The New York Merchantile Exchange's (NYMEX) first Storage Area Network (both Local & Multi-Site Distance SAN)

Architect.....: Noel Milton Vega
Implementation...: Noel Milton Vega
Documentation....: Noel Milton Vega

05/2004

The New York Mercantile Exchange's (NYMEX)
inaugural Local & Wide Area/Distance SAN

Primary Site <---------- redundant DWDM ----------> DR Site

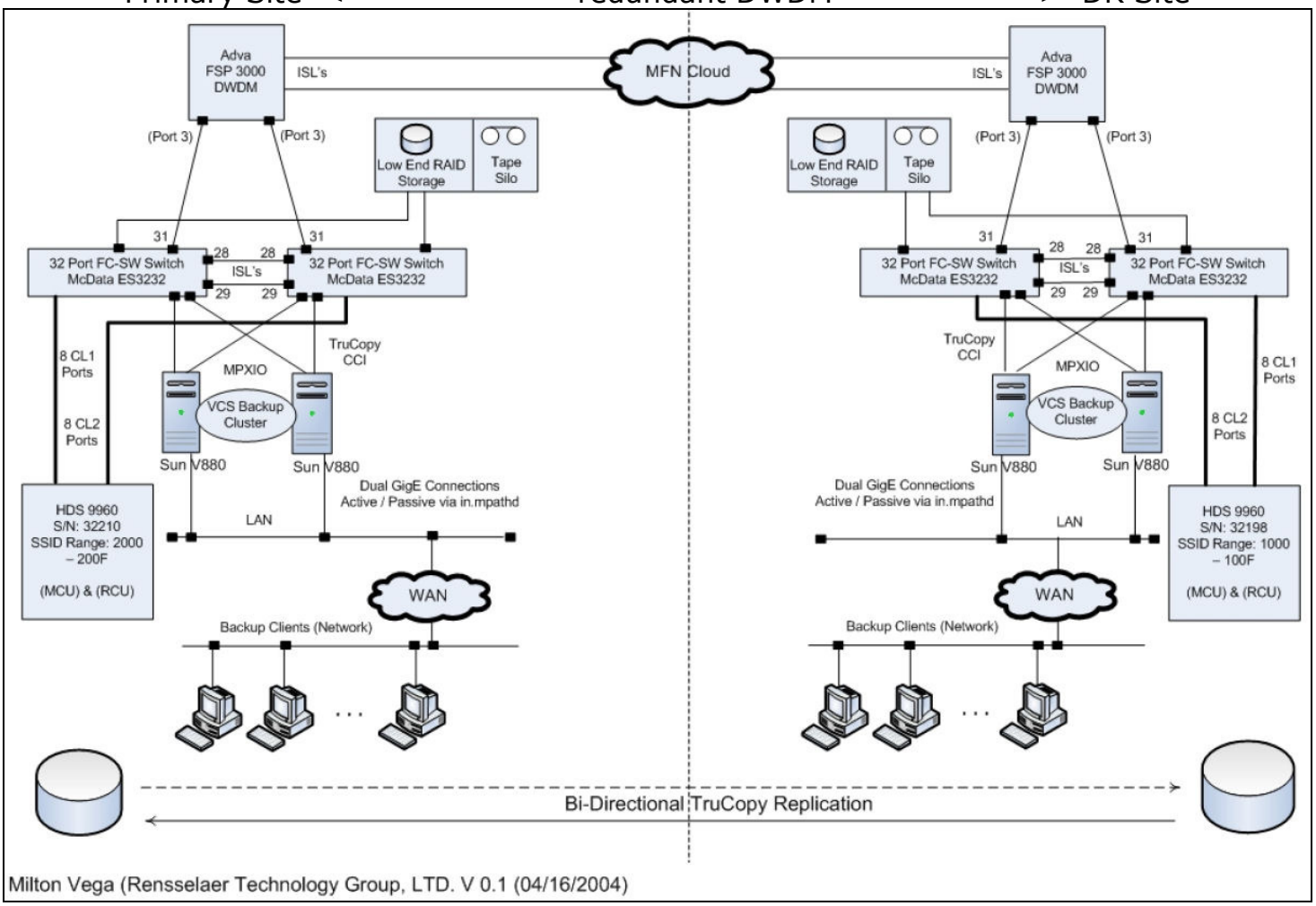


Figure1: Solution: Designed for local clusters, and simultaneous remote Sync/Async data replication for Disaster Recovery/Disaster Tolerance. The two horizontal lines at the top of the diagram represent diverse 2Gbps Dense Wave Division Multiplexing (DWDM) optic paths that carried the TrueCopy traffic. From start to end, each optical path took divergent geographical routes by design. Because the paths were asymmetric, their fibre optic distances differed, and this had several implications that needed to be accounted for: (1) higher end-to-end latency over the longer path, which can be noticeable for synchronous replication (but asynchronous as well); (2) The FibreChannel switches purchased (4 of them) would need to have enough Buffer Credits for the longest of the distances involved (i.e. for the worst case, since enough buffer credits ensures continuous streaming of data); and (3) asymmetric path (request over one path; response over the opposite path).

