Storage Decisions
New York 2015
November 3-4th

New Storage Interfaces for Your Next Datacenter Build Out

Dennis Martin ◈ **Demartek**®

1

Agenda

- ◆ About Demartek

- ◆ Increased Bandwidth Needs for Storage

- ◆ Storage Interface Technology & Futures
  - − Ethernet, Fibre Channel, SAS, SATA, Thunderbolt, USB, NVMe

- ◆ Demartek Free Resources

Demartek Video

Demartek SERVICES

Click to view this one minute video
(available in 720p and 1080p)

Demartek YouTube Channel:

http://www.youtube.com/user/Demartek/videos

Demartek

About Demartek

◆ Industry Analysis and ISO 17025 accredited test lab

◆ Lab includes enterprise servers, networking & storage (DAS, NAS, SAN, 10GbE, 40GbE, 16GFC)

◆ We prefer to run real-world applications to test servers and storage solutions (databases, Hadoop, etc.)

◆ Demartek is an EPA-recognized test lab for *ENERGY STAR Data Center Storage* testing

◆ Website: www.demartek.com/TestLab

SNIA Emerald™
RECOGNIZED TESTER

Demartek®

The Need For More Bandwidth
► Server and Application Growth

Physical Server

Virtual Machines

- **Server Virtualization**
  - How many VMs per physical server do you deploy?
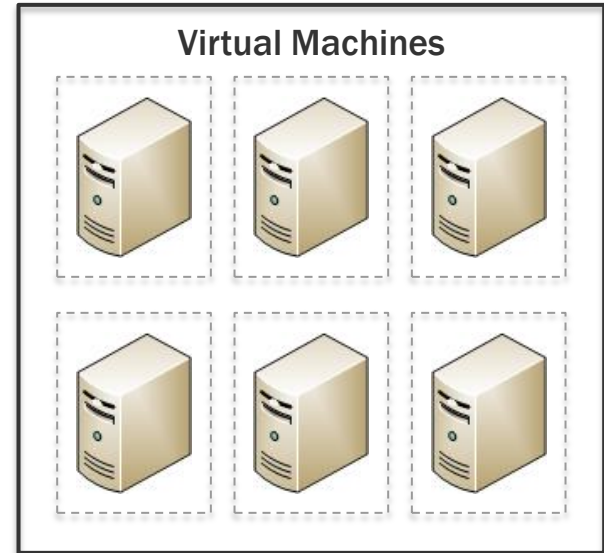 - Compare the number of VMs today vs. one and two years ago

- **Application Growth**
  - Applications processing more data today

- Bootstorm test with 90 VMs in one physical server
  www.demartek.com/Demartek_Analysis_of_VDI_Storage_Performance_during_Bootstorm.html

The Need For More Bandwidth
► Server and Application Growth

- ◆ **New Generations of Servers**
  - – PCI Express 3.0 since 2012
    - ◆ Up to 40 PCIe lanes per processor (Intel Xeon E5 series)
  - – New servers support 10GbE on the motherboard
  - – More cores per processor
  - – Larger memory support (up to 1.5TB/processor)

- ◆ **SSD**
 - – Are you deploying enterprise SSDs today?

PCI-Express

- Measured in gigatransfers/second (GT/s)
  - Bandwidth specified by indicating number of lanes such as "x1", "x2", etc., and generally spoken as "by 1", "by 2", etc.

PCIe Bus	GT/s	Encoding	x1	x2	x4	x8	x16
PCIe 1.x	2.5	8b/10b	250 MB/s	500 MB/s	1 GB/s	2 GB/s	4 GB/s
PCIe 2.x	5	8b/10b	500 MB/s	1 GB/s	2 GB/s	4 GB/s	8 GB/s
PCIe 3.x	8	128b/130b	1 GB/s	2 GB/s	4 GB/s	8 GB/s	16 GB/s

