Architecture guide

Building data center networks using HP and Alcatel-Lucent

Table of contents

Introduction....2
HP Data Center Interconnect—Connecting geographically dispersed data centers 2
Key considerations for DCI design 3
Ethernet Virtual Interconnect....6
Summary....15
Partnering architecture 16
HP DCI and the Alcatel-Lucent 1830 PSS 16
Additional links 22

Introduction

In today's fast paced critical data center environments, network admins are tasked with increasing the performance of their data centers as well as implementing disaster avoidance and disaster recovery mechanisms that increase availability. To that end, connecting geographically dispersed data centers is now more important than ever before. Data center interconnections (DCI) solutions provide disaster avoidance and enable shared resource utilization, optimization of application response, and allows the flexible mobility of workloads and services. However, many DCI solutions have design limitation and faults that make them overly complex. The HP DCI solutions discussed in this document are designed to address these limitations by delivering responsive, efficient, and resilient DCI solutions.

Many customers have seen the advantage of a dual vendor strategy within their networks. This reference architecture allows for, and incorporates, this dual vendor strategy specifically with Alcatel-Lucent.

This guide is intended for technology decision makers, solution architects, and other experts tasked with improving data center networking between data centers. It can serve as a baseline for network planning and design projects.

This document is complemented by three other documents. These documents if referenced all together, can provide a complete view of today's data centers and HP solutions within the data center:

- HP FFRA data center trends: Details trends and technology drivers that we are seeing in the data center.
- HP FFRA guide: Provides high-level recommended HP data center network designs, which are able to address and solve the data center trends and drivers listed in this document.
- HP FFRA deployment guide: Provides specific configuration examples and best practices that should be followed in the data center.

HP Data Center Interconnect—Connecting geographically dispersed data centers

As IT technology has been maturing, many organizations have been reducing costs and increasing efficiency by following the trend to consolidate their data centers. This consolidation has been driving organizations to heavily leverage virtualization in the data center as well as private clouds and multitenant environments that support different business units. This allows organizations to increase their availability and help them to move and deploy new workloads onto resources that can best serve them.

These advancements mean that connecting geographically dispersed data centers is now more important than ever before. Geographic data center interconnections allow the IT designer to put in place disaster avoidance and disaster recovery mechanisms that increase the availability of the applications. Geographic dispersion also enables shared resource utilization, optimization of application response, and allows the flexible mobility of workloads and services.

DCI features and benefits include:

- Mobile workloads and long distance vMotion:

  DCI establishes reliable connections between data centers that serve as a platform for fast and reliable vMotion between distant data centers. The added speeds of up to 80 percent improvement of vMotion enhance the value of workload mobility and make applications more available to users.

- Layer 2 extension

  Layer 2 extensions can be optimized by leveraging the integrated L2 functions of Alcatel-Lucent's DWDM solution, which provides interface aggregation and increases network resources utilization.

- Multitenant enabled Layer 2 extension solution:

  The HP Ethernet Virtual Interconnect (EVI) solution can leverage HP Multitenant Device Context (MDC) technology to partition a single HP switch into multiple logical devices, which provides up to 75 percent reduction of the number of physical platforms leading to CAPEX and OPEX reductions.

- Follow-the-sun support:

  Organizations can deliver a better experience with up-to-date information by adopting a follow-the-sun model, which would be leveraged by call centers, engineering, and development teams.

- Bursty and seasonal traffic patterns:

  Organizations could load balance heavy or bursty workloads among geographically dispersed data centers making more effective use of data center resources.

- Investment protection:

  DCI solutions can be customized to fit the customer's exact environment, from IP to MPLS to DWDM. DCI solutions can work with customer's existing IP networking infrastructure without requiring changes. This protects networking investments and allows easy and seamless deployment.

  The DWDM solution provided through the partnership with Alcatel-Lucent (1830 Photonic Service Switch [PSS] family), maximizes the utilization on links and minimizes the cost per bit transported, including the convergence of Storage Area Network (SAN) and Local Area Network (LAN) traffic.

- High availability:

  The DCI solution establishes active/active links between data centers. Along with HP IRF, DCI establishes reliable links with link aggregation and failover capabilities. This gives users uninterrupted access to applications in the event of disruption of service at one data center.

  This can be completed and further augmented by the combination of protection switching and redundancy mechanisms provided by the Alcatel-Lucent DWDM solution. The Alcatel-Lucent solution supports path diversity redundancy and optical section protection with less than 50 ms failover.

- Disaster recovery and data replication:

  Data can be replicated across data centers using DCI. In the event of failure, DCI provides the means to recover data from remote data centers and ensures business continuity.

  Alcatel-Lucent DWDM solution meets desired latency requirements offering a comprehensive range of multiprotocol interfaces and supports any network topology. End-to-end Fibre Channel transport solutions based on this platform are certified to be interoperable with major storage and FC switch vendors like HP.

- Secure data center connect:

  For distributed resources to work effectively and meet diverse end user requirements, applications require secure, low-latency real-time communications with guaranteed QoS. The DCI infrastructure and optical fibers are the key components of a holistic and systematic IT security program. The Alcatel-Lucent DWDM solution integrates on-the-fly Layer 1 encryption process with few microseconds of latency impairment, providing wire-speed encryption up to 10 Gbps for Ethernet, Fibre Channel, and InfiniBand traffic.

Key considerations for DCI design

Current network conditions

The network resources of a user between data centers will determine the solution to be used as follows:

- Ethernet Virtual Interconnect (EVI):

  This IP-based solution option is extremely useful in simplifying DCI. Transmission between data centers over DWDM or MPLS can be complex to manage and is often highly dependent upon costly dedicated and rigid service provider's infrastructures. In contrast, EVI runs over any IP infrastructure so it can be deployed without requiring changes to an existing infrastructure. This characteristic simplifies deployment by allowing Layer 2 connectivity across the network without having to deal with Layer 3 networking dependencies.

- Ethernet LAN extension:

  This option extends Ethernet natively over a dark fiber or DWDM optical transport. As such, this solution mostly applies to point-to-point deployments, where the sites are connected via dedicated dark fiber links or DWDM optical circuits.

- MPLS point-to-point or multipoint using MPLS or VPLS:

  This option uses MPLS technologies to provide L2 connectivity services over a L3 network service. Depending on the nature of the transport infrastructure between data center sites and the number of data center sites to be interconnected, different technologies can address the connectivity requirements.

- Optical data center connect:

  Large enterprises that require high-bandwidth interconnect between data centers over their private metro, regional, or national footprint can deploy the Alcatel-Lucent 1830 PSS to deliver the most advanced 100G technology, with Transponders and Muxponders optimized for metro, regional, or national reach. The 1830 PSS provides the converged aggregation capabilities for Ethernet, FC, and IB with client's speed of 1 Gbps, 10 Gbps, 40 Gbps, and 100 Gbps.

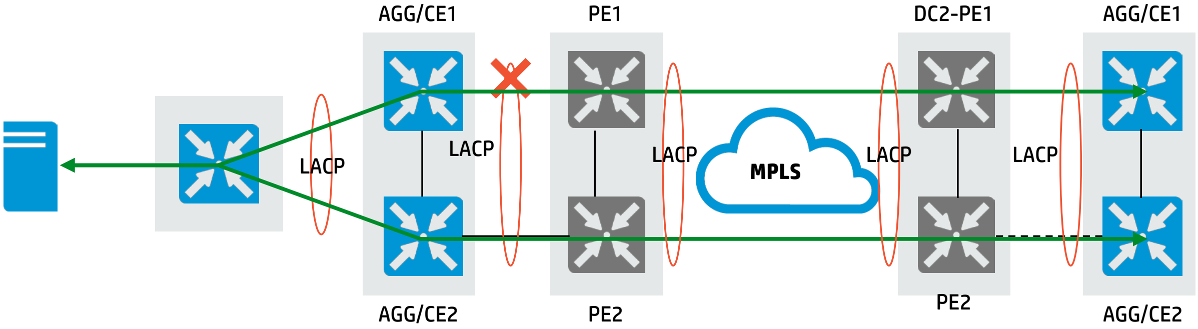
High availability (HA)

In the perspective of VM resource scheduling and remote cluster access, multiple interconnected data centers can be considered as a logical large-size data center. The links interconnecting data centers can be considered as the DCI backbone links of the large-size data center. More importantly, the backbone links interconnecting data centers transmit control signaling, in addition to vMotion or data packets. Therefore, once the links interconnecting the data centers fail, the large-size data center fails to work properly, and causes service interruption for users.

Therefore, a key consideration of Layer 2 DCI is to improve availability. The best way of improving HA is to design backup DCI links and backup nodes (DCI devices). To improve HA and increase the interconnecting bandwidth at the same time, you can design load sharing interconnection links, so that you can increase the bandwidth and enable the services to rapidly converge when the system encounters errors, thus improving HA.

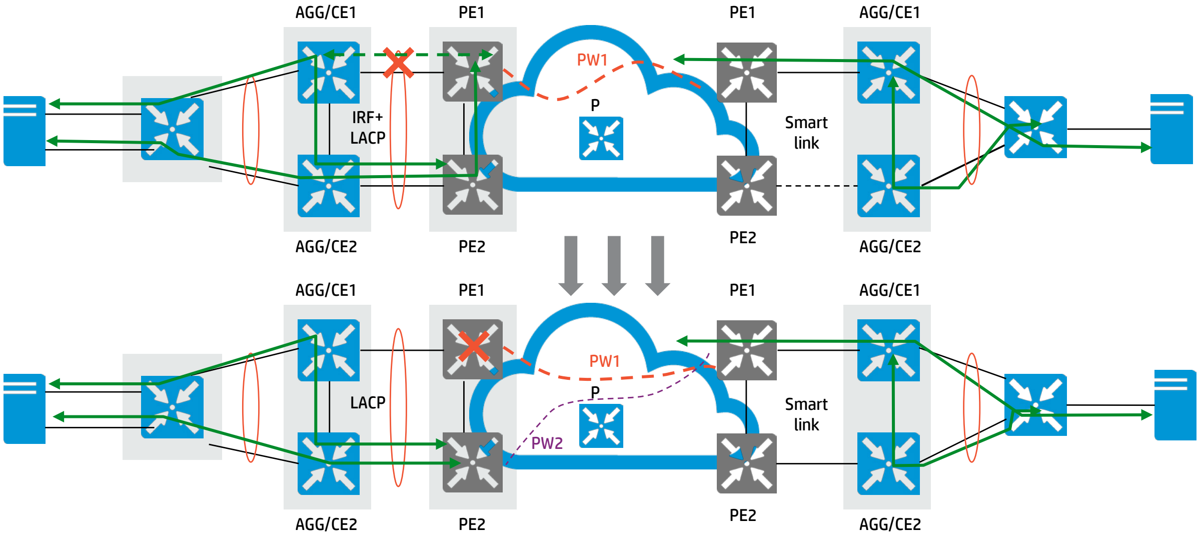
For example, whenever possible, HP recommends using IRF technology on the DCI devices as well as the CE core devices. Aggregating two or more links not only simplifies the DCI topology but provides HA and achieves load sharing. Use LACP to aggregate two or more links between the dedicated DCI devices into a logical link, so that the DCI topology is greatly simplified. At the same time, the bandwidth of the two HA links is optimized, and load sharing is achieved.

