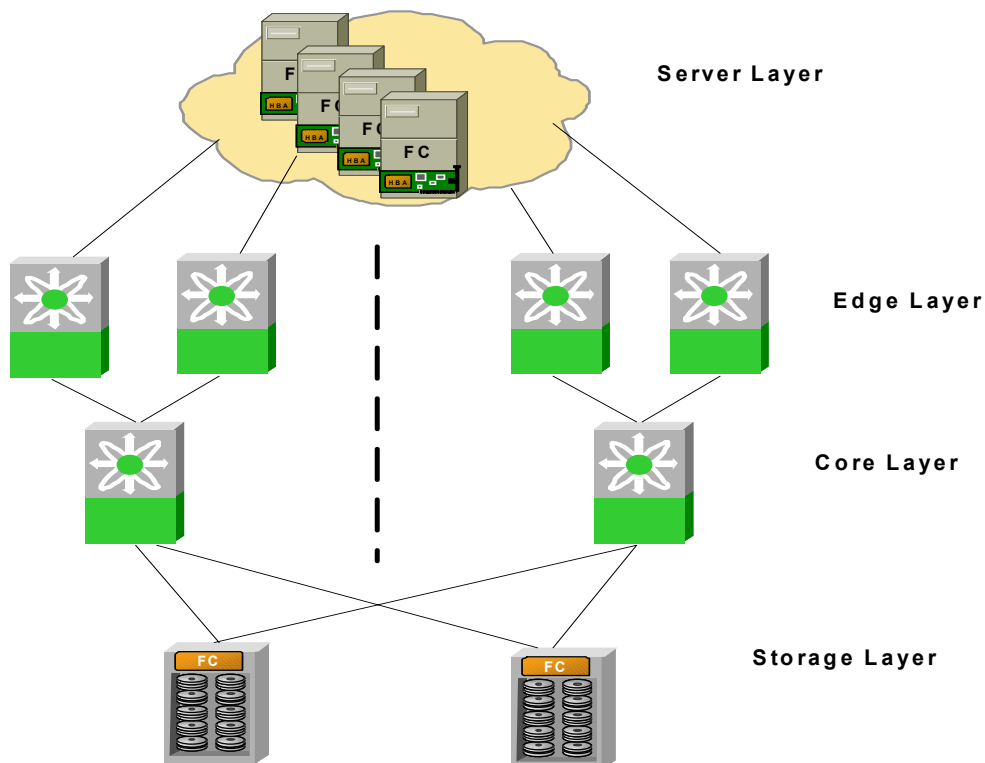


Figure 4: Simple Dual Core-Edge Topology

8.1.1 Fabric Naming

To minimize confusion the new fabrics will have different code names than the existing McData SAN Fabric. This will ensure no error can be made when documenting or discussing the fabrics. The new code names are:

- Jade = Host to storage Fabric 1
- Ruby = Host to Storage Fabric 2
- Gold = SRDF Fabric 1
- Silver = SRDF Fabric 2

This convention has been agreed for engineering SAN and Backup design documentation. As a point to note, there will be a separate Backup SAN fabric, but this is outside the scope of this document.

8.2 SAN Configuration for RDC using MDS 9513

The classic core edge topology has been chosen for the RDC. This design provides the scalability and flexibility required. As mentioned previously the priority applications are split across many different floors in RDC. The new fabric design will accomplish the ultimate goal of the program to migrate all applications onto it by Q3 2009. Edge switches will inevitably be required on every floor. The distance between the core switches which will be located on level 4 and other levels will not be an issue. The break down of what is possible on each floor is as follows.

- RDC Level 1 – No frame space available and very little cabinet space available.
- RDC Level 3 – Open Frame space available for Cisco switches next to existing SAN Infrastructure
- RDC Level 4 - Open Frame space available for Cisco switches next to existing SAN Infrastructure
- RDC Level 4 – Space for new DMX arrays to be installed here

RDC Level 1 presents the biggest issue in installing equipment as there is very little space available. It is likely, that the removal of the legacy Symm5 arrays will release floor space. This will accommodate, at the very minimum, a pair of MDS 9513 switches. The tables below show the breakdown of hardware SAN requirements for the Priority Applications at the RDC.

LINE CARD	NUMBER	SUM OF PORTS	FUNCTION
12 Port LC	2	24	ISL Connectivity
24 Port LC	4	96	Host Connectivity
48 Port LC	0	0	Host Connectivity

Table 6 Configuration of an Edge switch at the RDC for Ruby and Jade

LINE CARD	NUMBER	SUM OF PORTS	FUNCTION
12 Port LC	10	120	ISL and Storage Connectivity

Table 7 Configuration of a Core Switch at the RDC for Ruby and Jade

	Ruby Core	Ruby Edge	Jade Core	Jade Edge	Total
Switches	2	3	2	3	10
12 Port LC	20	6	20	6	52
24 Port LC	0	12	0	12	24
48 Port LC	0	0	0	0	0

Table 8 Breakdown of Switch and Line Card Requirement for RDC

The above figures represent the minimum requirement of hardware that should be in place for the first phase of migrations. These figures include a 60% growth potential that can be used by new applications that will be deployed in 2008, as well as other dependent applications that have not been identified at the time of writing. Further information on the SAN is given below:

Priority Application Host Ports = 108
 Total Port Count = 600 Per Fabric (Total SAN port count = 1200)
 No of Host Ports = 288
 No of storage Ports = 160 @ 4Gbps
 ISL Ports = 80
 Fan in Ratio is 10: 1
 Utilization = 38%

The following assumptions were made in reaching these figures

- The switches will be populated based on the tables above
- A storage array will have 32 front end ports evenly split between the fabrics
- The existing core is sized for a maximum of 10 storage arrays before an extra switch is required
- All edge switches will be populated with 24 port Line Cards
- Even though the above figures are based on partially populated frames, the ISL requirements are determined by the total number of host ports in a fully populated frame.
- Apart from 2 X 12 Port LC a populated frame was calculated with the remaining 8 slots evenly shared between 24 and 48 Port Line Cards. This results in a total of 288 ports per edge switch
- Total number of host ports in existing environment where extrapolated from ECC and Host lists created by the application analysis stream. These figures were used to determine port utilisation shown above as 38%.
- Host port per fabric requirement for production servers at RDC for Priority Application servers is 108 Ports. This figure includes a contingency of 20 ports for VMS servers and a minority of other servers that may not have host agents running. In some cases these servers might have 4 HBA's, which, is another factor taken into consideration.

- 2 Line Cards dedicated for Supervisors
- 2 Line Cards dedicated for 12-port ISL connectivity
- 1 Line Card dedicated for future SSM functionality
- The remaining 8 slots can be a combination of 24-port and 48-port Line Cards.

Hosts Per Edge Switch	288
Line Rate Per Host	2Gb
FAN:IN	12
ISL Port Line Rate	4
Number of ISL Ports Required	14
Number Of ISL Ports Per Core	7

Table 9: Number of ISL Ports Required Per EDGE Switch

For the initial install and support for priority applications, 6 ISL ports shared across core is more than sufficient. However, as all other applications are targeted for remediation by Q3 2009, it is advisable to install all necessary ISL ports in advance of their migration to the SAN. On that basis, the figure for ISL ports in table 9 reveals 14 ISL ports per switch split evenly in a port channel across both cores.

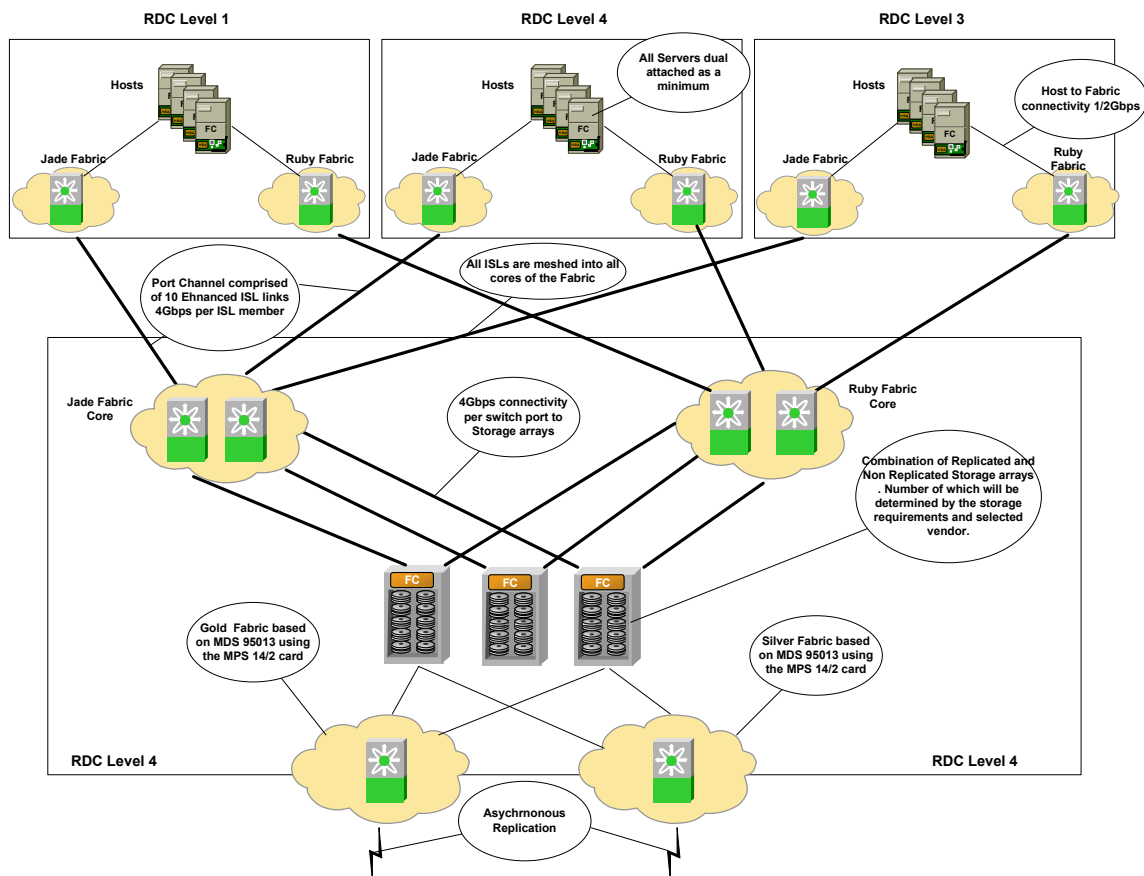


Figure 5: RDC SAN Topology

8.3 SAN Configuration for FDC using MDS 9513

FDC is a new data centre and has a different layout from the RDC. The preference is to mirror the technology and logical topology at the RDC in FDC. The topology will have a slight variation due to the floor layout, modular design and the large number of hosts the facility can accommodate. A three tier Topology will be deployed providing, Edge, Core and backbone connectivity. Hosts will be limited to the edge, storage to the core, and migration to other rooms through the backbone. The switch installed in all tiers will be the MDS9513.