- **PCIe 4.0** – In November 2011, the PCI-SIG announced the approval of 16 GT/s as the bit rate for PCIe 4.0.
  - PCIe 4.0 specification Rev 0.7 targeted for year-end 2015*
  - PCIe 4.0 specification Rev 0.9 targeted for 2H 2016* （* Source: PCI-SIG)
  - Products probably available in 2017 or 2018

Ethernet
► 1GbE and 10GbE
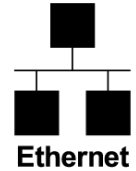
- **1GbE**
  - Not unusual to have 4, 6 or 8 NIC ports in a server (many cables)
  - Can be quad-port, dual-port or single-port
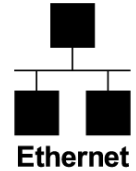
- **10GbE**
  - A dual-port 10GbE NIC provides bandwidth and failover
  - Good choice for 1U servers that have few I/O slots
  - Slot requirements
    - ◆ Quad-port 10GbE NIC – PCIe 3.0 x8
    - ◆ Dual-port 10GbE NIC – PCIe 3.0 x4 or PCIe 2.0 x8
  - Adoption: blade servers yes, rack servers not so much

- **Price drops: 10GBASE-SR SFP 2013: $165   2015: $75**

Ethernet
► 40GbE and 100GbE

- IEEE 802.3ba (40GbE & 100GbE) ratified June 2010

- The fastest Ethernet cables and connectors today are *10 Gbps* per lane or channel

- Higher speeds today are achieved by bundling
  - 40GbE today = 4 x 10 Gbps together
  - 100GbE today = 10 x 10 Gbps together

- 40 Gbps NICs require PCIe 3.0 x8 or x16 slot in the server
  - A PCIe 3.0 x8 slot cannot sustain two ports of 40GbE at full line rate
 - A PCIe 3.0 x16 slot can sustain two ports of 40GbE at full line rate

Ethernet
▶ 25GbE, 50GbE and 100GbE
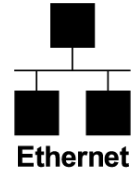
- ◆ **25 Gbps connectors will soon be available**
  - – 25Gbps over a single lane ("SFP28")
  - – 50GbE future = 2 x 25 Gbps together
  - – 100GbE future = 4 x 25 Gbps together ("QSFP100")
- ◆ **25G Ethernet Consortium Announcement – July 1, 2014**
  - – www.25GEthernet.org
- ◆ **IEEE P802.3by 25 Gb/s Ethernet Task Force**
  - – http://www.ieee802.org/3/by/index.html - standard by 2H 2016
- ◆ **25GbE end-user products expected to be available in 2016**

Ethernet
► Public Roadmap
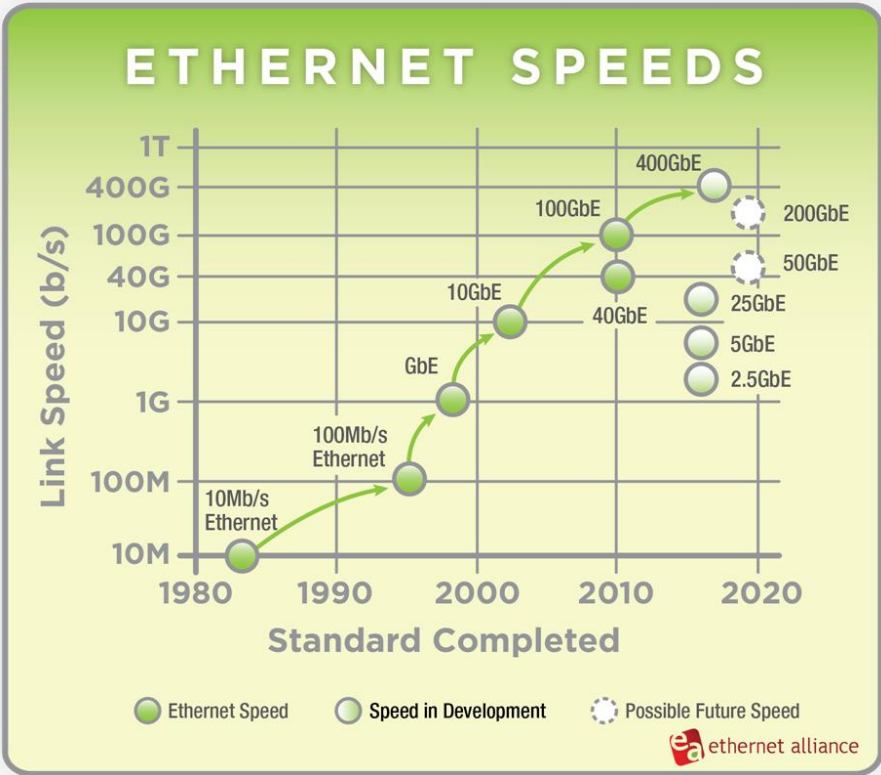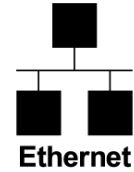
Development of four new speeds began in 2014:
2.5GbE, 5GbE, 25GbE, 400GbE

