BROCADE

TRILL

Creación de un DataCenter a nivel 2, escalable y Redundante

Julio Alba

IP SE Iberia

jalba@brocade.com

TRILL

Transparent Interconnection of Lots of Links

A proposed L2 protocol being developed by an Internet Engineering Task Force (IETF) workgroup

Mission

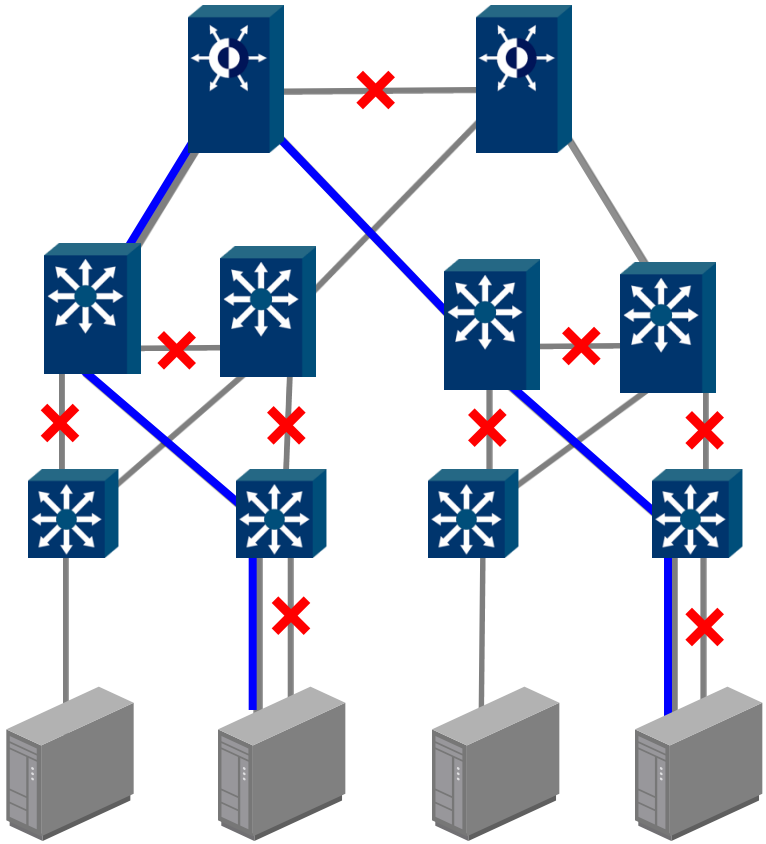
"*The TRILL WG will design a solution for **shortest-path frame routing** in multi-hop IEEE 802.1-compliant Ethernet networks with arbitrary topologies, using an existing link-state routing protocol technology*"

Scope

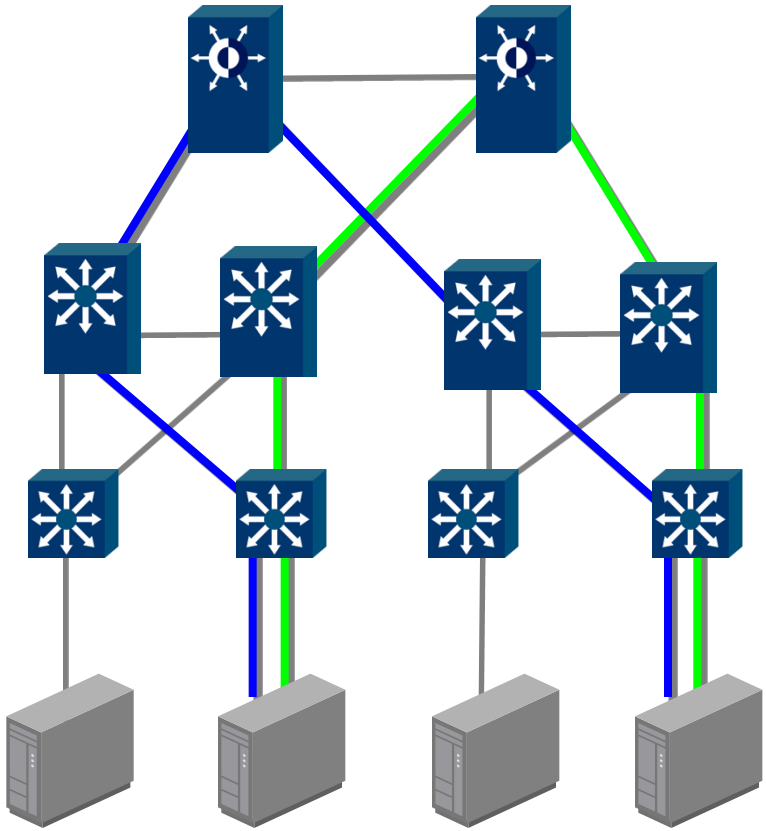
"TRILL solutions are intended to address the problems of …, inability to multipath, multihop… within a *single Ethernet link subnet*"

LANs Get L2 Multipathing

Today: STP Single Path

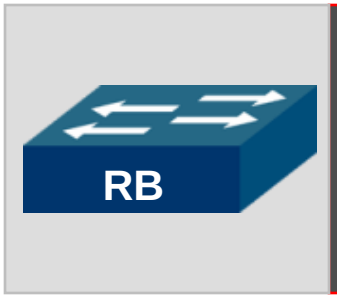


Next: L2 Multi Path



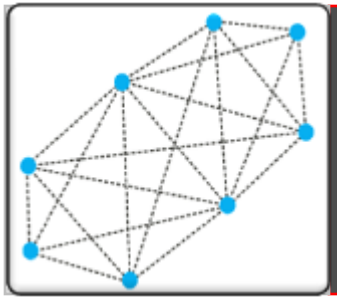
TRILL Solution

Overview



Devices are Routing Bridges (RBridges or Rbridges)

Data Plane is TRILL protocol

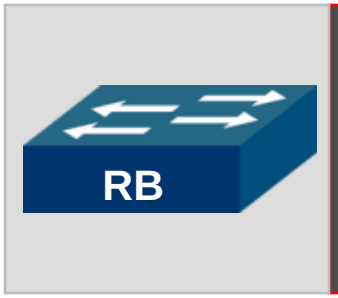


Control Plane is link state routing protocol

TRILL Solution

Functions

- Link state protocols
  - Flood configuration information to RBridges
  - Used for shortest path calculations
  - Used to distribute configuration database



- RBridges
  - Use link state Hellos to find each other
  - Calculate shortest paths to all other RBridges
  - Build routing tables

- TRILL
  - Ingress RBridges encapsulate TRILL data
  - Egress RBridges decapsulate TRILL data

