Data Center and Disaster Recovery Solution for Vietnam Operations

Design Option 1: Centralized Production DC with Active-Passive DR

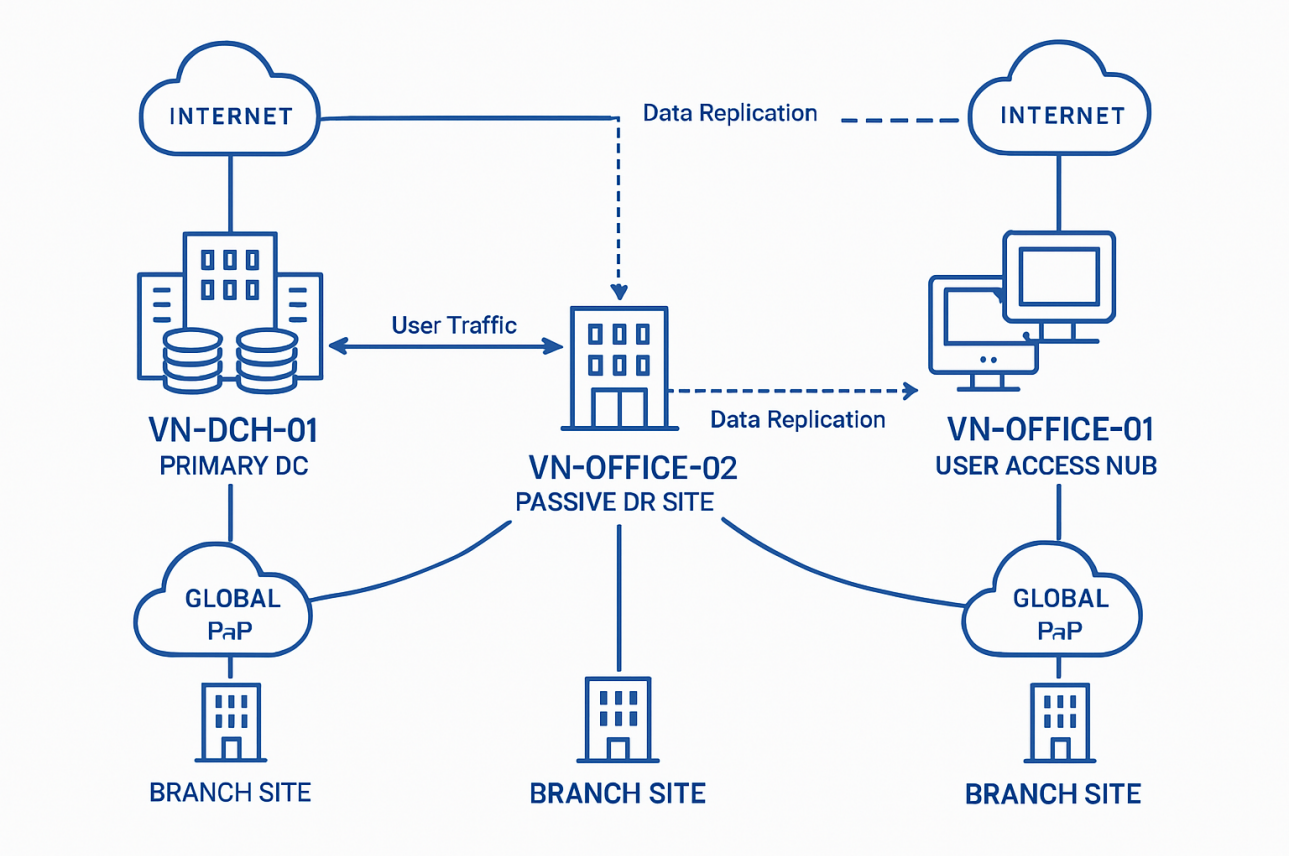
High-Level Strategy

This design option proposes a centralized production data center (DC) in a new colocation facility (VN-DCH-01) with an active-passive disaster recovery (DR) strategy. In this model, VN-DCH-01 will serve as the primary operational hub, hosting all core servers, applications, and critical network equipment. VN-OFFICE-01 will transition to a role primarily focused on user access and local services, while VN-OFFICE-02 will serve as the passive DR site, maintaining a synchronized replica of critical data and systems from VN-DCH-01. The smaller branch sites (Sites 3, 4, 5) will continue to connect to the main operational hubs for their network services.

The core concept behind this strategy is to consolidate primary IT infrastructure into a purpose-built, secure, and resilient colocation data center, thereby enhancing operational efficiency, security, and scalability. The active-passive DR approach at VN-OFFICE-02 ensures business continuity by providing a warm standby environment that can take over operations in the event of a major disaster affecting VN-DCH-01. This approach prioritizes cost-effectiveness and simplicity in disaster recovery, as the passive site does not actively serve traffic during normal operations, reducing the complexity of data synchronization and traffic management compared to an active-active setup. Data replication from VN-DCH-01 to VN-OFFICE-02 will be asynchronous to minimize performance impact on the primary site, while still ensuring a reasonably low Recovery Point Objective (RPO).

Network Architecture Diagram

Below is a conceptual network diagram illustrating the proposed centralized production DC with active-passive DR strategy. The diagram highlights the new colocation Data Center (VN-DCH-01), the re-purposed existing offices (VN-OFFICE-01, VN-OFFICE-02), the smaller branch sites, and the placement of global and internet connectivity.