The Modular layout is based around host classifications. There will be SAN attached modules, free standing modules for SUN 25Ks, IBM 595's and non-SAN attached modules. On the first floor there will be two rooms a Red room and a Green Room. A plan of the data centre layout can be viewed in Appendix B.

Below is a list of considerations that were taken into account during the design

- Fabrics will stretch across both rooms
- Backbone will only be used for migration between rooms
- Cluster nodes will be installed in the same module and room
- Blade servers with FC switches will connect edge switches
- 18 storage arrays per room
- Total SAN attached hosts will not exceed 1000 per room
- Host and storage provisioning will be limited to the same room
- Maximum front end connectivity per storage array is 32 ports
- VSAN's will be deployed to provide fault isolation due to increased size of the infrastructure
- Host bandwidth requirements fall within the following categories 1Gb, 2Gb and 4Gb.
- High port density provided by the MDS9513 will take advantage of the 48 port line cards.
- There is a physical limitation on the number of Edge switches per module. 4 in total.
Based on the present SAN design using high port density switches and the deployment of 48 port line cards in the future, should sufficiently meet port requirements
- Discussions held at the weekly Data Centre Strategy meeting, revealed the relocation of Servers classified as UAT and Development from London to the FDC in the near future. This has been factored into the planning of the FDC topology.

Some of the advantages of this solution are;

- Multiple system access to multiple storage devices, regardless of location. Resulting in flexibility
- With adequate procedures it will enhance the efficient utilization of storage resources
- Backbone layer provides connectivity between rooms enabling migration across the fabric.
- Choice of switch provides high port count at the Edge and Core reducing the total number of switches required
- High availability provided by two isolated fabrics Ruby and Jade
- VSAN's will provide isolation regardless of consolidation within high port density switches.
- Reduces hop count from host to storage from three in existing fabric to one.
- Design should meet all performance requirements

The following tables 10-13, provide a snapshot of the expected configuration of the Switches at each layer of the topology. The final table provides the sum of Switches and Line Cards required for the migration of Priority Applications.

LINE CARD	NUMBER	SUM OF PORTS	FUNCTION
12 Port LC	2	24	ISL Connectivity
24 Port LC	4	96	Host Connectivity
48 Port LC	0	0	Host Connectivity

Table 10: Configuration of an Edge Switch at the FDC for Ruby and Jade

LINE CARD	NUMBER	SUM OF PORTS	FUNCTION
12 Port LC	10	120	ISL and Storage Connectivity

Table 11: Configuration of a Core Switch at the FDC for Ruby and Jade

LINE CARD	NUMBER	SUM OF PORTS	FUNCTION
12 Port LC	2	24	ISL Connectivity

Table 12: Configuration of a Backbone Switch at the FDC for Ruby and Jade

	Ruby Core	Ruby Edge	Ruby Backbone	Jade Core	Jade Edge	Jade Backbone	Total
Switches	2	5	1	2	5	1	16
12 Port LC	20	10	2	20	10	2	64
24 Port LC	0	20	0	0	20	0	40
48 Port LC	0	0	0	0	0	0	0

Table 13: Breakdown of Switch and Line Card Requirement for FDC

The initial configuration is based on 2 core switches per fabric i.e. Jade and Ruby core will have 2 core switches each. In addition, a backbone layer will provide the ability to connect fabrics between rooms and enable functions such as migration. It is hoped, that storage provisioning for hosts will be restricted to storage arrays in the same room. This will reduce the traffic that traverses the backbone. In the event that the facility expands beyond the current rooms in the future, the backbone will help provide connectivity. Presently, the core has been sized to support the deployment of 10 storage arrays. To meet the maximum number of arrays provided for by DCIP, 18 per room, an additional core is required.

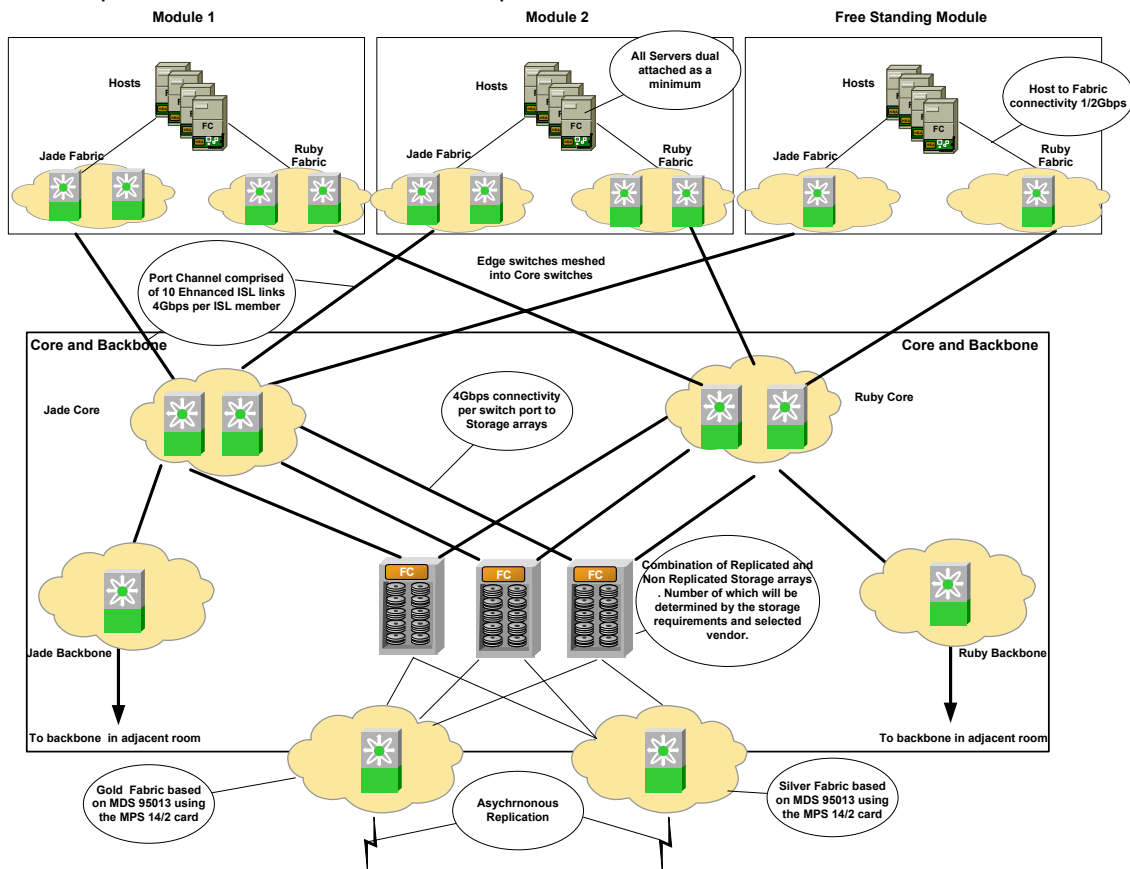


Figure 6: FDC SAN Topology

****NOTE**

The FDC SAN topology shows module1, 2 and free standing module. Initial implementation will only encompass module-1 as this is what will be available for the priority applications.

8.4 Scalability of FDC SAN using MDS 9513 Switches

Each room according to DCIP will contain a maximum of 972 SAN attached host. The edge switches will be shared with the backup SAN, freeing up there switch space for expansion. The number of hosts that need SAN backup connectivity across the entire FDC estate has been estimated to be within 200 clients. This low obligation is a factor that contributed to the decision to share the edge switches with the backup SAN. At this point it is hard to determine the proportion of backup hosts that will be installed into each room. As such, I have decided to add the backup requirement of 200 hosts to a fabric in a room on reaching the number of host ports required per room. The constraining factor on the design is a limitation on the number of edge switches per fabric that can be installed into each room. The limitation of 6 switches is down to sharing the infrastructure racks in each module with other kit such as networks. The Freestanding, Core and Backbone layers have ample room to grow. The maximum number of storage arrays per room is 18. The figures below are for one fabric in a room.

1. Port Count Using eight 24 Port Line cards per switch

No of Edge switches per Room = 6 per fabric
 No of Ports per Switch = 192 (24 Ports LC X 8 Slots)
 Total No of Ports = 1,152 Ports per Fabric

Ports Required = Total number of Host Ports per fabric + Total number of backup ports per fabric

Ports Requirement = 972 + 200 = 1172
 Difference between potential and required = -20 (1152 - 1172)
 ISL requirement fulfilled with 2 slots of 12 Port LC

2. Port Count with an even number of 24 and 48 port LC, 4 each.

No of Edge switches per Room = 6 per fabric
 No of Ports per switch = 288 (24 Port LC x 4 + 48 Port LC x 4)
 Total No of Ports = 1728 Ports per Fabric
 Difference between potential and required = 556 (1728 -1172)
 ISL requirement fulfilled with 2 slots of 12 Port LC

3. Port Count for Freestanding Area.

No of ports per fabric for IBM 595 = 48
 No of ports per fabric for SUN 25K = 18
 Total number of ports required for 6 x IBM and 5 x 25K = 374
 No of freestanding switches per fabric = 2
 No of Ports per Switch = 192 (24 Port LC x 8)
 Total no of ports = 384
 Difference between potential and required = 10 (384 – 374)
 Port Channel Bandwidth = $192 * 2 / 10 = 38.4$
 Number of port channel members = $96 / 4 = 9.6$

4. Port Count Per Core Switch

No of Ports per Switch = 120 (12 Port LC x 10 Slots)
 Ports Required = Total No of ISL Ports + Total Number of storage Ports
 ISL ports per Edge switch = 14 (table 9 section 8.2)
 ISL ports per Freestanding switch = 10
 Storage Ports for one array per fabric = 16

Total number OF Storage Arrays Per room = 18
 Ports required = $392 (14 \times 6) + (10 \times 2) + (16 \times 18)$
 No of Core switches required without backbone ports = $3.2 (392 / 120)$
 Backbone ISL for migration = 5 ports per Core Switch (Migration between rooms)
 Assuming 4 Core switches per fabric, total backbone ISL ports = $20 (5 \times 4)$
 No of Core switches required with Backbone ISL ports = $3.4 ((392 + 20) / 120)$

Summary

What these figures reveal for one fabric, is that, edge switches will require a mix of 24 port and 48 port line cards. A maximum number of 4 core switches per fabric will be required and 2 freestanding switches can meet connectivity for a combination of 595's and 25K's. The intention to deploy virtualization technology within the production environment with tools such as VMWARE, might significantly reduce the number of host ports required.

