



Schneller Interconnect für Spectrum Scale

März 2017

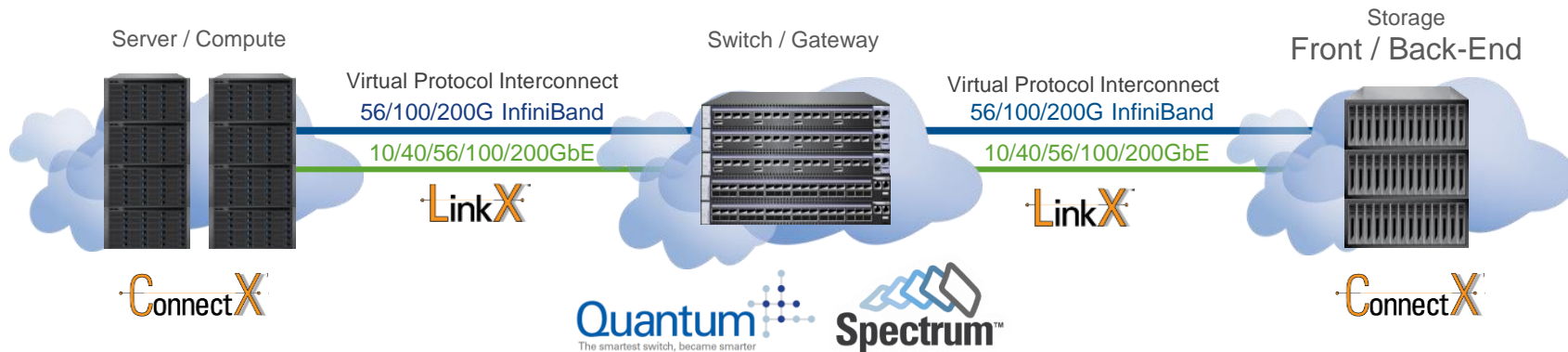
 **Mellanox**
TECHNOLOGIES
Connect. Accelerate. Outperform.™

- **Leading provider of high-throughput, low-latency Server and Storage Interconnect**
 - EDR 100Gb/s (5th generation) InfiniBand and 10/25/40/50/100GbE – HDR 200G coming later this year
 - Reduces application wait-time for data
 - Dramatically increases ROI on data center infrastructure

- **Company headquarters:**
 - Yokneam, Israel; Sunnyvale, California
 - ~ 3,000 employees worldwide

- **Solid financial position**
 - Record revenue last 3 years >\$1B
 - On a run rate close to \$1B annual Sales





Comprehensive End-to-End InfiniBand and Ethernet Portfolio (VPI)

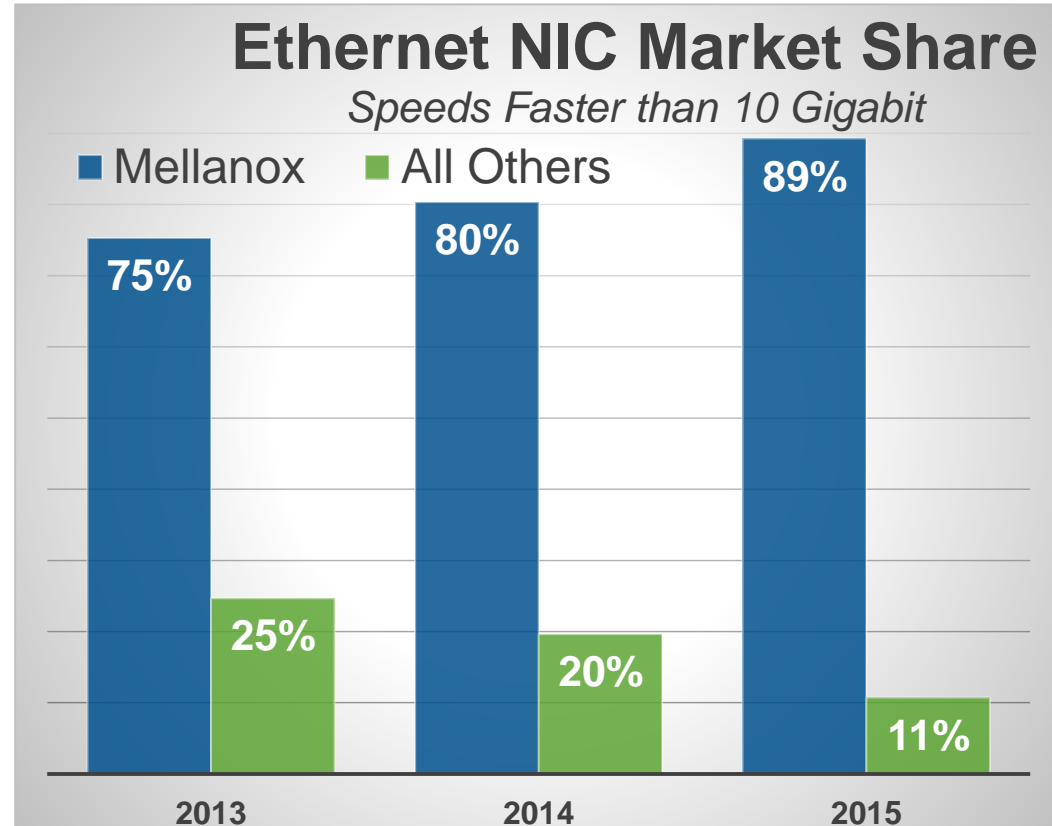
ICs	Adapter Cards	NPU & Multicore	Switches/Gateways	Software	Metro / WAN	Cables/Modules

■ Why Mellanox

- Established technology leader
- Dominant in High Speed Ethernet
- First to market with 25, 50, and 100GbE
- Fastest growing Ethernet switch vendor
- Only End to End Ethernet Solution
- Trusted supplier to all major OEMS

■ Mellanox Ethernet Switches

- Twice the performance at half the price
- Unique form factors
 - Ideal for hyper-converged networking
- Best in-class performance
 - Throughput
 - Latency
 - Power consumption
 - Highest Value/\$

