Cisco Expo 2011

Построение Современных ЦОД

Эльдар Женсыкбаев

Системный инженер-консультант

ezhensyk@cisco.com

Апрель 2011

Cisco Expo 2011

Построение Современных ЦОД

Эльдар Женсыкбаев

Системный инженер-консультант

ezhensyk@cisco.com

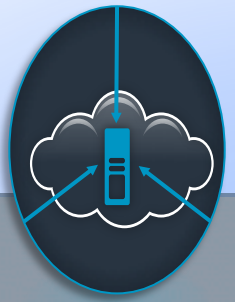
Апрель 2011

Содержание

- Эволюция ЦОД
- Эволюция LAN
  - Fabric Extender
  - VPC
  - Fabric Path
  - VM-FEX
- Унифицированная Сеть (Unified Fabric)
  - FCoE

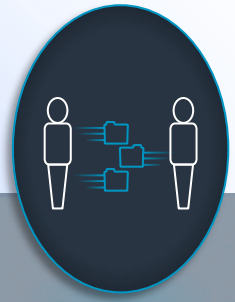
Основные Тренды, влияющие на ЦОД

Виртуализация
Серверов



Конвергенция
LAN и SAN

Поддержка
VM



Внедрение
Приложений

Доступность
Приложений

Drive for Green
Энергия,
Охлаждение, Место

Снижение Расходов
и/или Увеличение
Прибыли

IT как
Помощник
для Бизнеса

Консолидация Серверов, Виртуализация и новая Атомная Единица

Новая Парадигма

- Виртуальная Машина – новая “Атомная Единица”
- Динамическое передвижение VM / Приложений
- Новые возможности: VDI, Облака, Динамичность Нагрузки

Инфраструктура

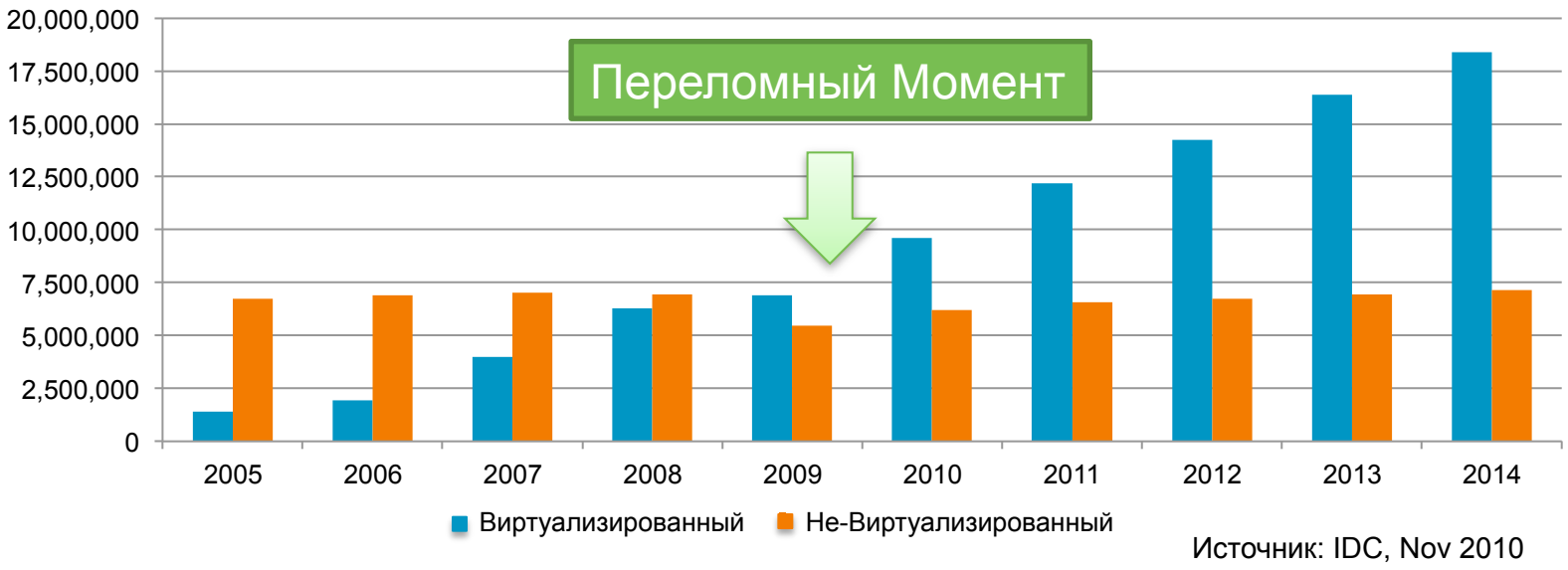
- Требуются сервисы для каждой VM
- Многоядерные CPU, Большая полоса B/B
- Новый акцент на Безопасность, QoS

Организация

- Ломается текущая Организационная Модель
- Снижает видимость ‘Скрытых’ Ресурсов
- Требует Непрерывной Доступности/Внедрения

Эволюция Архитектуры ЦОД

Разрушитель Технологий - Виртуализация



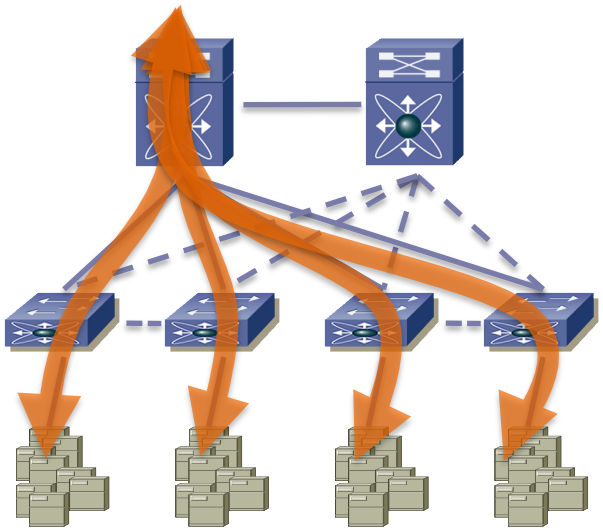
Источник: IDC, Nov 2010

Cisco Expo 2011

Эволюция LAN

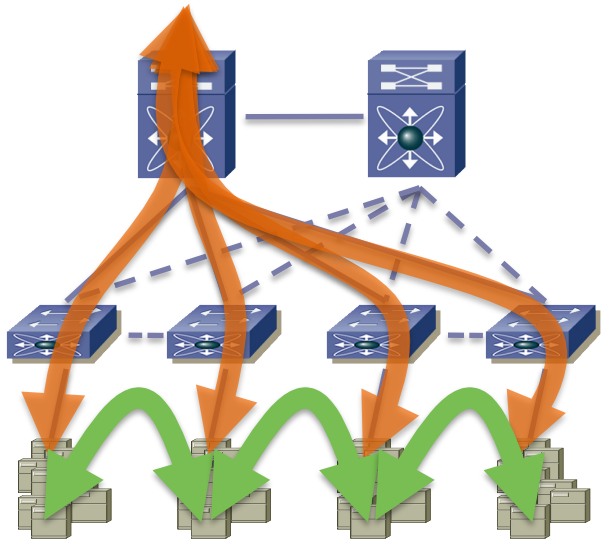
Допустимость переподписки?

Кампусная Сеть



- Основной трафик Север-Юг
- Переподписка приемлима для приложений клиент-сервер

ЦОД

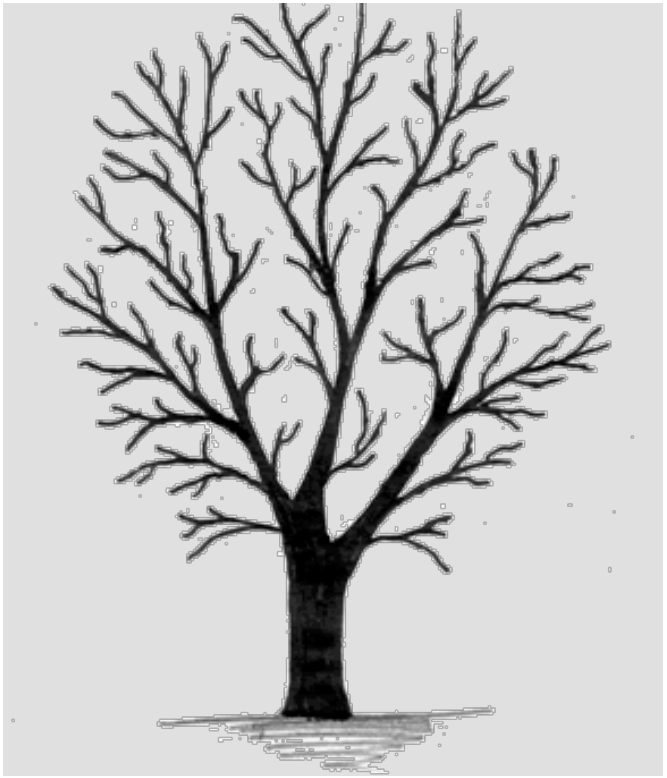


- Смесь трафиков Север-Юг и Восток-Запад
- Дизайн требует особого внимания для минимизации ограничений полосы пропускания из-за переподписки

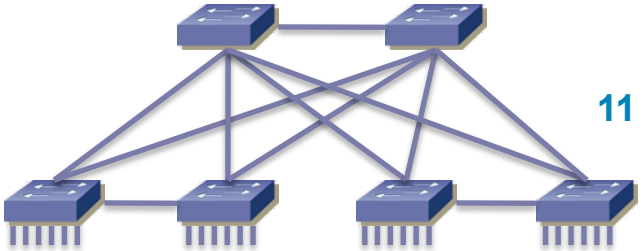
Требования к Сети L2 внутри ЦОД

- Максимизация полосы пропускания
- Масштабируемость L2 домена
- Высокая Доступность
  - Устойчивость control-plane
  - Быстрая сходимость
  - Изоляция доменов неисправности
- Внедрение Приложений с учетом особенностей
  - Мобильность приложений, Кластеризация, и т.д.
- Multi-Pathing/Multi-Topology

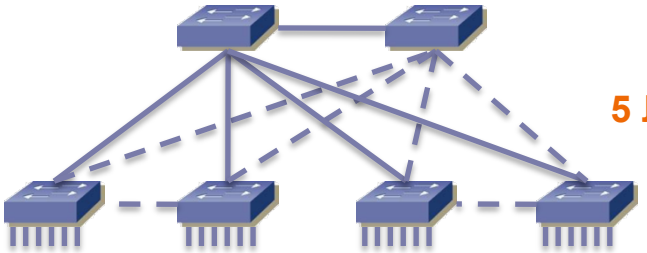
Spanning Tree и Переподписка



- Ветви дерева никогда не пересекаются (нет петель!!!)



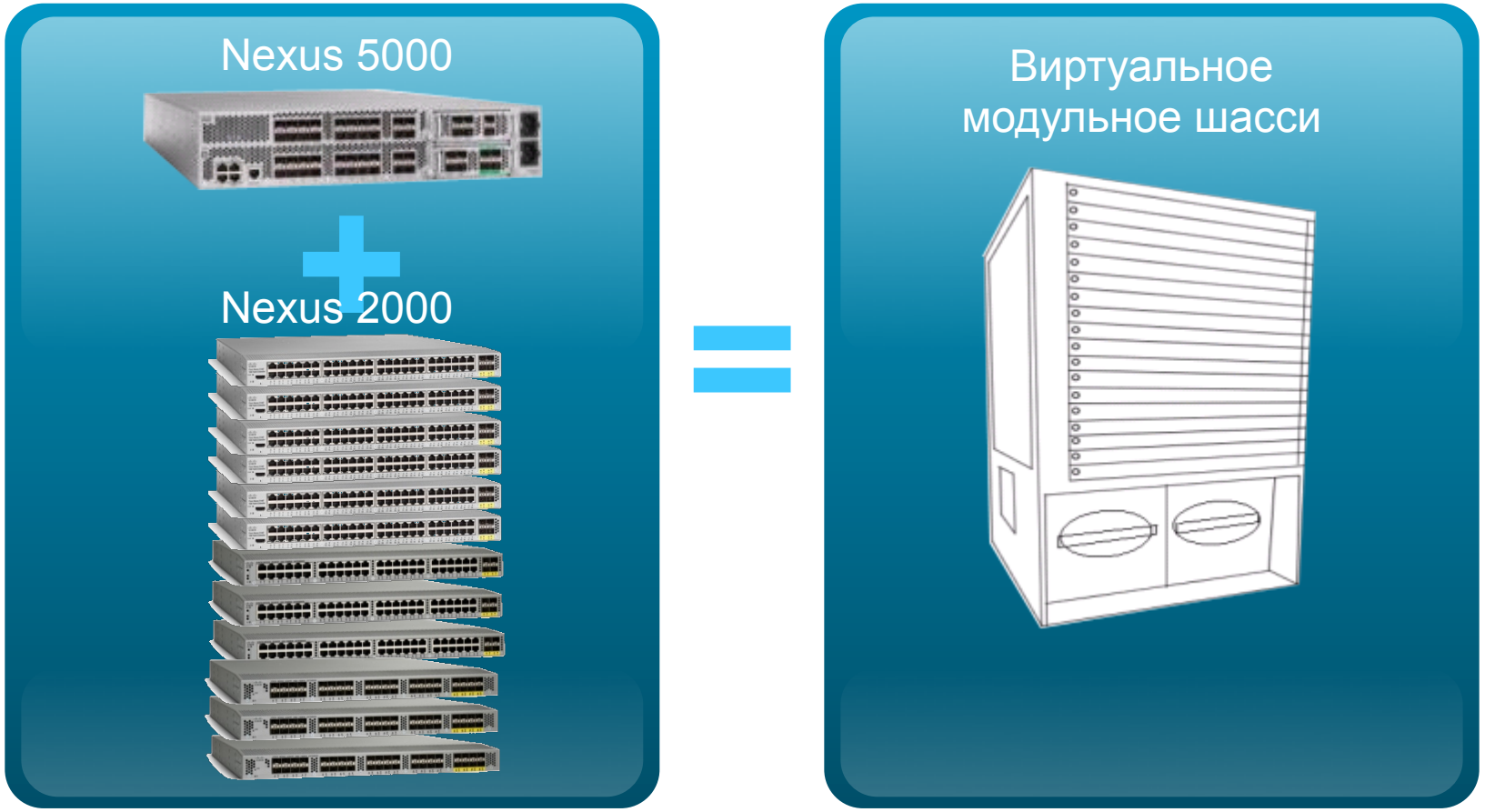
11 Физических Линков



5 Логических Линков

- Spanning Tree Protocol (STP) использует тот же подход для построения loop-free топологии L2
- **Переподписка усиливается STP**

Виртуальное модульное шасси Nexus 5000 + Nexus 2000

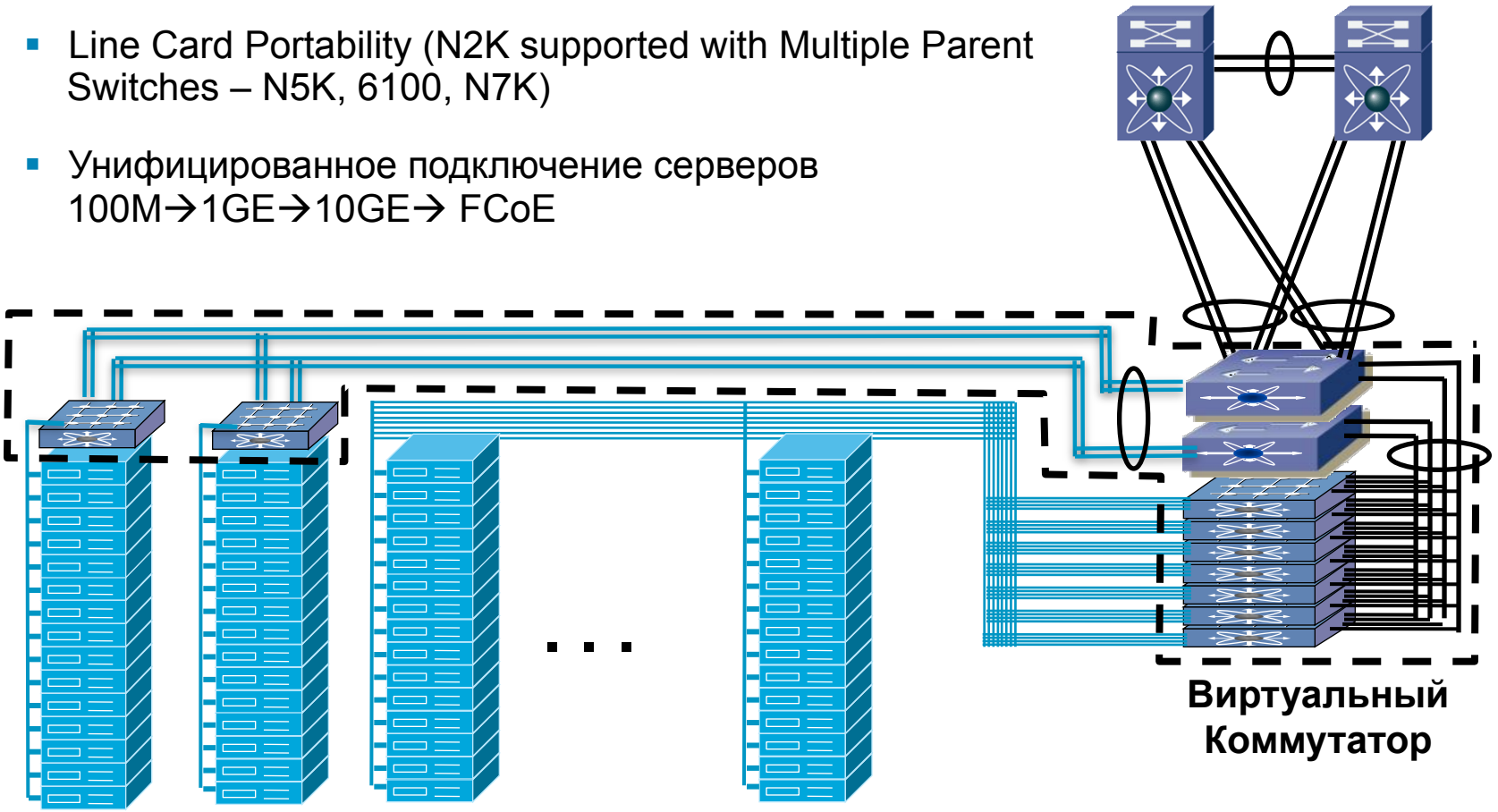


- Nexus 2000 FEX выполняет роль виртуальной удаленной карты для Nexus 5000/7000
- Единый конфигурационный файл на Nexus
- Между FEX и Nexus не используется STP