8.5 SAN Management

Each site will have its own management servers. The same management server will be used for managing both Ruby, Jade, Silver and Gold fabrics at each site. The Fabric Manager GUI limits you to perform a zoning operation one VSAN at a time. Since fabric Ruby and Jade are different VSANs, this effectively means that the administrator can only be changing one Fabric at a time. If multiple administrators are logged on simultaneously then this safeguard does not work. To prevent that, there should be a policy that only one administrator is allowed to configure zoning at one time. There are no hard limits in the software for the number of switches it can manage. The soft limit is listed below

- 200 Switches
- 5000 end devices
- 24 Fabrics

The management server will support four simultaneous Fabric Manager Sessions. Fabric Manager is fully supported to run on a VMWARE partition and that makes it a candidate for consolidation. In the event of a software or server failure, the CLI, which is a superset of the GUI, can be used from any workstation (provided it is configured in the IP ACL) even as the management server is being rebuilt.

GDSE has certified windows 2003 SOE Server as the platform of choice. Access will be through Windows Remote Desktop Client and SSH Client. The environment should be configured in line with recommendations prescribed in the CISCO MDS Switch Solutions Guide located at <http://engineering.citigroup.net/gdse/products/platforms/ciscomds/proposed/07002-sg.pdf>. It Provides information on how to integrate the switches to interact with other management tools, such as Sansscreen, etrust Audit, Tivoli Infrastructure and Citigroup Corporate Network. The Figure below is an overview of the Cisco MDS Network Architecture.

Cisco MDS Network Architecture

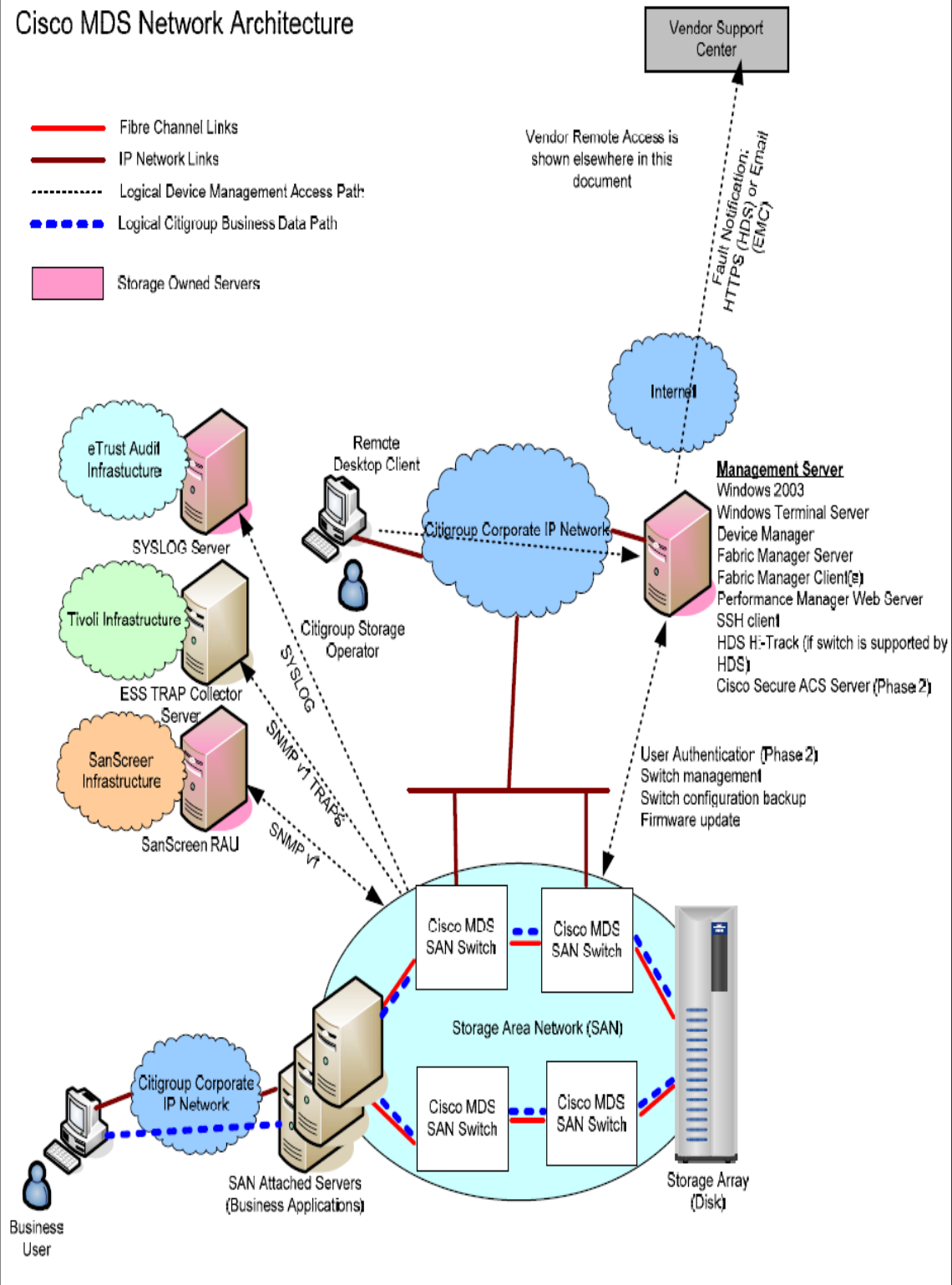


Figure 7: Cisco MDS Network Architecture

9 Storage Requirements

The specific storage requirements of each application have already been specified. A breakdown of the storage requirement for each host can be obtained from the following document [Priority Apps Capacity v0.1.xls](#). The file contains details such as, if the application is replicated or non-replicated storage, if it is a production or contingency server and finally its location. Please be aware that the plan to switch all production services to RDC within the next couple of months, will impact on the classification of servers as COB or Prod. What this breakdown does not show is the requirement of Timefinder BCV/Clone devices at the remote location for point-in-time copies of data.

It is likely that the Raid protection that will be deployed on the new Cisco SAN, will comprise of a combination of Raid-1 and Raid-5 protection levels. Please see below:

Storage Array	Disk Type	Protection Level	Number of Arrays	Capacity per Array GB
EMC DMX 2500	146GB	RAID-5 7+1	1	61320
EMC DMX 2500	300GB	RAID-5 7+1	1	126000
EMC DMX 2500	146GB	RAID-1	1	35040
EMC DMX 2500	300GB	RAID-1	1	72000
HDS USP	146GB	RAID-5 7+1	1	47000
HDS USP	300GB	RAID-5 7+1	1	TBC

Table 6: Storage Array Capacity

9.1 Timefinder BCVs or Timefinder Clones at the Remote Site

The use of BCVs at the remote site has been recommended by GDSE as a best practice to ensure that during a resynchronisation that a point-in-time copy is still available at the remote end. See below an extract from the SRDF/A solutions guide;

“A BCV device needs to be configured in the bin file for every R2 device on the SRDF/A target DMX. These BCVs will be used to take a point-in-time (PiT) copy of the R2 volumes prior to resynchronisation of the SRDF pairs after SRDF/A has been suspended. This is required to ensure that there is an I/O consistent copy of production data at all time at the target site. They can also be used, if required, for COB testing.

In the case where BCVs are needed at the SRDF/A target site for both purposes, it will be recommended to:

- a) Mirror (RAID1) the SRDF/A R2 device*
- b) Share a single BCV for each R2 volume for both the COB and for the SRDF/A PIT copy requirement.”*

The requirement for BCVs needs to be determined by the Application Analysis project stream. The configuration of BCVs will lead to a 33% increase in the amount of storage required in addition to the replicated storage for each application. This has cost implications and this additional cost needs to be weighed against the Recovery Point Object of the applications. For additional information on the configuration of the BCVs please refer to the SRDF/A solutions guide.

9.2 Planned Configurations Replicated

There are two planned target configurations. The plan is to have four EMC Symmetrix DMX3 (or DMX4 if released and certified on time) source arrays located at RDC and four target arrays at FDC. They will be DMX3 24S-config35 3 bay configurations or DMX3 24S-config42 5 bay configurations which appear as follows in the document **DMX configuration Document** listed above:

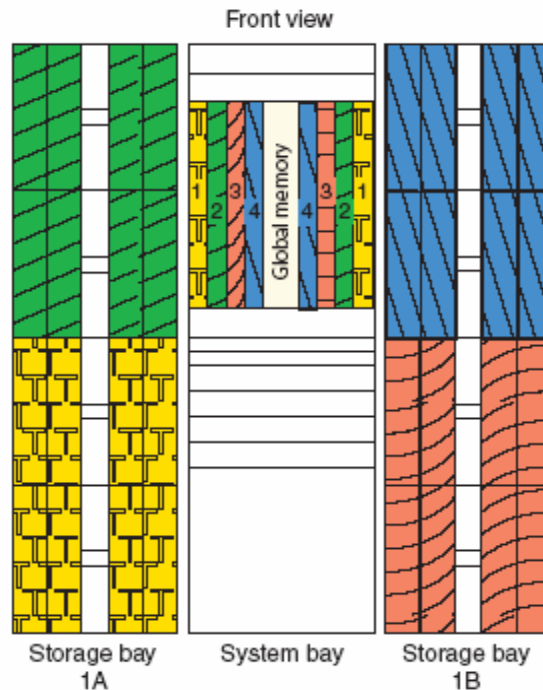


Figure 8: Planned 480 drives three bay Symmetrix array

The chosen configuration (see table 4) consists of three (3) three bay Symmetrix arrays with four (4) pairs of disk adapters and eight (8) FA cards two (2) of which will be reserved for RA ports giving a total of 16 RA ports. This configuration gives thirty two (32) fibre channel loops with fifteen (15) disks each giving a maximum of 480 disks per Symmetrix. Each of the production arrays to be located in RDC will be populated with three hundred and twenty (320) disks and each of the Disaster Recovery (DR) arrays to be located in FDC (Contingency R2 target arrays) will have four hundred and eighty (480) disks to allow for one hundred and sixty (160) production RAID 1 disks and one hundred and sixty (160) unprotected BCV point in time copy devices. Host connectivity will be through 6x4 (16) FA 4Gbit host ports. An alternative configuration consists of a DMX3 4500 with five (5) Bays, four (4) DA pairs giving max 30 disks per loop in 32 loops and a total of nine hundred and sixty (960) disk devices. Because this particular array will be used for Flexcube and Flexcube use clone technology it is recommended that we install nine hundred and sixty (960) disk devices on both the production and CoB side.