TRILL

All You Need to Know

Definition
L2 shortest path frame routing solution in multi-path, multi-hop IEEE 802.1 compliant networks

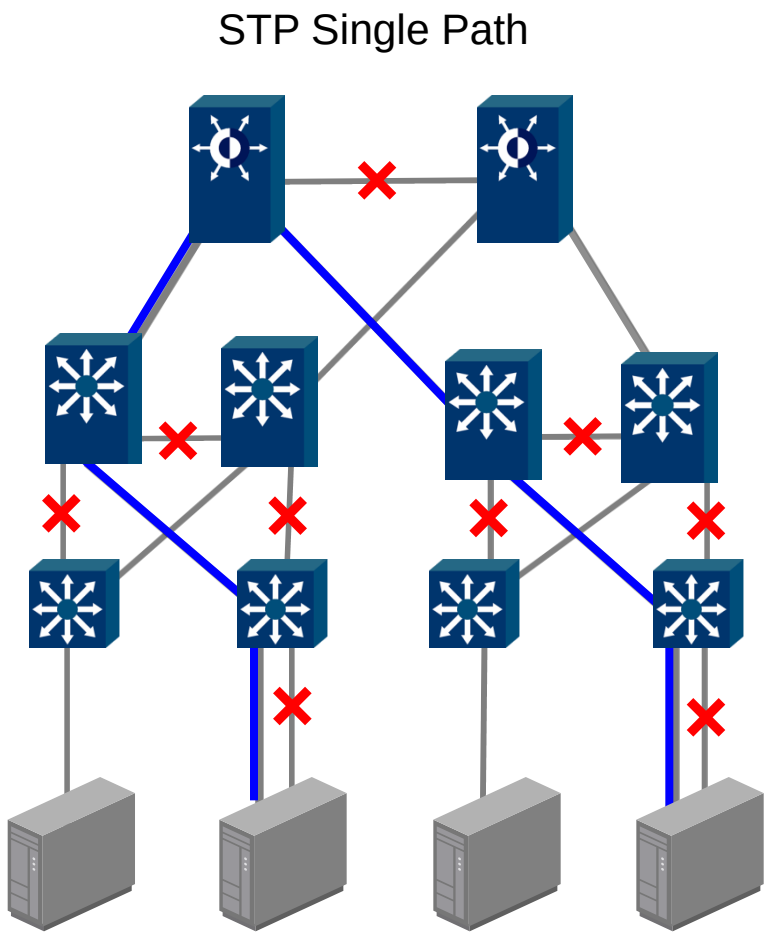
Features
Use Routing Bridges & existing link state routing protocols for discovery and creating routing tables

Benefits
Layer 2 multi path v. STP single path
Can use ECMP in L2

The STP Effect

Data Center with Spanning Tree Protocol (STP)

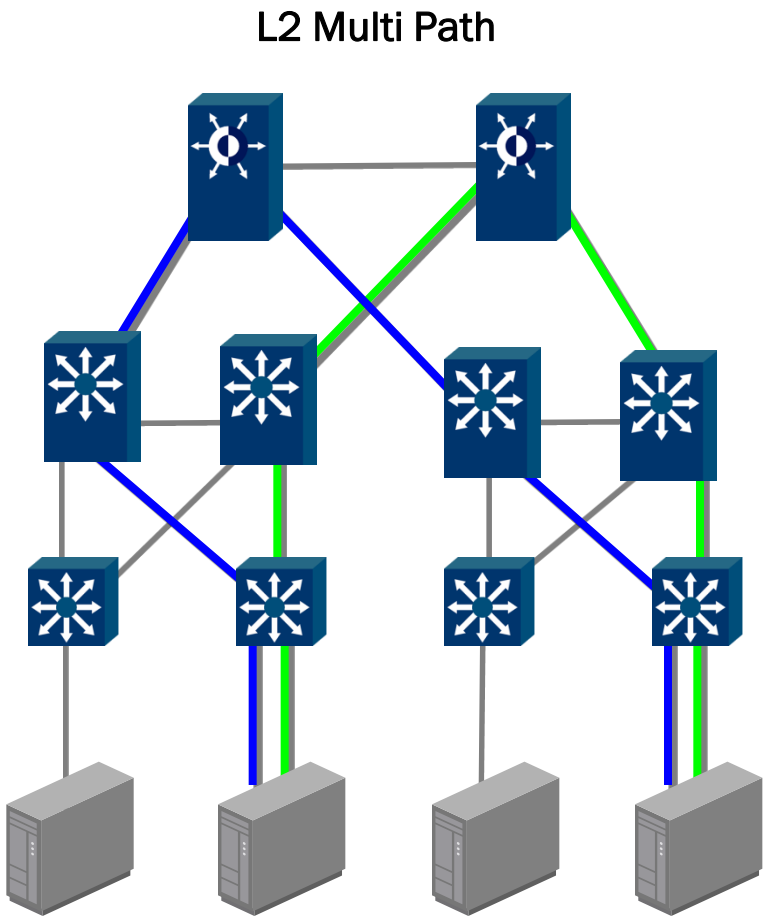
- STP is an Ethernet protocol that establishes and maintains a single loop free spanning tree among all the bridges on a VLAN
- All alternate paths are blocked
- Inefficient use of available links reduces aggregate bandwidth
- Reacts to small topology changes
- VLANs may partition due to connectivity changes



The TRILL Solution

L2 multipathing

- Enables L2 multiple paths via load splitting among paths
- Reclaims network bandwidth and improves utilization
- Improves efficiency with L2 shortest path and ECMP
- Faster response to failures
- TRILL is backward-compatible with existing infrastructures



BROCADE

TRILL Details

TRILL Solution

New Concepts

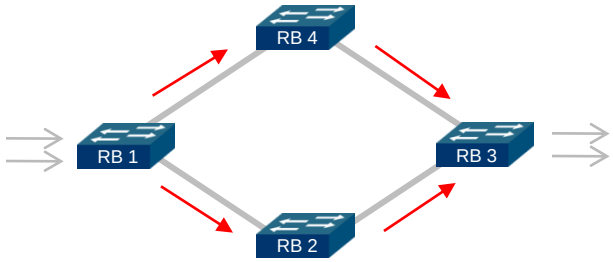
- TRILL Encapsulation
  - TRILL frame
  - TRILL EtherType
  - TRILL header
- TRILL Header
  - 64-bit field
  - Contains ingress and egress RBridge nicknames