Fabric Extender (FEX)

Унификация архитектуры уровня доступа

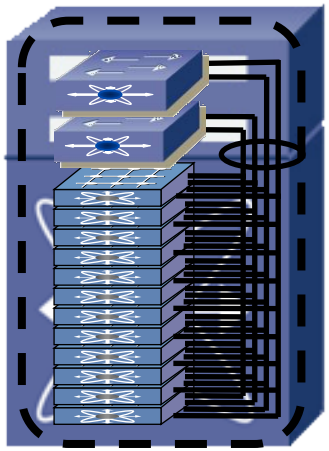
- Развязка Layer 1 и Layer 2 Топологий
- Упрощенное управление, plug-n-play внедрение, централизованная настройка
- Line Card Portability (N2K supported with Multiple Parent Switches – N5K, 6100, N7K)
- Унифицированное подключение серверов 100M→1GE→10GE→ FCoE



Виртуализированный Коммутатор

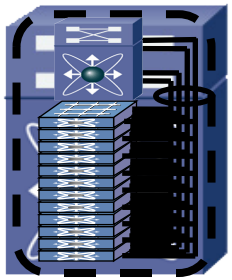
Основной Коммутатор ~= Supervisor

- Основной коммутатор Nexus обеспечивает коммутацию для всего Виртуализированного Коммутатора
- Апгрейд основного коммутатора изменяет возможности всего Виртуализированного Коммутатора

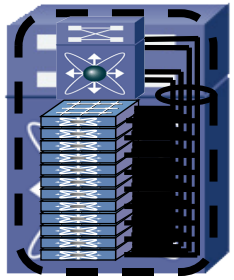


Nexus 5000
12 FEX, DCB, Layer 2 Ethernet, FCoE, FC

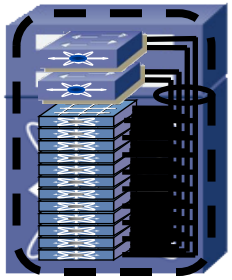
Замена Основного Коммутатора



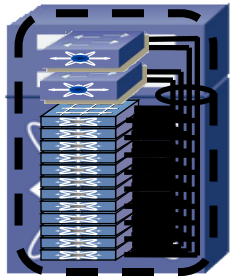
Nexus 7000
M1 Line Card - 32 FEX, Layer 2 & 3 Ethernet



Nexus 7000 w/F2 Line Card - DCB, Ethernet, FC, FCoE, NIV, Layer 2 & Layer 3, FabricPath (2HCY11)



Nexus 5500 16 FEX, DCB, Ethernet, FC, FCoE, NIV, Layer 2 & Layer 3, FabricPath



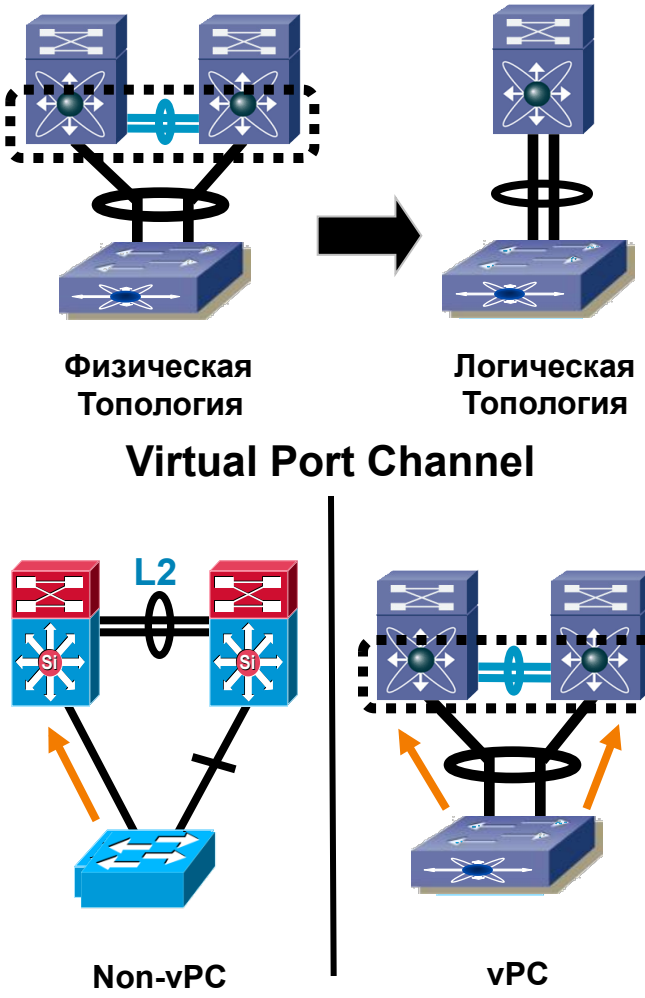
Будущие версии Коммутаторов

Архитектура Коммутации в ЦОД

vPC – MultiChassisEtherChannel

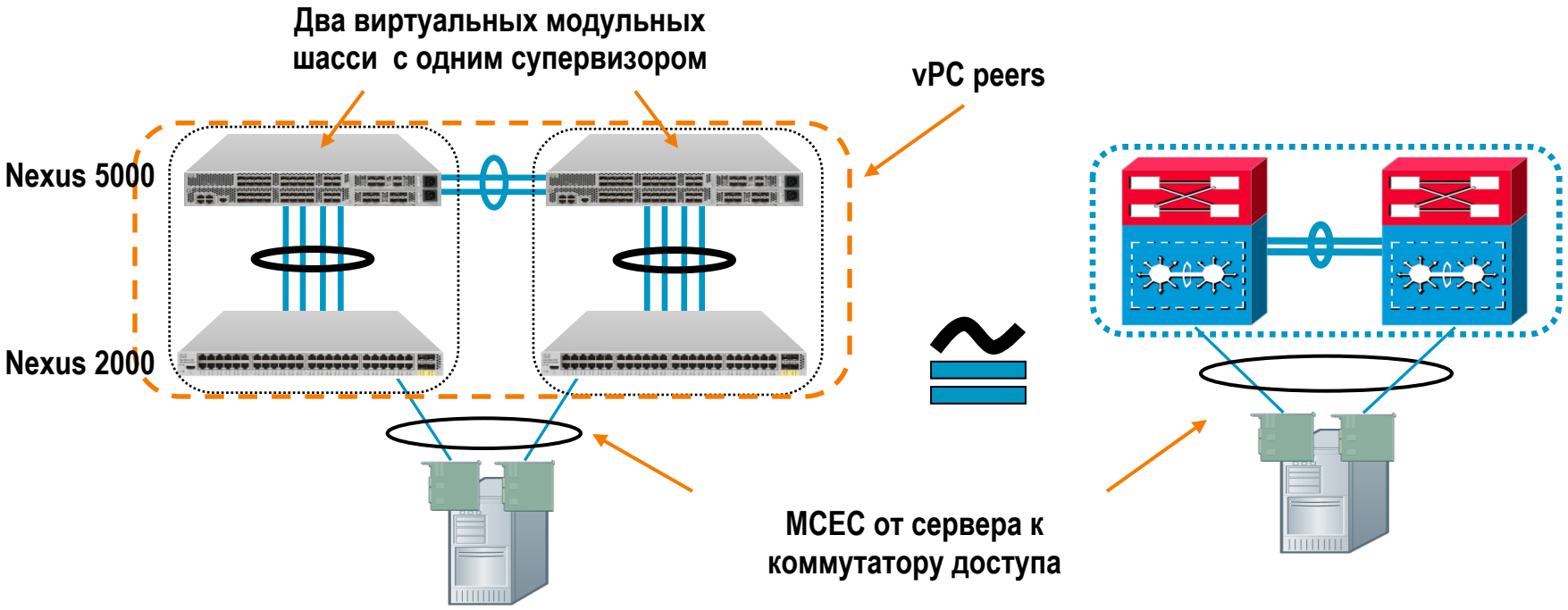
- Возможность организации агрегированного канала (port channel) приходящего на два разных коммутатора
- Позволяет создавать устойчивые L2 топологии, используя Агрегацию Линков
- Использование полосы всех имеющихся соединений

  Все линки активные
- vPC сохраняет независимость control plane



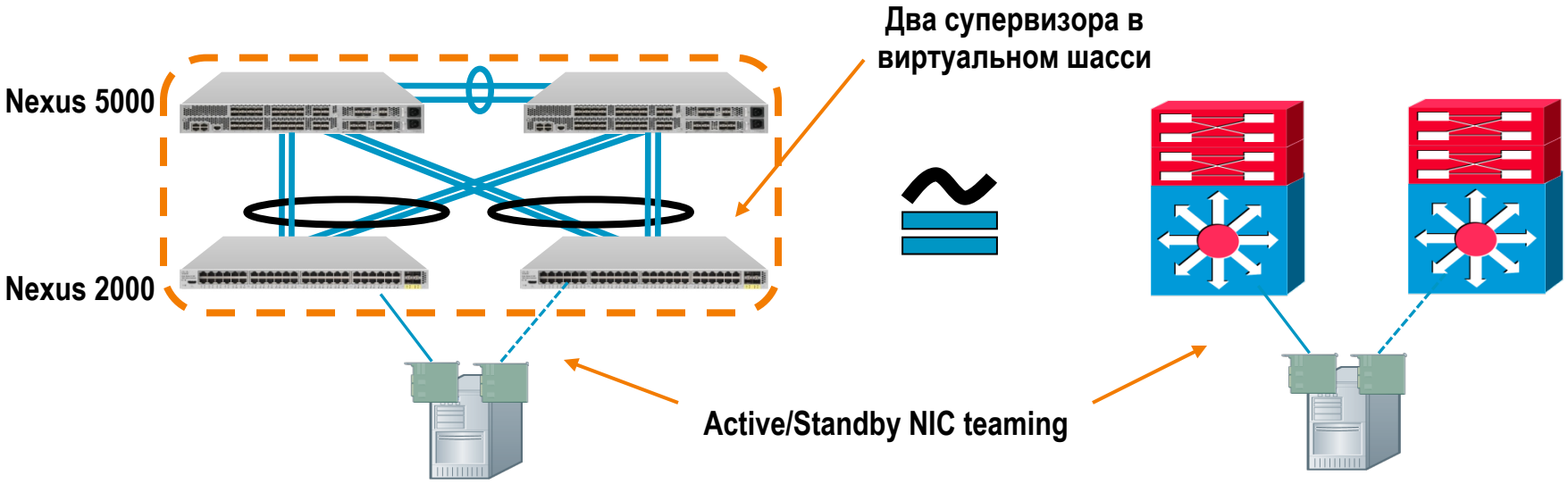
Резервирование с помощью vPC
Два шасси

- Вариант 1 – Multi-chassis EtherChannel для подключения сервера
- Nexus 2000 подключен к одному Nexus 5000
- Защита от отказа супервизора, интерфейсной карты, кабеля, NIC

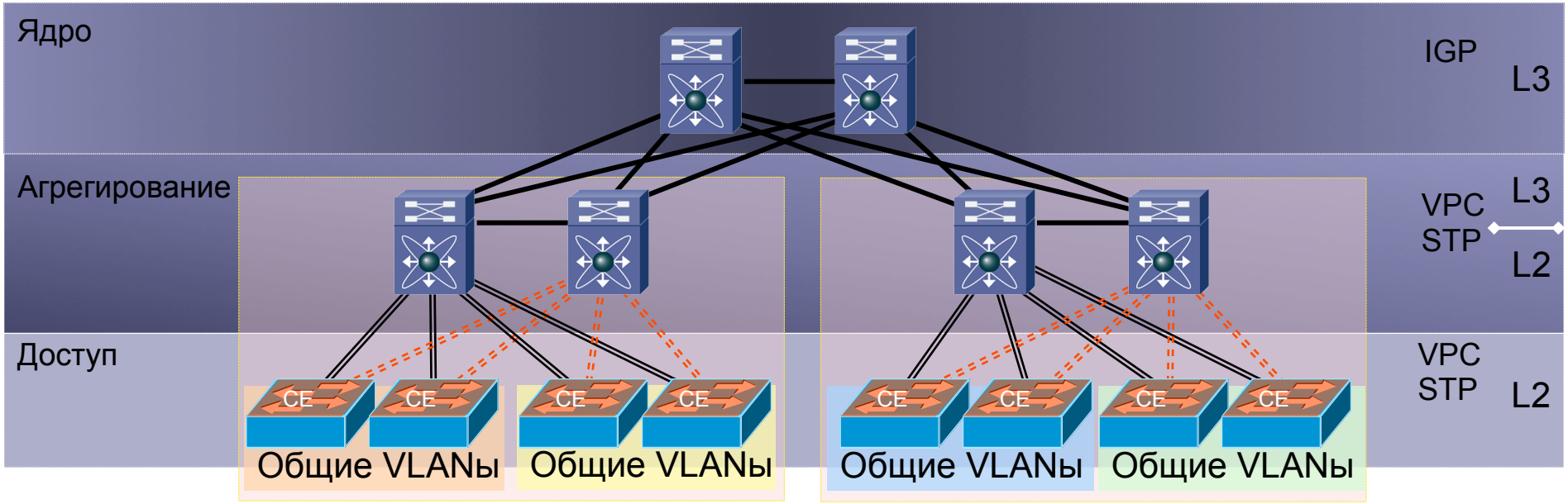


Резервирование с помощью vPC
Два супервизора

- Вариант 2 – Nexus 2000 подключены к двум Nexus 5000
- Сервер подключен к коммутатору с резервным супервизором
- Защита от отказа супервизора, фабрики
- При использовании Active/standby NIC Teaming – защита от отказа интерфейсной карты, кабеля или NIC



Традиционная иерархическая модель



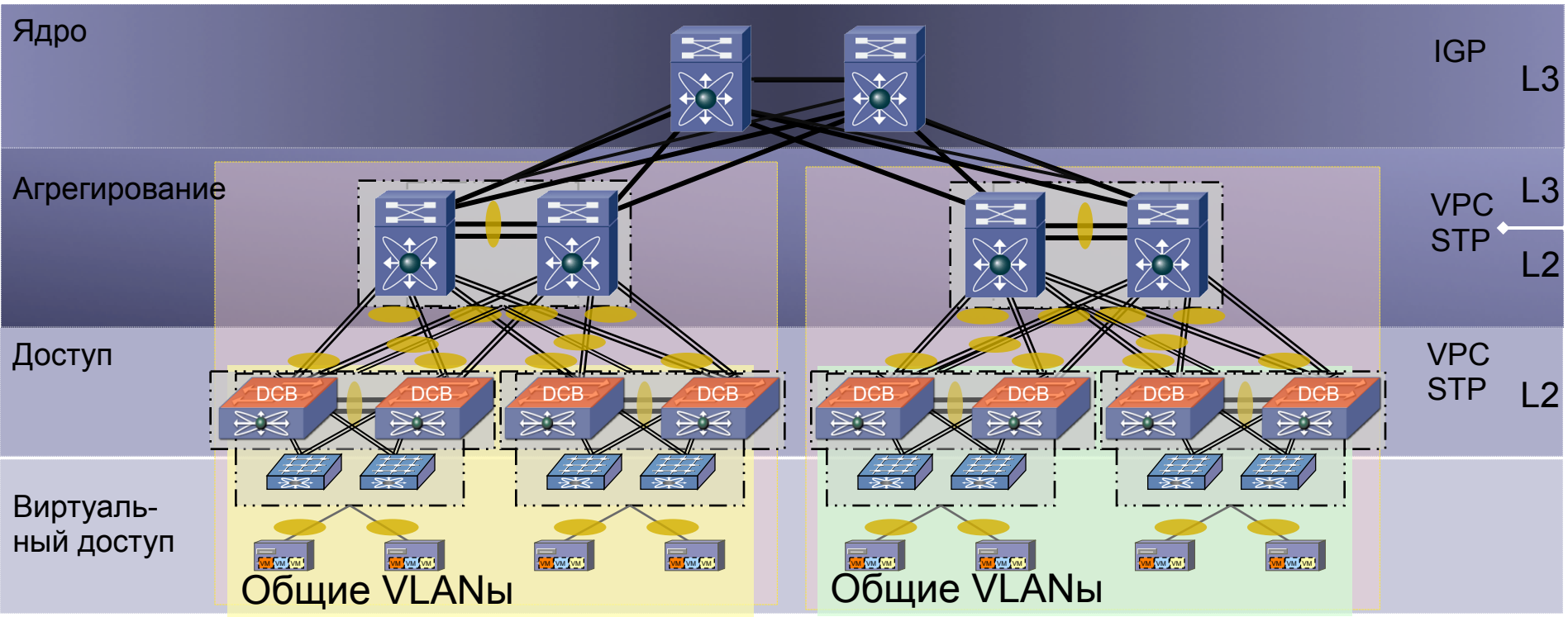
- Классическая сеть с STP
- Модульный дизайн:

  Воспроизводимые сетевые модули с предсказуемым масштабом и производительностью: блоки («POD»)

  Каждый уровень архитектуры сети реализует свои функции: упрощается проектирование, эксплутация и диагностика

  Размер определяется спецификой задачи и требованиями приложений: стандартные приложения, Grid Computing, VDI, SAN, потребность расширения ЦОД и т.д.