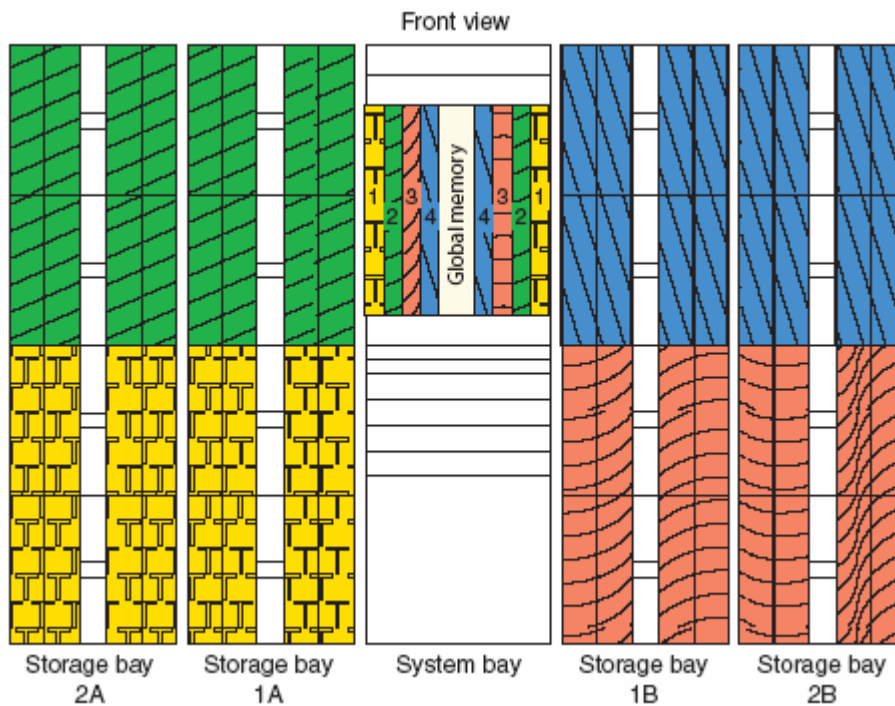


Figure 9: Planned 960 drives five bay Symmetrix array

RDC			FDC					
SID	SID	Site	SID	SID	Site	Model		Comment
1	DMX01RDC	RDC	1	DMX01FDC	FDC	DMX3 4500 24S-config35	SRDF/A	
2	DMX02RDC	RDC	2	DMX02FDC	FDC	5 Bay DMX3 4500 24S-config42 OR DMX3 4500 24S-config35	SRDF/A	FLEX
3	DMX03RDC	RDC	3	DMX03FDC	FDC	DMX3 4500 24S-config35	SRDF/A	
4	DMX04RDC	RDC	4	DMX04FDC	FDC	DMX3 4500 24S-config35	SRDF/A	Possibly SRDF/star ready
			5	Array5 either USPV from HDS Or DMX from EMC	FDC	### >20TB		Non-rep

Table 7: Planned RDC/FDC DMX array configurations.

Planned Configurations Non-Replicated

HDS OR EMC

USPV
 Number of cabinets: 3
 Type of Disks: 300G 15K
 Number of Disks: 11 + 2 spares
 Cache: 32GB
 Control memory: 20GB
 4 BED cards
 4x 8 port 4GBps adapters



Symmetrix DMX
 24S-CONFIG07
 DMX 3 2500 or DMX 4
 Number of Disks: 15
 Type of disks: 300G 15k



10 Replication SAN

The design decision has been made to implement a physically separate fabric for array based replication of data between RDC and FDC. The design decision was based on a number of factors:

- The MPS 14/2 card is a Generation 1 line-card and as such there are limitations around its use when combined with Generation 2 line-cards. The former have an architecture limit of 256 interfaces compared with 1024 interfaces for Generation 2 Cards
- Segregation of the storage and replication fabrics allows additional ports in the core switches to be used as array (disk) connectivity.
- The Host to Storage fabrics traditionally have a high frequency of zoning changes, approx 1 every week. The replication fabric in comparison has a very low frequency of change, approx 1 every few months.
- Segregation of the fabrics is in line with Citigroup best practices.
- Segregation of the fabrics opens the possibility to accommodate mainframe replication traffic⁷ to use the Storage Wide Area Network (SWAN).

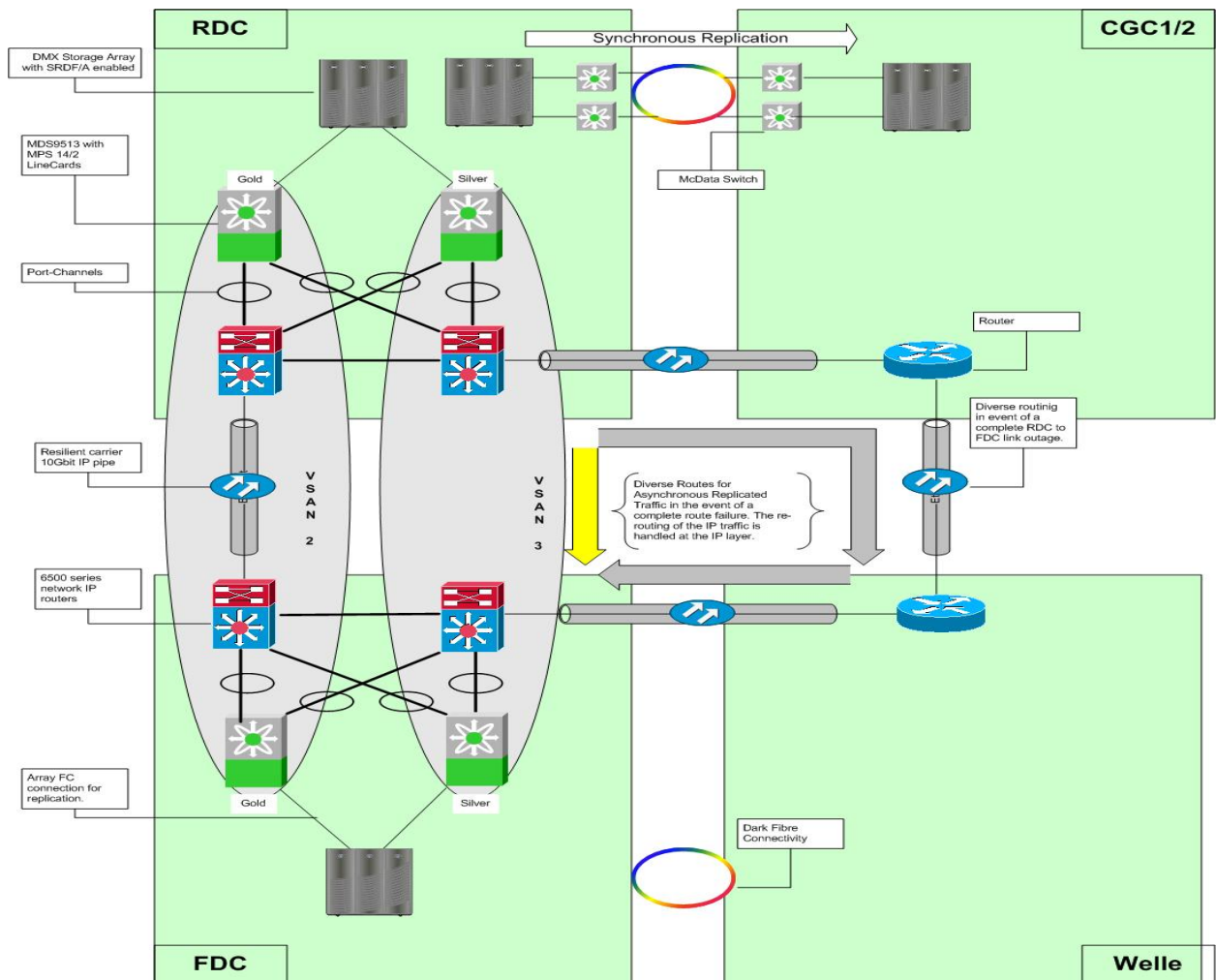


Figure 8: Phase 1 SRDF/A SWAN Network

⁷ Due to the low frequency of change of the replication fabric mainframe technologists are more comfortable in allowing sharing in this area.

The replication SAN comprises of two independent fabrics, Gold and Silver. The Gold and Silver fabrics will be comprised of Cisco MDS 9513 switches that will connect directly into 6500 series network routers. The SWAN is an IP based network which will carry FCIP data between RDC and FDC. A topology diagram of the SWAN phase 1 is shown above in Figure 8.

The SWAN will provide two network routes:

1. A protected 10Gbps IP circuit between RDC → FDC.
2. A protected 10Gbps IP circuit between RDC → CGC → Welle → FDC.

In normal operation replication traffic will flow from RDC → FDC in the event of a complete protected circuit failure between these locations the replicated traffic will be re-routed using network protocol around the alternative circuit RDC → CGC → Welle → FDC.

The Cisco MDS switches have a technical feature called Port Channels. Port Channels allow the aggregation of Virtual Inter-Switch links (VISLs) into a trunk, the bandwidth of the being the sum of each individual links bandwidth. In this case the available bandwidth of 10Gbps is split between the Gold and Silver fabrics. The intention is to create two port channels per Gold and Silver Fabric see Figure 9.