TRILL Ethertype	V	R	M	OpLng	Hop
Egress RBridge Nickname	Ingress RBridge Nickname				

- Routing Bridges
  - Identified by a 16-bit “nickname”
  - Nicknames are auto configured local names
  - Ingress RBridge encapsulates TRILL frames
  - Egress RBridge decapsulates the TRILL frames
- Link state protocols
  - Discover configurations
  - Calculate shortest paths

RBridges

Overview

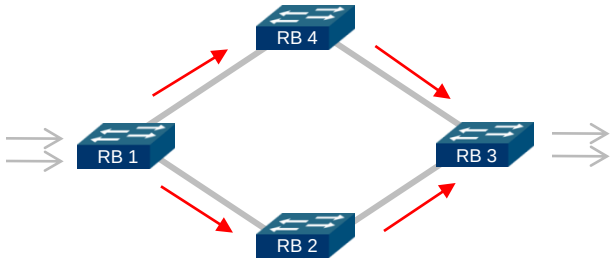


- Implement TRILL protocol
- Perform L2 forwarding
- Use link state routing
- Provide point-to-point forwarding with zero configuration
- Can auto configure themselves
- RBridges forwarding tables scale with the number of RBridges
- RBridges know what options other RBridges support

- Support multi-pathing for unicast and multicast traffic
- Compatible with classic bridges and can be deployed in bridged LANs
- Ingress RBridge adds *TRILL & outer MAC headers* to frames
- Outer MAC header is modified hop-by-hop as with routing
- Egress RBridge decapsulates the frame and learns the association of the “Inner MAC SA” with the Source RBridge nickname

RBridges

Personality & Behavior



Routers?

- Decrement a hop count in TRILL frames on each hop
- Swap the outer addresses on each RBridge hop from ingress to egress
- Use routing protocols, not STP
- Optionally learn MAC addresses by distribution through the control messages
- Use IP multicast control messages such as IGMP and restrict the distribution of IP multicast frames

Bridges?

- Deliver frames from the source RBridge to the destination RBridge without modification
- Support restricting frames to VLANs like IEEE 802.1Q bridges
- Support frame priorities like IEEE 802.1Q bridges
- By default, learn MAC addresses from the data frames they receive
- Can operate with zero configuration and auto configure themselves

Role of Link State Routing

Discovery & Shortest Path

- Link-state routing protocols are used to
  - Discover RBridge peers
  - Determine RBridge VLAN topology
  - Establish L2 delivery using shortest path calculations
  - Routers tell every router on the network about their closest neighbor
  - The routers only distribute parts of the routing table containing its neighbors

- Link-state routing neighbor information
  - Gathered continuously
  - The list is flooded to all neighbors
  - Neighbors in turn send it to all of their neighbors and so on
  - Flooded whenever there is a (routing-significant) change
  - Allows routers to calculate the best path to any router on the network

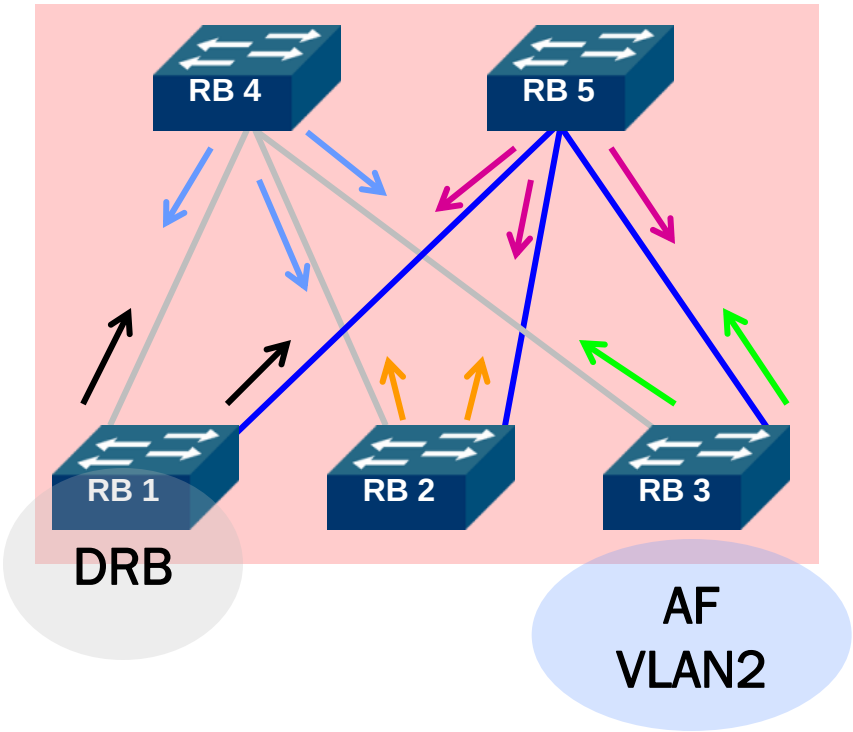
How RBridges Work?

Building the Routing Table

- RBridges need to maintain information about
  - Peer information
  - Topology information
  - Forwarding information : unicast, flooded, and multicast
- Link state protocols used to carry routing information about MAC addresses devices connected to VLANs
- Each RBridges uses the flooded information to
  - Construct a map of the VLAN
  - Calculates the shortest path from it to every RBridge on the VLAN
- The collection of the next best hop maps form the RBridge Routing Table
- Each RBridge has a copy of the global “link state” database

How RBridges Work?

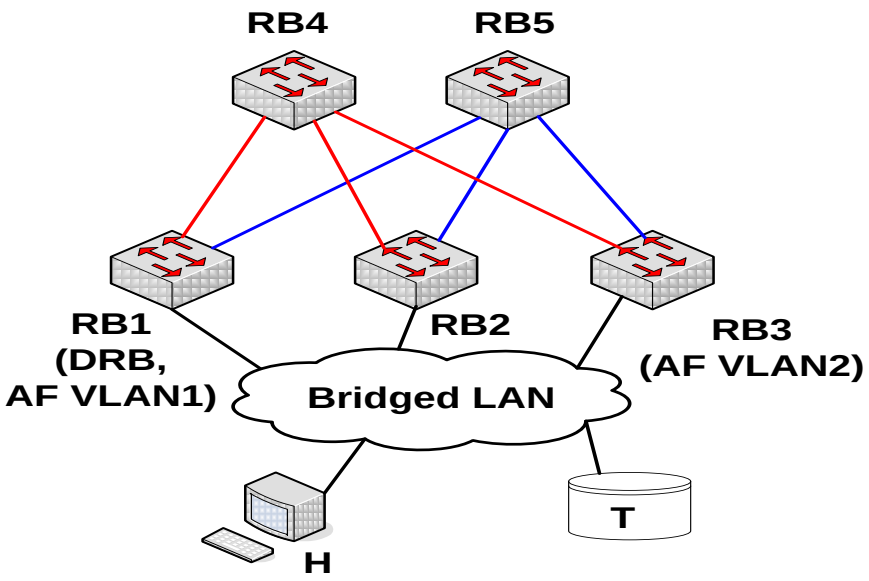
Designated RBridge - DRB



- RBridges discover each other by exchanging TRILL Hello frames
  - TRILL Hellos are sent to the All-IS-IS-RBridges multicast address
- Using link state protocol a single Designated RBridge (DRB) is elected from among all RBridges on the LAN
  - The DRB specifies the Appointed Forwarder (AF) for each VLAN

How RBridges Work?

