BROCADE

25GBE SERVER TO SWITCH ARCHITECTURES

Scott Kipp

skipp@brocade.com

September 2014

kipp_25GE_01_0914a.pdf

Supporters

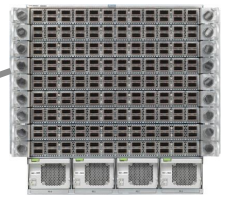
- Jonathan King – Finisar
- Andy Moorwood – Infinera
- Steve Swanson – Corning
- John Abbott – Corning
- Gary J. Bernstein – Leviton
- Paul Kolesar – Commscope
- Jack Jewell – Green VCSEL
- Doug Coleman – Corning
- Rick Pimpinella – Panduit
- Nathan Tracy – TE Connectivity
- Megha Shanbhag – TE Connectivity
- Robert Lingle Jr. – OFS
- Scott Sommers – Molex
- Dave Lewis – JDSU
- Chris Cole – Finisar
- Dave Chalupsky – Intel
- John D'Ambrosia - Dell
- Rick Rabinovich – Alcatel Lucent
- Tom Palkert - Molex

Server to Switch Architectures

- Server to switch architectures/deployments depend on 5 main characteristics
  1. Application(s)
  2. Server
  3. Cabling
  4. Switch
  5. IT Departments
- With an amazing variety within each of these categories, an amazing variety of solutions results – like the Cambrian Period of life on Earth
- The 25GbE standard should support the widest variety of deployments as possible so an optical PMD is needed

Five Components of Switch To Server Architectures

	Low End	Mid Range	High End
Application	Web Server	Enterprise Apps – payroll, HR, CAD	Financial Transactions
Server	1U Server 	Multi-U Server 	Mainframe
Cabling	Simple Patchcord	AOC or Simple Cabling	Structured Cabling
Switch	1U or less Switch 	Multi-U Switch 	Modular Switch
IT Staff	1 Guy or Gal	Small Team	Large Teams



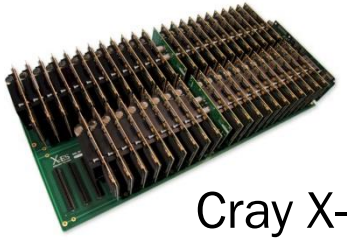
Application Variety

Applications drive bandwidth

- Single applications
  - Low bandwidth – few users or small data rates
  - High bandwidth – many users or high data rates
- Multiple applications via virtualization
- Massive web applications with millions or billions of users
- File Server – NAS
- Financial transactions

Server Designs

- Microservers – ARM Servers
- Blade Servers
- 1/2U Servers
- 1U Servers
- 2U Servers
- 4-12U Servers
- Rack and multi-rack Servers