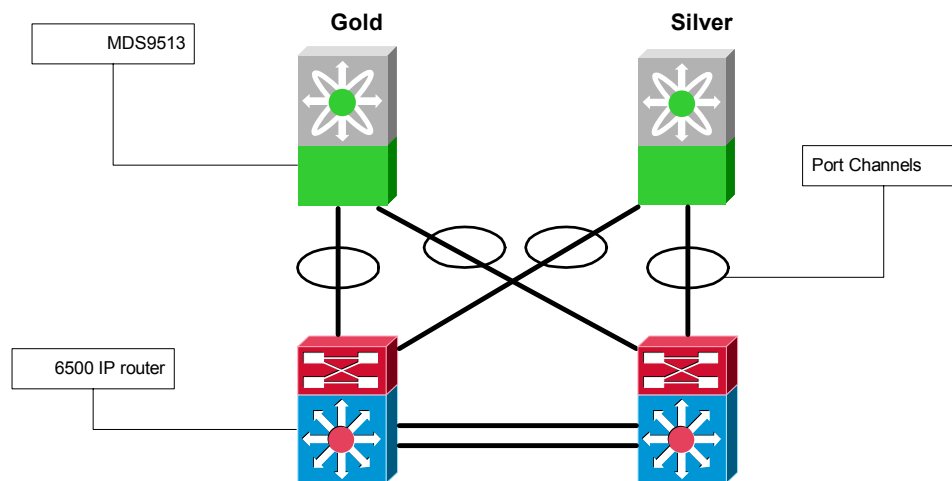


Figure 9: Port-Channel Distribution

Each port channel will consist of a number of VISL members. Each VISL has a cost associated with it:

- 1Gbps = 1000
- 2Gbps = 500
- 4Gbps = 250

The Fabric Shortest Path First (FSPF) cost is calculated by the FSPF link cost divided by the number of links. To ensure that the replicating data always flows via the shortest route from RDC to FDC a combination of FSPF tuning and network Open Shortest Path First (OSPF) tuning needs to be configured.

10.1 Predicted Bandwidth Utilisation

The SWAN network will carry both mainframe replicated traffic and distributed replicated traffic. The SWAN will provide 10Gbit end to end connectivity for this traffic to traverse. For details on how to optimise the SWAN FCIP traffic please refer to the following GDSE solutions guide.
<http://engineering.citigroup.net/gdse/products/platforms/ciscomds/stdsdocs/07015-SG.pdf>

10.1.1 Mainframe Estimated Bandwidth Utilisations

As provided by David Kelvin the following are estimates on the peak bandwidth utilisations. These figures do not take into account compression.

Corporate (phase 1):

Peak Write DASD I/O (FC over IP): 70 MBytes/sec uncompressed
 Disk I/O will NOT be compressed before transmission - maybe the Cisco 95xx can ~ 2:1

Peak Write Tape I/O (native IP): 150 MBytes/sec uncompressed
 Tape I/O will be compressed before transmission ~ 3:1, giving a peak of approximately 50 MBytes/sec

Consumer (phase 2):

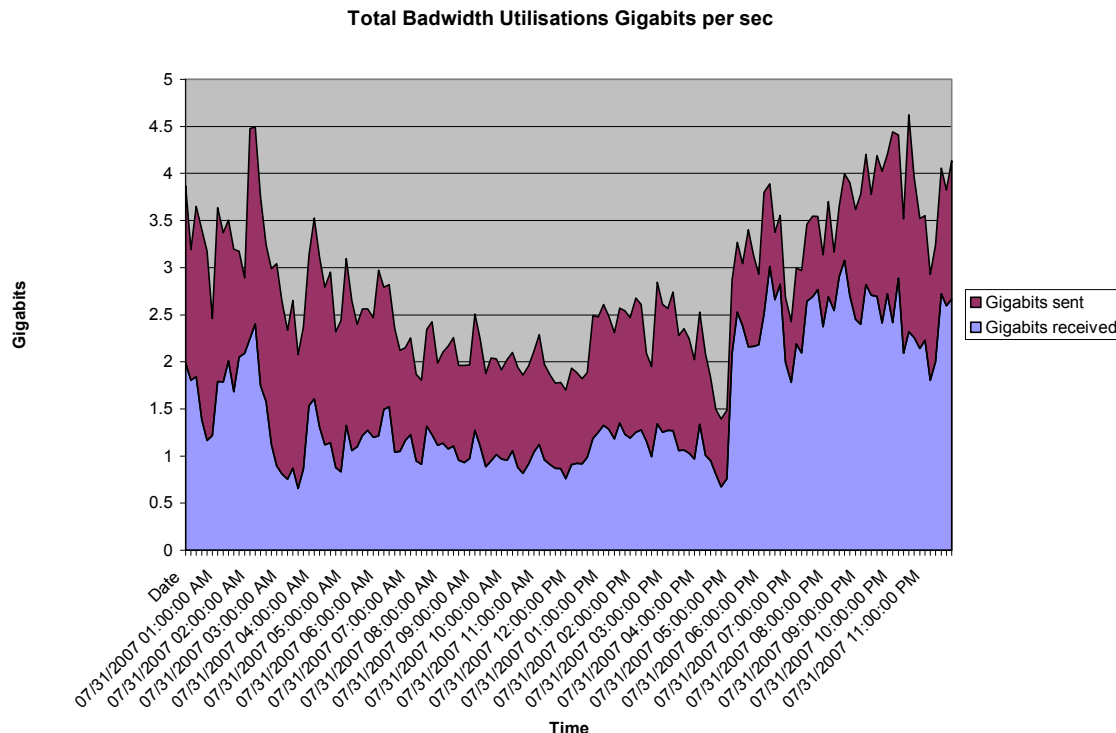
Peak Write DASD I/O (FC over IP): 305 MBytes/sec uncompressed (expected to increase significantly due to additional workload)
 Disk I/O will NOT be compressed before transmission - maybe the Cisco 95xx can ~ 2:1

Peak Write Tape I/O (native IP): 215 MBytes/sec uncompressed (expected to increase significantly due to additional workload)
 Tape I/O will be compressed before transmission ~ 3:1, giving a peak of approximately 72 MBytes/sec

The tape traffic however will not utilise the SWAN network. The tape traffic will use the corporate LAN.

10.1.2 Distributed Measured Bandwidth Utilisations

Through the use of EMC ControlCenter and Performance Manager it has been possible to gather statistics regarding the existing SRDF/S traffic flows. The graph that follow show utilisations over a period of 24hours, note that all bandwidth utilisations here are for uncompressed block data.



Graph 2: Total Bandwidth Utilisations CGC

As shown for the above graph the peak bandwidth utilisations for all arrays in the CGC environment peak at 4.5Gbps. Initially only the priority applications will be moving on the SRDF/A environment, it is likely that the bandwidth utilisations for phase 1 will be much less. In order to properly size the storage arrays for the SRDF/A traffic storage engineering personnel are currently working with the vendor to generate straw-man models of the new storage arrays, this activity is quite time intensive and will take a number of weeks to complete.

10.2 Replication SAN Hardware

The hardware that will need to be procured for the replication SAN is as follows:

- 2x9513 for RDC
- 2x9513 for FDC
- 12xMPS 14/2 Line Card for RDC
- 12xMPS 14/2 Line Card for FDC

It is preferable that the new 18/4 MPS card is procured for actual deployment but this has not been GA'd yet by the vendor however it is likely that it will be available later this month for testing and is planned to be certified for Q4 2007.

11 Cabling Requirements

The cabling requirements for the SAN at the FDC are shown in table 13 below. The assumptions drawn in reaching the figures outlined are also listed. The requirements are broken down into the following

- Edge frame to Core frame
- Storage Array to Core frame
- Backbone to Backbone frame

ASSUMPTIONS							
Day 1 requirements are based on deploying 1 x SAN attached Cabinet module, 1 x Freestanding module and 1 SAN storage area to accommodate the priority applications							
Cabling is duplex, 50 Micron, Multimode with the exception of the DWDM links(TBA)							
Excludes any back bone links to Citi Network Switches/L2&3 or SAC/PAC/SISpod connectivity.							
Day 1 storage array totals are scaled for dual vendor strategy and are based on day 1 maximum amounts, these may Hall 1Buce if Raid 5 configurations are deployed							
The DWDM requirements are based on the DWDM open frames having interlinks inbetween Hall 1B and Hall 1A Rooms							
EDGE FRAME TO SAN CORE FRAME - SAN Fibre requirements							
Room	Cabinet or Freestanding Module	To	SAN Fabric / Edge Frame	Total Number of ISL per Fabric	Minimum Fibre channels Required - Edge to Core Frame	Ideal amount	FIBRES ARE 50Micron Multimode - Duplex
Hall 1A	Cabinet Module #1	-	Ruby	20	20	36.48	
Hall 1A	Freestanding Module #1	-	Ruby	20	20	36.48	
Hall 1A	Cabinet Module #1	-	Jade	20	20	36.48	
Hall 1A	Freestanding Module #1	-	Jade	20	20	36.48	
STORAGE ARRAY TO SAN CORE FRAME - SAN Fibre Requirements							
Room	Storage Array - locations TBA	TO	SAN Fabric / Core Frame	RequiHall 1B Day one?	Minimum Fibre channels required - SAN Storage to Core Frame	Ideal amount	FIBRES ARE 50Micron/Multimode - Duplex
Hall 1A	SAN STORAGE ARRAY #01	-	Ruby/Silver	Yes	16	24.48	
Hall 1A	SAN STORAGE ARRAY #01	-	Jade/Gold	Yes	16	24.48	
Hall 1A	SAN STORAGE ARRAY #02	-	Ruby/Silver	Yes	16	24.48	
Hall 1A	SAN STORAGE ARRAY #02	-	Jade/Gold	Yes	16	24.48	
Hall 1A	SAN STORAGE ARRAY #03	-	Ruby/Silver	Yes	16	24.48	
Hall 1A	SAN STORAGE ARRAY #03	-	Jade/Gold	Yes	16	24.48	
Hall 1A	SAN STORAGE ARRAY #04	-	Ruby/Silver	Yes	16	24.48	
Hall 1A	SAN STORAGE ARRAY #04	-	Jade/Gold	Yes	16	24.48	
Hall 1A	SAN STORAGE ARRAY #05	-	Ruby/Silver	Yes	16	24.48	
Hall 1A	SAN STORAGE ARRAY #05	-	Jade/Gold	Yes	16	24.48	
Hall 1A	SAN STORAGE ARRAY #06	-	Ruby/Silver	Yes	16	24.48	
Hall 1A	SAN STORAGE ARRAY #06	-	Jade/Gold	Yes	16	24.48	
Hall 1A	SAN STORAGE ARRAY #07	-	Ruby/Silver	Yes	16	24.48	
Hall 1A	SAN STORAGE ARRAY #07	-	Jade/Gold	Yes	16	24.48	
Hall 1A	SAN STORAGE ARRAY #08	-	Ruby/Silver	Yes	16	24.48	
Hall 1A	SAN STORAGE ARRAY #08	-	Jade/Gold	Yes	16	24.48	
Hall 1B ROOM SAN BACKBONE FRAME to Hall 1A ROOM SAN BACKBONE FRAME - Fibre requirements							
Room	Core Frame	TO	Core Frame	Room	Minimum Fibre channels required - Core to Core Frame	Ideal amount	ARE 50Micron/multimode - Duplex
Hall 1A	Ruby/Silver - CORE FRAME	-	Ruby/Silver - CORE FRAME	Hall 1B	40	48.72	
Hall 1A	Jade/Gold - CORE FRAME	-	Jade/Gold - CORE FRAME	Hall 1B	40	48.72	
SAN CORE FRAMES to WAN/DWDM AREAS							
Room	Area	TO	Area	Room	Minimum Fibre channels - Core to Core Frame	Ideal amount	DUPLEX FIBRE CHANNELS - 9 or 50MICRON TBA
Hall 1B	Ruby/Silver - CORE FRAME	-	DWDM AREA	Hall 1B	6	12	
Hall 1B	Jade/Gold - CORE FRAME	-	DWDM AREA	Hall 1B	6	12	
Hall 1A	Ruby/Silver - CORE FRAME	-	DWDM AREA	Hall 1A	6	12	
Hall 1A	Jade/Gold - CORE FRAME	-	DWDM AREA	Hall 1A	6	12	
The DWDM requirements are based on the DWDM open frames having interlinks inbetween Hall 1B and Hall 1A Rooms							