Appointed Forwarder - AF

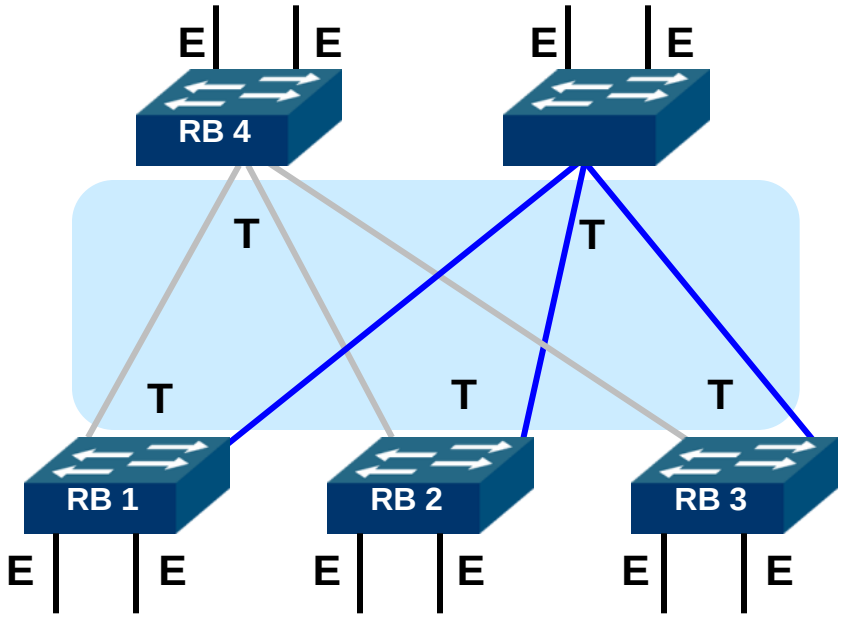


- The DRB specifies the Appointed Forwarder (AF) for each VLAN
  - DRB can also be the AF
- Only ONE AF can be appointed per VLAN; One VLAN - One AF
- The AF is in charge of handling all native frames in the VLAN
  - Ingress RBridge function: Encapsulates TRILL data frame
  - Egress RBridge function: Decapsulates TRILL data frames

TRILL Ports & Processing

Ethernet & TRILL Ports

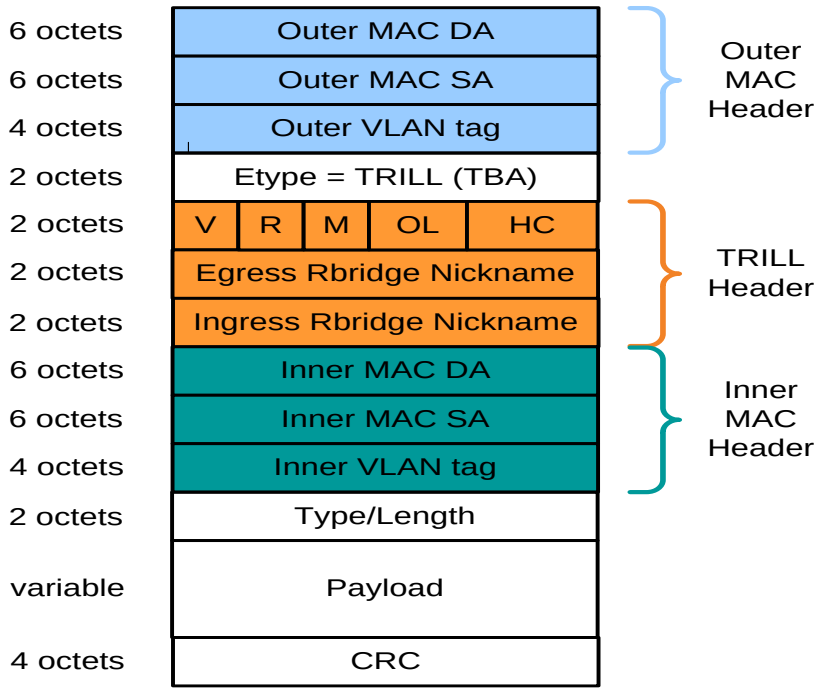
- Ports designations
  - Ethernet ports; E ports
  - TRILL ports; T ports
- Processing categories
  - Ingress: from E port to T port; E-T
  - Core: between T ports; T-T
  - Egress: from T port to E port; T-E



TRILL is layered at the upstream ports of RBridges

TRILL Frame Format

Header length: 64 bits



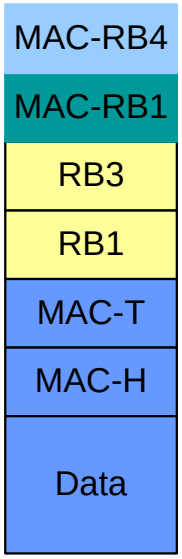
- Nickname: auto-configured 16-bit local names for RBridges
- V = Version (2 bits)
- R = Reserved (2 bits)
- M = Multi-destination (1 bit)
- OL = Options Length of TRILL options (5 bits)
- HC = Hop Count (6 bits)
- If M = 0, egress Nickname is the egress Rbridge
- If M = 1, egress Nickname is that of the RBridge that is the root of the tree