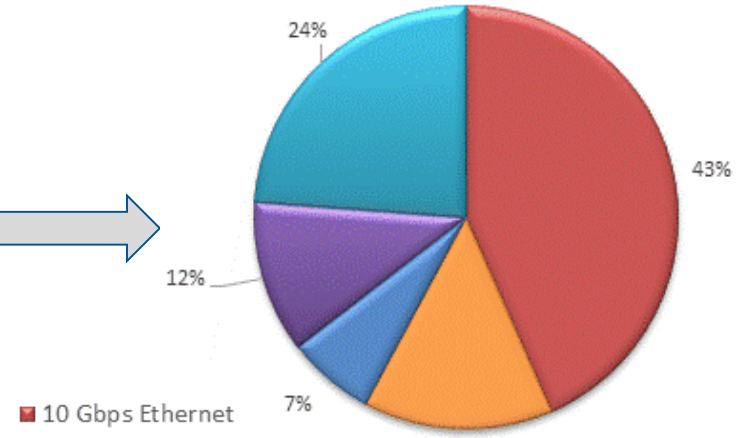
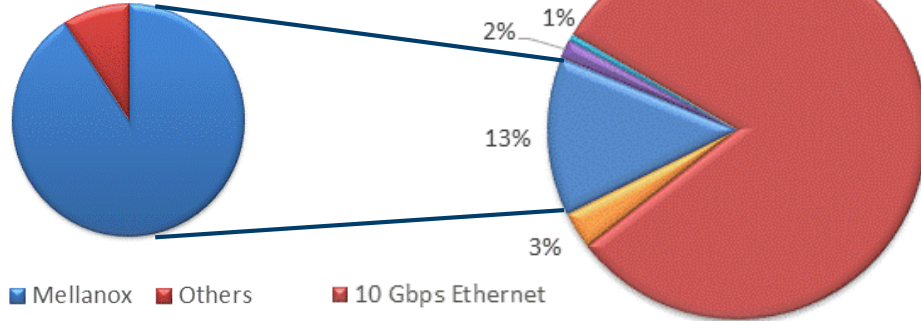


Source: Crehan Research, March 2016

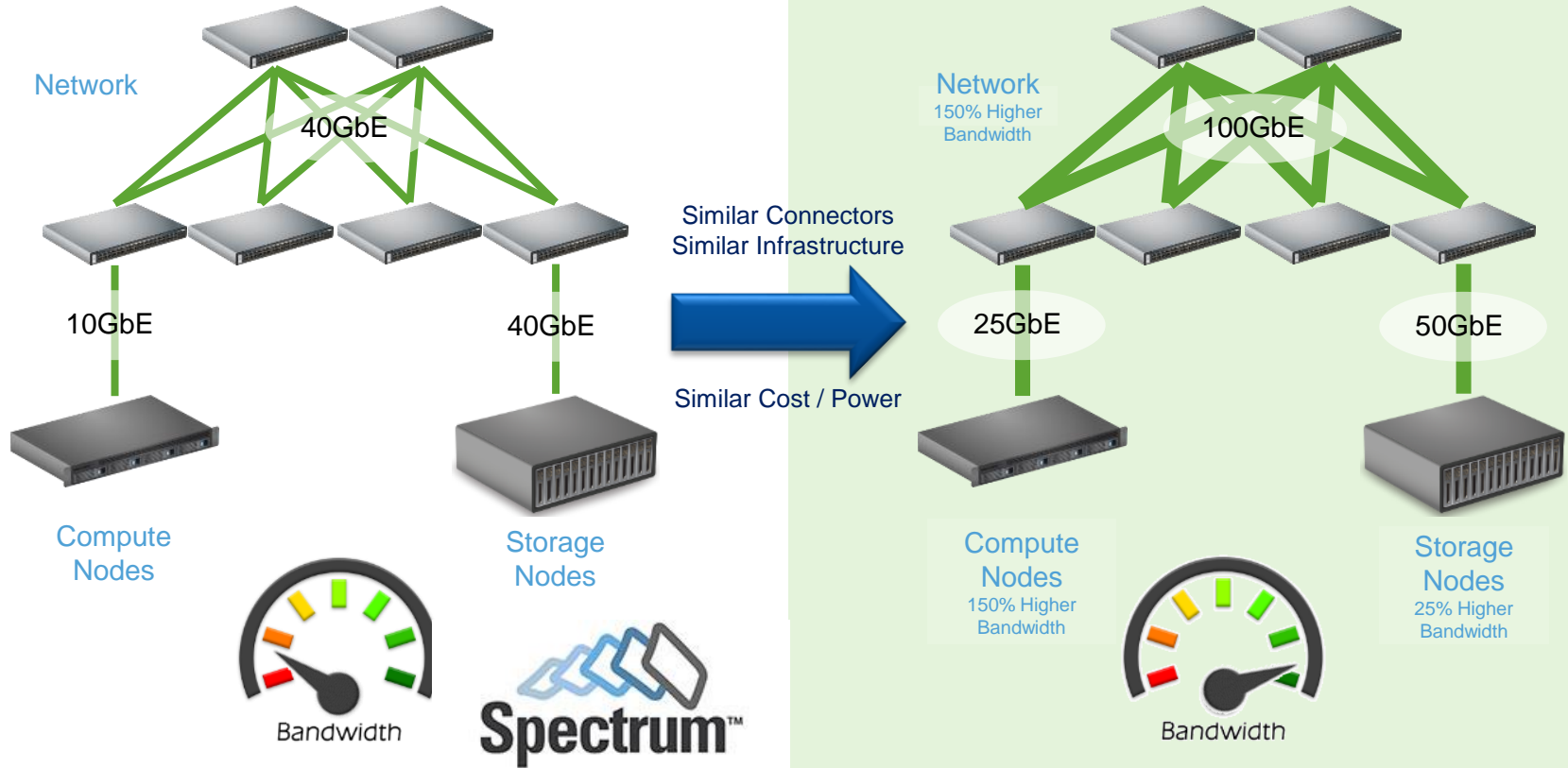
High Speed Adapter Market Forecast 2020 (\$1.8B)

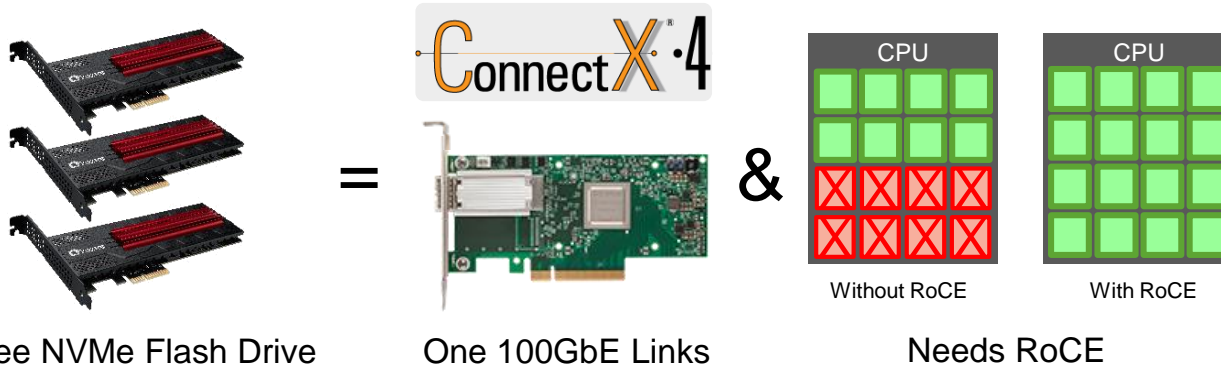
High Speed Adapter Market Forecast 2016 (\$1.4B)