While designing the distance portion of this SAN, I wrote a white paper to myself that talked about the effects that physical and logical distances, bit-insertion rates, etc., imposed on switch Buffer credit requirements and performance. That paper can be viewed here: http://doc5.computingarchitects.com

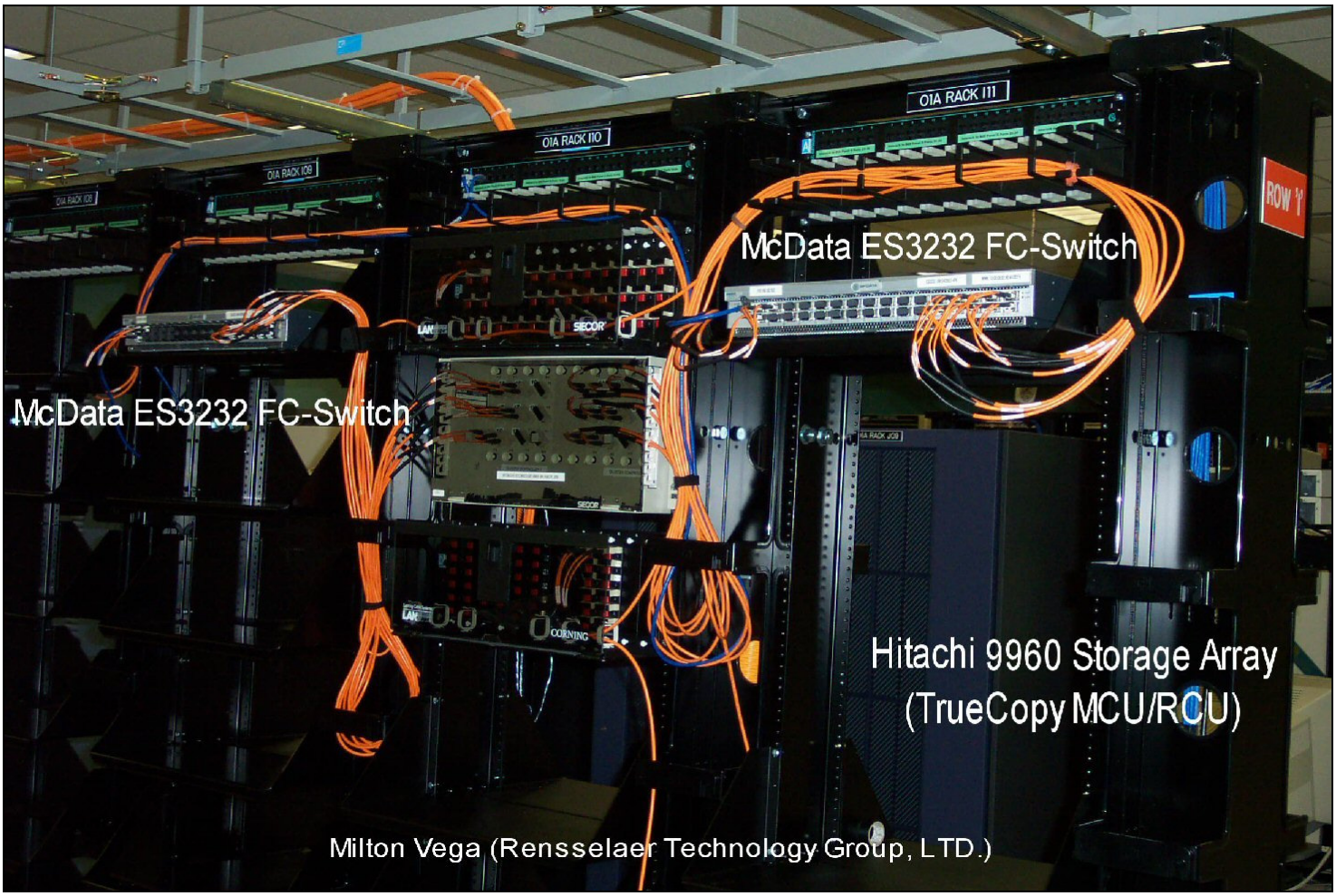


Figure 2: A photo of one datacenter (of two), in this TrueCopy enabled multi-site SAN solution. The photo illustrates the redundant ES3232 McData switches (which were temporarily connected via ISL's), as well as the HDS9960 array (the bluish-purple frame in the background towards the right). The design at the opposite end of the DWDM/ISL link (i.e. at the other datacenter 92Km away) is configured symmetrically, right down to the ports and LUN's used (as suggested in the diagram above

Here, the integration of two McData ES3232 Fibre Channel switches serves the following purposes:

(1) Front-ending the Hitachi 9960 to efficiently use Array ports by fanning them out in a 1 – N configuration to clients through the Fibre Channel switch. The switch on the left services the Cluster-1 ports of the array, while the switch on the right covers the Cluster-2 ports of the array. The switch provides F-Port to N-Port FC BB_Credits.

(2) Ports 29 and 30 on each switch are configured as 2Gbps E-Ports to connect them to each other as local ISL's. Ports 31 of each switch, also configured as E-Ports, each connect to physically different DWDM equipment and circuit ID's (for redundancy – see diagram above). These DWDM connections create remote ISL's with two corresponding ES3232 FC switches at the downtown Manhattan site. Each Fibre Channel switch provides 60 EE_Credits, sufficient to cover the 96km linear fibre distance between the two sites, for uninterrupted streaming of data. The final result is a 4 switch, 128 port highly available multi-switch fabric, replicating mission critical data between sites for both Active/Active purposes, as well as disaster recover (DR) purposes.