http://www.ethernetalliance.org/roadmap/

Why 2.5GbE and 5GbE?

Ethernet

▶ 1GbE, 2.5GbE and 5GbE – "Multigigabit"

Ethernet
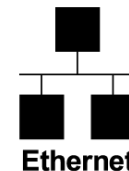
WiFi	802.11n	802.11ac Wave 1	802.11ac Wave 2
Band	2.4 GHz & 5 GHz	5 GHz	5 GHz
PHY Rate*	Up to 600 Mbps	Up to 1.3 Gbps	Up to 6.9 Gbps

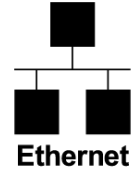* Effective data rates are less than the PHY rate

The new "multigigabit" switches can handle the 802.11ac Wave 2 speeds and are expected to be a low-cost way to provide network speeds > 1 Gbps with existing cabling.

Cable Type	1 Gbps	2.5 Gbps	5 Gbps	10 Gbps
Cat 5e	✔	✔	✔	✘
Cat 6	✔	✔	✔	✔ (55m)
Cat 6a	✔	✔	✔	✔

Some pre-standard "multigigabit" switches were introduced in 2015

Demartek

Ethernet
▶ Future

Short-term (next year or two)

- ◆ Next set of data center switch ASICs will support:
  - – 32 x 100 GbE, 64 x 40/50 GbE and/or 128 x 25 GbE
- ◆ Some switch vendors are already shipping 32 x 100 GbE
  - – These use the new "QSFP28" or "QSFP100" technology

Long-Term (several years)

- ◆ 40GbE and 50GbE will share common transceivers, pricing
- ◆ Single-lane speeds will increase to 50Gbps
  - – Speed doubling will follow: 100, 200, 400, 800, 1600 Gbps

Fibre Channel
► History

◆ **Fibre Channel deployed in 90% of Fortune 1000 data centers**

– Purpose-built for data center storage

– Meets enterprise-class performance and reliability requirements

– Proven technology

– High-bandwidth

◆ **History**

– Introduced in 1997

– Doubled in speed every 3 or 4 years

Fibre Channel Speed History

Source: Fibre Channel Industry Association (FCIA)
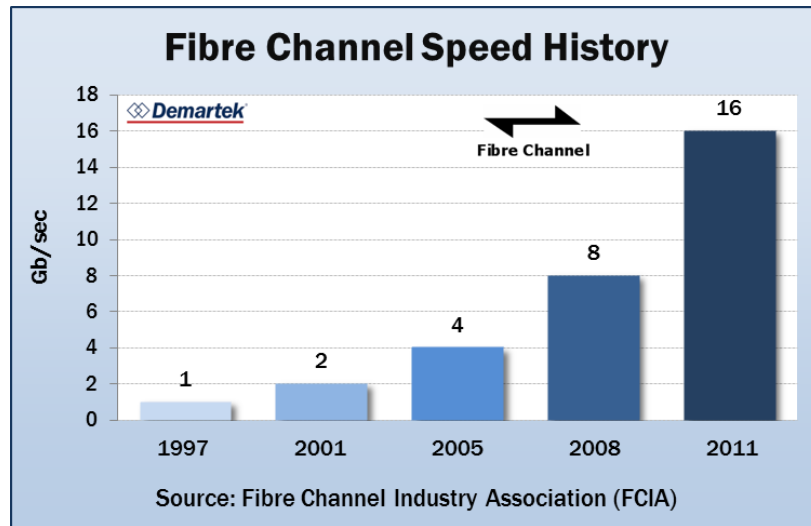
Fibre Channel
▶ 16 Gigabit (16GFC)