40GbE Adapter Market Share, Q3'15 (\$23.9M)



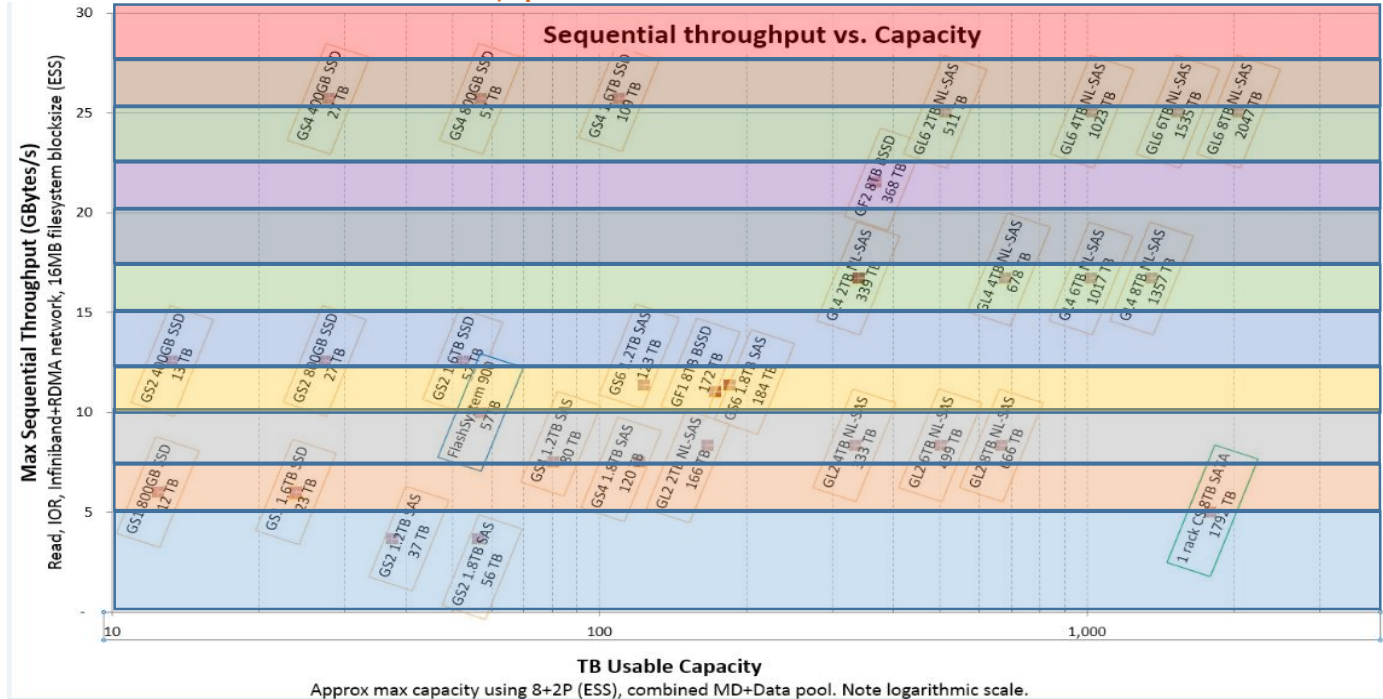
Key Messages:
 25 is the New 10!
 50 & 100 are Now!
 Future Proof Your Network!
 With 10/25Gb/s Ethernet!





- Just three NVMe Flash can saturate 100 Gb/s Link
 - Needs 100GbE ConnectX-4 & RDMA
- RDMA
 - Burn Rubber! Not CPU Cycles.