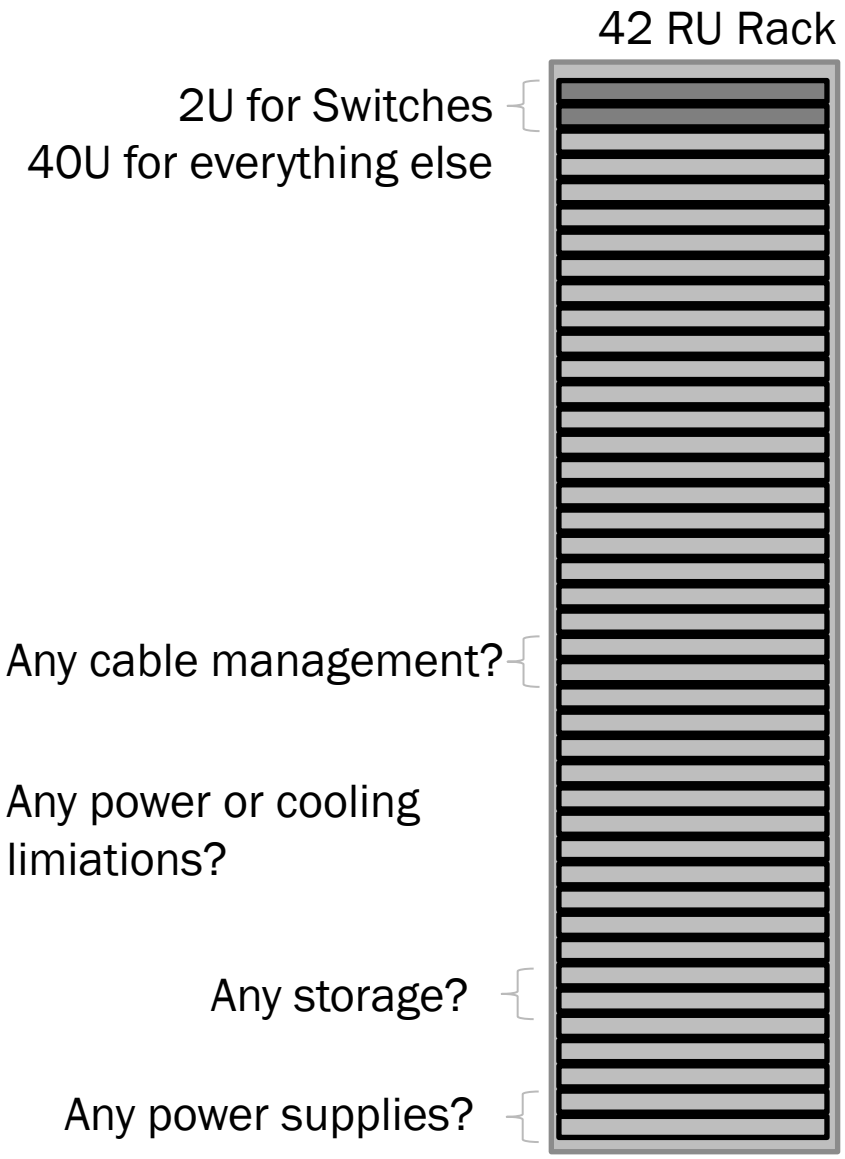
Cray X-ES Microserver

8U Storage Server

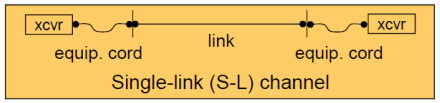
Mainframe

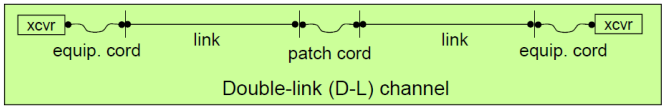
Rack Space

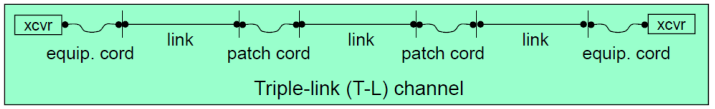
	Max Servers / 40RU
Micro-Server	>100
Blade Server	>100
1/2U Server	80
1U Server	40
2U Server	20
4U Server	10
8U Server	5
12U Server	3
Mainframe	<1



Cabling Variety



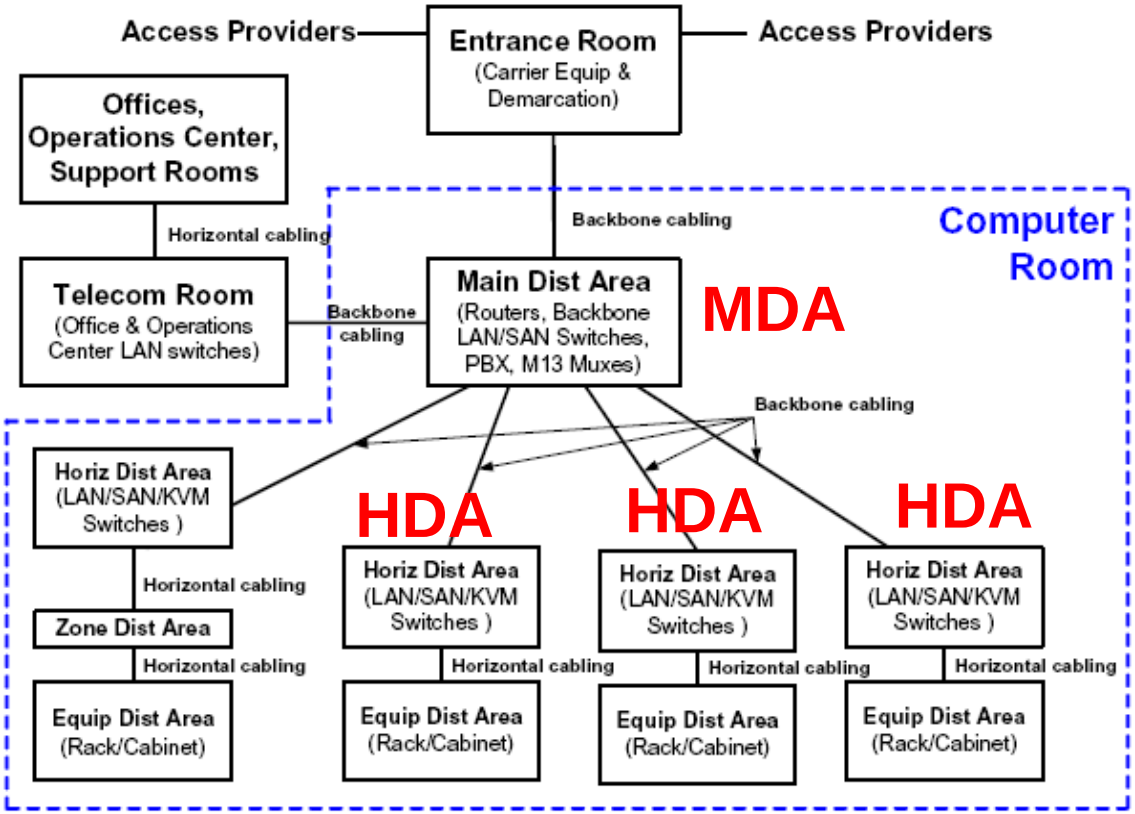




- From lowest cost, complexity and scalability
  - Patchcord
  - Direct Attach Cabling
    - AOC
  - Structured Cabling
    - Single-Link Channel
    - Double-Link Channel
    - Triple-Link Channel
    - Inter-building Channel