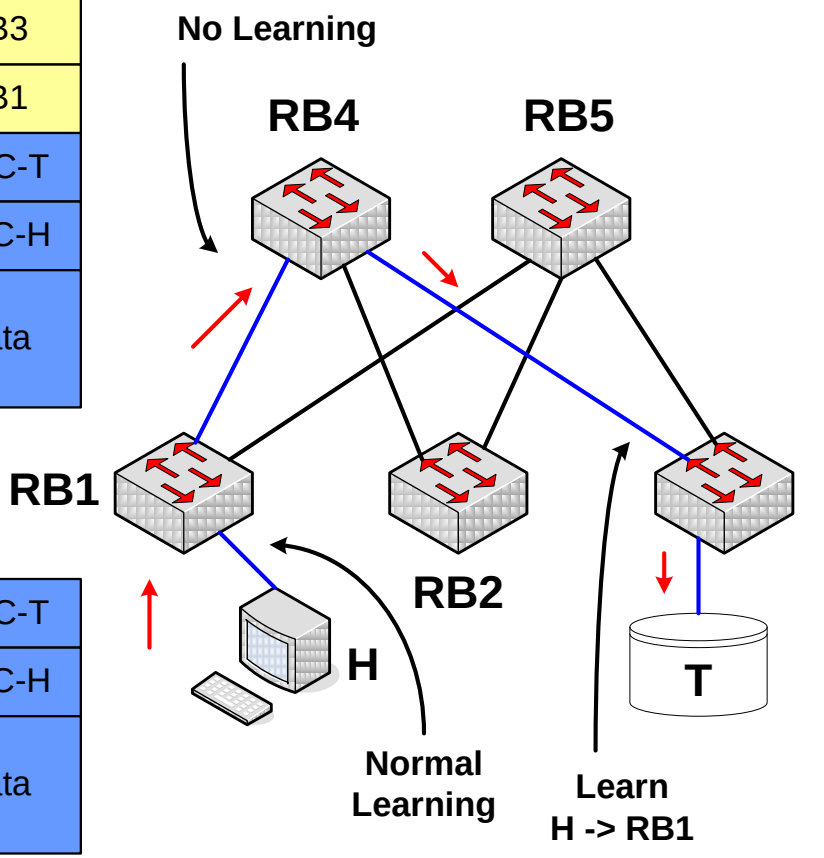
TRILL Encapsulation

Unicast data path

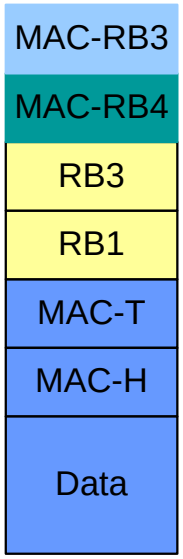
Known unicast destinations are forwarded RBridge hop by RBridge hop toward the egress RBridge

2. RB1 adds TRILL header and outer MAC header

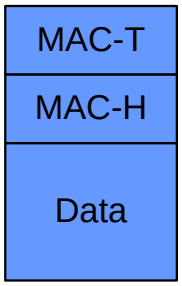




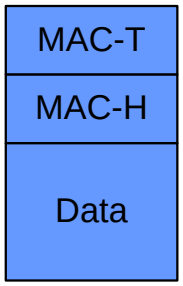
3. RB3 removes TRILL header



1. Host sends frames

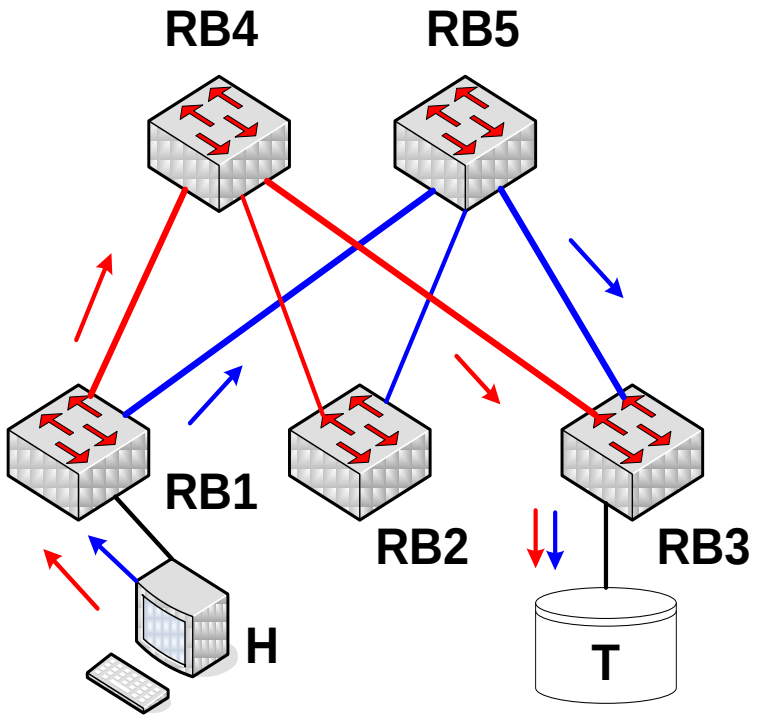


4. Target receives frames



Utilizing ECMP Paths & Reordering

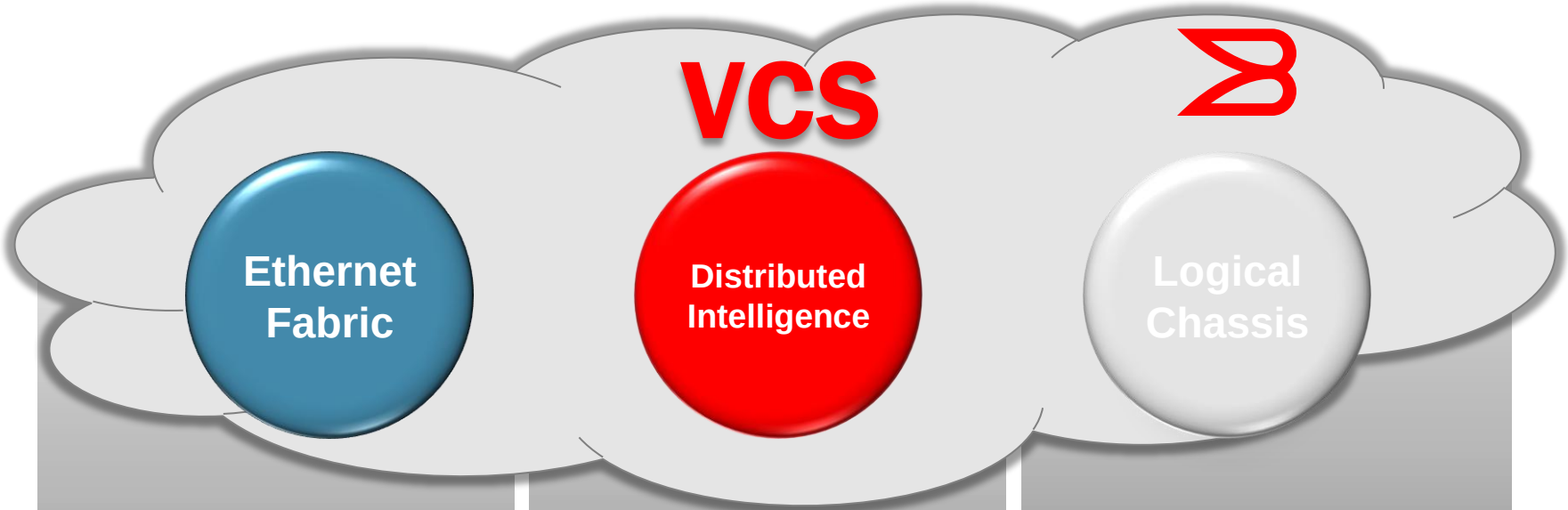
- TRILL supports up to 64 ECMP paths
  - Packet (frames) ordering maintained within flows
- RBridges are required to maintain frame ordering internally
- When multi-pathing is used, all frames for an order-dependent flow must be sent on the same path if unicast or the same distribution tree if multi-destination
- Re-ordering can occur when
  - A destination address transitions between being known and unknown
  - A topology change occurs