IBM Mellanox Infrastructure for ESS/Spectrum Scale

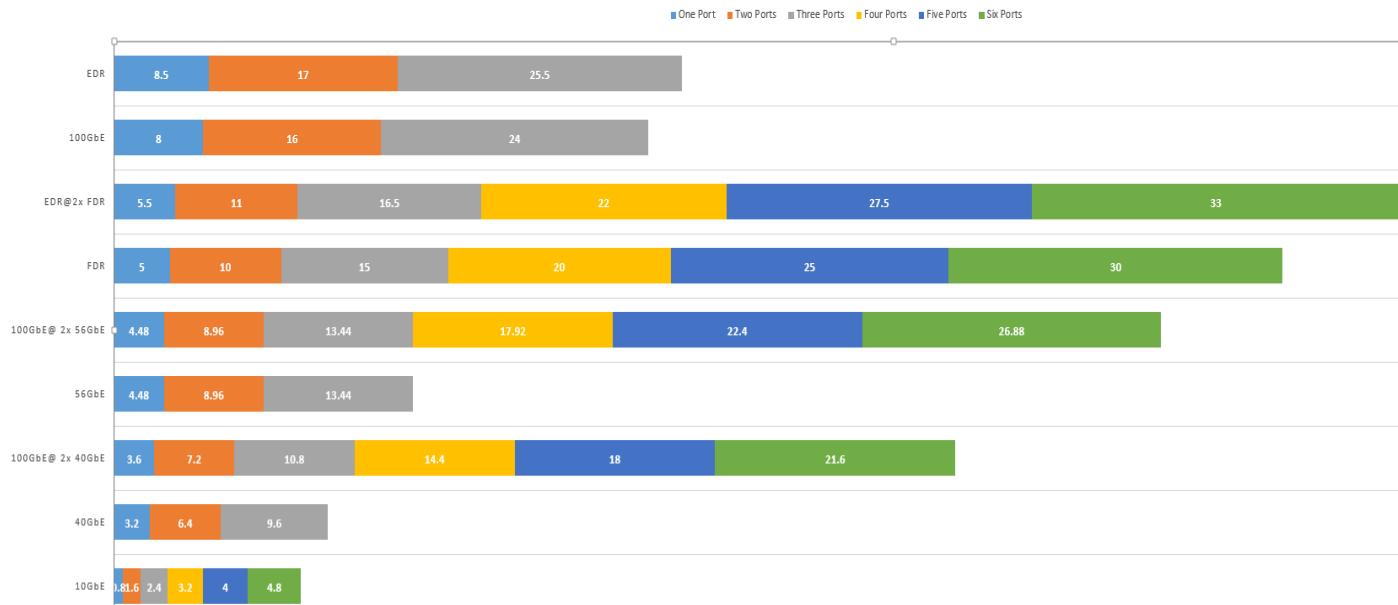


Dual NSD Port Bandwidth options

Ports	10GbE	40GbE	100GbE@ 2x 40GbE	56GbE	100GbE@ 2x 56GbE	FDR	EDR@2x FDR	100GbE	EDR
One Port	1.6	6.4	7.2	8.96	8.96	10.0	11.0	16.0	17.0
Two Ports	3.2	12.8	14.4	17.92	17.92	20.0	22.0	32.0	34.0
Three Ports	4.8	19.2	21.6	26.88	26.88	30.0	33.0	48.0	51.0
Four Ports	6.4		28.8		35.84	40.0	44.0		
Five Ports	8.0		36		44.80	50.0	55.0		
Six Ports	9.6		43.2		52.76	60.0	66.0		

IBM Mellanox Infrastructure for ESS/Spectrum Scale

GB Bandwidth per Port per Speed for single NSD



SINGLE NSD Port Bandwidth options

Ports	10GbE	40GbE	100GbE@ 2x 40GbE	56GbE	100GbE@ 2x 56GbE	FDR	EDR@2x FDR	100GbE	EDR
One Port	0.8	3.2	3.6	4.48	4.48	5.0	5.5	8.0	8.5
Two Ports	1.6	6.4	7.2	8.96	8.96	10.0	11.0	16.0	17.0
Three Ports	2.4	9.6	10.8	13.44	13.44	15.0	16.5	24.0	25.5
Four Ports	3.2		14.4		17.92	20.0	22.0		
Five Ports	4.0		18.0		22.4	25.0	27.5		
Six Ports	4.8		21.6		26.88	30.0	33.0		

- <http://www.mellanox.com/oem/ibm/>



IBM and Mellanox Solutions



[Home](#)

Overview

Today's modern data centers need an agile, high performance and scalable infrastructure that is durable, simple and future ready. It needs to incorporate continuous improvements in computer, storage, networking, and application technologies. It needs to empower IT managers to deliver this infrastructure in a changing business environment, and it needs to be backed by a trusted and reliable partner. Mellanox's networking solutions based on InfiniBand, Ethernet, or RoCE (RDMA over Converged Ethernet) provide the best price, performance, and power value proposition for network and storage I/O processing capabilities.

Advanced data centers can utilize 56/100 Gb/s InfiniBand, 10/25/40/50/100Gb/s Ethernet, with RDMA/RoCE to consolidate I/O to a single wire and enable IT managers to deliver significantly higher application service levels, while reducing Capex and Opex related to I/O infrastructures. Mellanox provides deployment, manageability and performance tools for Ethernet and InfiniBand fabrics on a wide range of operating systems, supporting a diverse set of software environments to fine tune solutions for customer requirements and deliver tomorrow's data center today.

Mellanox intelligent interconnect solutions increase data center efficiency by providing the highest throughput and lowest latency, delivering data faster to applications and unlocking system performance.

Mellanox IBM Contact:
Jim Lonergan
OEM Business Development Mgr.
Mellanox Technologies
Tel: (512) 897-8245
james@mellanox.com



[Mellanox Technologies Overview](#)



Main Benefits

MAXIMIZE THE NETWORK

- Maximize datacenter connectivity ROI through: Density, Scale, Performance

OPEN THE NETWORK

- Leverage new technologies that increase functionality, investment & ROI
- Freedom from vendor lock-in

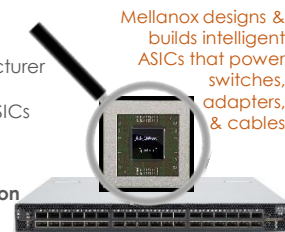
Unique Capability

MELLANOX: THE DIFFERENCE IS IN THE CHIP

- Founded as a state-of-the-art silicon chip (ASIC) manufacturer
- Intelligence built directly onto our own chip
- Other switch vendors are forced to source expensive ASICs from third parties such as Broadcom
- Mellanox uses own chip & passes savings to customers

SwitchX.2 6th Generation
10/40/56GbE & 40/56 Gb IB

SwitchIB 7th Generation
40/56/100 Gb IB



Mellanox designs & builds intelligent ASICs that power switches, adapters, & cables

Key Differentiators



VALUE & PERFORMANCE

The Enterprise Integrated model is familiar to those with traditional SAN deployments. Adding ESS/Spectrum Scale will not only eliminate the data silos, but can also improve performance and reduce data bottleneck.

The most common deployment is using Network Shared Disks, whose modular design scales performance and capacity independently.

For those familiar with HDFS, or other scale-out software-defined storage, we support shared nothing clusters that provide the native locality APIs for HDFS, but work like centralized parallel storage for other protocols. Using commodity storage rich servers can be the most economical way to scale out your storage needs.

Market Opportunities

SCALE-OUT STORAGE

Combines compute & storage, easier to manage & lowers cost – top of rack switch with density at lower price point most

attractive CLOUD

Create economies of scale through shared services – open switch platform with fairness best for software-defined DC.

MEDIA & ENTERTAINMENT

Video streaming & post-production on 4k/8k workflows –needs extreme high bandwidth to support real-time frame-rates

BIG DATA

Improved analytics for better business decisions – needs non-blocking architecture to speed data ingestion.