Key elements and their roles:

- **VN-DCH-01 (Primary DC):** This will be the new colocation data center, serving as the primary production site. All critical applications, servers, and data will reside here. It will have a dedicated internet link and one of the Global P2P links terminating here.
- **VN-OFFICE-01 (User Access Hub):** This office will primarily serve as a user access point and local services hub. It will retain an internet link for local internet breakout and user connectivity. User traffic from VN-OFFICE-01 and the branch sites will be routed to VN-DCH-01.
- **VN-OFFICE-02 (Passive DR Site):** This office will function as the passive disaster recovery site. It will house a replicated set of critical data and systems from VN-

DCH-01. The second Global P2P link will terminate here, providing connectivity to the global corporate network during a DR event. Data replication traffic will flow from VN-DCH-01 to VN-OFFICE-02.

- **Branch Sites (Sites 3, 4, 5):** These smaller sites will connect to either VN-OFFICE-01 or VN-DCH-01 for their network and application access.
- **Global P2P Links:** Two dedicated private Point-to-Point circuits connect the Vietnam network to the global corporate network. One terminates at VN-DCH-01, and the other at VN-OFFICE-02, ensuring redundant global connectivity.
- **Internet Links:** Dedicated internet links are present at VN-DCH-01 and VN-OFFICE-01, providing internet access for the respective sites and users.
- **Data Replication Traffic:** Unidirectional traffic from VN-DCH-01 to VN-OFFICE-02, ensuring the DR site is kept up-to-date with the primary site's data.
- **User Traffic:** Traffic originating from users at VN-OFFICE-01 and branch sites, destined for applications and services hosted in VN-DCH-01.
- **Global Network Traffic:** Traffic flowing between the Vietnam network and the global corporate network via the Global P2P links.

Recommended Interconnection Technology:

For interconnecting VN-DCH-01, VN-OFFICE-01, VN-OFFICE-02, and the branch sites, a **Software-Defined Wide Area Network (SD-WAN)** solution is recommended. SD-WAN offers several advantages for this architecture:

- **Centralized Management:** Simplifies network management and configuration across all sites from a single pane of glass.
- **Application-Aware Routing:** Enables intelligent routing of traffic based on application requirements, ensuring optimal performance for critical business applications.
- **Enhanced Security:** Provides integrated security features, including encryption and segmentation, to protect data in transit.
- **Cost Optimization:** Can leverage a mix of transport services (e.g., MPLS, broadband internet) to reduce overall connectivity costs while maintaining performance.
- **Improved Resilience:** Offers dynamic path selection and failover capabilities, enhancing network resilience and ensuring continuous connectivity even during

link failures.

Detailed Perimeter Component Breakdown

This section outlines the necessary network components at each key location (VN-DCH-01, VN-OFFICE-01, VN-OFFICE-02) for Design Option 1, structured by network perimeter.

VN-DCH-01 (Primary DC)

Perimeter	Component Type	Description	Quantity
Edge	Next-Generation Firewalls (NGFWs)	High-throughput firewalls with advanced security features (IPS/IDS, URL filtering, application control) for internet and global P2P connectivity.	2 (Active/Standby)
	Edge Routers	High-performance routers for BGP peering with ISPs and global corporate network, handling internet and P2P link terminations.	2 (Active/Standby)
	SD-WAN Appliances	Centralized SD-WAN orchestrator and local SD-WAN edge devices for intelligent traffic management and secure overlay network.	2 (Active/Standby)
	Global P2P Link Termination	Devices (e.g., routers, media converters) for terminating the dedicated private P2P circuit from the global corporate network.	2 (Primary/Backup)
	Internet Link Termination	Devices (e.g., routers, modems) for terminating the dedicated internet link.	2 (Primary/Backup)
Core	Core Switches	High-density, low-latency Layer 3 switches forming the backbone of the DC network, supporting server infrastructure and SAN connectivity.	2 (Redundant Pair)
	Server Infrastructure	Rack servers, blade servers, or hyper-converged infrastructure (HCI) hosting all	As required

Perimeter	Component Type	Description	Quantity
		critical applications and services.	
	Storage Area Networks (SANs)	High-performance, redundant storage systems (e.g., Fibre Channel, iSCSI) for primary data storage.	2 (Redundant Arrays)
Distribution/Access	Top-of-Rack (ToR) Switches	Layer 2/3 switches providing connectivity to individual servers and network devices within each rack.	As required per rack

VN-OFFICE-01 (User Access Hub)

Perimeter	Component Type	Description	Quantity
Edge	Next-Generation Firewalls (NGFWs)	Firewalls for local internet breakout and secure connectivity to VN-DCH-01 via SD-WAN.	2 (Active/Standby)
	Edge Routers	Routers for internet link termination and SD-WAN connectivity.	2 (Active/Standby)
	SD-WAN Appliances	Local SD-WAN edge devices for intelligent traffic management and secure overlay network to VN-DCH-01.	2 (Active/Standby)
	Internet Link Termination	Devices for terminating the dedicated internet link.	2 (Primary/Backup)
Core	Core Switches	Layer 3 switches providing connectivity for local servers (if any) and aggregating traffic from access layer.	2 (Redundant Pair)
Distribution/Access	Access Switches	Layer 2 switches providing connectivity for end-user devices (desktops, laptops, IP phones) and local network segments.	As required per floor/area

VN-OFFICE-02 (Passive DR Site)