Figure 1. IRF + LACP load sharing design



Because many routers do not support IRF, there may still be an end-to-end loop between DCI devices of each DC, thus smart link or Virtual Router Redundancy Protocol (VRRP) should be used to obtain HA and loop avoidance. As shown in the figure below, if the PE or link between PE and CE devices fails, the flow can achieve fast convergence by using smart link and quickly switch the traffic path to another one.

Figure 2. Smart link design



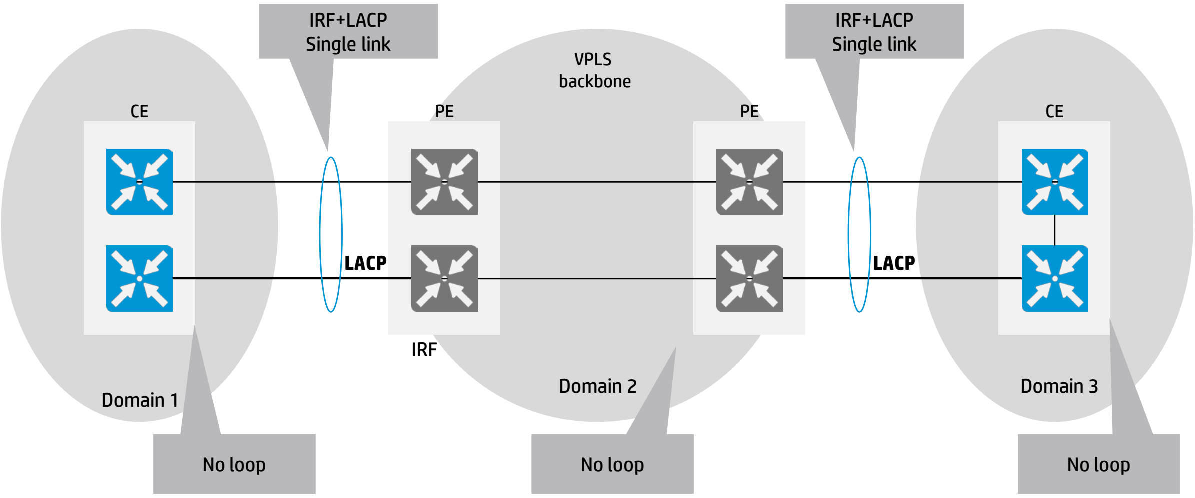
Layer 2 topology management (STP domain isolation and loop management)

When you deploy STP in a Layer 2 DCI network, HP recommends that you isolate STP domains to each data center. On each port connected to the DCI link, enable BPDU drop (edge port) and disable STP.

Additionally, end-to-end loop management needs to be addressed to manage the loops across multiple interconnected Layer 2 domains. Since STP domains will be isolated at each data center, you will need to ensure no physical loops exist.

Of course, this could be done by utilizing only one single link between data centers, but a better solution would be to use IRF and LACP, or smart link if the devices do not support IRF.

Figure 3. End-to-end loop management design



Path optimization for DCI

Typical traffic models of data centers show that the incoming traffic and the outgoing traffic are seriously asymmetrical. The request traffic entering the data center is usually very light and the response traffic from the servers to clients is usually very heavy.

If a large volume of data traffic passes through the DCI core network links, the data traffic consumes heavy bandwidth and degrades the quality of transmitted control data. This causes a loss in control data packets and affects the migration or disaster recovery process.

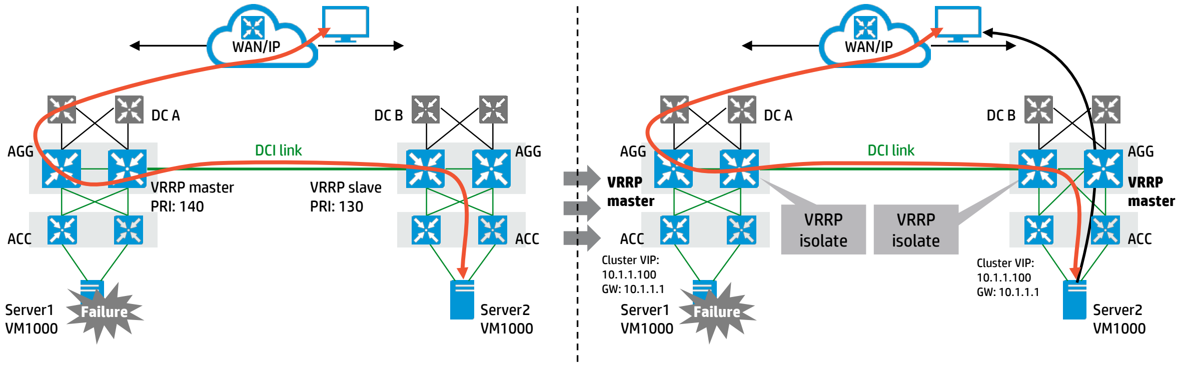
Therefore, in the event of a DC failure, you must prevent the response data traffic from passing back through the DCI core network link. The key to data path optimization is drawing the response traffic of servers away from the original DCI core network path and towards the data center local gateway.

To draw the response traffic of servers away from the original incoming DCI core network link, you can configure a VRRP group on each data center gateway, as shown in figure 4.

More specifically:

- Configure the same VRRP group (the same virtual IP address and priority) on the gateways for data center A and data center B.
- Configure an ACL on the outgoing interface that connects the gateway to the DCI core network link to prevent the VRRP packets from entering the other data center through the DCI link.

Figure 4. Data path optimization for data centers



With these configurations performed, when a VM is moved from data center A to data center B, its IP address does not change, and data center B is configured with the same VRRP gateway. As a result, the response traffic of servers can be directly sent to the clients through data center B's gateway. This prevents a large volume of response traffic from passing back through the DCI link.

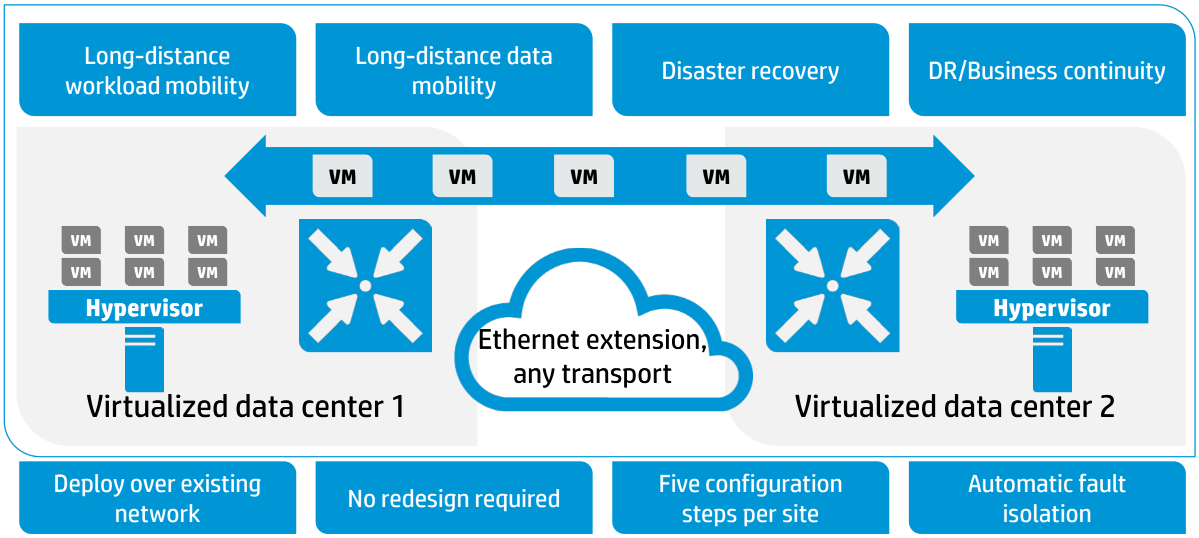
Ethernet Virtual Interconnect

Along with existing DCI solutions based on DWDM and MPLS/VPLS, HP has developed DCI solution with multiple technologies to help organizations quickly and easily extend Layer 2 networking to multiple data centers. EVI runs over any Internet Protocol (IP) transport and extends Layer 2 domains across a WAN network, typically between data centers. By virtualizing and automating the link layer domain across data centers, EVI delivers the elements necessary to enable a Software Defined Networking (SDN) data center infrastructure. It enables several data centers to work as one that is more responsive, with higher efficiency and solid high availability for business resiliency.

With EVI, enterprises are able to:

- Accelerate the delivery workload mobility with remote vMotion
- Increase application's performance with multipathing and load balancing
- Allow organizations to scale to up to 16 geographically dispersed data centers without requiring them to change the underlying network
- Simplify the Layer 2 connection by encapsulating traffic over GRE and automatically isolate Spanning Tree Protocol domains
- Achieve optimum degrees of high availability and disaster recovery for valuable data
- Allows clients to have a simple set of Layer 2 routing extensions that can provide data interconnectivity in minutes rather than months with legacy approaches like VPLS

Figure 5. HP Ethernet Virtual Interconnect


Long-distance workload mobility
Long-distance data mobility
Disaster recovery
DR/Business continuity
VM
Hypervisor
Virtualized data center 1
Ethernet extension, any transport
Virtualized data center 2
Deploy over existing network
No redesign required
Five configuration steps per site
Automatic fault isolation

When used along with the IRF switch virtualization technology, EVI delivers greatly enhanced reliability, resilience, and faster remote vMotion capabilities. The combination of EVI and MDC brings multitenancy to cloud-ready and remotely connected data centers.