BROCADE

Brocade Solutions

Virtual Cluster Switching (VCS)



No Spanning Tree Protocol

Multi-path, deterministic

Auto-healing, non-disruptive

Lossless, low latency

Convergence-ready

Self-forming

Arbitrary topology

Fabric is aware of all members, devices, VMs

Masterless control, no reconfiguration

Logically flattens and collapses network layers

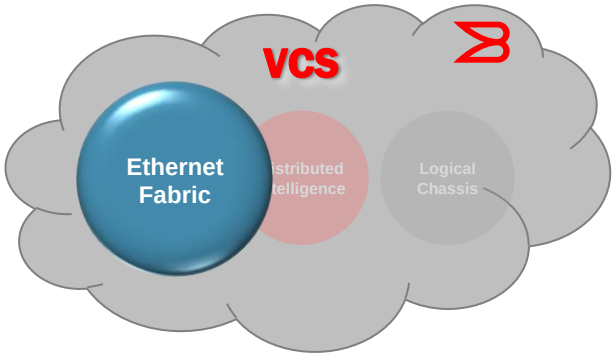
Scale edge and manage as if single switch

Auto-configuration

Centralized or distributed mgmt

Dynamic Services — Connectivity over Distance, Native Fibre Channel, Security Services, Layer 4-7, etc.

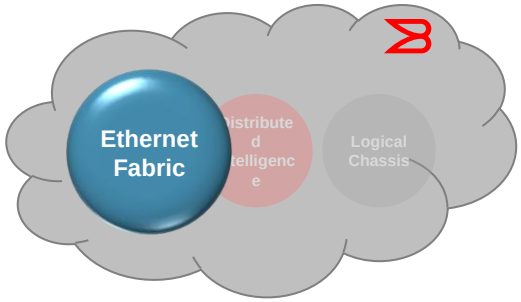
Ethernet Fabric Details



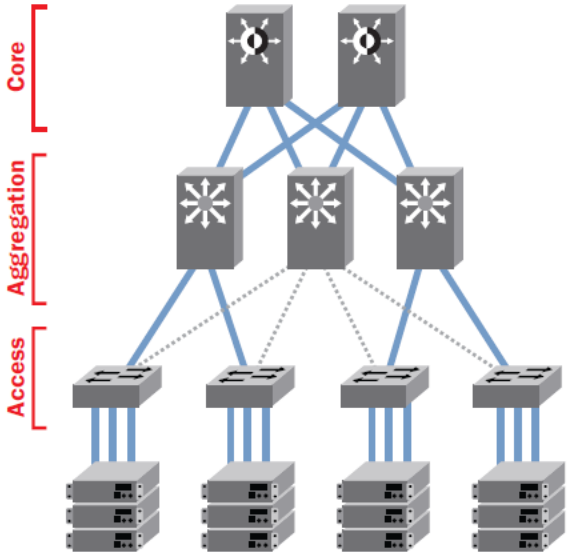
- 1st true Ethernet fabric
  - Layer 2 technology
- Link speed agnostic
- Data Center Bridging (DCB)
  - Lossless, deterministic
  - Priority-based Flow Control (PFC)
  - Enhanced Transmission Selection (ETS)
  - Data Center Bridging Exchange (DCBX)
- Transparent Interconnection of Lots of Links (TRILL)
  - Active multi-path
  - Multi-hop routing
  - Highly available, rapid link recovery
- LAN/SAN Convergence Ready
  - FCoE and iSCSI traffic
- Standards-based
  - Extends existing Ethernet infrastructure

Ethernet Fabrics

A New Network Architecture

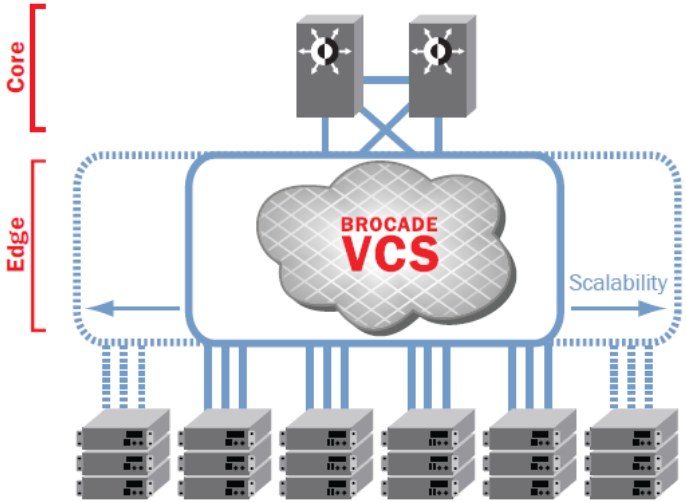


Classic Hierarchical Ethernet Architecture



- Classic architectures often require three tiers in the physical network
- STP disables links in the fabric to prevent loops, limiting network utilization
- Each switch has to be managed individually

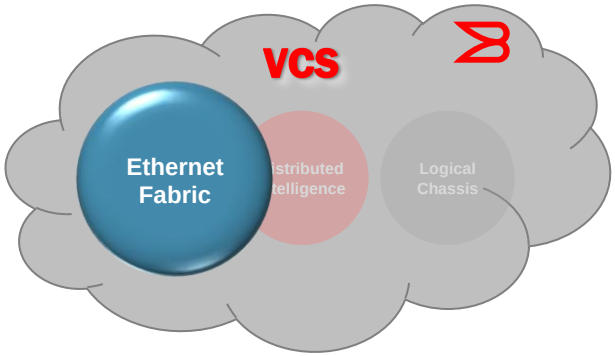
Ethernet Fabric Architecture



- Fabric architectures flatten and seamlessly scale out the Layer 2 network at the edge
- All links in the VCS fabric are active and it is managed as one
- Switches in the VCS fabric are managed at one