Perimeter	Component Type	Description	Quantity
Edge	Next-Generation Firewalls (NGFWs)	Firewalls for secure connectivity to VN-DCH-01 (for replication) and to the global corporate network via P2P link.	2 (Active/Standby)
	Edge Routers	Routers for global P2P link termination and SD-WAN connectivity (if used for DR traffic).	2 (Active/Standby)
	SD-WAN Appliances	Local SD-WAN edge devices for secure and optimized data replication traffic from VN-DCH-01.	2 (Active/Standby)
	Global P2P Link Termination	Devices for terminating the dedicated private P2P circuit from the global corporate network (passive termination).	2 (Primary/Backup)
Core	Core Switches	Layer 3 switches supporting replicated server infrastructure and SAN connectivity for DR purposes.	2 (Redundant Pair)
	Server Infrastructure	Replicated rack servers, blade servers, or HCI for DR environment, mirroring critical systems from VN-DCH-01.	As required (matching primary)
	Storage Area Networks (SANs)	Replicated high-performance, redundant storage systems for DR data.	2 (Redundant Arrays)

Perimeter	Component Type	Description	Quantity
Distribution/Access	Top-of-Rack (ToR) Switches	Layer 2/3 switches providing connectivity to replicated servers and network devices within each rack in the DR environment.	As required per rack

Traffic Flow and Failover Process

Normal Operations Traffic Flow

During normal operations, all primary production traffic is directed to VN-DCH-01, which serves as the active data center. The SD-WAN solution plays a crucial role in optimizing and securing these traffic flows.

- **Internal Vietnam Traffic (User to DC):** Users at VN-OFFICE-01 and the smaller branch sites (Sites 3, 4, 5) access applications and services hosted in VN-DCH-01. Their traffic is routed over the SD-WAN overlay network, which provides secure and optimized paths to VN-DCH-01. The SD-WAN intelligently selects the best path based on application policies, network conditions (latency, jitter, packet loss), and available bandwidth. This ensures a consistent and high-performance user experience.
- **Global Network Traffic:** The Global P2P link terminating at VN-DCH-01 is the primary path for all traffic between the Vietnam operations and the global corporate network. This includes access to global applications, shared services, and inter-regional data exchange. The SD-WAN can also be configured to prioritize critical global application traffic over this link.
- **Internet Traffic:** Internet access for VN-DCH-01 is provided by its dedicated internet link. For VN-OFFICE-01, local internet breakout is maintained through its dedicated internet link, reducing latency for general internet browsing and cloud-based services not hosted in VN-DCH-01. Branch sites' internet traffic can either be backhauled to VN-DCH-01 or utilize local internet breakout if available and configured within the SD-WAN policy.

- **Data Replication Traffic (DC to DR):** Data replication from VN-DCH-01 to VN-OFFICE-02 is a continuous, asynchronous process. This traffic flows over a dedicated secure channel, potentially leveraging the SD-WAN for optimized and encrypted transport. Asynchronous replication ensures that the performance of the primary VN-DCH-01 is not impacted by the replication process, while still maintaining a reasonably low Recovery Point Objective (RPO) for the DR site. This includes replication of databases, virtual machine images, and file systems.

Failover Process (Primary DC Failure)

In the event of a catastrophic failure at VN-DCH-01 (e.g., power outage, natural disaster, major equipment failure), the failover process to VN-OFFICE-02 (Passive DR Site) would be initiated. The goal is to restore critical business operations with minimal downtime and data loss.

1. **Detection and Notification:** Network monitoring systems and application performance monitoring tools detect the outage at VN-DCH-01. Automated alerts are triggered, notifying the IT operations team.

2. **DR Plan Activation:** The IT operations team, following predefined disaster recovery runbooks, formally declares a disaster and activates the DR plan. This involves a decision to failover to VN-OFFICE-02.

3. **DNS/GSLB Update:** For external and internal services, DNS records (or Global Server Load Balancing - GSLB if implemented) are updated to redirect user and application traffic from VN-DCH-01's IP addresses to VN-OFFICE-02's IP addresses. This is a critical step for re-routing traffic.

4. **Application and Service Activation at VN-OFFICE-02:**

    - **Virtual Machines/Servers:** Replicated virtual machines and servers at VN-OFFICE-02 are powered on and brought online. This may involve a final synchronization of data from the last available replication point to ensure data consistency.
    - **Databases:** Replicated databases are brought online and recovered to the last consistent state. Depending on the replication method, this might involve applying transaction logs.
    - **Network Services:** Network services (e.g., DHCP, DNS, Active Directory) at VN-OFFICE-02 are verified to be operational and serving requests.

5. **Network Re-routing:**
    - **SD-WAN:** The SD-WAN orchestrator detects the unavailability of VN-DCH-01 and automatically re-routes internal Vietnam traffic (from VN-OFFICE-01 and branch sites) to VN-OFFICE-02. This is facilitated by the dynamic routing capabilities of SD-WAN.
    - **Global P2P:** The Global P2P link at VN-OFFICE-02 becomes the active path for global corporate network traffic. Routing protocols (e.g., BGP) will converge to direct traffic to VN-OFFICE-02.
    - **Internet:** Internet access for the now-active VN-OFFICE-02 will be through its local internet link. If VN-DCH-01's internet link was also used for other sites, those sites might temporarily use VN-OFFICE-01's internet link or be re-routed via VN-OFFICE-02.
6. **Verification and Testing:** Once services are online at VN-OFFICE-02, comprehensive testing is performed to ensure all critical applications and services are functioning correctly and accessible to users.
7. **Communication:** Regular communication updates are provided to stakeholders regarding the status of the recovery and estimated time to full restoration.