EVI is currently supported on the HP 12500 Switch Series. All technology references and design considerations focus on positioning these switches in the data centers to establish Layer 2 connectivity between locations. EVI is a feature supported within HP Comware 7.

EVI basics

EVI is a Layer 2 routing technology that uses EVI Links and GRE tunnels to extend VLANs across up to 16 locations. Each EVI network has a unique network ID, extends a unique list of VLANs, and has separate control and forwarding planes. When used in conjunction with MDC, each MDC can support 32 EVI networks with each MDC running up to 4K VLANs.

EVI control plane

The EVI control plane is responsible for the discovery of new EVI nodes, the connection of these nodes, and maintaining MAC learning and advertisement.

EVI Neighbor Discovery Protocol (ENDP)

ENDP is the protocol responsible for setting up and maintaining the EVI Links between end devices. There are two types of entities used with ENDP:

- ENDP Server (ENDS):

  An ENDS is responsible for receiving registration requests, maintaining the ENDC database, and propagating member information.
- ENDC:

  These are the locations that communicate with the ENDS. Requests are sent from ENDCs and are received by the ENDS.

Once EVI has been configured the discovery process is autonomous. The high-level process is described below:

- The main data center switch(es) have EVI server (ENDS) enabled
- When a new location is added, the switches need to be configured with EVI client enabled
- The new location sends a registration request to ENDS
- The main data center sends a response back
- The two locations exchange EVI member information

EVI MAC address learning and advertisement

The standard MAC learning behavior is not affected by the configuration of EVI. In fact, EVI adds enhanced functionality by advertising known MAC addresses to other EVI enabled devices as well as other features that are highlighted in the EVI features section. EVI uses IS-IS to propagate MAC addresses reachability information. The EVI IS-IS packets are transported across the underlying carrier network.

Figure 6 demonstrates how a MAC address is learnt at Site 1 (IP1) and sent to the other locations.

- Site IP1 learns about new MAC entries mac1 and mac2 to VLAN 100
  - The EVI IS-IS process creates an LSP update that includes the mac addresses and VLAN info
  - The Edge Device (ED) at Site IP1 sends the LSP to its neighbors
- Each neighbor reads the LSP and delivers the information to the local EVI IS-IS process. Locally the MAC address looks like it was learnt by the EVI tunnel interface. When a packet is destined for either MAC at any location other than IP1, it is the EVI IS-IS process that routes this back to IP1
- The LSP updates can also include MAC addresses that have been aged out

EVI data plane

The EVI data plane is separate from the EVI control plane and is used to forward unicast, multicast, and broadcast traffic across the EVI network. This section covers the forwarding of traffic in an EVI network in more detail.

Local forwarding

Local packet flow is unchanged and not influenced by EVI.

EVI data plane unicast forwarding

When it is determined that a packet is destined for a device at another location, a lookup is done and the correct EVI interfaces are identified. The packet is encapsulated in a GRE tunnel and flows across the EVI network from the source ED to the destination ED. The destination ED removes the encapsulation and forwards the Ethernet packet to the end device as if the original sender were local.

EVI data plane multicast forwarding

In the current release, multicast is supported. The EVI ED closest to the multicast traffic source will unicast the frames across the EVI links. The local site will propagate the multicast packet based on the EVI control plane learning mentioned previously.

In the future, the end device will map the multicast frame to the multicast tree on the transport. The transport then replicates the frame for multicast member sites. This will require the devices to support Any Source Multicast (ASM) and source specific multicast (SSM).

EVI data plane broadcast forwarding

By default, broadcast traffic is allowed across the EVI network. Unknown unicast and multicast packets are dropped at the ED.

EVI features

The following features are used to optimize the control plane traffic thus increasing efficiency.

Selective MAC routing

This feature stops unknown unicast and multicast frames from flooding the EVI Links. The internal interfaces are capable of flooding to the internal interfaces while the EVI-tunnel interfaces will drop these frames. Similar to selective flooding, it is possible to permit or deny a MAC route.

Automatic loop prevention

EVI has the following loop prevention mechanisms built in and enabled by default on both the data and control planes:

- EVI-split horizon:

  This is enabled on the data plane to prevent frames received from EVI tunnels from being forwarded to the transport layer (EVI Links). Its primary function is to prevent loops among EDs.

- STP domain boundary:

  This disables STP on the EVI Links. STP domains and BPDUs are not extended across sites keeping topology changes local. BPDUs are blocked from one location to another so that STP changes are contained within a site. This also allows each site to run different versions of STP. The following versions of STP are supported 802.1d, 802.1s, and 802.1w.

- Selective flooding:

  By default unknown unicast and multicast are dropped at the EVI Links. If an application uses a special MAC address for traffic identification it would break. Selective flooding enables the ED to flood frames with a certain unknown destination MAC to an EVI tunnel interface.

An example of this would be Microsoft® Network Load Balancer (NLB) that uses a special MAC address (cluster MAC) to identify a cluster. If cluster members are located in multiple sites, selective flooding would be used to propagate the cluster MAC address on the EVI-tunnel interface(s).

Automatic VRRP isolation

To provide routing path optimization, EVI enables VRRP isolation by default. This allows each DC to have an active Layer 3 gateway that leverages the VRRP protocol to hide the specific HW details (IP and MAC addresses). All data centers will run separate sets of VRRP instances. VRRP isolation stops the population of VRRP keep-a-lives of each VLAN over EVI Links, so that each data center always has active/active gateways.

ARP flooding suppression

This feature will reduce the number of broadcasts that traverse the EVI network. When an ARP request is made and initially flooded across the EVI network, the EVI ED listens to ARP responses on the EVI Link. These ARPs are cached for the remote MACs so that subsequent ARP requests can be handled directly by the ED.

In the figure above:

1. Host B sends an ARP for Host A’s MAC
2. The ED at site IP3 floods the request out of all ports, including the EVI-tunnel interfaces (vLinks). The remote EDs decapsulate the packet and flood at their respective sites
3. Host A responds to the ARP request
4. Site IP3 caches the entry for host A and forwards the response to host B
5. Any subsequent ARPs for host A are now handled locally by the RD at site IP3 rather than repeating the process

EVI with IRF

At each location, HA is achieved by configuring IRF on the HP 12500 Switch Series to simplify the network topology and increase uptime. IRF can also be configured in conjunction with MDC.

EVI and MDC

MDCs can be configured with IRF in an EVI environment to create completely secure isolation between tenants. Up to 32 EVI networks can be deployed in each MDC and each MDC also has a completely functional L2 and L3 environment.

Figure 6. HP EVI and MDC key takeaways

Transport agnostic

- No change on the customer infrastructure
- No specific dependency on the transport network

Multitenant enabled Layer 2 DC extension

- Up to four MDC per physical system
- Each MDC operates independently 32 EVI networks

Simplified operation and management

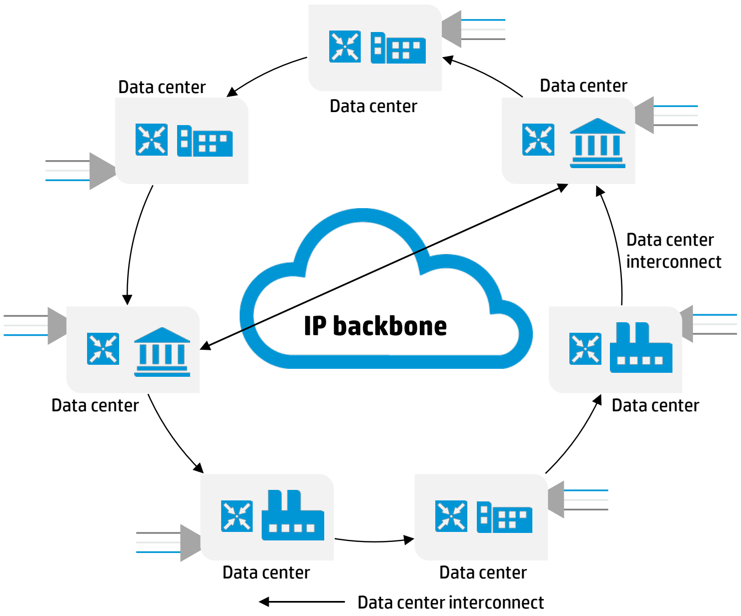
- Five steps to enable EVI
- Reduce the workload to from months to minutes
- Integration of L3 GW and L2 extension

Automatic HA and failure isolation

- Native IRF supports for HP

Scalability

- Scale up to eight data centers
- Scale up to 4K VLANs
- Scale up to 32 EVI networks per MDC



DCI design guidelines for DWDM

Optical Wave Division Multiplexing (WDM) is the technology of choice to address the high-bandwidth, low-latency, and multiprotocol transport requirements of mission-critical data across enterprise metro and long-haul DCI networks. Satisfying these requirements with versatility and scale, allows enterprises to realize lower IT costs, while improving network reliability and performance for DCI.

DCI built on DWDM provides the ultimate solution to address the concerns of data center evolution. DWDM with T-ROADM (tunable/reconfigurable optical add/drop multiplexer) provides the highest level of scale and flexibility for switching data center traffic, along with metro and long-haul reach, high reliability and operations, and administration and maintenance (OA&M). DWDM delivers low latency across a predictable infrastructure of high-speed optical channels from 10G to 100G and beyond, and provides the lowest overall TCO of any DCC solution.

Starting with a simple point-to-point fixed configuration with low-cost optical interfaces for small scale DCI needs, the solution can be further enhanced by providing tunable optics and colorless optical transponders. This can be for any wavelength to be connected to and from any direction across the optical infrastructure with multidegree any direction Reconfigurable Optical Add-Drop Multiplexing (ROADM). The service provisioning can be automated using Generalized Multi protocol Label Switching (G-MPLS) control signaling, which addresses the real-time dynamic nature required for connectivity between optical DWDM sites. This feature list can then be further enhanced by providing full Ethernet aggregation capabilities.

Security, another critical element in DCI, must be supported by providing encryption capabilities for mission-critical traffic.