Fibre Channel

◆ **16GFC is backward compatible with 4GFC & 8GFC**

◆ **Uses 14 Gbps single-lane connectors**

    – Doubles speed of 8GFC due to newer 64b/66b encoding

◆ **First 16GFC switches and HBAs shipped in 2011**

    – Some of these HBAs can function as 10 Gb NICs

◆ **FC speeds and server slot requirements (dual-port)**

    – 4 Gb: PCI-X 2.0, PCIe 1.0

    – 8 Gb: PCIe 2.0 x4 or PCIe 1.0 x8

    – 16 Gb: PCIe 3.0 x4 or PCIe 2.0 x8   (quad-port needs PCIe 3.0 x8)

Demartek

Fibre Channel
▶ 32 Gigabit and 128 Gigabit ("Generation 6")

Fibre Channel

- In February 2014, "Gen 6" Fibre Channel was announced

- 32 Gbps single-lane connection ("32GFC")
  - OM4 fiber-optic cable expected distance: 100m
  - Uses same SFP28 transceiver technology as 25GbE

- 128 Gbps parallel connection (4 x 32, "128GFCp")
  - Initially used for switch-to-switch connections

- Backward compatible with 16GFC and 8GFC

- 32GFC products expected to be available in 2016

Fibre Channel
► New Features

- **Energy Efficiency**
  - Power at transceiver is reduced when not in use ("dimmer switch")

- **Forward Error Correction (FEC)**
  - Additional error checking required at speeds of 16 Gbps and higher

- **New Diagnostics**

- **Fabric-assigned WWNs (Brocade switches)**
  - Allows WWNs to be created and assigned in advance in the fabric rather than discovered in the adapters

- **New "Enhanced Gen 5" adapters (32G features, 16G optics)**

Fibre Channel

Converged Networks

Fibre Channel over Ethernet

- ◆ **Combined LAN and SAN networks**
  - – Lossless features of Fibre Channel with ubiquity of Ethernet
- ◆ **Data Center Bridging (DCB)**
  - – Enhanced Ethernet to support FC storage traffic and more
  - – Required for RDMA over Converged Ethernet (RoCE)
- ◆ **FCoE – Fibre Channel over Ethernet**
  - – First major application for DCB – runs FC at 10 Gbps
- ◆ **CNA – Converged Network Adapter**
  - – Supports 10 Gb Ethernet and 10 Gb FCoE at the same time on the same cable