TIA-942 – Data Center Cabling and Design

- TIA-942 - Telecommunications Infrastructure for Data Centers defines:
- MDA (Main Distribution Area) that fans out to
- HDAs (Horizontal Distribution Areas)



25GbE Switches

25GbE will be like 10GbE today

- Switch ASICs are increasing speed from 10GbE to 25GbE and more than doubling the port counts from 64 ports to 128+ ports

64 Ports of 10G

64 10GbE port ASIC enables
48 SFP+ and 4 QSFP+
640Gb/s of Throughput

128+ Ports of 25G

128 25GbE port ASIC enables
32 QSFP+
3.2 Tb/s of Throughput

10GbE Switch Designs

- Blade Switches
- 1/2U Switches
- 1U Switches
- 2U Switches
- 4-12U Modular Switches

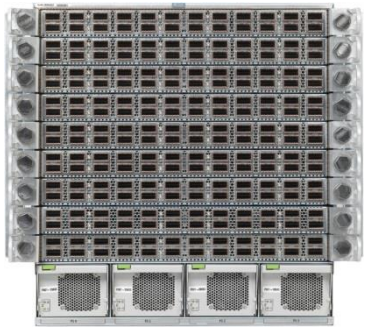
4 SFP+

12 QSFP+ = 48 25GbE

64 SFP+

36 QSFP+
= 144 25GbE

96 SFP+



216 QSFP+
= 864 25GbE

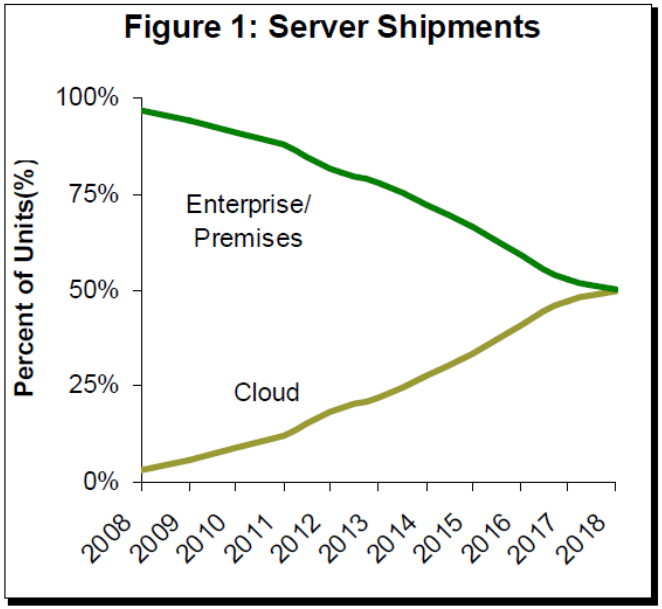
IT Departments

- One Person Shops
- Small teams
  - Applications
  - Servers
  - Storage
  - WAN
  - Cabling
  - Power
  - Cooling
  - Consultants
  - Contractors
  - …
- Large Teams/Departments

From here
To there…

25GbE Deployments

- First deployments in the cloud data center, but they are still the smaller segment of the market
  - ToR done by many
- Second deployments in enterprises, and they will gladly do it for the right applications if it saves money and time
  - Wide variety of deployment models



Enterprise and Premises dominate server shipments through forecast – Revenue is even more lopsided