Table 13: FDC Cable Requirements

The topology for the RDC is a Core to Edge architecture and the cable requirements for the RDC are listed in the table below. The requirements are broken down into

- Storage To Core
- Edge To Core
- Core To WAN

ASSUMPTIONS							
Day 1 requirements are based on deploying the same amount of Day 1 arrays as In FDC							
Day 1 storage array totals are scaled for dual vendor strategy and are based on day 1 maximum amounts, these may reduce if Raid 5 configurations are deployed							
Cisco Switches and new infrastructure will be mounted adjacent to existing SAN infrastructure in Open frames AU Row RDC3/4 - RDC1 TBA							
RDC1 Frames TBA - due to power and space limitations - DCIP to advise location							
Cabling is duplex, 50 Micron, Multimode - with IBM SCDC panels for interlinks & LC connector at the storage array end							
STORAGE ARRAY TO SAN FRAME - SAN Fibre Requirements							
Floor/ Room	Storage Array - locations TBA	TO	SAN Fabric / Core Frame	Required Day one?	Minimum Fibre channels required - SAN Storage to Core Frame	Ideal amount	FIBRES ARE 50Micron/MultiMode - Duplex
RDC4	SAN STORAGE ARRAY #01	-	AU15 to 10	Yes	16	24.48	
RDC4	SAN STORAGE ARRAY #01	-	AU30 to 35	Yes	16	24.48	
RDC4	SAN STORAGE ARRAY #02	-	AU15 to 10	Yes	16	24.48	
RDC4	SAN STORAGE ARRAY #02	-	AU30 to 35	Yes	16	24.48	
RDC4	SAN STORAGE ARRAY #03	-	AU15 to 10	Yes	16	24.48	
RDC4	SAN STORAGE ARRAY #03	-	AU30 to 35	Yes	16	24.48	
RDC4	SAN STORAGE ARRAY #04	-	AU15 to 10	Yes	16	24.48	
RDC4	SAN STORAGE ARRAY #04	-	AU30 to 35	Yes	16	24.48	
RDC4	SAN STORAGE ARRAY #05	-	AU15 to 10	Yes	16	24.48	
RDC4	SAN STORAGE ARRAY #05	-	AU30 to 35	Yes	16	24.48	
RDC4	SAN STORAGE ARRAY #06	-	AU15 to 10	Yes	16	24.48	
RDC4	SAN STORAGE ARRAY #06	-	AU30 to 35	Yes	16	24.48	
RDC4	SAN STORAGE ARRAY #07	-	AU15 to 10	Yes	16	24.48	
RDC4	SAN STORAGE ARRAY #07	-	AU30 to 35	Yes	16	24.48	
RDC4	SAN STORAGE ARRAY #08	-	AU15 to 10	Yes	16	24.48	
RDC4	SAN STORAGE ARRAY #08	-	AU30 to 35	Yes	16	24.48	
InterFloor Links							
Floor/ Room	SAN Frame	TO	SAN Frame	Floor/Room	Minimum Fibre channels required - Core to Core Frame	Ideal amount	FIBRES ARE 50Micron/multimode - Duplex
RDC1	Ruby/Silver - FRAME TBA	-	Ruby/Silver - FRAME AU15 to 10	RDC4	40	48-72	
RDC1	Jade/Gold - FRAME TBA	-	Jade/Gold - FRAME AU30 to 35	RDC4	40	48-72	
RDC3	Ruby/Silver - FRAME AU15 to 10	-	Ruby/Silver - FRAME AU15 to 10	RDC4	40	48-72	
RDC3	Jade/Gold - FRAME AU30 to 35	-	Jade/Gold - FRAME AU30 to 35	RDC4	40	48-72	
SAN CORE FRAMES to WAN/DWDM AREAS							
Floor/ Room	Area	TO	Area	Floor/Room	Minimum Fibre channels - SAN to DWDM cabinets	Ideal amount	DUPLEX FIBRE CHANNELS - 50MICRON TBC
RDC1	Ruby/Silver - CORE FRAME	-	RDC1 - Q Row - Lower number cabinet	RDC1	6	12	
RDC1	Jade/Gold - CORE FRAME	-	RDC1 - Q Row - Lower number cabinet	RDC1	6	12	
RDC1	Ruby/Silver - CORE FRAME	-	RDC1 - Q Row - Higher number Cabinet	RDC1	6	12	
RDC1	Jade/Gold - CORE FRAME	-	RDC1 - Q Row - Higher number Cabinet	RDC1	6	12	

Table 14: RDC Cable Requirements

Built into both tables is an additional contingency requirement for any faulty cables that are discovered during implementation.

12 References

DOCUMENT	LOCATION
DMX Series Solution Guide	http://engineering.ti.citigroup.net/gdse/products/platforms/symmetrix/stdsdocs/03038-sg.pdf
HDS USP Solutions Guide	http://engineering.ti.citigroup.net/gdse/products/platforms/tagmastore/stdsdocs/05015-sg.pdf
Cisco Solutions Guide V0.3	http://engineering.ti.citigroup.net/gdse/products/platforms/ciscomds/proposed/07002-sg.pdf
SRDF/A for Long Distance Replications Solutions Guide	http://engineering.ti.citigroup.net/gdse/products/platforms/symmetrix/stdsdocs/05002-sg.pdf
WAN Throughput optimization for SRDF/A	http://engineering.ti.citigroup.net/gdse/products/platforms/symmetrix/stdsdocs/06034-sg.pdf

14 Appendix B using MDS 9509

DCIP have advised that it would be preferable to use the lower port density switch in the designs rather than the higher density MDS 9513. The reasons that DCIP have given for this are:

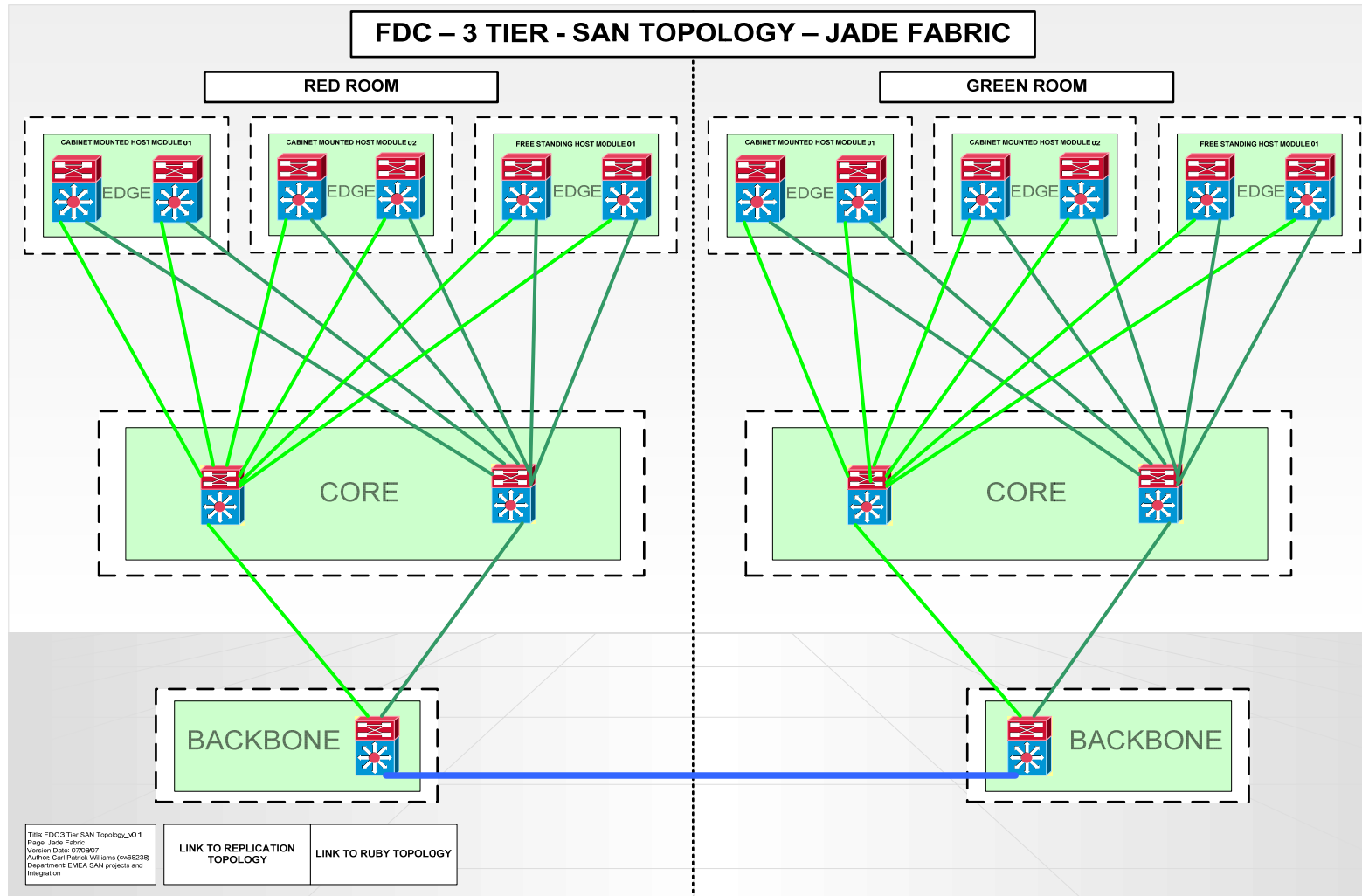
1. The MDS 9509 is fully supported in a 2-post rack
2. There are health and safety concerns regarding the installation of the MDS 9513
3. No alteration of the 2-post rack configuration is required