(3) Inter-site Hitachi TrueCopy (i.e. remote data replication) traffic commute through the DWDM ISL links. At least 2 logical paths between two Hitachi 99xx arrays are needed for reliable TrueCopy traffic going in one direction. Reliable Bi-Directional TrueCopy traffic therefore requires four. Since initially only two physical DWDM paths were allotted for FibreChannel traffic, the switches also serve to fan-out these remote ISL connections to create 4 logical inter-Site Fibre Channel paths out of two physical DWDM lines; thus allowing for (i.e. facilitating) Bi-Directional TrueCopy traffic.

Determining CH Port WWPN's of Hitachi HDS 7700E & 9900 series storage arrays.

General Format (Hex number): 500060E8[01|02][SSSS][0|1][0-F]

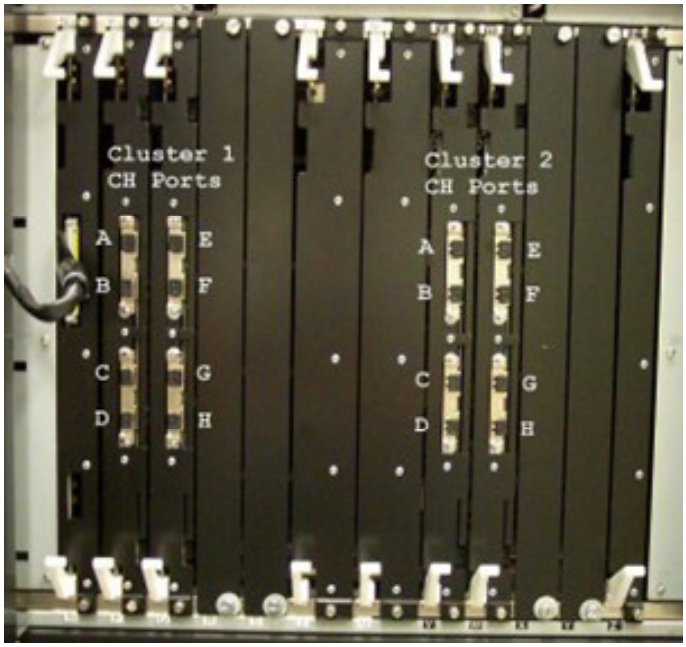
Fixed 8 digit hex prefix (constant):

2 digits as follows:
7700E Series = 01
9900 Series = 02

4 digit Serial number of array in hex:
(Example: 32198_{10} = $7DC6_{16}$)

1 digit as folows:
Cluster 1 = 0
Cluster 2 = 1

1 digit as follows:
0 = CH Port A
1 = CH Port B
2 = CH Port C
3 = CH Port D
4 = CH Port E
5 = CH Port F
6 = CH Port G
7 = CH Port H
8 = CH Port I
9 = CH Port J
A = CH Port K
B = CH Port L
C = CH Port M
D = CH Port N
E = CH Port O
F = CH Port P



Noel Milton Vega

Figure 3: The WWPN of each HDS Array front-end CH ports can be determined using the algorithm above. Although it is possible to also determine the WWPN's by connecting the ports to a FibreChannel switch, when designing and building a SAN (two distinct steps), it is useful to know such information before hand: (1) its allows you to design and debug the Zone-Port diagram on paper first (see below); and (2) it allows you to pre-configure the FibreChannel switch Zone databases with Zones before anything is connected.