Integrated Packet Transport (IPT) over WDM provides LAN extension across multiple data centers, using Ethernet Ring Protection (ERP) G.8032 protocol for resilience.

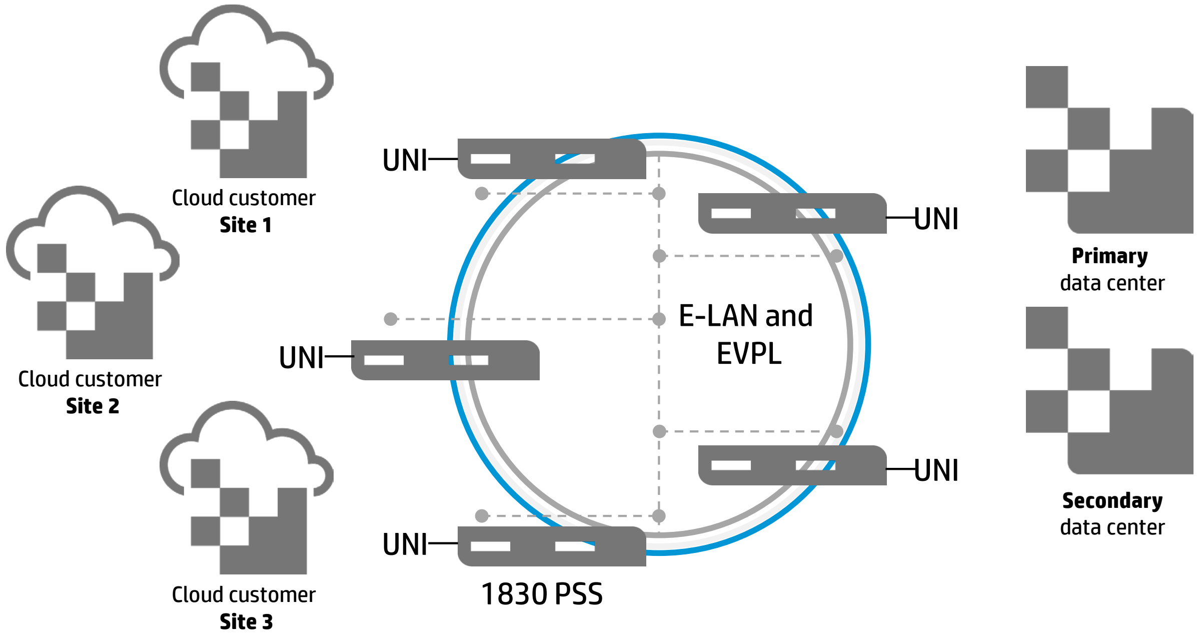
Packet-optimized WDM is used to interconnect cloud consumers to private and hybrid clouds. Metro Ethernet Forum (MEF)-defined CE services supported on the Alcatel-Lucent 1830 PSS provide E-LAN and E-Tree to offer more scalable cloud services with disaster recovery and flexible service reconfiguration. These features and capabilities support the on-demand, bursty dynamic nature of cloud-based services in terms of service provisioning/configuration flexibility, class of service differentiation, security, and reliability.

The integrated packet switching capability enables increased operational efficiency, scalability, network optimization, and bandwidth utilization when compared to point-to-point/multipoint interconnection of external switches/routers. Allows network optimization in terms of bandwidth and lowers the number of interfaces on hub nodes.

MEF-based services can be used to address different cloud models:

- E-LAN for private cloud WAN solution
- E-Line or E-Tree for private cloud, community cloud, or hybrid cloud WAN solution
- E-LAN and E-Tree facilitate data replication, disaster recovery, and cloud bursting—making it simpler to add additional cloud customers as opposed to using E-Lines

Figure 7. Integrated packet transport on DWDM for cloud services: E-LAN for private cloud WAN



The highest level of resilience is reached using:

- ERP with sub 50 ms traffic recovery both for fiber cut or complete node failure
- Multichassis Link Aggregation (MC-LAG) for link level redundancy, card level redundancy, and node level redundancy of the interconnection between WDM and CE devices. MC-LAG doesn't require specific feature support on CE devices but the standard link aggregation protocol

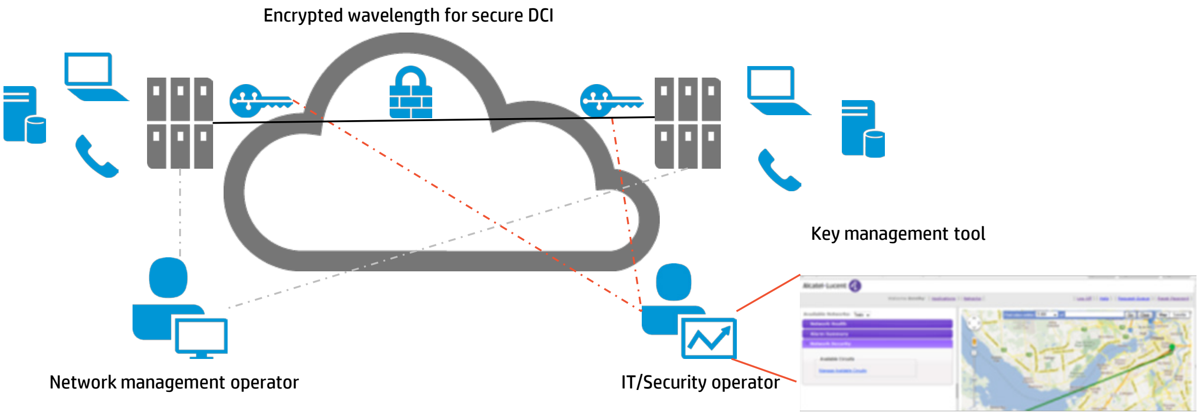
Design guidelines for secure DCI with DWDM

While encryption in the higher layers of the Open Systems Interconnection (OSI) network stack can be effective in certain situations, such encryption can be complex—resulting in high CPU utilization, increased latency and overhead—and can suffer from problems with compatibility between OSI network layers. L1 physical layer encryption is, therefore, the preferred method for securing data across the DCC WAN.

The Alcatel-Lucent 1830 PSS addresses these needs thanks to embedded security features:

- In-flight and ultra-low latency AES 256 encryption @10G applicable to any protocol (Ethernet, FC, and IB)
- Optical intrusion detection (OID) used to monitor attempts of fiber tapping
- Role separation between optical network infrastructure administrator and security administrator responsible of encryption management

Figure 8. Alcatel-Lucent secure DCI


Encrypted wavelength for secure DCI
Key management tool
Network management operator
IT/Security operator

The Alcatel-Lucent 1830 PSS implements the National Institute of Standards and Technology (NIST) Advanced Encryption Standard (AES) block cipher to perform symmetric L1 encryption. This cipher can encrypt data very quickly, and it is extremely difficult to break when large key sizes are used. The Alcatel-Lucent 1830 PSS uses integrated hardware and robust 256-bit AES keys to encrypt data flows and deliver securely transported information. Because encryption and decryption of the blocks is done using the same key in the devices and keys, the algorithm is called symmetric.
The Alcatel-Lucent 1830 PSS encryption module was designed and tested using FIPS 1402 standards, including detailed requirements for strong cryptographic algorithms and physical device protection from NIST.

DCI design guidelines for VPLS

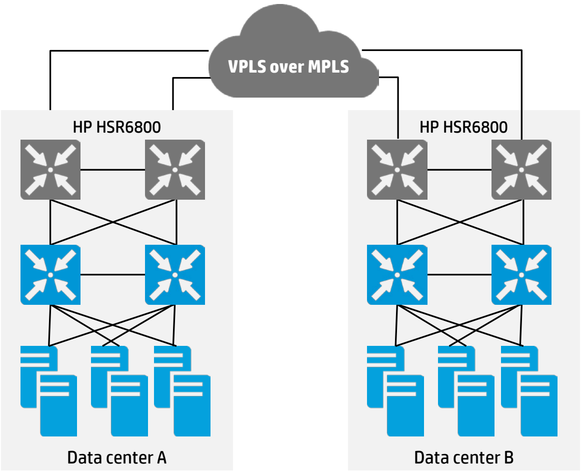
VPLS overview

VPLS controls packet forwarding by using two layers of labels and can implement point-to-multipoint DCI connections. With VPLS, you can simulate an Ethernet switch among multiple data centers in the MPLS network to make inter-L2 port forwarding decisions based on MAC address or the combination of MAC address + VLAN ID. A VPLS instance for implementing L2 DCI contains multiple data centers, which are connected to multiple DCI devices (could be viewed as customer-owned PE devices). Data center aggregation layer switches directly communicate with the other aggregation layer switches associated with the VPLS instance.

Such a design is relevant in these two scenarios:

- Enterprise customer owns or manages their own L1 and VPLS network
- Enterprise customers acquire a L1 or L2 type of service from a provider, and VPLS is run between the enterprise PE devices at the edge of the provider's cloud

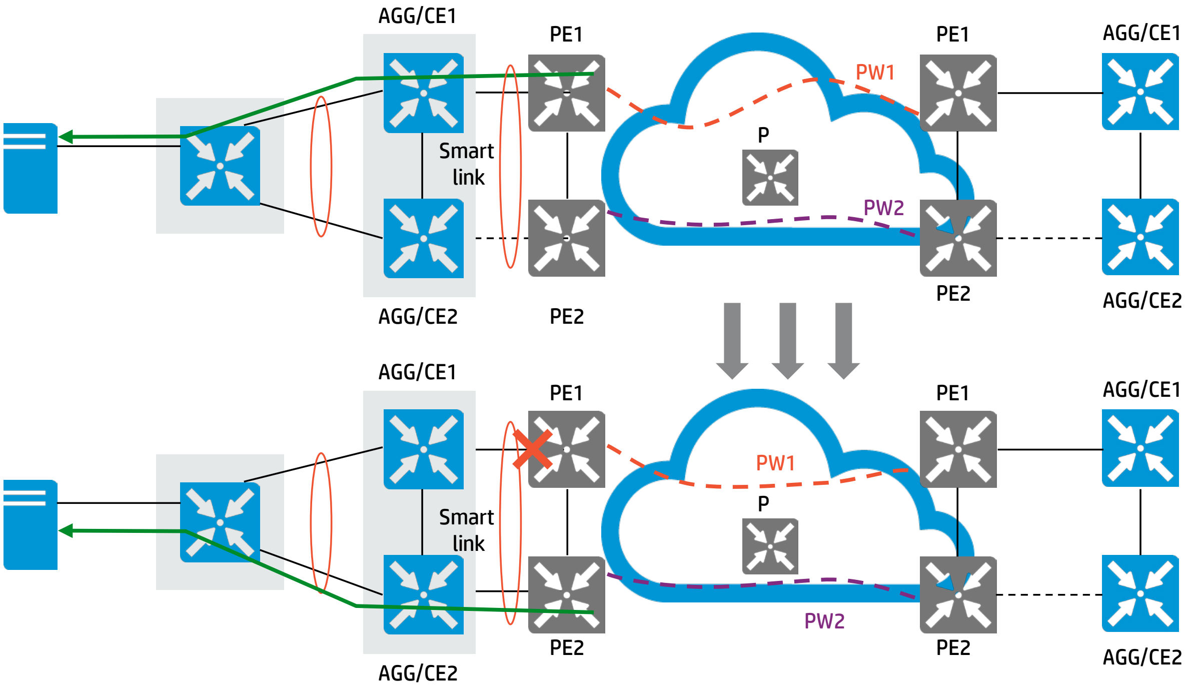
Figure 9. L2 DCI solution for interconnecting two data centers by using MPLS and VPLS