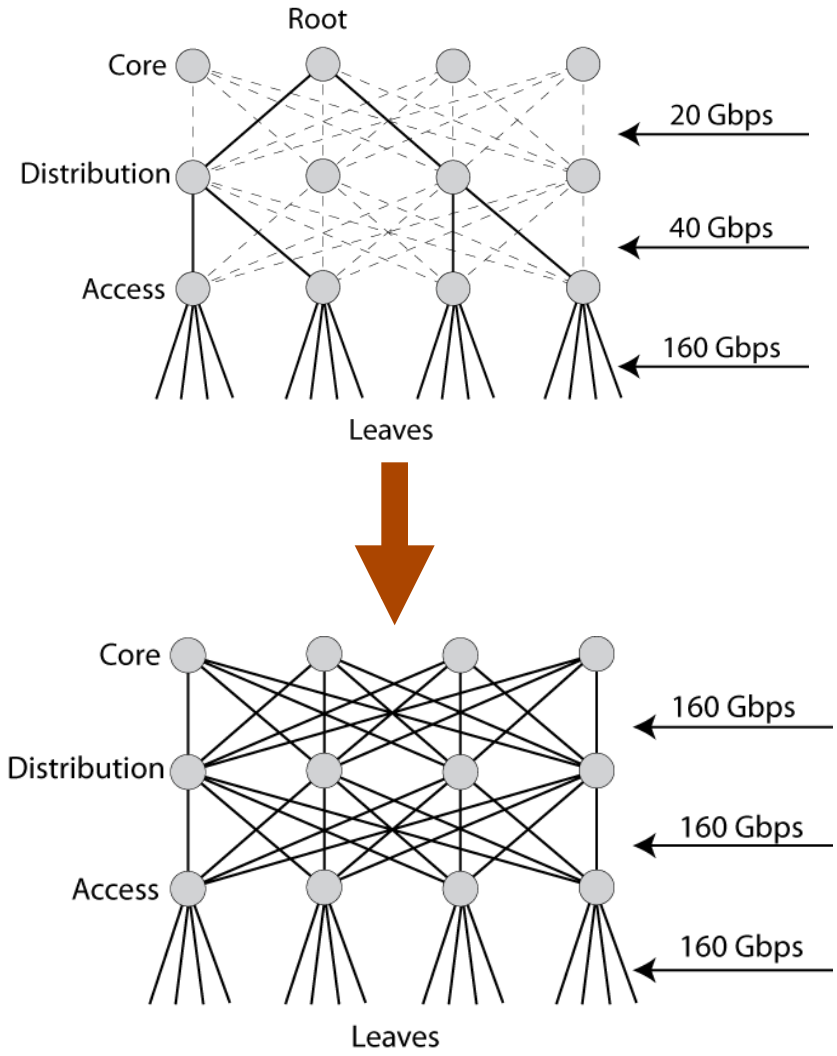
Развитие иерархической модели



- Что стало возможным сейчас
- Модульный дизайн:
  - Снижение «переподписки»- увеличение доступной полосы
  - Увеличение размера блока с ростом производительности и плотности портов
  - Уход от опоры на STP, сокращение числа точек управления
- Эволюция технологий:
  - VDC, VPC, Fabric Extenders, FCoE & DCB (PFC, ETS, DCBX)

Потребность в L2MP: STP превращает многосвязную сеть в дерево

- Нужно альтернативное решение, позволяющее:
  - Задействовать все соединения
  - Наращивать производительность путем увеличения числа связей
  - Принципиально исключить возможность бесконечных «петель»
  - Обеспечить быструю и надежную сходимость
- ...и всё это – для коммутации на втором уровне!



Представляем Cisco FabricPath

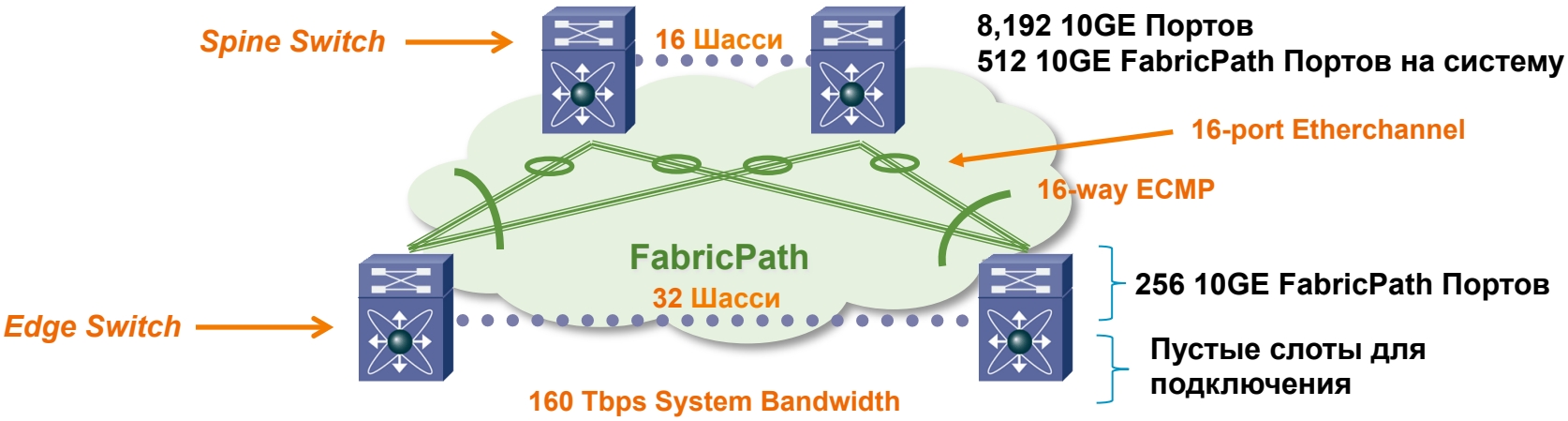
Инновации NX-OS для Layer 2 Сетей

"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

T · · Systems ·

Cisco FabricPath
Архитектура

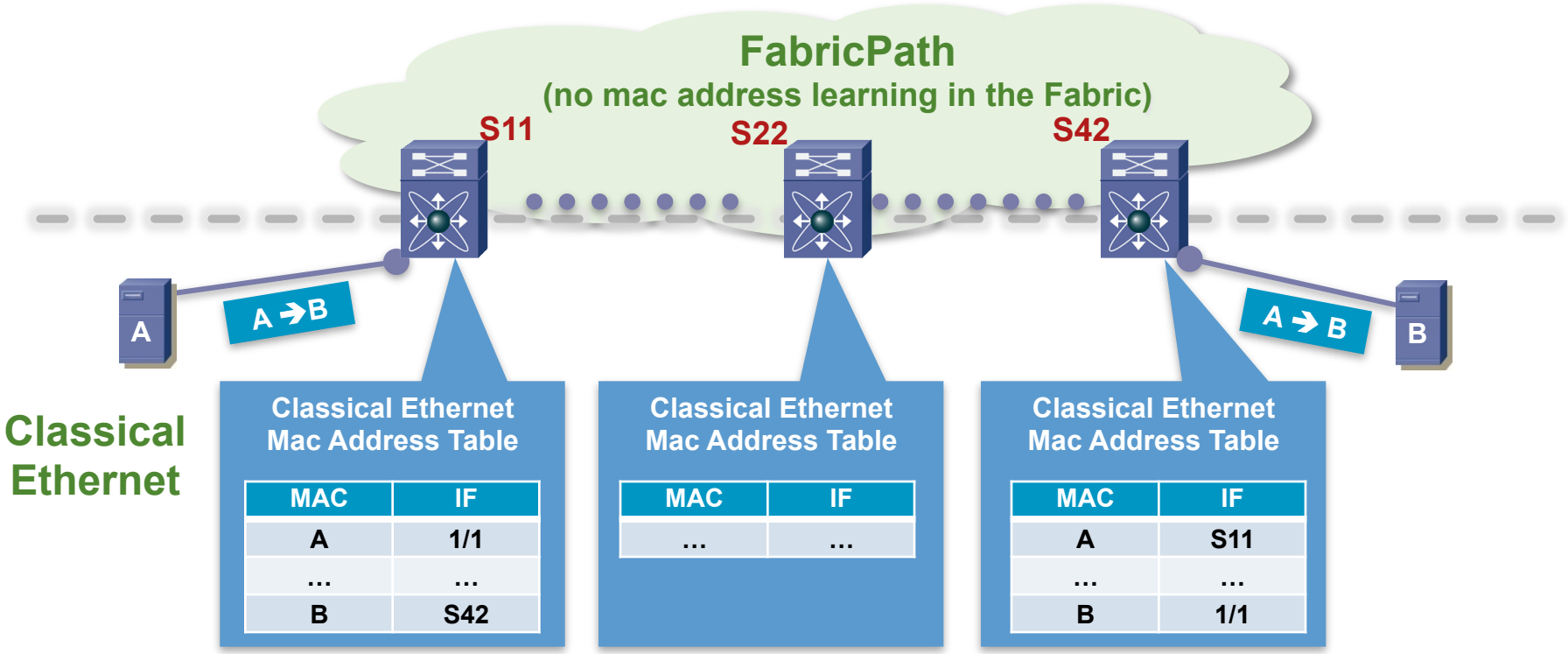


Преимущества FabricPath

- Неблокируемый транспорт за счёт FabricPath ECMP и агрегирования портов (port-channel)
- Устраняет недостатки Spanning Tree
- Высокая устойчивость, быстрое схождение сети
- Любой VLAN, в любой точке фабрики

Масштабируемость FabricPath

- Заголовок FabricPath: иерархическая адресация со встроенным предотвращением «зацикливания» (RPF, TTL)
- Выучивание MAC «по диалогам»: эффективное использование аппаратных ресурсов

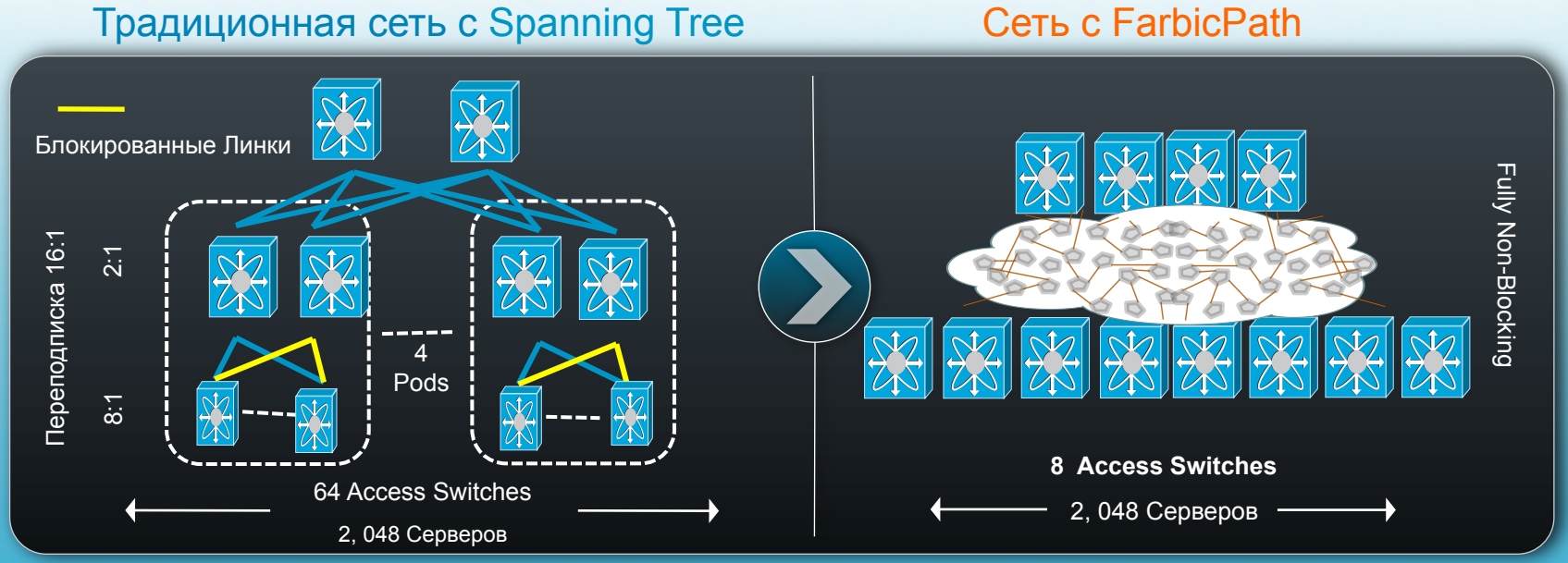


Масштабируемость FabricPath

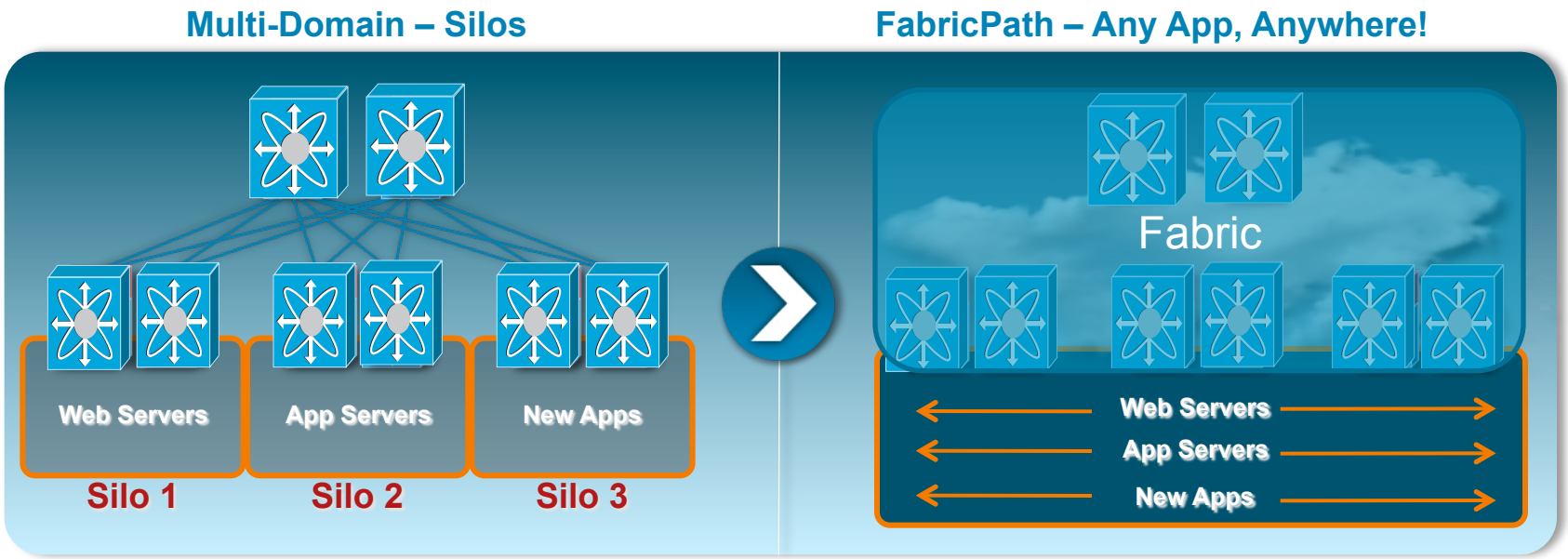
Пример: Дизайн 2,048 X 10GE Серверов

- 16X улучшение пропускной способности
- От 74 управляемых устройств до 12

  Меньше энергии, меньше задержки
- 2X+ увеличение доступности сети
- Упрощение задач IT и гибкость внедрения приложений