Ethernet Fabric Details

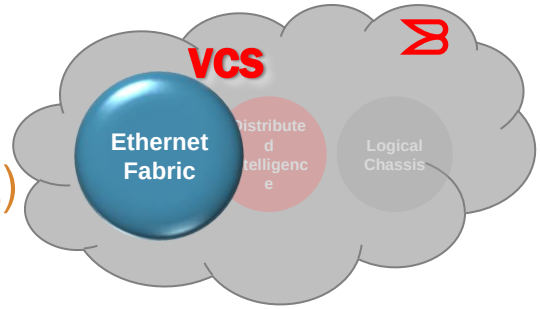
Data Center Bridging (DCB)

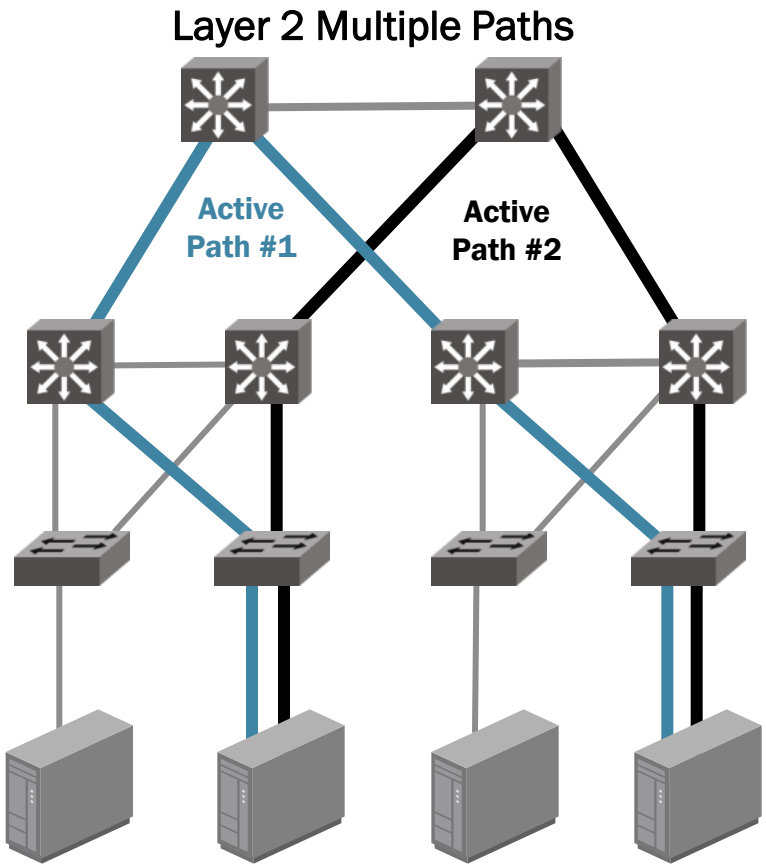


- Making Ethernet Lossless
- 802.1Qbb – Priority-Based Flow Control
  - PFC: Allows Identification and prioritization of traffic
- 802.1Qaz – Enhanced Transmission Selection/Data Center Bridging Exchange
  - ETS: Allows grouping of different priorities and allocation of bandwidth to PFC groups
  - DCBX: Discovery and initialization protocol to discover resources connected to DCB-enabled network

Ethernet Fabric Details

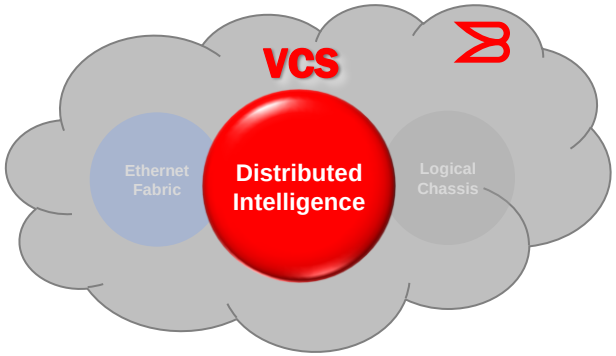
Transparent Interconnection of Lots of Links (TRILL)





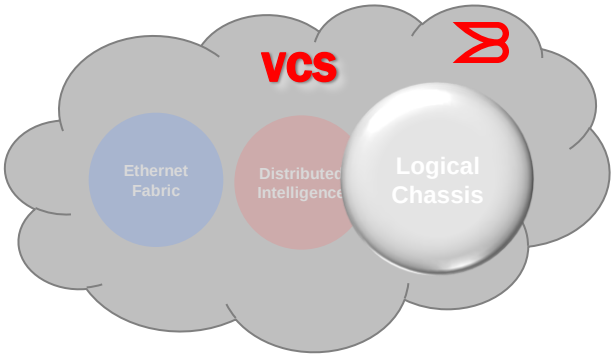
- Multi-path Layer 2 switching
  - All paths are active and traffic is distributed across all paths
  - Fully utilize all fabric bandwidth
- Establishes shortest paths through the Layer 2 fabric
- Uninterrupted response to link failures
- Backward-compatible and connects into existing infrastructures
- Delivers multiple hops for all traffic types (including FCoE)

Distributed Intelligence Details



- Distributed Fabric Services
  - Fabric is self-forming
  - Information shared across all fabric members
  - Fabric is aware of all devices connected
- Masterless Control
  - Switch or link failure does not require full fabric reconvergence
- Shared Port Profiles information
  - Automatic Migration of Port Profiles (AMPP)
  - Enables seamless VM migration without compromise