Recovery Time Objective (RTO) and Recovery Point Objective (RPO)

- **Recovery Time Objective (RTO):** For this active-passive setup, the RTO is estimated to be in the range of **2-4 hours**. This accounts for the time required to detect the failure, activate the DR plan, bring up replicated systems at VN-OFFICE-02, update DNS/GSLB, and re-route network traffic. The exact RTO will depend on the automation level of the failover process and the complexity of the applications.
- **Recovery Point Objective (RPO):** Given asynchronous data replication, the RPO is expected to be in the range of **15-30 minutes**. This means that in the event of a disaster, there might be up to 15-30 minutes of data loss, representing the data that was not yet replicated from VN-DCH-01 to VN-OFFICE-02 at the time of the failure. The RPO can be improved with more frequent replication intervals or by implementing near-synchronous replication for extremely critical data, though this would increase complexity and potentially impact primary site performance.

Pros and Cons

Advantages (Pros)

- **Cost-Effectiveness:** Compared to an active-active setup, an active-passive DR solution generally incurs lower operational costs. The passive site does not require the same level of continuous resource utilization or complex load balancing, leading to reduced hardware, software, and power consumption expenses during normal operations.
- **Simplicity in Management:** The active-passive model is less complex to design, implement, and manage. Data synchronization is typically unidirectional (from active to passive), simplifying data consistency challenges. Network routing is also straightforward, as traffic is primarily directed to a single active site.
- **Clear Roles and Responsibilities:** The distinction between the primary production site and the passive DR site provides clear roles for each location, simplifying operational procedures and troubleshooting.
- **Reduced Data Contention:** With only one site actively handling write operations, there is no risk of data contention or conflicts that can arise in active-active environments where multiple sites are simultaneously writing to the same datasets.
- **Established Technology:** Active-passive DR is a mature and widely adopted strategy, with well-understood technologies and best practices for implementation and management.
- **Scalability of Primary DC:** Consolidating primary operations into a dedicated colocation data center (VN-DCH-01) allows for better scalability and resource optimization within that single, focused environment.

Disadvantages (Cons)

- **Higher Recovery Time Objective (RTO):** The failover process in an active-passive setup involves bringing up systems at the passive site, which inherently takes more time than an active-active setup where both sites are already operational. This can lead to a longer period of downtime during a disaster.
- **Potential for Data Loss (RPO):** Asynchronous replication, while efficient, means there will always be some degree of data loss (RPO) in the event of a primary site

failure. The amount of data loss depends on the replication interval and the volume of changes.

- **Underutilized Resources:** The resources at the passive DR site are largely idle during normal operations, representing a significant investment that is not actively contributing to daily business functions. This can be seen as inefficient resource utilization.
- **Complexity of Failback:** The process of failing back from the DR site to the primary site (once the primary is restored) can be complex and requires careful planning to ensure data consistency and minimize disruption.
- **Single Point of Failure (Conceptual):** While the DR site provides resilience against a primary DC failure, the entire system still relies on the successful activation and operation of the passive site. Any issues during the failover process itself can lead to extended downtime.
- **Testing Challenges:** Regular and comprehensive testing of active-passive DR solutions is crucial but can be disruptive, as it often requires taking the passive site offline or simulating a failover, which can be resource-intensive and complex to execute without impacting production.