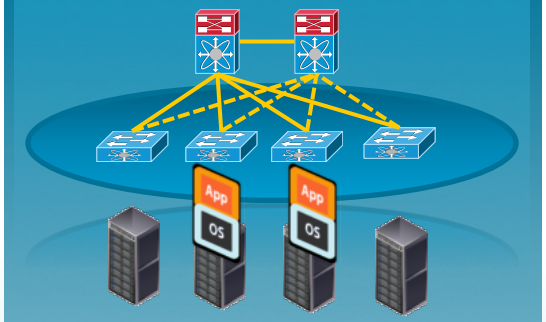
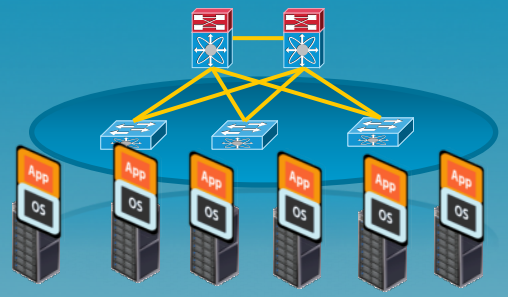
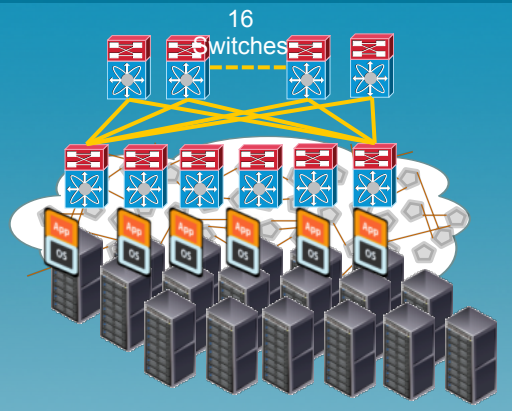


FabricPath – Простота снаружи



- Сеть выглядит как один Коммутатор → Мобильность приложений, максимальная гибкость
- Снижение OPEX путем упрощения работы администраторов серверов → Снижение зависимости от администраторов сети

Архитектурная Гибкость с NX-OS

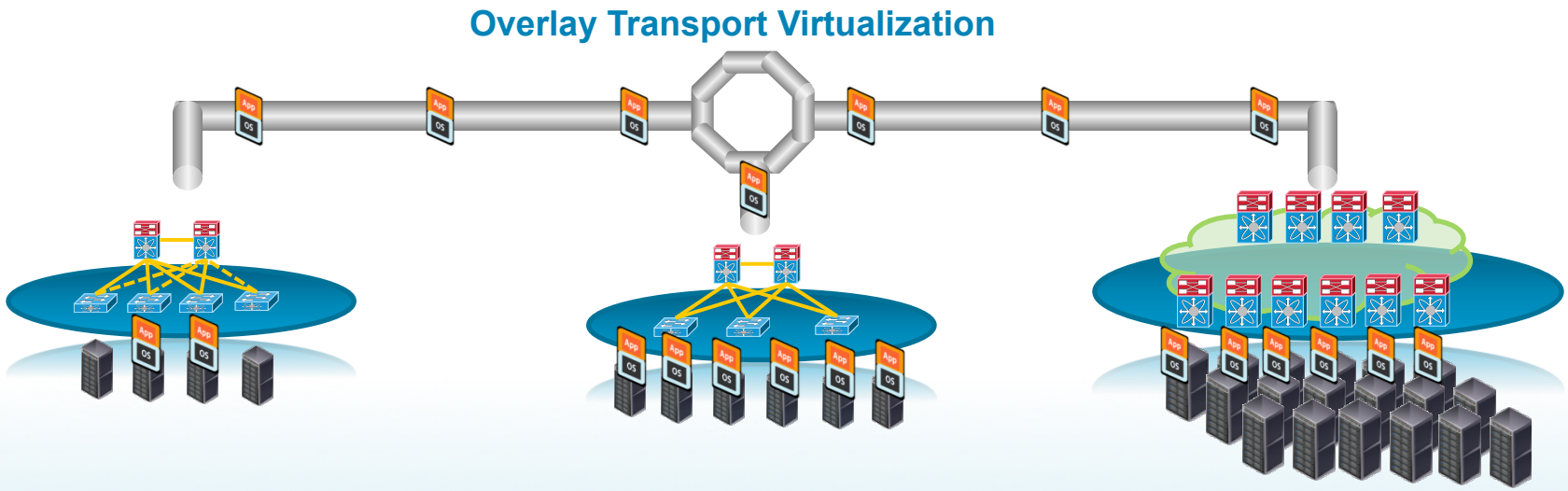
Spanning-Tree	vPC	FabricPath
		
Активных Путей	Single	Dual	16 Way
POD. Полоса пропускания	До 10 Gbps	До 20 Gbps	До 160 Tbps

Масштабируемость Layer 2

Виртуализация и Емкость Инфраструктуры

Архитектурная Гибкость с NX-OS

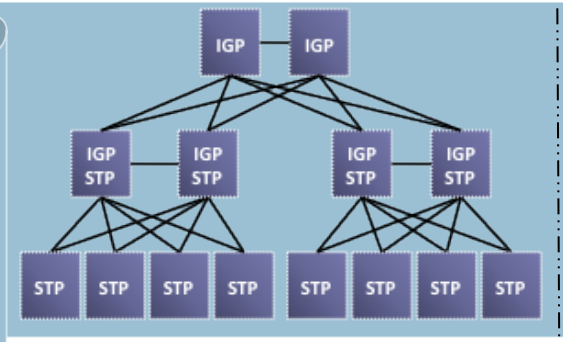
Общая инфраструктура для разных сценариев



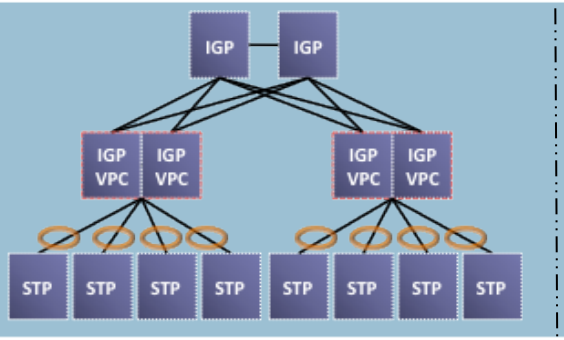
OTV – Мобильность Приложений

FabricPath – Производительность и Гибкость Приложений

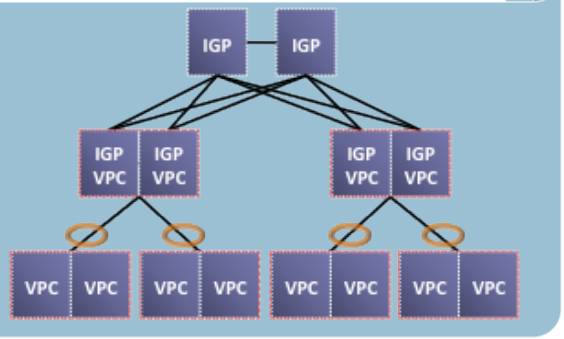
Развитие топологий сетей ЦОД



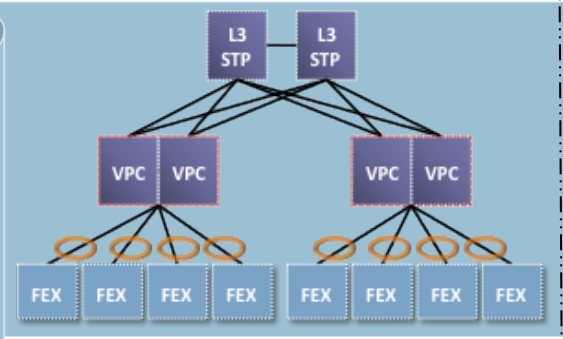
Classic Ethernet



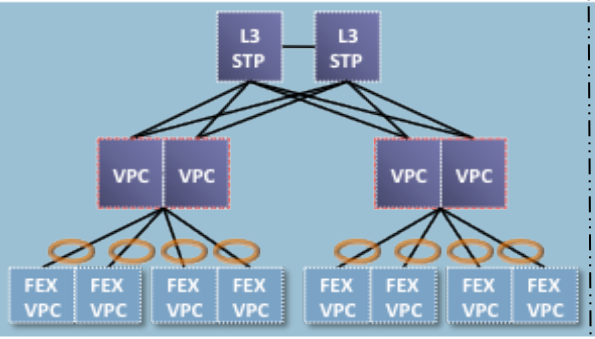
VPC на уровне агрегирования



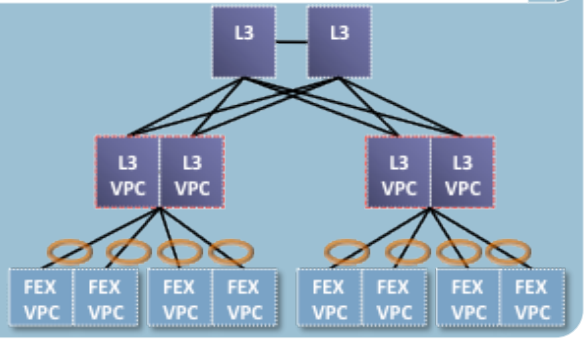
VPC на уровне агрегирования и доступа



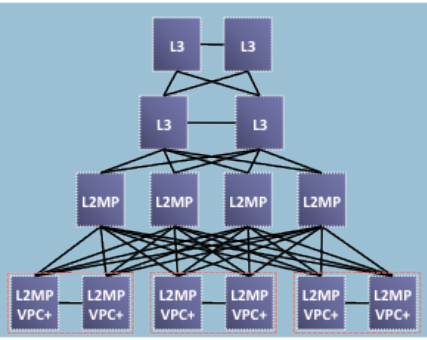
FEX



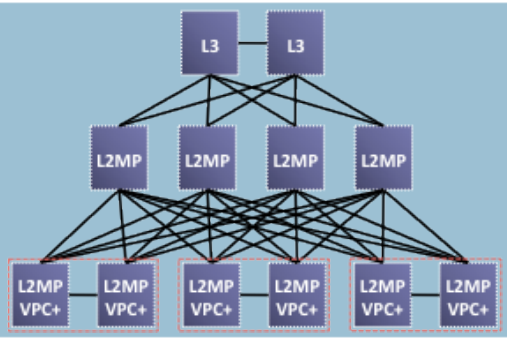
VPC до FEX



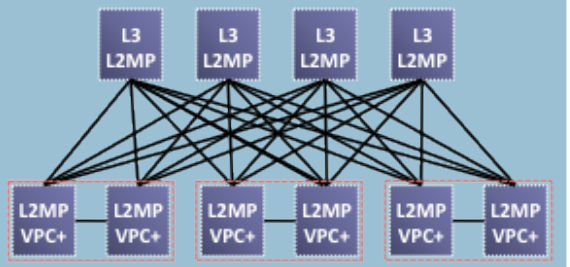
L3 доступ и VPC до FEX



L2MP подключение к L3 сети



L2MP подключение к «коллапсированной» L3 сети



Совмещение агрегирования и Spine уровня

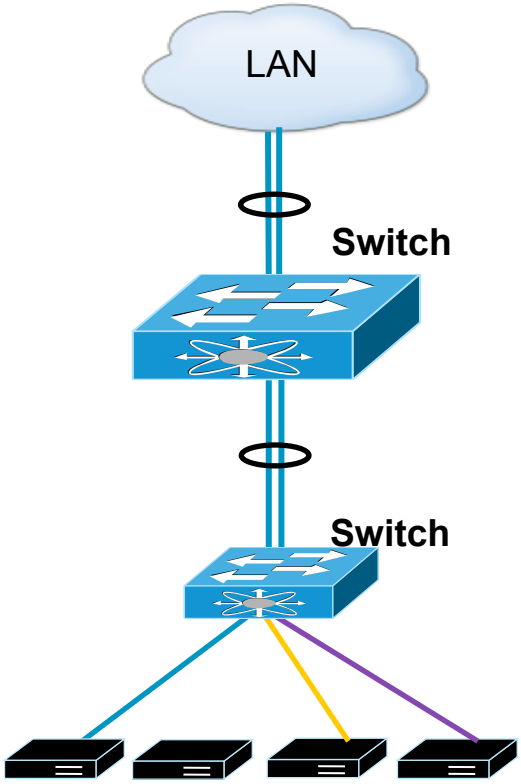
Cisco Expo 2011

VM-FEX

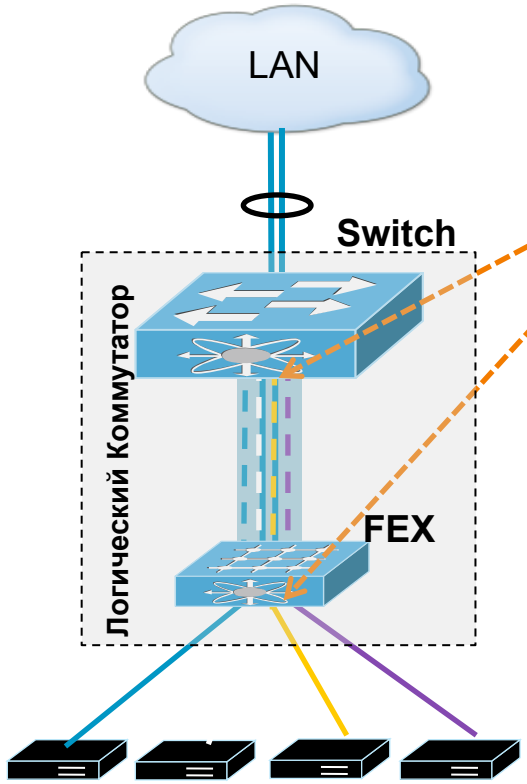
Концепция Cisco Fabric Extender

Port Extender (Pre-standard 802.1Qbh)

Устаревшая архитектура



FEX архитектура

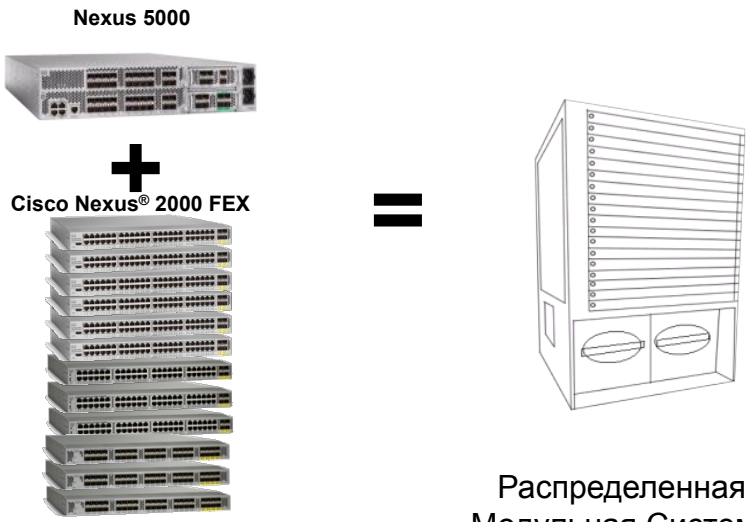


Порты коммутатора продлеваются через Fabric Extender

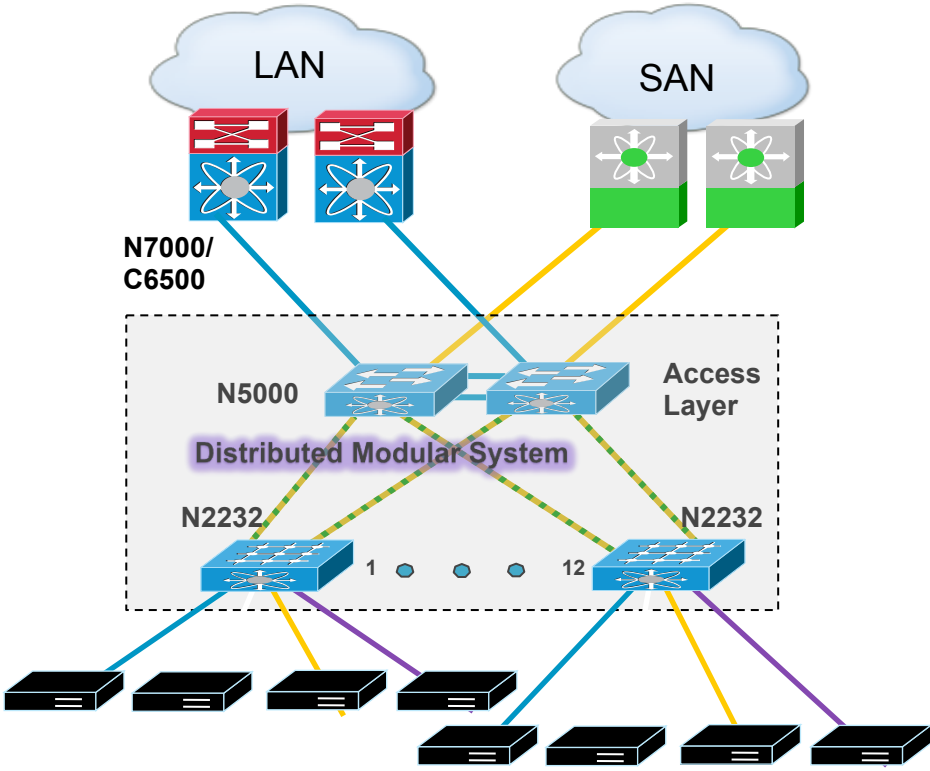
Объединяя уровни сети, снижаем число точек управления!!!