Source: Dell'Oro Controller and Adapter Forecast Summary – July 2014

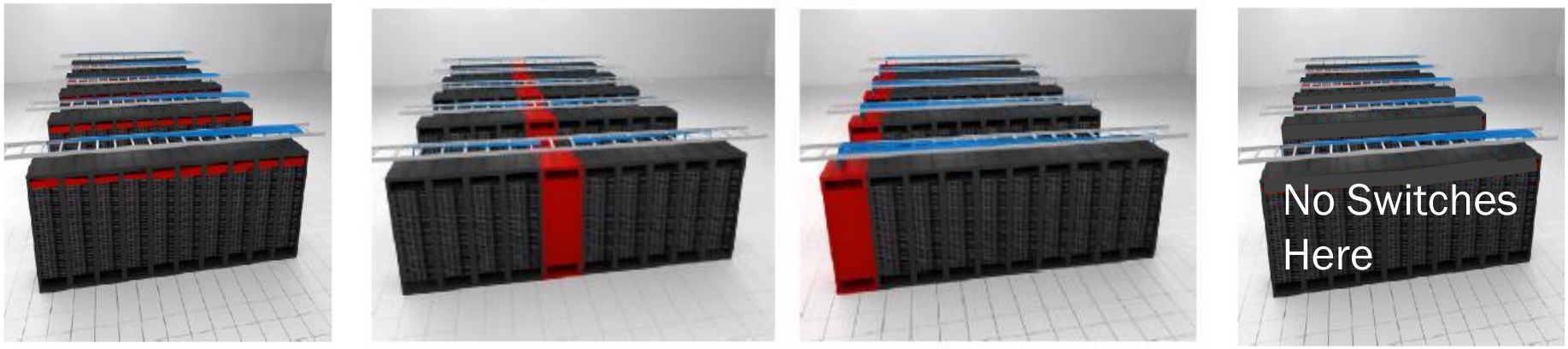
ToR, MoR, EoR

ToR
(3 to 5 m)

MoR
(3 to 15 m)

EoR
(3 to 50 m)

Home Run
(50 to 100+ m)