Design Option 2: Geo-Clustered Active-Active DC

High-Level Strategy

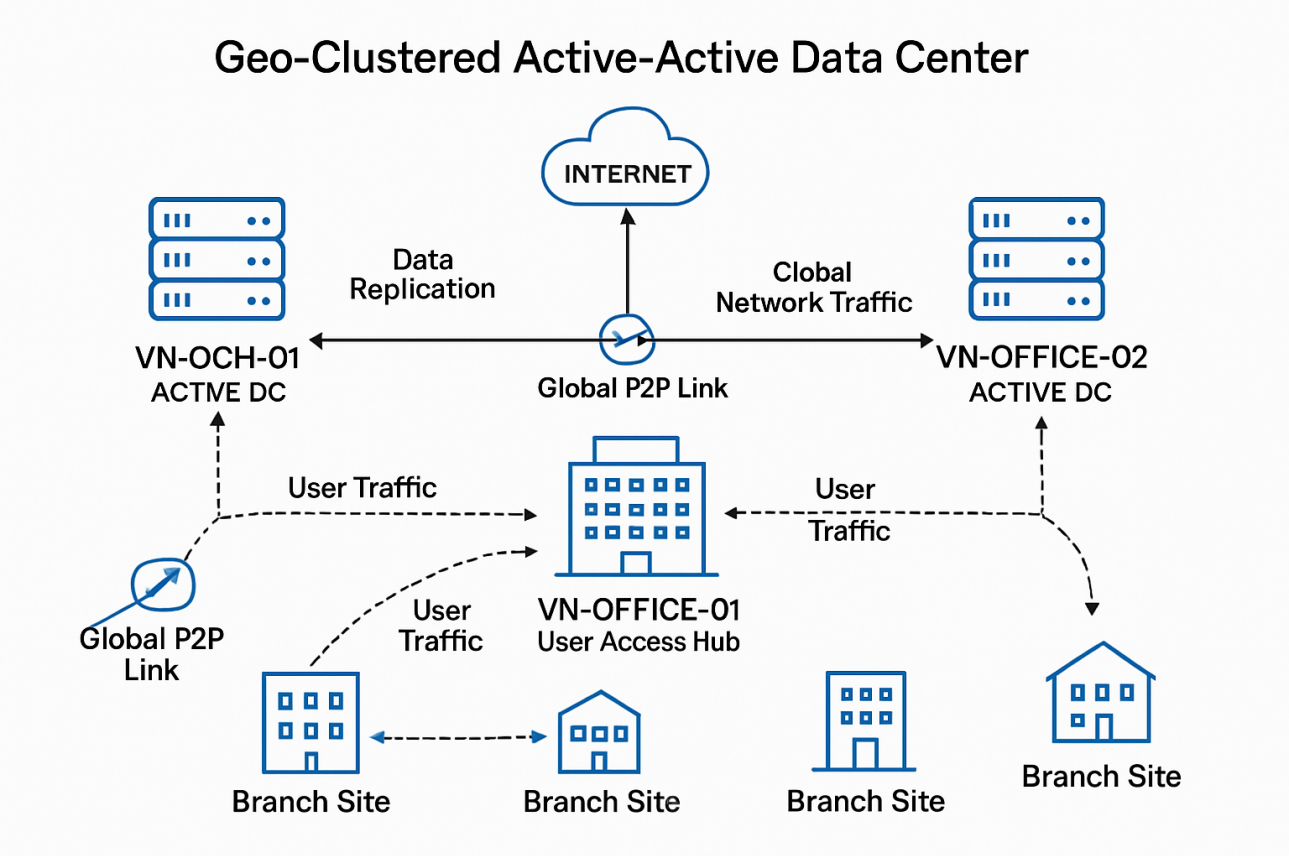
This design option proposes a Geo-Clustered Active-Active Data Center (DC) solution, leveraging both the new colocation facility (VN-DCH-01) and VN-OFFICE-02 as active production sites. In this model, both VN-DCH-01 and VN-OFFICE-02 will simultaneously host and serve applications and data, distributing the workload and providing immediate failover capabilities. VN-OFFICE-01 will continue to function as a user access hub, similar to Design Option 1, with its traffic intelligently routed to the nearest or most optimal active DC. The smaller branch sites (Sites 3, 4, 5) will also be configured to connect to the active DCs, ensuring continuous access to resources.

The core concept of this strategy is to achieve maximum business continuity and disaster avoidance by eliminating a single point of failure and ensuring that both data centers are fully operational and capable of handling production traffic at all times. This approach significantly reduces Recovery Time Objective (RTO) to near-zero, as services are already running at the alternate site. Data synchronization between VN-

DCH-01 and VN-OFFICE-02 will be synchronous or near-synchronous for critical applications to ensure data consistency across both active sites, which is crucial for an active-active setup. This design requires advanced technologies such as Global Server Load Balancing (GSLB) for intelligent traffic distribution and Data Center Interconnect (DCI) for high-speed, low-latency communication between the two active sites.

Network Architecture Diagram

Below is a conceptual network diagram illustrating the proposed Geo-Clustered Active-Active Data Center solution. The diagram highlights VN-DCH-01 and VN-OFFICE-02 operating as active data centers, with bidirectional data replication between them. VN-OFFICE-01 serves as a user access hub, and the smaller branch sites connect to the active DCs.



Key elements and their roles:

- **VN-DCH-01 (Active DC):** The new colocation data center, actively hosting applications and serving traffic. It has its own dedicated internet link and one of the Global P2P links.
- **VN-OFFICE-02 (Active DC):** The existing secondary office, upgraded to function as an active data center, also hosting applications and serving traffic. It has its

own dedicated internet link and the second Global P2P link.

- **VN-OFFICE-01 (User Access Hub):** Primarily a user access point. User traffic from this office and branch sites is intelligently routed to either VN-DCH-01 or VN-OFFICE-02 based on proximity, load, or other defined policies.
- **Branch Sites (Sites 3, 4, 5):** These smaller sites connect to the active data centers, with traffic routed for optimal performance and resilience.
- **Global P2P Links:** Two dedicated private Point-to-Point circuits connect the Vietnam network to the global corporate network, one terminating at each active DC. This provides redundant and load-balanced global connectivity.
- **Internet Links:** Dedicated internet links are present at VN-DCH-01 and VN-OFFICE-01. For VN-OFFICE-02, internet access can be provided via the DCI or a dedicated link, depending on requirements.
- **Bidirectional Data Replication:** Continuous, synchronous or near-synchronous data replication between VN-DCH-01 and VN-OFFICE-02. This is crucial for maintaining data consistency across both active sites.
- **User Traffic:** Traffic from users at VN-OFFICE-01 and branch sites is intelligently distributed to the active data centers.
- **Global Network Traffic:** Traffic flowing between the Vietnam network and the global corporate network, distributed across both active DCs.

Recommended Interconnection Technology:

For interconnecting VN-DCH-01 and VN-OFFICE-02, a **Data Center Interconnect (DCI)** solution is essential. DCI provides high-speed, low-latency, and highly resilient connectivity between the two active data centers. Technologies like Dark Fiber, Wavelength Services (DWDM/CWDM), or Carrier Ethernet can be used for DCI. This ensures seamless data replication and application synchronization.

For traffic distribution and intelligent routing, **Global Server Load Balancing (GSLB)** is recommended. GSLB distributes incoming user requests across the active data centers based on various factors such as geographic proximity, server load, and application health. This ensures optimal performance and provides automatic failover in case one of the active DCs becomes unavailable.