Nexus 5000 + FEX

Единый уровень доступа

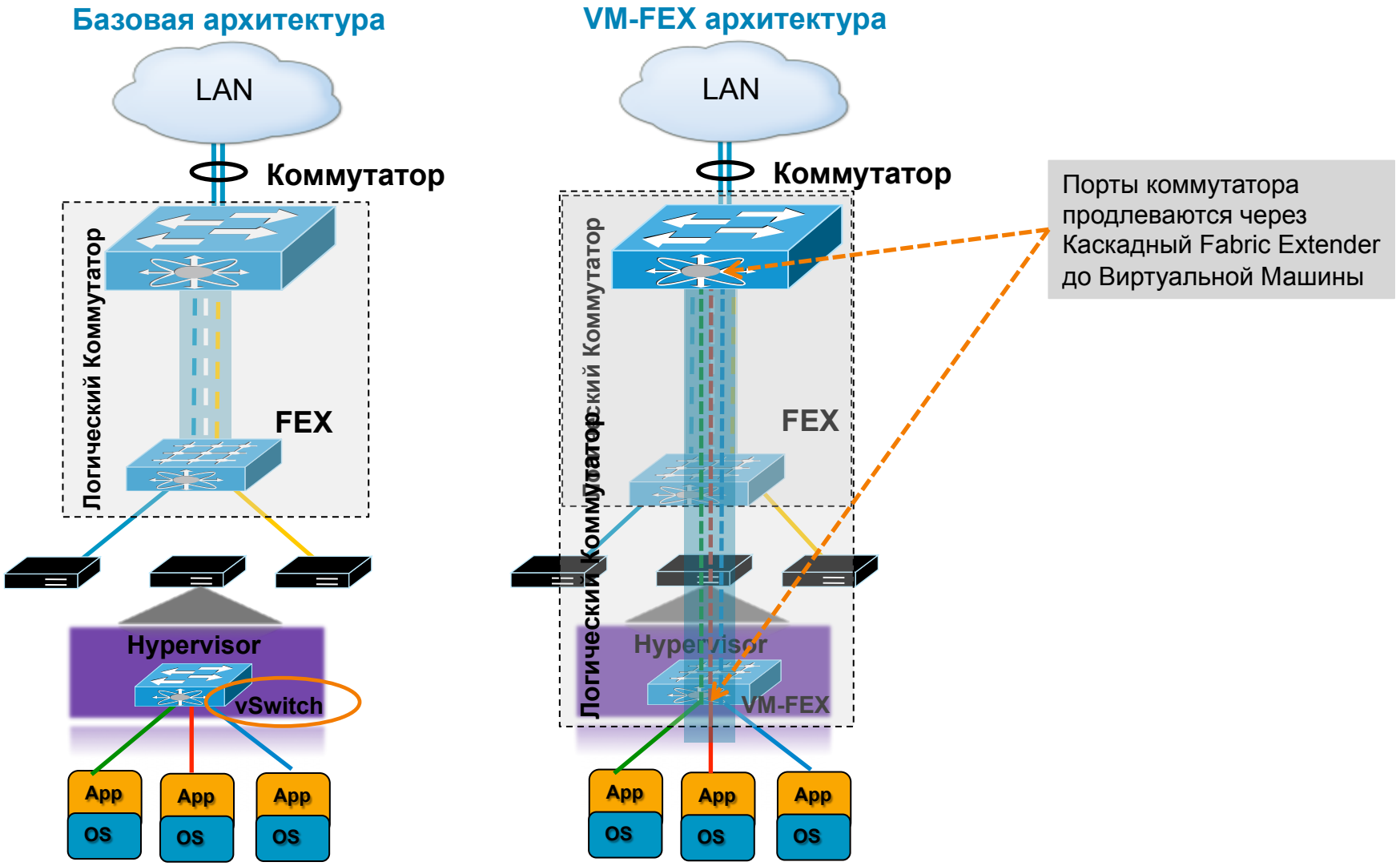


Распределенная
Модульная Система



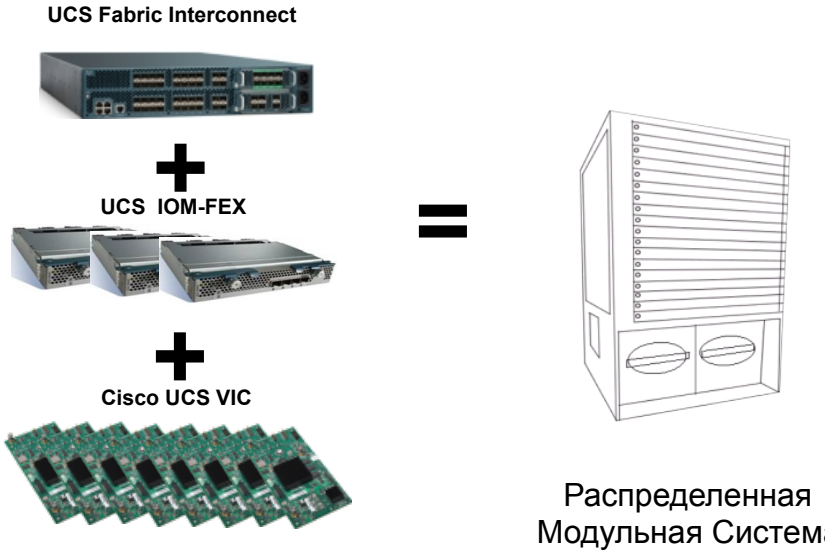
Расширяем FEX архитектуру до VM

Каскадируем Port Extender (Pre-standard 802.1Qbh)

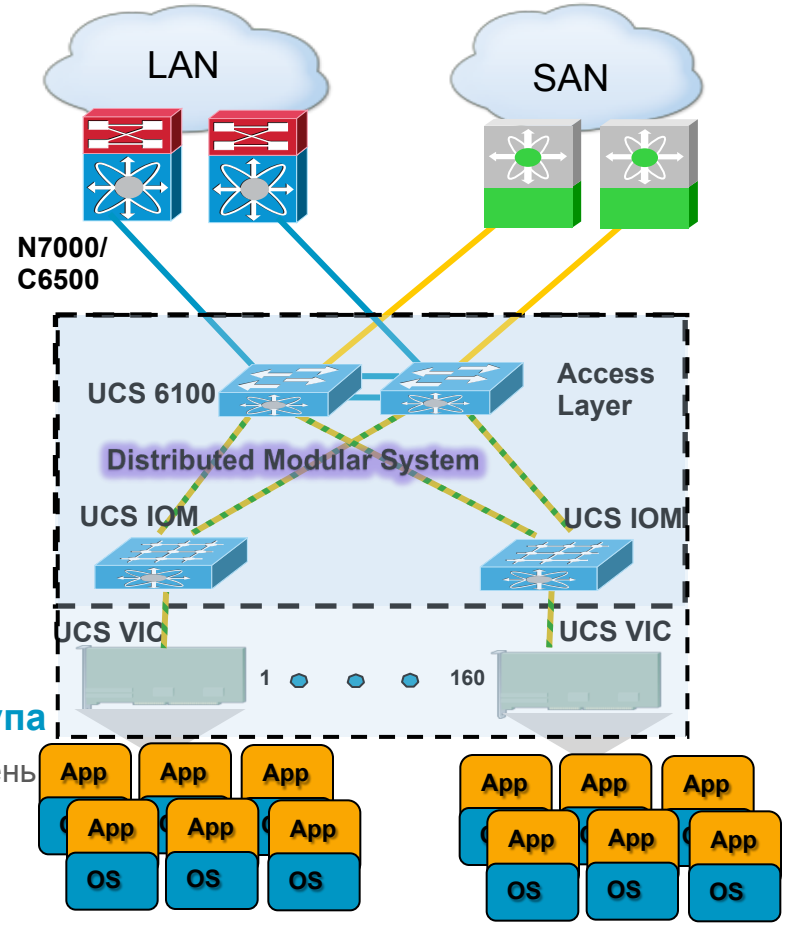


UCS VM-FEX

Расширяем FEX Архитектуру на Виртуальный Уровень



Распределенная
Модульная Система



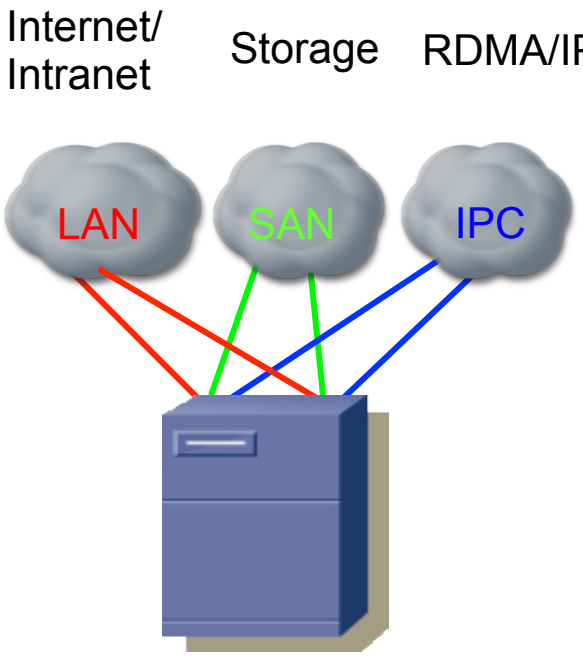
VM-FEX: Единый Виртуально-Физический Уровень Доступа

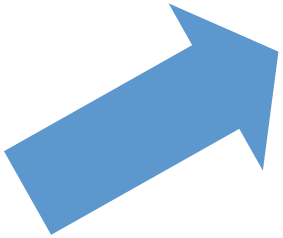
- Объединяем виртуальную и физическую коммутацию в единый уровень доступа
- VIC Виртуальная Линейная Карта для UCS Fabric Interconnect
- Fabric Interconnect содержит все настройки и управление
- Виртуальный и Физический трафик рассматривается одинаково

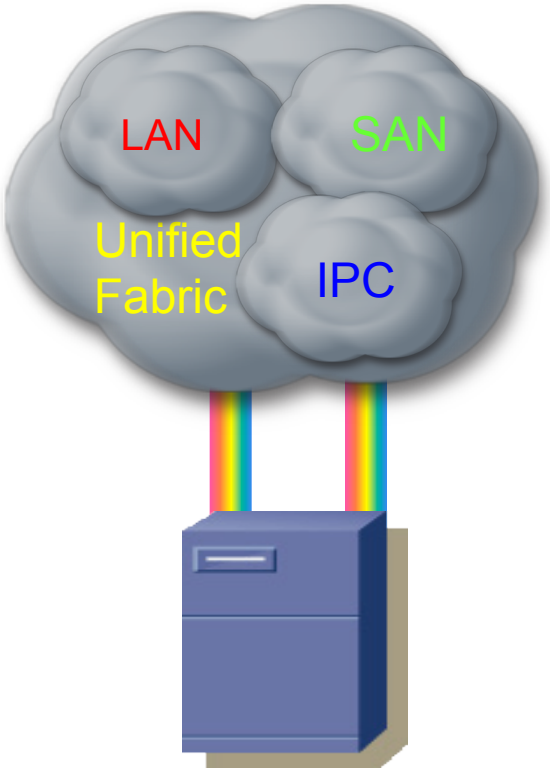
Cisco Expo 2011

Унифицированная Сеть

Консолидация ввода-вывода: Объединенный транспорт FCoE/IEEE DCB







Сегодня

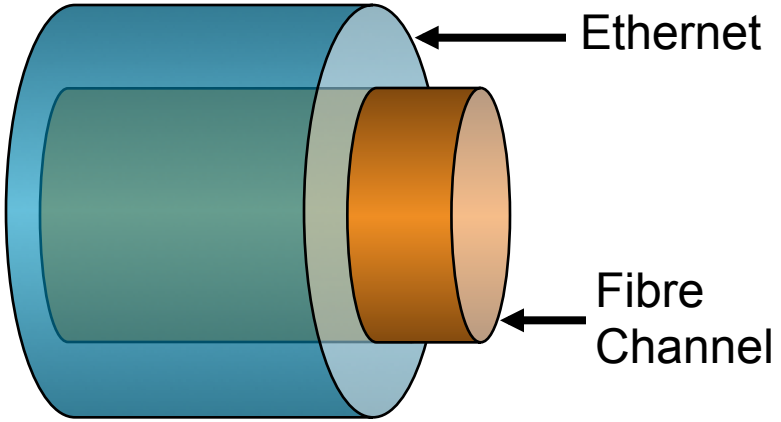
- Много портов ввода-вывода
- Высокие расходы на оборудование и эксплуатацию

Используя DCB

- Общий транспорт
- Сокращение числа кабелей, портов, адаптеров
- Обеспечение совместимости

Fibre Channel over Ethernet (FCoE)

- Метод передачи фреймов FC по Ethernet
  - Выглядит как FC для серверов и сети
  - Сохраняет текущую инфраструктуру и управление FC
  - Фрейм FC остается неизменным

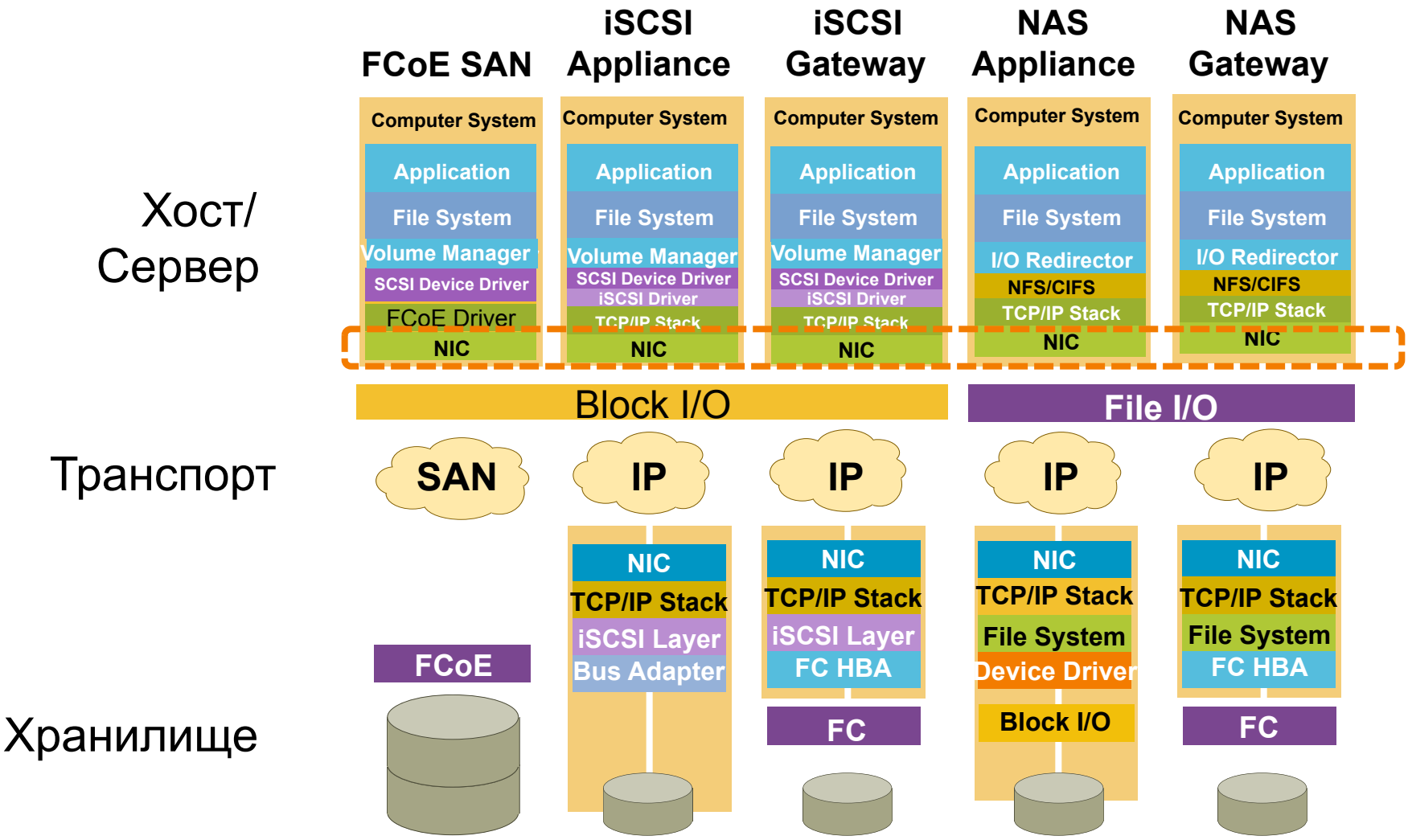


- Может работать на стандартных коммутаторах (с jumbo фреймами)
- Priority Flow Control обеспечивает отсутствие потерь
  - Имитирует систему буферных кредитов FC
- Стандарт утвержден 3 июня 2009 года (ANSI T11 FC-BB-5)

