DC Design for Small and Mid-Size Data Center

Višnja Milovanović, CSE
Data Center and Virtualization
South East Europe
Data Center Designs for Midmarket and SMBs: 100 – 300 ports

We will focus on:

- The recommended N5K/N2K design scenarios for SMB customer (from 100 to 300 ports)
- L2 and L3 capabilities (vPC, first-hop redundancy in N5K)
- Servers connectivity: CNAs/FCoE, and classical (NIC+HBA) scenarios
- SAN (iSCSI, FCoE and FC) and NAS deployment scenarios
- DC Services insertion (ACE appliance, WAAS, ASA, VSG, IPS, NAM ...etc)
- Server Access layer security for virtual machines
DC Designs
100 – 300 ports

- Rack Mount Servers
  Self contained Virtualized Data Center

- Leverage Campus Core (Small Modular) as Collapsed Core & Backbone

- Differentiation – FCoE, FEX, EvPC, Adapter-FEX, Unified Port, N1KV, vPATH, VXLAN, FabricPath

- Dependent on ASR WAN router for DCI and other transport services
DC Designs
Virtualized Data Center Switch

- 2 x Nexus 5548 + Nexus 2200 access switches
  - 2 x Nexus 5548
  - 2 x L3 Module
  - L3 License & Storage License
  - 2 x Nexus 2232
  - 4 x Nexus 2248TP-E

Nexus/NX-OS Features

Scale

Low
High

Solution Pricing

Low
High

Nexus 5548UP with L3 module
Nexus 2232 and 2248TP-E

Catalyst 4507 Sup7E

WAN

FCoE

NAS iSCSI

FC

VM #1

VM #2

VM #3

VM #4
DC Design Details
vPC & FEX Options

- Core Design Criteria
  Server NIC teaming requirements (Single NIC, ALB/TLB, 802.3ad, …)

- Design Options
  Straight Through FEX
  Dual Homed FEX
  EvPC (new capability supported with 5.1(3)N1)

- Considerations
  FCoE (isolation of SAN ‘A’ & SAN ‘B’)

- Design Strengths
  EvPC enables all server NIC redundancy options simultaneously
Redundancy model – Dual Switch with redundant fabric
- Provides isolation for Storage topologies (SAN ‘A’ and ‘B’)

vPC Supported with up to 2 x 8 links

vPC Supported with up to 2 x 8 links

- Redundancy model – Single switch with dual ‘supervisor’ for fabric, data control & management planes
- No SAN ‘A’ and ‘B’ isolation

Local Etherchannel with up to 8 links

FCoE Adapters supported on 10G N2K interfaces
A vPC orphan port is an non-vPC interface on a switch where other ports in the same VLAN are configured as vPC interfaces.

With the NX-OS release 5.0(3)N2 the orphan ports on the vPC secondary peer will also be shut down triggering NIC teaming recovery for all teaming configurations.

Configuration is applied to the physical port.
In an Enhanced vPC (EvPC) configuration any and all server NIC teaming configurations will be supported on any port.

Requires NX-OS 5.1(3)N1 (shipping Q4 CY11).

No ‘orphan ports’ in the design.

All components fully redundant in a MCEC environment.

Supported with *Nexus 5500 only*.

*Not* required to support a mixed NIC teaming environment, use case is restricted to a mix of all three server NIC configurations (single NIC, ALB/TLB and 802.3ad).
DC Designs
Nexus 5500 Topologies starting with NX-OS 5.1(3)N1

• In an Enhanced vPC (EvPC) SAN ‘A/B’ isolation is configured by associating each FEX with either SAN ‘A’ or SAN ‘B’ Nexus 5500

• FCoE & FIP traffic is forwarded only over the links connected to the specific parent switch

• Ethernet is hashed over ‘all’ FEX fabric links
Data Center Designs for Midmarket and SMBs: 100 – 300 ports

This update to this section will focus on:

- The recommended N5K/N2K design scenarios for SMB customer (from 100 to 300 ports)
- L2 and L3 capabilities (vPC, first-hop redundancy in N5K)
- Servers connectivity: CNAs/FCoE, and classical (NIC+HBA) scenarios
- SAN (iSCSI, FCoE and FC) and NAS deployment scenarios
- DC Services insertion (ACE appliance, WAAS, ASA, VSG, IPS, NAM ...etc)
- Server Access layer security for virtual machines
DC Design Details
Layer 3 Considerations

• Core Design Considerations
  Support for a maximum of 8 x FEX connected to a Nexus 5500 with an L3 daughter-card

• Key Support Considerations
  vPC configuration for L3 interfaces needs to be consistent between peers
  ISSU is not currently supported with Layer 3 enabled

• Key Design Strengths
  Cisco NX-OS routing capabilities (EIGRP, OSPF, RIPv2, MBGP, HSRP, VRRP)
  VRF Support
**DC Designs**

**Nexus 5500 Layer 3 Topologies – 8 FEX**

**Standard vPC**
- 8 FEX attached to ‘each’ Nexus 5548
- 16 total FEX in the fabric
- Maximum of 768 x 1G ports

**Enhanced vPC**
- 8 FEX are attached to ‘both’ Nexus 5548
- 8 total FEX in the fabric
- Maximum of 384 x 1G ports

**NOTE:** In a dual homed (EvPC) FEX configuration the FEX is counted against the maximum supported FEX count on ‘both’ parent switches
DC Design Details
Layer 3 Design Considerations

- HSRP tuning not required with vPC dual active default gateway
- Remember to configure L3 peering between vPC peers
- Passive all SVI other than this L3 peering and the routed uplink interfaces
- EIGRP stub requires some tuning to compatible with vPC L3 design
  Need to provide failover to a secondary stub (vPC peer) – Leverage floating static default to point to loopback on peer switch
- Layer 3 + FEX + vPC designs are limited to 1K multicast groups
DC Design Details
Layer 3 Considerations

• QoS policy ‘must’ use a set cos value to ensure that queueing is correctly performed for traffic passing to and from the L3 daughter-card

• Leverage vPC enhancements
  “peer-gateway” – 5.0(2)N1
  “arp table sync” – 5.1(3)N1
  “auto-recovery” – 5.0(2)N1

• Leverage vPC+ if PIM-SSM support is required (FabricPath license)

Layer 3 Design Guide:
This update to this section will focus on:

- The recommended N5K/N2K design scenarios for SMB customer (from 100 to 300 ports)
- L2 and L3 capabilities (vPC, first-hop redundancy in N5K)
- Servers connectivity: CNAs/FCoE, and classical (NIC+HBA) scenarios
- SAN (iSCSI, FCoE and FC) and NAS deployment scenarios
- DC Services insertion (ACE appliance, WAAS, ASA, VSG, IPS, NAM ...etc)
- Server Access layer security for virtual machines
DC Design Details
Rack and Server Considerations

- Core Design Considerations
  - Power Allocated Per Rack
  - Server NIC Density
  - Migration from 1G to 10G
- Cisco Nexus and NX-OS Infrastructure Strengths:
  - FEX provides migration to 10G and cost effective 1G (cable reduction)
  - B22 or pass-thru options support 3rd party Blade chassis
  - FCoE, Adapter-FEX
- Issues to be aware of
  - 10GBaseT FEX (2232TM) does ‘not’ support FCoE
DC Design Details
Optimize the port density per rack

Three racks configured as a redundancy pair - FEX ‘A’ & ‘B’ racked in common center rack

Two racks configured as a redundancy pair - FEX ‘A’ & ‘B’ split between 2 racks

<table>
<thead>
<tr>
<th>FEX Model</th>
<th>Servers/Rack</th>
<th>Ports/Server</th>
</tr>
</thead>
<tbody>
<tr>
<td>2224TP</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>2248TP-E</td>
<td>16</td>
<td>2</td>
</tr>
<tr>
<td>2232TM</td>
<td>12</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FEX Model</th>
<th>Servers/Rack</th>
<th>Ports/Server</th>
</tr>
</thead>
<tbody>
<tr>
<td>2224TP</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>2248TP-E</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>2232TM</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>