Demartek®

SAS
► Serial Attached SCSI
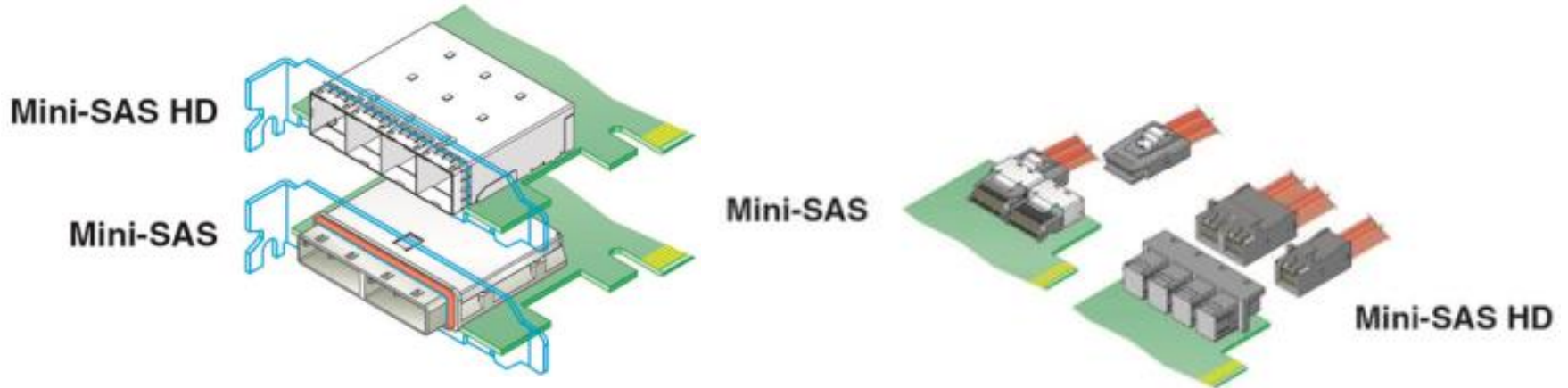
- Used for enterprise SSDs and HDDs and for connections to JBOD shelves and external array shelves

- Current specification is 12 Gbps SAS

- SCSI command protocol used in FC, FCoE, iSCSI and SAS

- Supports up to 16K devices on single "fabric"

- Roadmap for 24 Gbps SAS, at or near the same time as PCIe 4.0, approximately 2017

SAS

▶ Mini-SAS HD Connectors

- Cables can be copper or fiber-optic

- See larger versions of these diagrams and information for other storage interfaces on the Demartek Storage Interface Comparison page: http://www.demartek.com/Demartek_Interface_Comparison.html

SATA
▶ Serial ATA

- **Device (drive) types**
  - Common for client (consumer) SSDs & HDDs
  - Sometimes used for enterprise SSDs & HDDs
- **Mostly for inside the case connections**
  - eSATA allows for short external distances
- **SATA is point-to-point, single device per cable or connector**
- **Traditional SATA has no roadmap beyond 6 Gbps**
  - Some new enterprise features planned
  - Unclear if SATA Express will gain acceptance in the market

Thunderbolt 3

- **Thunderbolt 3 announced in June 2015**
  - Increases speed to 40 Gbps (active copper & optical cables)

- **Target audience is media creators and editors who use premium laptops, desktops, workstations and peripherals that connect to them.**
  - Devices include video cameras and storage devices, especially SSDs