Addressed by
25Gb/s CFI

Not addressed by 25Gb/s CFI

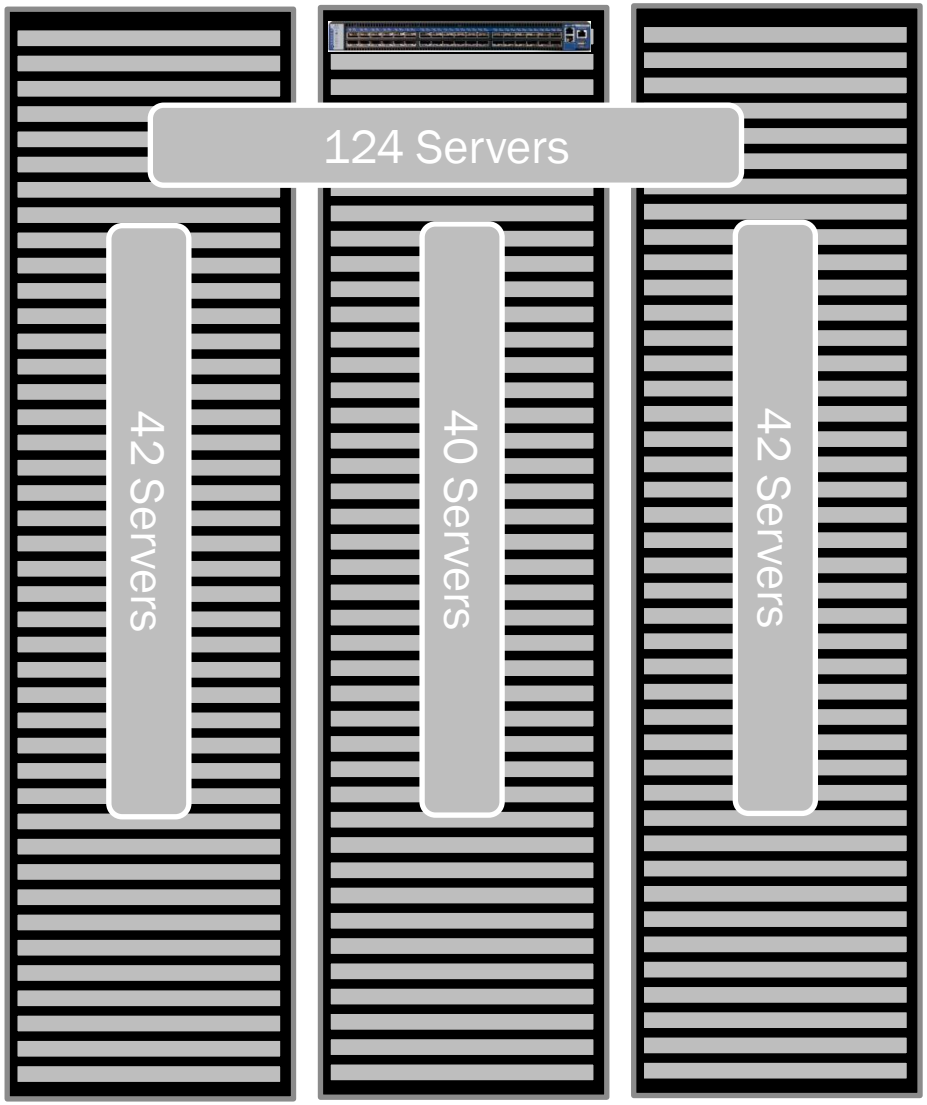
Source of graphics: jimenez_3bq_01_0711

1U Server ToR Designs

5m is Fine

36 QSFP+
= 144 25GbE

42 RU Racks



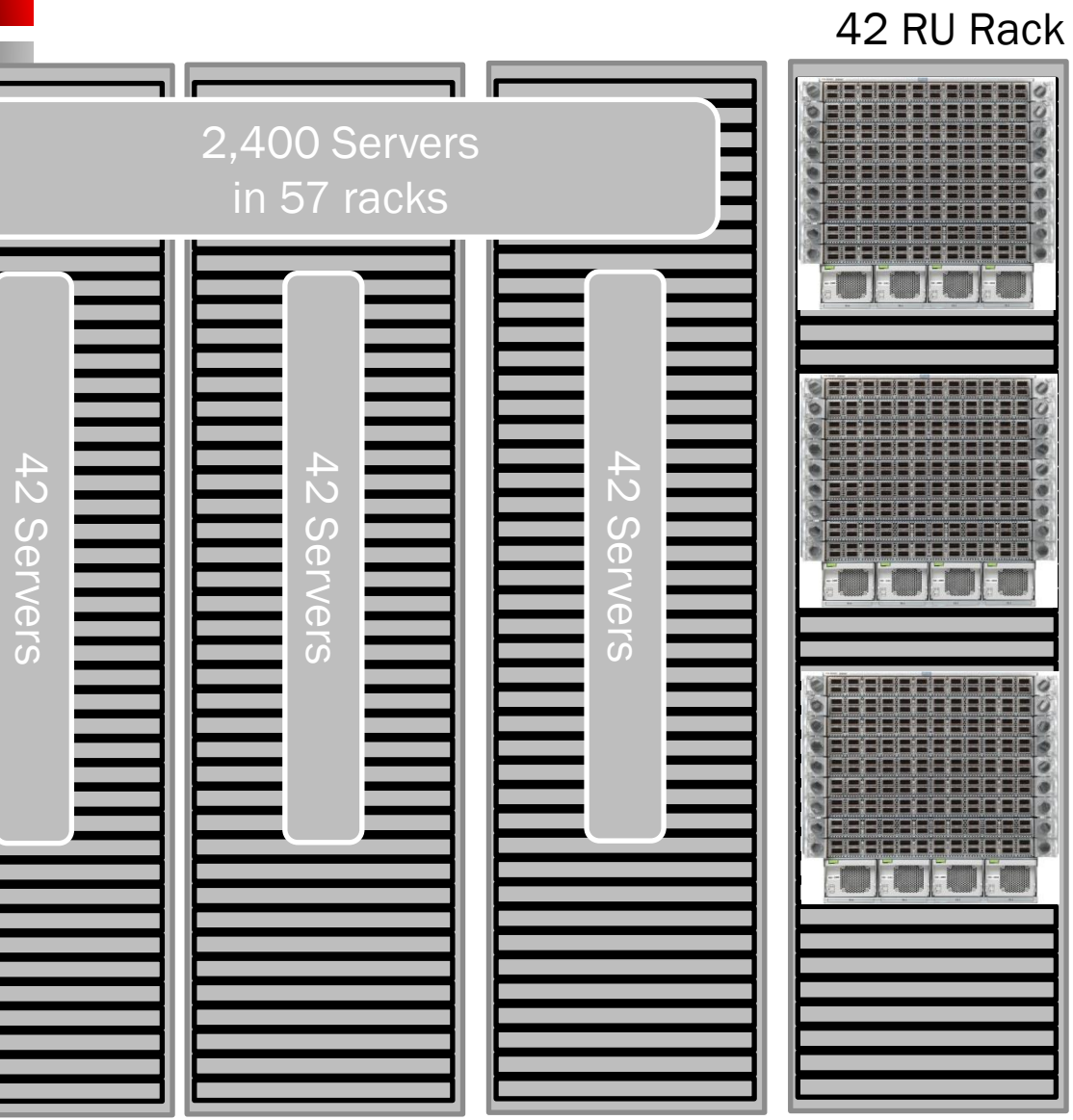
3m for Intra-rack
5m for Adjacent racks

20 ports (500G) left for uplinks…

One ToR switch can support multiple racks of 1U servers

If the servers are 2U and not fully filled, the switch can support many more racks and 5m is not enough