GENOMICS

Extreme scalability using a building - block approach: Capacity, bandwidth and a single name space expand as more building blocks are added, resulting in near-linear performance gains

SCALE OUT DATABASE

Scale out of DB2 PureScale, Oracle RAC, SAP HANA

Company Background





➤ Established 1999 * NASDAQ: MLNX

➤ End-to-end Ethernet Connectivity Solutions – Adapters, Switches, Cables, Software, Support

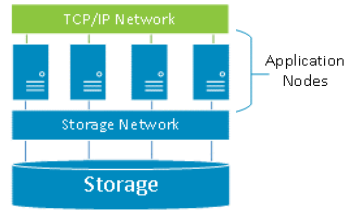
➤ World-class, non-outsourced technical support

➤ Trusted as switch manufacturer for every major server OEM

IBM Mellanox Infrastructure for ESS / Spectrum Scale

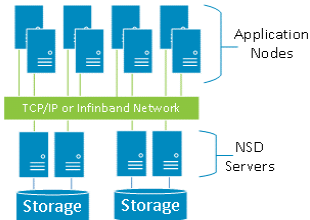
Speed	Switch	Cabling	Adapter	Optics*
FDR	SX6036 – 8831-F36 	See lists on right	EL3D	
EDR	SB7700 – 8828-E36 		EC3E EC3T	
40 GbE	SX1710 – 8831-NF2 		EC3A	EB27 + EB2J or EB2K
10/40 GbE	SX1410 – 8831-S48 		EL3X EL40 (SR) EC3A	EB28 + ECBD or ECBE

Enterprise Integrated Model



Unify and parallelize storage silos

Network Shared Disk (NSD) Model



Modular High-Performance Scaling

Shared Nothing Cluster (SNC) Model



Span storage rich servers for converged architecture or HDFS deployment

Choice of Cabling


40GbE / FDR Cabling

Length	Description	FC
0.5m	40GbE / FDR Copper Cable QSFP	EB40
1m	40GbE / FDR Copper Cable QSFP	EB41
2m	40GbE / FDR Copper Cable QSFP	EB42
3m	40GbE / FDR Optical Cable QSFP	EB4A
5m	40GbE / FDR Optical Cable QSFP	EB4B
10m	40GbE / FDR Optical Cable QSFP	EB4C
15m	40GbE / FDR Optical Cable QSFP	EB4D
20m	40GbE / FDR Optical Cable QSFP	EB4E
30m	40GbE / FDR Optical Cable QSFP	EB4F
50m	40GbE / FDR Optical Cable QSFP	EB4G

EDR Cabling

Length	Description	FC
0.5m	EDR Copper Cable QSFP28	EB50
1m	EDR Copper Cable QSFP28	EB51
2m	EDR Copper Cable QSFP28	EB52
1.5m	EDR Copper Cabling QSFP28	EB54
3m	EDR Optical Cable QSFP28	EB5A
5m	EDR Optical Cable QSFP28	EB5B
10m	EDR Optical Cable QSFP28	EB5C
15m	EDR Optical Cable QSFP28	EB5D
20m	EDR Optical Cable QSFP28	EB5E
30m	EDR Optical Cable QSFP28	EB5F
50m	EDR Optical Cable QSFP28	EB5G
100m	EDR Optical Cable QSFP28	EB5H

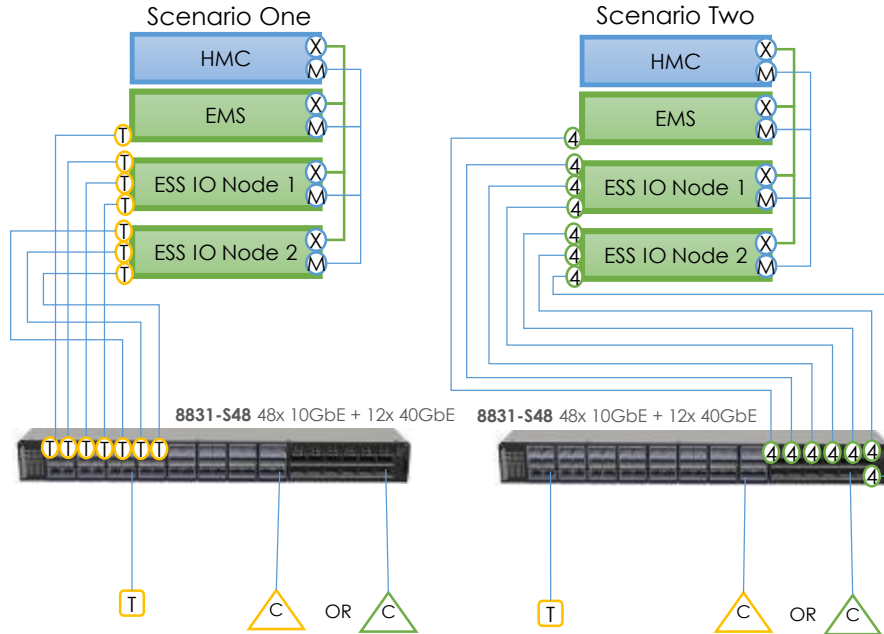
* Optics are IBM Parts only

Speed	Switch	Cabling	Adapter	Optics*
10/40 GbE	SX1410 – 8831-S48 	See list on right	EC37 / EL3X EC2M / EL40 (SR)	EB28 + ECBD or ECBE
			EC3A	EB27 + EB2J or EB2K


Choice of Cabling

40GbE / FDR Cabling

Length	Description	FC
0.5m	40GbE / FDR Copper Cable QSFP	EB40
1m	40GbE / FDR Copper Cable QSFP	EB41
2m	40GbE / FDR Copper Cable QSFP	EB42
3m	40GbE / FDR Optical Cable QSFP	EB4A
5m	40GbE / FDR Optical Cable QSFP	EB4B
10m	40GbE / FDR Optical Cable QSFP	EB4C
15m	40GbE / FDR Optical Cable QSFP	EB4D
20m	40GbE / FDR Optical Cable QSFP	EB4E
30m	40GbE / FDR Optical Cable QSFP	EB4F
50m	40GbE / FDR Optical Cable QSFP	EB4G

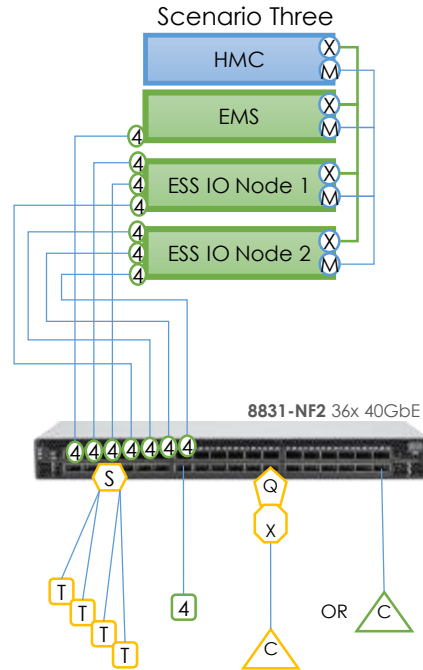












* 10GbE & Optics are IBM Parts

Speed	Switch	Cabling	Adapter	Optics*
40 GbE	SX1710 – 8831-NF2 	See list on right	EC3A	EB27 + EB2J or EB2K