SD-WAN will continue to be a key technology for connecting VN-OFFICE-01 and the branch sites to the active data centers, providing application-aware routing, enhanced security, and simplified management across the distributed network.

Detailed Perimeter Component Breakdown

This section outlines the necessary network components at each key location (VN-DCH-01, VN-OFFICE-01, VN-OFFICE-02) for Design Option 2, structured by network perimeter.

VN-DCH-01 (Active DC)

Perimeter	Component Type	Description	Quantity
Edge	Next-Generation Firewalls (NGFWs)	High-throughput firewalls with advanced security features (IPS/IDS, URL filtering, application control) for internet and global P2P connectivity, and DCI.	2 (Active/Active)
	Edge Routers	High-performance routers for BGP peering with ISPs and global corporate network, handling internet and P2P link terminations. Also, for DCI routing.	2 (Active/Active)
	SD-WAN Appliances	Centralized SD-WAN orchestrator and local SD-WAN edge devices for intelligent traffic management and secure overlay network.	2 (Active/Active)
	Global P2P Link Termination	Devices (e.g., routers, media converters) for terminating the dedicated private P2P circuit from the global corporate network.	2 (Primary/Backup)
	Internet Link Termination	Devices (e.g., routers, modems) for terminating the dedicated internet link.	2 (Primary/Backup)
	Global Server Load Balancer (GSLB)	Appliance or software-based solution for intelligent traffic distribution across VN-DCH-01 and VN-OFFICE-02.	2 (Active/Active)
Core	Core Switches	High-density, low-latency Layer 3 switches forming the backbone of the DC network, supporting server	2 (Redundant Pair)

Perimeter	Component Type	Description	Quantity
		infrastructure, SAN connectivity, and DCI.	
	Server Infrastructure	Rack servers, blade servers, or hyper-converged infrastructure (HCI) actively hosting applications and services.	As required
	Storage Area Networks (SANs)	High-performance, redundant storage systems (e.g., Fibre Channel, iSCSI) for primary data storage, with synchronous replication capabilities.	2 (Redundant Arrays)
Distribution/Access	Top-of-Rack (ToR) Switches	Layer 2/3 switches providing connectivity to individual servers and network devices within each rack.	As required per rack

VN-OFFICE-01 (User Access Hub)

Perimeter	Component Type	Description	Quantity
Edge	Next-Generation Firewalls (NGFWs)	Firewalls for local internet breakout and secure connectivity to active DCs via SD-WAN.	2 (Active/Standby)
	Edge Routers	Routers for internet link termination and SD-WAN connectivity.	2 (Active/Standby)
	SD-WAN Appliances	Local SD-WAN edge devices for intelligent traffic management and secure overlay network to active DCs.	2 (Active/Standby)
	Internet Link Termination	Devices for terminating the dedicated internet link.	2 (Primary/Backup)
Core	Core Switches	Layer 3 switches providing connectivity for local servers (if any) and aggregating traffic from access layer.	2 (Redundant Pair)
Distribution/Access	Access Switches	Layer 2 switches providing connectivity for end-user devices (desktops, laptops, IP phones) and local network segments.	As required per floor/area

VN-OFFICE-02 (Active DC)

Perimeter	Component Type	Description	Quantity
Edge	Next-Generation Firewalls (NGFWs)	High-throughput firewalls with advanced security features (IPS/IDS, URL filtering, application control) for internet and global P2P connectivity, and DCI.	2 (Active/Active)
	Edge Routers	High-performance routers for BGP peering with ISPs and global corporate network, handling internet and P2P link terminations. Also, for DCI routing.	2 (Active/Active)
	SD-WAN Appliances	Centralized SD-WAN orchestrator and local SD-WAN edge devices for intelligent traffic management and secure overlay network.	2 (Active/Active)
	Global P2P Link Termination	Devices for terminating the dedicated private P2P circuit from the global corporate network.	2 (Primary/Backup)
	Internet Link Termination	Devices for terminating the dedicated internet link (if applicable, or via DCI).	2 (Primary/Backup)
	Global Server Load Balancer (GSLB)	Appliance or software-based solution for intelligent traffic distribution across VN-DCH-01 and VN-OFFICE-02.	2 (Active/Active)
Core	Core Switches	High-density, low-latency Layer 3 switches forming the backbone of the DC network, supporting server	2 (Redundant Pair)

Perimeter	Component Type	Description	Quantity
		infrastructure, SAN connectivity, and DCI.	
	Server Infrastructure	Rack servers, blade servers, or hyper-converged infrastructure (HCI) actively hosting applications and services.	As required
	Storage Area Networks (SANs)	High-performance, redundant storage systems (e.g., Fibre Channel, iSCSI) for primary data storage, with synchronous replication capabilities.	2 (Redundant Arrays)
Distribution/Access	Top-of-Rack (ToR) Switches	Layer 2/3 switches providing connectivity to individual servers and network devices within each rack.	As required per rack

Traffic Flow and Failover Process

Normal Operations Traffic Flow

In a Geo-Clustered Active-Active architecture, both VN-DCH-01 and VN-OFFICE-02 are simultaneously active, serving production traffic. This design aims to optimize performance, distribute load, and provide continuous availability. Key technologies like GSLB and DCI are critical for managing traffic and data synchronization.

