Agenda

• Unified Fabric Overview
  Nexus Portfolio
  Feature Offerings

• FCoE Design
  Topologies
  Design Criteria

• Fabric Path
  Overview
  Technical Introduction
Cisco Nexus and MDS Data Center
Industry Leading Technologies and Solutions
Nexus 5500 Series
2nd generation Nexus 5000 with FabricPath & L3

5548: 48 Ports capable of 10GE/1GE/FCoE/FC

- Flexible port configurations
- Unified Ports (Q1CY11)
  - 10GE/FCoE/DCB, 2/4/8 G FC, 1GE
  - Support on all ports of 5596 and expansion module for 5548
- Layer 2 / Layer 3 support
- Increased number of FEX per N5500
- 4K VLANs & 32K Station Table
- 40G Uplinks (future)

5596: 96 Ports capable of 10GE/1GE/FCoE/FC and Unified Port

- 48 N5K Port Channels + 384 N2K port channels + 768 vPC port channels
- Unicast latency: <2µs
- FabricPath - Layer 2 Multi-Pathing Support
  - IETF TRILL & Cisco L2MP
- Standards-based T11 FCoE
- Hardware support for IEEE 1588 (Precision Time Protocol – µs accuracy & timestamp)
Nexus 7000 F-Series SFP+ Module
10GbE supporting Multiprotocol Storage

- Multiprotocol Storage Connectivity FCoE, iSCSI, NAS
- Standards Compliant T11 FCoE, DCBX, IEEE PFC, IEEE ETS
- Scalable 512 Ports Per System, 230Gbps Per Slot
- Low Latency Sub 5µs Port to Port
- Feature Rich Enables Cisco FabricPath Layer 2 Multipathing
- Cost Effective Lowest Price 10GbE Nexus 7000 Port

FCoE License Orderable Q1CY11

32-port 1/10 GbE for Server Access and Aggregation
Direct Attached FC Storage

Customer benefits
- Support to connect FC and FCoE storage to 6100
- Lower hops and reduced latency to access the storage
- End to end FCoE topologies possible

Feature details
- Support for NetApp and EMC direct attached storage
- Ability to turn On/Off FibreChannel limited switching
- Zoning configuration not supported, but zoning may be inherited from upstream switch
- Not a general purpose FC switch – very little switching configuration possible
- Ethernet and FC switching modes are independent
- Note: No management integration with FC Storage
Single Hop Design
Introduction of 10Gig/FCoE Fabric Extender

Nexus 2232

- 32 server facing 10Gig/FCoE ports with 8 10Gig/FCoE uplinks to the Nexus Parent Switch
- Management and configuration handled by the Nexus Parent Switch
- Support for Converged Enhanced Ethernet including PFC
- Supported on Nexus 5000/5500 Platforms for both Ethernet and FCoE
- Support for connection to Nexus 7k in Dehli for Ethernet Only and FCoE in Freetown/F2
- Must be used in “straight-through” mode for FCoE environments

Remote Line Card of the Nexus 5000
**Virtual Expansion Ports (VE_Ports)**

*The beginning of multi-hop FCoE*

VE_Ports refer to an FCoE ISL that runs between two FCoE capable switches:
- Defined by the FC-BB-5 standard as an entity that emulates a Fibre Channel E_Port over a non Fiber Channel link
- The FC-BB_E protocol provides mechanisms to create VE_Port to VE_Port virtual links
- No further standards or protocols necessary for implementing “multi-hop” FCoE

VE_Ports offers the following benefits…
- *Expand* the FCoE fabric beyond a first hop access solution
- Provide multi-hop FCoE connections between FCoE capable switches
FCoE Multi-Tier Fabric Design
VE_Ports Support

Supported on N5k with NX-OS 5.0(2)N2
Supported on N7k/MDS with NX-OS 5.2

VE_Ports are run between switches acting as Fibre Channel Forwarders (FCFs)

- Unified Wire support between N5ks
- Dedicated Wire required when connecting to N7k and MDS platforms

VE_Ports are bound to the underlying 10G infrastructure

- vFC can be bound to a single 10GE port
- vFC can be bound to a port-channel interface consisting of multiple 10GE links
**FCF vs FIP Snooping**

- **What is an FCF???** An FCoE capable device that forwards FCoE frames based on FSPF.
  - Nexus 5k/7k and MDS can all be FCFs.
  - Offer the same visibility for FC traffic that Fibre Channel Switches offer.