Version 10.0										
McData3232-SN:S405392-IPK			IP Address: 192.168.102.191 / 255.255.254.0 / Defrouter: 192.168.102.52							
		INPUTS				FAN IN(S)/OUT(S)				
SW Port#	Speed 1/2Gbps	Device Name/ID	Device Type	Device Adapter Id	Device Adapter WWPN	Device Name/ID	Device Type	Device Adapter Id	Device Adapter WWPN	SW Port#
0	F-Port / 2Gbps	S/N:32210	HDS9960	CL1P-A (mode00)	500060e8027dd200	db02.prod.nymex.com	Sun v880	PCI Slot 7	21:00:00:E0:8B:07:6F:AC	McData3232-SN:S405392-IPK (port 8)
1	F-Port / 2Gbps	S/N:32210	HDS9960	CL1P-B (mode00)	500060e8027dd201	c2202.prod.nymex.com	IBM x445	Adapter1 - Slot5	21:00:00:E0:8B:0F:D4:75	McData3232-SN:S405392-IPK (port 10)
2	F-Port / 2Gbps	S/N:32210	HDS9960	CL1P-C (mode00)	500060e8027dd202					
3	F-Port / 2Gbps	S/N:32210	HDS9960	CL1P-D (mode00)	500060e8027dd203					
4	F-Port / 2Gbps	S/N:32210	HDS9960	CL1Q-E (mode00/RCU-T)	500060e8027dd204	S/N:32198	HDS9960	CL1Q-E (mode00)	500060e8027dc604	McData3232-SN:S405762-1NE (port 4)
5	F-Port / 2Gbps	S/N:32210	HDS9960	CL1Q-F (mode00/RCU-T)	500060e8027dd205	S/N:32198	HDS9960	CL1Q-F (mode00)	500060e8027dc605	McData3232-SN:S405762-1NE (port 5)
6	F-Port / 2Gbps	S/N:32210	HDS9960	CL1Q-G (mode00/Init)	500060e8027dd206	S/N:32198	HDS9960	CL1Q-G (mode00)	500060e8027dc606	McData3232-SN:S405762-1NE (port 6)
7	F-Port / 2Gbps	S/N:32210	HDS9960	CL1Q-H (mode00/Init)	500060e8027dd207	S/N:32198	HDS9960	CL1Q-H (mode00)	500060e8027dc607	McData3232-SN:S405762-1NE (port 7)
8	F-Port / 1Gbps	db02.prod.nymex.com	Sun v880	PCI Slot 7	21:00:00:E0:8B:07:6F:AC	S/N:32210	HDS9960	CL1P-A (mode00)	500060e8027dd200	McData3232-SN:S405392-IPK (port 0)
9	F-Port / 1Gbps	db02.prod.nymex.com	Sun v880	PCI Slot 8	21:00:00:E0:8B:07:55:AE	S/N:32210	HDS9960	CL2V-A (mode00)	500060e8027dd210	McData3232-SN:S405392-IPK (port 22)
10	F-Port / 2Gbps	c2202.prod.nymex.com	IBM x445	Adapter1 - Slot5	21:00:00:E0:8B:0F:D4:75	S/N:32210	HDS9960	CL1P-B (mode00)	500060e8027dd201	McData3232-SN:S405392-IPK (port 1)
11	F-Port / 2Gbps	c2202.prod.nymex.com	IBM x445	Adapter0 - Slot6	21:00:00:E0:8B:0F:4D:76	S/N:32210	HDS9960	CL2V-B (mode00)	500060e8027dd211	McData3232-SN:S405392-IPK (port 23)
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										
22	*F-Port / 2Gbps*	*S/N:32210*	*HDS9960*	*CL2V-A (mode00)*	*500060e8027dd210*	db02.prod.nymex.com	Sun v880	PCI Slot 8	21:00:00:E0:8B:07:55:AE	McData3232-SN:S405392-IPK (port 9)
23	*F-Port / 2Gbps*	*S/N:32210*	*HDS9960*	*CL2V-B (mode00)*	*500060e8027dd211*	*c2202.prod.nymex.com*	*IBM x445*	*Adapter0 - Slot6*	*21:00:00:E0:8B:0F:4D:76*	*McData3232-SN:S405392-IPK (port 11)*
24	*F-Port / 2Gbps*	*S/N:32210*	*HDS9960*	*CL2W-E (mode00/RCU-T)*	*500060e8027dd214*	*S/N:32198*	*HDS9960*	*CH2W-E (mode00)*	*500060e8027dc614*	*fcsw02-1NE (port 24)*
25	*F-Port / 2Gbps*	*S/N:32210*	*HDS9960*	*CL2W-F (mode00/RCU-T)*	*500060e8027dd215*	*S/N:32198*	*HDS9960*	*CH2W-F (mode00)*	*500060e8027dc615*	*fcsw02-1NE (port 25)*
26	*F-Port / 2Gbps*	*S/N:32210*	*HDS9960*	*CL2W-G (mode00/Init)*	*500060e8027dd216*	*S/N:32198*	*HDS9960*	*CH2W-G (mode00)*	*500060e8027dc616*	*fcsw02-1NE (port 26)*
27	*F-Port / 2Gbps*	*S/N:32210*	*HDS9960*	*CL2W-H (mode00/Init)*	*500060e8027dd217*	*S/N:32198*	*HDS9960*	*CH2W-H (mode00)*	*500060e8027dc617*	*fcsw02-1NE (port 27)*
28	E-Port / 1Gbps	*A3-UNYMAR-B120 (6th. Fl.)*	*FSP-3000*	*Adva Port-3*	*N/A (DWDM ISL)*	???	*FSP-3000*	*Adva Port-3*	*N/A (DWDM ISL)*	*McData3232-SN:S405762-1NE (port 30)*
29	E-Port / 2Gbps	ES3232-SerNumTBD-IPK	ES3232	Port 29	N/A (LOCAL IPK ISL)					
30	E-Port / 2Gbps	ES3232-SerNumTBD-IPK	ES3232	Port 30	N/A (LOCAL IPK ISL)					
31	E-Port / 1Gbps	A3-UNYMAR-B300 (2nd. Fl.)	FSP-3000	Adva Port-3	N/A (DWDM ISL)	???	FSP-3000	Adva Port-3	N/A (DWDM ISL)	*McData3232-SN:S405762-1NE (port 31)*

Figure 4: Zone-Port for one of the four (qty. 4) ES3232 McData switches. (Zoom in to see detail).