Choice of Cabling

40GbE / FDR Cabling		
Length	Description	FC
0.5m	40GbE / FDR Copper Cable QSFP	EB40
1m	40GbE / FDR Copper Cable QSFP	EB41
2m	40GbE / FDR Copper Cable QSFP	EB42
3m	40GbE / FDR Optical Cable QSFP	EB4A
5m	40GbE / FDR Optical Cable QSFP	EB4B
10m	40GbE / FDR Optical Cable QSFP	EB4C
15m	40GbE / FDR Optical Cable QSFP	EB4D
20m	40GbE / FDR Optical Cable QSFP	EB4E
30m	40GbE / FDR Optical Cable QSFP	EB4F
50m	40GbE / FDR Optical Cable QSFP	EB4G
1m*	Passive QSFP+ 4x Break-out Cable	EB24
3m*	Passive QSFP+ 4x Break-out Cable	EB25
5m*	Passive QSFP+ 4x Break-out Cable	EB26



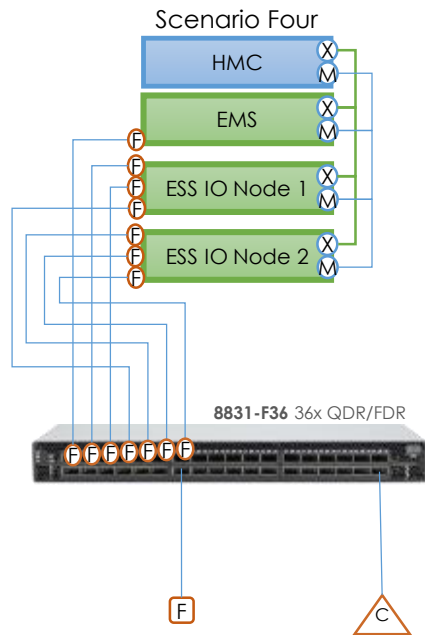
-  10 GbE Customer Network
 -  1GbE HMC Network
 -  40 GbE Data Network
 -  Passive QSFP+ Break-out Cable
 -  40 GbE Client
 -  10 GbE Client
 -  40 GbE Customer Network
 -  QSFP to SFP+ Adapter (QSA)
 -  * SFP+ DAC or Transceiver
 -  10 GbE Customer Network
- * 10GbE & Optics are IBM Parts






Speed	Switch	Cabling	Adapter	Optics
FDR	SX6036 – 8831-F36 	See list on right	EC32 / EL3D	NA

Choice of Cabling

40GbE / FDR Cabling

Length	Description	FC
0.5m	40GbE / FDR Copper Cable QSFP	EB40
1m	40GbE / FDR Copper Cable QSFP	EB41
2m	40GbE / FDR Copper Cable QSFP	EB42
3m	40GbE / FDR Optical Cable QSFP	EB4A
5m	40GbE / FDR Optical Cable QSFP	EB4B
10m	40GbE / FDR Optical Cable QSFP	EB4C
15m	40GbE / FDR Optical Cable QSFP	EB4D
20m	40GbE / FDR Optical Cable QSFP	EB4E
30m	40GbE / FDR Optical Cable QSFP	EB4F
50m	40GbE / FDR Optical Cable QSFP	EB4G

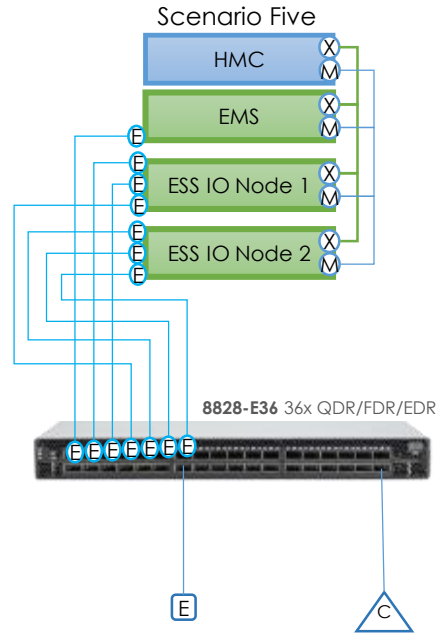






-  1GbE xCAT Network
-  1GbE HMC Network
-  FDR Data Network
-  QDR/FDR Client
-  QDR/FDR Customer Network


Speed	Switch	Cabling	Adapter	Optics
EDR	SB7700 – 8828-E36 	See list on right	EC3E EC3T	NA

Choice of Cabling

EDR Cabling		
Length	Description	FC
0.5m	EDR Copper Cable QSFP28	EB50
1m	EDR Copper Cable QSFP28	EB51
2m	EDR Copper Cable QSFP28	EB52
1.5m	EDR Copper Cabling QSFP28	EB54
3m	EDR Optical Cable QSFP28	EB5A
5m	EDR Optical Cable QSFP28	EB5B
10m	EDR Optical Cable QSFP28	EB5C
15m	EDR Optical Cable QSFP28	EB5D
20m	EDR Optical Cable QSFP28	EB5E
30m	EDR Optical Cable QSFP28	EB5F
50m	EDR Optical Cable QSFP28	EB5G
100m	EDR Optical Cable QSFP28	EB5H

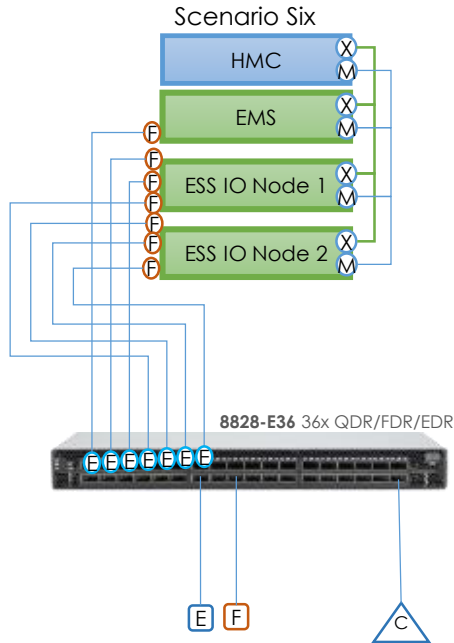









-  1GbE xCAT Network
-  1GbE HMC Network
-  EDR Data Network
-  QDR/FDR/EDR Client
-  QDR/FDR/EDR Customer Network