Cisco первой представила основанный на стандартах коммутатор FCoE Cisco Nexus 5000

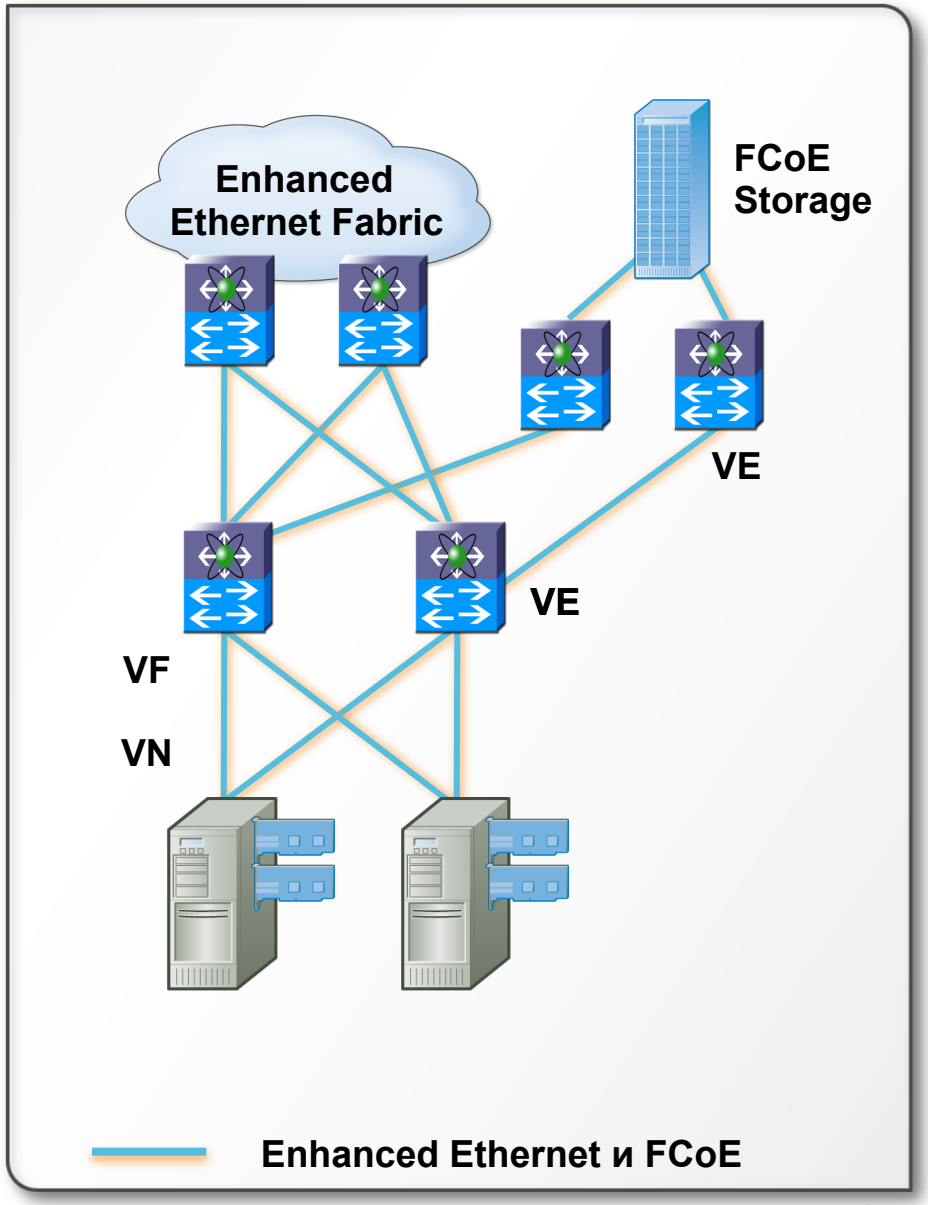
Почему FCoE?

Данные передаются по общей фабрике



FCoE: консолидация в масштабах сети!

- Расширение консолидации ввода-вывода на магистраль
- Поддержка разделения SAN фабрик для отказоустойчивости
- Поддержка систем хранения с подключением по DCB/FCoE



Продукты Cisco с поддержкой FCoE

Фиксированная конфигурация

Nexus 5596

Nexus 5020

Nexus 5548

Nexus 2232

Nexus 5010

Nexus 4000

Класс Директоров

Nexus 7000

MDS 9500

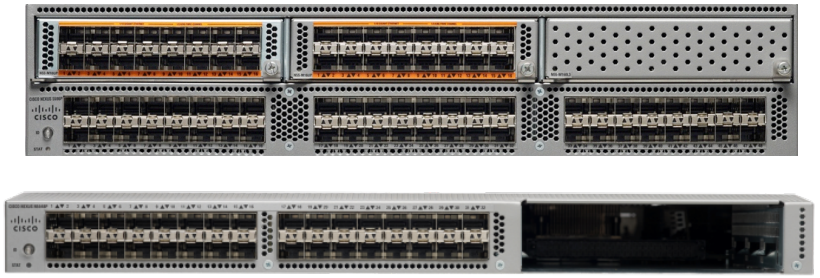
F1 32-port

FCoE 8-port

NX-OS & DCNM

Nexus 5596UP & 5548UP
Новое поколение уровня доступа ЦОД

- Коммутаторы ToR высокой плотности (1RU / 2RU)
- 10GE / 1GE / FCoE / 8G FC



Инновации

- Универсальные порты
- Поддержка маршрутзации
- Increased FEX scale (24/L2)
- FCoE Multi-hop
- Adapter-FEX (будущее)
- Cisco FabricPath (будущее)
- VM-FEX (будущее)

Преимущества

- Защита инвестиций!
- Продолжение семейства Nexus 5000, использование NX-OS
- Низкая предсказуемая задержка коммутации

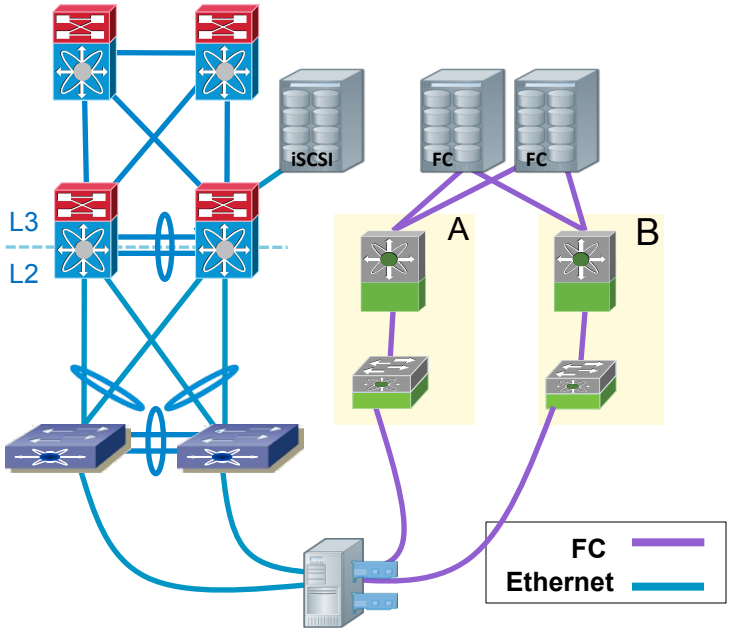
Изоляция SAN A / SAN B

Чего требуют заказчики?

Физическая изоляция

Что-то промежуточное

Логическая изоляция

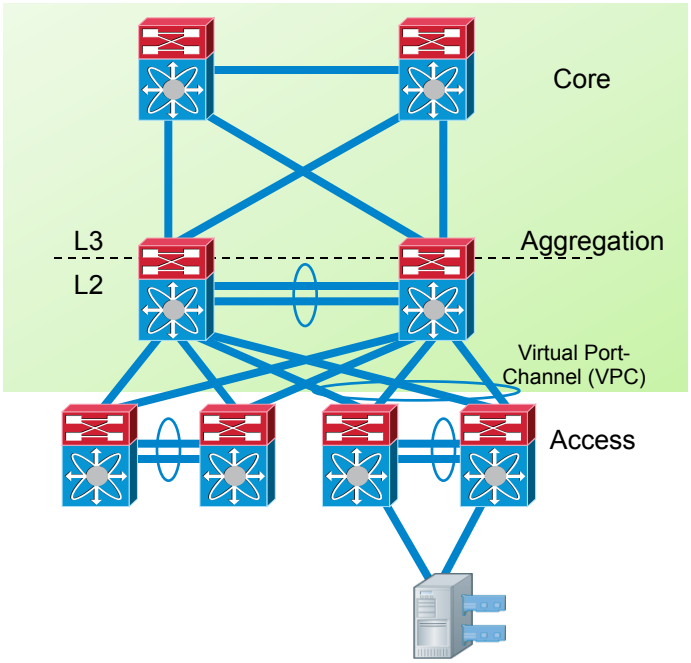


Изоляция на уровне...

Коммутатора

VDC

Кабеля

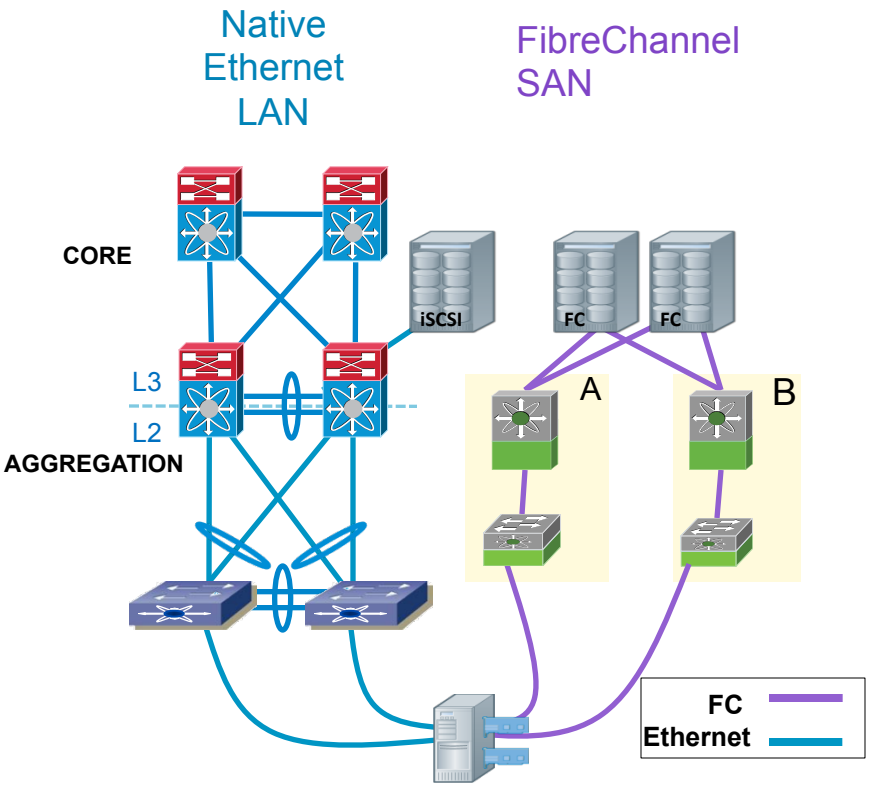


Отдельные физические сети для каждой фабрики

VLAN и VSAN для изоляции фабрик на едином наборе устройств и соединений

Традиционные сети
Выделенные LAN и SAN

- Полное разделение сетей
  - Отдельные домены эксплуатации
  - Неэффективное масштабирование
- Ethernet – основная технология передачи данных
  - Допускает потери
- Fibre Channel – стандарт для корпоративных SAN
  - Отсутвие потерь
  - Высокая доступность за счёт изоляции фабрик
- Увеличение числа интерфейсов сервера
  - Минимум 2xHBAs + 2xNICs !!

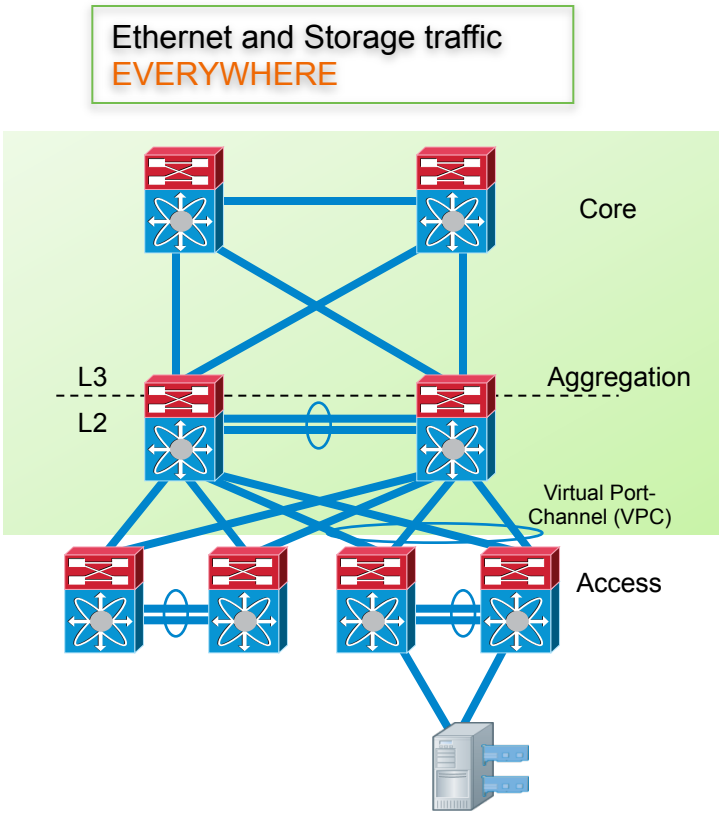


Полная консолидация транспорта
Другая крайность

- Единая сеть ЦОД
- Все соединения несут все виды трафика

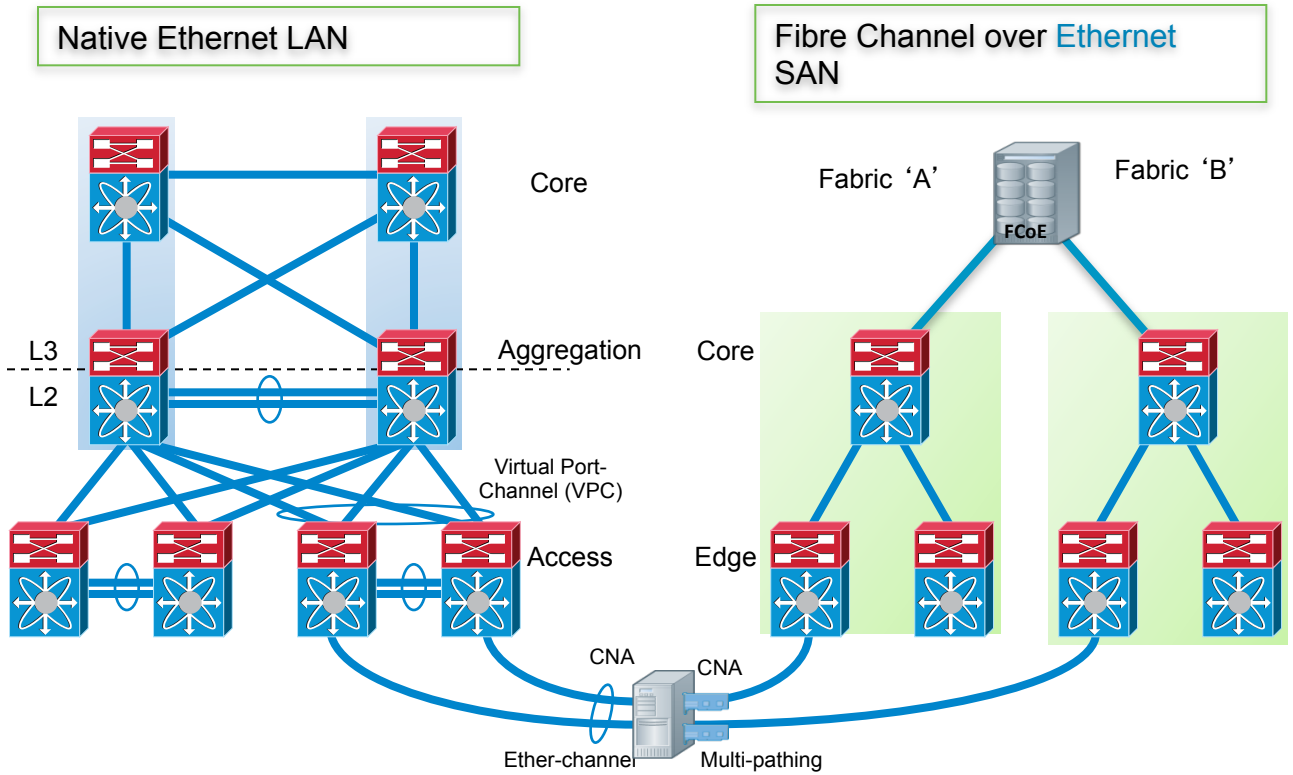
  Протоколы передачи и хранения данных
- Максимальное сокращение числа устройств и соединений
- Уход от фабрик A/B