Introduction to smart link: Router-based solution

To implement dual-homing from the aggregation layer to routers, which do not support IRF, you can use the smart link feature to connect to the routers through the links of a smart link group. If a router or an aggregation layer to router link fails, smart link can implement 50 ms convergence, and rapidly switch the traffic to the other link.

Figure 10. Smart link solution for dual-homing CEs

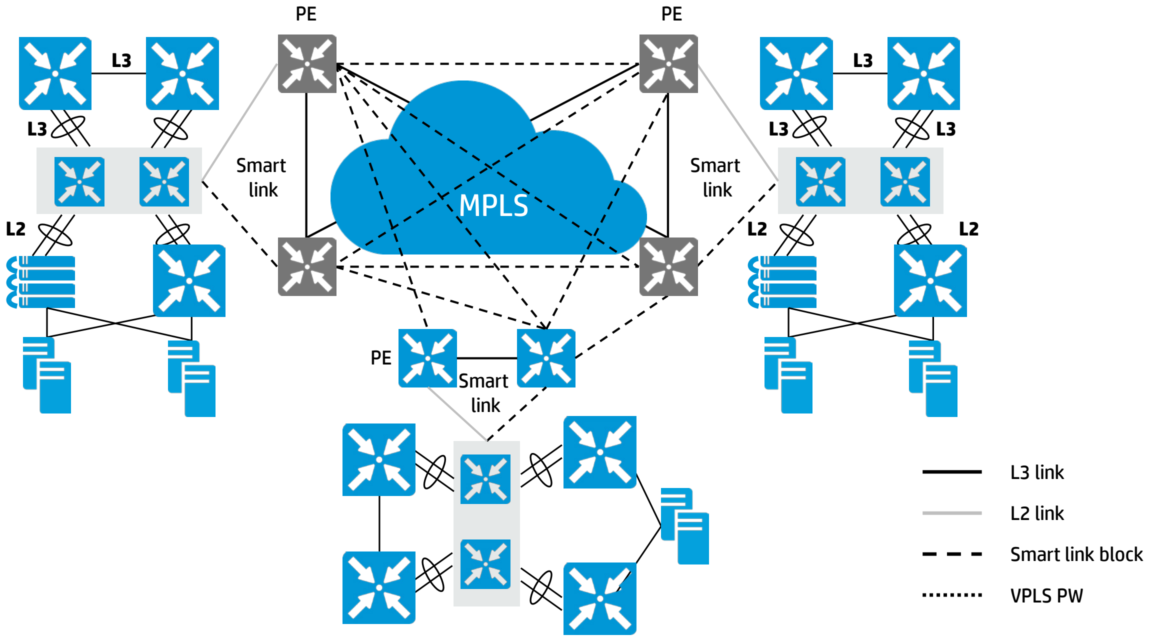


Interconnecting multiple data centers using VPLS + smart link
In the VPLS network, consider the following two factors with caution:

- Some MPLS networks do not support Ethernet interfaces. Therefore, in those scenarios you must use routers to connect to the network.
- To improve the HA performance, you must implement dual-homing for the routers.

Considering the two factors, HP recommends that you use the VPLS + smart link solution.

Figure 11. VPLS + smart link solution



In the VPLS + smart link solution, use the existing MPLS network as the core network, configure two routers (could be viewed as customer-owned PE devices) for each data center, and configure the aggregation layer devices (IRF-capable devices) of each data center. Configure smart link on the aggregation layer to implement 50 ms convergence for the two links connecting to the routers.

Set up full-mesh connections for the routers.

Note:
The HA performance of the VPLS network depends on the HA design of the MPLS network. Therefore, if the customer owns the MPLS network, HP recommends that you simplify the architecture to facilitate the traffic path planning.

The VPLS + smart link solution delivers the following benefits:

- High standardization level, interoperability, and compatibility with most MPLS networks
- Fast switchover of smart link, which improves the HA performance of the system

The VPLS + smart link solution has the following disadvantages:

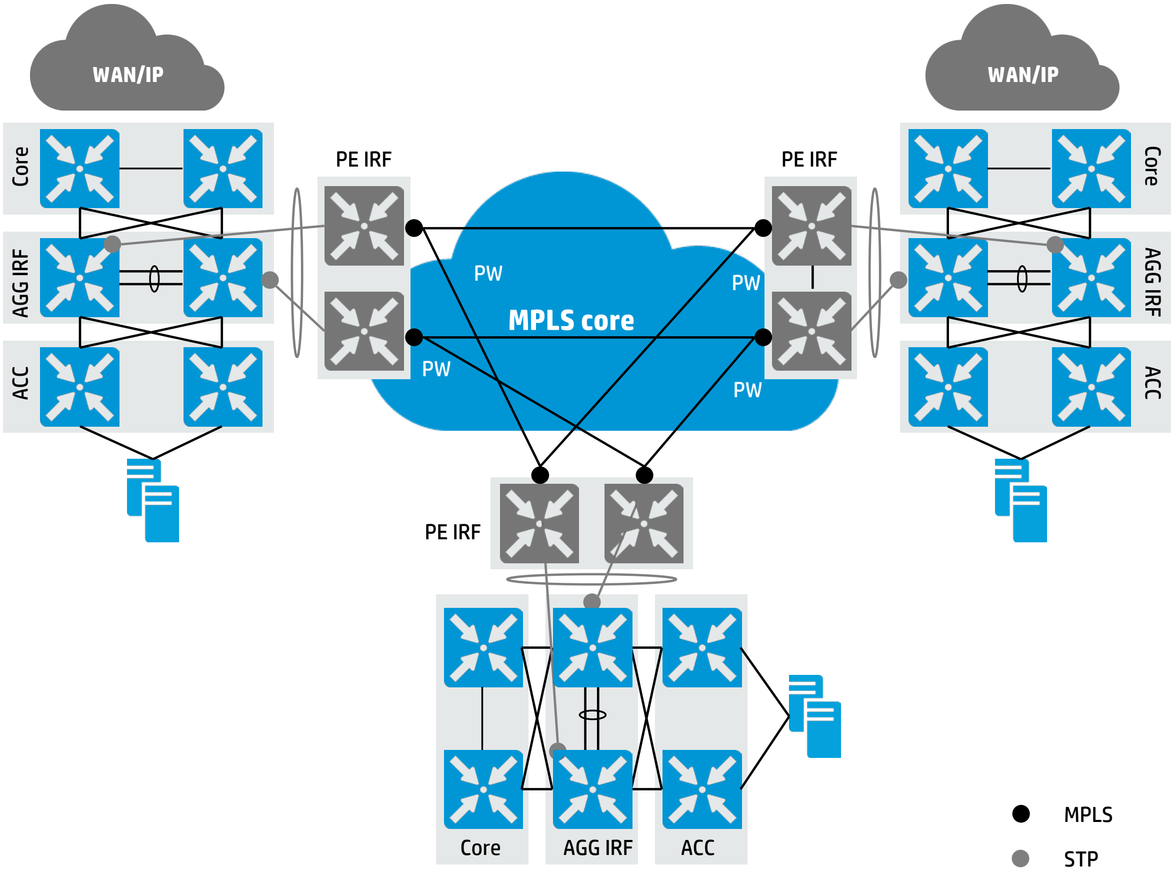
- The routers do not support IRF. As a result, there are many routers and a large volume of broadcast traffic in the network
- Smart link cannot sense the link failures at the outbound sides of the router PEs and may result in black hole routes

Interconnecting multiple data centers using VPLS + IRF + LACP
Suppose a VPLS network provides the following conditions:

- The MPLS network supports Ethernet interfaces. Therefore, you can use dedicated IRF-enabled switches (could be viewed as customer-owned PE devices) for connecting to the MPLS network.
- The users need dual-homing aggregation layer devices to improve the HA performance.

In this case, HP recommends that you use the VPLS + IRF+ LACP solution.

Figure 12. VPLS + IRF + LACP solution



In the VPLS + IRF + LACP solution, use the existing MPLS network as the core network, configure two IRF-capable switches acting as the dedicated DCI devices for each data center (could be viewed as customer-owned PE devices), and configure the IRF-capable aggregation layer devices of each data center. Configure link aggregation between the aggregation layer and dedicated DCI devices to implement millisecond-level convergence for the failure of any node or any link between the layers.

Set up full-mesh connections for the dedicated DCI devices.

Note:
The HA performance of the VPLS network depends on the HA design of the MPLS network. Therefore, if the customer owns the MPLS network, HP recommends that you simplify the architecture to facilitate the traffic path planning.