1U Server EoR Designs



216 QSFP+
= 864 25GbE/Switch

Each switch could support 800 servers with 64 ports (1.6Tb/s) of uplinks

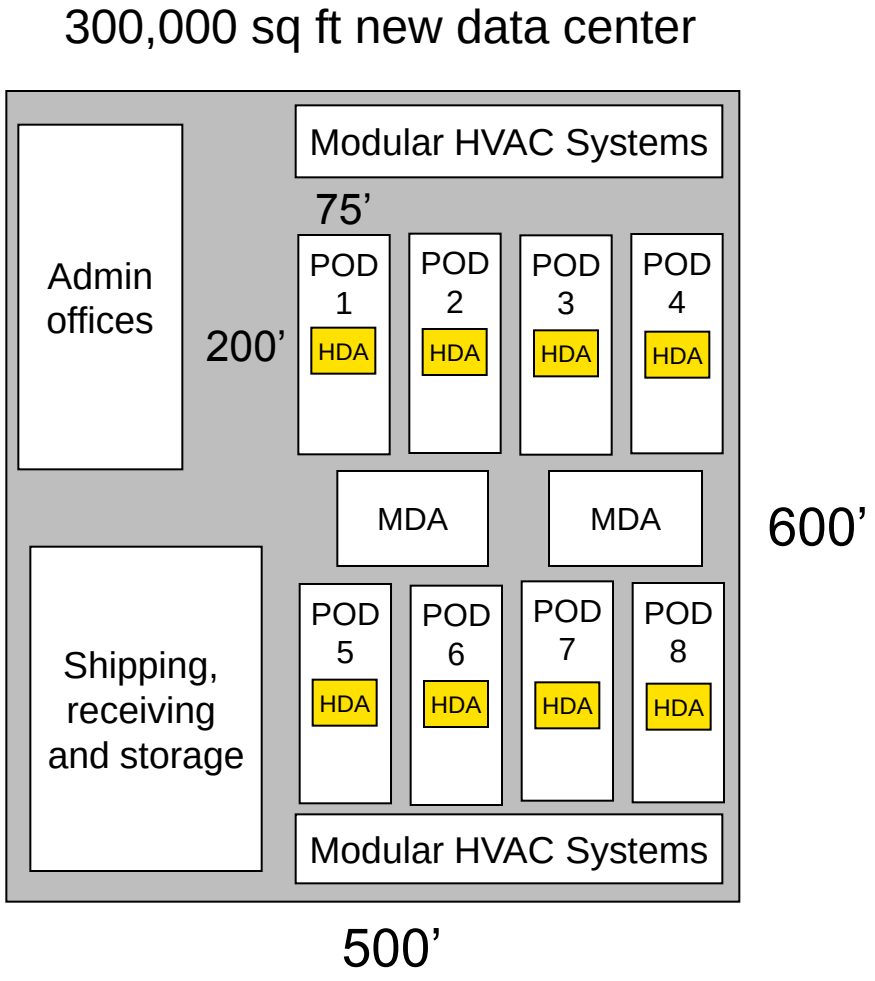
This rack of modular switches would support 2,400 servers

Home Run

- A home run is a server connect architecture where a server is connected straight into the core of the network
  - Common for storage servers or NAS that shares massive files – feedback to Ethernet Alliance Bird of Feather at HPC'13
  - Mainframes and large enterprise servers may connect straight into the core
  - These links need high speed
- Used when servers, storage and switches are consolidated into different areas
- Usually associated with structured cabling where every port is represented in the MDA