Speed	Switch	Cabling	Adapter	Optics
FDR / EDR	SB7700 – 8828-E36 	See list on right	EC32 / EL3D	NA
			EC3E EC3T	

Choice of Cabling

EDR Cabling		
Length	Description	FC
0.5m	EDR Copper Cable QSFP28	EB50
1m	EDR Copper Cable QSFP28	EB51
2m	EDR Copper Cable QSFP28	EB52
1.5m	EDR Copper Cabling QSFP28	EB54
3m	EDR Optical Cable QSFP28	EB5A
5m	EDR Optical Cable QSFP28	EB5B
10m	EDR Optical Cable QSFP28	EB5C
15m	EDR Optical Cable QSFP28	EB5D
20m	EDR Optical Cable QSFP28	EB5E
30m	EDR Optical Cable QSFP28	EB5F
50m	EDR Optical Cable QSFP28	EB5G
100m	EDR Optical Cable QSFP28	EB5H



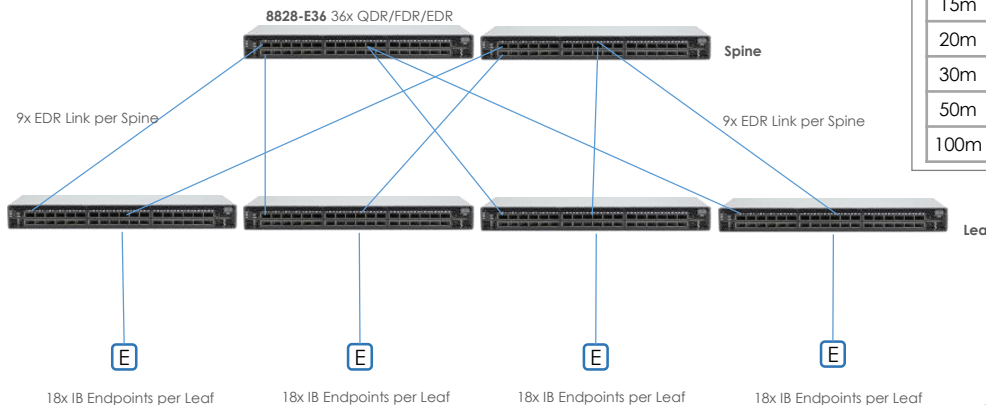
-  1GbE xCAT Network
-  1GbE HMC Network
-  EDR Data Network
-  FDR Data Network
-  QDR/FDR Client
-  QDR/FDR/EDR Client
-  QDR/FDR/EDR Customer Network








Speed	Switch	Cabling	Adapter	Optics
FDR / EDR	SB7700 – 8828-E36 	See list on right	EC3E EC3T	NA

Choice of Cabling

EDR Cabling		
Length	Description	FC
0.5m	EDR Copper Cable QSFP28	EB50
1m	EDR Copper Cable QSFP28	EB51
2m	EDR Copper Cable QSFP28	EB52
1.5m	EDR Copper Cabling QSFP28	EB54
3m	EDR Optical Cable QSFP28	EB5A
5m	EDR Optical Cable QSFP28	EB5B
10m	EDR Optical Cable QSFP28	EB5C
15m	EDR Optical Cable QSFP28	EB5D
20m	EDR Optical Cable QSFP28	EB5E
30m	EDR Optical Cable QSFP28	EB5F
50m	EDR Optical Cable QSFP28	EB5G
100m	EDR Optical Cable QSFP28	EB5H

Sample 72 Node Cluster



-  1GbE xCAT Network
-  1GbE HMC Network
-  EDR Data Network
-  FDR Data Network
-  QDR/FDR Client
-  QDR/FDR/EDR Client
-  QDR/FDR/EDR Customer Network

# Links to Spine	# Spines	#Leafs	#Ports
1	18	36	648
2	9	18	324
3	6	12	216
6	3	6	108
9	2	4	72

Some Rules:

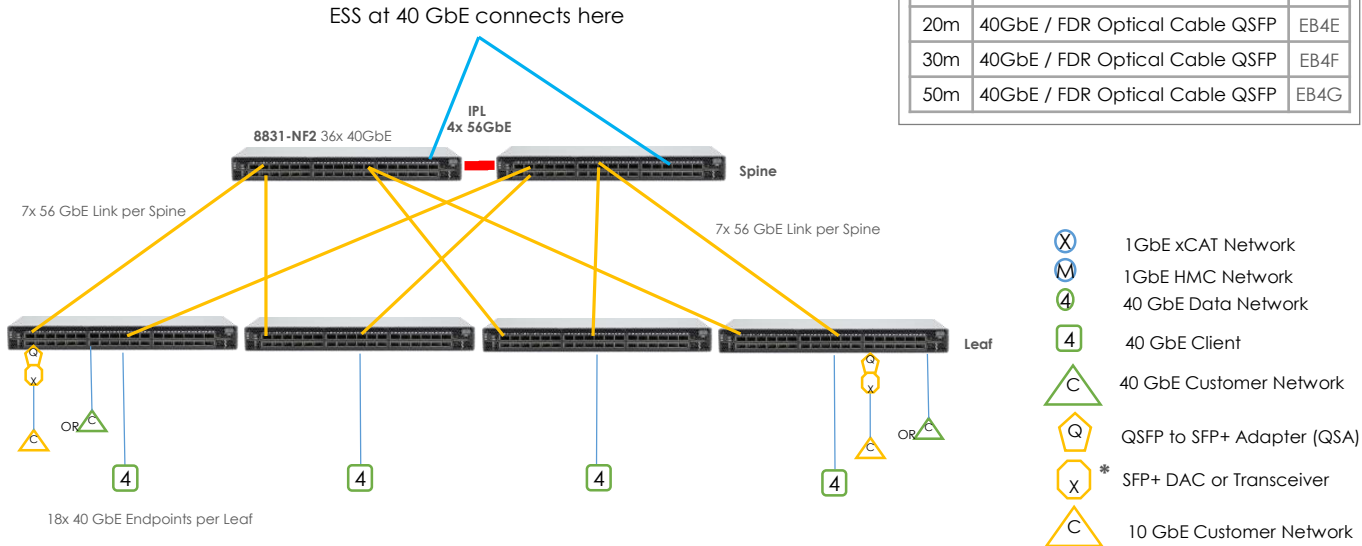
- Links from Leaf to Spine must be Modulo of 18 - 1,2,3,6,9
- Non-Blocking requires as many links down to servers from Leaf as up to Spine from Leaf
- Biggest two tier network is 648 Nodes, 18 Spines & 36 Leafs
- Think ahead. Add Spines at day 1 for expansion, so extra leafs can be added without re-cabling existing leafs