The VPLS + IRF solution delivers the following benefits:

- Per-layer IRF design, which has a simple architecture and is easy to maintain
- Per-layer IRF design, which provides high HA performance and implements millisecond-level convergence for the failure of any node or any link
- Implementing IRF on the dedicated DCI devices decreases the number of nodes in the MPLS network and the number of broadcast packets
- Per-layer IRF design implements load sharing for all links, improves the bandwidth utilization, and improves the performance of the whole network

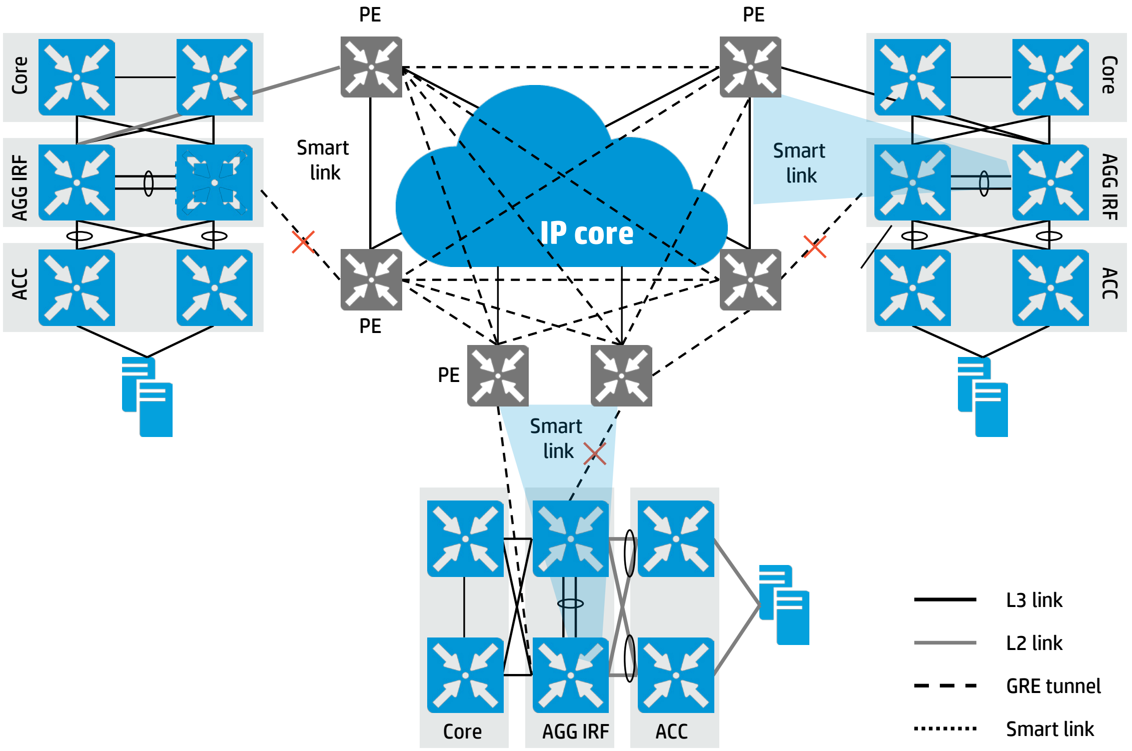
DCI design guidelines for IP-based solution

VPLS over GRE overview

When only IP networks are used for interconnecting data centers, you can use both EVI and VPLS over GRE (VPLSoGRE) to implement L2 DCI. This VPLSoGRE design solution is only relevant in those scenarios, where an enterprise acquires an IP type of service from a provider and VPLSoGRE is run between the enterprise PE devices at the edge of the provider's cloud.

The VPLSoGRE design requires a tunnel that can carry multiple pseudo wires (PWs) between the various dedicated DCI devices at each data center. The tunnel can be an MPLS tunnel or a GRE tunnel. The MPLS tunnel implementation of VPLS uses an outer MPLS label, while the GRE tunnel implementation of VPLS uses VPLS + GRE. Therefore, the network topology and L2 loop prevention of the VPLSoGRE solution are similar to those of the VPLSoMPLS solution.

Figure 13. The VPLSoGRE solution



In the VPLSoGRE solution, use the existing IP network as the core network, configure two routers as the dedicated DCI devices (could be viewed as customer-owned PE devices) at each data center, and configure the IRF-capable aggregation layer devices of each data center. Configure smart link on the aggregation layer devices to implement 50 ms convergence for the two links connecting to the dedicated DCI devices.

Set up full-mesh GRE tunnel connections for the dedicated DCI devices (can be viewed as customer-owned PE devices).

The VPLSoGRE solution delivers the following benefits:

- High standardization level, powerful compatibility, and compatibility with all IP networks
- Fast switchover of smart link, which improves the HA performance of the system

The VPLSoGRE solution has the following disadvantages:

- The dedicated DCI devices do not support IRF. As a result, there are many dedicated DCI devices and the PW configuration is complicated

The IP network provides low-service quality and you must configure network-wide, end-to-end QoS, which can be very difficult.

Summary

EVI is the preferred DCI method from HP and stands out when compared to other vendors like Brocade, Cisco, and Juniper.

Table 1. HP EVI—Unmatched in the market

DCI feature	Brocade	Juniper	Cisco	HP
Layer 2 routing extension	No	No	Yes—OTV	Yes—EVI
Layer 2 routing extension scaling	N/A	N/A	512 VLANs 6 data centers	4K VLANs 16 data centers
Layer 2 routing extension with multitenancy	No	No	No	Yes EVI and MDC
Multiple MPLS/VPLS-based and Layer 2 DCI on one switch	No	No	No	Yes HP 12500

For data center clients who have multiple data centers, Cisco and HP are the only vendors that offer a L2 DCI that doesn’t rely on MPLS and allows for multiple network instances.

Table 2. HP EVI advantages vs. Cisco OTV

DCI feature	VPLS	Cisco OTV	HP EVI
End-to-end loop-free without STP	√	√	√
Failure domain isolation	√	√	√
Multipathing and load balancing			√
Independent of infrastructure	MPLS required	Multicast required by default	√
Multi-DC interconnectivity	√	√ Up to 6	√ Up to 16
Active/active physical redundancy			√
Large number of VLANs	256	√ 512	√ 4K
Number of network instances		10 per Nexus 7000	32 per MDC

When utilizing DCI solutions other than EVI, select the proper solution by considering the performance, HA, Layer 2 management, and security factors.

Table 3. Traditional DCI performance, HA, Layer 2 management, and security factors

Your network resource	Solution	Performance	HA	Layer 2 management	Broadcast control	Security
DWDM	Integrated Packet Transport (IPT)–ALU	High	High	High	High	High
MPLS core network	VPLS + IRF	Medium	Medium	Medium	Medium	Low
	VPLS + smart Link	Medium	Medium	Medium	Low	Low
IP core network	VPLSoGRE	Medium	Low	Medium	Low	Low

Partnering architecture

HP DCI and the Alcatel-Lucent 1830 PSS

Today's enterprises expect anywhere, anytime access to storage and computing from their service providers. Providers are, therefore, deploying virtual and shared resources across geographically diverse data centers—all connected with a high-speed, redundant mesh of links. These metro and long-haul links must deliver a high-quality experience to enterprises over provider's data center connect (DCC) networks.

Additionally, today's enterprises recognize that their data centers are continuously at risk from internal and external security threats. More than simply deploying antivirus and firewall defenses, enterprises must establish comprehensive IT security programs that protect a virtual and distributed environment of computing and storage resources.

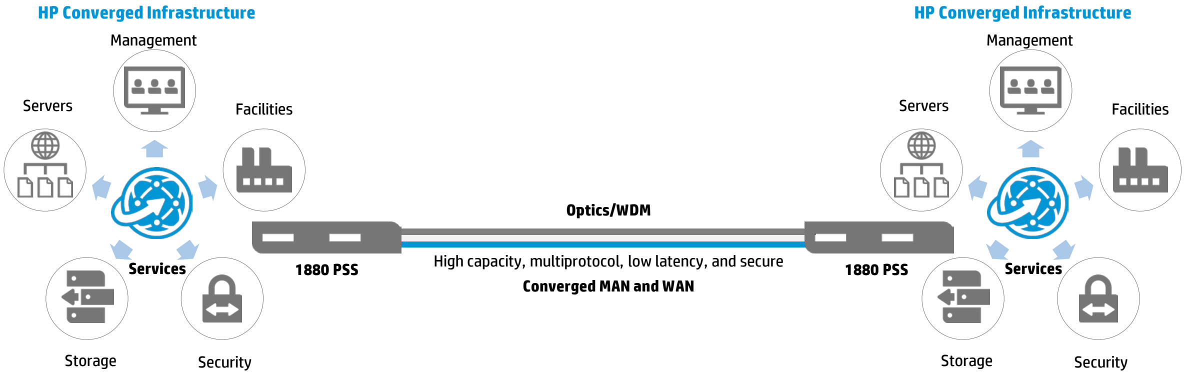
For these distributed resources to work effectively and meet diverse end-user requirements, applications require secure, low-latency real-time communications with guaranteed QoS. The DCI infrastructure and optical fibers used for DCI transport are key components of a holistic and systematic IT security program.

Alcatel-Lucent offers a versatile, scalable, secure, and easy to use photonic service fabric for the DCI and plays an important role in enabling cloud computing based services. The Alcatel-Lucent 1830 PSS is the networking product of choice for enterprise and communication service provider DCI solutions. Through versatile, scalable, and secure configurations, it allows enterprises to realize lower IT costs while improving network reliability, performance, and security.

With a unique ability to seamlessly "virtualize" data center and networking resources that facilitate a cloud services model, HP and Alcatel-Lucent are enabling enterprises and communication service providers to reduce overall IT costs, improve business continuity and agility while reducing space, power consumption, and management costs. The 1830 PSS is an integral part of the HP and Alcatel-Lucent alliance for DCI solutions.

Note:
Photonic sometimes also referred to as Optical or Wavelength Demultiplexing (WDM). WDM virtualizes the physical fiber infrastructure from one into many, and in case of DWDM up to 88 virtual fibers per physical fiber.

Figure 14. Converged MAN and WAN with ALU 1830 PSS



Key business problem solved

Enterprises face major inter data center networking challenges as bandwidth demands for storage applications continue to grow. There are four important market trends that shape the future of DCI and its challenges to enterprises:

- Continued improvement of business continuity and disaster recovery (BCDR) capabilities (solution needs to be latency optimized):

  Improving BCDR capabilities continue to be one of the top priorities for IT decision makers. Enterprises are also moving towards cloud infrastructure services to realize capital and operational cost savings. Some enterprises will pursue a strategy of implementing cloud infrastructure services delivered by communication service providers for increased elasticity. Therefore, although enterprises are moving towards public Cloud Infrastructure services, they continue to prefer dedicated infrastructures to protect their mission-critical data. Enterprises will continue to be challenged in meeting the performance, reliability, and bandwidth requirements of BCDR for mission-critical data using their private data center interconnect. As public cloud services mature, they could become a viable low-cost alternative for enterprise's less critical applications and systems (increasing the elasticity).