- **What is NOT and FCF???** An “DCB Lossless” device that forwards FCoE frames based on Ethernet characteristics (MAC address).
  - Simply and Ethernet Bridge that supports lossless transport.
  - Commonly called “FCoE pass-through” or “DCB capable”.
  - FIP Snooping Bridges/FCoE NPV devices are NOT FCF’s.
  - Limits visibility and control over the FC topology.
Agenda

• Unified Fabric Overview
  Nexus Portfolio
  Feature Offerings

• FCoE Design
  Topologies
  Design Criteria

• Fabric Path
  Overview
  Technical Introduction
FCoE Building Blocks
Unified Wire vs Unified Dedicated Wire

- **Unified Wire** to the access switch
  - Cost savings in the reduction of required equipment
  - “Cable once” for all servers to have access to both LAN and SAN networks

- **Unified Dedicated Wire** from access to aggregation
  - Separate links for SAN and LAN traffic - both links are same I/O (10GE)
  - Advanced Ethernet features can be applied to the LAN links
  - Maintains fabric isolation
Unified Access Designs
Consolidating LAN/SAN at the Access Layer

- Cable once, repurpose for any storage based on server need
- Supports connectivity to existing SAN
  Can be connected via Native FC or FCoE with VE_Ports
- Take advantage of next-gen storage without complete SAN upgrade
- Maintains storage fabric segregation
- Nexus 5000 Access Switch can be in NPV or Switch mode
  NPV is important for interoperability with Brocade
Unified Aggregation Designs
Consolidating LAN/SAN at the Agg-Device Level

• Storage traffic utilizing high-performance, highly available aggregation switch
• Is compatible with traditional Edge-Core-Edge SAN topologies
• Increased scalability at the aggregation layer
  Large FCoE deployments while providing access to existing FC storage (through MDS or N5k)
• Maintaining isolation for Storage traffic
  Dedicated FCoE links
  Dedicated Storage VDCs within the Nexus 7000
FCoE Configuration Considerations

The FCoE VLAN

- Each FCoE VLAN and VSAN count as a VLAN HW resource – therefore a VLAN/VSAN mapping accounts for TWO VLAN resources
- FCoE VLANs are treated differently than native Ethernet VLANs: no flooding, broadcast, MAC learning, etc.
- **BEST PRACTICE**: use different FCoE VLANs/VSANs for SAN A and SAN B
- The FCoE VLAN must not be configured as a native VLAN
- Unified Wires connecting to HOSTS must be configured as **trunk ports** and **STP edge ports**
- **Remember**: STP does not run on FCoE vlans between FCFs (VE_Ports)

```
! VLAN 20 is dedicated for VSAN 2 FCoE traffic
(config)# vlan 20
(config-vlan)# fcoevsan2
```
FCoE Configuration Considerations

FCoE and vPC together

- vPC with FCoE are **ONLY** supported between hosts and N5k or N5k/2232 pairs... **AND they must follow specific rules**
  - A ‘vfc’ interface can only be associated with a single-port port-channel
  - While the port-channel configurations are the same on N5K-1 and N5K-2, the FCoE VLANs are different

- FCoE VLANs are ‘not’ carried on the vPC peer-link (automatically pruned)
  - FCoE and FIP ethertypes are ‘not’ forwarded over the vPC peer link either

- vPC carrying FCoE between two FCF’s is **NOT** supported

- vPC with FCoE from host to N7k is **NOT** supported at FCS
Multi-Hop Design
Extending FCoE with FIP Snooping

Nexus 4000 is a Unified Fabric capable Blade Switch
- DCB enabled (PFC support)
- FIP Snooping Bridge (added security mentioned in the FC-BB-5 Annex)

Servers IP connection to the Nexus 4000 is Active/Standby

MCEC is not currently supported from blade server to Nexus 4000

Options 1: Dedicated Wires from Nexus 4000 to Nexus 5000

Options 2: Unified Wire (Port Channel) from Nexus 4000 to Nexus 5000
Does attaching FCoE targets to a shared Core/Agg make sense?