Speed	Switch	Cabling	Adapter	Optics
40 GbE	SX1710 – 8831-NF2 	See list on right	EC3A	EB27 + EB2J or EB2K

40GbE / FDR Cabling



Length	Description	FC
0.5m	40GbE / FDR Copper Cable QSFP	EB40
1m	40GbE / FDR Copper Cable QSFP	EB41
2m	40GbE / FDR Copper Cable QSFP	EB42
3m	40GbE / FDR Optical Cable QSFP	EB4A
5m	40GbE / FDR Optical Cable QSFP	EB4B
10m	40GbE / FDR Optical Cable QSFP	EB4C
15m	40GbE / FDR Optical Cable QSFP	EB4D
20m	40GbE / FDR Optical Cable QSFP	EB4E
30m	40GbE / FDR Optical Cable QSFP	EB4F
50m	40GbE / FDR Optical Cable QSFP	EB4G

Sample 72 Node L2 (VMS) Cluster



* 10GbE & Optics are IBM Parts

IBM Mellanox Infrastructure for 10 GbE Cluster ESS/Spectrum Scale Choice of Cabling

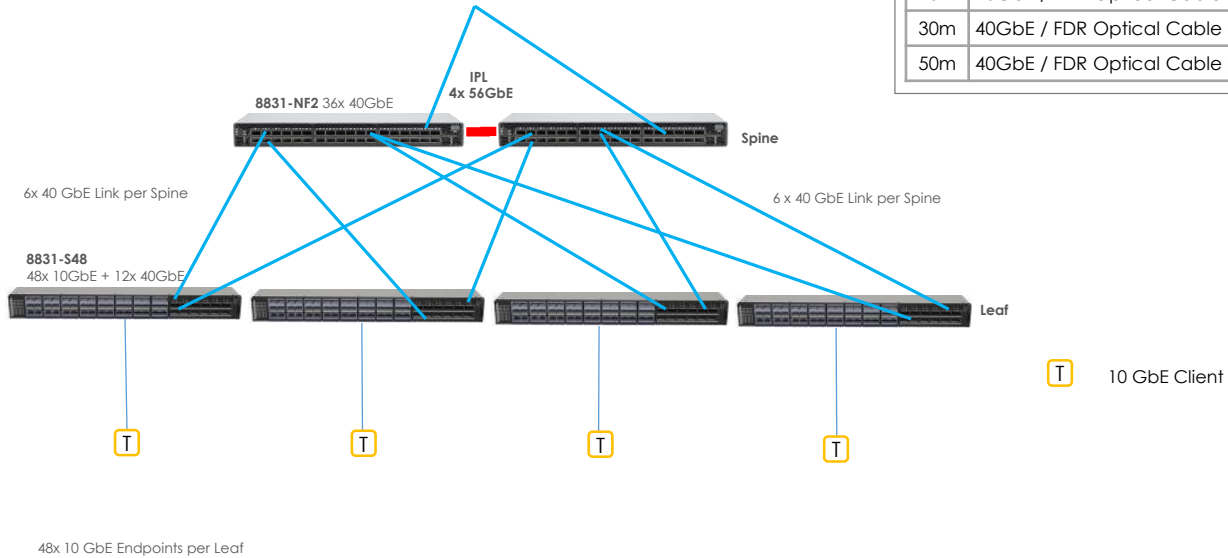
Speed	Switch	Cabling	Adapter	Optics
40 GbE	SX1710 – 8831-NF2 	See list on right	EC3A	EB27 + EB2J or EB2K
10/40 GbE	SX1410 – 8831-S48 		EC37 / EL3X EC2M / EL40 (SR)	EB28 + ECBD or ECBE

40GbE / FDR Cabling

Length	Description	FC
0.5m	40GbE / FDR Copper Cable QSFP	EB40
1m	40GbE / FDR Copper Cable QSFP	EB41
2m	40GbE / FDR Copper Cable QSFP	EB42
3m	40GbE / FDR Optical Cable QSFP	EB4A
5m	40GbE / FDR Optical Cable QSFP	EB4B
10m	40GbE / FDR Optical Cable QSFP	EB4C
15m	40GbE / FDR Optical Cable QSFP	EB4D
20m	40GbE / FDR Optical Cable QSFP	EB4E
30m	40GbE / FDR Optical Cable QSFP	EB4F
50m	40GbE / FDR Optical Cable QSFP	EB4G

Sample 192 Node L2 (VMS) Cluster

ESS at 40 GbE connects here



Some Guidelines

Bandwidth

10GbE = ~800MB/s – Both ports can be used and will provide ~1.6GB per Card

40GbE = ~ 3.2GB/s – Both ports can be used and will provide ~4.3GB per Card, Max Bandwidth per adapter is ~52Gb/4.3GB

FDR (56Gb) = ~5.6GB/s – Both ports can be used and will provide ~11.2GB per Card, More cards will provide redundancy and more access

to disk up to maximum disk i/o capacity

EDR (100Gb) = ~10GB/s – 2 Ports per Node will exceed Disk i/o capacity, third provides redundancy

One port active per Card – Max Bandwidth per adapter is 100Gb/10GB

Load Balancing TCP/IP Aggregated Links

Recommended number of adapters in aggregation for IP traffic = TWO (2)

Reason as there are only two IP addresses on the ESS, hashing algorithms will only hash to max of two adapters.

If more than 800MB/s per client / server is required for R/W then use 40GbE instead of 10GbE.

Load Balancing in Infiniband

Infiniband has no issue and can Hash across all ports.

RoCE

Today supports only one active port. (Work under way to make Active/Active on one card available)

@10GbE will see CPU SYS time reduction, so more application CPU available

@40GbE will see CPU SYS time reduction and throughput improvement

Storage Connection

To ensure that all nodes have same distance to/from storage the best recommendation is a storage and management switch with no nodes on it.

Best option storage only option is 12 Port 40GbE/56GbE/FDR or 16 Port 10/25/40/50/100 GbE with

- 7x 40Gbe Ports down and 5 56GbE ports up.
- 6x FDR ports up and down, with EMS connected to other switches
- 8x 100GbE ports up and down.
- This today can be sourced via Mellanox BP and Distributors.

The available e-Config option is use of NF2 / F36 / E36 switches with only management nodes and storage nodes on them.