- Growth of mission-critical data (solution needs to provide dynamic and flexible bandwidth):

  Enterprises recognize that more and more applications are critical. Enterprises consider email and collaborative applications such as SharePoint also critical applications because applications ecosystems have become more complex and developed interdependencies that force less critical application data to have the same degree of availability as mission-critical application data.

  As storage data will continue to grow and the percentage of data considered critical increases, the growth of critical data will grow at a higher rate. The growth of mission-critical data increases the bandwidth required by enterprises to replicate the information synchronously within the boundary of a Metropolitan Area Network (MAN).

- Secure mission-critical data (solution needs to support in-flight encryption and intrusion monitoring):

  In a world filled with electronic security threats, companies are recognizing that their data centers are at continuous risk. Security threats to the data center arise not just from traditional malware or hobbyist hackers, but increasingly from criminal organizations that are directly targeting the enterprise. Multisite data center security is not just about technical countermeasures such as antivirus and firewalls used inside the data center, but a much more systematic and holistic approach to enterprise-wide security. Enterprises must establish a comprehensive and coordinated counterattack, implementing solutions that provide detection and mitigation of security threats.

  Detection of security breaches requires the deployment of embedded security monitoring technology in network devices that can detect intrusions even when the traffic traversing the network is undisturbed.

  Encryption of data is required to mitigate the impact of security breaches. It is the most effective method of mitigation as it renders stolen data useless to the intruder. Encryption is the algorithmic process of transforming data into unreadable cryptographic text. Encryption is no longer an exotic mechanism whose use is limited to secret organizations, it is a common tool used as part of normal business workflow for security.

- Consolidation of data centers (solution needs to support MAN and WAN distance optimization):

  Data center consolidation is one of the most fundamental ways of lowering the cost of IT operations. Larger data centers are simply more cost-effective on a per-unit basis. Enterprises are consolidating their existing data centers to new strategic locations outside major cities where real estate costs are lower.

Enterprises are also searching for proximity to greener energy sources (hydro dams and improved solar or wind locations) to realize lower energy costs. These larger data centers hold larger volumes of storage data that need to be shared with remote data center sites geographically apart.

Enterprises cannot look at the benefits of data center consolidation without considering the networking challenges they bring. The need to communicate effectively over the MAN and WAN becomes a critical factor in achieving the lower costs of IT operations promised by data center consolidation.

- Introduction of Federated Storage technology (solution needs to support meshed networks):

  Federated Storage is the collection of autonomous storage resources that form a scale-out cluster governed by common software that manages the rules of the federation: how disk capacity is shared, joined, and presented to the hosts.

  Storage federation is implemented in the storage arrays or specialized appliances that sit in the data path between hosts and storage ports. With this technology, enterprises can scale-out storage resources over distance to create networks of virtually unlimited capacity, improve BCDR, eliminate disruptive migrations, and allow applications to share a storage volume over local, metro, or global distances.

  Federated Storage solutions have similar interconnect requirements to BCDR solutions in terms of latency and multiprotocol support. In addition, data migration applications introduce the challenge to dynamically allocate bandwidth for the duration of the data migration process.

Use cases and potential customer profile

The joint HP and Alcatel-Lucent target market will typically consist of larger enterprises and communication service providers looking to HP and Alcatel-Lucent leadership to offer both intra and inter data center networking solutions. Cost reduction isn’t their only driver for data center consolidation. They want to create competitive advantage by moving to a cloud services model and will be looking for single/minimal suppliers to achieve that aim. The following use cases, among others, have been identified:

Data center interconnect

Enterprise

- Own and manage your private transport
- Own and manage your secure private transport

Communication service provider

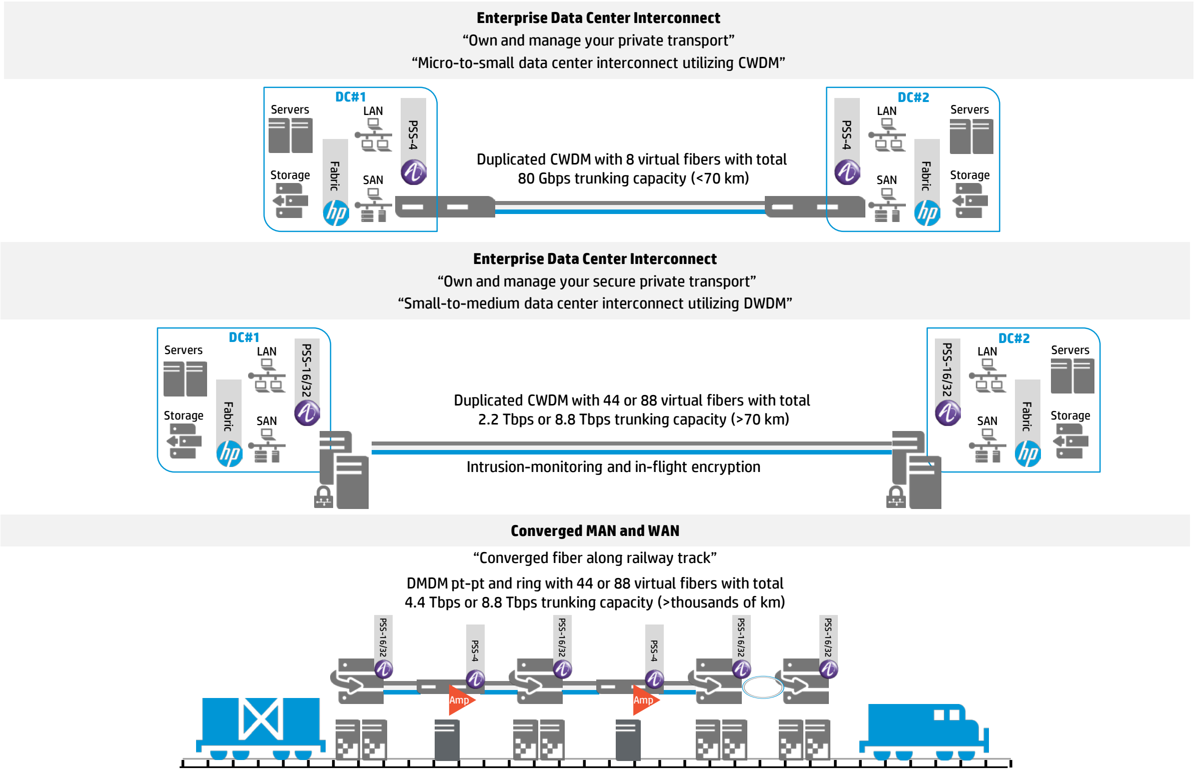
- Managed private transport services (in case enterprise only wishes to own the private transport network)
- Managed transport services (in some cases known as leased Lamda services or managed wavelength services)
- Managed secure private transport services (in case enterprise only wishes to own the secure private transport network)
- Managed secure transport services

Converged MAN and WAN

The Converged MAN and WAN is essentially a broader set of use cases not necessarily relating to pure DCI. Example in the utility and vertical space:

- Power utility in need for reliable mission-critical and multiprotocol transport (such as the Alcatel-Lucent reference case Statnett in Norway)
- Railway utility with separated physical fiber infrastructure (could be a different company) in need of flexible mission-critical and multiprotocol transport
- Road utility in need for Converged IP and WDM low-latency transport (such as the Alcatel-Lucent reference case Highways Agency in England)

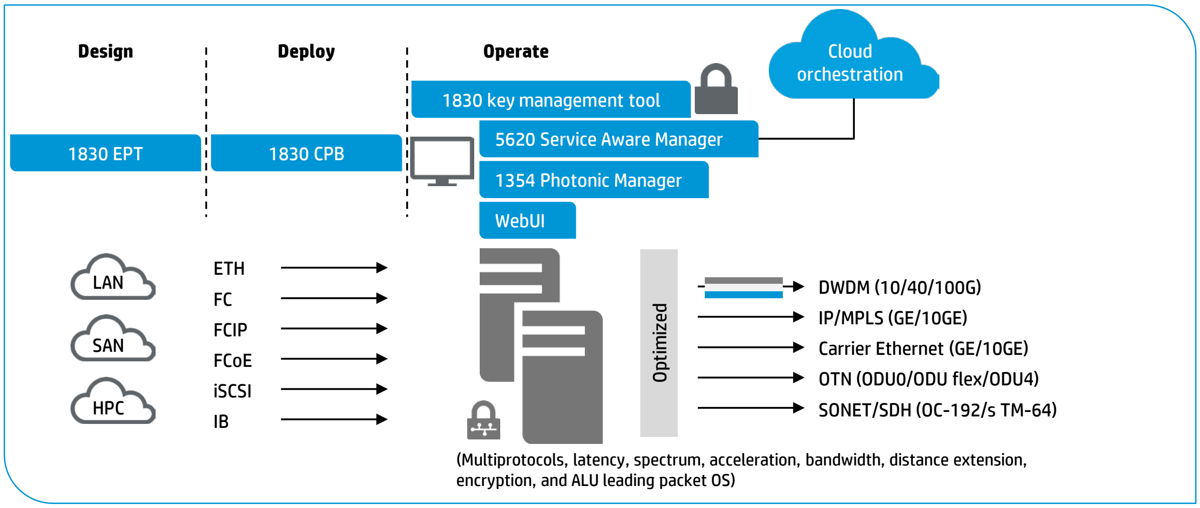
Figure 15. ALU 1830 PSS use cases


Enterprise Data Center Interconnect
"Own and manage your private transport"
"Micro-to-small data center interconnect utilizing CWDM"
DC#1
Servers
LAN
PSS-4
Storage
Fabric
SAN
Duplicated CWDM with 8 virtual fibers with total
80 Gbps trunking capacity (<70 km)
DC#2
PSS-4
LAN
Servers
SAN
Fabric
Storage
Enterprise Data Center Interconnect
"Own and manage your secure private transport"
"Small-to-medium data center interconnect utilizing DWDM"
DC#1
Servers
LAN
PSS-16/32
Storage
Fabric
SAN
Duplicated CWDM with 44 or 88 virtual fibers with total
2.2 Tbps or 8.8 Tbps trunking capacity (>70 km)
Intrusion-monitoring and in-flight encryption
DC#2
PSS-16/32
LAN
Servers
SAN
Fabric
Storage
Converged MAN and WAN
"Converged fiber along railway track"
DMDM pt-pt and ring with 44 or 88 virtual fibers with total
4.4 Tbps or 8.8 Tbps trunking capacity (>thousands of km)
PSS-16/32
PSS-4
PSS-16/32
PSS-4
PSS-16/32
PSS-16/32
Amp
Amp

What is an Alcatel-Lucent 1830 PSS?