Example Rack/Server layouts
10GBaseT Key Benefits:

- 10 Gigabit bandwidth requirements for 1 Gig NIC consolidation at the server access and virtual environments
- Ease of 1GBASE-T to 10GBASE-T migration
- Flexible, scalable cabling with standard RJ-45 connector, at distances up to 100m
- Reuse of existing structured cabling
- Economics of 1Gigabit Ethernet versus 10 Gigabit Ethernet
- Prepare for Server LOM
- Investment protection with 1/10G capabilities

Source: Crehan Research (Q4CY10)
DC Design Details
Migration from 1G to 10G servers

**Nexus 2248TP**
List Price: $9,000
List Price w/ FET: $11,000

**Nexus 2248TP-E**
List Price: $10,000
List Price w/ FET: $12,000

**Nexus 2232TM**
List Price: $11,500
List Price w/ FET: $15,500

- Nexus 2000 provides an option for a per ‘line card’ migration of 1G and 10G
- Nexus 2232TM and 2232PP provide a per ‘port’ migration path for 1G to 10G
DC Design Details
10GBaseT – Power and EMI Considerations

- Undesired coupling of signal between adjacent cables
- Main electrical parameter limiting the performance of 10G
- Cannot be cancelled
- Re-Training is the major barrier to use of 10GBaseT for block level storage (FCoE)
- Can be prevented or mitigated by:
  - Space (Cat6a solution)
  - Shield (Cat6/Ca6a/Cat7 shielded solutions)

<table>
<thead>
<tr>
<th>Technology</th>
<th>Cable</th>
<th>Distance</th>
<th>Power (each side)</th>
<th>Transceiver Latency</th>
</tr>
</thead>
<tbody>
<tr>
<td>2232PP</td>
<td>SFP+ CU Copper</td>
<td>1-10m</td>
<td>~0.1 - 1W</td>
<td>~0.25ms</td>
</tr>
<tr>
<td>10GBASE-T – 65nm</td>
<td>Cat6/6a/7</td>
<td>100m</td>
<td>~6W</td>
<td>~3ms</td>
</tr>
<tr>
<td></td>
<td>Cat6/6a/7</td>
<td>30m</td>
<td>~4-5W</td>
<td>~3ms</td>
</tr>
</tbody>
</table>
DC Design Details
Blade Server Considerations

- Blade Chassis
  - Migration to 2 x 10G NIC/CNAs per blade
  - Typical Sizing: SMB for HP is up to 4 x c7000 enclosures (64 blade servers)
  - 2 to 4 x 10G uplinks per chassis

- Cisco Strengths:
  - FEX + Pass Thru enables a Nexus capable host
    - FCoE capabilities & Adapter-FEX
  - B22-HP and Nexus 4000 (IBM)
DC Design Details – Blade Chassis
Nexus B22 Series Fabric Extender

- B22 extends FEX connectivity into the HP blade chassis
- Cisco Nexus 5000 Switch is a single management point for all the blade chassis I/O modules
- 66% decrease in blade management points*
- Blade & rack networking consistency
- Interoperable with Nexus 2000 Fabric Extenders in the same Nexus parent switch
- End-to-end FCoE support
- Support for 1G & 10G, LOM and Mez
- Dell supports Pass-Thru as an alternative option to directly attaching Blade Servers to FEX ports
DC Design Details
Mixed 1/10G, FC/FCoE, Rack and Blade

- Consolidation for all servers both rack and blade onto the same virtual switch
- Support for 1G, migration to 10G, FC and migration to FCoE

1G server racks are supported by 1G FEX (2248TP, 2224TP) or future proofed with 1/10G FEX (2232PP or 2232TM)

10G server racks are supported by the addition of a new 10G FEX (2232PP or 2232TM)

Support for direct connection of HBA to Unified Ports on Nexus 5500UP

Support for NPV attached blade switches during FC to FCoE migration

1G, 10G and FCoE connectivity for HP Blade Chassis
DC Design Details
Virtualized Interfaces – Adapter-FEX