- Different requirements
  - Traditionally, SAN Core provides ports for targets
  - LAN Aggregation traditionally does not attach end devices

- Factors that will influence this use case
  - Port density
  - Operational roles and change management

- Potentially viable for smaller environments

- Larger environments will need dedicated FCoE ‘SAN’ devices providing target ports
Unified SAN Core Designs
Dedicated Ethernet SAN Core

- Utilize Common Platforms for LAN and SAN
- Utilize higher performance and capacity Ethernet switches for SAN core
- Takes advantage of Ethernet Economics and Roadmap
- Provides Organizational Separation
vPC+FCoE is **NOT SUPPORTED** upstream from Nexus 5k/2232 to next-hop switch
- Breaks SAN A / SAN B physical isolation
- Can **NOT** segregate traffic types on FEX uplinks to parent switch

Nexus 2232 with FCoE supported on Nexus 5X00 only
- Nexus 7000 supports Nexus 2232 for Ethernet

Unified Wire **NOT** supported on VE_Ports between Nexus 5k and Nexus 7k/MDS

Nexus 4k connections to 2232, N7k or MDS for FCoE are **NOT SUPPORTED** (not shown)
FCoE Configurations Considerations

**UNSUPPORTED Designs**

- Optimal layer 2 LAN design may not meet FC high availability and operational design requirements
  - This is where Unified Wire and Dedicated Wire comes in…
- Features such as vPC & MCEC are not viable and not supported beyond the direct attached server
- L2MP and TRILL provide options to change the design paradigm and come up with potential solutions
  - FCoE over L2MP/TRILL is not currently supported

![Diagram showing network configurations and labels such as SAN A, SAN B, FIP, and FCoE frames]
Agenda

• Unified Fabric Overview
  Nexus Portfolio
  Feature Offerings

• FCoE Design
  Topologies
  Design Criteria

• Fabric Path
  Overview
  Technical Introduction
L2 Provides Flexibility in the Data Center

- Layer 2 required by data center applications
- Layer 2 is “plug and play”
- Layer 2 is Layer 3 agnostic
- With Layer 2:
  - Server mobility does not require interaction between Network/Server teams
  - Theoretically, no physical constraint on server location
MAC Address Scaling

- MAC addresses encode no location or network hierarchy
- Default forwarding behavior in bridged network is flood
- MAC filtering database limits scope of flooding
- Ultimately, does not scale – every switch learns every MAC
L2 Requires a Tree

- Spanning Tree Protocol (STP) typically used to build this tree
- Tree topology implies:
  - Wasted bandwidth → increased oversubscription
  - Sub-optimal paths
  - Conservative convergence (timer-based) → failure catastrophic (fails open)
Introducing Cisco Fabric Path
An NX-OS Innovation of L2 Networks

"The FabricPath capability within Cisco's NX-OS offers dramatic increases in network scalability and resiliency for our service delivery data center. FabricPath extends the benefits of the Nexus 7000 in our network, allowing us to leverage a common platform, simplify operations, and reduce operational costs."

Mr. Klaus Schmid, Head of DC Network & Operating, T-Systems International GmbH
Architecture Flexibility Through NX-OS

- **Spanning-Tree**
  - Single
  - Up to 10 Tbps

- **vPC**
  - Dual
  - Up to 20 Tbps

- **FabricPath**
  - 16 Way
  - Up to 160 Tbps

**Layer 2 Scalability**

**Infrastructure Virtualization and Capacity**
FabricPath - Simplicity from the Outside

- Benefits server team by providing a network Fabric that looks like a single switch → Breaks down silos, permits workload mobility, provides maximum flexibility

- Lowers OPEX by simplifying server team operation → Reduces dependency on/interaction with network team
FabricPath - Simplicity from the Inside

Benefits network team by:

- Reducing number of switches
  - Higher port density
  - Lower oversubscription
- Isolating network from the users
  - No impact due to topology changes
  - Fabric can be upgraded/reconfigured live
- Utilizing an open protocol
  - Unicast, multicast, broadcast, VLAN pruning all controlled by single control protocol
  - Maintenance and troubleshooting equivalent to L3 network
  - Easy to extend, providing standards-compliance with Cisco value-add
FabricPath IS-IS

- FabricPath IS-IS replaces STP as control-plane protocol in FabricPath network
- Introduces link-state protocol with support for ECMP for Layer 2 forwarding
- Exchanges reachability of Switch IDs and builds forwarding trees
- Improves failure detection, network reconvergence, and high availability
- Minimal IS-IS knowledge required – no user configuration by default
  Maintains plug-and-play nature of Layer 2
Why IS-IS?

A few key reasons:

• Has no IP dependency – no need for IP reachability in order to form adjacency between devices

• Easily extensible – Using custom TLVs, IS-IS devices can exchange information about virtually anything

• Provides SPF routing – Excellent topology building and reconvergence characteristics
FabricPath- versus Classic Ethernet Interfaces

Classic Ethernet (CE) Interface
- Interfaces connected to existing NICs and traditional network devices
- Send/receive traffic in 802.3 Ethernet frame format
- Participate in STP domain
- Forwarding based on MAC table

FabricPath Interface
- Interfaces connected to another FabricPath device
- Send/receive traffic with FabricPath header
- No spanning tree!!!
- No MAC learning
- Exchange topology info through L2 ISIS adjacency
- Forwarding based on ‘Switch ID Table’
In FabricPath system, each VLAN identified as either a CE VLAN (default) or a FabricPath VLAN.