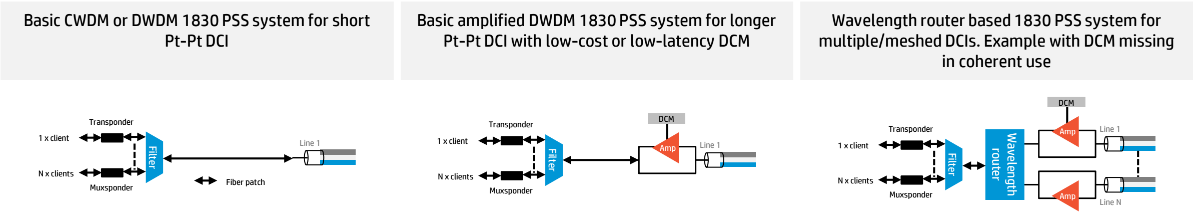
The Alcatel-Lucent 1830 PSS is a versatile, scalable, secure, and easy-to-use photonic service fabric, but how does the configuration provide for this?

Figure 16. Alcatel-Lucent 1830 PSS



The client (SAN, LAN, or HPC) signal is connected to a universal Transponder/Muxsponder that converts and aggregates the multiprotocol client signals to a CWDM or DWDM wavelength. The wavelength is connected to a wavelength multiplexer (Filter)/Tuneable ROADM port that multiplexes all wavelengths onto a line fiber connected to the other data center. In case of longer distances, all wavelengths are amplified and typically compensated for dispersion. All connections between the discrete modules are provided by fiber optics patches supporting an unmatched scalability and flexibility in terms of configuration.

Figure 17. Alcatel-Lucent 1830 PSS deployment options



Basic CWDM or DWDM 1830 PSS system for short Pt-Pt DCI

Basic amplified DWDM 1830 PSS system for longer Pt-Pt DCI with low-cost or low-latency DCM

Wavelength router based 1830 PSS system for multiple/meshed DCIs. Example with DCM missing in coherent use

Why Alcatel-Lucent 1830 PSS?

Versatile:

- Multilayer cloud service fabric agility facilitates ease of response to unpredictability
- Low-latency, optimized enabling application monetization
- Protocol and bit-rate agnostic for a variety of LAN, SAN, and HPC DCI applications
- Unconstrained topology support including Pt-Pt, ring, and mesh configurations
- De-risked deployments with certified partner solutions
- Contentionless, colorless, directionless multidegree optical switching (Wavelength Router) supporting any-to-any data center interconnectivity
- Enable seamless mobility and high availability of storage data across data centers
- Hands-free operation based on autonomous programmability of the network elements using controller-based administration and management
- Evolutionary cloud modes of operation offering investment protection
- Optimized TCO through 1.5x density and 2/3x power at 100G

Scalable:

- Support for 10G to 40G to 100G to 200G with a path to 400G with the Photonic Service Engine (PSE) for growth
- Flexibility of choice in networking capacity with C/DWDM, ROADM, flexible grids, OTN, GMPLS, and packet technologies
- Right-sized platform commonality across any premise from branch to HQ to DC
- Cost-optimized cloud service levels enabled by flexible configuration of increasing network availability
- Cost-optimized metro, regional, and global network implementations supported
- High-performance fiber capacity up to 8.8 Tb (17.6 Tb @400G)
- Lower cost interconnection through transparent and statistical multiplexing
- Mixed 10G/40G/100G/200G wavelength compatibility within same fiber pair
- Enhanced application performance enabled by Ethernet fabric extension into the packet optical transport

Secure:

- In-flight security mitigation through AES 256 encryption
- Security breach detection through optical intrusion detection (OID)
- Security attack prevention through secure configuration operational mode
- Assurance of specification and proven implementation via CC/FIPS certifications
- Assured secure networking achieved with fully disassociated user roles
- Customer controllable key management enabling differentiated services

Easy to use:

- Plug and play, self-restoring managed photonics
- Lower TCO due to FCAPS/FAB commonality derived from extended packet OS
- Elastic environment delivering next-generation cloud services employing pre-integrated orchestration northbound interfaces for flow-through provisioning
- Common practices derived from a unified network management platform
- Measurement, monitoring, recording, and reporting of SLA attributes through continuous monitoring on QoS enabling a cloud-ready service fabric
- Simplified engineering and planning resulting from equipment supporting multiprotocols and multiple bit-rates in addition to streamlining equipment and sparing costs

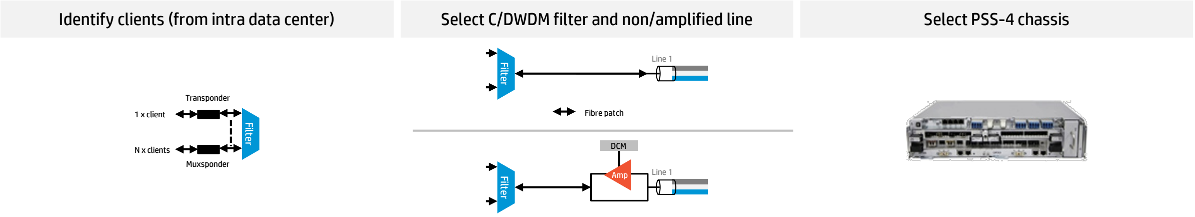
Which Alcatel-Lucent 1830 PSS product to position?

Right-sized platform (Scalable)

- 1830 PSS-4: "Micro-to-small DCI utilizing CWDM":

The 1830 PSS-4 offers four half-height slots or two full-height slots that support C/DWDM fixed OADM point-to-point applications. It supports AC or DC power options. The PSS-4 shelf can be stacked in multishelf applications for increased low-end scalability. It can also be used as a small amplifier in case of a longer distance Converged WAN.

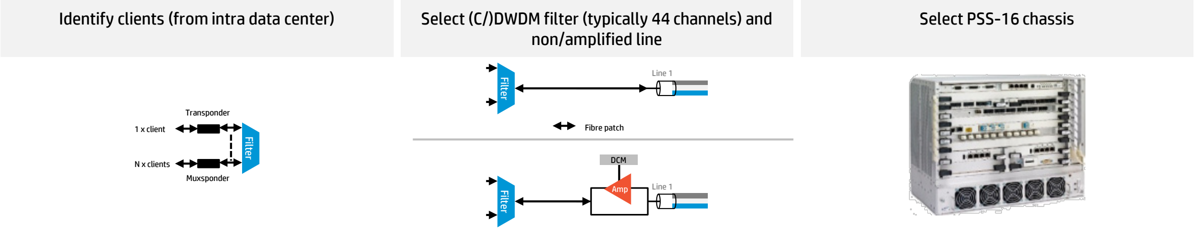
Figure 18. Micro-to-small DCI utilizing CWDM



- 1830 PSS-16: "Small-to-medium DCI utilizing DWDM":

The 1830 PSS-16 platforms offer sixteen half-height slots or eight full-height slots. It typically is used for DWDM FOADM ring or Pt-to-Pt applications. It can also be used as a large amplifier in case of a longer distance converged WAN.

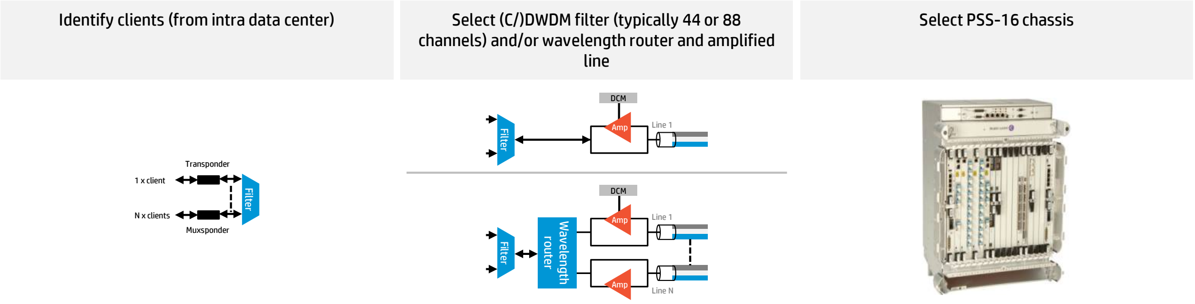
Figure 19. Small-to-medium DCI utilizing DWDM


Identify clients (from intra data center)
Select (C/)DWDM filter (typically 44 channels) and non/amplified line
Select PSS-16 chassis
Transponder
1 x client
N x clients
Muxsponder
Filter
Filter
Line 1
Fibre patch
DCM
Amp
Filter
Line 1

- 1830 PSS-32: "Medium-to-large DCI utilizing DWDM":

The 1830 PSS-32 platforms offer 32 half-height slots or 16 full-height slots. It typically is used for DWDM reconfigurable OADM ring, mesh, or Pt-Pt applications.

Figure 20. Medium-to-large DCI utilizing DWDM


Identify clients (from intra data center)
Select (C/)DWDM filter (typically 44 or 88 channels) and/or wavelength router and amplified line
Select PSS-16 chassis
Transponder
1 x client
N x clients
Muxsponder
Filter
DCM
Amp
Filter
Line 1
DCM
Amp
Line 1
Filter
Wavelength router
Amp
Line N

Note:
For more information on the 1830 PSS chassis and variants, find datasheets on alcatel-lucent.com by searching for "1830 PSS".

Additional links

HP Networking

hp.com/go/networking

HP Data Center Solutions

hp.com/go/flexfabric

Alcatel-Lucent

alcatel-lucent.com/solutions/data-center-connect

Learn more at
hp.com/networking/dci

Sign up for updates
hp.com/go/getupdated

Share with colleagues

★ Rate this document

4AA5-5624ENW, November 2014