- Adapter-FEX presents standard PCIe virtual NICs (vNICs) to servers
- Adapter-FEX virtual NICs are configured and managed via Nexus 5500
- Forwarding, Queuing, and Policy enforcement for vNIC traffic by Nexus 5500
- Adapter-FEX connected to Nexus 2000 Fabric Extender - Cascaded FEX-Link deployment
- Forwarding, Queuing, and Policy enforcement for vNIC traffic still done by Nexus 5500
**Nexus 5500 Adapter-FEX**

**P81E Virtual Interface Card & BCM57712 CNA**

- vNICs are presented to the host like standard PCIe devices
- In A-FEX mode: supports up to 16 Eth vNIC and 2 FC vHBA
- Adapter Failover feature: in failure scenarios, the vNIC is mapped to the other port transparently to the OS
- In VM-FEX mode: supports up to 96 Virtual Interfaces (vNICs + vHBAs)
- No need of trunking all VLANs to the server interface (improving security and scalability)

- 3rd Party adapter supporting VN-TAG
- vNICs are presented to the host like standard PCIe devices
- In A-FEX mode supports up to 8 Virtual Interfaces total
  - Max of 8 vEth
  - Max of 2 vHba
- No adapter failover
DC Designs

UCS Option

- 5500 providing support for disjoint L2 domains and/or L3
- 2 x Nexus 5548 Data Center Core Virtualized Switch
- Fabric Differentiation – FCoE, FEX, Adapter-FEX, VM-FEX, Unified Port, N1KV, vPATH, VXLAN, FabricPath, UCSM
- iSCSI Boot supported Q1CY12 targeted at SMB customer base migrating to NFS/CIFS
- Dependent on ASR WAN router for DCI and other transport services (WCCP, LISP, MPLS, OTV)
Data Center Designs for Midmarket and SMBs: 100 – 300 ports

This update to this section will focus on:

• The recommended N5K/N2K design scenarios for SMB customer (from 100 to 300 ports)
• L2 and L3 capabilities (vPC, first-hop redundancy in N5K)
• Servers connectivity: CNAs/FCoE, and classical (NIC+HBA) scenarios
• SAN (iSCSI, FCoE and FC) and NAS deployment scenarios
• DC Services insertion (ACE appliance, WAAS, ASA, VSG, IPS, NAM ...etc)
• Server Access layer security for (virtual machines and bare-metal servers)
DC Design Details

Storage Considerations

• Core Design Considerations
  • Storage I/O capacity required per server
  • Storage protocol(s) required
  • Boot from FC/FCoE/iSCSI

• Cisco Nexus and NX-OS

Infrastructure Strengths:

QoS capabilities to support iSCSI, NFS, NAS

Nexus 5500 supports FC and FCoE fabric services

• Issues to be aware of
  • 10GBaseT FEX (2232TM) does ‘not’ support FCoE
DC Design Details
FC & FCoE Considerations

- Unified Ports supported on Nexus 5548UP and 5596UP
- Nexus 5500 supports FC and FCoE fabric services
- Co-existence with layer 3 and layer 2 IP forwarding (vPC, STP or FabricPath)
- FCoE Supported on Nexus 5000/5500 and 2232PP and B22 FEX ports
- Issues to be aware of
  - 10GBaseT FEX (2232TM) does ‘not’ support FCoE
  - IVR is not supported on Nexus 5500
### DC Design Details

**FCoE/FC Storage Considerations**

- EvPC supports FCoE with dual homed FEX as of NX-OS 5.1(3)N1
- Boot from SAN with over an 802.3ad port channel using LACP is supported with NX-OS 5.1(3)N1
  - The ‘vfc’ is bound directly to physical port that is a member of the vPC port channel
  - Prior to 5.1(3)N1 the ‘vfc’ was bound to the port channel which could result in a race condition on boot from SAN
- The FCoE VLAN must not be configured as a native VLAN
- Ports with CNA must be configured as trunk ports and STP edge ports