- **Uses a USB type-C cable**
  - Supports USB 3.1 (10 Gbps), Display Port (dual 4K displays), four lanes of PCI Express 3.0
  - Provides 15 watts for bus-powered devices and supports USB power delivery to charge laptop computers up to 100 watts.
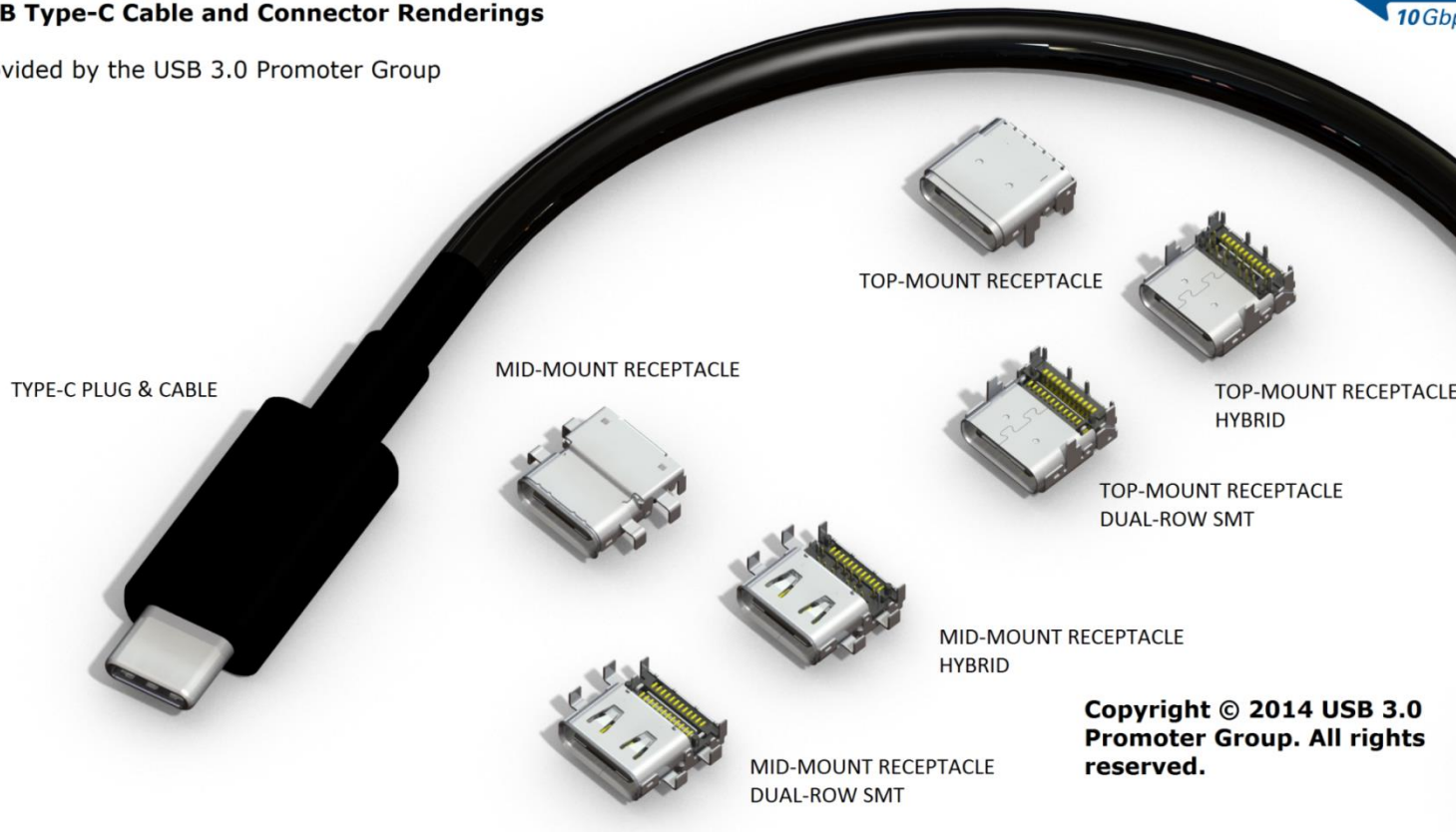
- **Bridge devices to 10GbE**

USB 3.1

- **USB 3.1 specification completed July 2013**
  - Doubles speed to 10 Gbps (USB 3.0 is rated for 5 Gbps)
  - Works with existing USB 3.0 and 2.0 products

- **USB 3.1 Power Delivery**
  - Can deliver up to <u>100 watts</u>, bi-directionally
  - Can deliver audio/video, data and power concurrently

- **Media Agnostic USB protocol (USB over WiFi)**
  - Allows wireless devices and docking stations to communicate using the USB protocol

- **New USB Type-C bi-directional connector**
  - Similar in size to existing USB 2.0 micro-B

USB Type-C Cable and Connector Renderings

Provided by the USB 3.0 Promoter Group

TYPE-C PLUG & CABLE

MID-MOUNT RECEPTACLE

TOP-MOUNT RECEPTACLE

TOP-MOUNT RECEPTACLE
HYBRID

TOP-MOUNT RECEPTACLE
DUAL-ROW SMT

MID-MOUNT RECEPTACLE
HYBRID

MID-MOUNT RECEPTACLE
DUAL-ROW SMT

Copyright © 2014 USB 3.0
Promoter Group. All rights
reserved.