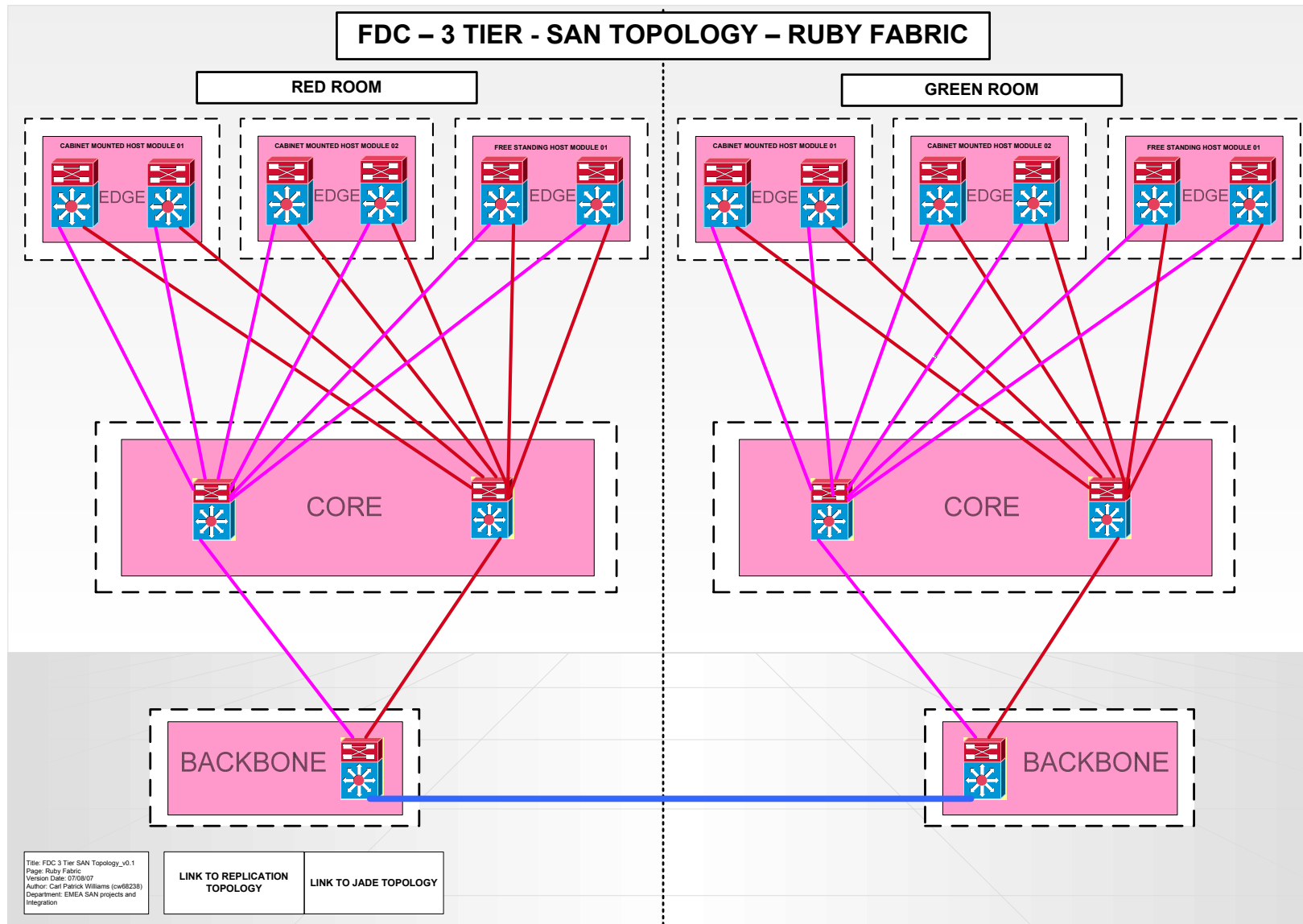
While these are valid concerns, there are some distinct disadvantages in installing the MDS 9509 as the preferred switch for the core-edge fabric and these are:

- The MDS 9509 has a lower port density count i.e. 384 hosts ports falls to 270 useable ports for the 9509.
- The MDS 9509 is the same height as the MDS 9513 and therefore due to cabinet space restrictions, only 2 edge switches per fabric (Jade/Ruby) per module would be allowed.
- The MDS 9509 used as a core switch would dramatically reduce the number of storage arrays connecting to the core from 11 to 4. This could be remedied with a doubling of the number of core switches but the design would begin to look more like a meshed fabric.

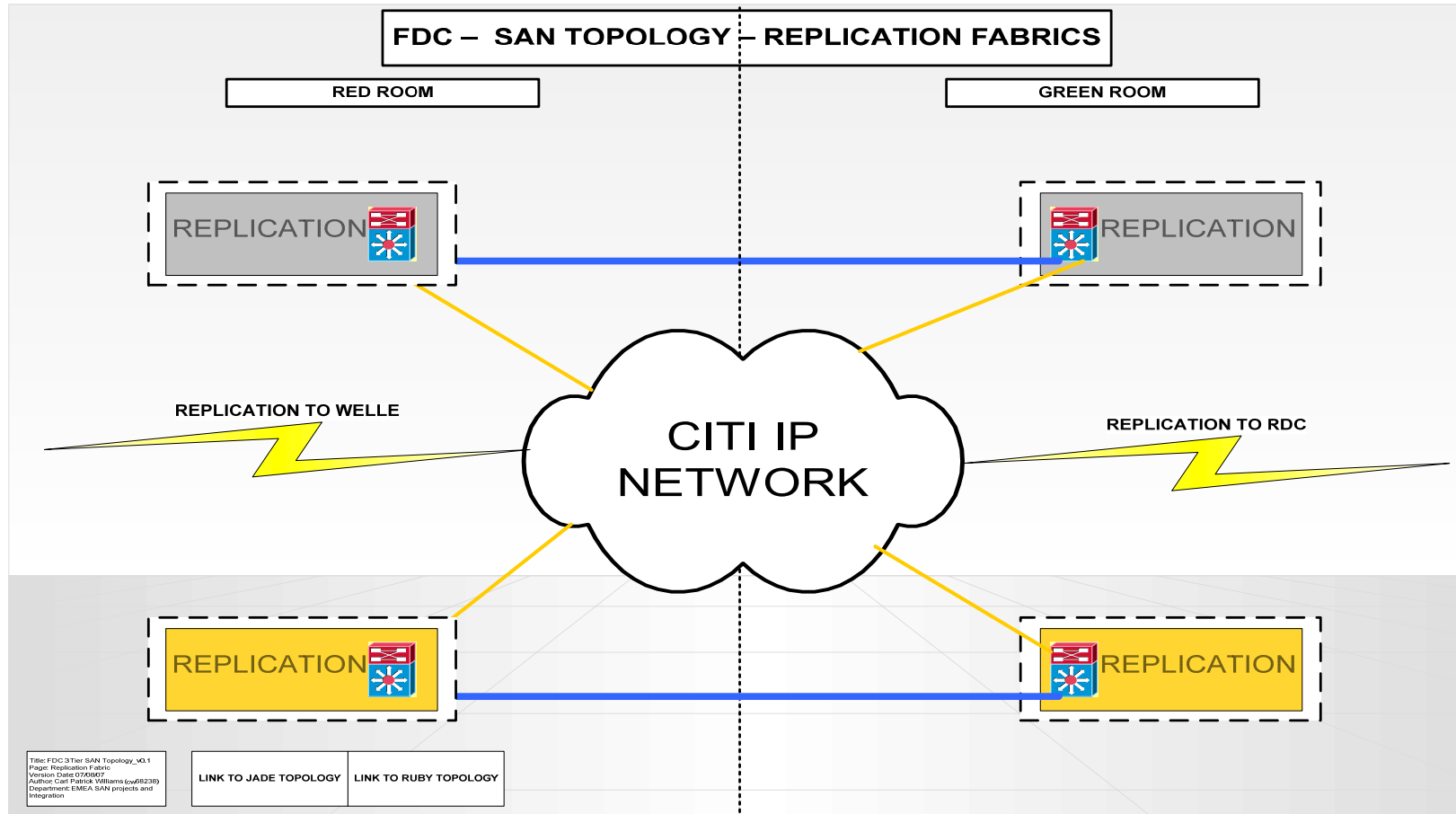
The prohibiting factor and chief concern, regard the need for a supporting plate to accommodate the MDS 9513 switch. There is a strong design case for the use of the MDS 9513 at the backbone, core and edge of the fabrics.