  Единая фабрика с резервированием сервисов



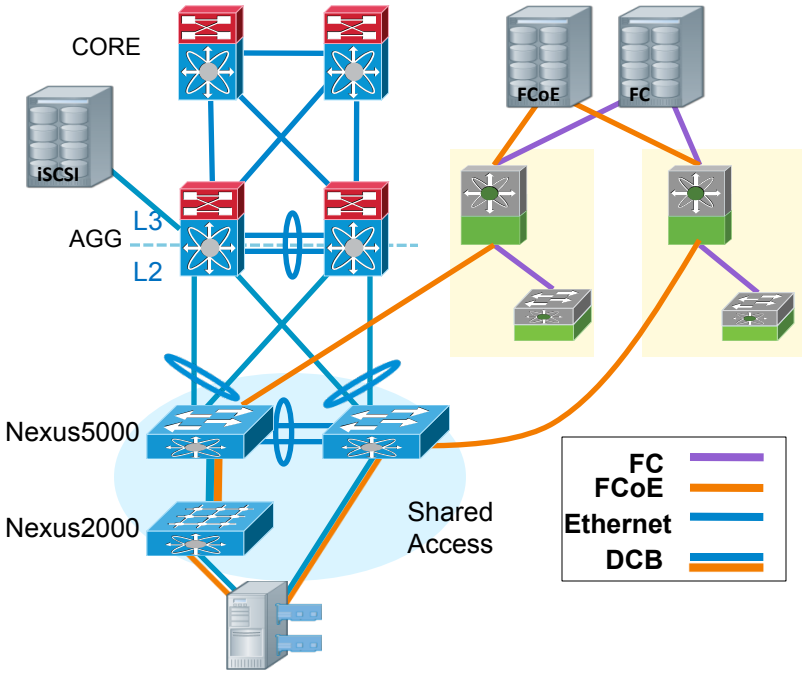
Ethernet SAN
Консолидация технологий

- LAN и SAN используют те же элементы: коммутаторы, кабели, адаптеры, трансиверы...
- Полная изоляция построения и эксплуатации
- Использование эволюции Ethernet (10G→40G→100G)



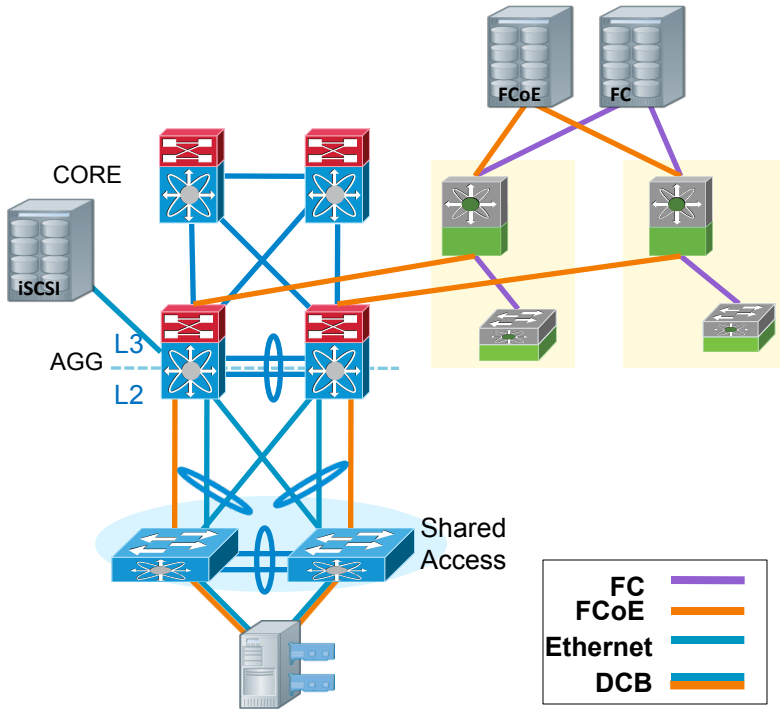
Консолидация на уровне доступа

- Единая СКС и топология доступа
- Подключение к существующим SAN
  - К FC сети или к FCoE по VE портам
- Плавная миграция к конвергентное сети
- Сохранение изоляции SAN фабрик
- Обеспечение совместимости с использованием NPV режима



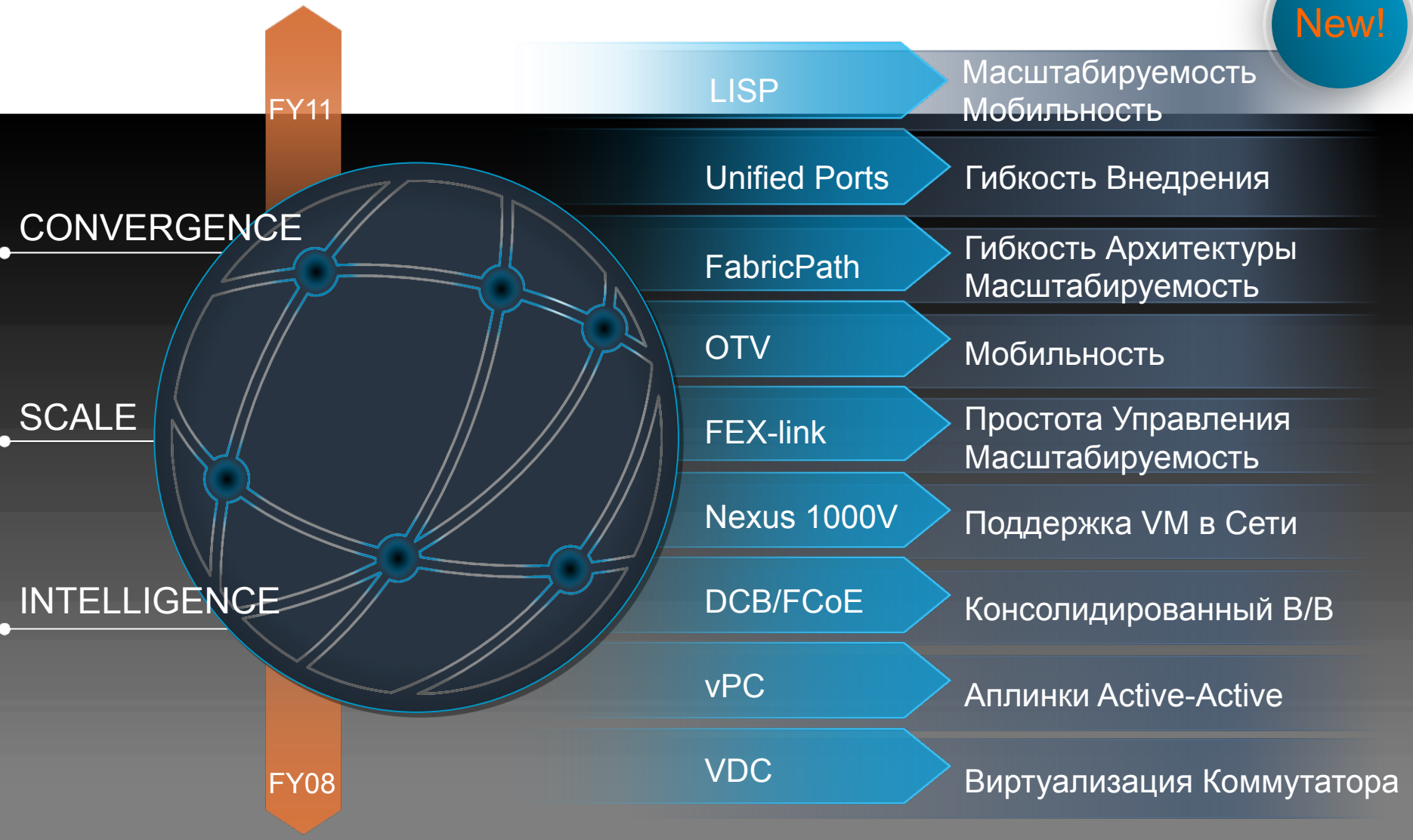
Консолидация на уровне агрегирования

- Транспорт сети хранения через высокопроизводительный конвергентный коммутатор высокой доступности
- Аналог традиционных SAN топологий «Edge-Core-Edge»
- Повышение производительности на уровне агрегирования
  - Построение крупных FCoE сетей с сохранением доступа к СХД с подключением по Fibre Channel
- Сохранение изоляции для трафика хранения
  - Выделенные FCoE соединения
  - Выделенные Storage VDC на Nexus 7000



Cisco Unified Fabric

Архитектурные Инновации



Cisco Expo 2011

Практика построения собственных ЦОД Cisco Systems

Немного об IT службе Cisco...

- 300 объектов в 90 странах
- 400 зданий
- 52 ЦОД и серверных комнат
  - –14: бизнес-приложения
  - –38: разработка
- 65,000+ сотрудников
- 20,000 м2 площадей ЦОД
- 20+ MW мощности до оборудования

Более 20,000 м2 площадей и 20 MW мощности в ЦОД Cisco

Развитие ЦОД Cisco

В перспективе на ~FY13

Global Disaster Recovery Strategy

Short to Mid-Term: Leverage Current Assets

Long-Term: Part of Decision Process (Make/ Buy)

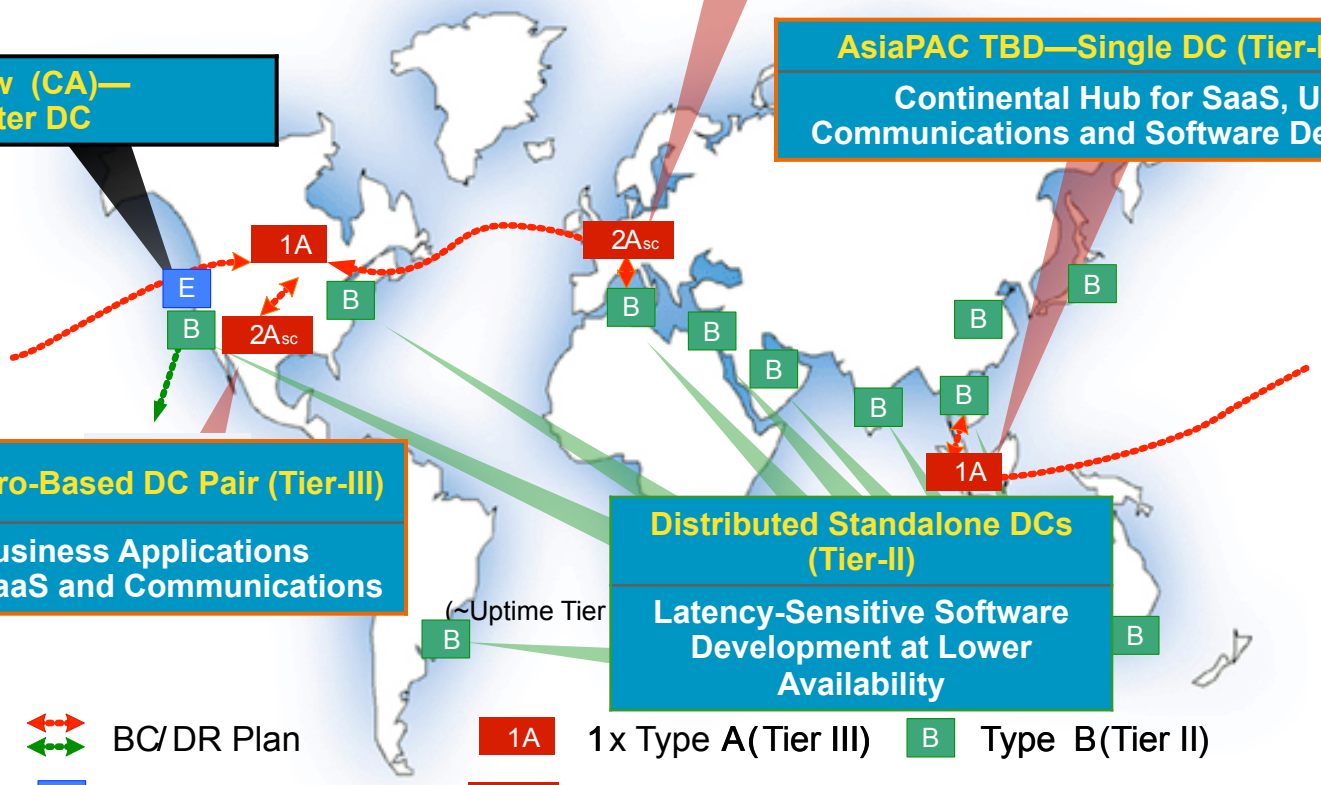
Netherlands— Metro-Based DC Pair (Tier-III)

Single-Instance Order Management (OM/AR)

Mountain View (CA)— Early Adopter DC

AsiaPAC TBD—Single DC (Tier-III) + Land

Continental Hub for SaaS, Unified Communications and Software Development



Richardson (TX)—Metro-Based DC Pair (Tier-III)

Global Hub for Business Applications
Continental Hub for SaaS and Communications

Distributed Standalone DCs (Tier-II)

Latency-Sensitive Software Development at Lower Availability

BC/DR Plan

1A 1x Type A (Tier III)

B Type B (Tier II)

E Early Adopter DC

2Asc 2x Type A at Synchronous Capable Distance

Новый сервис-ориентированный ЦОД Cisco Systems в Ричардсоне, Техас

Цель: Обеспечение текущих и будущих нужд бизнес-приложений Cisco

Питание и охлаждение: самый большой рабочий ЦОД Cisco, ~3,000 квадратных метров в зонах ЦОД ; два резервных ввода по 10 MW

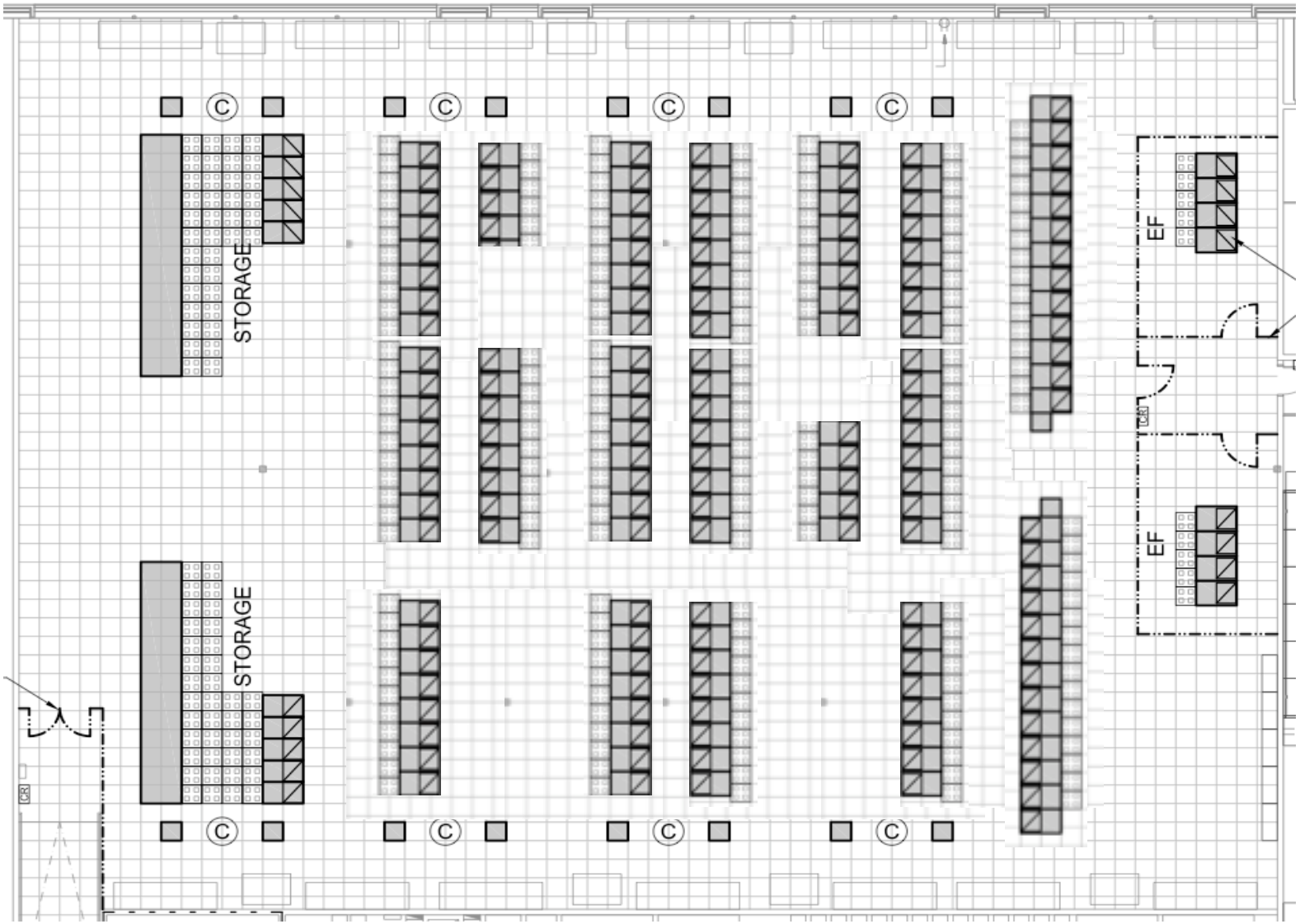
Оборудование: место для ~750 серверных стоек;