- **Internal Vietnam Traffic (User to DC):** User traffic from VN-OFFICE-01 and the smaller branch sites (Sites 3, 4, 5) is intelligently routed to the nearest or most optimal active data center (VN-DCH-01 or VN-OFFICE-02) by the GSLB. The SD-WAN solution ensures secure and optimized paths, dynamically selecting the best route based on factors such as latency, link utilization, and application health. This ensures that users always connect to the most responsive data center, improving overall user experience.

- **Global Network Traffic:** Global P2P links terminate at both VN-DCH-01 and VN-OFFICE-02. Global network traffic is distributed across these two links, leveraging the active-active nature of the DCs. GSLB can also be used to direct global traffic to the most appropriate active DC based on origin or application requirements. This provides redundancy and load balancing for global connectivity.

- **Internet Traffic:** Internet access is available at both VN-DCH-01 and VN-OFFICE-01. For VN-OFFICE-02, internet access can be provided either directly via a dedicated link or through the DCI to VN-DCH-01, depending on the network design and traffic patterns. GSLB can also be used to distribute internet-bound traffic from users across the available internet egress points.

- **Data Replication Traffic (DC to DC):** Bidirectional data replication between VN-DCH-01 and VN-OFFICE-02 is continuous and, for critical applications, synchronous or near-synchronous. This replication occurs over the high-speed, low-latency Data Center Interconnect (DCI). The DCI ensures that data changes made at one active site are immediately or almost immediately reflected at the other active site, maintaining data consistency across the cluster. This is crucial for applications that require strong consistency and zero data loss.

Failover Process (One Active DC Failure)

One of the primary advantages of an active-active architecture is its inherent ability to handle failures with minimal disruption. In the event of a failure at one of the active data centers (e.g., VN-DCH-01), the failover process is largely automated and transparent to users.

1. **Detection:** Monitoring systems (network, application, and infrastructure) immediately detect the unavailability of VN-DCH-01. This includes loss of heartbeats, application health checks, and network connectivity failures.

2. **GSLB Redirection:** The GSLB, continuously monitoring the health of both active data centers, detects the failure of VN-DCH-01. It automatically updates its routing tables and DNS responses to direct all incoming user and application traffic exclusively to the remaining active data center, VN-OFFICE-02. This redirection is typically very fast, often within seconds or a few minutes, depending on DNS TTL settings and GSLB configuration.

3. **Application and Service Continuity at VN-OFFICE-02:** Since VN-OFFICE-02 is already active and running all critical applications and services, there is no need

to

power on or recover systems. The applications continue to run seamlessly, and users experience minimal to no interruption.

1. **Network Re-routing:**
    - **SD-WAN:** The SD-WAN orchestrator detects the failure of VN-DCH-01 and automatically re-routes all internal Vietnam traffic (from VN-OFFICE-01 and branch sites) to VN-OFFICE-02. This is a dynamic process, leveraging the intelligence of the SD-WAN to maintain connectivity.
    - **Global P2P:** All global network traffic that was previously load-balanced across both P2P links will now be directed solely to the Global P2P link at VN-OFFICE-02. Routing protocols will converge quickly to reflect this change.
    - **Internet:** Internet traffic will be handled by the remaining active internet links, primarily at VN-OFFICE-01 and potentially VN-OFFICE-02 (if it has a dedicated link).
2. **Data Consistency:** Due to synchronous or near-synchronous data replication over the DCI, data consistency is maintained. Any in-flight transactions at VN-DCH-01 at the time of failure are either committed at VN-OFFICE-02 or rolled back, ensuring no data corruption or loss.
3. **Monitoring and Recovery of Failed DC:** The IT team focuses on diagnosing the cause of the VN-DCH-01 failure and initiating recovery procedures. Once VN-DCH-01 is restored, it can be brought back into the active-active cluster, and traffic can be re-distributed.

Recovery Time Objective (RTO) and Recovery Point Objective (RPO)

- **Recovery Time Objective (RTO):** For this Geo-Clustered Active-Active setup, the RTO is **near-zero (seconds to minutes)**. Since both data centers are active and serving traffic, and GSLB automatically redirects users, there is virtually no downtime from a user perspective. The time taken is primarily for DNS propagation and GSLB health checks to update.
- **Recovery Point Objective (RPO):** With synchronous or near-synchronous data replication over the DCI, the RPO is also **near-zero (seconds to a few minutes)**. This means that in the event of a failure, there is minimal to no data loss, as data

is replicated in real-time or near real-time between the active sites. This is a significant advantage for applications requiring strict data consistency and minimal data loss tolerance.

Pros and Cons

Advantages (Pros)

- **Near-Zero Recovery Time Objective (RTO):** This is the most significant advantage. Since both data centers are active and serving traffic, failover is almost instantaneous, resulting in minimal to no downtime for users and applications. This is critical for business-critical applications that cannot tolerate any interruption.
- **Near-Zero Recovery Point Objective (RPO):** With synchronous or near-synchronous data replication, data loss is virtually eliminated. This ensures the highest level of data integrity and consistency, which is crucial for financial transactions, critical databases, and other sensitive data.
- **Enhanced Performance and Load Distribution:** By distributing workloads across two active data centers, the overall performance of applications can be improved. Traffic can be routed to the closest or least-loaded data center, optimizing user experience and resource utilization.
- **Improved Resource Utilization:** Unlike active-passive setups where the DR site's resources are largely idle, in an active-active model, both data centers are actively contributing to production, maximizing the return on investment for infrastructure.
- **Disaster Avoidance:** The active-active setup provides true disaster avoidance rather than just disaster recovery. If one site experiences an issue, traffic is seamlessly redirected to the other active site before a full outage occurs.
- **Simplified Testing:** Testing an active-active setup can be less disruptive, as one site can be taken offline for maintenance or testing while the other continues to serve traffic.