	A	B	C	D	E	F	G
1	**CU:LDEV**	**EMUL**	**TYPE**	**~SIZE (MB)**	**Host Usage**	**CHA Port / LUN Mapping**	**Filesystem**
2	00:00	OPEN-9	REGULAR	7042.5	db01.prod (Sun)	CH1A & 2A as LUN 00 (LUSE Head)	/u01
3	00:01	OPEN-9	REGULAR	7042.5	Luns 00-09	CH1A & 2A as LUN 00 (LUSE Disk)	/u01
4	00:02	OPEN-9	REGULAR	7042.5		CH1A & 2A as LUN 00 (LUSE Disk)	/u01
5	00:03	OPEN-9	REGULAR	7042.5		CH1A & 2A as LUN 00 (LUSE Disk)	/u01
6	00:04	OPEN-9	REGULAR	7042.5		CH1A & 2A as LUN 00 (LUSE Disk)	/u01
7	00:05	OPEN-9	REGULAR	7042.5	C22PROD (x445MP)	CH1B/2B as LUN 05 (LUSE Head)	G:/
8	00:06	OPEN-9	REGULAR	7042.5	Luns 04-07	CH1B/2B as LUN 05 (LUSE Disk)	G:/
9	00:07	OPEN-9	REGULAR	7042.5	C22QA (x445MP)	CH1D/1B/2B as LUN 00 (LUSE Head)	F:/
10	00:08	OPEN-9	REGULAR	7042.5	Luns 00-03	CH1D/1B/2B as LUN 00 (LUSE Disk)	F:/
11	00:09	OPEN-9	REGULAR	7042.5		CH1D/1B/2B as LUN 00 (LUSE Disk)	F:/
12	00:0a	OPEN-9	REGULAR	7042.5		CH1D/1B/2B as LUN 00 (LUSE Disk)	F:/
13	00:0b	OPEN-9	REGULAR	7042.5		CH1D/1B/2B as LUN 00 (LUSE Disk)	F:/
14	00:0c	OPEN-9	REGULAR	7042.5		CH1D/1B/2B as LUN 00 (LUSE Disk)	F:/
15	00:0d	OPEN-9	REGULAR	7042.5			
16	00:0e	OPEN-9	REGULAR	7042.5			
17	00:0f	OPEN-9	REGULAR	7042.5			
18	00:10	OPEN-9	REGULAR	7042.5			
19	00:11	OPEN-9	REGULAR	7042.5			
20	00:12	OPEN-9	REGULAR	7042.5			
21	00:13	OPEN-9	REGULAR	7042.5			
22	00:14	OPEN-9	REGULAR	7042.5			
23	00:15	OPEN-9	REGULAR	7042.5			
24	00:16	OPEN-9	REGULAR	7042.5			
25	00:17	OPEN-9	REGULAR	7042.5			
26	00:18	OPEN-9	REGULAR	7042.5			
27	00:19	OPEN-9	REGULAR	7042.5			
28	00:1a	OPEN-9	REGULAR	7042.5			
29	00:1b	OPEN-9	REGULAR	7042.5			
30	00:1c	OPEN-9	REGULAR	7042.5			
31							

CL1, PG1 | CL1, PG2 | CL1, PG3 | CL1, PG4 | CL1, PG5 | CL1, PG6 | CL1, PG7 | CL1-PG8 | CL1, PG9

Figure 5: In the HDS9900 series (9960/9970/9980/9990) upon initial configuration, RAID5 PARITY GROUPS are created using either (3+1 = Basic 4) or (7+1 = Basic 8) configurations. Next, from those PARITY GROUPS, smaller logical devices are created by splitting each up according to the Open Emulation desired. Above we see a parity group formed by using four 72GB drives, which was subsequently split (at array initialization time) using an OPEN-9 Emulation. The resulting CU:LDEV devices could be assigned to hosts by mapping them to CH ports, and implementing appropriate LUN masks (via Santinel).

Some Tools Used:

- San Pilot & CLI (McData ES3232).
- Remote Console 4.x and HDS/StoreEdge 9960 Service Processor Laptop.
- Santinel for LUN Masking with the HDS 9960.
- Hitachi Command Control Devices (CCD) and horcm (similar to EMC's gatekeeper and ECC commands)
- Hitachi DataLink Manager (HDLM) (also MPXIO)
- Hitachi Cruise Control (and optimizer similar to EMC's SymOptimizer).
- LUSE devices (Logical Unit Storage Extension) (similar to EMC's Metadevice).
- Qlogic 23xx HBA's (Some I flashed with FCODE for use with Sun SPARC servers; and others I flashed with BIOS Code for use with Windows & Linux x86 based servers).

Parenthetical note: The HDS9960 arrays above where acquired through Sun, as StoreEdge 9900 arrays. The firms plans were to transform to a Windows based trading platform, and because they did not recognize that these Sun re-badged Hitachi arrays could also be used for Windows, Linux (etc.) hosts, the firm's development roadmap had no plans to use these arrays, even though they were practically brand new, and cost a significant amount to acquire. With no documentation (except those on a micro-code CD I found) and with no support contract from Sun or Hitachi, I convinced the, then VP of Technology, to allow me to divert some of my consulting time to investigating and building the local and distance SAN solution above.

To do this, it was necessary for me to take a proof-of-concept approach with these arrays, before I could convince the firm to purchase FibreChannel switches, use DWDM lines, or do anything that would incur collateral cost to implementing these arrays (which, again, they had no plans for). Towards that end I had to do things like, get the two arrays communicating/replicating via FC-AL using a simple back-to-back 50µm cable; then get them replicating via NISHAN IPS4000 FCIP switches (i.e. so the IP infrastructure could be used to test site-to-site replication without using DWDM lines), etc. Once the proof-of-concept tests were demonstrated to work, the firm allowed me to design the FibreChannel fabric based infrastructure above.

Calculating Fibre Channel switch port Buffer Credit requirements for Distance SAN's (v6)

Noelle Milton Vega
Rensselaer Technology Group, LTD.

Introduction

When purchasing a fibre channel switch (and licenses) for the purpose of implementing a multi-site SAN fabric, one switch feature you will need to consider is its buffer credit capability. In particular, how much of it will you need to ensure the optimal streaming of data over the distance covered by the two sites? This is important to know, since having too few buffers can cause application I/O to block as it waits for acknowledgment from the far-end; while having too many buffers will cause excess I/O to queue up in the switch (more on that later), not to mention unnecessarily increase the cost of the switch. The aim of this white paper is to demonstrate how to right size a FC switch with respect to buffer credits.