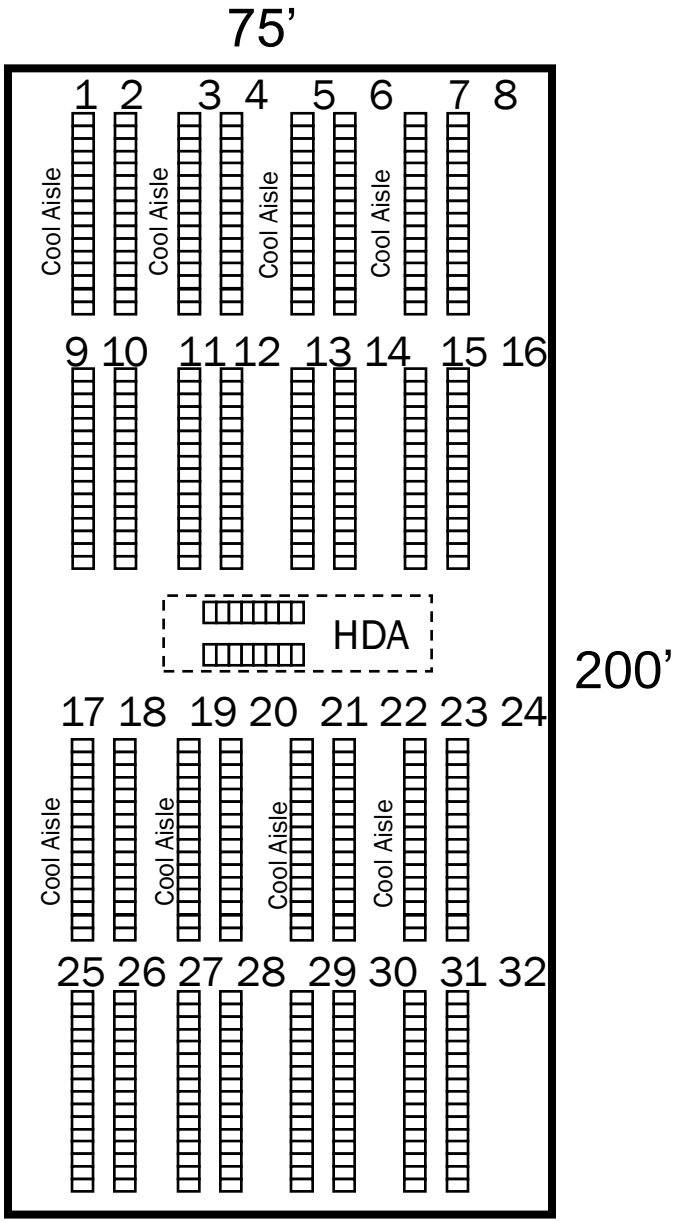
Large Data Center Design

- Most new data centers are being designed with a Pod (CoLo or Cell) Architecture
- Pods usually 15-20,000 sq ft
- HDA (Horizontal Distribution Area) is where distribution switches are located
- The Main Distribution Area (MDA) interconnect PODs and connects to the WAN and telecom networks



POD Architecture

- 15,000 sq ft POD
- Up to 5,000 servers / POD
- 512 Racks possible
  - 32 Rows of racks
  - Each row has 16 racks
- Horizontal Distribution Area (HDA) connects all of the racks