Only traffic in FabricPath VLANs can traverse FabricPath domain.

Bridging between M1 and F1 ports possible only on CE VLANs.

```plaintext
n7k(config)# vlan 10
n7k(config-vlan)# mode ?
    ce    Classical Ethernet VLAN mode
    fabricpath    FabricPath VLAN mode

n7k(config-vlan)# mode
```
Ingress FabricPath switch determines destination Switch ID and imposes FabricPath header.

Destination Switch ID used to make routing decisions through FabricPath core.

No MAC learning or lookups required inside core.

Egress FabricPath switch removes FabricPath header and forwards to CE.
FabricPath Encapsulation
16-Byte MAC-in-MAC Header

Classical Ethernet Frame

- **Switch ID** – Unique number identifying each FabricPath switch
- **Sub-Switch ID** – Identifies devices/hosts connected via VPC+
- **Port ID** – Identifies the destination or source interface
- **Ftag** (Forwarding tag) – Unique number identifying topology and/or multidestination distribution tree
- **TTL** – Decremented at each switch hop to prevent frames looping infinitely
FabricPath MAC Table

- Edge switches maintain both MAC address table and Switch ID table
- Ingress switch uses MAC table to determine destination Switch ID
- Egress switch uses MAC table (optionally) to determine output switchport

<table>
<thead>
<tr>
<th>MAC</th>
<th>IF/SID</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>e1/1</td>
</tr>
<tr>
<td>B</td>
<td>e1/2</td>
</tr>
<tr>
<td>C</td>
<td>S101</td>
</tr>
<tr>
<td>D</td>
<td>S200</td>
</tr>
</tbody>
</table>
Show mac address-table dynamic

<table>
<thead>
<tr>
<th>VLAN</th>
<th>MAC Address</th>
<th>Type</th>
<th>age</th>
<th>Secure NTFY Ports/SWID.SSID.LID</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0000.0000.0001</td>
<td>dynamic</td>
<td>0</td>
<td>F</td>
</tr>
<tr>
<td>10</td>
<td>0000.0000.0002</td>
<td>dynamic</td>
<td>0</td>
<td>F</td>
</tr>
<tr>
<td>10</td>
<td>0000.0000.0003</td>
<td>dynamic</td>
<td>0</td>
<td>F</td>
</tr>
<tr>
<td>10</td>
<td>0000.0000.0004</td>
<td>dynamic</td>
<td>0</td>
<td>F</td>
</tr>
<tr>
<td>10</td>
<td>0000.0000.0005</td>
<td>dynamic</td>
<td>0</td>
<td>F</td>
</tr>
<tr>
<td>10</td>
<td>0000.0000.0006</td>
<td>dynamic</td>
<td>0</td>
<td>F</td>
</tr>
<tr>
<td>10</td>
<td>0000.0000.0007</td>
<td>dynamic</td>
<td>0</td>
<td>F</td>
</tr>
<tr>
<td>10</td>
<td>0000.0000.0008</td>
<td>dynamic</td>
<td>0</td>
<td>F</td>
</tr>
<tr>
<td>10</td>
<td>0000.0000.0009</td>
<td>dynamic</td>
<td>0</td>
<td>F</td>
</tr>
<tr>
<td>10</td>
<td>0000.0000.000a</td>
<td>dynamic</td>
<td>0</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>0000.0000.000b</td>
<td>dynamic</td>
<td>0</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>0000.0000.000c</td>
<td>dynamic</td>
<td>0</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>0000.0000.000d</td>
<td>dynamic</td>
<td>0</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>0000.0000.000e</td>
<td>dynamic</td>
<td>0</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>0000.0000.000f</td>
<td>dynamic</td>
<td>0</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>0000.0000.0010</td>
<td>dynamic</td>
<td>0</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>0000.0000.0011</td>
<td>dynamic</td>
<td>0</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>0000.0000.0012</td>
<td>dynamic</td>
<td>0</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>0000.0000.0013</td>
<td>dynamic</td>
<td>0</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>0000.0000.0014</td>
<td>dynamic</td>
<td>0</td>
<td>F</td>
</tr>
</tbody>
</table>

S100#
FabricPath IS-IS manages Switch ID (routing) table

All FabricPath-enabled switches automatically assigned Switch ID (no user configuration required)