The key to understanding Fibre Channel buffers, and how many are required to accommodate adjacent FC ISL connected ports, is to first understand the concept of "line frame-length". **Line frame-length** is the number of frames (of whatever type) that can exist in transit, at any one time, along the fibre optic path that connects two adjacent nodes/sites. As we'll see, although the linear distance between ends of a fibre optic path is fixed, it's line frame-length actually varies as follows:

(1) It increases linearly as the transmission speed of the equipment in use **increases** (e.g. 1.0 $Gbps_{fc}$ -> 2.0 $Gbps_{fc}$ -> 10.0 $Gbps_{fc}$), and vice versa,

(2) It increases as the size of the frame decreases, and vice versa. For FC frames, the frame size is determined by its payload and frame header sizes.

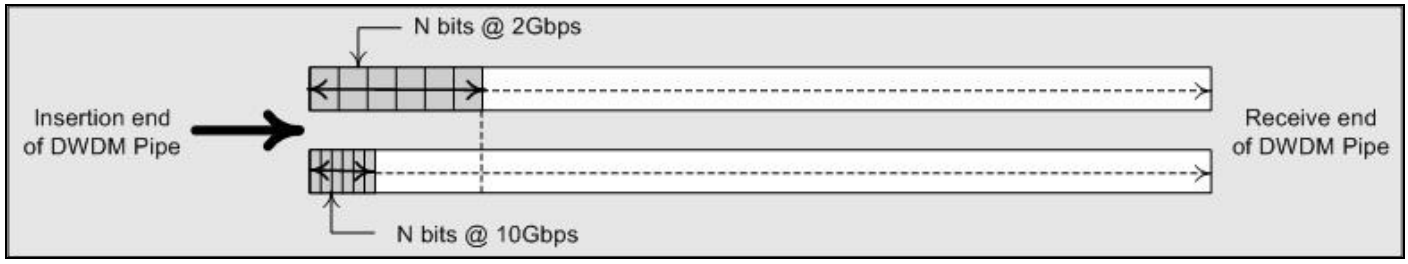
If a fibre optic path happens to represent an FC ISL link, then, in knowing its **line bit-length**, we can calculate the maximum number Fibre Channel frames that can simultaneously fit (source to destination and back) along that link. And that is the number of buffers your ISL switch ports will require, and for the following reasons:

(1) To safely store-and-forward frames at the source end, while they are in transit (on a piece of glass) to the destination end, and ultimately acknowledged as having been received,

(2) To, in the meantime, provide an immediate "successfully sent" I/O acknowledgment to the sending application so it does not have to block waiting for the frame to reach the opposite end, and then for the acknowledgment coming back.

Consider the following practical example. A company decides to locate its disaster recovery site 100 miles away from its primary operating site. It purchases Fibre Channel Switch and DWDM equipment with $2Gbps_{fc}$ interfaces, and with an appropriate number of buffers to accommodate continuous streaming of data between end ISL ports. Several months after the implementation, monitoring of the ISL links reveals congestion between the two sites. As a remedy, the company considers upgrading its end-to-end Switch and DWDM equipment to support $10Gbps_{fc}$ transmission rates. However, doing so, as indicated previously, will increase the "effective distance" or **line bit-length** between the two sites by 5 fold. Stated differently, it's as if the DR site were moved from being 100 miles away to being 500 miles away.

Why? Because, as illustrated in the diagram below, relative to the $2Gbps_{fc}$ speed, at $10Gbps_{fc}$, the physical distance between each inserted bit is 5 times smaller. This *relative* ***compression*** is due the to faster rate at which optical variations (bits) are now being introduced into the front end of the DWDM pipe by the FC switch ISL port (5 times as fast in our example).



Note that the diagram above is not drawn to scale.

Above, the gray boxes for the $2Gbps_{fc}$ and $10Gbps_{fc}$ cases contain the same number of optical variations (bits) in transit. However, the gray box for the $10Gbps_{fc}$ case consumes $1/5^{th}$ the physical fibre distance relative to its $2Gbps_{fc}$ counterpart. Stated differently, relative to the $2Gbps_{fc}$ case, at $10Gbps_{fc}$, 5 times the amount of bits (data frames) can fit along the same fibre optic path.

This means that for the $10Gbps_{fc}$ case, the same frame, while traveling the same physical distance, will now have 5 times as many frames behind it (in transit on the glass) when it reaches the opposite end (again, because of the compression relative to the $2Gbps_{fc}$ case). If, for example, the data in the gray boxes happened to represent one Fibre Channel frame then, relative to the $2Gbps_{fc}$ case, at $10Gbps_{fc}$ the application could issue the equivalent of 5 times as many FC frames before ever receiving an I/O acknowledgement for the first frame. That's because, at the input end, each frame is being inserted and acknowledged locally in $1/5^{th}$ the time. The resulting additional frames must be buffered at the source end to prevent the application from blocking while it waits for I/O acknowledgment.

Note that for a variety of reasons, the remote (destination) site storage array and/or host equipment can cause the source site application I/O to block. For example, if the destination site employed slower storage and/or host equipment relative to the source site, then during data bursts the destination equipment may not have enough horsepower to process the additional incoming I/O's fast enough. So, while allocating an appropriately sufficient number of switch buffer credits cannot guarantee continuous streaming of data between two sites, it does guarantee that the end-to-end ISL based fabric is not the cause of such a shortcoming.