15 Appendix C FDC room 1A SAN Topology

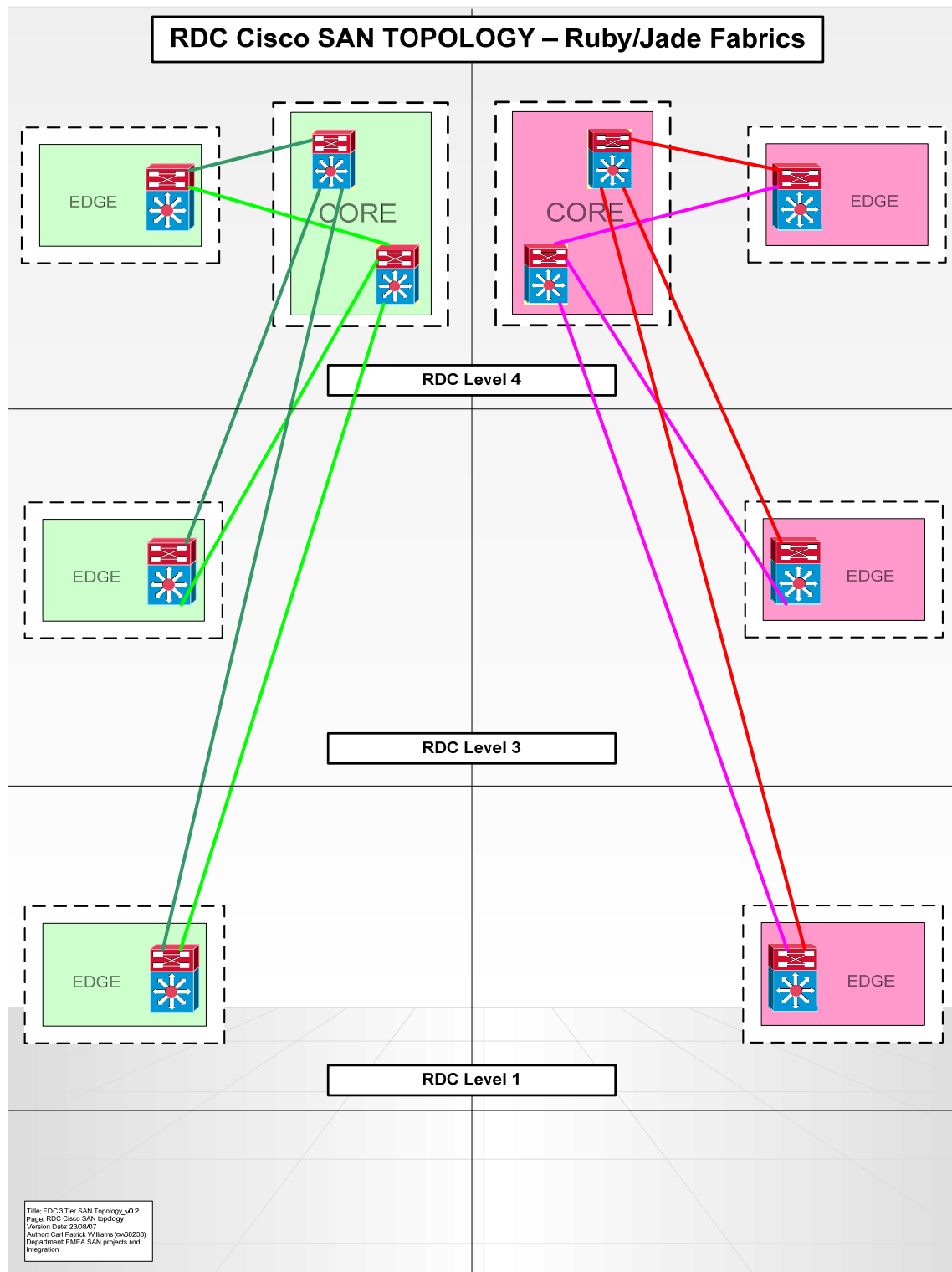




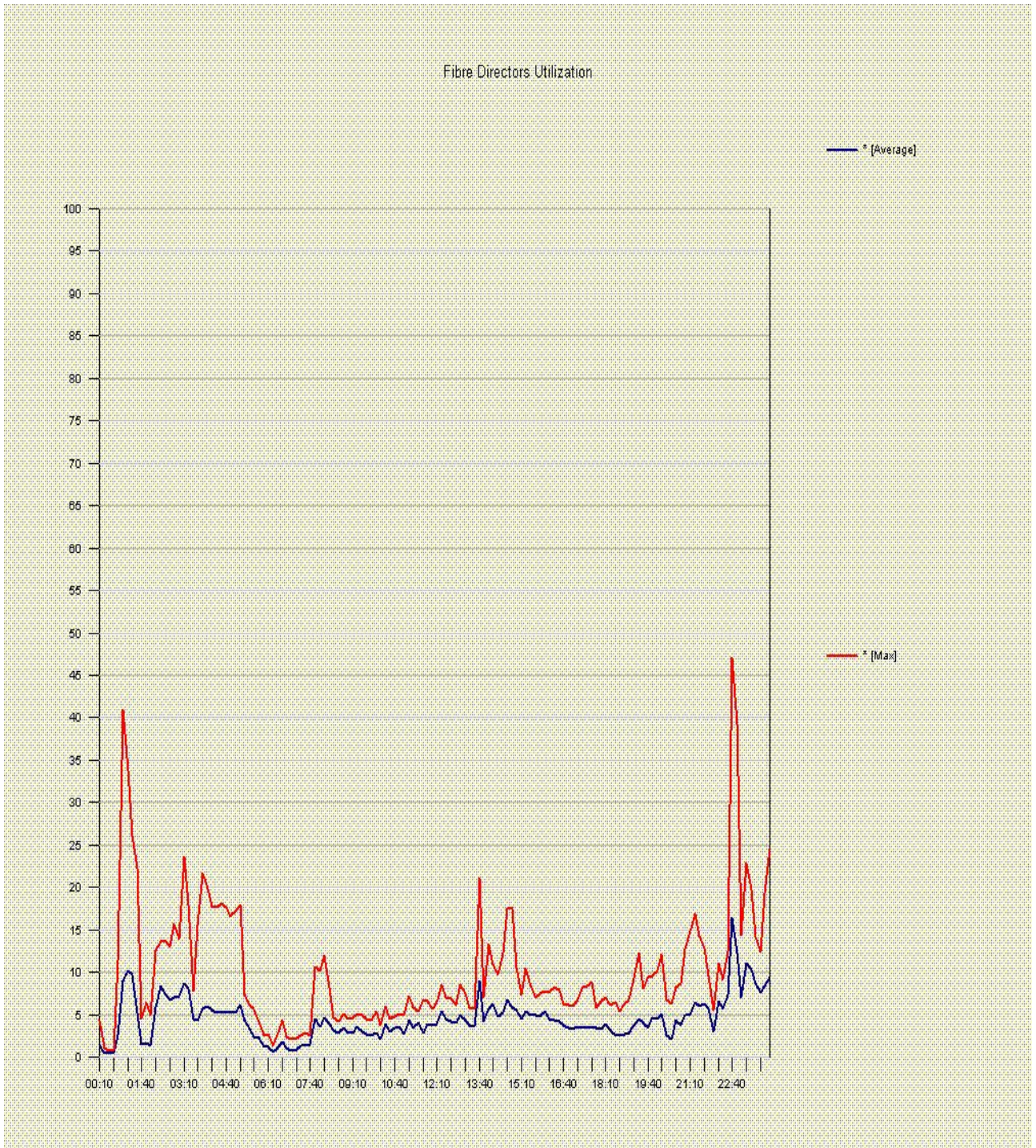
16 Appendix D Replication Fabric Topology



17 Appendix E RDC SAN Topology

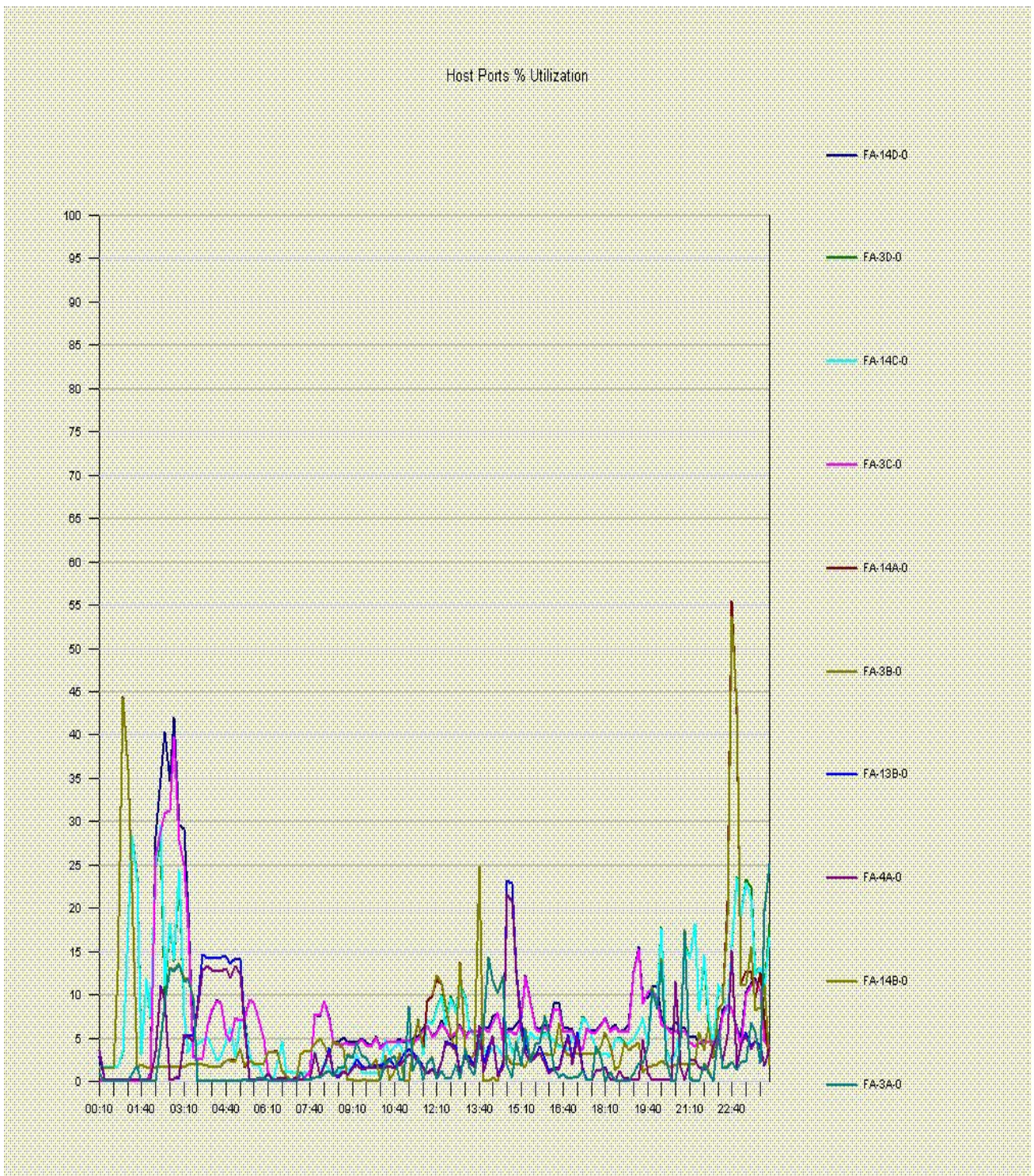


18 APPENDIX F



Fibre Director Utilisation Graph

19 APPENDIX G



FA Port Utilisation Graph