Algorithm computes shortest (best) paths to each Switch ID based on link metrics

Equal-cost paths supported between FabricPath switches
Building the FabricPath Routing Table

Switch | IF
-------|-----
S20    | L1, L5, L9
S30    | L1, L5, L9
S40    | L1, L5, L9
S100   | L1
S101   | L5
...    | ...
S200   | L9

Switch | IF
-------|-----
S10    | L4, L8, L12
S20    | L4, L8, L12
S30    | L4, L8, L12
S100   | L4
S101   | L8
...    | ...
S200   | L12

Switch | IF
-------|-----
S10    | L1
S20    | L2
S30    | L3
S40    | L4
S101   | L1, L2, L3, L4
...    | ...
S200   | L1, L2, L3, L4

Switch | IF
-------|-----
S10    | L9
S20    | L10
S30    | L11
S40    | L12
S100   | L9, L10, L11, L12
S101   | L9, L10, L11, L12
...    | ...

Switch | IF
-------|-----
S10    | L9
S20    | L10
S30    | L11
S40    | L12
S100   | L9, L10, L11, L12
S101   | L9, L10, L11, L12
...    | ...

MAC A   | L1, L2, L3, L4
MAC B   | L1, L2, L3, L4
MAC C   | L1, L2, L3, L4
MAC D   | L1, L2, L3, L4

FabricPath

FP/IS-IS
show fabricpath route

S100# sh fabricpath route
FabricPath Unicast Route Table
'a/b/c' denotes ftag/switch-id/subswitch-id
'[x/y]' denotes [admin distance/metric]
ftag 0 is local ftag
subswitch-id 0 is default subswitch-id

FabricPath Unicast Route Table for Topology-Default
0/100/0, number of next-hops: 0
  via ---- , [60/0], 5 day/s 18:38:46, local
1/10/0, number of next-hops: 1
  via Po1, [115/10], 0 day/s 04:15:58, isis_l2mp-default
1/20/0, number of next-hops: 1
  via Po2, [115/10], 0 day/s 04:16:05, isis_l2mp-default
1/30/0, number of next-hops: 1
  via Po3, [115/10], 2 day/s 08:49:51, isis_l2mp-default
1/40/0, number of next-hops: 1
  via Po4, [115/10], 2 day/s 08:47:56, isis_l2mp-default
1/200/0, number of next-hops: 4
  via Po1, [115/20], 0 day/s 04:15:58, isis_l2mp-default
  via Po2, [115/20], 0 day/s 04:15:58, isis_l2mp-default
  via Po3, [115/20], 2 day/s 08:49:51, isis_l2mp-default
  via Po4, [115/20], 2 day/s 08:47:56, isis_l2mp-default

S100#
FabricPath ECMP

- When multiple forwarding paths available, path selection based on ECMP hash function
- Up to 16 next-hop interfaces for each destination Switch ID
- Number of next-hops installed in U2RIB controlled by maximum-paths command under FabricPath IS-IS process (default is 16)
- Path selection based on hash function
Conversational MAC Learning

- MAC learning method designed to conserve MAC table entries on FabricPath edge switches
  - FabricPath core switches do not learn MACs at all
- Each forwarding engine distinguishes between two types of MAC entry:
  - Local MAC – MAC of host directly connected to forwarding engine
  - Remote MAC – MAC of host connected to another forwarding engine or switch
- Forwarding engine learns remote MAC only if bidirectional conversation occurring between local and remote MAC
  - MAC learning not triggered by flood frames
- Conversational learning enabled in all FabricPath VLANs
Conversational MAC Learning

FabricPath MAC Table on S100

<table>
<thead>
<tr>
<th>MAC</th>
<th>IF/SID</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>e1/1 (local)</td>
</tr>
<tr>
<td>B</td>
<td>S200 (remote)</td>
</tr>
</tbody>
</table>

FabricPath MAC Table on S200

<table>
<thead>
<tr>
<th>MAC</th>
<th>IF/SID</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>S100 (remote)</td>
</tr>
<tr>
<td>B</td>
<td>e12/1 (local)</td>
</tr>
<tr>
<td>C</td>
<td>S300 (remote)</td>
</tr>
</tbody>
</table>

FabricPath MAC Table on S300

<table>
<thead>
<tr>
<th>MAC</th>
<th>IF/SID</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>S200 (remote)</td>
</tr>
<tr>
<td>C</td>
<td>e7/10 (local)</td>
</tr>
</tbody>
</table>
For more information and registration:

http://www.ciscolive.com/