**Configuration Example**

```plaintext
interface vfc 1
bind interface eth 101/1/1
```
DCBX Switch

DCBX CNA Adapter

iSCSI and DCB
DCBX Overview - 802.1Qaz

- Negotiates Ethernet capability's: PFC, ETS, CoS values between DCB capable peer devices
- Simplifies Management: allows for configuration and distribution of parameters from one node to another
- Responsible for Logical Link Up/Down signaling of Ethernet and Fibre Channel
- DCBX is LLDP with new TLV fields
- The original pre-standard CIN (Cisco, Intel, Nuova) DCBX utilized additional TLV's
- DCBX negotiation failures result in:
  - per-priority-pause not enabled on CoS values
  - vfc not coming up – when DCBX is being used in FCoE environment

```
dc11-5020-3# sh lldp dcbx interface eth 1/40
Local DCBXP Control information:
  Operation version: 00  Max version: 00  Seq no: 7  Ack no: 0
  Type/Subtype    Version    En/Will/Adv Config
  006/000        000        Y/N/Y      00
<snip>
```

iSCSI and DCB
PFC (802.1Qbb) & ETS 802.1Qaz

- Nexus 5000/5500 and F1/F2 line cards support DCB (PFC, ETS and DCBX)
- Which mechanisms are beneficial for iSCSI
- Core requirement is ensuring capacity (ETS)
DC Design Details
NAS/NFS Storage Considerations

- Speed mismatch between 10G NAS and 1G server requires QoS tuning
- **Nexus 2248TP-E** utilizes a 32MB shared buffer to handle larger traffic bursts
- Hadoop, NAS, AVID are examples of bursty applications
- You can control the queue limit for a specified Fabric Extender for egress direction (from the network to the host)
- You can use a lower queue limit value on the Fabric Extender to prevent one blocked receiver from affecting traffic that is sent to other non-congested receivers ("head-of-line blocking")

```
N5548-L3(config)# fex 100
N5548-L3(config-fex)# hardware N2248T queue-limit 356000
N5548-L3(config-fex)# hardware N2248TPE queue-limit 4000000 rx
N5548-L3(config-fex)# hardware N2248TPE queue-limit 4000000 tx
N5548-L3(config)# interface e110/1/1
N5548-L3(config-if)# hardware N2348TP queue-limit 4096000 tx
```

Tune 2248TP to support a larger burst (NAS, …)
Tune 2248TP-E to support a extremely large burst (Hadoop, AVID, …)
DC Design
SAN Storage Extension – MDS 9221i

- 1/2/4/8 FC interfaces
- Up to 4095 BB_credits on a FC port
- Integrated FCIP
- Traffic Engineering with VSANs

SAN Services:
- Encryption (FCIP)
- Compression (FCIP)
- I/O Acceleration (FC and FCIP)
- Remote Site Data Migration (DMM)
### SAN Extension

Buffer Credits – Distance & Link Speed

<table>
<thead>
<tr>
<th>Speed (Gbps)</th>
<th>Distance (Km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>250</td>
</tr>
<tr>
<td>2</td>
<td>125</td>
</tr>
<tr>
<td>4</td>
<td>62</td>
</tr>
<tr>
<td>8</td>
<td>31</td>
</tr>
</tbody>
</table>

MDS 9148 Fibre Channel Ports

<table>
<thead>
<tr>
<th>Speed (Gbps)</th>
<th>Distance (Km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>128</td>
</tr>
<tr>
<td>2</td>
<td>64</td>
</tr>
<tr>
<td>4</td>
<td>32</td>
</tr>
<tr>
<td>8</td>
<td>16</td>
</tr>
</tbody>
</table>

Nexus 5000/5500 Fibre Channel Ports
SAN Extension
Unified Fabric Links

- Converged Fabric, sharing a common 10G wire (or non vPC port channel) for FCOE and IP traffic is supported in a square topology (FCoE links can be up to 3 km on Nexus 5500)

- vPC not currently certified for the ‘DCI’ links
This update to this section will focus on:

- The recommended N5K/N2K design scenarios for SMB customer (from 100 to 300 ports)
- L2 and L3 capabilities (vPC, first-hop redundancy in N5K)
- Servers connectivity: CNAs/FCoE, and classical (NIC+HBA) scenarios
- SAN (iSCSI, FCoE and FC) and NAS deployment scenarios
- DC Services insertion (ACE appliance, WAAS, ASA, IPS, NAM ...etc)
- Server Access layer security for (virtual machines and bare-metal servers)
Services Design Guidance
Smart Business Architecture (SBA)

- Updated twice a year (last one was August 2011)
- Tested in SBA labs as a complete architecture
- Overview, configuration files and deployment guides

Data Center Designs for Midmarket and SMBs: 100 – 300 ports

This update to this section will focus on:

- The recommended N5K/N2K design scenarios for SMB customer (from 100 to 300 ports)
- L2 and L3 capabilities (vPC, first-hop redundancy in N5K)
- Servers connectivity: CNAs/FCoE, and classical (NIC+HBA) scenarios
- SAN (iSCSI, FCoE and FC) and NAS deployment scenarios
- DC Services insertion (ACE appliance, WAAS, ASA, IPS, NAM ...etc)
- Server Access layer security for virtual machines
DC Design Details
Virtual Machine Security

- Provisioning and Security Policy for virtual machines tightly coupled with the Nexus 1000v
  
  Nexus 1000v provides for NX-OS port features at the connection of the Virtual Machine
  
  vPath provides a port level redirection capability to service appliances (VSG, vASA, vWAAS)
  
  VXLAN defines secure closed user groups amongst virtual machines
Cisco Nexus 1000V
Consistent Architecture across Hypervisors

Consistent architecture, feature-set, network services & Nexus 1010 ensures operational transparency across multiple hypervisors.
Cisco Nexus 1000V
What is vPATH?

- Intelligence build into Virtual Ethernet Module (VEM) of Cisco Nexus 1000V virtual switch (version 1.4 and above);

- vPATH has the following main functions:
  - Intelligent Traffic interception for Virtual Service Nodes (VSN): vWAAS & VSG;
  - Offload the processing of Pass-through traffic (from vWAAS, for instance);
  - ARP based health check;
  - Maintain Flow entry table.

- vPATH is Multitenant Aware
- Leveraging vPATH can enhance the service performance by moving the processing to hypervisor
- The VSN interested packets are redirected by vPATH by encapsulating the original frame and forwarding to service appliance
VSG and vASA
Policy Enforcement

1. VSG Security Policy leverages
   - Network Attributes
   - vCenter VM Attributes
   - Custom Attributes
   - Zones

NAT
IPSecVPN (Site-to-Site)
Default Gateway
DHCP
Static Routing
Basic Stateful Protocol
VM-FEX Portfolio
VMware, KVM and Hyper-V (future)
## DC Designs
### Summary of 100 - 300 ports

<table>
<thead>
<tr>
<th>Feature</th>
<th>Opt A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Server form factor</td>
<td>Rack/Blade</td>
</tr>
<tr>
<td>FEX &amp; FET</td>
<td>✓</td>
</tr>
<tr>
<td>Adapter-FEX &amp; VM-FEX</td>
<td>✓</td>
</tr>
<tr>
<td>vPC</td>
<td>✓</td>
</tr>
<tr>
<td>Multi-hop FCoE, iSCSI, NAS</td>
<td>✓</td>
</tr>
<tr>
<td>Unified Ports &amp; FC Targets</td>
<td>✓</td>
</tr>
<tr>
<td>Virtualized Security (vPath, VXLAN)</td>
<td>✓</td>
</tr>
<tr>
<td>Services (ACE, ASA, WAAS)</td>
<td>Compatible</td>
</tr>
<tr>
<td>MPLS, OTV &amp; LISP</td>
<td>ASR1000</td>
</tr>
</tbody>
</table>

- **Design Strengths:** Leverages a single Data Center switch to support ‘all’ storage and compute connectivity
- **Considerations:** May need to consolidate the LAN and Data Center L3 into a Nexus 5596 if budget constraints exist