Before we begin an example of how to calculate buffer requirements, it is important to know the numerical definition of a Fibre Channel Gigabit, as well as to understand the structure of a Fibre Channel Frame.

In the Fibre Channel world, one gigabit is defined to be 1,062,500,000 bits (which is not $(1024)^3$). Other Fibre Channel gigabit values are then derived from this reference definition. For example, two (Fibre Channel) gigabits = 2 x 1,062,500,000 bits = 2,125,000,000 bits. To avoid confusion with the traditional (non Fibre Channel) definition of a gigabit, throughout this document I will use the symbol Gb_{fc} to mean "1,062,500,000 bits", or 1 Fibre Channel Gigabit.

In summary:

1 Gb_{fc} = 1,062,500,000
2 Gb_{fc} = 2,125,000,000
10 Gb_{fc} = 10,625,000,000

Next, we show the anatomy of a Fibre Channel Frame with notes.

Start of Frame	4 bytes	32 bits
***Standard* Frame Header**	24 bytes	192 bits
Data (payload)	[0 – 2,112] bytes	[0 – 16,896] bits
CRC	4 bytes	32 bits
End of Frame	4 bytes	32 bits
TOTAL (Nbr bits/frame):	[36 – 2,148] bytes	288 – 17,184 bits

Notes:

The term byte used here, and in the Table 3 above means 8 bits (not the 10 bits that result from 8/10 bit encoding).

The maximum Fibre Channel frame size is 2,148 bytes.

The final frame size must be a multiple of 4 bytes. Thus the Data (payload) segment will, as necessary, be padded with 1 to 3 "fill-bytes" to achieve an overall 4 byte frame alignment.

The standard Frame Header size is 24 bytes. However, up to 64 additional bytes (for a total of an 88 byte header) can be included for applications that need extensive control information. Since the total frame size cannot exceed the maximum of 2,148 bytes, these additional Header bytes will subtract from the Data segment size by as much as 64 bytes (per frame). This is why the maximum Data (payload) size is 2,112 (because [2,112 – 64] = 2,048, which is exactly 2K-bytes of data).

The final frame, once constructed, is passed through the 8byte to 10byte conversion process.

In the FC world, 1 Word = 4 x 8/10 bit encoded bytes (40 bits).

Our Example

Let say, for the purposes of discussion, that we have redundant (i.e. two) fibre optic paths between a primary and secondary site.[1] The linear distance (as opposed to displacement) is 91.732608 Km (or about 57 miles) for one path, and 43.452288 Km (or about 27 miles) for the other path.

We start first with the speed at which all electrical variations (i.e. baud)[2] propagate through transmission mediums. This is the speed of light:

299,792.458 km / s == .00000333564095 s / km

Speed of Light

Next we take the linear distance between the two sites and determine:

(1) How many seconds it takes for 1-bit to travel the one-way distance linear between the two sites; that is, express the "distance" between the two sites in seconds. (Column 2 below). This is determined by the speed of light.

(2) Having determined the one-way distance in seconds between the two sites (a fixed number), we can now determine the maximum number of bits that can exist, in transit, between the two sites at any one time. In other words, we can calculate the "distance" between the two sites in bits (as opposed to miles). This is determined by the speed of light, as well as by the rate at which the transmitting equipment (e.g. a fibre channel port) can create electrical variations onto the medium (i.e. fibre optic line). In other words, how fast can it push bits onto one end of the fibre optic line (usually 1 Gb_{fc}, 2 Gb_{fc}, or 10 Gb_{fc}). (Columns 3 & 4).

Table 1

Distance (Km)	Distance (secs)	# of Bits (1Way)	8/10 FC "bytes"
91.732608 (57mi)	0.000305987044	325,111.234	32,511.1234
43.452288 (27mi)	0.000144941231	154,000.058	15,400.0058

Values for bits, bytes, & frames were calculated assuming a **1** Gb_{fc} bit insertion rate (i.e. the rate at which bits are introduced into the frond end of the DWDM fibre optic line).

Table 2

Distance (Km)	Distance (secs)	# of Bits (1Way)	8/10 FC "bytes"
91.732608 (57mi)	0.000305987044	650,222.468	65,022.2468
43.452288 (27mi)	0.000144941231	308,000.116	30,800.0116

Values for bits, bytes, & frames were calculated assuming a **2** Gb_{fc} bit insertion rate (i.e. the rate at which bits are introduced into the frond end of the DWDM fibre optic line).

[1] In DWDM terminology, redundant fibre optic paths is sometimes referred to as path protection.

[2] Note: 1-Baud (electrical variation) will represent a different number of bits depending on the compression codecs used. In our example, 1 baud represents 1 bit.

Calculations notes for the tables above:

Column 2 indicates the amount of time (in seconds) it takes one **OPTICAL VARIATION** (however many number of bits that optical variation happens to represent) to travel the one-way distance specified in column one. It comes from multiplying the number of seconds it takes **light** to travel 1 km (i.e. .00000333564095 s / km), by the number of km traveled which, again, is specified in Column 1.

Column 3 is the product of column 2 (i.e. the line distance in seconds) and the FC bit insertion rate (either 1,062,500,000 for Table 1, or 2,125,000,000 for Table 2). This column effectively represents the amount of additional I/O that could have been processed at the host, had the response to that I/O been instantaneous; (actually twice that amount, since acknowledgment time for those I/O's, coming back the other way, have to be accounted for as well).