NVM-Express (NVMe)

- Scalable host controller interface designed for enterprise and client systems that use PCI Express SSDs
  - Designed with Flash memory and technologies coming after Flash memory in mind (non-volatile memory)
  - Much faster (lower latency) software stack than existing storage stacks such as SAS and SATA
- In-box drivers for Windows and Linux now
- Faster individual devices than other interfaces
  - PCIe card and drive form factor (SFF-8639 → U.2)
 - Not as well-established, but ramping up quickly

NVM-Express (NVMe)

- Demartek test experience with NVMe
  - Some of our recent Ethernet storage testing with NVMe required 40GbE – 10GbE was too slow
  - We've seen 3+ GB/sec (yes, GigaBytes/sec) from a single NVMe SSD with a real-world database workload

- I shared NVMe performance results in my Storage Decisions presentation yesterday, Nov. 3

- Additional comments and explanation:
  http://www.demartek.com/FMS2015

NVM-Express (NVMe)

- PCI Express (PCIe) projected to be the leading enterprise SSD interface by 2018

- First NVMe products were enterprise SSDs

- Client (consumer) SSDs began shipping in 2015

Enterprise SSD by Interface

Source: IDC

- NVMe over Fabrics development underway. Goal is to run NVMe over network of choice within ~10 μs latency of local.
  - NVMe works well with RDMA
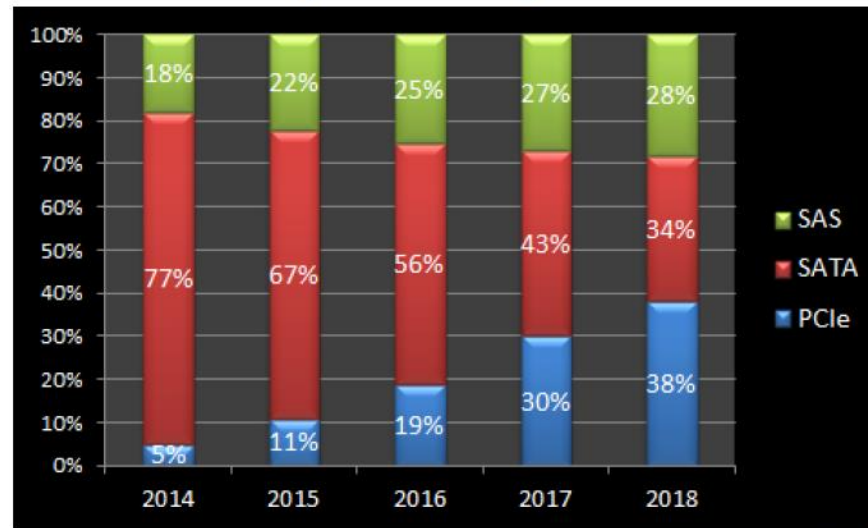
Storage Interface Comparison Reference Page

◆ Demartek Storage Interface Comparison page

  – Search engine: *Storage Interface Comparison*

  – Includes new interfaces such as 25GbE, 32GFC, Thunderbolt 3

  – HTML version and downloadable interactive PDF

  – Periodic updates

http://www.demartek.com/Demartek_Interface_Comparison.html

Demartek Free Resources

- Demartek SSD Zone
  www.demartek.com/SSD

- Demartek iSCSI Zone
  www.demartek.com/iSCSI

- Demartek FC Zone – www.demartek.com/FC

- Demartek SSD Deployment Guide
  www.demartek.com/Demartek_SSD_Deployment_Guide.html

- Demartek commentary: "Horses, Buggies and SSDs"
  www.demartek.com/Demartek_Horses_Buggies_SSDs_Commentary.html

- Demartek Video Library - http://www.demartek.com/Demartek_Video_Library.html

Performance reports, Deployment Guides and commentary available for free download.

Thank You!

Demartek public projects and materials are announced on a variety of social media outlets. Follow us on any of the above.

Sign-up for the Demartek monthly newsletter, *Demartek Lab Notes*.
www.demartek.com/newsletter