POD Design with Consolidation

200 Server Racks

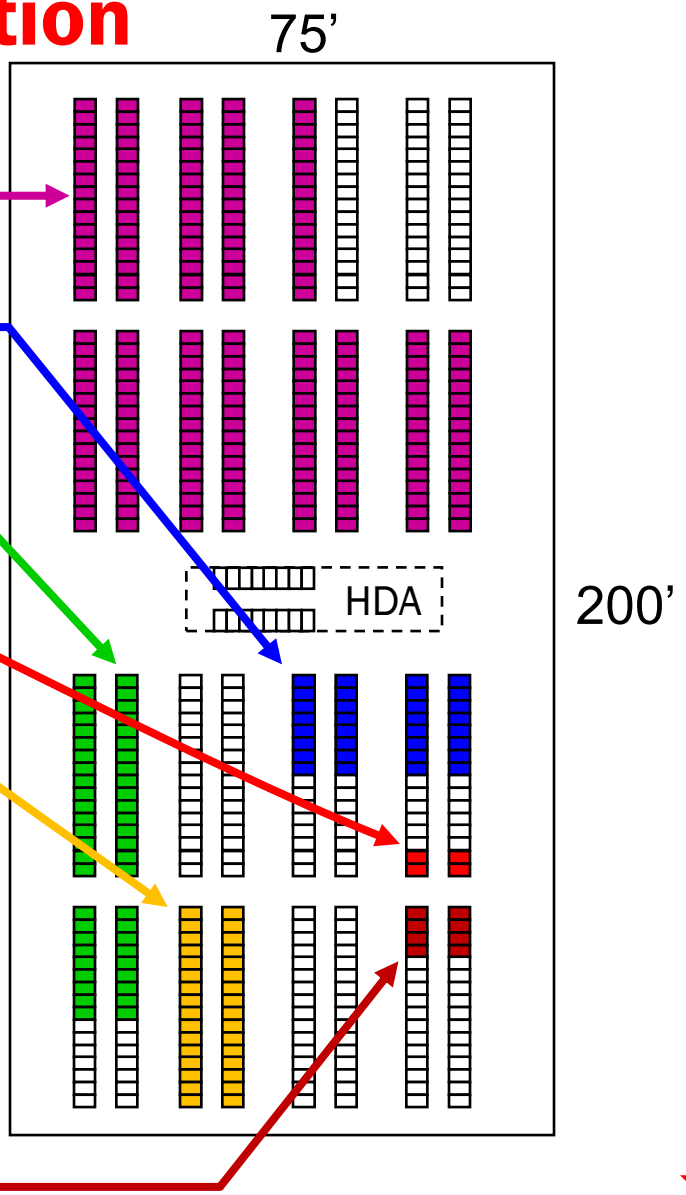
32 Switch Racks

50 Storage Racks

4 Router Racks

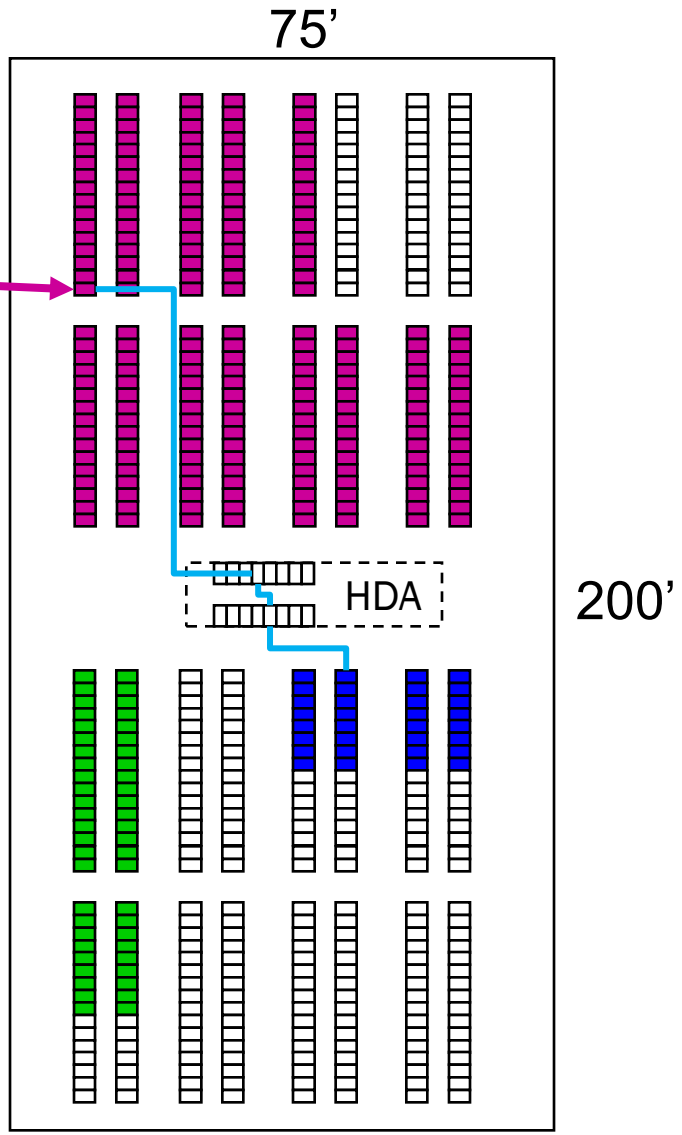
32 Tape Racks

8 Mainframe Racks



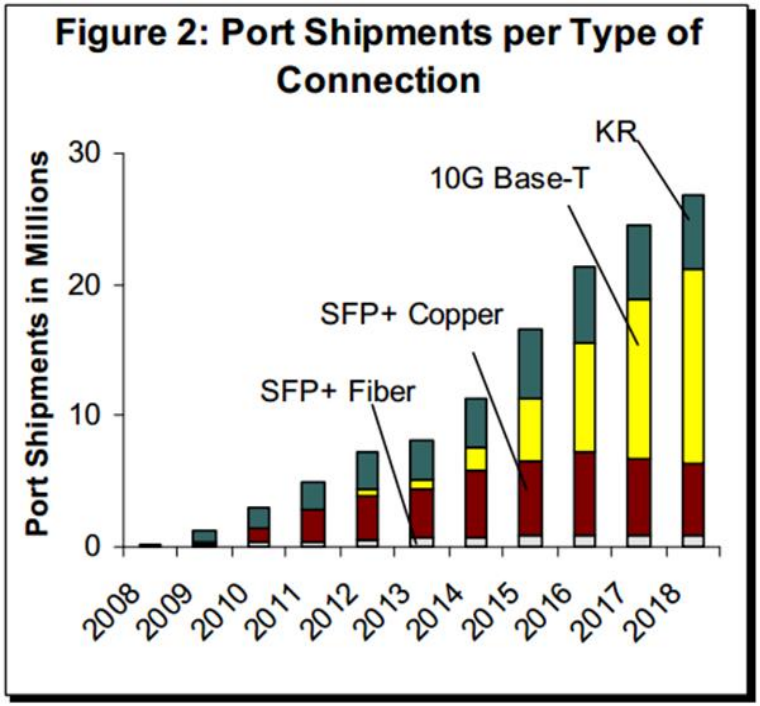
Home Runs

- Home Runs go
  - From each server
  - To HDA
  - Patchcord within HDA
  - To Switch
- 100 meters required



How 10GbE Servers Connect

- Dell'Oro tracks how servers connect to switches by media type, but not topology
- Millions of 10GbE servers are currently connecting to switches with optics



10GbE Optical ports expected to be >1M ports in 2016

Source: Dell'Oro Controller and Adapter Forecast Summary – January 2014

Enable Multiple Solutions with 25GbE Optics

We are in the Cambrian Period of IT with many things possible

- With incredible variation in bandwidth requirements of servers as shown in the 25GbE CFI, we should enable as many applications as possible to increase broad market potential
- Limiting IEEE standards to 3 to 5 meters will limit broad market potential and restrict users to ToR configurations
- Let's enable MoR, EoR and Home Runs with and optical PMD to 100 meters

BROCADE

THANK YOU