Column 4 is Column 3 divided by 10. This is done to group single bits in transit into equivalent 8/10 bit "byte" quantities. The Fibre Channel protocol converts every 8 bit byte into a 10 bit equivalent (via the 8/10 bit encoding algorithm) before transmitting it. So the value in this Column 4 will determine the number of buffers needed for an ISL switch port.

Comparing the two tables above, we can see that the faster the transmitting Fibre Channel port is, the more total number of bits that particular port can "insert" into the line before the very first one it inserted reaches the opposite end. That's because the time in between bit insertions is reduced for faster ports (i.e. an increased rate at which bits are being introduced into the front end of the DWDM pipe). Therefore, relatively speaking, a fixed length fibre optic path *effectively* gets longer (bit length wise) as the speed of the trasmit/receive equipment increases. The line (bit) length for 2 $Gbps_{fc}$ xmit/recv equipment is twice as long as it is for 1 $Gbps_{fc}$ xmit/recv equipment. Similarly, the line (bit) length for 10 $Gbps_{fc}$ xmit/recv equipment is five times as long as it is for 2 $Gbps_{fc}$ xmit/recv equipment. The net/practical effect of this is that the faster the FC port, the larger the number of buffers that will be required for your switch ISL ports.

Given the **linear distance** between the two datacenters (in Kilometers) and the **speed of the fibre channel** equipment (1 $Gbps_{fc}$, 2 $Gbps_{fc}$, 10 $Gbps_{fc}$), we have thus far been able calculate **"line bit length"** of the link between the two sites (i.e. the number of bits that can fit on the line, end to end, at any one time). Knowing the line bit length, we can divide this value by the number of bits in a Fibre Channel frame to determine the equivalent **"line Fibre Channel frame length"** (in other words, the number of FC frames that can fit on the line, end to end, at any one time). As we saw from table 1, the number of bits contained in a FC frame varies with: (1) the Data/Payload size, (2) padding, which ensures a FC fame whose final size is a multiple of 32 bits (4 bytes), and (3) the size of the Frame header, which can range from 24 to 88 btyes.

Frame Length8 (in km) = (299,792.458 ***km/s***) x
(Seconds-Between-Inserted-Bits ***secs/bit***) x
(Number-Of-Bits-Per-Frame ***bits***)

Frame Length10 (in km) = (299,792.458 ***km/s***) x
(Seconds-Between-Inserted-Bits ***secs/bit***) x
(Number-Of-Bits-Per-Frame ***bits***) x 10/8

Where:

Seconds-Between-Inserted-Bits = 1/1,062,500,000 (for 1Gbps FC)
= 1/2,125,000,000 (for 2Gbps FC)

Number-Of-Bits-Per-Frame = Variable depending on Data payload and/or Header size. See table and notes above. In most cases the Header will be the Standard size of 24 bytes.

Table 4

Number of bits / frame (**after** 8/10 bit encoding)	Frame Length (km) @ 1 Gbps FC (.000282157607529411 km/bit)	Number of in-transit frames (1-way) for 91.732608 km DWDM leg.
17,184 / 21,480	6.0607 km (2112 PL-bytes)	15.1355 frames (buffers)
16,672 / 20,840	5.8801 km (2048 PL-bytes)	15.6003 frames (buffers)
8,480 / 10,600	2.9908 km (1024 PL-bytes)	30.6708 frames (buffers)
4,384 / 5,480	1.5462 km (0512 PL-bytes)	59.3268 frames (buffers)

1Gbps (1,062,500,000 bps) data input rate.

Table 5

Number of bits / frame (**after** 8/10 bit encoding)	Frame Length (km) @ 2 Gbps FC (.000141078803764705 km/bit)	Number of in-transit frames (1-way) for 91.732608 km DWDM leg.
17,184 / 21,480	3.0303 km (2112 PL-bytes)	30.2170 frames (buffers)
16,672 / 20,840	2.9400 km (2048 PL-bytes)	31.2006 frames (buffers)
8,480 / 10,600	1.4954 km (1024 PL-bytes)	61.3417 frames (buffers)
4,384 / 5,480	0.7731 km (0512 PL-bytes)	118.6537 frames (buffers)

2Gbps (2,125,000,000 bps) data input rate.

Notes:

- Column 1 represents the total number of 8/10 bits in a Fibre Channel frame (with a standard header size of 24 bytes) at varying Data payloads (**PL**). From top to bottom, the payloads used to calculate each row of column 1 are, respectively: 2112 bytes, 2048 bytes, 1024 bytes, and 512 bytes.

- Column 2 represents the product of the 10 bit value in Column 1, and (1/1,062,500,000) for the 1Gbps (Table 4), and (1/2,125,000,000) for the 2Gbps (Table 5). Thus, this Column (Column 2) essentially represents the linear distance that 1 (one) single frame consumes for the specified payload (PL).

- Column 3 represents the quotient derived by dividing the longest (worst case) DWDM line distance (in kilometers), by the number of kilometers per frame (calculated in Column 2). Thus, this column essentially indicates how many additional ONE-WAY frames worth of data could have been processed by the host/application, had the response to the first frame been instantaneous. In other words, this is how many ONE-WAY (not round trip) switch buffers you would need to allow non-stop transmission. Double (i.e. round trip) the values in this column 3 to yield the number of buffers required of your ISL switch ports.