Использование: размещение основных бизнес-приложений Cisco с обеспечением требований к ЦОД Tier III+

Виртуализация серверов

- Снижение затрат на ЦОД ($$$/ сервер)
- Повышение утилизации физических серверов
- Снижение времени запуска сервисов
- Увеличение доступности

Традиционный дизайн: 1MW, 1000 м^2



Дизайн нового поколения

Общие данные

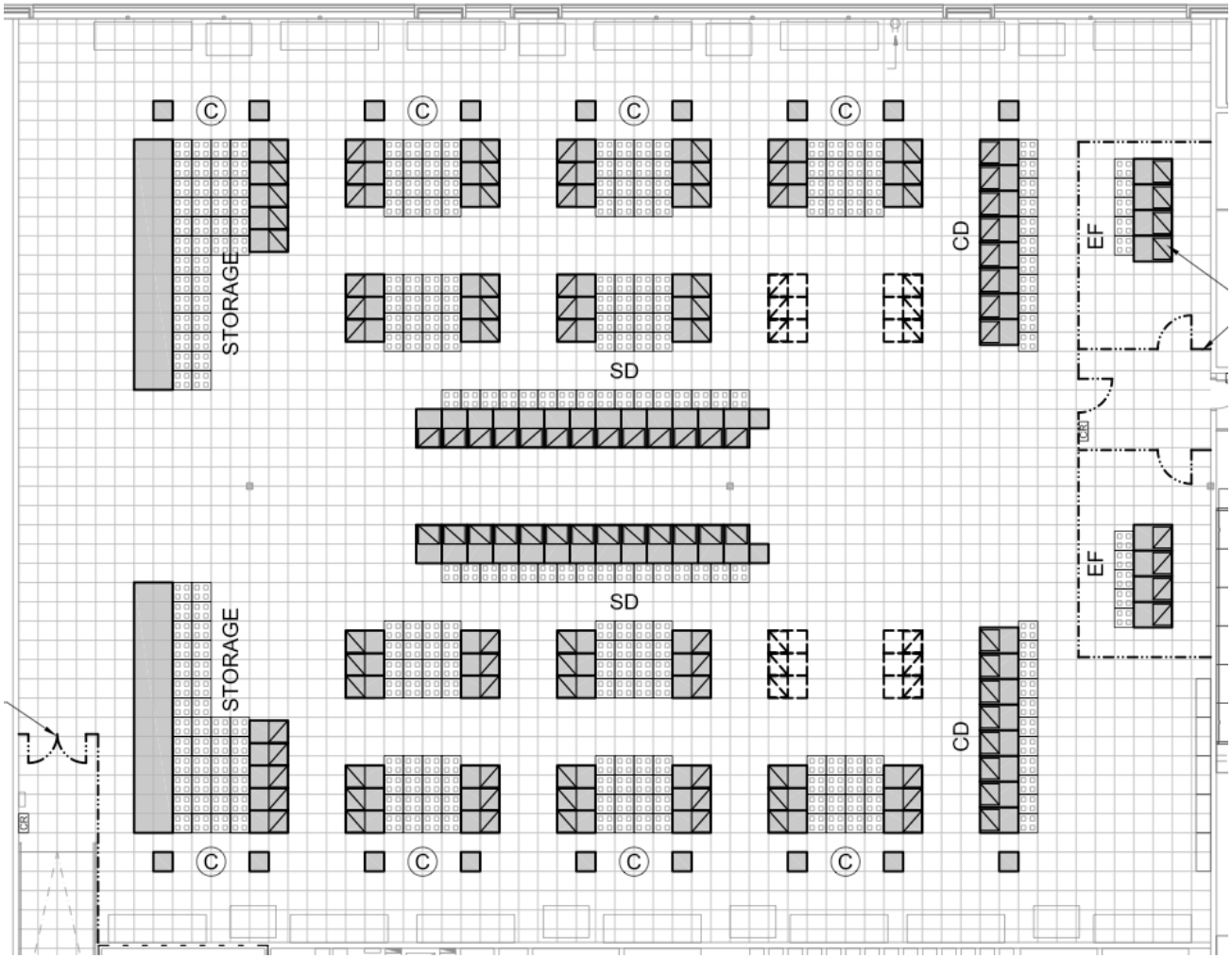
- 1MW UPS на этаж
- 12 kW/стойку
- Охлаждение в стойку («дымоход»)
- TO3R
- 21 Pods

To3R

- Nexus 5000
- Консоль
- Nexus 2000

Ядро

- Nexus 7000
- Catalyst 6500
- MDS



Эволюция ЦОД: пример Cisco Systems

	Раньше	Unified Fabric	UCS
Эффективность ЦОД	100%	130-150%	130% 170-200%
1,000 м^2, 1 MW			
Кабельная система	$2.7 million	$1.6 million	$1.6 m
Число физических серверов	720	930 -1080	1200-1400
Число VM	7200	9300-10800	12000-28000

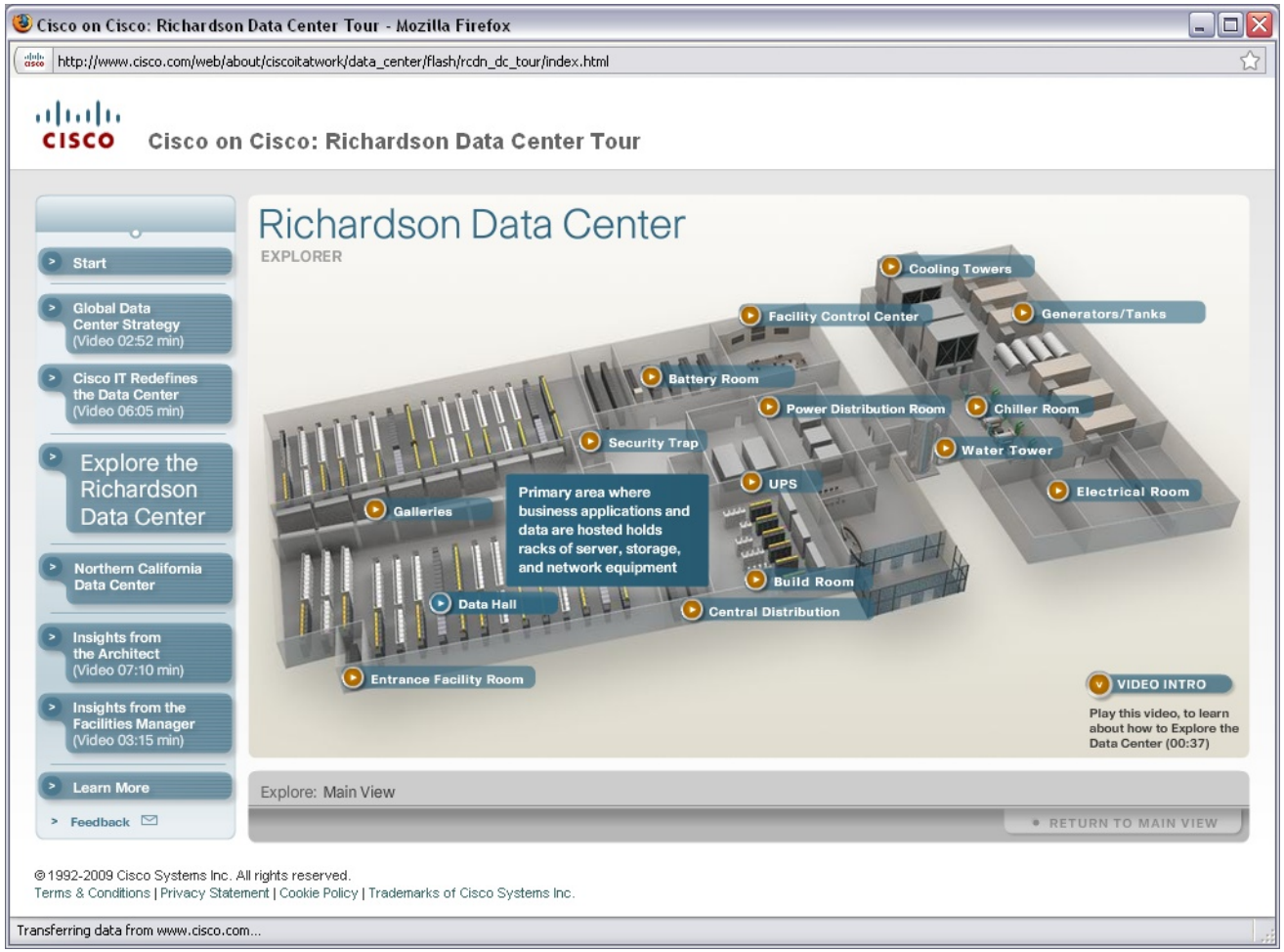
Повышение плотности

~40% экономия на СКС

В ЦОД того же размера!

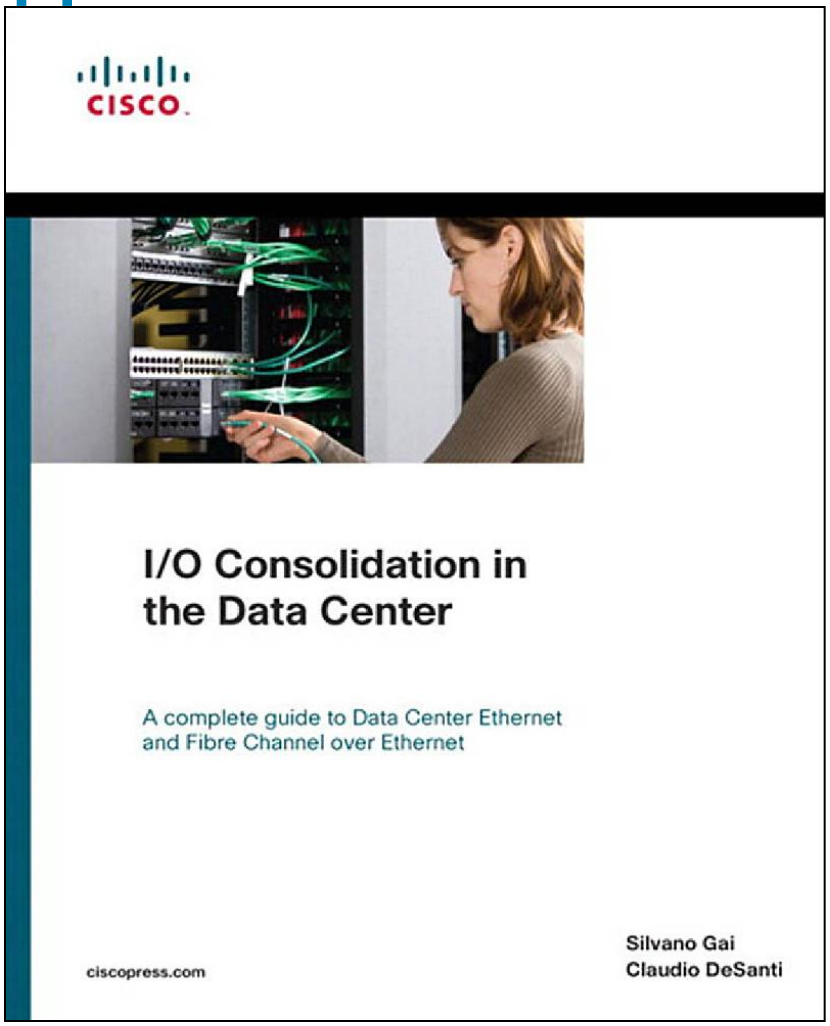
Notes: Assumes pre-UCS average V2P ratio of 10 to 1 and post UCS average ratio of 20 to 1 due to the memory expansion technology. Unified Fabric efficiency gains result from power optimization. UCS efficiency gains result from additional power benefits of UCS.

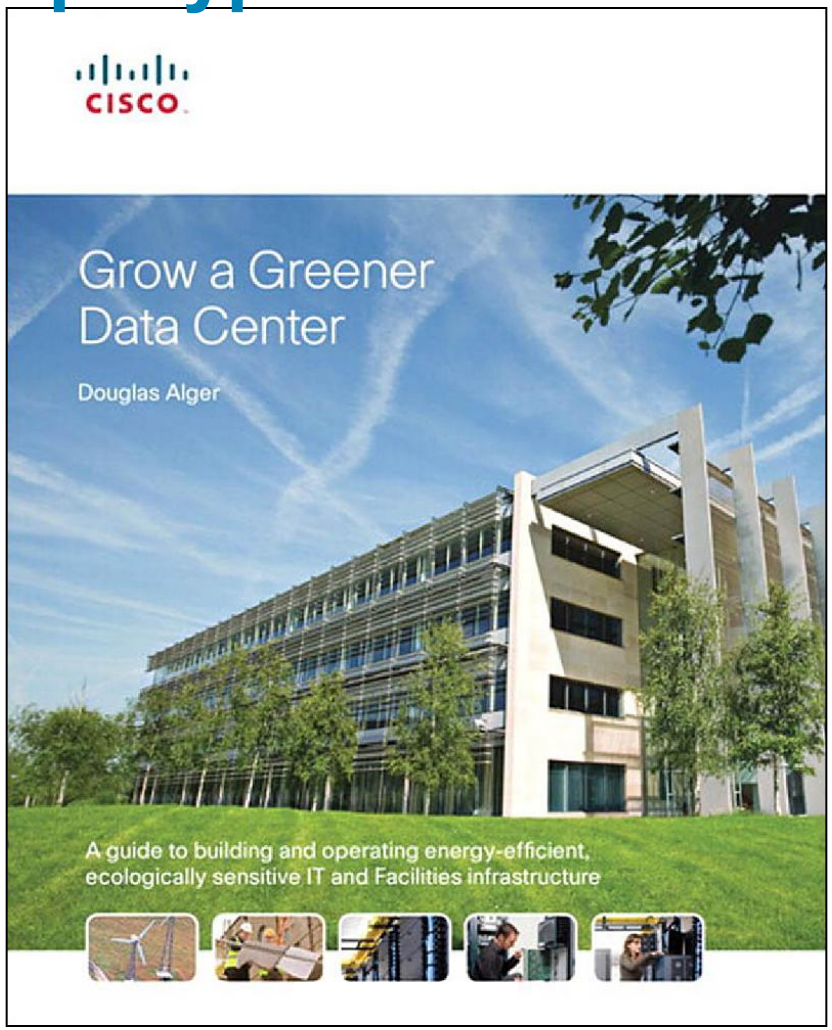
Виртуальный тур по ЦОД Cisco



www.cisco.com/go/virtualdatacenter

Дополнительная литература





Cisco Expo 2011

Вопросы?

Thank you.

Eldar Zhensykbaev
Consulting Systems Engineer

ezhensyk@cisco.com

Fibre Channel Speed Roadmap

	Product Naming	Throughout (MBps)	Line Rate (Gbaud)	T11 Spec Technology Completed (Year)	Market Availability (Year)
Base 2	1GFC	200	1.0625	1996	1997
	2GFC	400	2.125	2000	2001
	4GFC	800	4.25	2003	2005
	8GFC	1600	8.5	2006	2008
	16FC	3200	14.025	2009	2011
Parallel Optics	32GFC	6400	28.5	2012	Market Demand
	64GFC	12800	57	2016	Market Demand
	128GFC	25600	114	2020	Market Demand

Обзор функционнальности
Сравнение Nexus vPC и Catalyst VSS

Функция	VSS (Virtual Switching System)	vPC (Virtual Port Channel)
Multi-Chassis Port Channel	✓	✓
Топология без петель	✓	✓
STP как «запасной вариант»	✓	✓
Уровень управления	Один	Два, оба активны
Поддержка для L3 Portchannel	✓	✗
Управляющие протоколы	Один экземпляр	Экземпляры на каждом узле
Портов 10GE на Etherchannel	8	16
Конфигурация устройств	Единая конфигурация	Отдельные конфигурации (с контролем совместимости)
ISSU без потерь	✗	✓
Порты для связи узлов (Inter-switch Link)	Virtual Switching Supervisor 720-10G, карты 6708, 6716	Любые модули 10GE