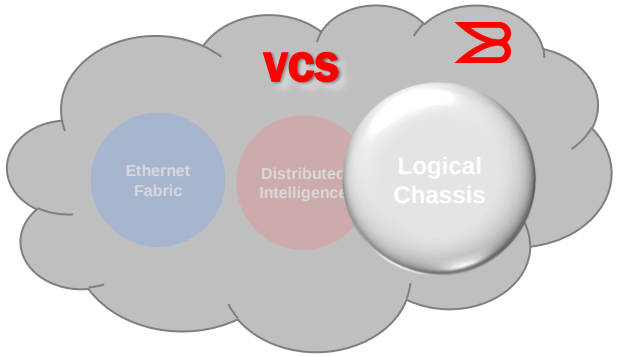
Logical Chassis Details

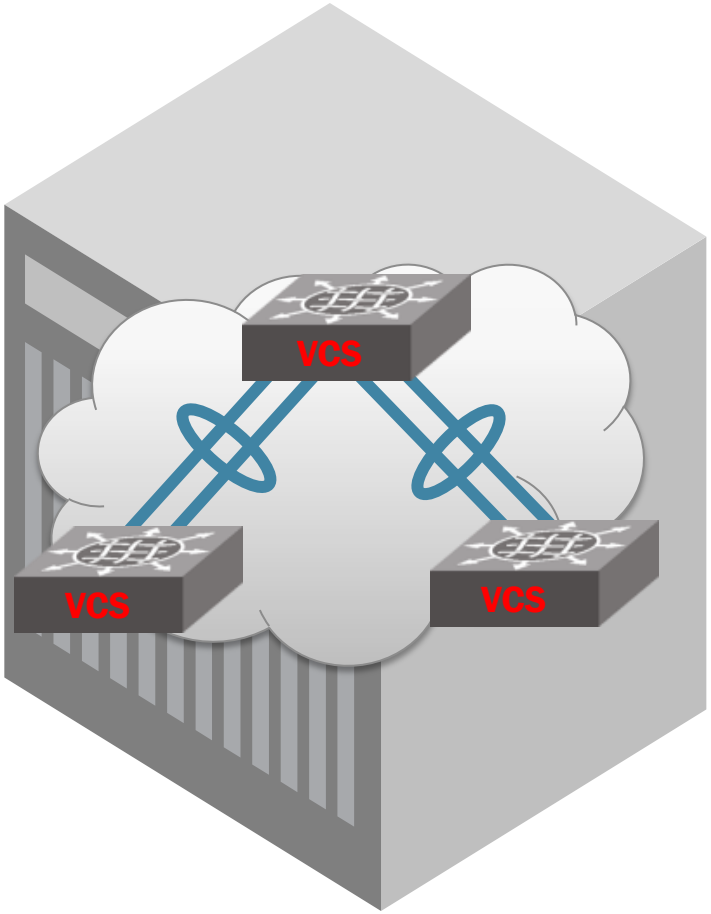


- Fabric auto-configures
  - Once VCS is enabled, no fabric necessary
- Fabric behaves/managed as a single logical chassis
  - Aggregation (or Core) layer sees one logical element
  - Fabric members act like a blade in a chassis
- Logically flattens and collapses network layers
  - Fabric is self-aggregating
  - Flexible fabric topologies
- Scales without added management complexity

Logical Chassis Details

Auto-Configuration

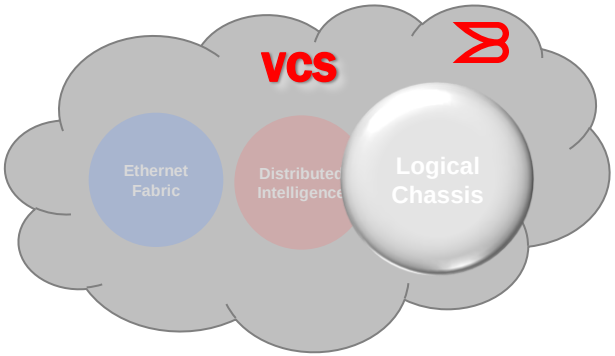


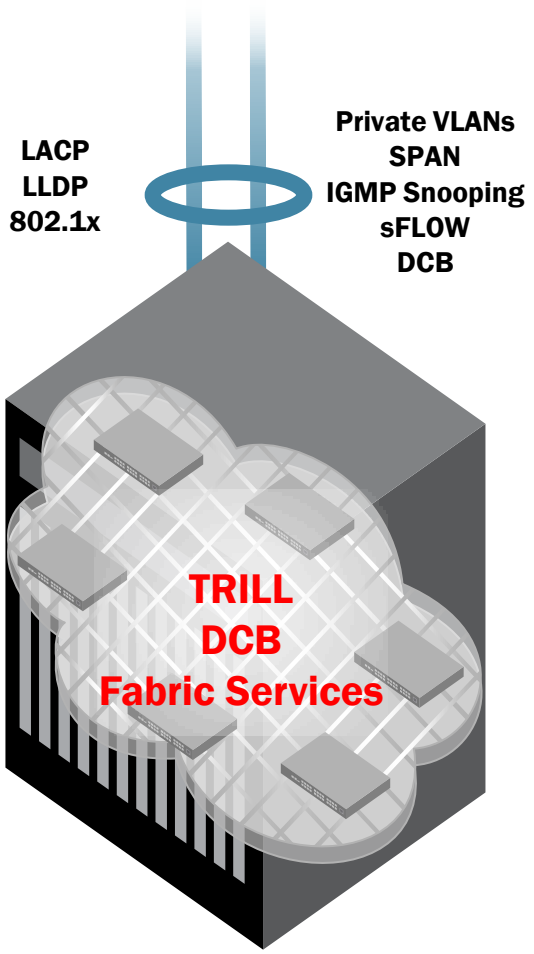


- VCS simplified deployment, scalability, and management of the network
- Enable VCS on each switch
- Connect the switches
- Fabric automatically forms
  - Common configuration across all switches
  - vLAGs auto-configure
- Managed as a single logical chassis

Logical Chassis Details

Single Logical Switch Behavior





- VCS behaves like a transparent LAN service
  - For example, BDPUs in STP environments are passed through the fabric
- Fabric protocols used within the fabric
  - TRILL, DCB, Fabric Services, etc.
- Industry-standard protocols used to communicate outside the fabric
  - LACP, 802.1x, sFLOW, etc.

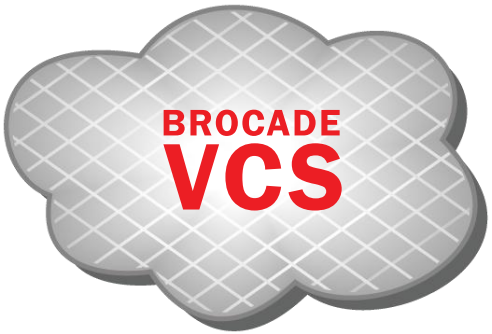
Brocade VDX 6720 Data Center Switches

Revolutionizing the way data center networks are built

Simplifies network architectures, dramatically reduces operating expenses

Allows the virtualized data center to scale while reducing complexity and enabling seamless application mobility

Increases network performance, utilization, and resiliency with Ethernet fabrics



Summary of Key Messages

Superior Layer 2 Brocade solutions

- TRILL solution
  - Layer 2 multipathing and ECMP
- Brocade will support and exceed TRILL
  - Extending FOS storage services to every hop
    - FSPF over DCB brings Fibre Channel fabric technology to Ethernet
- Brocade delivers unified & superior solution
  - All Ethernet and Fibre Channel resources unified under same fabric
  - New FOS services based on proven Fibre Channel technology