Disadvantages (Cons)

- **Higher Cost:** Implementing and maintaining an active-active solution is significantly more expensive than an active-passive one. This includes higher costs for infrastructure (hardware, software), network connectivity (DCI, GSLB),

and operational expenses (power, cooling, staffing) due to the need to duplicate resources and maintain complex synchronization.

- **Increased Complexity:** The design, implementation, and management of an active-active architecture are considerably more complex. This involves managing bidirectional data replication, ensuring data consistency across multiple active sites, and configuring advanced traffic management solutions like GSLB and DCI. Troubleshooting can also be more challenging.
- **Data Consistency Challenges:** While synchronous replication aims for zero data loss, ensuring perfect data consistency across geographically dispersed active sites can be a significant challenge, especially for applications with high write volumes. Latency between sites can impact synchronous replication performance.
- **Application Compatibility:** Not all applications are designed to run in an active-active configuration. Applications must be state-aware and capable of handling concurrent writes from multiple locations. Legacy applications may require significant re-architecture or may not be suitable for this model.
- **Network Latency Sensitivity:** The performance of synchronous data replication and distributed applications is highly sensitive to network latency between the active data centers. Long distances or poor network infrastructure can severely impact performance and data consistency.
- **Increased Operational Overhead:** Managing an active-active environment requires a highly skilled team and robust monitoring tools to ensure continuous operation, troubleshoot issues, and manage the complex interplay of various technologies.

Recommendation

Based on the current infrastructure, objectives, and the detailed analysis of both design options, I recommend **Design Option 1: Centralized Production DC with Active-Passive DR** for your Vietnam operations.

Justification

While Design Option 2 (Geo-Clustered Active-Active DC) offers superior RTO and RPO, providing near-zero downtime and data loss, its complexity and significantly higher

cost make it less suitable for the initial phase of enhancing business continuity for a company with an *expanding presence* in Vietnam. The prompt emphasizes enhancing business continuity, minimizing risk, and creating a scalable architecture for *future growth*, but also mentions being *open to renting rack space in a commercial colocation data center*. This suggests a need for a robust yet pragmatic solution that balances resilience with practical considerations.

Here's a detailed justification for the recommendation:

1. **Cost-Effectiveness and Resource Utilization:** Design Option 1 is considerably more cost-effective to implement and operate. The active-passive model means that the DR site (VN-OFFICE-02) does not require the same level of continuous resource utilization as an active-active setup. This aligns better with the initial investment in a new colocation facility (VN-DCH-01) and allows for a more phased approach to infrastructure scaling. While the resources at the passive site are underutilized during normal operations, the cost savings in hardware, software licenses, power, and cooling are substantial compared to maintaining two fully active production environments.

2. **Simplicity and Manageability:** The active-passive architecture is inherently less complex to design, deploy, and manage. Data synchronization is unidirectional, simplifying data consistency challenges. This reduces the operational overhead and the need for highly specialized personnel required to manage the intricate data replication and traffic management of an active-active setup. For a company with an expanding presence, a simpler, more manageable solution allows the IT team to focus on core business growth rather than managing overly complex infrastructure.

3. **Achieving Core Objectives:** Design Option 1 effectively addresses the primary goals outlined in the request:

    - **Enhance Business Continuity:** By consolidating primary operations into a dedicated, resilient colocation DC (VN-DCH-01) and establishing a clear DR site (VN-OFFICE-02), the solution significantly improves business continuity compared to the current distributed setup. The RTO of 2-4 hours and RPO of 15-30 minutes, while not near-zero, are acceptable for many enterprise applications and represent a substantial improvement in resilience.
    - **Minimize Risk:** Centralizing the primary DC reduces the attack surface and simplifies security management. The active-passive DR provides a robust

mechanism to recover from major disasters affecting the primary site.

- **Scalable Architecture for Future Growth:** The new colocation DC (VN-DCH-01) provides a dedicated environment for scalable growth. As the company's presence in Vietnam expands, the primary DC can be scaled vertically and horizontally without impacting the DR strategy. Furthermore, the active-passive model can serve as a foundational step, allowing for a future transition to an active-active model if business criticality and budget allow.

4. **Leveraging Existing Infrastructure:** Design Option 1 effectively re-purposes VN-OFFICE-01 as a user access hub and VN-OFFICE-02 as a dedicated DR site. This optimizes the use of existing assets while introducing a new, purpose-built primary data center.

5. **Key Technologies:**

   - **SD-WAN:** This technology is crucial for both options but is particularly beneficial in Design Option 1 for optimizing user traffic from VN-OFFICE-01 and branch sites to the centralized VN-DCH-01. It provides application-aware routing, enhanced security, and simplified management across the distributed network, ensuring efficient and secure connectivity to the primary DC.
   - **Asynchronous Data Replication:** This is a key enabler for the active-passive DR strategy. It allows for continuous data synchronization to VN-OFFICE-02 without impacting the performance of the primary VN-DCH-01, providing a balance between data currency and operational efficiency.

In conclusion, while an active-active solution is the gold standard for maximum availability, the active-passive approach offers a more balanced and practical solution for your current needs, providing significant improvements in resilience and business continuity at a more manageable cost and complexity. It also lays a solid foundation for potential future upgrades to an active-active model as the company's operations in Vietnam mature and critical application requirements evolve.